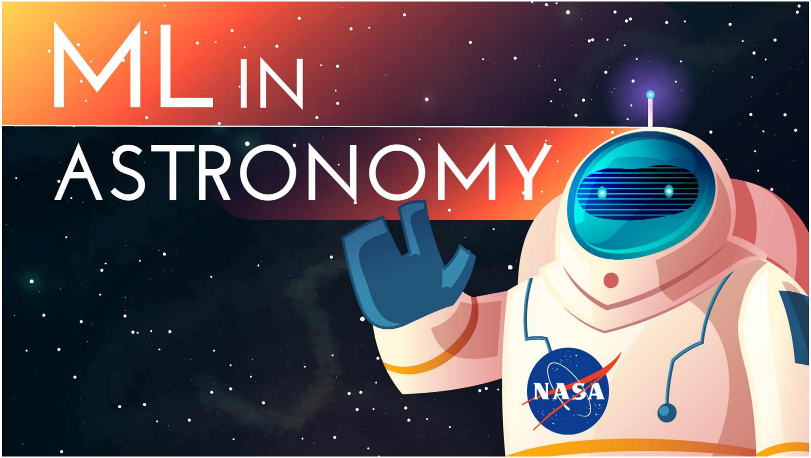
Machine Learning in Astronomy

Reza Monadi

UC Riverside

May 14, 2020



credit: 365datascience.com

- How astronomy is tied to **BIG DATA**?

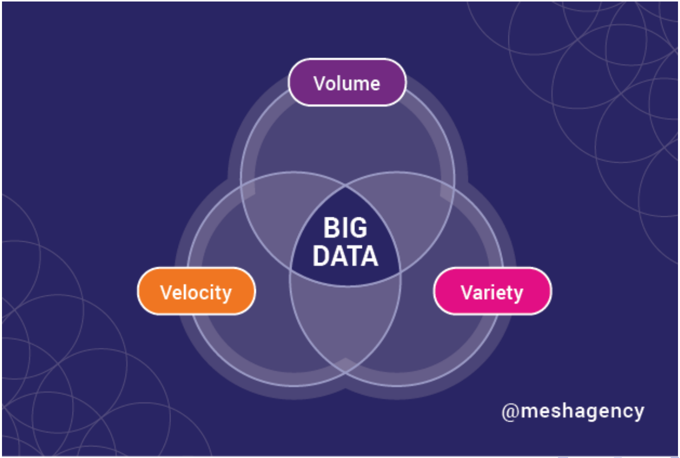
- How astronomy is tied to **BIG DATA**?
- What is **ML**?

- How astronomy is tied to **BIG DATA**?
- What is **ML**?
- How to implement **ML** in astronomy?

- How astronomy is tied to **BIG DATA**?
- What is **ML**?
- How to implement **ML** in astronomy?
- How **ML** helps **SKA**?

- How astronomy is tied to **BIG DATA**?
- What is **ML**?
- How to implement **ML** in astronomy?
- How **ML** helps **SKA**?
- What are the pitfalls of **ML** in astronomy?

What is BIG DATA?



VVV in astronomy

- Volume: larger quantities of data by better facilitates

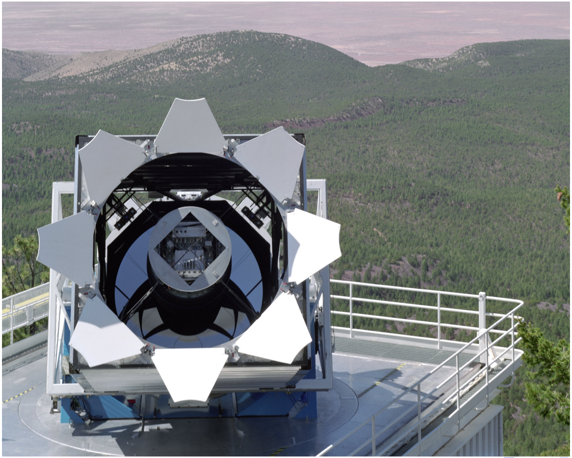
VVV in astronomy

- Volume: larger quantities of data by better facilitates
- Velocity: Higher speed of getting data

VVV in astronomy

- Volume: larger quantities of data by better facilitates
- Velocity: Higher speed of getting data
- Verity: More complex structures of data

Sloan Digital Sky Server $\Rightarrow$ 40 TB



Sloan Digital Sky Server $\Rightarrow$ 40 TB

Imaging statistics

Total unique area covered	14,555 square degrees
Total area of imaging (including overlaps)	31,637 square degrees (excluding supernova runs)
Individual image field size	1361x2048 pixels (0.0337 square degrees)
Number of individual fields	938,046 (excluding supernova runs)
Number of catalog objects	1,231,051,050
Number of unique detections	932,891,133
Median PSF FWHM, *r*-band	1.3 arcsec
Pixel scale	0.396 arcsec
Exposure time per band	53.9 sec
Time difference between observations of each band	71.72 sec (in *riuzg* order)
Global astrometric precision	0.1 arcsec rms (absolute)

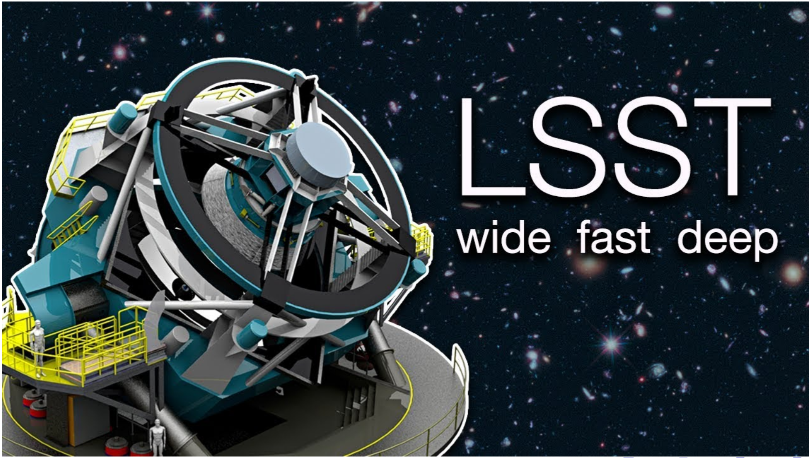
Sloan Digital Sky Server $\Rightarrow$ 40 TB

Optical spectroscopy data statistics

All programs combined

Total spectra	5,789,200
Useful spectra	4,846,156
Galaxies	2,863,635
Quasars	960,678
Stars	1,021,843
Sky	475,531
Standards	108,603
Unknown	352,320

Large Synaptic Survey Telescope $\Rightarrow$ 200 PB



Large Synaptic Survey Telescope $\Rightarrow$ 200 PB

- 800+ panoramic images each night

Large Synaptic Survey Telescope $\Rightarrow$ 200 PB

- 800+ panoramic images each night
- 3.2 billion-pixel camera

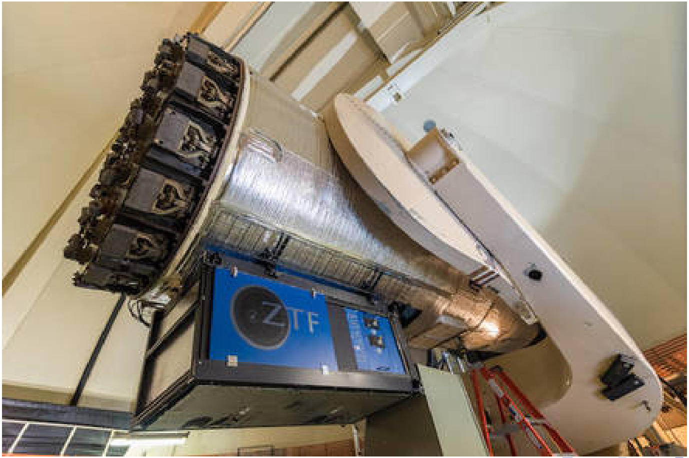
Large Synaptic Survey Telescope $\Rightarrow$ 200 PB

- 800+ panoramic images each night
- 3.2 billion-pixel camera
- Recording the entire visible sky twice each week

Large Synaptic Survey Telescope $\Rightarrow$ 200 PB

- 800+ panoramic images each night
- 3.2 billion-pixel camera
- Recording the entire visible sky twice each week
- Each patch of sky will be visited 1000 times

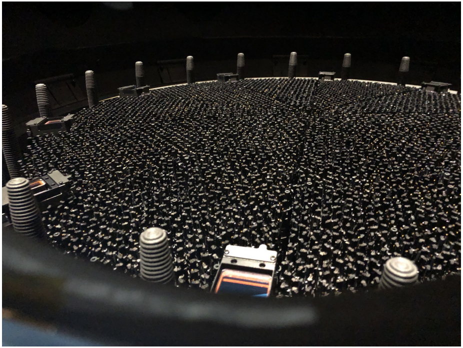
Zwicky Transient Facility

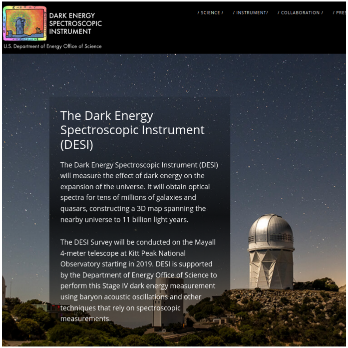


Zwicky Transient Facility

Filter(s)	#PSFcat-*sci* sources	#Aperturecat-*sci* sources	#PSFcat-*ref* sources	#Aperturecat-*ref* sources
g	34,799,157,939	22,008,868,967	2,053,507,697	656,772,487
r	64,698,158,346	40,249,465,415	2,954,247,127	1,013,085,916
i	868,042,655	495,178,759	557,831,277	163,240,538
$g+r+i$	100,365,358,940	62,753,513,141	5,565,586,101	1,833,098,941

Dark Energy Spectroscopic Instrument





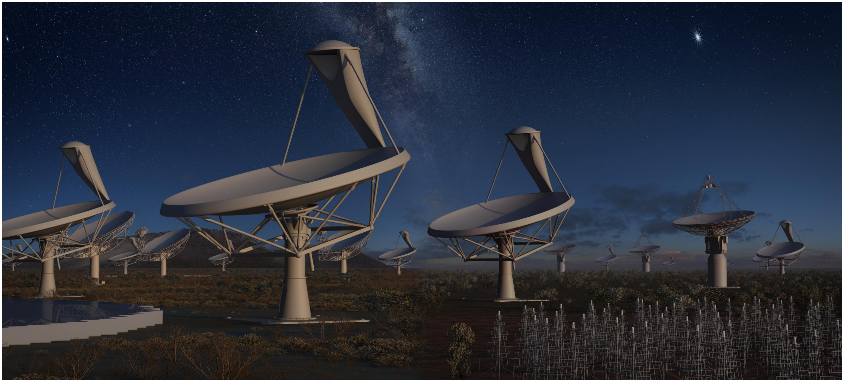
Dark Energy Spectroscopic Instrument

- Spectra of 25 M galxies, quasars, stars.

Dark Energy Spectroscopic Instrument

- Spectra of 25 M galxies, quasars, stars.
- 5000 spectra per exposure

Square Kilometer Array $\Rightarrow$ 4.6 EB



Large surveys

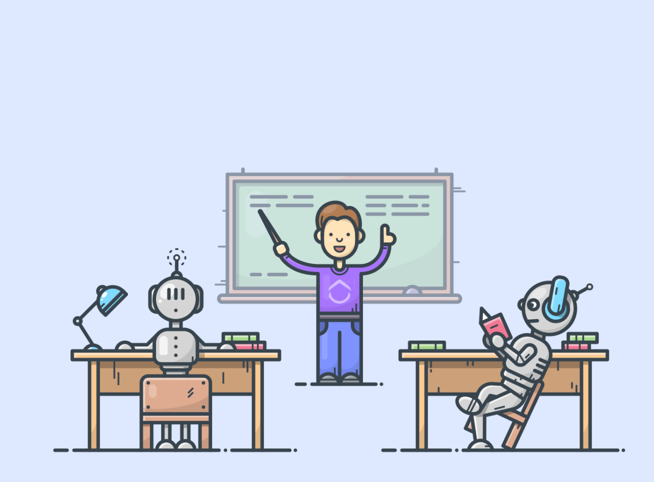
Large surveys $\Rightarrow$

Large surveys $\Rightarrow$ Big Data

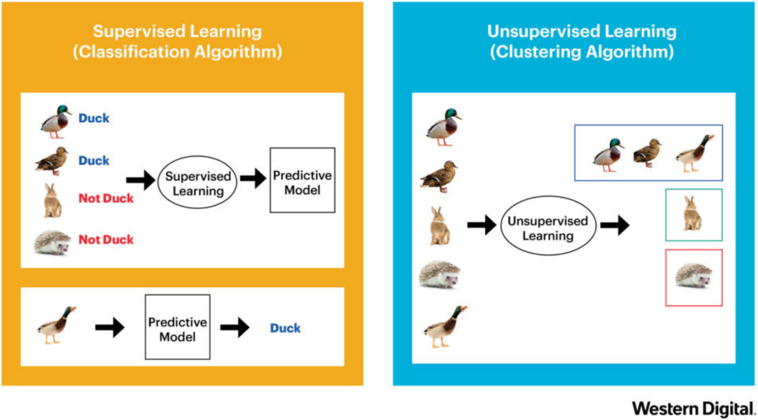
Large surveys $\Rightarrow$ Big Data $\stackrel{ML}{\Longrightarrow}$

Large surveys $\Rightarrow$ Big Data $\xRightarrow{ML}$ Astronomy Knowledge

Supervised ML vs. Unsupervised ML



Supervised ML vs. Unsupervised ML



Stages of Supervised Learning

- Training:

Stages of Supervised Learning

- Training:
    1. Select a model

Stages of Supervised Learning

- Training:
  1. Select a model
  2. Set up hyper-parameters of model

Stages of Supervised Learning

- Training:
  1. Select a model
  2. Set up hyper-parameters of model
  3. Teach the machine by training set

Stages of Supervised Learning

- Training:
  1. Select a model
  2. Set up hyper-parameters of model
  3. Teach the machine by training set
- Validation:

Stages of Supervised Learning

- Training:
  1. Select a model
  2. Set up hyper-parameters of model
  3. Teach the machine by training set
- Validation:
  1. Change the hyper-parameters

Stages of Supervised Learning

- Training:
  1. Select a model
  2. Set up hyper-parameters of model
  3. Teach the machine by training set
- Validation:
  1. Change the hyper-parameters
  2. Select the optimum hyper-parameters

Stages of Supervised Learning

- Training:
    1. Select a model
    2. Set up hyper-parameters of model
    3. Teach the machine by training set
- Validation:
    1. Change the hyper-parameters
    2. Select the optimum hyper-parameters
- Testing:

Stages of Supervised Learning

- Training:
  1. Select a model
  2. Set up hyper-parameters of model
  3. Teach the machine by training set
- Validation:
  1. Change the hyper-parameters
  2. Select the optimum hyper-parameters
- Testing:
  1. Test learned model by an unseen part of the data-set.

Stages of Supervised Learning

- Training:
  1. Select a model
  2. Set up hyper-parameters of model
  3. Teach the machine by training set
- Validation:
  1. Change the hyper-parameters
  2. Select the optimum hyper-parameters
- Testing:
  1. Test learned model by an unseen part of the data-set.
  2. Select the best model and use it for predictions.

Supervised Learning vs Traditional Model Fitting ?

Similarities

Supervised Learning vs Traditional Model Fitting ?

Similarities

- Both need a set of labeled measurements and a model

Supervised Learning vs Traditional Model Fitting ?

Similarities

- Both need a set of labeled measurements and a model

Differences

Supervised Learning vs Traditional Model Fitting ?

Similarities

- Both need a set of labeled measurements and a model

Differences

- SML: model gets adapted by data

Supervised Learning vs Traditional Model Fitting ?

Similarities

- Both need a set of labeled measurements and a model

Differences

- SML: model gets adapted by data
- SML: model can be very nonlinear/complex

Supervised Learning vs Traditional Model Fitting ?

Similarities

- Both need a set of labeled measurements and a model

Differences

- SML: model gets adapted by data
- SML: model can be very nonlinear/complex
- TMF: model is predefined and has limited adaptivity

Supervised Learning vs Traditional Model Fitting ?

Similarities

- Both need a set of labeled measurements and a model

Differences

- SML: model gets adapted by data
- SML: model can be very nonlinear/complex
- TMF: model is predefined and has limited adaptivity
- SML: designed for predicting unseen data

Supervised Learning vs Traditional Model Fitting ?

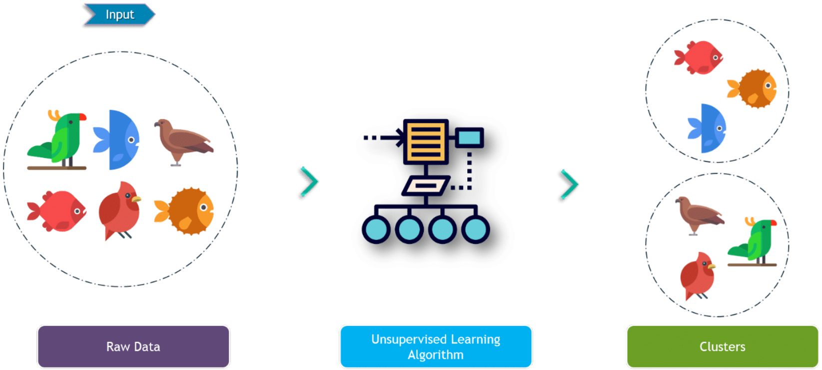
Similarities

- Both need a set of labeled measurements and a model

Differences

- SML: model gets adapted by data
- SML: model can be very nonlinear/complex
- TMF: model is predefined and has limited adaptivity
- SML: designed for predicting unseen data
- TMF: infers relationships between features

How unsupervised learning works?



Unsupervised Learning and knowledge discovery

- Finding unknown patterns in the data set

Unsupervised Learning and knowledge discovery

- Finding unknown patterns in the data set
- Distinguishing similar and dissimilar objects

Unsupervised Learning and knowledge discovery

- Finding unknown patterns in the data set
- Distinguishing similar and dissimilar objects
- Helping for visualization in lower dimensions

Unsupervised Learning and knowledge discovery

- Finding unknown patterns in the data set
- Distinguishing similar and dissimilar objects
- Helping for visualization in lower dimensions
- Recognizing odd objects (unknown unknowns)

Unsupervised Learning and knowledge discovery

- Finding unknown patterns in the data set
- Distinguishing similar and dissimilar objects
- Helping for visualization in lower dimensions
- Recognizing odd objects (unknown unknowns)

$$\text{DATA} \xrightarrow{ML} \text{Knowledge}$$

text

text