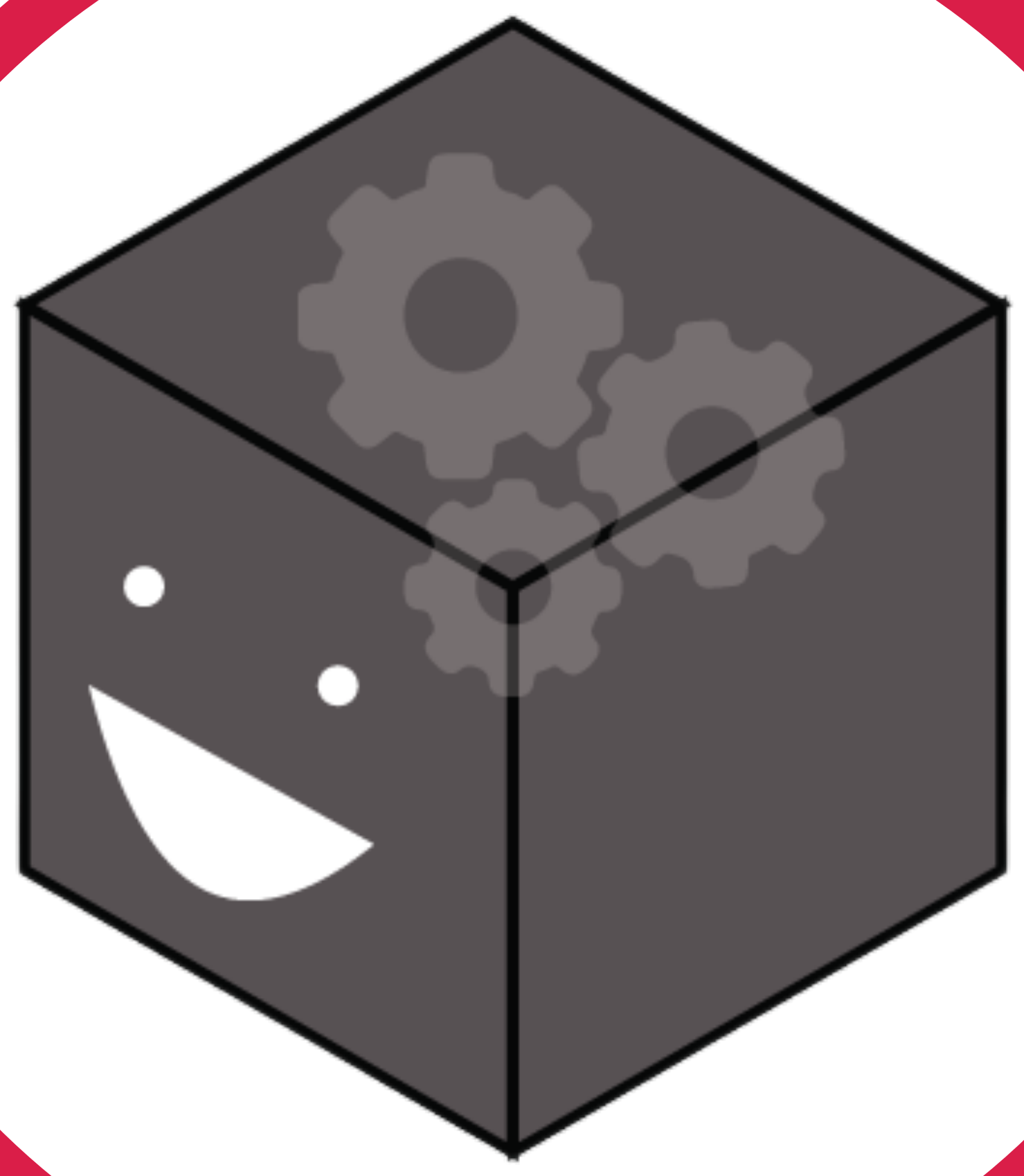
Connecting the Dots: Harnessing Explainable AI to better understand Climate Extremes

Building models is hard
Explaining models is harder

My focus is on creating intuitive visualizations and research software tools that not only enhance the interpretability of complex climate models but also empower scientists and policymakers to make informed decisions.

Black Box Models Are Needed

Modern machine learning models enable researchers to let the system define decision processes without predefined functional forms.



Use XAI Methods For Explanations

As high performing models become more complex, there's a growing demand to understand the decision-making process, shifting the focus to the question, "How did the model arrive at this decision?".

Who Am I?

Janith Wanniarachchi

PhD student of the Econometrics and Business Statistics Department at **Monash University**

Supervised by **Di Cook, Kate Saunders, Patricia Menendez and Thiyanga Talagala**

Appsilon

R/ Shiny Developer at **Appsilon**

BSc. (Hons.) in Statistics from University of Sri Jayewardenepura, **Sri Lanka**

What Am I Doing?

Bushfires are becoming a major concern

The Australian bushfires, notably during the 2019/2020 season, have emerged as a significant global concern, showcasing the escalating threat of climate change. The devastating impact on ecosystems, wildlife, and communities emphasizes the critical role of both immediate firefighting efforts and long-term climate resilience measures.

As bushfire ignitions become more frequent and intense even during the present year, building models on predicting where the next fire is going to ignite is not enough. *By obtaining explanations from model we can generate new insights into the key features and patterns to monitor to detect bush fires.*

Gathering climate and anthropogenic activity data

Currently my efforts are in building models for the state of Victoria using the following data.

- ✓ Previously identified bushfires and their causes
- ✓ Temperature, rainfall, humidity and wind speed data
- ✓ Forest coverage
- ✓ Powerline distribution

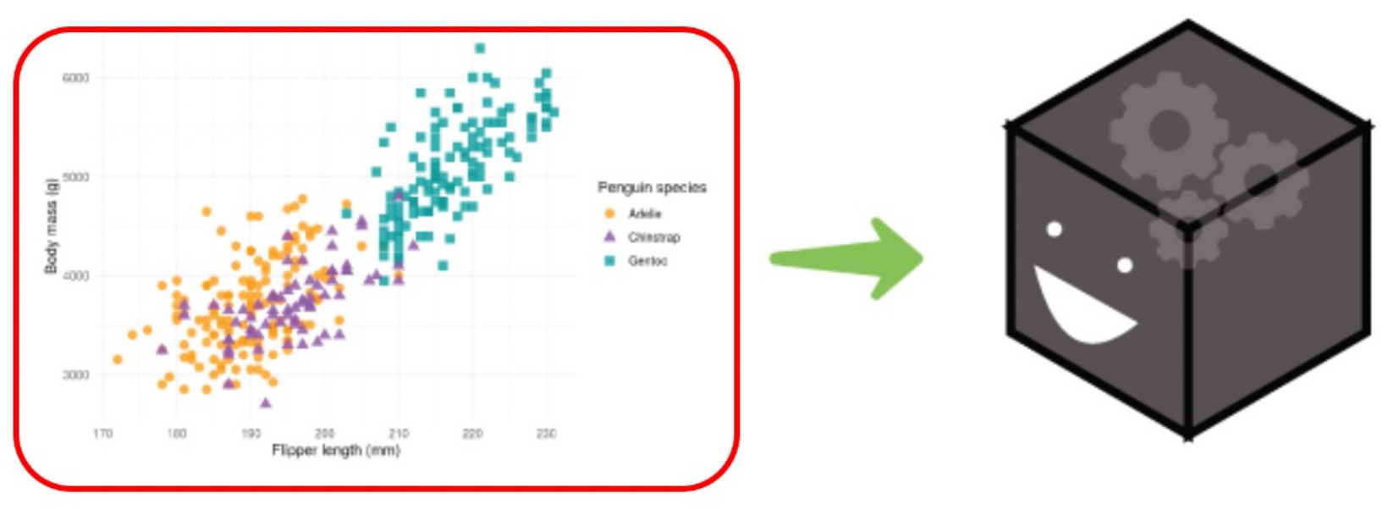
After using these variables to build predictive models, I'm hoping to use eXplainable AI methods and visualize the underlying decision process in an intuitive method. *Thereby bridging the gap between the black box model and the general public to explain the effect of different climate and anthropogenic factors towards the ignition of bushfires.*

How Does XAI Work?

Simply put,

Most model agnostic XAI methods work by applying different variations of data to the black box model to examine it's behavior and generate explanations by uncovering patterns in the observed behaviour.

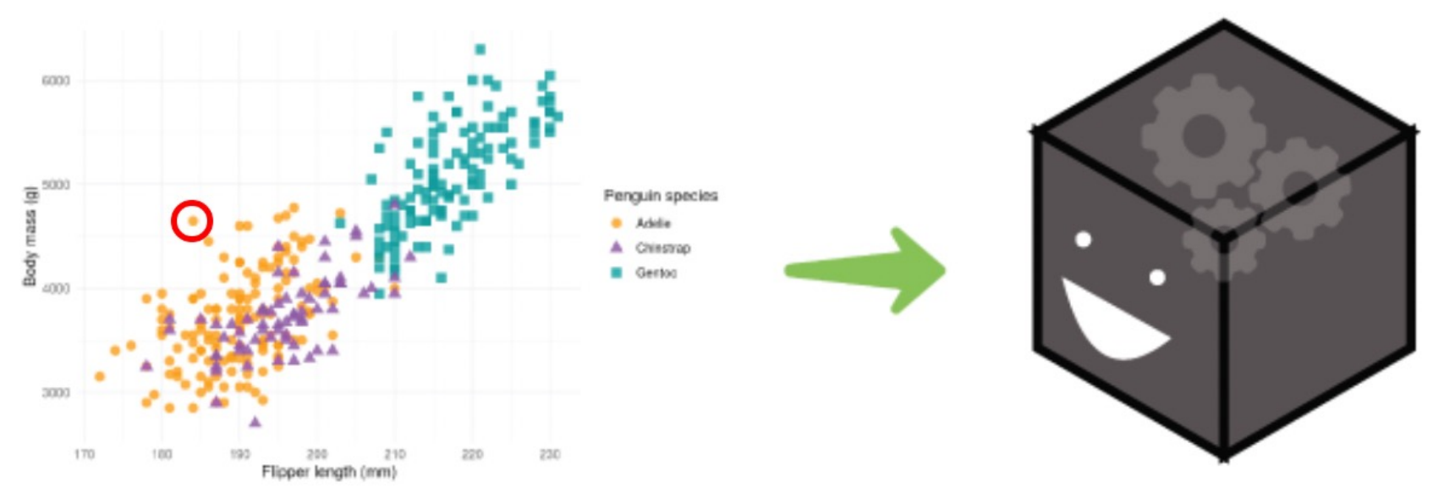
There are two types of XAI methods



Global Interpretability Methods

Gives an overall bird's eye view of the entire dataset and model behaviour.

Examples include
- Partial Dependency Plots (PDP)
- Inidividual Conditional Expectation plot (ICE)
- Accumulated Local Effects plot (ALE)



Local Interpretability Methods

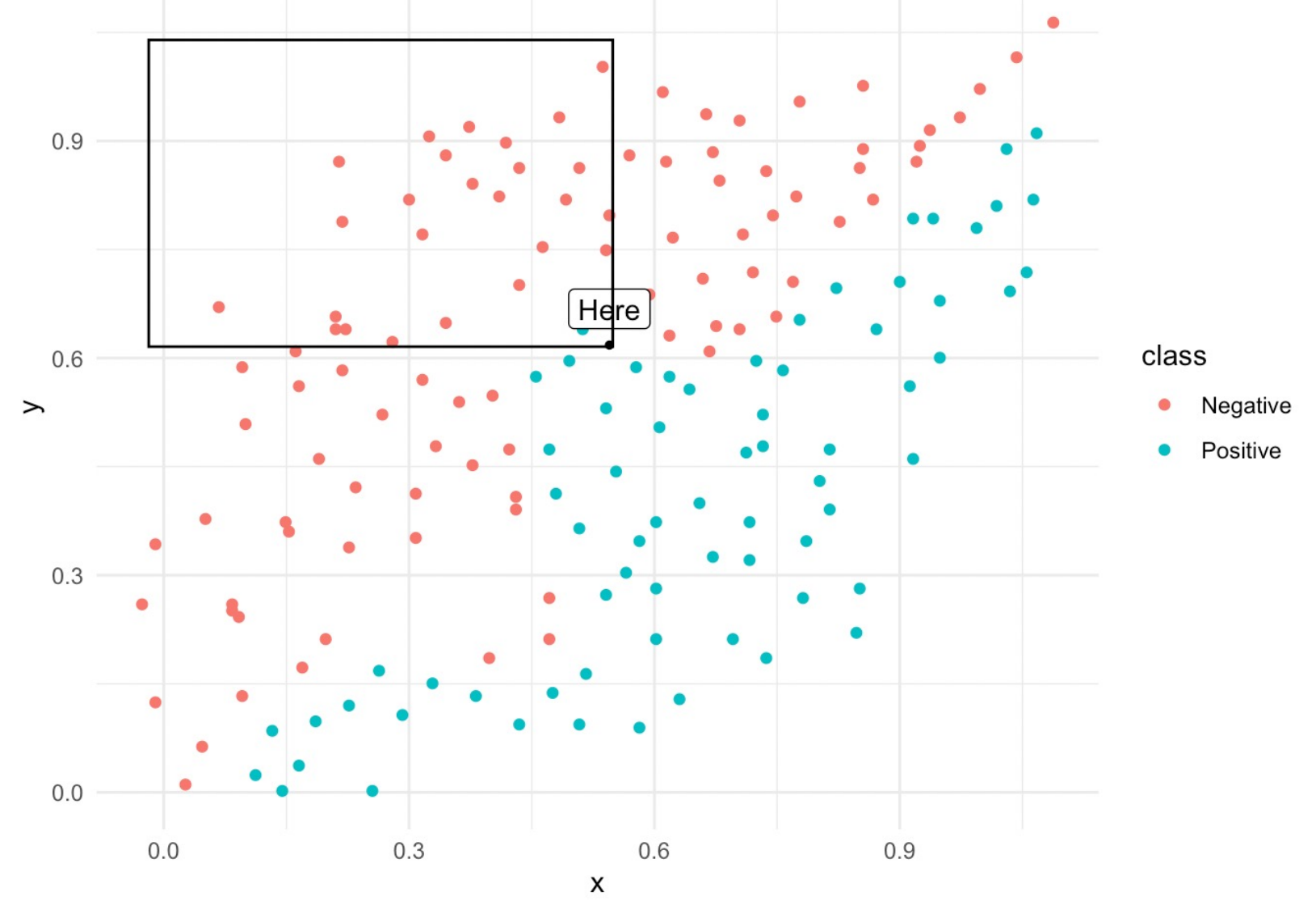
Explains the reasoning behind a single instance.

Examples include
- Anchors
- Local Interpretable Model agnostic Explanations (LIME)
- Shapley explanations (SHAP)
- Permutation based feature importances (PFI)

*From the above methods, **Anchors** (Ribeiro, Singh, and Guestrin 2018) is a recently introduced method that has an easy to understand underlying concept while also giving easy to understand explanations of the decision process of black box models for a specific instance.*

How does Anchors work

Anchors are defined as,
- A model agnostic local interpretability method that generates human readable decision rules that explain the decision process for one single instance
- A rule or a set of predicates that satisfy the given instance and is a sufficient condition for $f(x)$ (i.e. the model output) with high probability



Anchors are found by, solving the following optimization problem so that *We can find a big enough boundary box in the feature space containing other points that would have the same model prediction as the anchoring point.*

$$\max_{\mathcal{A} \text{ s.t. } \Pr(\text{Prec}(\mathcal{A}) \geq \tau) \geq 1-\delta} \text{cov}(\mathcal{A})$$

$$\text{cov}(\mathcal{A}) = \mathbb{E}_{\mathcal{D}(z)}[\mathcal{A}(z)]$$

$$\text{Prec}(\mathcal{A}) = \mathbb{E}_{\mathcal{D}(z|\mathcal{A})}[\mathbb{1}_{f(x)=f(z)}]$$

This is achieved by formulating the problem as a Multi Armed Bandit problem to purely explore the feature space and generate a large enough bounding box based on the coverage and the precision of the bounding box.

MONASH University

climate extremes
ARC centre of excellence