



**Vilnius
universitetas**

Duomenų Mokslo ir Skaitmeninių Technologijų Institutas



Informatikos inžinerijos krypties doktorantų konferencija

Veiklos ataskaita už 2023 m. rugsėjo 26d. – 2024 m. kovo 28d.

Dalia BRESKUVIENĖ – Informatikos inžinerija T 007 doktorantė

Darbo vadovas – prof. habil. dr. Gintautas DZEMYDA

Doktorantūros pradžios ir pabaigos metai: 2021.12.01 – 2025.11.30

2021–2025m.

Vilniaus
universitetas

STUDIJŲ PLANAS IR JO VYKDYMO SUVESTINĖ

Studijų metai	Egzaminai ¹		Dalyvavimas konferencijose ²				Publikacijos ³					
			Tarptautinėse		Nacionalinėse		Su citav. rodikliu			Be citav. rodiklio		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė ⁴	Planas	Įvykdyta	Būklė ⁴
I (2021/2022)	3	3			1	1						
II (2022/2023)	1	1	1	1		1		1	Publikuota	1	1	Publikuota
III (2023/2024)			1			1	1					
IV (2024/2025)							1					
Iš viso:	4	4	2	1	1	2	2	2		1	1	

2023/2024 – I pusmetis

Vilniaus
universitetas

Dalyvavimas Konferencijose 2023/2024 (I pusmetis)		
Planas	Įvykdyta	Konferencijos tipas
DAMSS: 14th conference on data analysis methods for software systems, Druskininkai	D.Breskuvienė, G. Dzemyda. „What is a concept drift, and does it affect machine learning performance?“ DAMSS: 14th conference on data analysis methods for software systems, Druskininkai	Nacionalinė

Doktorantūros studijų pasiekimai

Vilniaus
universitetas

Aprašas

D.Breskuvienė, G.Dzemyda „Clustering-based optimization in fraud detection classifier training“
EURO2022 2022.07.03 /06 ESPOO, FINLAND

Publikacijos (tik su citavimo rodikliu)

Bibliografinis aprašas

Būklė

D. Breskuvienė, G. Dzemyda, “Categorical Feature Encoding Techniques for Improved Classifier Performance when Dealing with Imbalanced Data of Fraudulent Transactions”,
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL, vol.
18, no. 3, Art. no. 3, May 2023, doi: 10.15837/ijccc.2023.3.5433.




Publikuota

Mokslinių tyrimų ir disertacijos rengimo planas:

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

Mokslinių tyrimų ir disertacijos rengimo planas:

	Darbo pavadinimas	Atlikimo terminai	
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; 3.2. Analitinės disertacijos dalies parengimas; 3.3. Teorinės disertacijos dalies parengimas; 3.4. Eksperimentinės disertacijos dalies parengimas; 3.5. Bendrųjų išvadų formulavimas.	2024 m. spalio mėn. – 2025 m. vasario mėn. 2024 m. kovo mėn. – 2025 m. rugpjūčio mėn. 2024 m. kovo mėn. – 2025 m. rugpjūčio mėn. 2024 m. kovo mėn. – 2025 m. rugpjūčio mėn. 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:

Nesubalansuotų duomenų klasifikavimas finansiniams nusikaltimams identifikuoti.

- Tikslas:

Sukurti arba patobulinti jau egzistuojantį mašininio mokymosi algoritmą, siekiant pagerinti klasifikavimo rezultatus nesubalansuotiems duomenims.

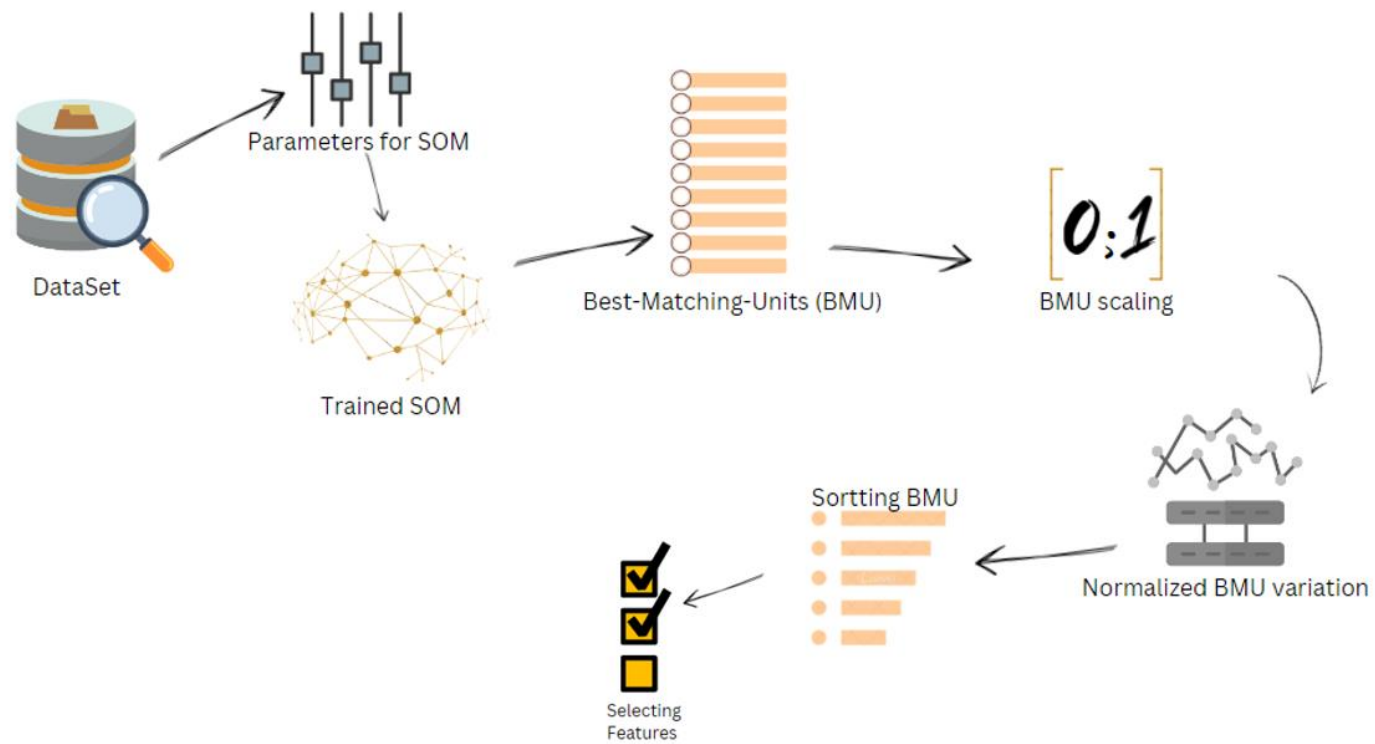
Tyrimo uždaviniai

- 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 nesubalansuotiems duomenims ir atlikti gautų duomenų analizę, rezultatų apibendrinimą, išvadų parengimą.

Atlikti darbai

- Pasiūlytas metodas sprendžiantis požymių atrinkimo uždavinį scenarijuose, kuriems būdingi labai nesubalansuoti duomenys
- Įdentifikuotos esamos nesubalansuotų duomenų požymių atrankos spragos
- Siūlomame metode integruotas SOM žemėlapis, kuris gali apdoroti požymių triukšmą ir nustatyti dėsningumus nesubalansuotuose duomenyse, kad pagerintų klasifikavimo uždavinius.
- Įvardintos problemos publikacijose susijusiose su finansinių nusikaltimų aptikimu, fokusuojantis labiausiai į sukčiavimus susijusius su kredito kortelėmis.

Grafinė metodo santrauka



Pasiūlyto metodo palyginimas su kitais klasikiniais metodais ir publikacijomis

Table 9: Comparison with other papers splitting data in a time based manner

Paper	Year	F1-Score	Recall	Precision
[50]	2019	0.82	0.73	0.93
[51]	2023	0.84	0.74	0.97
FIDSOM*	2024	0.85	0.76	0.97

*FIDSOM with XGB classifier selecting 23 features. Data split is done by selecting 70% of the first data points for training and 30% remaining data points for testing. For SOM training, 500 iterations were used with a sigma of 15 and a learning rate of 1.8.

[50] Fiore, U., De Santis, A., Perla, F., Zanetti, P., Palmieri, F.: Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences* **479**, 448–455 (2019)

[51] Fanai, H., Abbasimehr, H.: A novel combined approach based on deep autoencoder and deep classifiers for credit card fraud detection. *Expert Systems with Applications* **217**, 119562 (2023)

KITO PUSMEČIO DARBO PLANAS

- Pabaigti ruošti publikaciją „Enhancing Credit Card Fraud Detection: Highly Imbalanced Data Case“ apie atliktus uždavinius žurnalui „Journal of Big Data“
 - Paruošti pranešimo „Adapt or Fall Behind: A Deep Dive into Machine Learning Techniques for Detection of the Evolving Fraud in the Financial Realm“ medžiagą konferencijai „13th Counter Fraud, Cybercrime and Forensic Accounting Conference“ Liepos 12-13 2024. Portsmouth, Didžioji Britanija
-



**Vilnius
universitetas**

Ačiū už dėmesį!

Klausimai?