



PREDIKCIA OBLASTI VÝSKYTU RÝB POMOCOU METÓD STROJOVÉHO UČENIA

Vedúci práce: doc. RNDr. Ľubomír Antoni, PhD.

Konzultant: RNDr. Mariana Zapotoková, PhD.

Bc. Daniela Pillárová

Obhajoba diplomovej práce, 26.5.2026



Globálny problém

Nezákonný, nenahlásený a neregulovaný rybolov:

- Ekonomické straty odhadom 10 až 23,5 miliárd \$ ročne (11 – 26 miliónov ton rýb). [1]
- V rozvojových krajinách reálny úlovok preyšuje oficiálne nahlásený až o 40 %. [2]
- Export morských plodov v Nórsku predstavuje 1,4 miliardy € ročne (3. najvýznamnejší sektor po rope a plyne). [3]

Tlak na udržateľnosť:

- Cieľ Nórska znížiť emisie z rybolovu o 50 % do roku 2030. [3]

Motivácia

Lokalizácia oblastí s vysokým potenciálom výskytu rýb prináša niekoľko vylepšení:

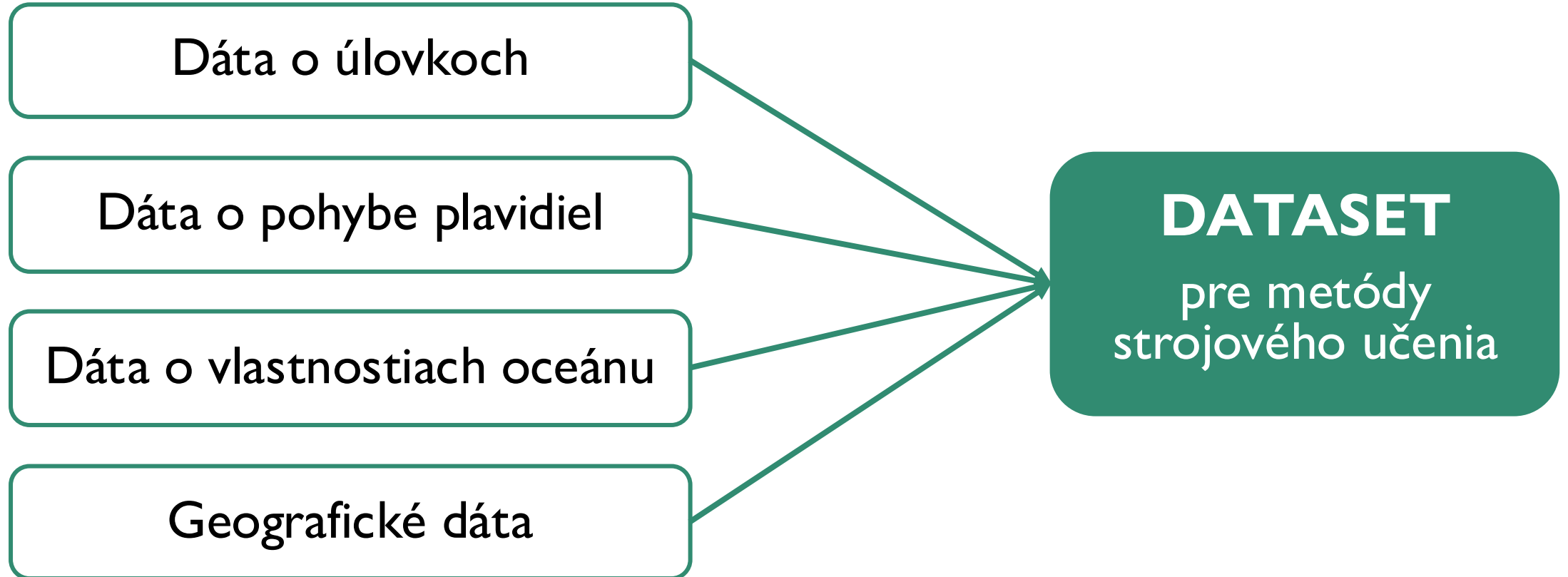
Optimalizácia komerčného rybolovu

Zvýšenie bezpečnosti námornej dopravy

Zefektívnenie kontroly a monitoringu



Návrh datasetu



SAFRUDDIN, Safruddin; HIDAYAT, R.; ZAINUDDIN, M, 2018. Effects of environmental factors on anchovies *Stolephorus* sp. distribution in Bone Gulf, Indonesia. In: AACL Bioflux.

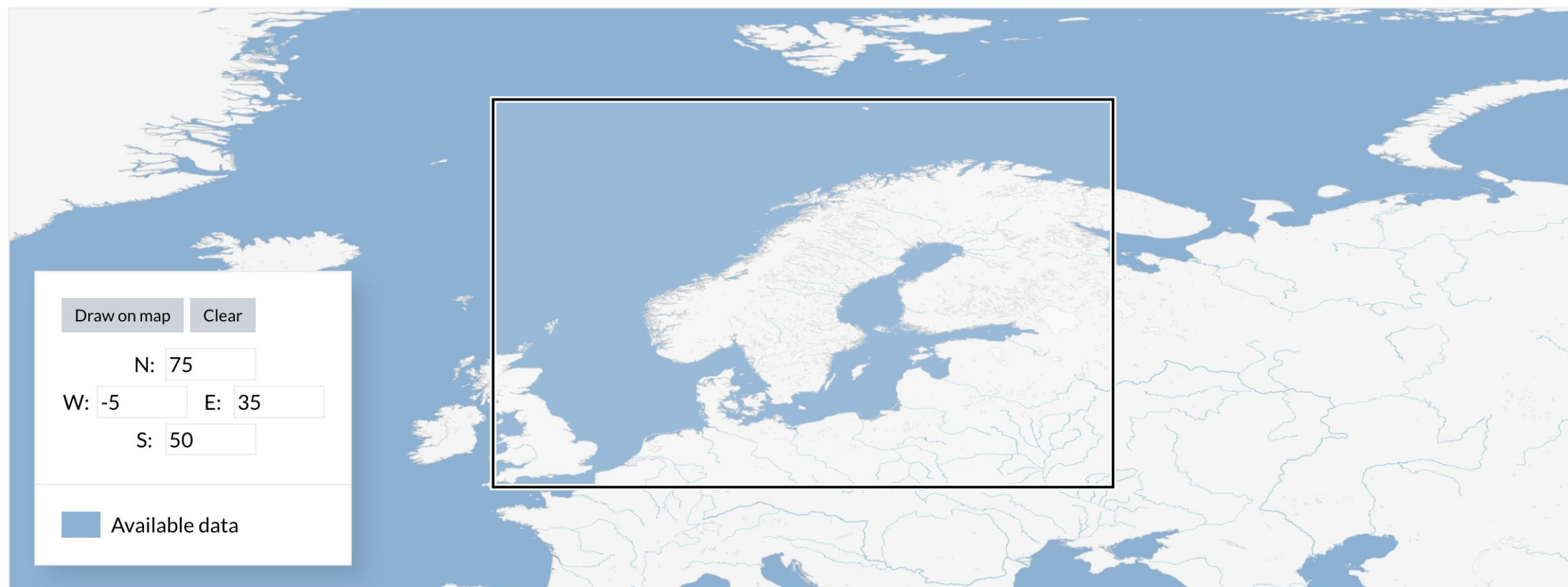
NINGSIH, Wiwik Andriyani Lestari, et al., 2021. Analysis of the relationship between chlorophyll-a and sea surface temperature on marine capture fisheries production in Indonesia: 2018. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing, p. 012057. DOI: 10.1088/1755-1315/944/1/012057

HIDAYAT, Rachmat, et al., 2022. The use of statistical models in identifying skipjack tuna habitat characteristics during the Southeast Monsoon in the Bone Gulf, Indonesia. In: Biodiversitas: Journal of Biological Diversity, 23.4. DOI: 10.13057/biodiv/ d230459

Skúmaná geografická oblasť a časový interval

- Oblasť Severnej Európy –
okolie Nórska

- Roky: **2023, 2024 a 2025**



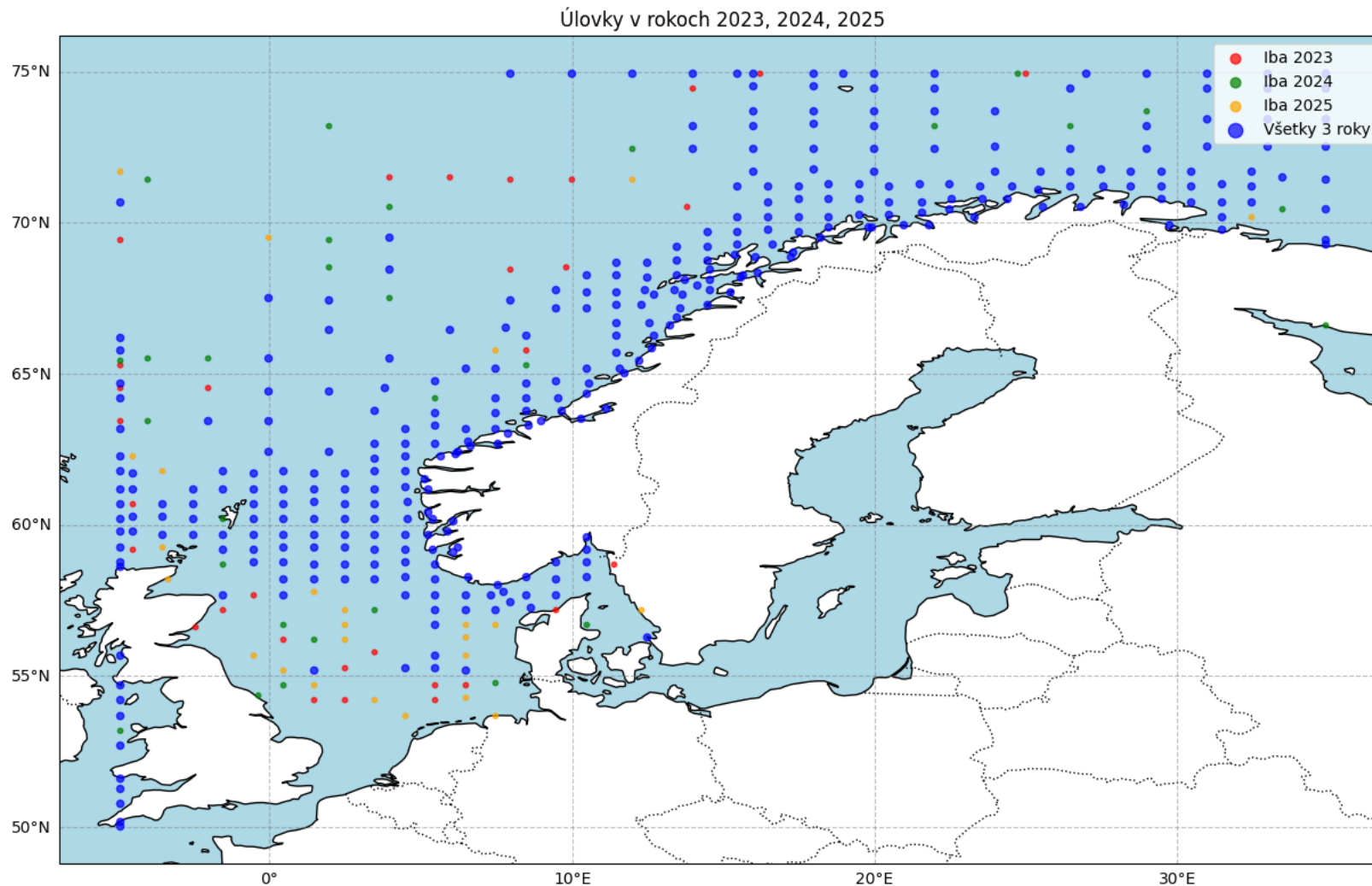
Dáta o úlovkoch



Fiskeridirektoratet

Nórske riaditeľstvo rybolovu

- Záznamy o registrovaných úlovkoch
- **Dôležité atribúty:**
 - Súradnice GPS
 - Dátum úlovku
 - Hmotnosť úlovku (v kg)
- **Agregácia dát** podľa dátumu a miesta
- **Modré bodky** na obrázku: stále loviská počas 3 rokov





Dáta o pohybe plavidiel (VMS)

- Údaje o nórskejších rybárskych plavidlách s dĺžkou **15 metrov a viac**
- Agregácia na **týždenné hodnoty**
- Dôležité atribúty:

1. Počet hlásených plavidiel

boat_count_k

2. Počet unikátnych plavidiel

unique_boats_k

3. Priemerná rýchlosť

$$\text{avg_speed}_k = \frac{\sum_{i=1}^{\text{boat_count}_k} v_i}{\text{boat_count}_k}$$

4. Pomer lodí s nízkou rýchlosťou

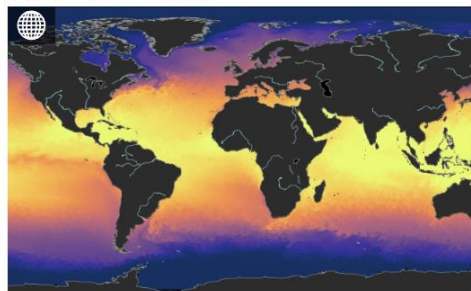
$$\text{slow_speed_ratio}_k = \frac{\sum_{i=1}^{\text{boat_count}_k} [v_i < 2.0]}{\text{boat_count}_k}$$

kde k predstavuje k -ty týždeň, v_i predstavuje rýchlosť plavidla

a $[v_i < 2.0]$ vráti hodnotu 1.0 ak sa plavidlo pohybuje rýchlosťou menšou ako 2 uzly (námorná míľa/h), inak 0.0



MOST POPULAR



Global Ocean Physics Analysis and Forecast

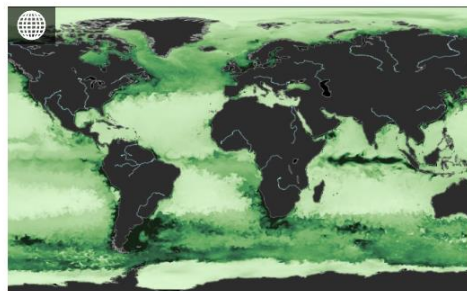
GLOBAL_ANALYSISFORECAST_P... 001_024

Models

Global, 0.083° × 0.083° × 50 levels

1 Nov 2020 to 4 Apr 2026, hourly, daily,...

Temperature, salinity, sea surface height, velocity, mixed layer thickness, wave...



Global Ocean Biogeochemistry Analysis and Forecast

GLOBAL_ANALYSISFORECAST_B... 001_028

Models

Global, 0.25° × 0.25° × 50 levels

1 Oct 2021 to 3 Apr 2026, daily, monthly

Plankton, nutrients, oxygen, carbonate system, optics



Copernicus Marine Service

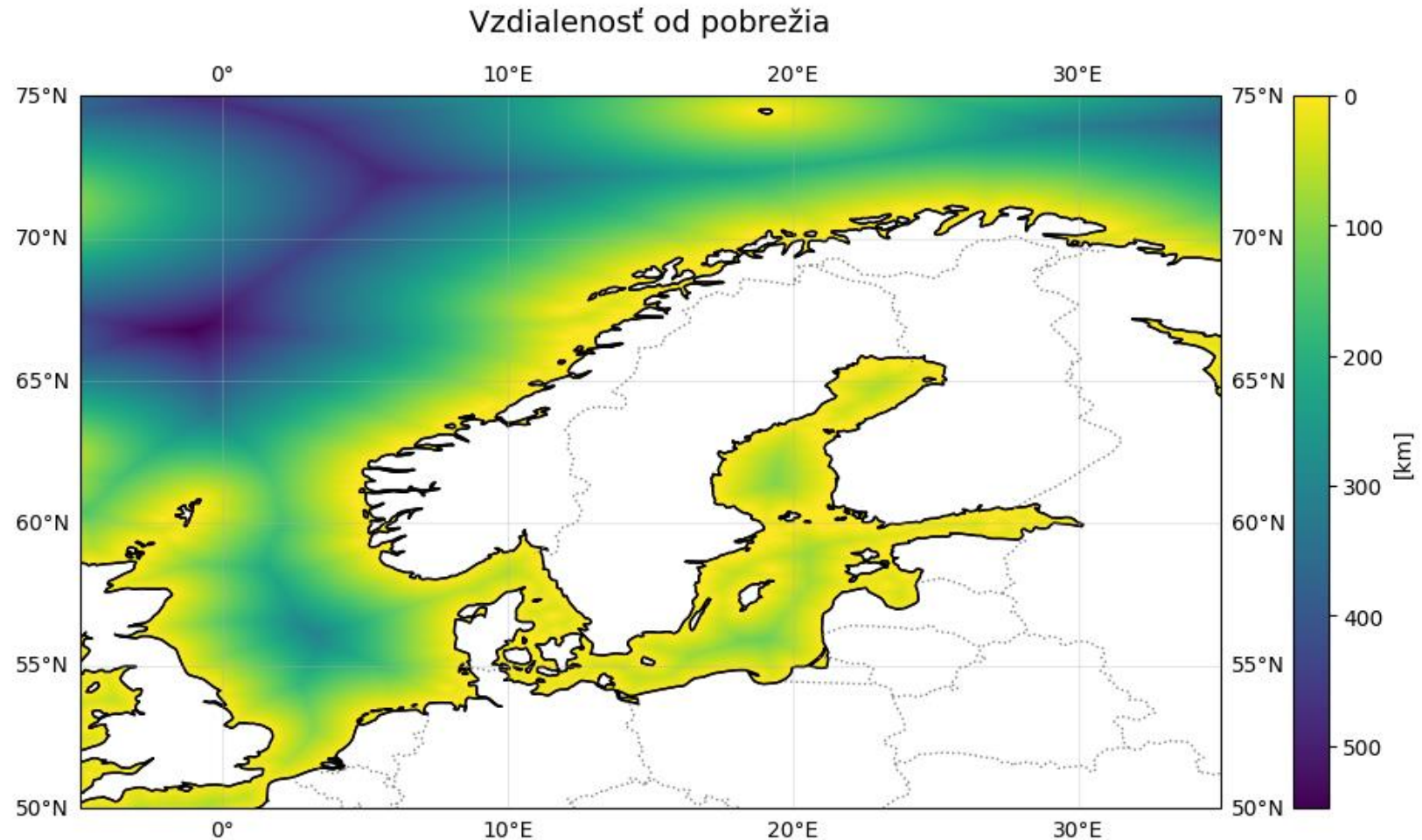
Satelitné monitorovanie Zeme

Dáta o oceáne

- **SLANOSŤ** (sal): vody je indikátor morských prúdov a hustoty vody
- **TEPLOTA** (sst): vody indikuje výskyt konkrétnych druhov rýb
- **CHLOROFYL** (chl): vody je ukazovateľ fytoplanktónu, potravy pre ryby
- Agregácia na **týždenné priemery**
- **Interpolácia** chýbajúcich hodnôt – **chlorofylu**
- **Mriežkový systém** – rozdelenie oceánu na „grid“ s rozlíšením:
 - **0,083° x 0,083°**
 - **9,2 x 4,3 km**
 - **301 x 480 bodov**

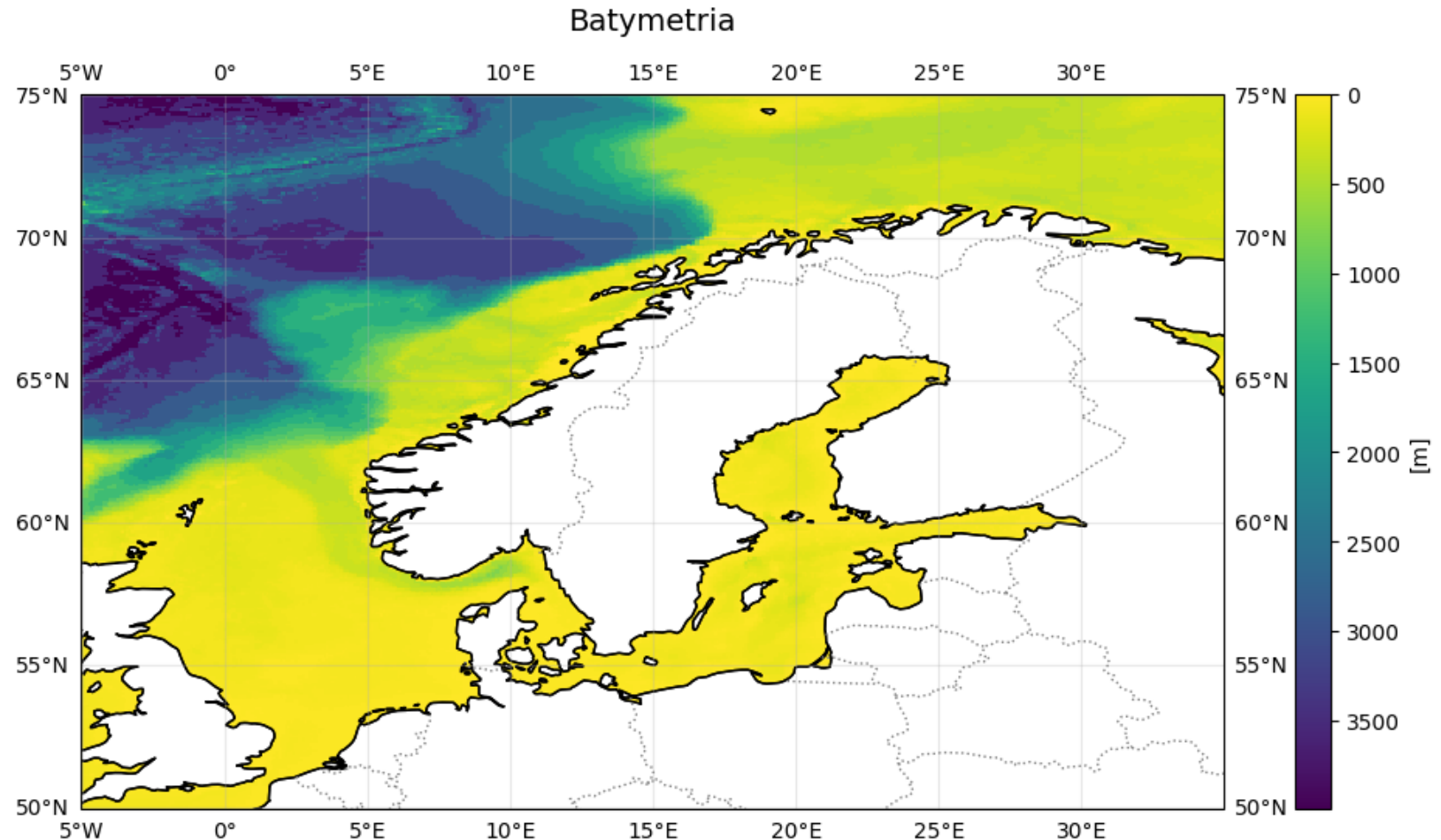
Vzdialenosť od pobrežia

- Vzdialenosť konkrétneho bodu na mape **od najbližšieho brehu**



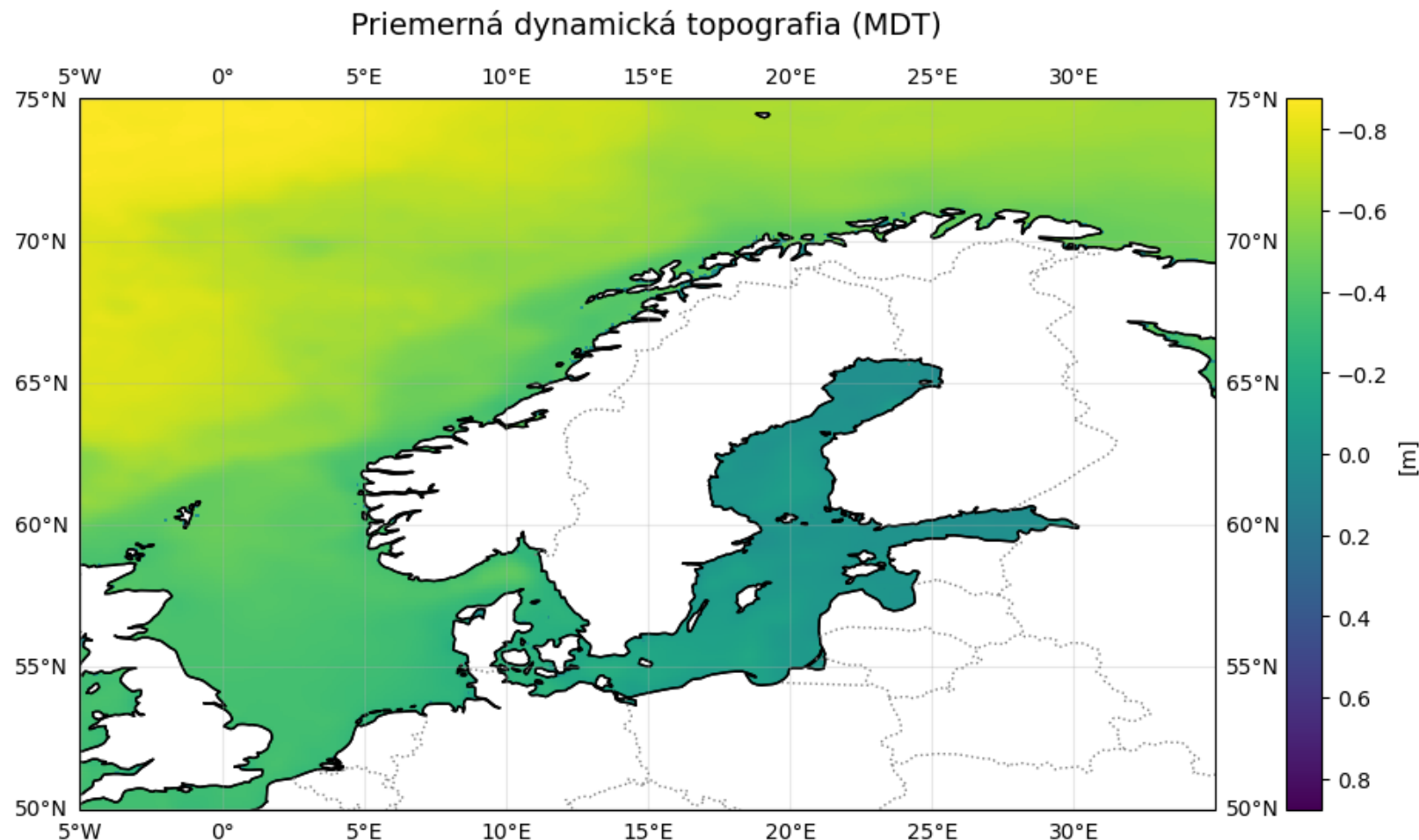
Batymetria

- Údaj o tom ako hlboko sa nachádza morské dno pod hladinou
- Súvislosť s teplotou a tlakom vody v rôznych vrstvách



Priemerná dynamická topografia

- **Tvar a nerovnosti povrchu morského dna pod vodou**
- **Vynášanie živín z hlbín smerom nahor pri náraze na terén – súvis s chlorofylom**



Náš cieľ

Pre jednu **GPS oblasť** a **jeden rok** vieme:

Historické dáta

- Denné **úlovky** v kg
- Týždenné **VMS** dáta
- Týždenné **oceánske** vlastnosti
 - **Geografické** dáta



Chceme predikovať:

Bude v danej GPS oblasti
úlovok d'alsí týždeň?

Klasifikácia:

- **1**: ak úlovok bude väčší ako 0 kg
- **0**: inak

Úlovky (kg) - metóda posuvných okien

	1.1.2023	2.1.2023	3.1.2023	1.1.2023	...	25.1.2023	26.1.2023	27.1.2023	28.1.2023			
	day_1	day_2	day_3	day_4	...	day_25	day_26	day_27	day_28			
	0	0	41	3 462	...	37 836	452	0	0			
	0	0	41	3 462	897	...	452	0	0	54 146		
	0	0	41	3 462	897	29 978	...	0	0	54 146	1 075	
	0	0	41	3 462	897	29 978	446	...	0	54 146	1 075	0
	4.1.2023	5.1.2023	6.1.2023	7.1.2023	...	28.1.2023	29.1.2023	30.1.2023	31.1.2023			

Pre rok 2023:

- **331** posuvných okien (**28-dňových intervalov**) z jednej GPS oblasti (šírka x výška)
- posledné okno začína 27.11.2023

Úlovok budúci týždeň

3.1.2023	1.1.2023	...	25.1.2023	26.1.2023	27.1.2023	28.1.2023	Suma úlovkov počas: 29.1. - 4.2.2023	Bude úlovok? (1/0)
day_3	day_4	...	day_25	day_26	day_27	day_28	catch_next_week	CNW_binary
41	3 462	...	37 836	452	0	0	55 221	1
3 462	897	...	452	0	0	54 146	1 075	1
897	29 978	...	0	0	54 146	1 075	0	0
29 978	446	...	0	54 146	1 075	0	0	0
6.1.2023	7.1.2023	...	28.1.2023	29.1.2023	30.1.2023	31.1.2023	1.2. - 7.2.2023	

- Atribút **catch_next_week** je **suma hmotnosti úlovkov nasledujúcich siedmych dní** po dni reprezentovanom v atribúte day_28
- **CNW_binary** je **binárne ohodnotenie** atribútu catch_next_week podľa podmienky:
 - 1: ak úlovok väčší ako 0 kg, inak 0

Spojenie dát = dataset pre jeden rok

GRID	WINDOW START	DAYS	SAL	SST	CHL	VMS	DIST, BATHY, TOPO,	CATCH NEXT WEEK	CNW binary
grid_i	date	day_1		week_1		boat_count w1-w4	distance_to_coast	xxx kg	0/1
grid_j	month	...		week_2		unique_boats w1-w4	bathymetry		
grid_lon		day_28		week_3		avg_speed w1-w4	topography		
grid_lat				week_4		slow_speed_ratio w1-w4			
				week_5 (next week values)					
4	2	28	5	5	5	16	3	1	1

70 atribútov v datasete

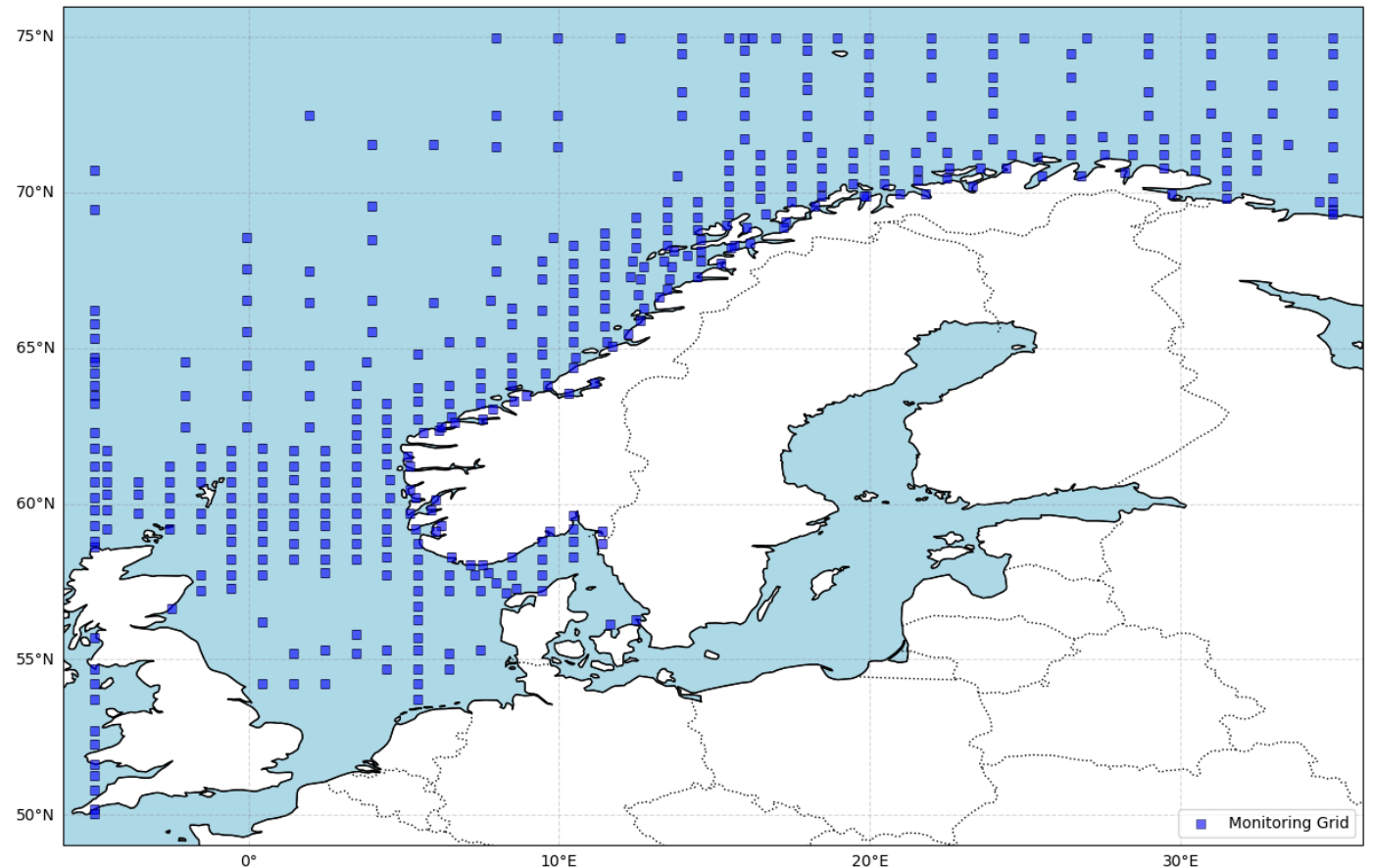
Rozmery datasetov

- **2023: (141 999, 70)**
- **2024: (141 999, 70)**
- **2025: (141 006, 70)**
 - menej oblastí s úlovkom

Rozbor pre rok **2023**:

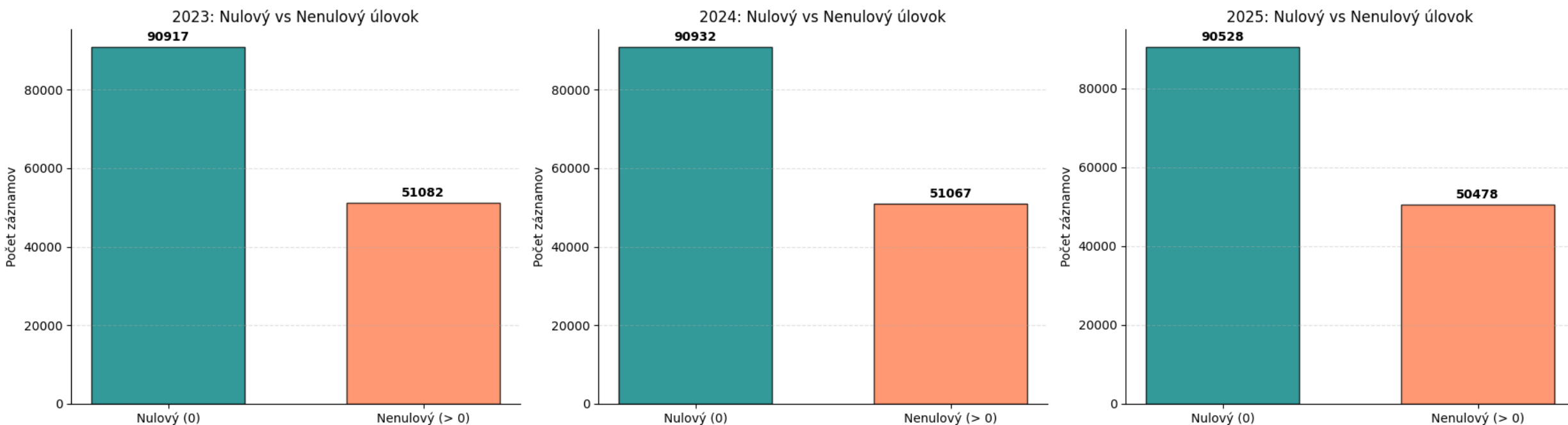
- **331** posuvných okien z jednej GPS oblasti
 - **429** GPS oblastí, kde boli zaznamenané úlovky
- 331 x 429 = 141 999**

2023 Grid structure (Number of unique grids: 429)



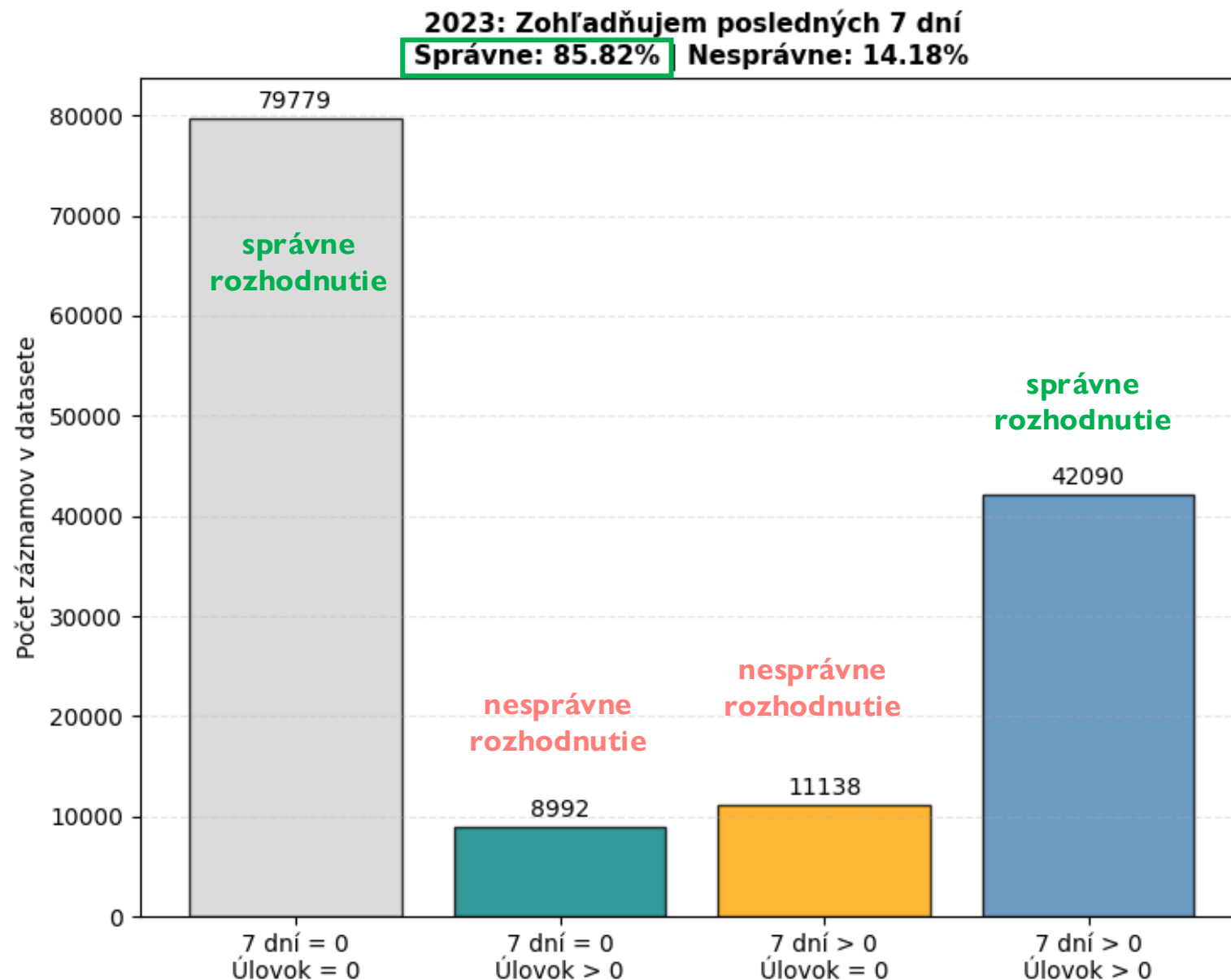
Analýza atribútu catch_next_week

- **Nevybalancované** datasety – prevažka nulových úlovkov
- **Náhodne podvzorkovanie** počas tréningu modelu, aby obidve triedy mali približne 51 000 záznamov (posuvných okien)



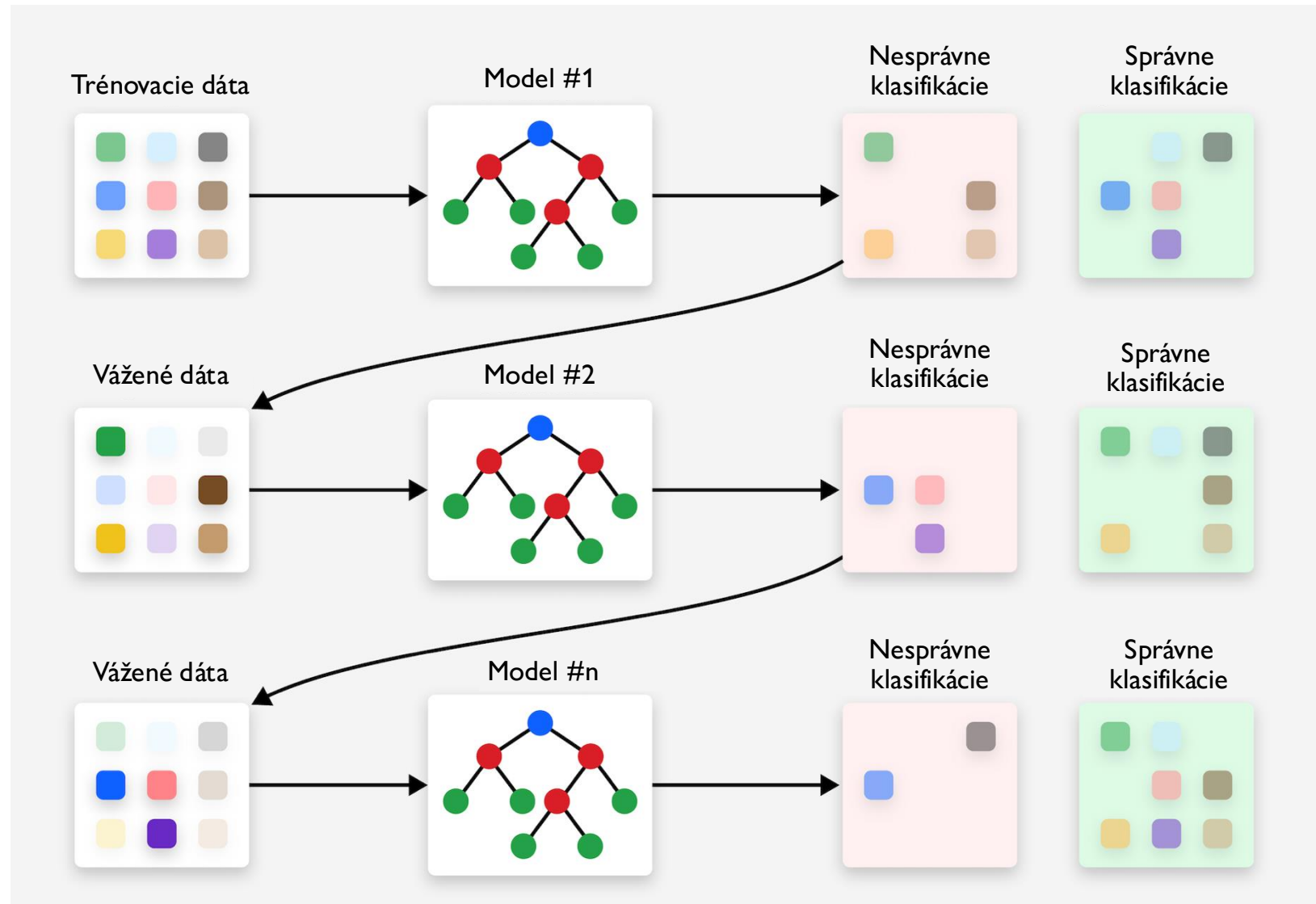
Baseline model

- Simuluje **rozhodovanie rybára**
- Ak v sledovanom historickom období (7 dní dozadu) bol úlovok väčší ako 0 kg, bude aj tento týždeň, to znamená, že **idem loviť**



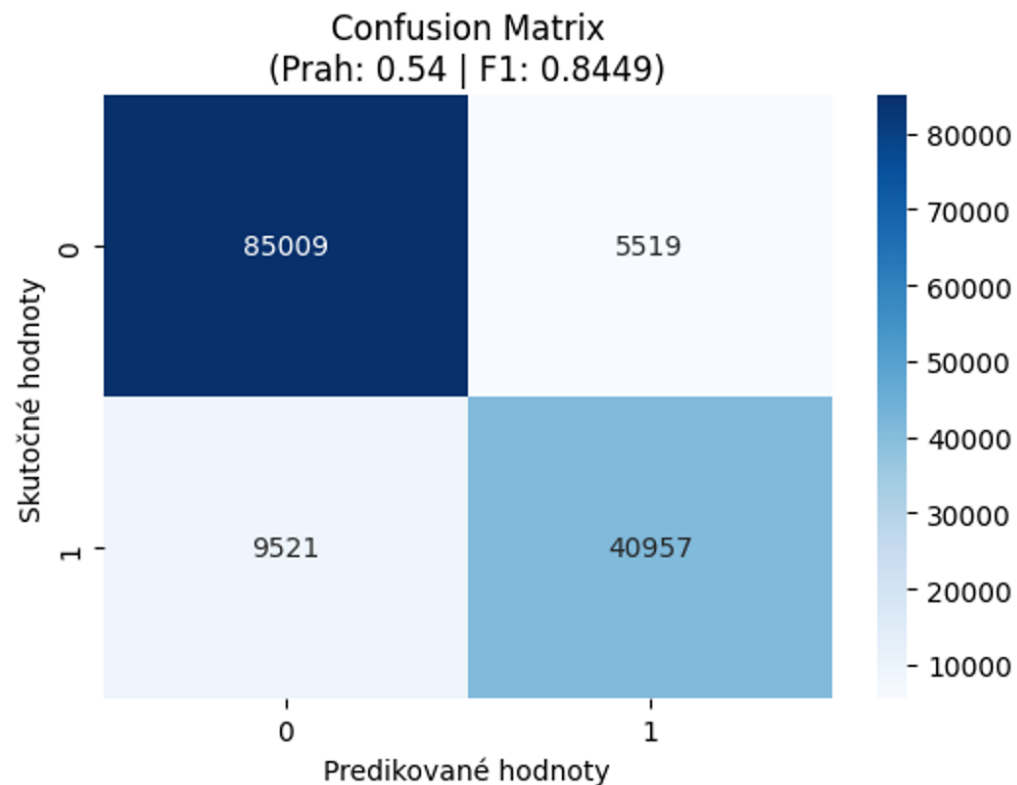
Metódy strojového učenia založené na gradient boostingu

- XGBoost
- LightGBM

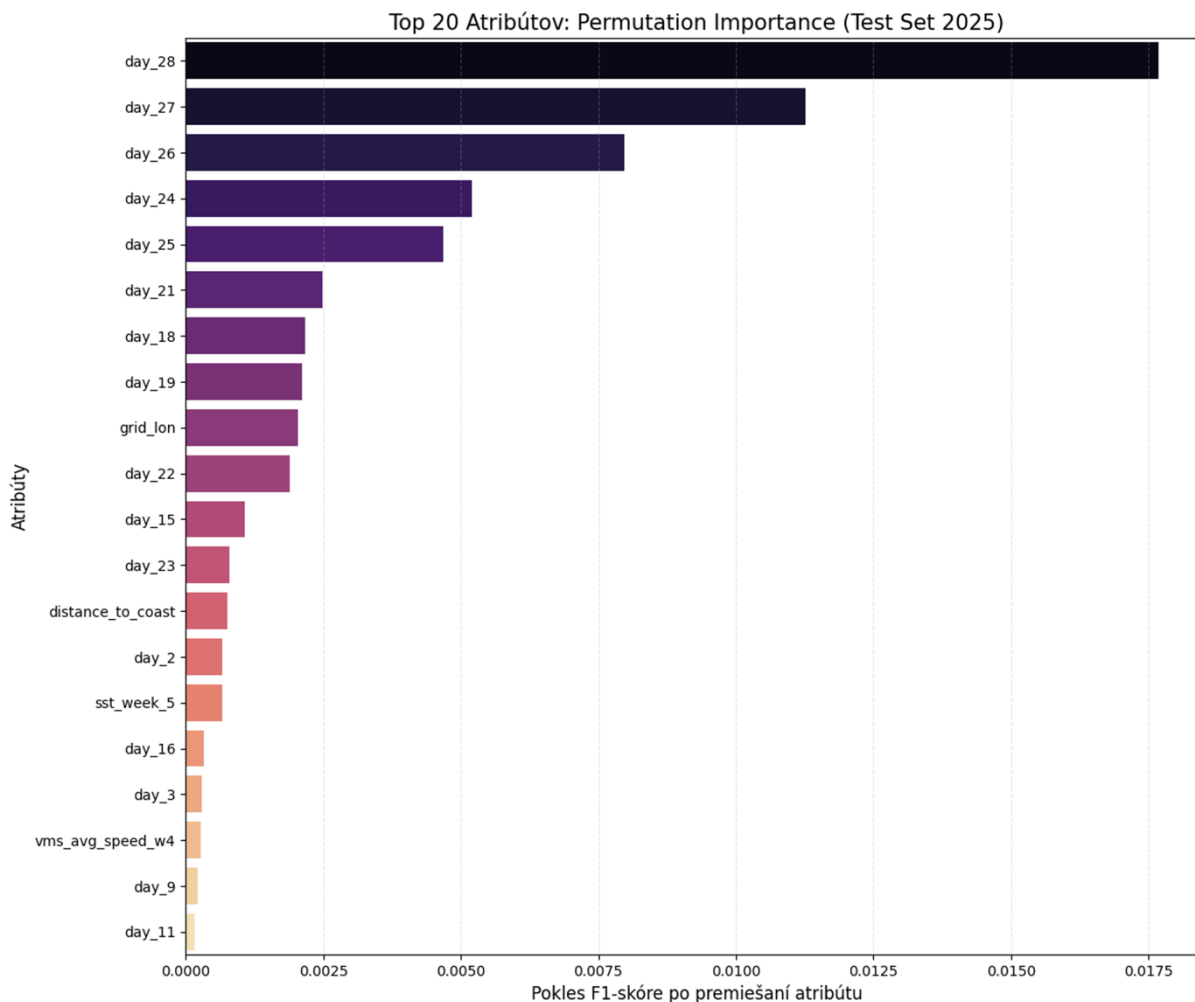


Výsledky XGBoost modelu

- Trénovacia množina: 2023 a 2024
- Testovacia množina: 2025

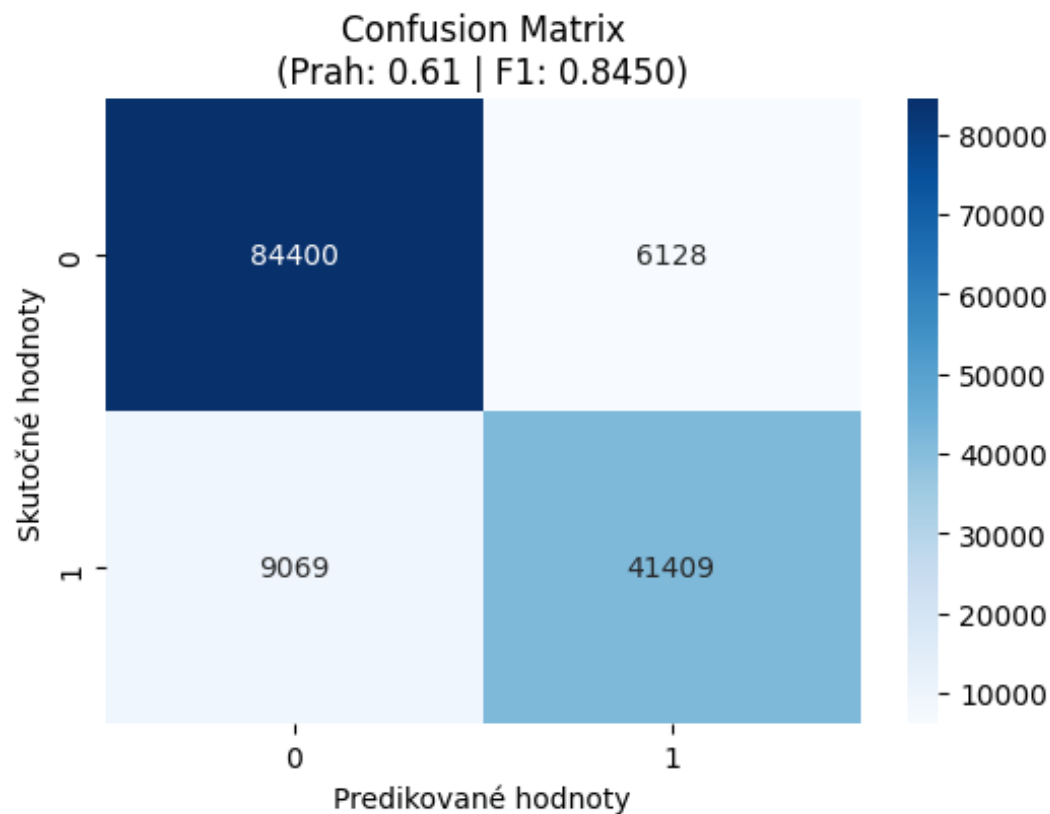


Trieda	Presnosť	Citlivosť	F_1 -skóre	Podpora
0 (Bez úlovku)	0.90	0.94	0.92	90 528
1 (Úlovok)	0.88	0.81	0.84	50 478
Úspešnosť		0.89		141 006

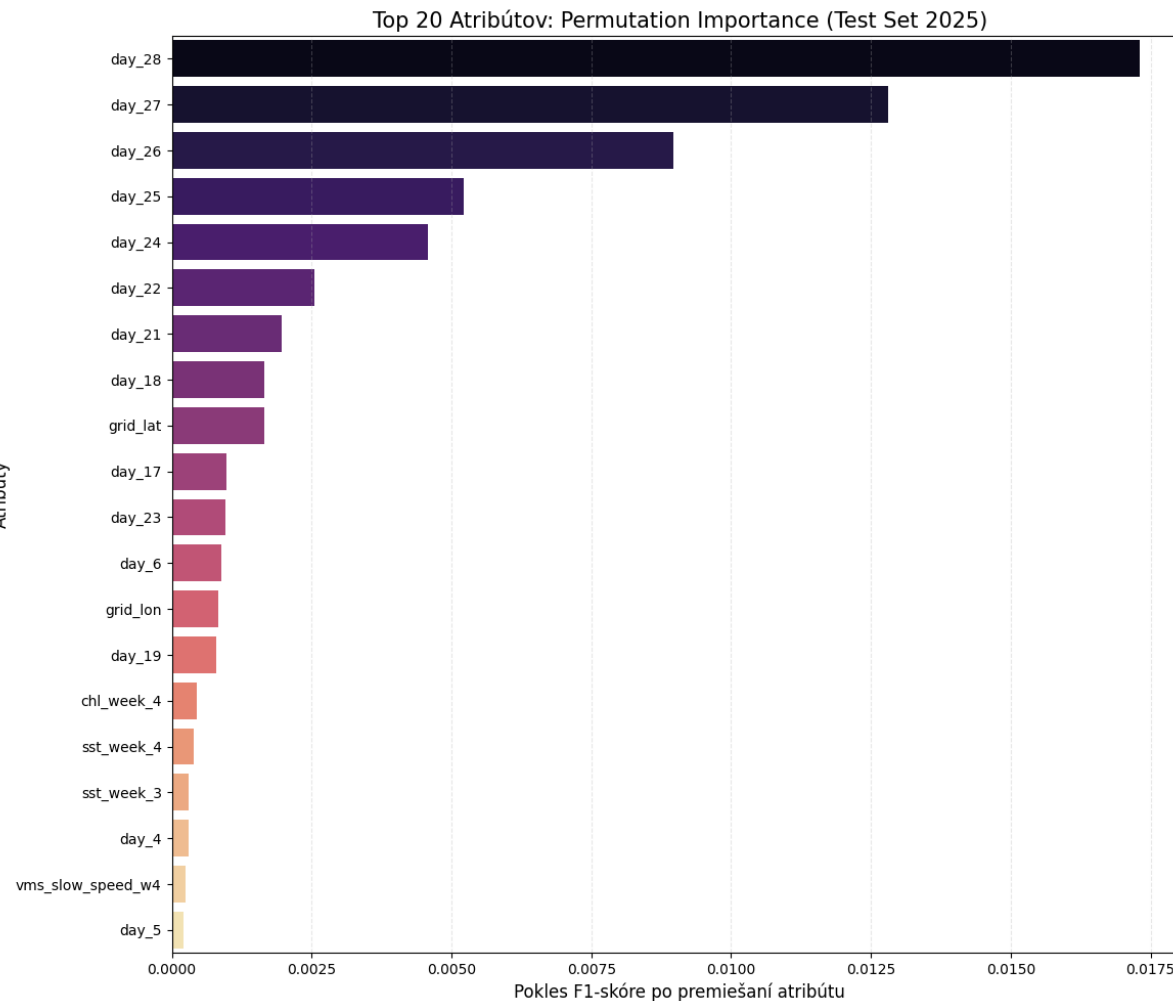


Výsledky LightGBM modelu

- Trénovacia množina: 2023 a 2024
- Testovacia množina: 2025

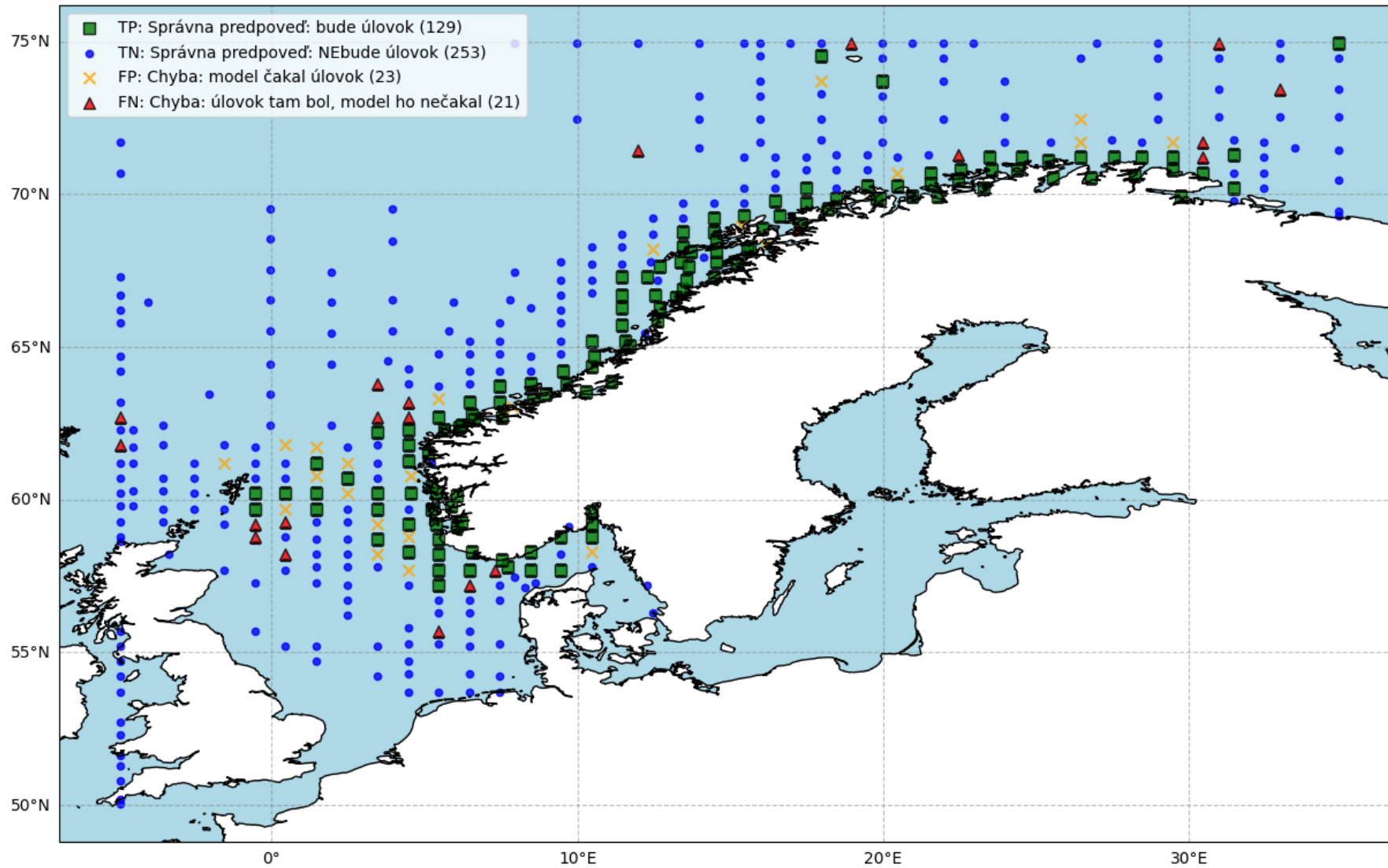


Trieda	Presnosť	Citlivosť	F_1 -skóre	Podpora
0 (Bez úlovku)	0.90	0.93	0.92	90 528
1 (Úlovok)	0.87	0.82	0.85	50 478
Úspešnosť		0.89		141 006



Vizualizácia úspešnosti modelu LightGBM

Porovnanie predikcie s realitou (Úspešnosť: 89.67 %)
Vstupy z: 01.09.2025 | Predikovaný týždeň: 29.09.2025 - 05.10.2025



Porovnanie výsledkov s inými prácami

Práca	Problém	Dáta	Model	Výsledok
Scene-Based Ensemble [4]	Lokalizácia tuniaka v Pacifiku	Teplota, chlorofyl, výška hladiny	Scénicky orientovaný ensemble	Úspešnosť: 83 %
FishMAZE [5]	Regresia súradníc lovu	Úlovky, teplota, salinita	Deep Learning / Ensemble	RMSE: 8,68
Lodestar Platform [6]	Predikcia lokalít lovu (týždeň dopredu)	Úlovky, historické trasy	XGBRegressor	Úspešnosť: 5 % (baseline model: 67 %)
Náš baseline model	Klasifikácia výskytu úlovku (týždeň vopred)	Úlovky	Prístup: zohľadňujem úlovky týždeň dozadu	Úspešnosť: 85 %
Naše metódy strojového učenia	Klasifikácia výskytu úlovku (týždeň vopred)	Úlovky, VMS, teplota, salinita, chlorofyl, geo dáta	XGBoost / LightGBM	Úspešnosť: 89 %, F1-skóre: 0,84

Ďalšie možnosti výskumu

- **Druhovo-špecifická** predikcia
- **Geografické rozšírenie** oblasti
- Implementácia **regresných modelov**



Literatúra

- [1] AGNEW, David J., et al., 2009. Estimating the worldwide extent of illegal fishing. In: PloS one, 4.2: e4570. DOI: 10.1371/journal.pone.0004570
- [2] PAULY, Daniel; ZELLER, Dirk, 2016. Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. In: Nature communications, 7.1: 10244. DOI: 10.1038/ncomms10244
- [3] LINKIÖ, Visa; LAHTINEN, Kristian; KOLMONEN, Juho, 2022. Clusters and traveling fisherman. In: Nordic Machine Intelligence, 2.2: 13-15. DOI: 10.5617/nmi.9930
- [4] ALFATINAH, Adillah, et al., 2023. Fishing area prediction using scene-based ensemble models. In: Journal of Marine Science and Engineering, 11.7: 1398. DOI: 10.3390/jmse11071398
- [5] LAMBON, Aldrin, et al., 2022. FishMAZE: Fish Monitoring and AI-based Zone Evaluation. In: Nordic Machine Intelligence, 2.2. DOI: 10.5617/nmi.9941
- [6] BREKKE, Åsmund, et al., 2022. FishAI: The Lodestar fishing platform. DOI: 10.5617/nmi.9931
- [7] SAFRUDDIN, Safruddin; HIDAYAT, R.; ZAINUDDIN, M, 2018. Effects of environmental factors on anchovies *Stolephorus* sp. distribution in Bone Gulf, Indonesia. In: AACL Bioflux.
- [8] NINGSIH, Wiwik Andriyani Lestari, et al., 2021. Analysis of the relationship between chlorophyll-a and sea surface temperature on marine capture fisheries production in Indonesia: 2018. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing, p. 012057. DOI: 10.1088/1755-1315/944/1/012057
- [9] HIDAYAT, Rachmat, et al., 2022. The use of statistical models in identifying skipjack tuna habitat characteristics during the Southeast Monsoon in the Bone Gulf, Indonesia. In: Biodiversitas: Journal of Biological Diversity, 23.4. DOI: 10.13057/biodiv/ d230459
- [10] KAISER, Michel J., et al., 2002. Modification of marine habitats by trawling activities: prognosis and solutions. In: Fish and fisheries, 3.2: 114-136. DOI: 10.1046/j.1467-2979.2002.00079.x
- [11] MORATO, Telmo, et al., 2016. Fishing down the deep. In: Fish and fisheries, 7.1: 24-34. DOI: 10.1111/j.1467-2979.2006.00205.x
- [12] SRINIVASAN, Margaret; TSONTOS, Vardis, 2023. Satellite altimetry for ocean and coastal applications: A review. In: Remote Sensing, 15.16: 3939. DOI: 10.3390/rs15163939

ĎAKUJEM ZA POZORNOST

Vedúci práce: doc. RNDr. Ľubomír Antoni, PhD.

Konzultant: RNDr. Mariana Zapotoková, PhD.

Bc. Daniela Pillárová

Obhajoba diplomovej práce, 26.5.2026



Otázka #1

Vedúci práce:

„Modely boli trénované na dátach zo severnej Európy. Do akej miery by bolo podľa Vás možné navrhnutý prístup generalizovať aj na **iné geografické oblasti** alebo **iné druhy rýb**?“

Iné geografické oblasti závisia od dostupnosti dát o úlovkoch.

- Napr. Afrika – veľký problém s nezákonným rybolom.
- Predikcia na základe vlastností mora získaných z Copernicus Marine Service.

Iný druh rýb – je to možné, avšak musel by sa rozšíriť dataset.

Otázky #2 a #3

Oponent (RNDr. Viktor Pristaš, PhD.):

„Na konci kapitoly 1 popisujete **techniku EFB**, ktorá sa ďalej v práci explicitne nevyskytuje. Použili ste ju niekde v rámci experimentov, alebo je táto technika len súčasťou všeobecného opisu modelu **LightGBM**?“

Exclusive Feature Bundling (EFB) je vlastnosť modelu LightGBM.

- Identifikácia atribútov s vysokou mierou exkluzivity a ich zoskupenie.

„Vaše predikcie slúžia na odhad oblastí, v ktorých možno v najbližšom období očakávať úlovok. V reálnom použití by mohla byť veľmi cenná aj **informácia o skutočnom biologickom výskyte rýb nezávisle od aktivity rybárov**. Existujú podľa Vás datasey s takýmto typom informácií, ktoré by bolo možné v budúcnosti do riešenia zahrnúť?“

- **Database of Trawl Surveys** – náhodné lovy, spisovanie štatistík.
- **Animal Telemetry Network** – migrácia rýb v závislosti od vlastností oceánu.