AI EXPERT NETWORK

Explainable AI (XAI)

Mike Salem

Associate Director of Data Science – QA

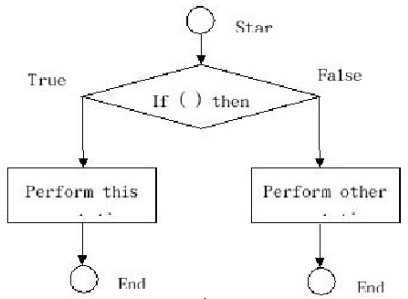
Gilead Sciences

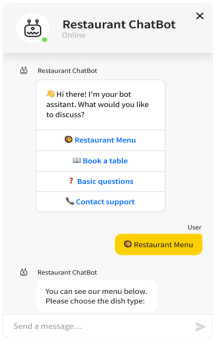
AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

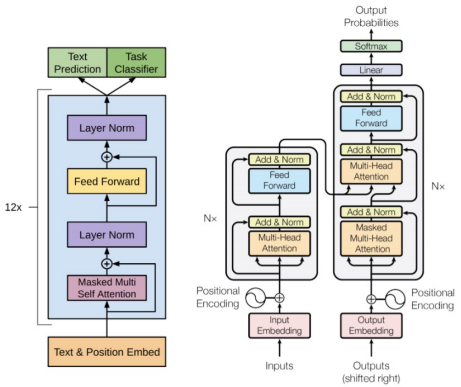
Inspiring Collaboration. Leading Innovation. Making a difference.

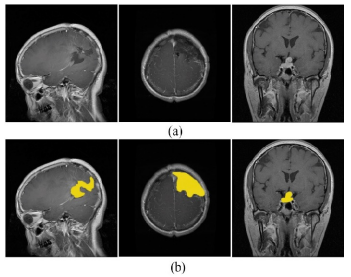
AI EXPERT NETWORK

AI is Transforming at a Rapid Rate









AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

Inspiring Collaboration. Leading Innovation. Making a difference.

AI EXPERT NETWORK

What is even real anymore?

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

Inspiring Collaboration. Leading Innovation. Making a difference.

Not Everyone is an AI Expert

- Stakeholders are becoming more accountable for their products that use AI
- Company data teams are being questioned more about the inner workings of their models prior to their use in production systems
- Simple probabilities are not sufficient in systems where life and death are involved
- People want to know more about the “why” of what is being produced so they can:
  - Stand behind the decision that is made by backing it up with domain expertise
  - To ensure the RIGHT information is being used in the first place (first principles in chemical composition)
  - To weed out pure correlations and look to understand causal applications

Singularity or Nightfall?

- As these tools continue to evolve and show promise, they will start to show up in more and more applications
- However, if these tools are used incorrectly, in some cases catastrophic events can occur :
  - Drugs passing inspection with contaminants
  - Large amounts of adverse effects
  - False positives passing through systems (e.g. cancer)
- This has led to many industries attempting to regulate AI use in systems

AI EXPERT NETWORK

Is it possible for humans to understand how an AI model makes a prediction?

This is the concept of explainable AI

(XAI)

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

Inspiring Collaboration. Leading Innovation. Making a difference.

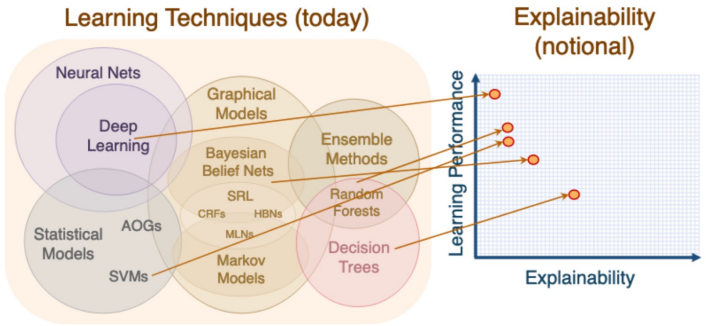
About XAI

- Aims to gives context to what is being generated
- Explains “what” is driving the decisions in ways humans can interpret
- Helps stakeholders and developers remove the “Black box” approach to AI algorithms
- Active area of research
- Lots of open source and free tools

AI EXPERT NETWORK

XAI can be broken into multiple stages

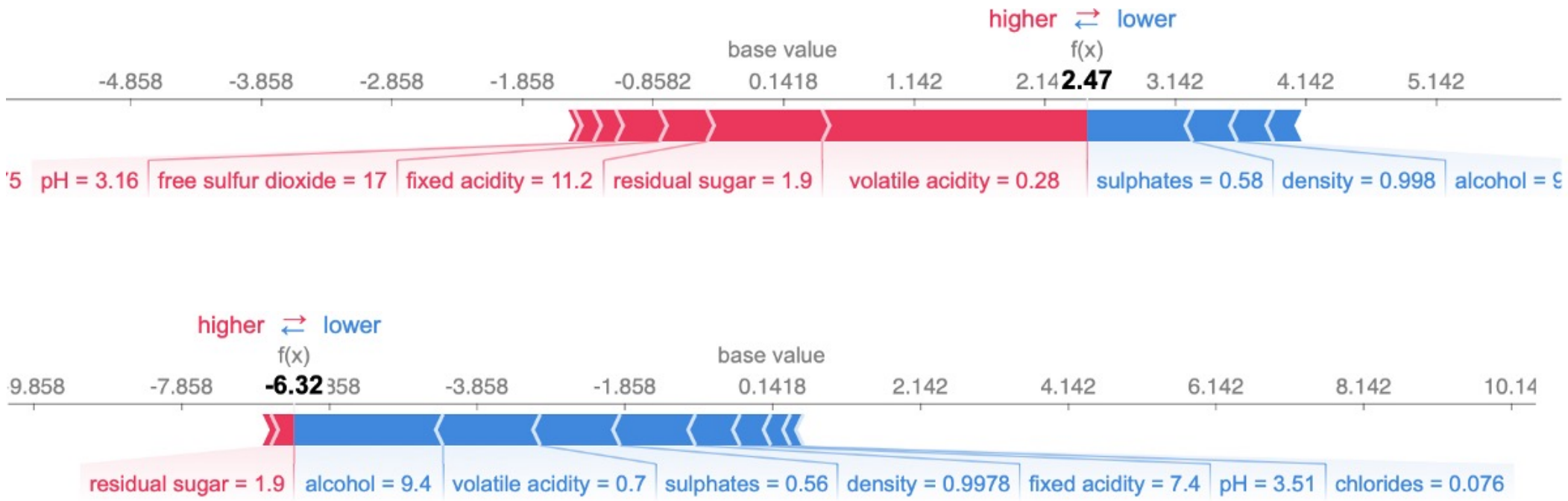
- **Pre-modeling Explainability – clear understandable data and features**
- **Modeling Explainability – choose models that are easy to understand and interpret**
- **Post-model Explainability – use approaches that allow you to understand the effect of changing the models features such as SHAP values**



http://report898.web.fc2.com/article/paper-20166161636/

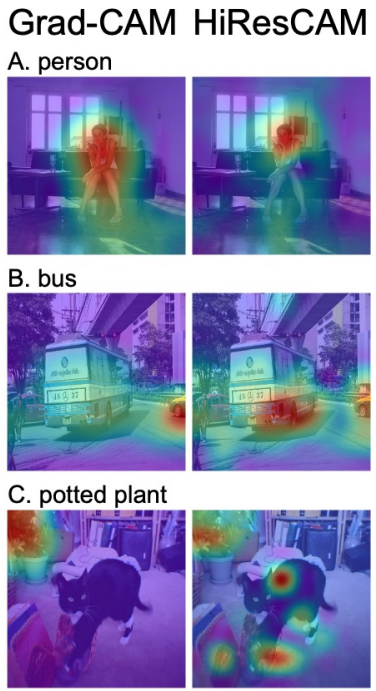
https://www.kdnuggets.com/2023/01/explainable-ai-10-python-libraries-demystifying-decisions.html

Example of XAI on Tabular Data to Classify Wine (SHAP)



https://betterdatascience.com/shap/

Example of XAI on Images (Grad-CAM / HiResCAM)



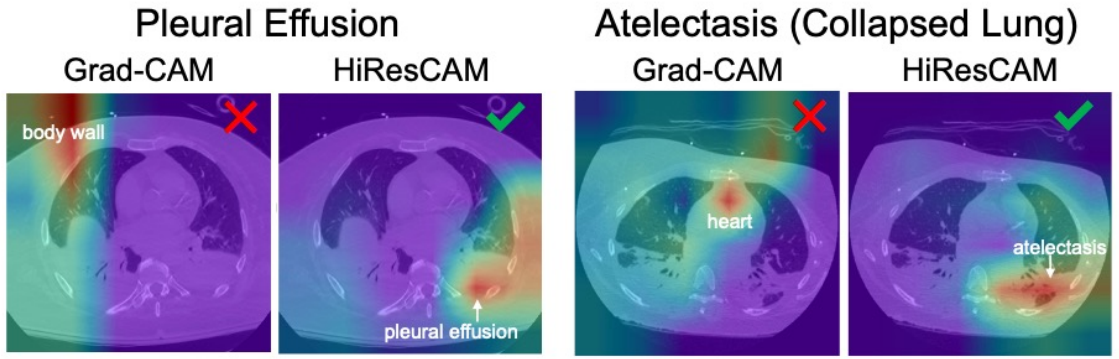
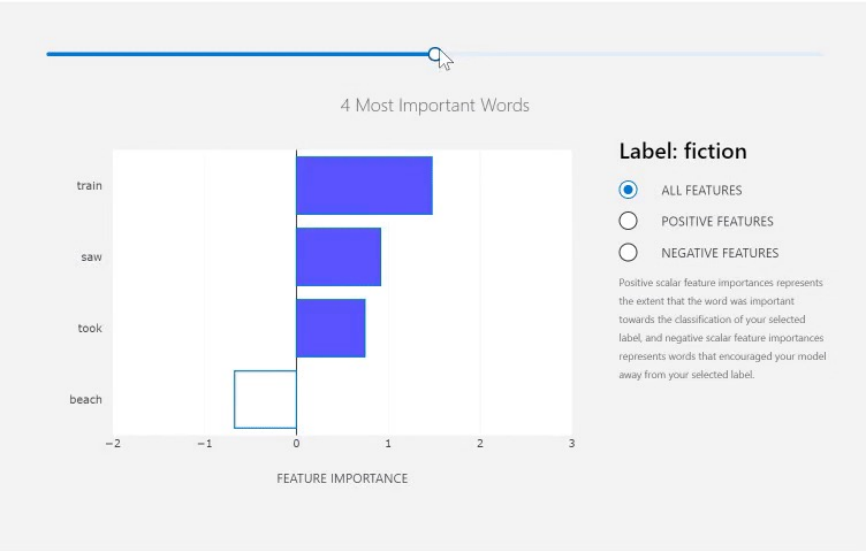


Figure 8: Examples of Grad-CAM creating the incorrect impression that an AxialNet model focused on the wrong anatomical structure. The HiResCAM and Grad-CAM explanations were generated using exactly the same model on the same input CT volume. The only difference is the explanation method. Both of the abnormalities shown here are lung findings, and HiResCAM indicates that the model used the lung fields to predict these lung findings. However, Grad-CAM creates the impression that the model predicted these lung abnormalities based on the body wall and heart, which are irrelevant. Best viewed in color; text annotations added for clarity.

https://arxiv.org/pdf/2011.08891v4.pdf

AI EXPERT NETWORK

Example of XAI for Text (Interpret-Text)



https://github.com/interpretml/interpret-text

AFDO RAPS HEALTHCARE PRODUCTS COLLABORATIVE

Inspiring Collaboration. Leading Innovation. Making a difference.

Examples of Other XAI Tools

- Local Interpretable Model-Agnostic Explanations (LIME)
- Testing with Concept Activation Vector (TCAV)
- moDel Agnostic Language for Exploration and eXplanation (DALEX)
- Anchors
- Explainerdashboard
- Alibi

Questions for the group

- Are you using any XAI tools? If so, which one(s)?
- Should we provide guidance / suggestions on tools of preference?
- How do we partner with other entities to continue the advancement of XAI for healthcare?