

---

# PREDIKCIA OBLASTI VÝSKYTU RÝB POMOCOU METÓD STROJOVÉHO UČENIA

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

Konzultant: Mariana Zapotoková, Vissim

**Bc. Daniela Pillárová**

Diplomový seminár, 26.3.2026



# Globálny problém

- Každý rok je približne **20 %** celosvetových úlovkov rýb nelegálnych (**11 až 26 miliónov ton rýb**).
- Najviac postihnuté sú africké pobrežné krajiny. Ročný pokles príjmov vedie k potravinovej neistote, strate zdrojov obživy a poškodeniu morskej biodiverzity.

# Riešenie

Predpoved' rybolovných oblastí s predstihom niekoľkých dní

## Motivácia

- **Prevencia** nelegálneho rybolovu v rizikových zónach
- **Optimalizácia nákladov** firiem a vyššia efektivita úlovku

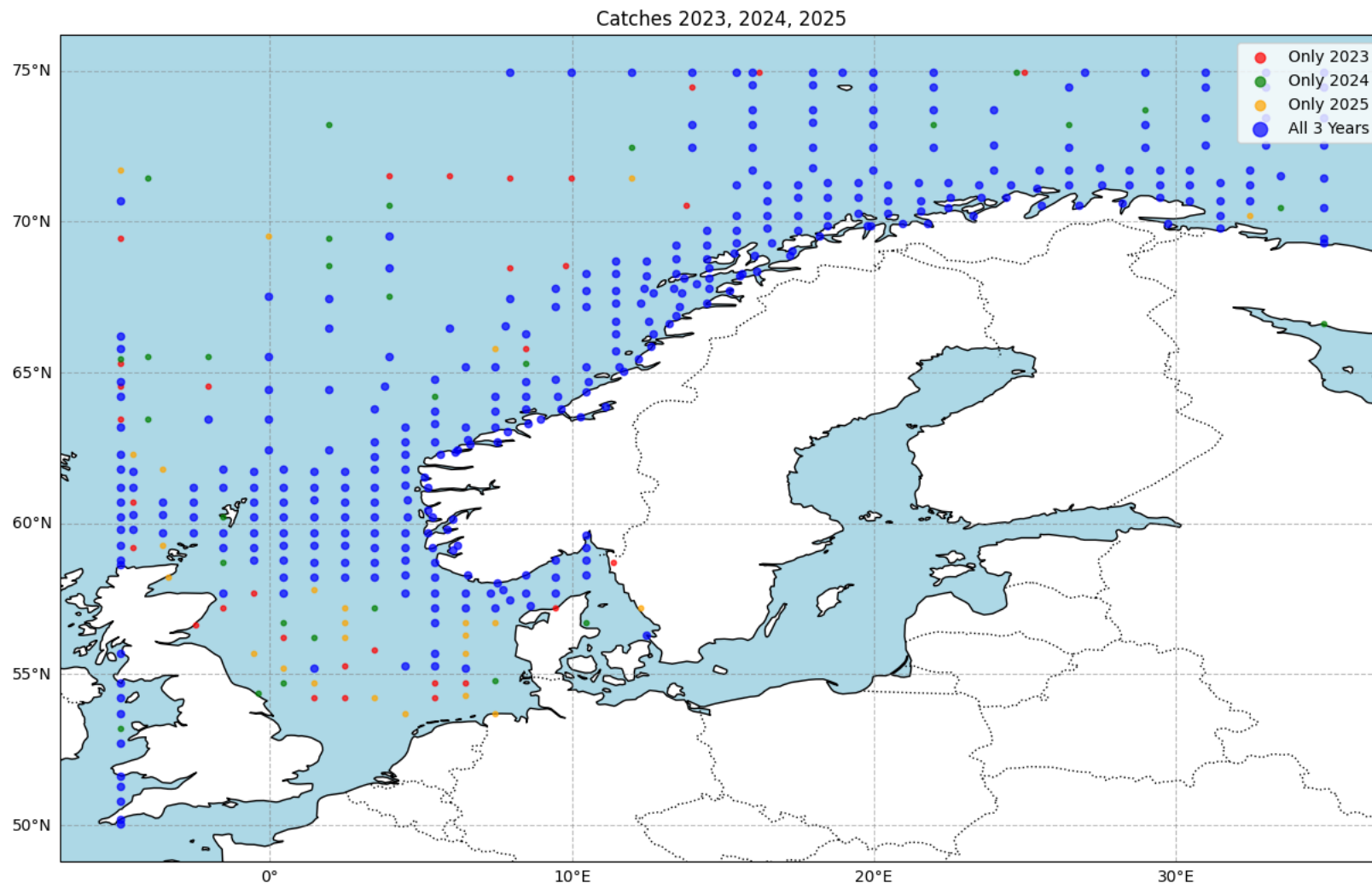
# Dataset?

- ... musíme si vytvoriť 😊
- Využitie verejných **historických dát** (úlovky, oceánske vlastnosti, dáta o polohe nórskeho plavidiel)
- Zameranie na severnú Európu a vody v **okolí Nórska**



# Dáta o úlovkoch

- Zdroj dát: Nórske riaditeľstvo rybolovu
- Roky: **2023, 2024, 2025**
- **Súradnice GPS, dátum, hmotnosť úlovku (kg)**
- Agregácia dát podľa **dátumu a miesta** (v jeden deň v rovnakej GPS oblasti môže byť viacero úlovkov)
- Modré body v grafe ako stále loviská počas 3 rokov



# VMS dáta



Fiskeridirektoratet

- Údaje o nórskejších rybárskych plavidlách s dĺžkou **15 metrov a viac** (povinnosť v Nórsku hlásiť svoju polohu)

**Priestorová** (podľa GPS oblasti) a **časová** (týždenné hodnoty) agregácia:

- **počet hlásených lodí, počet unikátnych lodí**
- **rýchlosť plavenia:**
  - nad 10 uzlov (námorná míľa/h) = tranzit plavidla
  - menej ako 2 uzly = rybolov

## Priemerná rýchlosť

$$avg\_speed = \frac{\sum \text{všetky rýchlosti za týždeň}}{\text{počet lodí za týždeň}}$$

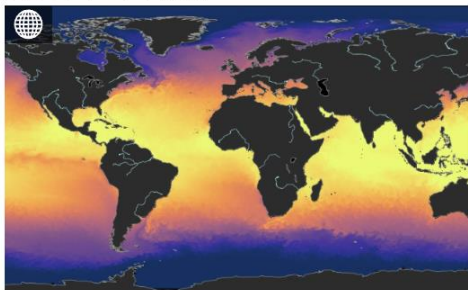
## Pomer lodí s nízkou rýchlosťou

$$slow\_speed\_ratio = \text{lambda } x: (x < 2.0).sum() / len(x)$$

**0.0:** Žiadne plavidlo neloví v tomto týždni

**1.0:** Všetky zachytené plavidlá pravdepodobne lovia v tomto týždni

## MOST POPULAR



### Global Ocean Physics Analysis and Forecast

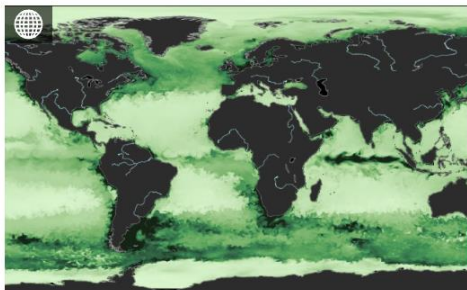
GLOBAL\_ANALYSISFORECAST\_P... 001\_024

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

# Dáta o oceáne

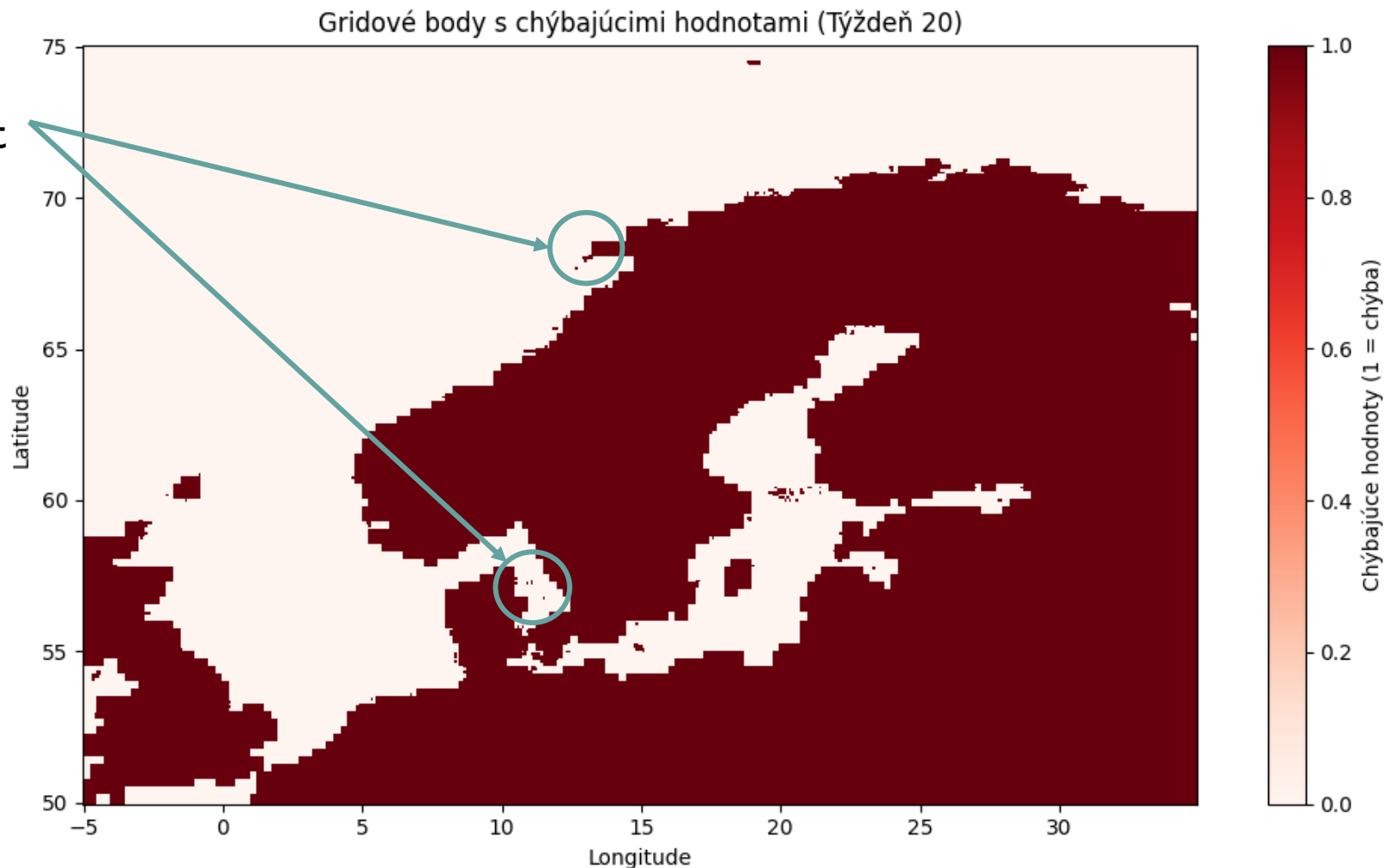
- Zdroj dát: **Copernicus Marine Service** – satelitné monitorovanie Zeme
- **SLANOSŤ** (sal) vody je indikátor morských prúdov a hustoty vody
- **TEPLOTA** (sst) vody indikuje výskyt konkrétnych druhov rýb
- **CHLOROFYL** (chl) vody je ukazovateľ fytoplanktónu, potravy pre ryby

# Spracovanie dát o oceáne

- Týždenné priemery
- Chýbajúce dáta najmä pri pobreží, najviac neúplných dát pri **chlorofyle**

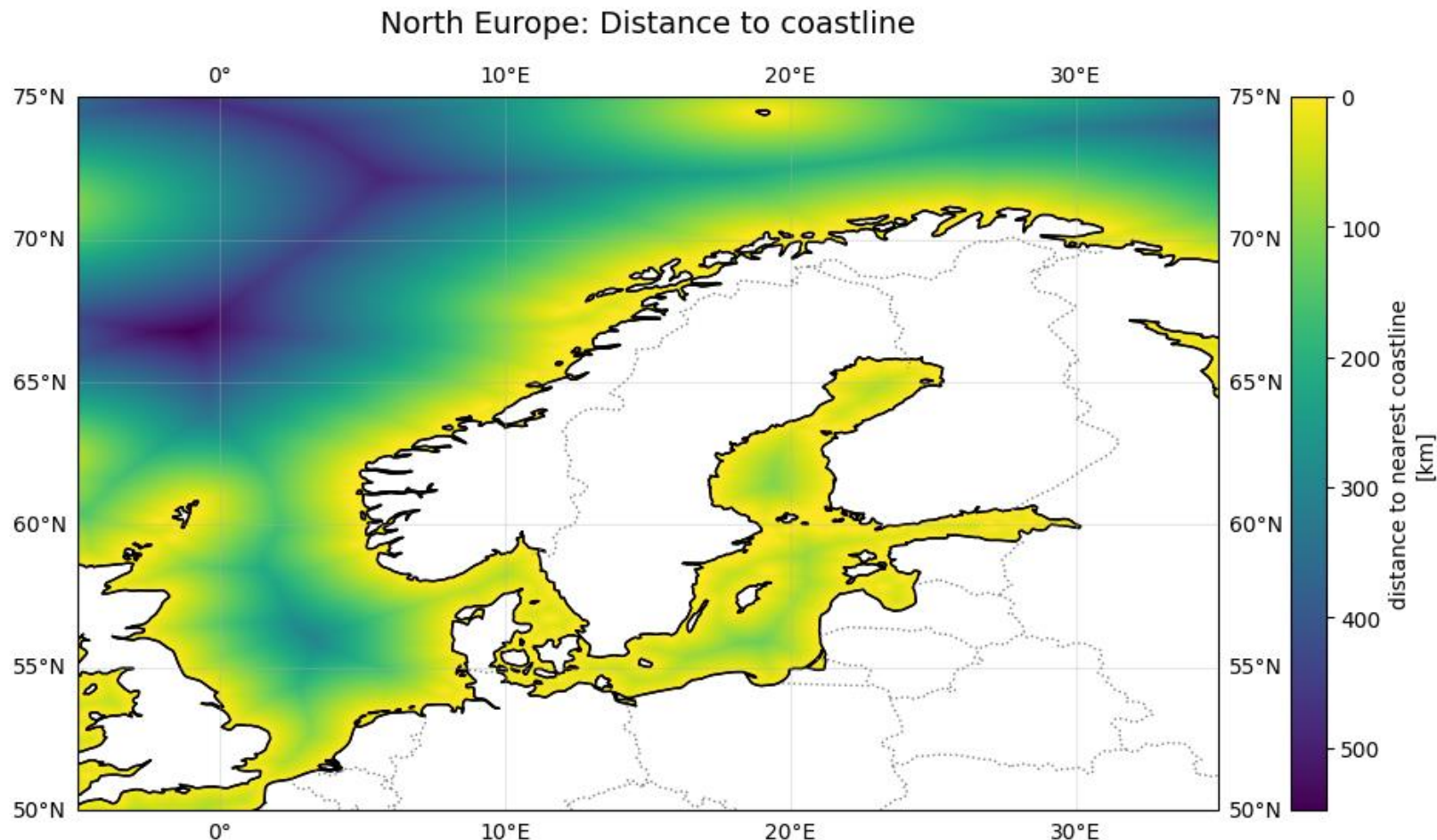
## Interpolácia chýbajúcich hodnôt:

- „**nearest**“ – vezmem hodnotu suseda
  - nefungujúca stratégia keď chýba naraz veľa okolitých hodnôt
- „**linear**“ - hodnotu vypočítam ako priemer viaceru susedov
  - fungujúca stratégia - podľa ocean masky



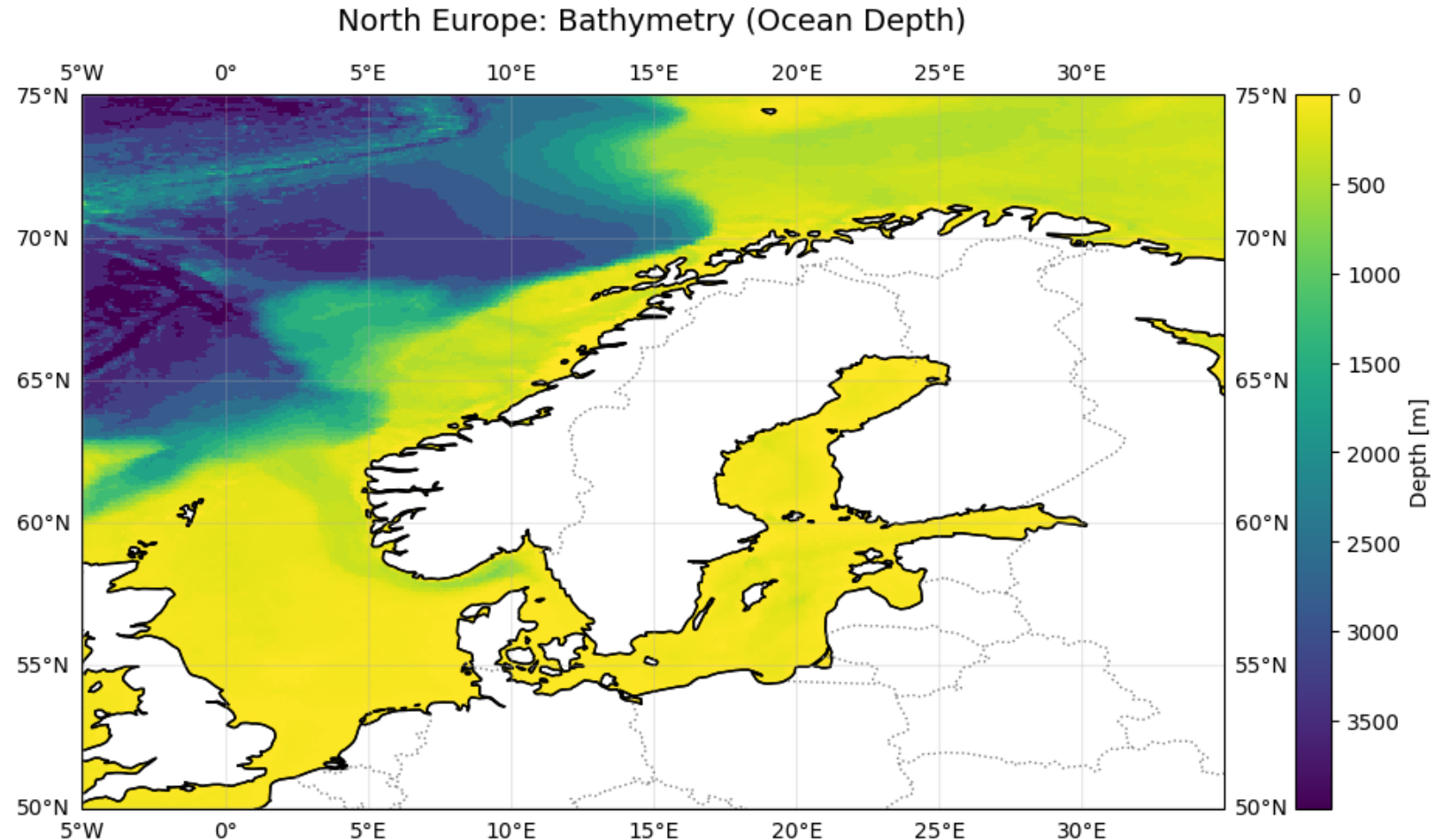
# Vzdialenosť od pobrežia

- **Vzdialenosť konkrétneho bodu na mape od najbližšieho brehu**
- Identifikácia plytkých pobrežných lovísk a hlbokého mora
- Rozlíšenie medzi rôznymi druhmi rýb podľa ich výskytu



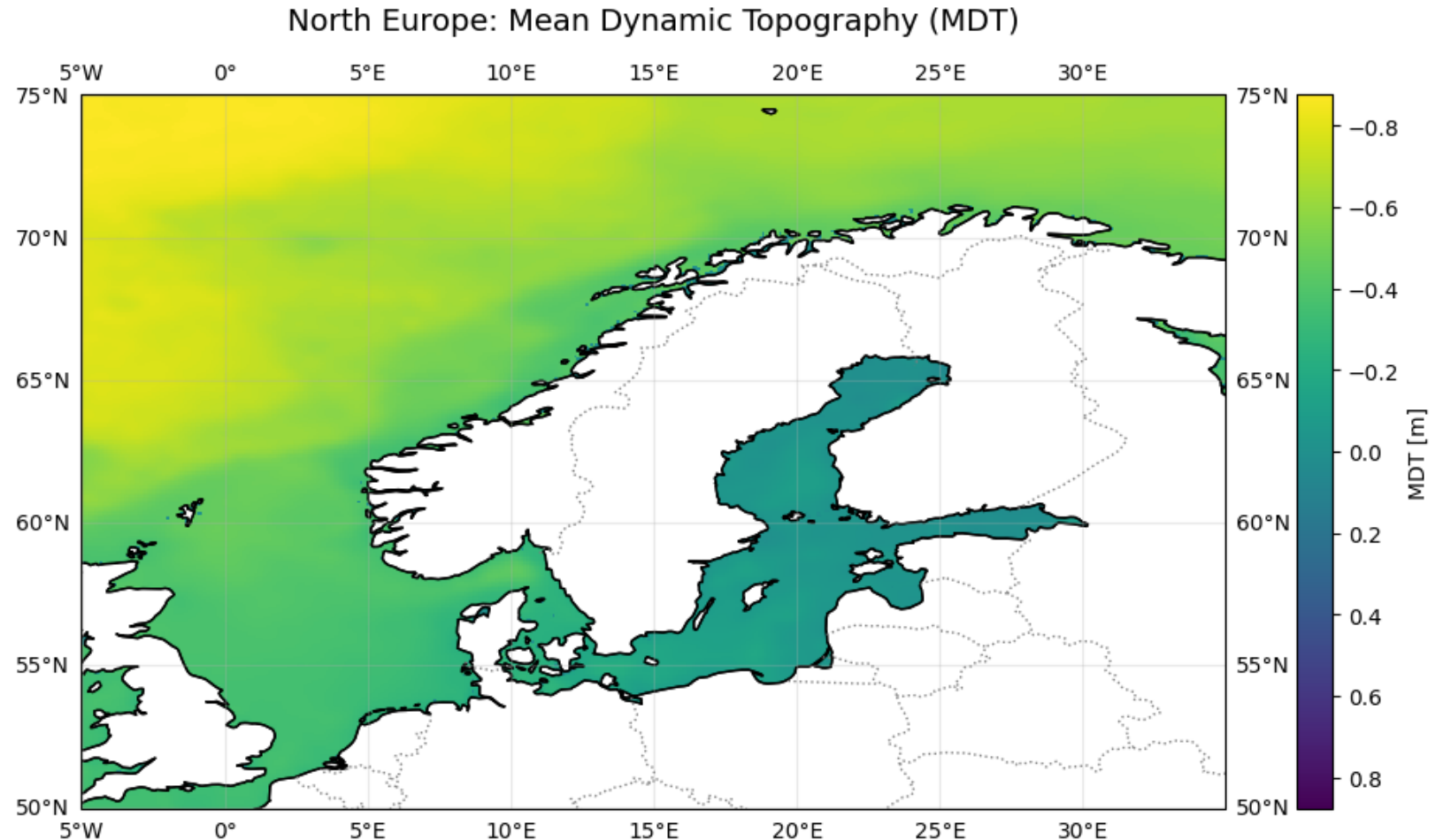
# Bathymetria

- Údaj o tom **ako hlboko sa nachádzaorské dno pod hladinou**
- Kľúčový faktor pre výskyt rýb žijúcich pri dne
- Súvislosť s **teplotou** a tlakom vody v rôznych vrstvách



# Topografia

- **Tvar a nerovnosti povrchu morského dna pod vodou**
- Vplyv na smer prúdenia vody a vznik morských vírov
- **Vynášanie živín** z hlbín smerom nahor pri náraze na terén
- Vytváranie prirodzených bariér a úkrytov pre morské živočíchy



# Rekapitulácia dát

**catches**

denné úlovky rýb (kg)

**VMS data**

týždenné údaje o pohybe lodí

**ocean data**

týždenné hodnoty slanosti,  
teploty a chlorofylu vody

**distance to coast**

vzialenosť od pobrežia (km)

**bathymetria**

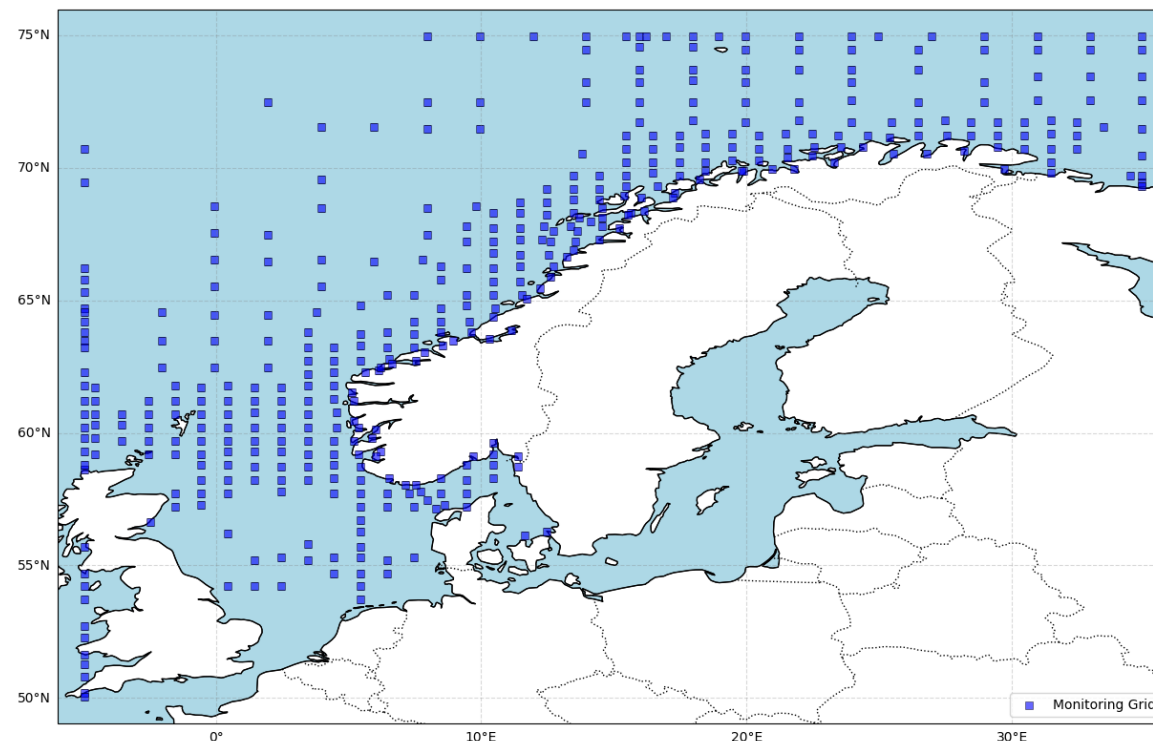
hĺbka mora (m)

**topografia**

typ reliéfu morského dna (m)

## Oblasti o ktorých máme všetky dáta

2023 Grid structure (Number of unique grids: 429)



Definované pomocou (**grid\_lon, grid\_lat**) a (**grid\_i, grid\_j**)

# Vytvorenie datasetu (sliding windows)

GRID	WINDOW START	DAYS	SAL	SST	CHL	VMS	DIST, BATHY, TOPO,	CATCH NEXT WEEK
grid_i	date	day_1		week_1		boat_count w1-w4	distance_to_coast	xxx kg
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

**69 stĺpcov – atribútov v datasete**

grid_i	grid_j	grid_lon	grid_lat	window_start_day	window_start_month	catch_next_week	day_1	day_2	day_3	day_4	day_5	day_6	day_7	day_8	day_9	day_10	day_11	
0	225	233	14.458344	68.791672	2023-01-01	1	79618	0	0	41	3462	897	29978	446	26881	3935	1287	1281
1	225	233	14.458344	68.791672	2023-01-02	1	25472	0	41	3462	897	29978	446	26881	3935	1287	1281	100
2	225	233	14.458344	68.791672	2023-01-03	1	24397	41	3462	897	29978	446	26881	3935	1287	1281	100	2102
3	225	233	14.458344	68.791672	2023-01-04	1	19959	3462	897	29978	446	26881	3935	1287	1281	100	2102	0
4	225	233	14.458344	68.791672	2023-01-05	1	17168	897	29978	446	26881	3935	1287	1281	100	2102	0	64840

GRID	WINDOW START	DAYS	SAL	SST	CHL	VMS	DIST, BATHY, TOPO,	CATCH NEXT WEEK	
grid_i	date	day_1	week_1			boat_count w1-w4		distance_to_coast	xxx kg
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	

**69 stĺpcov – atribútov v datasete**

grid_i	grid_j	grid_lon	grid_lat	window_start_day	window_start_month	catch_next_week	day_1	day_2	day_3	day_4	day_5	day_6	day_7	day_8	day_9	day_10	day_11	
0	225	233	14.458344	68.791672	2023-01-01	1	79618	0	0	41	3462	897	29978	446	26881	3935	1287	1281
1	225	233	14.458344	68.791672	2023-01-02	1	25472	0	41	3462	897	29978	446	26881	3935	1287	1281	100

day_12	day_13	day_14	day_15	day_16	day_17	day_18	day_19	day_20	day_21	day_22	day_23	day_24	day_25	day_26	day_27	day_28	chl_week_1	sst_week_1	sal_week_1
100	2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	0.062157	5.847397	33.033215
2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	54146	0.062157	5.847397	33.033215
0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	54146	1075	0.062157	5.847397	33.033215
64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	54146	1075	4438	0.062157	5.847397	33.033215
1241	413	42	93477	0	0	0	0	0	37836	452	0	0	54146	1075	4438	2791	0.062157	5.847397	33.033215

grid_i	date	day_1	week_1			boat_count w1-w4	distance_to_coast	xxx kg
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

**69 stĺpcov – atribútov v datasete**

grid_i	grid_j	grid_lon	grid_lat	window_start_day	window_start_month	catch_next_week	day_1	day_2	day_3	day_4	day_5	day_6	day_7	day_8	day_9	day_10	day_11	
0	225	233	14.458344	68.791672	2023-01-01	1	79618	0	0	41	3462	897	29978	446	26881	3935	1287	1281
1	225	233	14.458344	68.791672	2023-01-02	1	25472	0	41	3462	897	29978	446	26881	3935	1287	1281	100

day_12	day_13	day_14	day_15	day_16	day_17	day_18	day_19	day_20	day_21	day_22	day_23	day_24	day_25	day_26	day_27	day_28	chl_week_1	sst_week_1	sal_week_1
100	2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	0.062157	5.847397	33.033215
2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	54146	0.062157	5.847397	33.033215

chl_week_2	sst_week_2	sal_week_2	chl_week_3	sst_week_3	sal_week_3	chl_week_4	sst_week_4	sal_week_4	chl_week_5	sst_week_5	sal_week_5	vms_boat_count_w1
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	1152.0
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	978.0
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	978.0
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	978.0
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	978.0

grid_lon	grid_lat	day_28	week_3			avg_speed w1-w4	topography	
grid_lon	grid_lat	day_28	week_4			slow_speed_ratio w1-w4	topography	
grid_lon	grid_lat	day_28	week_5 (next week values)			slow_speed_ratio w1-w4	topography	
4	2	28	5	5	5	16	3	1

**69 stĺpcov – atribútov v datasete**

grid_i	grid_j	grid_lon	grid_lat	window_start_day	window_start_month	catch_next_week	day_1	day_2	day_3	day_4	day_5	day_6	day_7	day_8	day_9	day_10	day_11	
0	225	233	14.458344	68.791672	2023-01-01	1	79618	0	0	41	3462	897	29978	446	26881	3935	1287	1281
1	225	233	14.458344	68.791672	2023-01-02	1	25472	0	41	3462	897	29978	446	26881	3935	1287	1281	100

day_12	day_13	day_14	day_15	day_16	day_17	day_18	day_19	day_20	day_21	day_22	day_23	day_24	day_25	day_26	day_27	day_28	chl_week_1	sst_week_1	sal_week_1
100	2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	0.062157	5.847397	33.033215
2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	54146	0.062157	5.847397	33.033215

chl_week_2	sst_week_2	sal_week_2	chl_week_3	sst_week_3	sal_week_3	chl_week_4	sst_week_4	sal_week_4	chl_week_5	sst_week_5	sal_week_5	vms_boat_count_w1
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	1152.0
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	978.0

unique_boats_w1	vms_avg_speed_w1	vms_slow_speed_w1	vms_boat_count_w2	vms_unique_boats_w2	vms_avg_speed_w2	vms_slow_speed_w2	vms_boat_count_w3	vms_unique_boats_w3
1.0	0.006632	1.0	978.0	1.0	0.015951	1.0	892.0	4.0
1.0	0.015951	1.0	892.0	4.0	0.046121	1.0	967.0	3.0
1.0	0.015951	1.0	892.0	4.0	0.046121	1.0	967.0	3.0
1.0	0.015951	1.0	892.0	4.0	0.046121	1.0	967.0	3.0
1.0	0.015951	1.0	892.0	4.0	0.046121	1.0	967.0	3.0

			week_5 (next week values)					
4	2	28	5	5	5	16	3	1

**69 stĺpcov – atribútov v datasete**

grid_i	grid_j	grid_lon	grid_lat	window_start_day	window_start_month	catch_next_week	day_1	day_2	day_3	day_4	day_5	day_6	day_7	day_8	day_9	day_10	day_11	
0	225	233	14.458344	68.791672	2023-01-01	1	79618	0	0	41	3462	897	29978	446	26881	3935	1287	1281
1	225	233	14.458344	68.791672	2023-01-02	1	25472	0	41	3462	897	29978	446	26881	3935	1287	1281	100

day_12	day_13	day_14	day_15	day_16	day_17	day_18	day_19	day_20	day_21	day_22	day_23	day_24	day_25	day_26	day_27	day_28	chl_week_1	sst_week_1	sal_week_1
100	2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	0.062157	5.847397	33.033215
2102	0	64840	1241	413	42	93477	0	0	0	0	0	37836	452	0	0	54146	0.062157	5.847397	33.033215

chl_week_2	sst_week_2	sal_week_2	chl_week_3	sst_week_3	sal_week_3	chl_week_4	sst_week_4	sal_week_4	chl_week_5	sst_week_5	sal_week_5	vms_boat_count_w1
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	1152.0
0.057779	5.53537	33.143635	0.057307	6.085476	33.526974	0.079338	5.250828	33.244289	0.090782	5.935687	33.694298	978.0

unique_boats_w1	vms_avg_speed_w1	vms_slow_speed_w1	vms_boat_count_w2	vms_unique_boats_w2	vms_avg_speed_w2	vms_slow_speed_w2	vms_boat_count_w3	vms_unique_boats_w3
1.0	0.006632	1.0	978.0	1.0	0.015951	1.0	892.0	4.0
1.0	0.015951	1.0	892.0	4.0	0.046121	1.0	967.0	3.0

vms_avg_speed_w3	vms_slow_speed_w3	vms_boat_count_w4	vms_unique_boats_w4	vms_avg_speed_w4	vms_slow_speed_w4	bathymetry	mdt	distance_to_coast
0.046121	1.0	967.0	3.0	0.019473	1.0	130.666	-0.397951	1.0
0.019473	1.0	971.0	2.0	0.008651	1.0	130.666	-0.397951	1.0
0.019473	1.0	971.0	2.0	0.008651	1.0	130.666	-0.397951	1.0
0.019473	1.0	971.0	2.0	0.008651	1.0	130.666	-0.397951	1.0
0.019473	1.0	971.0	2.0	0.008651	1.0	130.666	-0.397951	1.0



# Sliding windows dataset

- Shape 2023: (141 999, 69)
- Shape 2024: (141 999, 69)
- Shape 2025: (141 006, 69)
  - menej úlovkov, dáta sa na CMS stránke dodatočne dopĺňajú

Rozbor pre rok 2023:

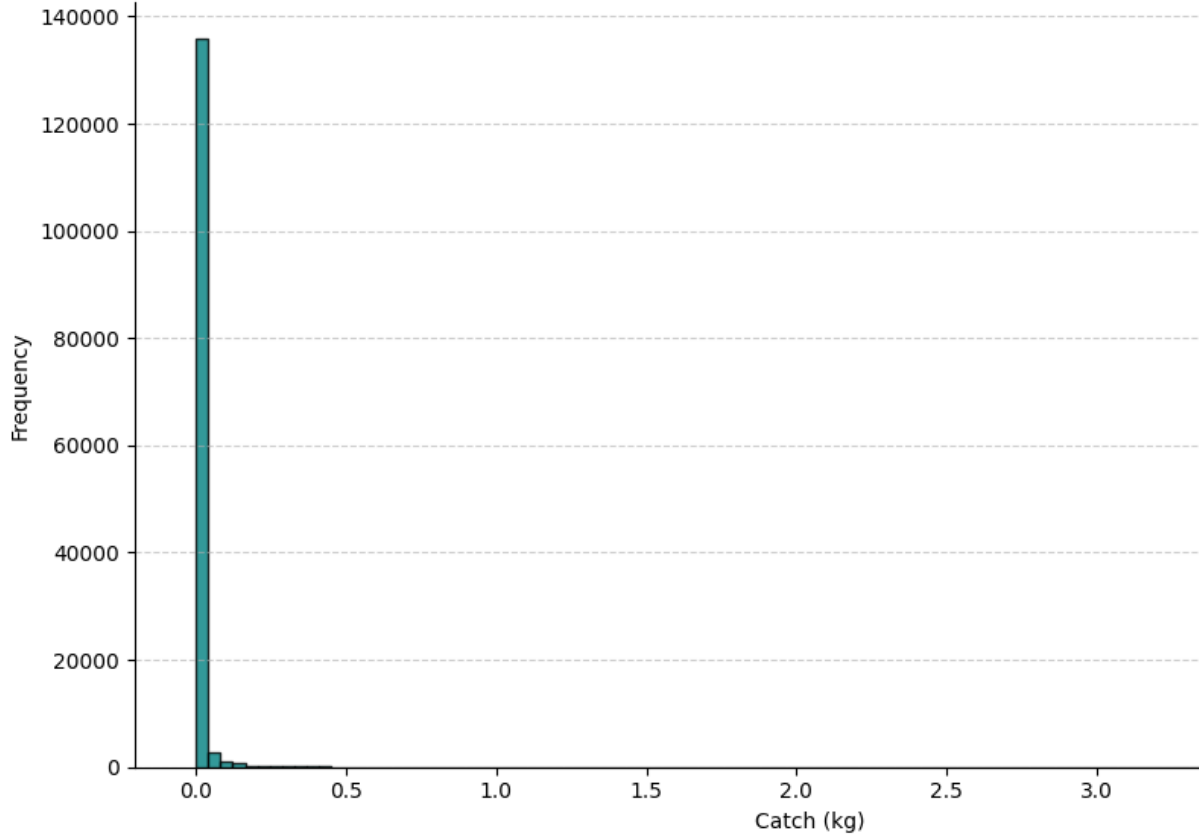
- **331** sliding windows z jednej GPS oblasti (posledný window začína 27.11.)
- **429** GPS oblastí, kde boli zaznamenané úlovky

$$331 \times 429 = 141\ 999$$

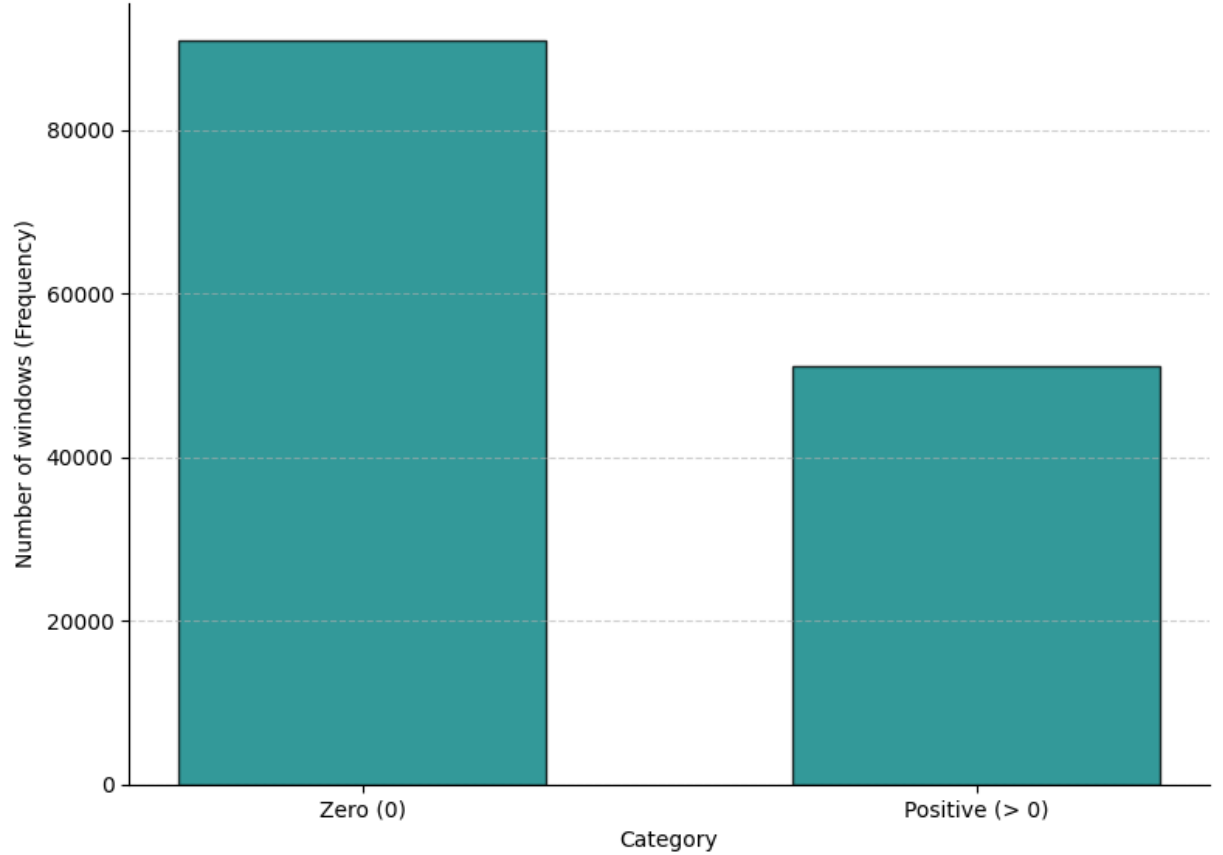
# catch\_next\_week

```
Distribúcia pre catch_next_week:  
Trieda 0 (Nulový úlovok): 90917  
Trieda 1 (Pozitívny úlovok): 51082
```

2023: Catch next week histogram



2023: Catch: Zero vs Non-Zero

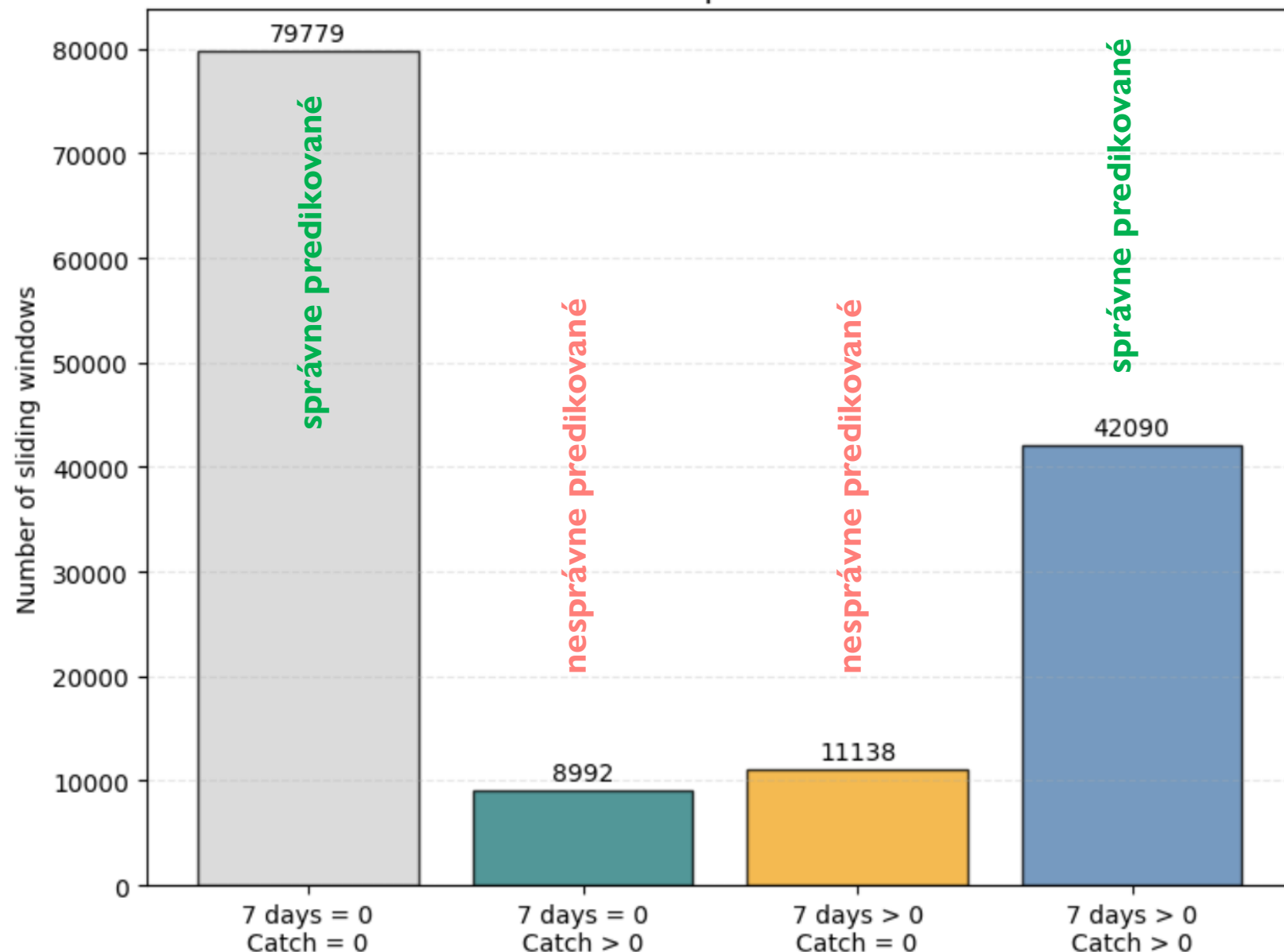


- Počítame ako súčet úlovkoch nasledujúcich 7 dní od posledného dňa (28.) v sliding window

# Baseline model

- „ľudské zhodnotenie“
- Pozriem si úlovky **7 dní dozadu**, ak tam bol aspoň jeden úlovok  $> 0$ , bude aj teraz, inak nie
- Úspešnosť približne **86%**
- Chceme ju prekonať a nájsť závislosti v dátach, ktoré o úlovku najviac rozhodujú

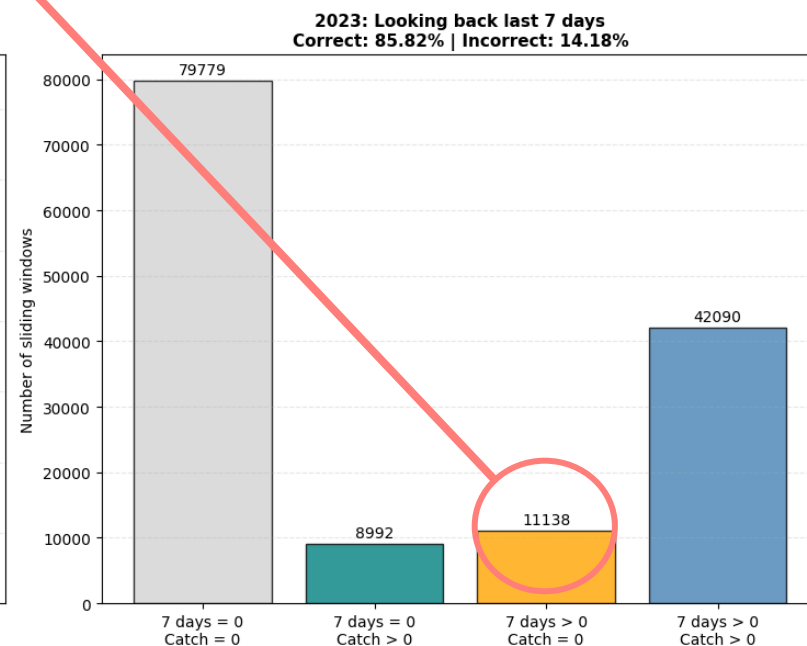
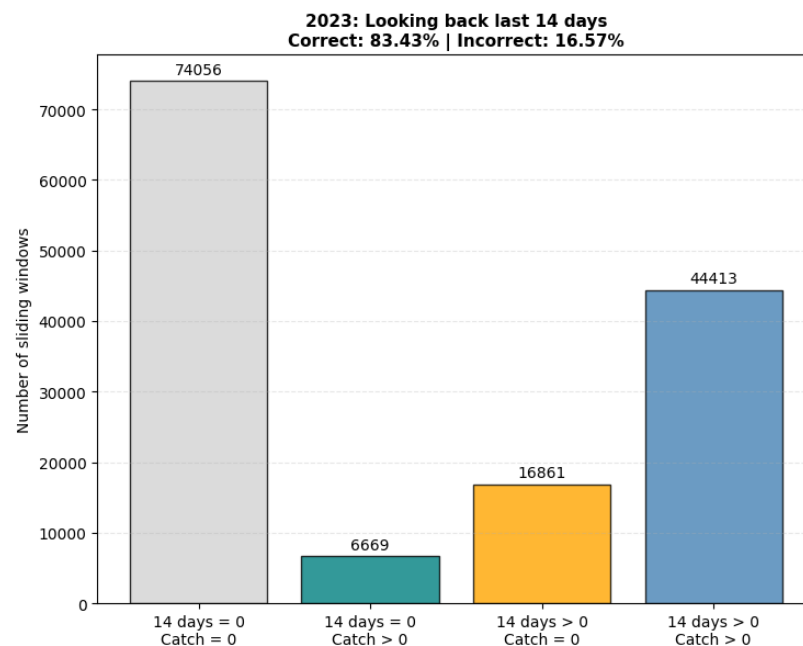
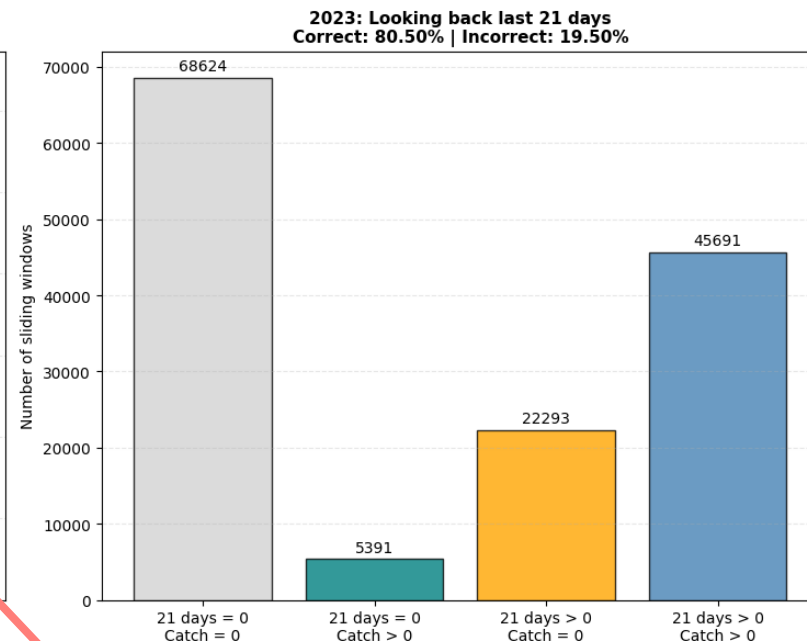
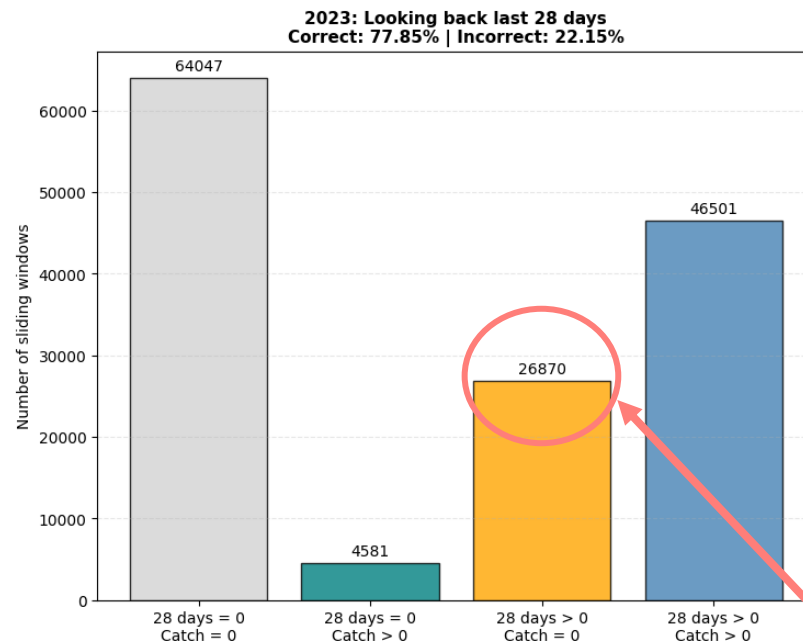
2023: Looking back last 7 days  
Correct: 85.82% | Incorrect: 14.18%



# Baseline model

Zaujímavosť:

- Ak zhodnotíme viac dní z minulosti, napr. až 28, tak celková úspešnosť tohto prístupu klesá
- o cca **10%** vieme správnejšie predpovedať ak úlovok bol aj bude (modrá)
- Avšak predpoveď ak úlovok bol a o týždeň už nebude sa **zhorší výrazne o 141%** (žltá)



# Klasifikácia (bude alebo nebude o týždeň úlovok?)

## Binárna:

- 0: žiaden úlovok
- 1: úlovok  $> 0.0$  kg

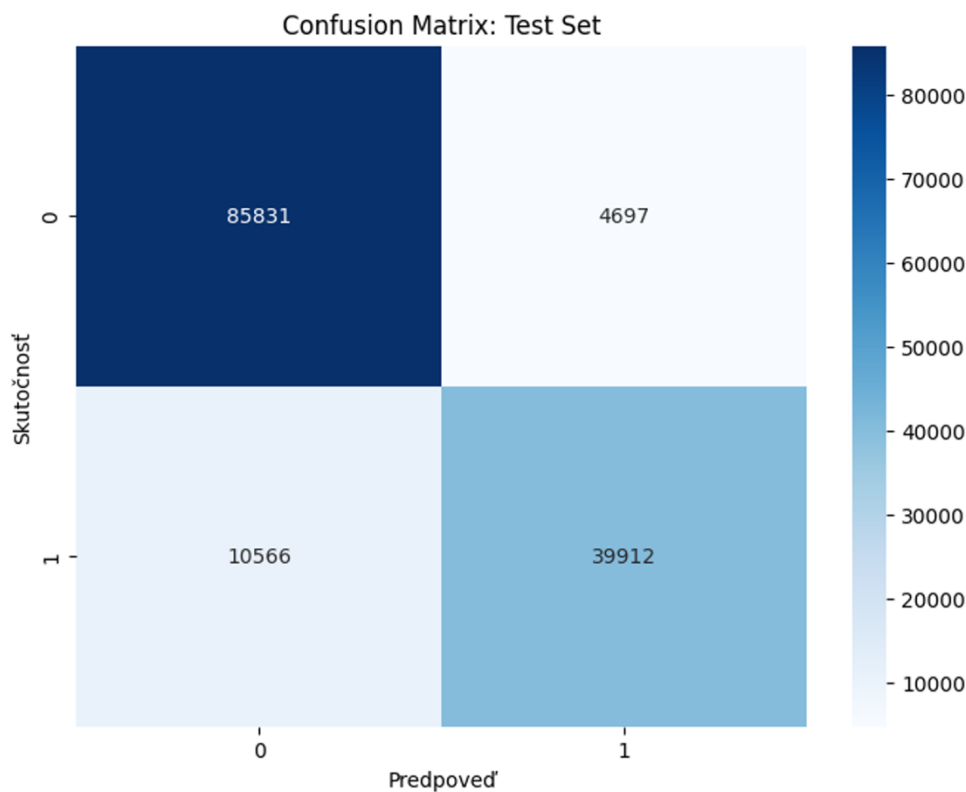
```
Distribúcia pre catch_next_week:  
Trieda 0 (Nulový úlovok): 90917  
Trieda 1 (Pozitívny úlovok): 51082
```

## Terciárna:

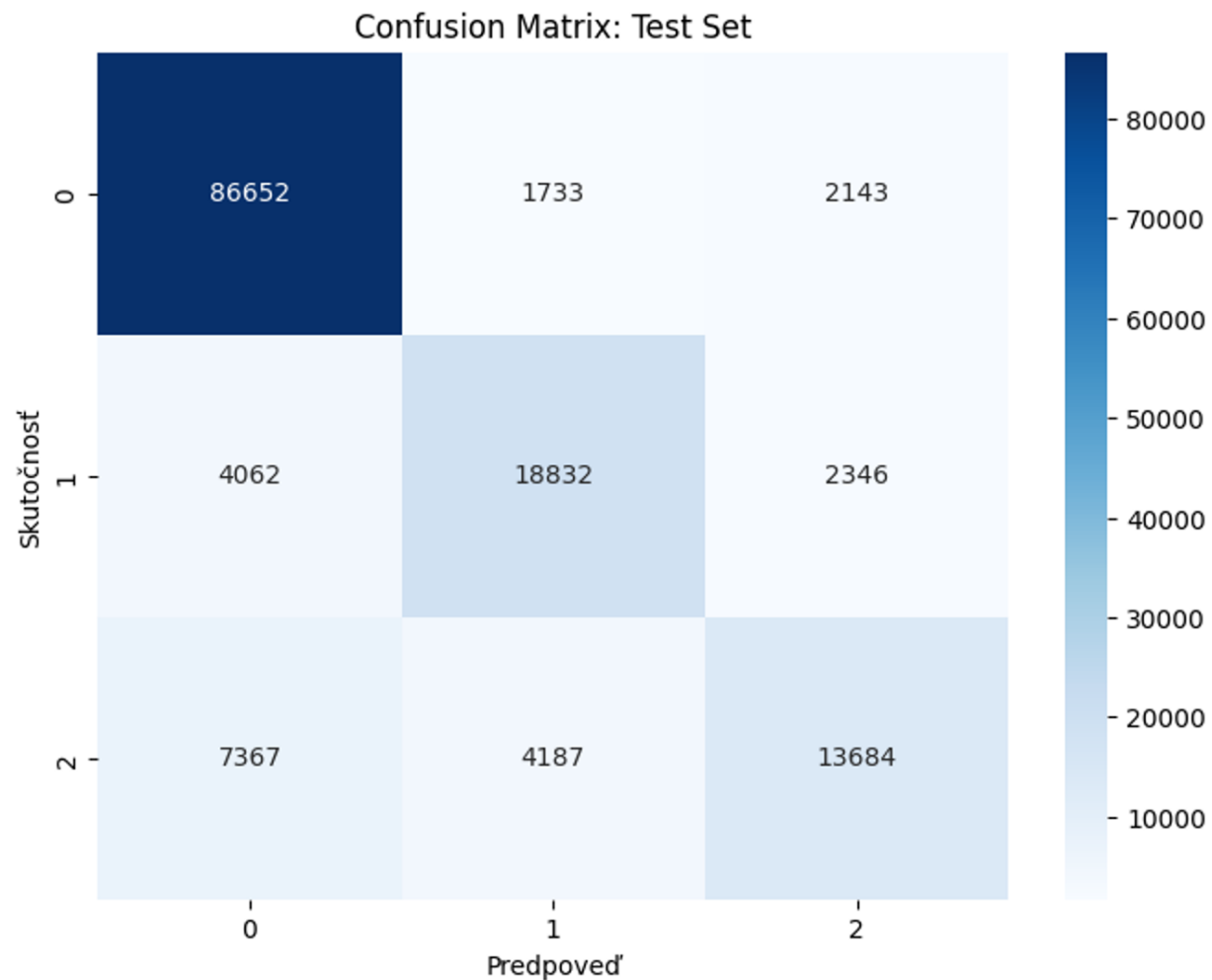
- 0: žiaden úlovok
- 1: menší úlovok až stredný
- 2: stredný úlovok až väčší

```
Hranica pre 'Jackpot' (Medián nenulových): 20922.50 kg  
CNWB_012  
0    90917  
1    25541  
2    25541
```

# XGBoost model



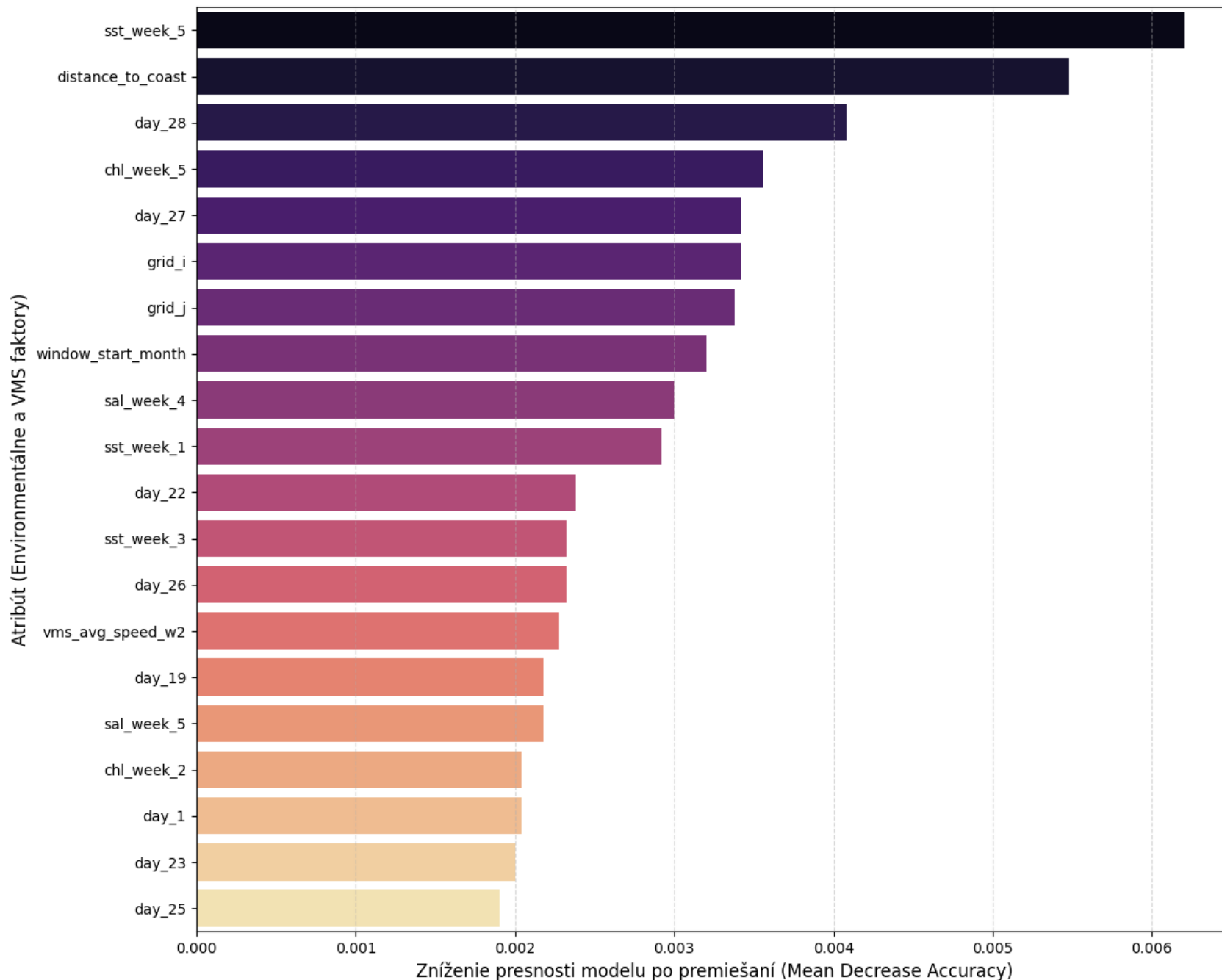
Binárna klasifikácia:  
Zatiaľ dosiahnutá accuracy **89%**



Terciárna klasifikácia:  
Zatiaľ dosiahnutá accuracy **84%**

**train na 2023 + 2024 a test na 2025**

Top 20 Atribútov: Permutation Importance (Test Set 2025)



3 atribúty ktoré majú pravdepodobne najväčší vplyv na úlovok budúci týždeň:

- **teplota v predpovedanom týždni**
- **vzdialenosť od pobrežia**
- **úlovok v 28. dni**

# Nasledujúce kroky

- Vylepšiť parametre **XGBoost modelu**
- Skúsiť **iné modely** a porovnať (RandomForest, LightGBM...)
- Spustiť modely na **dátach bez:**
  - VMS dát
  - oceánskych vlastností
  - bathymetrie, topografie, vzdialenosti od pobrežia

---

# ĎAKUJEM ZA POZORNOST

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

Konzultant: Mariana Zapotoková, Vissim

**Bc. Daniela Pillárová**

Diplomový seminár, 26.3.2026

