Data Science in Earth Observation

AI4EO Hackathon - Regression

WS 2024/25

Andrés Camero

Andres.CameroUnzueta@dlr.de

Acting Co-Leader

Department EO Data Science, DLR

Agenda

Part 1

1. Learn about the Helmholtz AI CountMeIn challenge,
2. Meet the data,
3. Explore a random forest regressor-based solution, and
4. Propose a simple artificial neural network-based solution.

Part 2

5. We will use data augmentation to improve the performance, and
6. Explore data fusion to benefit from all data sources.


SUSTAINABLE DEVELOPMENT GOALS
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS

DLR TUM

55% of World's population lives in urban areas, and by 2030 it will increase to 66%

United Nations, 2016. The world's cities in 2016. http://www.un.org/

Census

Expensive

U$ 15.6B spent in US-2020 census

Manipulation

Political pressure postponed census in Madagascar since 2003

Not real time

Since 1932 no national census in Lebanon

Incomplete

Puerto Rico's census respone rate in 2010 was 53.8%

Availability

Official population data is often available only at administrative level

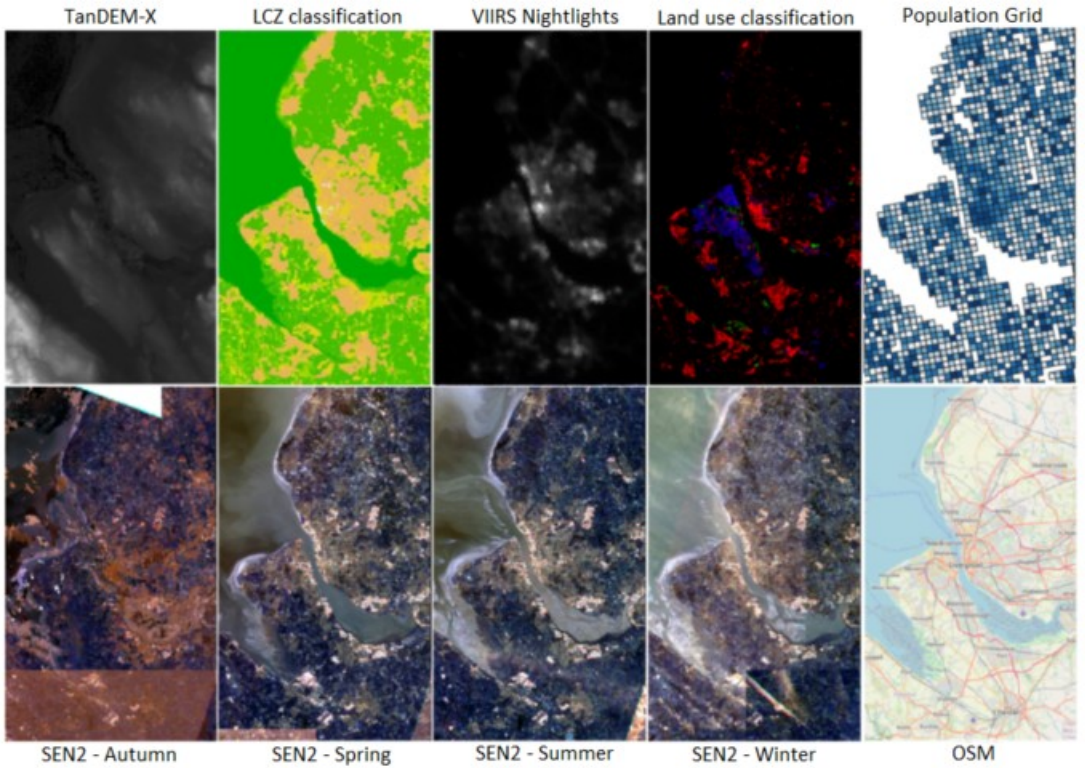
Population estimation is important since it helps people make decisions about the future, e.g., where to invest in education, health, urbanization.

DLR

So2Sat POP

98

European cities



6

Data sources

Doda, S., Wang, Y., Kahl, M., Hoffmann, E.J., Taubenböck, H. and Zhu, X.X., 2022. So2Sat POP--A Curated Benchmark Data Set for Population Estimation from Space on a Continental Scale. arXiv preprint arXiv:2204.08524.

Helmholtz AI CountMeIn

GOGREEN

USE AS LITTLE ENERGY AS POSSIBLE!

Given the CountMeIn problem and a maximum RMSE of 1111, participants should propose a solution that minimizes the environmental impact (while it achieves the target performance). Participants in this track need to use the HAICORE resources at Karlsruhe Institute of Technology (KIT) as the Computer HoreKa features a very accurate power measurement facilities. The impact will be computed using HAICORE logs, and your submission will include the Job ID.

GOFAST

USE AS LITTLE TIME AS POSSIBLE!

Given the CountMeIn problem and a RMSE of 1111, participants should propose a solution that minimizes the training and prediction time. Participants in this track can choose to use the HAICORE resources at KIT and the HAICORE resources at Forschungszentrum Jülich (FZJ). At KIT, a maximum of 56 A100 GPUs can be used, while at JUWELS Booster at FZJ all 3744 A100 GPUs are available. For the submission, the compute center, as well as the Job ID must be included.

HELMHOLTZAI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

DLR JÜLICH Forschungszentrum KIT Karlsruher Institut für Technologie

DLR

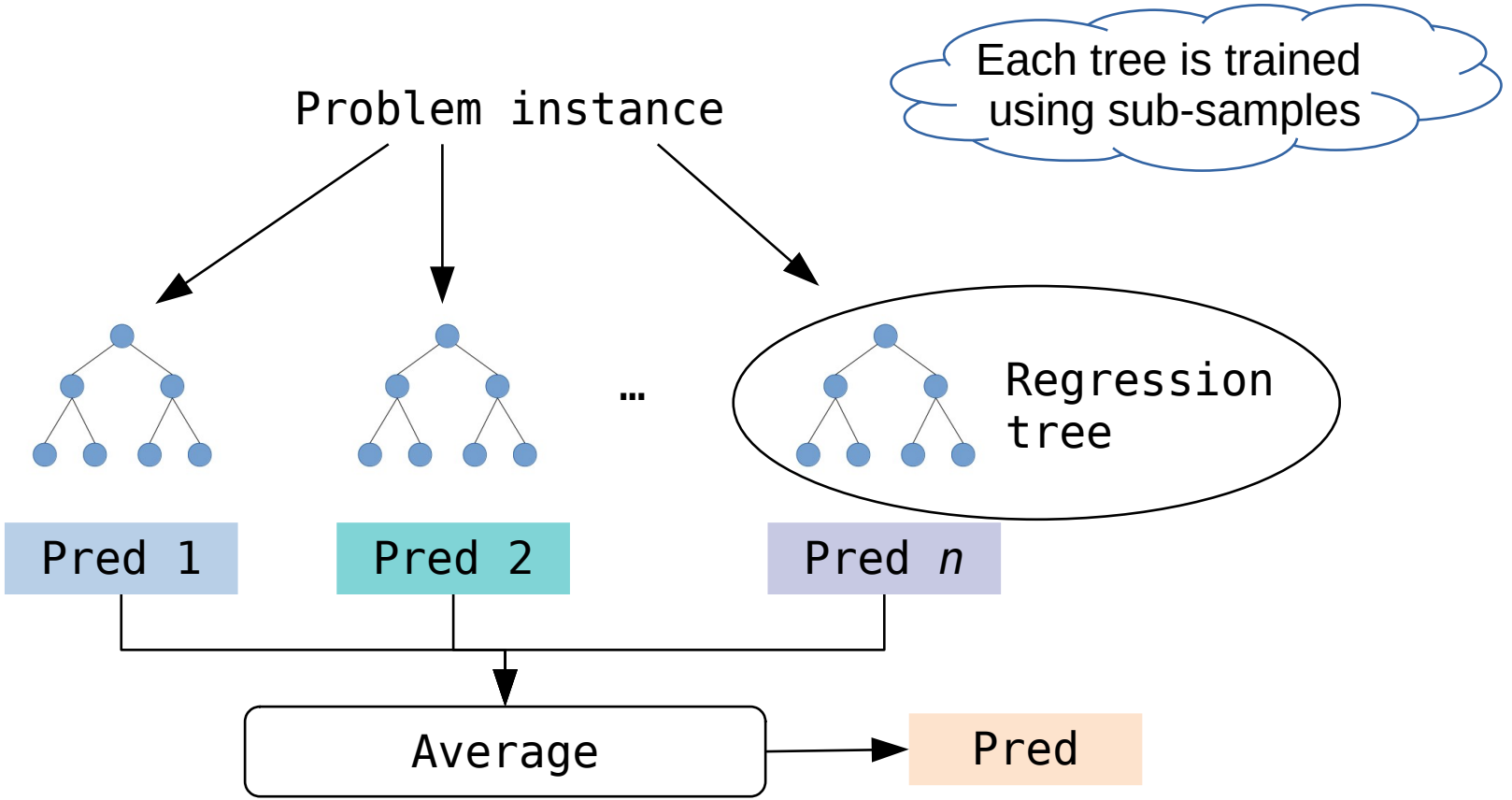
https://github.com/acamero/data-science-eo-regression

Machine learning

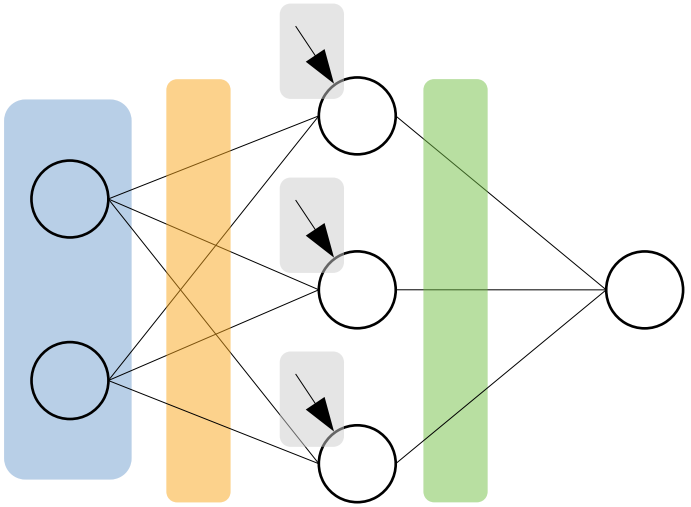
A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Tom M. Michell, 1997

Random forest regressor



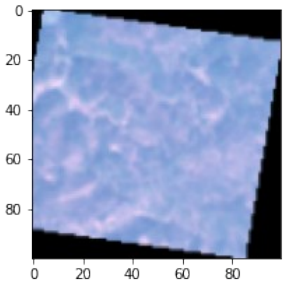
Artificial neural networks



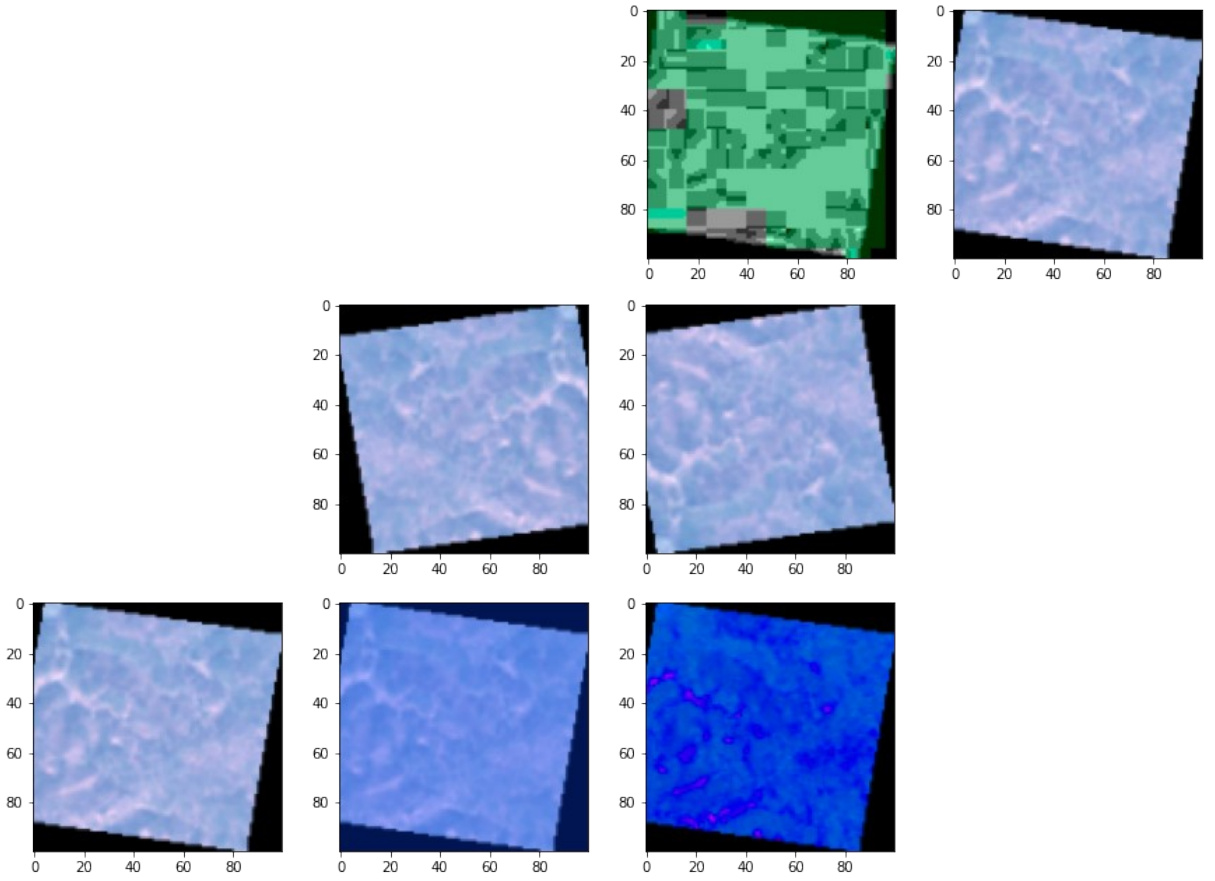
$$\hat{y} = f\left(W \cdot x + b\right)$$

activation function

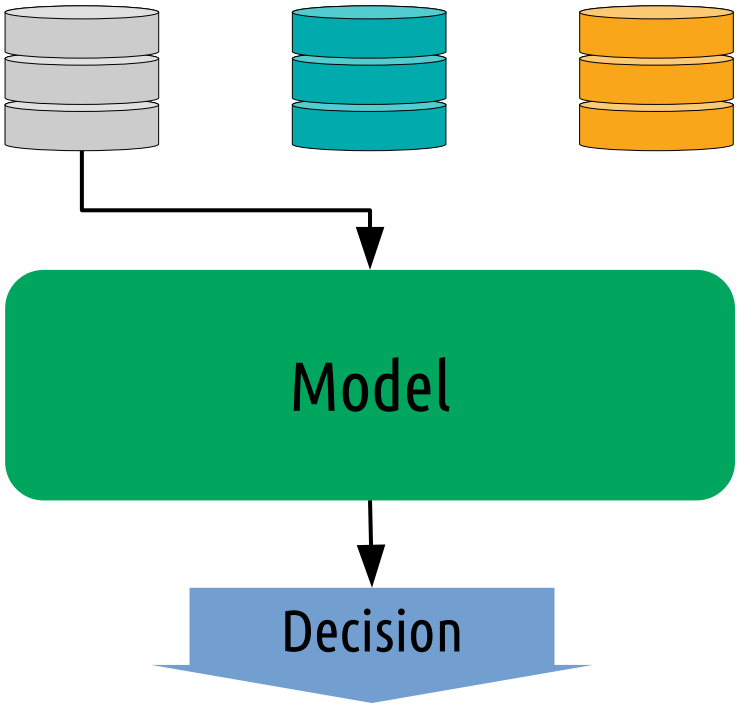
Data augmentation



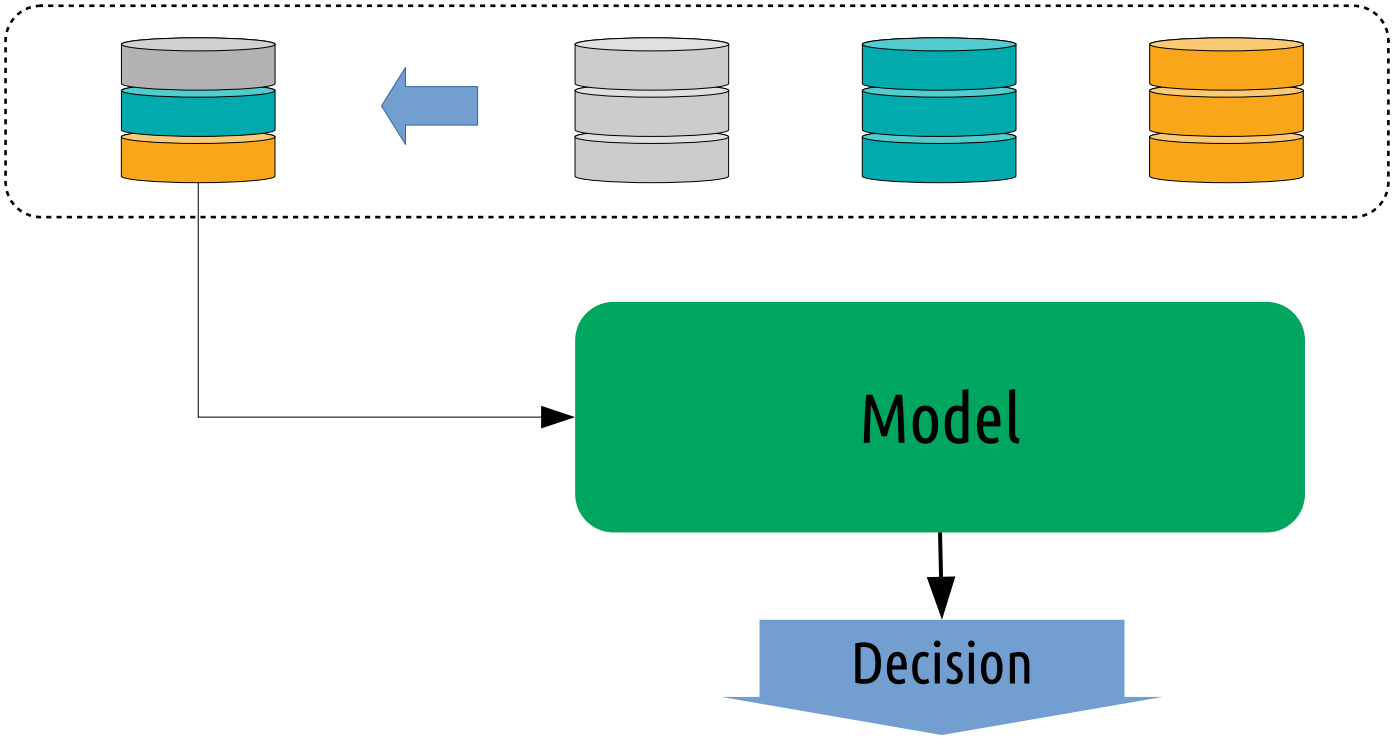
Original image



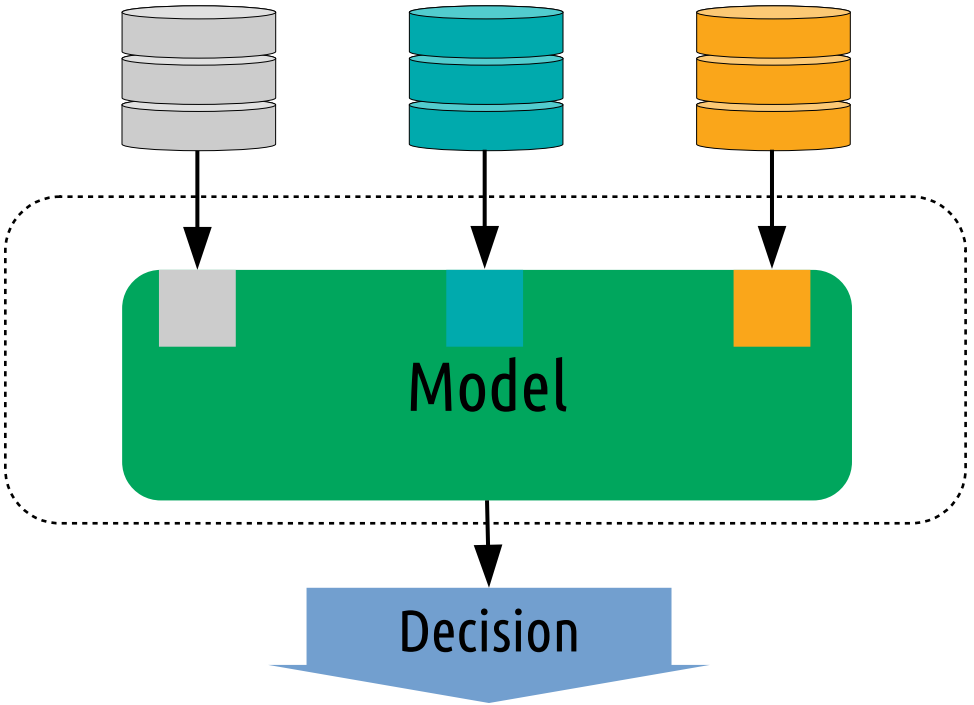
Data fusion



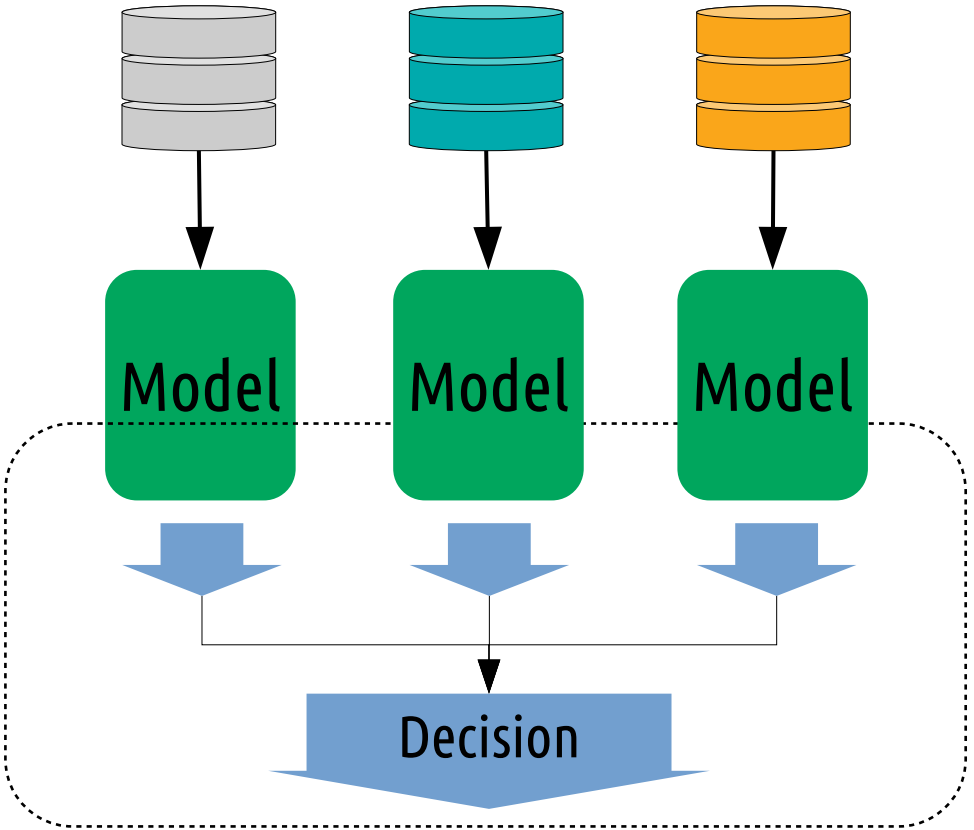
Data fusion



Data fusion



Data fusion



References

1. Doda, S., Wang, Y., Kahl, M., Hoffmann, E.J., Taubenböck, H. and Zhu, X.X., 2022. So2Sat POP--A Curated Benchmark Data Set for Population Estimation from Space on a Continental Scale. arXiv preprint arXiv:2204.08524.
2. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.
3. McCarthy, J., 2007. What is artificial intelligence?
4. Mohri, M., Rostamizadeh, A. and Talwalkar, A., 2018. Foundations of machine learning. MIT press.
5. Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F. and Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), pp.8-36.