Human in the Loop:
Adaptive learning with astro examples

9th BCD ISHEP Cargèse School - part II
28 March 2024 - Cargèse, France

Emille E. O. Ishida

Laboratoire de Physique de Clermont, CNRS - Université Clermont-Auvergne
Clermont Ferrand, France

Acknowledgements

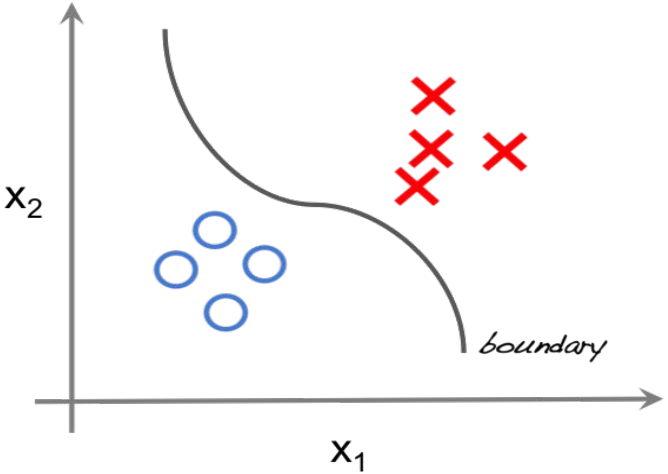
Alberto Krone-Martins (UCI)
Rafael S. de Souza (Hertfordshire)
and the entire Cosmostatistics Initiative (COIN)

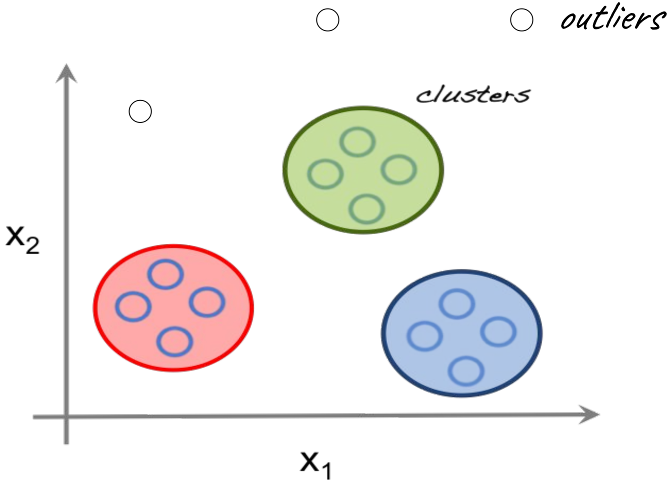
Konstantin Malanchev (CMU)
Maria Pruzhinskaya (LPC)
and everyone in the SNAD collaboration

Anais Moller (Swinburne)
Julien Peloton (IJCLab)
and everyone in the Fink broker

Alex I. Malz (CMU)
Mi Dai (JHU)
Kara Ponder
Amanda Wasserman (UIUC)
and all those working in the RESSPECT team

Supervised x Unsupervised





Training sample:

features + labels

Target sample:

features

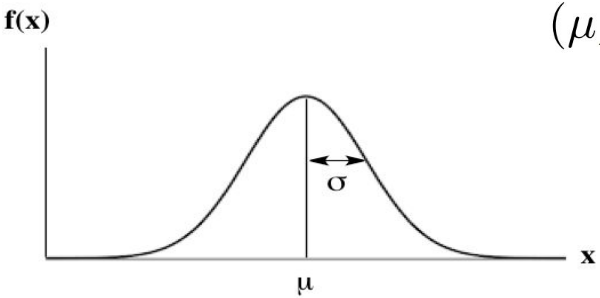
Data sample:

features

From yesterday …

Representativeness

Probability distribution, *P*



(μ_P, σ_P)

This is why it works!

Training

$(\mu_{S_1}, \sigma_{S_1})$

Target

$(\mu_{S_2}, \sigma_{S_2})$

Ideal Supervised learning situation

Learn

Training sample

Images, colors, light curves, etc.

\+

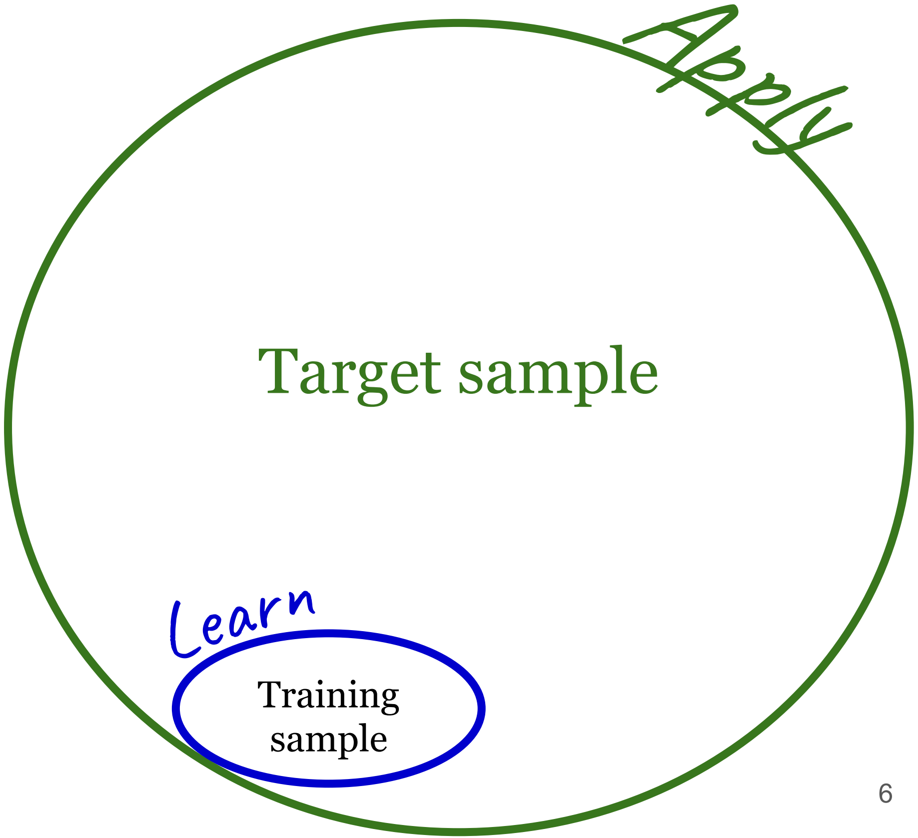
Classes

Apply

Target sample

Images, colors, light curves, etc.

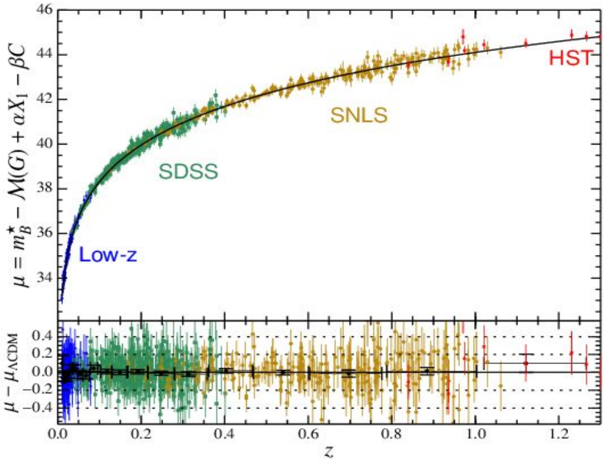
Most-likely practical science situation



Type Ia supernova cosmology

Standard candles used to measure cosmological distances

https://supernova.eso.org/exhibition/images/1111_E_549779main_pia14095_full/

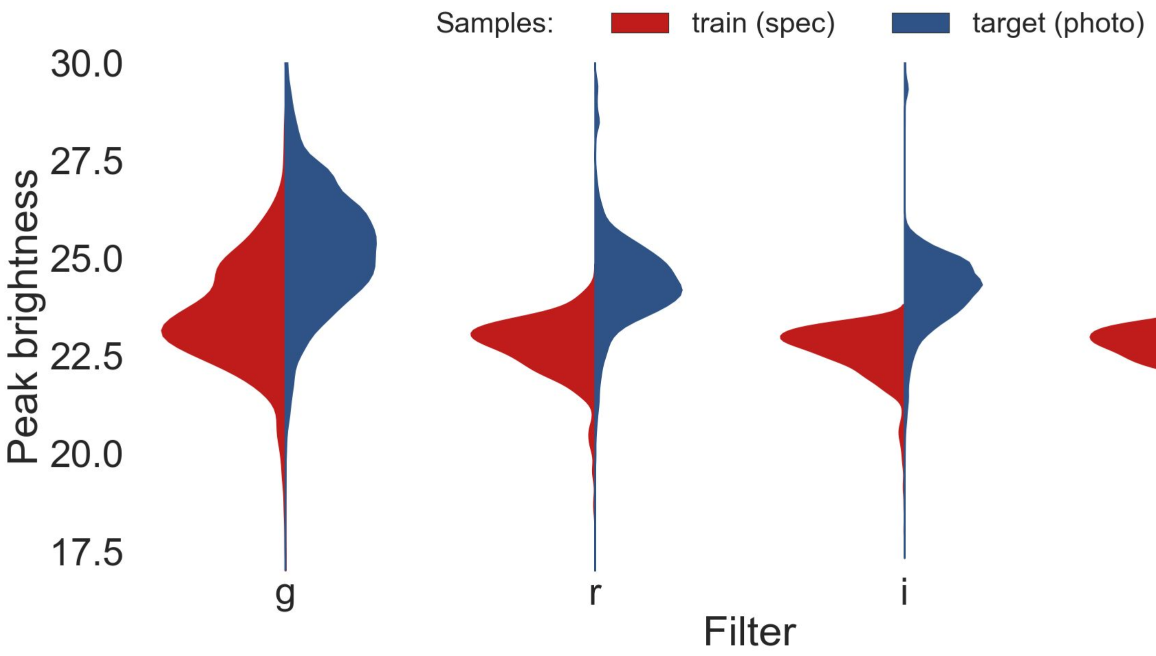


http://supernovae.in2p3.fr/sdss_snls_jla/ReadMe.html

Ω

Λ

Real astro-learning situation



From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

Very common situation

Labels are often far too expensive!

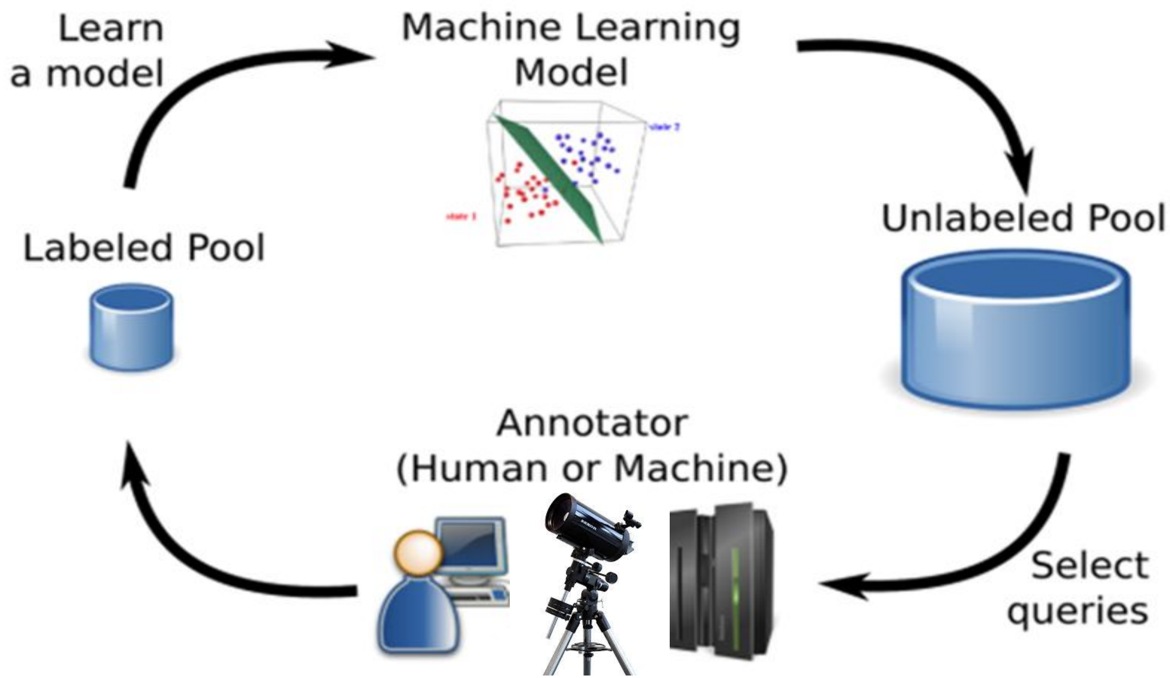
Given limited resources, we need recommendation systems!

35% of Amazon's revenue are generated by it's recommendation engine.

75% of users select movies based on Netflix's recommendations.

Active Learning

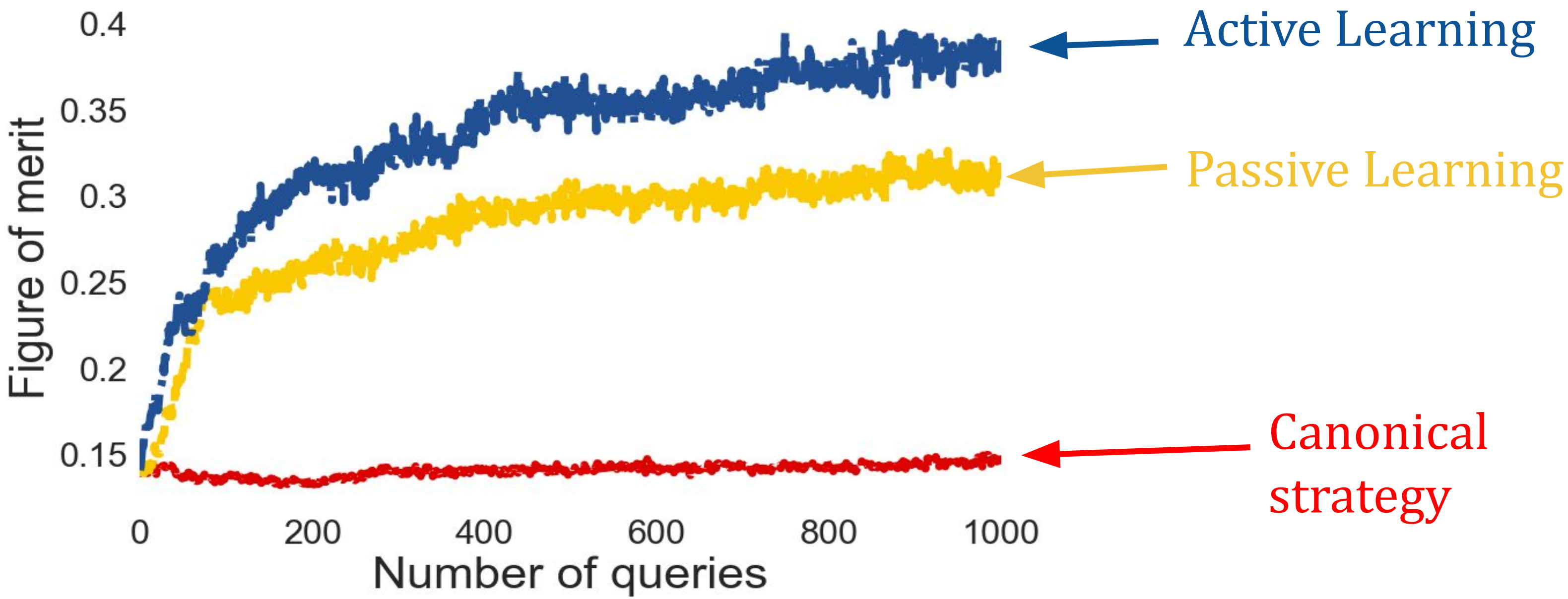
Optimal classification, minimum training



Plot modified from [Chowdhury et al., 2021, SPIE Medical Imaging]

AL for SN classification

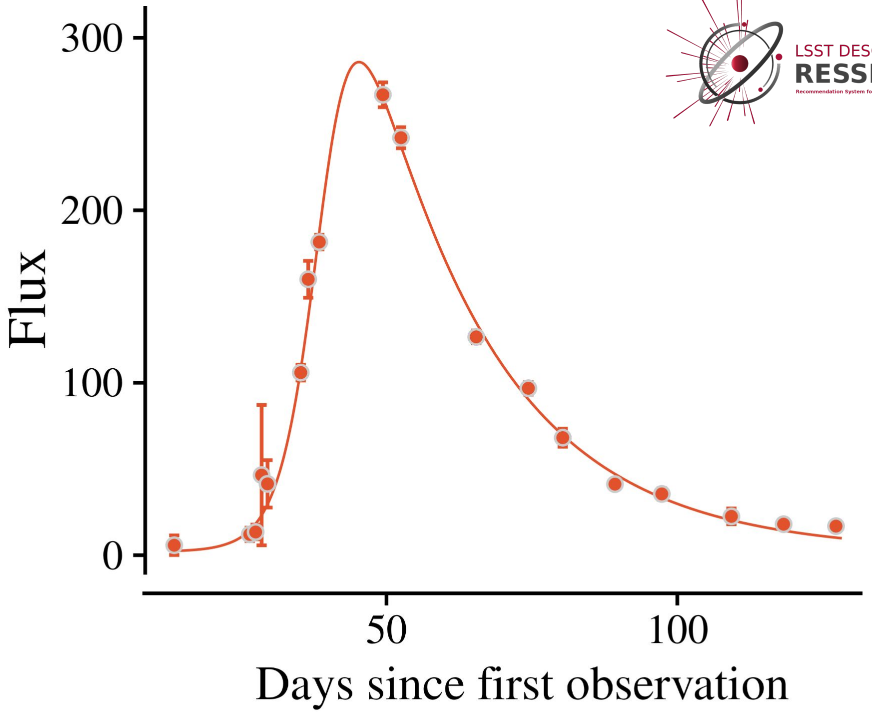
Static results



From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

If only it were that simple ...

- Window of Opportunity for Labelling
- Evolving Samples
  - We must make query decisions before we can observe the full LC
- Multiple Instruments for labelling
- Evolving budget
  - Other people want to use the telescope
- Evolving Costs
  - Observing costs for a given object changes as it evolves.



LSST DESC & COIN
RESSPECT

[Kennamer, Ishida et al., 2020 - arXiv:astro-ph/2010.05941](https://arxiv.org/abs/2010.05941) - *The RESSPECT team: LSST-DESC and COIN, IEEE Symposium on Computational Intelligence for Astroinformatics, 2020, Canberra, Australia*

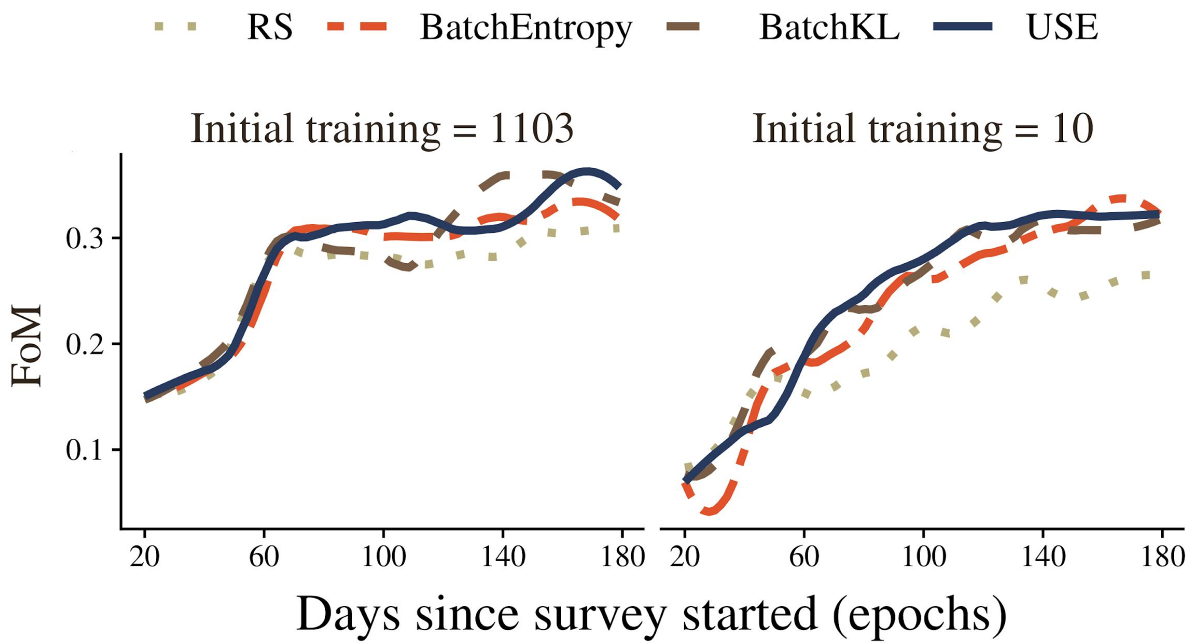
If only it were that simple …

List of the complications that prevent you from getting labels in your science case

Illustrative plot of your data

Your references …

Start from scratch, do not overcomplicate



Kennamer, Ishida et al., 2020 - arXiv:astro-ph/2010.05941 - The RESSPECT team: LSST-DESC and COIN, IEEE Symposium on Computational Intelligence for Astroinformatics, 2020, Canberra, Australia

The difficult part is data treatment/gathering

- The power of machine learning is in its connection with domain knowledge
- There are caveats in using machine learning and we should avoid off-the-shelf and black bloxes applications
- ML for science must be **personalized**

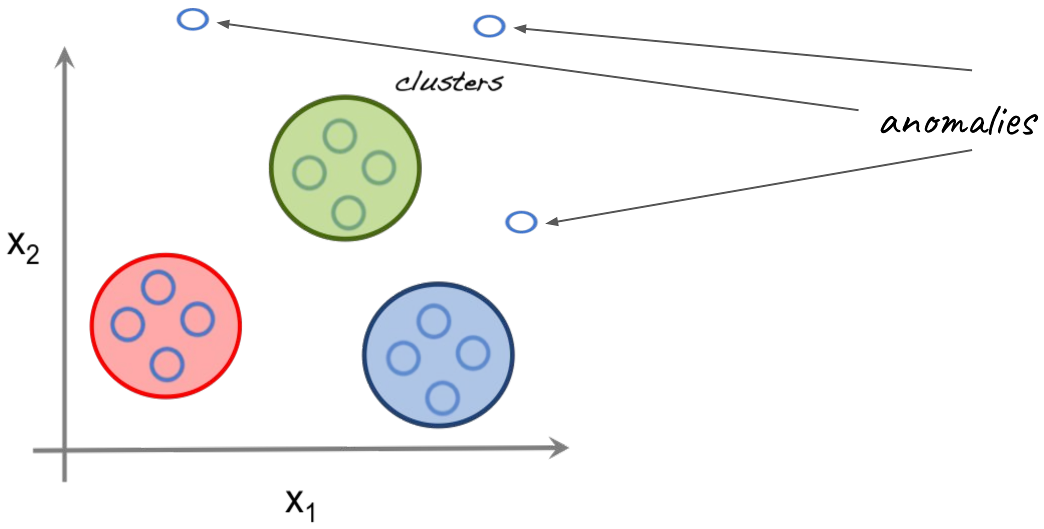
The beauty of an observational science

"… telescopes that merely achieve their stated science goals have probably failed to capture the most important scientific discoveries available to them."

Norris, R. (2017). Discovering the Unexpected in Astronomical Survey Data. Publications of the Astronomical Society of Australia, 34, E007. doi:10.1017/pasa.2016.63

Statistically,

Anomaly Detection

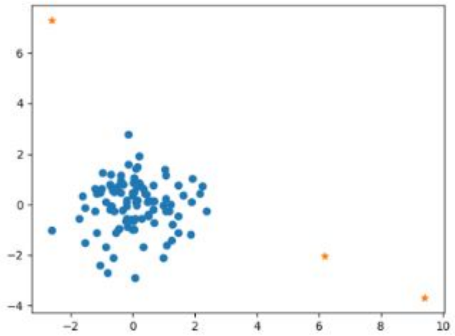


"An anomaly is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism"

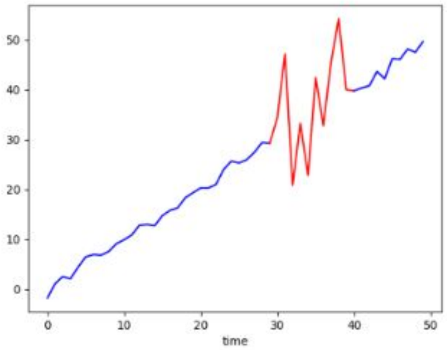
Statistically,

Anomaly Detection

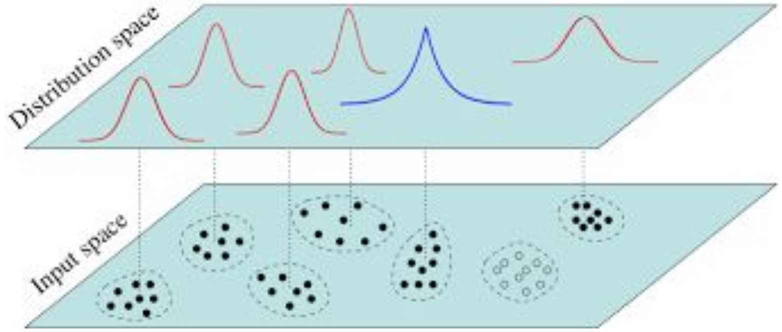
"An anomaly is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism"
Hawkins, 1980



isolation



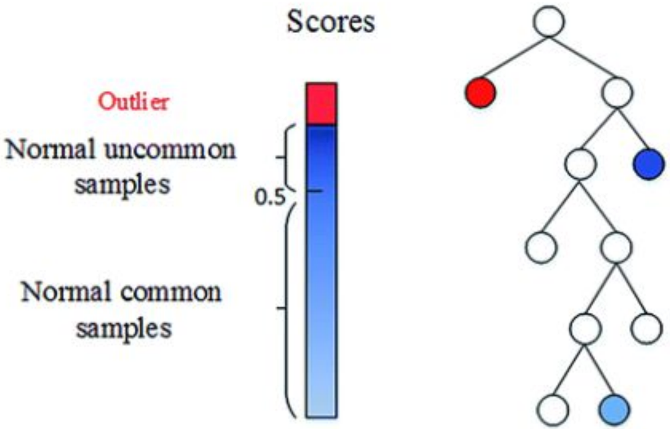
behaviour



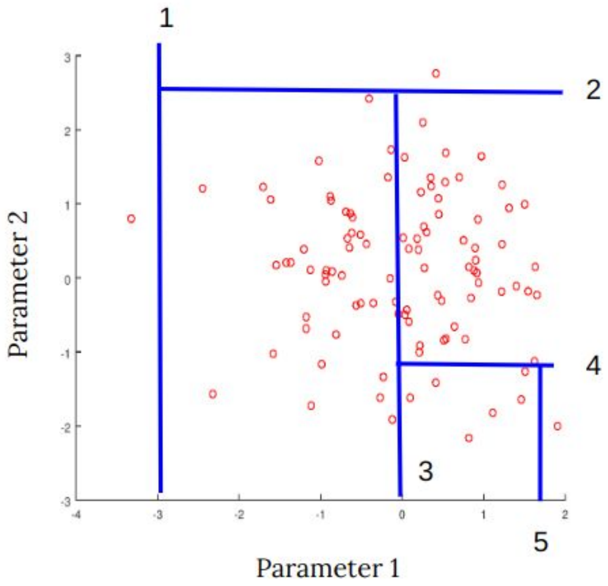
Plot from Muandet and Scholkopf, 2013 - arXiv:stat.ML/1303.0309

group

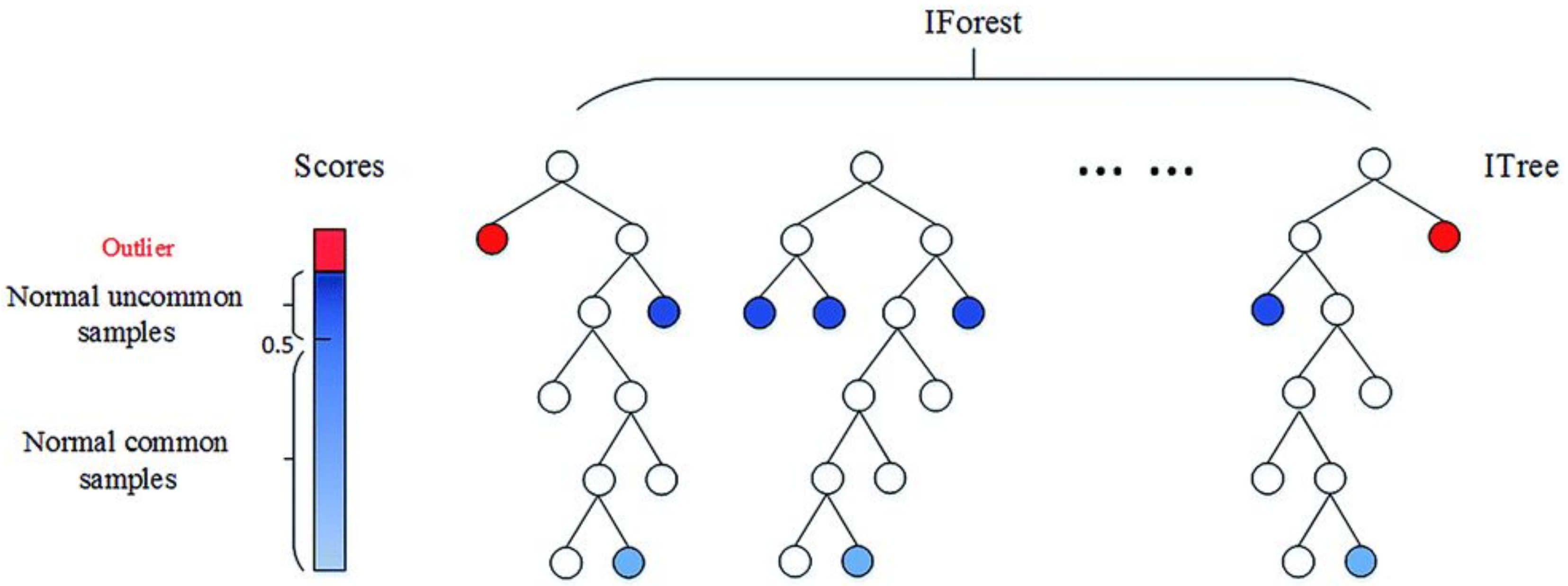
Isolation tree



Plot from https://donghwa-kim.github.io/iforest.html

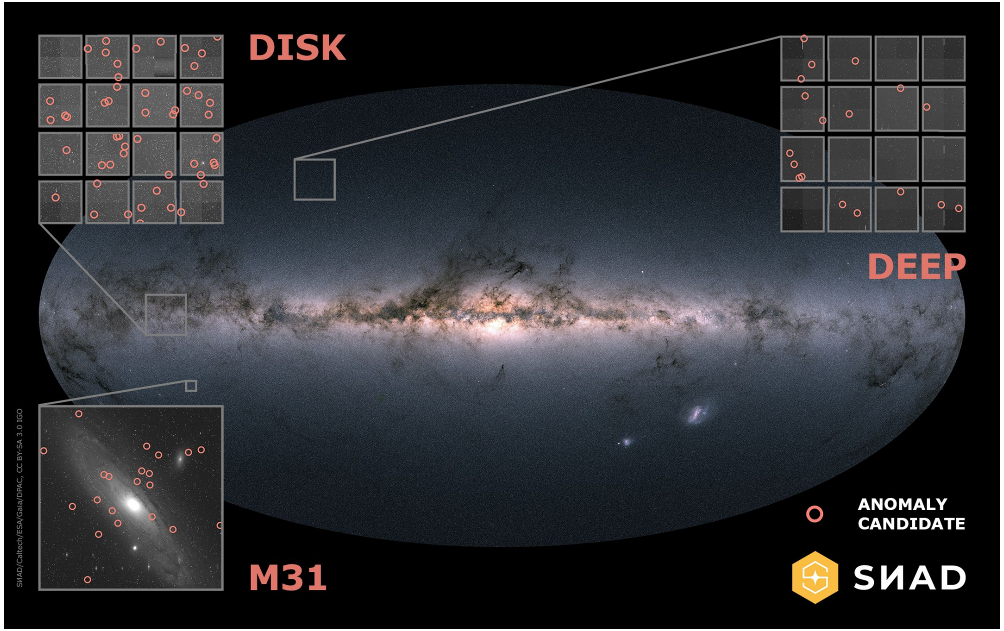


Isolation forest



Plot from https://donghwa-kim.github.io/iforest.html

Zwicky Transient Facility DR3



- Survey currently in operation, telescope in California
- 3 fields from Dara Release 3 (DR3)

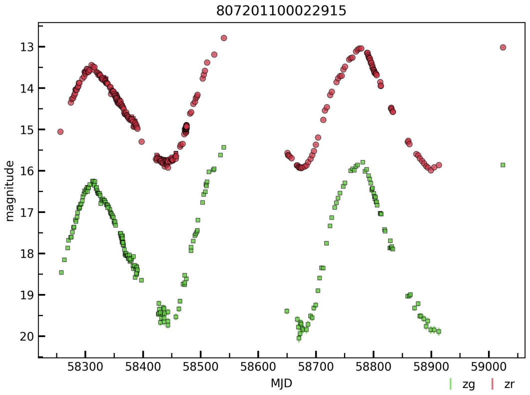
*After selection cuts and feature extraction, **2.25 million objects***

Malanchev et al., 2021 - MNRAS - https://arxiv.org/abs/2012.01419

Figure by Maria Pruzhinskaya

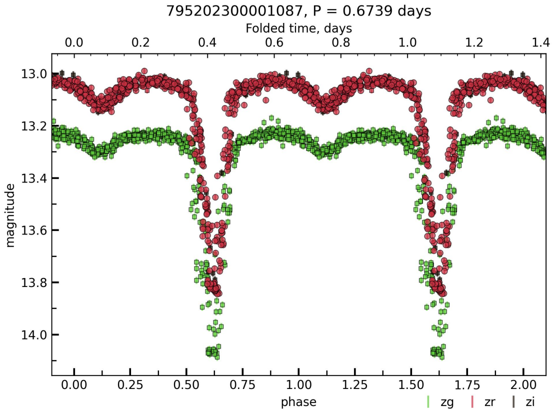
SNAD

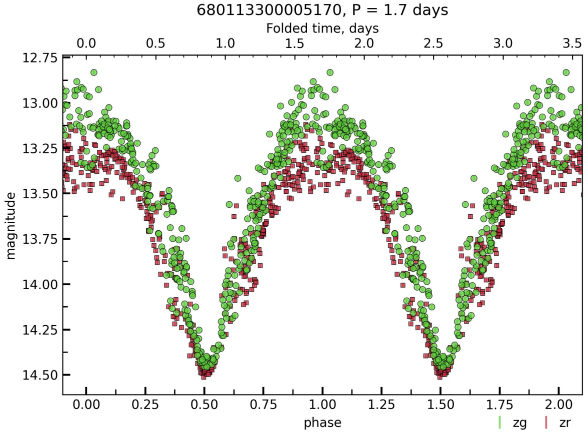
Zwicky Transient Facility DR3



expected to contain stars and periodic variables (no transients)

Visualization generated with the SNAD ZTF viewer: https://ztf.snad.space/

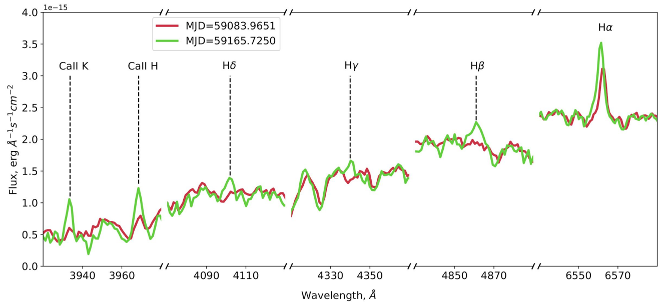




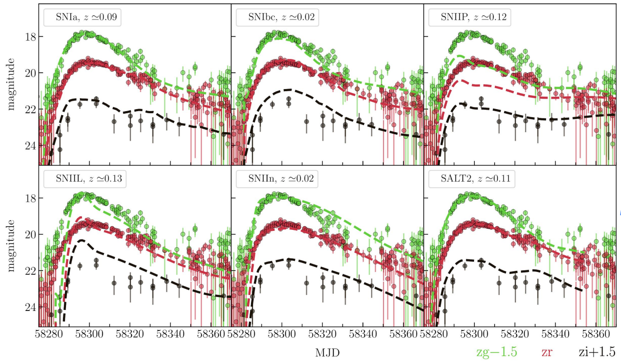
SNAD

Zwicky Transient Facility DR3

- Feature extraction
- Anomaly detection algorithms:
  - *Isolation Forest*
  - *Local Outlier Factor*
  - *Gaussian Mixture Model*
  - *One-Class Support Vector Machine*
- Initial data: 2.25 million objects
- Expert analysis: 277 objects



- 1 RS Canum Venaticorum star
- 1 red dwarf flare
- 4 Supernova candidates



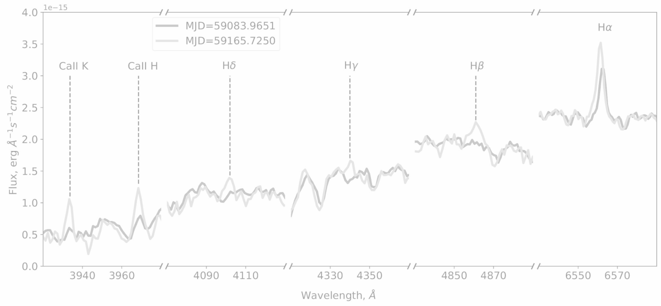
Results:

- 68 % (188) - artifacts, bogus
- 24 % (66) - previously cataloged
- 8 % (23) - discoveries

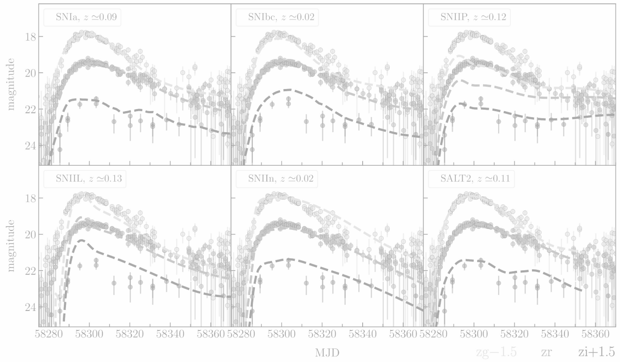
Malanchev et al., 2021 - MNRAS - https://arxiv.org/abs/2012.01419

Zwicky Transient Facility DR3

- Feature extraction
- Anomaly detection algorithms:
  - *Isolation Forest*
  - *Local Outlier Factor*
  - *Gaussian Mixture Model*
  - *One-Class Support Vector Machine*
- Initial data: 2.25 million objects
- Expert analysis: 277 objects



- 1 RS Canum Venaticorum star
- 1 red dwarf flare
- 4 Supernova candidates



Results:

- **68 % (188) - artifacts, bogus**
- 24 % (66) - previously cataloged
- 8 % (23) - discoveries

Malanchev et al., 2021 - MNRAS - https://arxiv.org/abs/2012.01419

It is about Discovery

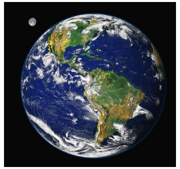
"An anomaly is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism"

Stages of discovery in astronomy:

- Detection
- Interpretation
- Understanding
- Acceptance

Which mechanism?
Is it something we are familiar with but fail to proper model or recognise?
Is it something we have never seen before?

Is there something new for us to learn?

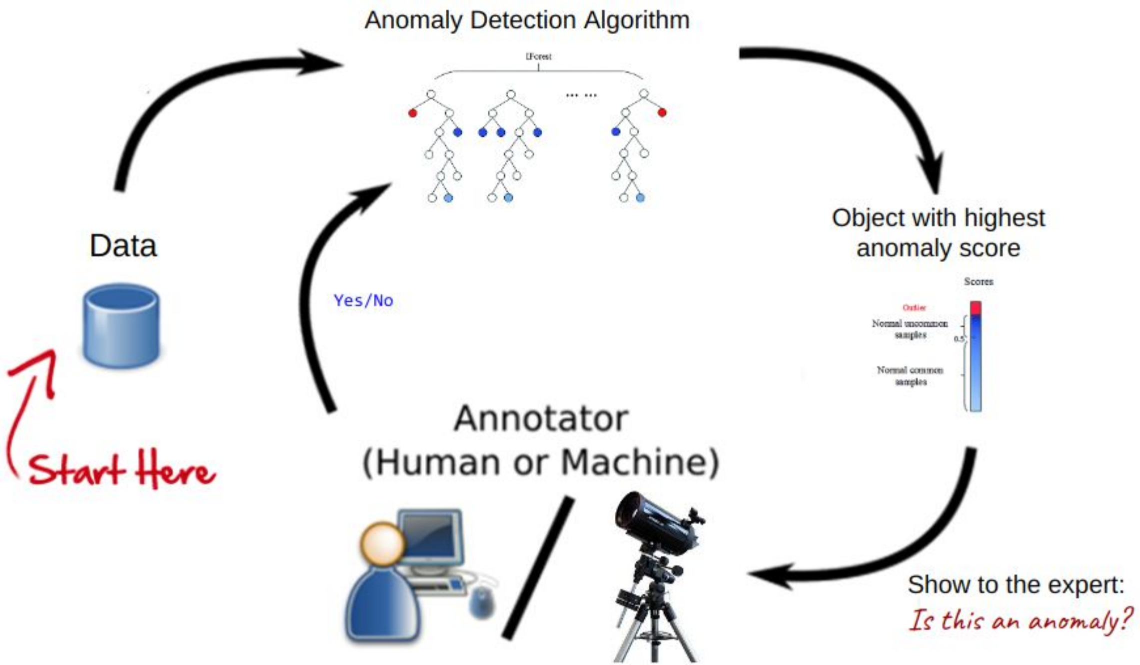


In order to identify the unusual we need to have a clear ideal of what is usual …

.. and that is a social construct. It changes and adapts with time!

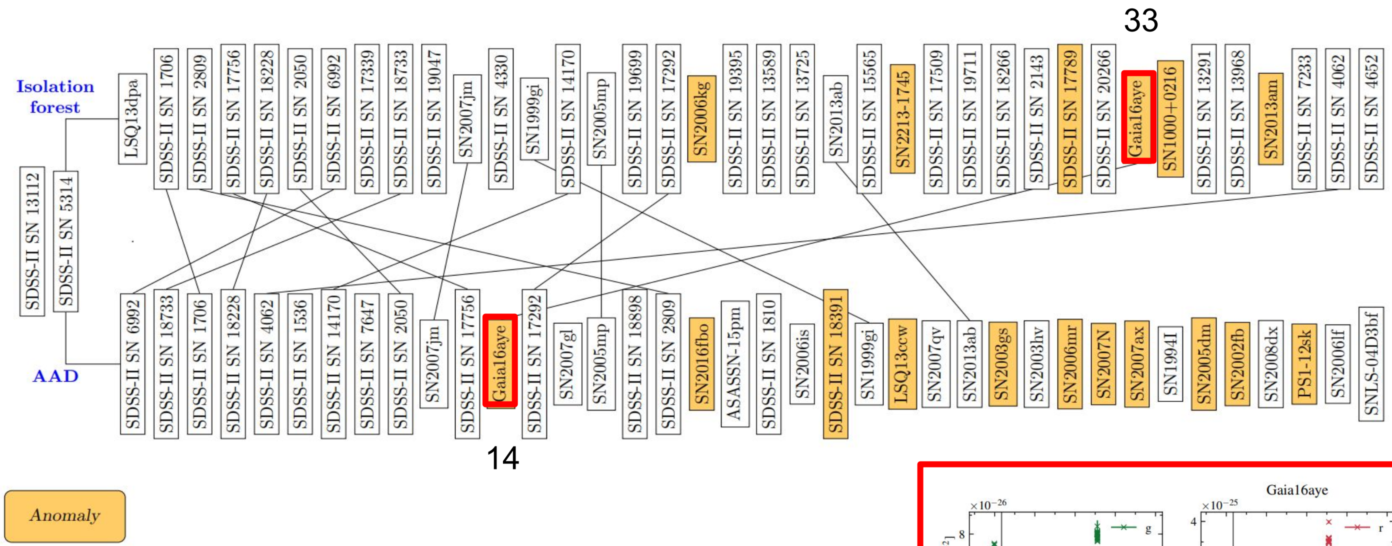
Discovery and Classification in Astronomy - by Steven Dick - Cambridge University Press (2013)

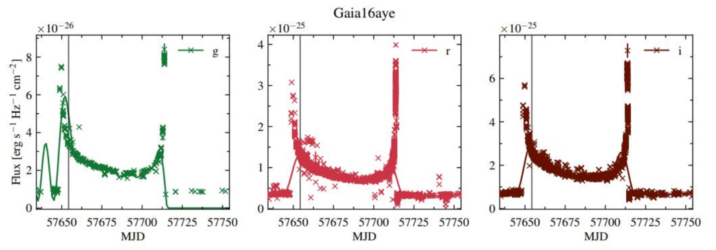
Active Anomaly Detection



Plot modified from [Chowdhury et al., 2021, SPIE Medical Imaging]
Algorithm from Das, S., et al., 2017, in Workshop on Interactive Data Exploration and Analytics (IDEA'17), KDD workshop, [arXiv:cs.LG/1708.09441]
Try the SNAD implementation: [https://coniferest.readthedocs.io/en/latest/tutorial.html]

AAD on real data: The Open Supernova Catalog

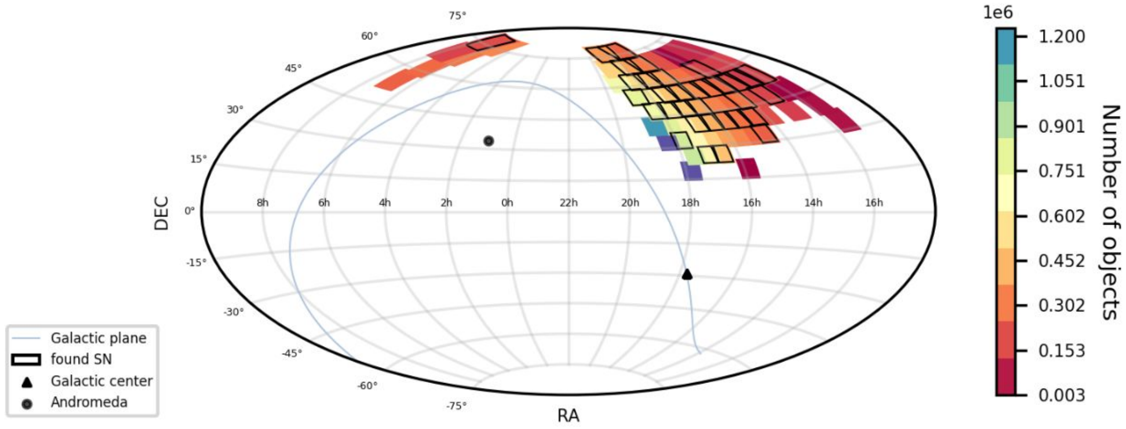




Fast identification of binary microlensing event

Ishida et al., 2021, Astronomy & Astrophysics, Volume 650, id.A195, 9 pp, arXiv:astro-ph/1909.13260

AAD on real data: ZTF data releases



- March - December/2018
- 70 fields
- 30 objects/field
- Total 2100 objects inspected

Found:

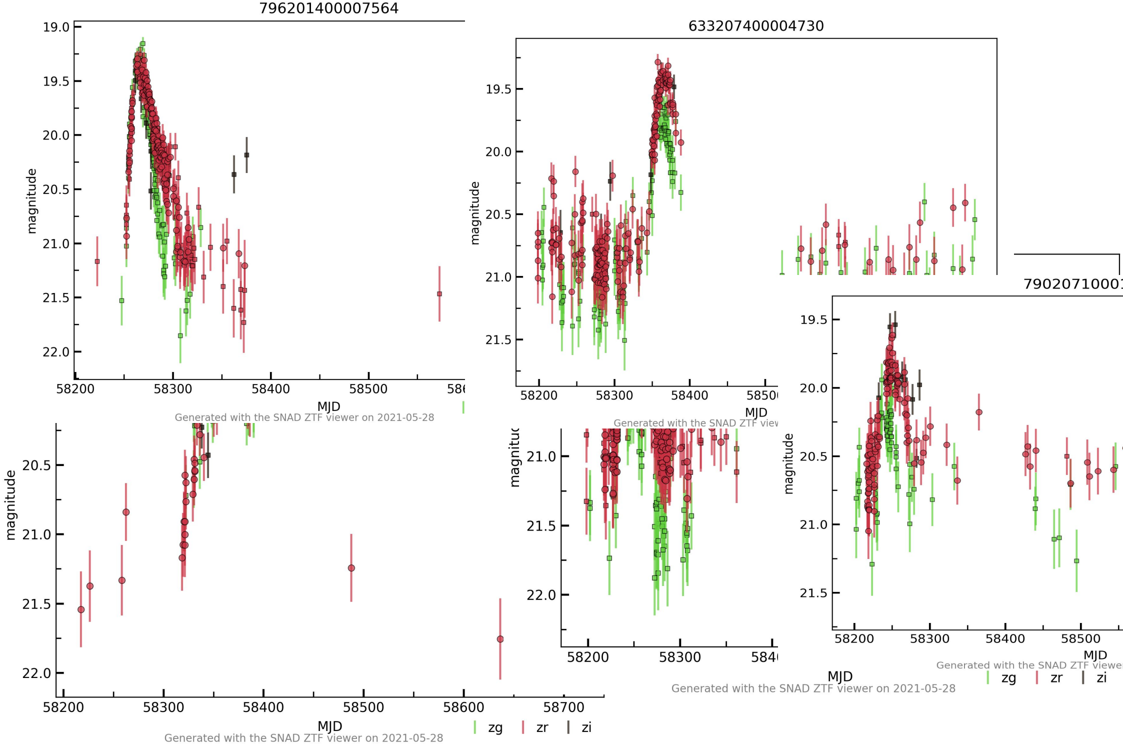
- 100 SN-like objects
  - 46 already catalogued
  - 54 newly discovered
- The SNAD catalog: https://snad.space/catalog/

Pruzhinskaya et al., 2023, A&A, 672, id.A111, arXiv:astro-ph/2208.09053

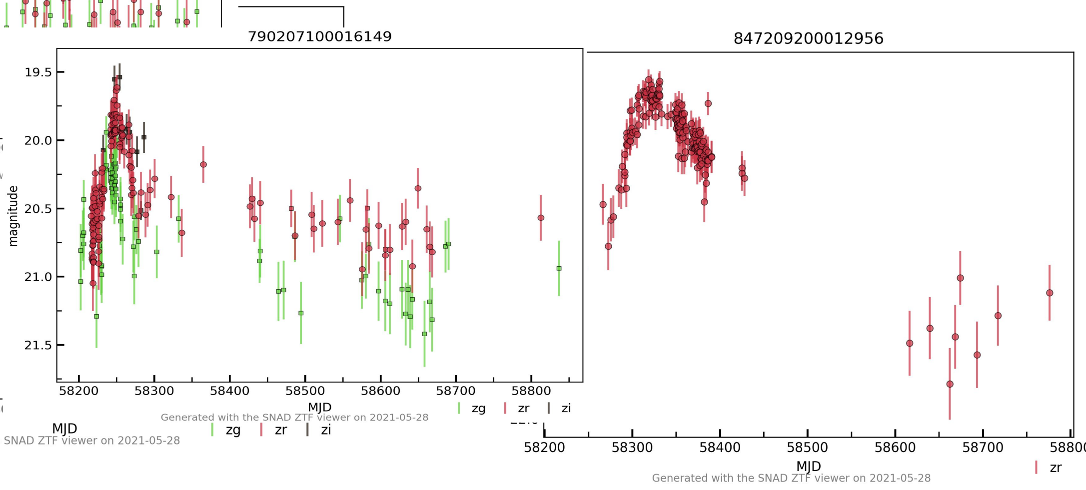
AAD on real data: ZTF DR3

"There are no new supernova-like objects in ZTF DR"

Basically everyone to whom we mentioned we were looking for them.



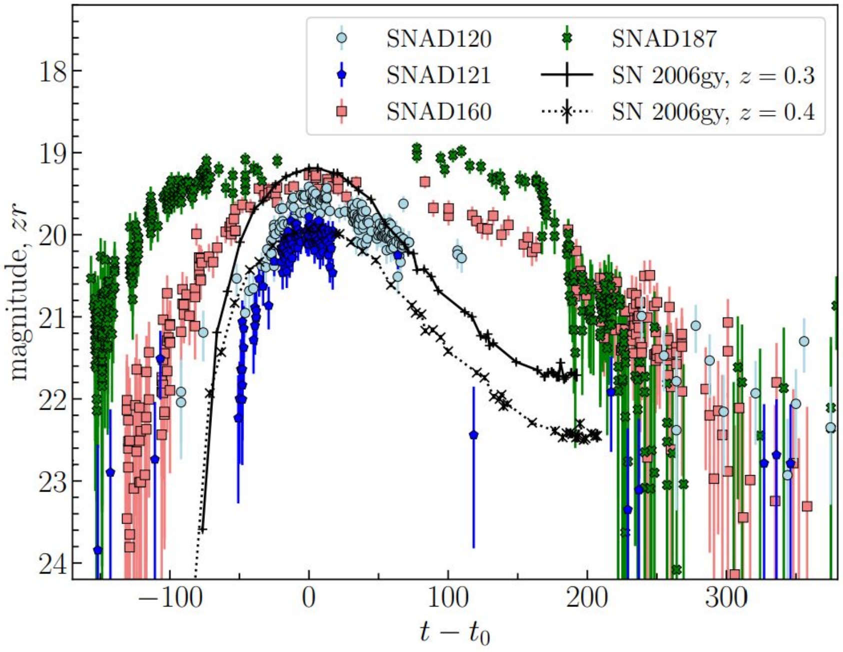
Initial data set: **26.5 million objects**
Visually inspected: **2100**
Confirmed scientifically interesting: **104**
New discoveries: **47**

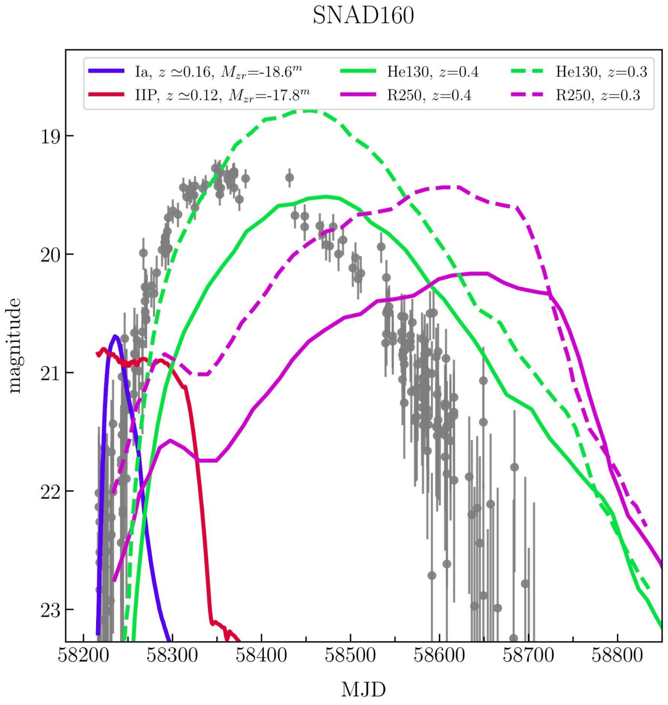


Catalog of lost transients: https://snad.space/catalog/

Pruzhinskaya et al., 2023, A&A - [arXiv:astro-ph/2208.09053](https://arxiv.org/abs/2208.09053)

Interesting SLSN candidates





Pruzhinskaya et al., 2023, A&A 672, A111 (2023), [arXiv:astro-ph/2208.09053]()

Explore the boundaries of your knowledge

- In the era of Big Data, serendipitous discoveries will not happen
- Domain experts **must be included** in the development of new techniques **from the first stages**. They should supervise the first prototypes.

Explore the boundaries of your knowledge

- In the era of Rubin, serendipitous discoveries will not happen

- Domain experts must be included in the development of new techniques from the first stages. They should supervise the first prototypes.

- Scientific discovery is an **intrinsically human** endeavor within a social construct. Advances in the era of Big Data means filtering meaningful small data which people can analyze.

It is crucial to know what you are looking for


TIME TO THINK

Deductive and Plausible Reasoning

Suppose some dark night a policeman walks down a street, apparently deserted. Suddenly he hears a burglar alarm, looks across the street, and sees a jewelry store with a broken window. Then a gentleman wearing a mask comes crawling out through the broken window, carrying a bag which turns out to be full of expensive jewelry. The policeman doesn't hesitate at all in deciding that this gentleman is dishonest.

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Deductive and Plausible Reasoning

Join at menti.com with code: 4697 0711

Deductive Reasoning ≠ Plausible Reasoning

- Deductive reasoning is based on **strong syllogisms**:

  if A is true, then B is true.
  if B is true, then A is true.

No ambiguities!

Deductive Reasoning ≠ Plausible Reasoning

- Deductive reasoning is based on **strong syllogisms**:

  if A is true, then B is true.
  if B is true, then A is true.

No ambiguities!

- Plausible reasoning is constructed from **weak syllogisms**:

  if A is true, then B is true.
  if B is true, then A is more plausible.

Plausible reasoning is the art of making decisions with incomplete, uncertain, messy information

Deductive Reasoning ≠ Plausible Reasoning

- Deductive reasoning is based on **strong syllogisms**:

  if A is true, then B is true.
  if B is true, then A is true.

No ambiguities!

- Plausible reasoning is constructed from **weak syllogisms**:

  if A is true, then B is true.
  if B is true, then A is more plausible.

Plausible reasoning is the art of making decisions with incomplete, uncertain, messy information

The evidence did not make the gentleman's dishonesty certain, but it did make it extremely plausible

Deductive Reasoning ≠ Plausible Reasoning

$A \equiv$ it will start to rain by 10 AM at the latest;

$B \equiv$ the sky will be cloudy before 10 AM.

Join at menti.com with code: 4697 0711

Deductive Reasoning ≠ Plausible Reasoning

if A is true, then B is true.
if B is true, then A is more plausible.

$A \equiv$ it will start to rain by 10 AM at the latest;

$B \equiv$ the sky will be cloudy before 10 AM.

Question:
What is the logical connection? $A \Rightarrow B$ or $B \Rightarrow A$

Answer:
if rain (A) then clouds (B): which is **certain**, though not causal.

Deductive Reasoning ≠ Plausible Reasoning

- Deductive reasoning is based on **strong syllogisms (first kind)**:

  if A is TRUE, then B is TRUE.
  if B is TRUE, then A is TRUE.

- Plausible reasoning is constructed from **weak syllogisms (second kind)**:
  if A is TRUE, then B is TRUE.
  if B is TRUE, then A is *more plausible*

- There is also the possibility of **weak syllogisms (third kind)**:
  if A is TRUE, then B is TRUE.
  if B is FALSE, then A is *less plausible*

Deductive Reasoning ≠ Plausible Reasoning

Given the same description of the situation, what could lead the policeman to arrive in a different conclusion?

Artificial Thinking

Will computers ever be able to think?

Probability Theory: the logic of Science - by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

A robot that can reproduce at least part of what a brain does would have to make decisions based on weak syllogisms

Probability Theory: the logic of Science – by E. T. Janes– Cambridge University Press (2003)

Artificial Thinking

Consider that we are building a robot who follows the basic desiderata:

I. *Degrees of plausibility are represented by real numbers.*

II. *Qualitative correspondence with common sense.*

III. (a) *If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.*

(b) *The robot always takes into account all of the evidence it has relevant to a question. It does not arbitrarily ignore some of the information, basing its conclusions only on what remains. In other words, the robot is completely nonideological.*

(c) *The robot always represents equivalent states of knowledge by equivalent plausibility assignments. That is, if in two problems the robot's state of knowledge is the same (except perhaps for the labeling of the propositions), then it must assign the same plausibilities in both.*

Our robot will make decisions base solely in degrees of plausibility, which by definition are real numbers. Humans normally take into account many different aspect of a question when making a decision. Can you think of a class (es) of issues where our robot would most closely mimic the behavior of a human?

Probability Theory: the logic of Science – by E. T. Janes– Cambridge University Press (2003)

Artificial Thinking

Analytical part of the robot is already here … because this is the part of “thinking” we can better describe

Artificial Thinking

Analytical part of the robot is already here … because this is the part of "thinking" we can better describe

It is embedded in the foundations of our logical system

Artificial Thinking

Analytical part of the robot is already here … because this is the part of "thinking" we can better describe

It is embedded in the foundations of our logical system

Boolean algebra (TRUE/FALSE)

Artificial Thinking

Analytical part of the robot is already here … because this is the part of "thinking" we can better describe

It is embedded in the foundations of our logical system

Boolean algebra (TRUE/FALSE)

Roots in Organon, of Aristotle, 400 BC.

Artificial Thinking

The robot is presented with this situation:

$B \equiv$ An urn contains N balls, identical in every respect except that they carry numbers (1, 2, . . . , N) and M of them are colored red, with the remaining $(N - M)$ white, $0 \leq M \leq N$. We draw a ball from the urn blindfolded, observe and record its color, lay it aside, and repeat the process until n balls have been drawn, $0 \leq n \leq N$.

$R_i \equiv$ Red ball on the i-th draw.

$W_i \equiv$ White ball on the i-th draw.

For the first draw,

$$P(R_1|B) = \frac{M}{N}, \tag{1}$$

$$P(W_1|B) = 1 - \frac{M}{N}. \tag{2}$$

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The robot is presented with this situation:

For the first draw,

$$P(R_1|B) = \frac{M}{N}, \tag{1}$$

$$P(W_1|B) = 1 - \frac{M}{N}. \tag{2}$$

What do equations 1 and 2 tell you about the content of the urn?

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The robot is presented with this situation:

For the first draw,

$$P(R_1|B) = \frac{M}{N}, \tag{1}$$

$$P(W_1|B) = 1 - \frac{M}{N}. \tag{2}$$

What do equations 1 and 2 tell you about the content of the urn?

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The situation described in before can be identified as *sampling without replacement*. Consider now the situation of *sampling with replacement*. Meaning that every time we draw a ball from the urn we record its color and put it back before drawing again.

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The situation described in before can be identified as *sampling without replacement*. Consider now the situation of *sampling with replacement*. Meaning that every time we draw a ball from the urn we record its color and put it back before drawing again.

Which of the two scenarios would the robot consider more complex for estimating the probability of a given color in the subsequent draw?

Join at menti.com with code: 4697 0711

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The situation described in before can be identified as *sampling without replacement*. Consider now the situation of *sampling with replacement*. Meaning that every time we draw a ball from the urn we record its color and put it back before drawing again.

Which of the two scenarios would the robot consider more complex for estimating the probability of a given color in the subsequent draw?

With replacement

Join at menti.com with code: 4697 0711

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The situation described in before can be identified as *sampling without replacement*. Consider now the situation of *sampling with replacement*. Meaning that every time we draw a ball from the urn we record its color and put it back before drawing again.

How can make the sampling with replacement simpler for this task?

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The situation described in before can be identified as *sampling without replacement*. Consider now the situation of *sampling with replacement*. Meaning that every time we draw a ball from the urn we record its color and put it back before drawing again.

How can make the sampling with replacement simpler for this task?

Probability Theory: the logic of Science – by E. T. Janes- Cambridge University Press (2003)

Artificial Thinking

The situation described in before can be identified as *sampling without replacement*. Consider now the situation of *sampling with replacement*. Meaning that every time we draw a ball from the urn we record its color and put it back before drawing again.

How can make the sampling with replacement simpler for this task?

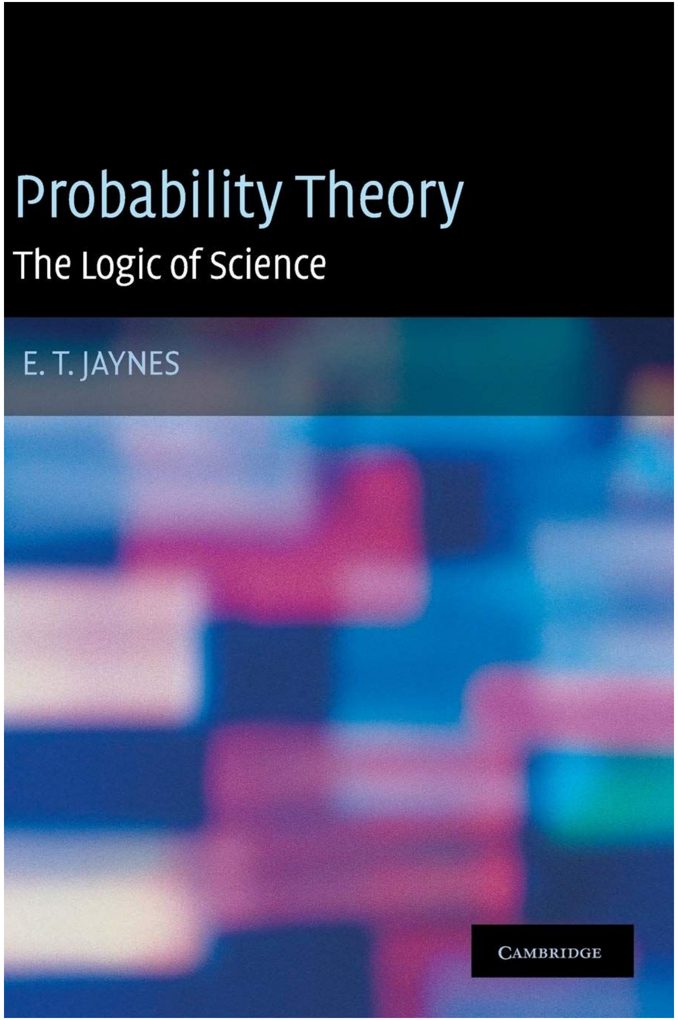
Shaking changes your state of knowledge, you are basically saying: I am not able anymore to follow the position of the ball that was put back in because I do not have all necessary information (the position of the ball that was put back changed and I do not know where it is, so my problem returns to the mathematical state of symmetry – I know nothing about all balls). From the point of view of the urn, nothing change.

Probability Theory: the logic of Science - by E. T. Janes- Cambridge University Press (2003)

Constructing the future will take a lot of math

- All machine learning is based on probability
- We will be able to construct a computer that "thinks" when we can describe what "thinking" means
- We already started

To use machine learning responsibly we need to, at least, have an idea about the theory behind it



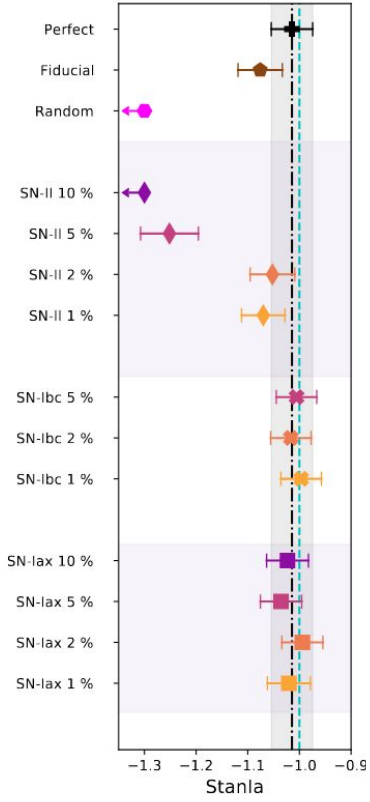
This digression was a very bad summary of the first few pages of the best statistics book ever written …

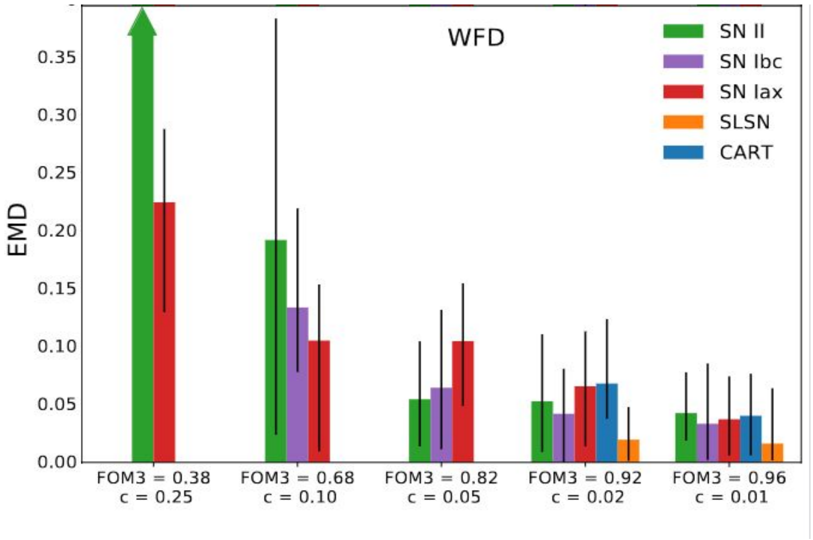
[Probability Theory: the logic of Science - by E. T. Janes- Cambridge University Press (2003)]

THANK YOU

Extra slides

Good classification might not be enough

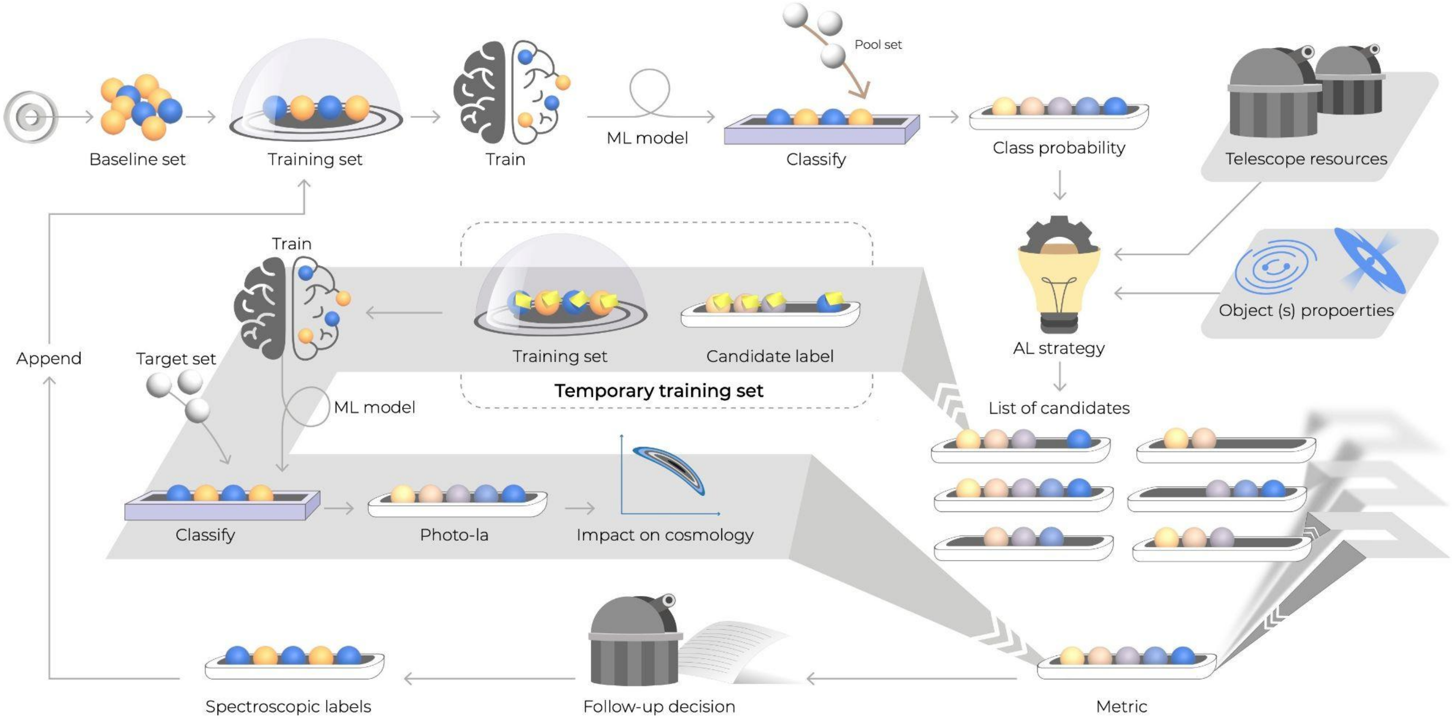




LSST DESC & COIN
RESSPECT
Recommendation System for Spectroscopic Follow-up

Malz et al., 2023 - [arXiv:astro-ph/2305.14421](https://arxiv.org/abs/2305.14421) - The RESSPECT team: LSST-DESC and COIN,
Are classification metrics good proxies for SN Ia cosmological constraining power? -- submitted to A&A

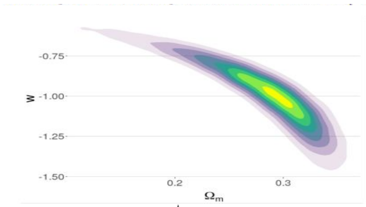
The RESSPECT workflow



https://github.com/COINtoolbox/RESSPECT

What about science?

Cosmology results from photometrically classified SN IA



2. Impacts this

Photo-classified SN Ia

SN candidates

Trained machine learning classifier

learning algorithm

+

Training sample

1. Different choices of this!

LSST DESC & COIN RESSPECT

Recommendation System for Spectroscopic Follow-up