



**HAL**  
open science

## Using machine learning to predict drownings in surf beaches of southwest France

David Carayon, Bruno Castelle, Eric Tellier, Bruno Simmonet, Jeffrey Dehez

► **To cite this version:**

David Carayon, Bruno Castelle, Eric Tellier, Bruno Simmonet, Jeffrey Dehez. Using machine learning to predict drownings in surf beaches of southwest France. World Conference on Drowning Prevention, Surf Life Saving Australia; Life Saving Society Australia, Dec 2023, Perth (Australia), Australia. hal-04342633

**HAL Id: hal-04342633**

**<https://hal.inrae.fr/hal-04342633>**

Submitted on 13 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open licence - etalab



World Conference on  
Drowning Prevention



# Using machine learning to predict drownings in surf beaches of southwest France

**David**   
**Carayon**<sup>1</sup>

<sup>1</sup>INRAE | ETTIS Research  
Unit

**Bruno**   
**Castelle**<sup>2</sup>

<sup>2</sup>CNRS | EPOC Research  
Unit

**Eric**   
**Tellier**<sup>3</sup>

<sup>3</sup>INSERM | AheaD Research  
Unit

**Bruno**   
**Simmonet**<sup>4</sup>

<sup>4</sup>Bordeaux University  
Hospital

**Jeoffrey**   
**Dehez**<sup>1</sup>

<sup>1</sup>INRAE | ETTIS Research  
Unit

# Some context

*Should France be your next vacation destination 🌞 ?*

# Southwest France : The place to be 😊



OCEAN ATLANTIQUE  
Côte d'Argent

Saint

Vieux-Bou



Anglet, Adour

# Some (scary 🤨) numbers



Castelle et al. (2019)

- One of the most dangerous coasts in the world ([Castelle et al. 2019](#)) : heavy rip current (Baïnes), shore break
- Surveillance mainly during summer : Thousands of rescues each year
- 20 to 30 fatal drownings each year (👊0 between the flags)
- About 1/2 of swimmers bathe outside of the surveillance zone ([Dehez and Lyser 2021](#))

Can we prevent  
drownings ?



Can we prevent  
**predict** drownings ?



# Goal of this work

- Same philosophy as previous work (**Tellier et al. 2021**) : Daily emergency calls prediction based on weather and beach crowd
- Better (and cleaner) data
- New statistical methods
- **The SWYM Project**





# Predicting drownings using Machine Learning



# Predictors and outcome

$$\text{Risk} = \text{Hazard} + \text{Exposure}$$



## OUTCOME

- Water inhalation & respiratory impairments which lead to emergency calls

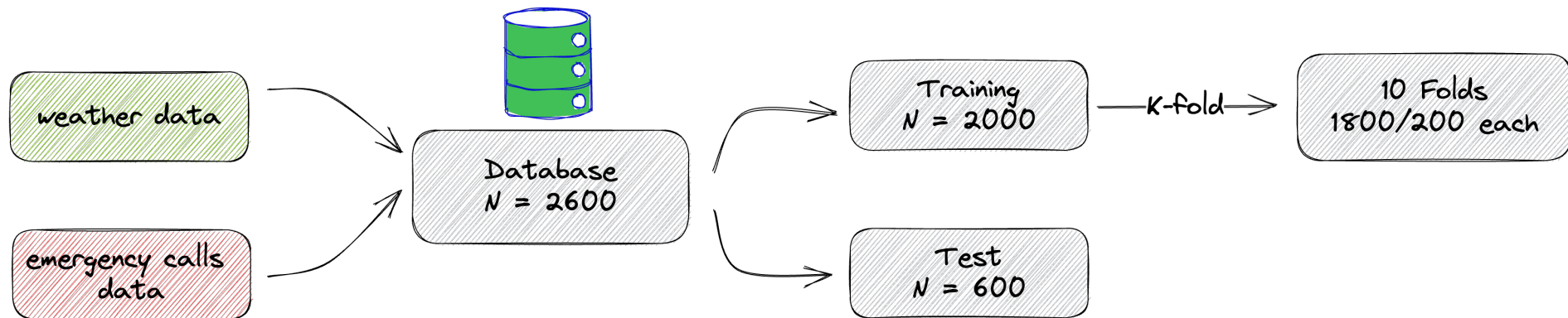
*Binary daily data  
2011-2022*



## PREDICTORS

- Wave Incidence Factor (cos4h)
- Wave Factor (HsTp)
- Air temperature
- Day (ex : 6 for 6th of July)
- Month
- Wday (ex : 1 for Monday)

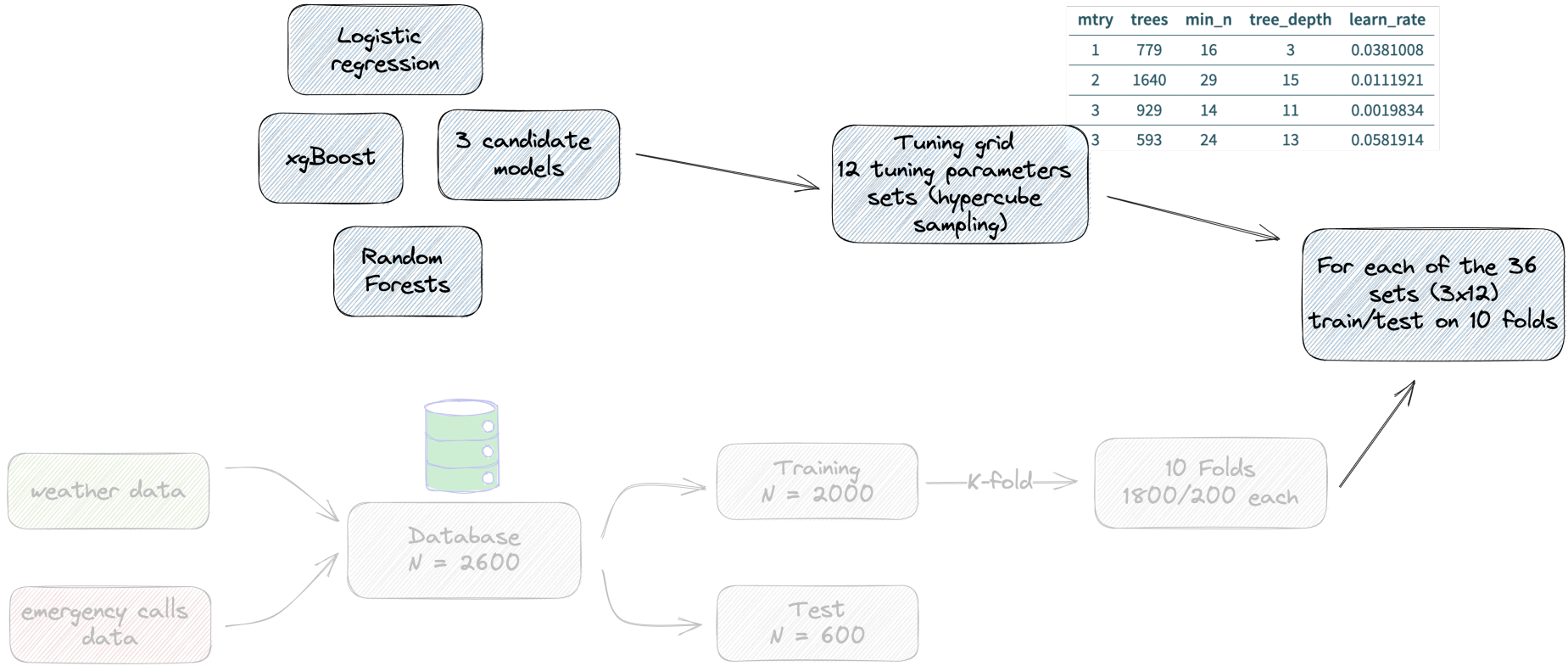
# Modelling workflow



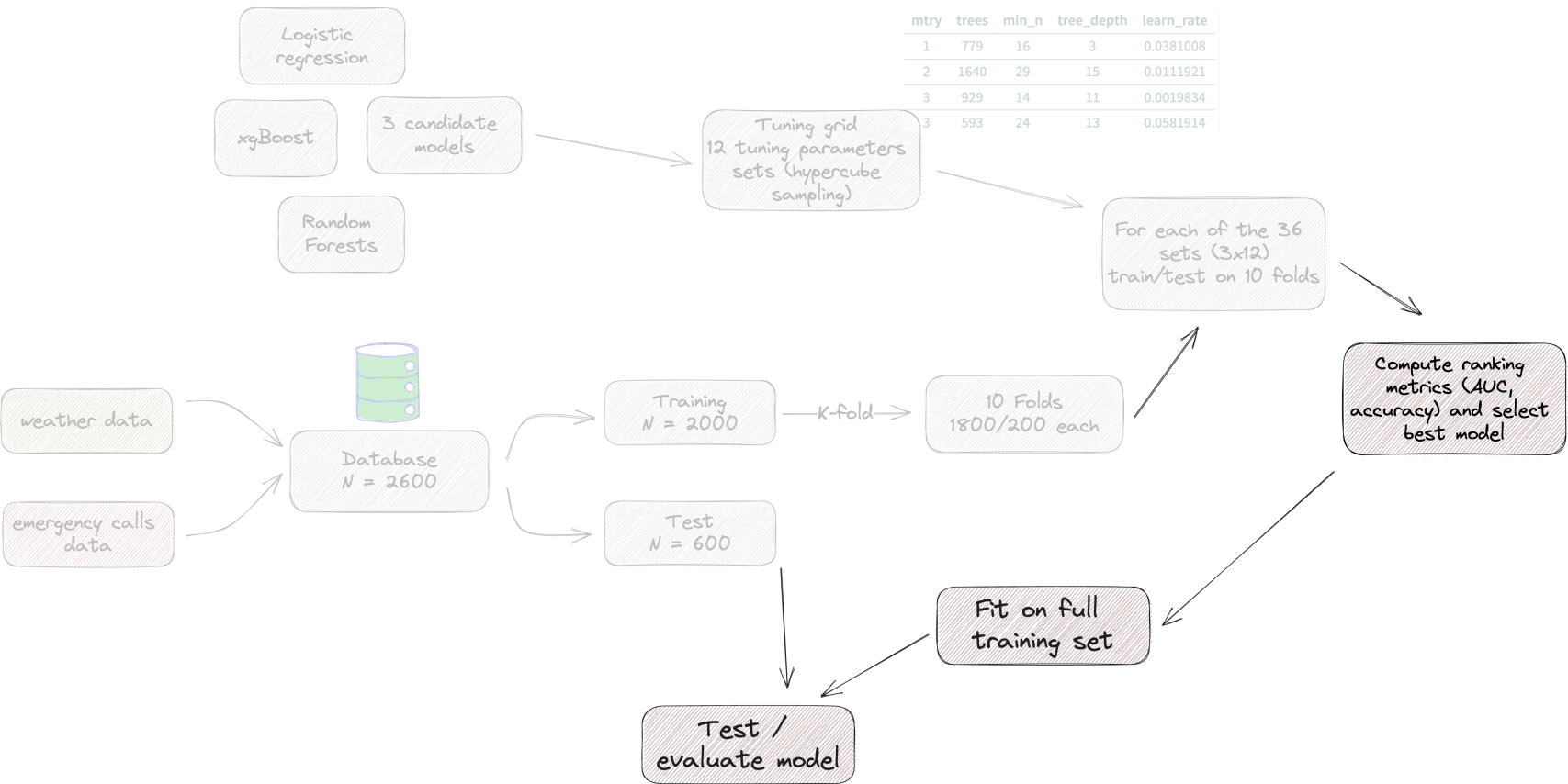
## Data pre-processing

- Centering, [scaling](#), dummy-coding
- **S**ynthetic **M**inority **O**ver-sampling **T**Echnique ([SMOTE](#)) (Chawla et al. 2002) for the outcome

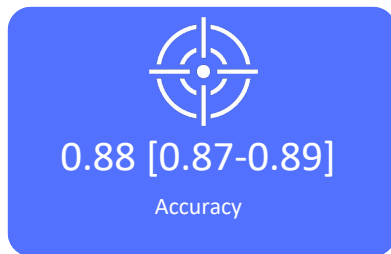
# Modelling workflow



# Modelling workflow



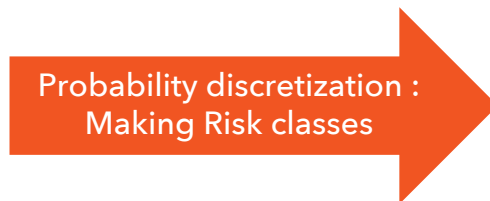
# Our winner : XGBoost



Reality

		Reality	
		Drown	No drown
Model	Drown	True Pos = <b>58</b>	False Pos = <b>60</b>
	No drown	False Neg = <b>23</b>	True Neg = <b>522</b>

Confusion matrix



Risk class	# of Drownings	# of "No Drownings"	Total
1	31	502	533
2	20	44	64
3	4	20	24
4	16	11	27
5	7	8	15

N = 663 testing set

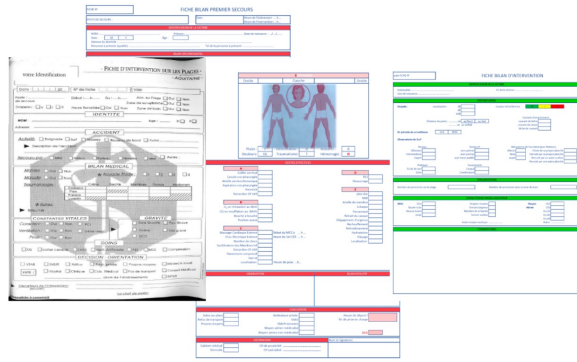
# Discussion

“all models are wrong, but some are useful”

*George Box*

- **What is the best drowning prediction model ?** Lowering false negatives *or* false positives ? → Risk management and political decisions
- Low improvements over previous models → dataset limitation ?
- Emergency call database only covers a **minority** (👮) of all rescues

# Perspectives



Majority of rescues

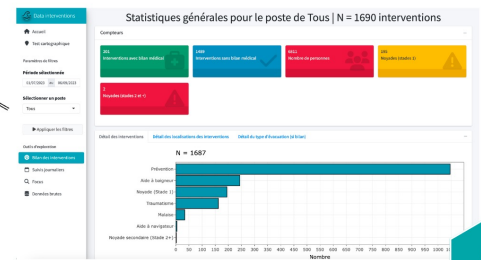


iPad data collection  
offline mode available

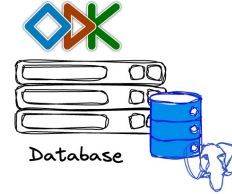
Summer 2023  
experiment



New Model !



Monitoring Dashboard





# Thank you !

## Want to try this at home ?

## Interested in the technical stuff ?

Contact us :

 [david.carayon@inrae.fr](mailto:david.carayon@inrae.fr)

 @david\_carayon



# References

Castelle, Bruno, Tim Scott, Rob Brander, Jak McCarroll, Arthur Robinet, Eric Tellier, Elias de Korte, Bruno Simonnet, and Louis-Rachid Salmi. 2019. “Environmental Controls on Surf Zone Injuries on High-Energy Beaches.” *Natural Hazards and Earth System Sciences* 19 (10): 21832205.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. “SMOTE: Synthetic Minority over-Sampling Technique.” *Journal of Artificial Intelligence Research* 16 (June): 321–57. <https://doi.org/10.1613/jair.953>.

Dehez, Jeffrey, and Sandrine Lyser. 2021. “Fréquentation des plages océanes et risques de baignade en Aquitaine en 2020. Une étude exploratoire.” Research Report. INRAE. <https://hal.science/hal-03549020>.

Sacks, Jerome, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. 1989. “Design and Analysis of Computer Experiments.” *Statistical Science* 4 (4). <https://doi.org/10.1214/ss/1177012413>.

Tellier, Éric, Bruno Simonnet, Cédric Gil-Jardiné, Marion Lerouge-Bailhache, Bruno Castelle, and Rachid Salmi. 2021. “Predicting Drowning from Sea and Weather Forecasts: Development and Validation of a Model on Surf Beaches of Southwestern France.” *Injury Prevention* 28 (1): 16–22. <https://doi.org/10.1136/injuryprev-2020-044092>.