

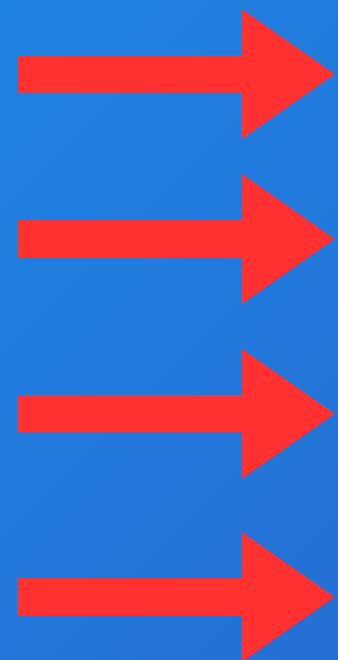


**SUPERVISED LEARNING IN ASTROPHYSICS:
RANDOM FOREST IN PRACTICE**

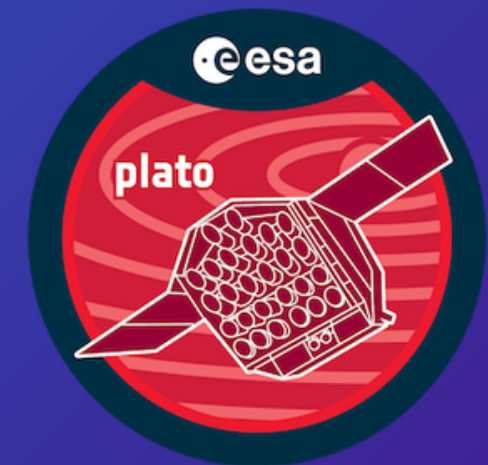
Maximiliano Blanco-Python Meeting-Instituto de Astrofísica
UC

DATA DOMAIN ASTRONOMY

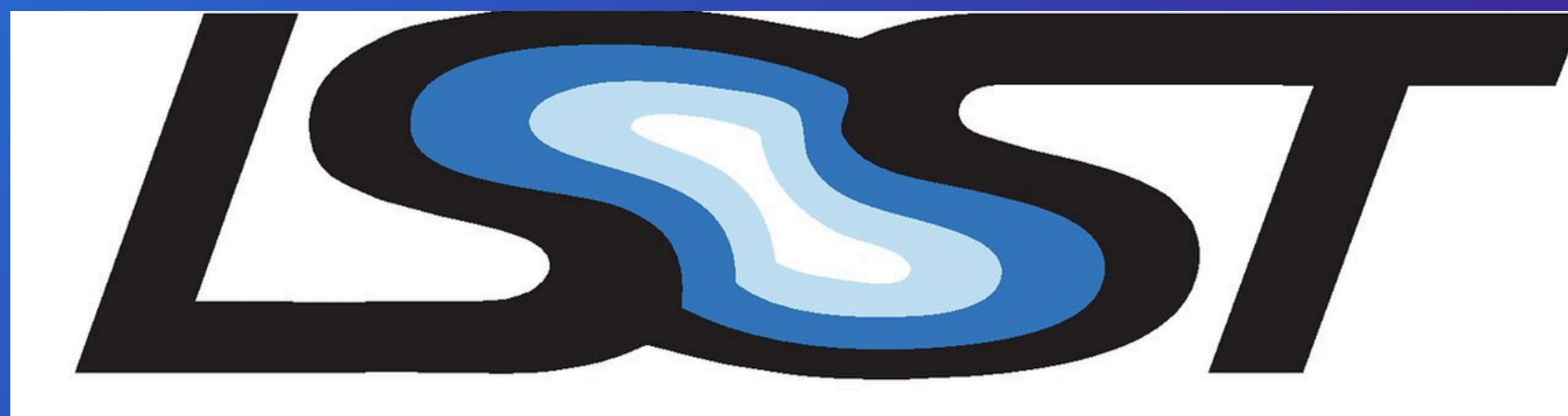
- We are entering in golden age for big surveys and new data.
- New facilities and surveys observing periodically a great portion of the sky will generate substantial archives of data.



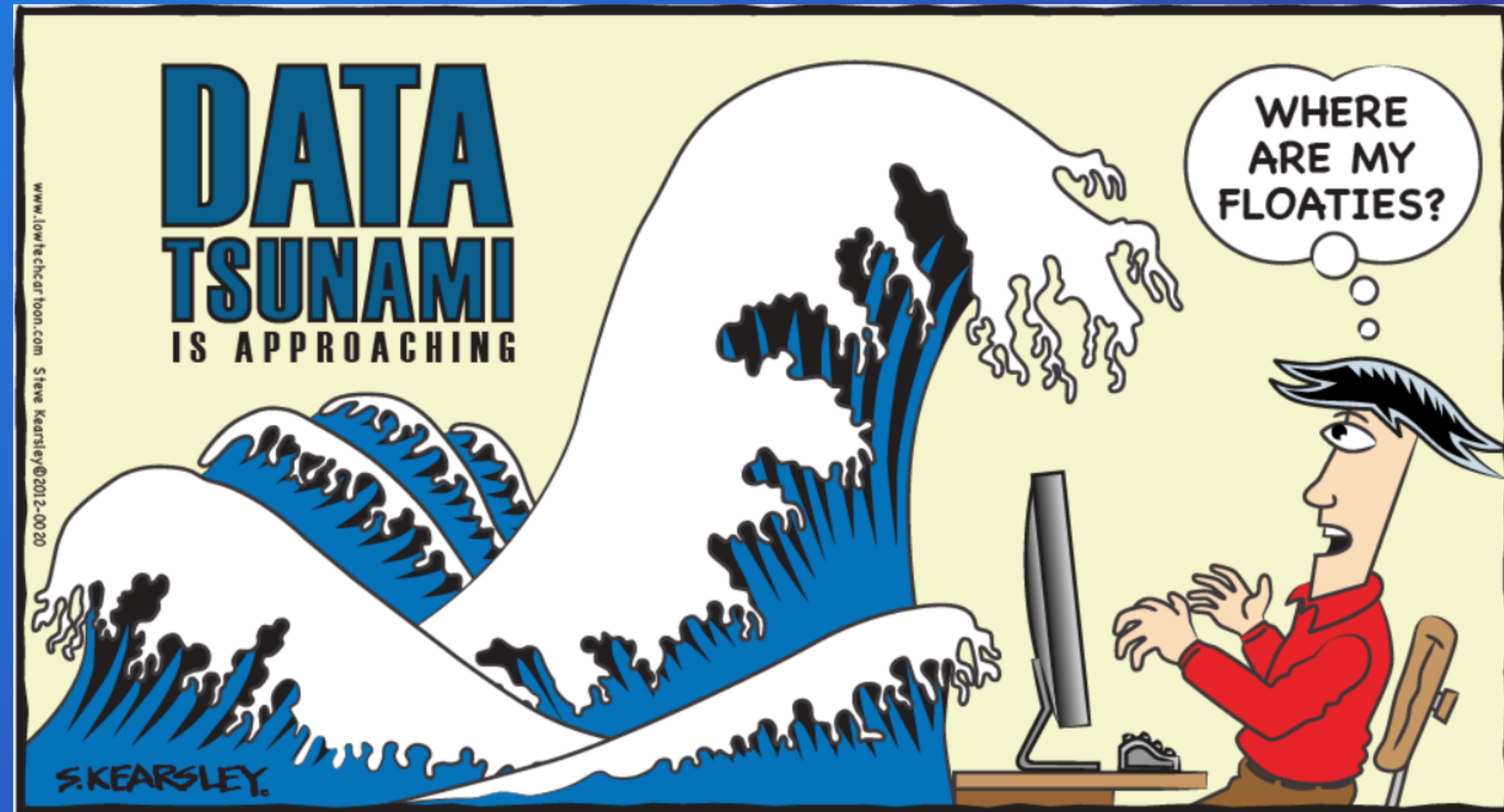
ZTF
GAIA
LSST
PLATO



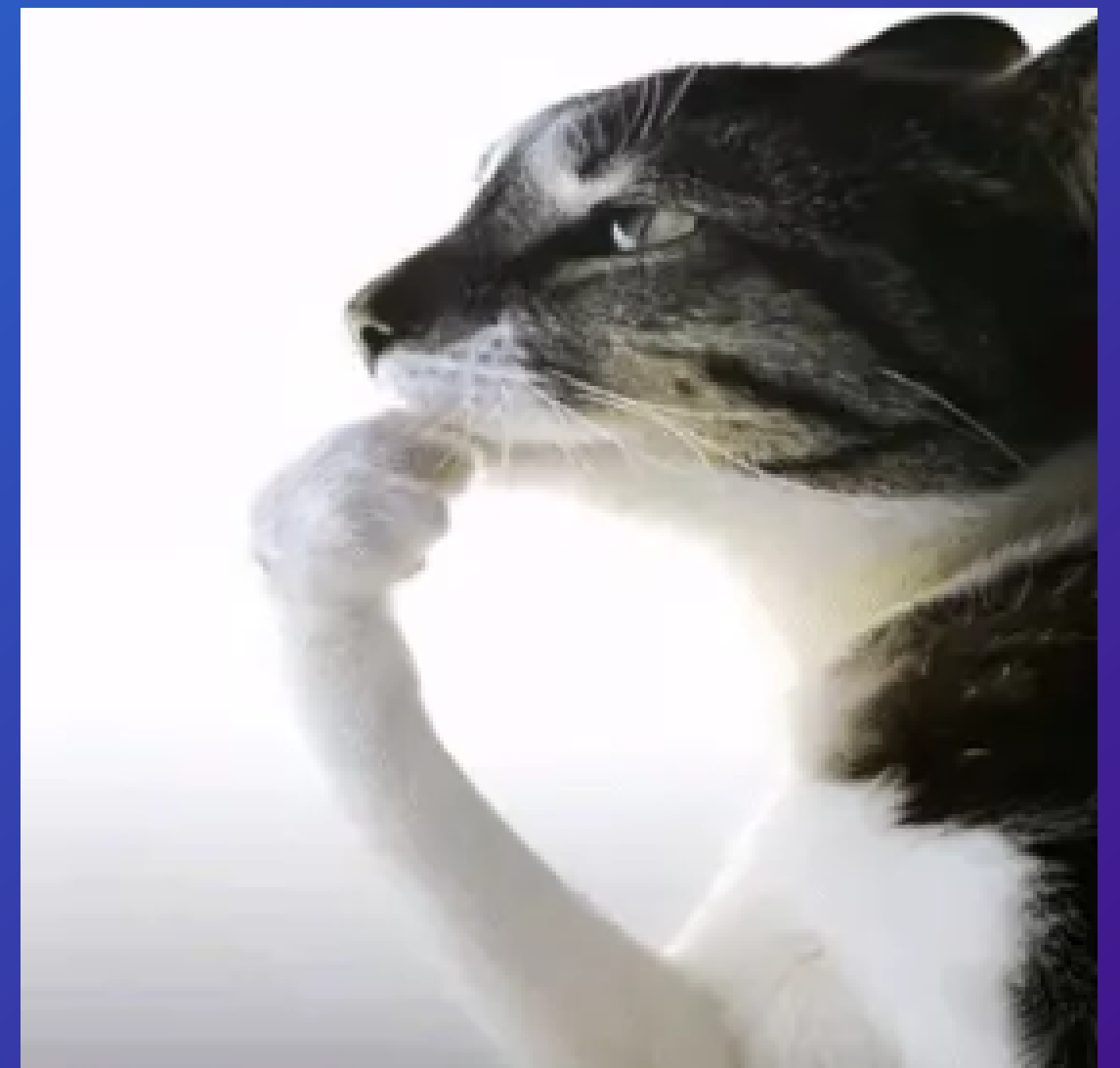
- ZTF and LSST will observe 1~40 billion objects. Trillions of measurements for these objects.
- LSST will produce 10 Terabytes of data each night!
- 500 petabytes generated by Rubin data processing in 10 years.



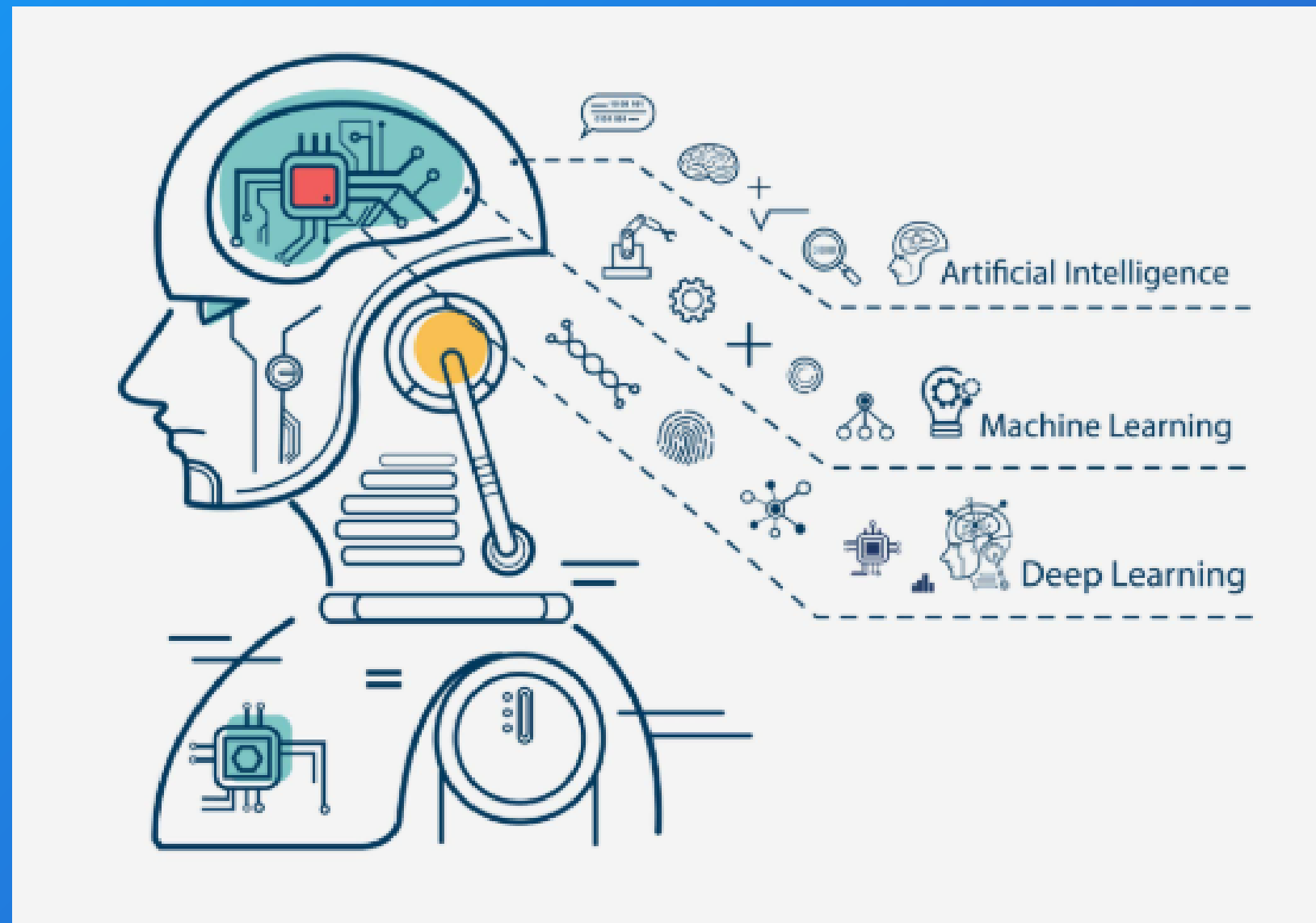
- ZTF and LSST will observe 1~40 billions objects. Trillions of measurements for this objects.
- LSST will produce 10 Terabytes of data each night!
- 500 petabytes generated by Rubin data processing in 10 years.



HOW DO WE OVERCOME THIS?

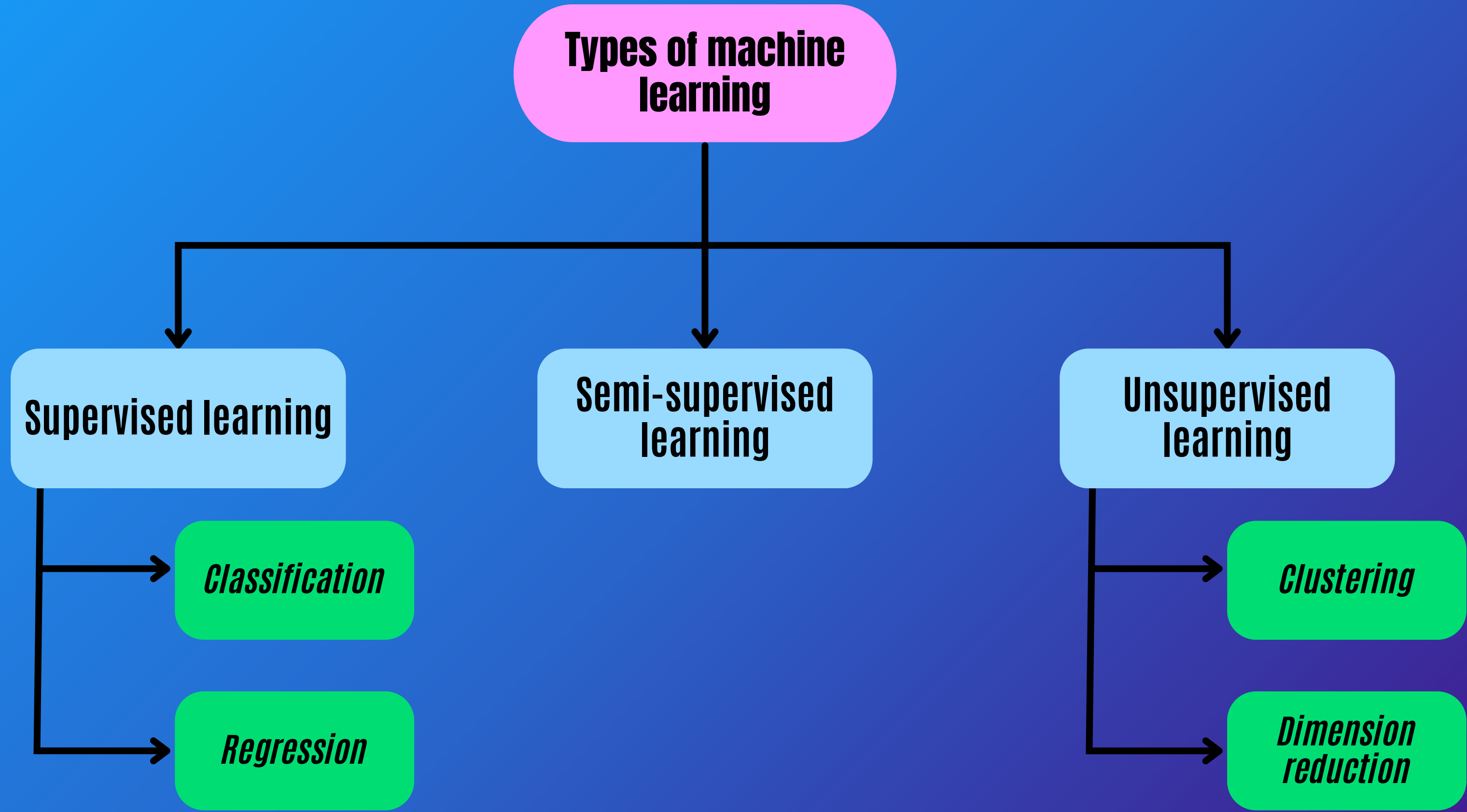


MACHINE LEARNING AS A SOLUTION

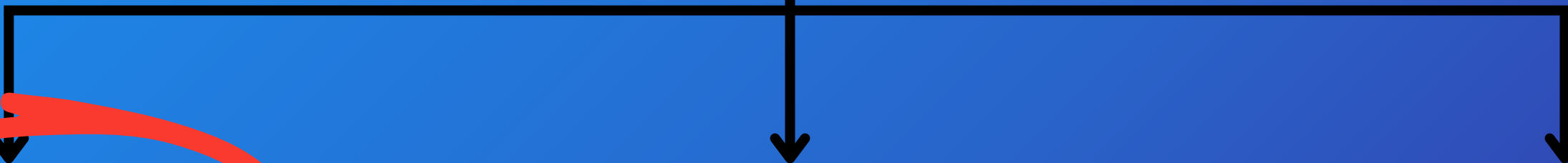


What is machine learning?

A subfield of computing science focused on building algorithms that can “learn” from data.



Types of machine learning



Supervised learning

Classification

Regression

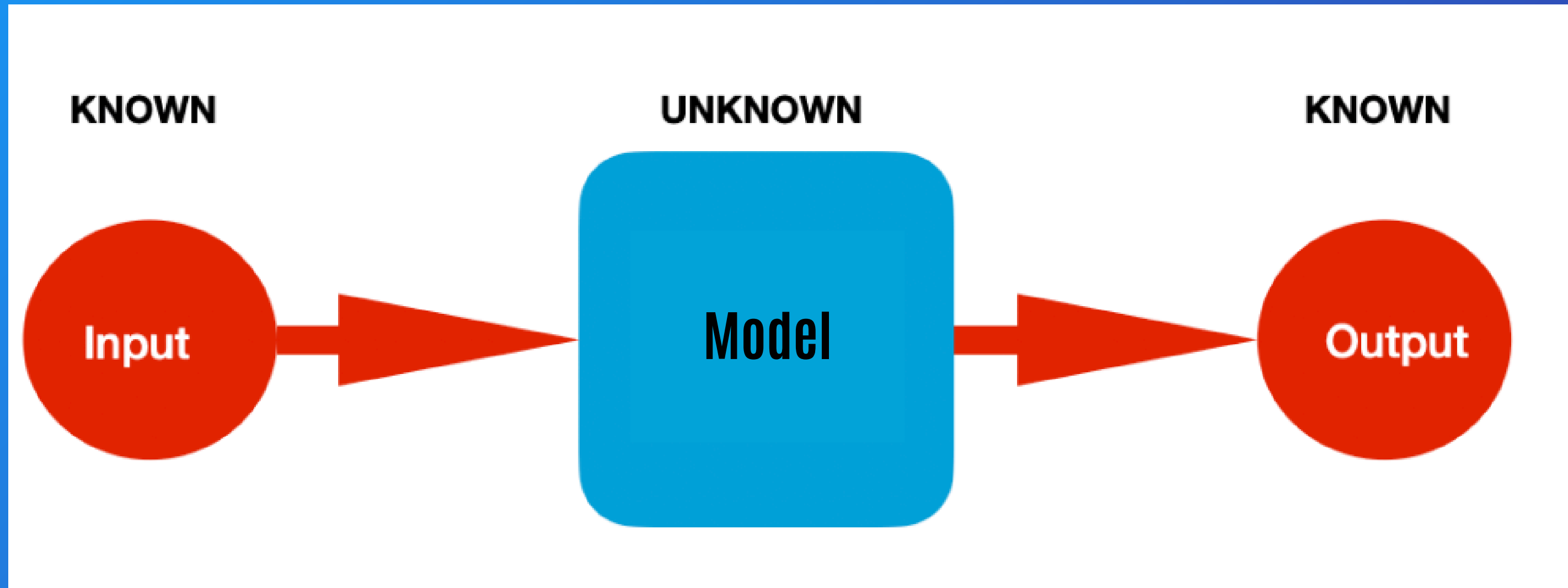
Semi-supervised learning

Unsupervised learning

Clustering

Dimension reduction

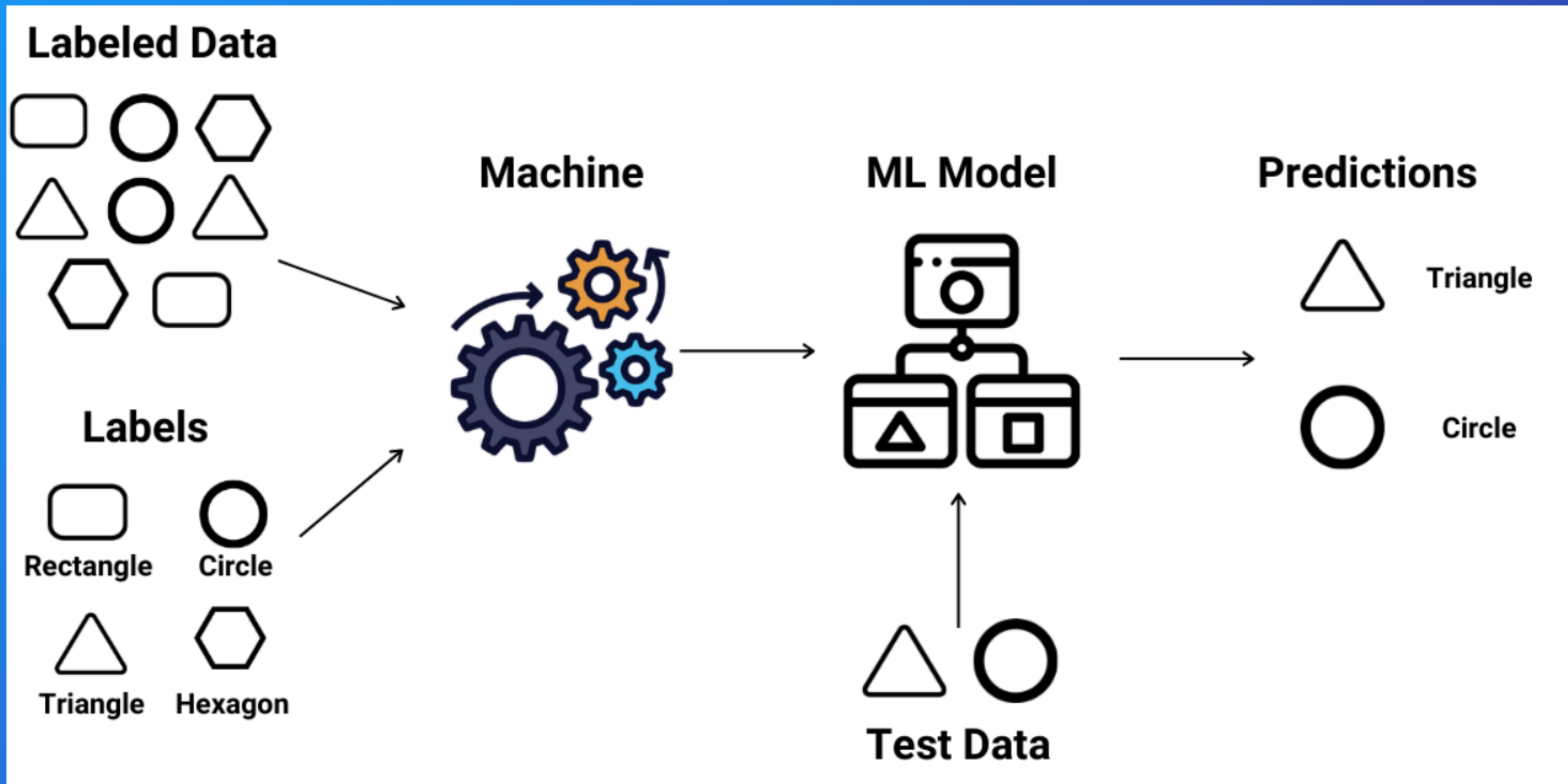
SUPERVISED LEARNING



<https://hshan0103.medium.com/what-is-supervised-semi-supervised-and-unsupervised-learning-12cd8160f99a>

- The data used to train the model is labeled.  We know both input and outputs.
- The algorithm learns how to map between the input and correct outputs.
- The downside is that having a big set of data labeled is complicated.

SUPERVISED LEARNING





























A&A, 693, A306 (2025)

<https://doi.org/10.1051/0004-6361/202451491>

© The Authors 2025

**Astronomy
&
Astrophysics**

Stellar atmospheric parameters and chemical abundances of ~5 million stars from S-PLUS multiband photometry




















C. E. Ferreira Lopes^{1,2,★}, L. A. Gutiérrez-Soto³, V. S. Ferreira Alberice^{4,5}, N. Monsalves⁶, D. Hazarika^{1,2},
M. Catelan^{7,8,2}, V. M. Placco⁹, G. Limberg¹⁰, F. Almeida-Fernandes^{5,11}, H. D. Perotoni⁵,
A. V. Smith Castelli^{3,12}, S. Akras¹³, J. Alonso-García^{14,2}, V. Cordeiro¹⁵, M. Jaque Arancibia^{6,16},
S. Daflon¹⁵, B. Dias¹⁷, D. R. Gonçalves¹¹, E. Machado-Pereira^{9,15}, A. R. Lopes³, C. R. Bom¹⁸,
R. C. Thom de Souza^{19,20}, N. G. de Isídio²¹, A. Alvarez-Candal^{22,23}, M. E. De Rossi^{24,25}, C. J. Bonatto²⁶,
B. Cubillos Palma²⁷, M. Borges Fernandes¹⁵, P. K. Humire⁵, G. B. Oliveira Schwarz^{4,5},
W. Schoenell²⁸, A. Kanaan²⁹, and C. Mendes de Oliveira⁵

(Affiliations can be found after the references)

Received 13 July 2024 / Accepted 18 November 2024



Alert Classification for the ALeRCE Broker System: The Light Curve Classifier

P. Sánchez-Sáez^{1,2,3} , I. Reyes^{1,4,5} , C. Valenzuela^{4,1,6,3} , F. Förster^{7,1,8} , S. Eyheramendy^{3,1} , F. Elorrieta^{7,1} ,
F. E. Bauer^{2,9,1,10} , G. Cabrera-Vives^{11,1} , P. A. Estévez^{5,1} , M. Catelan^{2,9,1} , G. Pignata^{12,1} , P. Huijse^{13,1} ,
D. De Cicco^{1,2} , P. Arévalo¹⁴ , R. Carrasco-Davis⁵ , J. Abril^{15,16}, R. Kurtev^{14,1} , J. Borissova^{14,1} , J. Arredondo¹ ,
E. Castillo-Navarrete^{1,4}, D. Rodríguez¹, D. Ruz-Mieres^{17,1,4}, A. Moya^{4,1} , L. Sabatini-Gacitúa^{4,1}, C. Sepúlveda-Cobo^{4,1}, and
E. Camacho-Iñiguez²

¹ Millennium Institute of Astrophysics (MAS), Nuncio Monseñor Sótero Sanz 100, Providencia, Santiago, Chile; pasanchezsaez@gmail.com

² Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile

³ Faculty of Engineering and Sciences, Universidad Adolfo Ibañez, Diagonal Las Torres 2700, Peñalolén, Santiago, Chile

⁴ Center for Mathematical Modeling, Universidad de Chile, Beauchef 851, North building, 7th floor, Santiago 8320000, Chile

⁵ Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, Santiago 8320000, Chile

⁶ Data Observatory Foundation, Santiago, Chile

⁷ Departamento de Matemáticas, Facultad de Ciencia, Universidad de Santiago de Chile, Av. Libertador Bernardo O'Higgins 3663. Estación Central, Santiago, Chile

⁸ Departamento de Astronomía, Universidad de Chile, Casilla 36D, Santiago, Chile

⁹ Centro de Astroingeniería, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile

¹⁰ Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, CO 80301, USA

¹¹ Department of Computer Science, University of Concepción, Edmundo Larenas 219, Concepción, Chile

¹² Departamento de Ciencias Físicas, Universidad Andrés Bello, Avda. República 252, Santiago, Chile

¹³ Informatics Institute, Universidad Austral de Chile, General Lagos 2086, Valdivia, Chile

¹⁴ Instituto de Física y Astronomía, Facultad de Ciencias, Universidad de Valparaíso, Gran Bretaña No. 1111, Playa Ancha, Valparaíso, Chile

¹⁵ European Southern Observatory (ESO), Alonso de Córdova 3107, Vitacura, Santiago, Chile

¹⁶ Centro de Estudios de Física del Cosmos de Aragón (CEFCA)—Unidad Asociada al CSIC, Plaza San Juan, 1, E-44001, Teruel, Spain

¹⁷ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

Received 2020 May 14; revised 2020 November 10; accepted 2020 November 11; published 2021 February 23

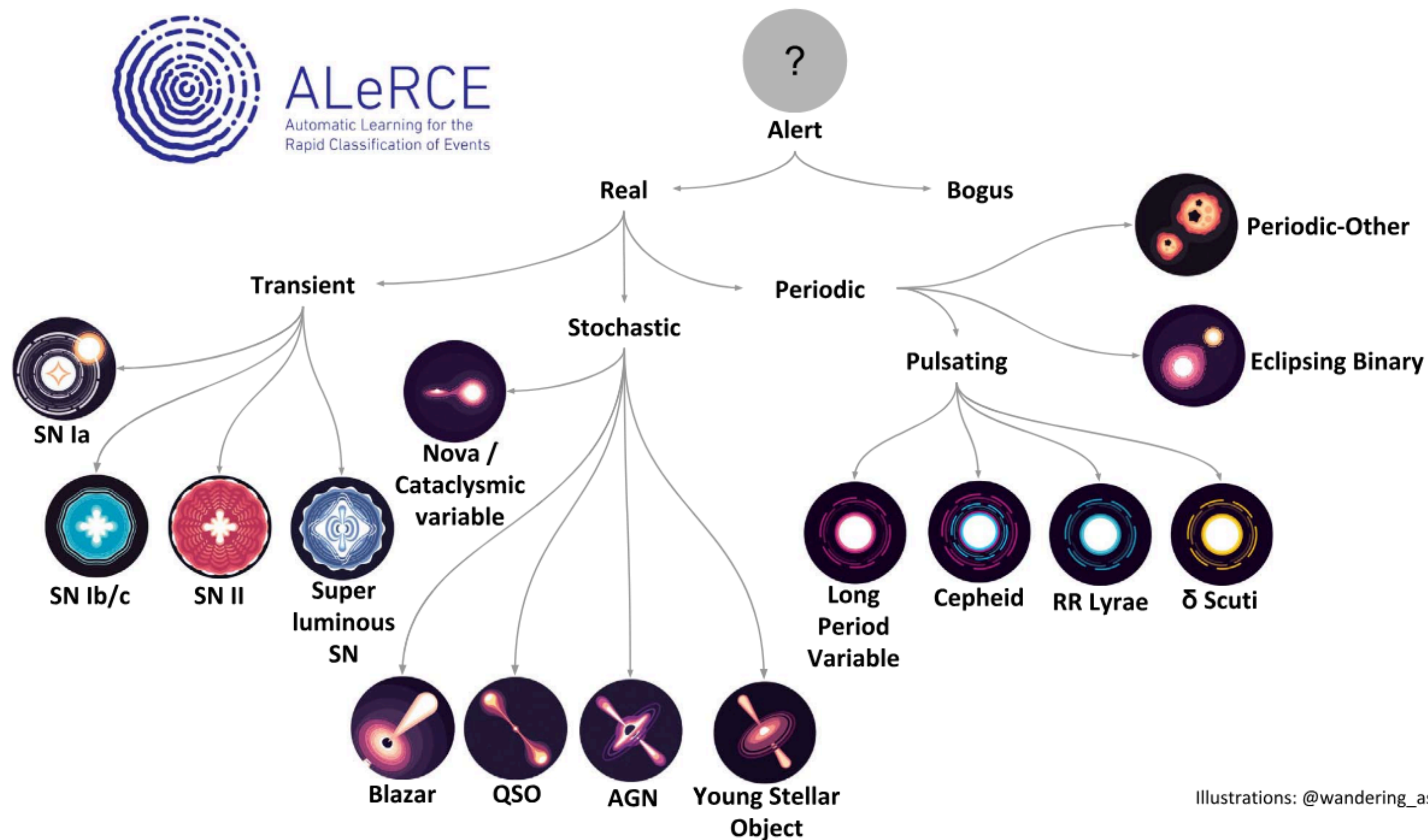
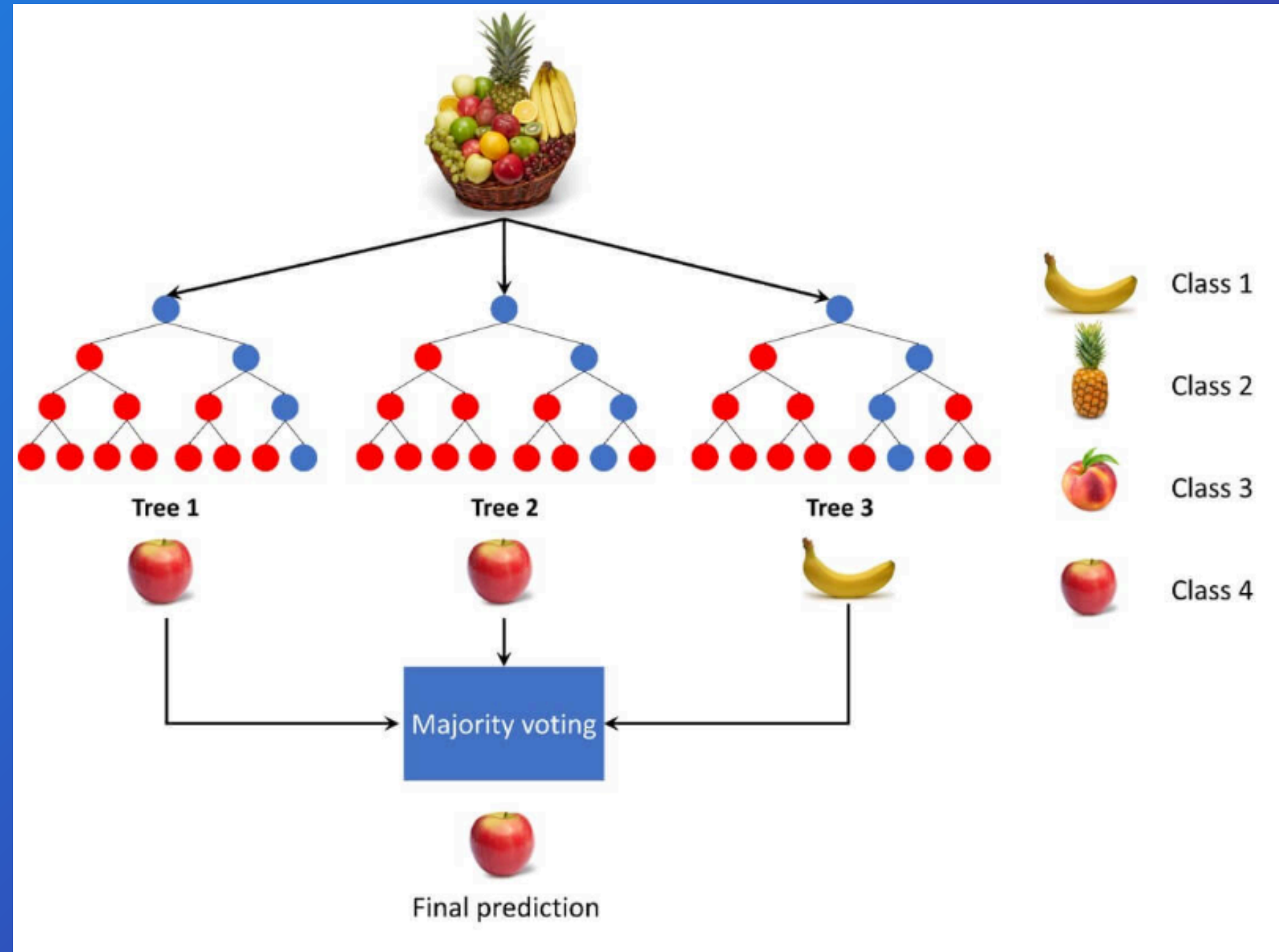


Figure 2. Taxonomy tree used in the current version of the ALeRCE light curve classifier.

RANDOM FOREST

- Goal: Learn a mapping between the input data and the prediction output.
- Build many different decision trees and combine them.
- Each tree is trained on a random subset and features of the data (bootstrap sampling).
- Each tree makes a classification → the forest outputs the majority vote.



<https://datahacker.rs/012-machine-learning-introduction-to-random-forest/>

TO MUCH TALKING



TO MUCH TALKING



LETS DO THE FUN PART!

