Anomaly Detection in Astronomy

Michelle Lochner

Senior Lecturer

University of the Western Cape/
South African Radio Astronomy
Observatory

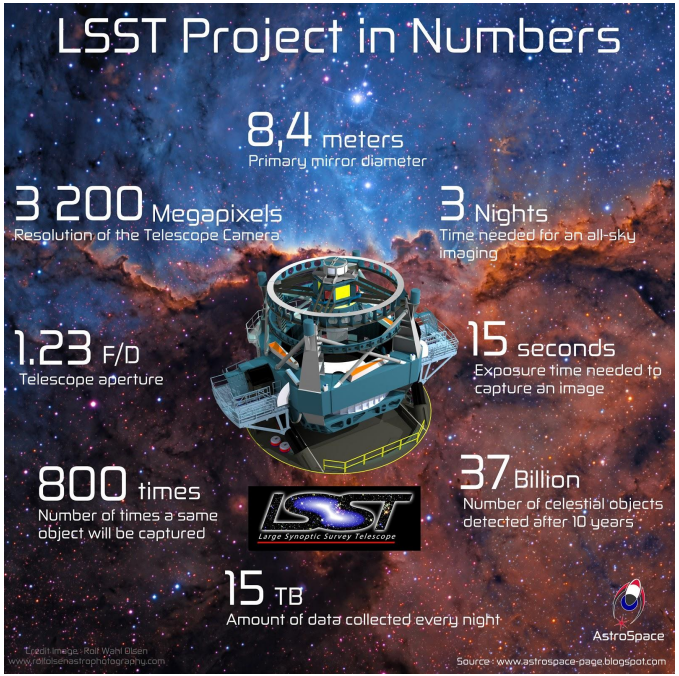
UNIVERSITY of the WESTERN CAPE

CRC

NRF National Research Foundation | SARAO South African Radio Astronomy Observatory

The Vera C. Rubin Observatory

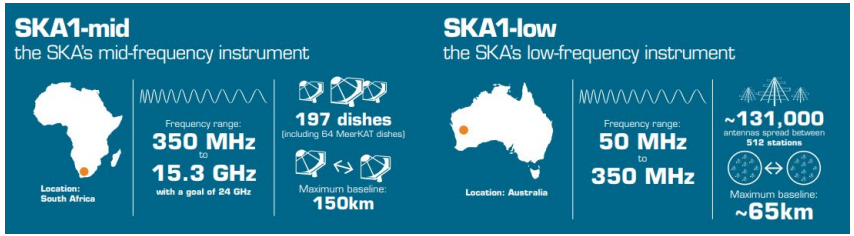
Legacy Survey of Space and Time (LSST)



Rubin Obs/NSF/AURA & Bruno C. Quint

Vera C. Rubin Observatory

The Square Kilometre Array



Square Kilometre Array Observatory

MeerKAT

SARAO

ASKAP

CSIRO

LOFAR

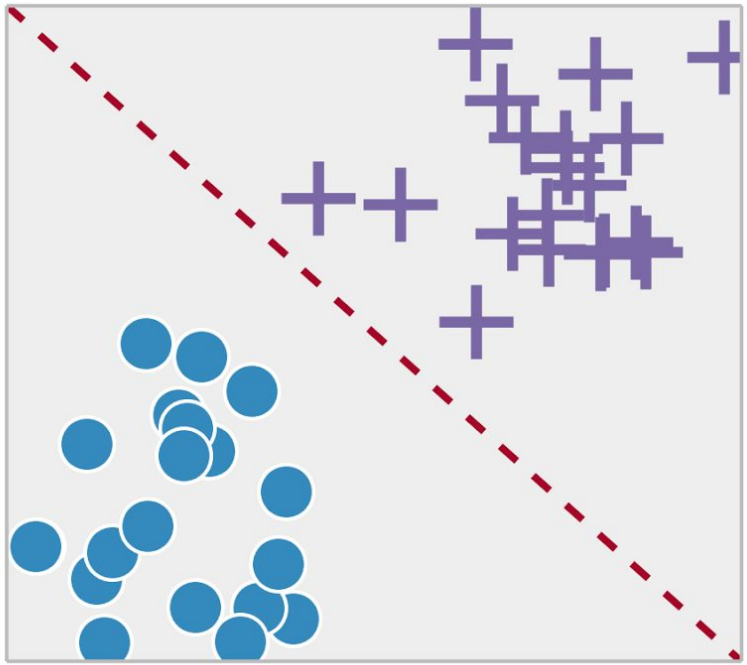
LOFAR

We’re facing a data explosion

Harvard Business Review

Machine Learning

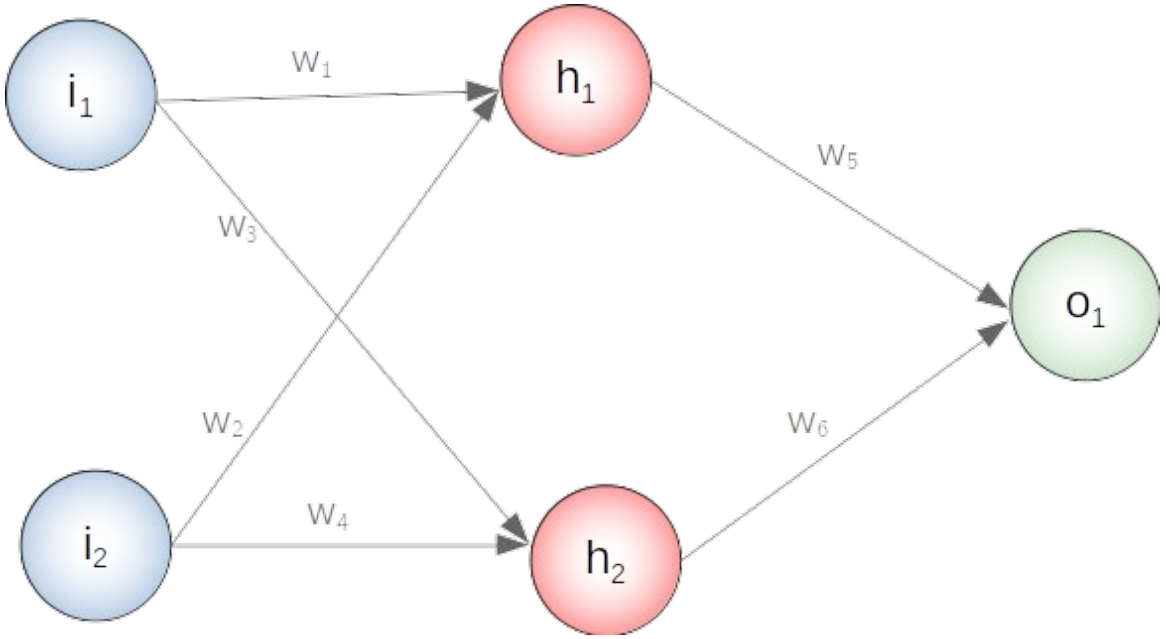
Supervised Machine Learning



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

https://towardsdatascience.com

Machine Learning

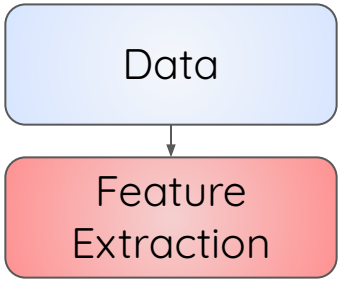


$$\tanh\left[w_5 \tanh(w_1 i_1 + w_2 i_2) + w_6 \tanh(w_3 i_1 + w_4 i_2))\right]$$

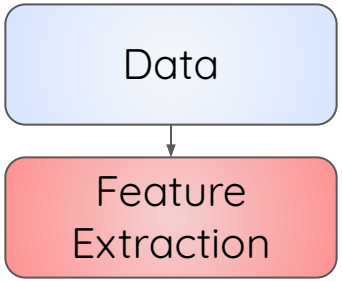
Some Definitions

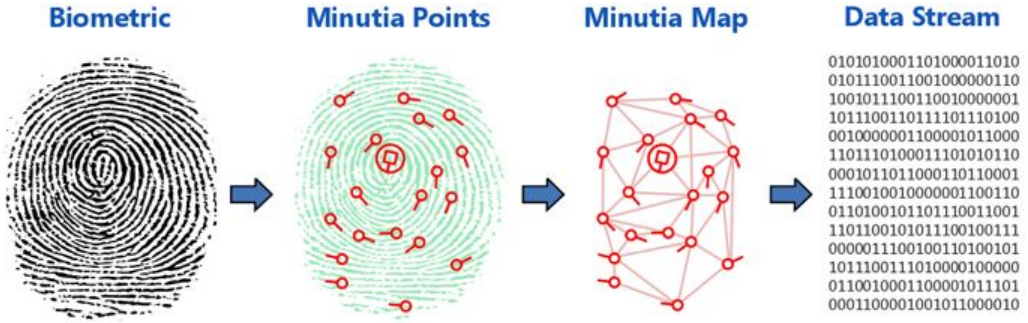
Some Definitions

Some Definitions

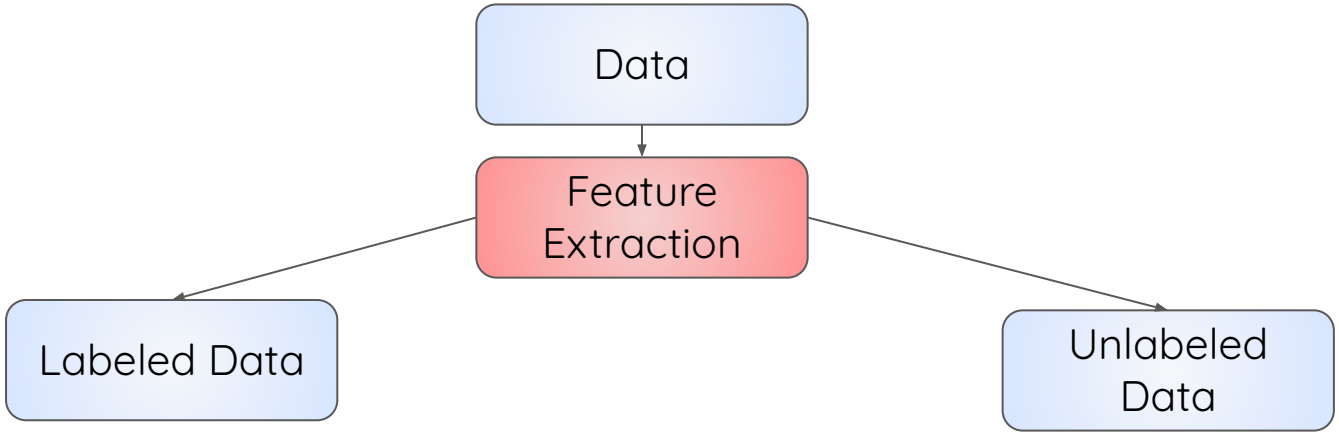


Some Definitions

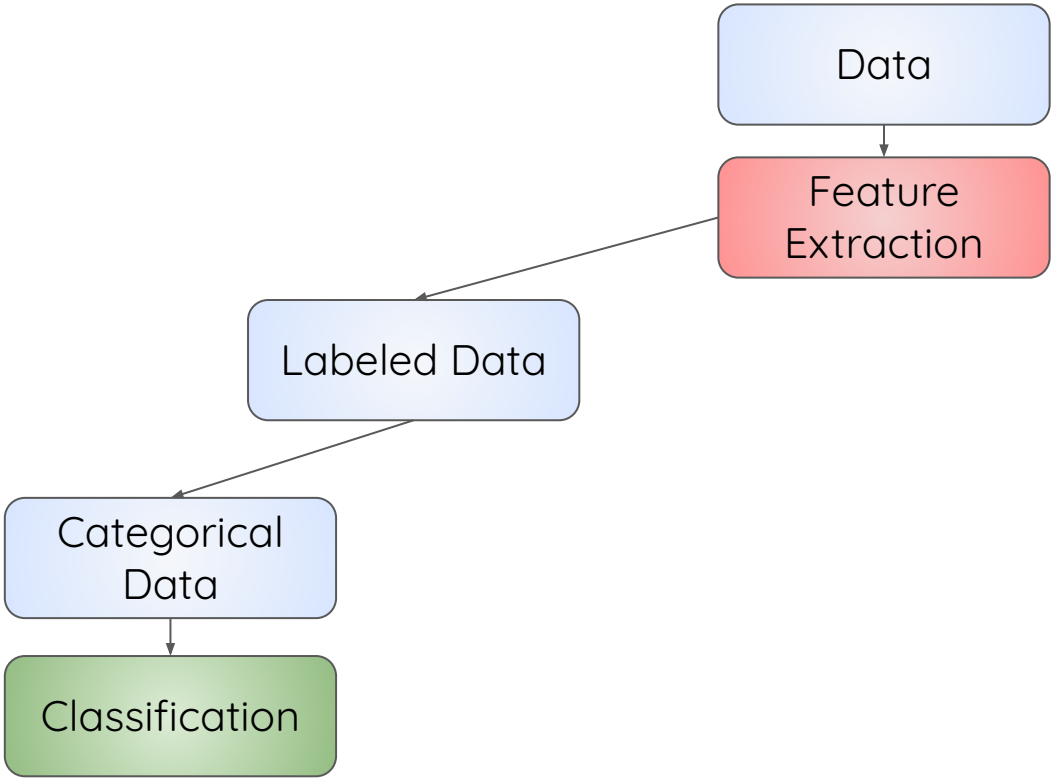




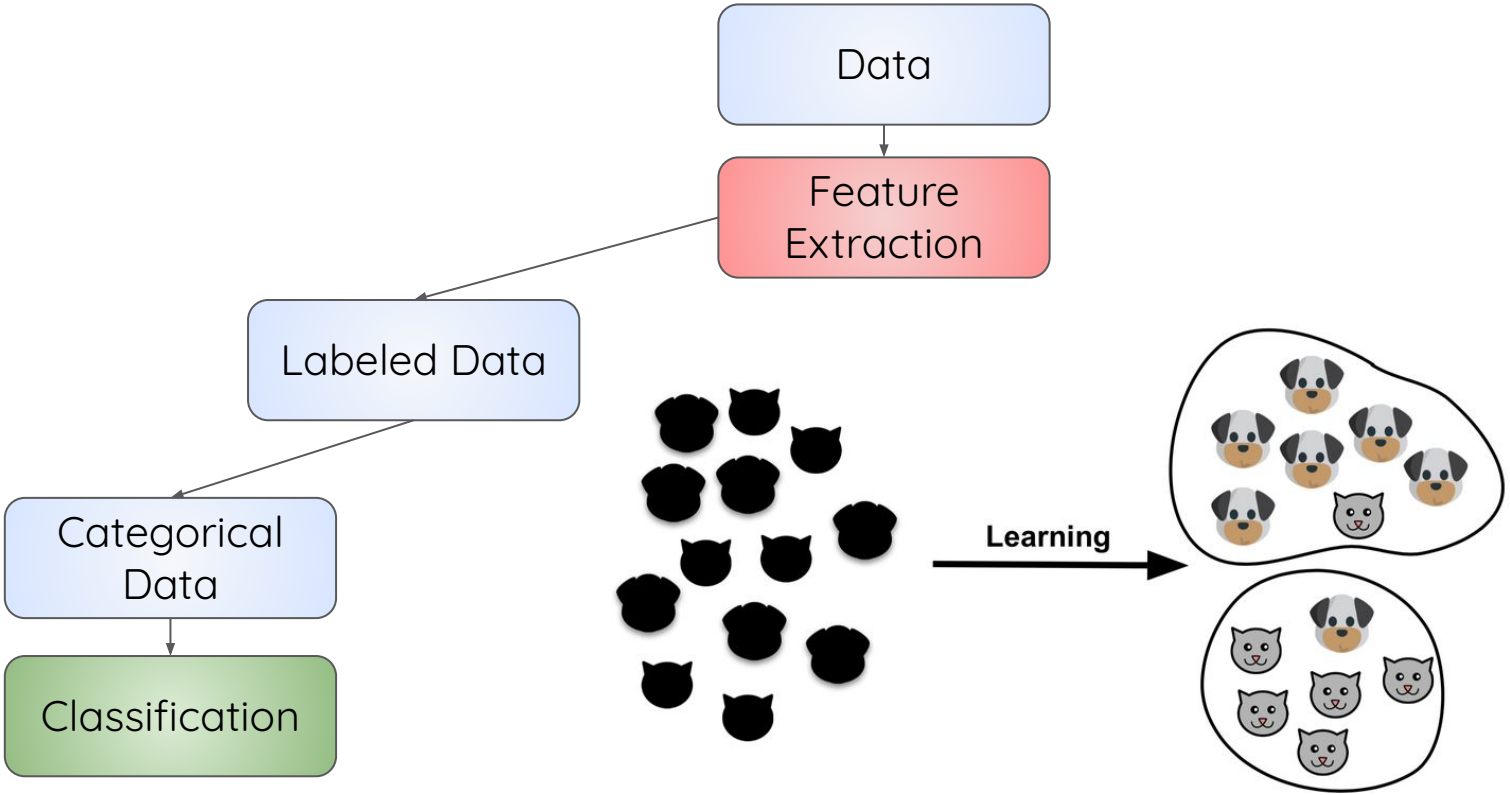
Some Definitions



Some Definitions

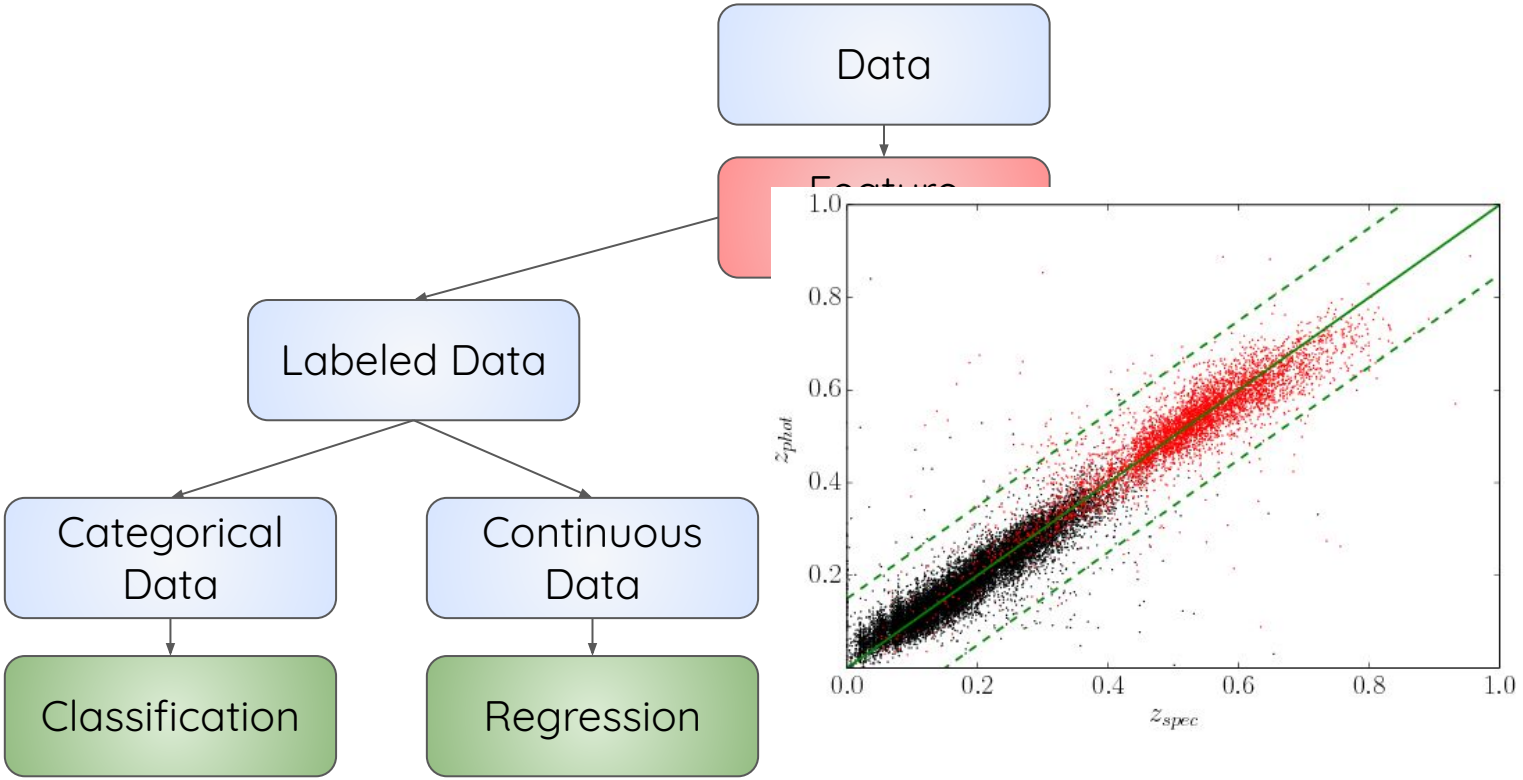


Some Definitions



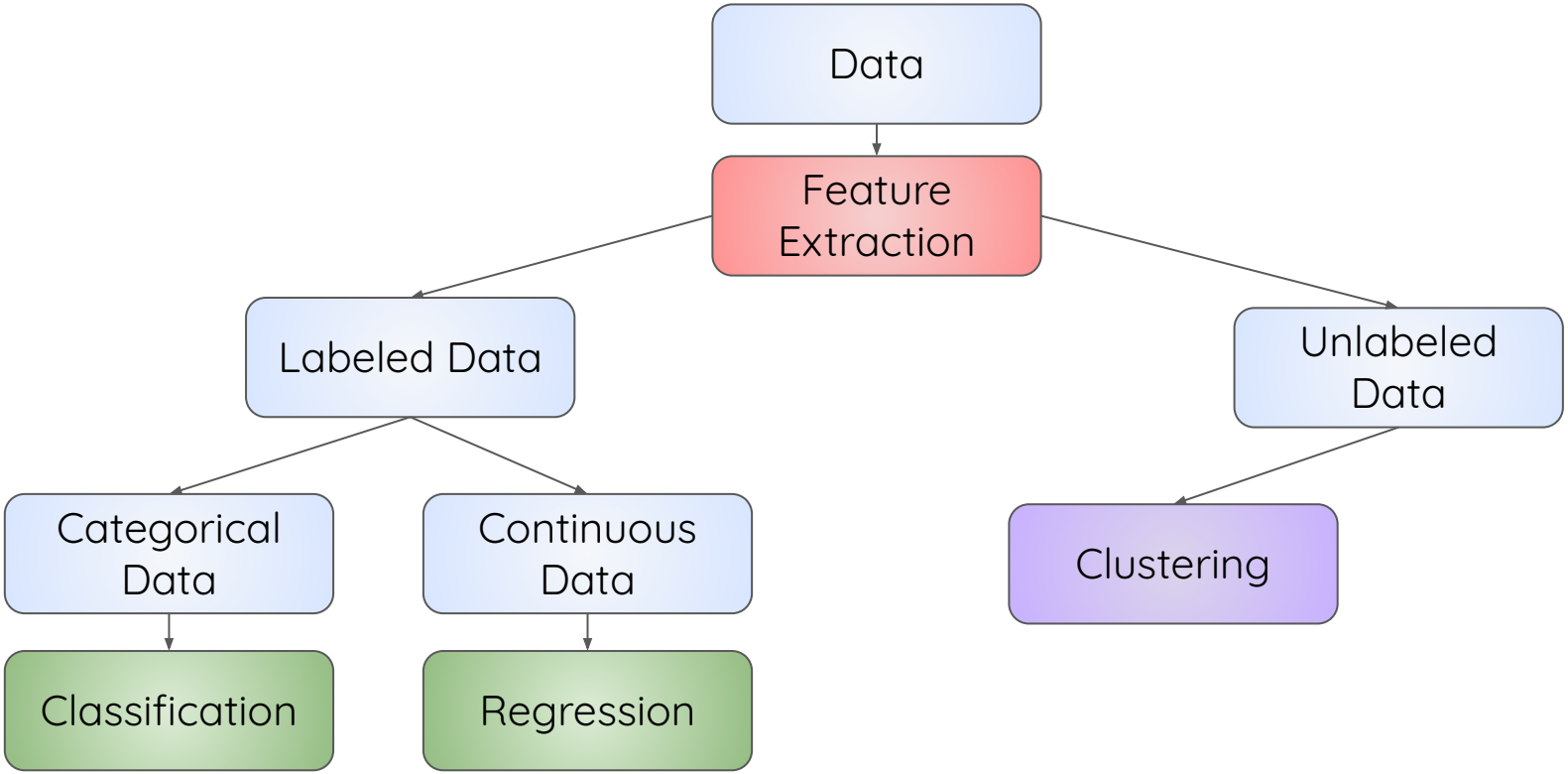
Goyal 2018

Some Definitions

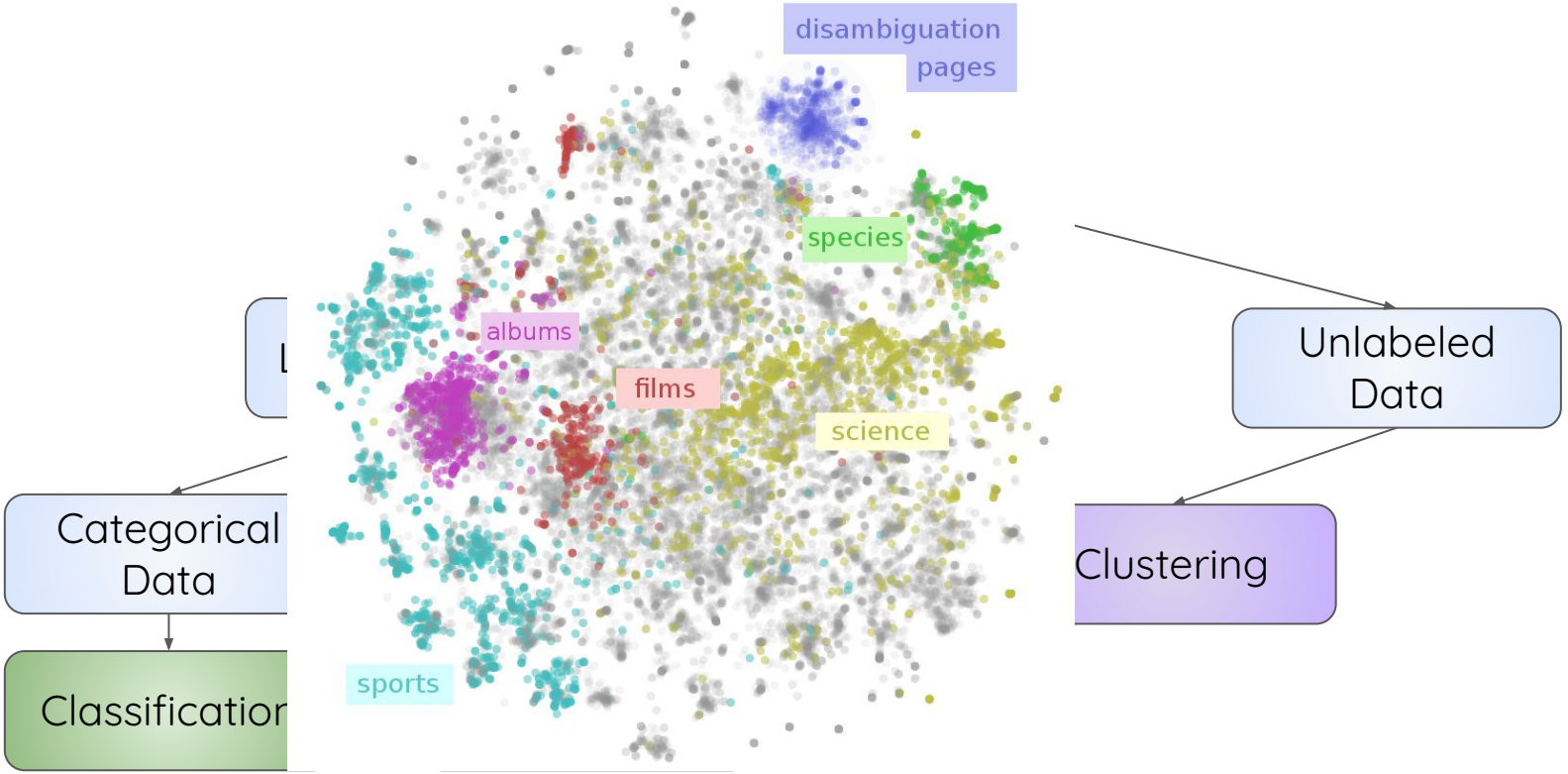


de Jong et al. (2015)

Some Definitions

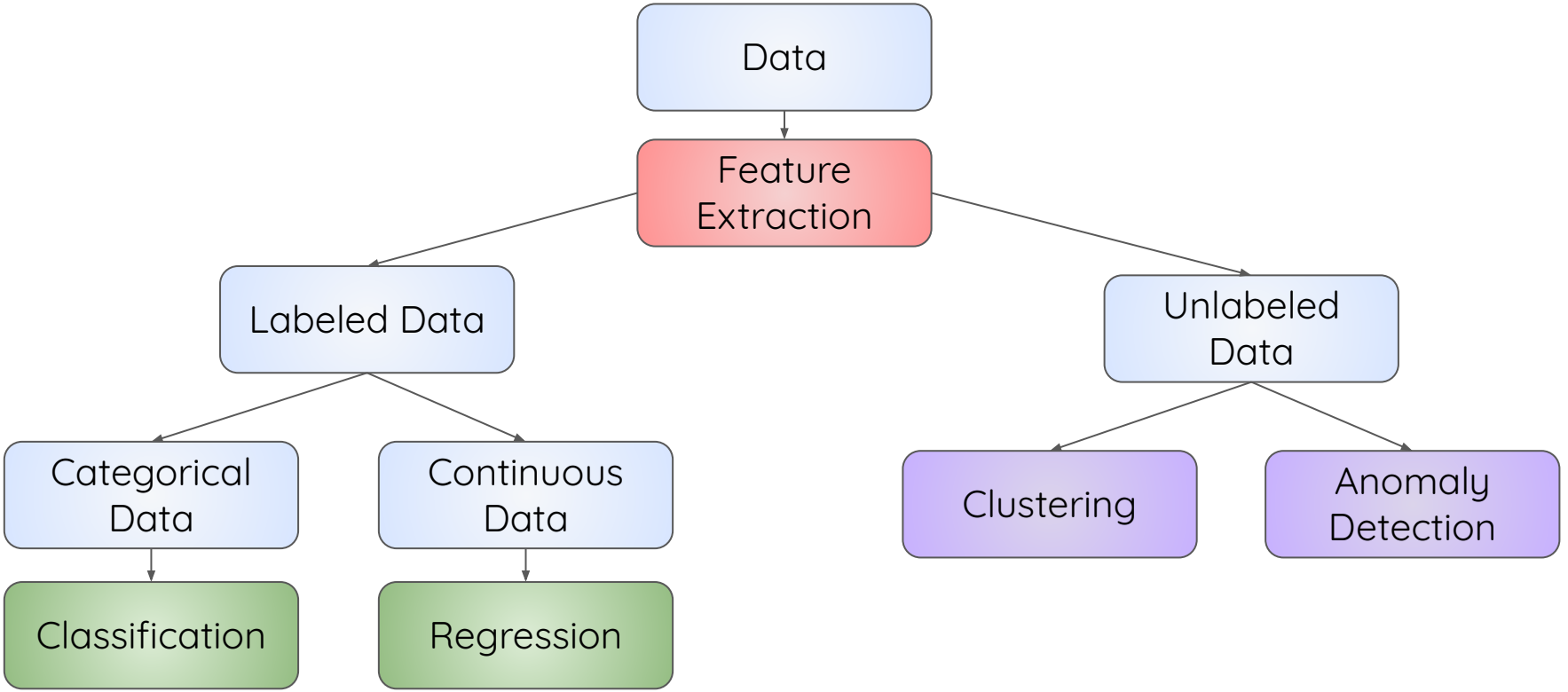


Some Definitions

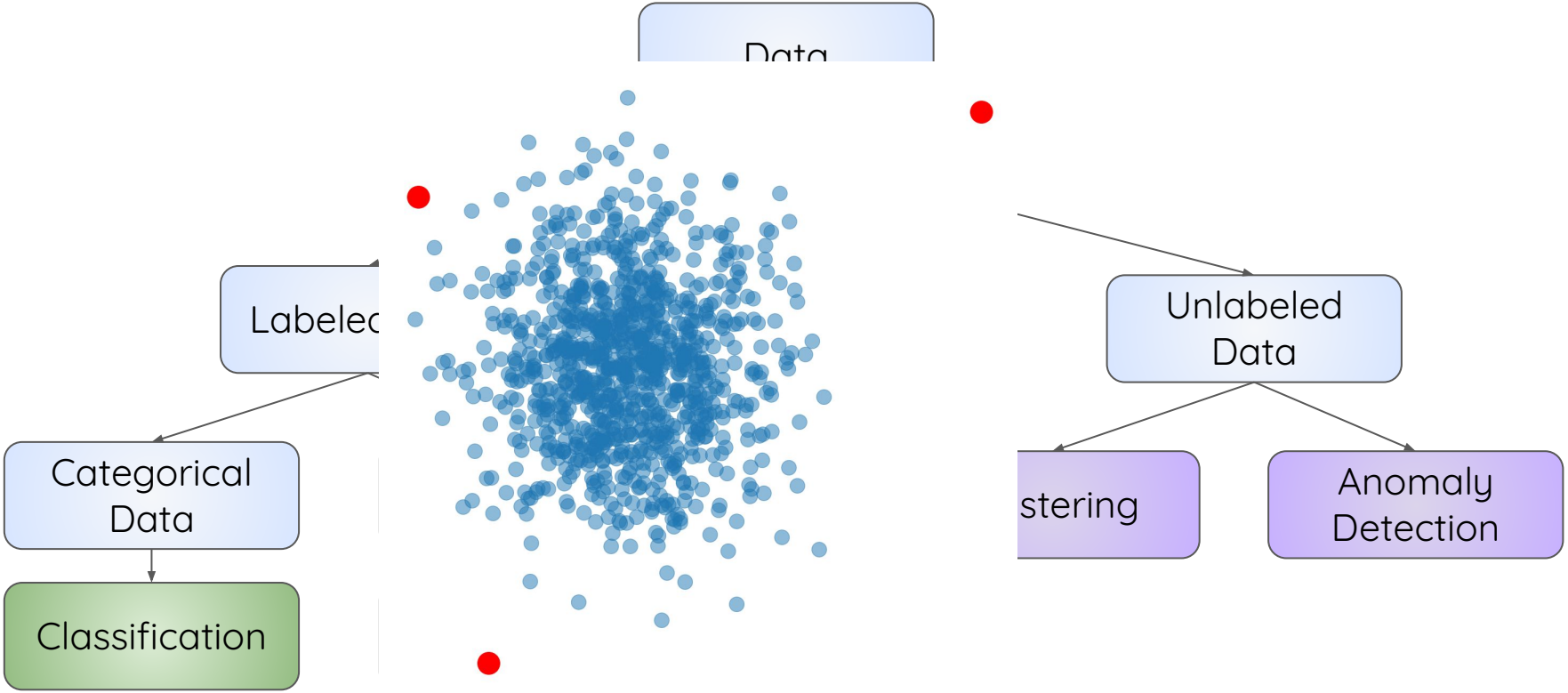


http://colah.github.io/

Some Definitions



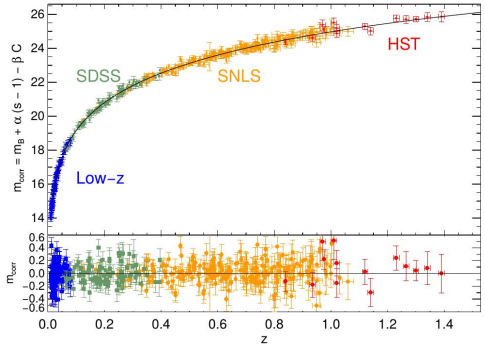
Some Definitions



Supernova Classification

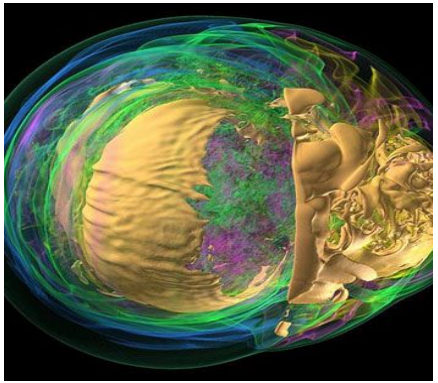
Supernova Classification

Type Ia

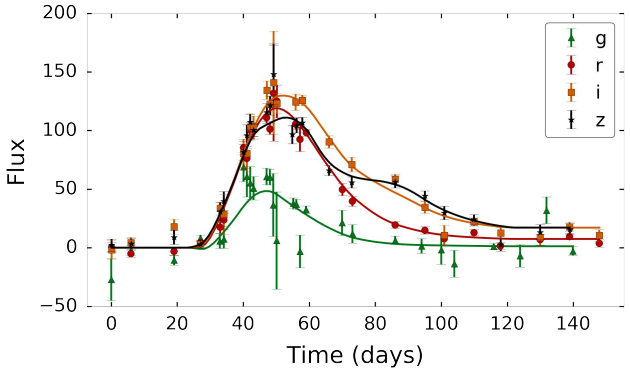


Conley et al. (2011)

Core Collapse

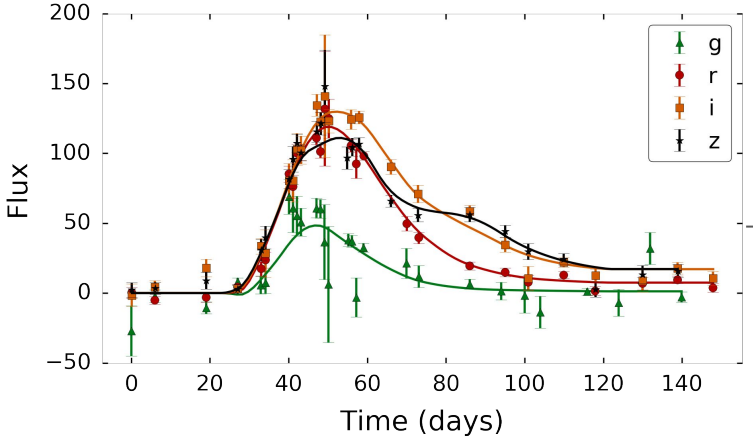


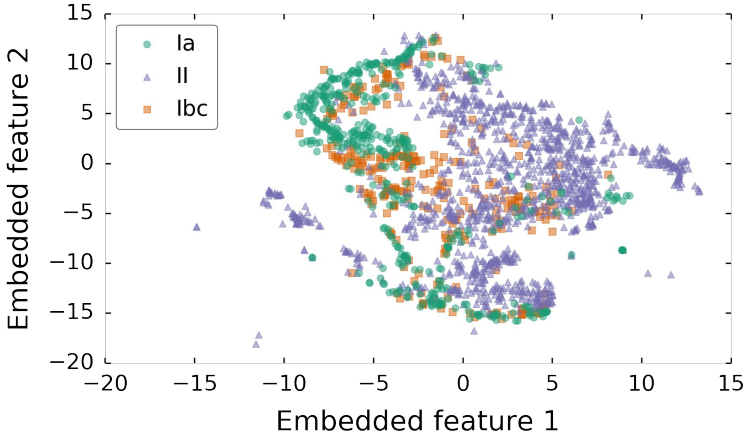
Supernova Classification



Lochner et al. (2016) - 1603.00882

Supernova Classification

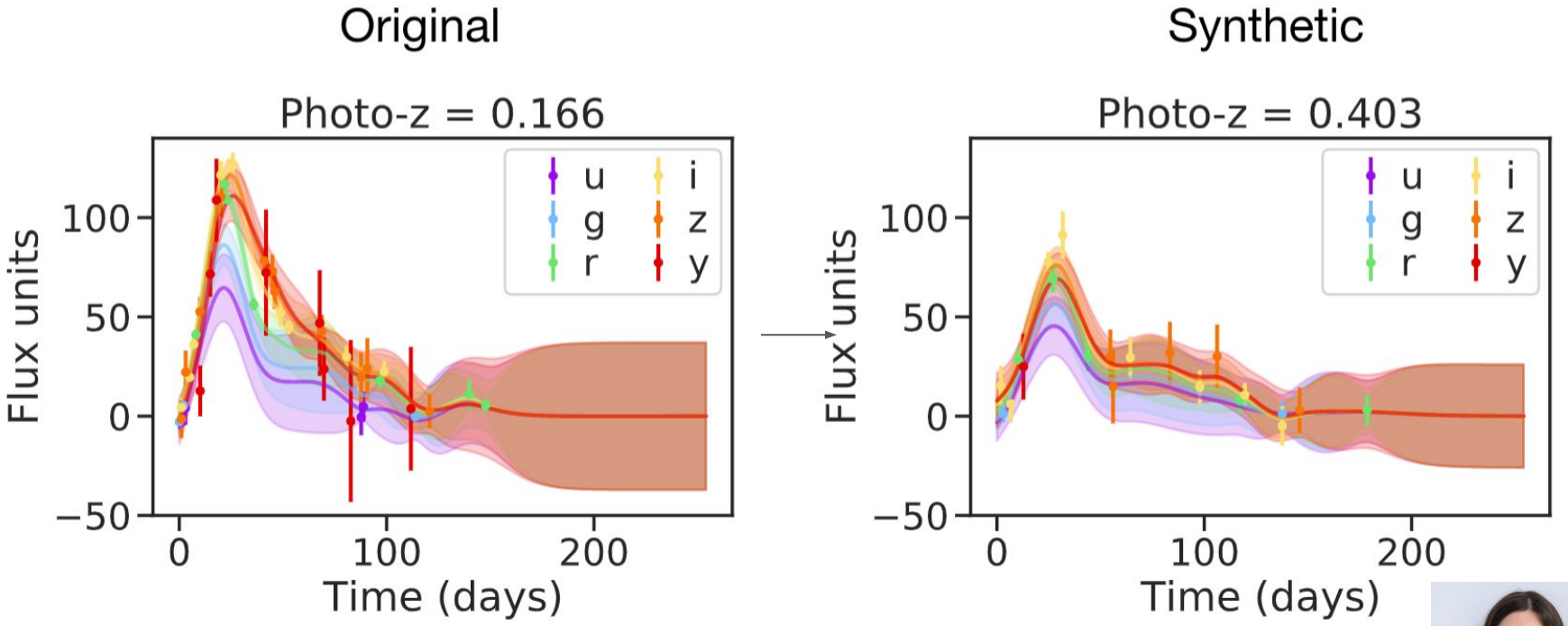




Lochner et al. (2016) - 1603.00882

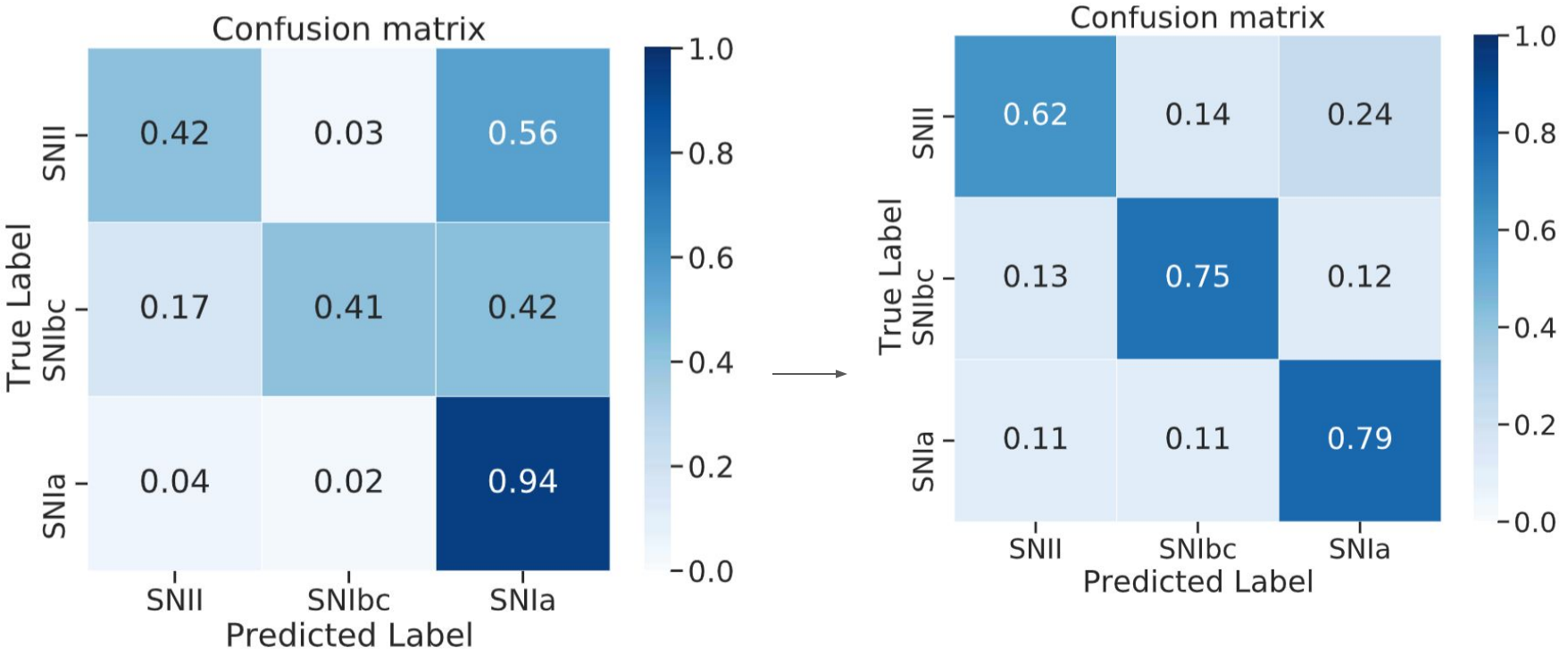
Sooknunan et al. (2018) - 1811.08446

PLAsTiCC



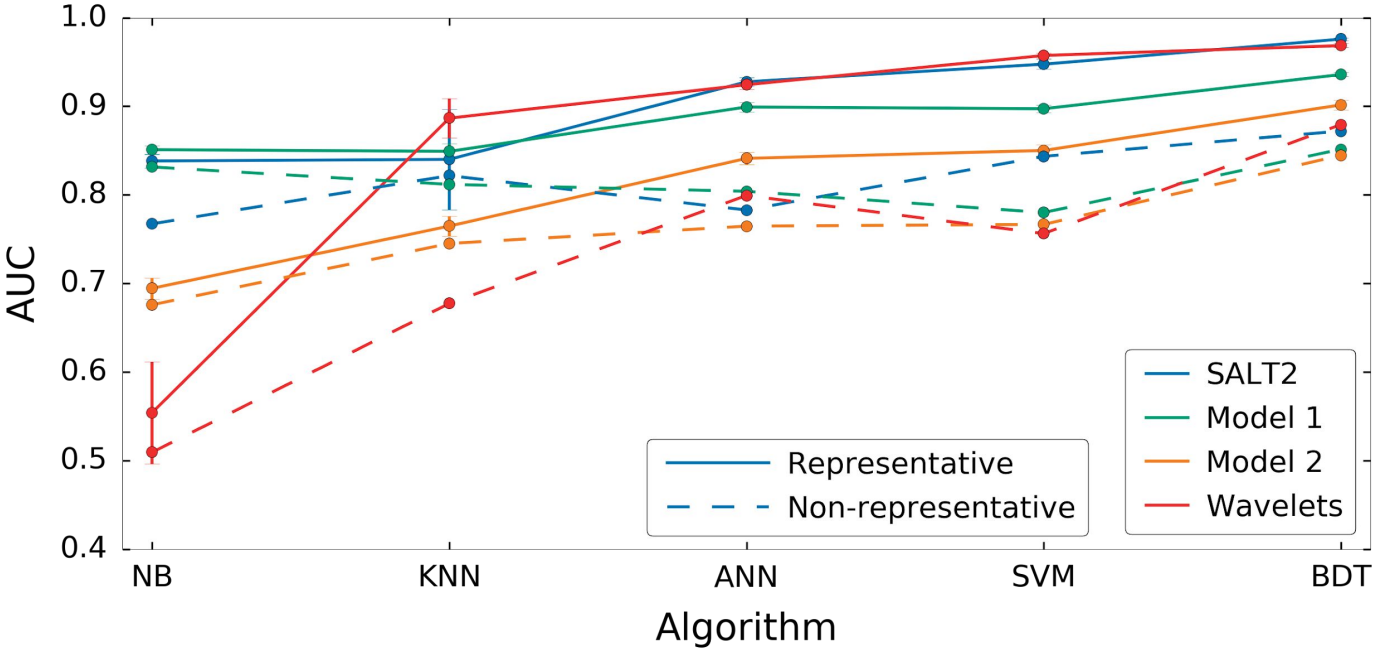
Catarina Alves et al. (2021) [2107.07531]

PLAsTiCC



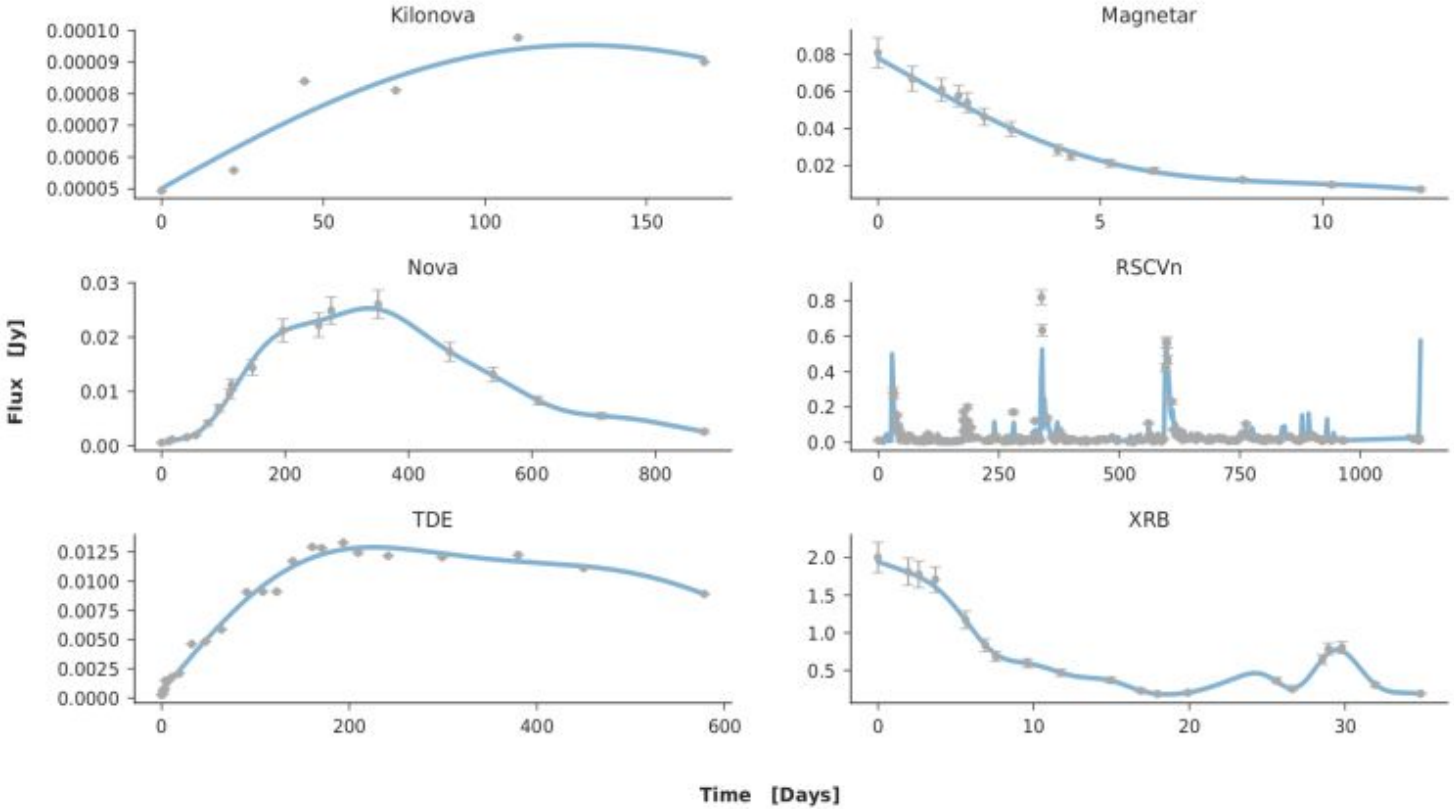
Catarina Alves et al. (in prep)

Supernova Classification



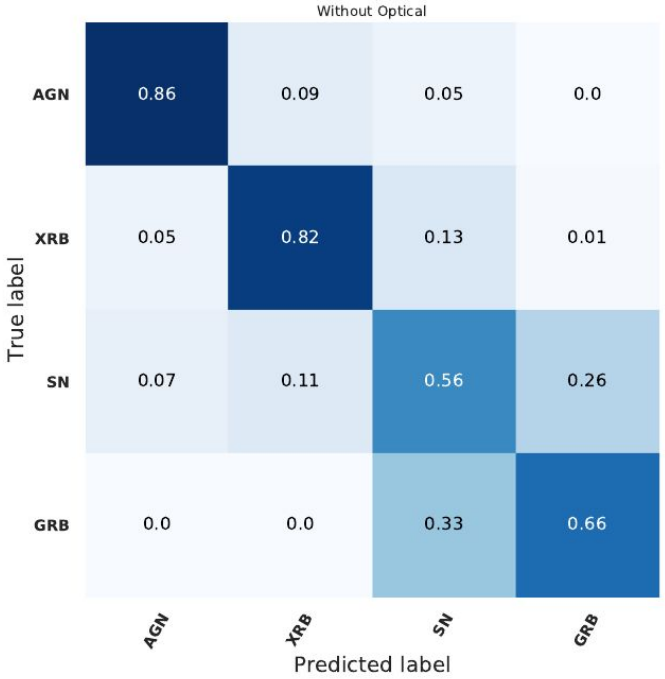
Lochner et al. (2016) - 1603.00882

Multiwavelength Transient Classification

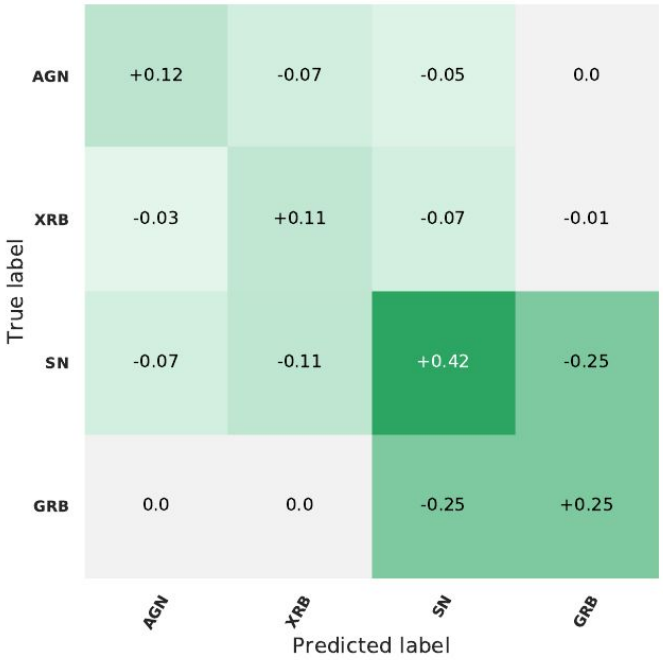


Sooknunan et al. (2018) - 1811.08446

Multiwavelength Transient Classification



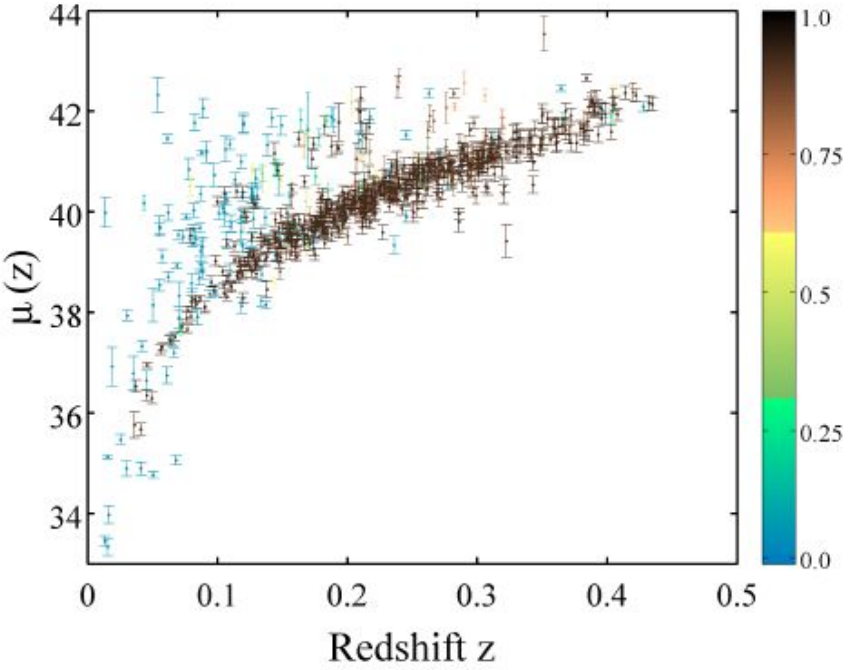
(a) Confusion matrix without optical feature

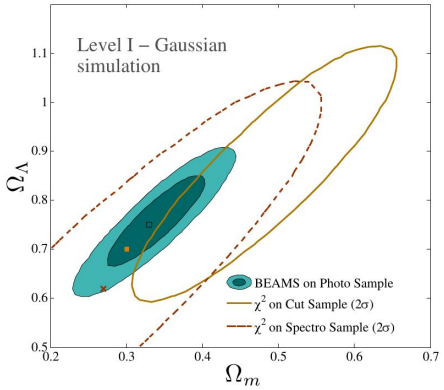


(b) Confusion matrix showing the difference when optical feature is added

Kimeel Sooknunan et al. (2018) - 1811.08446

Science with Imperfect Classification



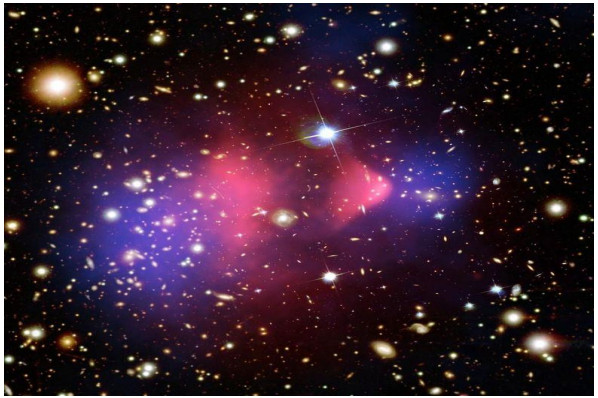


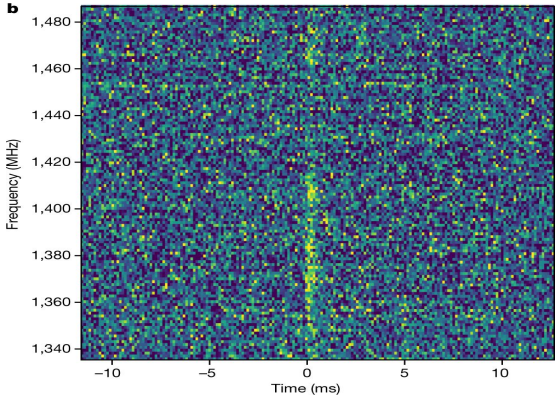
Bayesian Estimation Applied to Multiple Species (BEAMS)

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

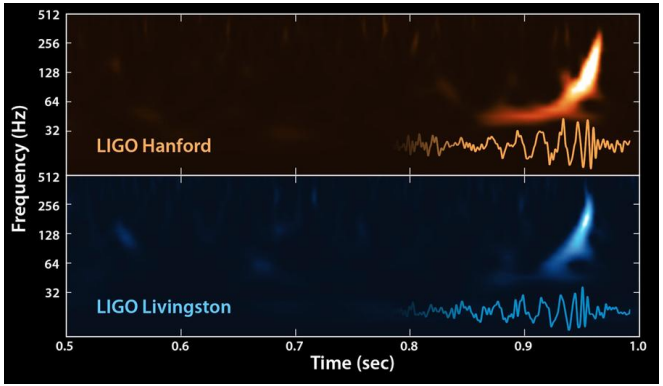
Anomaly Detection

Known Unknowns - rare events



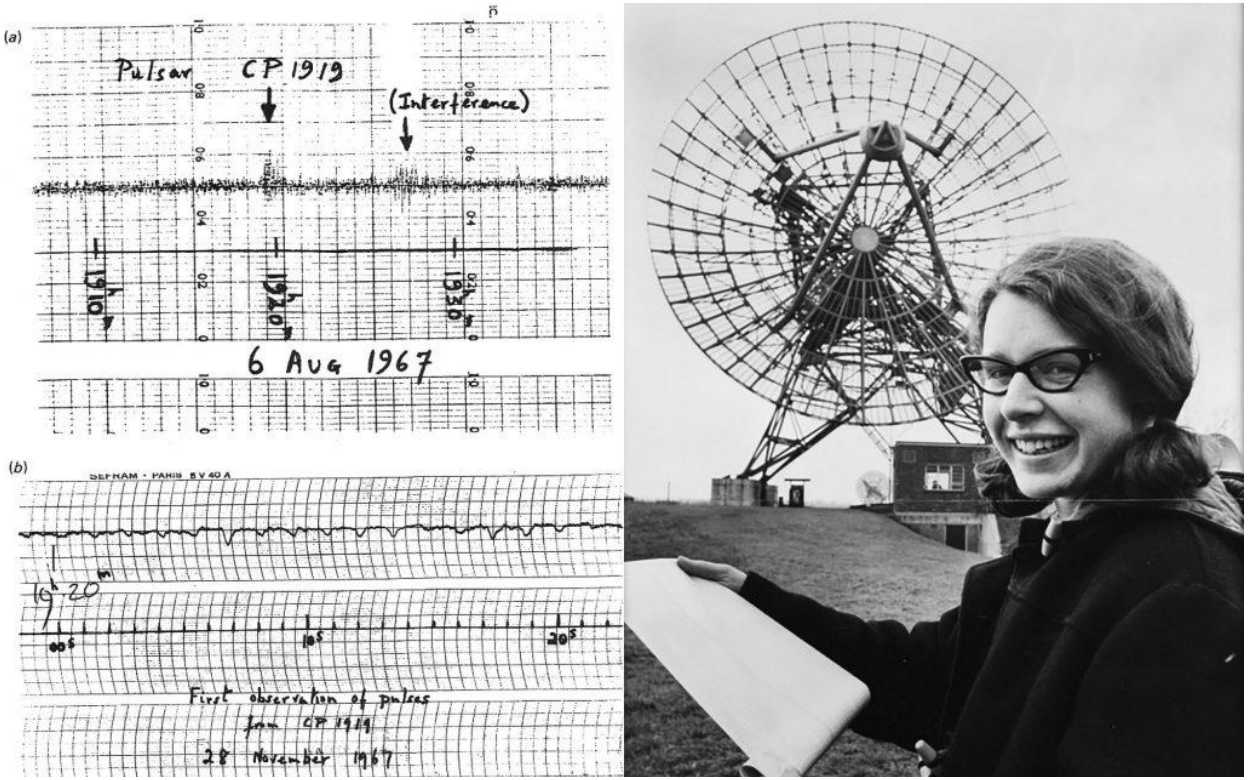


Ravi et al. (2019)



Hubble/ NASA/ ESA

Unknown Unknowns - new anomalies

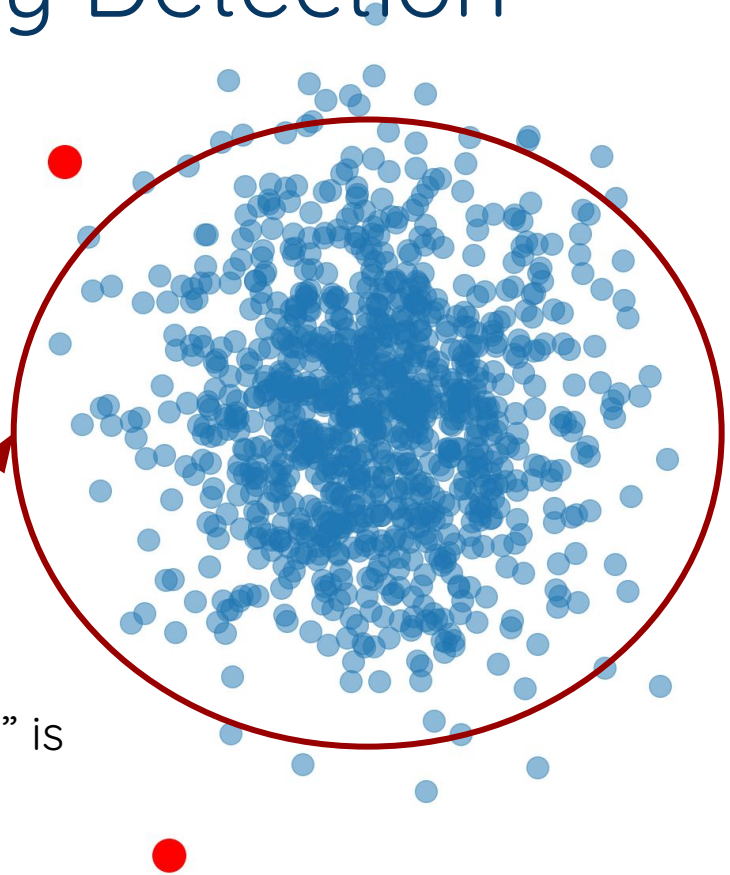


Daily Herald Archive / SSPL / Getty Images

How do we discover new phenomena...

...among 10 million possibilities?

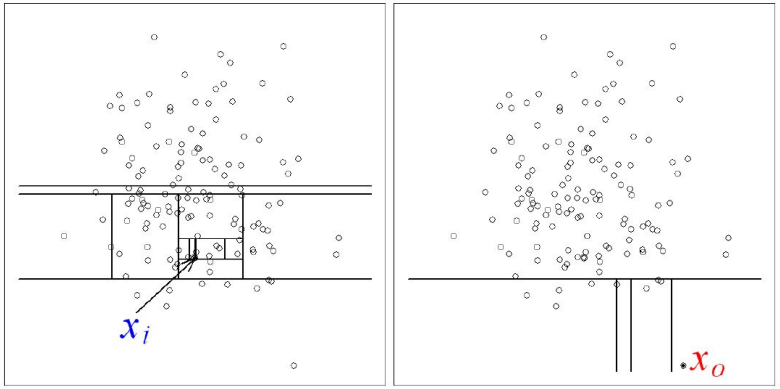
Anomaly Detection



Try to learn what “normal” is

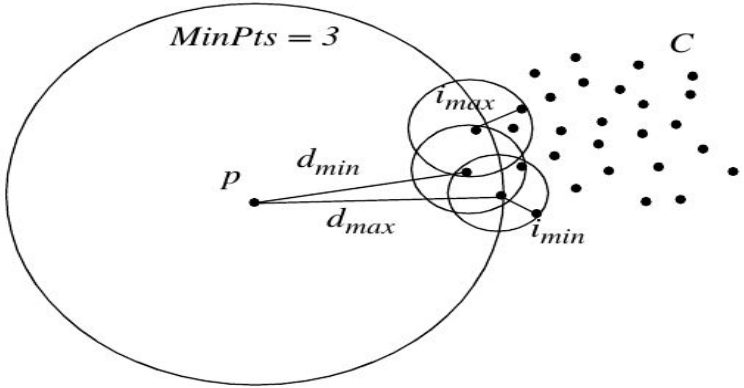
The farther away from normal the higher the anomaly score

Anomaly Detection Algorithms

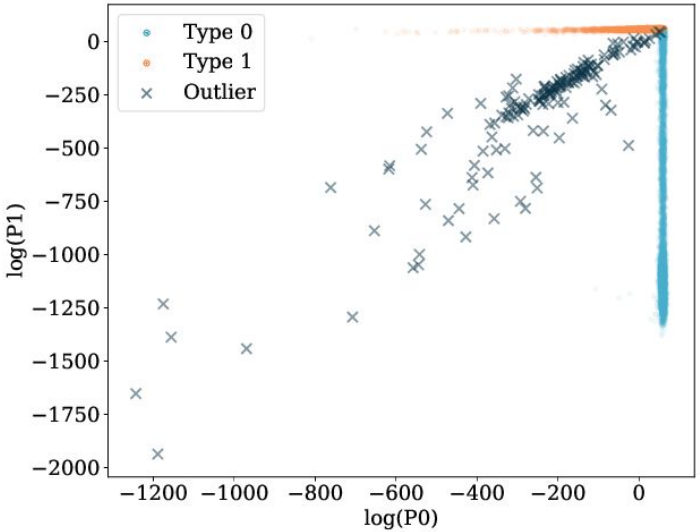


Isolation Forest (Liu, Ting & Zhou; 2008)

Local Outlier Factor (Breunig et al; 2008)

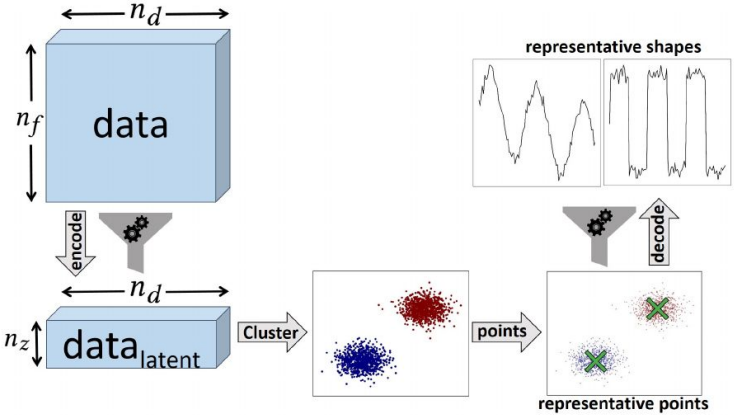


Anomaly Detection Algorithms



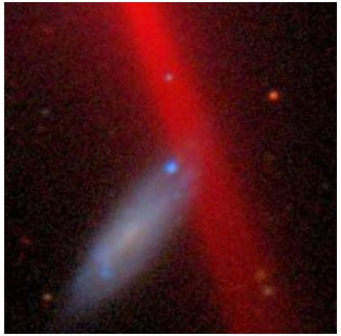
BADAC (Roberts, Bassett & Lochner - 1902.08627)

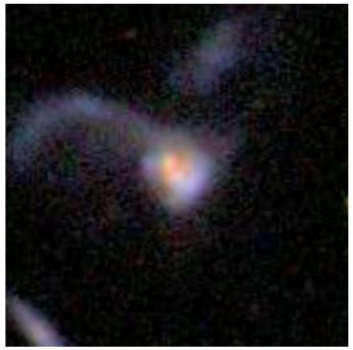
DRAMA (Vafaei Sadr, Bassett & Kunz - 1909.04060)

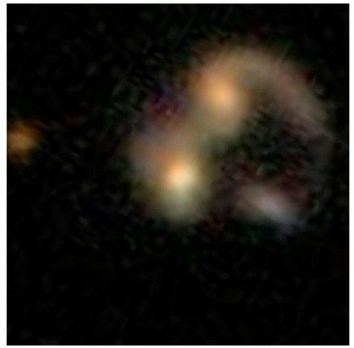


Anomaly Detection Isn’t Enough

Artefacts

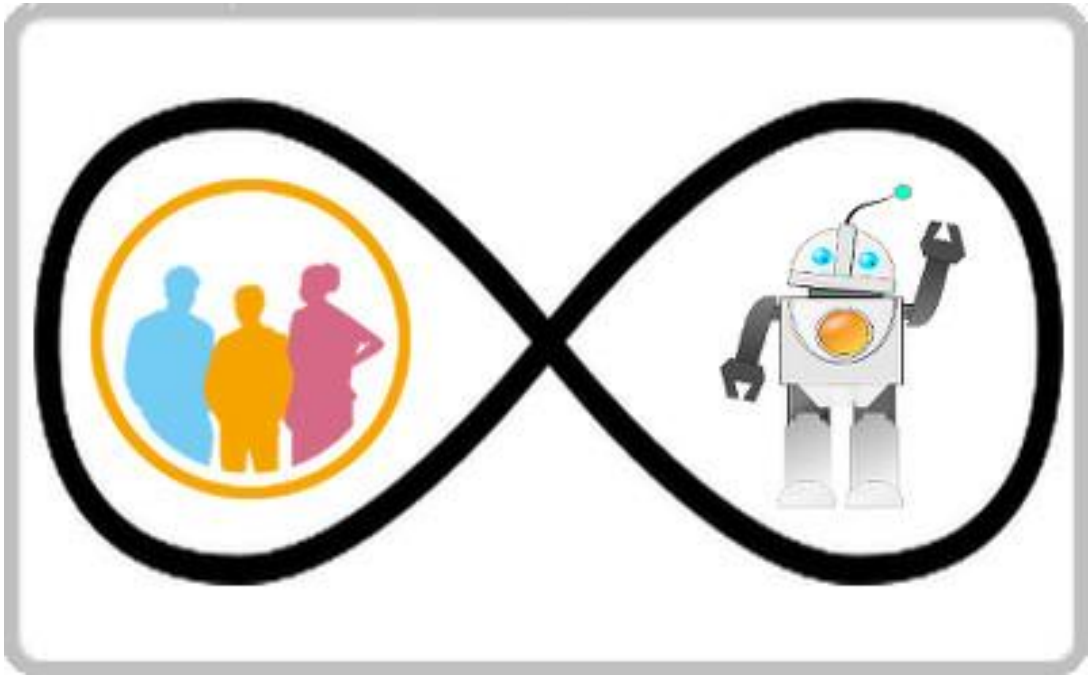




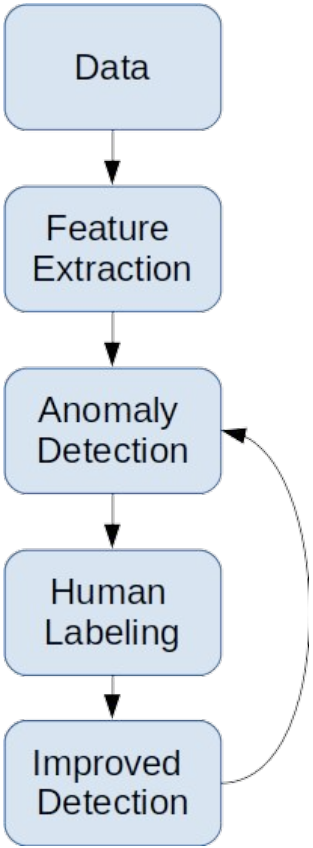


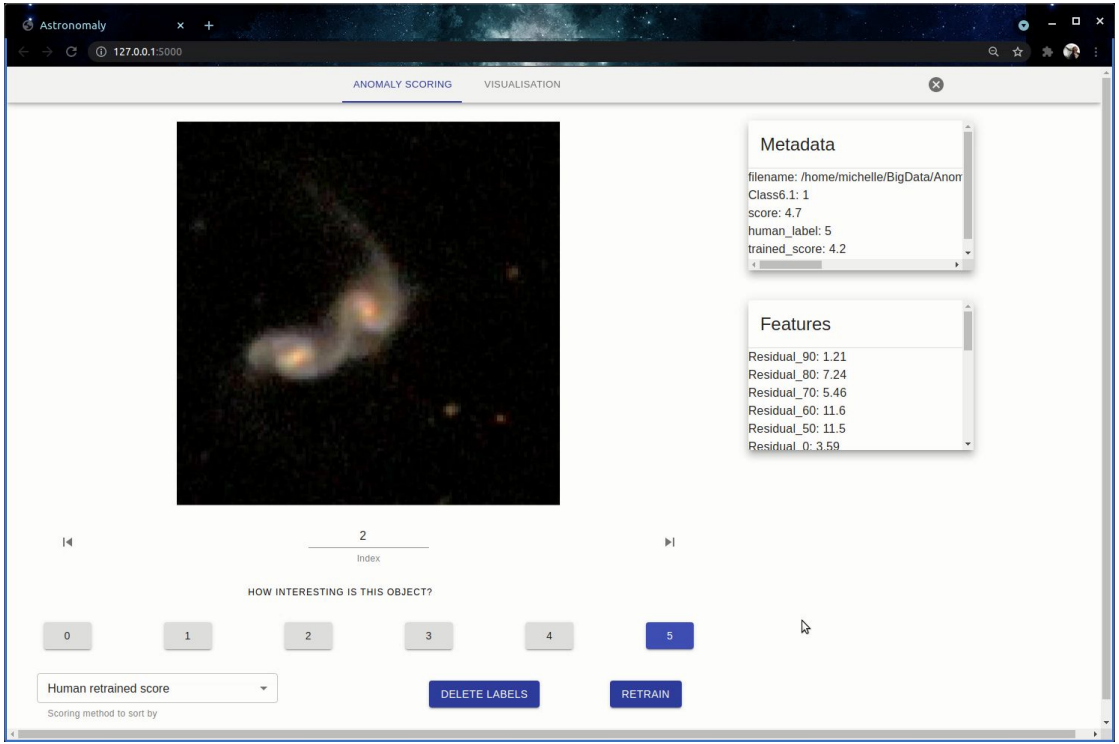
Real sources

Active Learning



Astronomaly

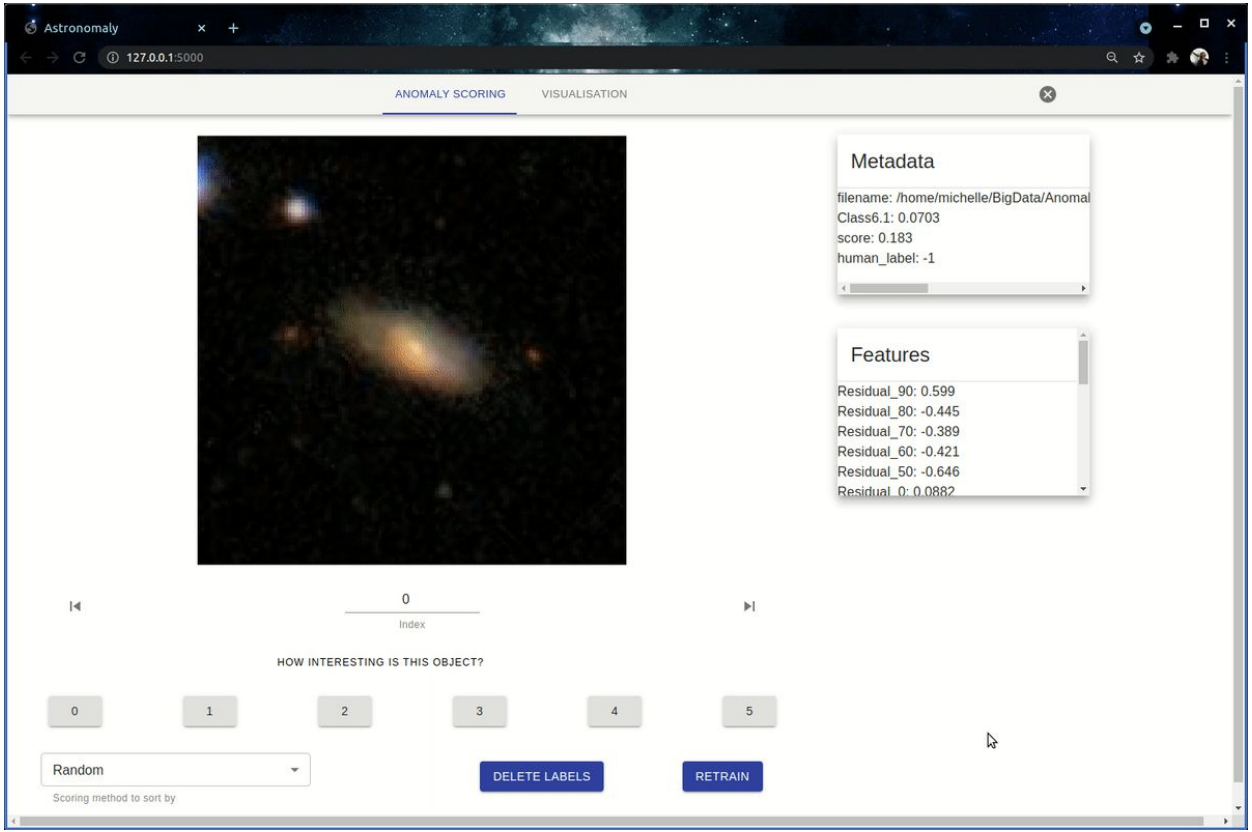




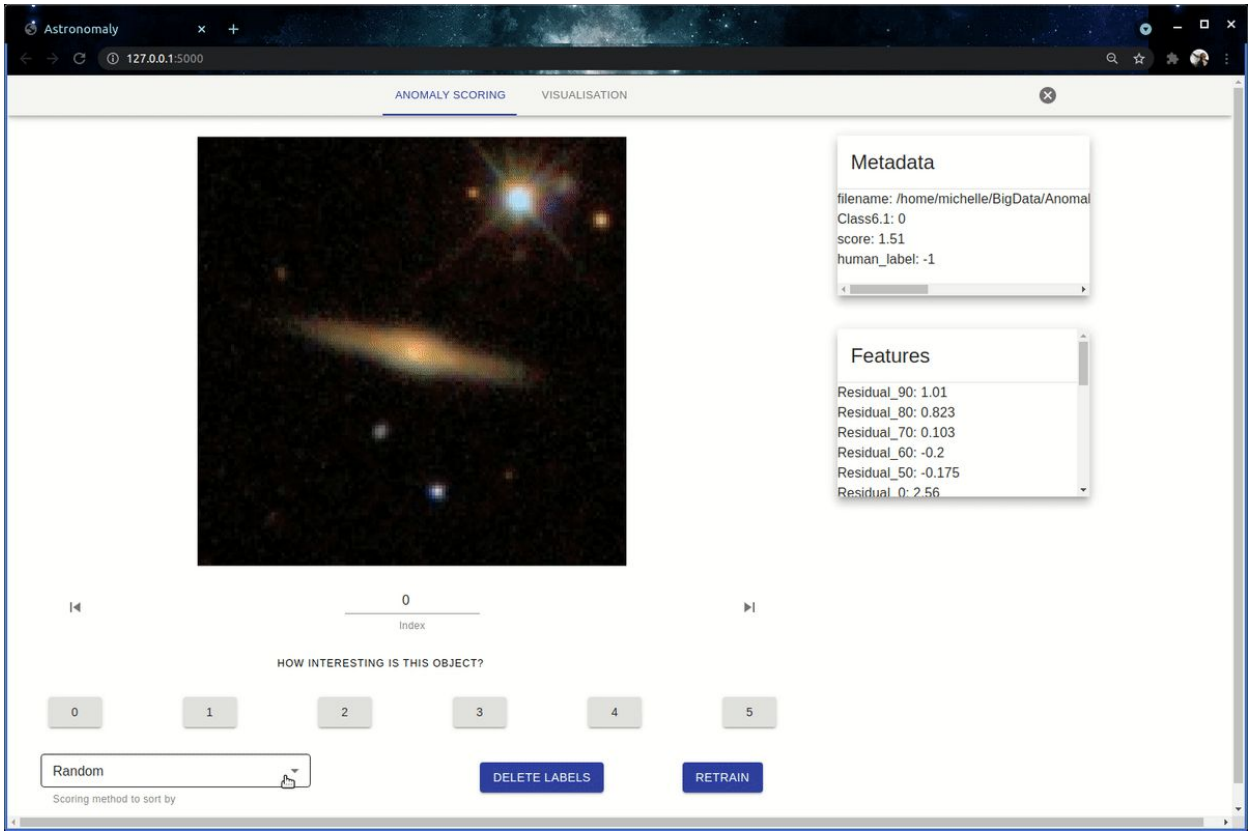
Lochner and Bassett (2020) - 2010.11202

https://github.com/MichelleLochner/astronomaly

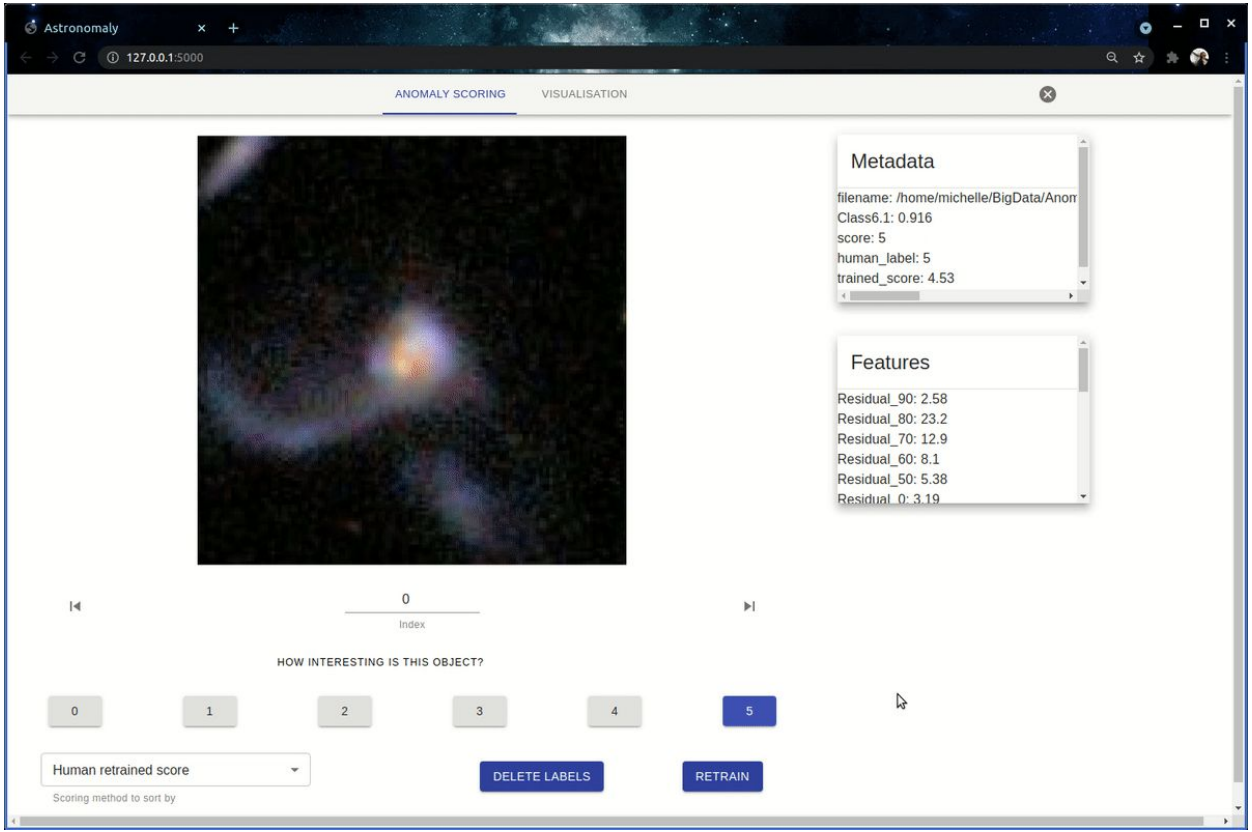
Galaxy Zoo - Random Examples



Galaxy Zoo - Machine Learning

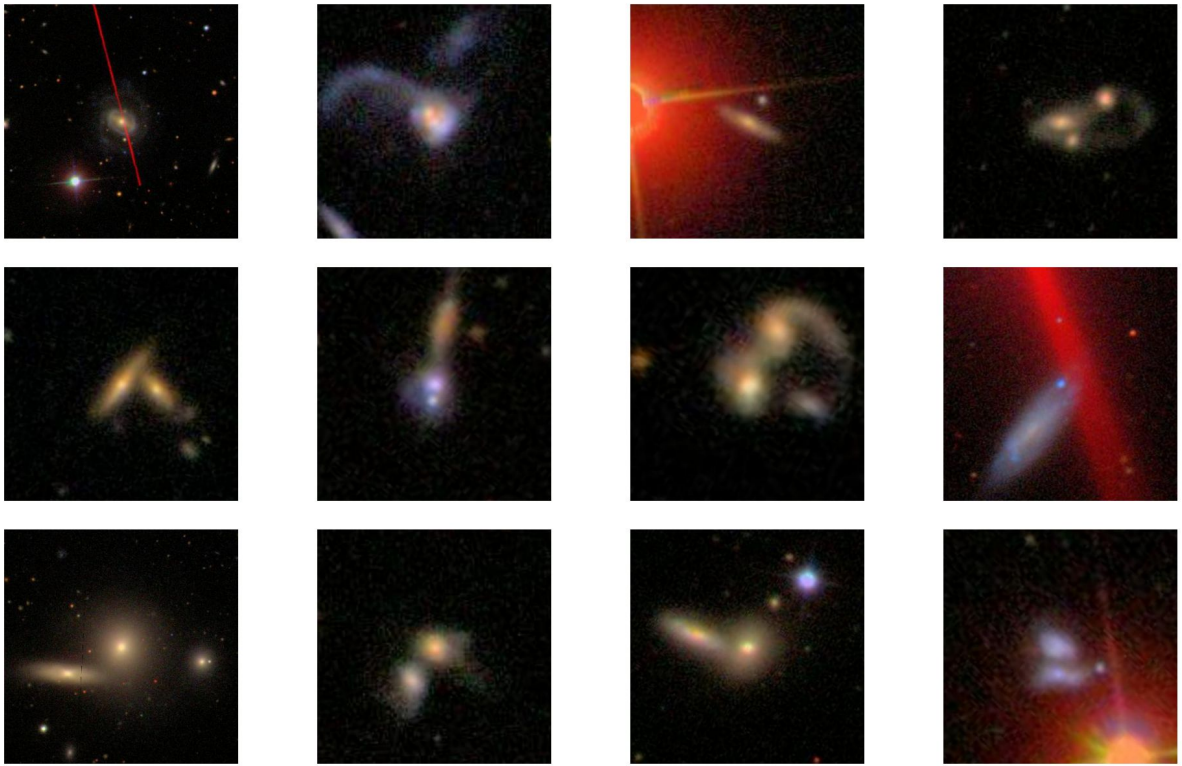


Galaxy Zoo - Active Learning



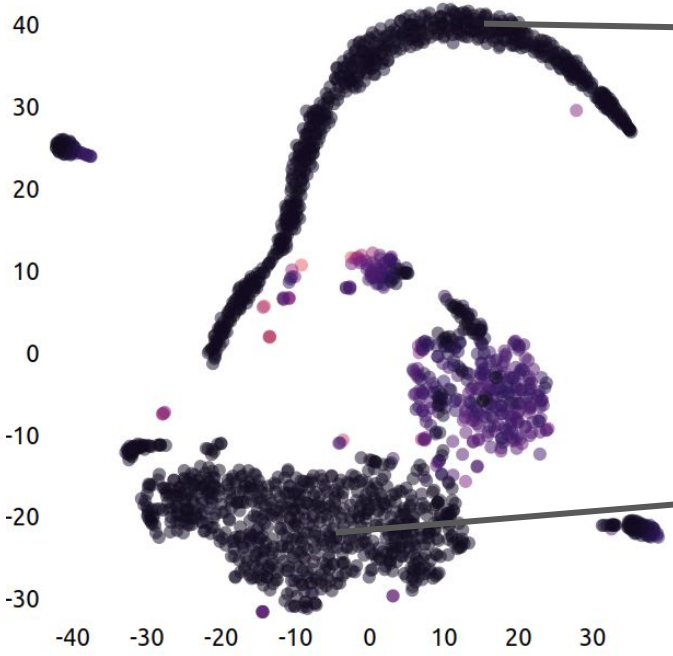
Galaxy Zoo - Random Examples

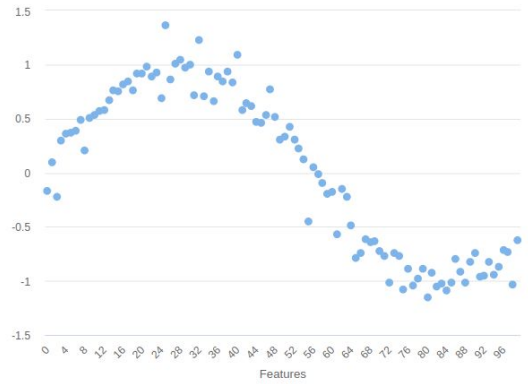
Galaxy Zoo - No active Learning



Galaxy Zoo - Active Learning

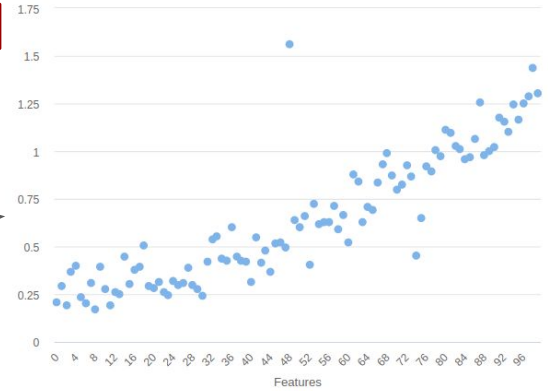
Synthetic data





Anomaly Score: 0.270122

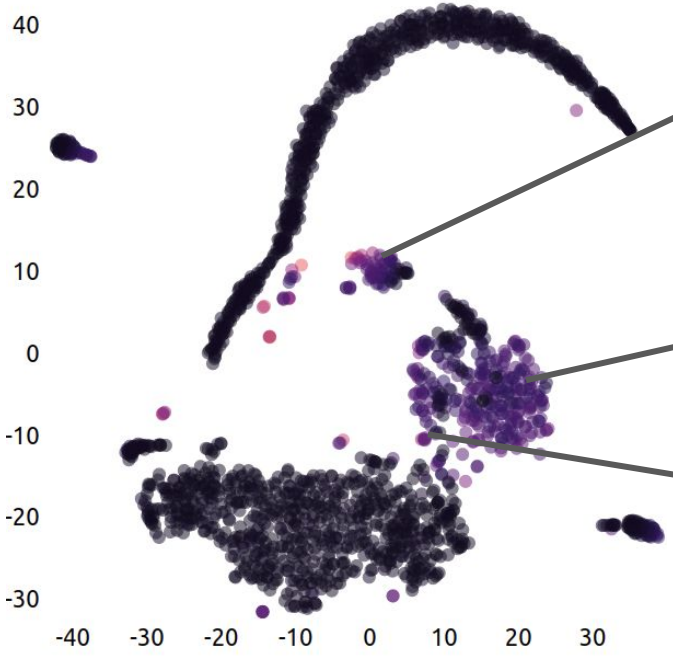
Normal



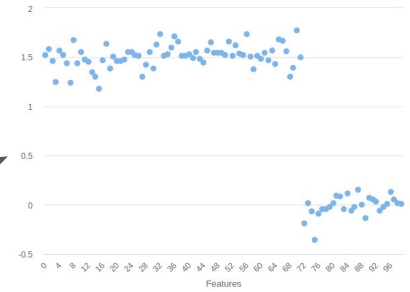
Anomaly Score: 0.241527

Lochner and Bassett (2020)

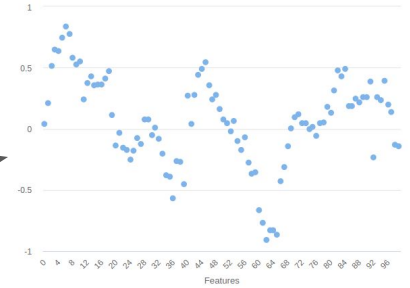
Synthetic data



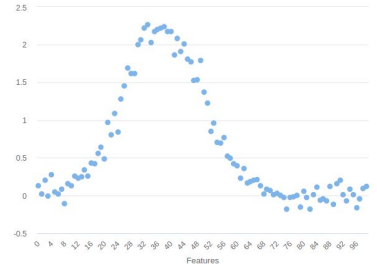
Anomalous



Anomaly Score: 3.112012



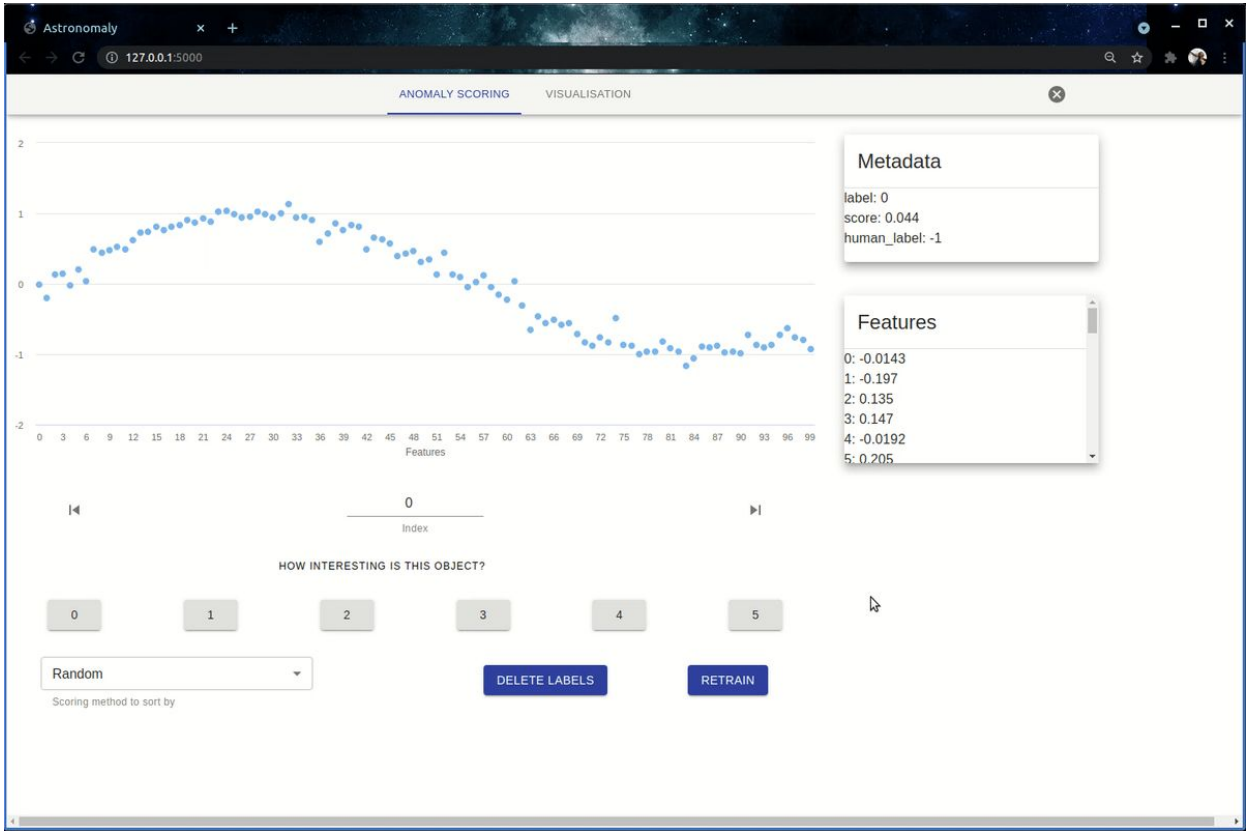
Anomaly Score: 1.303772



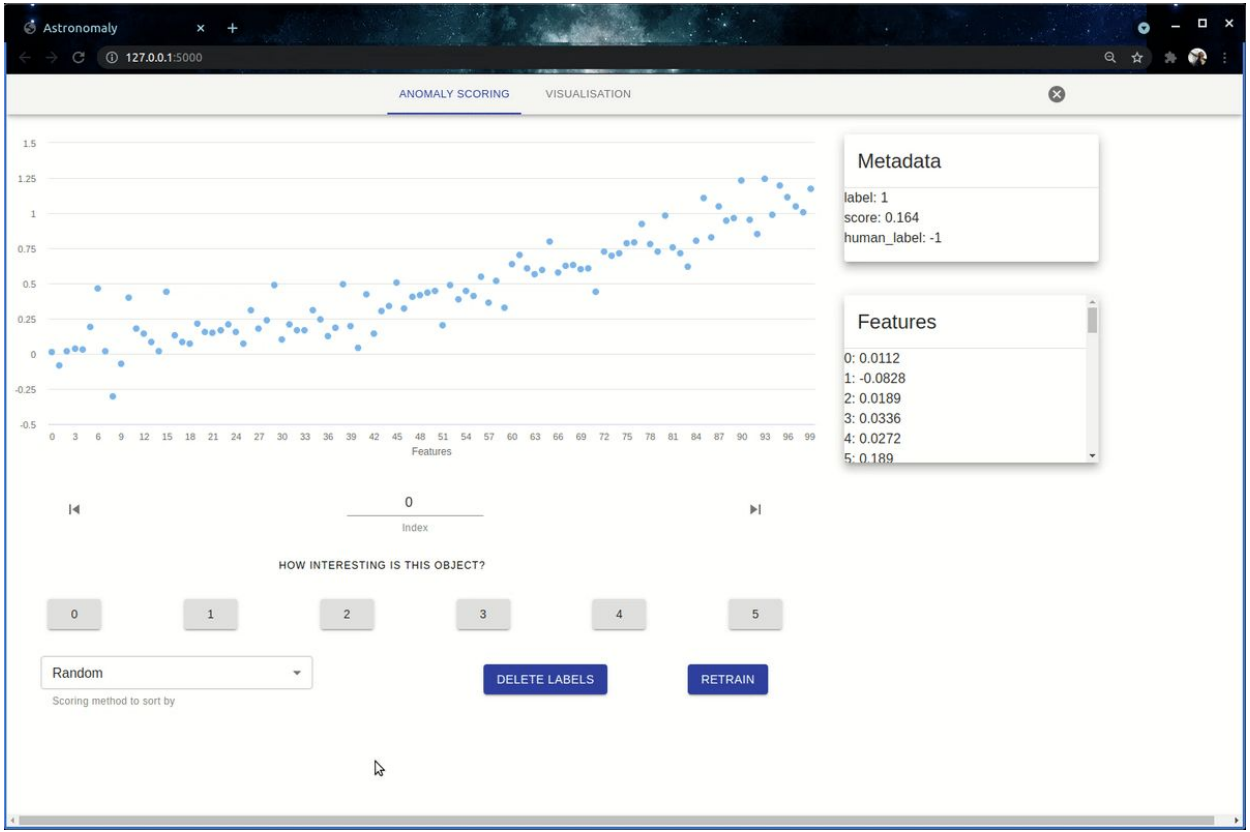
Anomaly Score: 3.92992

Lochner and Bassett (2020)

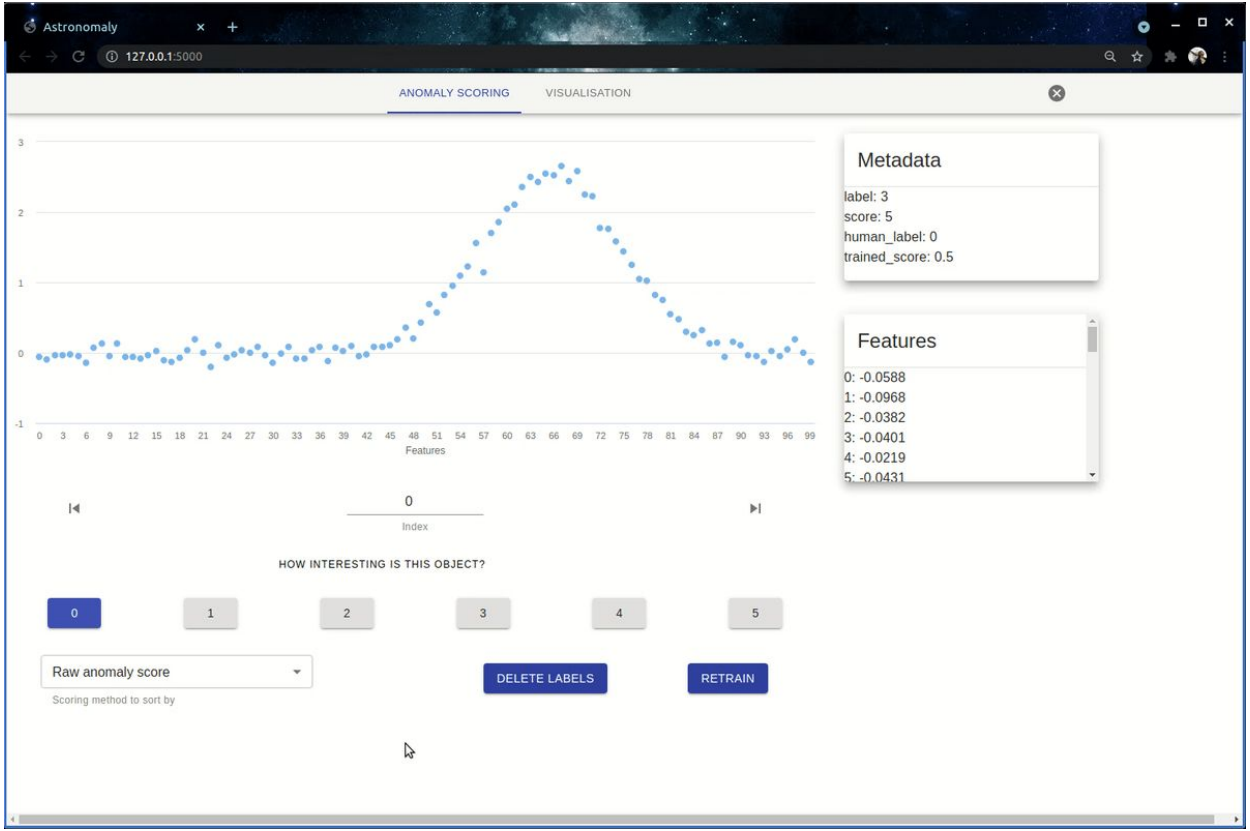
Synthetic data - Random



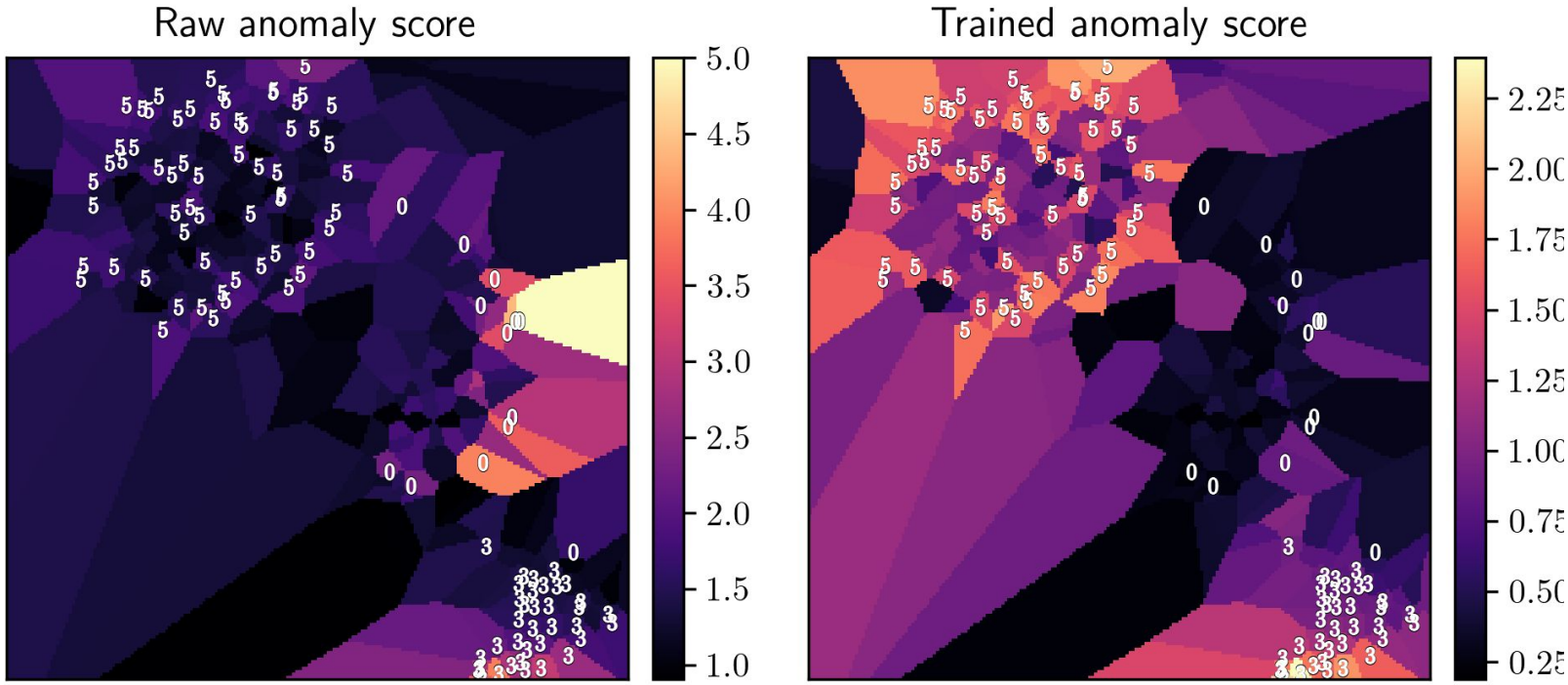
Synthetic data - No active Learning



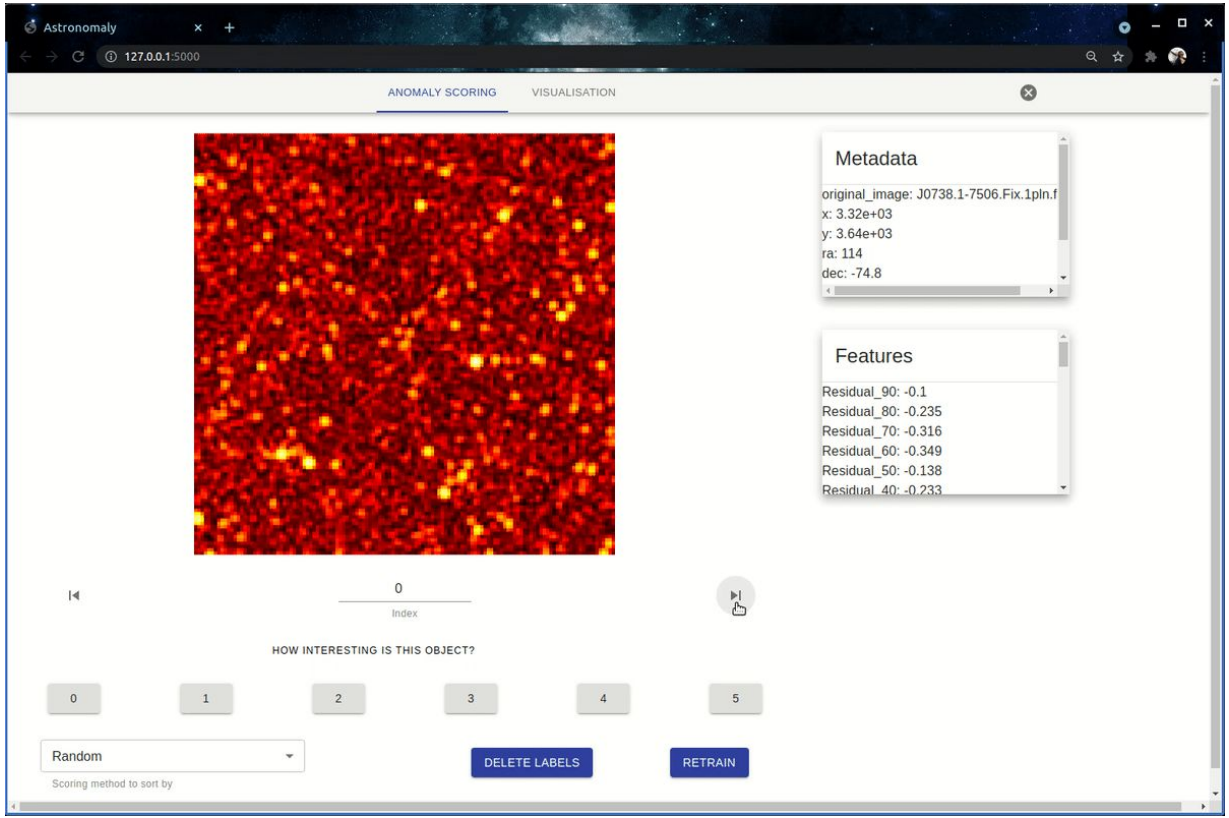
Synthetic data - Active Learning



Visualisation with Synthetic Data

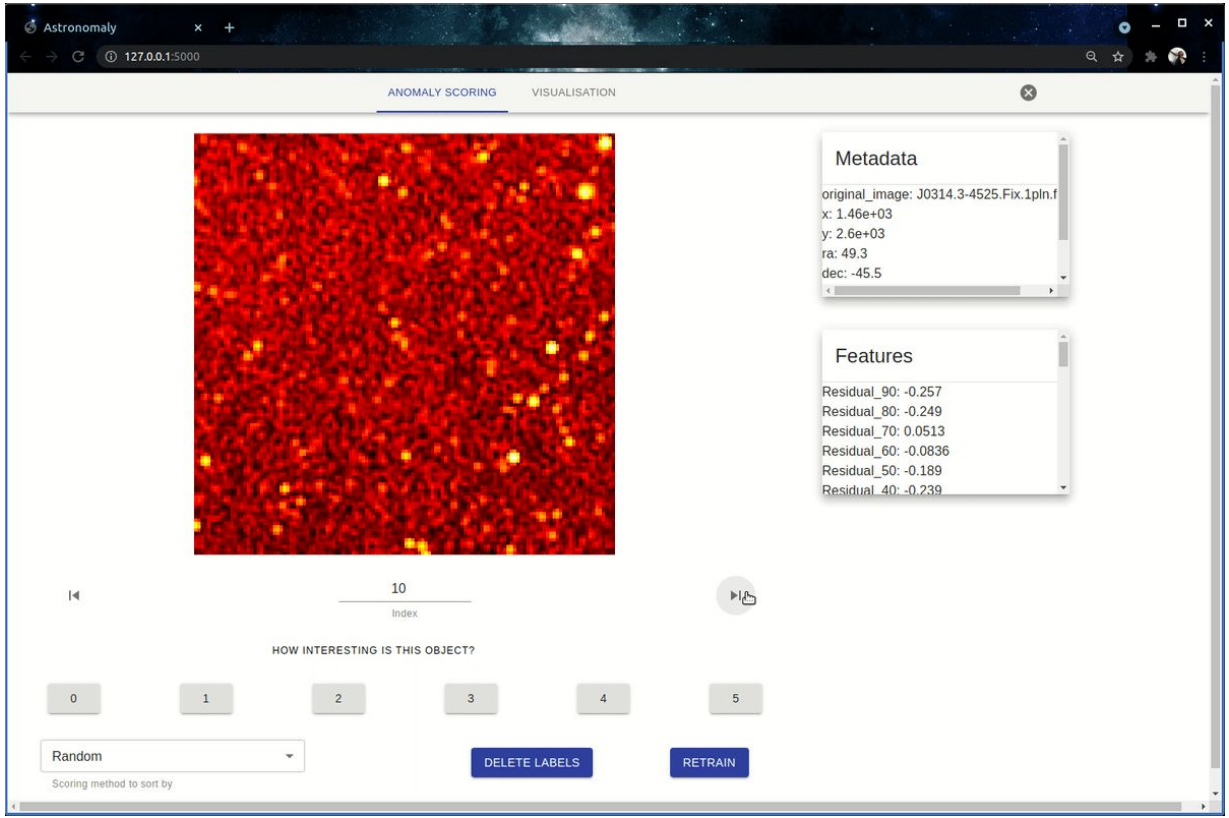


MeerKAT Data - Random Examples



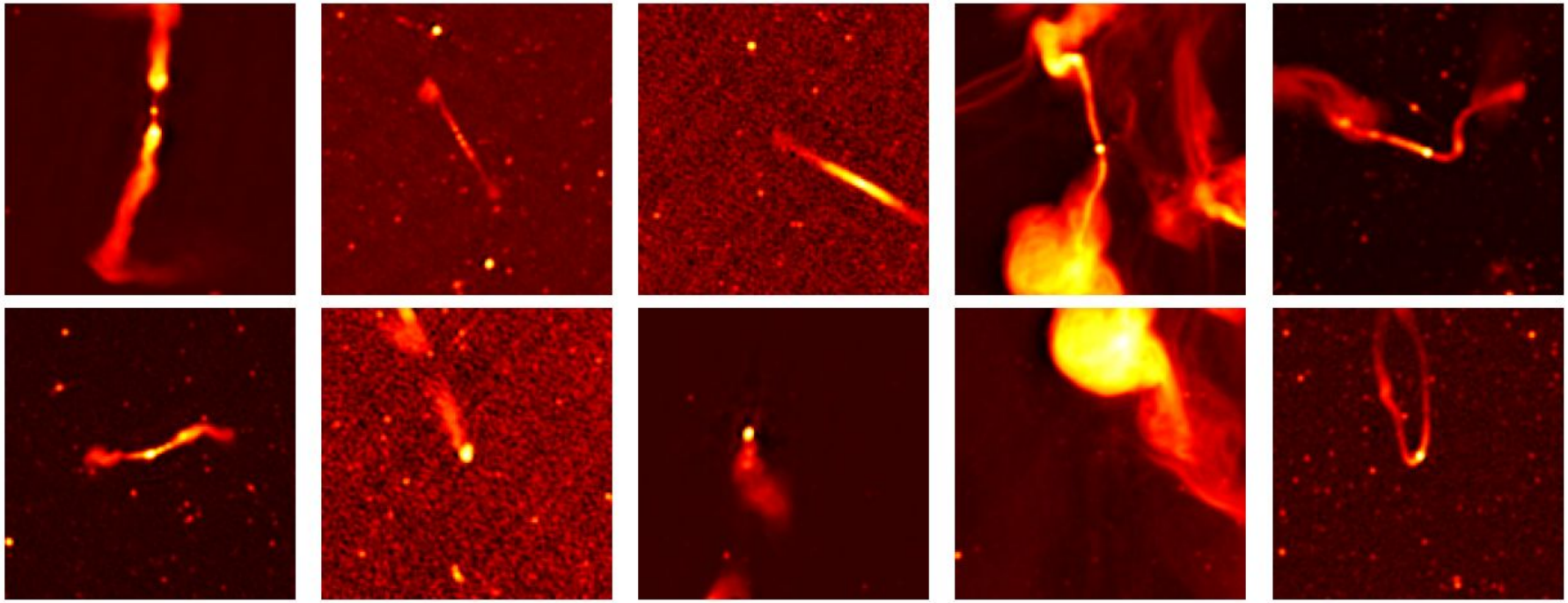
Knowles, K et al. (2021) [2111.05673]

MeerKAT Data - Anomalies



Knowles, K et al. (2021) [2111.05673]

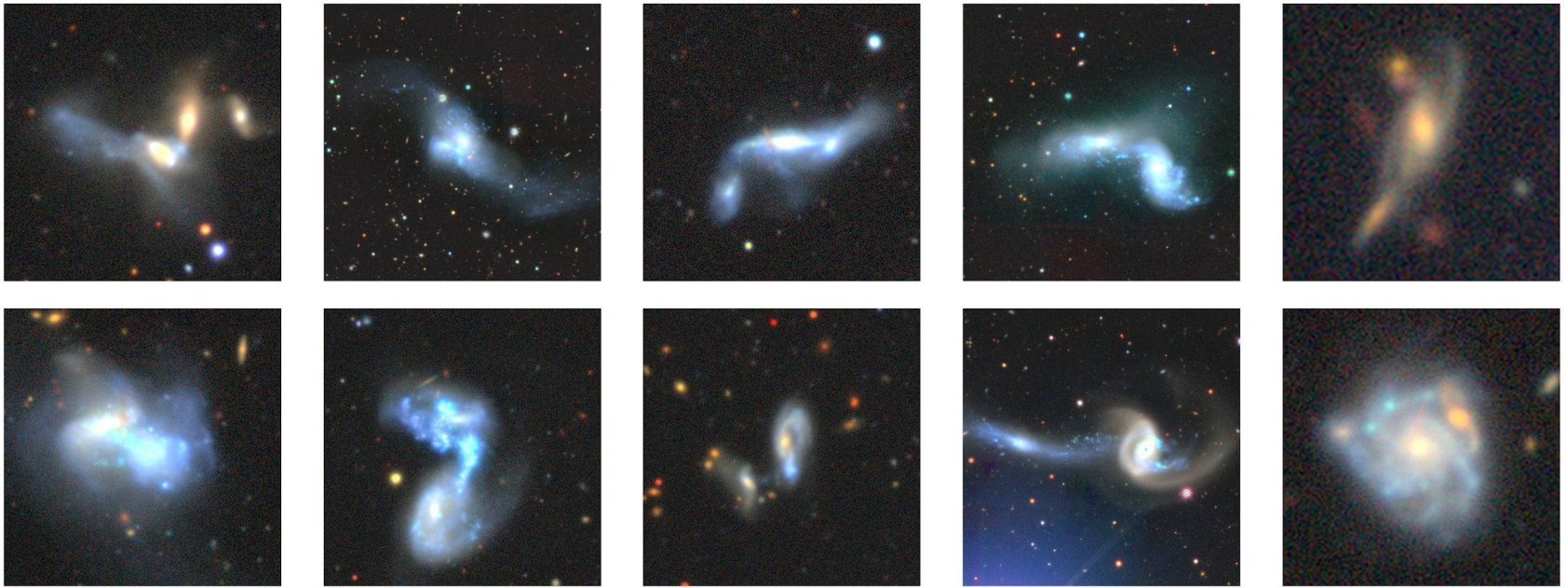
MeerKAT Galaxy Cluster Legacy Survey



Knowles, K et al. (2021) 2111.05673

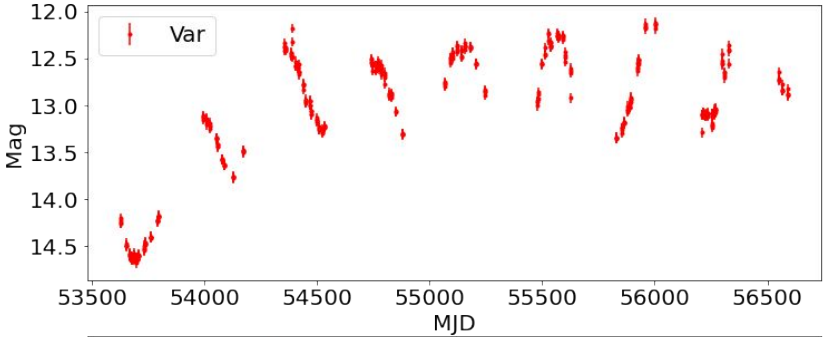
Anomalies in DECALS

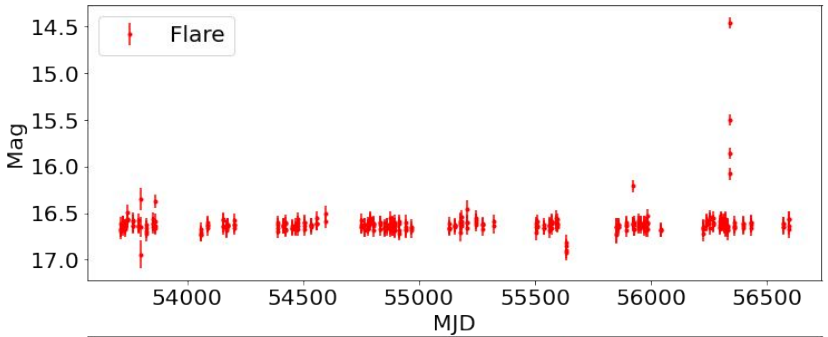
Verlon Etsebeth (MSc student)

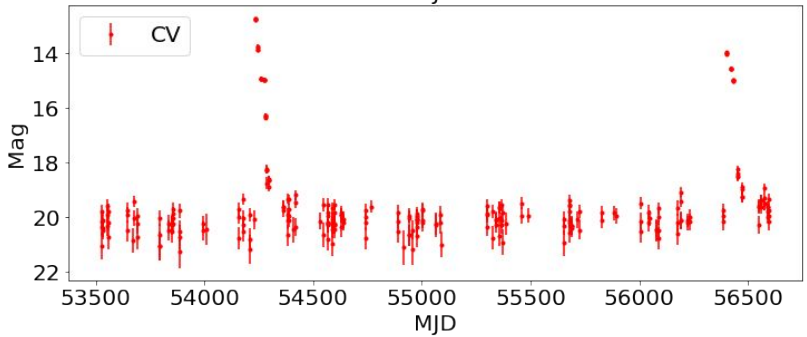


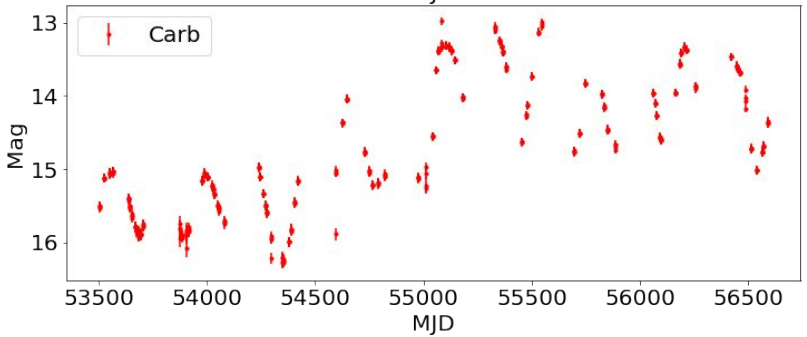
Anomalous Transients

Malema Ramonyai (MSc student)

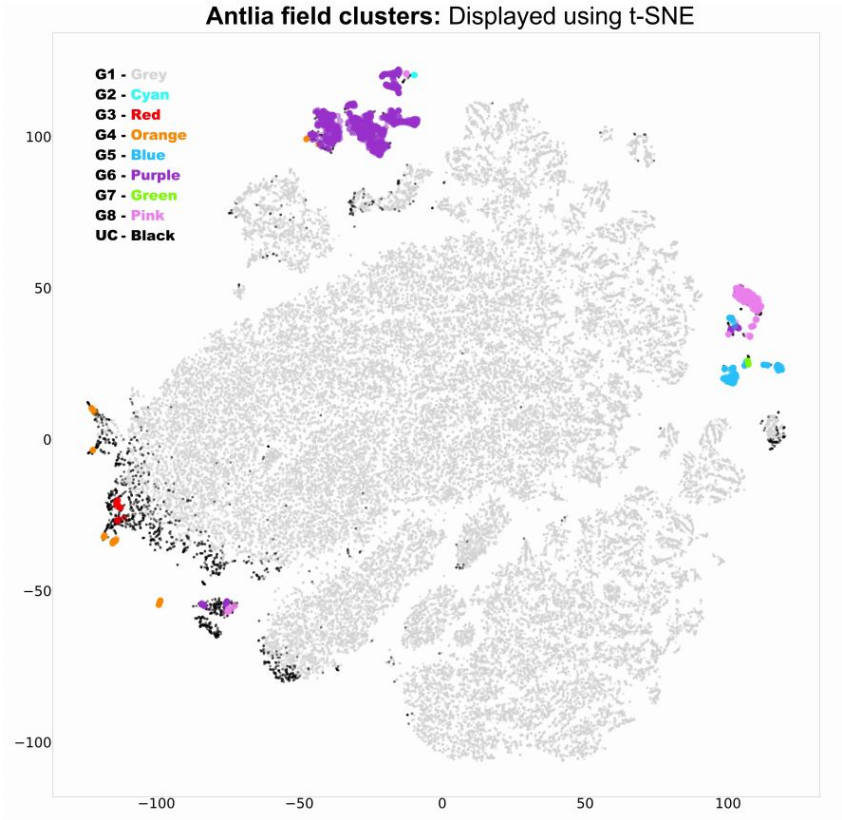






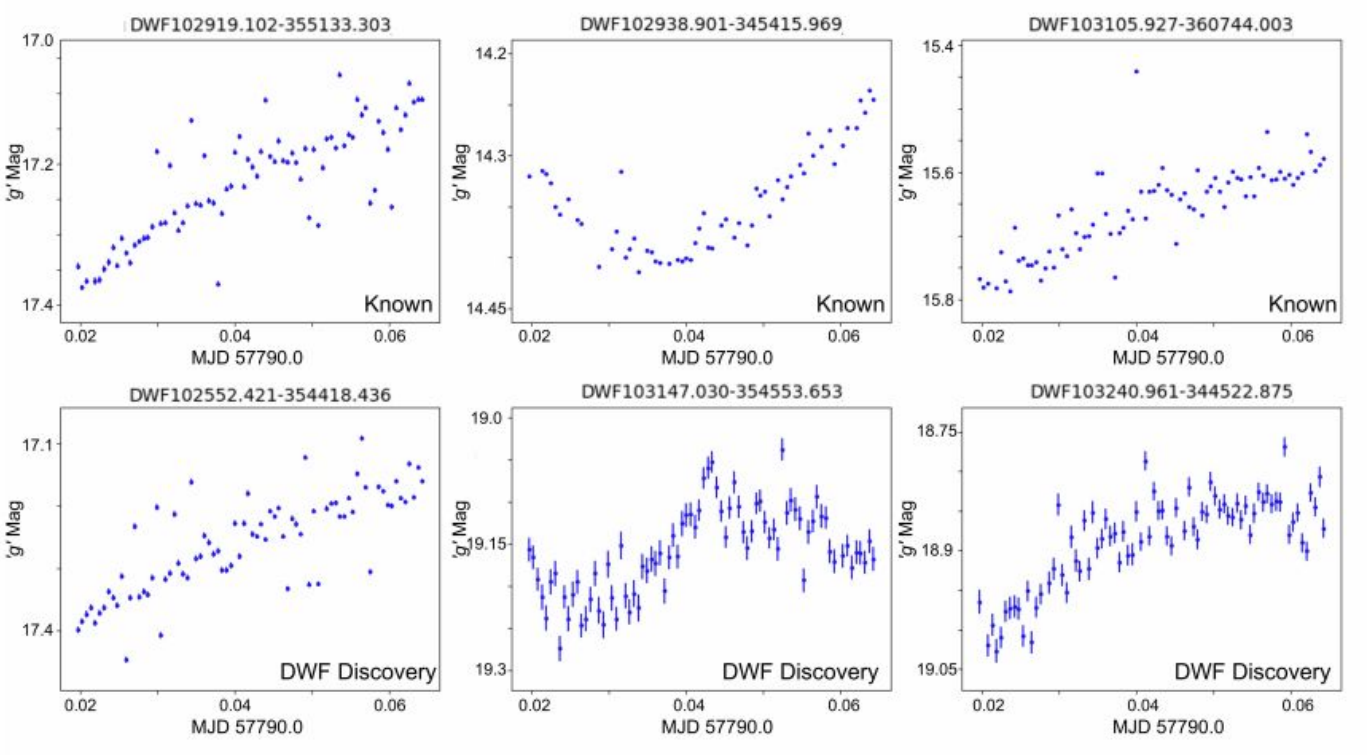


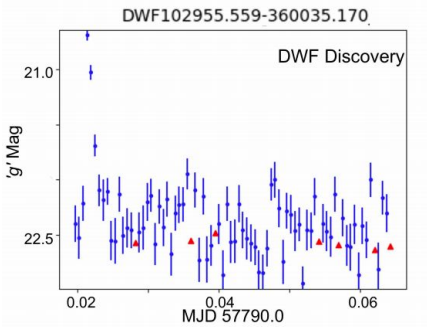
Astronomaly Applied to DWF



Webb et al. (2020) - 2008.04666

Astronomaly Applied to DWF

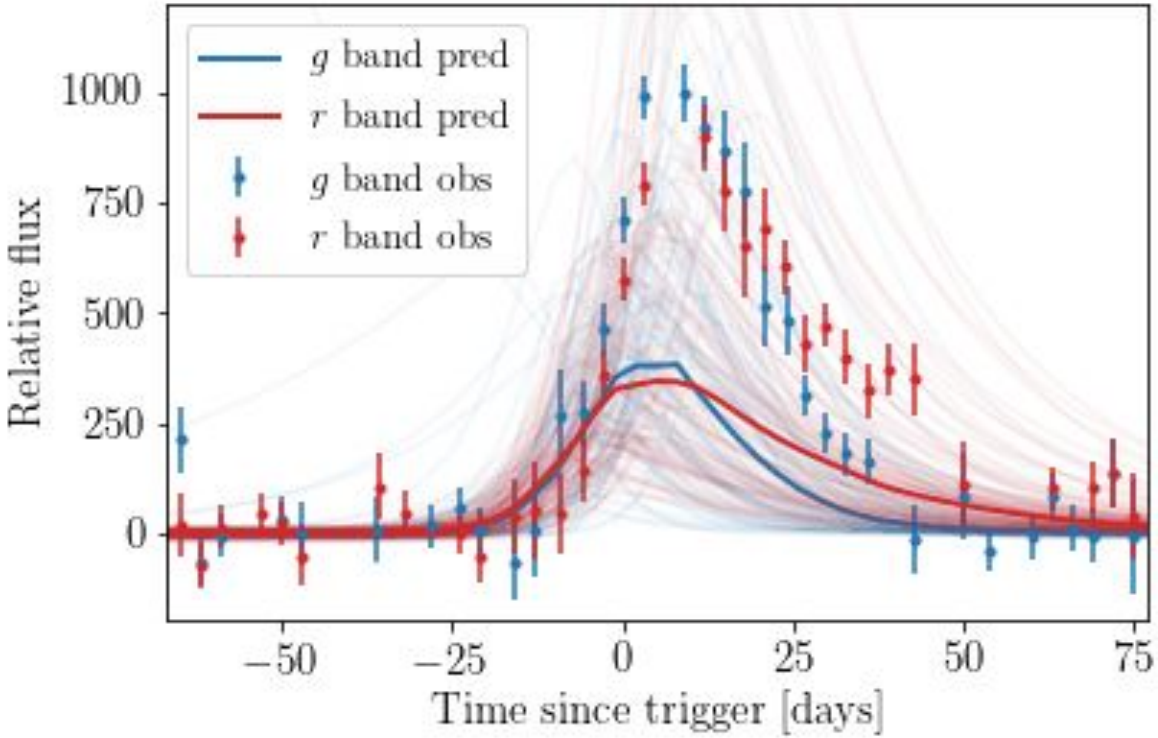




Webb et al. (2020) - 2008.04666

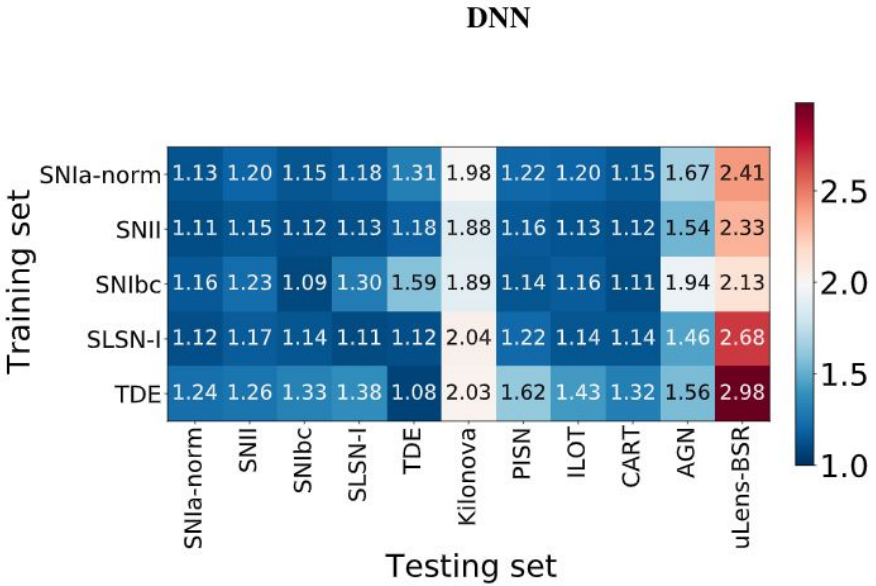
Real-time anomaly detection

Real-time anomaly detection



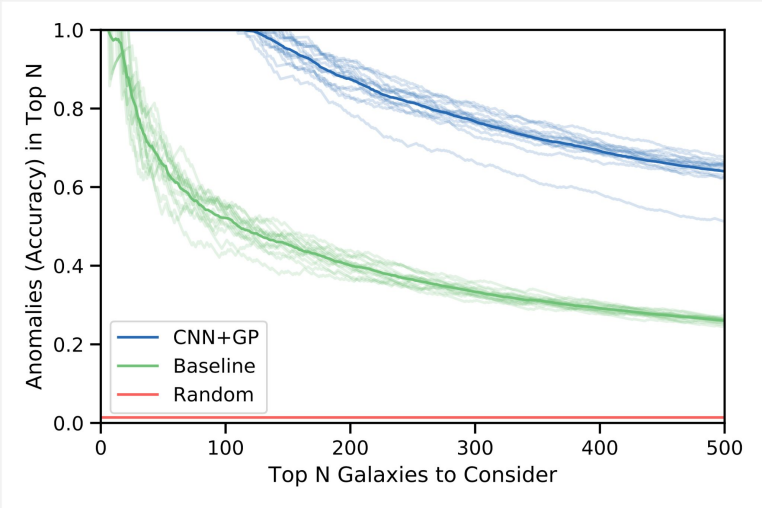
Muthukrishna et al. (2021) 2111.00036

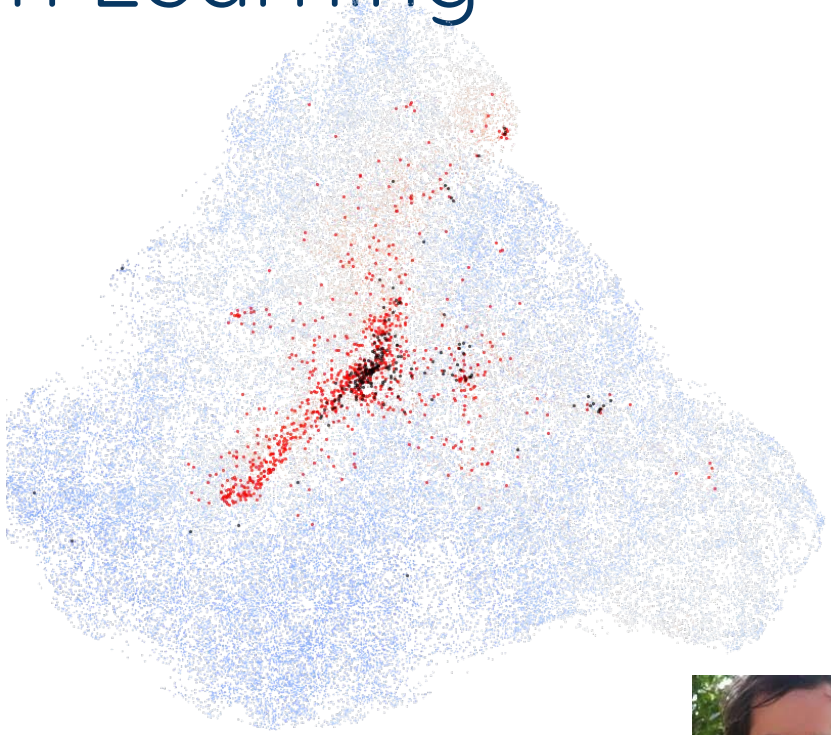
Real-time anomaly detection



Muthukrishna et al. (2021) [2111.00036]

Deep Representation Learning





Walmsley et al. (2021) [2110.12735]

www.supernovafoundation.org

THE SUPERNOVA FOUNDATION

Mentoring For Women In Physics

SANGEETA UJJWAL

Country: India
Research Field: Physics
Institute: University of Delhi
Position: Postdoc

SATYA GONTCHO A GONTCHO

Country: United States of America
Research Field: Astrophysics and Cosmology
Institute: Lawrence Berkeley National Laboratory
Position: Project Scientist

SHAZRENE S. MOHAMED

Country: South Africa
Research Field: Astronomy
Institute: South African Astronomical Observatory and University of Cape Town
Position: Faculty

SIPHEPHILE NCUBE

Country: South Africa
Research Field: Physics
Institute: University of the Witwatersrand
Position: Postdoc

SIYI ZHOU

Country: Sweden
Research Field: Theoretical Physics
Institute: Stockholm University
Position: Postdoc

SUDESHNA BORO SAIKIA

Country: Austria
Research Field: Astrophysics and Cosmology
Institute: University of Vienna
Position: Post doctoral researcher

SWARNAMALA SIRSI

Country: India
Research Field: Theoretical Physics
Institute: University of Mysore
Position: Associate Professor (retired)

VALERIA PETTORINO

Country: France
Research Field: Astrophysics and Cosmology
Institute: CEA
Position: STAFF Scientist

Conclusions

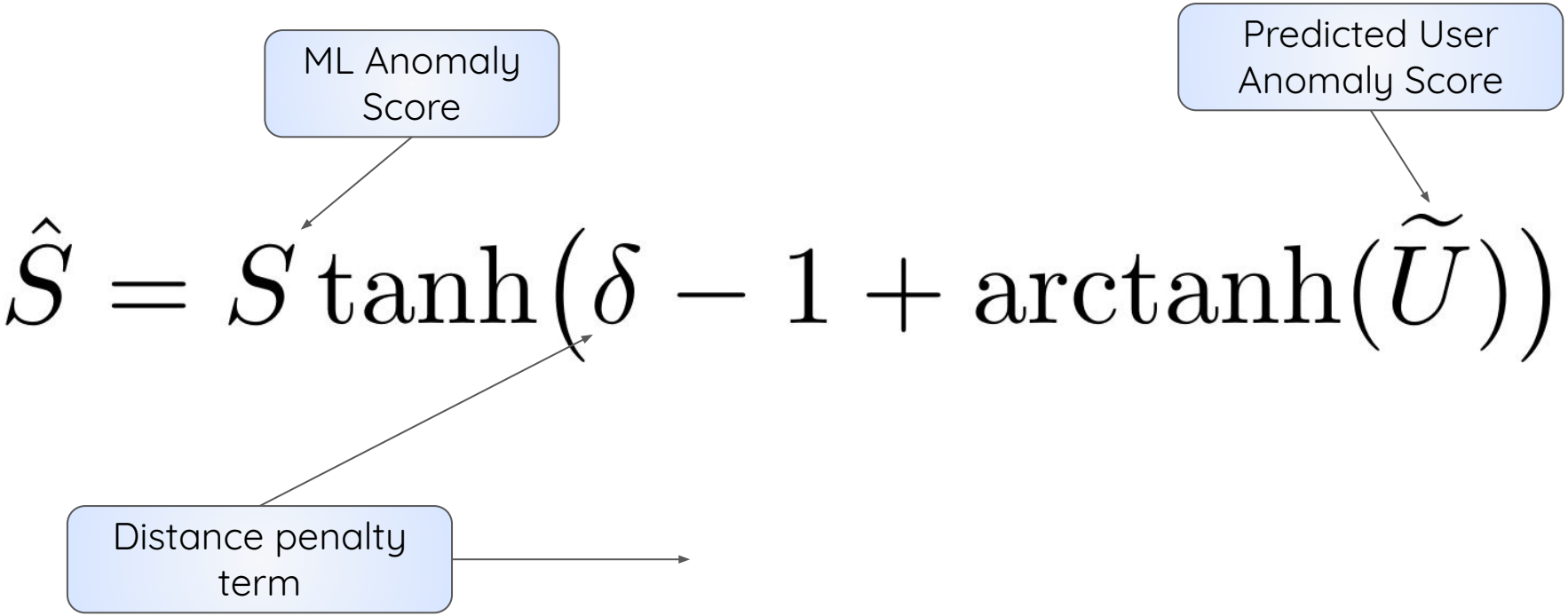
- Machine learning is critical in facing the data deluge
- We need automated anomaly detection to ensure scientific discoveries in datasets aren't missed
- Check out Astronomaly:
  - https://arxiv.org/abs/2010.11202
  - https://github.com/MichelleLochner/astronomaly
- And the Supernova Foundation:
  - https://www.supernovafoundation.org/

dr.michelle.lochner@gmail.com

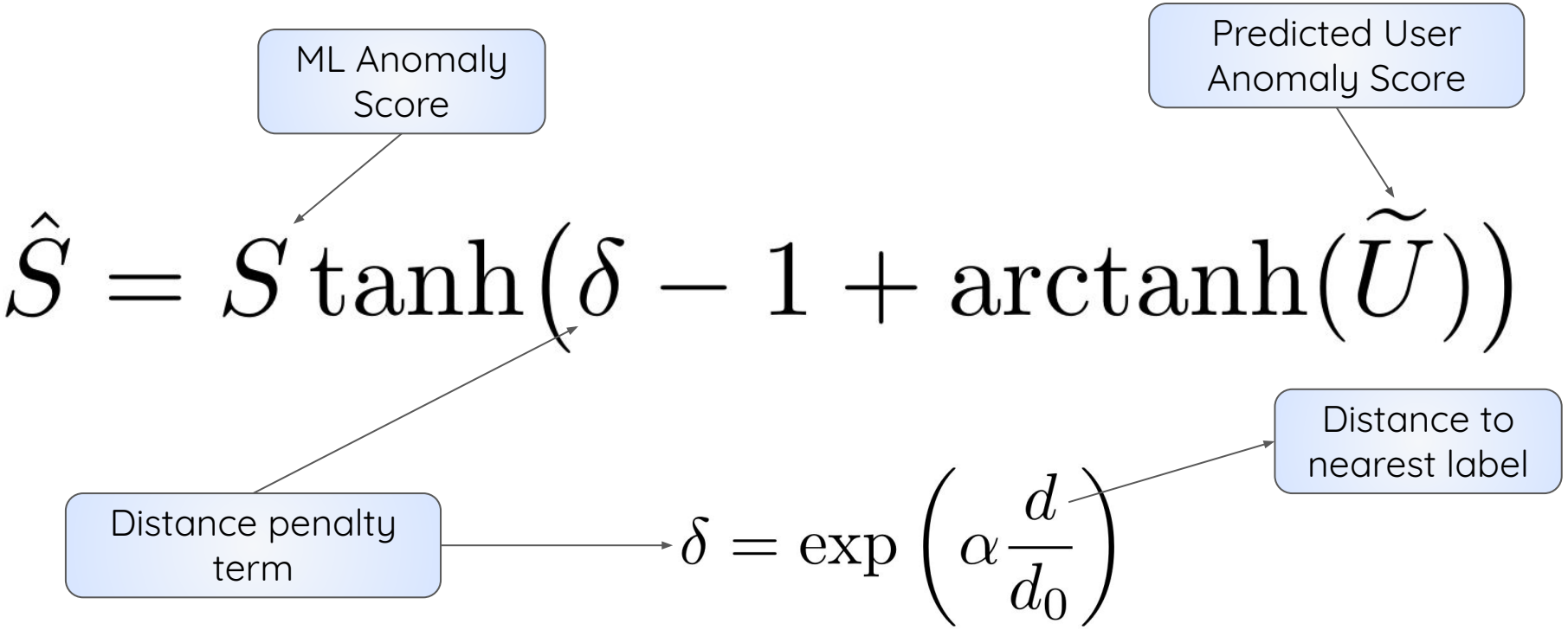
A Novel Active Learning Approach

$$\hat{S} = S \tanh\left(\delta - 1 + \operatorname{arctanh}(\widetilde{U})\right)$$

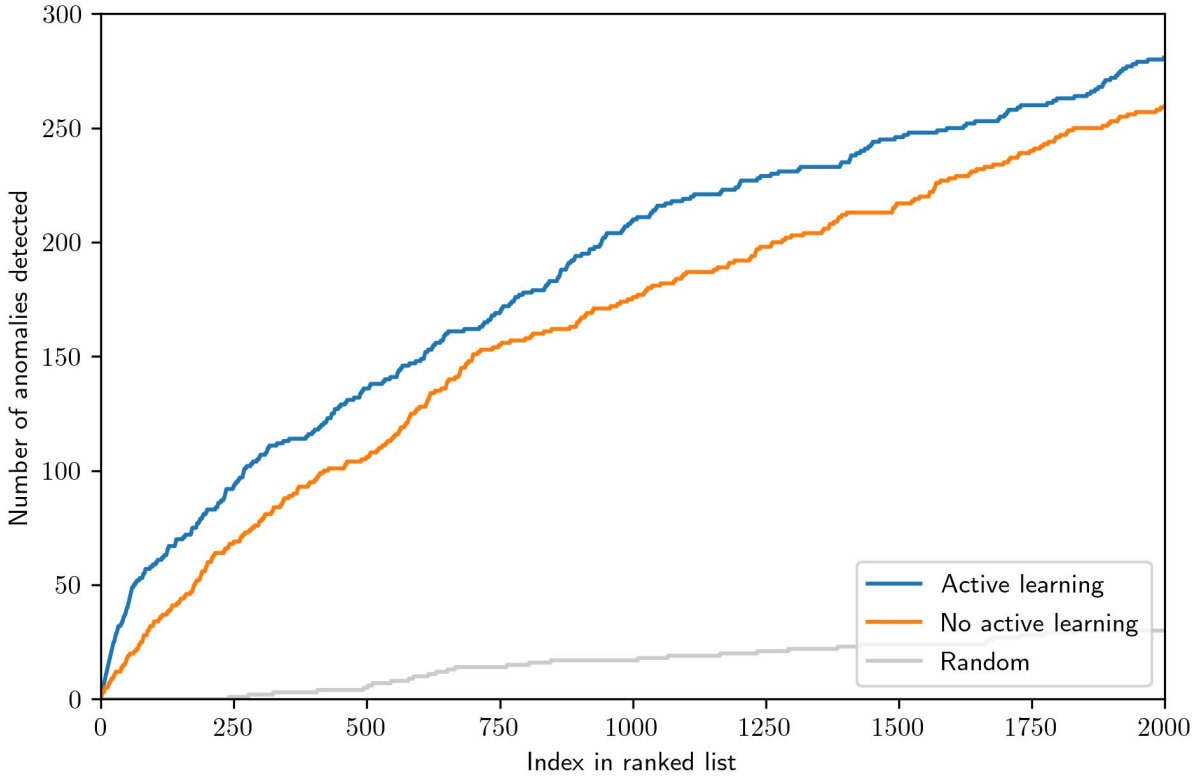
A Novel Active Learning Approach



A Novel Active Learning Approach



Galaxy Zoo - Active Learning



Lochner and Bassett (2020) - 2010.11202