Estimating the Photometric Redshifts of Galaxies Using Regression Techniques

A. Momtaz, M. H. Salimi, S.Shakeri

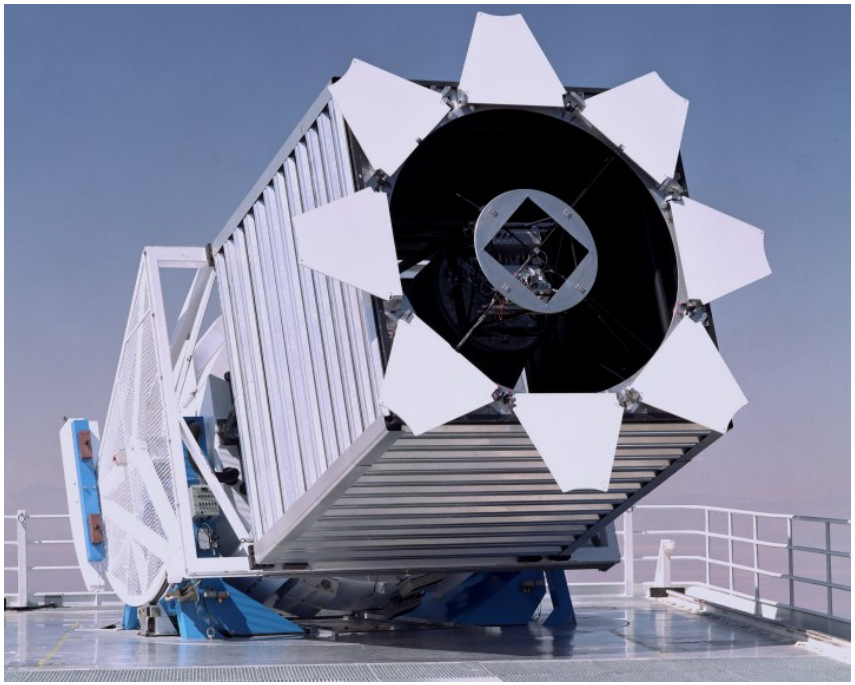
Department of Physics, Isfahan University of Technology, Isfahan 84156-8311, Iran.

aidinmomtaz@ph.iut.ac.ir
mhsalimi@ph.iut.ac.ir
s.shakeri@iut.ac.ir

MG-16, July 7, 2021

Overview

Astronomical Surveys, SDSS

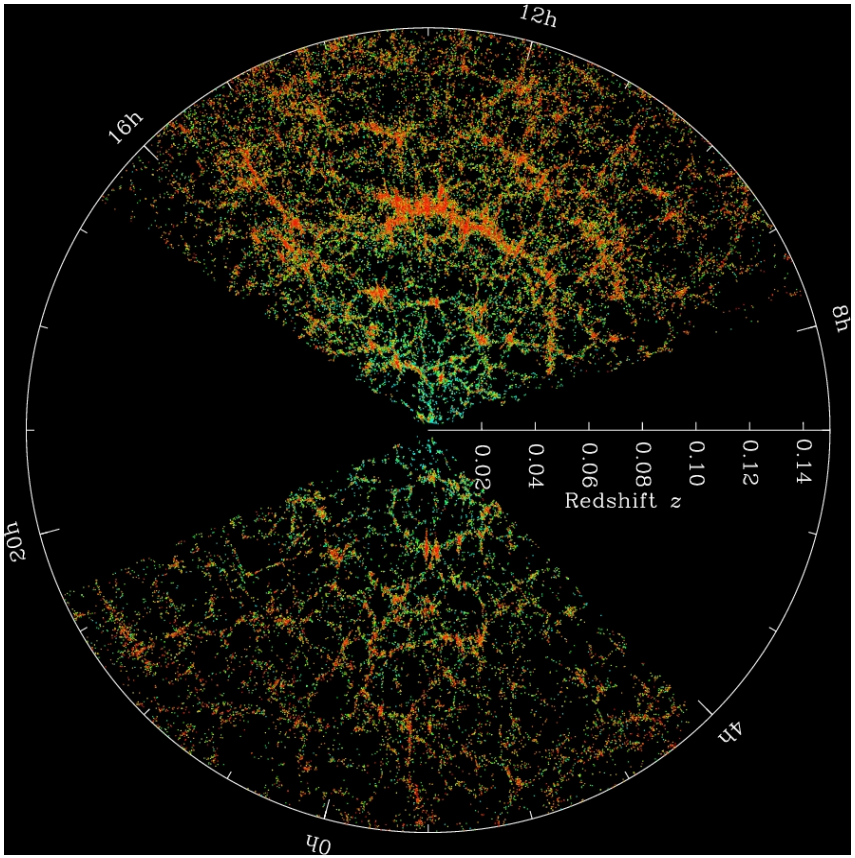


https://www.sdss.org/science/

SDSS Telescope

- Galactic Archaeology
- Gas in the Galaxy
- Star Forming Regions
- Multi-Star and Planetary Systems
- White Dwarfs

SDSS Data Release

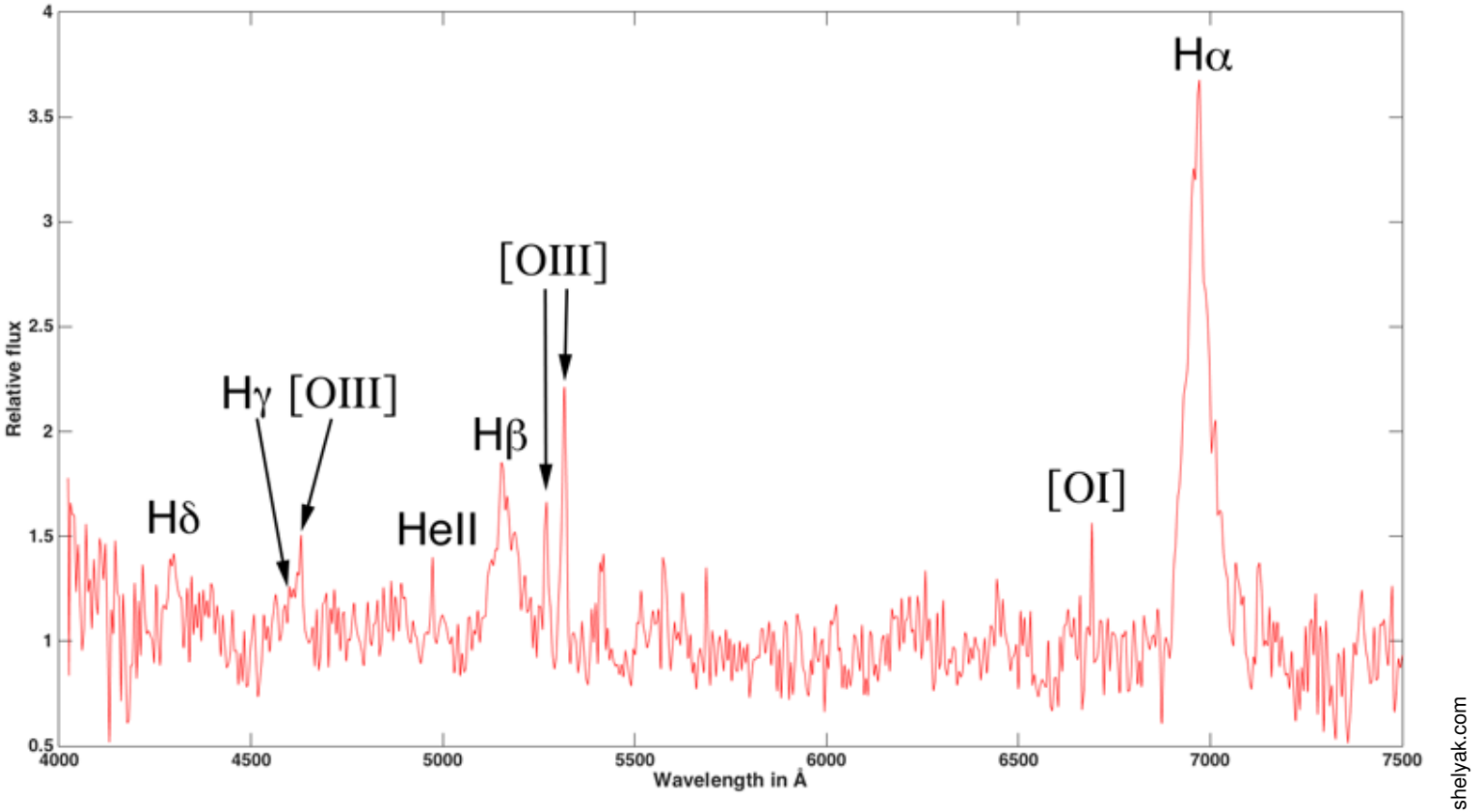


https://www.sdss.org/science/

SDSS Galaxy Mapping

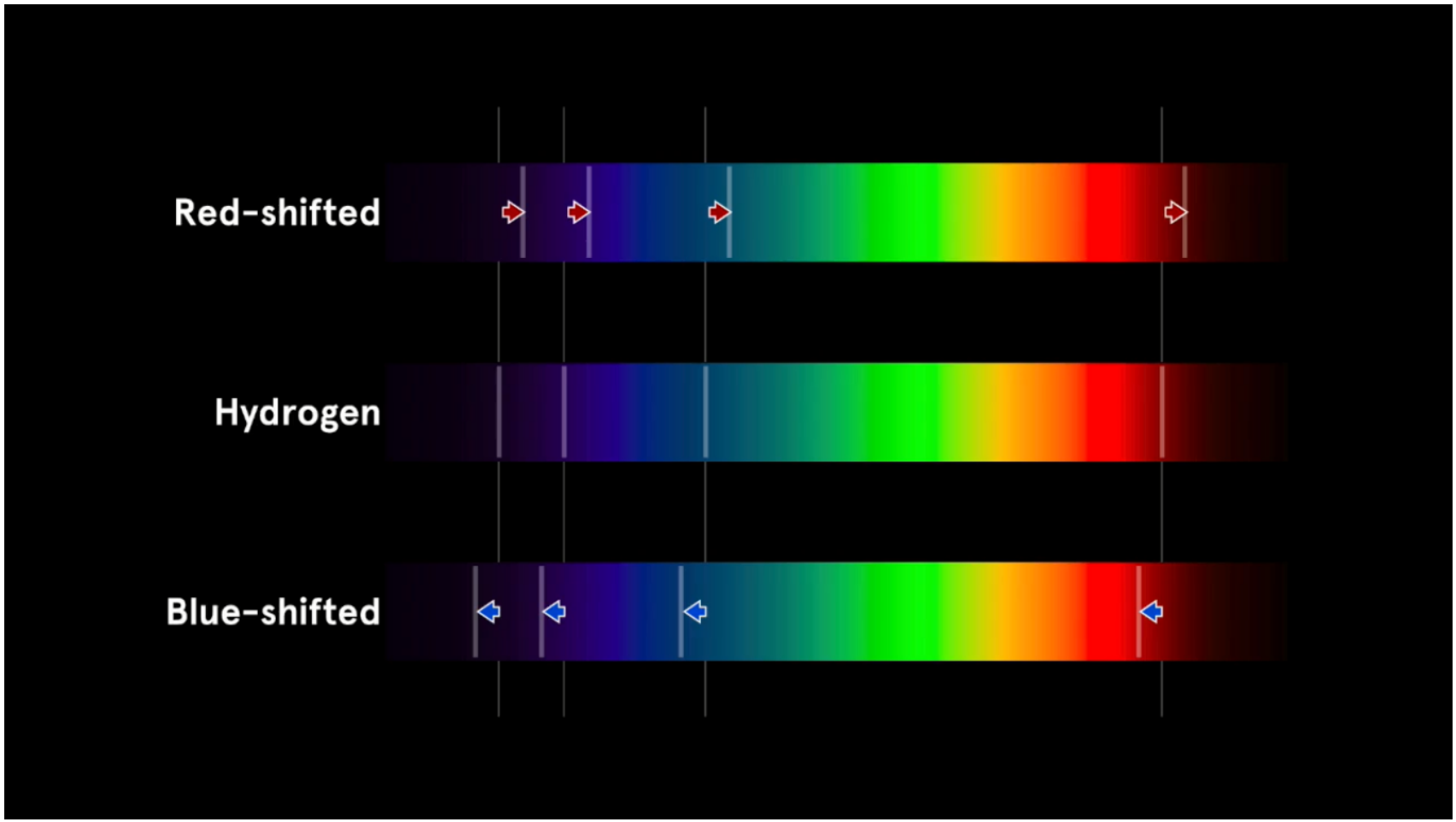
- Data Release 16 (August 2018)
- Infrared spectra
- Extragalactic spectra
- Integral field spectra for nearby galaxies
- Useful spectra: 4,846,156
- Galaxies: 2,863,635, QSO: 960,678

Measuring Redshift from Spectroscopy



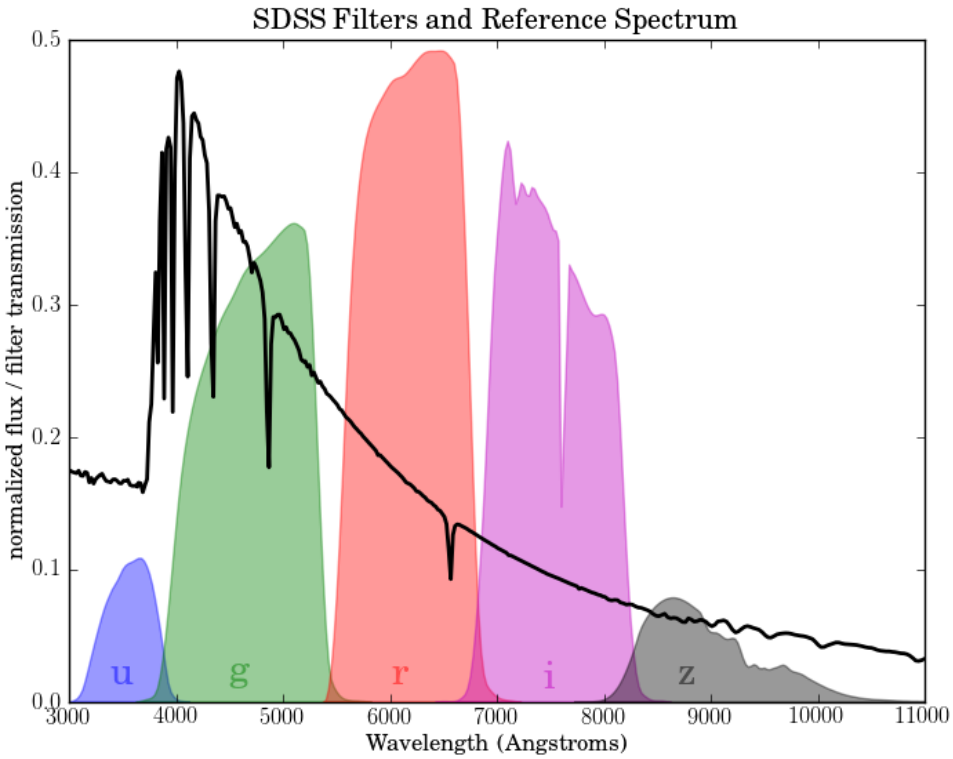
$$\lambda_{obs} = (1 + z)\lambda_{em} \tag{1}$$

Redshift and Blueshift



Coursera, Data-driven Astronomy

Measuring Redshift from Photometry



Coursera, Data-driven Astronomy

$$u = m_{\text{ref}} - 2.5 \log 10 \left[\int_0^{\infty} F(\lambda) S(\lambda) d\lambda \right] \quad (2)$$

Regression Algorithms

Decision Tree

Decision trees map a set of input features to their corresponding output targets. This is done through a series of individual decisions where each decision represents a node (or branching) of the tree.

Regression Algorithms

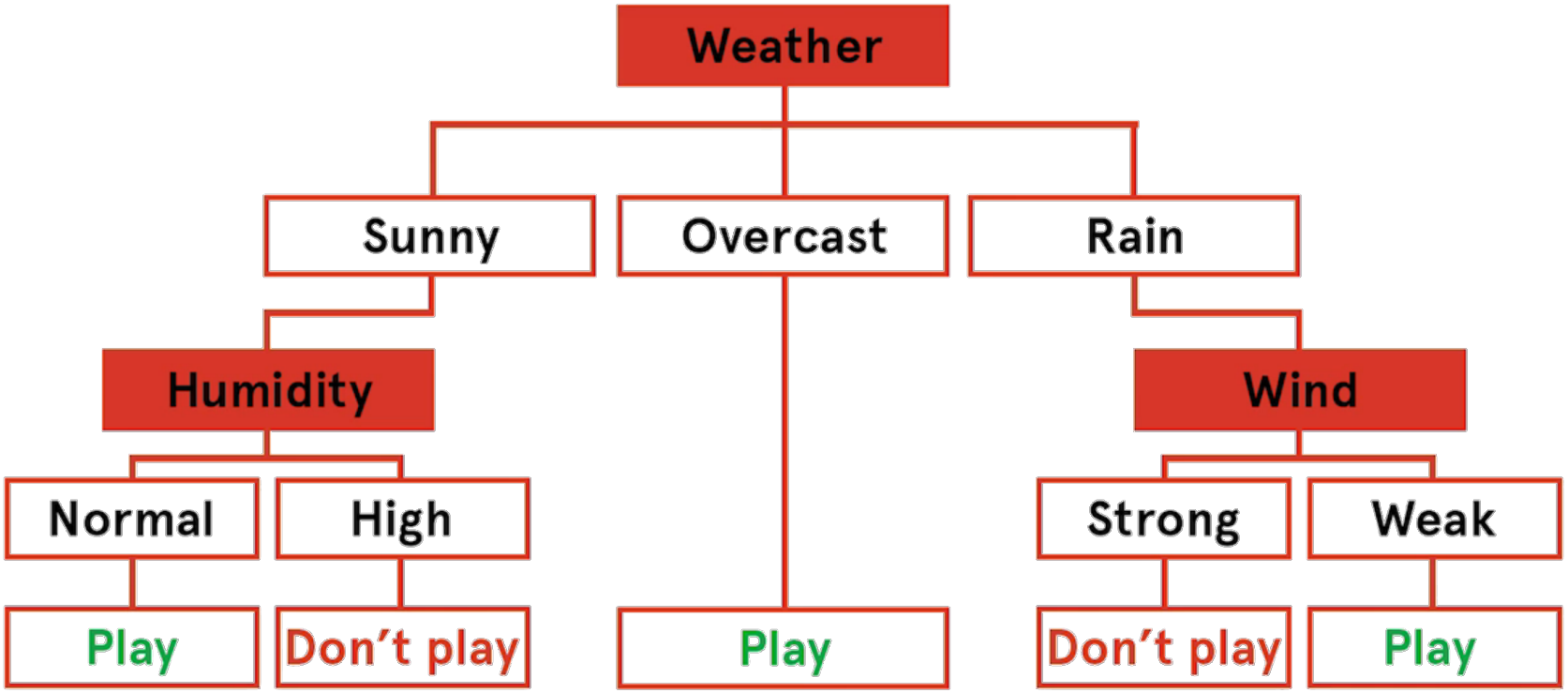
Decision Tree

Decision trees map a set of input features to their corresponding output targets. This is done through a series of individual decisions where each decision represents a node (or branching) of the tree.

Random Forest

Random forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

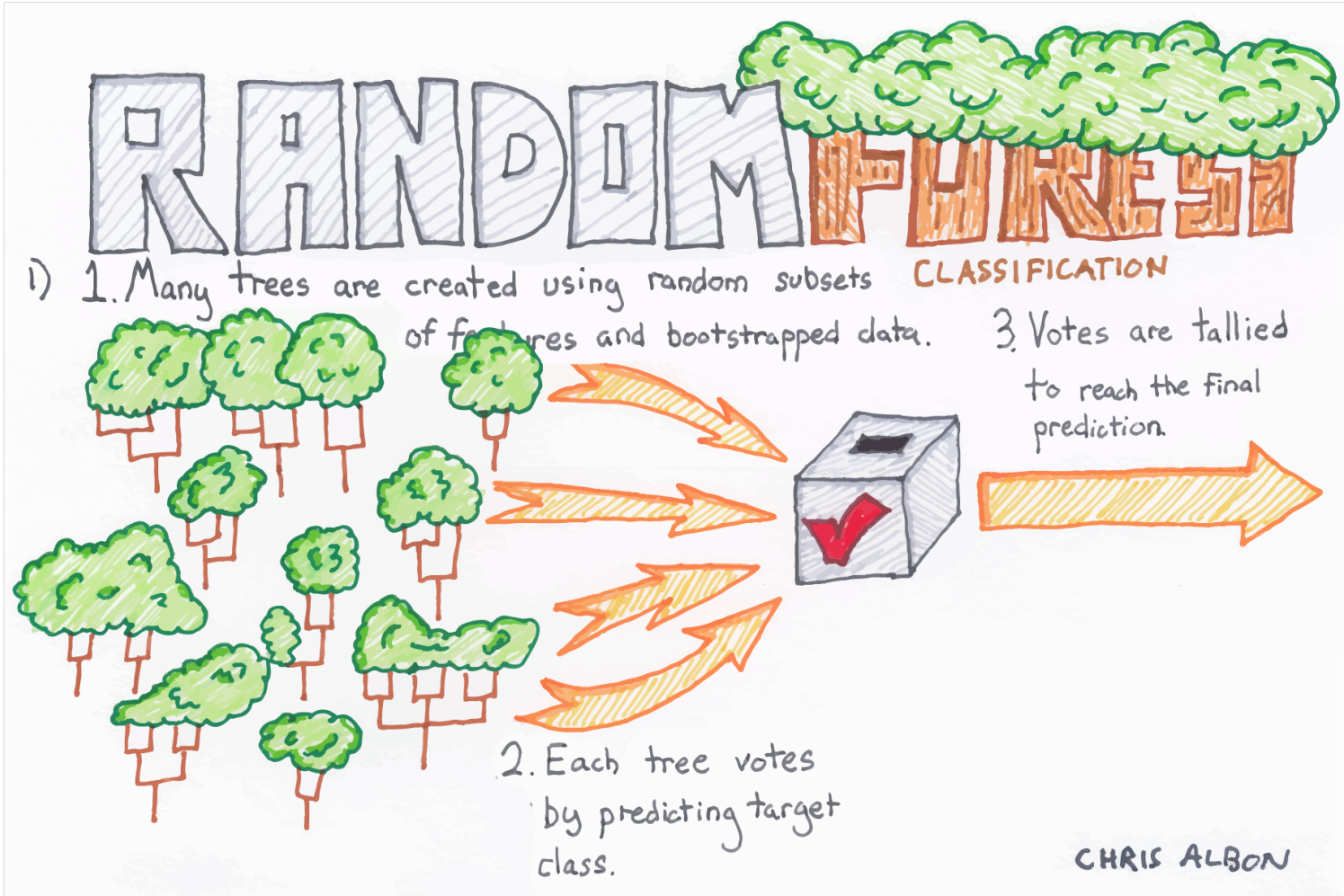
How does Decision Tree work?



Coursera, Data-driven Astronomy

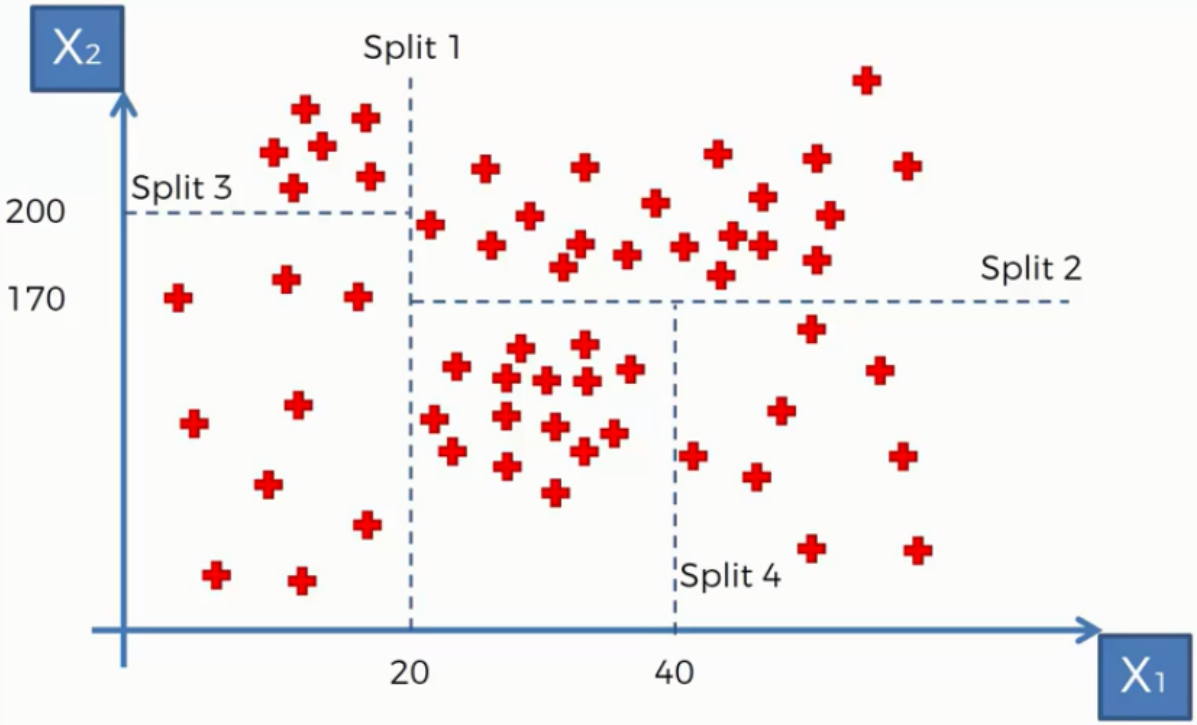
Schematic View of Decision Tree

How does Random Forest work?



Schematic View of Random Forest

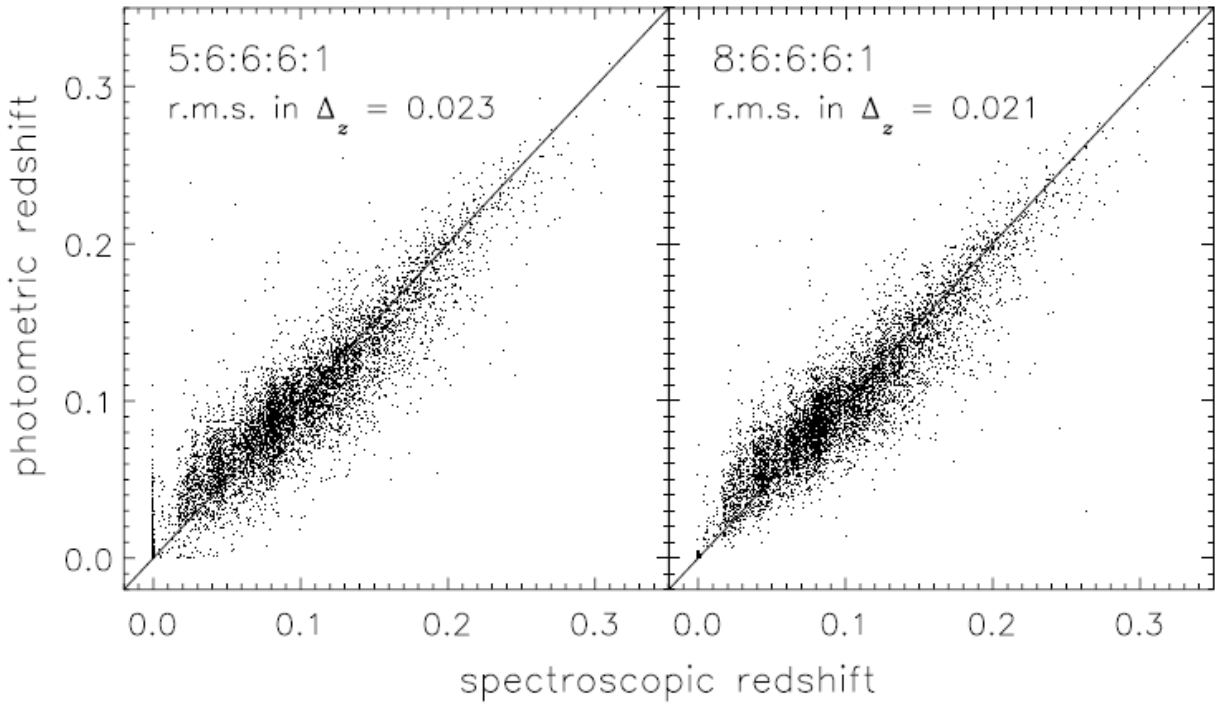
Geometrical View of Algorithm



Udemy-Machine Learning A-Z

Desicion Tree Intuition

ANN



Firth et al. Mon. Not. R. Astron. Soc. 339,(2003)

ugriz Photometry

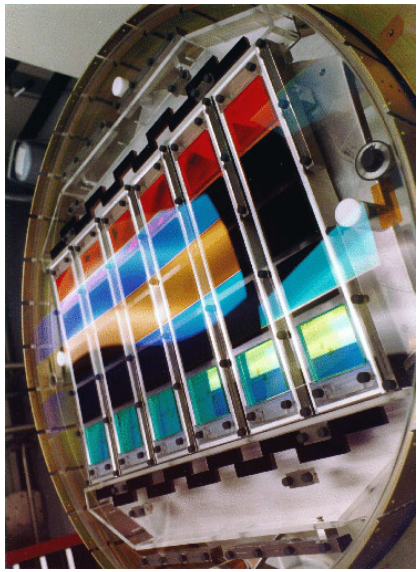
- Sample Size: 20000 Galaxies
- Testing Set: 7000 Galaxies

Similar Research Results

Reference Articles	Data Release	Training Set	$\overline{\Delta z_{norm}}$	ML Algorithm
Beck et al (2016) MNRAS	**12**	1'976'978	5.84×10^{-5}	**Local Linear Regression**
Paul et al (2018) Galaxies-MDPI	**12**	20'000	2×10^{-3}	**Random-Forest**
Baldeschi et al (2021) APJ	**16**	1'251'249	1×10^{-3}	**Random-Forest**
A. Momtaz, M. H. Salimi, S. Shakeri (2021) In preparation	**16**	400'000	1×10^{-3}	**Random-Forest**

$$\overline{\Delta Z}_{\text{norm}} = \frac{Z_{\text{spec}} - Z_{phot}}{1 + Z_{\text{spec}}} \tag{3}$$

Preparing the Dataset



https://www.sdss.org/instruments/

u 3551Å

g 4686Å

r 6166Å

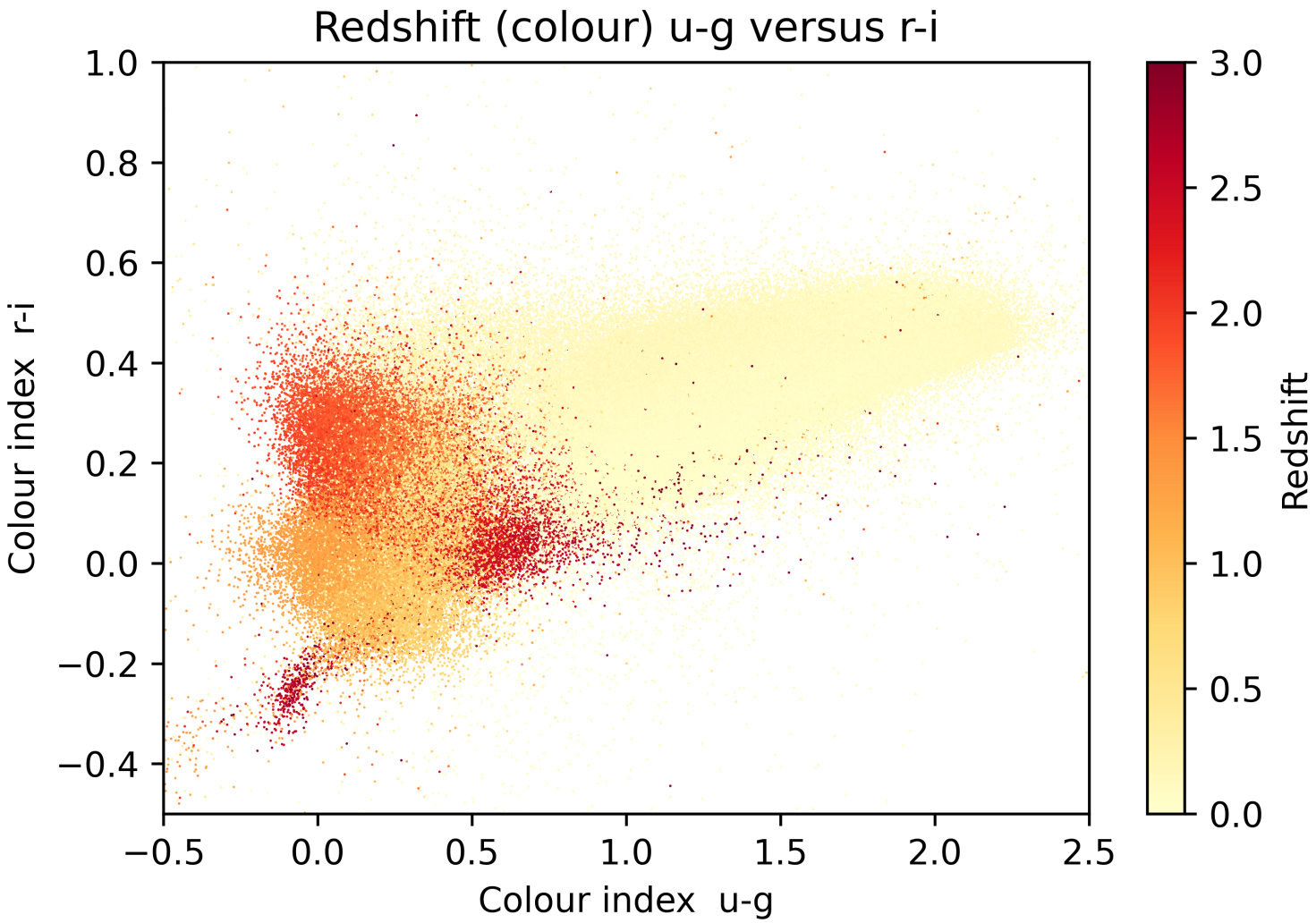
i 7480Å

z 8932Å

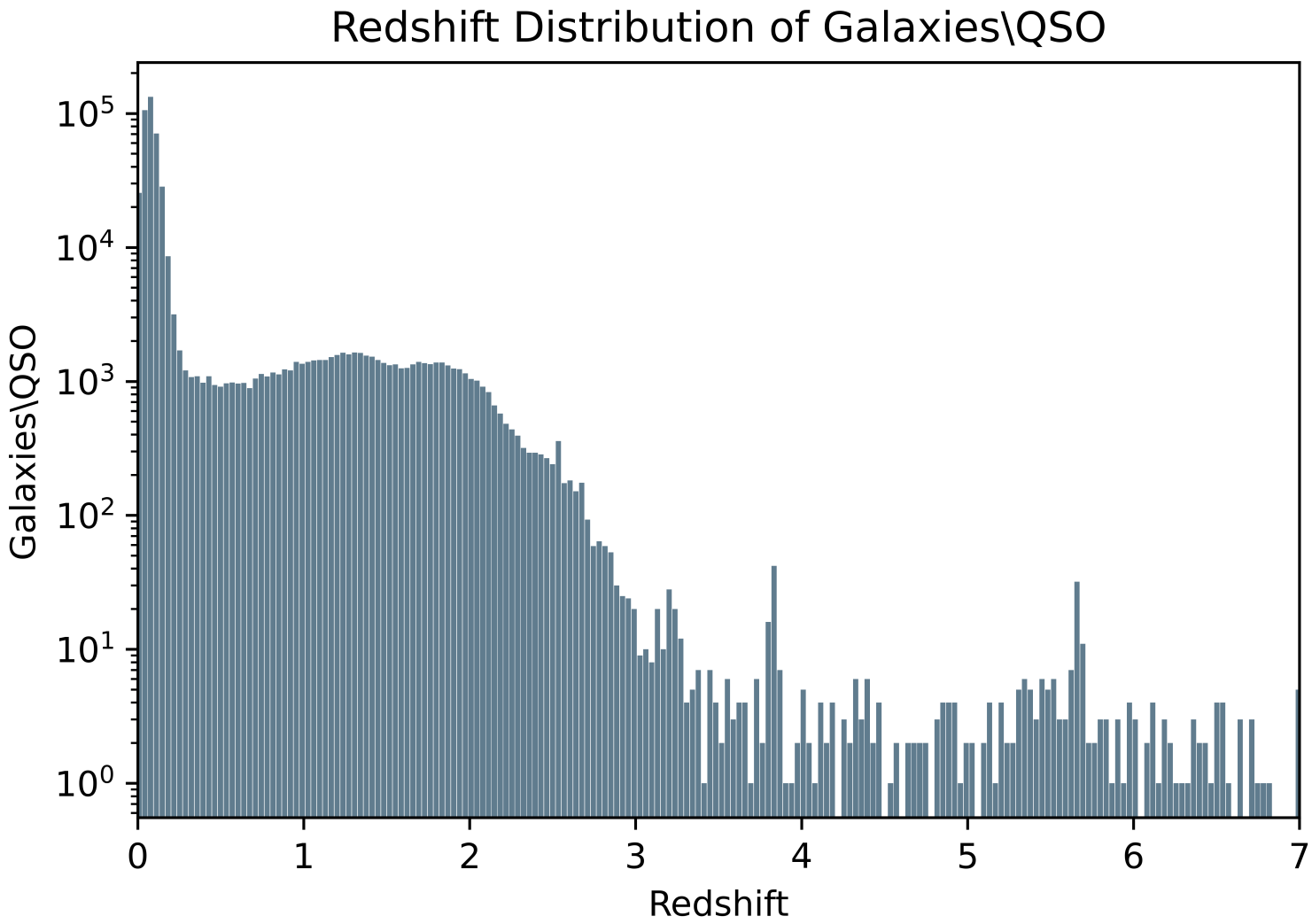
u	g	r	i	z	...	redshift
19.84	19.53	19.47	19.18	19.11	...	0.54
19.86	18.66	17.84	17.39	17.14	...	0.16
...	...	...	...	...	...	...
18.00	17.81	17.77	17.73	17.73	...	0.47

SDSS Data Sample

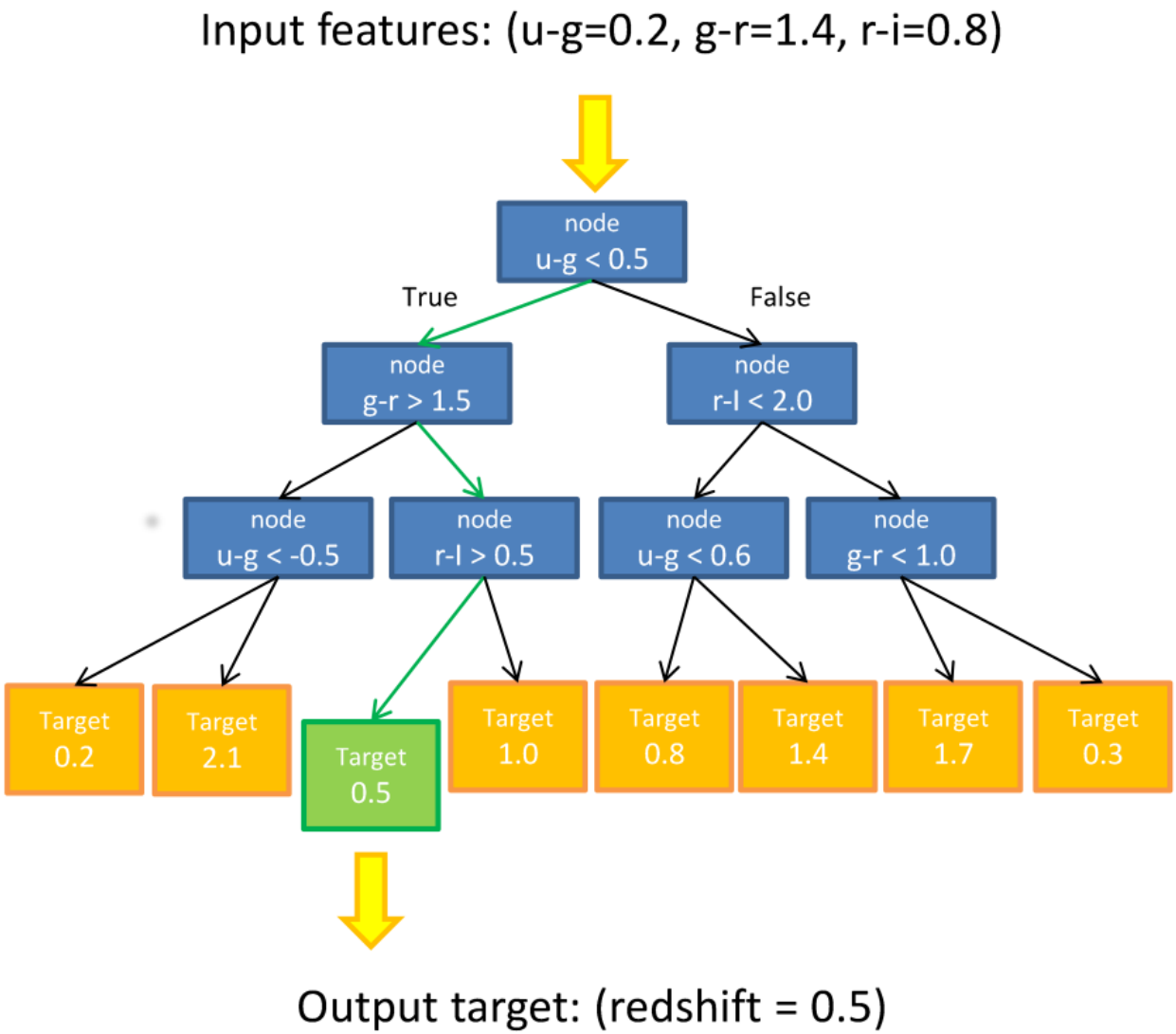
Contour map of the redshifts



Redshift Distribution of Galaxies

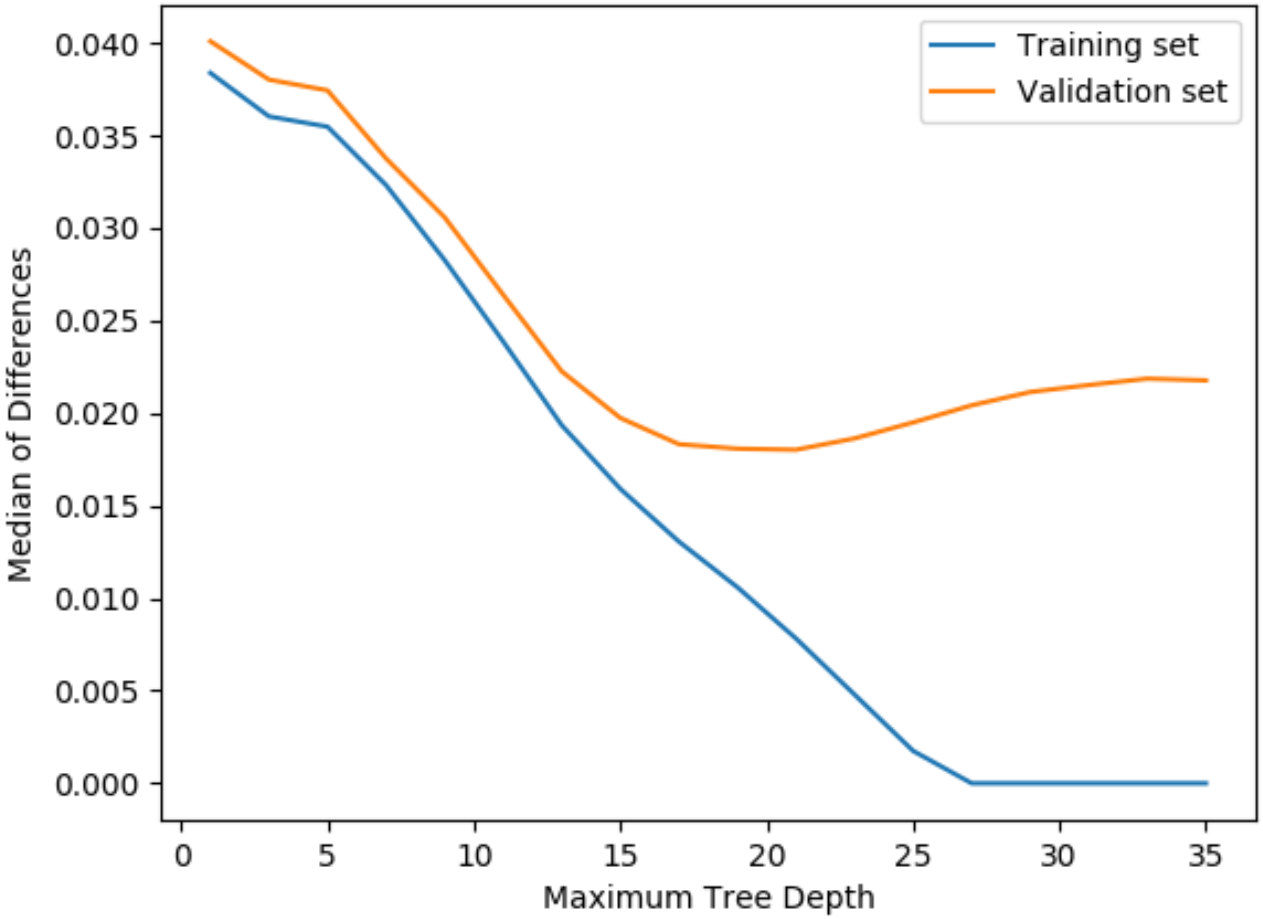


Detailed overview of Decision Tree



Coursera, Data-driven Astronomy

Optimizing process for Decision Tree



Setting a Maximum Depth

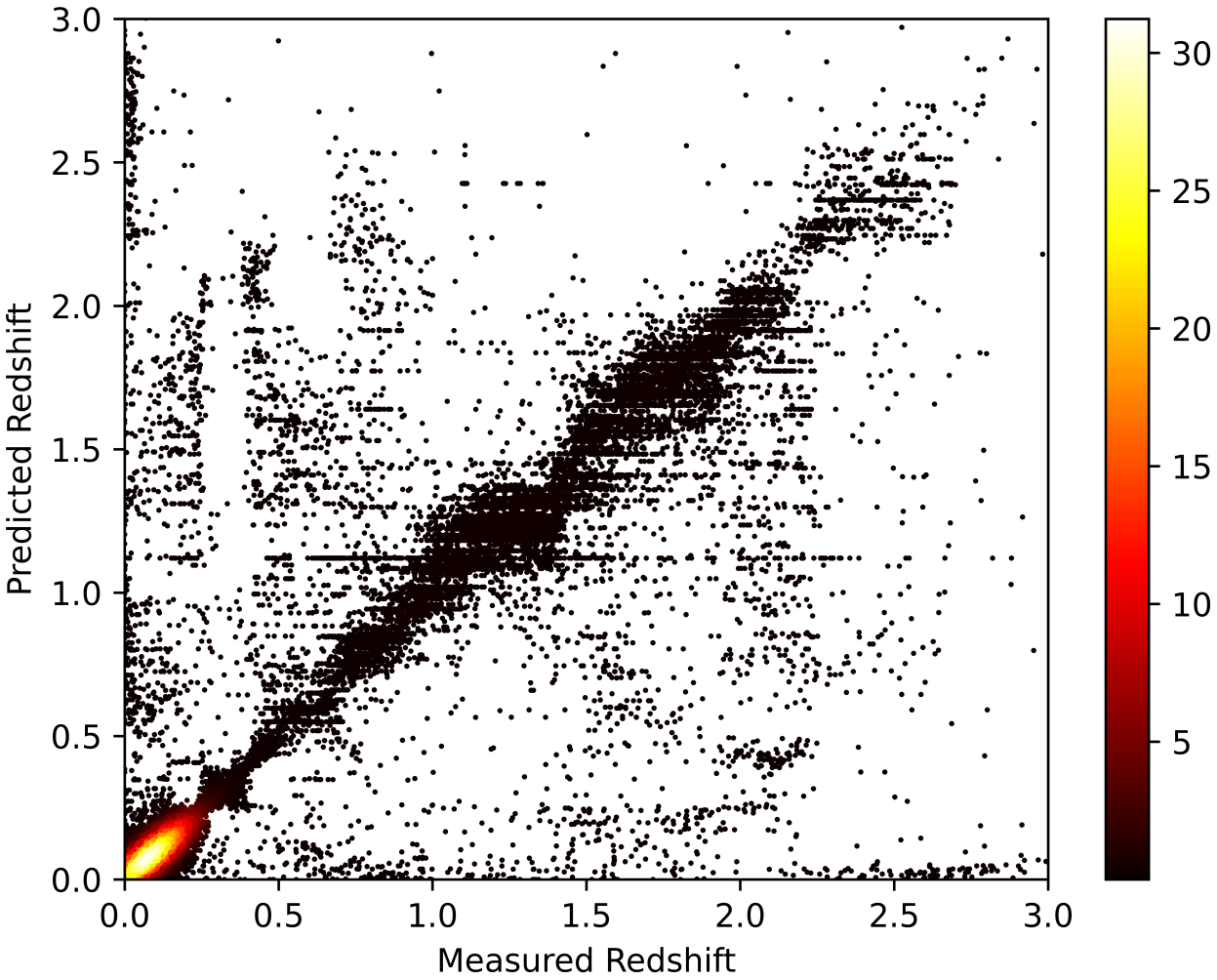
Accuracy of the model

The Effect of Training Set Size

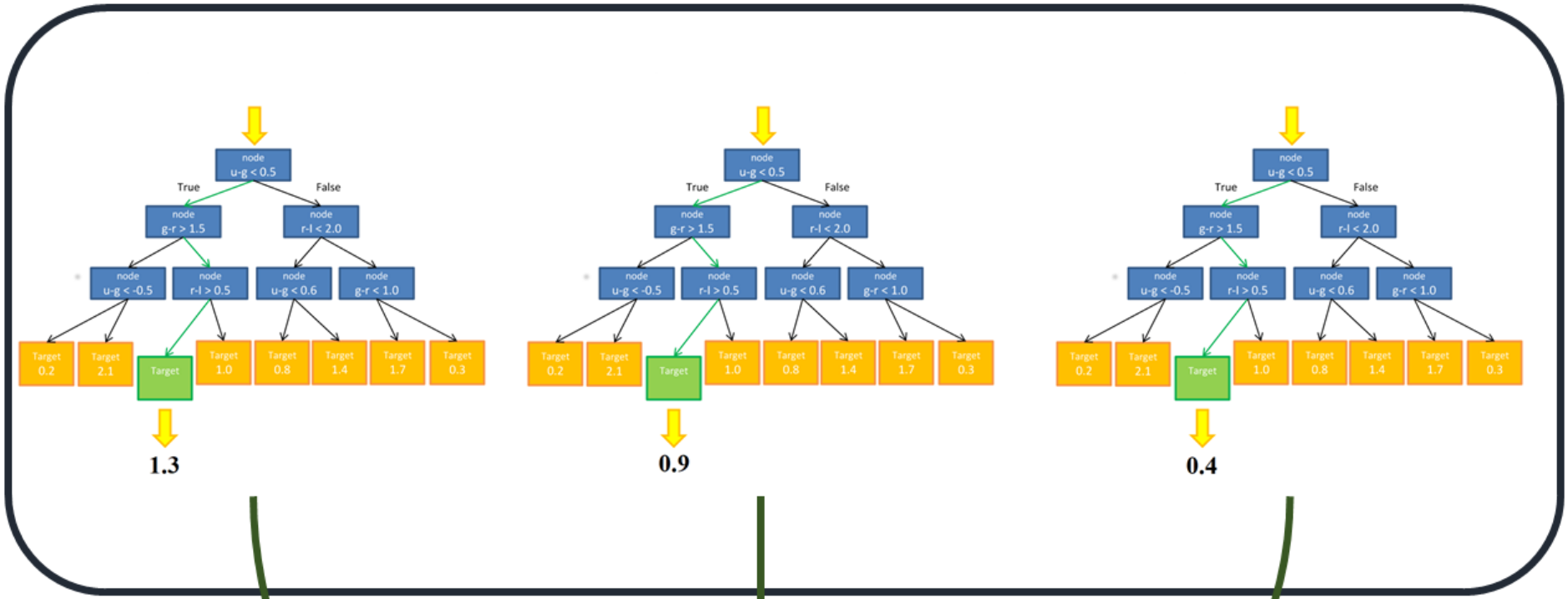
Training Galaxies	Median Diff
50	0.048
500	0.026
5000	0.023
50000	0.022
400000	0.017

$$med\_diff = \text{median}\left(|Y_{\text{i, pred}} - Y_{\text{i,act}}|\right) \tag{4}$$

Final Result of Decision Tree Algorithm

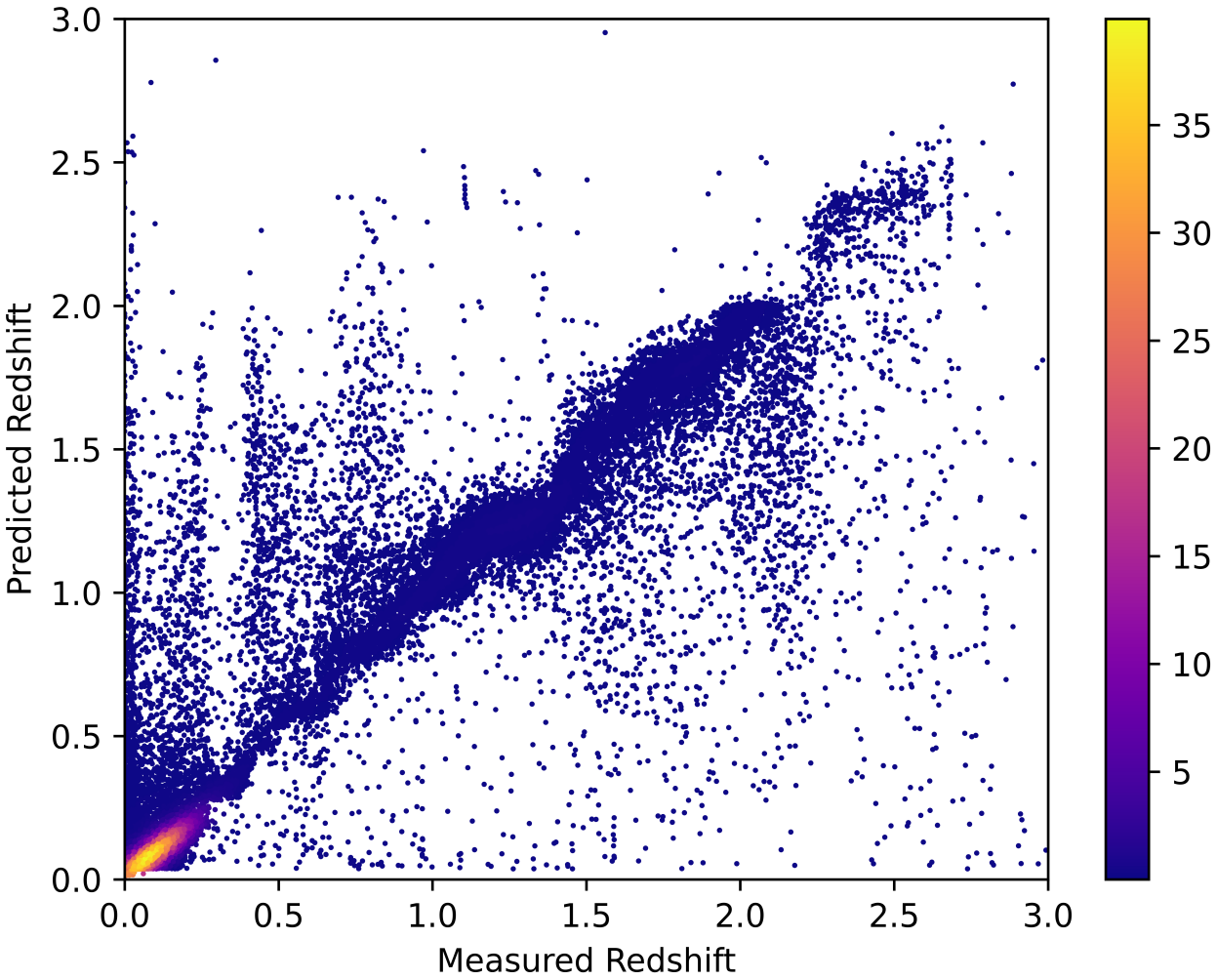


Steps of preparing Random Forest

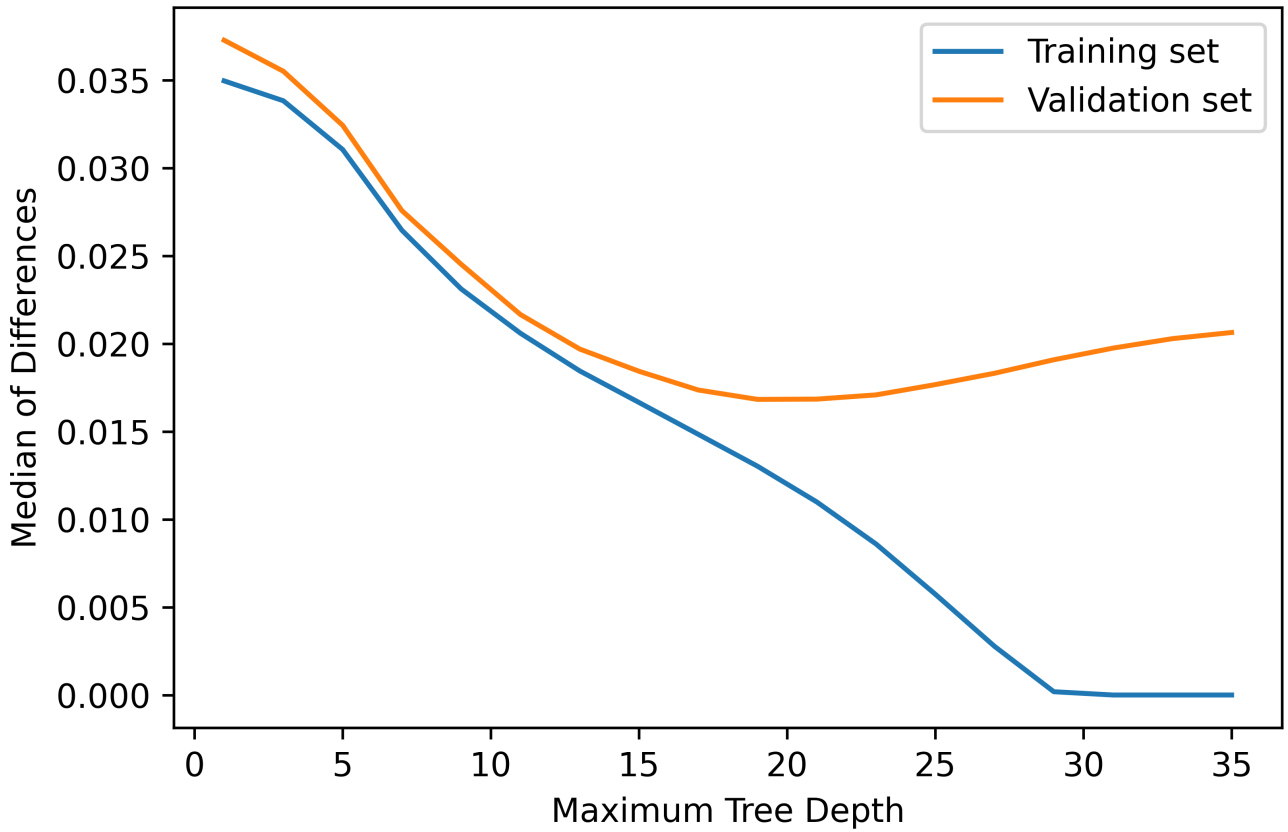


$$\hat{Y}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{y}_b(x) = \frac{1.3 + 0.9 + 0.4}{3} = 0.86$$

Final Result of Random Forest Algorithm



Optimization process



Analytical Overview

Parameters of Regression Modeling

	Decision Tree	Random Forest
Mean Accuracy	67.96%	81.01%
Best Accuracy	68.07%	81.02%
Max Depth	19	25
Number of Trees	1	600
Bootstrap	-	True
Standard Deviation	1.66%	0.69%
R2 Squared	0.68	0.81
Median Diff	0.017	0.016
$\overline{\Delta Z}_{\text{norm}}$	0.016	0.001

Comparison

Reference Articles	Data Release	Training Set	$\overline{\Delta z_{norm}}$	ML Algorithm
Beck et al (2016) MNRAS	**12**	1'976'978	5.84×10^{-5}	**Local Linear Regression**
Paul et al (2018) Galaxies-MDPI	**12**	20'000	2×10^{-3}	**Random-Forest**
Baldeschi et al (2021) APJ	**16**	1'251'249	1×10^{-3}	**Random-Forest**
A. Momtaz, M. H. Salimi, S. Shakeri (2021) In preparation	**16**	400'000	1×10^{-3}	**Random-Forest**

References

Andrew E. Firth, Ofer Lahav and Rachel S. Somerville

Estimating photometric redshifts with artificial neural networks

Mon. Not. R. Astron. Soc. 339, 1195–1202, 2003

Beck et al.

Photometric redshifts for the SDSS Data Release 12

MNRAS 460, 1371–1381, 2016

Baldeschi et al.

Photometric redshift estimation of galaxies in the Pan-STARRS

Draft version, submitted to APJ, 2021

Nicholas Paul, Nicholas Virag and Lior Shamir

A catalog of photometric redshift and the distribution of broad galaxy morphologies

Galaxies, MDPI, 2018

Thanks For Your Attention :)