FINK broker

XMM2Athena Workshop: XMM Newton survey legacy for Athena and beyond
IRAP, Toulouse - France
28 February 2024

Emille Ishida, Anais Möller and Julien Peloton
on behalf of the Fink Team

LPC Particules pLasmas Univers applicationS
Laboratoire de Physique de Clermont
CCIN2P3
cnrs

MMA and discovery with Rubin

- State of the art data sets in optical astronomy
- The Rubin broker ecosystem
- The role of Machine Learning
- The human factor

Big data happened

SDSS

120 TB

Primary mirror: 2.5 m

2000 - now

ZTF

1.4 TB/night

Primary mirror: 1.2 m

2018 - now

VERA C. RUBIN OBSERVATORY

15 TB/night

From 2025

Primary mirror: 8.4 m

The Vera Rubin Observatory
Large Survey of Space and Time (LSST)

In a nutshell:

- telescope: 8.4m primary mirror
- world's largest CCD camera:
  3.2 Gpixels

In numbers:

- 10-year survey, starting 2024+
- 1,000 images/night = 15TB/night
- 10 million transient candidates per night
  - Publicly available…
  - … but huge!

Slide from Julien Peloton (IJCLab)

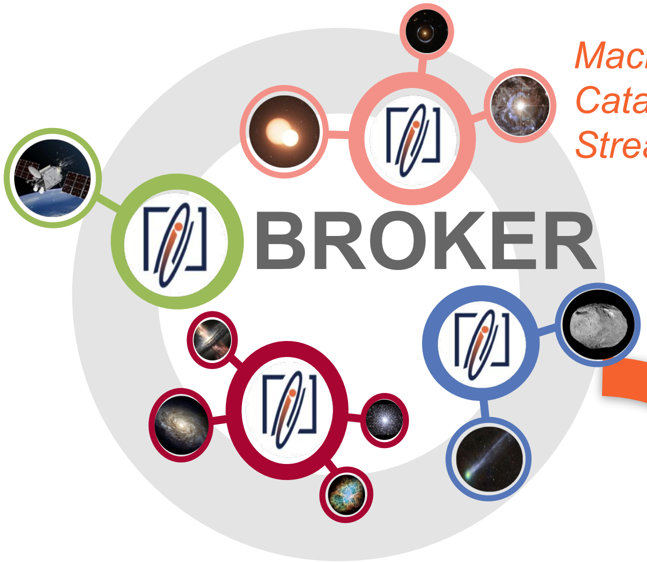
To keep in mind …

1) Very big data

2) Survey mode observation strategy

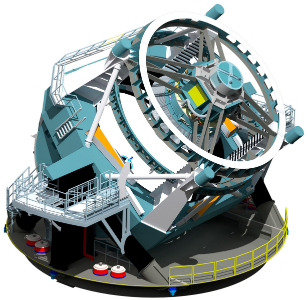
How to distribute this to people who will do science?

The data path



Machine learning
Catalog association
Streams join

every ~30 seconds down to mag ~24



10 million alerts per night…

We would like the ***interesting*** *ones ...*

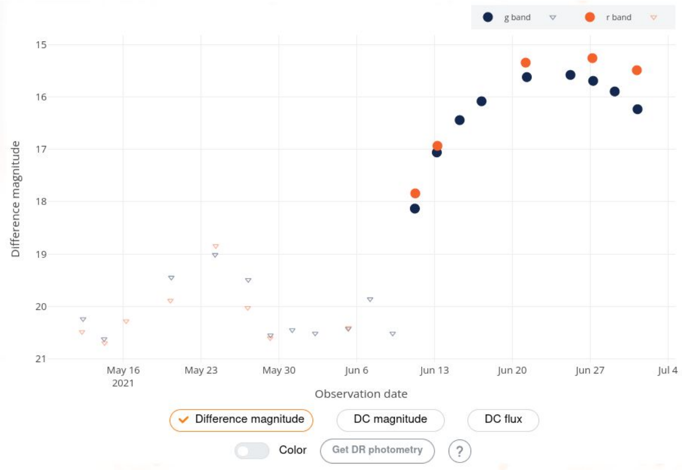
Rubin broker landscape

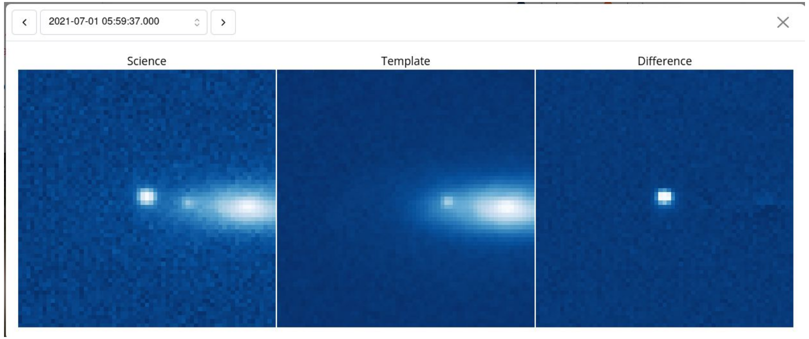
(What is an alert?)

Alerts based on Difference Image Analysis

Each alert contains

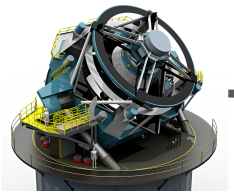
- Information about the new detection (magnitude, position, ...)
- Neighbours information (xmatches, etc)
- Historical information if the object has been seen previously
- Small images around the detection (60x60 pixels)

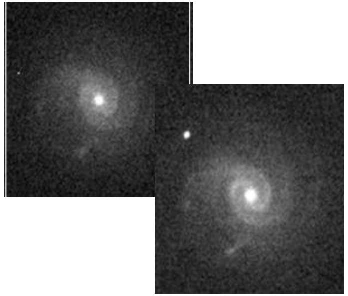




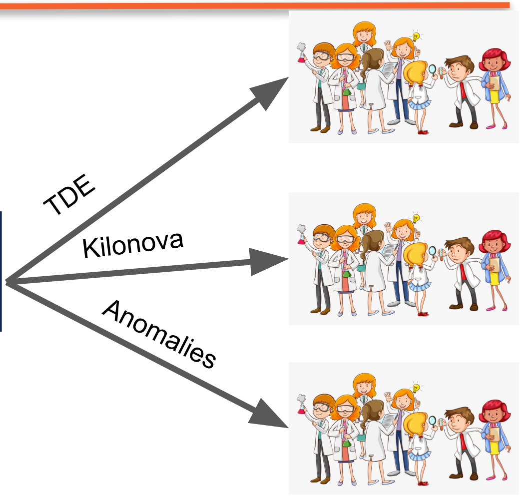
How to classify alerts?

Broker world





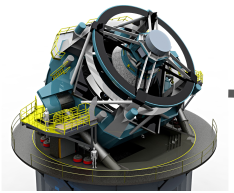
Brokers

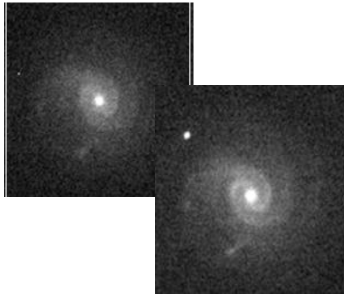


Domain specialist world (this is you)

How to classify alerts?

Broker world





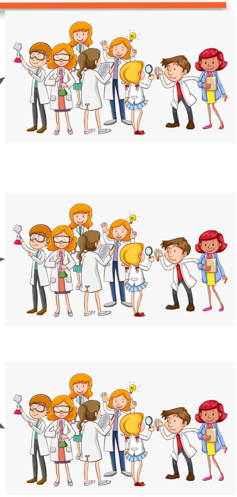
Brokers

FINK

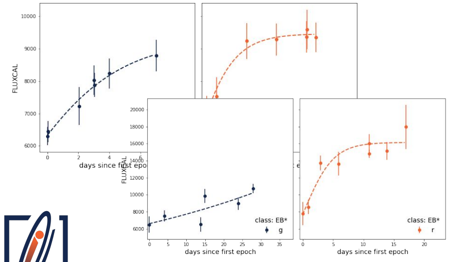
TDE

Kilonova

Anomalies



Domain specialist world (this is you)



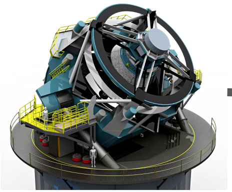
You can access this via de Fink Science portal or the API

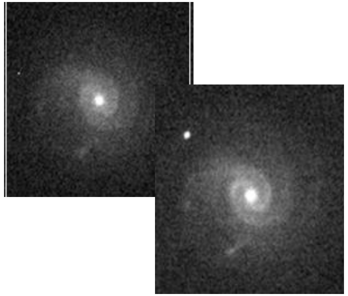
All alerts data is public!

https://fink-portal.org/

How to classify alerts?

Broker world





Brokers

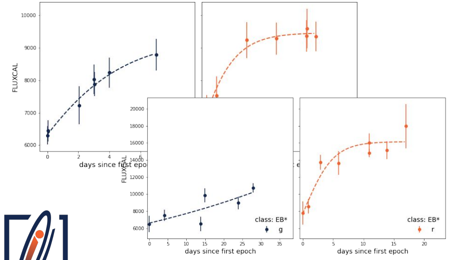
FINK

TDE

Kilonova

Anomalies

Domain specialist world (the expert)



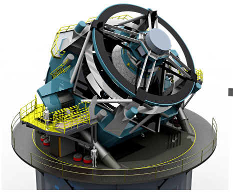
Filter

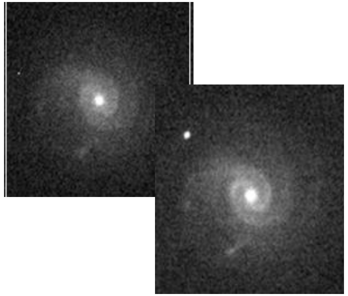
Catalog or stream xmatch

Taylored science module

How to classify alerts?

Broker world





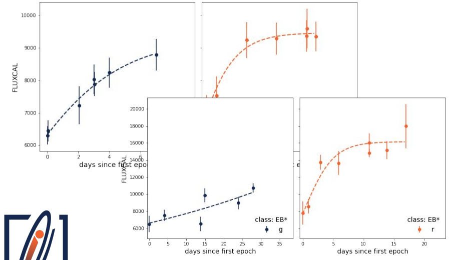
Brokers

TDE

Kilonova

Anomalies

Domain specialist world (the expert)



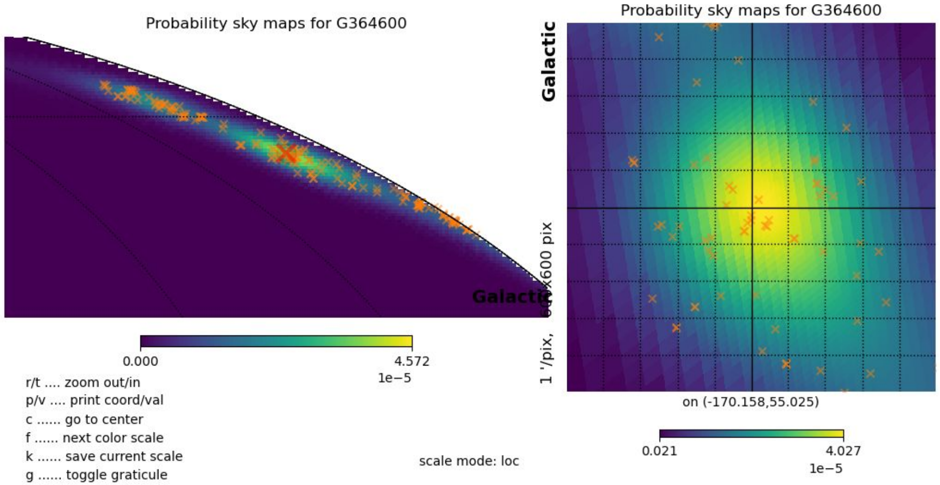
- Magnitude
- Position, or any column really
- X-match with
  - existing catalogs
  - other streams
  - probability maps

Xmatch with GW streams

O4 is coming – Fink has already some tool to play with GW sky maps

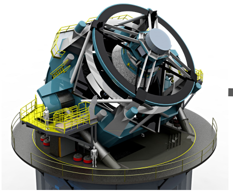
https://fink-portal.org/api → Gravitational Waves → tutorial!

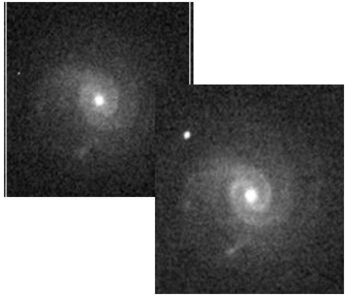
```python
# Query Fink
data = gzip.open(fn, 'rb').read()
r = requests.post(
    'https://fink-portal.org/api/v1/bayestar',
    json={
        'bayestar': str(data),
        'credible_level': credible_level,
        'output-format': 'json'
    }
)
```



How to classify alerts?

Broker world





Brokers

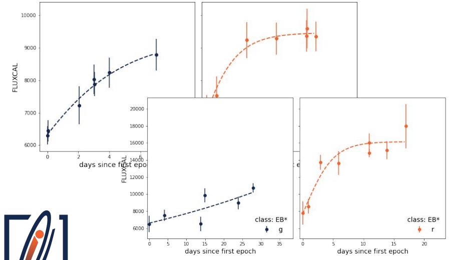
FINK

TDE

Kilonova

Anomalies

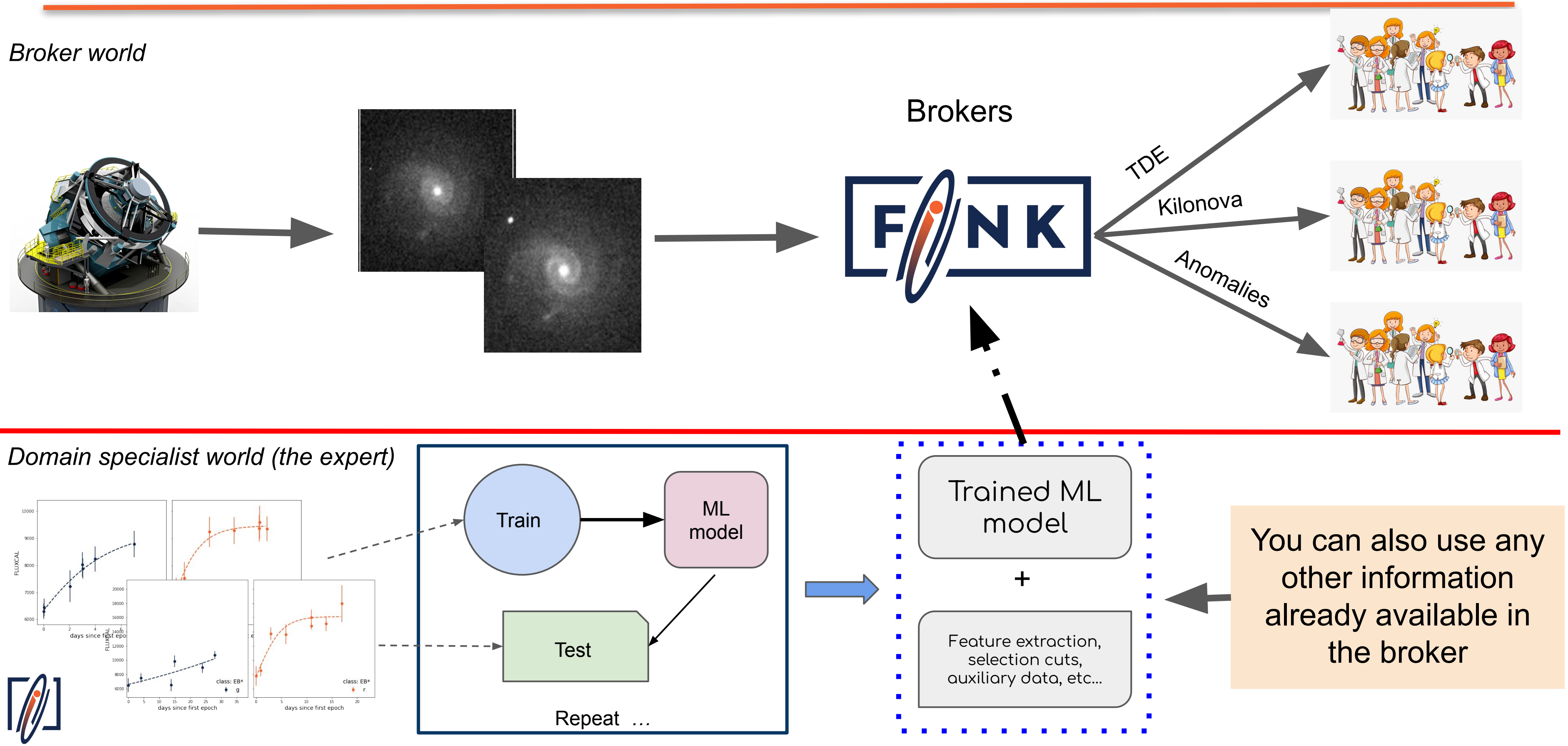
Domain specialist world (the expert)



Taylored science module

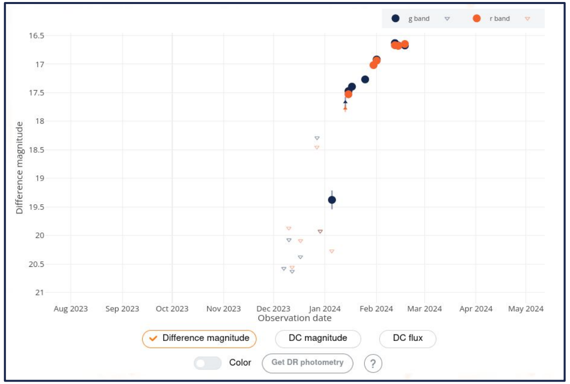
f(alerts; ++) => class scores
Boolean

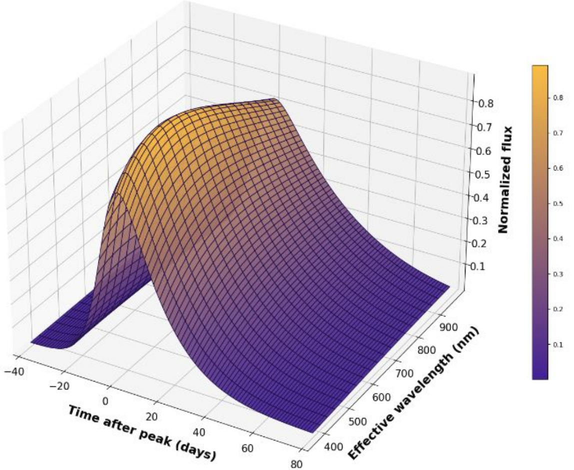
How to classify alerts?

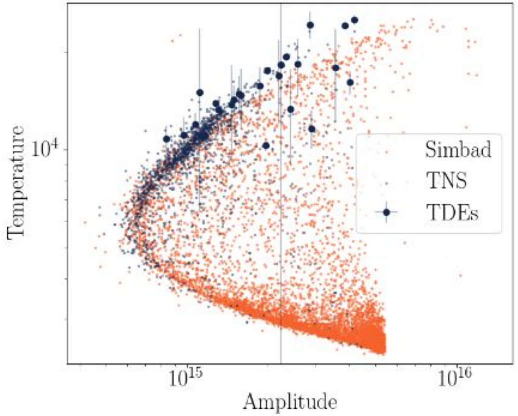


Example: TDE

- Focused on rising TDE examples
- Remove things we already know (xmatch with whatever possible)
- Multi-wavelength feature extraction with Rainbow
- Anomaly detection based ML-model







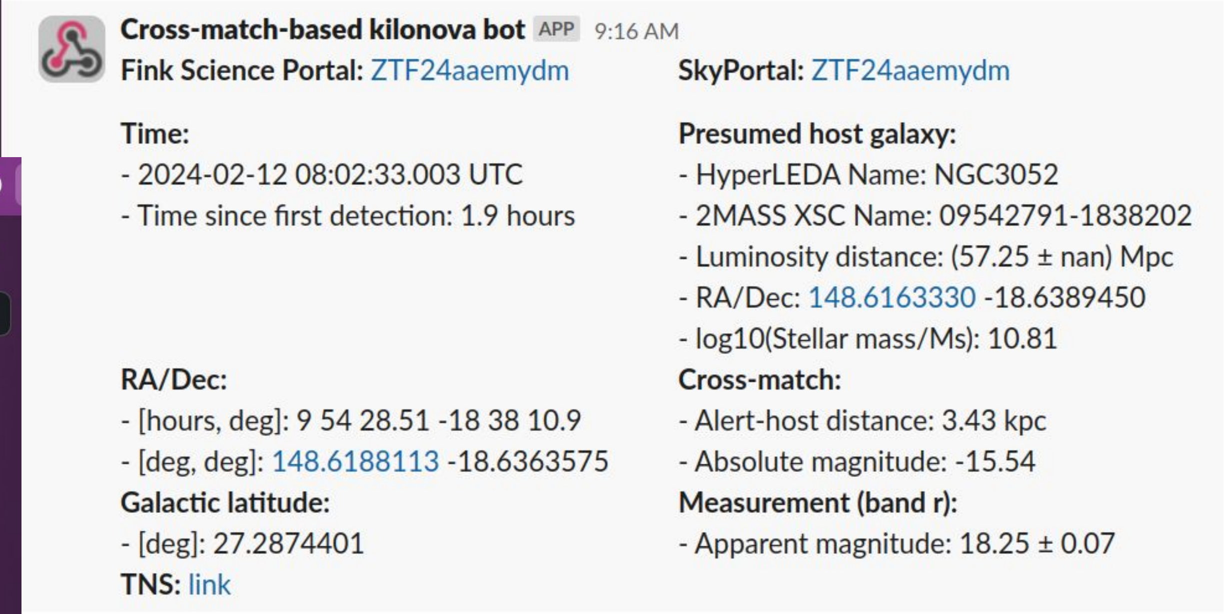
Work by Erwan Quintin, Miguel Llamas Lanza and Etienne Russeil

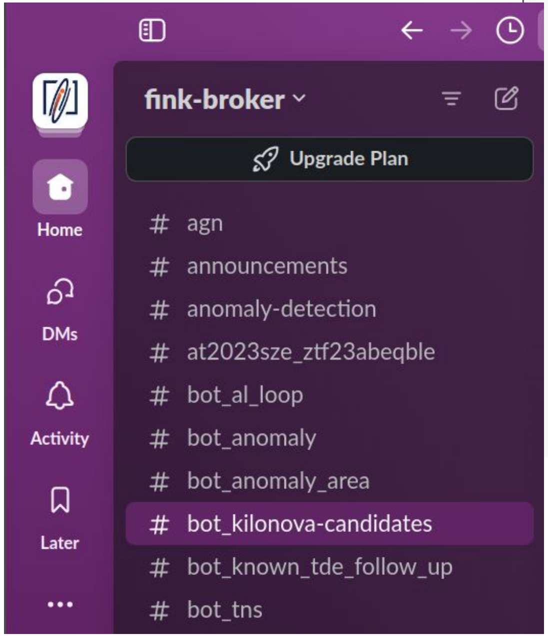
Receiving your candidates

- Kafka stream -- for on-the-fly notifications

Receiving your candidates

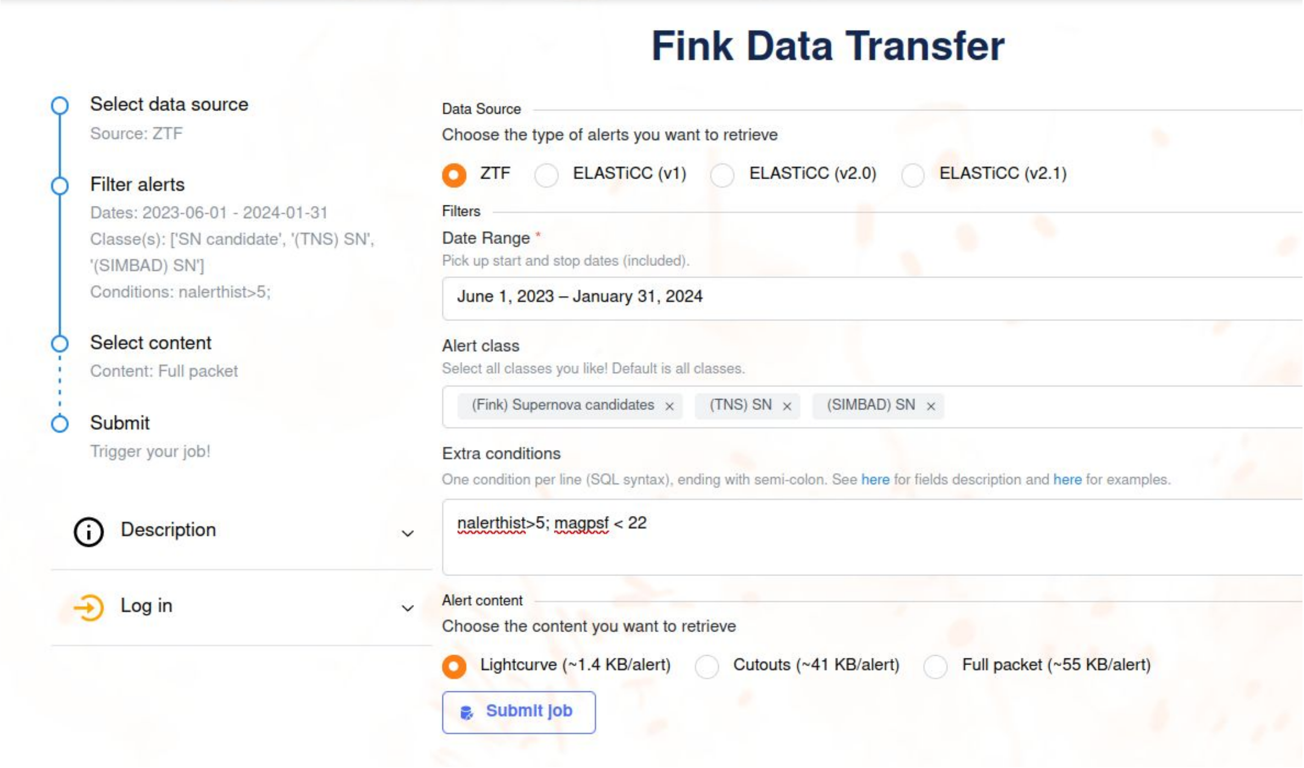
- Bots





Receiving your candidates

- Added value + download service



https://fink-portal.org/download

The beauty of an observational science

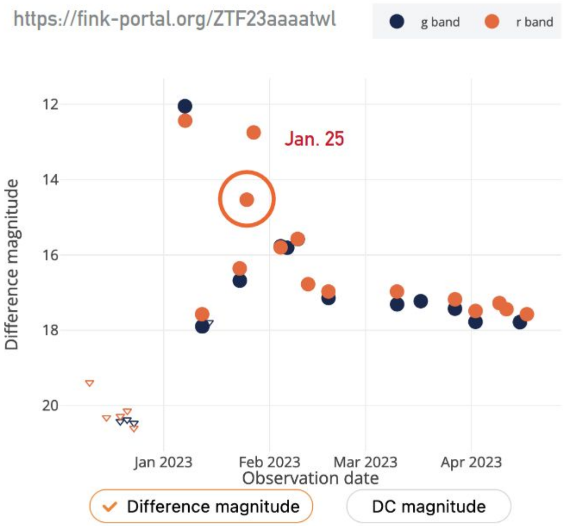
"… telescopes that merely achieve their stated science goals have probably failed to capture the most important scientific discoveries available to them."

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

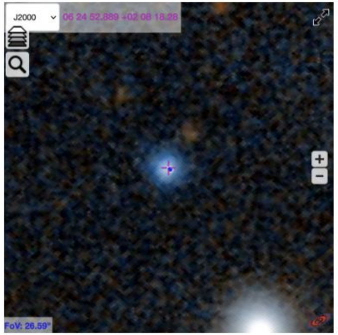
Fink Anomaly Detection module

- Started with extragalactic experts, now Fink-wide engagement
- Random forest model
- Quick reaction from the follow-up community:
  - 9.2m SALT (South Africa)
  - 0.6m and 2.5m KGO (Russia)
  - 0.25m FRAM-ORM (Spain)
  - 0.2M FOSC-ES32 (Italy)

FIRST NIGHT – FIRST ANOMALY

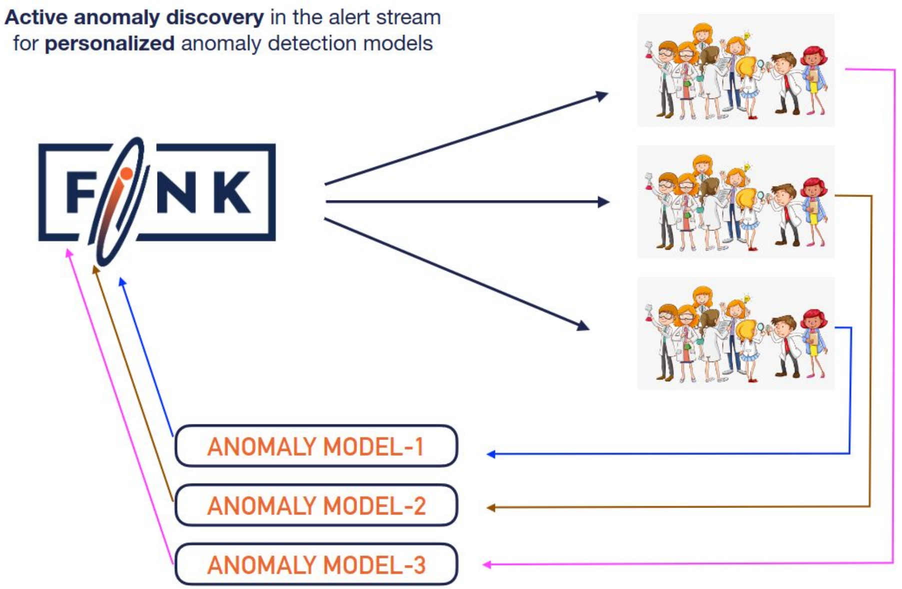


- ASASSN-23ac/PNV J06245297+0208207
- Simbad: WD candidate



REPORTED IN TNS AS AT 2023AWT

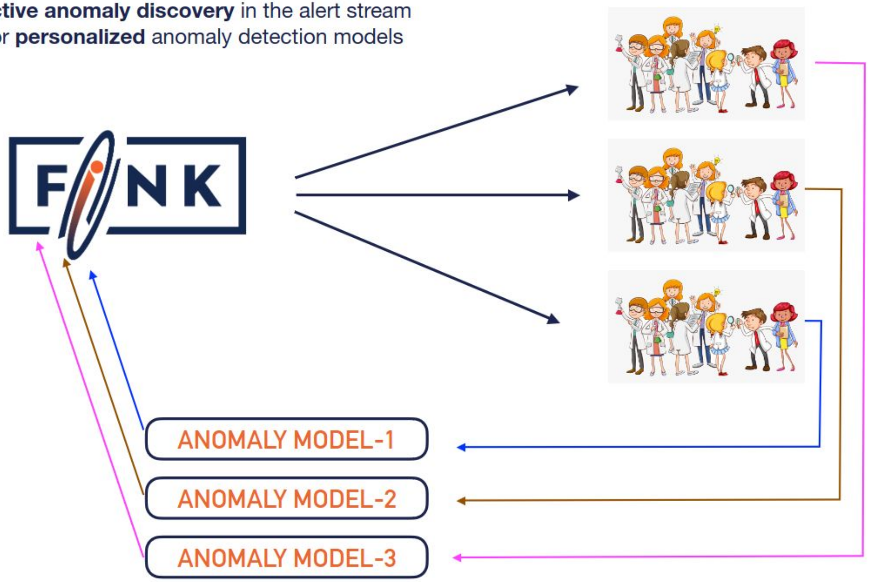
Fink Anomaly Detection module

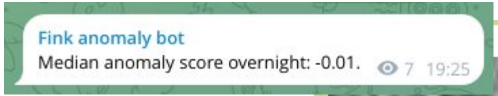


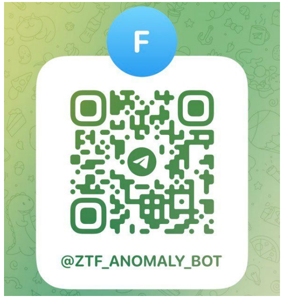
Algorithm from Das, S., et al., 2017, in DEA'17, KDD workshop, arXiv:cs.LG/1708.09441
Implementation by the SNAD team, via coniferest package

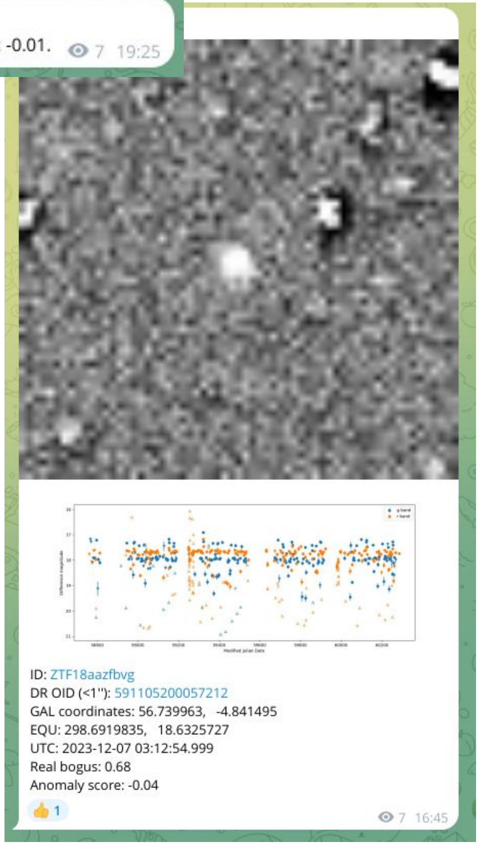
Fink Anomaly Detection module

Active anomaly discovery in the alert stream
for **personalized** anomaly detection models



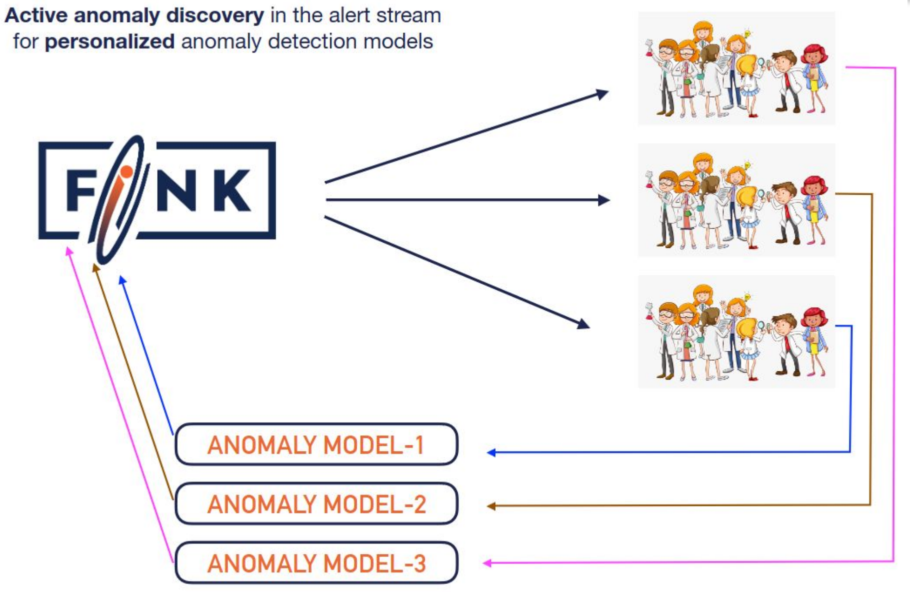


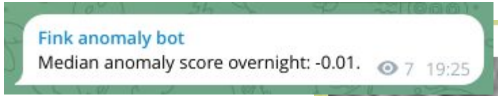


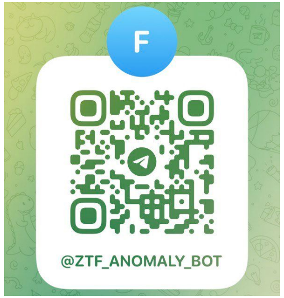


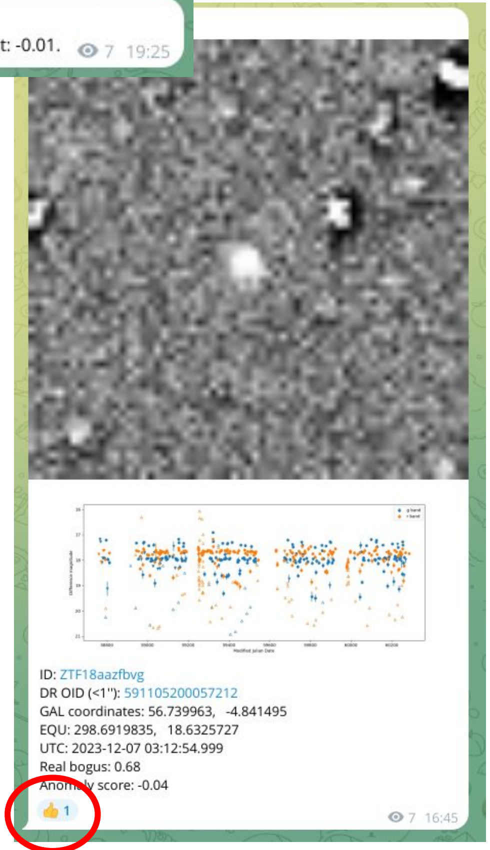
Algorithm from Das, S., et al., 2017, in DEA'17, KDD workshop, [arXiv:cs.LG/1708.09441]
Implementation by the [SNAD team], via [coniferest package]

Fink Anomaly Detection module





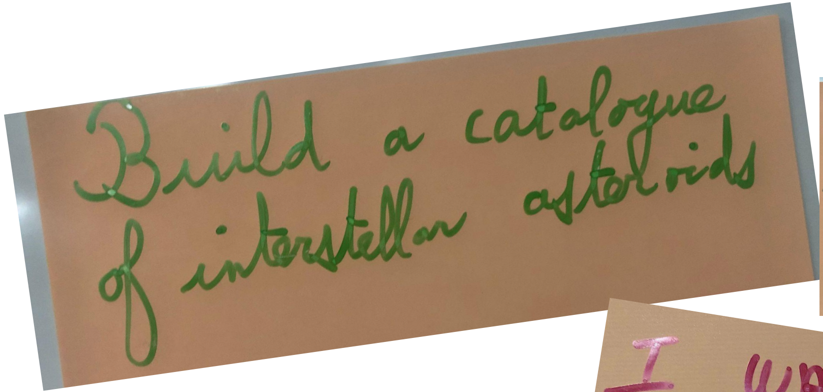


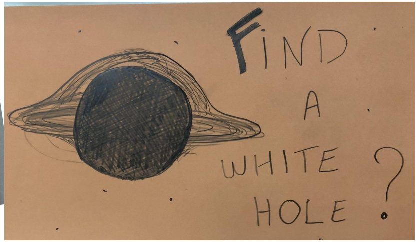


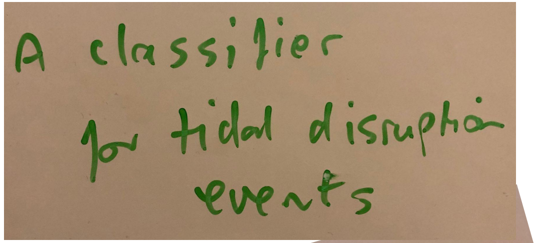
Algorithm from Das, S., et al., 2017, in DEA'17, KDD workshop, arXiv:cs.LG/1708.09441
Implementation by the SNAD team, via coniferest package

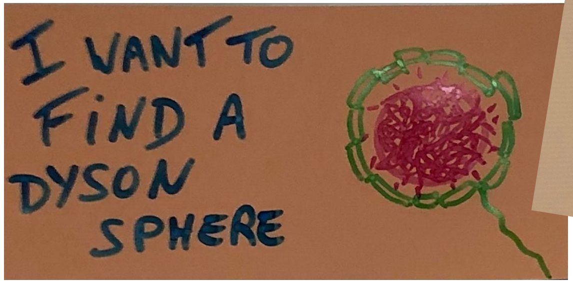
Personalized ML for big data

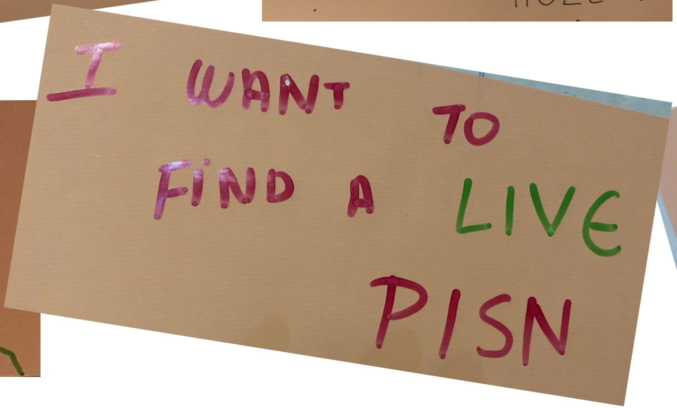
#FinkDreamShots

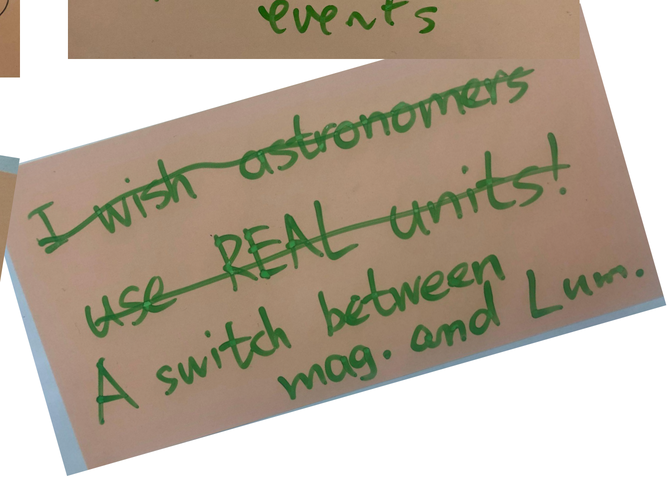












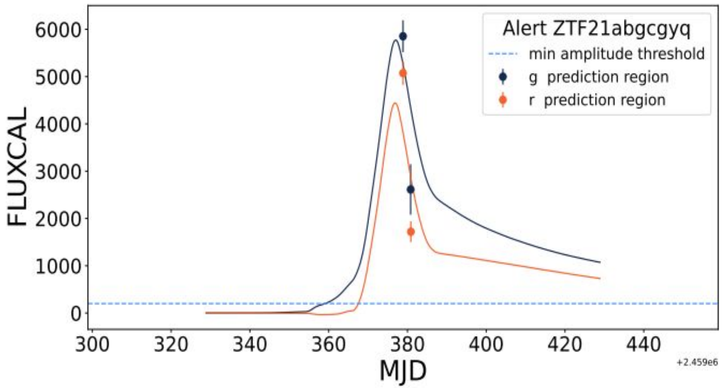
From OzFink 2023 - Melbourne, Australia - https://www.ozgrav.org/ozfink-workshop-2023.html

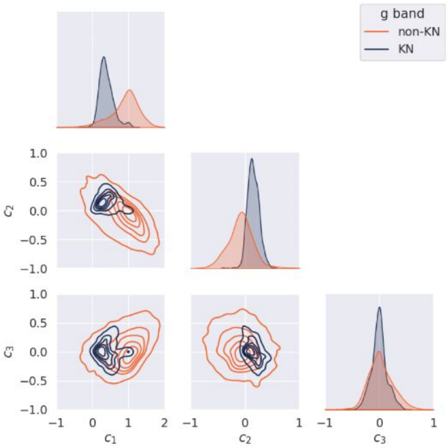
What do you want to see?

Case study: Kilonova

Problem 1: *<u>there are no labels</u>, only 1 confirmed detection- with a GW counterpart*

Problem 2: *we need to find it <u>fast</u>*





https://fink-portal.org/ZTF21abgcgyq

<u>Data set:</u>
Simulated ZTF light curves

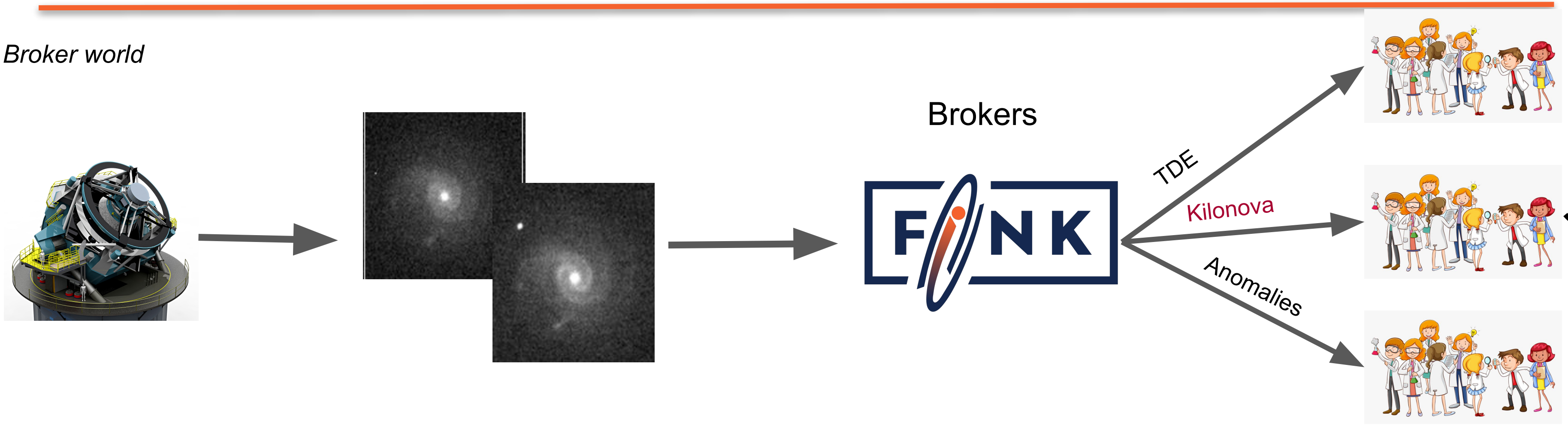
<u>Feature extraction:</u>
Principal components from perfect sims

<u>Classifier:</u>
Random Forest

Biswas et al., submitted to A&A, [arXiv:astro-ph/2210.17433](https://arxiv.org/abs/2210.17433)

Case study: **Kilonova**

Broker world



GRANDMA Observations of ZTF/Fink Transients during Summer 2021

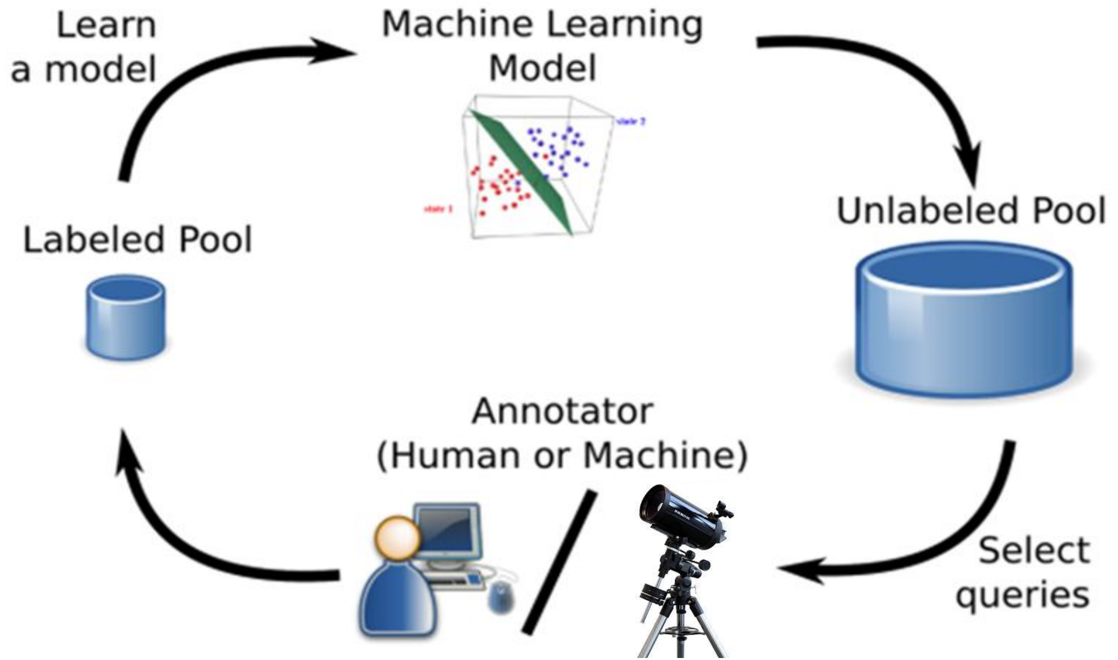
Aivazyan et al., 2021, arxiv:astro-ph/2202.09766

- 35 million candidate alerts
- 100 surviving selection cuts
- 6 followed-up by GRANDMA

Extra imaging from professional and amateur astronomers

Case study: Early SN Ia classification

Active Learning
Optimal experiment design

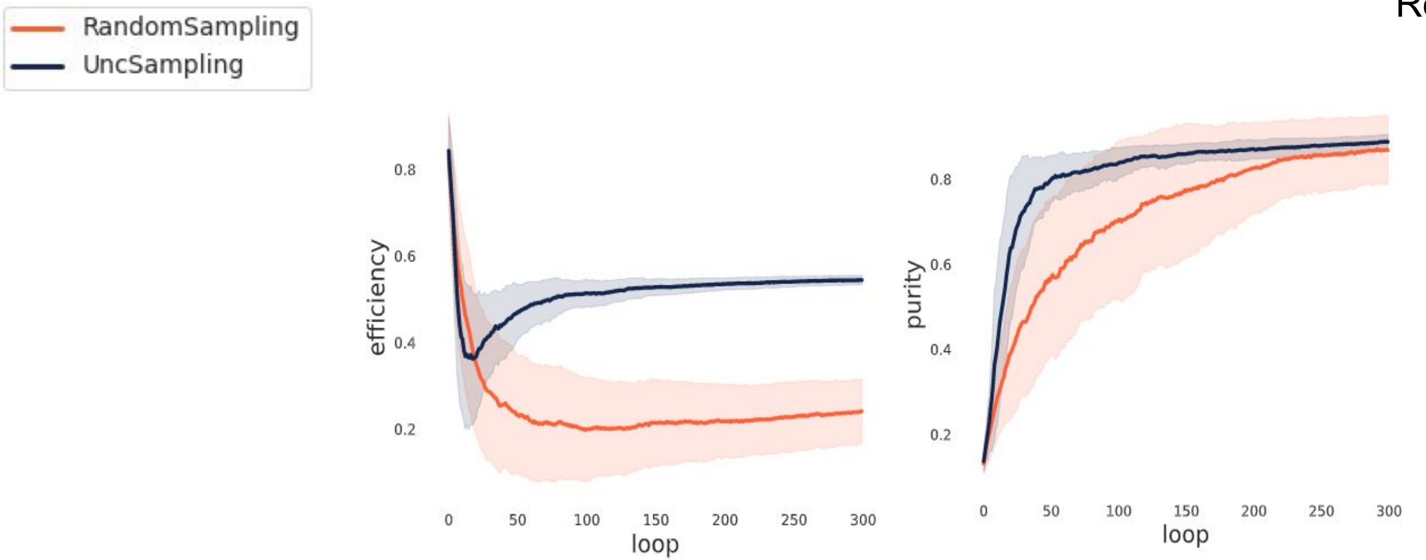


Plot modified from [Chowdhury et al., 2021, SPIE Medical Imaging]

Case study: Early SN Ia classification

Results after 300 loops:

Training: 310 alerts
Testing: > 52 000 alerts



Choose training sample which lead to better results and train a Random Forest classifier ...

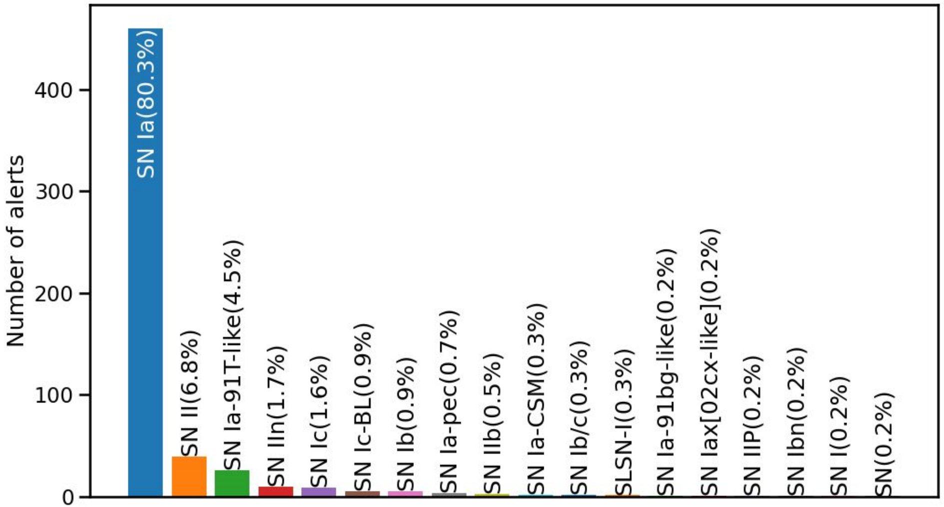
Trained ML model

Leoni et al., arxiv:2111.11438, in press A&A

Case study: Early SN Ia classification

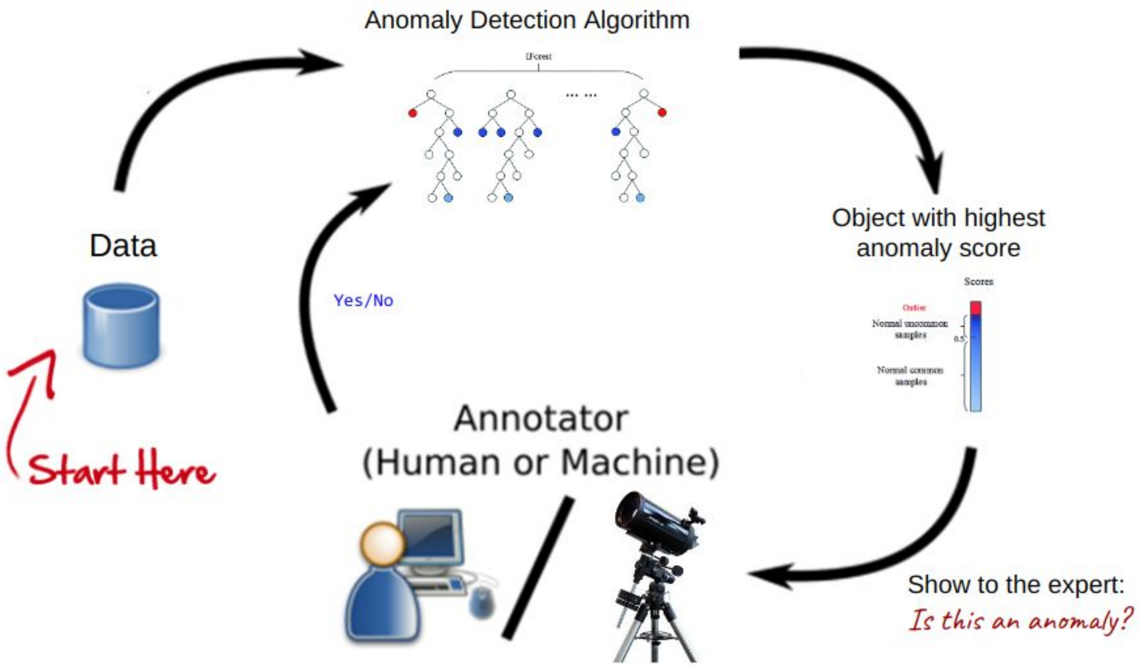
Fink Early SN Ia candidates reported to TNS from November/2020 - March/2022:

- 4 661 Early SN Ia candidates
- 573 spectroscopically classified
- Contaminants are mostly other SNe
  - 1 LBV



Plot by Julien Peloton (CNRS/IJCLab)

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]

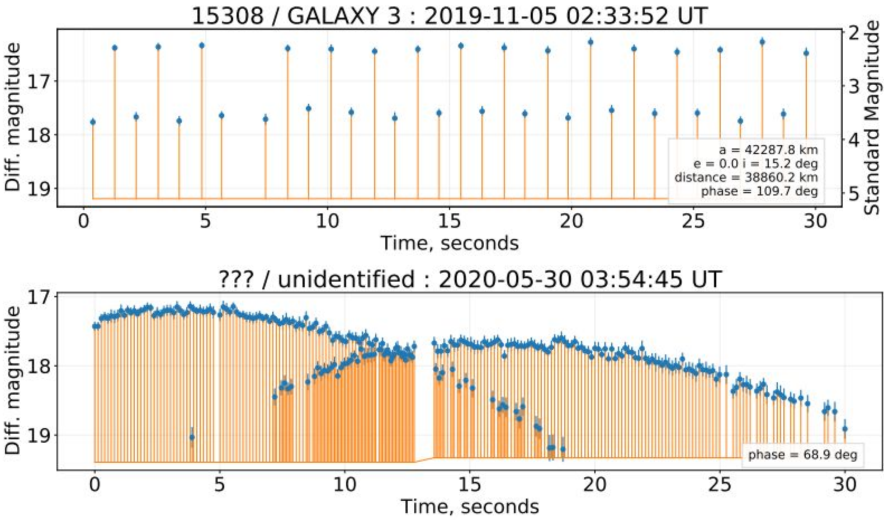
Case study: Satellite tracks

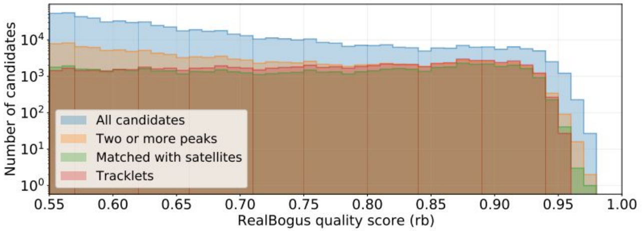
***Problem 1**: they hide in plein sight, labels must evolve*

Not mega-constellations!

***Problem 2**: they move fast and may confuse difference image analysis*

- Module to identify satellite glints
- 11.5 % all single-frame events
- 30% of those with real-bogus > 0.8
- 140 per night





Karpov and Peloton, 2021, arXiv:astro-ph/2202.05719