



Vilnius
universitetas

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
DUOMENŲ MOKSLO IR SKAITMENINIŲ
TECHNOLOGIJŲ INSTITUTAS

JŪROS EISMO VERTINIMAS NAUDOJANT GILIUOSIUS NEURONINIUS TINKLUS

MARITIME TRAFFIC AWARENESS EVALUATION USING DEEP NEURAL NETWORKS

2020 m. spalio mėn. 1 d. – 2024 m. rugsėjo mėn. 30 d.
IV – studijų metai (II pusmetis)

Doktorantas: Robertas Jurkus

Vadovas: prof. dr. Povilas Treigys

Vilnius, 2024



TURINYS

1. Studijų plano vykdymo informacija
2. Mokslinių tyrimų pasiekimai ir disertacijos rengimo etapai
3. Mokslinio tyrimo rezultatai

— Bendra visų studijų plano vykdymo suvestinė (1)

Studijų metai	Egzaminai ¹	
	Planas	Įvykdyta
I (2020/2021)	2	2
II (2021/2022)	2	1
III (2022/2023)		1 (skola iš II metų)
IV (2023/2024)		
Iš viso:	4	4

¹ Nurodamas egzaminų skaičius.

Bendra visų studijų plano vykdymo suvestinė (2)

Studijų metai	Dalyvavimas konferencijose							Publikacijos		
	Tarptautinėse ²		Nacionalinėse ³		Su citav. rodikliu ⁴			Be citav. rodiklio ⁵		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta ⁶	Būklė ⁷	Planas	Įvykdyta ⁶	Būklė ⁷
I (2020/2021)	1	1						1	1	Publikuota
II (2021/2022)			1	1						
III (2022/2023)	1	1	1	1	1	1	Publikuota	1	1	Publikuota
IV (2023/2024)					1		Įteikta (gautos pirmos recenzijos)			
Iš viso:	2	2	2	2	2	1		2	2	

² Nurodomas tarptautinių mokslo renginių, kuriuose pristatyti disertacijos rezultatai, skaičius.
³ Nurodomas nacionalinių mokslo renginių, kuriuose pristatyti disertacijos rezultatai, skaičius.
⁴ Nurodomas publikacijų, kuriuose pateikti pagrindiniai disertacijos rezultatai ir publikuoti Clarivate Analytics Web of Science (CA WoS) referuojamuose ir turinčiuose citavimo rodiklį (Impact Factor) žurnaluose, skaičius ir būklė.
⁵ Nurodomas kitų publikacijų skaičius ir būklė (žurnaluose be citavimo rodiklio arba straipsniai konferencijų medžiagoje).
⁶ Stulpelyje „Įvykdyta“ nurodomi skaičiai tik tuomet, jei publikacija yra priimta ar publikuota.
⁷ Nurodoma publikacijos, pažymėtos stulpelyje „Įvykdyta“, būklė: „Publikuota“, „Priimta“, „Įteikta“, „Įteikta (gautos pirmos recenzijos)“.

Ataskaitinis studijų pusmetis (IV: 2023/2024 – I pusmetis)

Publikacijos 2023/2024 (I pusmetis)			
Planas*	Įvykdyta	Būklė	Publikacijos tipas
Contributions to Statistics	Robertas Jurkus, Julius Venskus, Povilas Treigys. Extended Research on Categorical Data Encoding Techniques for Recursive Multi-Step Prediction of Vessel Trajectory.	Priimta <i>(planuojama publikuoti 2024 m. pab. ar 2025 m. pradž.)</i>	Theory and Applications of Time Series Analysis and Forecasting: Selected Contributions from ITISE 2023 - A Springer book series (Electronic ISSN: 2628-8966)

* Ši publikacija yra papildoma, nenumatyta studijų plane, kuri buvo išplėtota konferencijos temos pagrindu ir bus publikuota kaip Springer žurnalo straipsnis, patvirtinimas:



an 2023-12-26 21:29
ITISE <itise@ugr.es>
ITISE-2023 chapters for Springer book have already been sent to Springer publishers.

Kam

Kopija ☐ irojas@ugr.es; ☐ 'Fernando Rojas Ruiz'; ☐ olgavc@ugr.es; ☐ Héctor Pomares; ☐ 'Luis Javier Herrera'

Dear author of ITISE-2023,

First of all thank you for your interest in actively contributing to the conference, and again congratulations on the selection of your paper for the Springer book.

Please note that we have already sent all the information to Springer.
We hope that the whole process of editing the book will be done as quickly and smoothly as possible. If there is any news, we will contact the authors.

Please, do not hesitate to contact us for any additional questions or remark.
Wishing a Merry Christmas and a Happy New Year 2024
Best regards,
Ignacio Rojas

Ataskaitinis studijų pusmetis (IV: 2023/2024 – II pusmetis)

Publikacijos 2023/2024 (II pusmetis)			
Planas*	Įvykdyta	Būklė	Publikacijos tipas
Ocean Engineering	Jurkus, R., Venskus, J., Markevičiūtė, J., Treigys, P., 2024. <i>Enhancing Maritime Safety: Estimating Collision Probabilities with Prediction Boundaries using Deep Learning Models. Ocean Engineering.</i>	Įteikta (gautos pirmos recenzijos)	Turi cituojamumo rodiklį. Web of Science Category Quartile: Q1 Journal Impact Factor: 2023: 4.6 (five years: 4.8) Publisher name: PERGAMON-ELSEVIER SCIENCE LTD

Doktorantūros studijų pasiekimai

Dalyvavimas tarptautinėse konferencijose

Aprašas

1.

R. Jurkus, P. Treigys, J. Venskus: Investigation of Recurrent Neural Network Architectures for Prediction of Vessel Trajectory. In: Information and Software Technologies, Chapter No: 16, Chapter DOI:10.1007/978-3-030-88304-1_16 (2021).

International Conference on Information and Software Technologies (ICIST) 2021
ICIST 2021, 2021 spalio 14-16, KTU, Kaunas

2.

R. Jurkus, P. Treigys, J. Venskus: Categorical Data Encoding Techniques for Recursive Multi-Step Prediction of Vessel Trajectory. International Conference on Time Series and Forecasting (ITISE 2023), held in Gran Canaria, 12-14 July, 2023 (Spain).

Publikacijos (tik su citavimo rodikliu)

Bibliografinis aprašas

Būklė

1.

Jurkus, Robertas; Venskus, Julius; Treigys, Povilas. Application of coordinate systems for vessel trajectory prediction improvement using a recurrent neural networks // Engineering applications of artificial intelligence. Oxford : Pergamon-Elsevier Science Ltd. ISSN 0952-1976. eISSN 1873-6769. 2023, vol. 123, part C, art. no. 106448, p. [1-10]. DOI: [10.1016/j.engappai.2023.106448](https://doi.org/10.1016/j.engappai.2023.106448)

Publikuota

**Vilniaus
universitetas**

Ataskaitinis studijų pusmetis (IV: 2023/2024 – II pusmetis)

Doktorantūros mobilumo veiklos pagal studijų planą

DALYVAVIMAS KONFERENCIJOSE, SEMINARUOSE, KITOSE DOKTORANTŲ MOBILUMO VEIKLOSE

Planuojama veikla	Data	Būklė	Įvykdyta*
Disertacijos tyrimų pristatymas bent 2 juose kitų universitetų organizuojamuose renginiuose	2021-2024	Vilnius TEC, matematinio modeliavimo seminaras	2024-06-18
		KU Informatikos ir statistikos (ISK) katedros posėdis	2024-06-17

Tema: „Saugumo jūroje didinimas: susidūrimo tikimybių su prognozavimo ribomis vertinimas naudojant giliojo mokymosi modelius“

* Yra gauti dalyvavimą patvirtinantys raštai.

Mokslinių tyrimų ir disertacijos rengimo etapai (1)

Darbo pavadinimas		Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje): - Disertacijos darbo objekto konkretizavimas. - Jūros eismo būsenos sąvokos formalizavimas. - Jūros eismo būsenos duomenų tyrimas. - Giliaisiais rekurentiniais tinklais paremtų architektūrų ir metodologijų nagrinėjimas. - Atliktos apžvalgos apibendrinimas ir pateikimas disertacijos analitinės dalies aprašyme.	2020 m. spalio mėn. – 2021 m. rugsėjo mėn.	Šiame etape analizuoti giliai rekurentiniai neuroniniai tinklai ir atliktas tyrimas su skirtingomis jų architektūromis: fully-connected (simple) recurrent neural network, basic (vanilla), bidirectional, stacked Long Short-Term Memory network, autoencoder, ir gated recurrent unit, siekiant prognozuoti laivo judėjimo trajektoriją jūros eisme. Tyrimo eiga ir rezultatai aprašyti mokslinėje ataskaitoje, apimtis: 15 psl. Trajektorijos prognozės gerinimui atlikti nauji eksperimentai - trajektorijos judėjimo keitimas iš poliarinės koordinatės sistemos į: atstumą ir poslinkio kampą bei Dekarto sistemą (UTM). Rezultatai sisteminami ir rengiamas straipsnis. Paraleliai eksperimentams atliekama mokslinių tyrimų apžvalga Lietuvoje ir užsienyje.
2.	Mokslinio tyrimo vykdymas: 2.1. Tyrimo metodikos sudarymas: - Pagrindžiamas temos aktualumas, nurodoma darbo problema ir suformuluojamas tikslas. - Atliekamas uždavinių pasirinkimas bei formulavimas sprendžiamai problemai spręsti. - Taikomos metodologijos ar tyrimo metodo pateikimas ir aprašymas.	2021 m. spalio mėn. – 2021 m. gruodžio mėn.	Darbo aktualumas grindžiamas augančiu jūros eismo srautu (pagal IUMI ataskaitas) ir giliojo mašininio mokymo taikymo galimybėmis, atliekant jūros eismo vertinimą. Keliamas tikslas – laivo eigos prognozė. Sudarytos ir aprašytos 4 metodologijų strategijos eksperimentinėje dalyje, kurios leidžia nustatyti laivo padėties geolokaciją, išskiriant: AIS duomenų panaudojimą su geografinėmis koordinatėmis, delta išvestinių požymių augmentacija, polinės ir Dekarto sistemų taikymą, pastaroji išreiškta universaliąja skersine projekcija (UTM). Mokslinių tyrimų apžvalgos, teorinių ir praktinių žinių taikymas aprašomas mokslinėje publikacijoje, apimtis: 14 psl.
	2.2. Teorinis tyrimas: - Mokslinės ir kitos informacijos analizė ir sisteminimas. - Tyrimo metodologijos sudarymas. - Erdvės ir laiko duomenų tyrybos metodų taikymas jūros eismo prognozavimui. - Giliojo mašininio mokymo taikymas jūros eismo būsenos analizei.	2021 m. gruodžio mėn. – 2022 m. rugsėjo mėn.	Mokslinės literatūros analizė pateikta mokslinės analizės įvadinėje dalyje. Tyrimo metodologijų taikymas validuojamas Baltijos jūros regione – skirtingais metodais tiriant pozicionavimą erdvėje ir laike. Tyrimo eiga ir rezultatai aprašyti mokslinėje ataskaitoje, apimtis: 17 psl. Papildomai atlikti tyrimai: trajektorijos įvesties ir išvesties ilgio santykio nustatymas; kardinalaus celių didinimo ir sekų maišymo įtaka prognozės tikslumui.

Mokslinių tyrimų ir disertacijos rengimo etapai (2)

Darbo pavadinimas		Atlikimo terminai	Pastabos
2.	2.3. Empirinis tyrimas: 1. Eksperimentinės dalies aprašas. 2. Egzistuojančio metodo realizavimas kuriamo metodo palyginimui. 3. Metodų modifikacijos sukūrimas. 4. Sukurtų modifikacijų eksperimentinis tyrimas ir modifikacijų palyginimas su publikuotais egzistuojančiais metodais.	2022 m. spalio mėn. – 2023 m. kovo mėn.	Empiriniame tyrime įvertinama skirtingų laivo tipų įtaką jo judėjimo trajektorijai prognozuoti, naudojant giliuosius rekurentinius neuroninius tinklus. Išbandomos kelios skirtingos kategorinių duomenų kodavimo technikos: etikečių (angl. ordinal), fiktyvus (angl. one-hot) ir įterptinis (angl. embedding). Šiame etape naudojami skirtingų laivo tipų duomenys. Papildomai atliekami tyrimai tinklo architektūros lygmenyje – bandant taikyti skirtingas praradimo (angl. loss) funkcijas: MSE, MAE, RMSE, MAPE. Šiomis temomis renkama/nagrinėjama mokslinė literatūra bei rengiami publikacijų straipsniai.
	2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas: 1. Teorinio tyrimo apibendrinimas. 2. Empirinio tyrimo apibendrinimas. 3. Rezultatų apibendrinimas, išvadų parengimas.	2023 m. kovo mėn. – 2023 m. rugsėjo mėn.	RNN tinklo modeliai pritaikomi skaičiuojant prognozių intervalus. Šiame etape bandoma įvertinti tikimybinę laivų susidūrimo riziką realiuose istoriniuose incidentuose. Iš teorinių/empirinių tyrimų bei gautų rezultatų yra rengiamas straipsnis. Kartu su kategorijų kodavimu turima papildomų 10 psl. medžiagos.
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas: 1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas. 2. Analitinės disertacijos dalies parengimas. 3. Teorinės disertacijos dalies parengimas. 4. Eksperimentinės disertacijos dalies parengimas. 5. Bendrųjų išvadų suformulavimas.	2023 m. spalio mėn. – 2024 m. kovo mėn.	Išplėtota ir pateikta galutinė ITISE publikacija Springer leidiniui, apimtis: 13 psl. Pusmečio metu atlikti papildomi tyrimai konstruojant ir lyginant prognozavimo tikimybinės sritis, taikant: prognozavimo ir pasiklivimo intervalus, elipsinius prognozavimo regionus ir konforminius prognozavimo intervalus. To pagrindu beveik paruošta sekanti publikacija, apimtis: virš 10 psl.
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2024 m. balandžio mėn.	Parengtas disertacijos juodraštis (pirminis variantas), pateikta prie mokslinė ataskaitos (apimtis apie 80 psl.)
5.	Daktaro disertacijos gynimas	2024 m. rugsėjo mėn.	

Bendruosius gebėjimus stiprinančios ir kitos mobilumo veiklos

MOKSLINĖS INFORMACIJOS IŠTEKLIAI, PAIEŠKA, IR ĮRANKIAI (0,1 ECTS)

MOKSLINIŲ REZULTATŲ PUBLIKAVIMAS PAGAL FORMALAUS VERTINIMO REIKALAVIMUS (0,1 ECTS)

LIETUVOS AKADEMINĖ ELEKTRONINĖ BIBLIOTEKA (eLABa): autoriaus sąsaja (0.15 ECTS)

MOKSLINIŲ TYRIMŲ DUOMENŲ VALDYMO DIRBTUVĖS (0.25 ECTS)

LaTeX (1,25 ECTS)

TRUSTWORTHY AI IN PRACTICE (5 val. 0,2 ECTS)

Doktorantų vasaros mokykla (2 ECTS)

Doktorantūros studijos

Studijų apimtis – ne mažiau nei 33 kreditai (ne mažiau kaip 3 dalykus), 3 kreditai už dalyvavimą bendruosius gebėjimus stiprinančiose veiklose.

Nuolatinės studijos: **30** / **30** ECTS

Studijų dalykai: **4** / **4**

Bendrieji gebėjimai: **3** / **4,05** ECTS



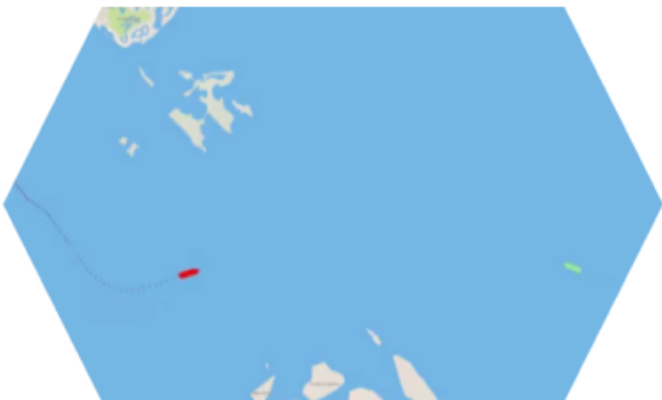
Mokslinio tyrimo rezultatai (paskutinio pusmečio)

- ❑ Įteikta ir koreguota publikacija Ocean Engineering žurnalui;
- ❑ Parengtas disertacijos juodraštis (pirminė versija);
- ❑ Išgauti ir prijungti orų meteorologiniai duomenys (tyrimai testavimo stadijoje);
- ❑ Atlikti pristatymai kitų 2-jų universitetų renginiuose.



Įvadas ir Dalykinė sritis

- Jūrų eismo situacijos suvokimas yra esminis jūrų transporto saugos veiksnys.
- Šiuolaikiniai tyrimai visų pirma tiria: autonominių laivų valdymą, susidūrimo rizikas ir anomalijas.
- Tarptautinė jūrų draudimo sąjunga (IUMI) skelbia, kad patiriami nuostoliai dėl įvykusių incidentų, kurių priežastys:



PRASTAI APMOKYTA ĮGULA

ĮRANGOS GEDIMAI

ORO SĄLYGOS

SUSIDŪRIMAI (~ 10 %)



Problematika

Prognozuojant laivo plaukimo eigos trajektoriją būtų galima įvertinti susidūrimo riziką ar pastebėti nejprastą eismą.

Tyrimo objektas

Jūros eismo duomenimis apmokyti gilieji rekurentiniai neuroninių tinklų modeliai.

Mašininio mokymo užduotis

Prižiūrimasis mokymas (angl. supervised learning) – regresijos užduotis.

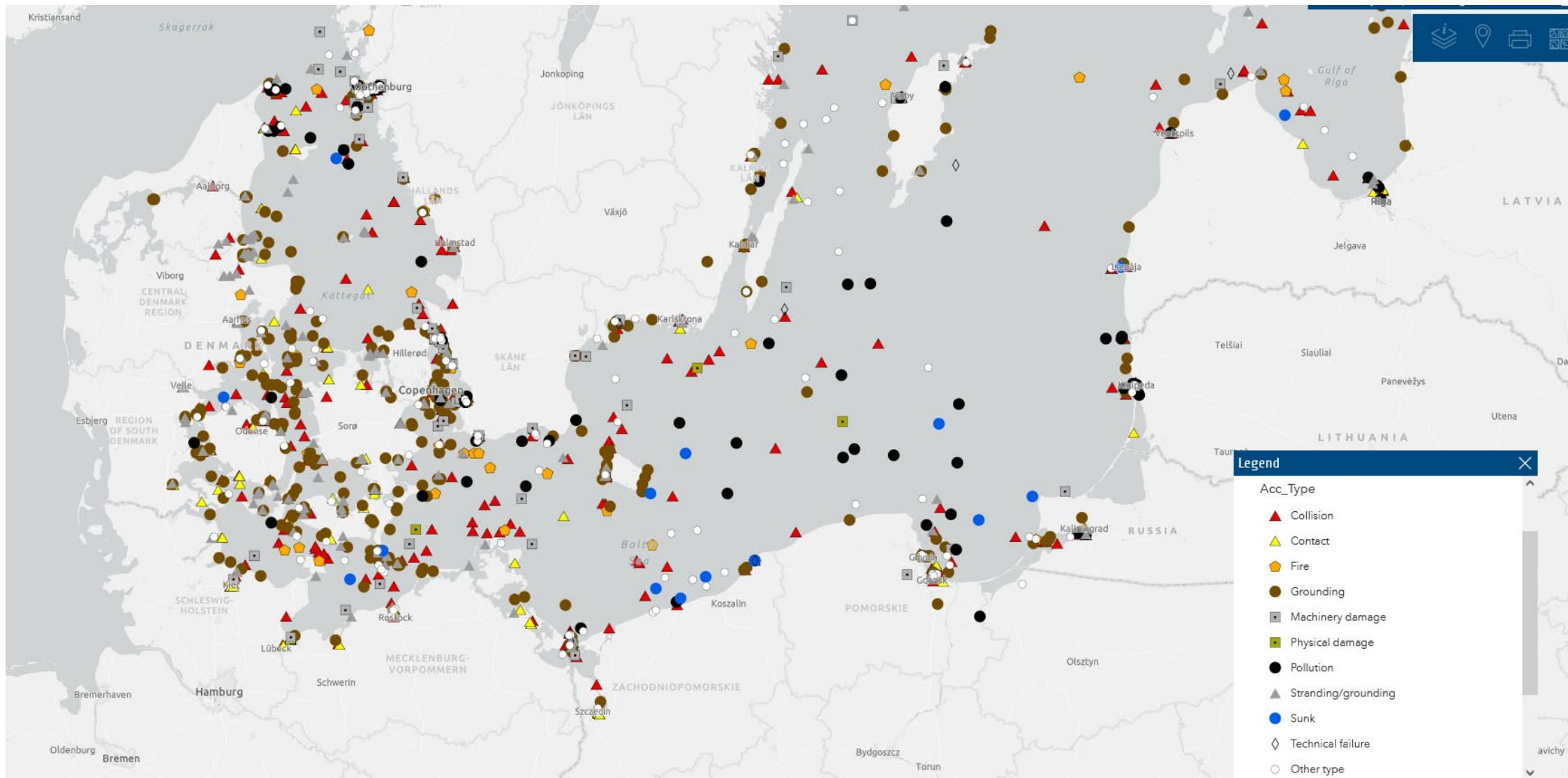
Tikslas

Ištirti giliuosius rekurentinius neuroninius tinklus, skirtus jūrų eismo stebėsenai: laivų judėjimo prognozavimas ir galimų laivų susidūrimų aptikimas bei vertinimas kaip prevencinė priemonė.

Uždaviniai:

1. Atlikti literatūros apžvalgą apie bendrą padėtį jūroje, įskaitant panašių tyrimų, kuriuos atliko kiti mokslininkai jūrų eismo, transporto ir saugos srityse, analizę.
2. Sukurti giliojo mokymosi metodais paremtus modelius, skirtus laivo trajektorijos prognozėms sudaryti.
3. Įvertinti prognozavimo tikslumo poveikį kategoriniams duomenims, atsižvelgiant į jūrinių duomenų tipus ir nustatyti geriausius metodus kategoriniams duomenims apdoroti.
4. Nustatyti tinkamiausius laivų susidūrimų statistinio ir tikimybinio vertinimo metodus.
5. Empiriškai verifikuoti metodus su tikru istoriniu incidentu jūroje.

HELCOM laivybos avarijų žemėlapis





Temos aktualumas

- Mokslinių straipsnių/publikacijų skaičius susijusia tematika



ScienceDirect

Trajectory prediction

Time series

Vessel



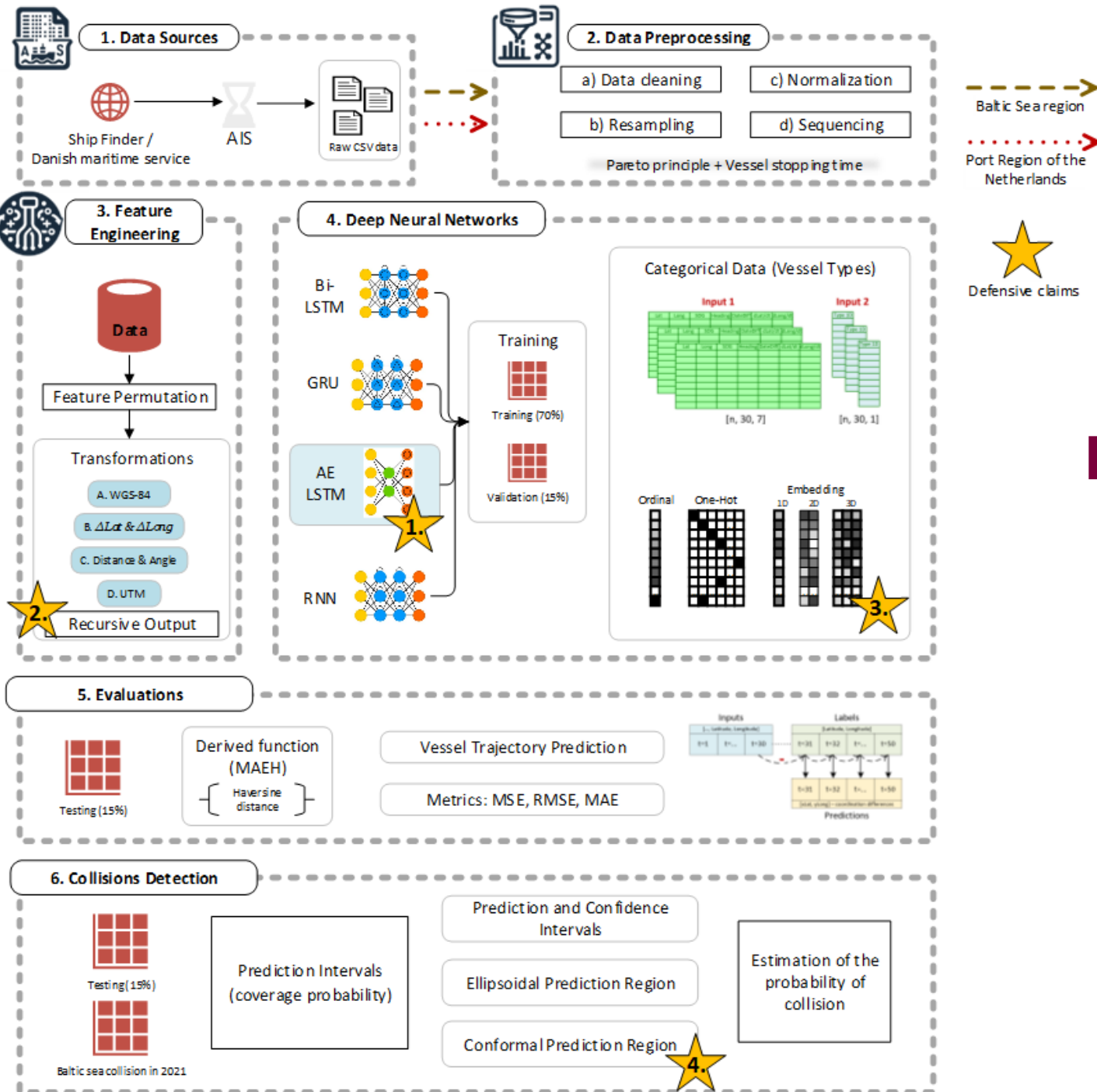
Dr. Andrius Daranda (2021 m.): Mašininis mokymusi grindžiamas laivybos eismo dalyvių elgsenos prognozavimas bei nestandartinių laivybos srauto situacijų atradimas.

Dr. Julius Venskus (2021 m.): Dalinai prižiūrimų ir neprižiūrimų mašininio mokymosi metodų tyrimas jūrų eismo anomalijoms aptikti.

**Vilniaus
universitetas**

Ginami teiginiai

1. LSTM architektūroje ląstelių skaičiaus pasirinkimas nuo tam tikros ribos nebeturi įtakos prognozavimo tikslumui arba turi labai minimalų poveikį, tačiau ląstelių skaičiaus didinimas gerokai sulėtina modelio mokymą.
2. Prognozuojant kelių žingsnių daugiamatę laivo trajektoriją, rekursinis prognozės laiko eilutės perskaiciavimas pagal ankstesnę įvesties koordinatę leidžia tiksliau nustatyti kitą prognozės tašką, ypač pradinuose prognozavimo etapuose.
3. Laivų tipų įtraukimas į LSTM rekurentinio kelių žingsnių daugiamatį duomenų rinkinį pagerina laivų trajektorijų prognozavimo tikslumą, kai taikomas įterptojo kodavimo (angl. embedding) neuroninio tinklo sluoksnis.
4. Konforminis (angl. conformal) prognozavimo regionas leidžia aptikti laivų susidūrimo ribas su didžiausia statistine tikimybe, taikant 95% pasiklivimo lygį, kai duomenys yra daugiamačiai ir juose nėra normaliojo skirstinio.

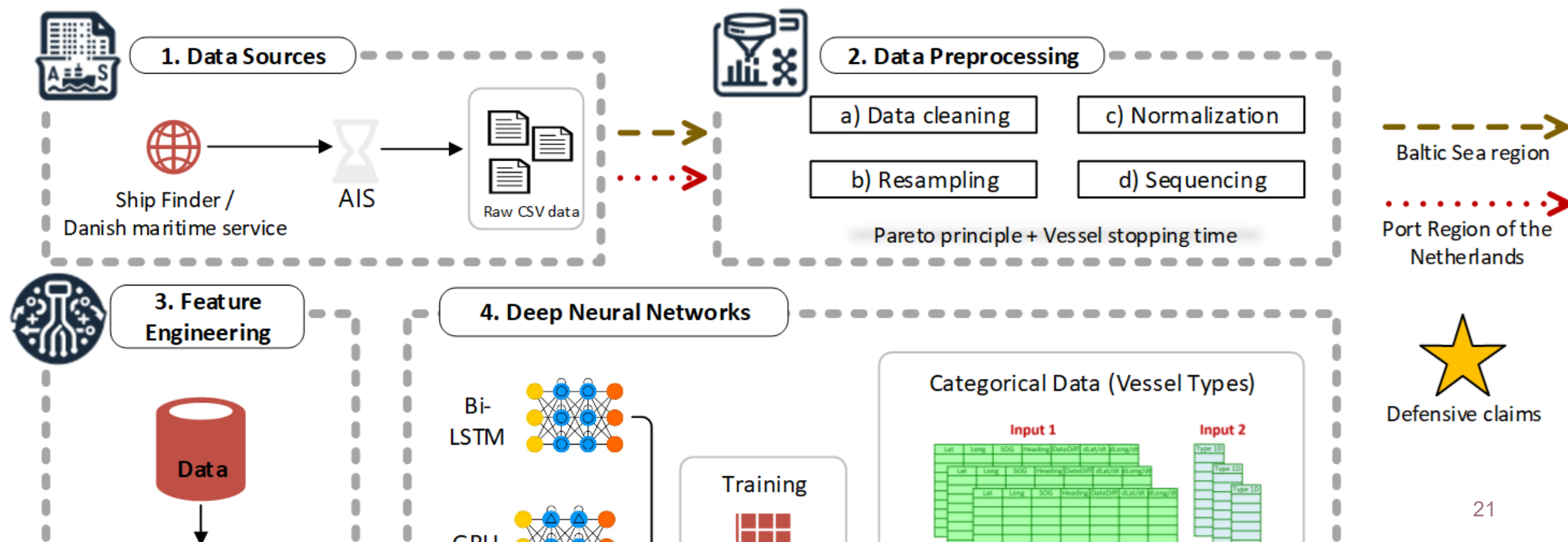


Preliminari disertacijos tema

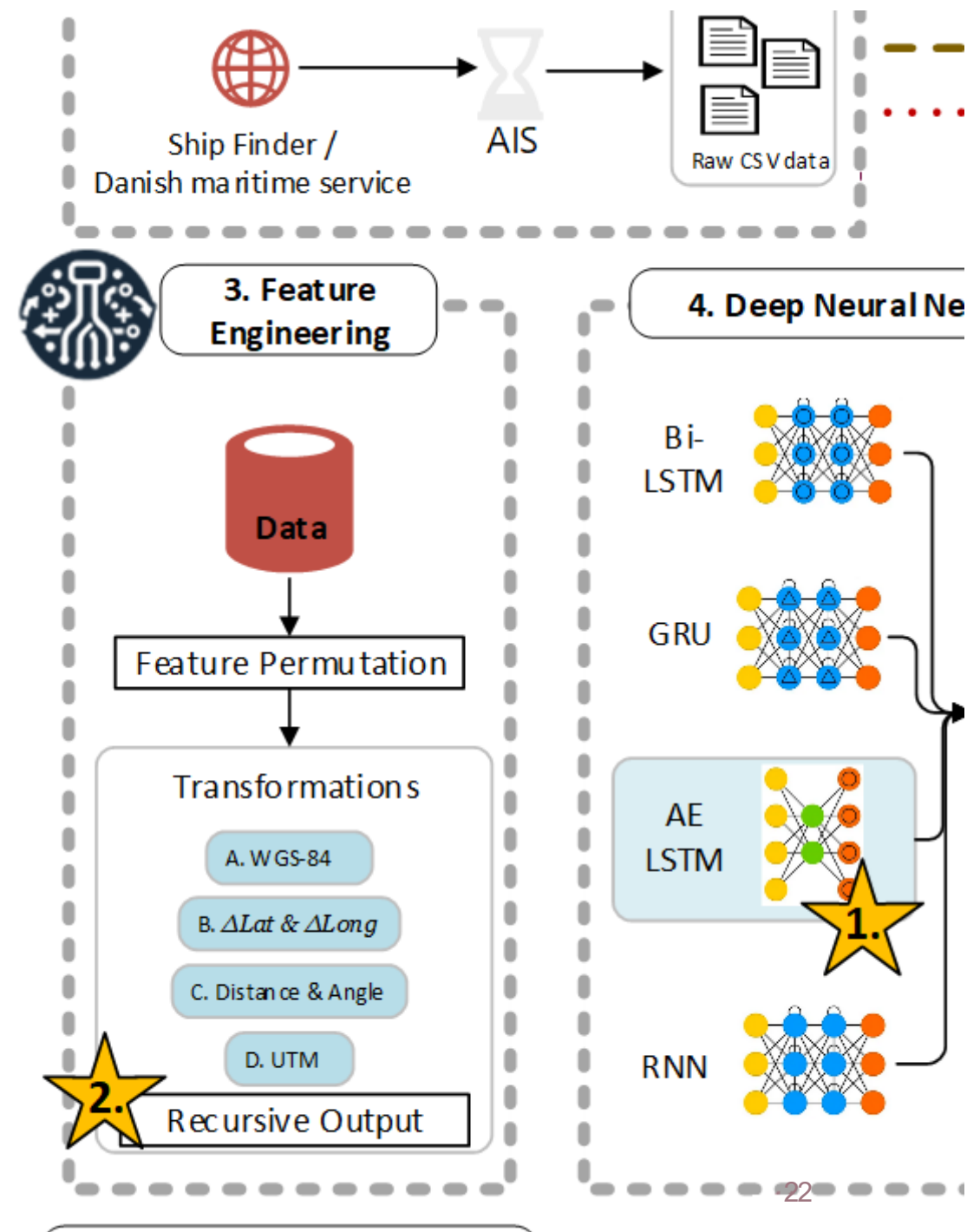
- LT: Jūros eismo informuotumo vertinimas naudojant giliuosius neuroninius tinklus
- EN: Maritime traffic awareness evaluation using deep neural network

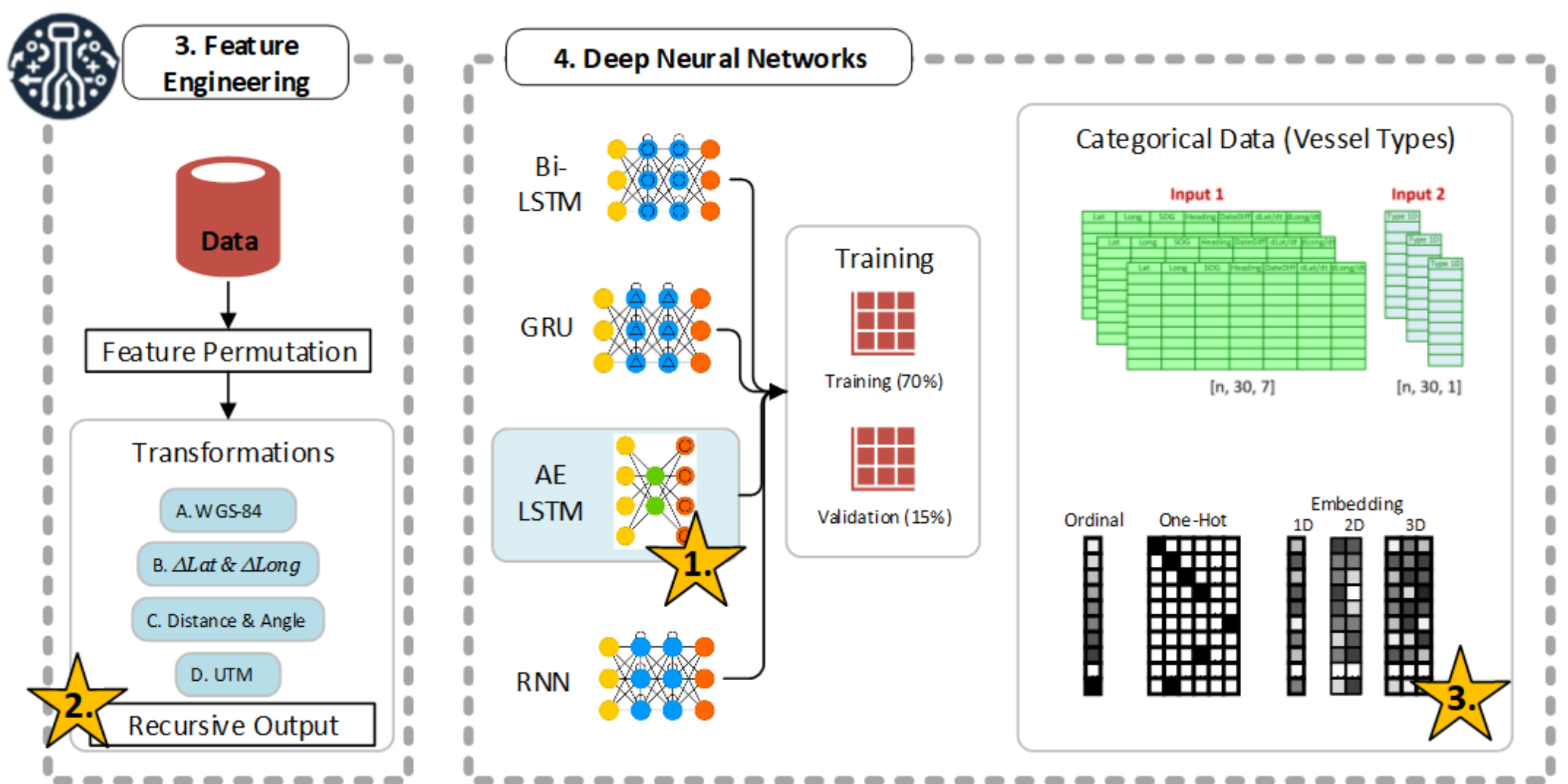
Tyrimų eigos diagrama

Gilusis mokymasis yra mašininio mokymosi, kuris pats savaime yra dirbtinio intelekto (DI) poaibis, dalis. Daugiausia dėmesio skiriama daugiasluoksnių neuroninių tinklų naudojimui, kad būtų galima modeliuoti ir suprasti sudėtingus duomenų dėsningumus.



1. **Robertas Jurkus**, Povilas Treigys, and Julius Venskus, “*Investigation of Recurrent Neural Network Architectures for Prediction of Vessel Trajectory*”. In: Information and Software Technologies, Springer, Chapter No: 16, vol 1486. Chapter DOI:10.1007/978-3-030-88304-1_16, KTU, Lithuania, 2021.
2. **Robertas Jurkus**, Julius Venskus, and Povilas Treigys, “*Application of coordinate systems for vessel trajectory prediction improvement using a recurrent neural networks*”. In Engineering applications of artificial intelligence, Oxford : Pergamon-Elsevier Science Ltd. ISSN 0952-1976. eISSN 1873-6769. 2023, vol. 123, part C, art. no. 106448, p. [1-10]. DOI: 10.1016/j.engappai.2023.106448.





- Robertas Jurkus, Julius Venskus, and Povilas Treigys, “Extended Research on Categorical Data Encoding Techniques for Recursive Multi-Step Prediction of Vessel Trajectory”. In: Contributions to Statistics, A Springer book series, ISSN: 2628-8966, ITISE 2023/2024.**

5. Evaluations



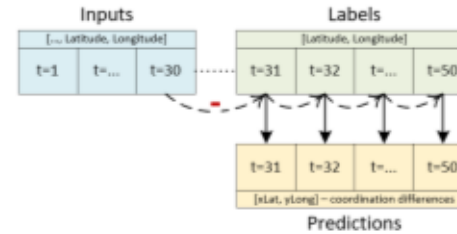
Testing (15%)

Derived function
(MAEH)

{
Haversine
distance
}

Vessel Trajectory Prediction

Metrics: MSE, RMSE, MAE



6. Collisions Detection



Testing (15%)



Baltic sea collision in 2021

Prediction Intervals
(coverage probability)

Prediction and Confidence
Intervals

Ellipsoidal Prediction Region

Conformal Prediction Region

Estimation of the
probability of
collision

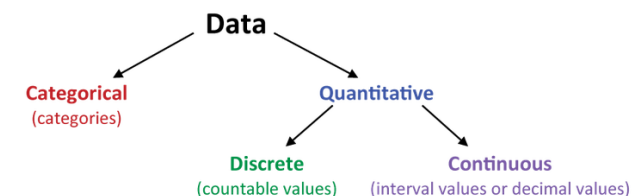
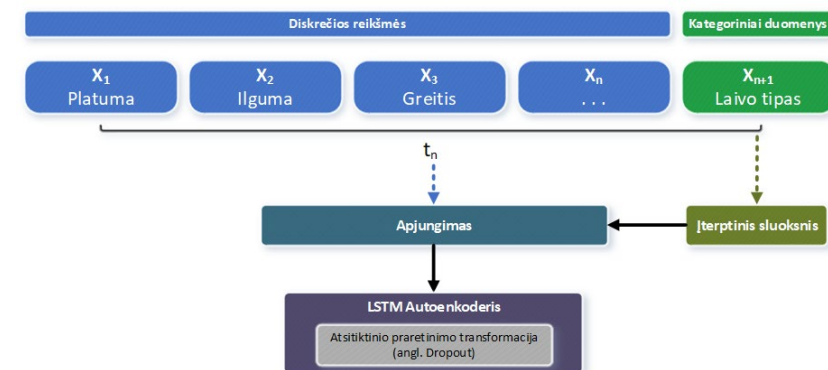
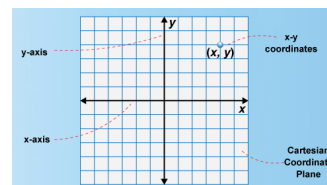
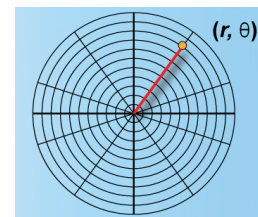
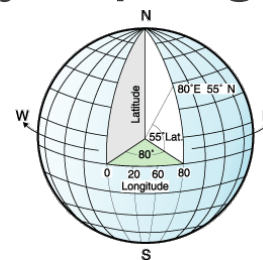
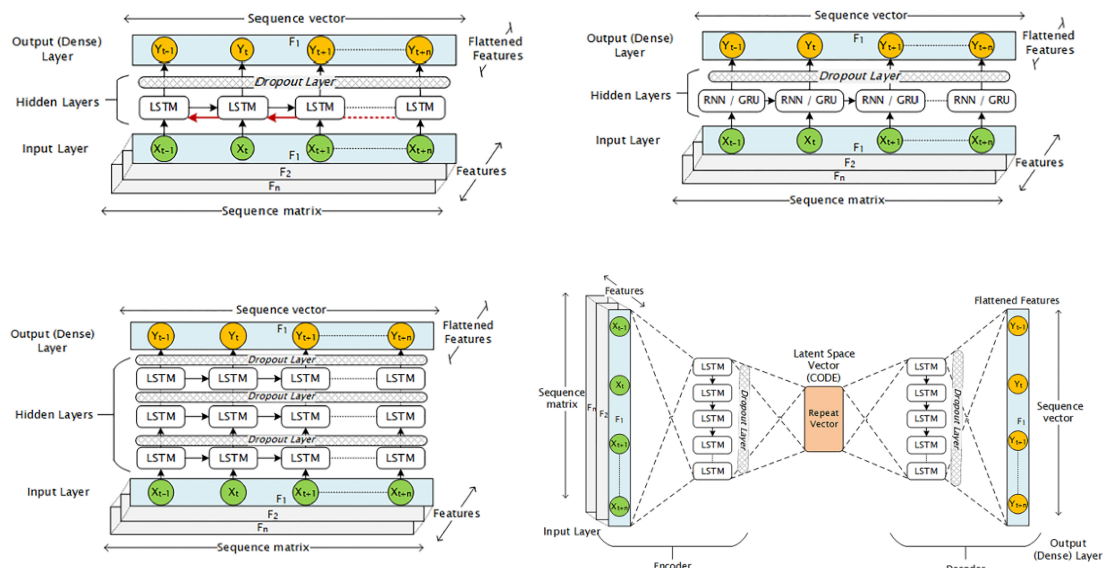


4. **Robertas Jurkus**, Julius Venskys, Jurgita Markevičiūtė, and Povilas Treigys, “*Enhancing Maritime Safety: Estimating Collision Probabilities with Prediction Boundaries using Deep Learning Models*”. In: Ocean Engineering.



Tyrimų/Studijų eiga

- Tirtos skirtingos RNN tinklų architektūros; skirtingos koordinatinių sistemas; kategoriniai duomenys, prognozių ribos/zonos.



2020/2021
1 Kursas

2021/2022
2 Kursas

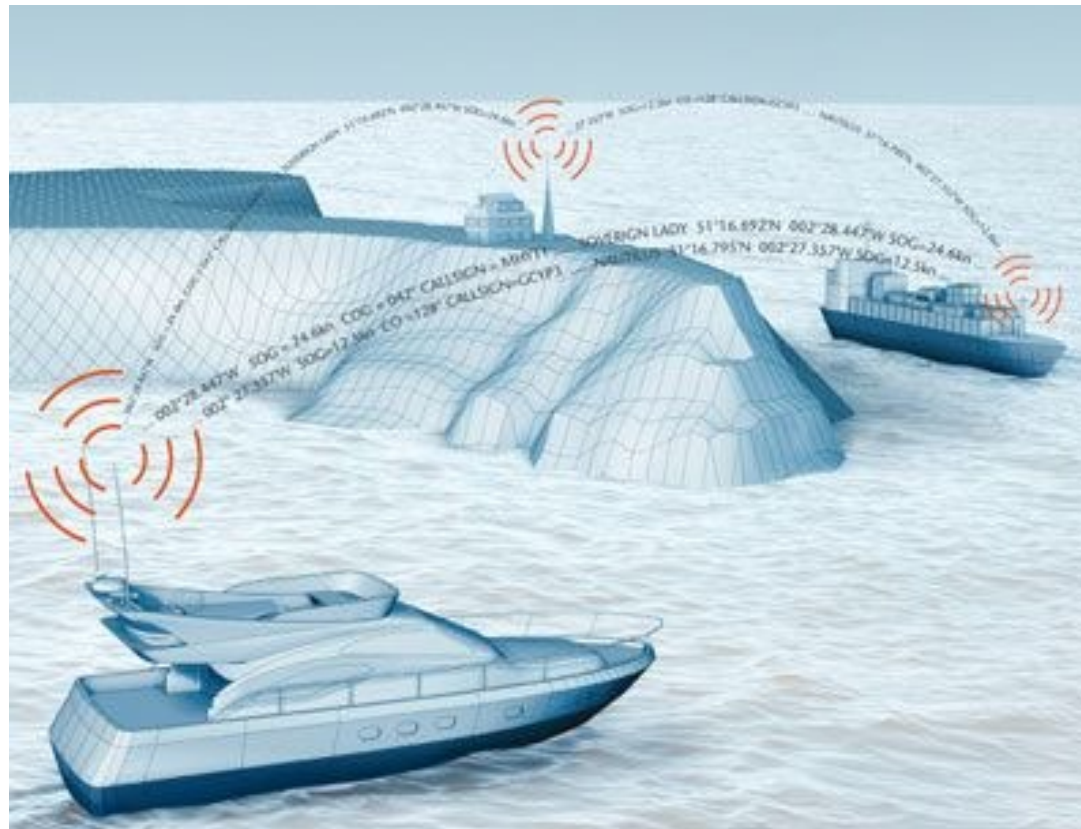
2022/2023
3 Kursas

2023/2024
4 Kursas

Vilniaus
universitetas

Automatinė indentifikavimo sistema (AIS)

- Laivų judėjimo stebėjimas atliekamas remiantis pakrantėse esančiomis AIS tinklo stotimis, kurie duomenis priima palydoviniais imtuvais.





■ AIS laivo tipų duomenys

➤ Duomenų šaltinis yra automatinė identifikavimo sistema (AIS).

➤ Naudojami laivo požymiai:

- geografinės koordinatės (platuma, ilguma);
- greitis (matuojamas mazgais = jūrmylė/val);
- plaukimo kryptis;
- laiko eilutės – skirtumas tarp dviejų laiko momentų trajektorijoje (sekundžių tikslumu);
- platumos ir ilgumos išvestinės (rodančios laiko ir erdvės pokyčius);
- **laivo tipas.**



Tiriamasis regionas Baltijos jūroje prie Bornholmo salos

➤ Tyrimai atlikti su 6 mėn. duomenimis (2021–06 iki 2021–12).

Laivų tipai duomenų imtyje

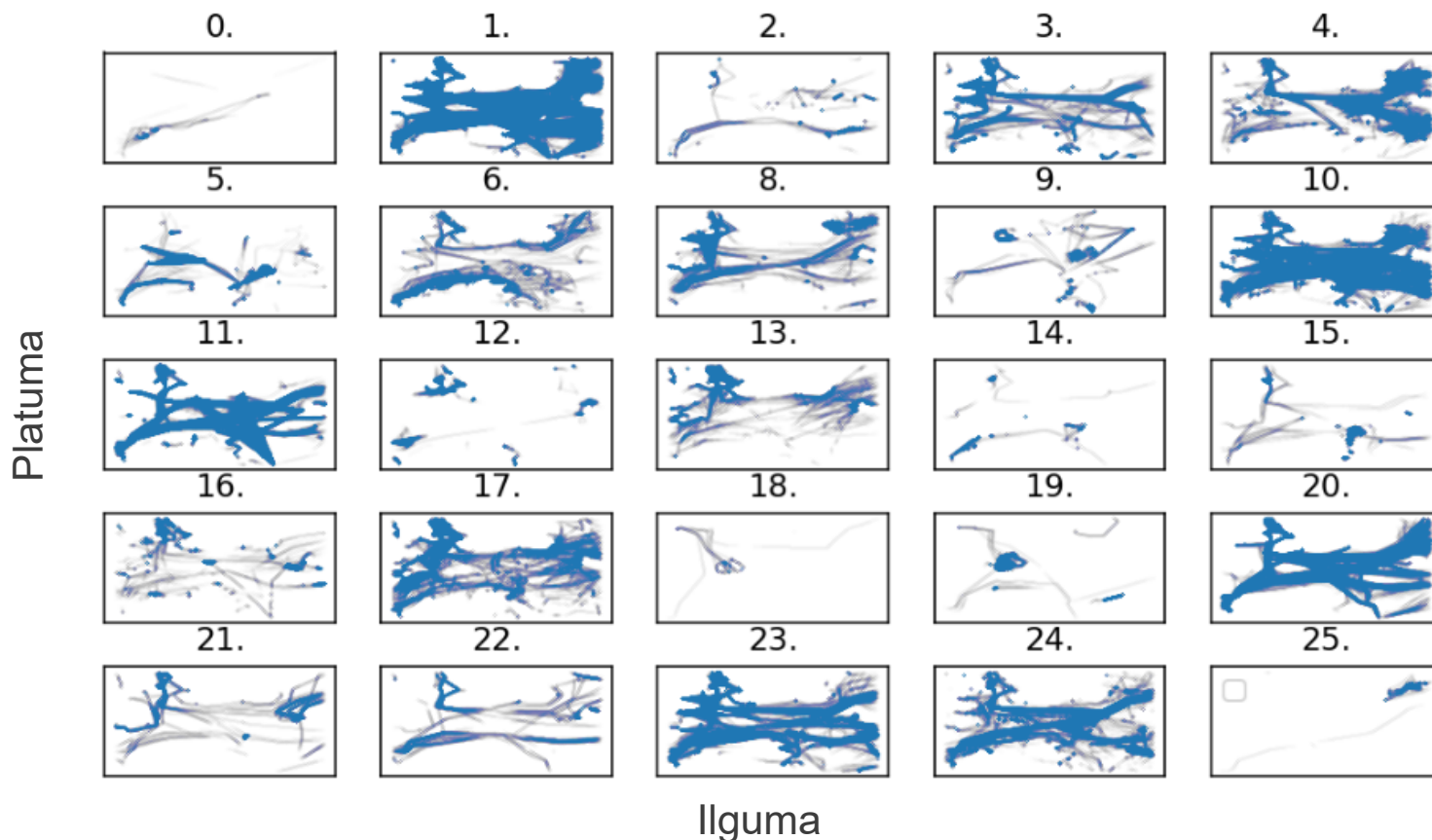
➤ Intensyviausias:

- 1. krovininiai;
- 10. kitų tipų laivai;
- 11. keleiviniai;
- 20. tanklaiviai.

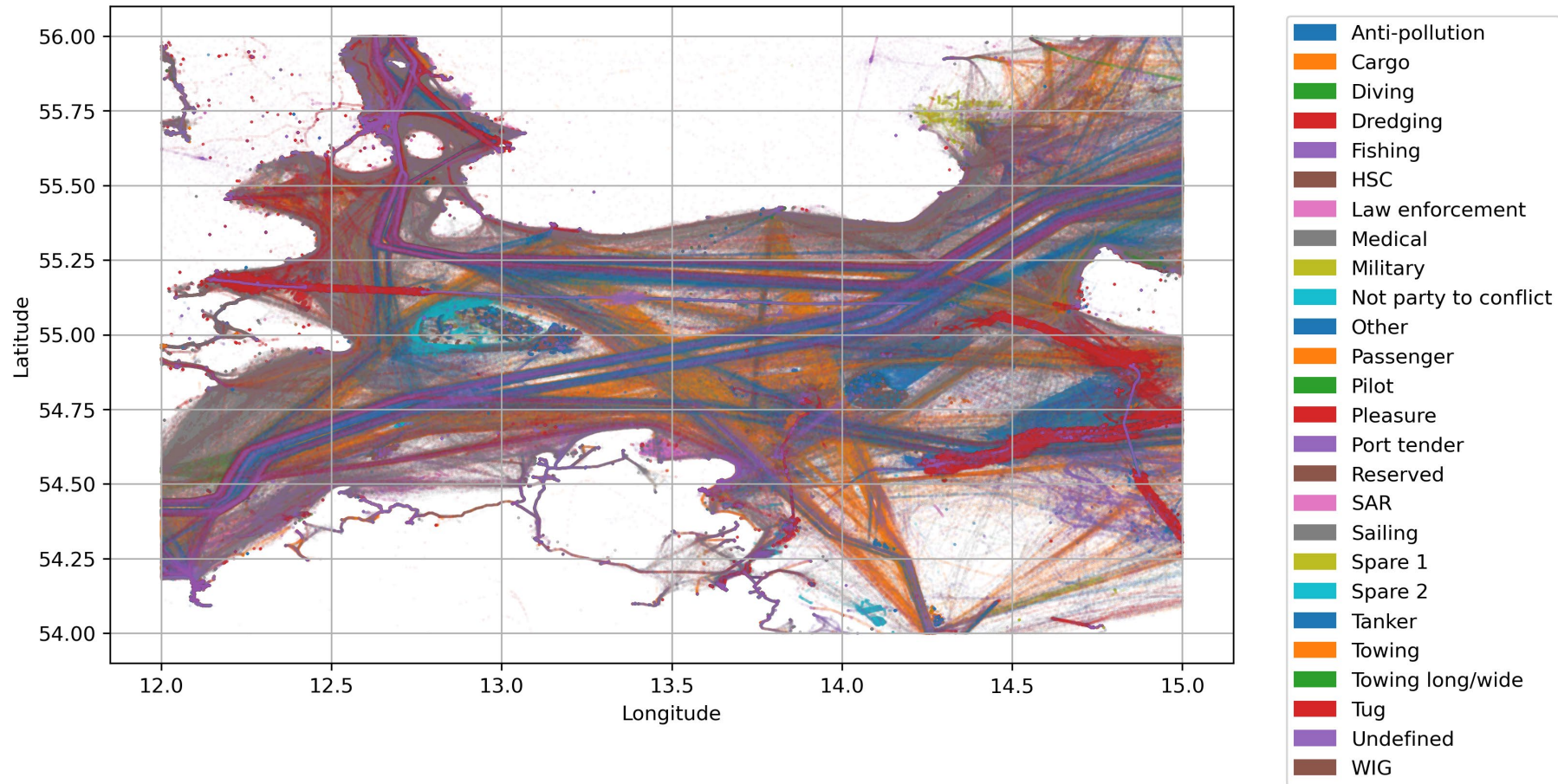
➤ Specifinis:

- 0. taršą valantys;
- 14. uosto priežiūros;
- 18. ir 19. atsarginiai;
- 25. ekranoplanai.

Eismo intensyvumas pagal laivo tipą

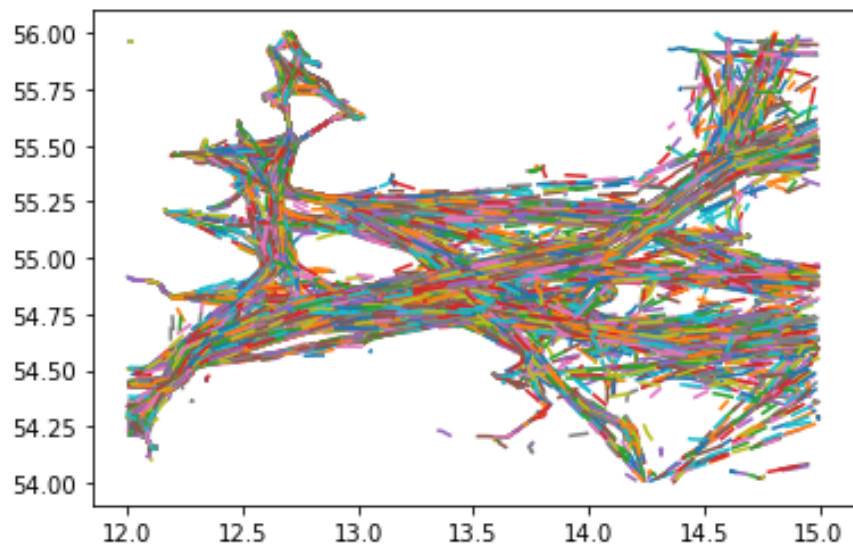
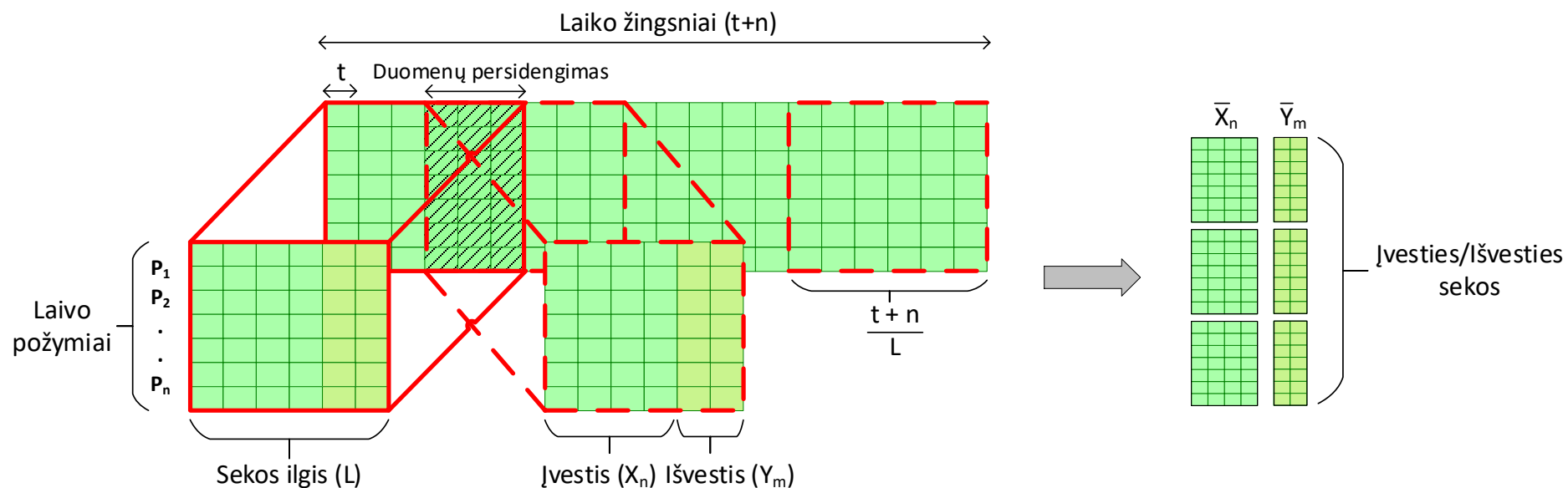


Laivų tipai duomenų imtyje (2)





Sekų generavimas



Požymis	Reikšmė
Visas sekos ilgis	50 laiko žingsnių
Įvesties trajektorijos ilgis	30 laiko žingsnių
Išvesties trajektorijos ilgis	20 laiko žingsnių
Vidutinis laiko žingsnio pokytis laike	~ 60 s
Vidutinis atstumo pokytis erdvėje	~ 320 m
Vidutinis vienos trajektorijos ilgis	~ 16 km
Sugeneruotų sekų skaičius	943 584

Duomenų imties
statistiniai sekų įverčiai

**Vilniaus
universitetas**

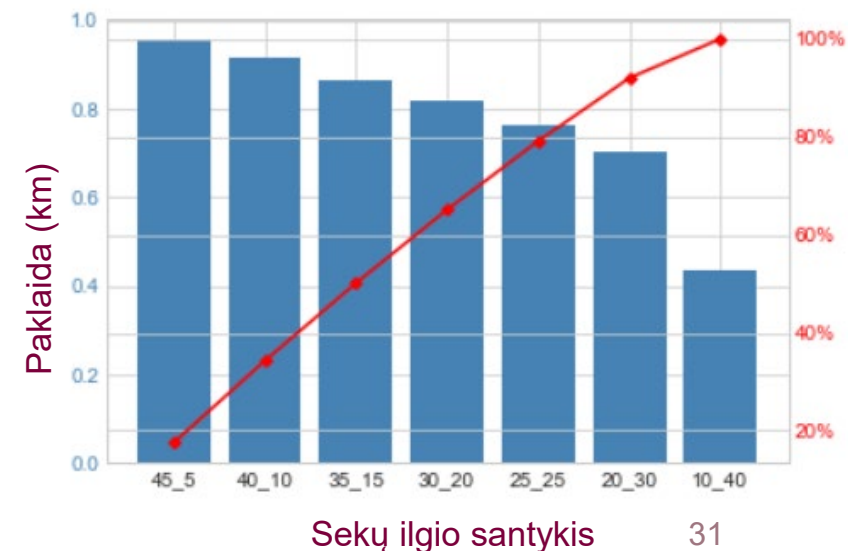


Laivo sustojimo laikas pagal tipą

- Stopping Distance, Turning Circle, Ships Manoeuvring (2020)
- Atliekant stabdymo manevrą laivas sustabdomas panaudojant atbulinės eigos jėgą.

SHIP TYPE	TRACK REACH	TIME TO STOP
Fine and fast ship	5 Ship's length	1.5 to 2 minutes
Cargo ships loaded, displacement 28000 tonnes L 145 m	11-13 ship's length	9-11 minutes
Loaded bulk carrier and tanker	15-20 ship's length	20-25 minutes

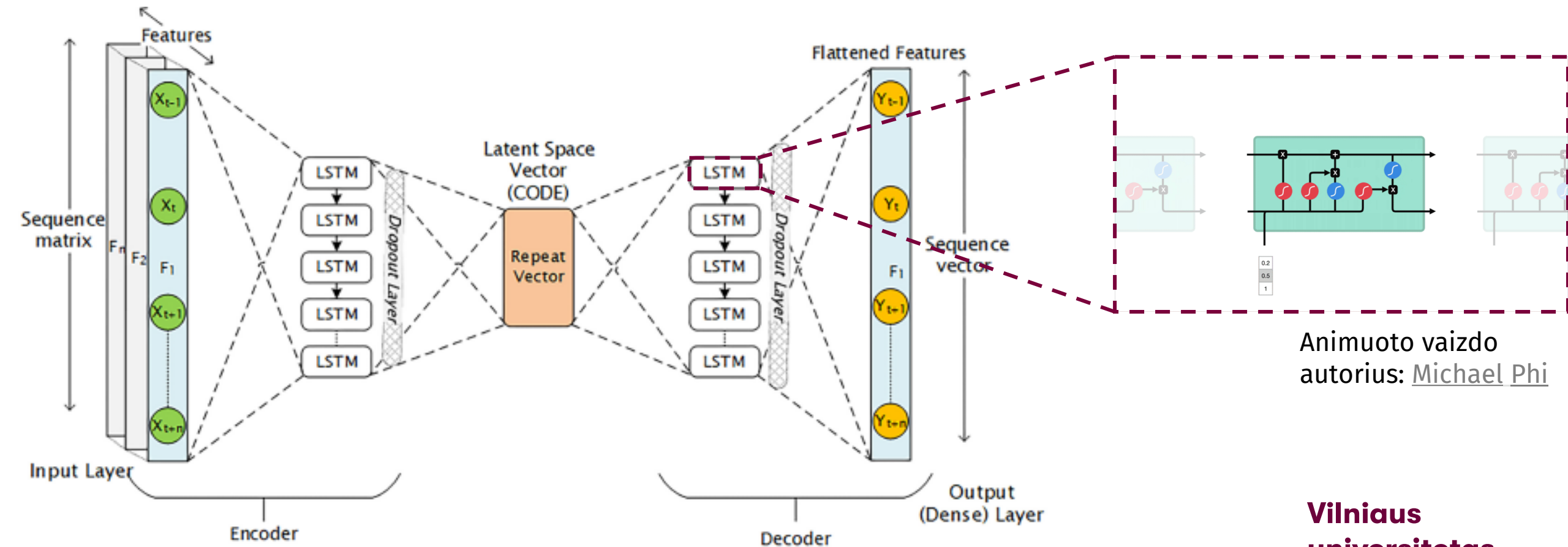
- Prognozės trajektorijos santykio nustatymas Pareto principu, žinoma kaip 80/20 taisyklė.
- Principas pabrėžia disbalansą tarp įvesties (pastangų, išteklių, priežasčių) ir išvesties (rezultatų, išdavų, padarinių). Jis rodo, kad maža dalis priežasčių paprastai lemia didelę dalį padarinių.





LSTM Rekurentinio neuroninio tinklo architektūra

- Tinklo įvestis: laivo charakteristikos požymiai;
išvestis: koordinatės.



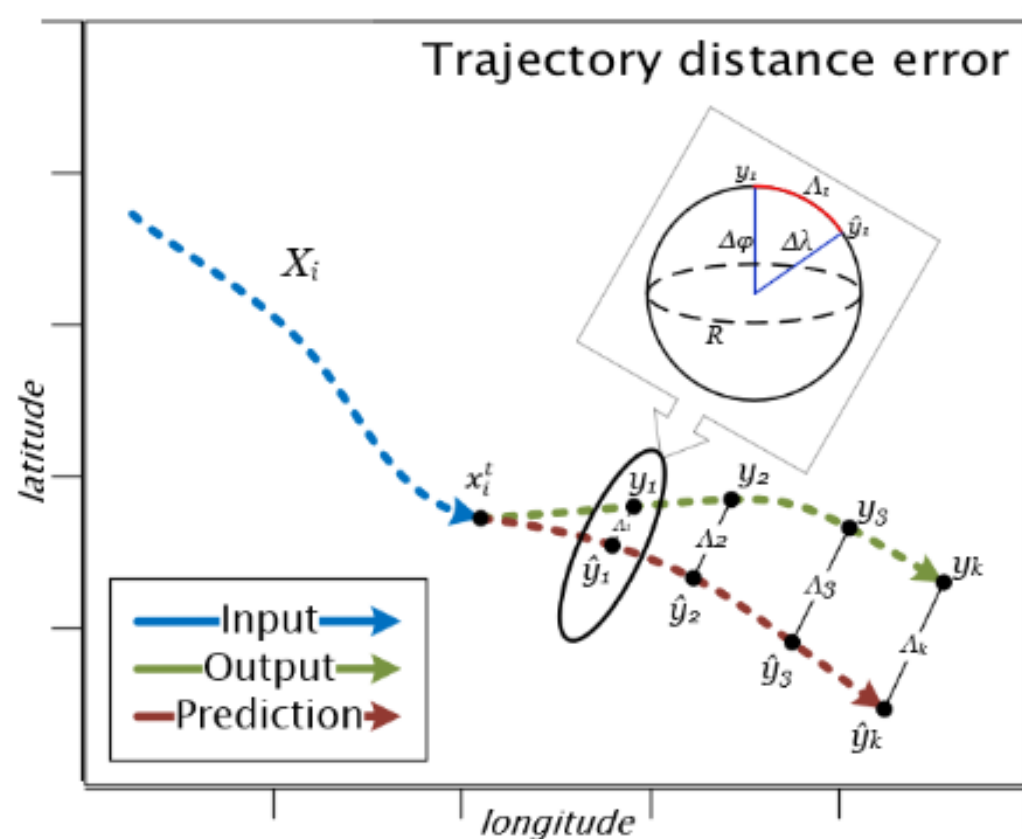
Animuoto vaizdo
autorius: [Michael Phi](#)

**Vilniaus
universitetas**

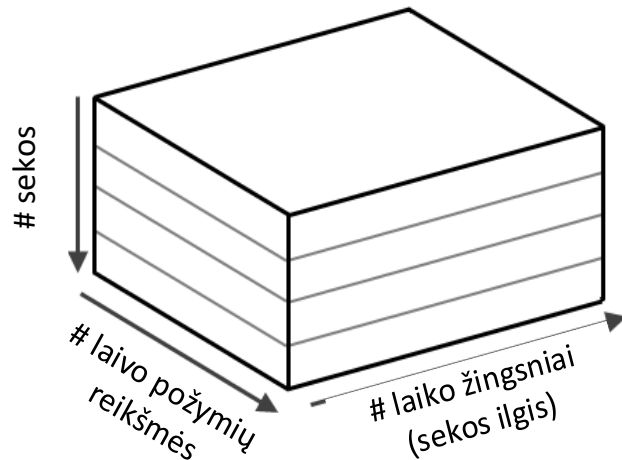
Modelių vertinimas

- MAEH – vidutinė haversinio atstumo absoliučioji paklaida
- Paskaičiuotas tikrosios ir prognozuojamos trajektorijos atstumo paklaida (haversinės funkcijos atstumu)

$$MAEH = \frac{1}{n} \sum_{i=1}^n |haversine(y_i - \hat{y}_i)|$$

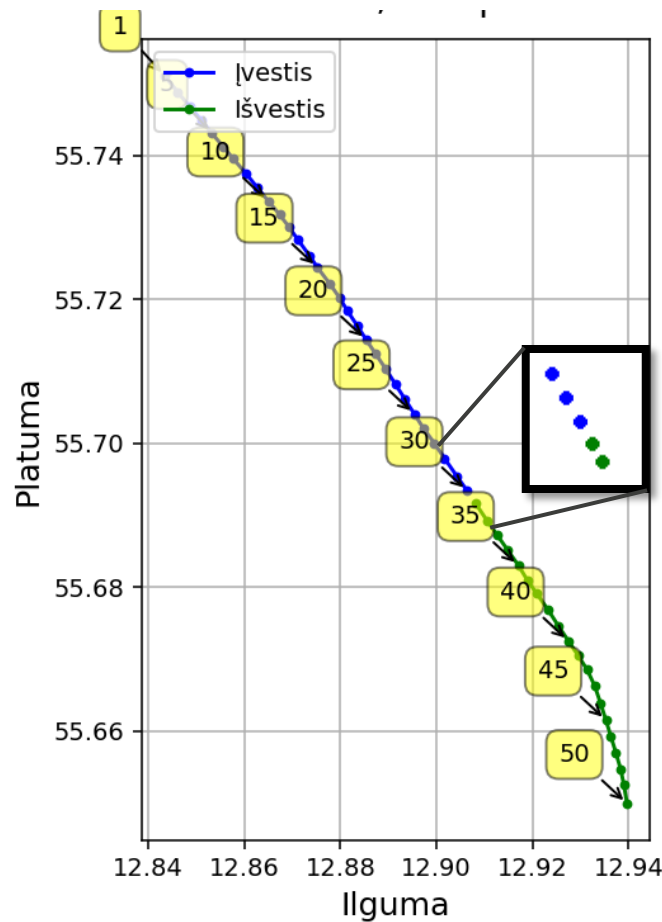


Duomenų apibendrinimas



Trimatis duomenų struktūros vaizdas

$$X_i = \begin{bmatrix} x_i^{(1,1)} & x_i^{(1,2)} & \dots & x_i^{(1,n)} \\ \dots & \dots & \dots & \dots \\ x_i^{(t,1)} & x_i^{(t,2)} & \dots & x_i^{(t,n)} \end{bmatrix} \quad Y_i = \begin{bmatrix} y_i^{(1,1)} & y_i^{(1,2)} \\ \dots & \dots \\ y_i^{(k,1)} & y_i^{(k,2)} \end{bmatrix}$$



Atsitiktinės sekos pavyzdys

➤ Įvestis:

- Ilguma
- Platuma
- Greitis
- Kryptis
- ...

➤ Išvestis:

- Ilguma
- Platuma

**Vilniaus
universitetas**

Požymių permutacija

Metodas (angl. permutation feature importance), naudojamas požymių svarbai prognozuojamajame modelyje įvertinti. Juo įvertinamas kiekvienos savybės poveikis modelio veikimui, permutuojant (permaišant) savybės reikšmes ir matuojant, kaip sumažėja veikimas.

[n, 20, f] – (samples, time steps, features)

Lat	Long	SOG	Heading	DateDiff	dLat/dt	dLong/dt
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...

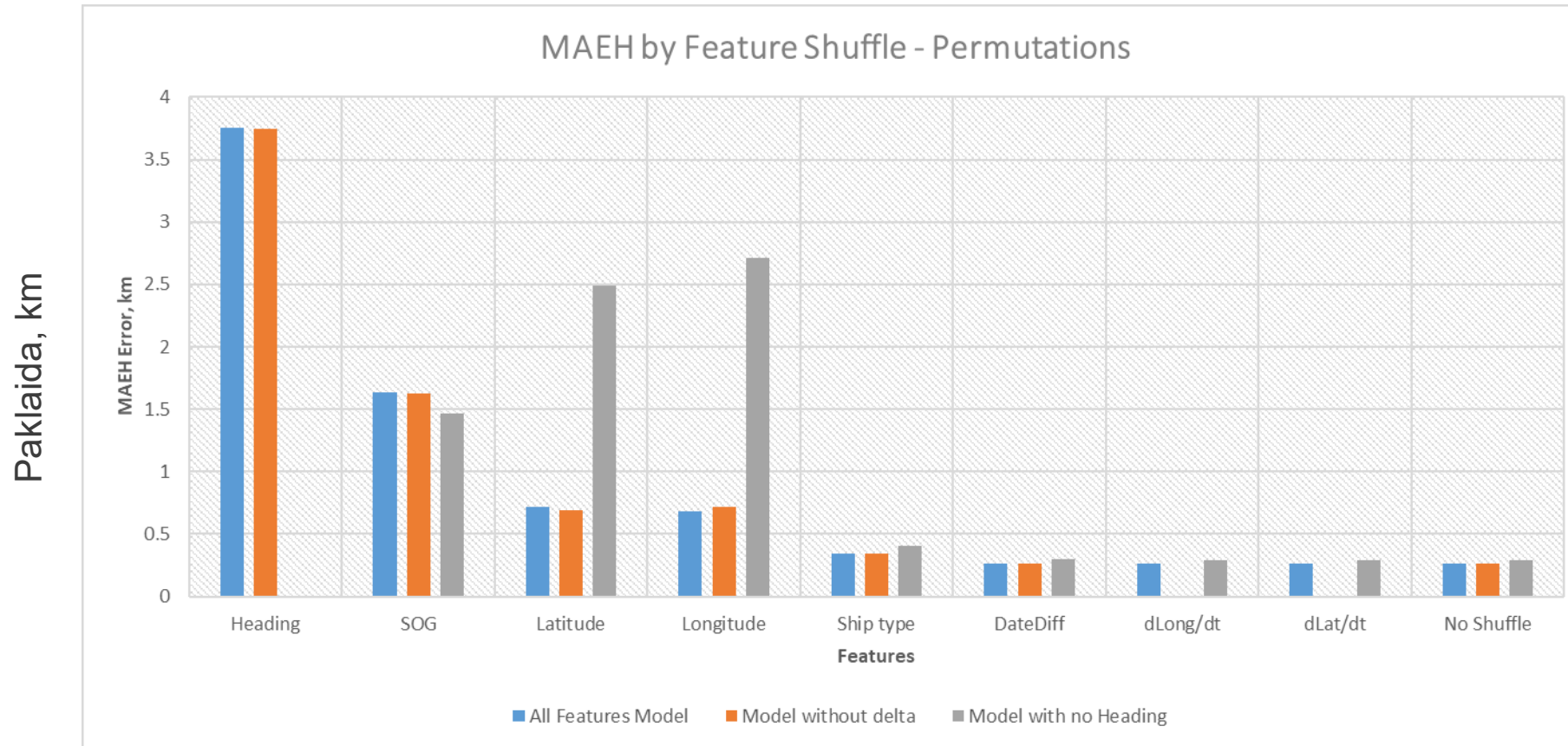
Lat	Long	SOG	Heading	DateDiff	dLat/dt	dLong/dt
0.10	0.24	0.55	0.6	0.4	0.72	0.41
0.10	0.29	0.67	0.6	0.4	0.77	0.12
0.10	0.29	0.60	0.5	0.4	0.22	0.02
...	...	...	...	...	...	...

Lat	Long	SOG	Heading	DateDiff	dLat/dt	dLong/dt
0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	0.2	0.3	0.4	0.5	0.6	0.7
...	...	...	...	...	...	...

Lat	Long	SOG	Heading	DateDiff	dLat/dt	dLong/dt
0.11	0.22	0.33	0.44	0.55	0.66	0.77
0.11	0.22	0.33	0.44	0.55	0.66	0.77
0.11	0.22	0.33	0.44	0.55	0.66	0.77
...	...	...	...	...	...	...

Požymių permutacija (2)

[n, 20, f] – (samples, time steps, features)



Duomenų rinkinio savybė

$$X_{\Delta t} = t_s - t_{s-1}$$

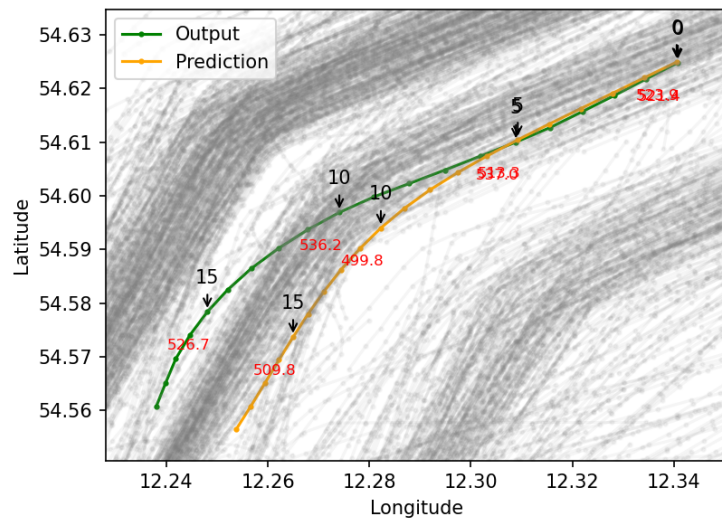
$$X_{\delta Lon} = \frac{X_{Lon,s} - X_{Lon,s-1}}{X_{\Delta t}}$$

$$X_{\delta Lat} = \frac{X_{Lat,s} - X_{Lat,s-1}}{X_{\Delta t}}$$

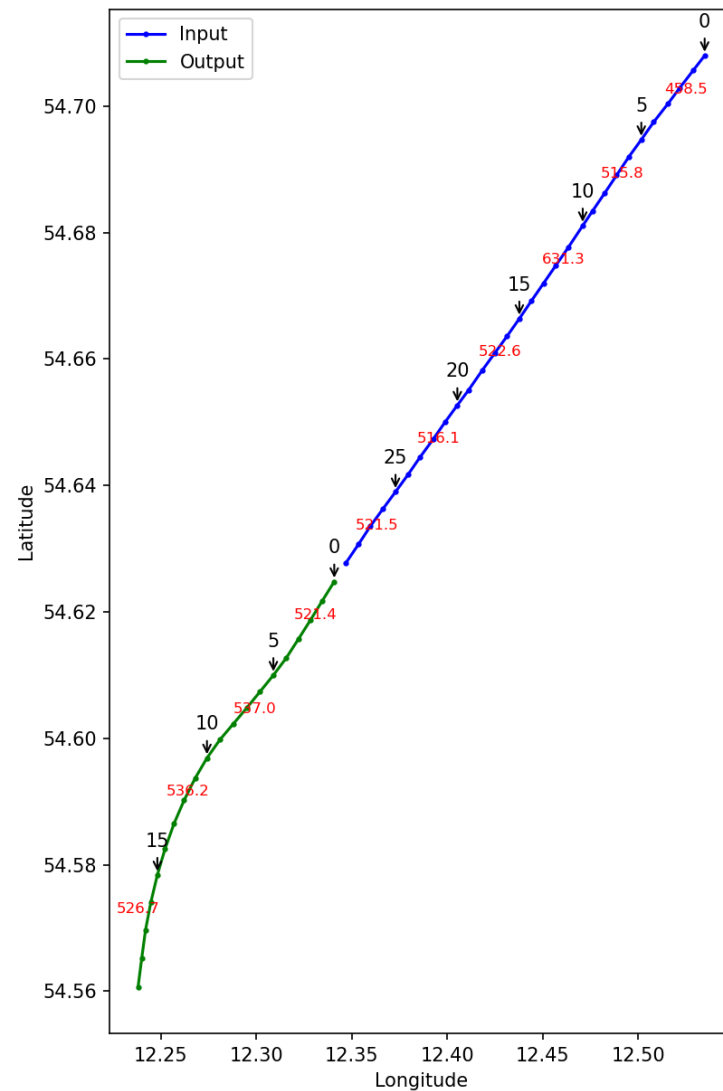
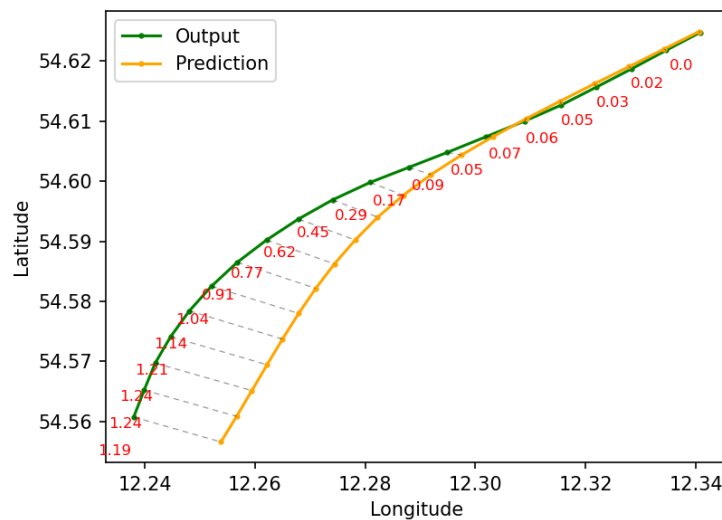
Prognozės pavyzdys

Actual and Predicted movement in the traffic (m)

Distances (m): AVG = 524.91, SUM = 25139.76

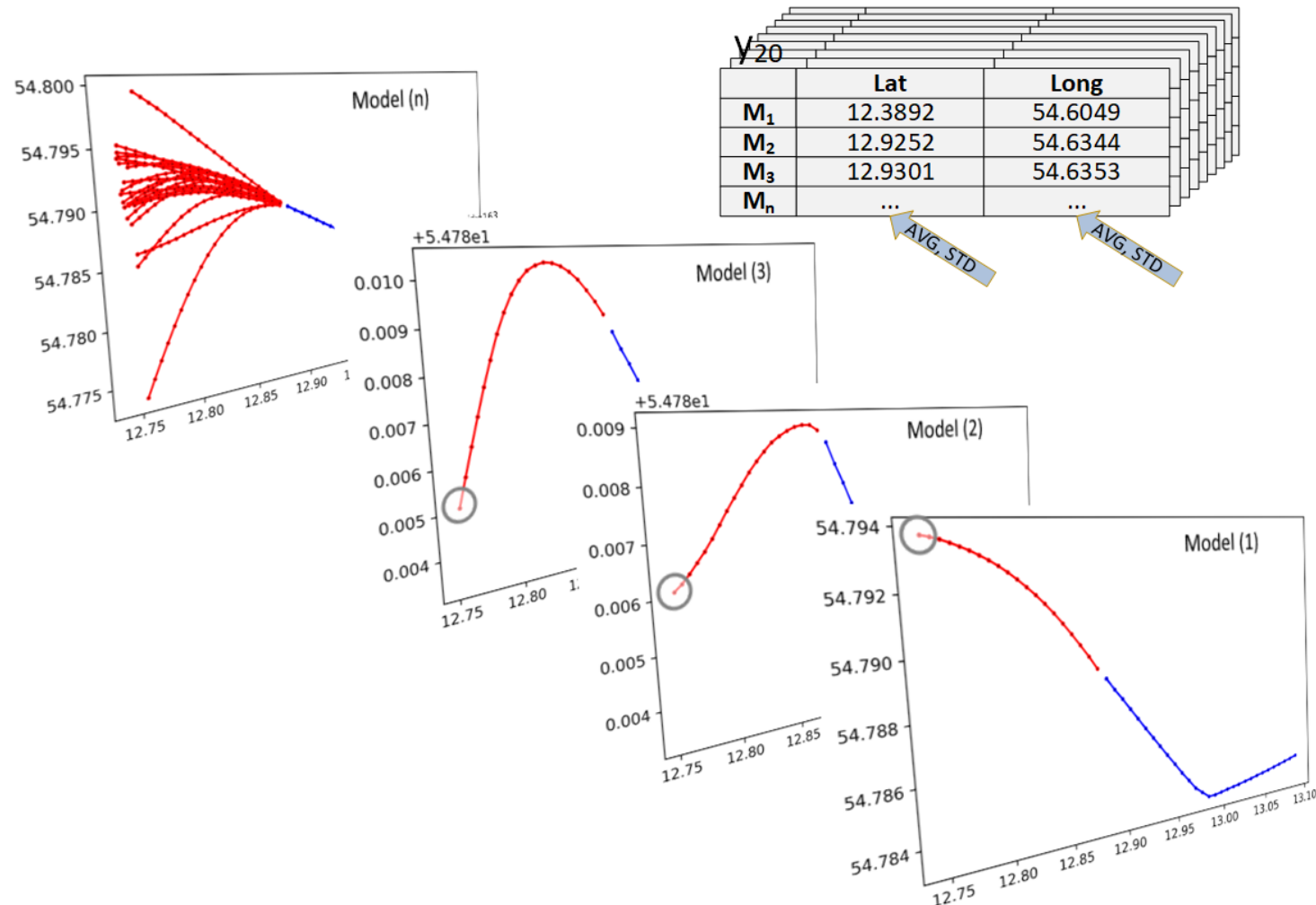


MAEH = 0.588 km



Prognozės intervalų skaičiavimas modeliuose

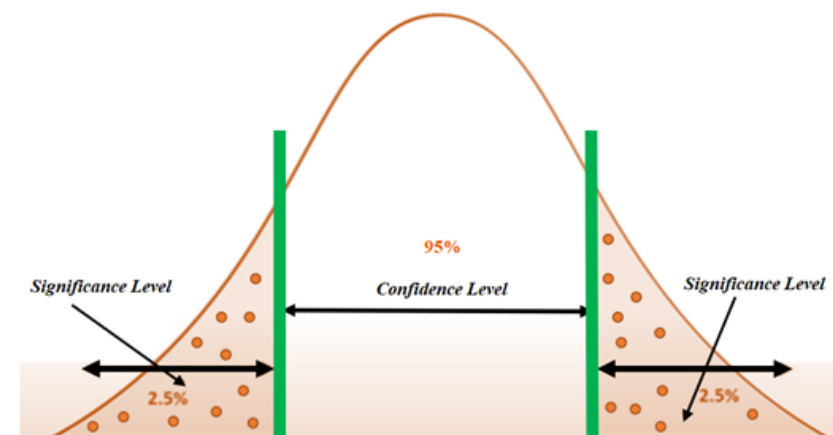
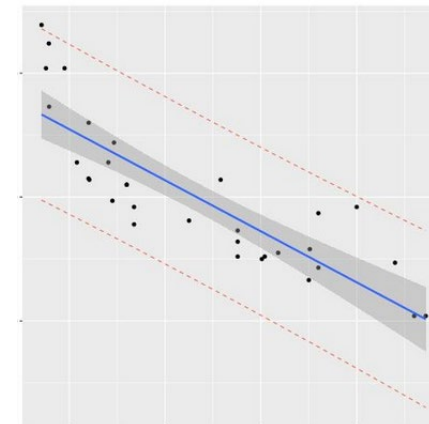
➤ Ekstrapoliuojamos trajektorijos prognozių sekos



Regionų nustatymo metodai

➤ Nagrinėti metodai (taikant pasiklivimo lygį su 95%):

Pavadinimas	Metodas	Tipas
Elipsoidinis prognozavimo regionas (EPR)	Geometrinis – atsižvelgiama į prognozuojamų trajektorijos taškų dispersiją ir erdvinį pasiskirstymą	Ne parametrinis (ar mažiau param.)
Prognozių intervalas (PI)	Statistinis - remiasi prielaidomis apie duomenų pasiskirstymą (pvz., normalusis skirstinys), naujo taško galimos ribos	Parametrinis
Pasiklivimo intervalas (CI)	Statistinis - remiasi prielaidomis apie duomenų pasiskirstymą (pvz., normalusis skirstinys), tikroji reikšmė aplink vidurkį	Parametrinis
Konforminis prognozavimo regionas (CPR)	Statistinis - labiau empirinis, sudarytas iš praeities duomenų (kalibravimo) be griežtų parametrinių prielaidų, neatitikties įvertinimas	Ne parametrinis



$$SE = \frac{SD}{\sqrt{n}}$$

$$(1) \quad W_{PI} = t\left(\frac{1+\alpha}{2}, n-1\right) \cdot SE \cdot \sqrt{1 + \frac{1}{n}} \quad (2)$$

$$W_{CI} = t\left(\frac{1+\alpha}{2}, n-1\right) \cdot \frac{SE}{\sqrt{n}} \quad (3)$$

EPR

Algorithm 1 EPR Calculation

function CALCULATEEPR(*points*, *center*)

Input: *points* - array of latitude and longitude points

Input: *center* - center point (mean) of the EPR

Output: *epr* - the calculated Ellipsoidal Prediction Region as a polygon

$hull \leftarrow \text{CONVEXHULL}(points)$

$hullVertices \leftarrow points[hull.vertices]$

$covMatrix \leftarrow \text{COVARIANCE}(hullVertices)$

$[eigValues, eigVectors] \leftarrow \text{EIGEN}(covMatrix)$

$confidenceLevel \leftarrow 0.95$

$radii \leftarrow \sqrt{\text{CHISQUAREDINVERSE}(confidenceLevel, 2)} * \sqrt{eigValues}$

$angleArray \leftarrow \text{array of angles from } 0 \text{ to } 2\pi$

$ellipsoidPoints \leftarrow \text{EMPTYARRAY}(size : 100, dimensions : 2)$

for $i \leftarrow 1$ **to** 100 **do**

$ellipsoidPoints[i][1] \leftarrow radii[1] * \cos(angleArray[i])$

$ellipsoidPoints[i][2] \leftarrow radii[2] * \sin(angleArray[i])$

end for

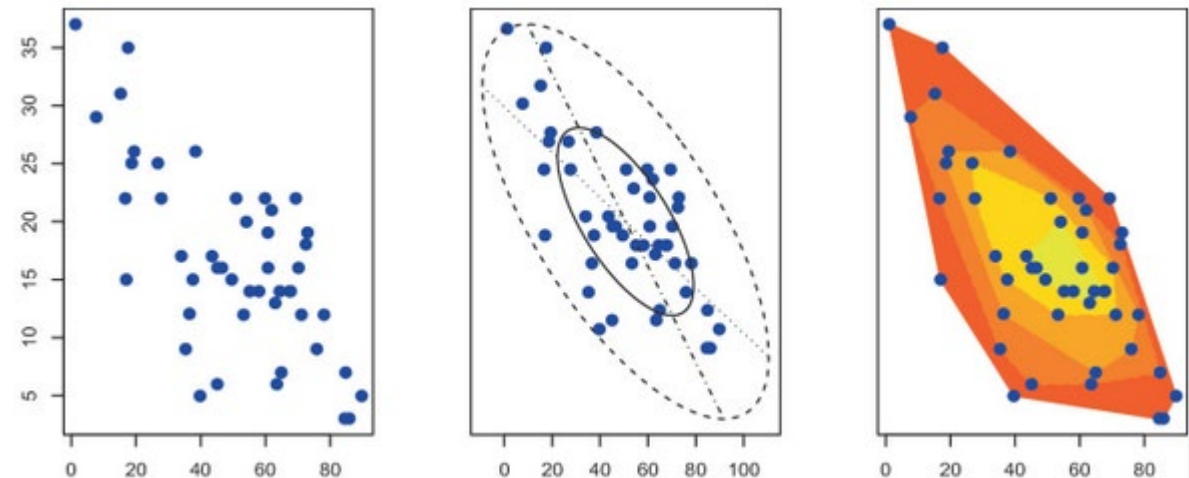
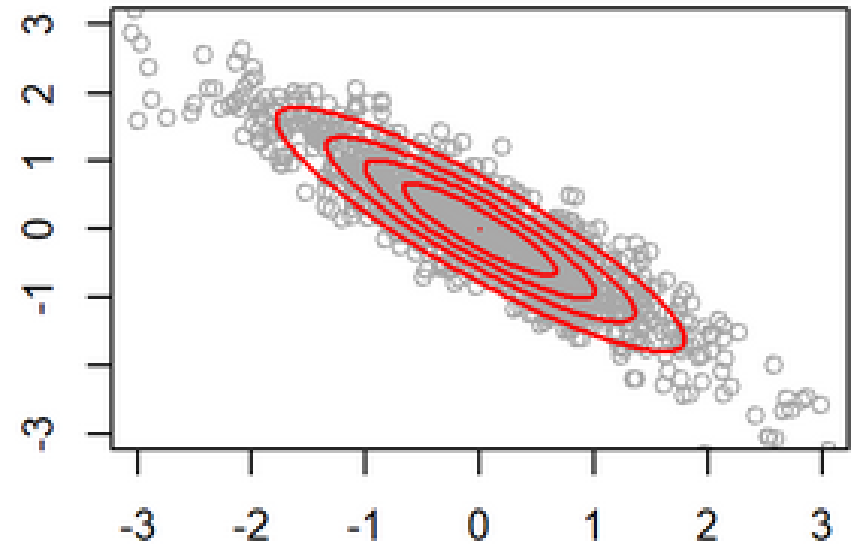
$rotatedPoints \leftarrow eigVectors \cdot ellipsoidPoints$

$translatedPoints \leftarrow rotatedPoints + center$

$epr \leftarrow \text{CREATEPOLYGON}(translatedPoints)$

return *epr*

end function



CPR

The model generates predictions for each instance in the calibration set, and the corresponding nonconformity scores are calculated. In a multi-model approach, these scores represent an average of the distances across all models at each time step. The problem of calibrating CPR is approached by calculating the average nonconformity scores across all models at each time step in the calibration set. For the i -th instance in the calibration set and time step t , the nonconformity score is defined $\mathcal{R}_t^{(i)}$ as the Euclidean distance between the actual value $\mathbf{Y}_t^{(i)}$ and the predicted value $\hat{\mathbf{Y}}_t^{(i)}$.

The average nonconformity score at time step t across all instances is given by:

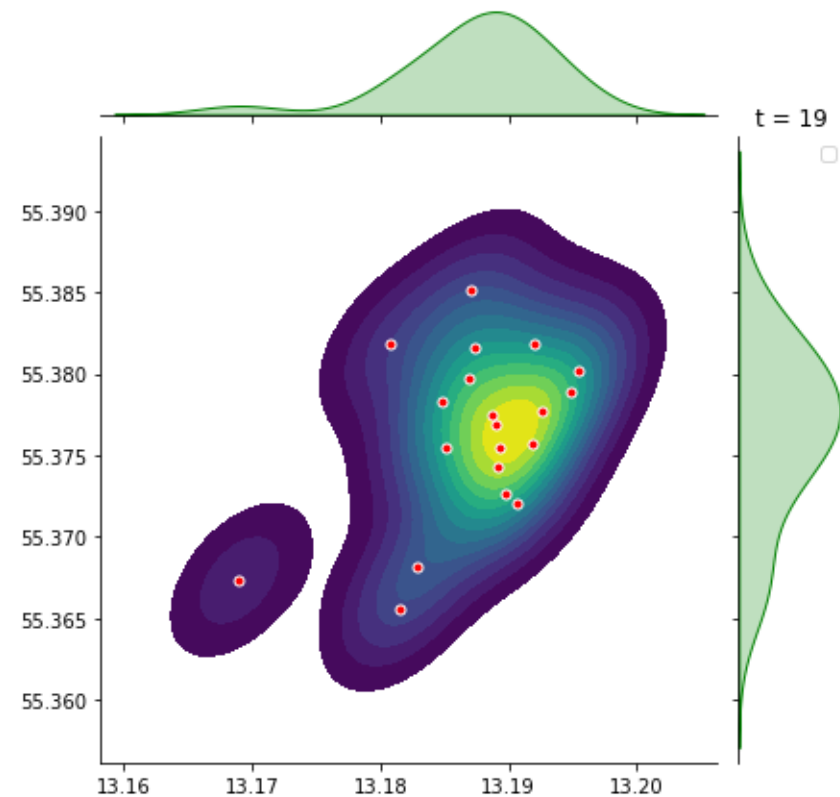
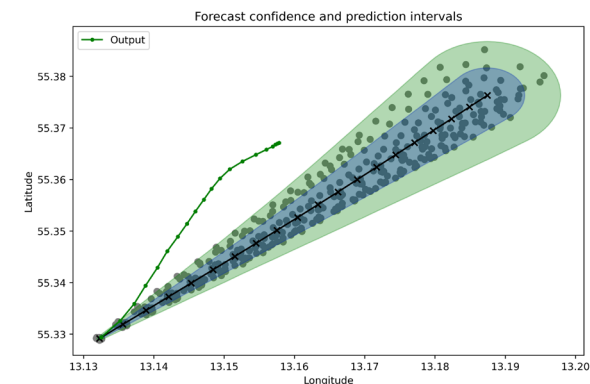
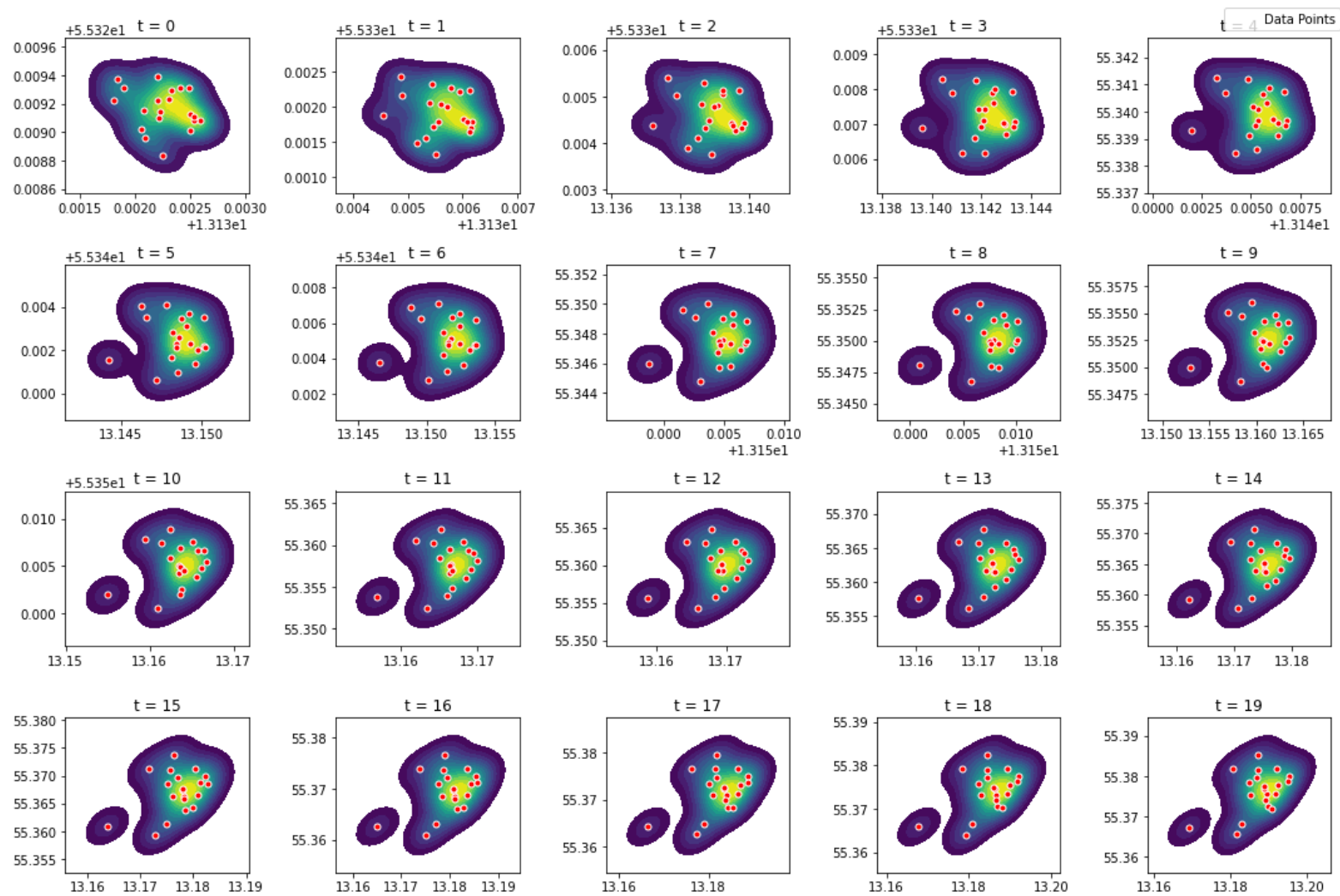
$$\bar{\mathcal{R}}_t = \frac{1}{n} \sum_{i=1}^n \mathcal{R}_t^{(i)},$$

where n denotes the number of instances in the calibration set. Consequently, we seek to optimize the following problem:

$$\begin{aligned} & \text{minimize} && \text{Quantile}(\{\bar{\mathcal{R}}_1, \dots, \bar{\mathcal{R}}_T\}, 1 - \delta) \\ & \text{subject to} && \sum_{t=1}^T \alpha_t \bar{\mathcal{R}}_t = 1 \\ & && \alpha_t > 0, \quad t = 1, \dots, T \end{aligned}$$

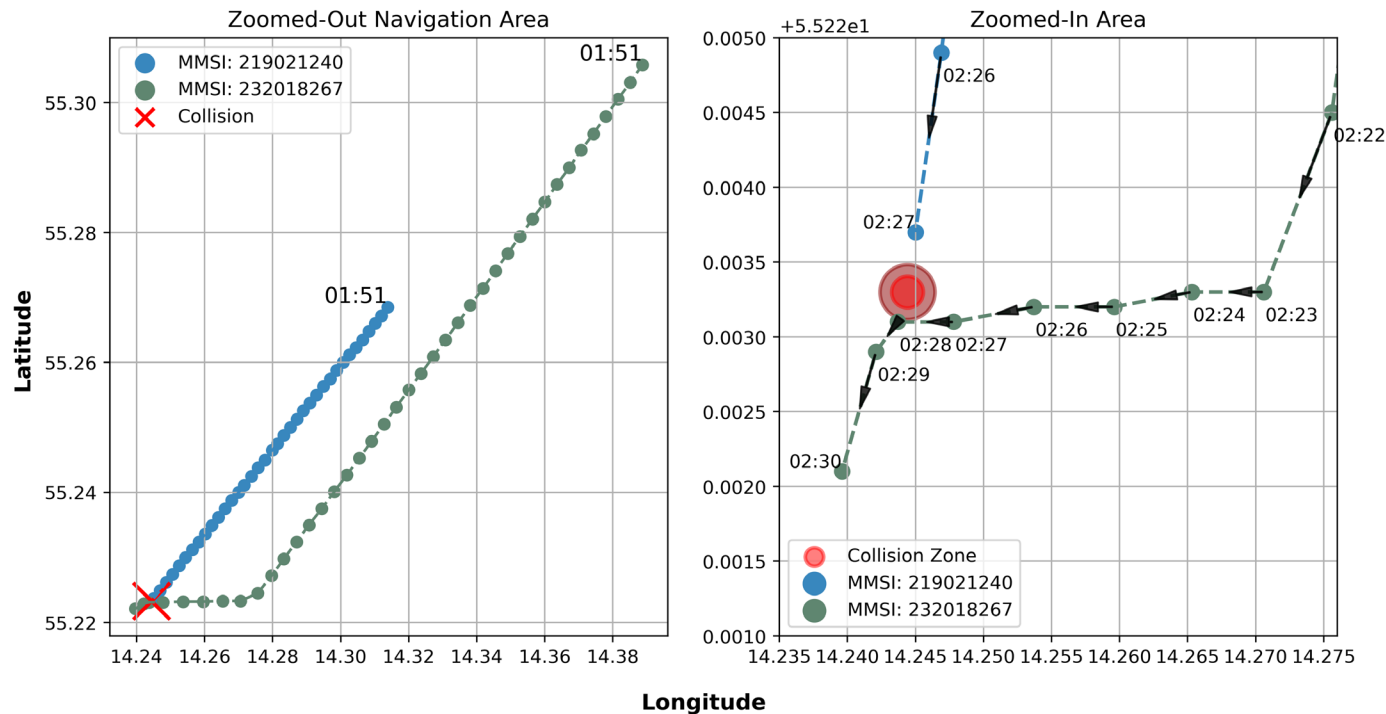
Here, α_t are the parameters to be optimized, which scale the nonconformity scores at each time step, and δ represents the desired confidence level for the CPR. The objective function minimizes the quantile of the averaged nonconformity scores, aligning with the confidence level to construct valid prediction intervals.

Koordinačių pasiskirstymas sekoje



EPR metodo verifikavimas realiame susidūrime

- Incidentas 2021-12-13 Baltijos jūroje
- Susidūrimo aptikimas remiantis sukurto modelio prognozavimo intervalu, kurio patikimumas 95 %



Cargo ships collide in Baltic Sea

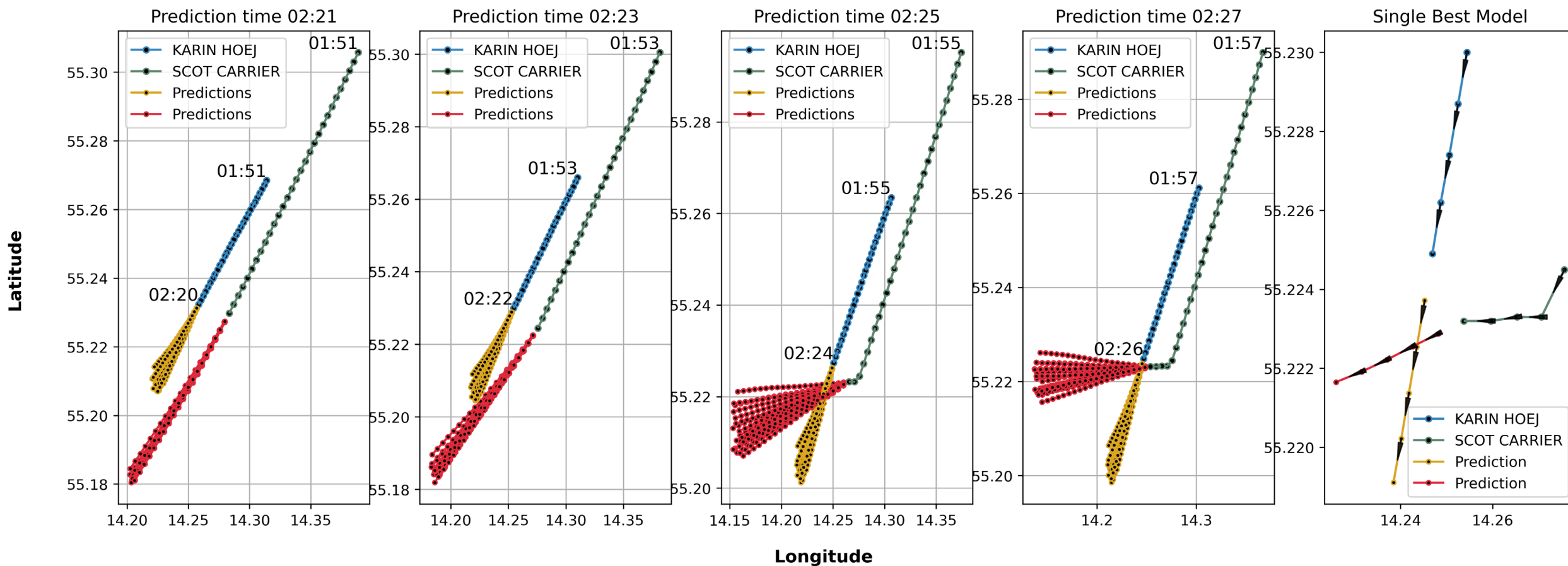
Two cargo ships collided on Monday during foggy conditions in the Baltic Sea between the Danish island of Bornholm and the southern Swedish city of Ystad, and a rescue operation was launched for two missing people, authorities said.



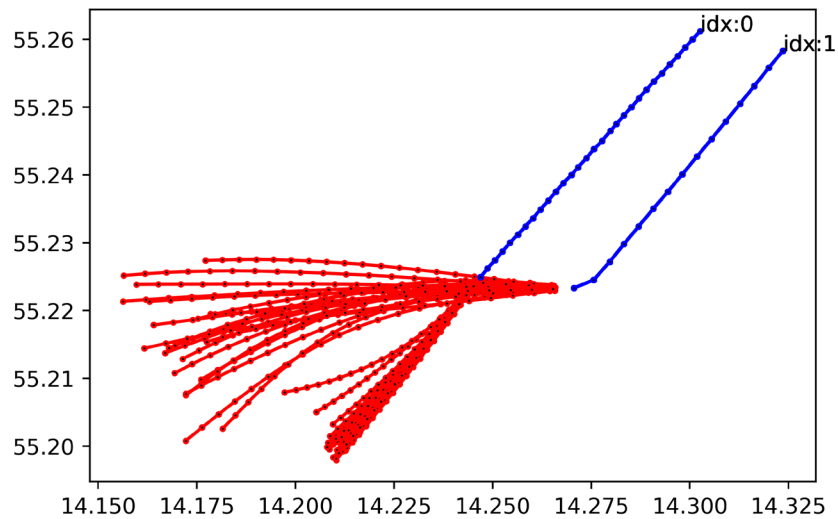
Vilniaus
universitetas

Anomalių aptikimas realiuose susidūrimuose

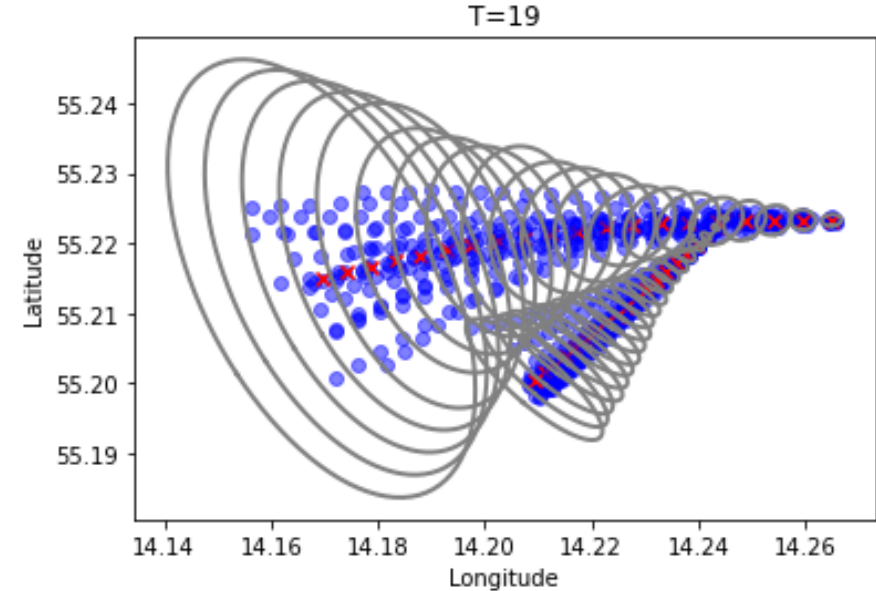
➤ Modelių prognozės skirtingu laiko momentu



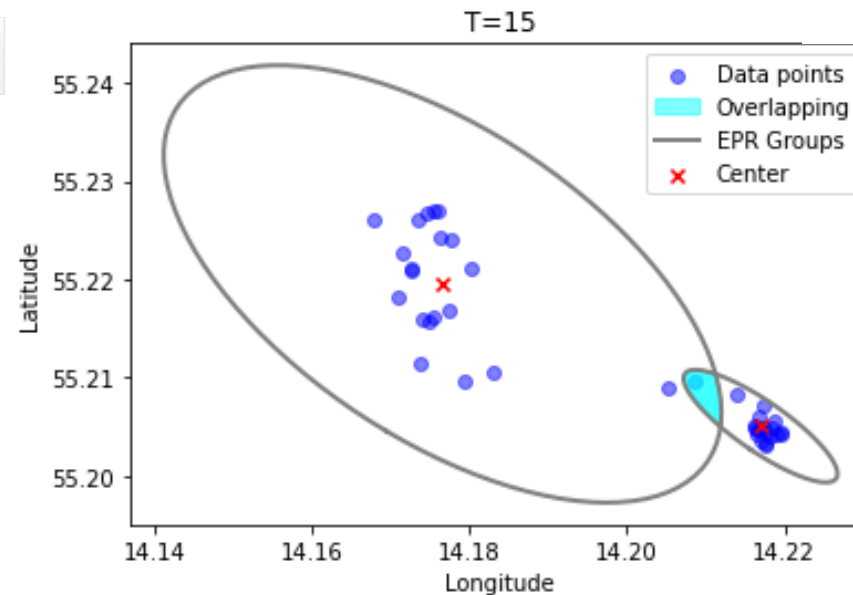
EPR metodo verifikavimas realiame susidūrime (2)



Modelių prognozė trajektorijoje

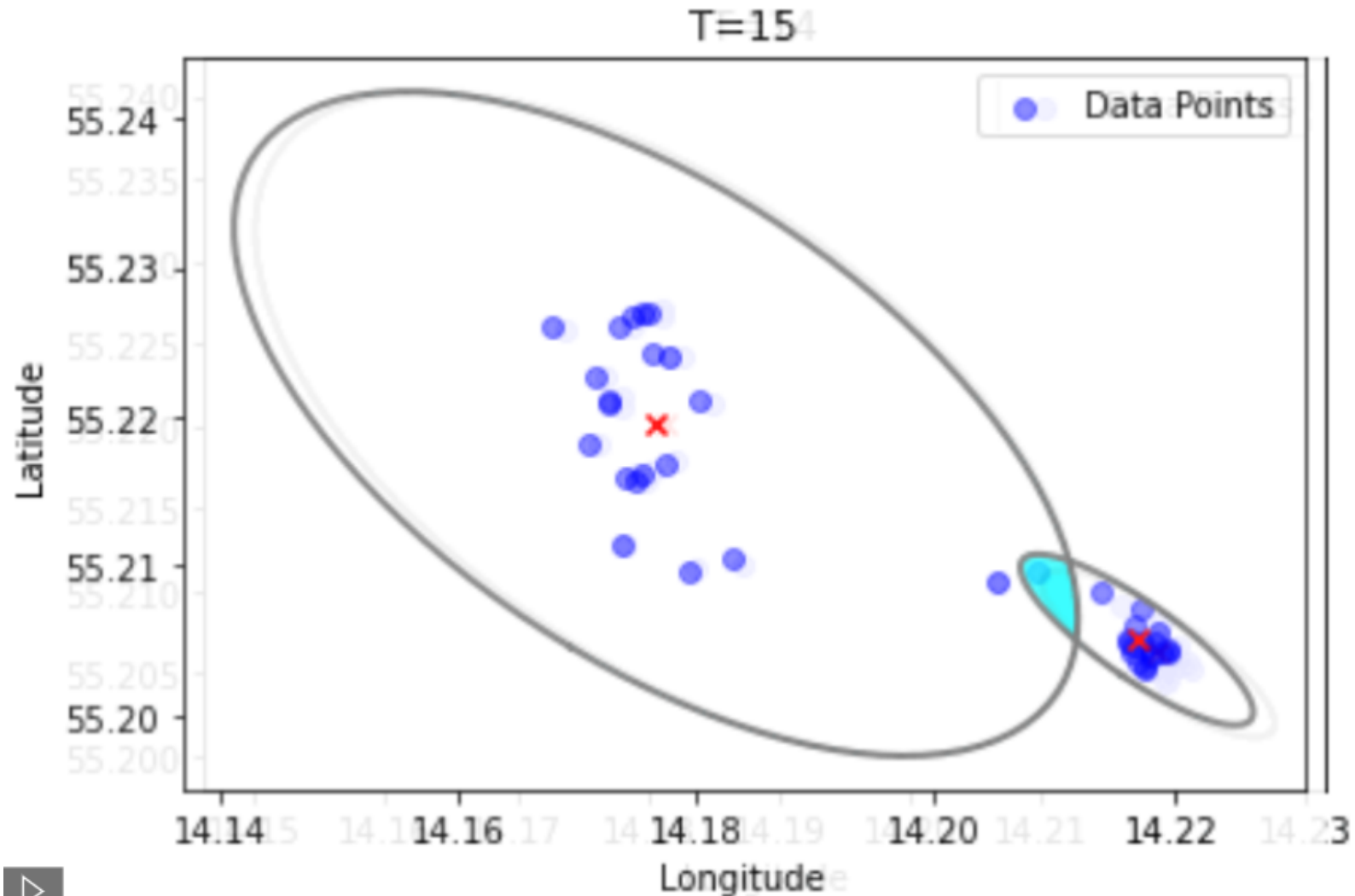


Prognozės intervalai T laiko momentu



Susidūrimo rizika T = 15

CASE #1 EPR – Realus susidūrimas (1)

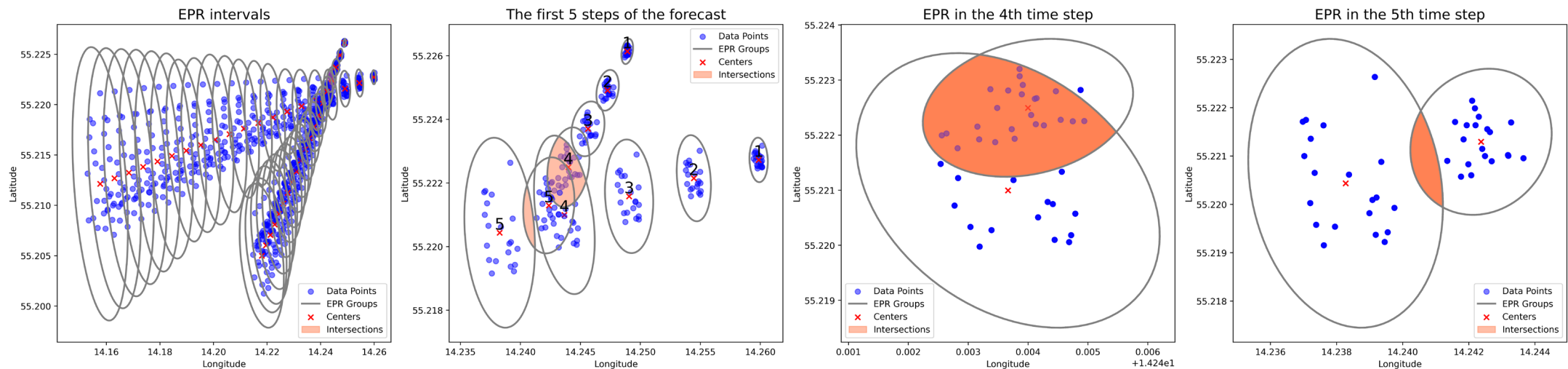


Tikimybiniis vertinimas

➤ EPR atveju

$$P(\text{laivų trajektorijos susikirs}) = \frac{V_{A \cap B}}{V_A + V_B - V_{A \cap B}},$$

Čia V_A ir V_B yra dviejų elipsių plotai, o $V_{A \cap B}$ yra jų susikirtimo plotas.



Time Step	EPR Area A, m ²	EPR Area B, m ²	Intersection Area, m ²	intersection_probability (%)
1	11 425.5	32 978.4	0.0	0.00%
2	35 960.2	117 517.0	0.0	0.00%
3	72 216.4	220 926.0	0.0	0.00%
4	132 867.0	368 967.0	96 986.7	23.96%
5	198 320.0	536 512.0	22 066.5	3.10%
6	274 578.0	762 444.0	0.0	0.00%



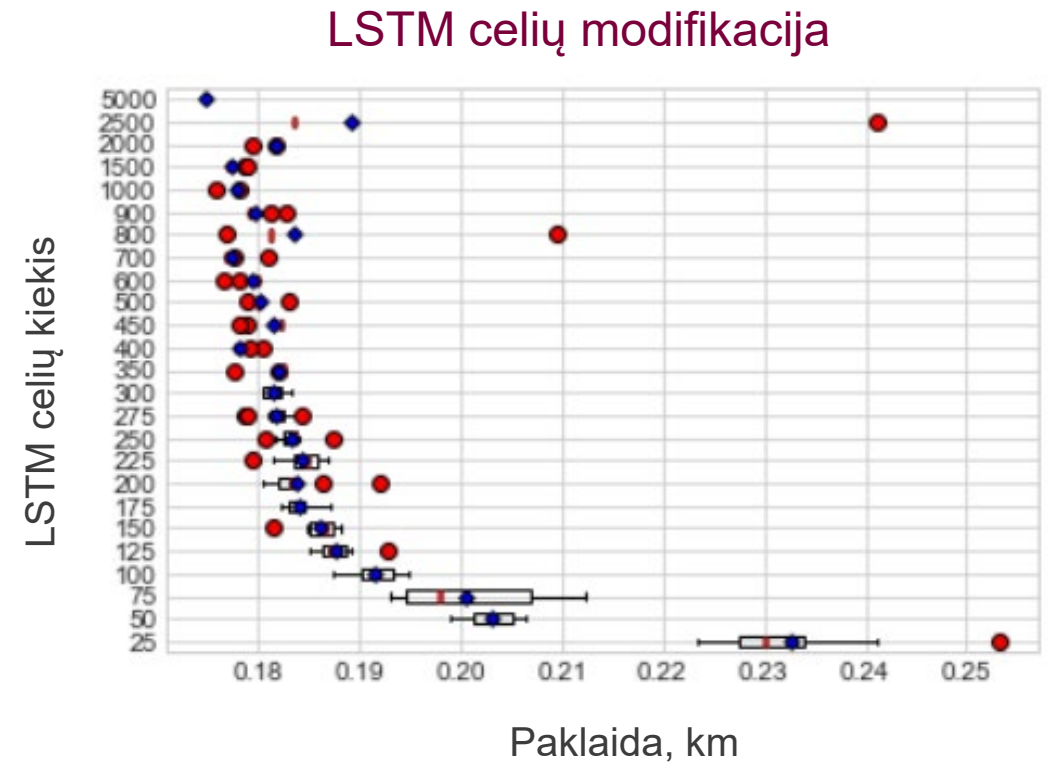
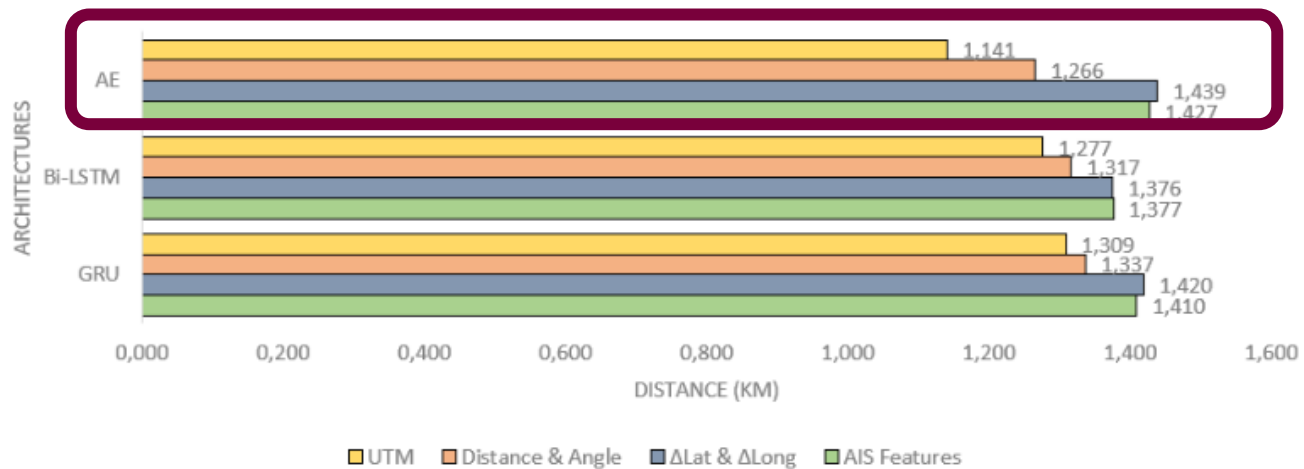
**Vilniaus
universitetas**

Ginamų teiginių rezultatai

Rezultatai (1)

1.

Rekurentinio tinklo architektūros ir celių kiekio įtaka LSTM tinkle

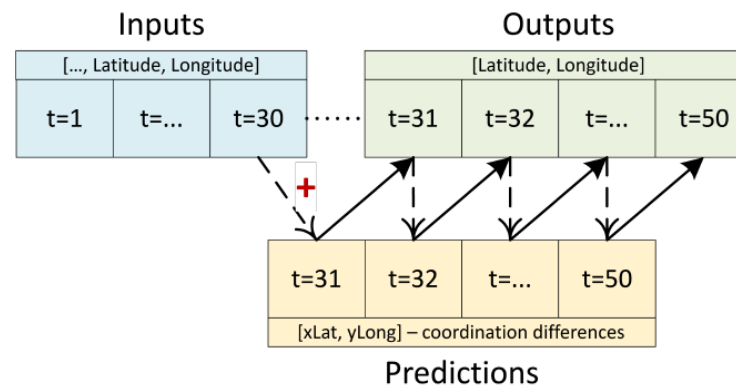
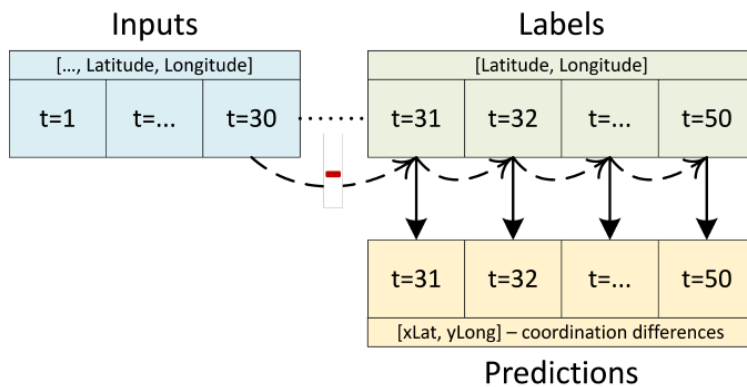
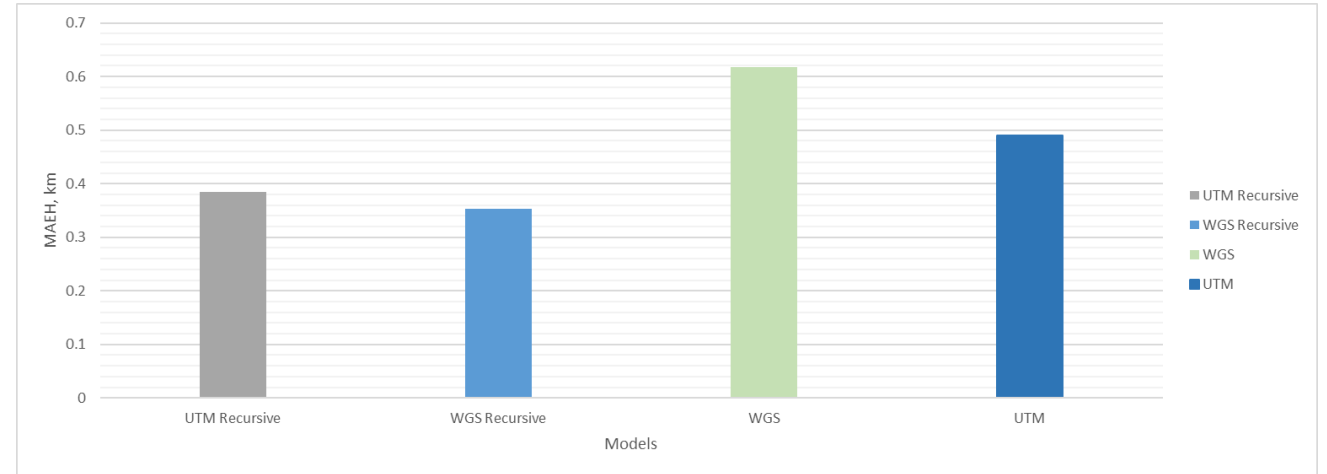


Vilniaus
universitetas

Rezultatai (2)

2.

Rekursinis prognozavimo
reikšmių perskaičiavimas

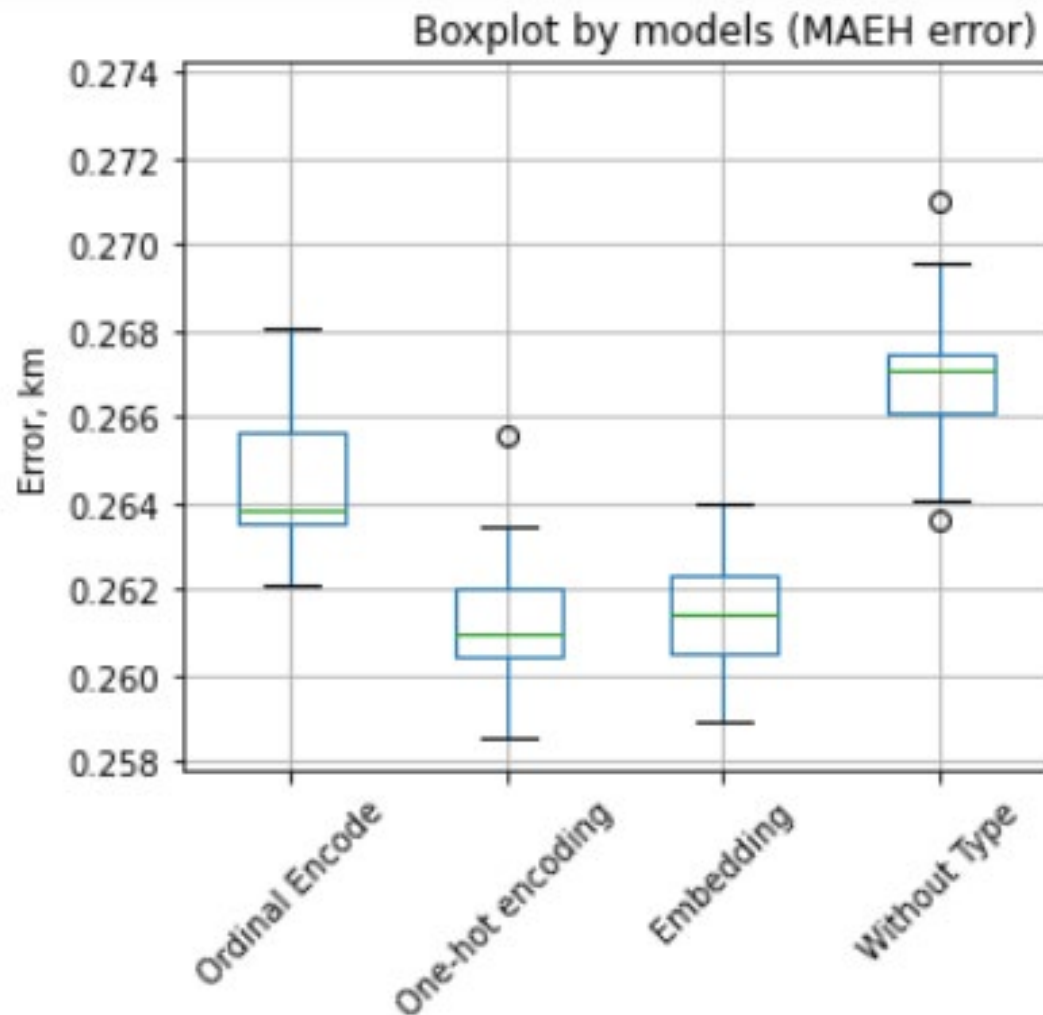


Recursive multi-step trajectory prediction model.

Rezultatai (3)

MAEH klaidos sklaida eksperimentuose (po 15 bandymų)

3. Kategorinių duomenų (laivo tipų) įtraukimas į duomenų rinkinį ir architektūrą

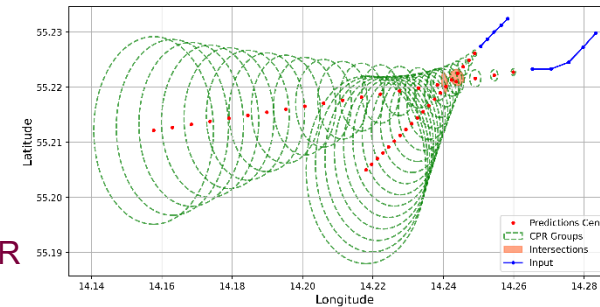


Rezultatai (4)

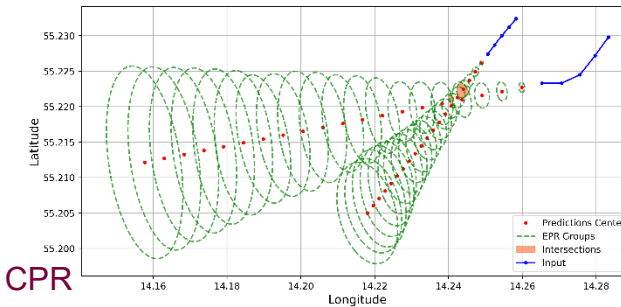
4.

Konforminio prognozavimo
regiono tinkamumas vertinant
susidūrimo tikimybes

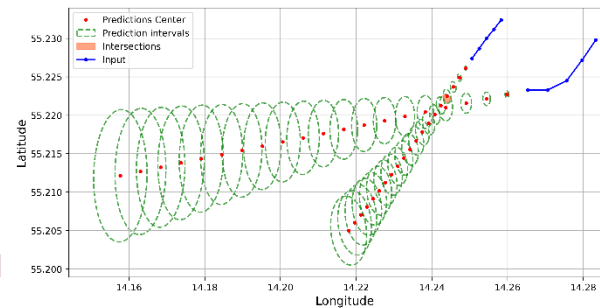
EPR



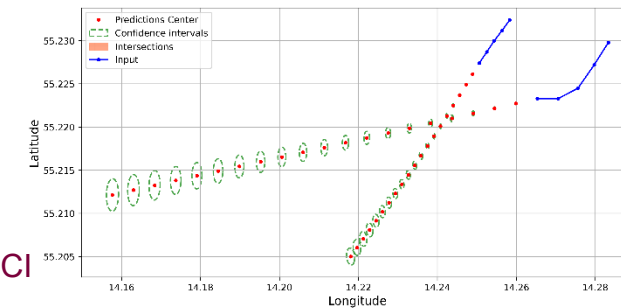
CPR



PI



CI



Comparison of EPR and CPR areas over time steps

Time Step	EPR				CPR			
	Area A, m ²	Area B, m ²	$A \cap B$, m ²	Prob (%)	Area A, m ²	Area B, m ²	$A \cap B$, m ²	Prob (%)
1	11,425.5	32,978.4	0.0	0.00	20,812.7	20,808.5	0.0	0.00
2	35,960.2	117,517.0	0.0	0.00	64,937.4	64,927.8	0.0	0.00
3	72,216.4	220,926.0	0.0	0.00	153,537.9	153,523.8	0.0	0.00
4	132,867.0	368,967.0	96,986.7	23.96	302,596.1	302,584.9	171,576.5	39.57
5	198,320.0	536,512.0	22,066.5	3.10	524,334.6	524,344.3	89,920.4	9.38
6	274,578.0	762,444.0	0.0	0.00	826,551.7	826,611.9	0.0	0.00

Time Step	Prediction Intervals				Confidence Intervals			
	Area A, m ²	Area B, m ²	$A \cap B$, m ²	Prob (%)	Area A, m ²	Area B, m ²	$A \cap B$, m ²	Prob (%)
1	4,717.5	14,151.0	0.0	0.00	224.6	673.9	0.0	0.00
2	14,568.8	53,199.1	0.0	0.00	693.8	2,533.3	0.0	0.00
3	31,642.8	112,766.8	0.0	0.00	1,506.8	5,369.8	0.0	0.00
4	53,492.5	190,868.5	8,420.9	3.57	2,547.3	9,089.0	0.0	0.00
5	79,293.0	288,027.5	0.0	0.00	3,775.9	13,715.6	0.0	0.00
6	107,950.1	405,907.8	0.0	0.00	5,140.5	19,328.9	0.0	0.00

1. Atlikus literatūros analizę nustatyta, kad gilieji rekurentiniai neuroniniai tinklai yra tinkami kuriant laivų trajektorijų prognozes iš istorinių AIS duomenų.
2. Tiksliausia prognozė su mažiausia MAEH paklaida gaunama naudojant LSTM autoenkoderio architektūrą su 275 celių dydžiu – bandymuose išlaiko stabiliausią paklaidos dispersiją, o LSTM ląstelės dydžio pasirinkimas tiesiogiai susijęs su rezultatais
3. Laivų tipų įtraukimas į LSTM rekurentinio kelių žingsnių neuroninio tinklo mokymo duomenų rinkinį padidina laivų trajektorijų prognozavimo tikslumą vidutiniškai apie 100 metrų.
4. Laivų susidūrimo atvejai randami apskaičiuojant prognozių sankirtų tikimybes, kai laivo trajektorijų prognozių ribos sudaromos prognozių, pasiklivimo intervalais bei elipsoidiniais ir konforminiais regionais.
5. Modelių prognozių ribos verifikuotos realiame Baltijos jūros incidente, kurio metu nustatyta, kad su konforminiu prognozavimo regionu avariją buvo galima identifikuoti didžiausia 39,6 % tikimybe.





**Vilniaus
universitetas**

Ačiū už dėmesį.

Robertas Jurkus

Doktorantas

robertas.jurkus@mif.stud.vu.lt