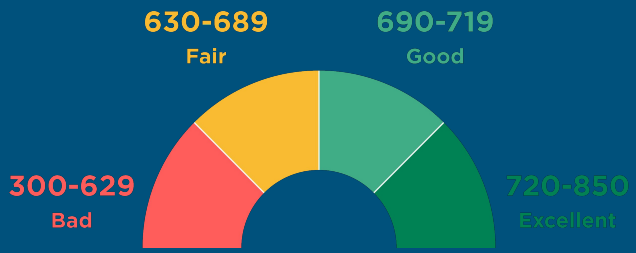
3.3 Summary and Farewell

- Trustworthy and human-centred AI and ML.
- FAT Forensics.
- Modular interpretability and transparency.
- Surrogate explainers.
- Extra learning resources.
- Stay in touch!

Raul Santos-Rodriguez

Trustworthy and human-centred AI and ML

- Understanding the models we use is of utmost importance.



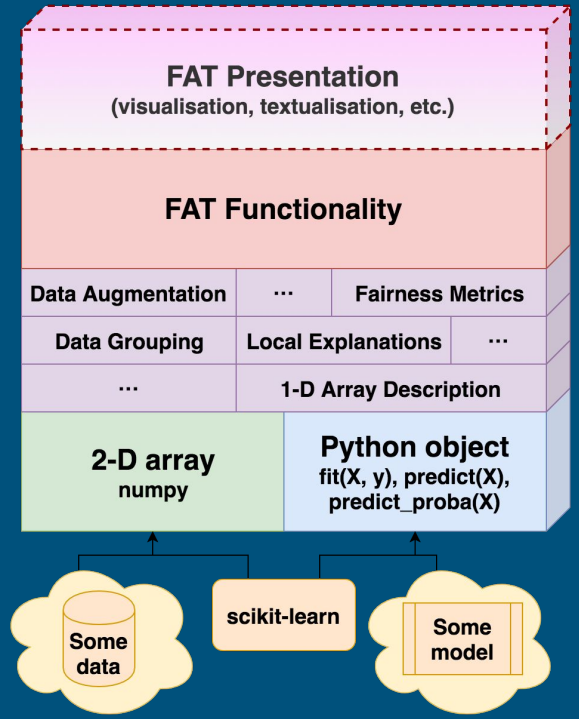
Credit Score

Prison Sentence

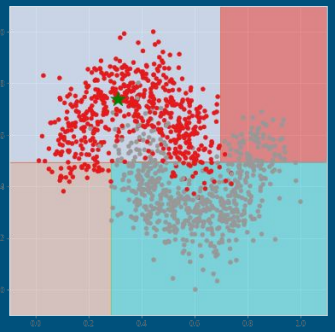
- **Humans** -- stakeholders and explainees -- must be able to **understand** and **trust** the models that affect their lives.
- We, as **engineers**, must be **confident** in the **tools** that we build and deploy.

FAT Forensics

- Software, not paperware.
- Fairness, Accountability and Transparency.
- Modular and flexible design.
- Two modes of operation:
  - research; and
  - deployment.
- Licenced under BSD 3-Clause.

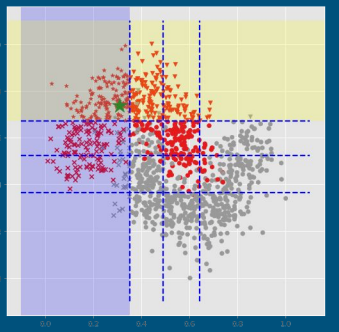


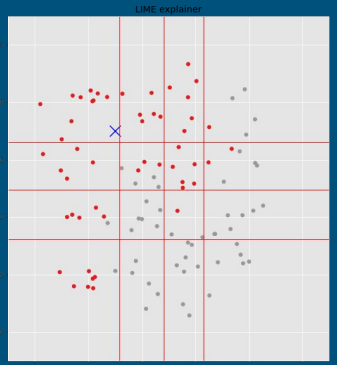
Modular interpretability

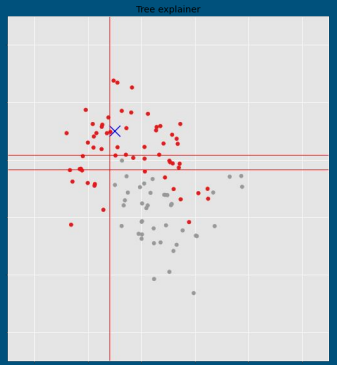


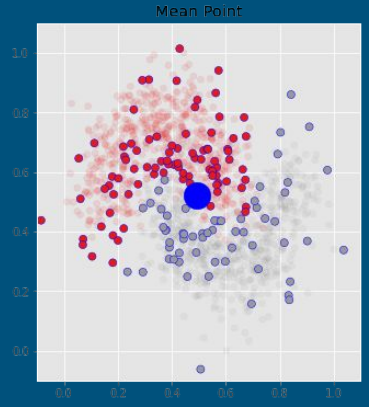
No Free Lunch:

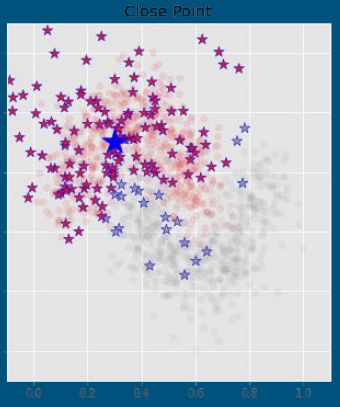
One explainer does not fit all!

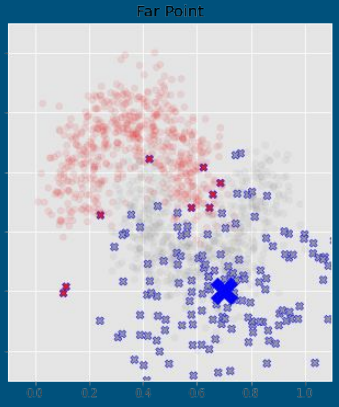






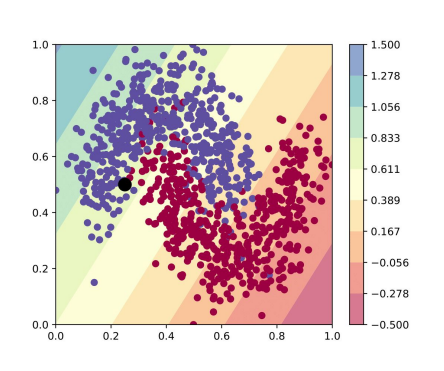




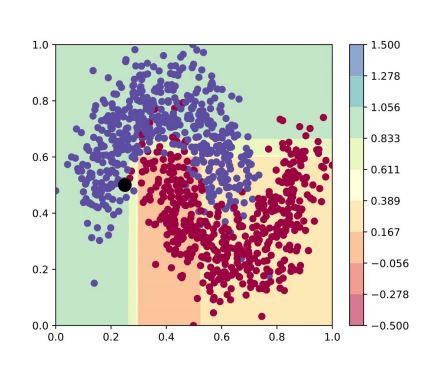


Surrogate explainers with bLIMEy

Linear

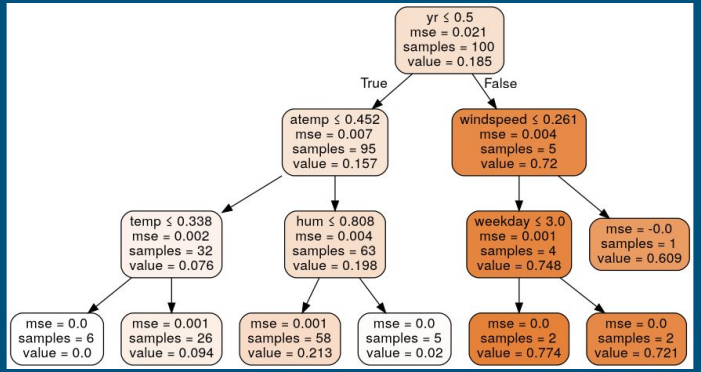


Decision Tree



build **LIME** **y**ourself

Surrogate explainers should be tailor-made from interoperable algorithmic modules.



Hands-on

- We have worked on the following notebooks:

`1-data-sets.ipynb`

`2-interpretable-representations.ipynb`

`3-data-sampling.ipynb`

`4-explanation-generation.ipynb`

- We covered how to:
  - understand trade-offs and measure them;
  - customise *data sampling* and *interpretable representation* generation to fit one's needs; and
  - build bespoke (local) surrogate explainers with the desired properties.

Next steps

- Write a **user guide** on building surrogate explainers based on this tutorial.
- Expand the analysis to **image** and **text** data.

- **Decompose** ANCHOR into algorithmic building blocks.
- **Implement** ANCHOR within FAT Forensics.

- Compose user guides for **Permutation Importance**, **Partial Dependence**, **Individual Conditional Expectation**, …

Worth checking out -- FAT Forensics

- GitHub: https://github.com/fat-forensics/fat-forensics
- Official Documentation: https://fat-forensics.org/
- arXiv and JOSS papers describing the package:
  - https://arxiv.org/abs/1909.05167
  - https://joss.theoj.org/papers/10.21105/joss.01904

JOSS

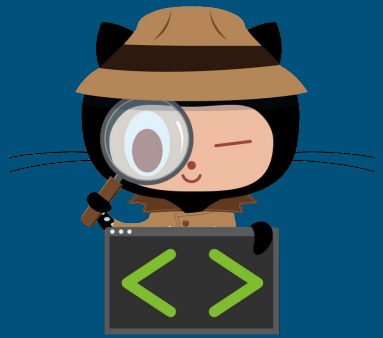
arXiv.org

Worth checking out -- Surrogate Explainers

- HCML 2019 workshop paper describing the bLIMEy algorithm:
  - https://arxiv.org/abs/1910.13016
- arXiv paper describing tree-based surrogates of image data:
  - https://arxiv.org/abs/2005.01427
- arXiv paper describing interpretable representations and their (computational) meaning:
  - https://arxiv.org/abs/2008.07007

Don’t be a stranger!

- Get in touch with email -- contact details available on the tutorial website.
  - https://events.fat-forensics.org/#instructors
- Reach out on Slack.
  - https://fatforensicsevents.slack.com/
- Report *issues* and submit *pull requests* via Github.
  - https://github.com/fat-forensics/



YOU DON'T HAVE TO GO HOME BUT YOU CAN'T STAY HERE