



**Vilniaus
universitetas**

**MATEMATIKOS IR INFORMATIKOS FAKULTETAS
DUOMENŲ MOKSLO IR SKAITMENINIŲ
TECHNOLOGIJŲ INSTITUTAS**

**SAUGUMO JŪROJE DIDINIMAS: SUSIDŪRIMO TIKIMYBIŲ SU
PROGNOZAVIMO RIBOMIS VERTINIMAS NAUDOJANT GILIOJO
MOKYMOSI MODELIUS**

Doktorantas: Robertas Jurkus

Vadovas: prof. dr. Povilas Treigys

Vilnius - Klaipėda,
2024



TURINYS

1. Aktualumas, dalykinė sritis ir problematika
2. Preliminari tema ir ginami teiginiai
3. Atlikti tyrimai ir publikacijos
4. Duomenų rinkinio analizė ir modelio kūrimas
5. Pasiklovimo intervalų taikymas realiose trajektorijose
6. Rezultatai ir apibendrinimas

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:



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.

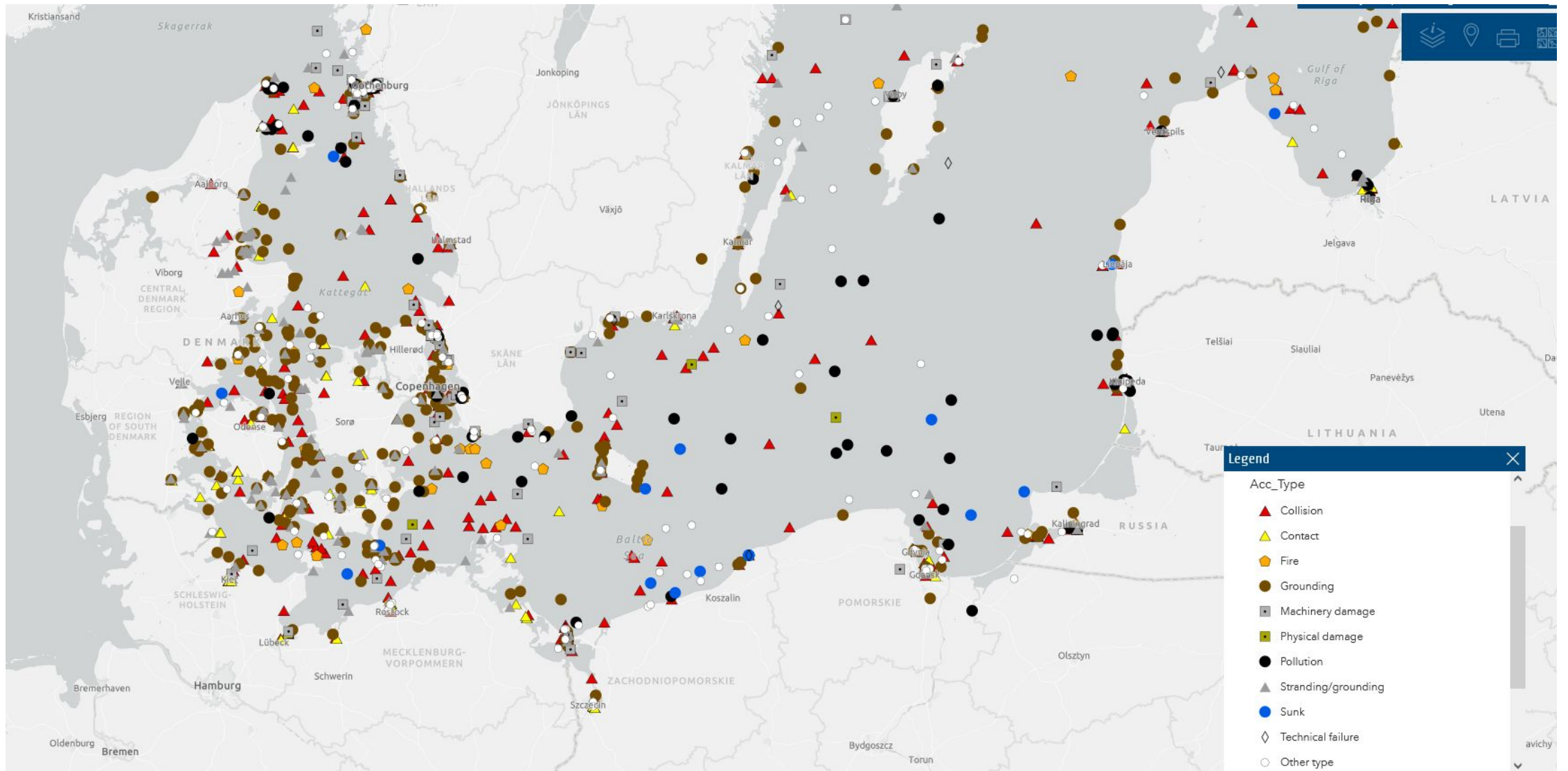
Tikslas

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



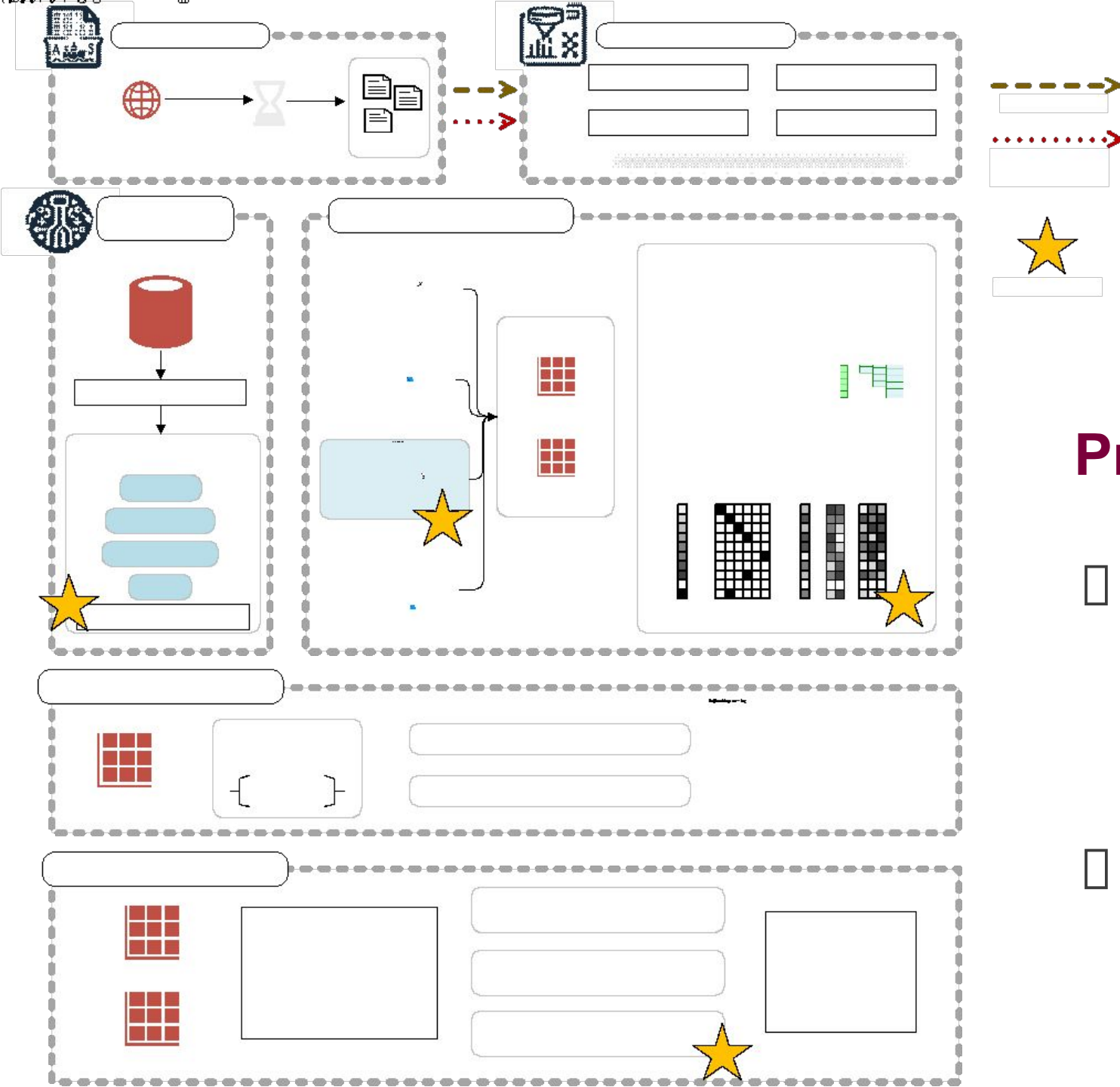
Dr. **Andrius Daranda** (2021 m.): Mašininis mokymusi grindžiamas laivybos eismo dalyvių elgsenos prognozavimas bei nestandartinių laivybos srauto situacijų atradimas.

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% pasikliovimo lygį, kai duomenys yra daugiamaciai ir juose nėra normaliojo skirstinio.

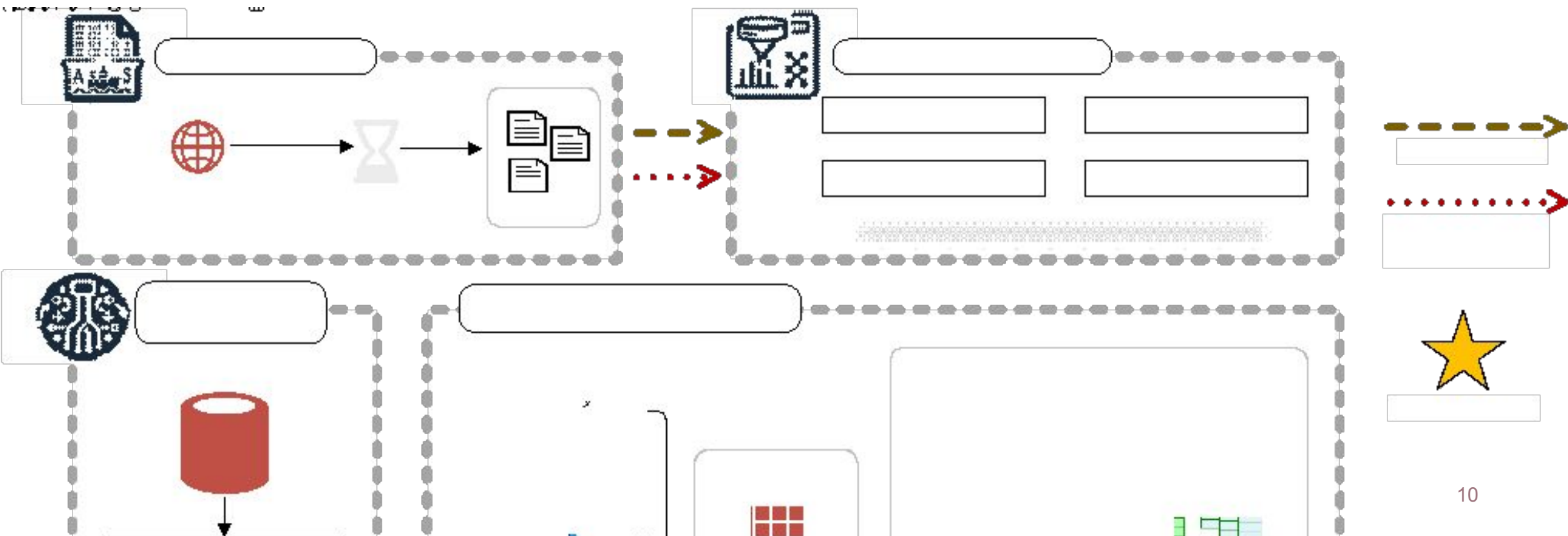


Preliminari disertacijos tema

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

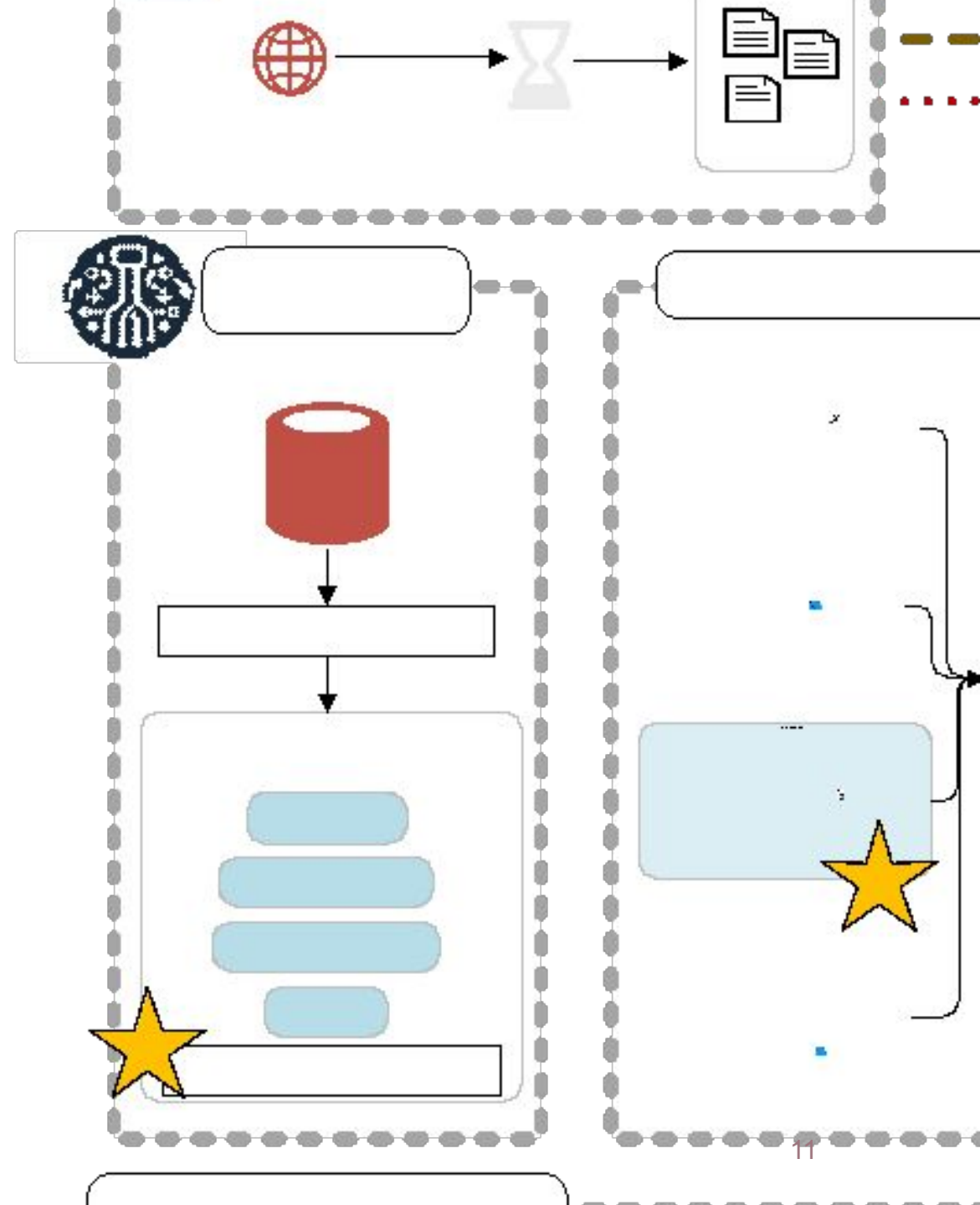
Tyrimų eigos diagrama

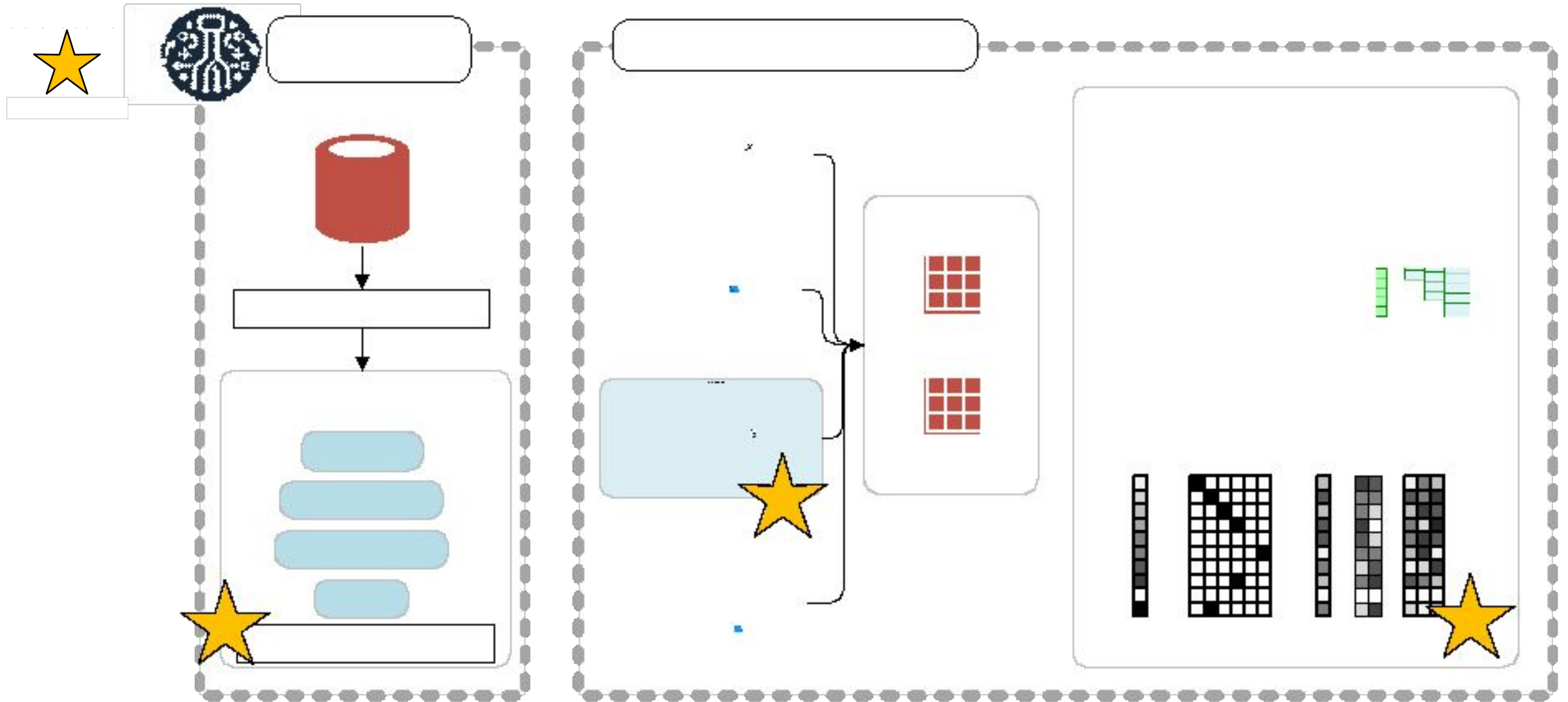
Gilusis mokymasis yra mašininio mokymosi, kuris pats savaime yra dirbtinio intelekto (DI) poaibis, dalis. Daugiausia dėmesio skiriama daugiasluoksnių neuroninių tinklų naudojimui, kad būtų galima modeliuoti ir suprasti sudėtingus duomenų dėsningumus.



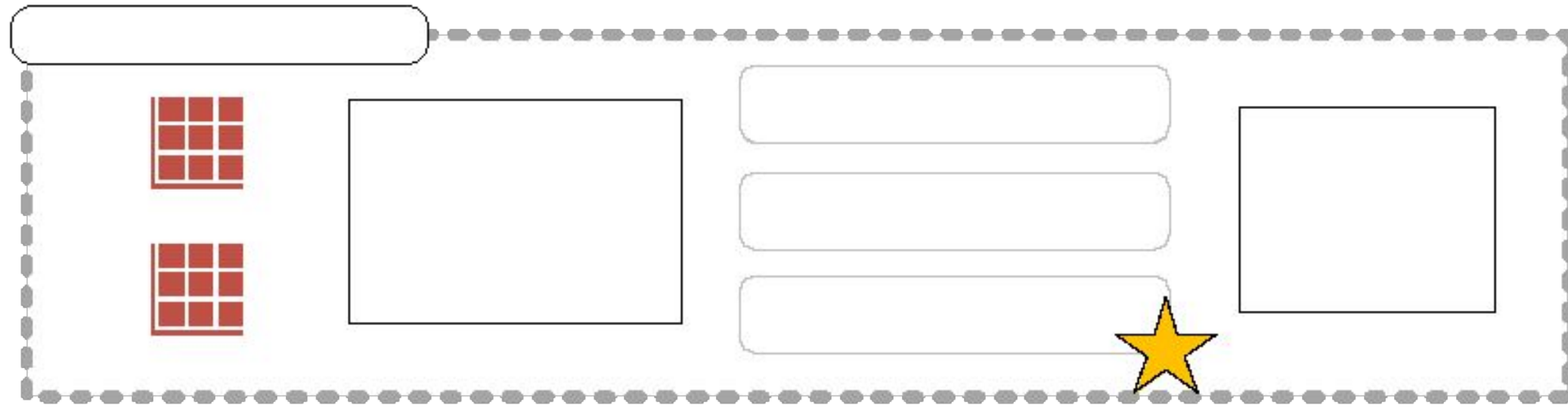
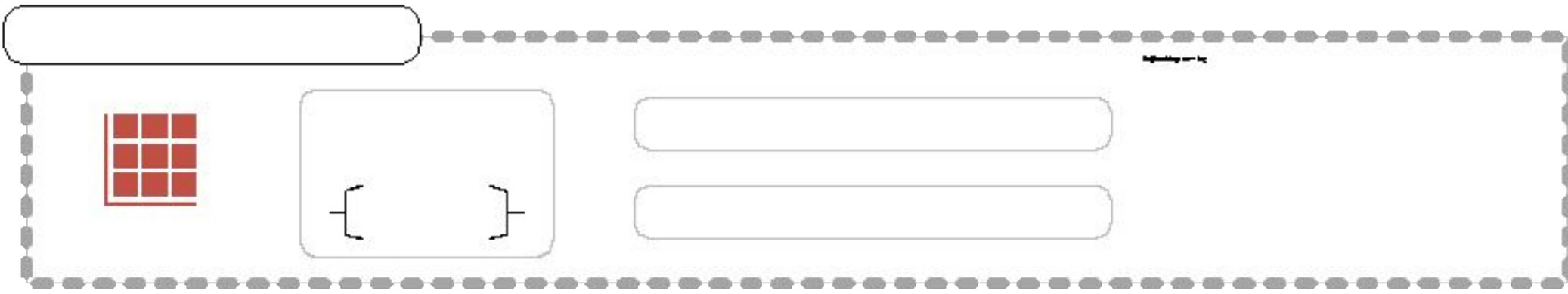
1. **Robertas Jurkus**, Povilas Treigys, and Julius Venskus, “*Investigation of Recurrent Neural Network Architectures for Prediction of Vessel Trajectory*”. In: Information and Software Technologies, Springer, Chapter No: 16, vol 1486. Chapter DOI:10.1007/978-3-030-88304-1_16, KTU, Lithuania, 2021.

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:





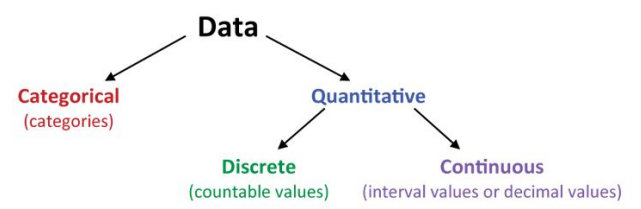
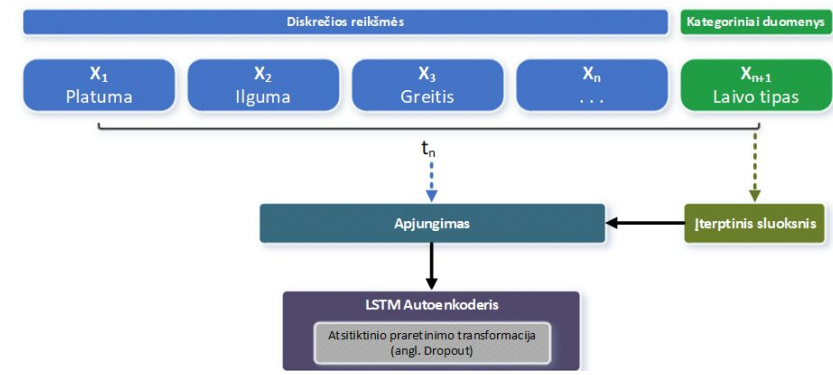
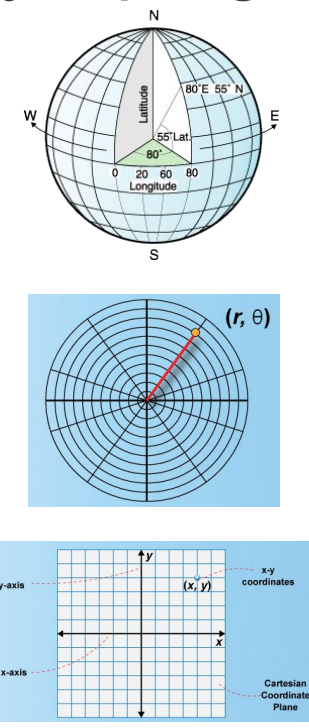
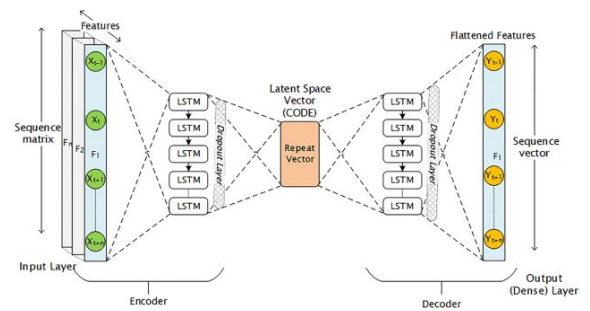
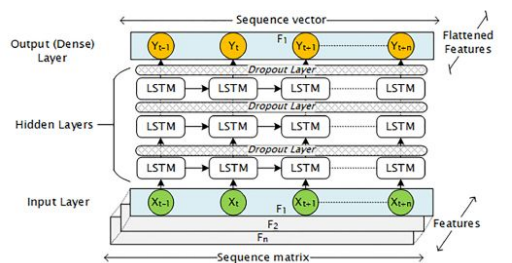
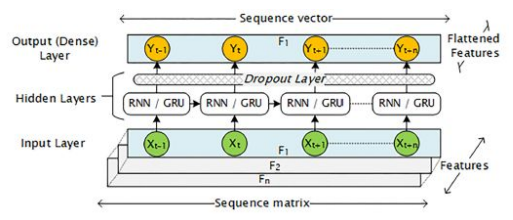
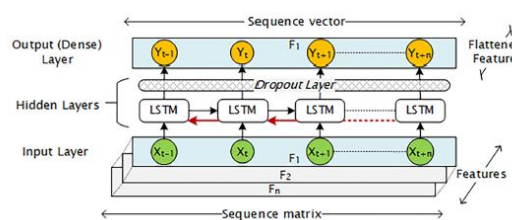
- Robertas Jurkus**, Julius Venskus, and Povilas Treigys, “*Extended Research on Categorical Data Encoding Techniques for Recursive Multi-Step Prediction of Vessel Trajectory*”. In: Contributions to Statistics, A Springer book series, ISSN: 2628-8966, ITISE 2023/2024.



4. **Robertas Jurkus**, Julius Venskus, 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 koordinacių sistemos; kategoriniai duomenys, prognozių ribos/zonos.



2020/2021
1 Kursas



2021/2022
2 Kursas



2022/2023
3 Kursas

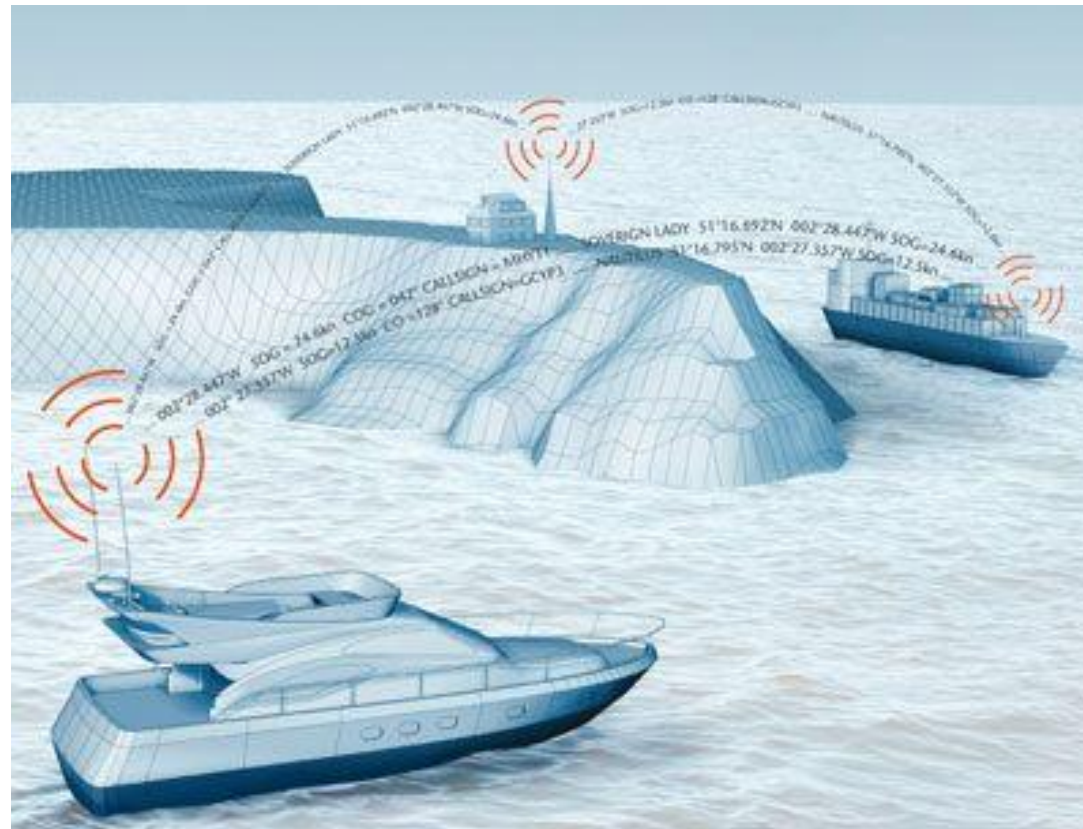


2023/2024
4 Kursas

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);

□ Tyrimo laivo tipas: 6 mėn. duomenimis (2021–06 iki 2021–12).



Tiriamasis regionas Baltijos jūroje prie Bornholmo salos

Laivų tipai duomenų imtyje

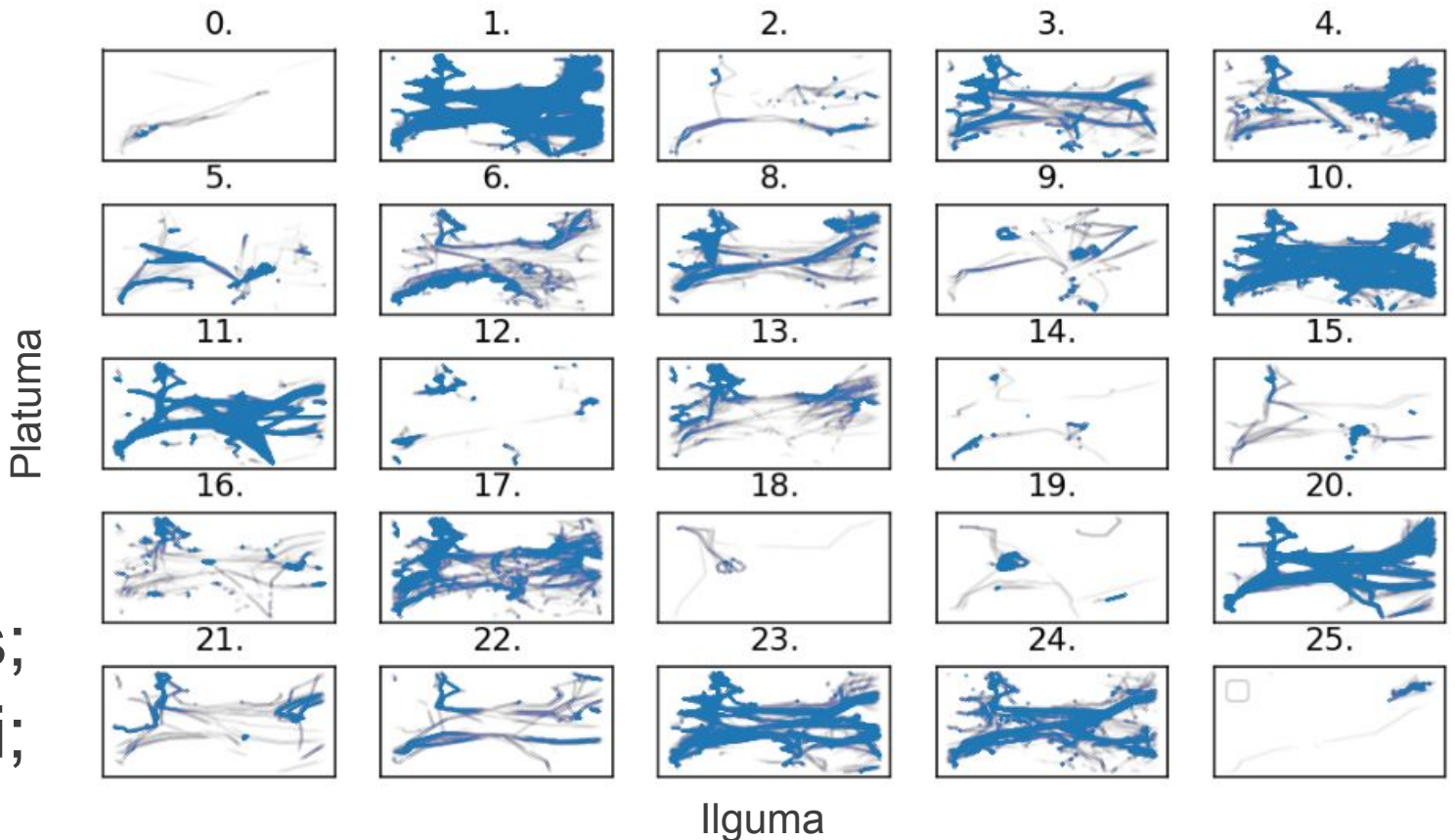
□ Intensyviausias:

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

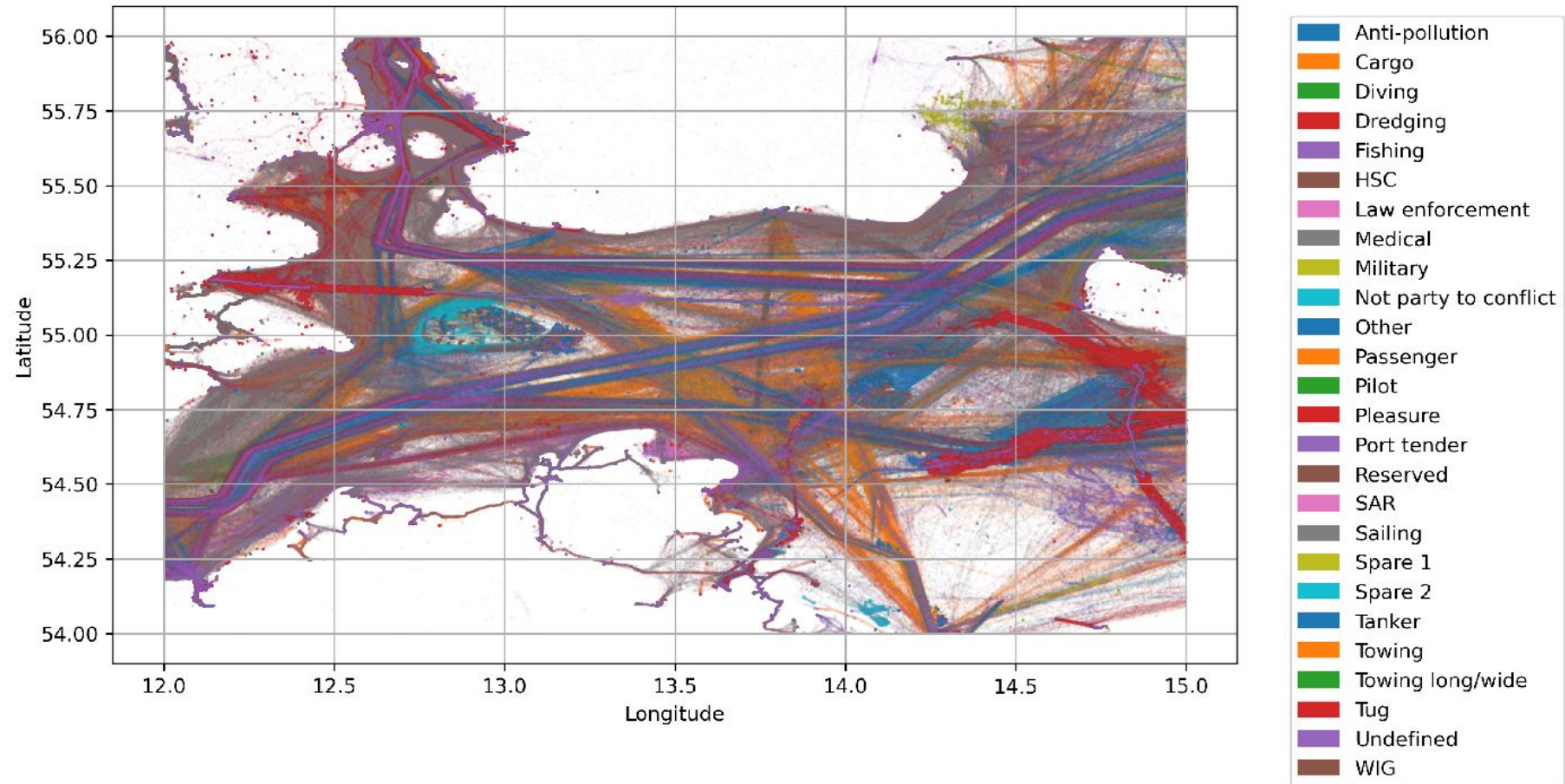
□ Specifinis:

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

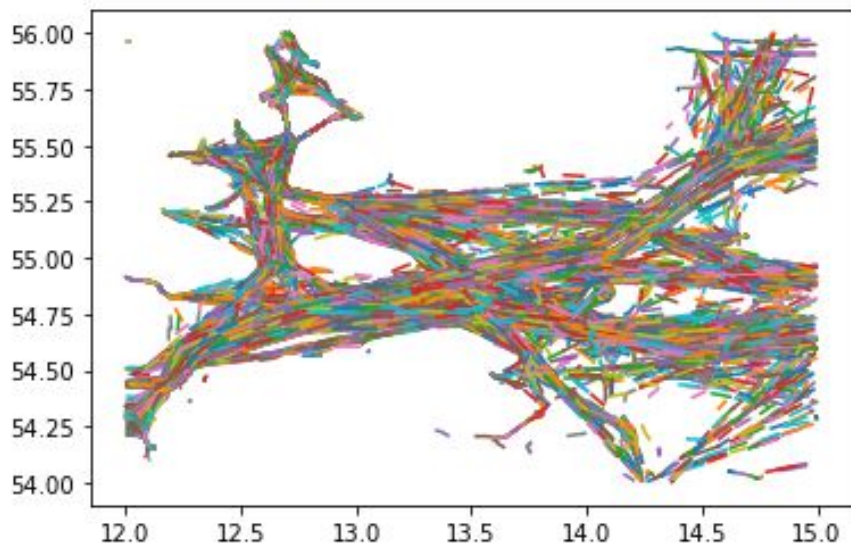
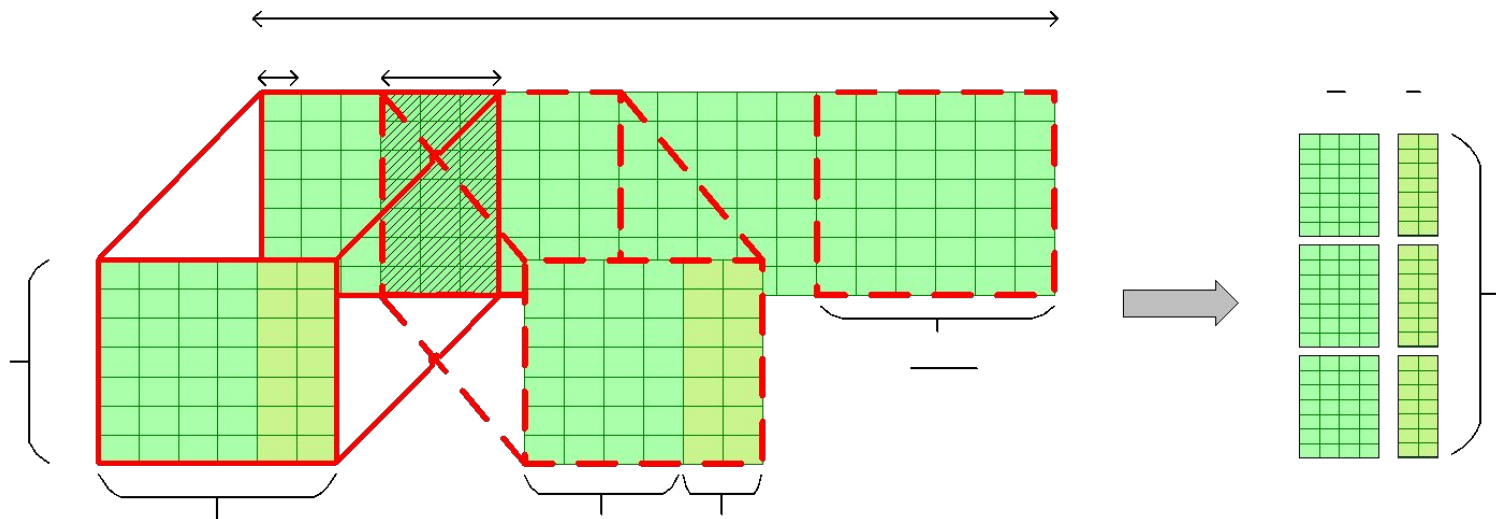
Eismo intensyvumas pagal laivo tipą



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.

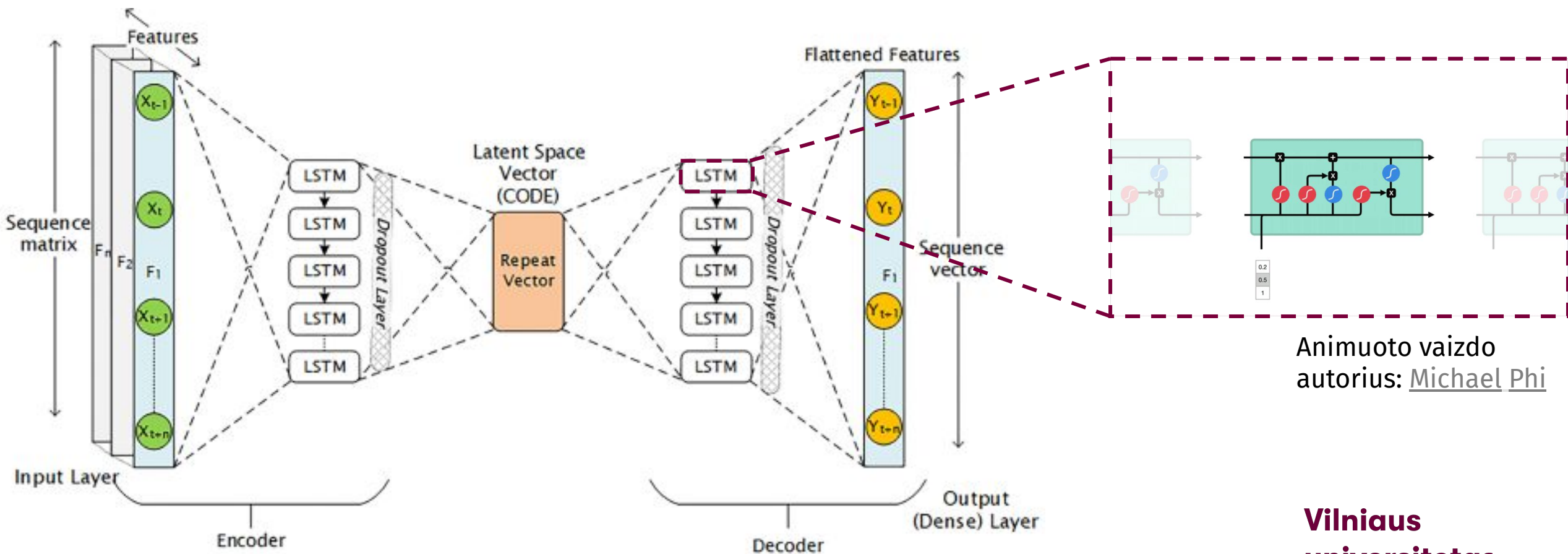
SHIP TYPE	TRACK REACH	TIME TO STOP
Fine and fast ship	5 Ship's length	1.5 to 2 minutes
Cargo ships loaded, displacement 28000 tonnes L 145 m	11-13 ship's length	9-11 minutes
Loaded bulk carrier and tanker	15-20 ship's length	20-25 minutes

- 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.

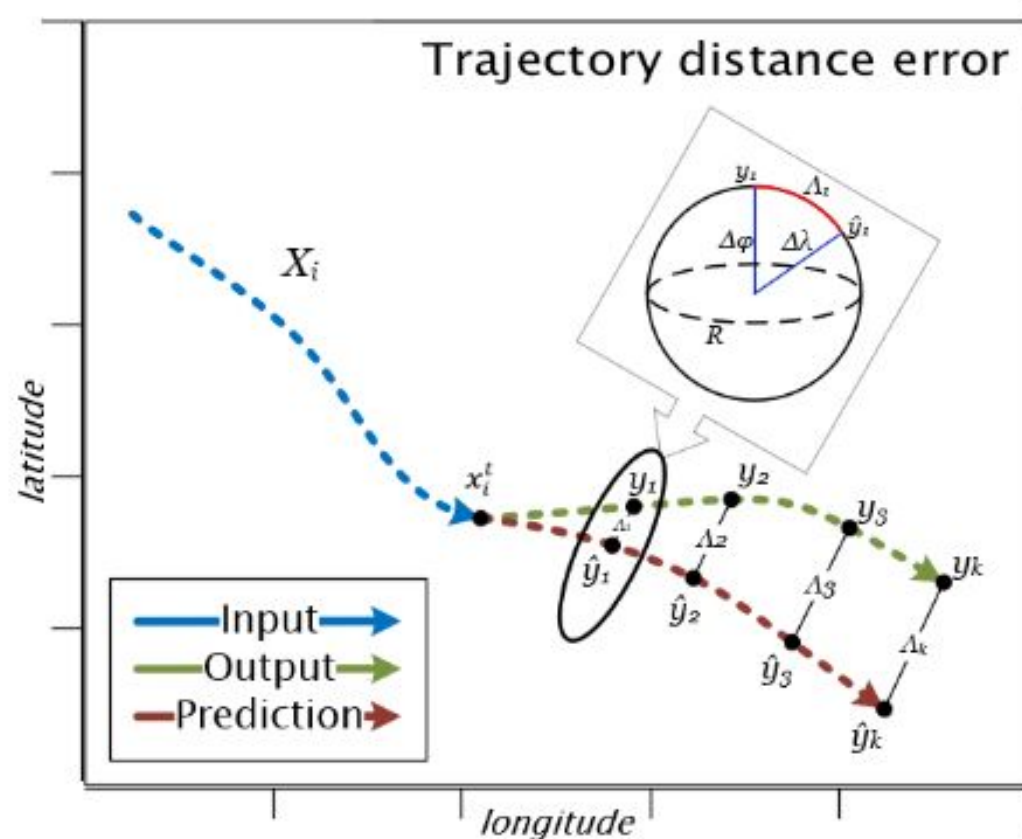


Animuoto vaizdo
autorius: [Michael Phi](#)

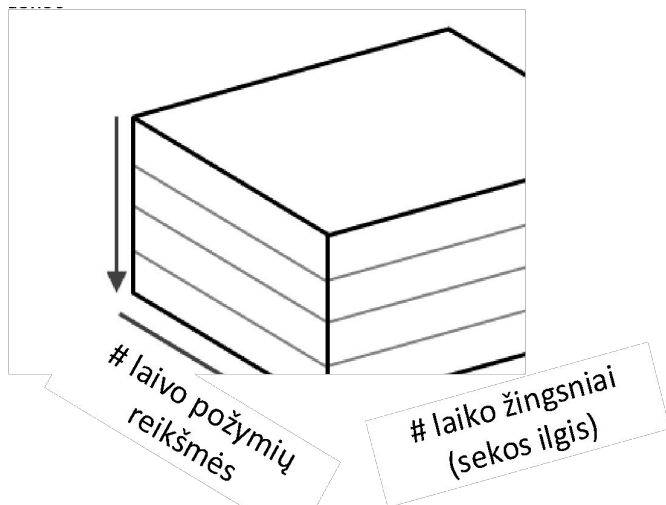
Modelių vertinimas

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

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

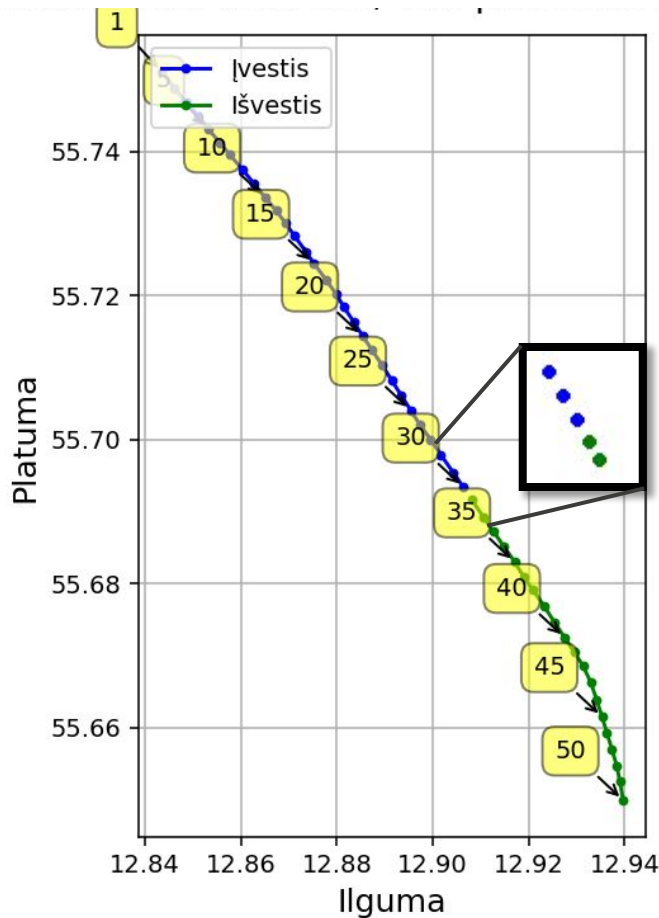


Duomenų apibendrinimas



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

□ Ivestis:

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

□ Išvestis:

- Ilguma
- Platuma

Požymių permutacija

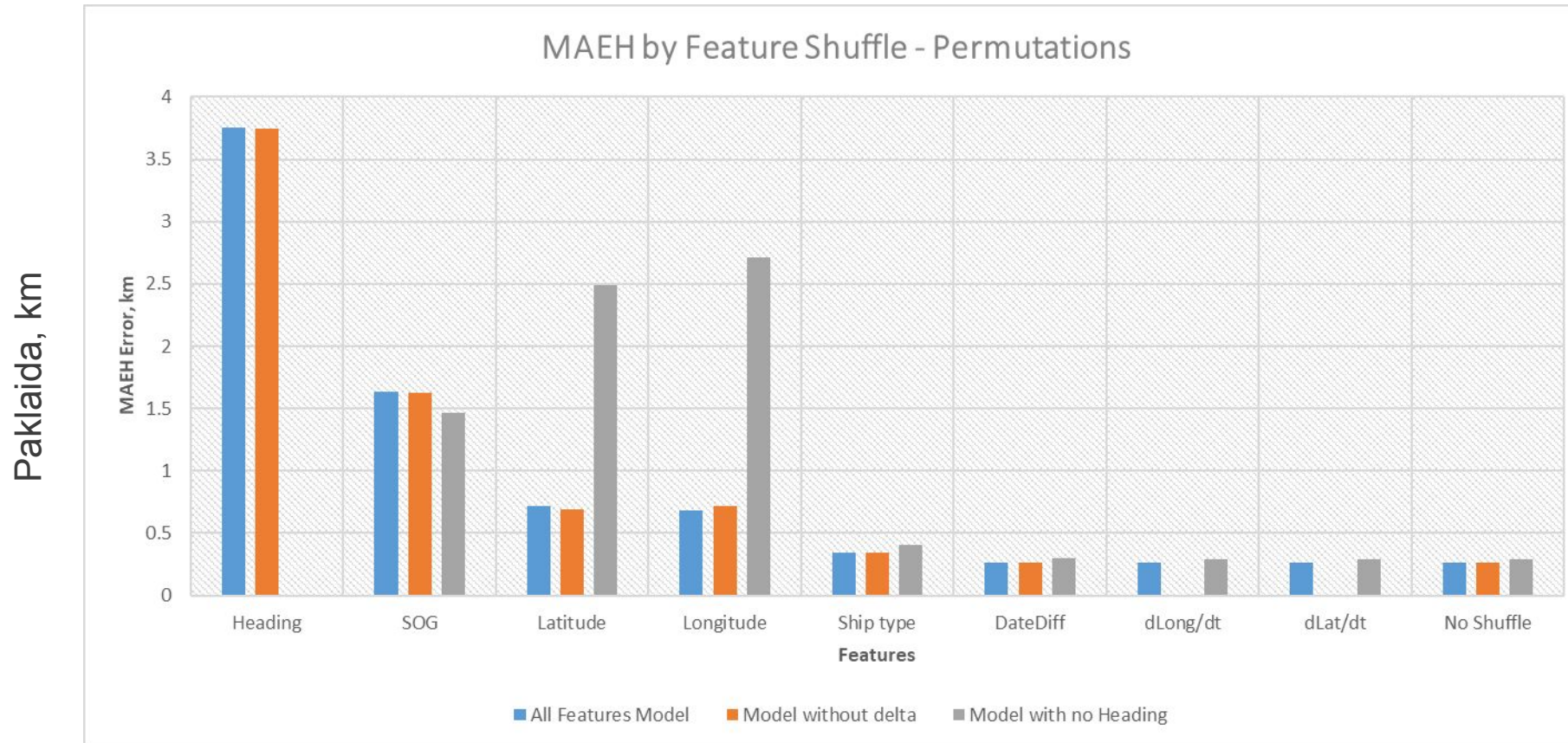
Metodas (angl. permutation feature importance), naudojamas požymių svarbai prognozuojamajame modelyje įvertinti. Juo įvertinamas kiekvienos savybės poveikis modelio veikimui, permutuojant (permaišant) savybės reikšmes ir matuojant, kaip sumažėja veikimas.

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

Lat	Long	SOG	Heading	DateDiff	dLat/dt	dLong/dt
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
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
0.11	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)

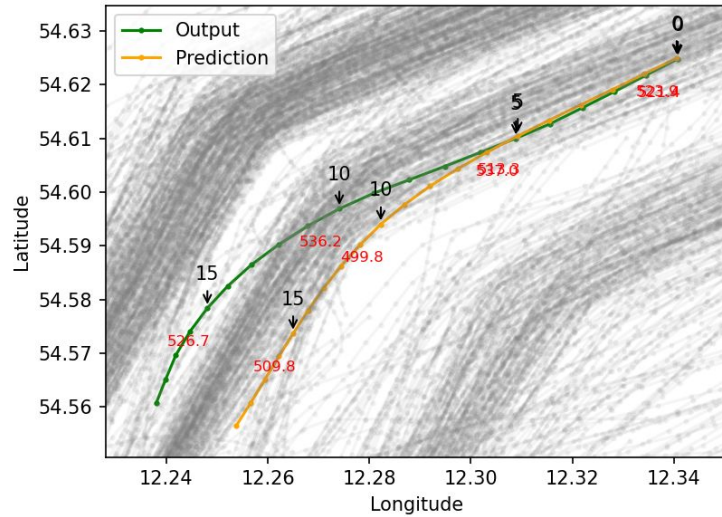


Duomenų rinkinio savybė

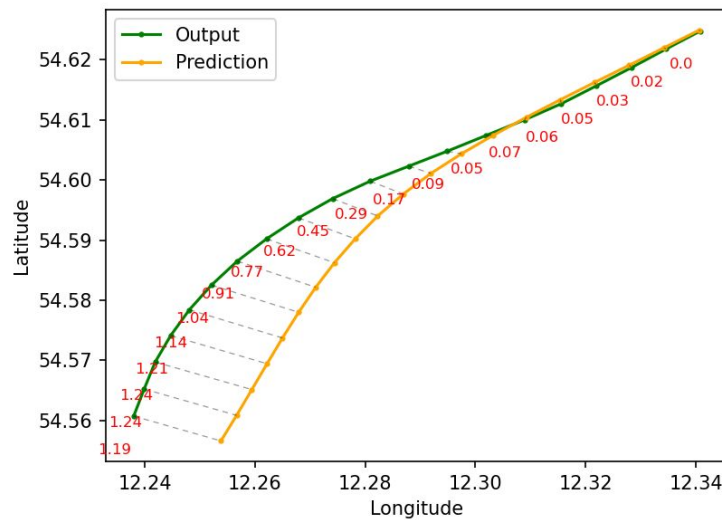
$$X_{\Delta t} = t_s - t_{s-1}$$
$$X_{\delta Lon} = \frac{X_{Lon,s} - X_{Lon,s-1}}{X_{\Delta t}}$$
$$X_{\delta Lat} = \frac{X_{Lat,s} - X_{Lat,s-1}}{X_{\Delta t}}$$

Prognozės pavyzdys

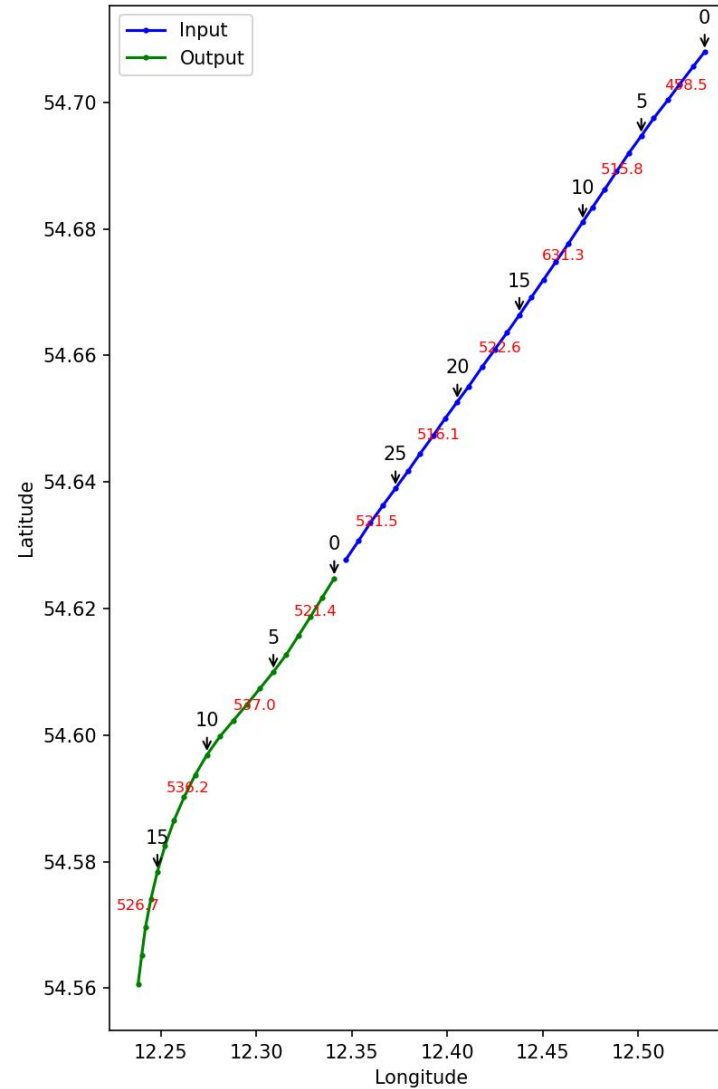
Actual and Predicted movement in the traffic (m)



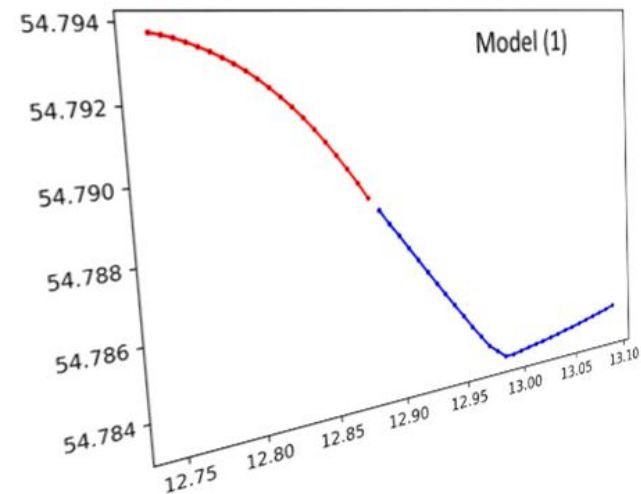
MAEH = 0.588 km



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



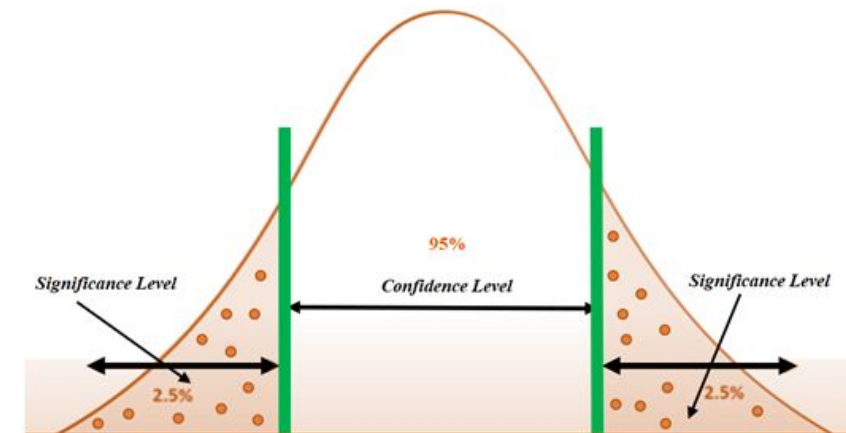
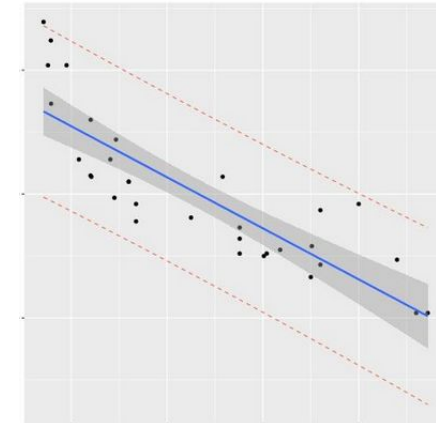
■ Prognozės ribų skaičiavimas modeliuose



Prognozės ribų skaičiavimas modeliuose

□ Nagrinėti metodai (taikant pasiklovimo lygį su 95%):

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



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

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

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

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 CONVEXHULL(*points*)

hullVertices \leftarrow *points*[*hull.vertices*]

covMatrix \leftarrow COVARIANCE(*hullVertices*)

[*eigValues*, *eigVectors*] \leftarrow EIGEN(*covMatrix*)

confidenceLevel \leftarrow 0.95

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

angleArray \leftarrow array of angles from 0 to 2π

ellipsoidPoints \leftarrow 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

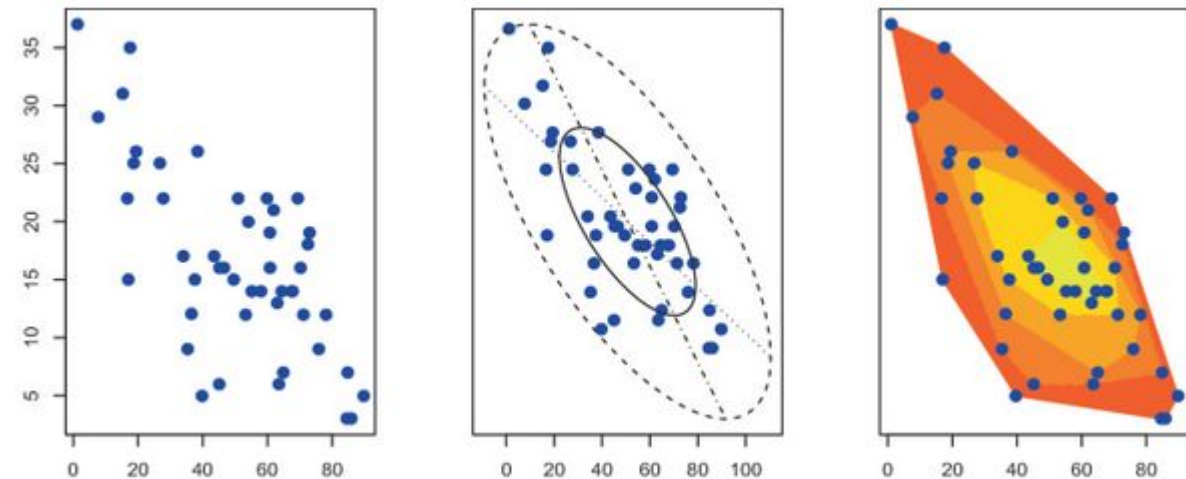
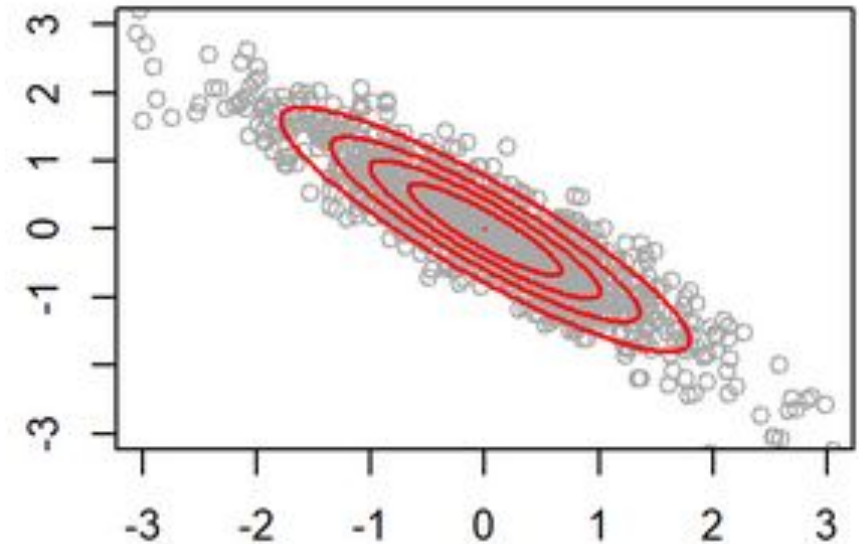
rotatedPoints \leftarrow *eigVectors* * *ellipsoidPoints*

translatedPoints \leftarrow *rotatedPoints* + *center*

epr \leftarrow 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:

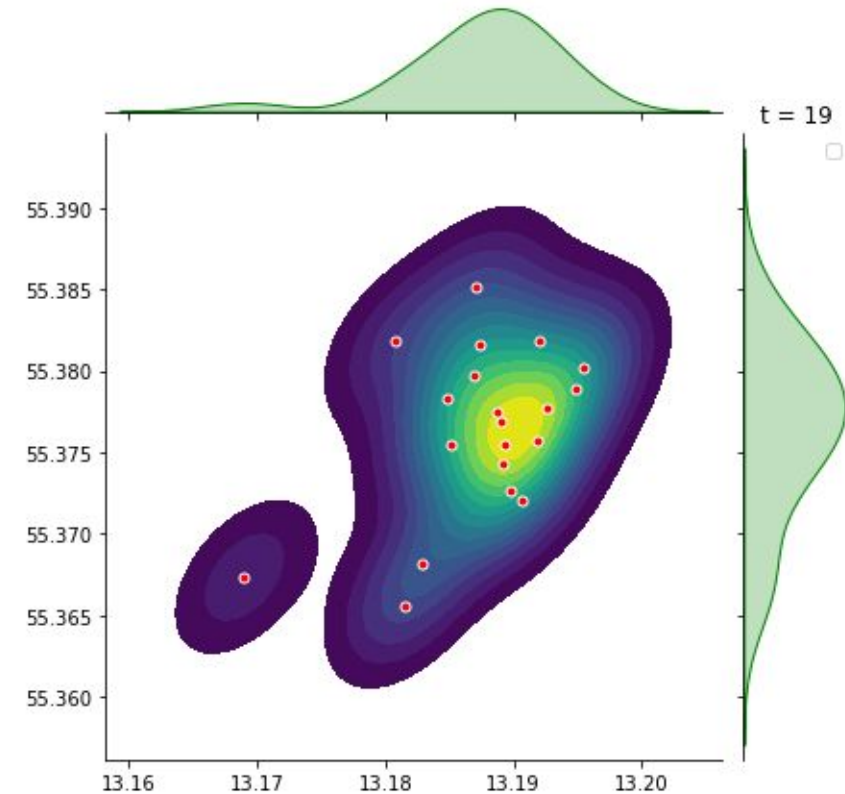
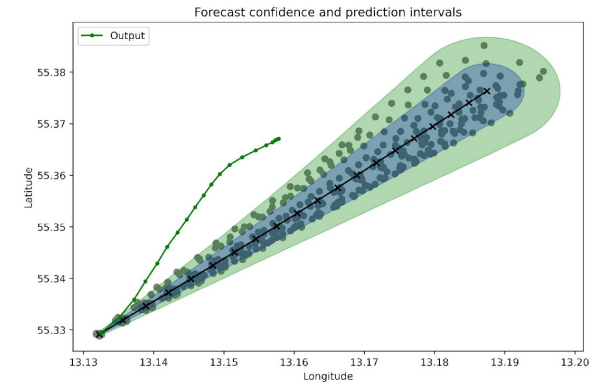
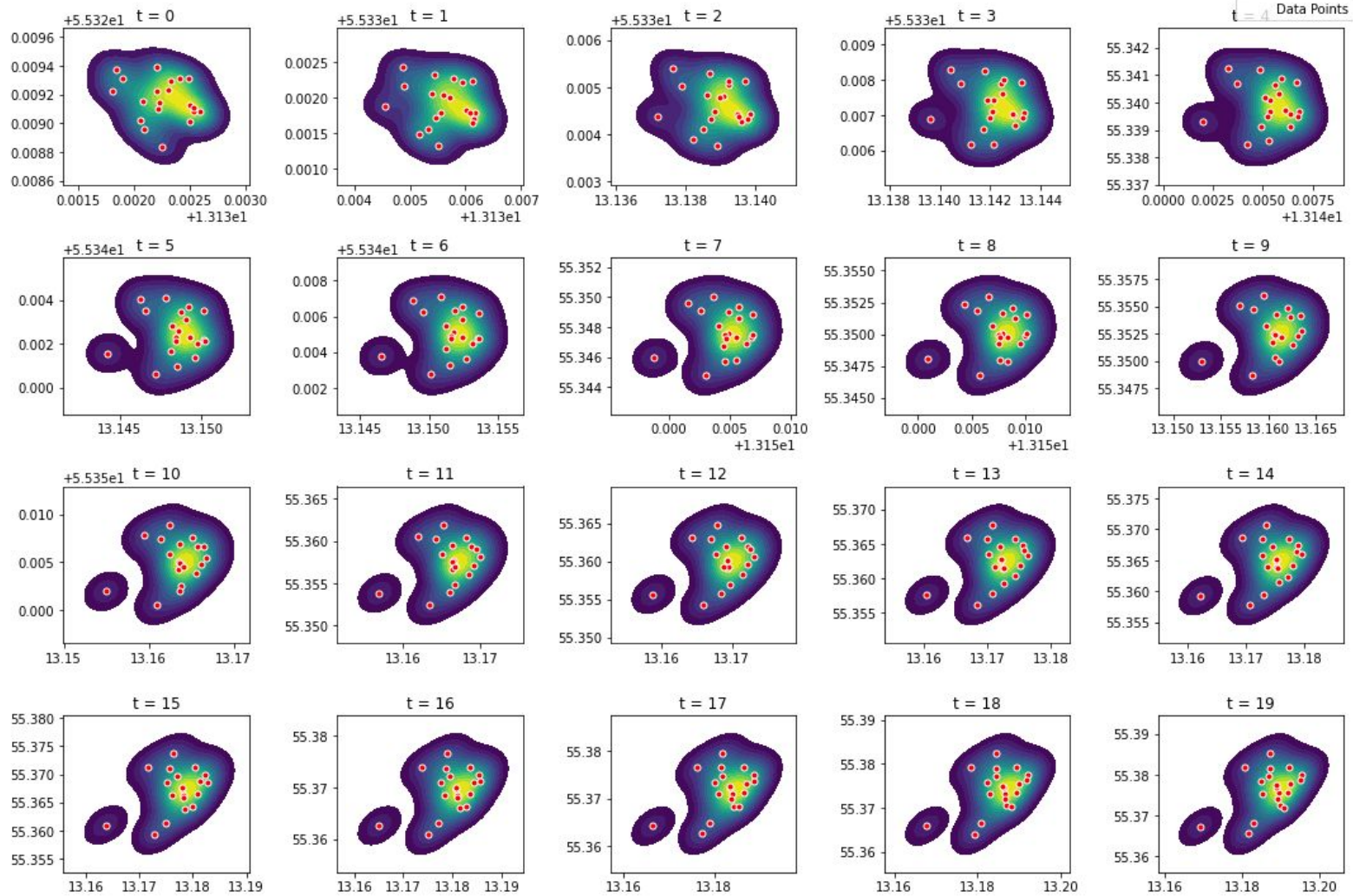
$$\bar{\mathcal{R}}_t = \frac{1}{n} \sum_{i=1}^n \mathcal{R}_t^{(i)},$$

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 susidūrimė

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

Cargo ships collide in Baltic Sea

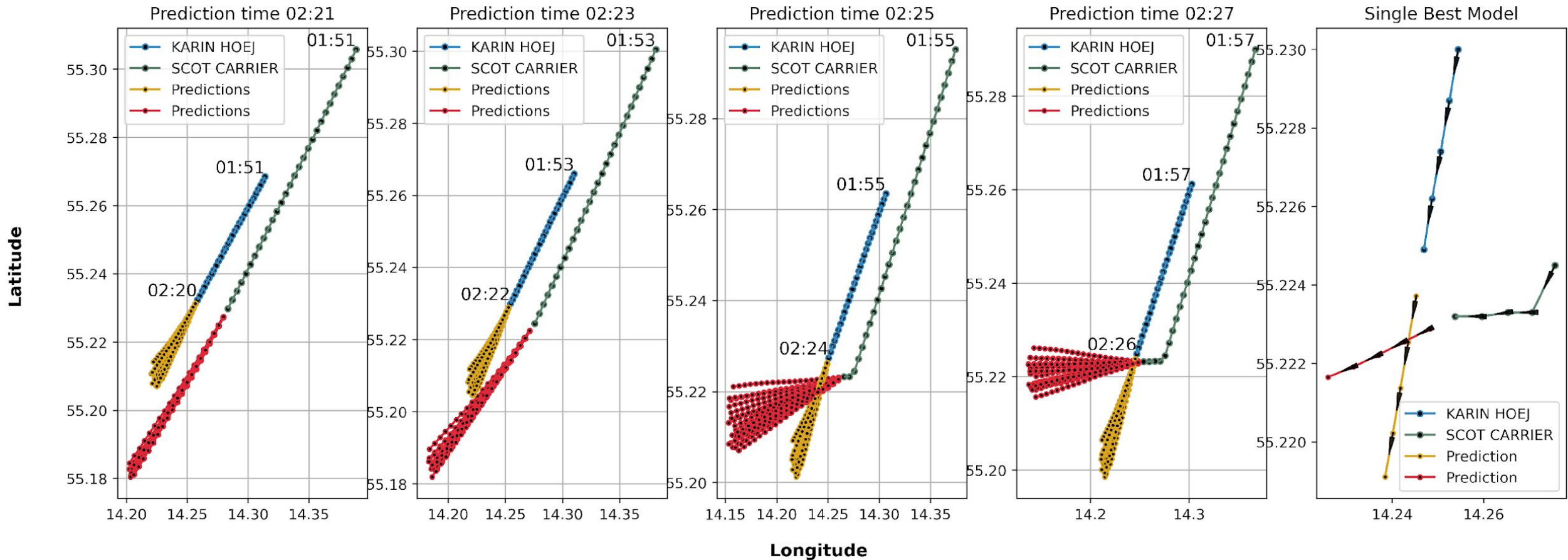
Two cargo ships collided on Monday during foggy conditions in the Baltic Sea between the Danish island of Bornholm and the southern Swedish city of Ystad, and a rescue operation was launched for two missing people, authorities said.



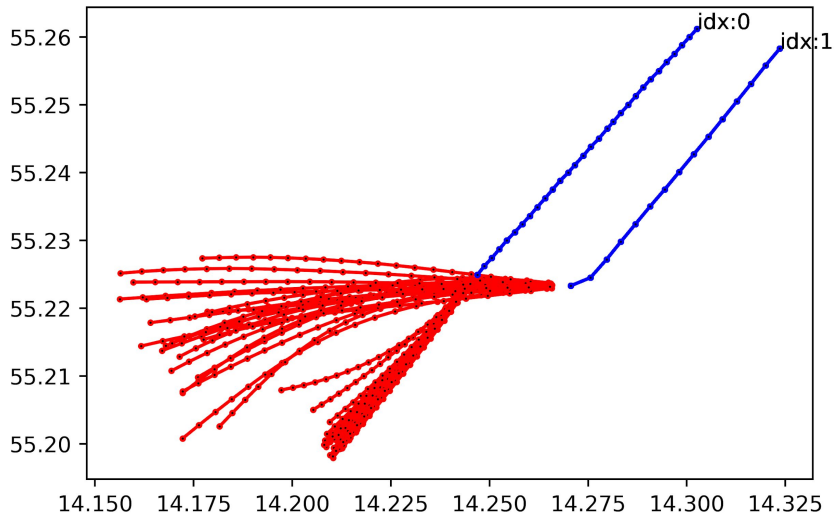
Source: Vessel Finder; Refinitiv Eikon; Reuters

Anomalių aptikimas realiuose susidūrimuose

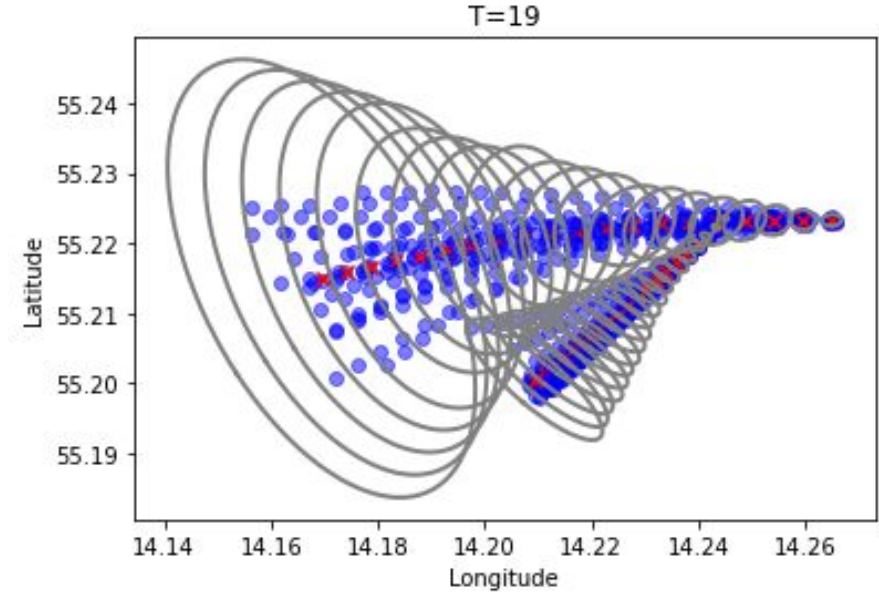
□ Modelių prognozės skirtingu laiko momentu



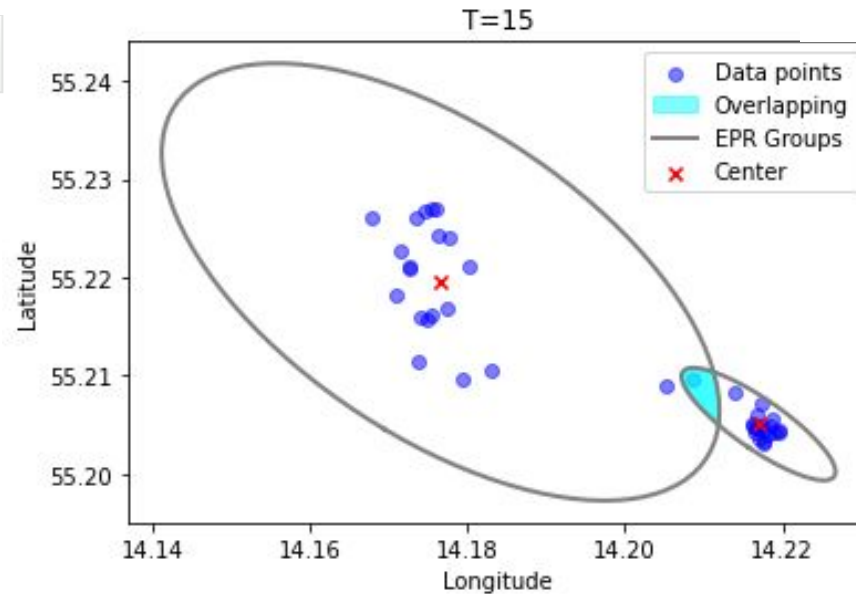
EPR metodo verifikavimas realiam susidūrimė (2)



Modelių prognozė trajektorijoje

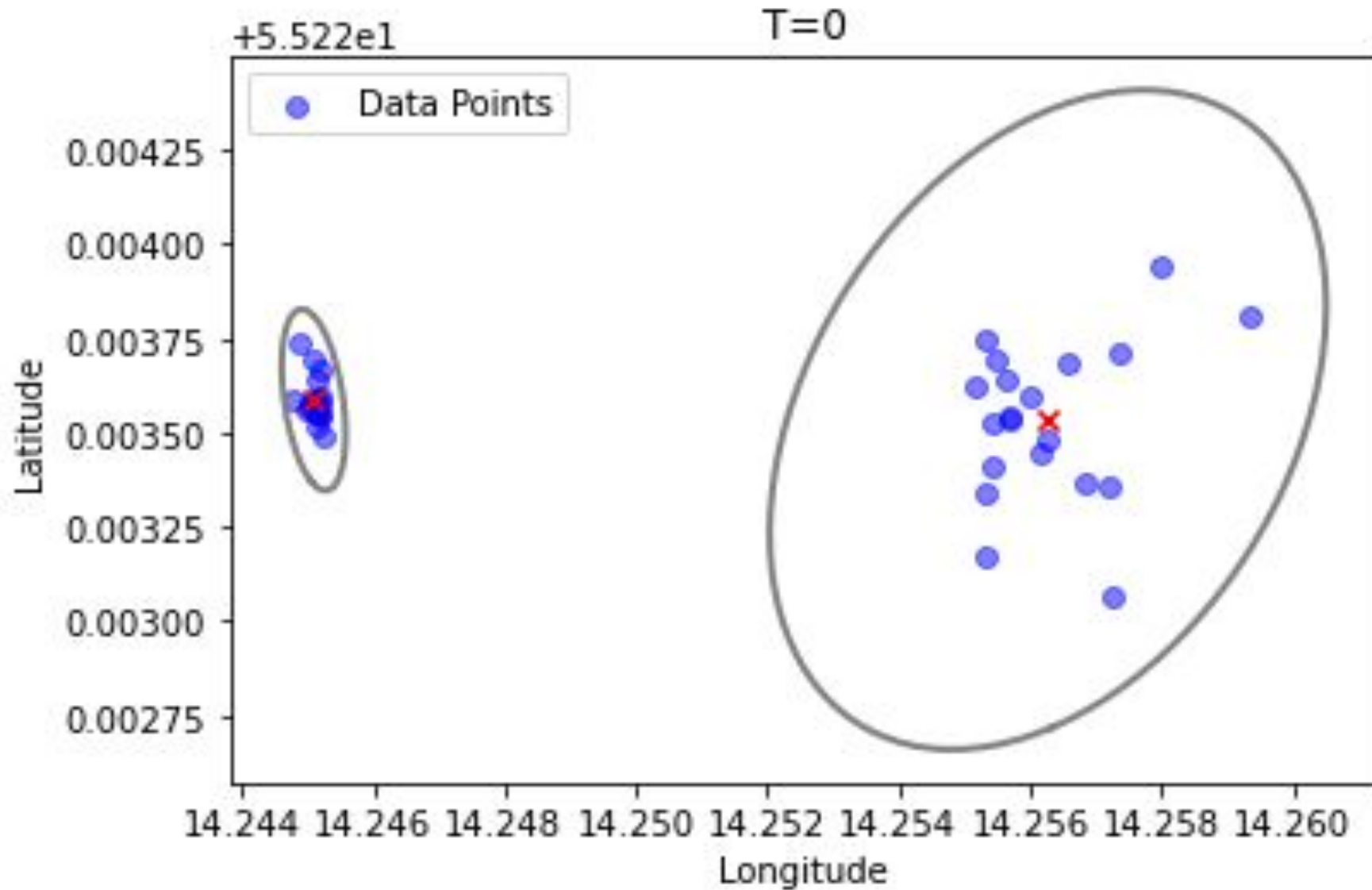


Prognozės intervalai T laiko momentu



Susidūrimo rizika T = 15

CASE #1 EPR – Realus susidūrimas (1)

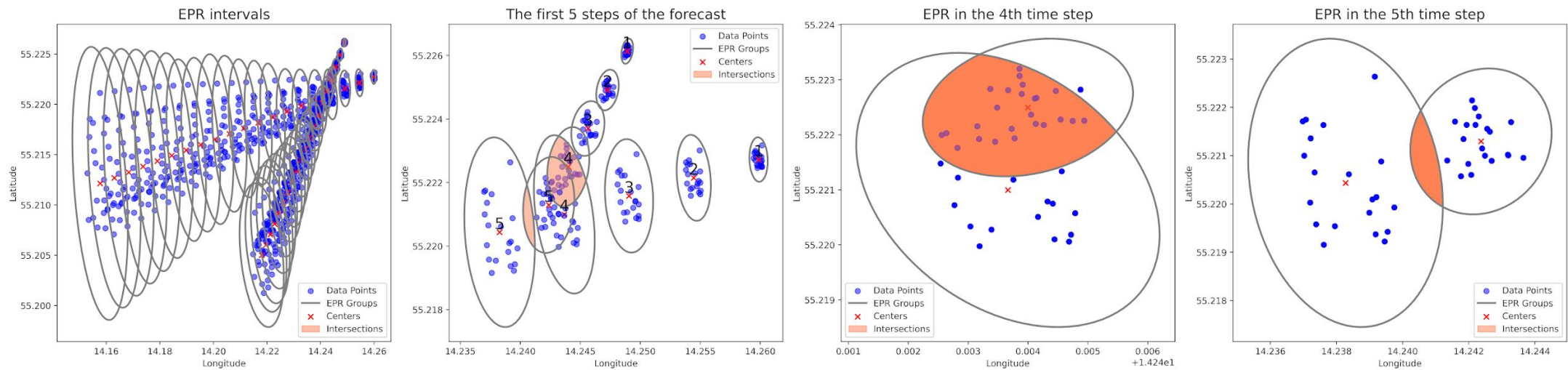


Tikimybinis vertinimas

□ 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.



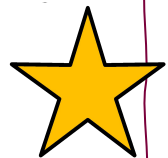
Time Step	EPR Area A, m ²	EPR Area B, m ²	Intersection Area, m ²	intersection_probability (%)
1	11 425.5	32 978.4	0.0	0.00%
2	35 960.2	117 517.0	0.0	0.00%
3	72 216.4	220 926.0	0.0	0.00%
4	132 867.0	368 967.0	96 986.7	23.96%
5	198 320.0	536 512.0	22 066.5	3.10%
6	274 578.0	762 444.0	0.0	0.00%



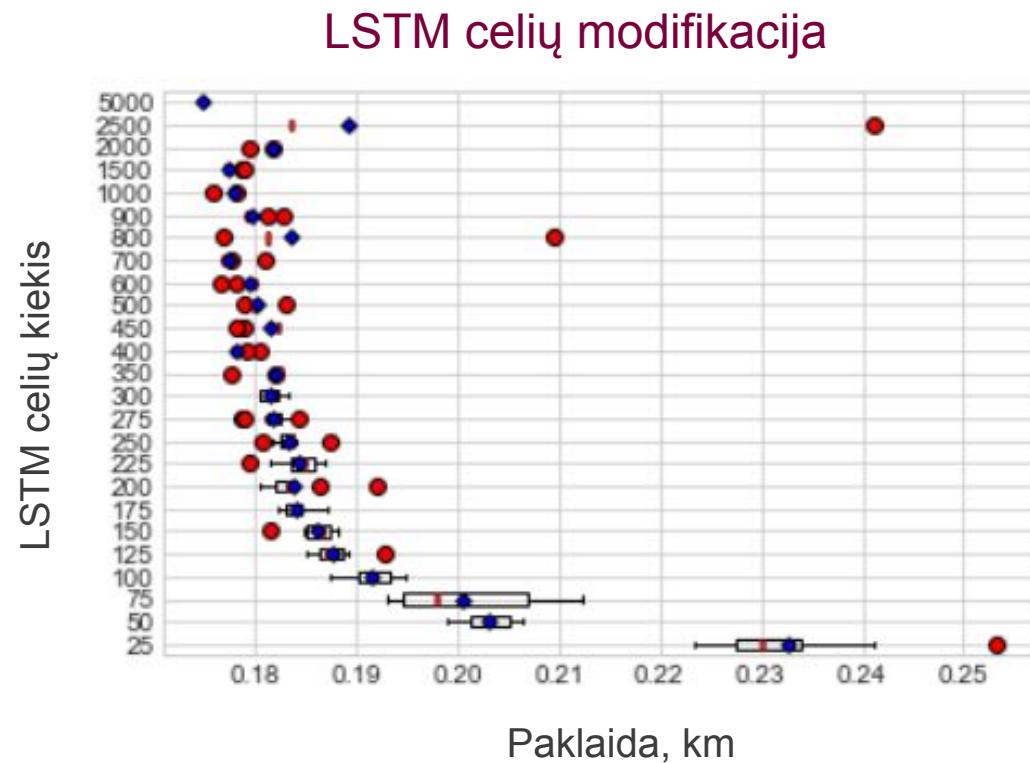
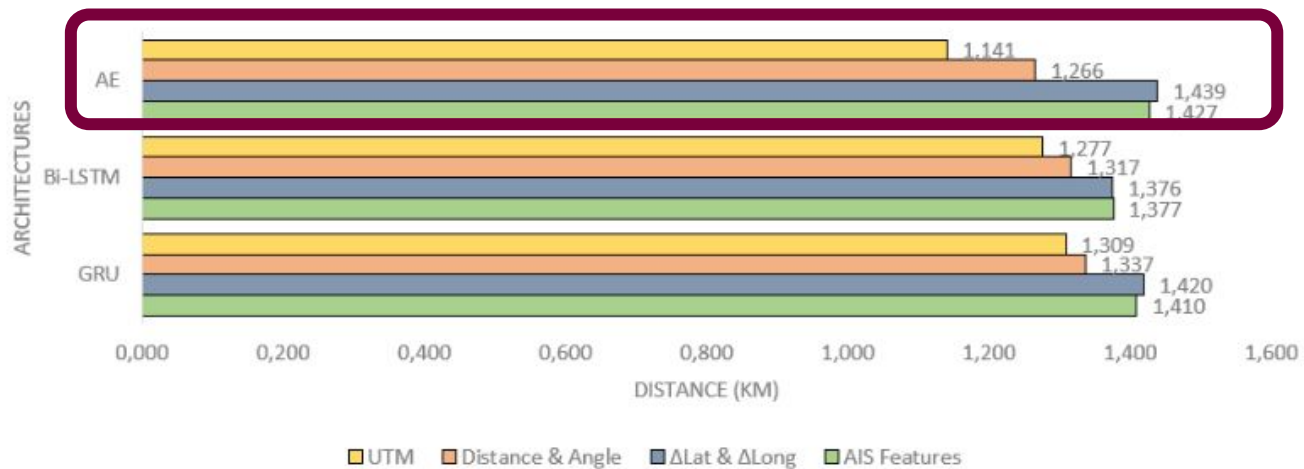
**Vilniaus
universitetas**

Ginamų teiginių rezultatai

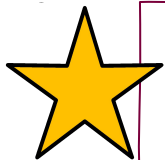
Rezultatai (1)



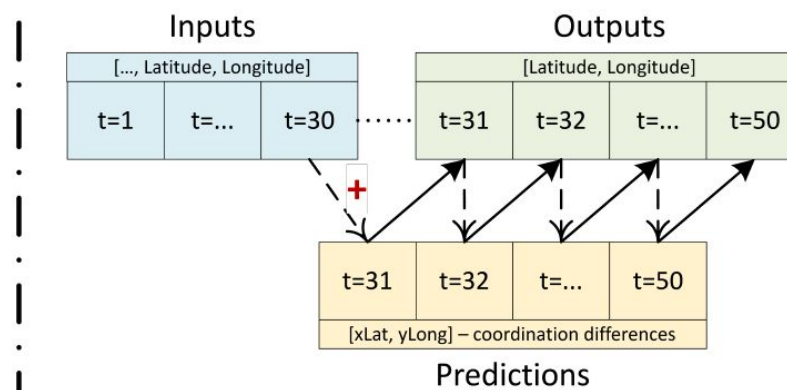
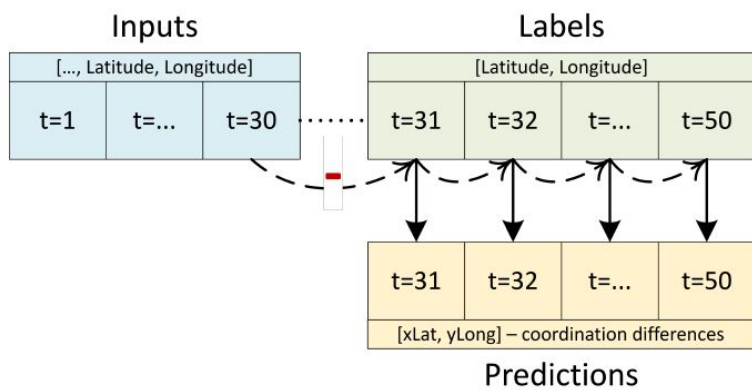
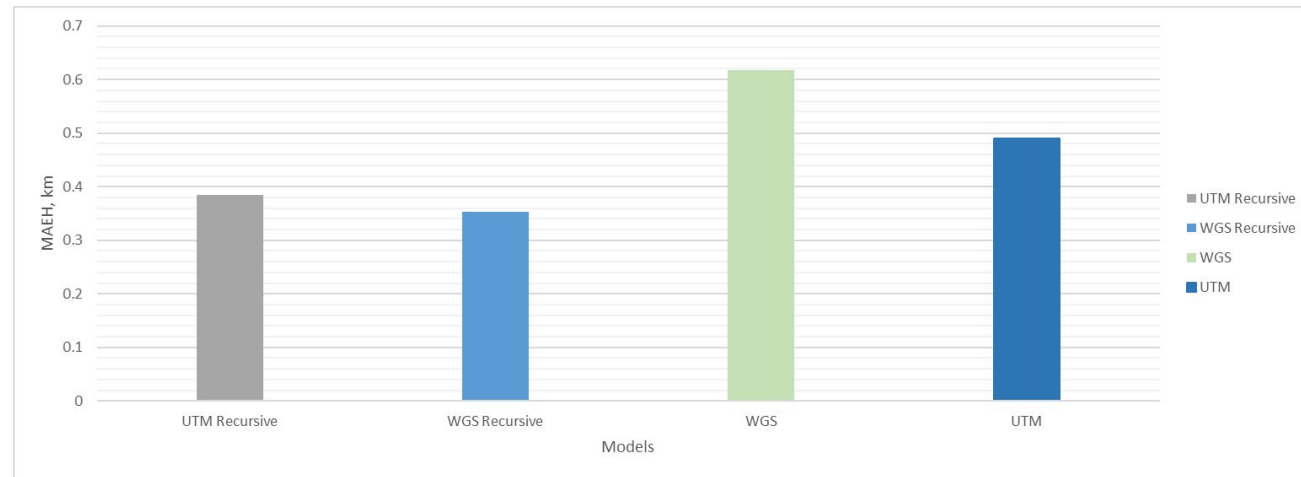
Rekurentinio tinklo architektūros ir celių kiekio įtaka LSTM tinkle



Rezultatai (2)



Rekursinis prognozavimo
reikšmių perskaičiavimas

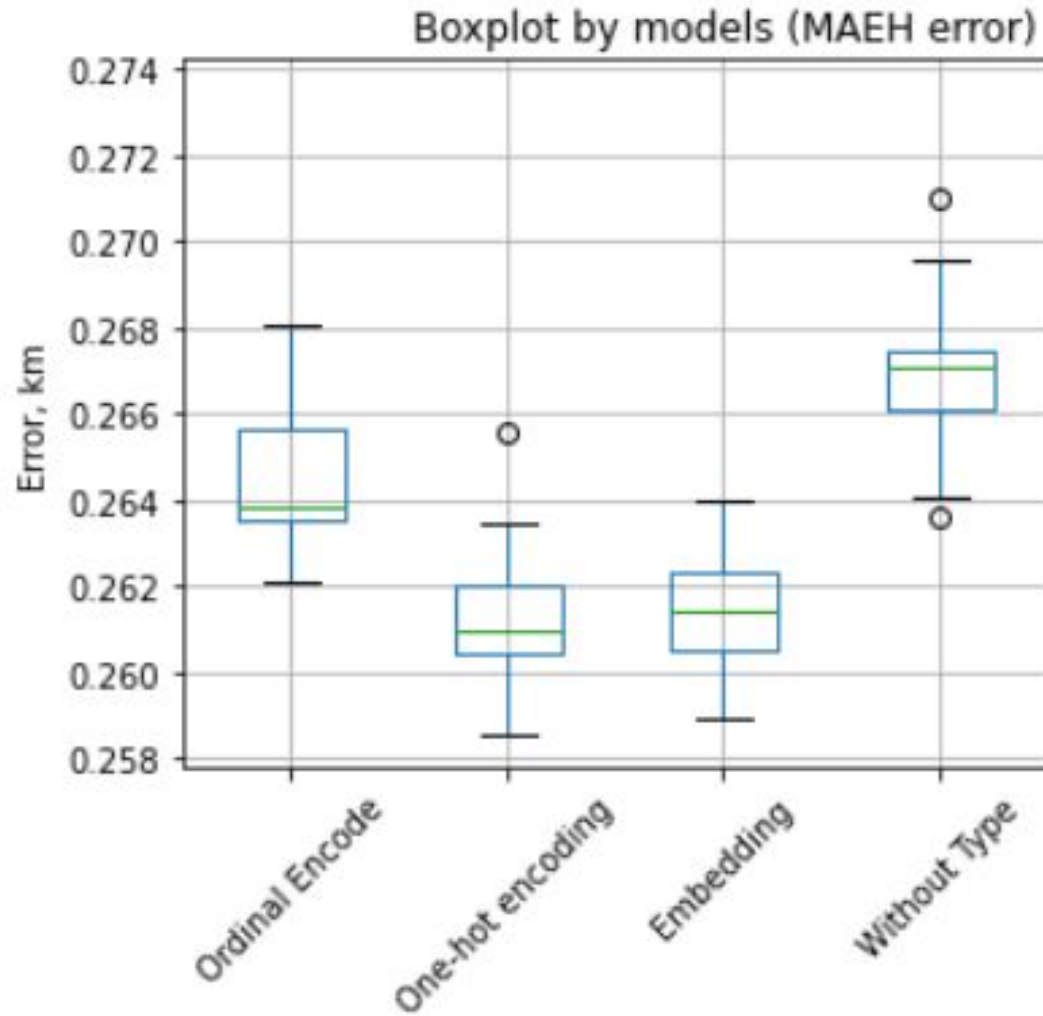


Recursive multi-step trajectory prediction model.

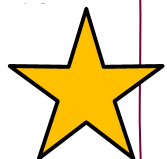
Rezultatai (3)

MAEH klaidos sklaida eksperimentuose (po 15 bandymų)

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

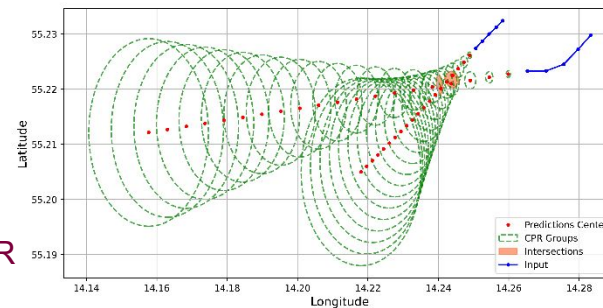


Rezultatai (4)

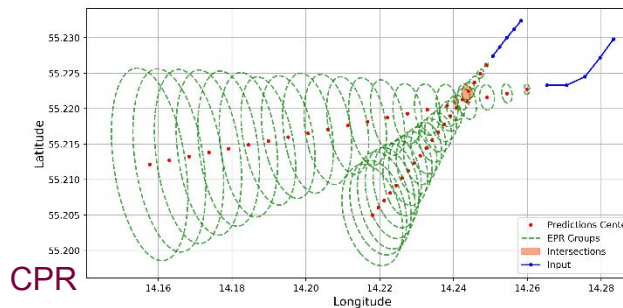


Konforminio prognozavimo
regiono tinkamumas vertinant
susidūrimo tikimybes

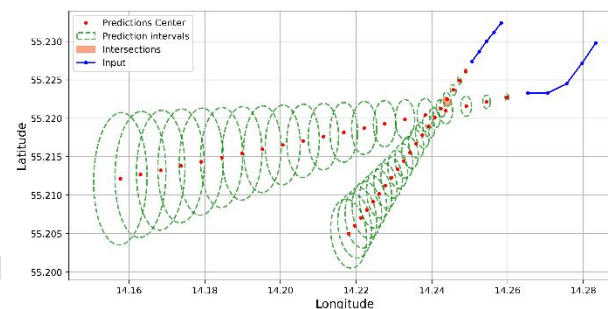
EPR



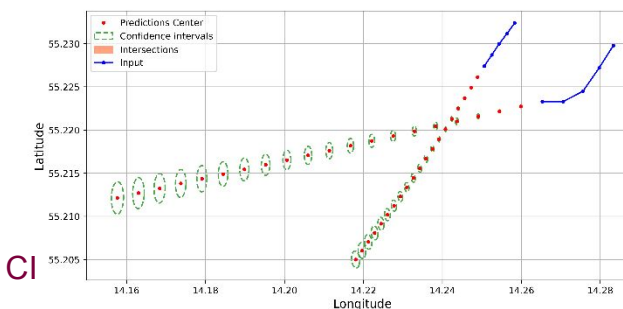
CPR



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.

$$one-hot = X_i^{(t,n+1)}$$

$$embedded = X_i^{(t,n+1)}$$





Dėkoju



Robertas Jurkus
VU Doktorantas
DMSTI, Vaizdų ir signalų analizės grupė
robertas.jurkus@mif.stud.vu.lt