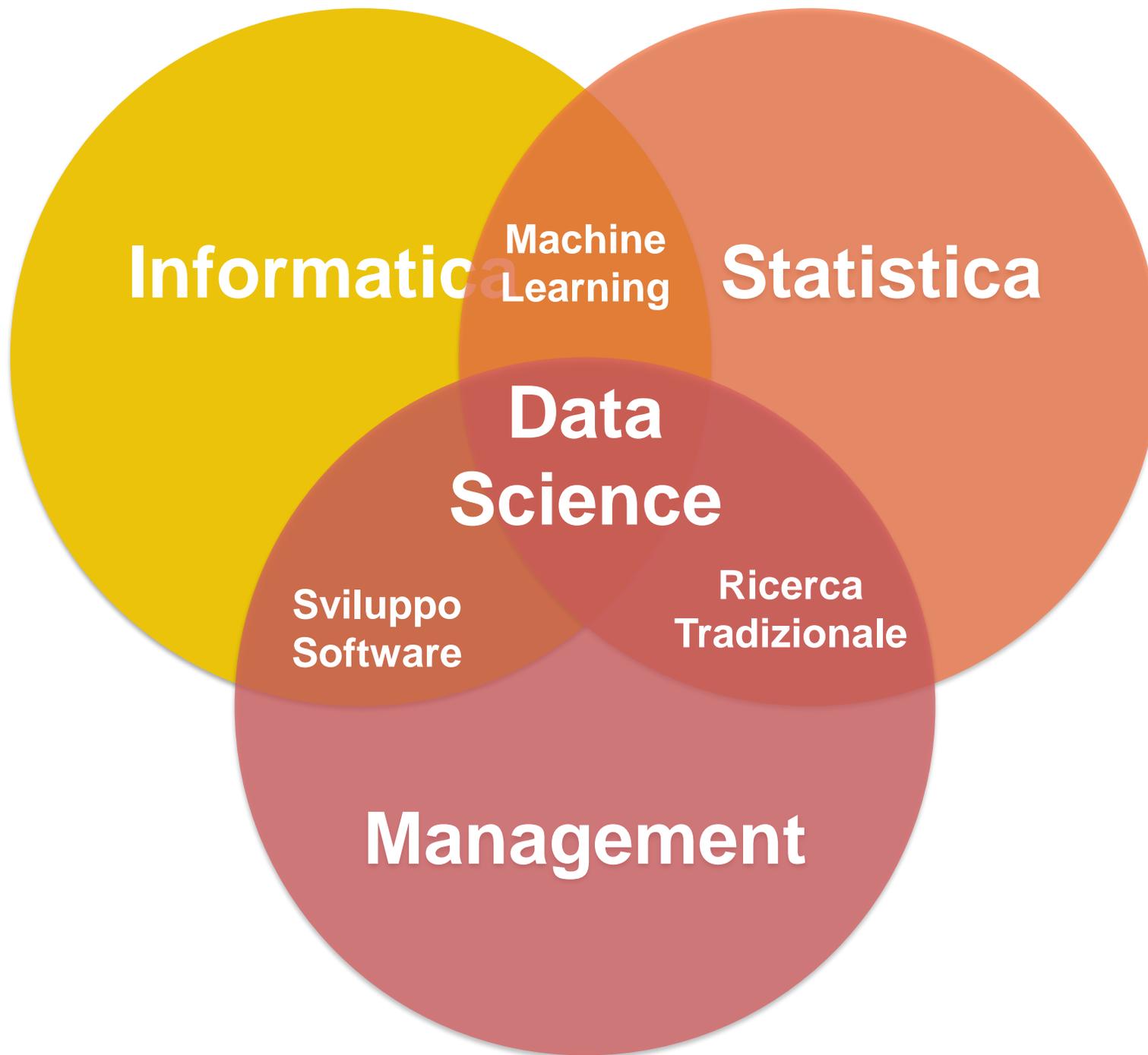




L'analisi dei dati: metodologie, trend e casi applicativi

Pierluca Lanzi – Politecnico di Milano

Che cos'è la data science?



Informatica

Machine
Learning

Statistica

**Data
Science**

Sviluppo
Software

Ricerca
Tradizionale

Management

Predittivo

Cosa succederà nel futuro?

Manutenzione predittiva: prevedere la probabilità di un guasto a partire dai record di funzionamento disponibili.

Descrittivo

Cos'è successo nel passato?

Root cause analysis: quale sequenza di eventi ha portato a un guasto o a una segnalazione di malfunzionamento.

Prescrittivo

Cosa ci consigliano di fare i modelli?

Sono trasversali, non si focalizzano tanto su uno scenario specifico quanto sul fornire delle indicazioni su cosa fare con la conoscenza estratta dai dati.

Il processo di analisi dei dati

Definizione del problema

```
graph TD; A[Definizione del problema] --> B[Acquisizione e integrazione dei dati]; B --> C[Preprocessamento ed esplorazione]; C --> D[Costruzione dei modelli]; D --> E[Validazione]; E --> F[Comunicazione];
```

Acquisizione e integrazione dei dati

Preprocessamento ed esplorazione

Costruzione dei modelli

Validazione

Comunicazione

70-90%



In passato doveva convincere della bontà dei risultati ottenuti e del vantaggio che portavano all'azienda.

La crescita esponenziale della complessità dei modelli rende praticamente impossibile spiegare come un risultato venga prodotto.

Comunicazione

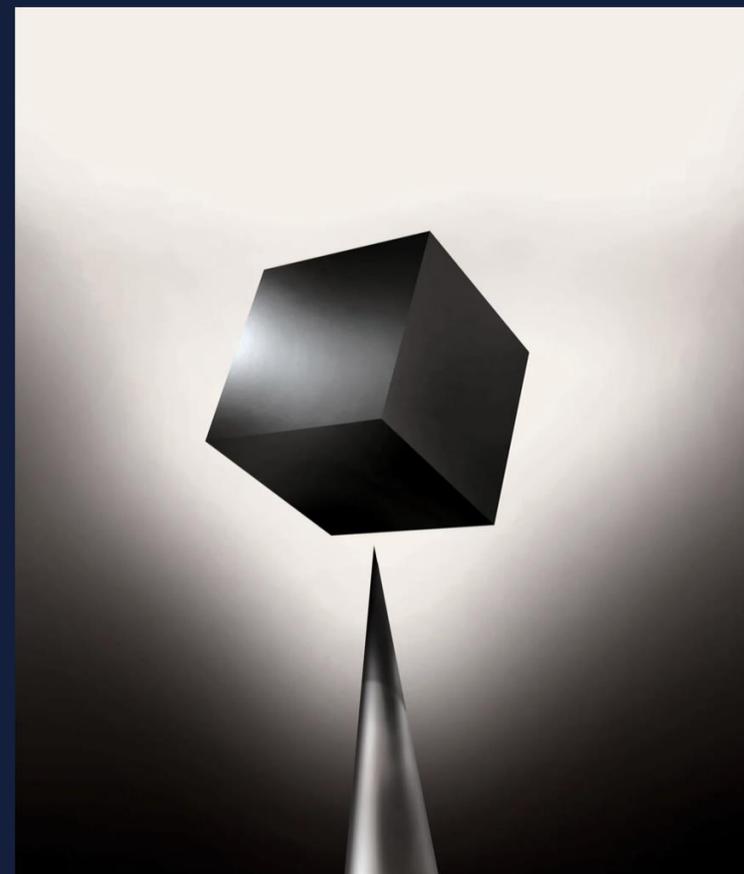
ARTIFICIAL INTELLIGENCE

The Dark Secret at the Heart of AI

No one really knows how the most advanced algorithms do what they do. That could be a problem.

By Will Knight

April 11, 2017



<https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>

“[...] The new tool offers no clue as to how it does this. [...]”

“[...] We can build these models,” Dudley says ruefully,
“but we don’t know how they work. [...] ”

Explainable artificial intelligence

Explainable machine learning

Cosa succede nel
rimanente 10-30%?

Costruzione dei modelli



Apprendimento Supervisionato

Apprendimento Non-Supervisionato

Apprendimento Per Rinforzo

Esiste un target da predire.

Linear regression
K-nearest neighbor
Naive Bayes
Decision trees
Random forests
Gradient boosting
Deep learning

Cerchiamo pattern interessanti.

Clustering
Regole di associazione
Regole sequenziali

Feedback qualitativo.

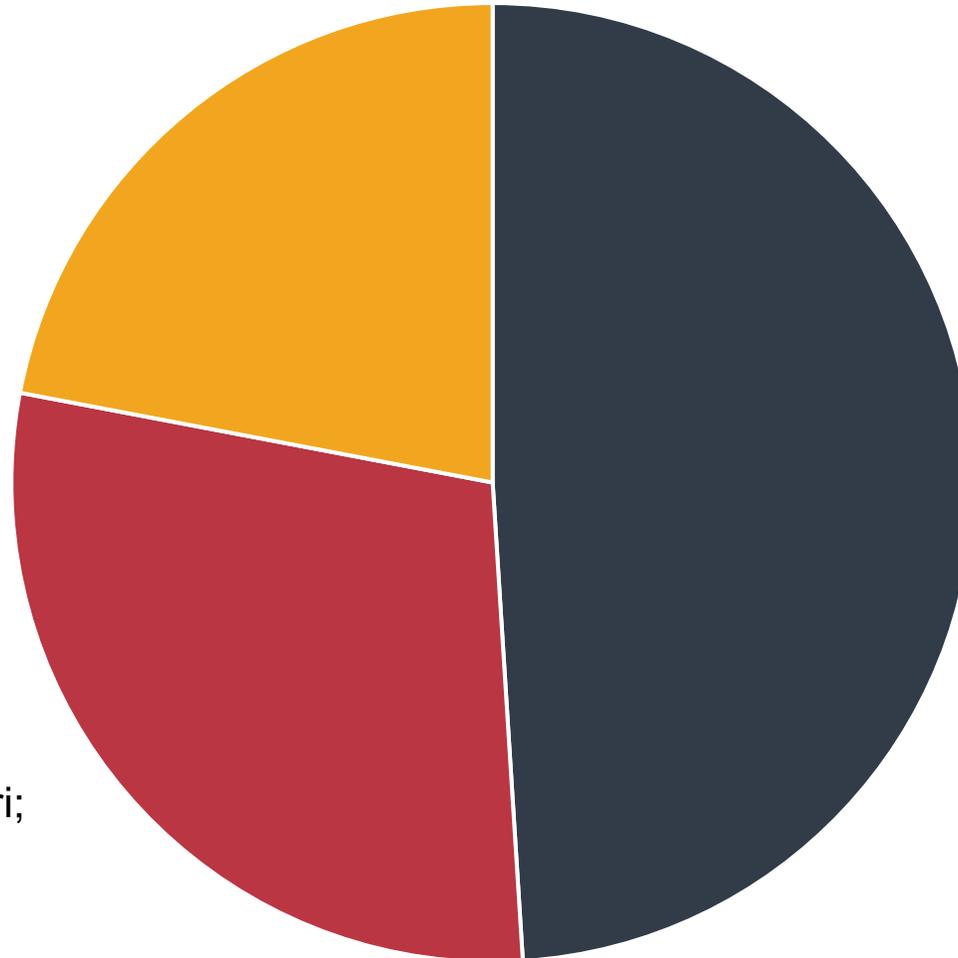
Q-learning (tabulare)
Deep Reinforcement Learning

E l'ambito ferroviario?

Applicazioni di Data Science in Ambito Ferroviario

- Individuare e caratterizzare fenomeni piuttosto che prevederne l'accadimento.
- Capire le cause e fattori che lo hanno influenzato, la frequenza, la gravità e valutare il rischio di un incidente
- Focus tipicamente su eventi molto rari con conseguenze gravi come, ad esempio, i deragliamenti.

- Controllo del traffico in tempo reale;
- Rilevamento di eventuali conflitti;
- Previsione del numero di passeggeri;
- Gestione delle interruzioni;
- Stima del ritardo dei treni.



- Ottimizzazione degli intervalli di manutenzione;
- Riduzione dei periodi di fermo macchina non pianificati;
- Identificazione cause dei guasti;
- Pianificazione delle procedure manutentive ottimali.

■ Manutenzione ■ Gestione Rete ■ Sicurezza

Per approfondire ...

Alcuni canali interessanti

- StatQuest
<https://www.youtube.com/c/joshstarmer>
- Introduction to Data Mining
https://www.youtube.com/playlist?list=PL8eNk_zTBST-gN6Y5E-5FZdARXjglYpyT
- Edureka!
<https://www.youtube.com/c/edurekaIN>
- Corey Shafer
<https://www.youtube.com/c/Coreyms>
- Simplilearn
<https://www.youtube.com/user/Simplilearn>

Grazie!

