Machine Learning & Anomaly Detection

Michelle Lochner

SARAO Data Science Team:

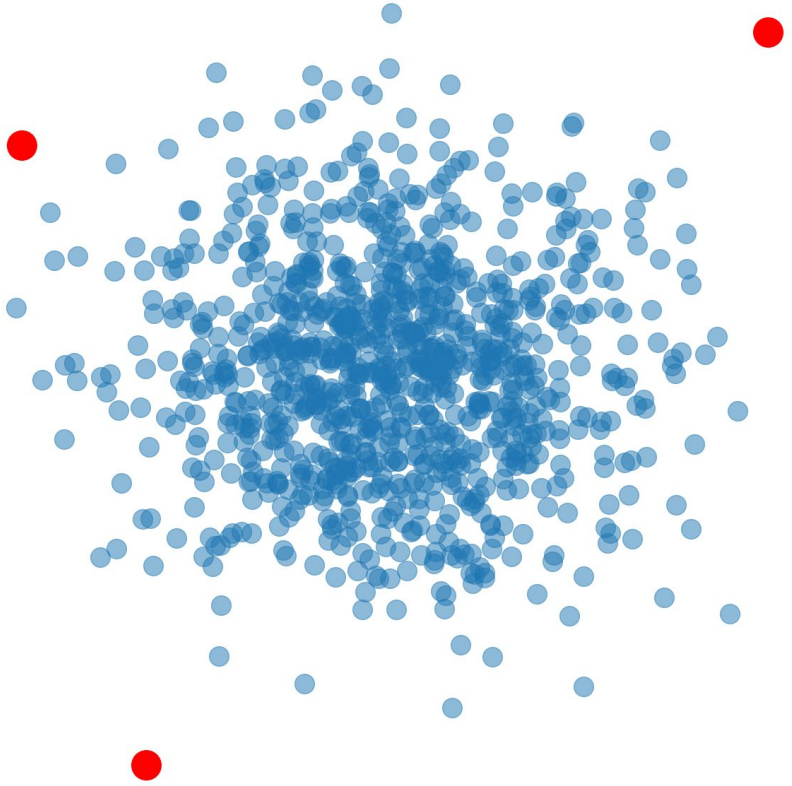
Bruce Bassett
Alireza Vafaei Sadr
Ethan Roberts

Kavli Summer Program Students:

Sara Webb
Daniel Muthukrishna

UCL/ Oskar Klein Centre:

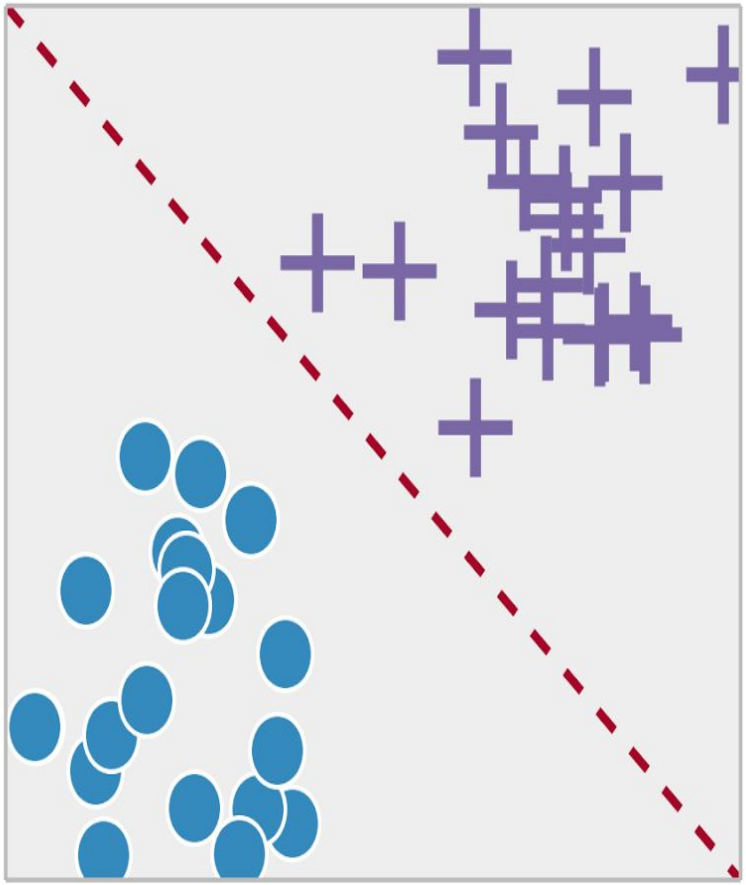
Hiranya Peiris
Catarina Alves
Rahul Biswas



AIMS | African Institute for Mathematical Sciences SOUTH AFRICA

NRF National Research Foundation | SARAO South African Radio Astronomy Observatory

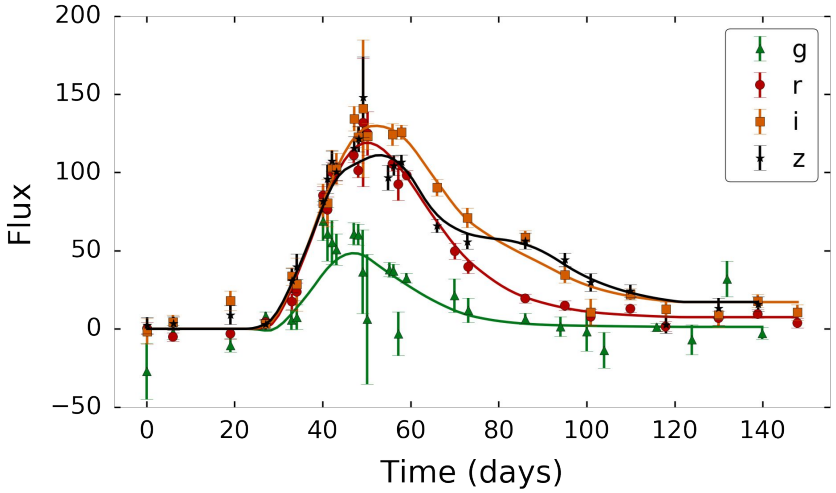
Supervised Machine Learning



Automatically learns a model to map inputs to outputs, using a training set.

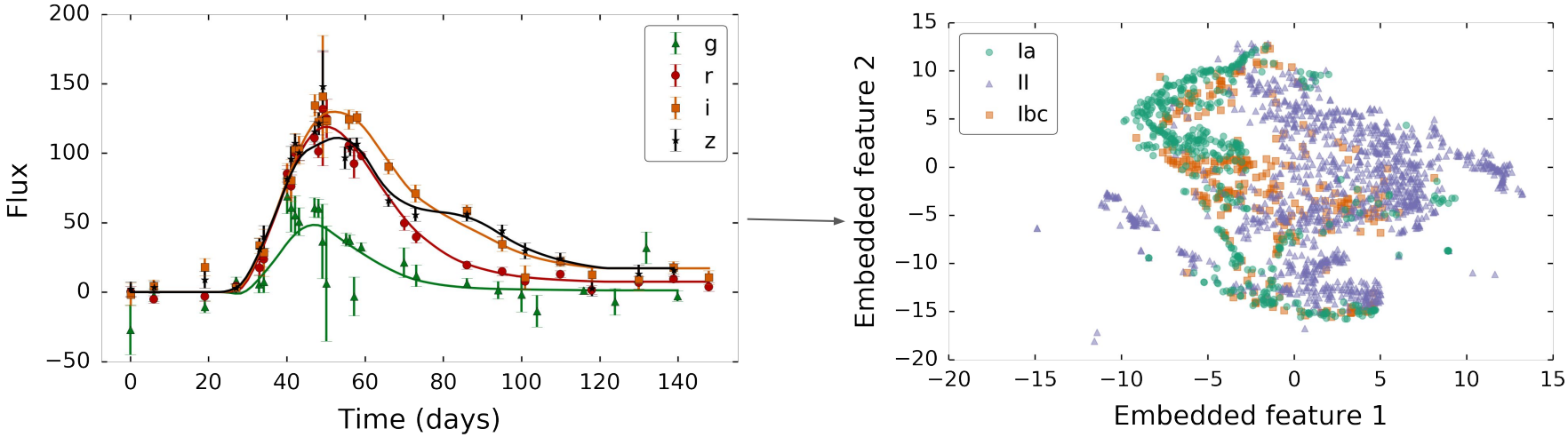
https://towardsdatascience.com

Transient Classification



Lochner et al. (2016) - 1603.00882

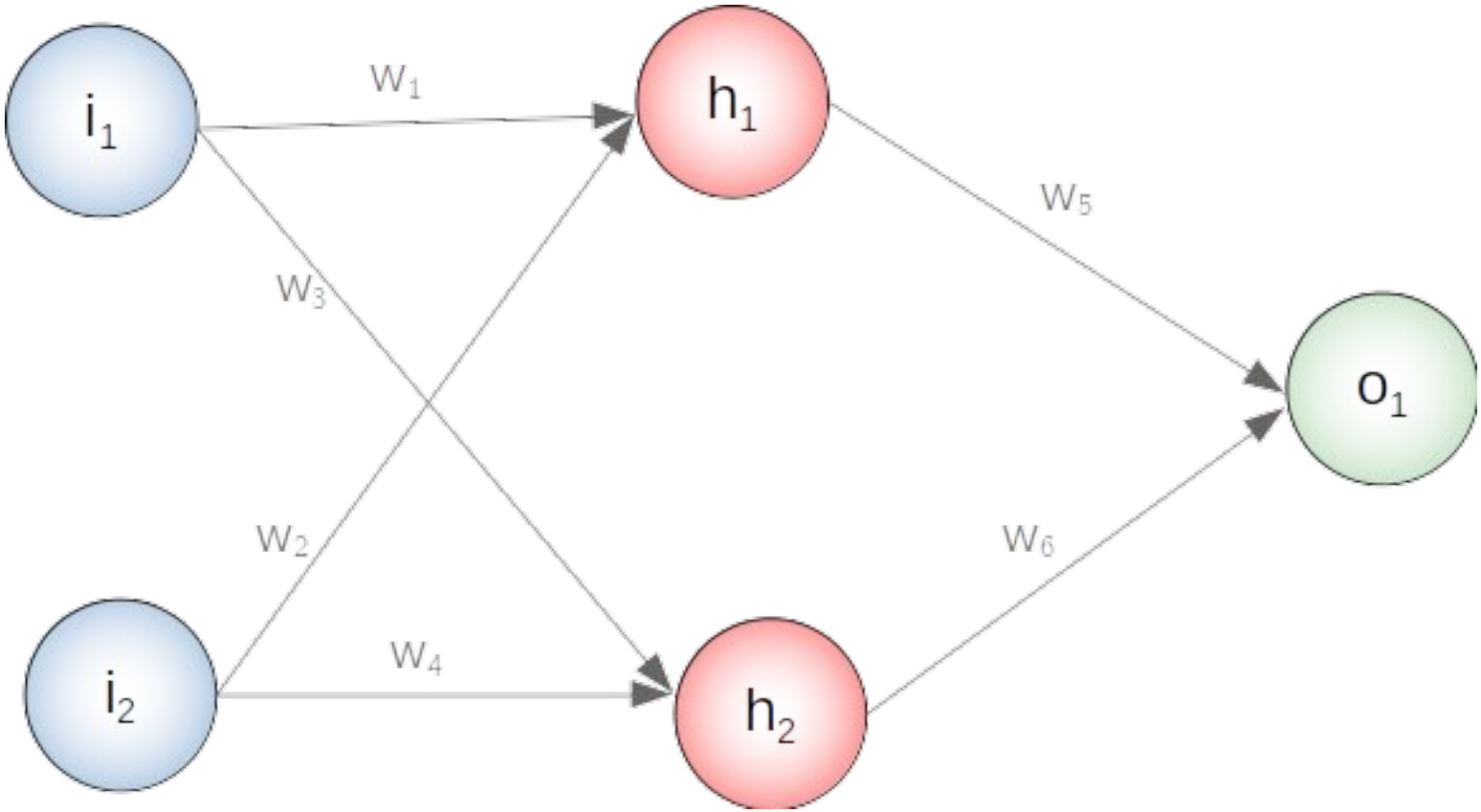
Transient Classification



Feature Extraction

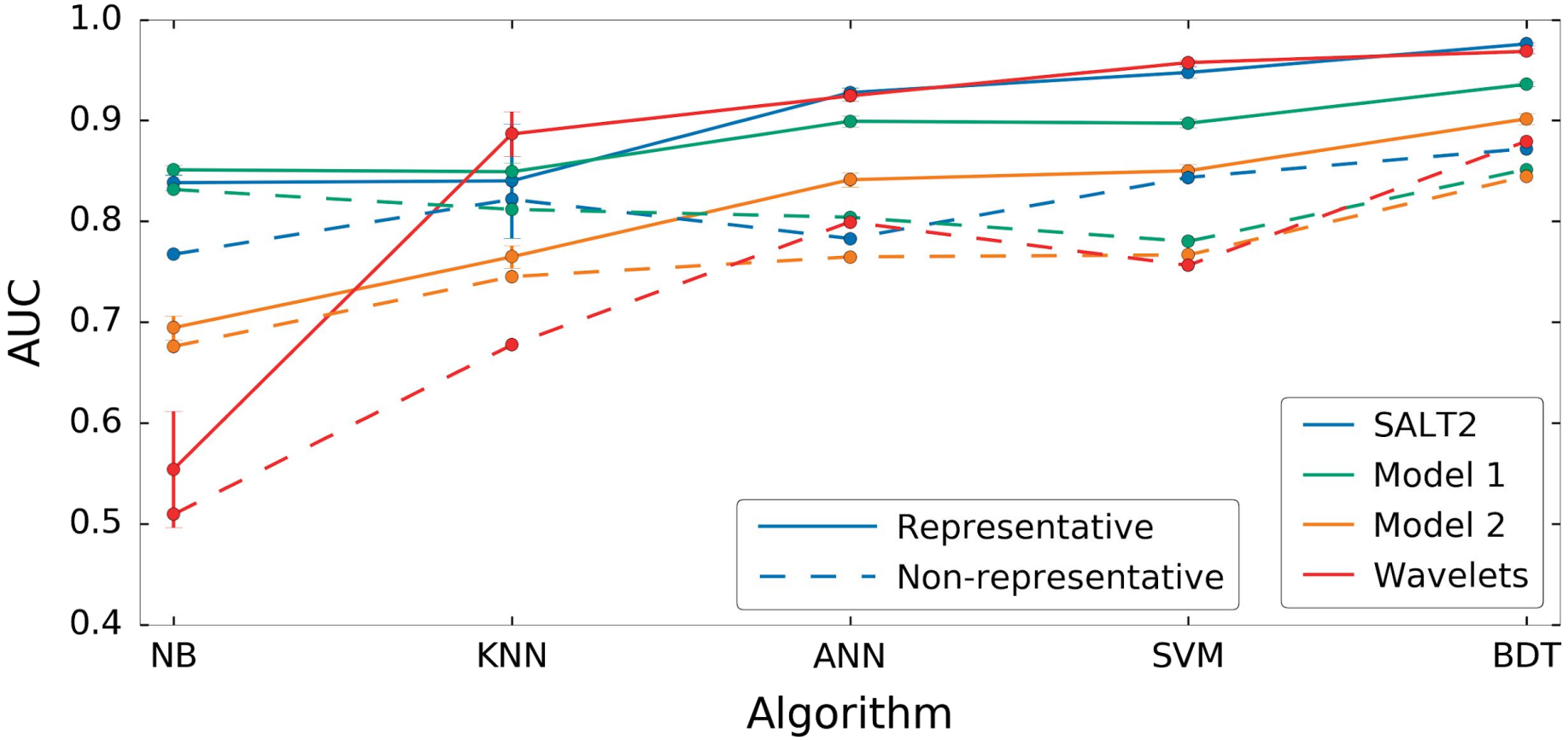
Lochner et al. (2016) - 1603.00882

Machine Learning



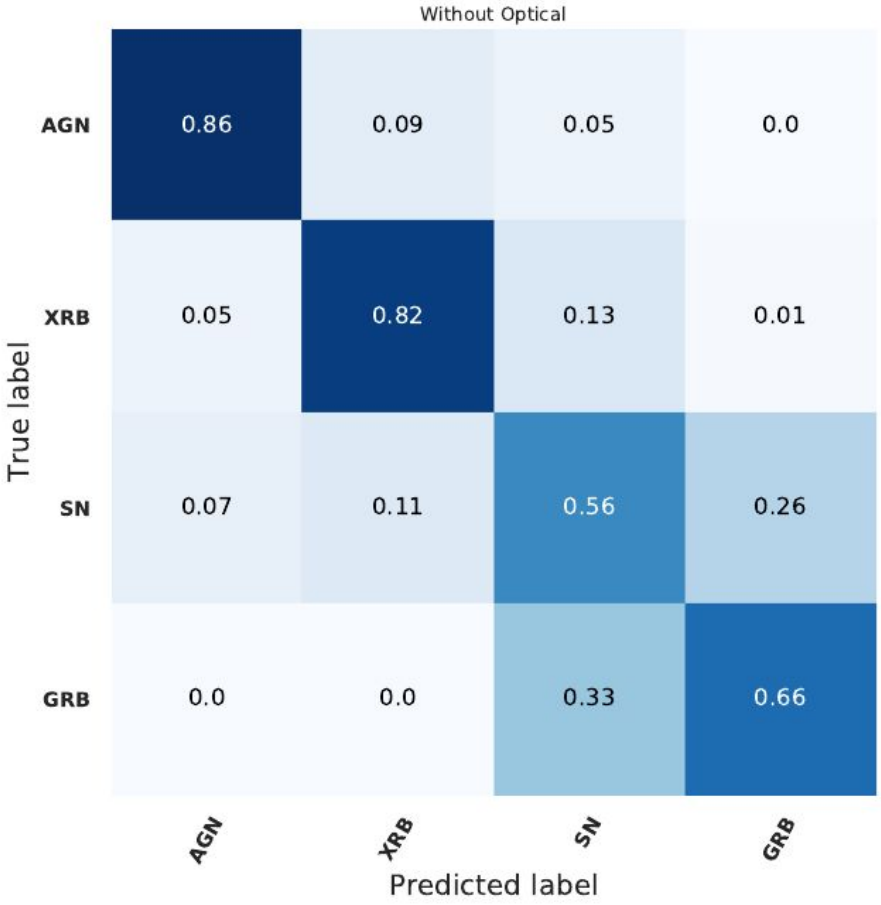
$$\tanh\left[w_5 \tanh(w_1 i_1 + w_2 i_2) + w_6 \tanh(w_3 i_1 + w_4 i_2)\right]$$

Transient Classification

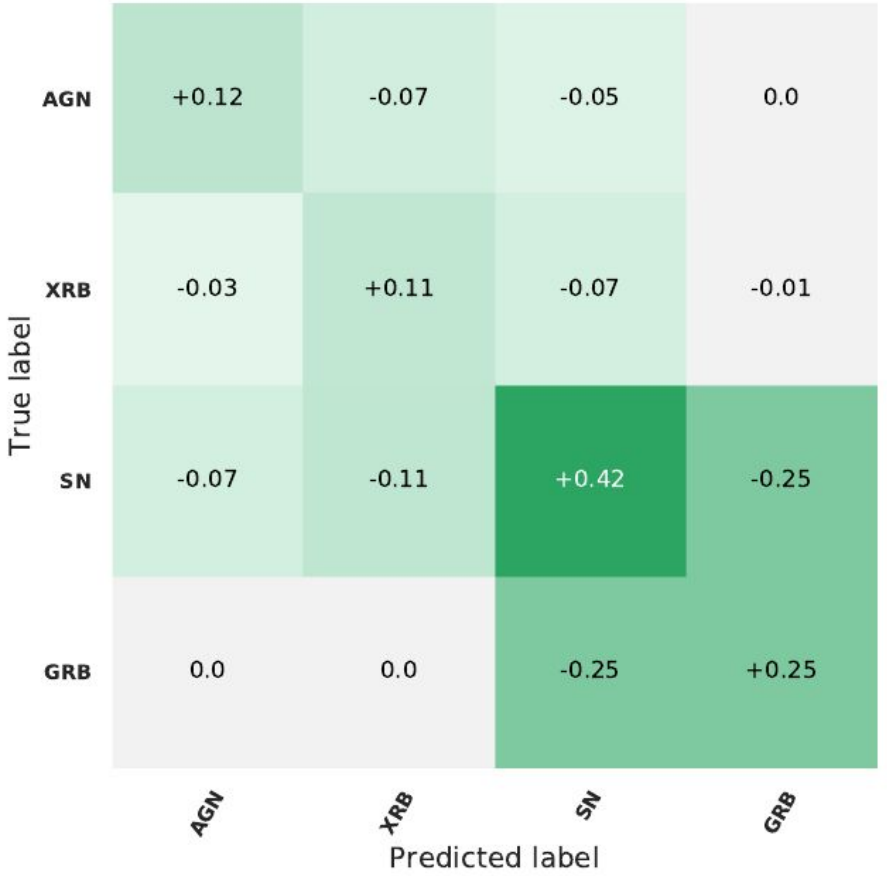


Lochner et al. (2016) - 1603.00882

Multiwavelength Transient Classification



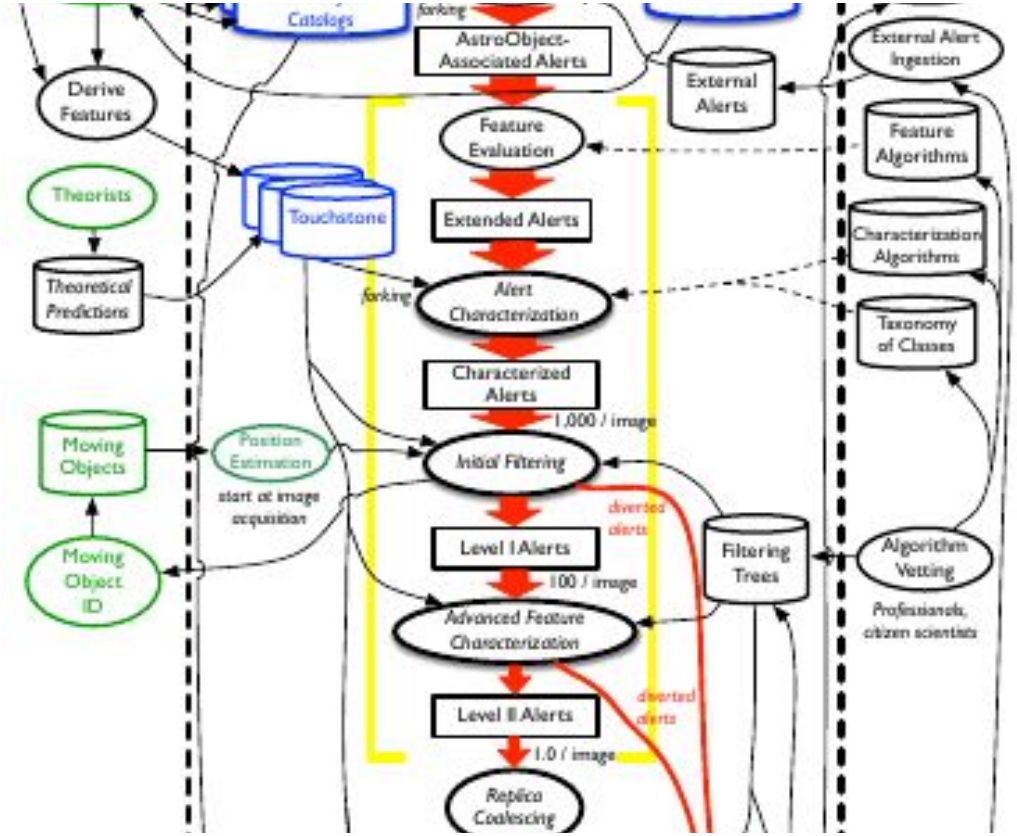
(a) Confusion matrix without optical feature



(b) Confusion matrix showing the difference when optical feature is added

Sooknunan et al. (2018) - 1811.08446

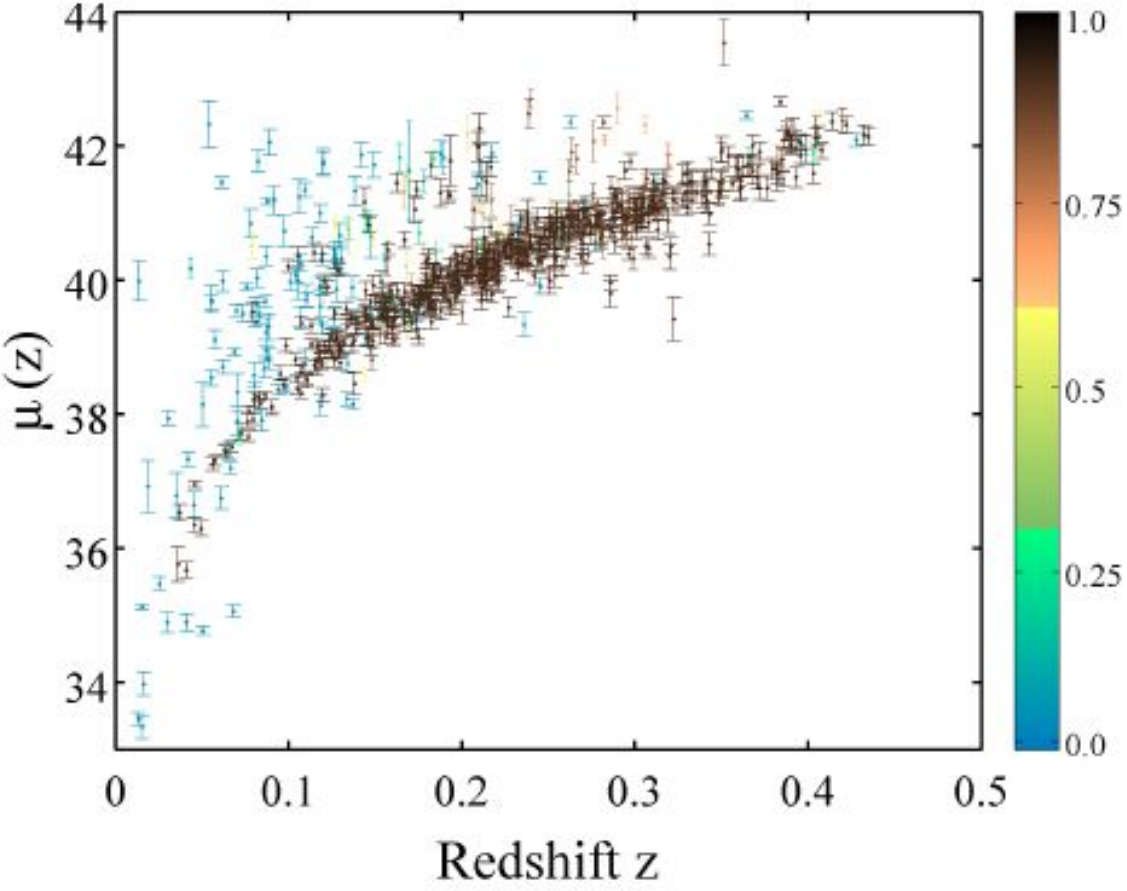
Online Learning with Brokers

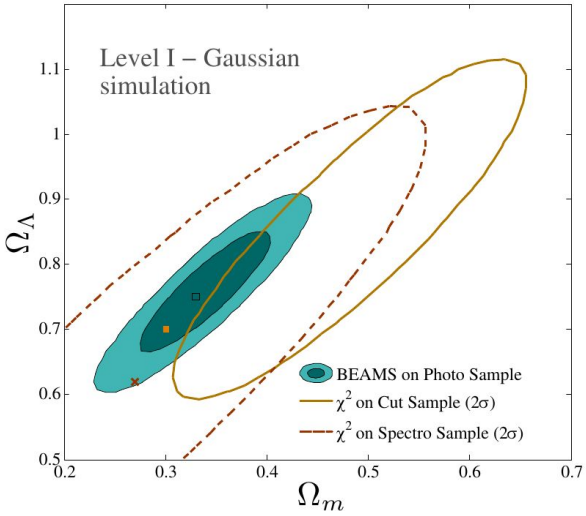


E.g: ANTARES (Saha et al. - 1409.0056,

Narayan et al. - 1801.07323)

Science with Imperfect Classification



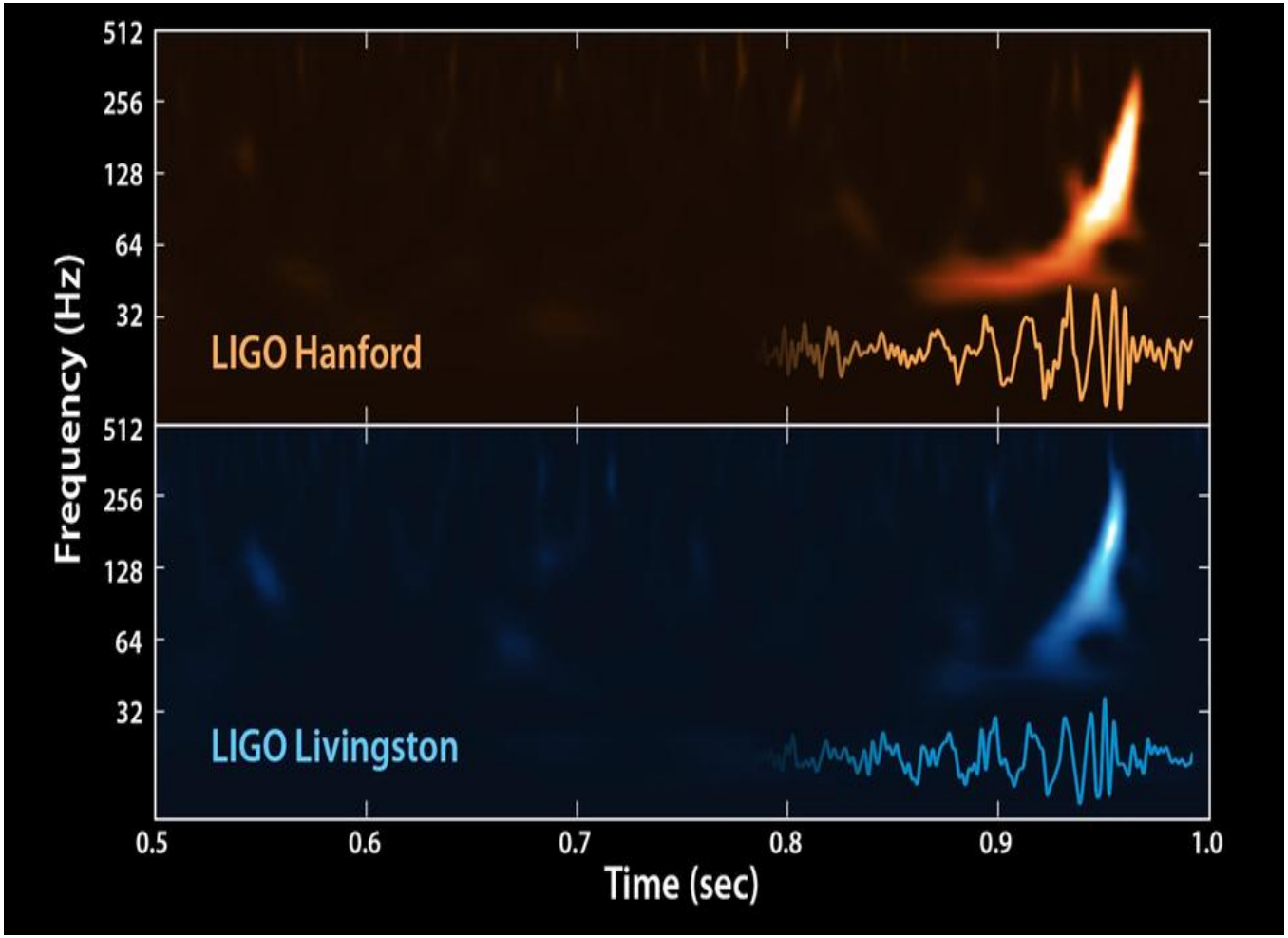


Bayesian Estimation Applied to Multiple Species (BEAMS)

Kunz, Bassett & Hlozek - 0611004
Newling et al. - 1110.6178
Hlozek et al. - 1111.5328
Lochner et al. - 1205.3493
Roberts et al. - 1704.07830

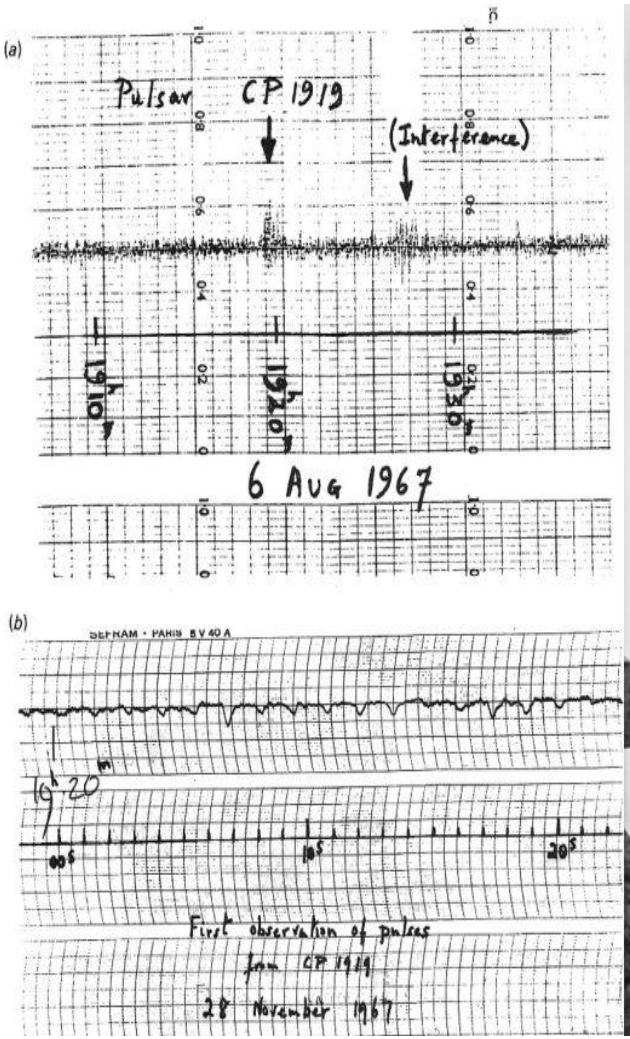
Anomaly Detection

Known Unknowns - rare events



Caltech/MIT/LIGO Lab

Unknown Unknowns - new anomalies

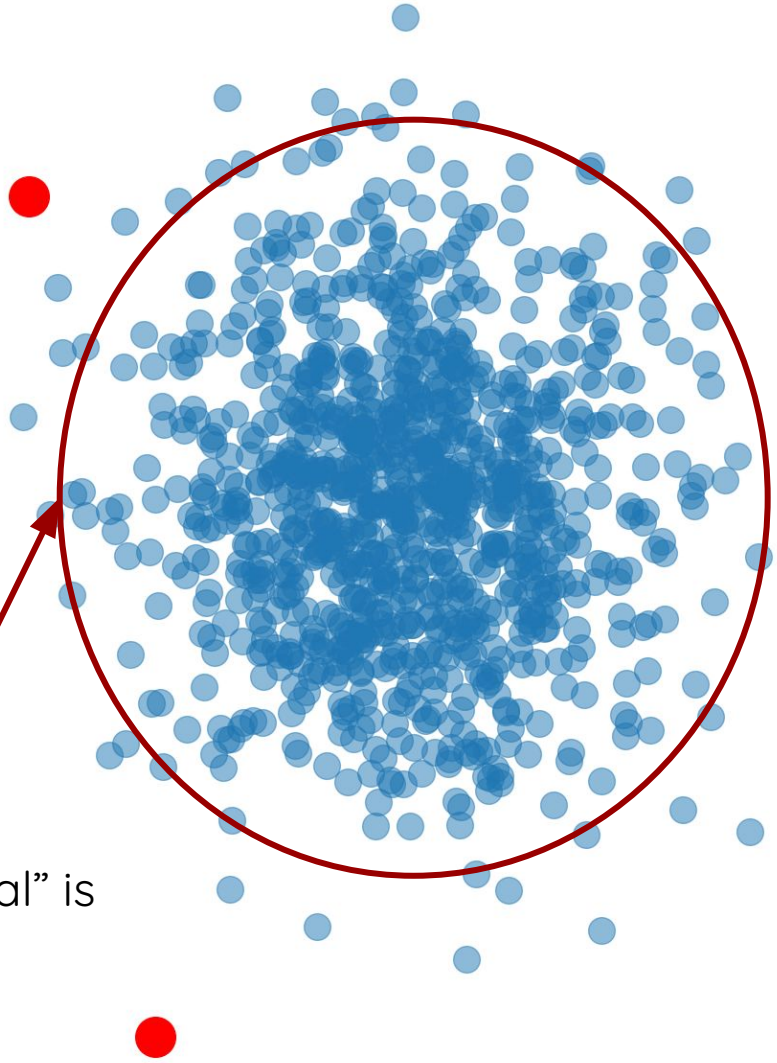


Daily Herald Archive / SSPL / Getty Images

How do we discover new phenomena…

…among 10 million possibilities?

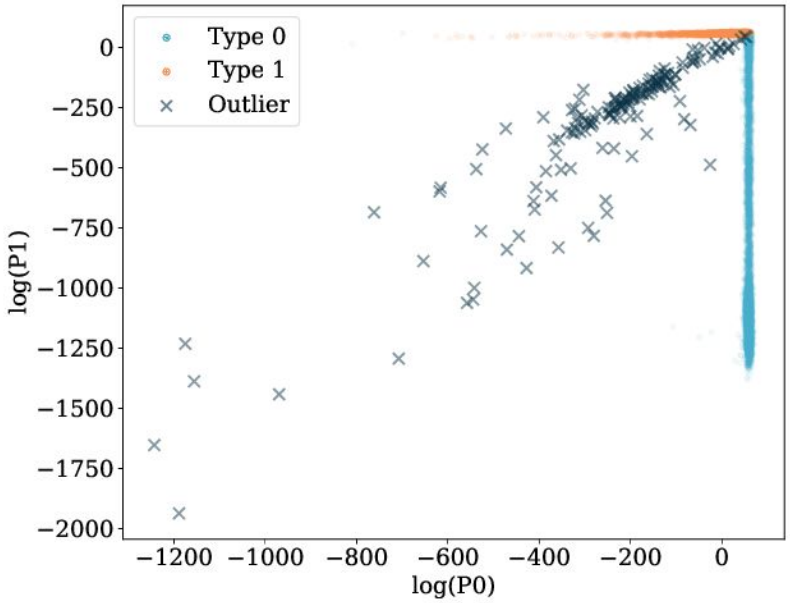
Unsupervised Learning



Try to learn
what “normal” is

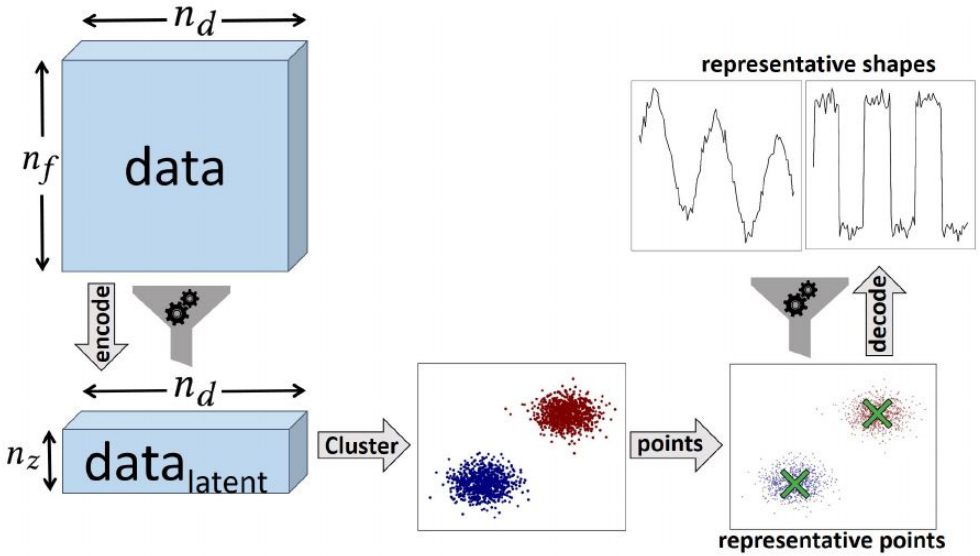
The farther away
from normal the
higher the
anomaly score

Anomaly Detection Algorithms

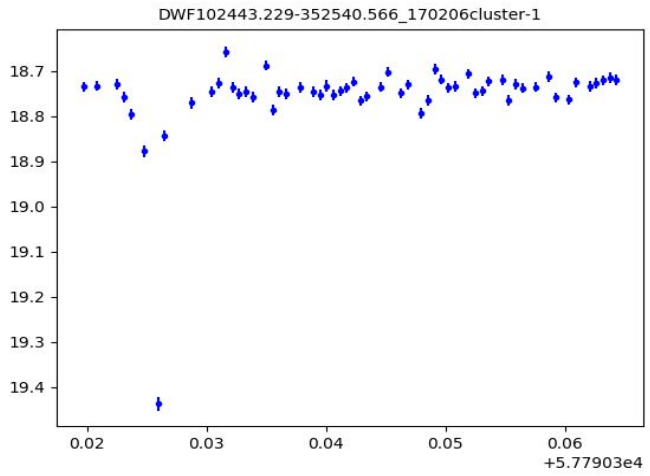


BADAC (Roberts, Bassett & Lochner - 1902.08627)

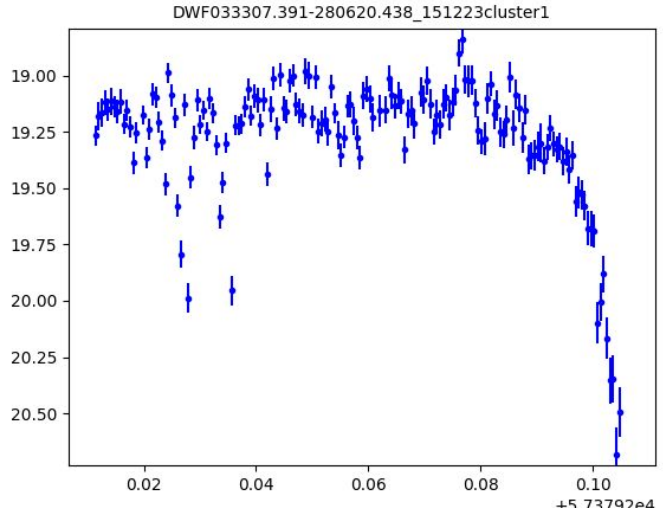
DRAMA (Vafaei Sadr, Bassett & Kunz - 1909.04060)

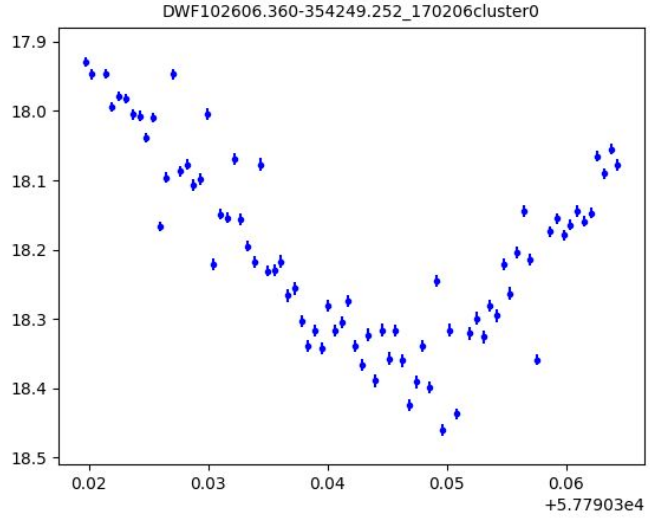


Anomaly Detection Isn't Enough

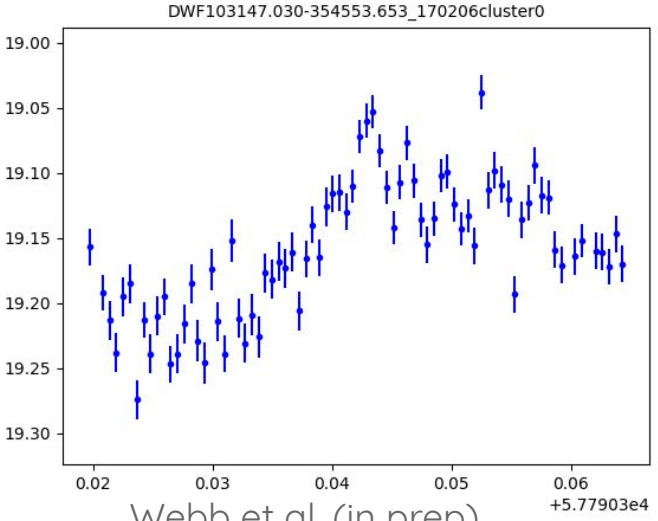


Instrumental effects



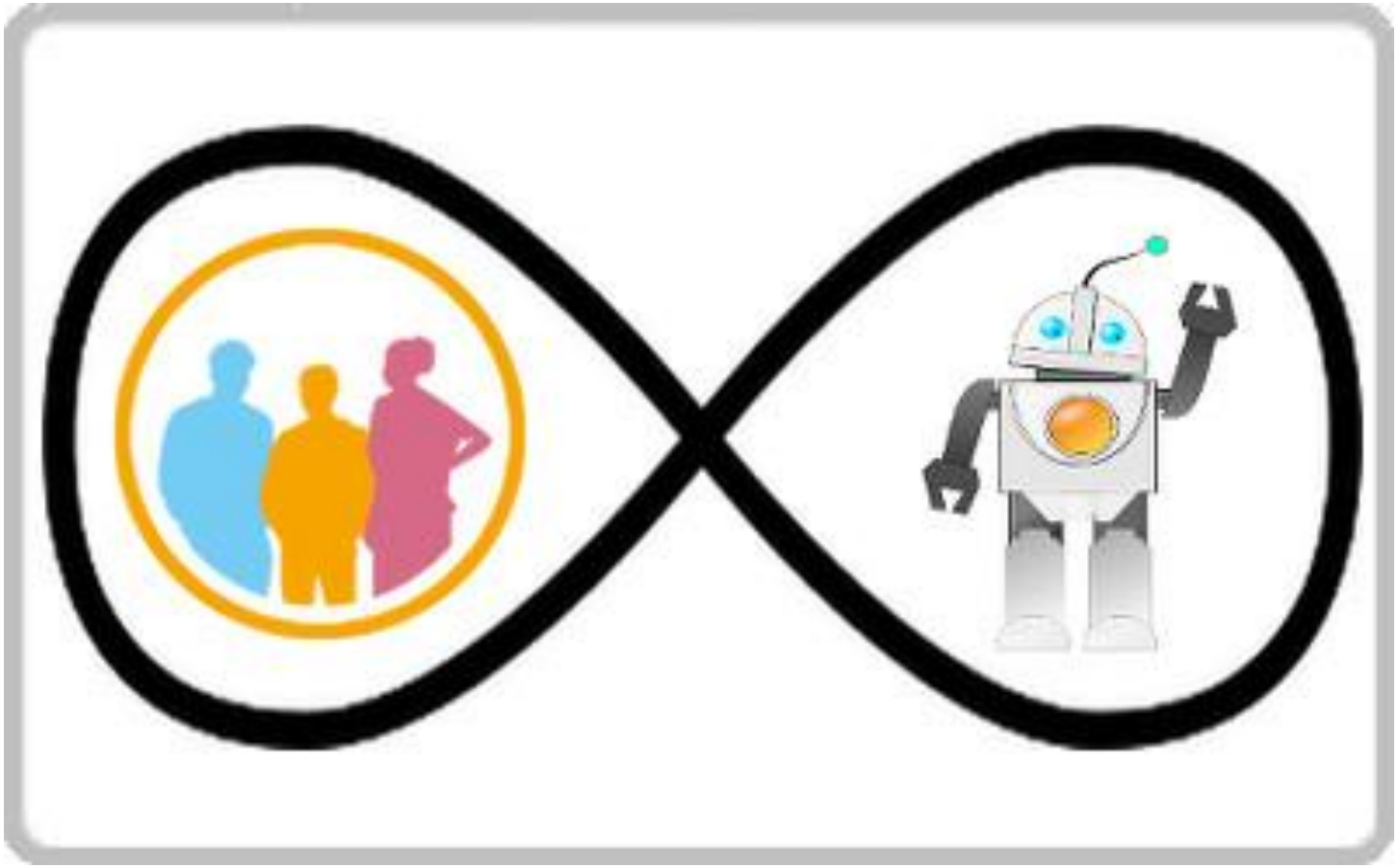


Real sources

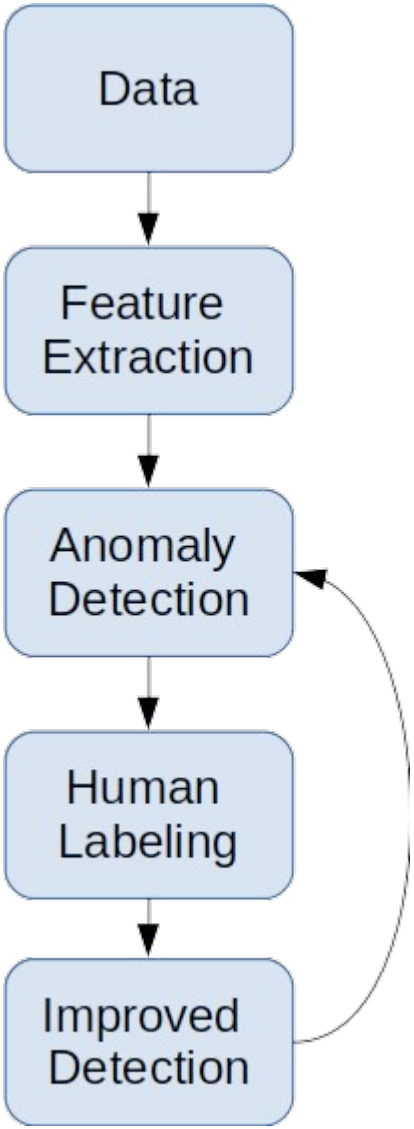


Webb et al. (in prep)

Human-in-the-loop Learning



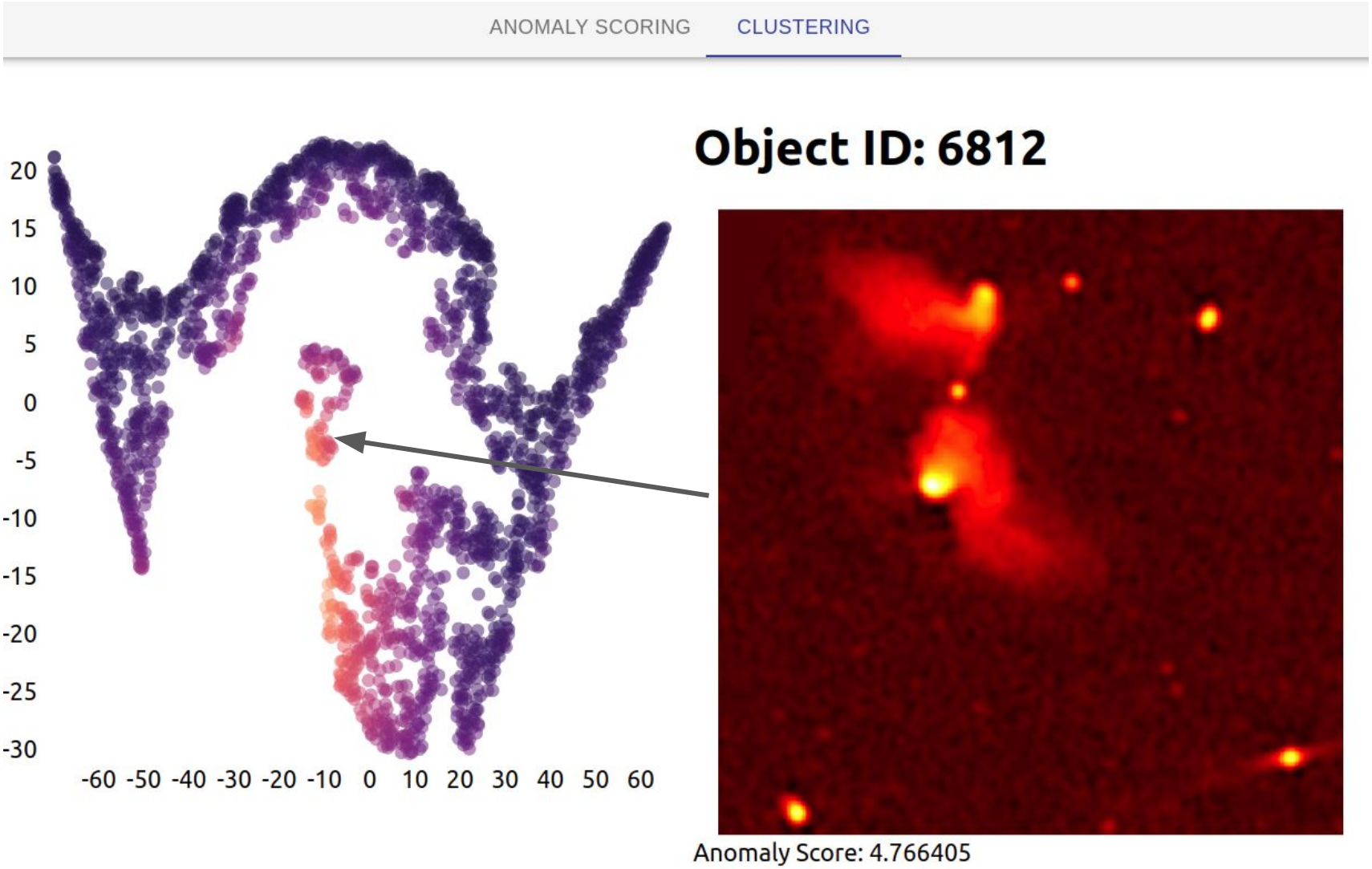
Astronomaly



Lochner and Bassett (in prep)

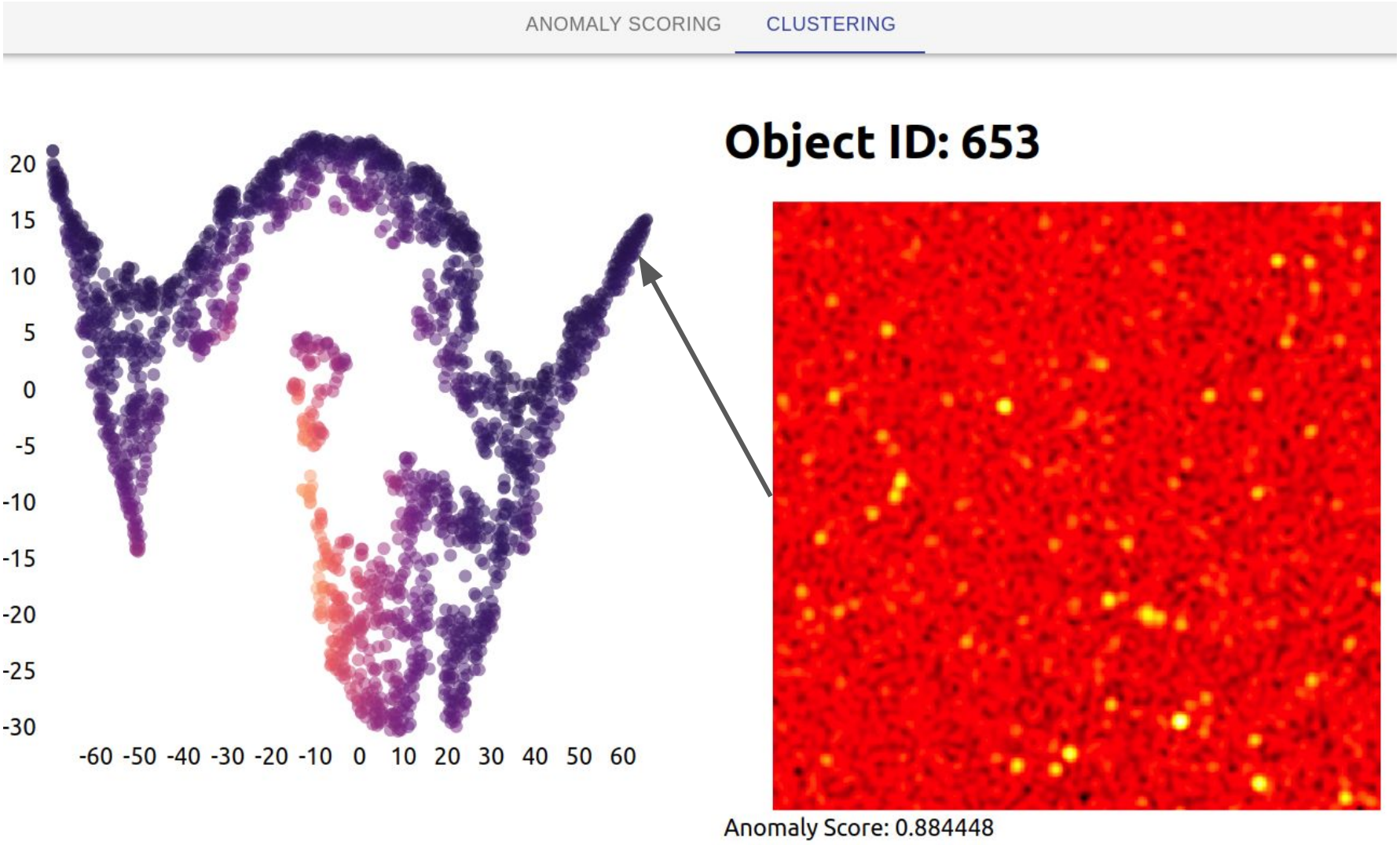
https://github.com/MichelleLochner/astronomaly

Astronomaly



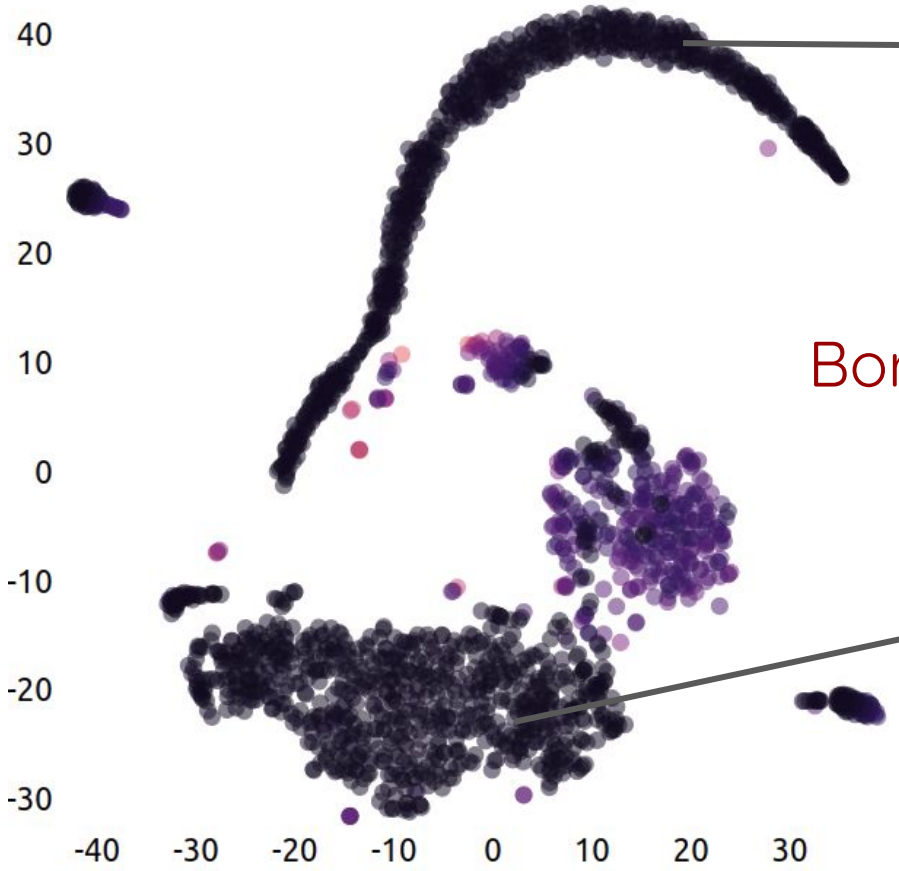
Lochner and Bassett (in prep)

Astronomaly

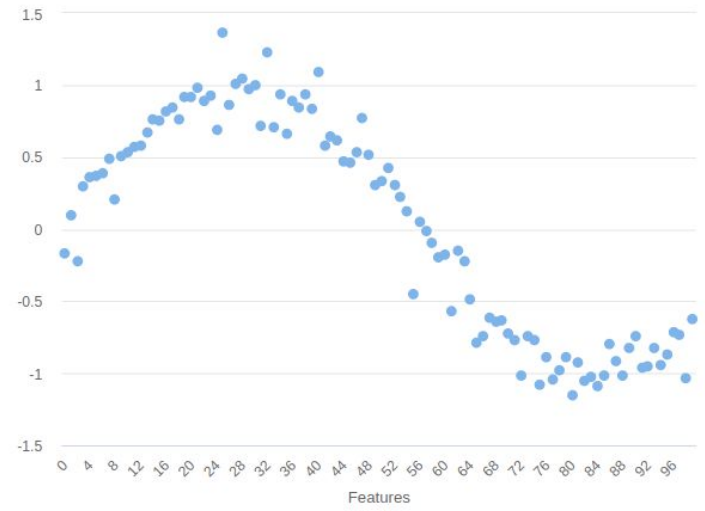


Lochner and Bassett (in prep)

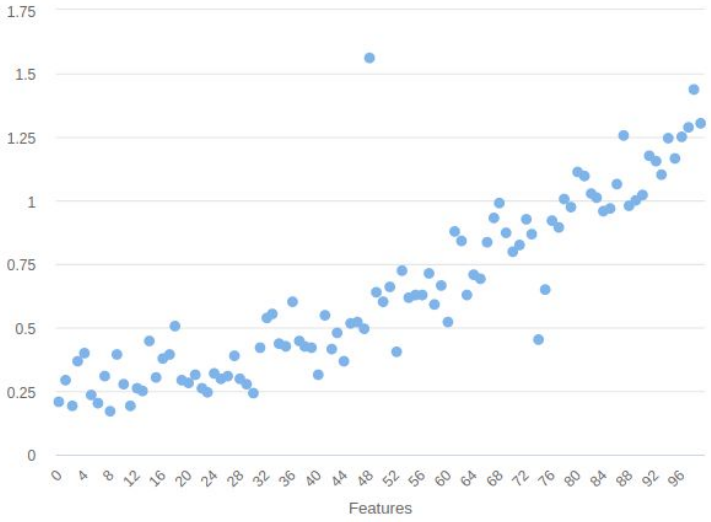
Synthetic data



Boring



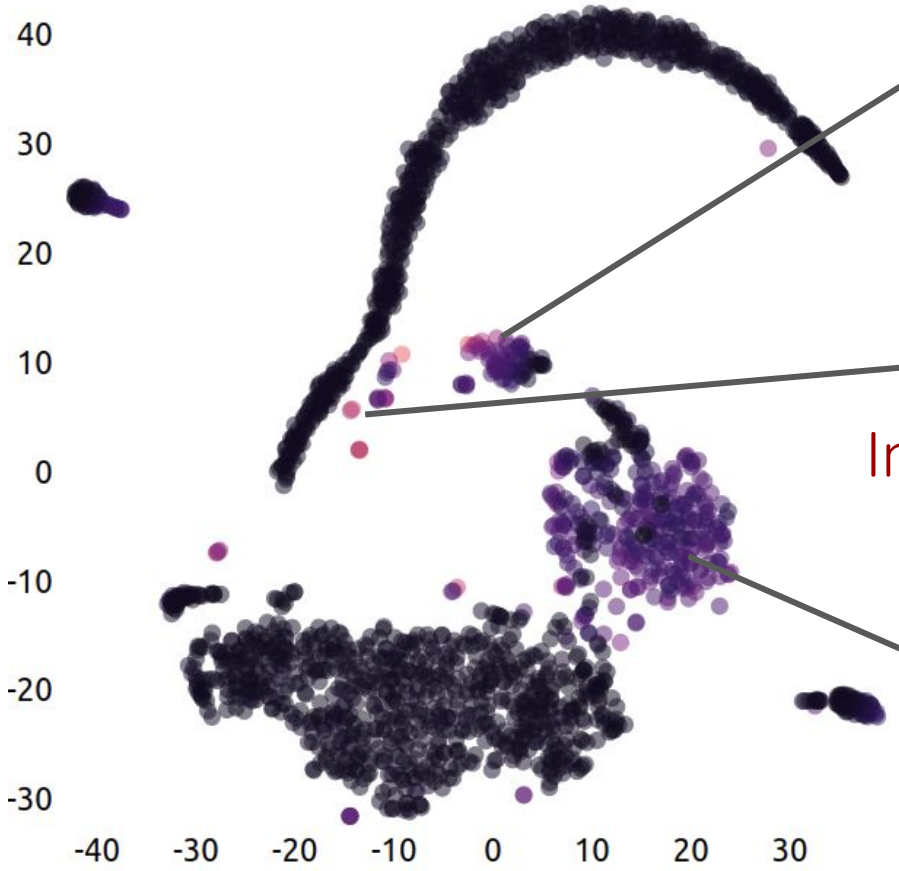
Anomaly Score: 0.270122



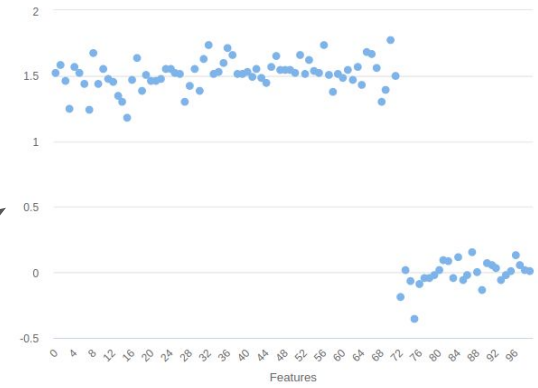
Anomaly Score: 0.241527

Lochner and Bassett (in prep)

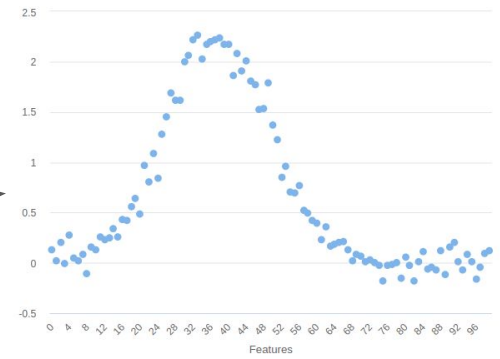
Synthetic data



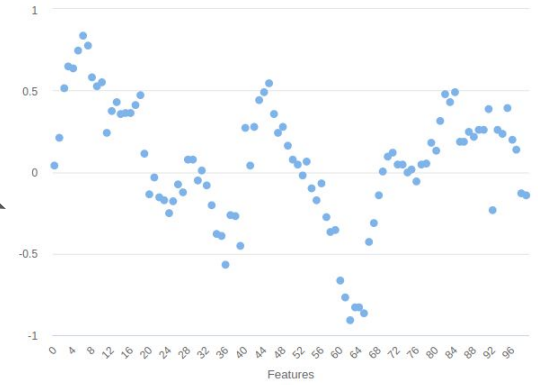
Interesting



Anomaly Score: 3.112012



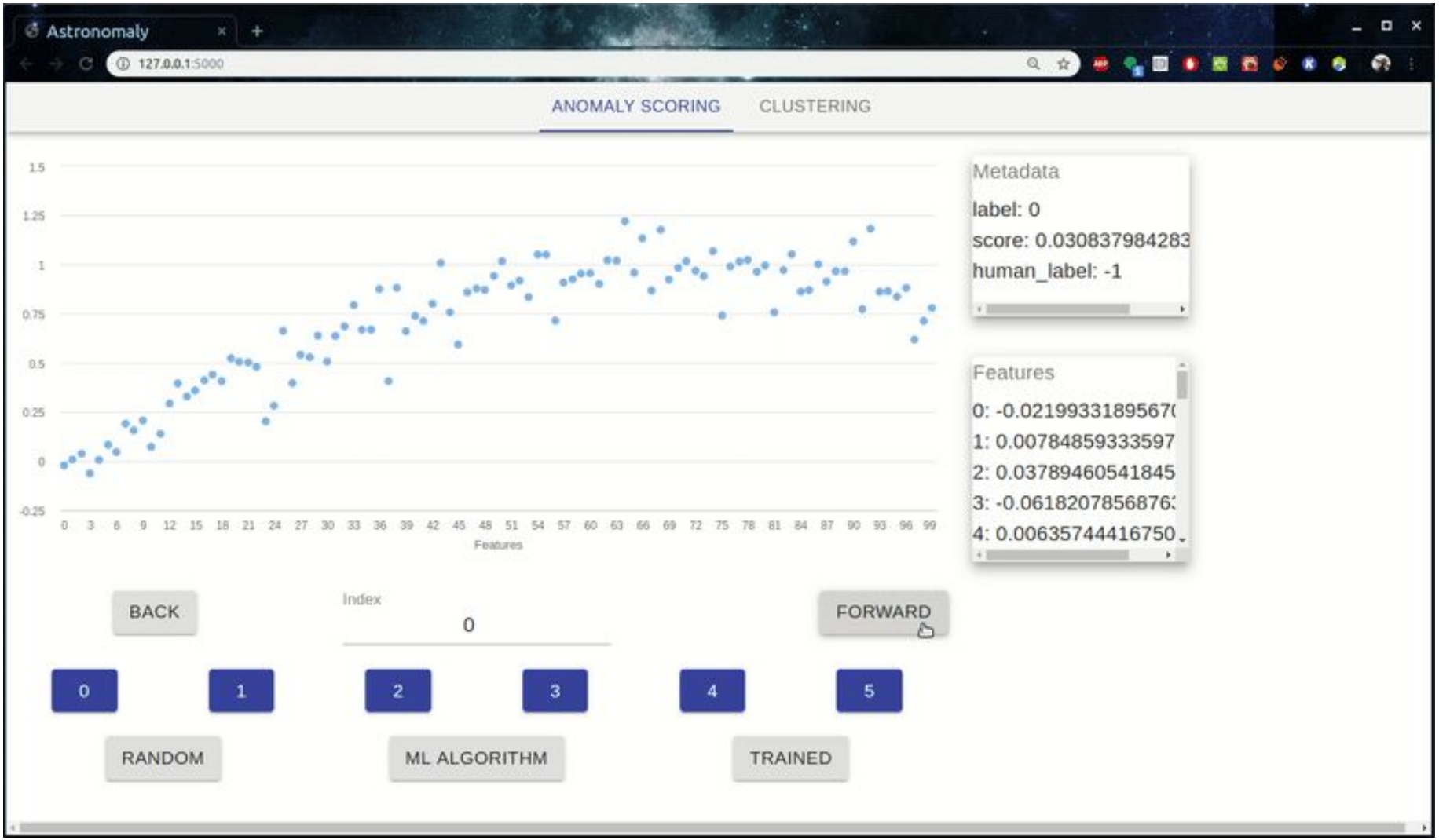
Anomaly Score: 3.92992



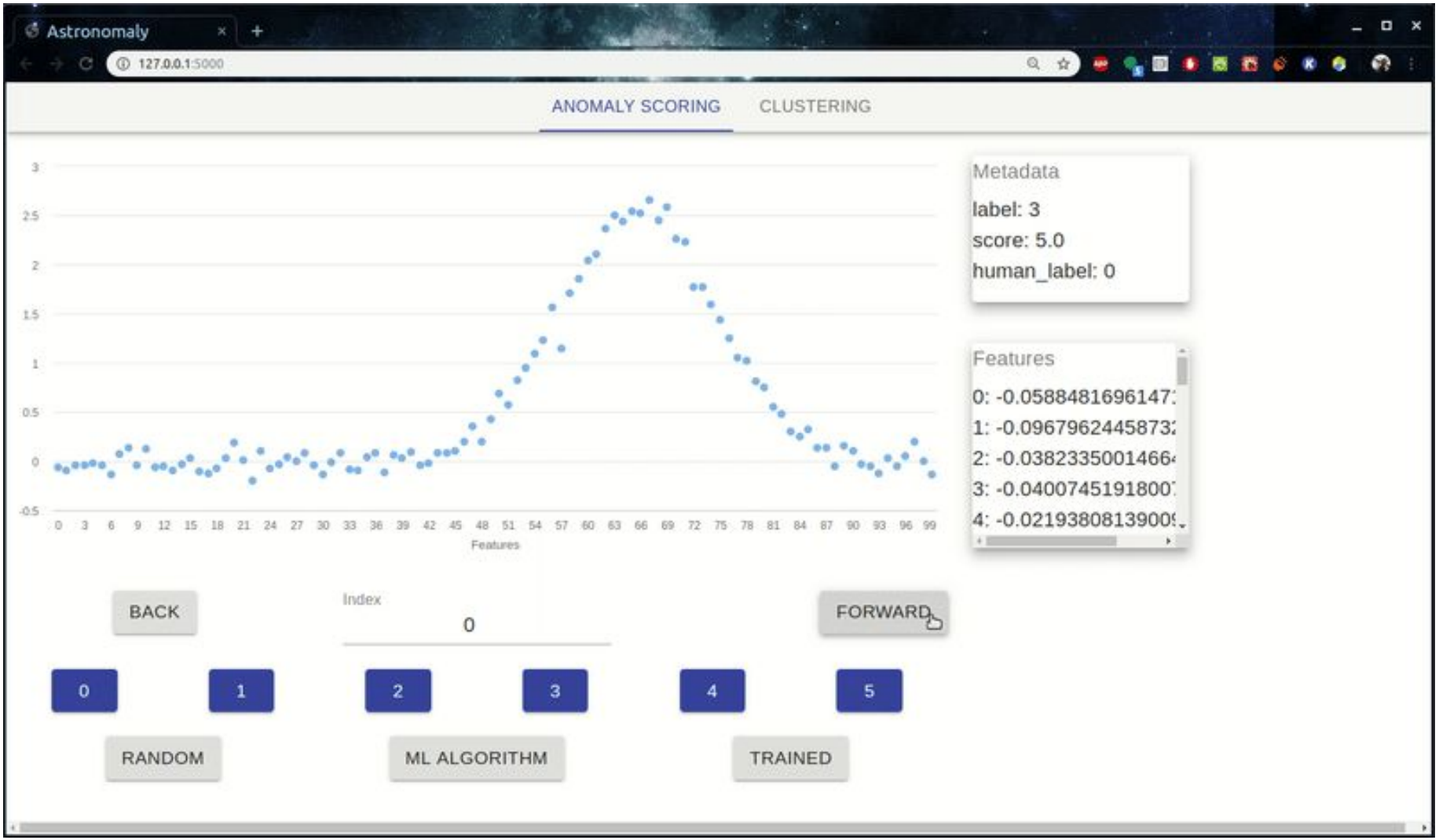
Anomaly Score: 1.303772

Lochner and Bassett (in prep)

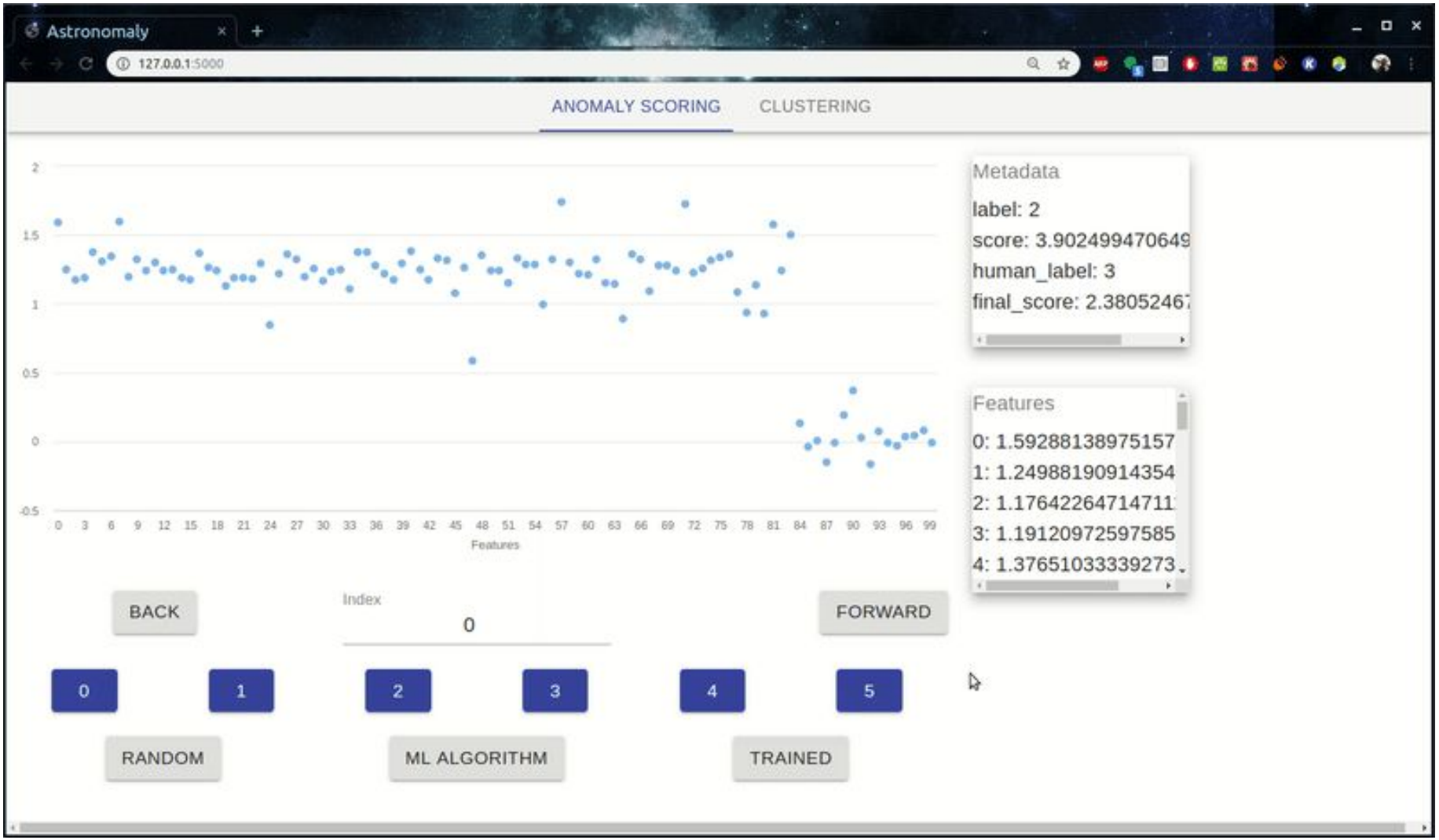
Astronomaly



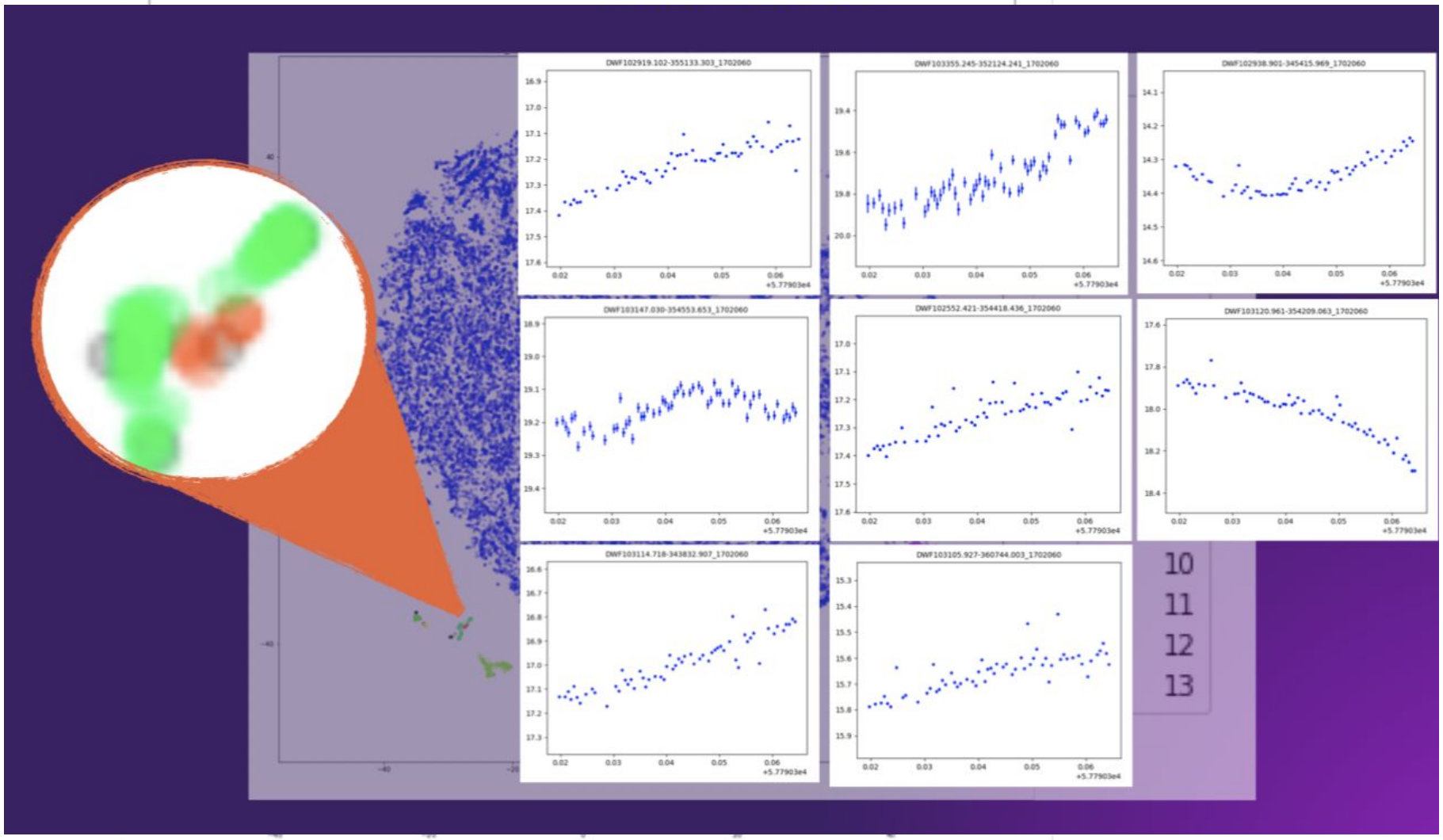
Astronomaly



Astronomaly



Astronomaly Applied to DWF

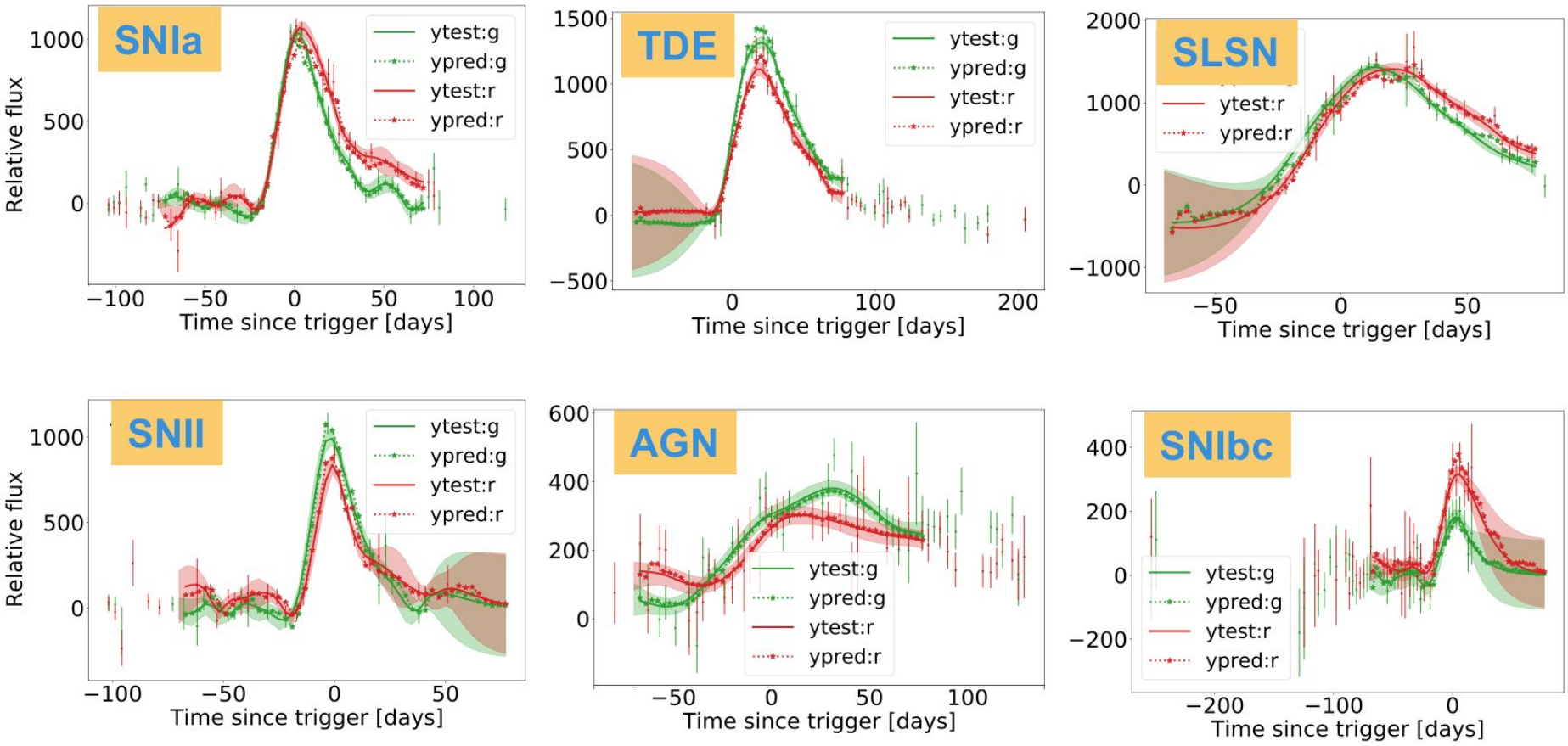


Webb et al. (in prep)

Real-time anomaly detection

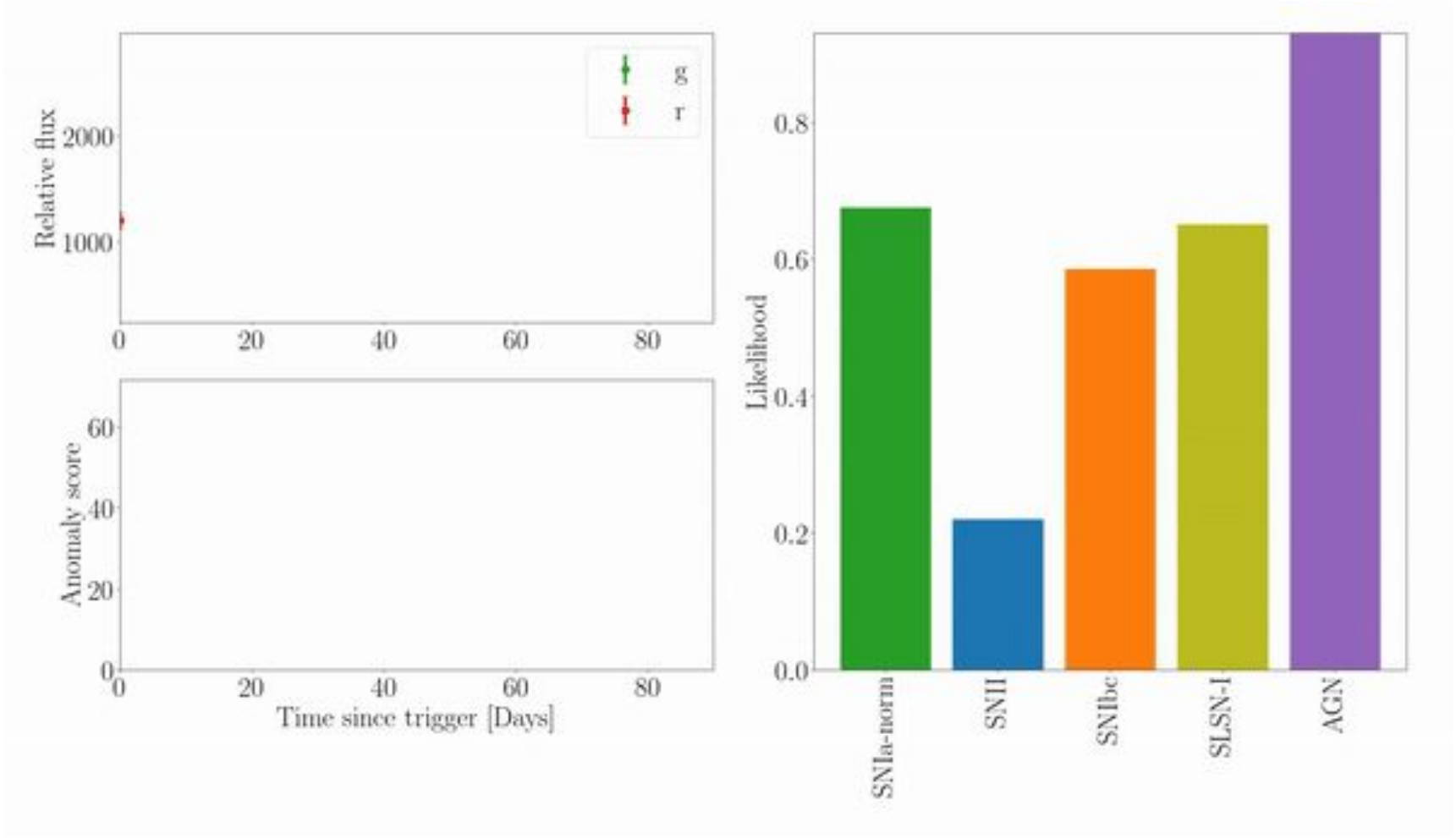
Muthukrishna et al. (in prep)

Real-time anomaly detection



Muthukrishna et al. (in prep)

Real-time anomaly detection



Muthukrishna et al. (in prep)

Talking Points

- No ML is perfect - how do we as a community move towards dealing with probabilistic catalogues for things like SN cosmology?
- Non-representivity remains a huge problem for ML algorithms - how do we build better training sets?
- How do we avoid a high false positive rate of anomaly detection in the low SNR regime?
- When 10 000 anomalous objects are detected each night, how do we coordinate the community?
- How do we use ML to automatically follow-up targets intelligently?