

# STRUMENTI DI ORCHESTRAZIONE E ANALISI DI WORKFLOW NEL MACHINE LEARNING

Analisi case study e illustrazione framework di orchestrazione di pipeline

Simone Boldrini

Alma Mater Studiorum - Università di Bologna  
Facoltà di Scienze

13 Ottobre 2021

**Obiettivo:** Ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio

**Obiettivo:** Ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio
- Analisi e confronto tra CS

**Obiettivo:** Ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio
- Analisi e confronto tra CS
- Workflow Generico

**Obiettivo:** Ricerca di un'astrazione di un'invariante di processo per i processi di Machine Learning

- Casi di studio
- Analisi e confronto tra CS
- Workflow Generico
- Framework

## Apprendimento Automatico

*L'**apprendimento automatico** raccoglie metodi in grado di migliorare la performance di un algoritmo nell'identificare pattern di dati, in modo autonomo.*

Gli algoritmi li suddividiamo in 3 categorie:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

# Apprendimento automatico nel Campo delle Reti

Insieme di problemi nel campo del Networking che permette l'integrazione con l'apprendimento automatico

## Supervised Learning

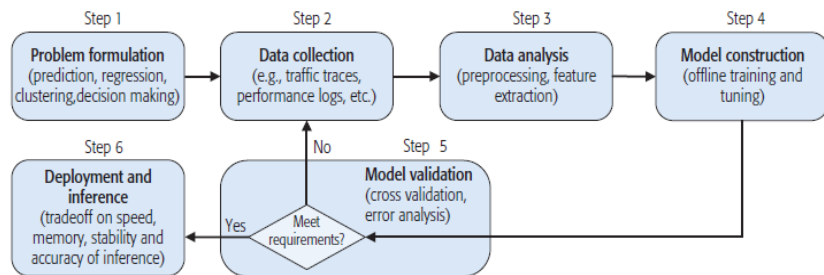
- Traffic Prediction
- Traffic Classification
- Routing Strategy
- Throughput prediction

## Unsupervised Learning

- Optimizing QoE

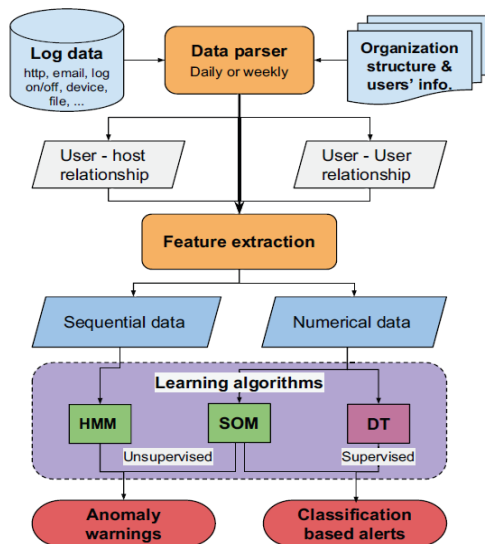
## Reinforcement Learning

- Resource Manager{Job Scheduler}
- TCP Congestion Control



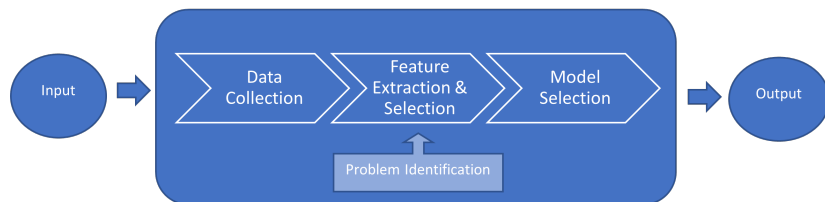


# Insider Threat Detection



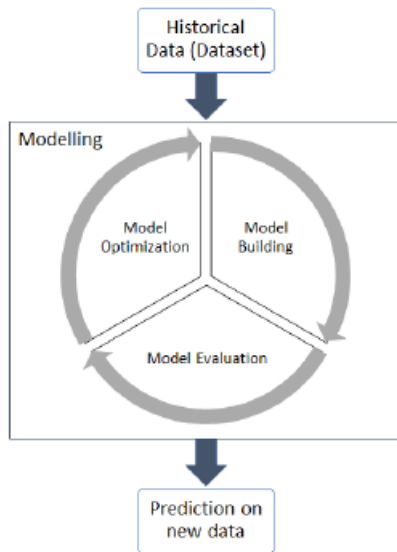
# Invariante di Processo

Analizzati i diversi casi di studio, abbiamo cercato di astrarre i problemi evidenziando un pattern generico di invariante.



La raccolta dati é fondamentale in ogni processo di apprendimento automatico. Viene diversificata a seconda del problema che andremo a trattare. Di solito i dati vengono suddivisi a seconda della **metodologia di raccolta**.

# Elaborazione del Modello



# Risultato

In quest'ultima fase definiamo l'output desiderato:

- Classificazione
- Regressione
- Clustering

## Classification

- Traffic Classification
- Image Recognition  
[Radiology]
- Speech Recognition
- Face2Face Traslation
- Classification based Alerts

## Regression

- Traffic Prediction
- Prediction/Forecasting

## Clustering

- Resource Management  
[Networking]
- Anomaly warning

## Strumenti di Orchestrazione

*Gli strumenti di orchestrazione di pipeline di ML hanno l'obiettivo di semplificare il processo di gestione e automatizzare l'implementazione dei modelli.*

Gli strumenti che andremo a mostrare di seguito sono tutti *open-source*, e si focalizzano su 3 punti chiave:

- raccolta dati
- creazione e implementazione del modello
- distribuzione

## ZenML

- Libreria Python
- Permette affiancamento ad altro strumento di orchestrazione
- Garantisce riproducibilità degli addestramenti
- Permette di memorizzare gli stati della pipeline nella cache

## Kedro

- GUI
- Modulare
- Favorisce il versioning

## Flyte

- Basato su python e K8s
- Estendibile attraverso diversi plug-in
- Gestisce piú di 10mila Workflow

## MLRun

- GUI
- Servizio Server-less
- Vasto reperto di plug-in