UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

Machine Learning

Session 24 - T

Multi-task and Multi-label Learning

Ciência de Dados Aplicada

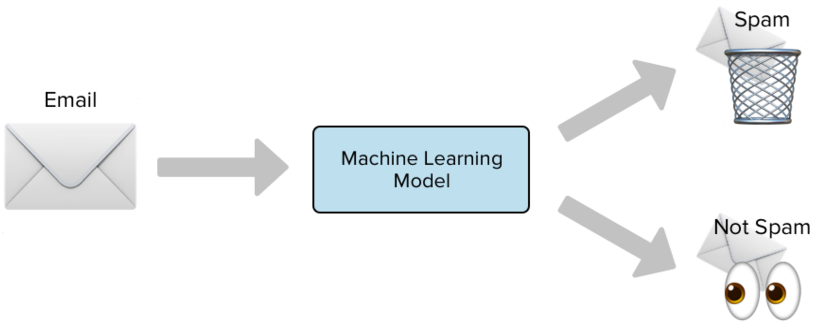
2023/2024

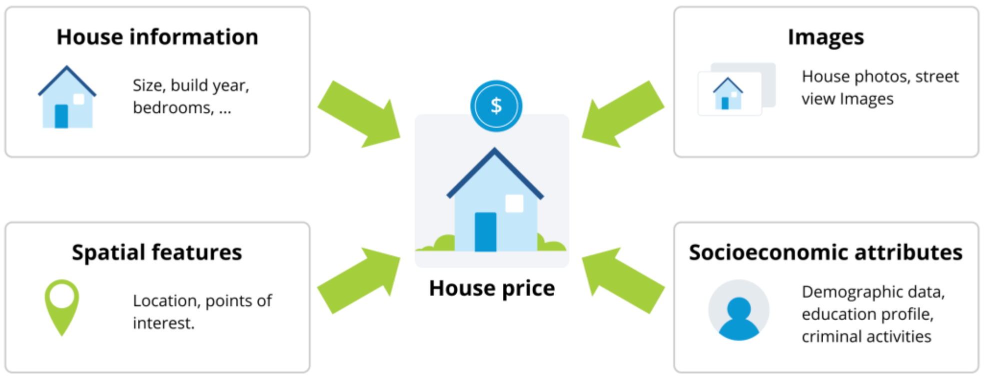
Concept of Tasks in Machine Learning

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- A task in machine learning is a specific **objective that the model aims to achieve**, such as **classifying images** or **predicting prices;**

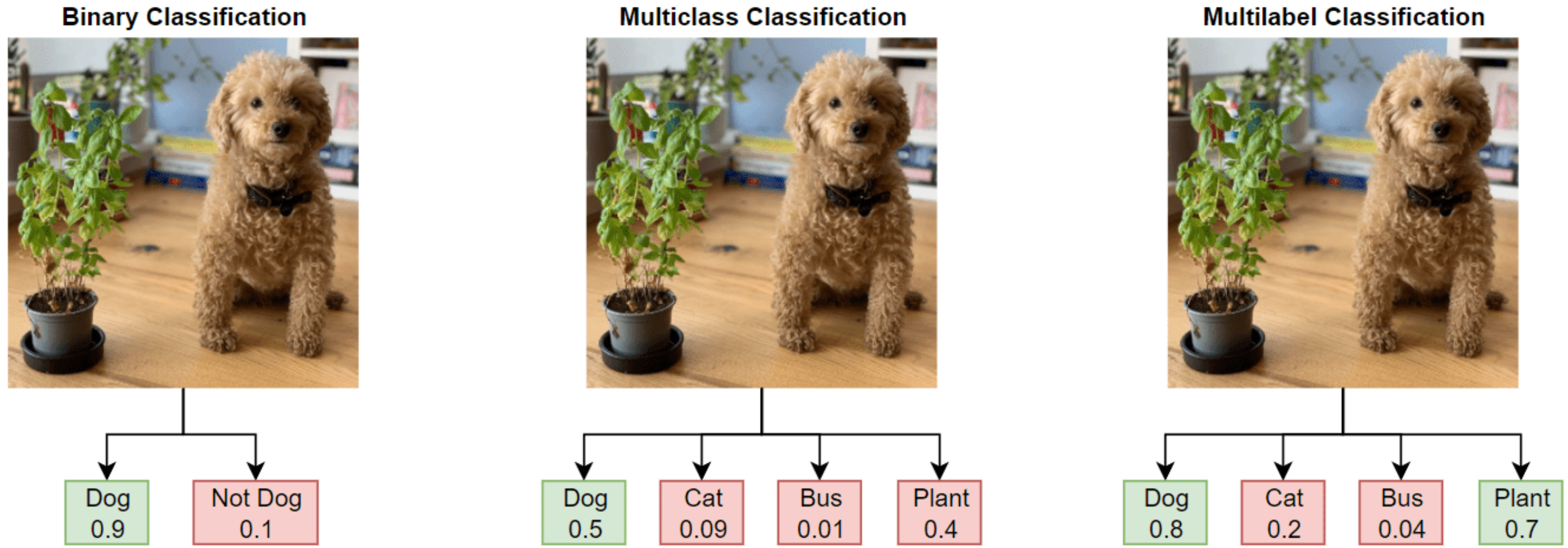
- Examples of Tasks:
  - **Classification** (e.g., image classification)
  - **Regression** (e.g., predicting house prices)





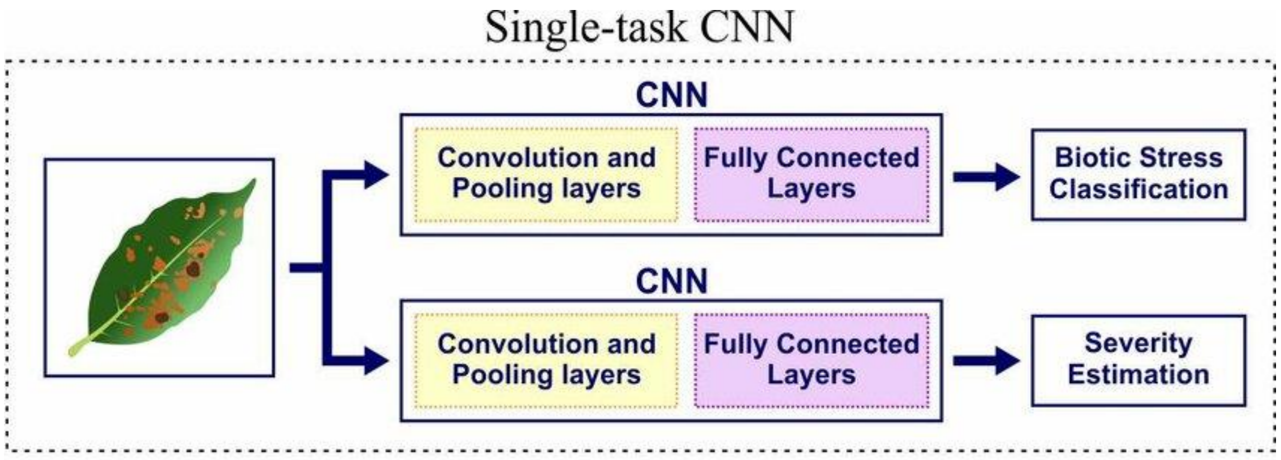
Concept of Label in Machine Learning

- The **output** or result associated with an input, can be **one or multiple per task**;
- Examples:
  - **Single label**: Classifying a image as a dog;
  - **Multi-label**: Classifying a image as both a dog and a plant.



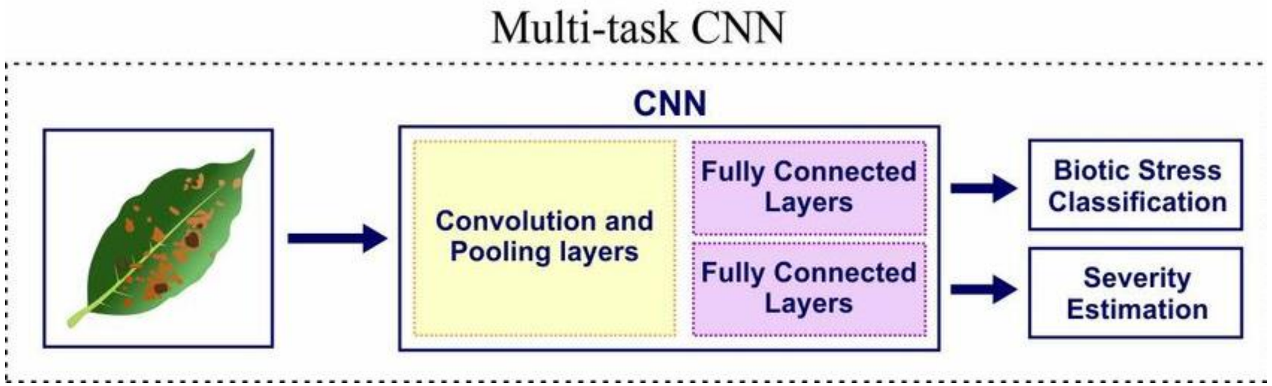
Single-Task Learning

- Models are trained to perform **one task at a time.**



Multi-Task Learning

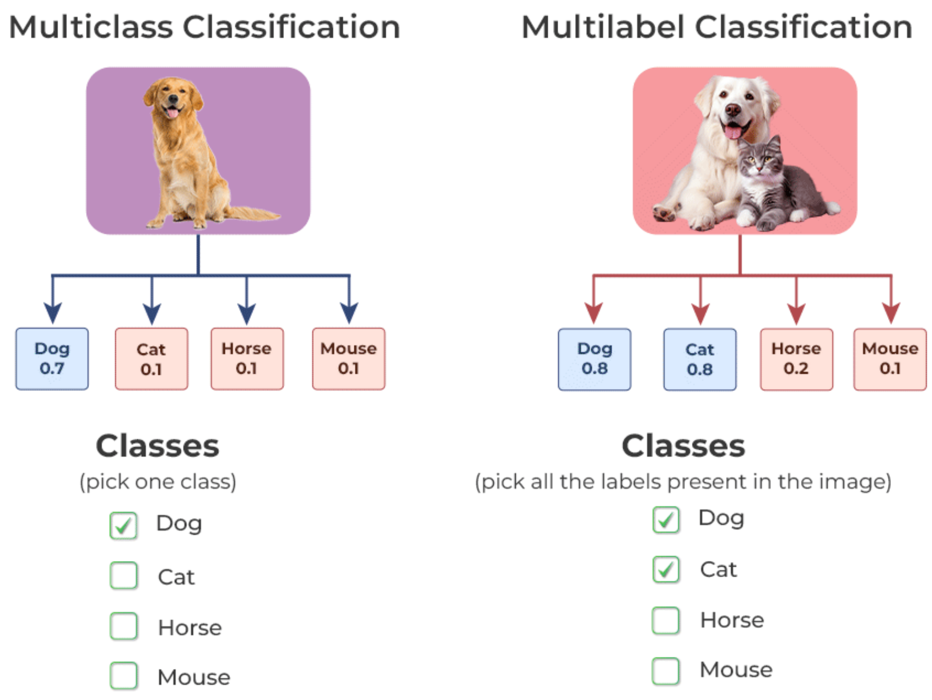
- An approach where a model **learns multiple tasks simultaneously**, sharing representations;
- Benefits:
  - **Improved generalization**;
  - **Efficiency** in learning;
  - **Shared information** among tasks.



Multi-Label Learning

- A **single task** where each instance can have **multiple labels;**

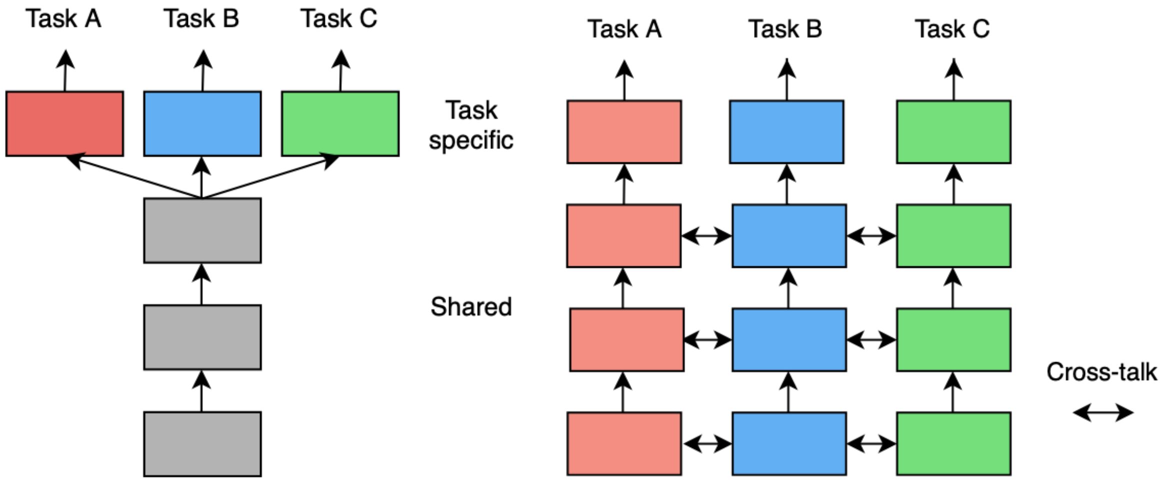
- Benefits:
  - Captures more **complex relationships** in data;
  - Reflects **real-world scenarios** where items bolong to multiple categories.



Types of Multi-Task Learning

- **Hard parameter sharing**: shared hidden layers with **task-specific output layers**;

- **Soft parameter sharing**: each task has its parameters but **regularization** is used to keep them **similar**;



(a) Hard parameter sharing (b) Soft parameter sharing

Resources

- Crawshaw, M. (2020). Multi-Task Learning with Deep Neural Networks: A Survey (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2009.09796

- Tarekegn, A. N., Ullah, M., & Cheikh, F. A. (2024). Deep Learning for Multi-Label Learning: A Comprehensive Survey (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2401.16549

UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

Machine Learning

Session 24 - T

Automated Machine Learning

Ciência de Dados Aplicada

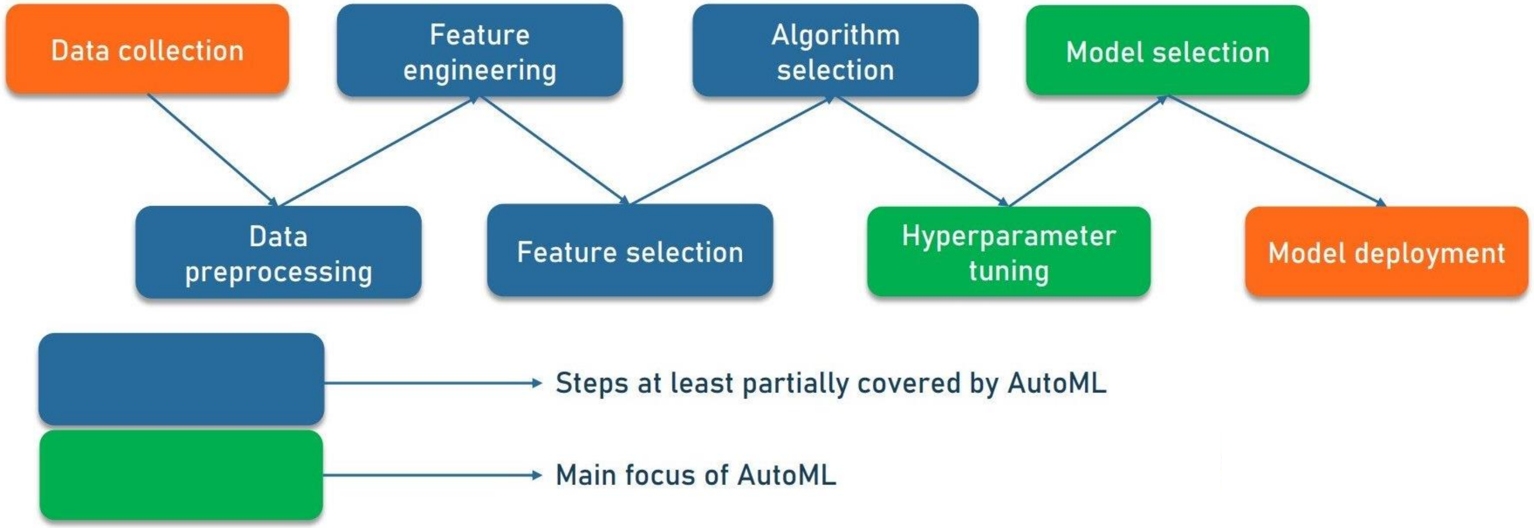
2023/2024

What is Automated Machine Learning (AutoML)?

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

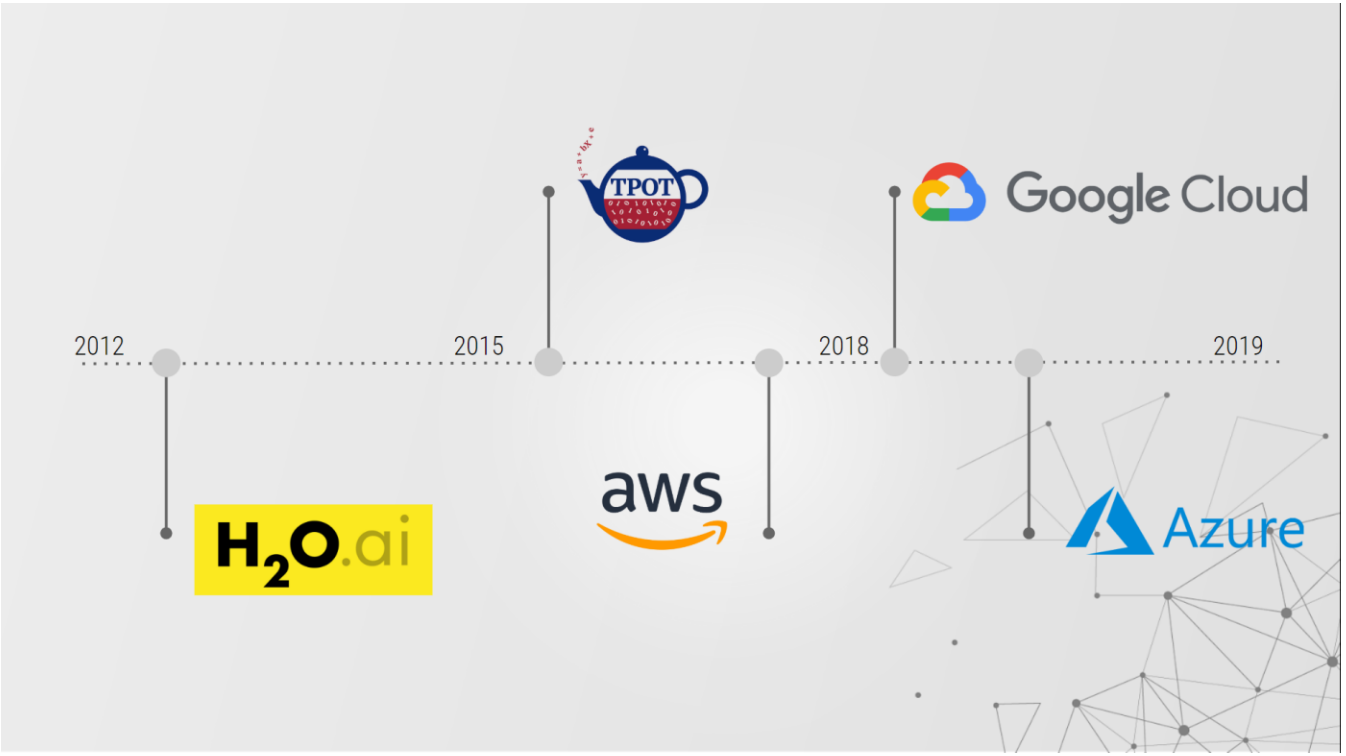
- AutoML is the process of automating the end-to-end process of applying machine learning to real-world problems;

- Simplify and speed up the development of machine learning models, making it accessible to non-experts.



AutoML – Tools and Frameworks

- Google AutoML
- H2O.ai
- Auto-sklearn
- TPOT
- Microsoft Azure AutoML
- …



Advanced Topics in AutoML

- **Neural Architecture Search** (NAS): Automating the design of neural network architectures.

- **Meta-Learning**: Learning how to learn; leveraging past experiences to improve future AutoML tasks.

- **Fairness and Ethics**: Addressing bias, transparency, and ethical considerations in automated systems.

Resources

- Automated Machine Learning. (2019). In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), The Springer Series on Challenges in Machine Learning. Springer International Publishing. https://doi.org/10.1007/978-3-030-05318-5

UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

Machine Learning

Session 24 - T

Model Deployment and Monitoring

Ciência de Dados Aplicada

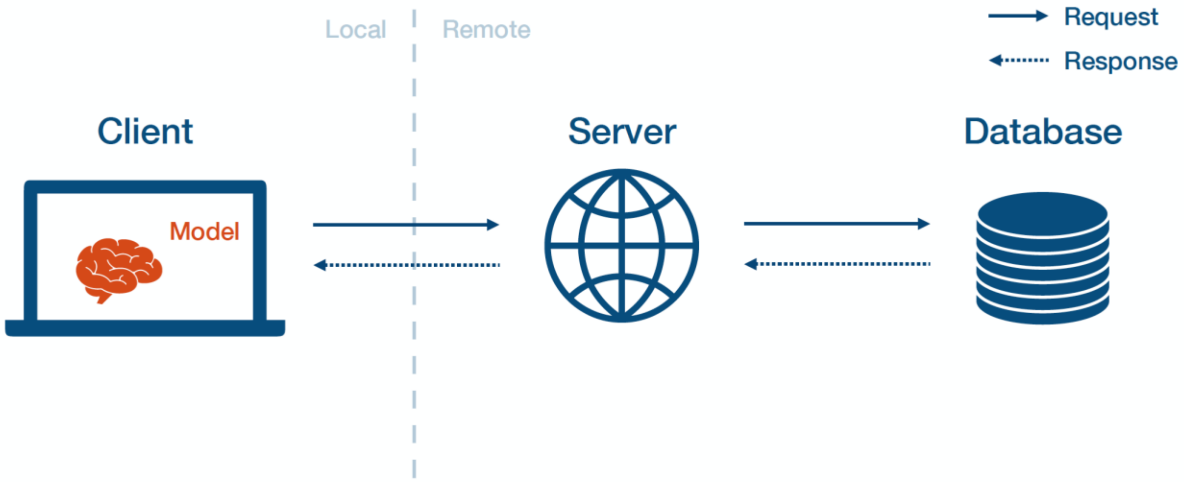
2023/2024

Model Deployment

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Model deployment is the process of making a machine learning model **available for use in production environments**;

- Models can be deployed in various environments, including on-premise **servers**, **cloud platforms**, and **edge devices**. Each scenario comes with its own challenges and considerations.



Model Optimization for Deployment

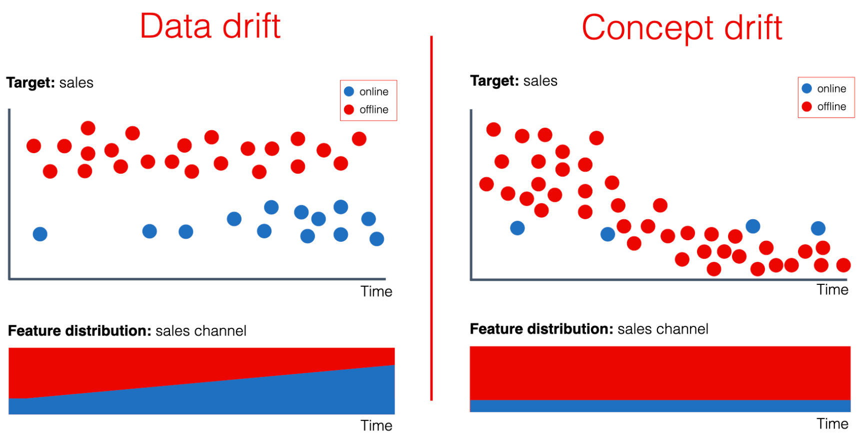
UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- **Model Compression Techniques**: To improve deployment efficiency, models can be compressed using techniques like **quantization**, **pruning**, and **knowledge distillation**, reducing their size and computational complexity.

- **Latency and Throughput**: Optimizing models for **low latency** and **high throughput** is essential for real-time applications. Techniques such as **hardware acceleration** and **architectural optimizations** can help achieve these goals.

Model Monitoring

- Monitoring ensures that deployed models **perform as expected over time**;
- **Track metrics** like accuracy, latency, and throughput to detect performance issues;
- Monitor for **concept drift** (changes in the data distribution) and **data drift** (changes in data characteristics).

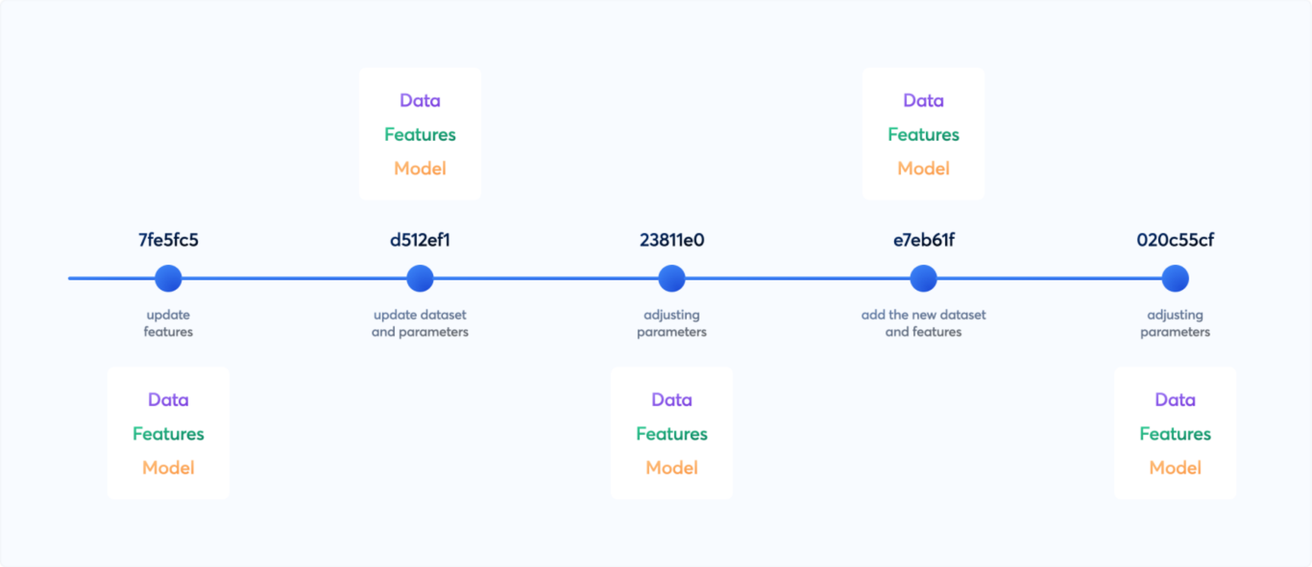


Model Versioning

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- **Managing Versions**: Keep track of different versions of deployed models to facilitate rollback if necessary.

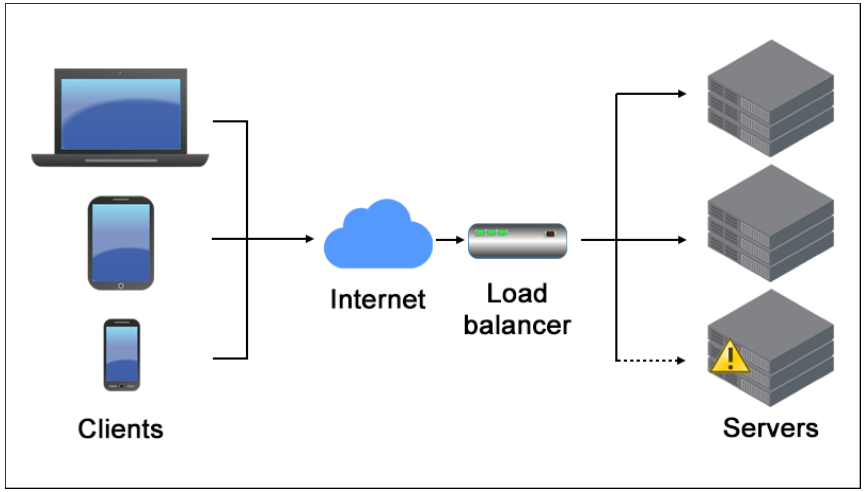
- **Rollback Strategies:** Plan for reverting to previous model versions in case of issues with new deployments.

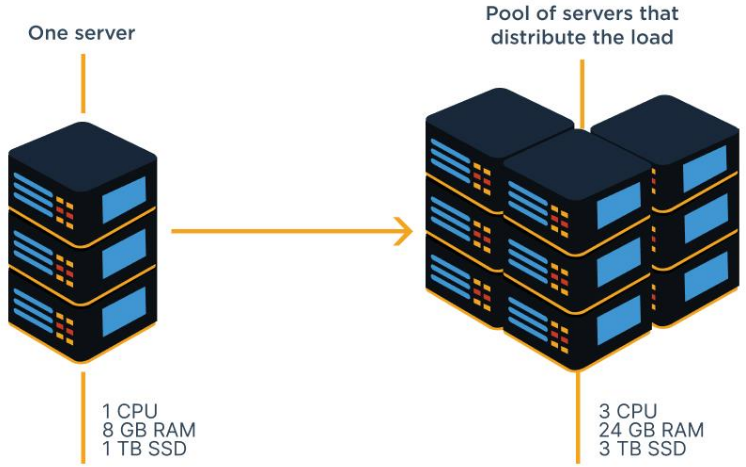


Scalability and Performance

- **Scaling Strategies**: Implement techniques like load balancing and horizontal scaling to handle increased demand;

- **Performance Optimization**: Optimize model inference speed and resource utilization for efficient deployment.





Resources

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Islam, J. (2022). Machine Learning Model Serving Patterns and Best Practices: A definitive guide to deploying, monitoring, and providing accessibility to ML models in production. Packt Publishing.