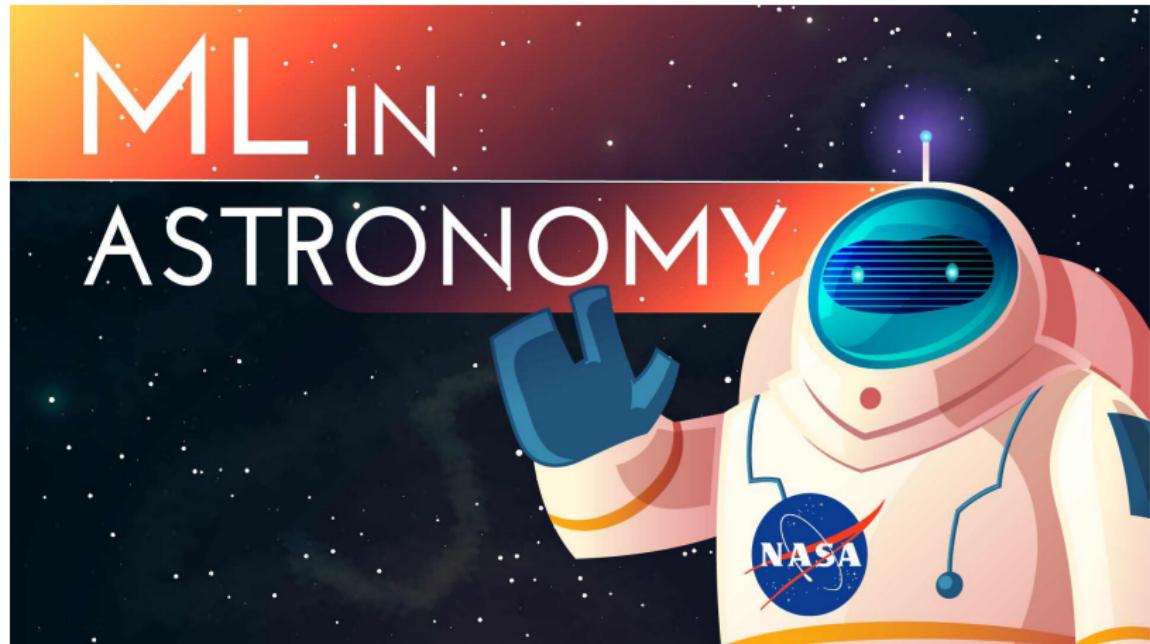


# Machine Learning in Astronomy

Reza Monadi

UC Riverside

May 14, 2020



credit: 365datascience.com

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

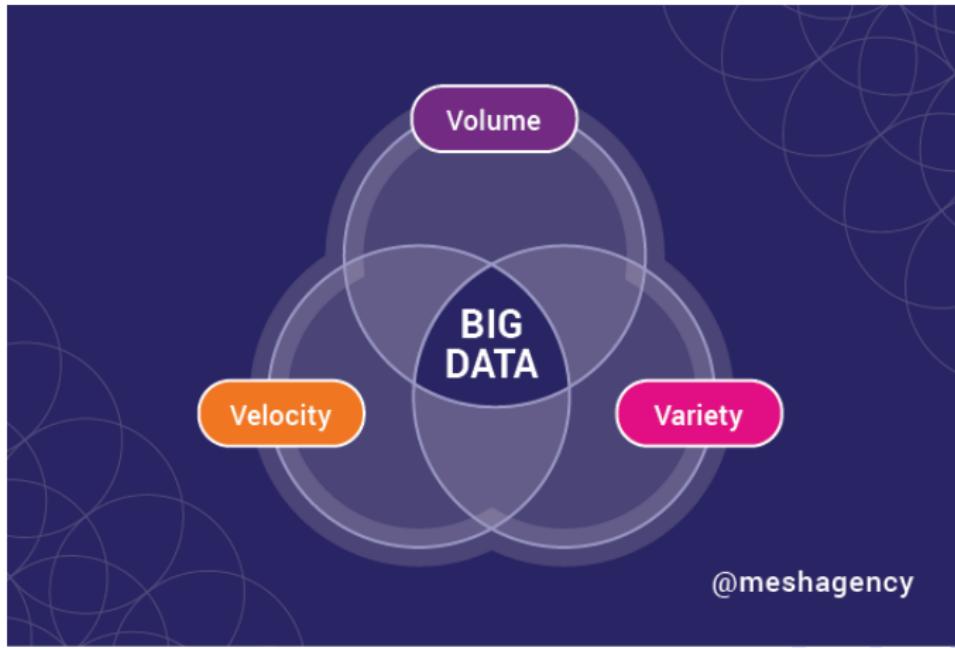
- How astronomy is tied to **BIG DATA?**
- What is **ML?**

- 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

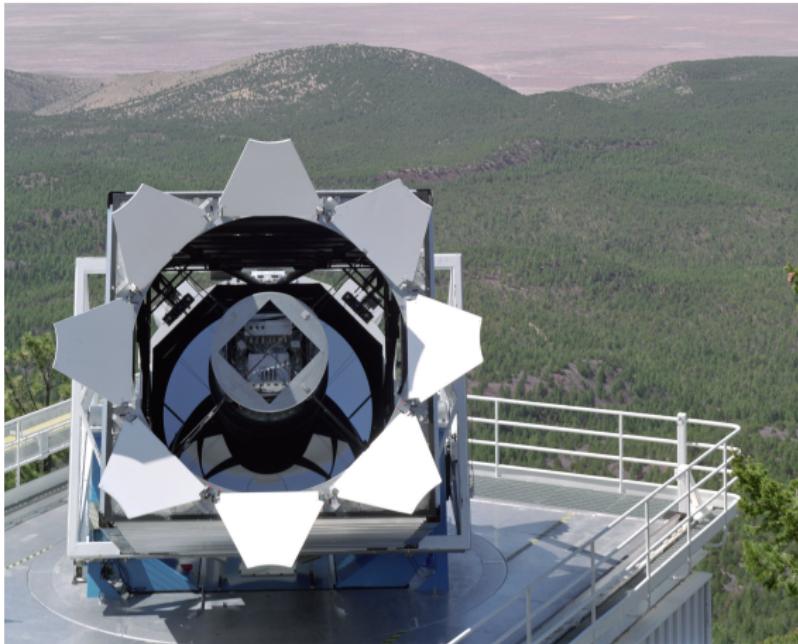
# VVV in astronomy

- 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

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

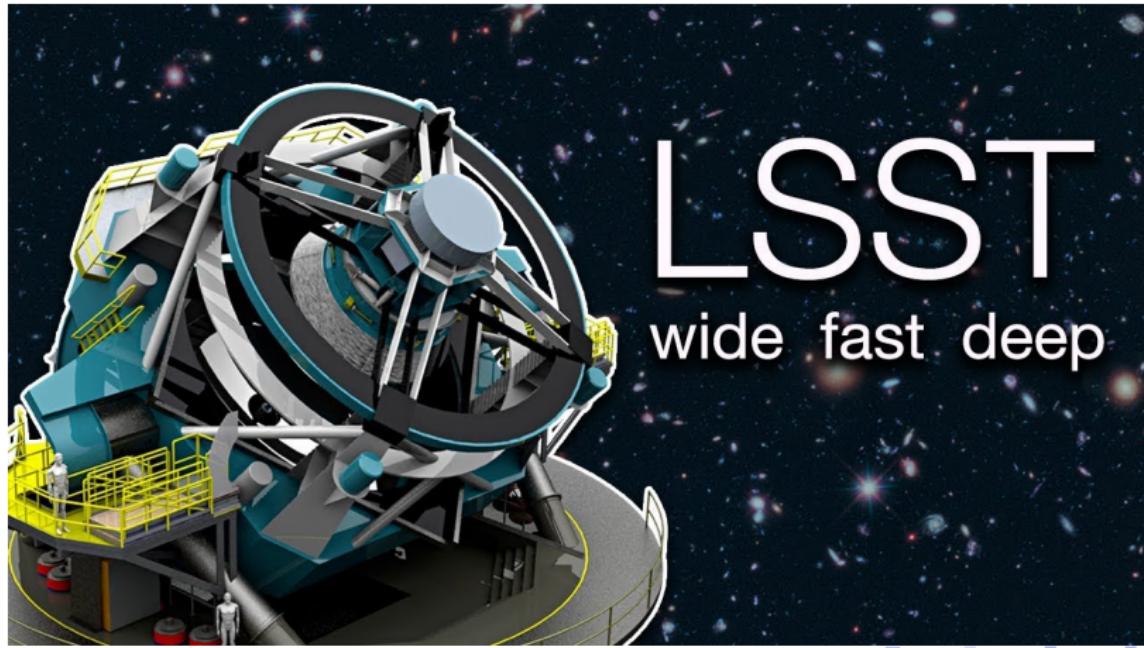
# Sloan Digital Sky Server $\Rightarrow$ 40 TB

## Optical spectroscopy data statistics

### All programs combined

<b>Total spectra</b>	5,789,200
<b>Useful spectra</b>	4,846,156
<b>Galaxies</b>	2,863,635
<b>Quasars</b>	960,678
<b>Stars</b>	1,021,843
<b>Sky</b>	475,531
<b>Standards</b>	108,603
<b>Unknown</b>	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 Synoptic Survey Telescope $\Rightarrow$ 200 PB

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

# Large Synoptic 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

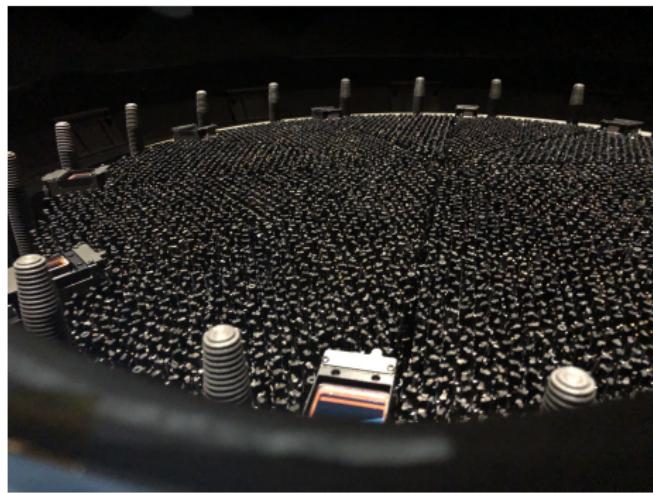
# Zwicky Transient Facility



# Zwicky Transient Facility

Filter(s)	#PSFcat- <i>sci</i> sources	#Aperturecat- <i>sci</i> sources	#PSFcat- <i>ref</i> sources	#Aperturecat- <i>ref</i> sources
<i>g</i>	34,799,157,939	22,008,868,967	2,053,507,697	656,772,487
<i>r</i>	64,698,158,346	40,249,465,415	2,954,247,127	1,013,085,916
<i>i</i>	868,042,655	495,178,759	557,831,277	163,240,538
<i>g + r + i</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

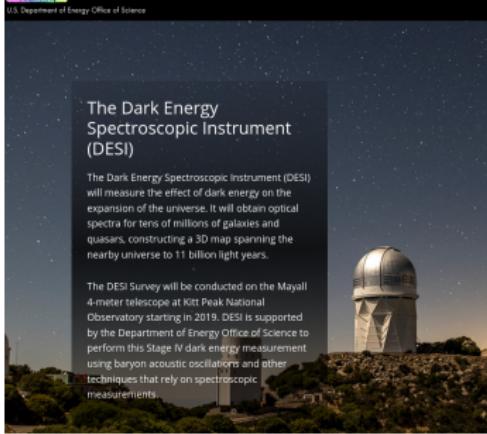
U.S. Department of Energy Office of Science

/ SCIENCE / / INSTRUMENT / / COLLABORATION / / FUNDING

The Dark Energy Spectroscopic Instrument (DESI)

The Dark Energy Spectroscopic Instrument (DESI) will measure the effect of dark energy on the expansion of the universe. It will obtain optical spectra for tens of millions of galaxies and quasars, constructing a 3D map spanning the nearby universe to 11 billion light years.

The DESI Survey will be conducted on the Mayall 4-meter telescope at Kitt Peak National Observatory starting in 2019. DESI is supported by the Department of Energy Office of Science to perform this Stage IV dark energy measurement using baryon acoustic oscillations and other techniques that rely on spectroscopic measurements.



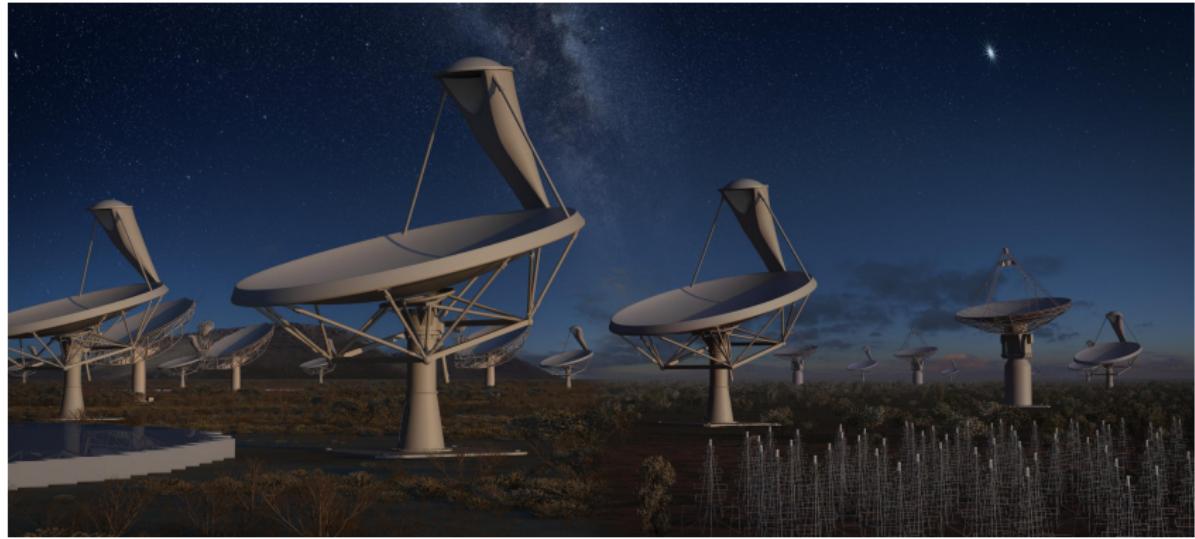
# Dark Energy Spectroscopic Instrument

- Spectra of 25 M galaxies, quasars, stars.

# Dark Energy Spectroscopic Instrument

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

# Square Kilometer Array $\Rightarrow$ 4.6 EB



# Large surveys

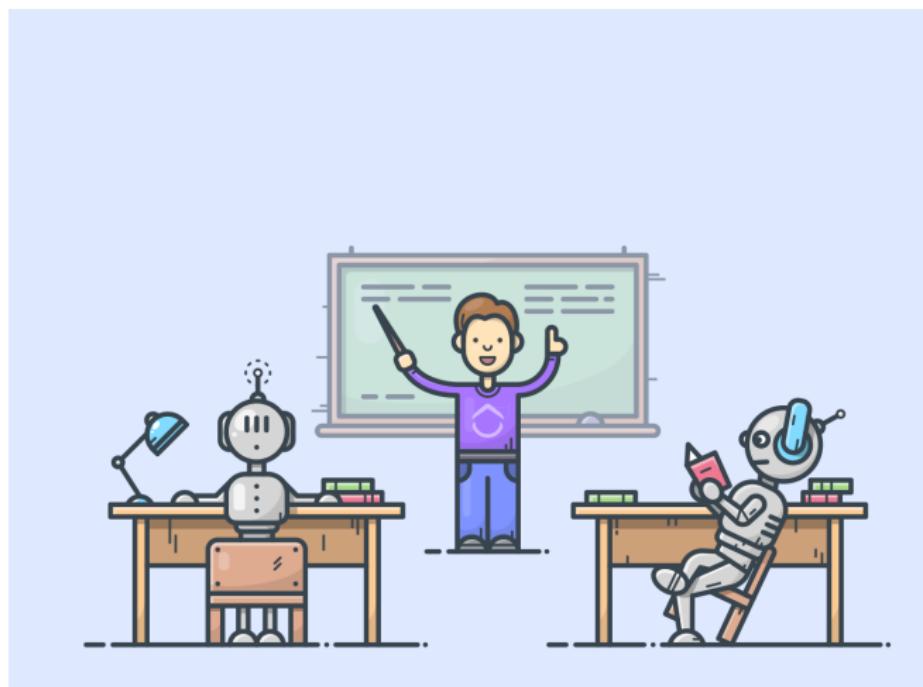
Large surveys ⇒

# Large surveys ⇒ Big Data

Large surveys  $\Rightarrow$  Big Data  $\xrightarrow{ML}$

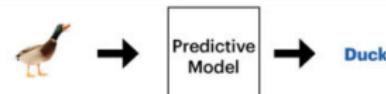
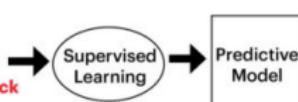
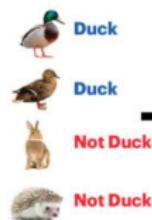
Large surveys  $\Rightarrow$  Big Data  $\xrightarrow{ML}$   
Astronomy Knowledge

# Supervised ML vs. Unsupervised ML

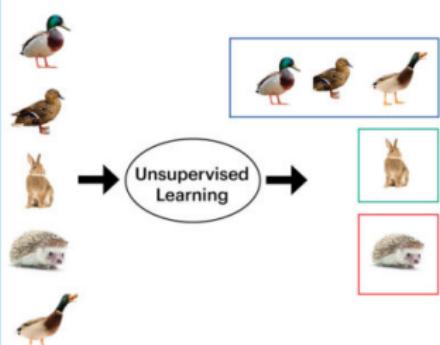


# Supervised ML vs. Unsupervised ML

## Supervised Learning (Classification Algorithm)



## Unsupervised Learning (Clustering Algorithm)



# 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

I

# Supervised Learning vs Traditional Model Fitting ?



## Similarities

- I • Both need a set of labeled measurements and a model

# Supervised Learning vs Traditional Model Fitting ?



## Similarities

- I • Both need a set of labeled measurements and a model



## Differences

# Supervised Learning vs Traditional Model Fitting ?



## Similarities

- I • Both need a set of labeled measurements and a model



## Differences

- I • SML: model gets adapted by data

# Supervised Learning vs Traditional Model Fitting ?



## Similarities

- I • 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

- I • 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

- I • 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

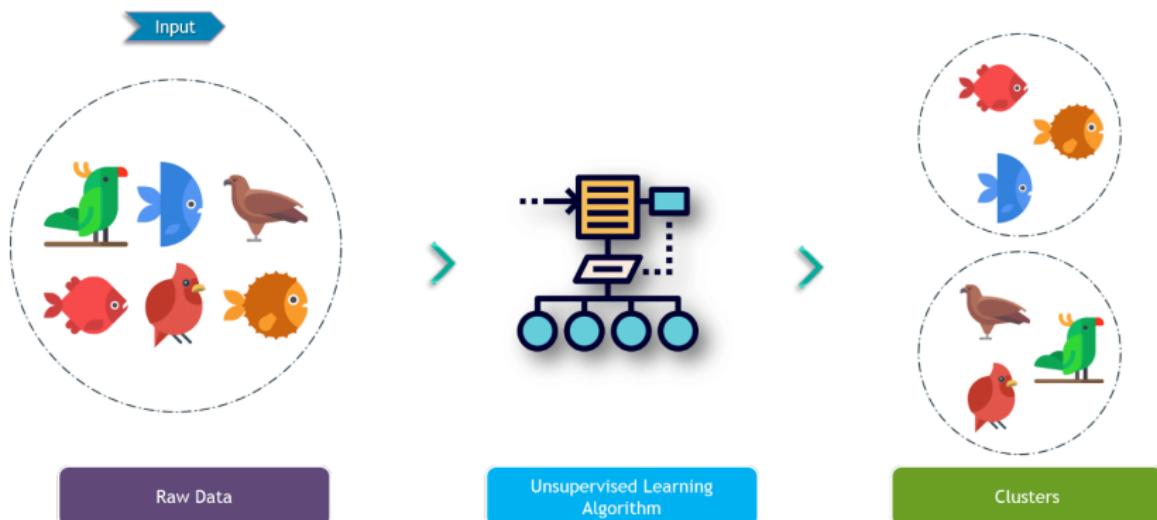
- I • Both need a set of labeled measurements and a model



## 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)

DATA  $\xrightarrow{ML}$  Knowledge

Outline  
Big Data  
ML Algorithms  
**Supervised ML in astronomy**  
Unsupervised ML in astronomy  
ML and SKA  
ML limitations

Classification  
Regression  
Neural Network

Outline  
Big Data  
ML Algorithms  
**Supervised ML in astronomy**  
Unsupervised ML in astronomy  
ML and SKA  
ML limitations

Classification  
Regression  
Neural Network

Outline  
Big Data  
ML Algorithms  
**Supervised ML in astronomy**  
Unsupervised ML in astronomy  
ML and SKA  
ML limitations

Classification  
Regression  
Neural Network

Outline  
Big Data  
ML Algorithms  
**Supervised ML in astronomy**  
**Unsupervised ML in astronomy**  
ML and SKA  
ML limitations

**Clustering**  
outlier detection  
dimensionality reduction  
Neural network

Outline  
Big Data  
ML Algorithms  
**Supervised ML in astronomy**  
**Unsupervised ML in astronomy**  
ML and SKA  
ML limitations

Clustering  
**outlier detection**  
dimensionality reduction  
Neural network

Outline  
Big Data  
ML Algorithms  
Supervised ML in astronomy  
Unsupervised ML in astronomy  
ML and SKA  
ML limitations

Clustering  
outlier detection  
**dimensionality reduction**  
Neural network

Outline  
Big Data  
ML Algorithms  
Supervised ML in astronomy  
Unsupervised ML in astronomy  
ML and SKA  
ML limitations

Clustering  
outlier detection  
dimensionality reduction  
Neural network

text

Outline  
Big Data  
ML Algorithms  
Supervised ML in astronomy  
Unsupervised ML in astronomy  
**ML and SKA**  
ML limitations

text

Outline  
Big Data  
ML Algorithms  
Supervised ML in astronomy  
Unsupervised ML in astronomy  
ML and SKA  
ML limitations

a  
b

Outline  
Big Data  
ML Algorithms  
Supervised ML in astronomy  
Unsupervised ML in astronomy  
ML and SKA  
ML limitations

a  
b