

ThoughtWorks®

Technology Led | Global Software Consultancy

MACHINE LEARNING FOR SCIENTIFIC DISCOVERY IN ASTRONOMY

SHRADDHA SURANA¹

SHRADDHA.SURANA@THOUGHTWORKS.COM

YOGESH WADADEKAR² AND DIVYA OBEROI²

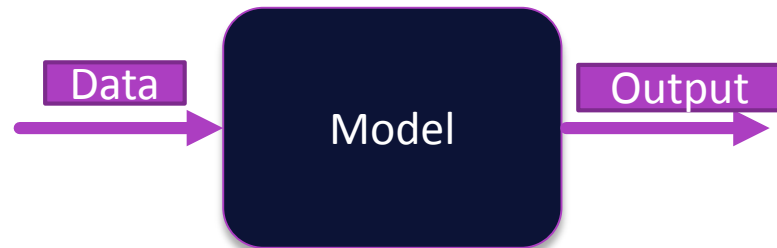
¹ THOUGHTWORKS

² NATIONAL CENTRE FOR RADIO ASTROPHYSICS

Training



Prediction



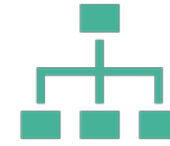
MACHINE LEARNING

- Don't code the logic
- Code the framework that will allow the algorithm to learn the logic. To extract information from the data.

TYPICAL USE CASES



Computer
vision



Classification



Clustering



Recommendation
systems



Natural
Language
Processing



Reinforcement
Learning

ML IN ASTROPHYSICS

- Star Formation Histories
- Solar Radio Imaging

STAR FORMATION HISTORIES

The background is a gradient from dark purple at the top to deep blue at the bottom, filled with a field of small, faint white stars. Overlaid on this are several faint, white geometric and astronomical diagrams. In the top right, there is a large circular scale with degree markings from 0 to 210 and concentric circles with arrows indicating rotation. In the bottom right, there is a diagram of two concentric circles with arrows showing a clockwise path. In the bottom left, there is a partial view of a similar circular diagram with an arrow.

STAR FORMATION HISTORIES OF GALAXIES

- Stellar population synthesis models
- Data from GAMA survey - a large multi wavelength survey using several different telescopes (76,455 galaxies)

STAR FORMATION HISTORIES



Deep learning



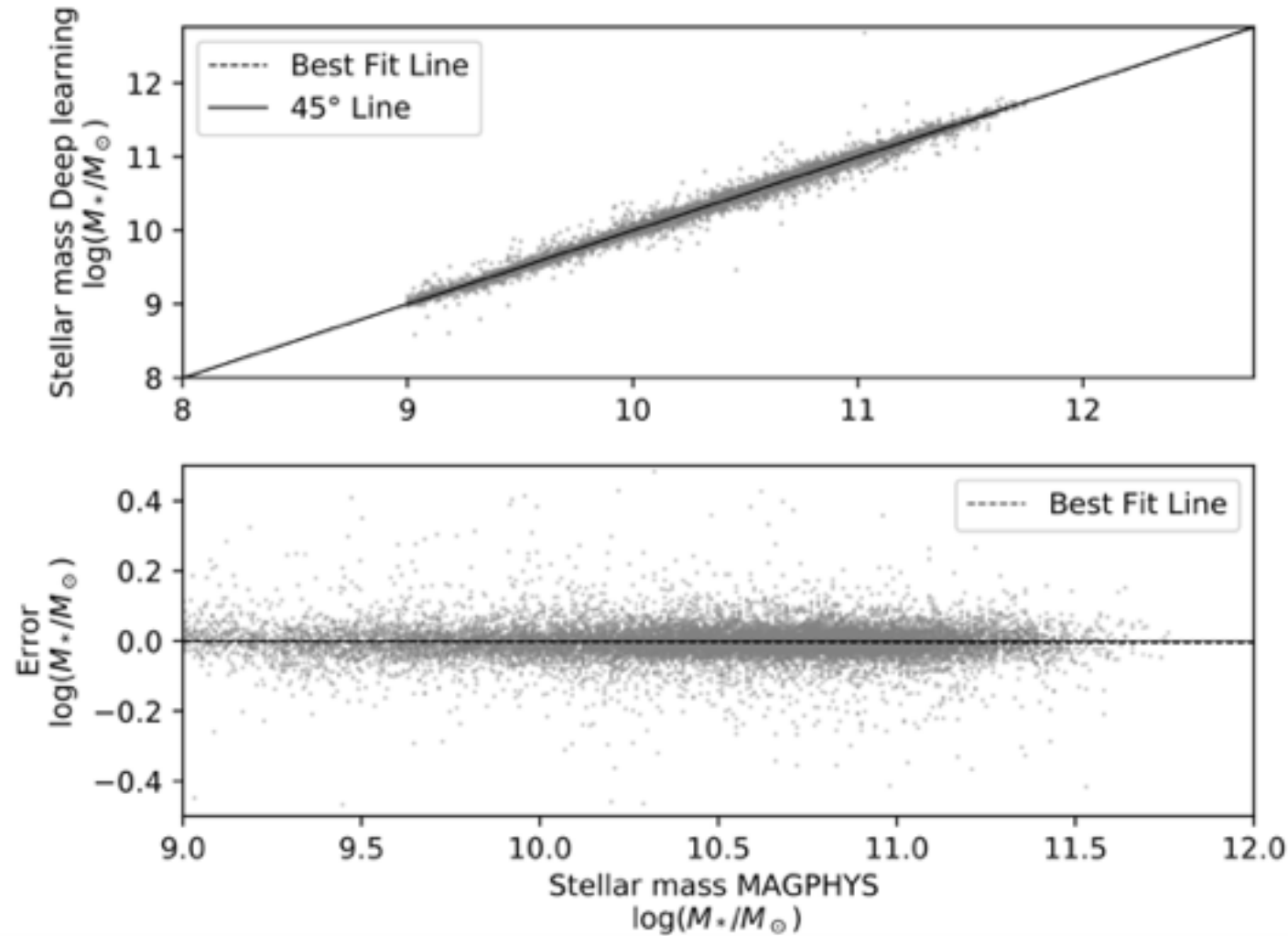
10mins per galaxy (~2.5 months for million galaxies) to seconds for millions of galaxies



Accelerates in our understanding of the galaxy evolution and hence the universe

STELLAR MASS

- Able to predict over 3 orders of magnitude
- Able to estimate within 0.06



SOLAR RADIO IMAGING



ABOUT SOLAR RADIO IMAGING

- Murchison Widefield Array (MWA) – It's a new generation radio interferometer. Precursor for the SKA (Square Kilometre Array)
- Generates up to $\sim 10^5$ images every minute.
- Traditional methods become infeasible
- We need creative ways of using existing ML algorithms
- Unsupervised learning is a complex challenge

SOLAR RADIO IMAGING



Unknown - unknown. We really don't know what we are looking for.

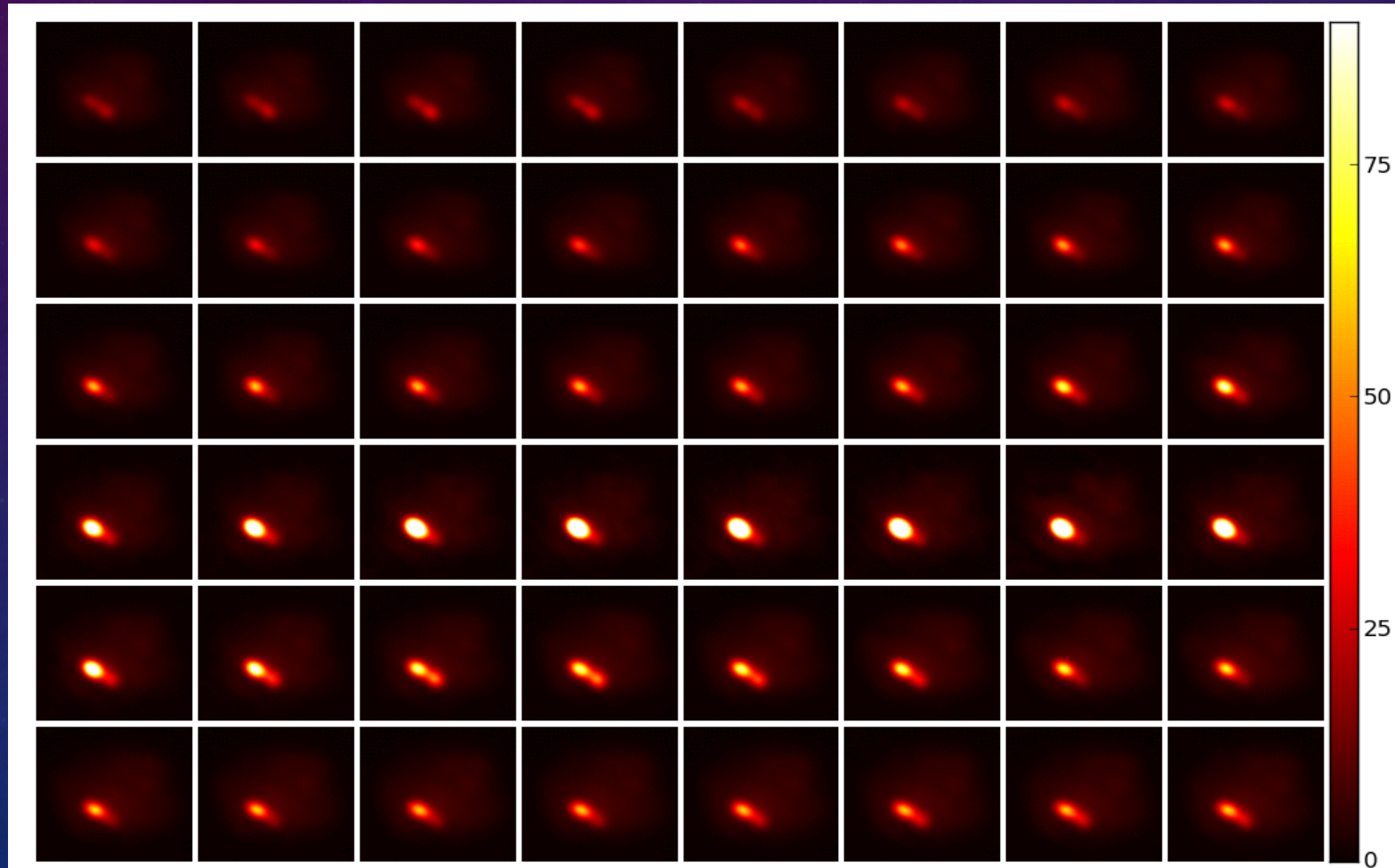


Discover patterns & structures in the dataset

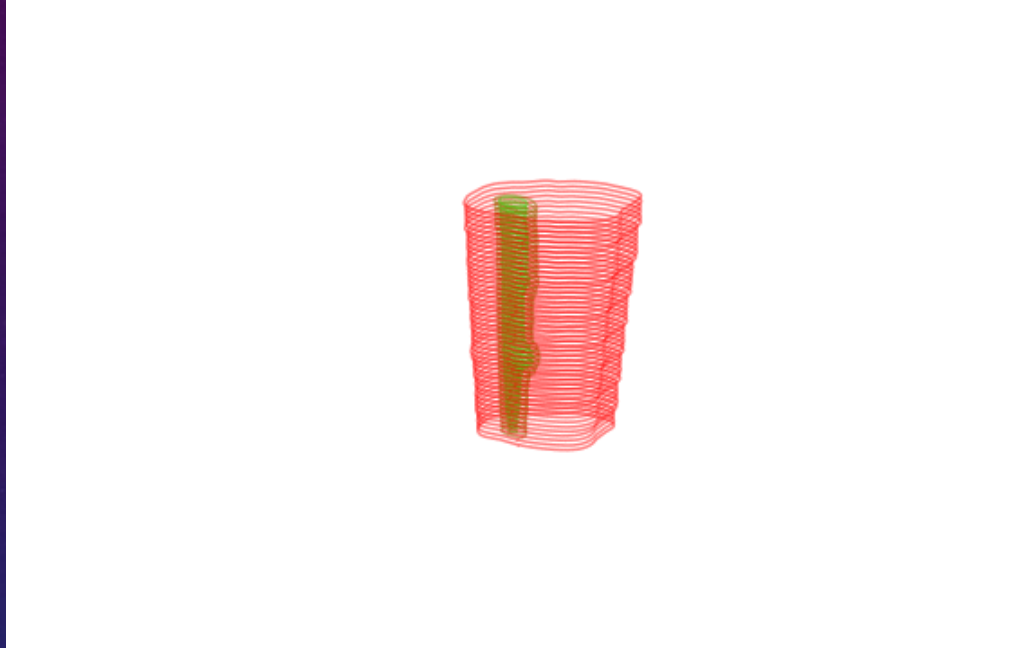


Unsupervised ML techniques.

RADIO IMAGES OF THE SUN IN 48 DIFFERENT SPECTRAL CHANNELS

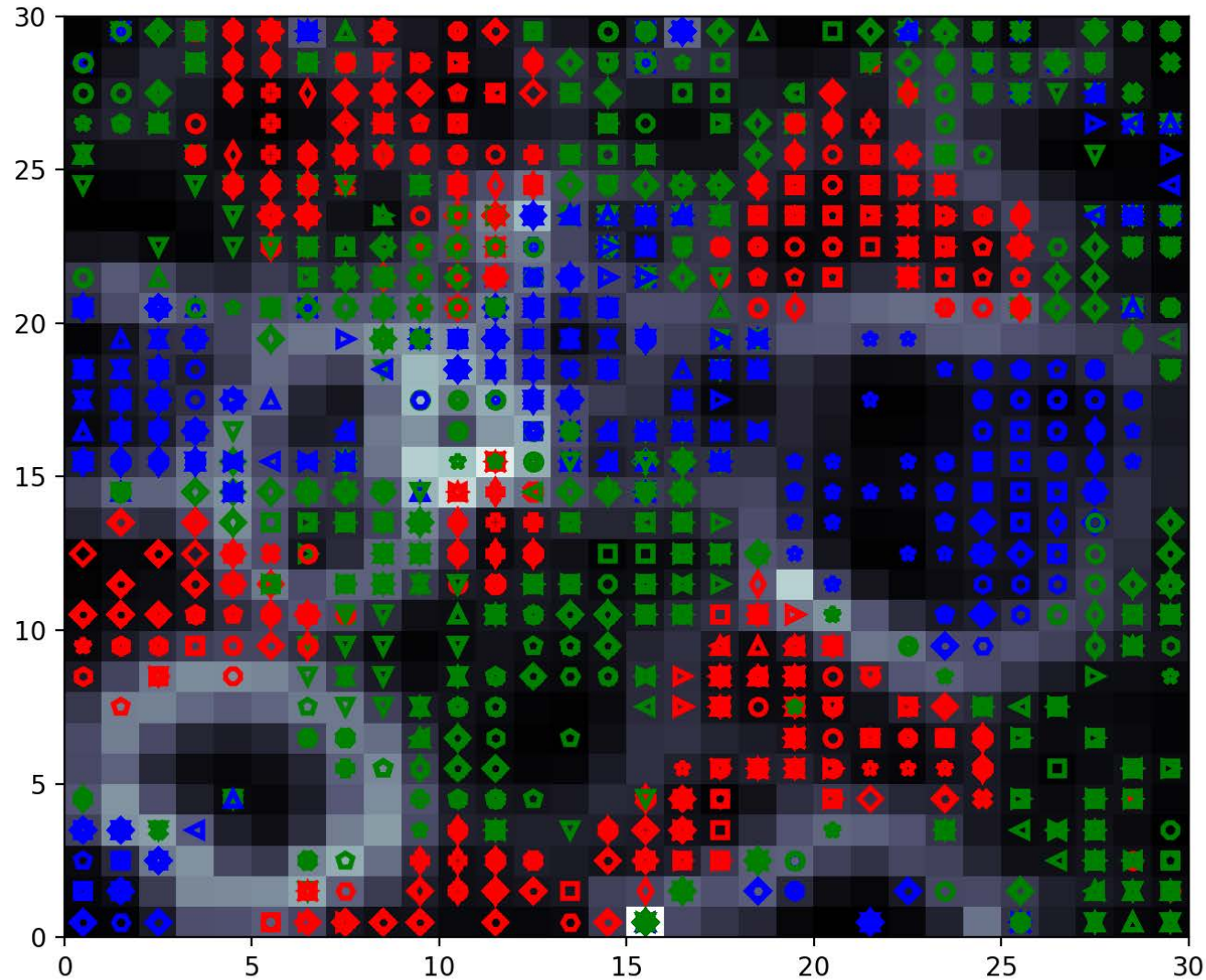


DATA HYPERCUBE I(Θ, Φ, N, T)



Size of the data cubes: 100 x 100 (800MB), 200 x 200 (3.3 GB), 500 x 500 (40 GB)

OUTPUT OF A SELF- ORGANIZING MAP



ADVANTAGE OF ML OVER CLASSICAL TECHNIQUES OF STATISTICAL ANALYSIS



Domain agnostic



Faster



Feasible

THANK YOU

SHRADDHA.SURANA@THOUGHTWORKS.COM



ENGINEERING FOR RESEARCH (E4R)

e4r@thoughtworks.com | thoughtworks.com

APPENDIX

