healthware international

Communicators
Connectors
Builders of Future Health

Master in Data Science | 21 MAY 2022

Nicola Procopio

A Gentle Introduction to XAI

healthware international

01

Healthware Overview

healthware international

Global Reach

healthware international

healthware international

The full-service healthcare agency of Healthware Group

We play at the intersection of science, creativity, boundless curiosity, and our understanding of human needs. That's how we design transformational healthcare experiences that engage, simplify and empower people's lives.

We are digital natives and multi-talented coders, connected and passionate to learn and innovate.

Our mission is to design and develop successful solutions and digital products.

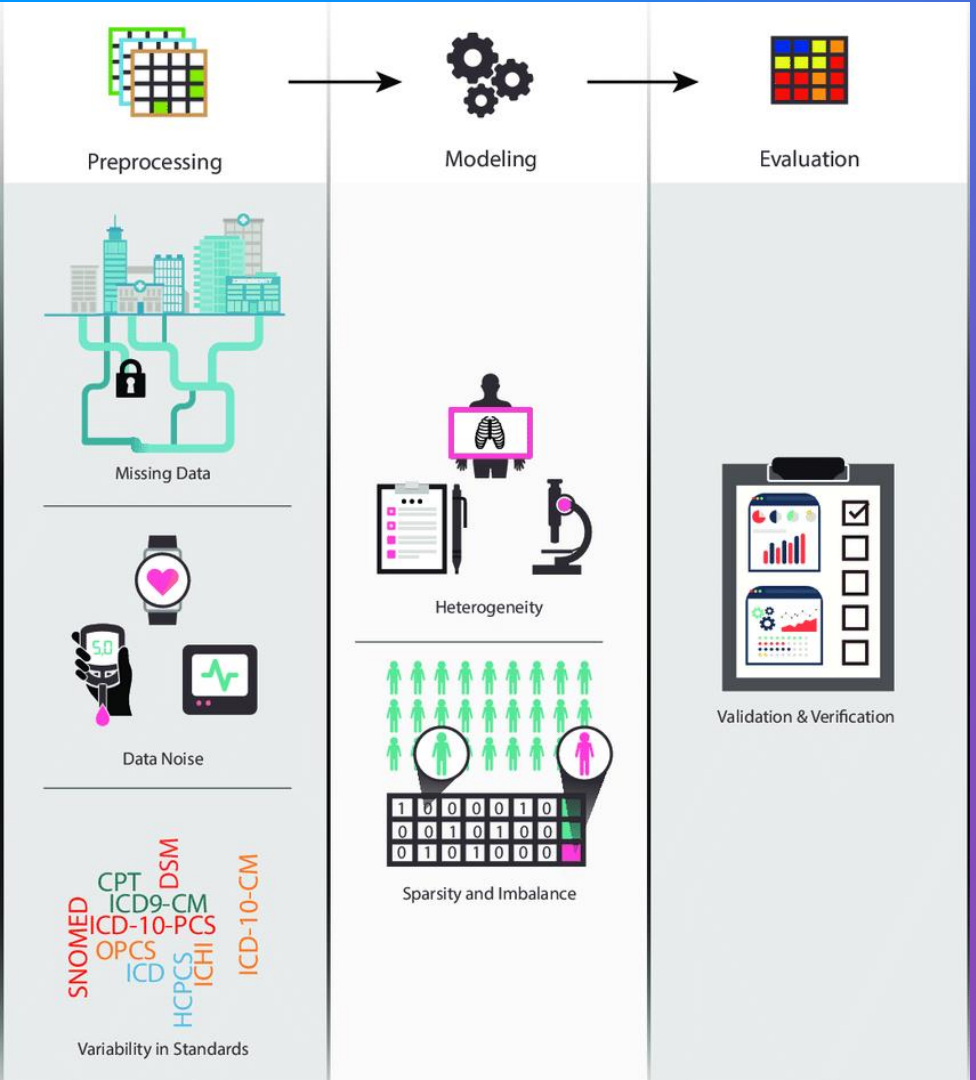
The Sila Valley

- **The Data Science team of Healthware covers with its expertise the entire design process**: Data Ingestion, Data Analysis and Analitycs, Algorithms (NLP, ML/AI, DeepLearning, Statistics), Data Visualization.
- The team has deep expertise not only on models and algorithms, but also on architectures: Big Data and Cloud in particular.
- **The Healthware Data Science Team, is located in a district of ICT particulary focused on Artificial Intelligence**: Cosenza is a very stimulating environment due to the presence of Universities and Research Centers, Startups and Communities, other companies in the Artificial Intelligence sector.
- The medical campus of the **Magna Graecia University** offers, among other research lines, also a research center **in neuroscience and medical science**.

Data Science: Our Philosofy

- **From Big to Smart Data**. Giving meaning to data is, therefore, the element that distinguishes us.
- **From Artificial to Augmented Intelligence**. Big data, NLP, machine learning, neural networks to support doctors to improve the quality of life.
- **Explainable Artificial Intelligence (XAI)**. XAI can improve the user experience of a product or service by helping end users trust that the AI is making good decisions.

healthware international



Getting from questions to answers

1. Define the problem to solve.

2. Define your approach:
    a. What data?
    b. What preprocessing strategy?
    c. What kind of models?
    d. What validation metrics?

3. Define and share short/long period strategies.

Who I am

Nicola Procopio

Senior Data Scientist

Contacts

nicola.procopio@healthwareinternational.com

https://it.linkedin.com/in/nicolaprocopio

https://github.com/nickprock

https://www.slideshare.net/NicolaProcopio

Education

Master in Applied Statistics for Economy and Finance.

UNIVERSITÀ DELLA CALABRIA

Actually @

healthware international

Background

integris

ICAR CNR

Affare

KDD Lab

UNIVERSITÀ DELLA CALABRIA

Community

AI DEMOCRATIZE

tag Talent Garden Cosenza

iaml ITALIAN ASSOCIATION FOR MACHINE LEARNING

GDG

SC STARTUP CALABRIA

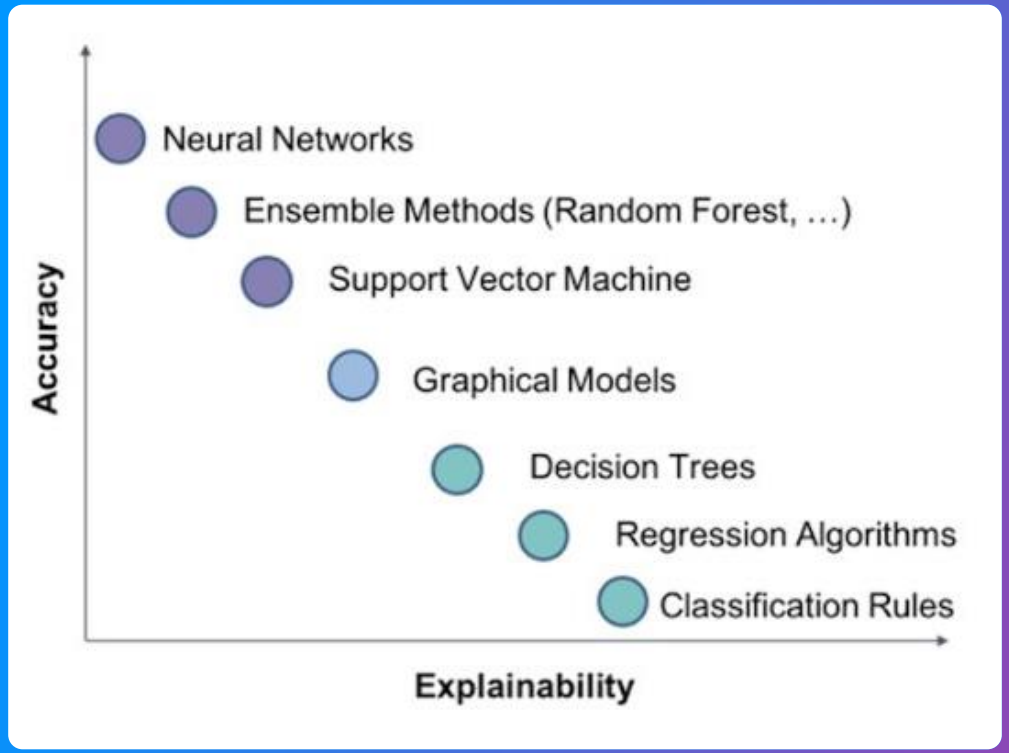
healthware international

02

A Gentle Introduction to XAI

healthware international

ML in Healthcare



- In recent years there has been an explosion of ML applications, mainly due to increasingly accurate predictions.
- The increase in the quality of the results clashes with the lowering of the explainability of the models.
- **This especially in the field of healthcare is not acceptable.**

ML in Healthcare

- Need for explainable medicine: medical education, research, and clinical decision making.
- Need for experts to understand and retrace thedecision-making process of the machine.
- **AI Trust and Medical Progress through XAI.**

It’s not a human move

- **Today the ML algorithms are powerful in terms of prediction but opaque for humans**
- *XAI is a set of methods and tools that can be adopted to make ML models understandable to human beings*
- ML models explainable, that is, to answer the questions “*What*”, “*How*” and “*Why*” on the result.
- European law through the GDPR sanctions the **right to know the reason of the decision taken by an algorithm**.

Can we trust?

We have powerful DNN why we need the XAI models?

- Problems in the **learning phase**
- Problems into the data: ***bias and fairness***
- Problems from humans**: adversarial attacks**

Problems in the Learning Phase

Explain the Prediction

Predicted: Wolf
True: Wolf

Predicted: Husky
True: Husky

Predicted: Husky
True: Husky

Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Wolf

Predicted: Husky
True: Wolf

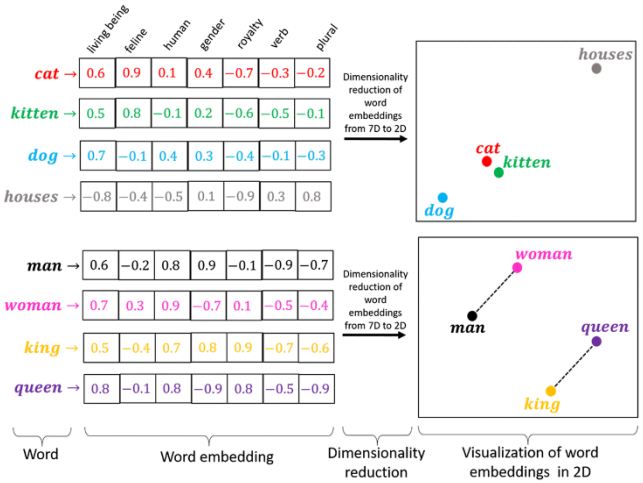
Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Husky

Predicted: Husky
True: Husky

healthware international

Problems into data



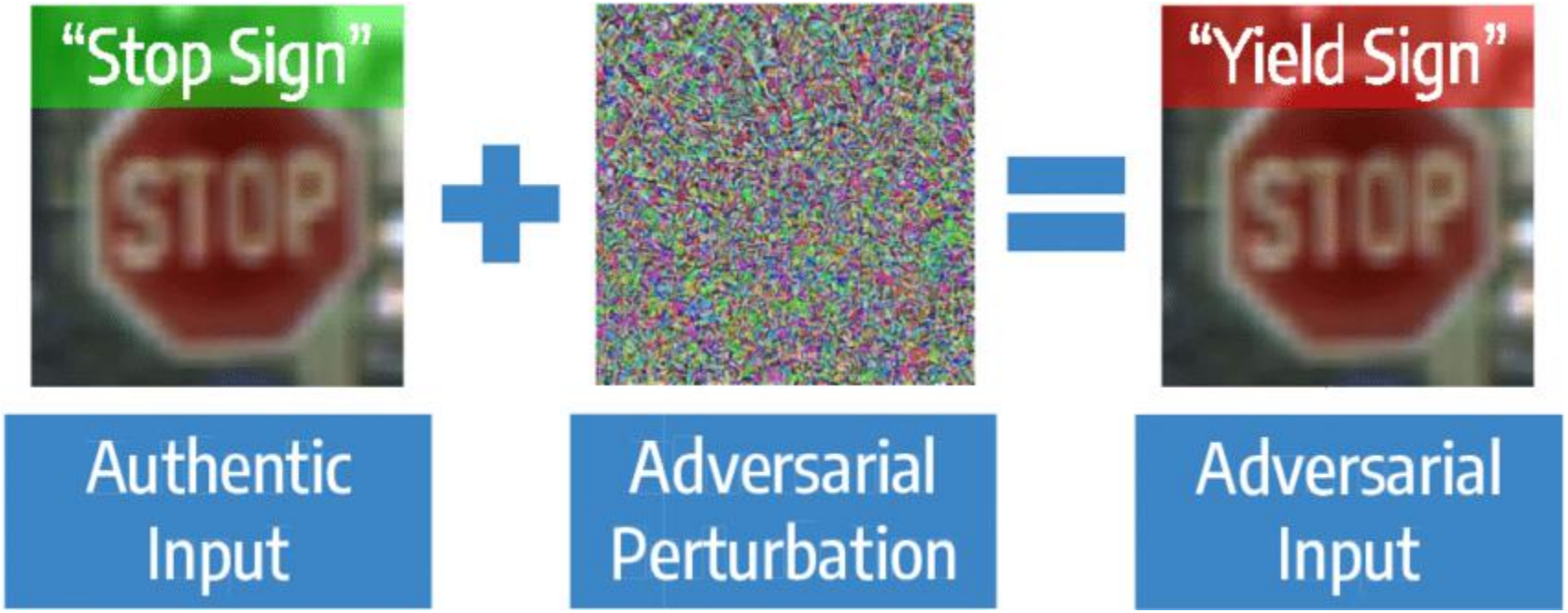
Masked Language Models (MLMs) have shown superior performances in numerous downstream NLP tasks when used as text encoders.

Unfortunately, MLMs also demonstrate significantly worrying levels of **social biases**.

"man is to programmer as woman is to <mask>"

"man is to programmer as woman is to housekeeper"

Problems from human



healthware international

XAI must be F.A.S.T.

- **Fair** and not negatively biased
- **Accountable** on its decision
- **Secure** to outside malevolent hacking
- **Transparent** in its internal

- *XAI isn't for depressing performance to the advantage of explanations*
- *Human in the loop is essential to XAI, performance and confidence*

How to explain?

To deal with the previous problems we need explanations and we can have them:

- *creating methodologies that explain already trained black box models:* ***EXPLAINABLE AI***
- *create transparent glass box models that allow their operation to be seen and easily interpreted:* ***INTERPRETABLE AI***

Interpretable A.I.

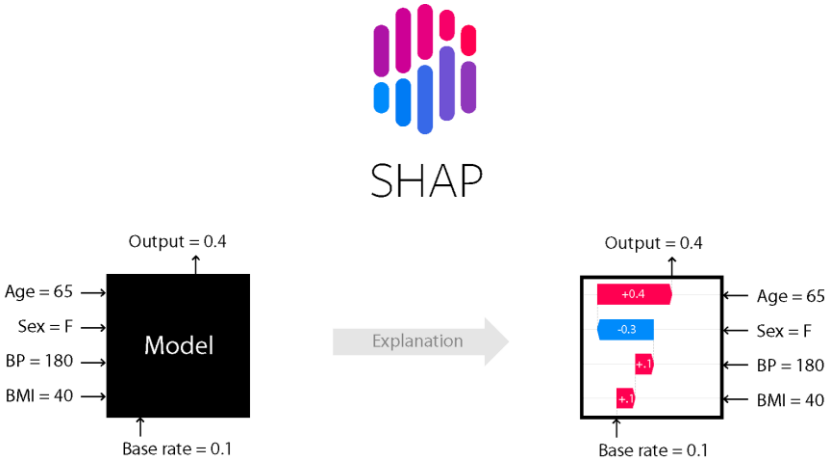
- ***Glass-box models*** are machine learning models designed for interpretability (ex: linear models, rule lists, generalized additive models)
- ***EBM*** is an interpretable model developed at *Microsoft Research*. It uses modern machine learning techniques like bagging, gradient boosting, and automatic interaction detection to breathe new life into traditional GAMs. This makes EBMs as accurate as state-of-the-art techniques and produce exact explanations and are editable by domain experts.

NOTEBOOK TIME!

Explainable A.I.

- As it is hard to gain a comprehensive understanding of their inner working after they have been trained, many ML systems (especially deep neural networks) are essentially considered ***black-boxes***. This makes it hard to understand and explain the behavior of a model.
- Explainable AI is used to describe an AI model, its expected impact and potential biases.
- Nowadays the state-of-the-art in XAI is **SHAP**.

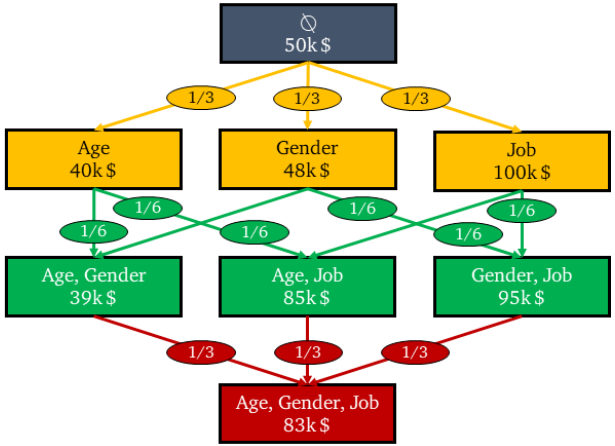
SHAP



- ***SHAP (SHapley Additive exPlanations)*** is a game theoretic approach to explain the output of any machine learning model (model-agnostic).
- It connects optimal credit allocation with local explanations using the *classic Shapley values from game theory* (they are a widely used approach from cooperative games).
- It's **a local method** and provides explanations of a single prediction through a linear combination of the underlying Shapely values.

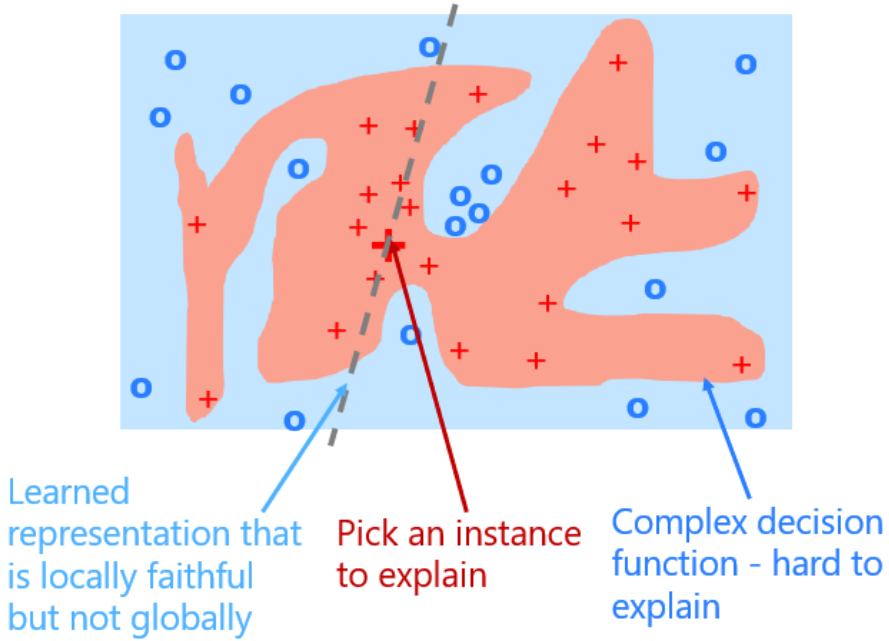
Demystifying the demystifier

	N. of Nodes $\binom{F}{f}$	N. of Edges $f \times \binom{F}{f}$
f = 0	1	
f = 1		3
	3	
f = 2		6
	3	
f = 3		3
	1	
Sum	$2^F=8$	$F \times 2^{F-1}=12$



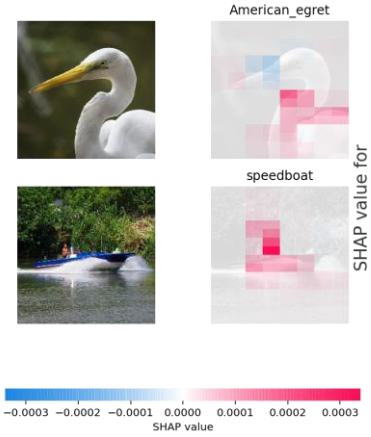
- Game theory needs at least two things: a **game** and some **players**:
  - the "game" **is reproducing the outcome of the model** *[One game: one observation]*
  - the "players" **are the features included in the model.**
- Shapley values are based on the idea that the outcome of **each possible combination of players should be considered to determine the importance of a single player**.
- **Shapely values are all of the same dimension**

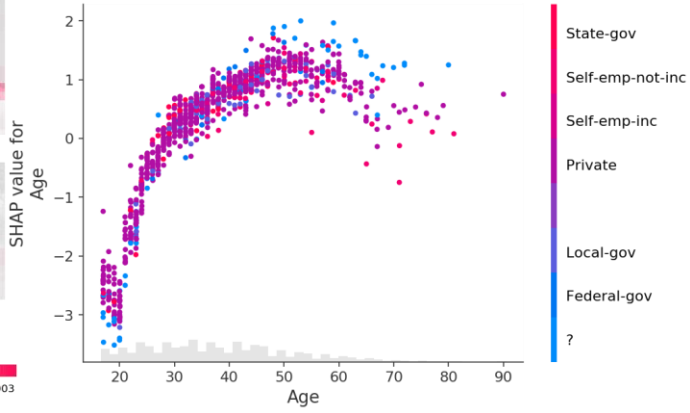
SHAP Speedups



- A standard calculation for large features is too time-consuming.
- **KernelShap** is an agnostic approximate linear approximation and works for every possible model you may train
- **TreeShap** only works on tree-based models
- Exists also a specialization for DNN called **DeepShap**

Some SHAP plots





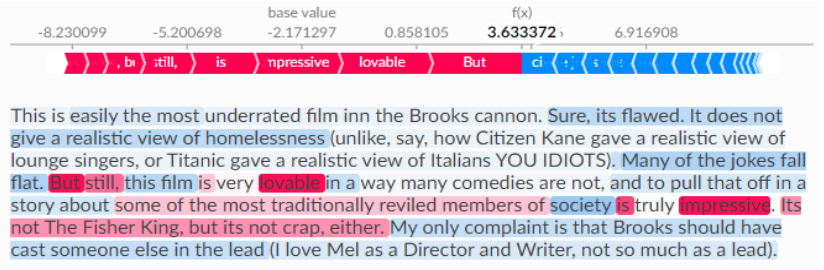
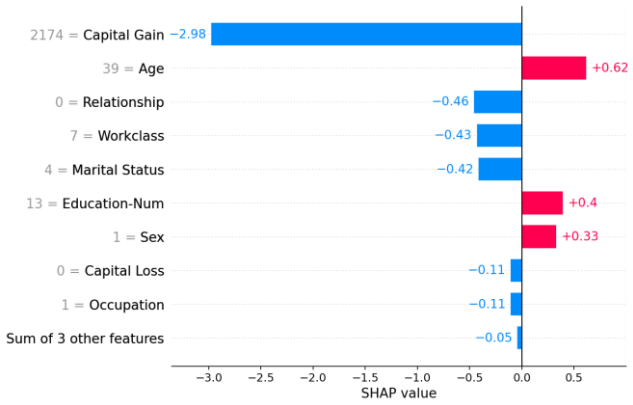
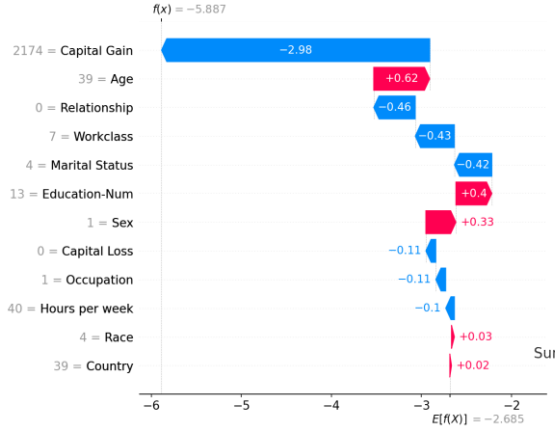
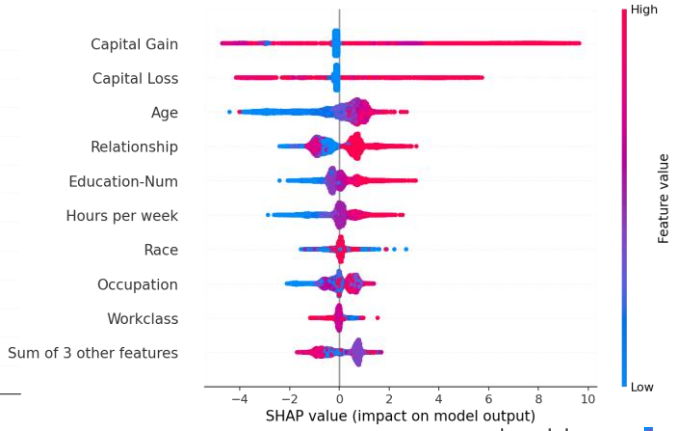


Image Plot



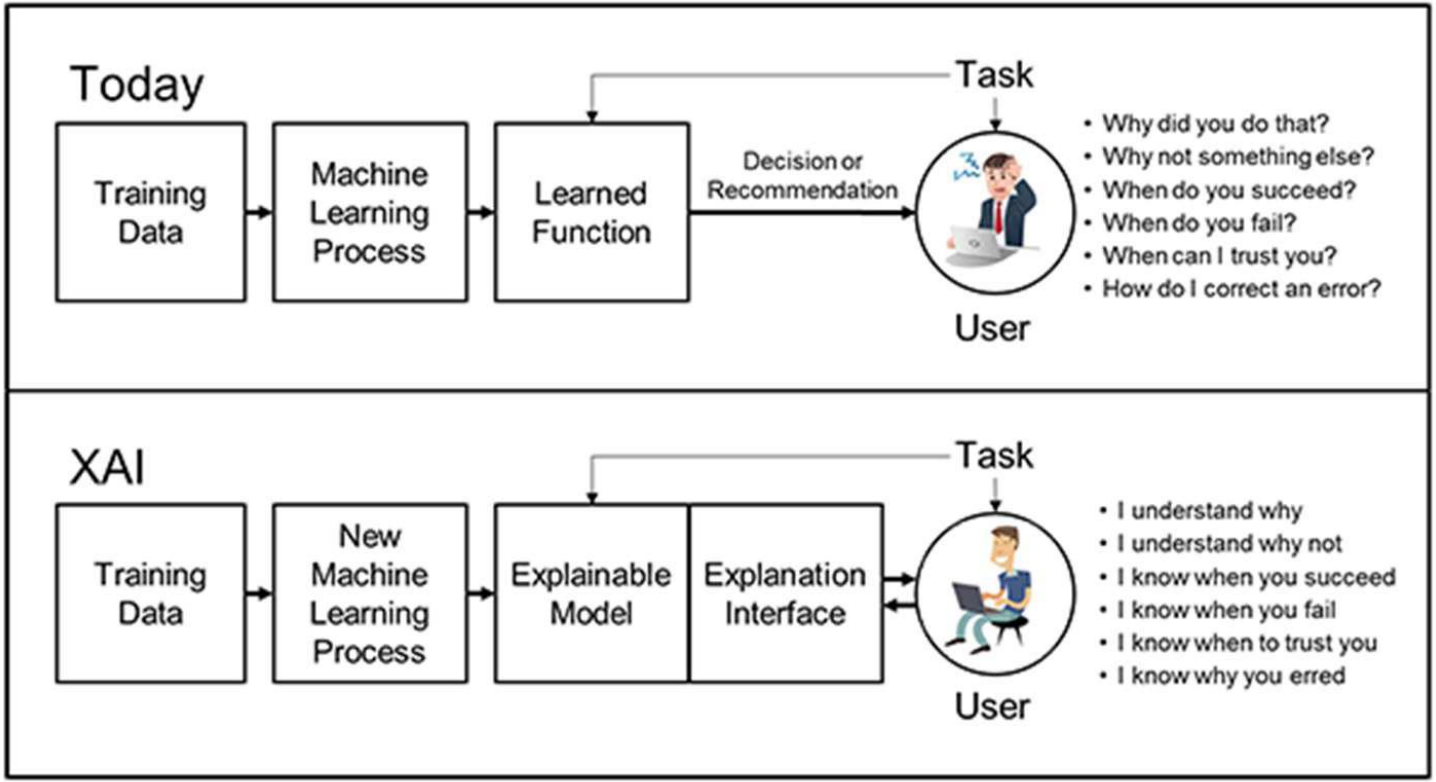




healthware international

NOTEBOOK TIME!

XAI Pipeline



healthware international

Conclusion

Use Black-Box Explanation (LIME, SHAP, Partial Dependence, …) when:

- *You don't have access to the training data*
- *Or model was pre-trained and given to you*
- *Or a specific black-box model was required (neural net, boosted trees, random forests, …)*
- *Or you're trying to understand a complex pipeline from beginning to end*

Must use black-box explanation methods

But Use Glass-Box Machine Learning when:

- You have access to the training data and you're the one training the model
- You're the one who needs to debug the model, retrain the model, improve model accuracy, …

Should use Glass-Box ML methods such as EBMs

- *Exact intelligibility, not approximate as with black-box explanation methods*
- *Better intelligibility leads to faster debugging and model development/improvement*
- *Models are editable to correct bias and errors*

"Friends, don't let friends deploy black-box models."

-Rich Caruana-

healthware international

Thank you.

@healthware_intl | @healthware_ita
linkedin.com/company/healthwareInternational/
facebook.com/healthwareintl
instagram.com/healthware_intl

www.healthwareInternational.com

healthware international

Communicators
Connectors
Builders of Future Health

Q&A

@healthware_intl | @healthware_ita
linkedin.com/company/healthwareInternational/
facebook.com/healthwareintl
instagram.com/healthware_intl

www.healthwareInternational.com

healthware international

Communicators
Connectors
Builders of Future Health