

Smithinstitute

Explainability in Machine Learning

Rachael Warrington

CMIH Academic Engagement Event, 26th July 2022

In the next 25 minutes or so...

- Who are Smith Institute?
- Why care about explainability in machine learning?
- How might we provide explainability?
- What might be of interest to explore in future?

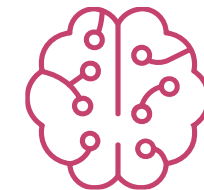
Who are Smith Institute?

- Mathematical consultancy, providing governments and industry with mathematical solutions to business challenges

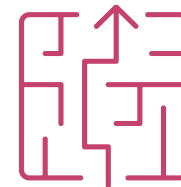
optimisation



verification



artificial
intelligence



forecasting

What have SI done around healthcare?

- RAMP: model review
- CASE projects:
 - RLUH & University of Liverpool – medical imaging
 - AstraZeneca & University of Leeds - [Mathematical modelling of cellular receptor signalling](#)
- Study groups

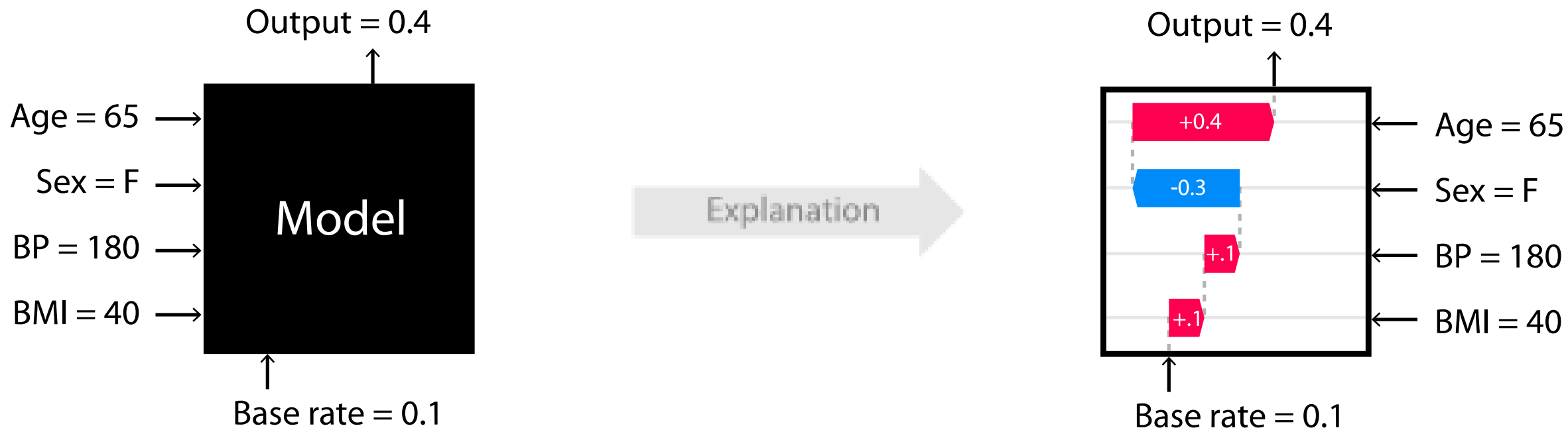
Why does explainability matter?

- Ethics
- Trust
- Understanding

Using SHAP to provide explainability

- Based on game theory
- Can be applied to any machine learning model
- Run ML model with particular feature values
- SHAP breaks down that model output into a baseline plus contributions from each feature

Using SHAP to provide explainability



Copyright (c) 2018 Scott Lundberg, used under MIT licence

Questions for future exploration

- What are SHAP's limitations, when does it no longer apply?
- Can it be extended to cover those cases?

Summary

- Explainability in machine learning matters, for:
 - Ethics
 - Trust
 - Understanding
- SHAP is one way to provide explainability...
- ...but it has limitations, and there are questions to ponder and avenues to explore