



**Vilniaus universitetas**  
**Duomenų mokslo ir skaitmeninių**  
**technologijų institutas**  
**LIETUVA**



**DOKTORANTŪROS PUSMEČIO ATASKAITA**

2021 m. spalio mėn. 1 d. – 2022 m. kovo mėn. 22 d.

**INFORMATIKOS STUDIJŲ PROGRAMOS**

**DOKTORANTĖ MARTA KARALIUTĖ**

**Disertacijos pavadinimas:** Erdvės-laiko duomenų klasifikavimas naudojant diskriminantines funkcijas

**Vadovas:** prof. dr. Kęstutis Dučinskas

**Konsultantas:** prof. habil. dr. Gintautas Dzemyda

**Doktorantūros laikotarpis:** 2017 – 2022

**Studijų metai:** IV (2021/2022)

► **Tyrimo objektas:**

Erdvės-laiko duomenų generatyvinio prižiūravimo klasifikavimo metodai

► **Tyrimo tikslas:**

Atlikti erdvės-laiko duomenų statistinį klasifikavimą naudojant diskriminavimo funkciją bei klasifikavimo klaidų tikimybes. Išvestų formulių pagrindu sukurti algoritmus ir juos pritaikyti realių duomenų analizei.

## **Tyrimo uždaviniai:**

- ▶ Bajeso ir įterptinių Bajeso diskriminantinių funkcijų pritaikymas Erdvės-laiko Gausinių duomenų generatyvinio prižiūrimo klasifikavimo uždaviniuose.
- ▶ Siūlomų klasifikavimų procedūrų tyrimas ir palyginimas, panaudojant klasių žymių skirstinius.

# Lentelė 1. Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2017/2018)	2	2					
II (2018/2019)	2	2	1	2	1	0	
III (2019 10/2020 08) (2021 09/2021 09)			1		1	1	Publikuota
<b>IV (2021/2022)</b>					<b>1 (skola iš II metų)</b>	<b>1</b>	<b>Įteikta</b>
Iš viso:	4	4	2	2	2	2	

## Lentelė 2. Ataskaitinių metų darbo planas ir jo įvykdymas

Publikacijos			
Planas	Įvykdyta	Būklė	Publikacijos tipas
<b>Pattern Recognition Letters</b>	Karaliutė, M.; Dučinskas, K. Supervised generative classification of areal spatio-temporal Gaussian data. 2022	Įteikta	<u>turi cituojamumo rodiklį</u> (impact factor) CA WoS duomenų bazėje

## Lentelė 3. Visų mokslinių tyrimų ir disertacijos rengimo etapai (I)

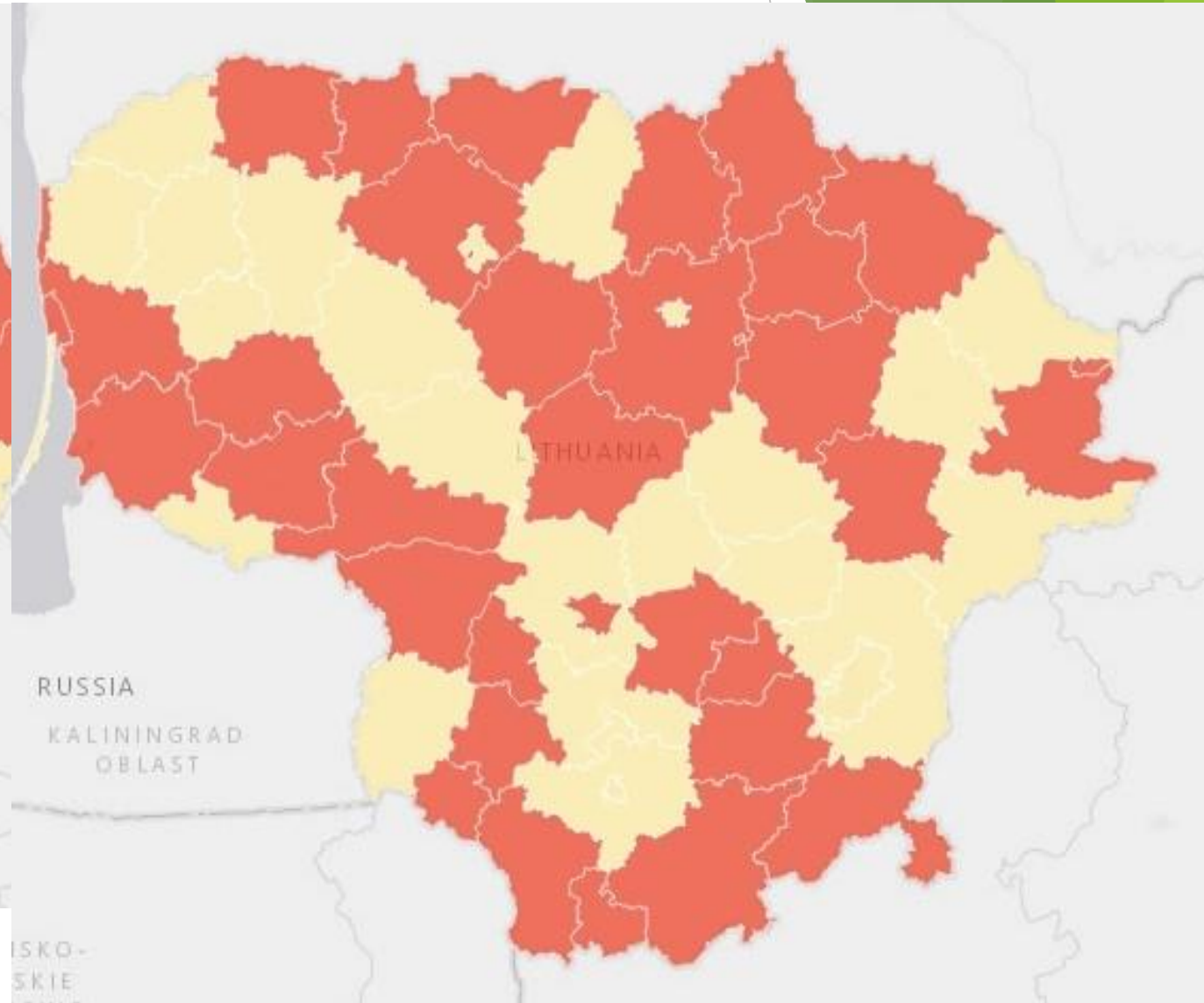
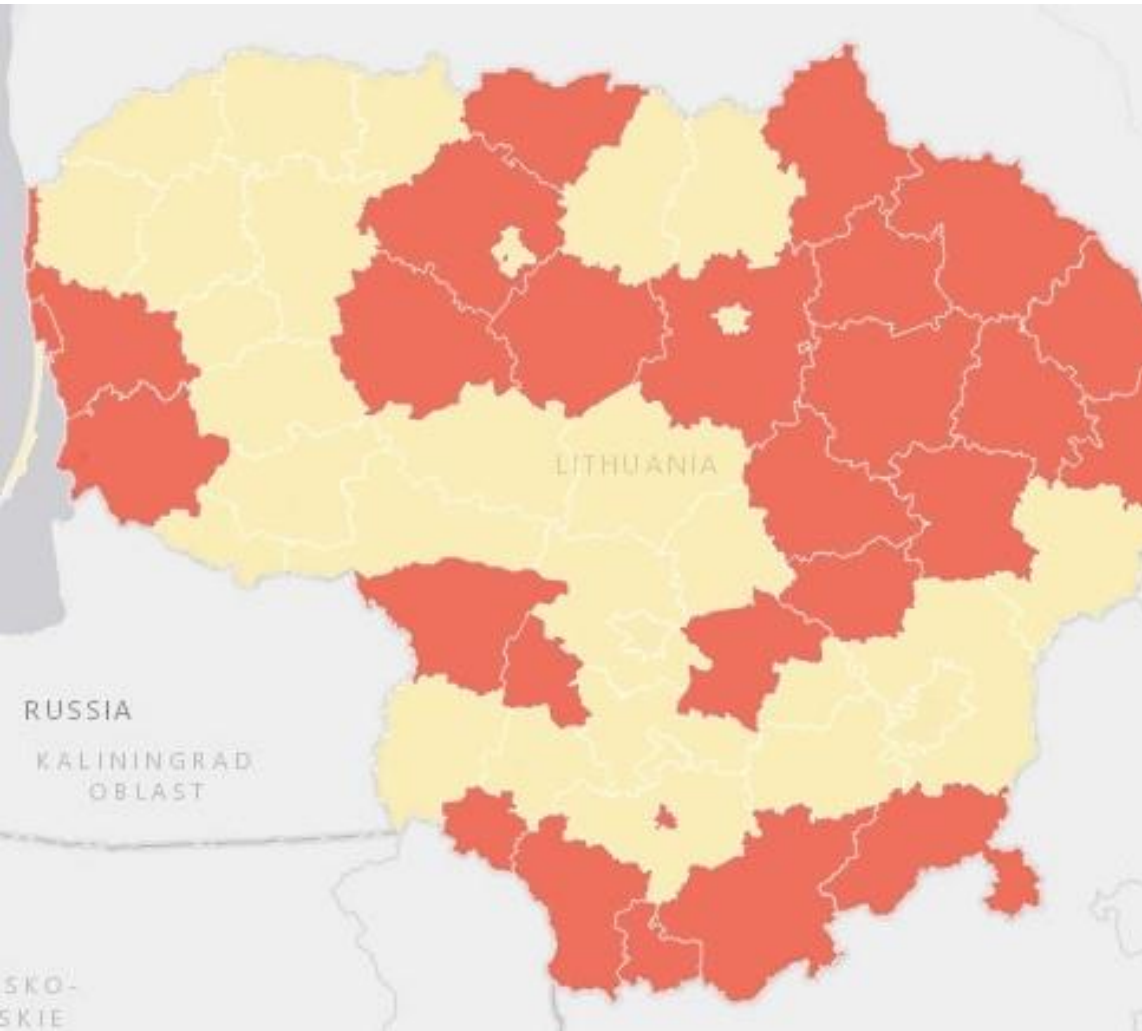
	Darbo pavadinimas	Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje).	2017 m. spalio – 2018 m. balandis	Parengta mokslinės literatūros disertacijos tema apžvalga.
2.	Mokslinio tyrimo vykdymas: 2.1. Tyrimo metodikos sudarymas:  2.2. Teorinis tyrimas:  2.3. Empirinis tyrimas:	2017 m. spalio  2018 m. gegužė - 2019 m. rugsėjis  2019 m. spalio – 2020 m. rugsėjis	Vykdoma.  Atlikta dalis teorinio tyrimo. Vykdoma.  Vykdoma. Publikacija (generuotų duomenų klasifikavimo uždavinys).

## Lentelė 3. Visų mokslinių tyrimų ir disertacijos rengimo etapai (II)

	<b>Darbo pavadinimas</b>	<b>Atlikimo terminai</b>	<b>Pastabos</b>
	2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas.	2020 m. spalio – 2020 m. gruodis	Vykdoma. Pateiktas realių duomenų klasifikavimo uždavinys.
<b>3.</b>	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:	2021 m. sausis – 2021 m. balandis	Planuojama 2022 m. balandis – 2022 m. rugpjūtis
<b>4.</b>	Daktaro disertacijos parengimas ir svarstymas padalinyje	2021 m. gegužė	Planuojama 2022 m. rugsėjis
<b>5.</b>	Daktaro disertacijos gynimas	2021 m. rugsėjis	



# Gauti rezultatai



Nagrinėjamos šios erdvės-laiko kovariacinės struktūros:

Erdvės koreliacijos matrica  $R = (I_n + \alpha H)^{-1}$ ,  
kur  $H = (h_{ij}: i, j = 1, \dots, n)$ ,

$$h_{ij} = \begin{cases} h_i = \sum_{j \in N_i} w_{ij}, & \text{kai } i = j \\ -w_{ij}, & \text{kai } s_i \in N_i \\ 0, & \text{kitais atvejais} \end{cases} .$$

Laiko kovariacijos matrica  $\Sigma$  skaičiuojama AR(1) modeliui.

Apriorinės tikimybės taške  $s_i$  laike  $t = m + 1$  yra apibrėžiamos:

1. Laiko svertinis vidurkis (Temporal Weighted Moving Average (TWMA)):

$$\pi_{1t}(s_i, m + 1) = \frac{\sum_{t=1}^m y_i^t t}{(1 + m)m/2}.$$

2. Laiko ir erdvės svertinis vidurkis (Spatial Temporal Weighted Moving Average (STWMA)):

$$\pi_{1ts}(s_i, m + 1) = \frac{\sum_{t=1}^m y_i^t t + \sum_{j:s_j \in N_i} \sum_{t=1}^m y_j^t t}{(1 + m)m/2 (1 + |N_i|)}$$

3. Laiko ir erdvės nesvertinis vidurkis (Spatial Temporal Unweighted Average (STUA)):

$$\pi_{1s}(s_i, m + 1) = \frac{\sum_{t=1}^m (y_i^t + \sum_{j:s_j \in N_i} y_j^t)}{m(1 + |N_i|)}.$$

4. Globalaus Morano I (Global Moran's I) atveju:

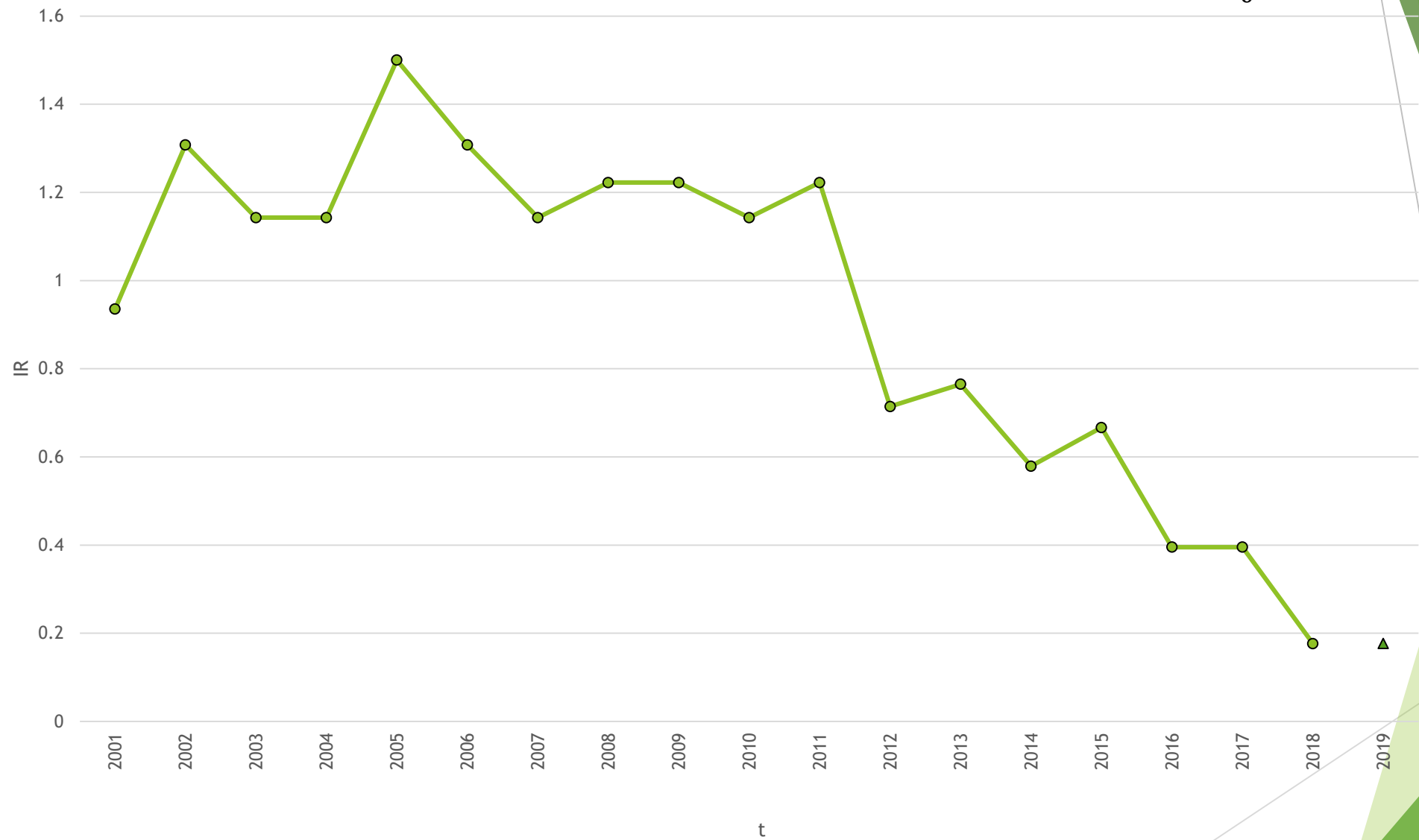
$$\pi_{1M}(s_i, m + 1) = \frac{1}{1 + \exp(-I(m)y_i^*(m))}$$

5. Geary's C atveju:

$$\pi_{1C}(s_i, m + 1) = \frac{1}{1 + \exp(-C(m)y_i^*(m))}$$

kur  $y_i^*(t) = 2y_i(t) - 1$ .

$$IR_t = \frac{n_1^t}{n_0^t}$$



	<b>TWMA</b>	<b>STWMA</b>	<b>STUA</b>	<b>Moran's I</b>	<b>Geary's C</b>
<b>ACC</b>	0.66666667	0.78333333	0.68333333	0.76666667	0.76666667
<b>AUC</b>	0.7124183	0.7810458	0.72222222	0.7254902	0.6797386

# Išvados

1. Ištirta Bajeso taisyklė Gauso erdvės-laiko klasifikavimo uždaviniui esant separabilaus kovariacijų modeliui.
2. Palyginus ACC skirtingoms apriorinėms tikimybėms nustatomas pranašumas laiko ir erdvės atžvilgiu.

# Kito pusmečio darbo planas

- ▶ Atskirų disertacijos dalių parengimas.



Ačiū už dėmesį