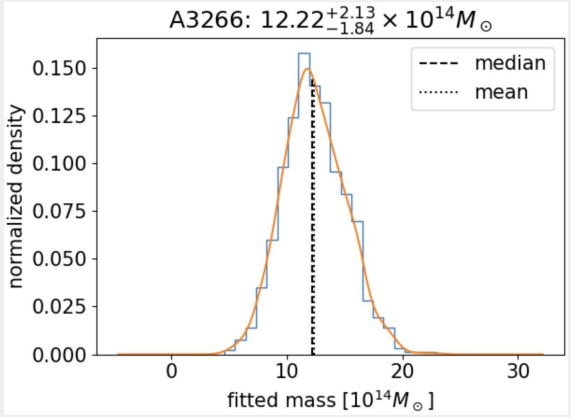
BROWN PHYSICS

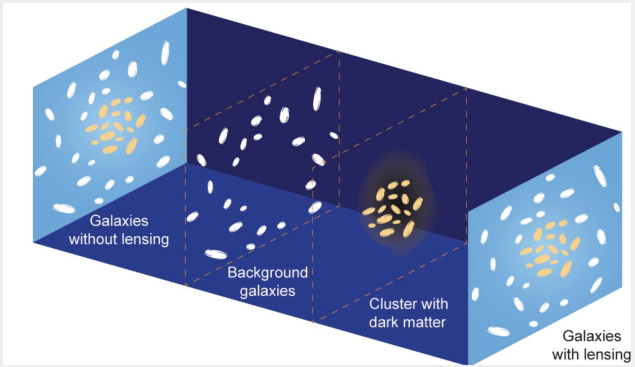
Binary Classification in a Galaxy Cluster Field Using ML

Zacharias Escalante
Brown University Department of Physics
October 23, 2024
GitHub

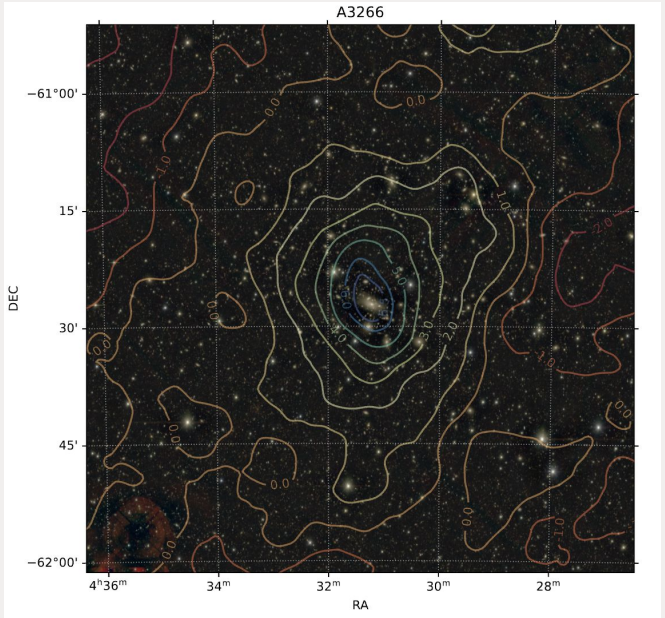
Introduction

- Astrolensing group
- Data products
- Catalogs of object magnitudes + shapes
- “Extendedness” → Classification!!
- Identify stars and galaxies in our fields to better calibrate weak lensing measurements





Credit: NASA/SSU/Aurore Simonnet

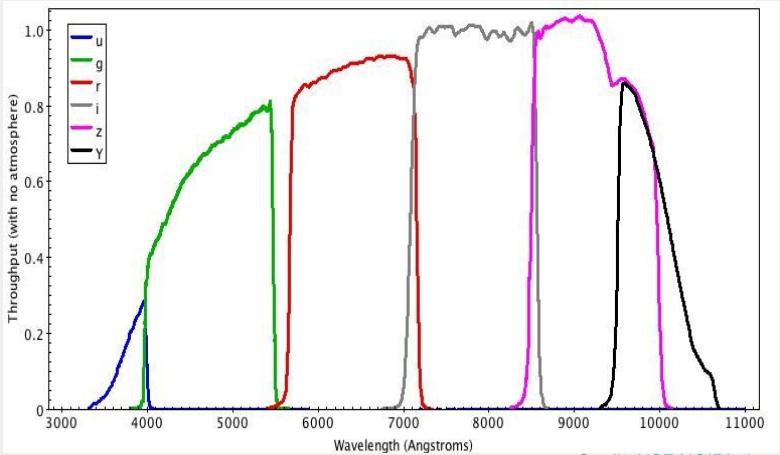


Full FOV
is 4 GB!

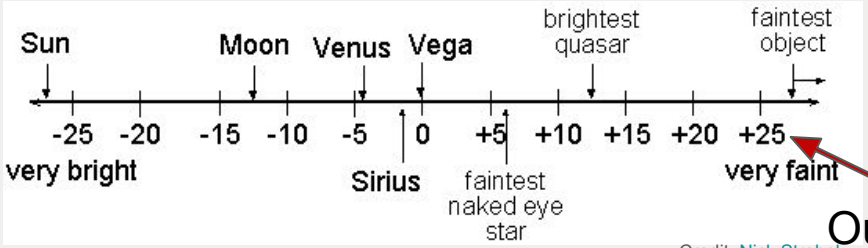
Abell 3266
combined irg coadd

Introduction

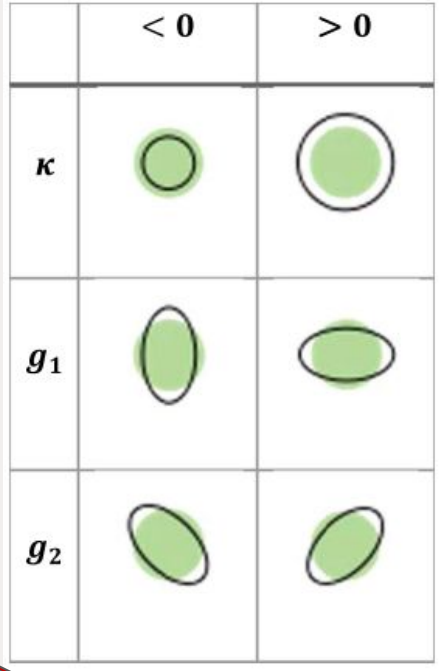
- Ellipticities
  - e1, e2 → [-1,1]
- Magnitudes
- ugriz photometric bands



Credit: NSF NOIRLab



Credit: Nick Strobel



Credit: Tung and Wright (2017)

Our source magnitudes

EDA

- Object catalog:

  ~1.8 Million rows, 41 features

- Majority of features:

  Ellipticities + magnitudes

- res, rkron, blendedness, psf_used

Target Variable

	ra	dec	x	y	e1	e2	res	sigmae	\
0	68.914455	-62.906044	17394.772375	3890.481991	NaN	NaN	NaN	NaN	
1	69.009523	-62.904732	16801.596400	3898.262765	NaN	NaN	NaN	NaN	
2	68.977504	-62.904551	17001.230648	3904.279632	NaN	NaN	NaN	NaN	

	rkron	extendedness	...	r_cmodel_mag	r_cmodel_magerr	i_psf_mag	\
0	2.177848	1.0	...	21.828693	0.031725	21.740775	
1	6.150654	1.0	...	19.786428	0.011056	27.238589	
2	3.431714	1.0	...	21.139035	0.020066	25.215983	

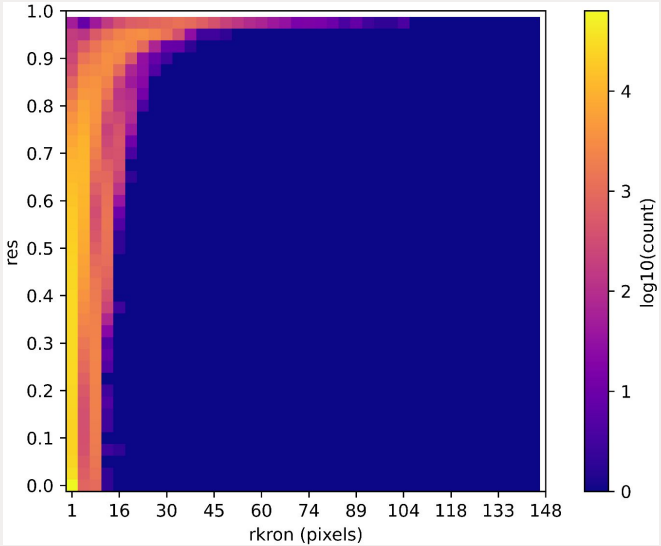
	i_psf_magerr	i_cmodel_mag	i_cmodel_magerr	z_psf_mag	z_psf_magerr	\
0	0.033023	21.712547	0.033202	21.778364	0.078916	
1	1.082054	NaN	NaN	25.915949	0.343267	
2	1.075873	NaN	NaN	NaN	-3.567432	

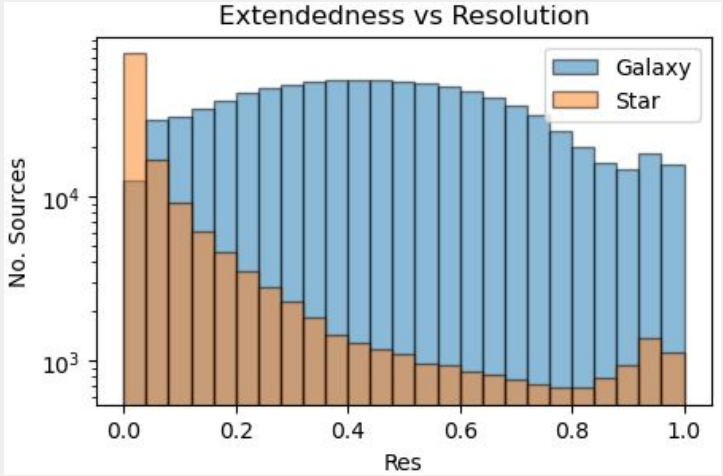
	z_cmodel_mag	z_cmodel_magerr
0	21.738405	0.078901
1	NaN	NaN
2	NaN	NaN

```
ra
dec
x
y
e1
e2
res
sigmae
rkron
extendedness
blendedness
psf_used
e1_sdss
e2_sdss
e1_psf_sdss
e2_psf_sdss
e1_hsm
e2_hsm
e1_psf_hsm
e2_psf_hsm
i_e1_psf_sdss
i_e2_psf_sdss
u_psf_mag
u_psf_magerr
u_cmodel_mag
u_cmodel_magerr
g_psf_mag
g_psf_magerr
g_cmodel_mag
g_cmodel_magerr
r_psf_mag
r_psf_magerr
r_cmodel_mag
r_cmodel_magerr
i_psf_mag
i_psf_magerr
i_cmodel_mag
i_cmodel_magerr
z_psf_mag
z_psf_magerr
z_cmodel_mag
z_cmodel_magerr
```

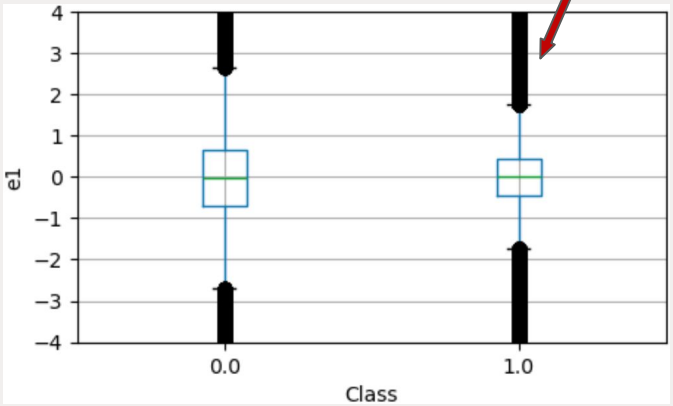
EDA

Interesting figures





All outliers!

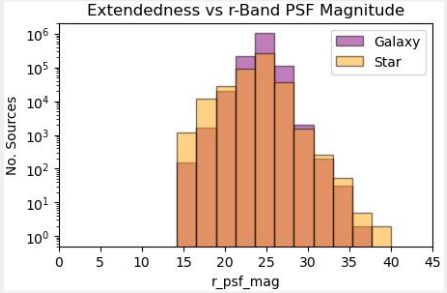


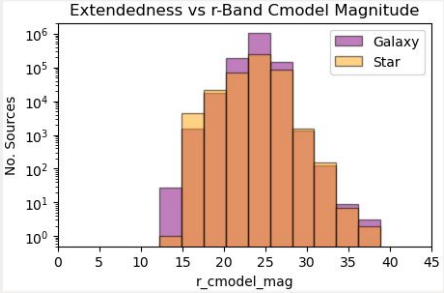
EDA

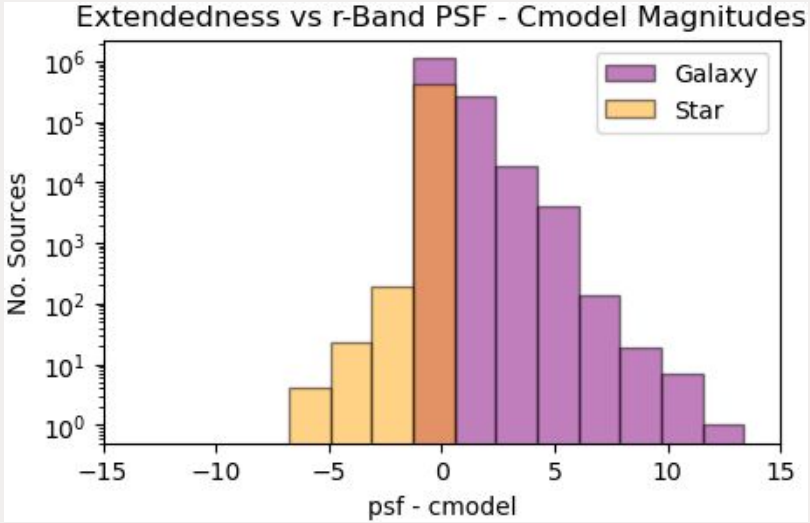
- Engineer a new feature:

  PSF - cmodel magnitudes

- Separation in r-band

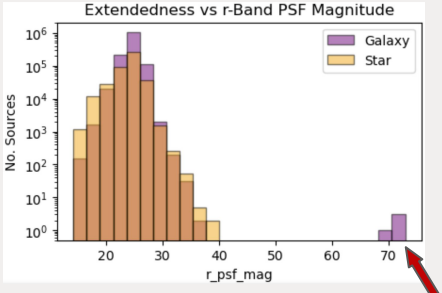


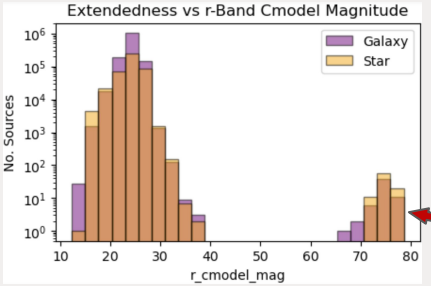


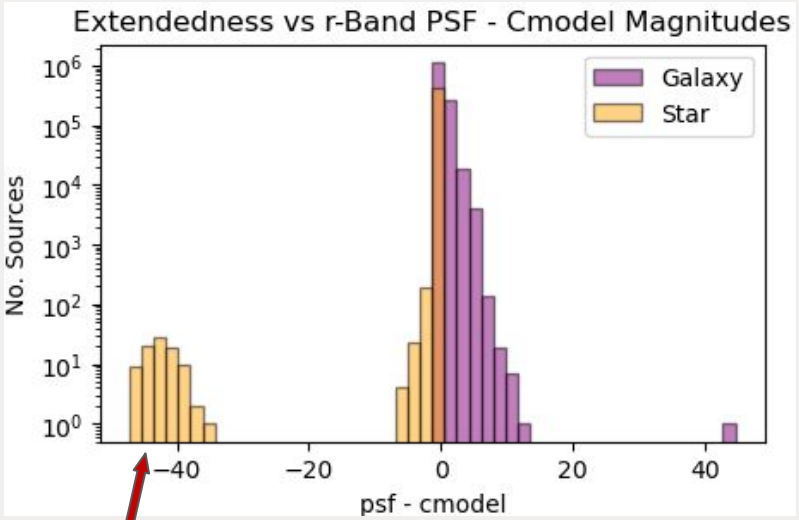


EDA

- Full x-range
- Outliers have unrealistic magnitudes





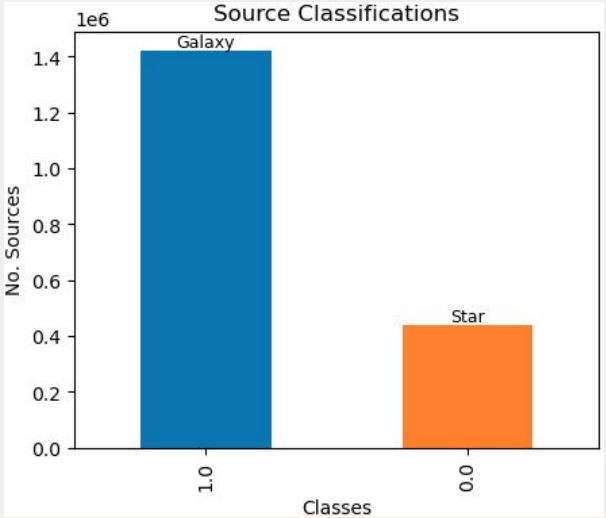


Not realistic. Error in flux measurements

Splitting & Preprocessing

- ~72% with missing values
- Stratified 60/20/20
- One categorical feature OHE
- MinMax → Ellipticities, coordinates, res, blendedness
- Standard → sigmae, rkron, all magnitudes

```
(array([0., 1.]), array([262635, 851976]))
(array([0., 1.]), array([ 87545, 283992]))
(array([0., 1.]), array([ 87545, 283992]))
```




```
Fraction of missing values in features:
e1                0.440272
e2                0.440272
res               0.440272
sigmae            0.440272
rkron             0.024337
blendedness       0.072675
e1_sdss           0.012683
e2_sdss           0.012683
e1_psf_sdss       0.000003
e2_psf_sdss       0.000003
e1_hsm            0.072675
e2_hsm            0.072675
i_e1_psf_sdss     0.001066
i_e2_psf_sdss     0.001066
u_psf_mag         0.552664
u_psf_magerr      0.497406
u_cmodel_mag      0.553069
u_cmodel_magerr   0.500693
g_psf_mag         0.013114
g_psf_magerr      0.001309
g_cmodel_mag      0.013168
g_cmodel_magerr   0.002261
r_psf_mag         0.001882
r_cmodel_mag      0.001966
r_cmodel_magerr   0.000039
i_psf_mag         0.010033
i_psf_magerr      0.001159
i_cmodel_mag      0.011375
i_cmodel_magerr   0.003586
z_psf_mag         0.025147
z_psf_magerr      0.002125
z_cmodel_mag      0.028290
z_cmodel_magerr   0.007605
u_mag_diff        0.563906
g_mag_diff        0.015623
r_mag_diff        0.002362
i_mag_diff        0.013289
z_mag_diff        0.032423
dtype: float64
Fraction of points with missing values: 0.719
```


Questions?