



Erdvinis kontekstinis klasifikavimas naudojant sąlyginius eliptinių ir eksponentinių šeimų skirstinius.

Doktorantės Eglės Zikarienės ataskaita už 2020/2021 mokslo metus

Vadovas: prof. dr. Kęstutis Dučinskas

Konsultantas: prof. dr. Julius Žilinskas

Doktorantūros pradžios ir pabaigos metai: 2013 – 2021

2021-09-30

Visų studijų planas ir jo vykdymo suvestinė:

Studijų metai	Egzaminai ¹		Dalyvavimas konferencijose ²		Publikacijos ³		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė ⁴
I (2013/2014)	2	2		1		1	Publikuota
II(2016 09/2017 03) (2019 03/2019 09)	2	2				1	Publikuota
III (2019/2020)			1	2	1		
IV (2020/2021)			1	(2)	1	(3)	1: Gautos recenzijos 2: Gautos recenzijos 3: Pateiktas

Ataskaitinių metų darbo planas ir jo vykdymas:

Egzaminai		Dalyvavimas konferencijose		Publikacijos	
Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta
-	-	Dalyvavimas tarptautinėje konferencijoje	2019 m. Dalyvauta Computer data analysis and modeling: stochasticity and data science. Baltarusija. Papildomai dalyvauta LMD konferencijoje 2021 06 16-17 „Kompiuterininkų dienos – 2021“ 2021 09 24-25	Straipsnis recenzuojamame mokslo žurnale.	2015 m. straipsnis Informatica žurnale. 2021 m. Pateikta į LMD darbai (gautos 1 recenzijos) 2021 m. Pateikta į Informatica (gautos 1 recenzijos) 2021 m. Pateikta į Spatial Statistics.

Visų mokslinių tyrimų ir disertacijos rengimo etapai:

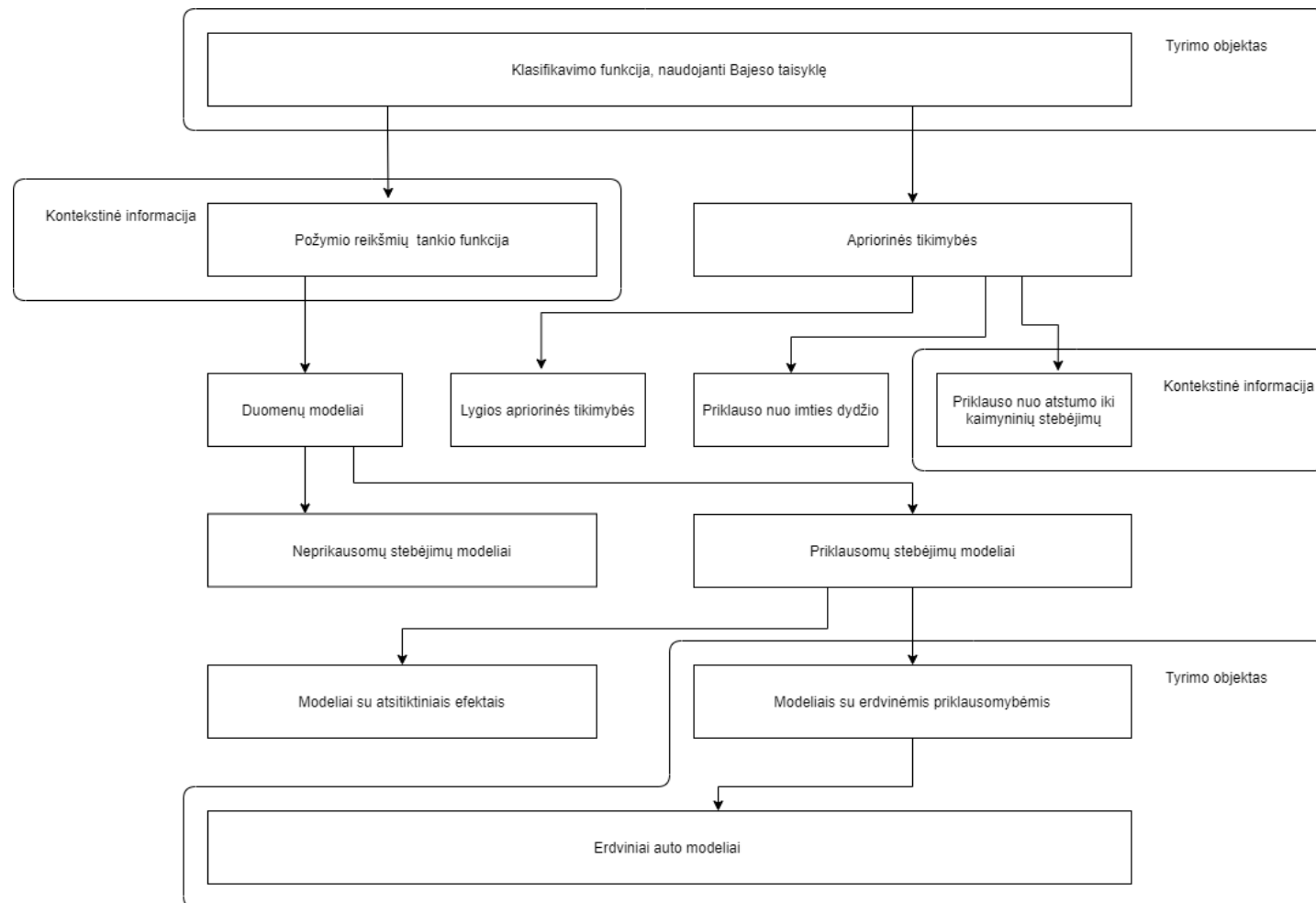
	Darbo pavadinimas	Atlikimo terminai	Pastabos
1	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje): 1. Apibendrintų tiesinių modelių taikymo erdvinių duomenų analizėje apžvalga.	2013 m. spalio – 2014 m. kovas.	Pristatyta (apimtis 18 psl.) atestacijoje 2014 m. kovo mėn.
	2. Erdvinių duomenų kontekstinio klasifikavimo metodų analizė.	2014 m. balandis - 2014 m. rugsėjis.	Ruošimas atestacijoje 2014 m. rugsėjo mėn.
2	Mokslinio tyrimo vykdymas: <u>2.1. Tyrimo metodikos sudarymas:</u> 1. Erdvinių duomenų klasifikavimo kokybės kriterijų analizė. 2. Originalių erdvinių duomenų kontekstinio klasifikavimo procedūrų pasiūlymas, jų teorinė analizė ir empirinis tyrimas.	2016 m. spalio – 2017 m. kovas.	Pristatyta (apimtis 11 psl.) atestacijoje 2017 m. kovo mėn.
	<u>2.2. Teorinis tyrimas:</u> 1. Erdvinių duomenų, aprašomų apibendrintais tiesiniais modeliais, kontekstinio klasifikavimo klaidos tikimybių analitinių išraiškų ir aproksimacijų išvedimas naudojant klasikinius ir Bajeso parametrų įvertinius. 1.1 Tolydžiuųjų požymių atveju.	2019 m. balandis – 2019 m. rugsėjis.	Pristatyta (apimtis 10 psl.) atestacijoje 2019 m. rugsėjo mėn.
	1.2 Diskrečiųjų požymių atveju.	2019 m. spalio – 2020 m. kovas	Pristatyta (apimtis 17 psl.) atestacijoje 2018 m. kovo mėn.

Visų mokslinių tyrimų ir disertacijos rengimo etapai:

	<p><u>2.3. Empirinis tyrimas:</u></p> <p>1.Siūlomų klasifikavimo procedūrų empirinis tyrimas, palyginimas ir optimizavimas naudojant tiek generuotus, tiek realius duomenis taikant programinę įrangą R ir OpenBUGS.</p>	2020m. balandis – 2020m. rugsėjis.	Pristatyta (apimtis 8 psl.) atestacijoje 2020 m. rugsėjo mėn.
	<p><u>2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas:</u></p> <p>1.Teorinio ir empirinio tyrimų rezultatų interpretacija ir apibendrinimas.</p> <p>2. Išvadų parengimas.</p>	2020 m. spalio – 2021 m. kovo	Atsiskaitoma atestacijoje 2021 m. kovo mėn. (apimtis 11 psl.)
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:	2013 m. spalio – 2014 m. rugsėjis	Atsiskaitoma atestacijoje 2014 m. rugsėjo mėn. (apimtis 15 psl.)
	1. Analitinė dalis		
	2.Teorinė dalis	2014 m. spalio – 2020 m. kovas	Atsiskaitoma atestacijoje 2019 m. kovo mėn. (apimtis 20 psl.)
	3. Eksperimentinė dalis	2020 m. balandis – 2021m. kovas	Atsiskaitoma atestacijoje 2021 m. kovo mėn. (apimtis 21 psl.)
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2021 m. gegužė	
5	Daktaro disertacijos gynimas	2021 m. rugsėjis	

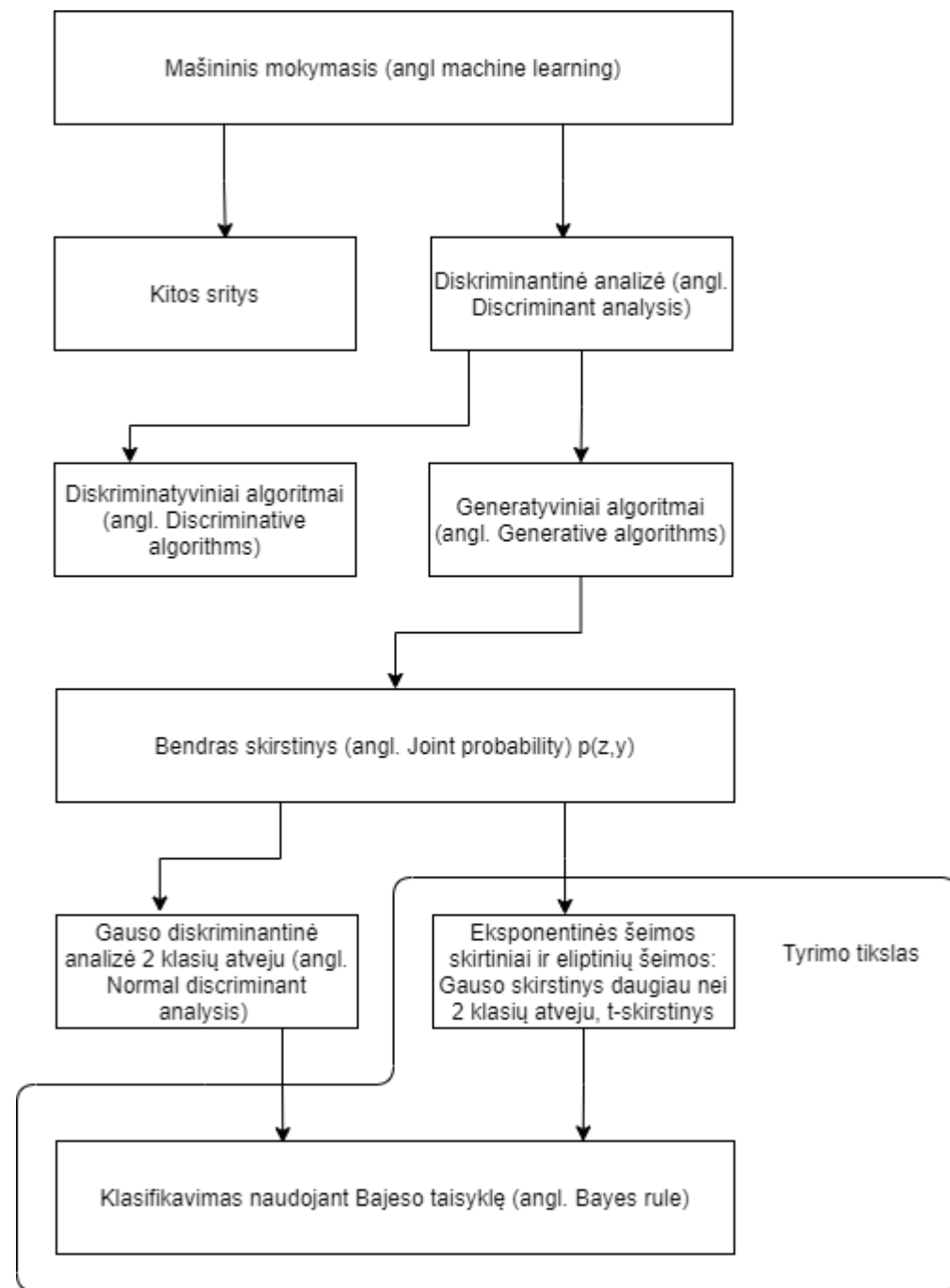
Tyrimo objektas

- Erdvinių duomenų modeliai,
- Klasifikavimo procedūra, naudojant Bajeso diskriminantinę funkciją (BDF).



Tyrimo tikslas

- Sudaryti ir ištirti BDF išraiškas eksponentinės šeimos skirstiniams ir eliptinių šeimos t skirstiniui, siekiant praplėsti klasifikavimo procedūros pritaikymo galimybes.
- (Pasiūlyti naują mašininio mokymosi metodą, skirtą erdvinių negausinių duomenų kontekstiniam Bajesiniam klasifikavimui).

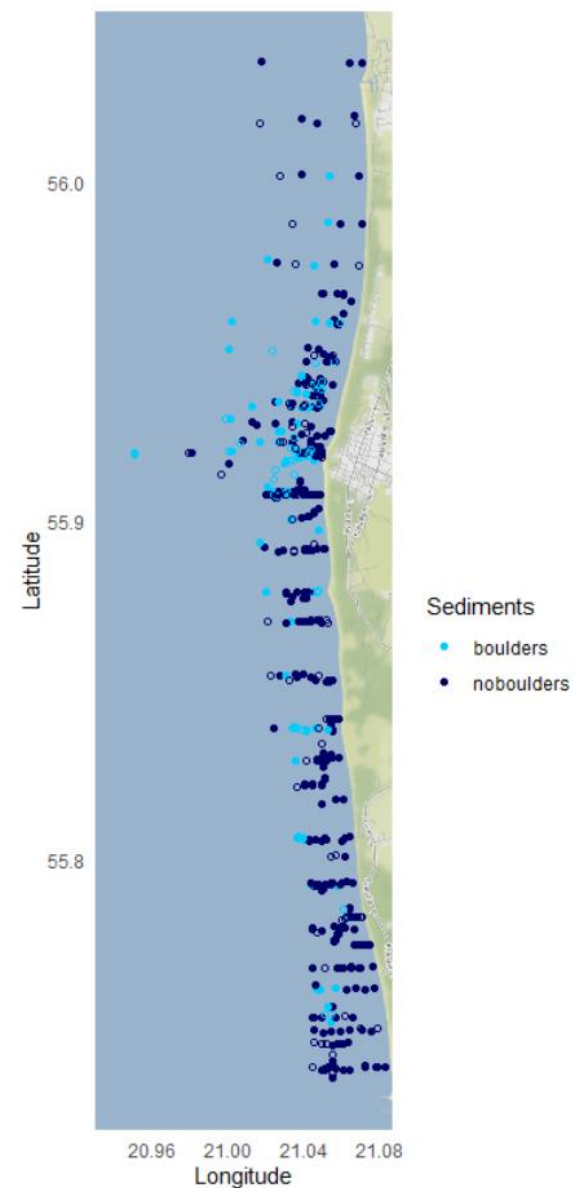


Planuojami doktorantūros rezultatai

- **Uždaviniai:**
- BDF išraiškos Bernulio, Binominis, Puasono, Gamma, Beta, t skirstiniams;
- Tikrųjų klasifikavimo klaidų tikimybių (angl. Actual error rate) formulės;
- Pasiūlytos procedūros teorinis pagrindas;
- Pasiūlytos procedūros algoritminė realizacija;

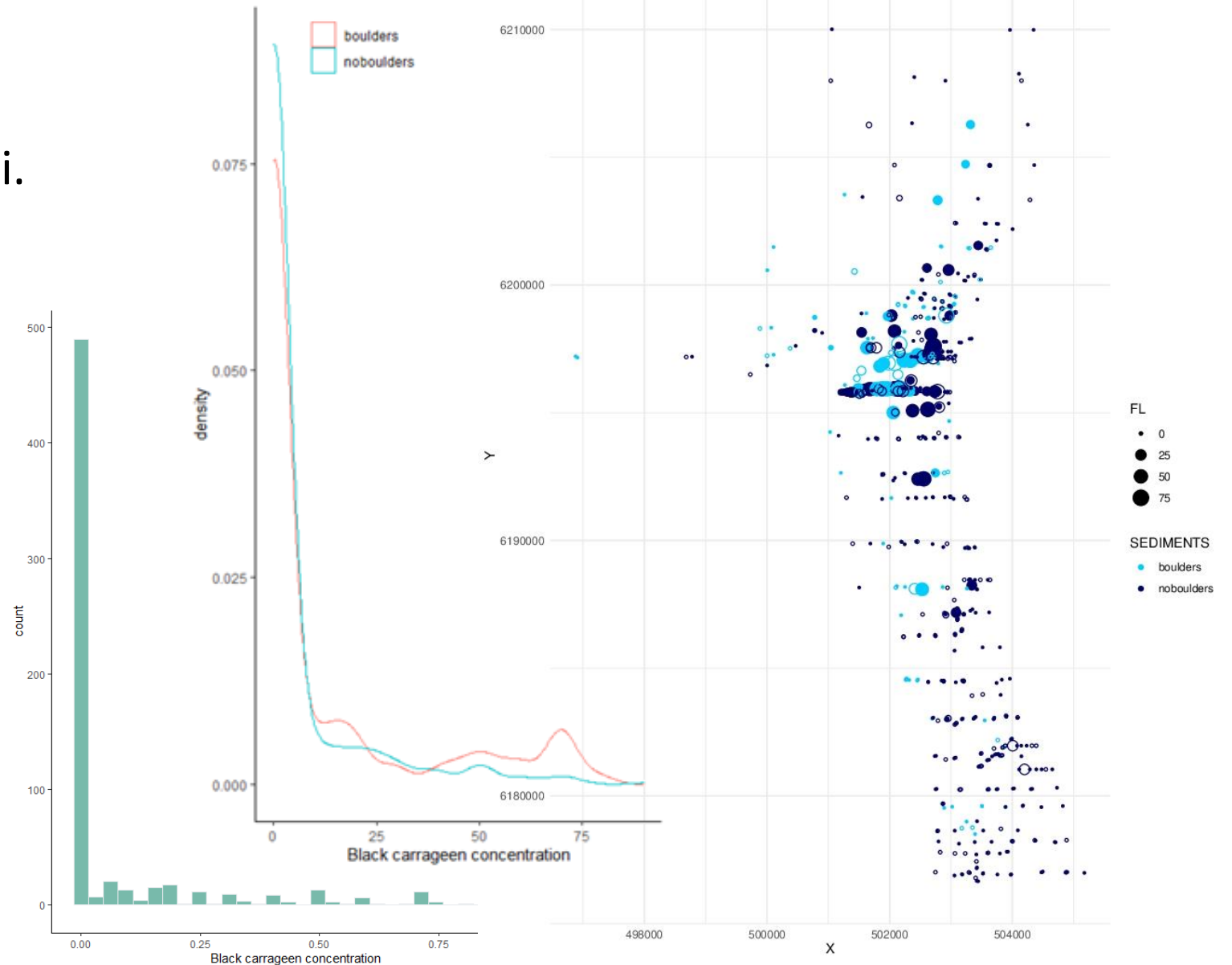
2021-03/2021-09 pusmečio rezultatai:

- Šakotojo Banguolio padengimas Baltijos jūros priekrantės zonoje.
- Klasifikavimo uždavinys (2 klasių atveju) – Baltijos jūros dugno tipo nustatymas (rieduliai, ne rieduliai) esantis po augmenija.
- Vienas stebėjimo taškas - tai 1 kv. m sritis, augmenijos padengimas, vertinamas procentais, požymio reikšmės Z.



2021-03/2021-09 pusmečio rezultatai:

- Duomenys: 641 stebėjimai.
- Mokymuisi:
 $n = 448$, $n_1 = 354$, $n_2 = 94$.
- 113 – nenulinės reikšmės;
335- nulinės reikšmės.
- Vertinimui:
 $m = 193$, $m_1 = 149$, $m_2 = 44$.
- 49– nenulinės reikšmės;
144 – nulinės reikšmės.



2021-03/2021-09 pusmečio rezultatai:

- Auto- Beta modelis: $Z_0 | T = t, Y_0 = l \sim \text{Beta}(a_{0l}, b_{0l})$

$$a_{0l} = \beta_1^l x_1 + \beta_2^l x_2 - \left(\eta \sum_{z_j \in NN_0^l, j \neq 0} \ln(1 - z_j) \right) + 1 \quad b_{0l} = \beta_3^l x_1 + \beta_4^l x_2 - \left(\eta \sum_{z_j \in NN_0^l, j \neq 0} \ln(z_j) \right) + 1$$

- Sąlyginė Bajeso diskriminantinė funkcija:

$$W(z_0; \Psi) = \ln \left(\frac{\pi_0^1 f_{01}}{\pi_0^2 f_{01}} \right) = (A_{01}^1 - A_{01}^2) \ln(z_0) + (A_{01}^1 - A_{02}^2) \ln(1 - z_0) + \gamma'_0(\Psi),$$

$$\gamma'_0(\Psi) = \ln \left(\pi_1^0 B(a_{02}, b_{02}) / \pi_1^0 B(a_{01}, b_{02}) \right), \quad \pi_l^0 - \text{apriorinės tikimybės.}$$

- Sąlyginė tiesinė diskriminantinė funkcija:

$$L(z_0, \hat{\Psi}) = \frac{(z_0 - (\mu_0^1 + \mu_0^2)/2)}{\pi_1^0 \sigma_{01}^2 + \pi_2^0 \sigma_{02}^2} (\mu_0^1 - \mu_0^2) + \gamma,$$

$$\mu_0^l = E(Z_0 | T = t, y_0 = l) = \frac{a_{0l}}{a_{0l} + b_{0l}}, \quad \gamma = \ln \left(\frac{\pi_1^0}{\pi_2^0} \right),$$

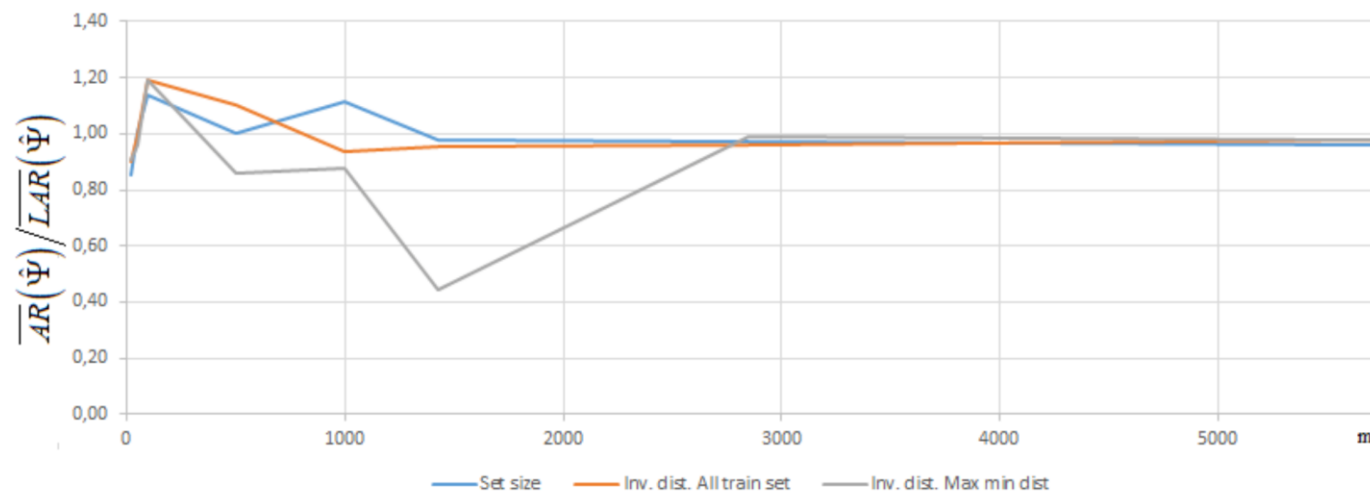
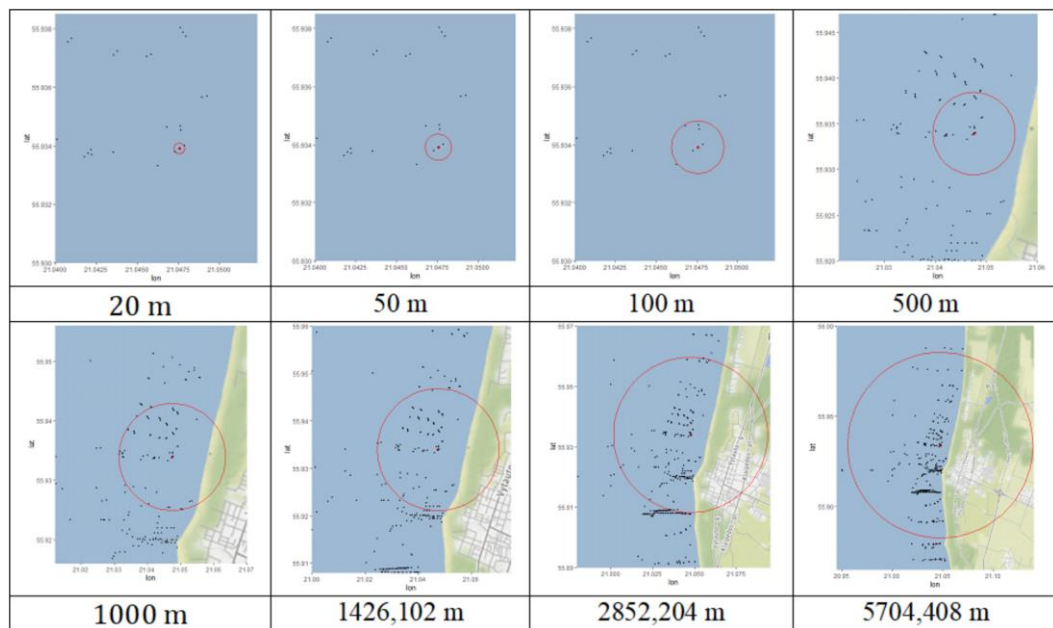
$$\sigma_{0l}^2 = \text{var}(Z_0 | T = t, y_0 = l) = \frac{a_{0l} b_{0l}}{(a_{0l} + b_{0l})^2 (a_{0l} + b_{0l} + 1)}.$$

2021-03/2021-09 pusmečio rezultatai:

- Apriorinės tikimybės:

$$\pi_l^0 = N_l / N$$

$$\pi_l^0 = \sum_{j \in NN_l^0} \frac{1}{d_{0j}} / \sum_{j \in NN_0} \frac{1}{d_{0j}}$$



Distance (m)	Prior probability					
	Sample size		Inv. dist. All train sample		Inv. dist. Max min dist	
	$\overline{AR}(\hat{\Psi})$	$\overline{LAR}(\hat{\Psi})$	$\overline{AR}(\hat{\Psi})$	$\overline{LAR}(\hat{\Psi})$	$\overline{AR}(\hat{\Psi})$	$\overline{LAR}(\hat{\Psi})$
20	8.52E-09	9.75E-09	1.09E-08	1.21E-08	1.60E-08	1.76E-08
50	0.001749	0.001715*	0.001859	0.001861	0.001272	0.001322
100	1.870E-7	1.63E-07*	1.83E-07	1.54E-07*	1.90E-07	1.59E-07*
500	0.001235	0.001218*	0.001549	0.001402*	0.001323	0.001542
1000	0.001090	0.001028*	0.001041	0.001108	0.000958	0.001095
1426.102	0.000545	0.000560	0.000558	0.000586	0.000270	0.000608
2852.204	0.001254	0.001288	0.001312	0.001369	0.001514	0.001530
5704.408	0.001314	0.001364	0.001389	0.001423	0.001588	0.001620

2021-03/2021-09 pusmečio rezultatai:

- Perteklinių nulių auto- Beta modelis (angl. Zero inflated auto-Beta model):

$$Z_0 | T = t, Y_0 = l \sim BEZI(a_{0l}, b_{0l}, c_{0l}) \quad a_{0l} = \beta_1^l x_1 + \beta_2^l x_2 - \left(\eta \sum_{z_j \in NN_0^l, j \neq 0} \ln(1 - z_j) \right) + 1$$

$$\text{logit}(c_{0l}) = \lambda^l + \sum_{z_j \in NN_0^l, j \neq 0} \rho^l I(z_j = 0) \quad b_{0l} = \beta_3^l x_1 + \beta_4^l x_2 - \left(\eta \sum_{z_j \in NN_0^l, j \neq 0} \ln(z_j) \right) + 1$$

- BEZI modelio klasifikavimo funkcija:

$$W(Z_0; \Psi) = \ln \left(\frac{\pi_0^1 g_{01}}{\pi_0^2 g_{02}} \right) = \ln \left(\frac{\pi_0^1}{\pi_0^2} \right) + \ln \left(\frac{c_{01}}{c_{02}} \right) I(z_0 = 0) + (1 - I(z_0 = 0)) \left(\ln \left(\frac{(1 - c_{01})}{(1 - c_{02})} \right) + \ln \left(\frac{f_{01}}{f_{02}} \right) \right)$$

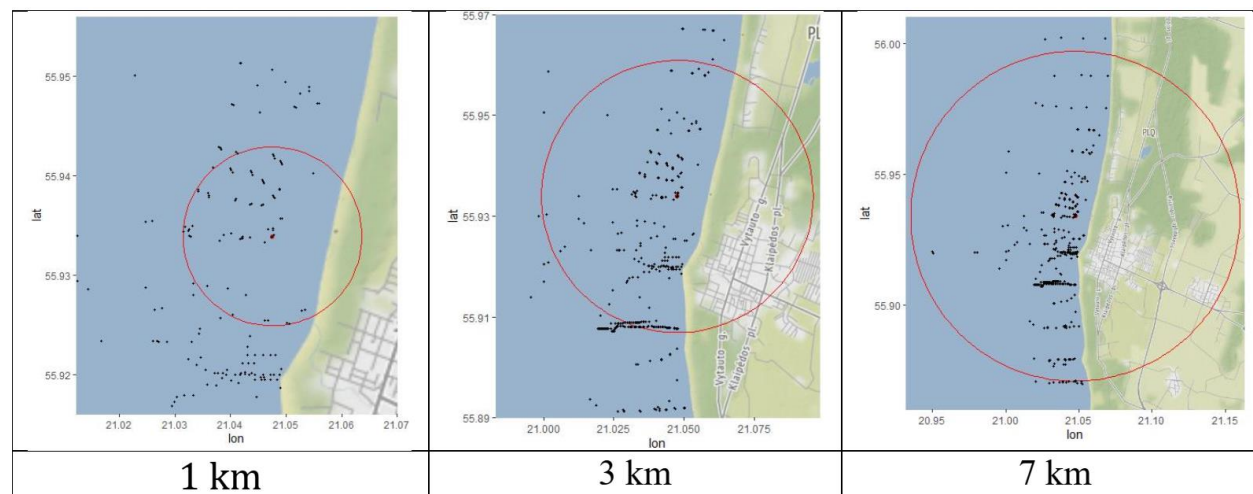
π_l^0 – apriorinės tikimybės, f_{0l} – sąlyginė beta skirstinio tankio funkcija.

2021-03/2021-09 pusmečio rezultatai:

- Apriorinės tikimybės:

$$\pi_l^0 = N_l / N$$

$$\pi_l^0 = \sum_{j \in NN_0^l} \frac{1}{d_{0j}} / \sum_{j \in NN_0} \frac{1}{d_{0j}}$$



	Prior probability		Nearest neighbour areas							
			1 km		3 km		7 km		~30 km	
	BETA	BEZI	BETA	BEZI	BETA	BEZI	BETA	BEZI	BETA	BEZI
Sample size	0,3550	0,2435	0,3783	0,2642	0,3377	0,2487	0,3837	0,2435		
Inv. dist. max min dist.	0,3550	0,2021	0,3783	0,2383	0,3377	0,2228	0,3837	0,2176		
Inv. dist. all training sample	0,3550	0,2228	0,3783	0,2487	0,3377	0,2435	0,3837	0,2383		
Inv. dist. 4 NN	0,3550	0,2073	0,3783	0,2280	0,3377	0,2124	0,3837	0,2021		

Visų straipsnių ir konferencijų sąrašas:

Tarptautinės konferencijos:

- Dučinskas K., Zikarienė E., Dreičienė L. Comparison of Performances of Plug-in Spatial Classification Rules Based on Bayesian and ML Estimators. 3rd International Conference on Pattern Recognition Applications and Methods 2014, Angers, Prancūzija, 2014 m. kovo 6-8 d.
- Zikarienė E., Dučinskas K. Application of spatial beta regression for modelling of the algae concentration index. Spatial statistics 2019, Sitges, Ispanija 2019 m. liepos 9-13 d.
- Zikarienė E., Dučinskas K. Implementation of generalized additive models for spatial beta regression. Computer data analysis and modeling: stochastics and data science. Minskas, Baltarusija 2019 m. rugsėjo 18-22 d.

Netarptautinės konferencijos:

- Zikarienė E., Dučinskas K. Erdvinių beta auto modelių taikymas klasifikavimo uždaviniuose. Lietuvos matematikų draugijos 62-oji konferencija. 2021 m. birželio 16-17 d.
- Zikarienė E., Dučinskas K. Prižiūrima aplinkos duomenų klasifikacija, pagrįsta erdviniais auto-beta modeliais. Kompiuterininkų dienos 2021, Klaipėda, Lietuva. 2021 m. rugsėjo 23-24 d.

Visų straipsnių ir konferencijų sąrašas:

Atspausdintos publikacijos:

- Dučinskas K., Dreišienė L., Zikarienė E. 2015. Multiclass classification of the scalar Gaussian random field observation with known spatial correlation function. *Statistics and probability letters*. p. 107-114. Available online 2014-12-16. (Impact factor Q2)
- Dučinskas K., Zikarienė E. 2015. Actual error rates in classification of the T-distributed random field observation based on plug in linear discriminant function. *Informatica*. (Impact factor Q2)

Konferencijų darbai:

- Zikarienė E., Dučinskas K. 2019. Implementation of generalized additive models for spatial beta regression. *Proceedings of the XII International Conference. Computer data analysis and modeling: stochastics and data science*. p. 341-343.
- Dučinskas K., Zikarienė E., Dreišienė L. 2014. Comparison of Performances of Plug-in Spatial Classification Rules Based on Bayesian and ML Estimators. *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods*. p. 161-166. 2021 m.

Pateiktos publikacijos ir gautos recenzijos:

- Dučinskas K., Zikarienė E. Application of spatial auto-beta models in statistical classification. *Lietuvos matematikos rinkinys. LMD darbai*.
- Dučinskas K., Zikarienė E. Performance evaluation for classifiers based on spatial auto-beta models. *Informatica*.

Pateikta publikacija:

- Dučinskas K., Zikarienė E. Discriminant analysis of environmental data based on zero inflated spatial auto-beta models. *Spatial statistics*.

Ačiū.

Eglė Zikarienė