MetalFin

An AI model for predicting heavy metal concentration in fish using machine learning.

By: Nate Oberdier

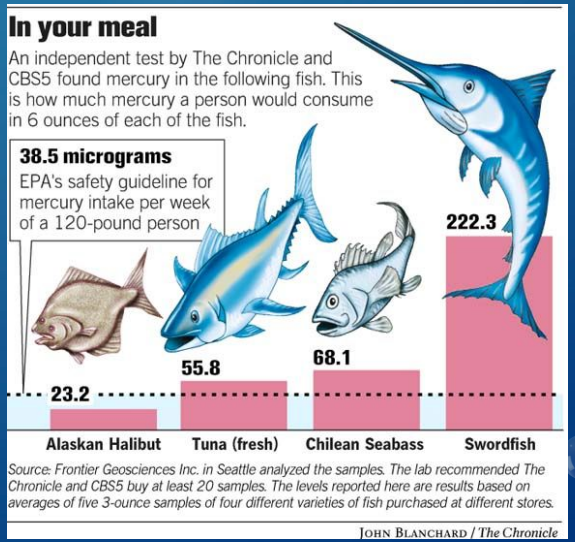
Problem Statement

The Problem:

- Heavy-metal contamination varies drastically by species and region
- Public datasets are incomplete and inconsistent
- Consumers receive generalized warnings instead of specific info

Why is it important?

- Helps consumers avoid unnecessary fear
- Provides a better understanding to heavy metal absorption
- Provides safer and more informed dietary selections



Introduction

What is MetalFin?

- Predicts heavy metals in a wide variety of seafoods
  - Mercury, Lead, Arsenic, and Cadmium
- Uses species traits, ecology, and location data
- Provides risk categories based on amounts of contamination

Interactive AI decision tool with multi-fish comparison (raw vs cooked, lb/kg/g, absorbed dose)

Species: Swordfish | Raised: wild | Prep: cooked | Amount: 1 | Unit: lb | + | Clear

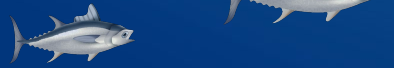
```
====================================
Species: Swordfish (wild, cooked)
Amount per week: 1.00 lb (~453.6 g)
------------------------------------
Model-predicted MAX metal (mg/kg): 0.3385

Estimated per-metal absorption and intake:
  Mercury  | conc ≈  0.3385 mg/kg | absorbed ≈    141.3 µg (92.0% of    153.5 µg)
  Lead     | conc ≈  0.0090 mg/kg | absorbed ≈      0.4 µg (10.0% of      4.1 µg)
  Arsenic  | conc ≈  0.0666 mg/kg | absorbed ≈     28.7 µg (95.0% of     30.2 µg)
  Cadmium  | conc ≈  0.0002 mg/kg | absorbed ≈      0.0 µg ( 5.0% of      0.1 µg)

Overall safety classification (based on absorbed load):
  >>> DANGEROUS <<<

Margin to change classification (approx):
  (Driven mostly by Mercury)
  • Remove ≈ 0.21 lb to drop to RISKY.
  • Remove ≈ 0.60 lb to drop to SAFE.
====================================
=== Comparison Table (absorbed µg per week) ===
```

	Species	Raised	Prep	Amount	Hg_abs_ug	Pb_abs_ug	As_abs_ug	Cd_abs_ug	Max_metal_mg_kg	Safety
0	Swordfish	wild	raw	1.00 lb	12.282908	0.408162	28.701771	0.005236	0.33849	SAFE
1	Swordfish	wild	cooked	1.00 lb	141.253448	0.408162	28.701771	0.005236	0.33849	DANGEROUS

Data

1. **EPA Fish Tissue Contamination Dataset**
2. **Seafood Mercury Database**
3. **Peer-Reviewed Scientific Datasets**
4. **Synthetic Expansion for Missing Traits**

Combined into a master dataset with 1000 individual entries across 54 unique species



Methods Pt. 1

Regression Model

- Random Forest Regressor
- Predicts and assigns concentration values
  - Hg, Pb, As, Cd in mg/kg
- Compares to FDA/EPA safety standards

Classification Model

- Random Forest Classifier
- Predicts and classifies risk categories
  - Low, Medium, and High
- Each group has an equal amount of species with fair distribution

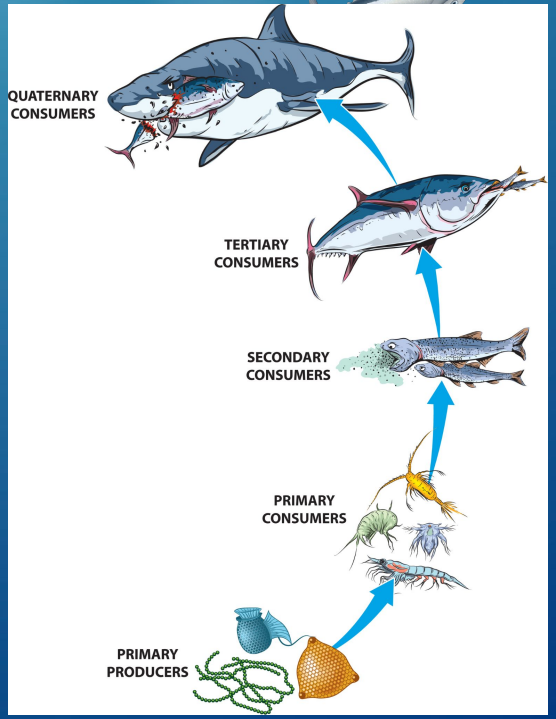
Methods Pt. 2

Key input features used for all models:

- Species
- Location (region/sea/ocean)
- Habitat (marine, freshwater, brackish)
- Diet Type (carnivore, omnivore, filter-feeder, etc.)
- Length (cm)
- Weight (g)
- Ecological traits (trophic level, feeding behavior)

Outcome:

- Contamination highly correlated with trophic level
- Larger species accumulate more metals
- Predators have higher levels of contamination



Methods Pt. 3

Data Processing:

- Unit standardization (wet weight vs. dry weight harmonized)
- Outlier detection and removal
- Handling missing values (imputation for ecological traits)
- Normalization of numerical features
- One-hot encoding of categorical variables
- Train/validation/test split: 70/15/15

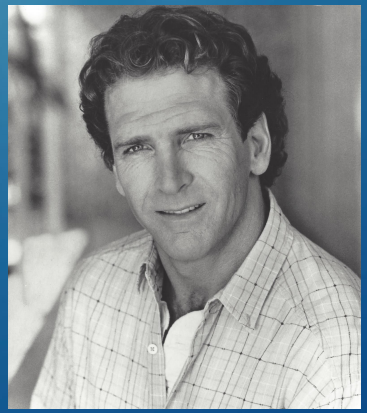
Evaluation Metrics:

- Regression:
  - R^2 score
  - Mean Absolute Error (MAE)
  - Root Mean Squared Error (RMSE)
- Classification:
  - Accuracy
  - Precision & recall
  - Confusion matrix (important for Medium risk overlap)

Methods Pt. 4

“Mercury In Fish; Do We Absorb It?”

- Blog post published by nutritionist Aajonus Vonderplanitz
- Study conducted in 1989 regarding mercury absorption in swordfish
- 8 dogs, 8 cats, equally tested between raw and cooked variants
- After 7 days, the raw group had discarded 12.25-13.03 µg/g of mercury
- In comparison, the cooked group discarded 1.1-1.64 µg/g of mercury
- The raw group only absorbed 2-8% of all mercury
- The cooked group absorbed 88-92% of all mercury
- Cooking swordfish causes mercury contamination



Uses:

- Set raw variants of seafood to only absorb 8% of mercury content
- Set cooked variants to absorb 92% of mercury content
- Somewhat experimental and controversial
- MetalFin does not advertise the consumption of raw seafood

Results

Regression Performance:

- Validation R^2 Score: 0.6384
- Test R^2 Score: 0.5591

Classification Performance:

- High accuracy for Low and High risk
- Some overlap in Medium risk (expected due to natural variability)

Absorption Rates:

- Mercury - 8%(raw) & 92%(cooked)
- Lead - 10%
- Arsenic - 95%
- Cadmium - 5%

MetalFin - Demo

Model Insights

Observations:

- Only a few species consistently show high contamination
- Many “dangerous” seafood species are actually low risk
- Ecological features alone reveal predictable patterns
- Raw seafood preparation had significantly lower amounts of absorbable mercury
- Arsenic was more threatening than mercury in raw forms, especially in swordfish
    - Despite swordfish containing 5x more mercury than arsenic!

=== Comparison Table (absorbed µg per week) ===

	Species	Raised	Prep	Amount	Hg_abs_ug	Pb_abs_ug	As_abs_ug	Cd_abs_ug	Max_metal_mg_kg	Safety
0	Shark	wild	raw	3.00 lb	21.590725	0.495142	35.05234	0.008521	0.198331	SAFE
1	Shark	wild	cooked	3.00 lb	248.293340	0.495142	35.05234	0.008521	0.198331	DANGEROUS

Limitations & Future Work

Limitations:

- Uneven sample sizes for some species
- Regional bias
- Medium-risk category overlap
- Minimal absorption rate information
    - Especially mercury & arsenic

Future Work:

- Add environmental pollution variables
- Build a public website
- Expand datasets
- Improve bioavailability modeling

THANKS!

Does anyone have any questions?
noberdier@oakland.edu