10 Settembre 2018

Data value e data sharing:
La prospettiva Cineca

David Vannozzi, Direttore Generale Consorzio Cineca

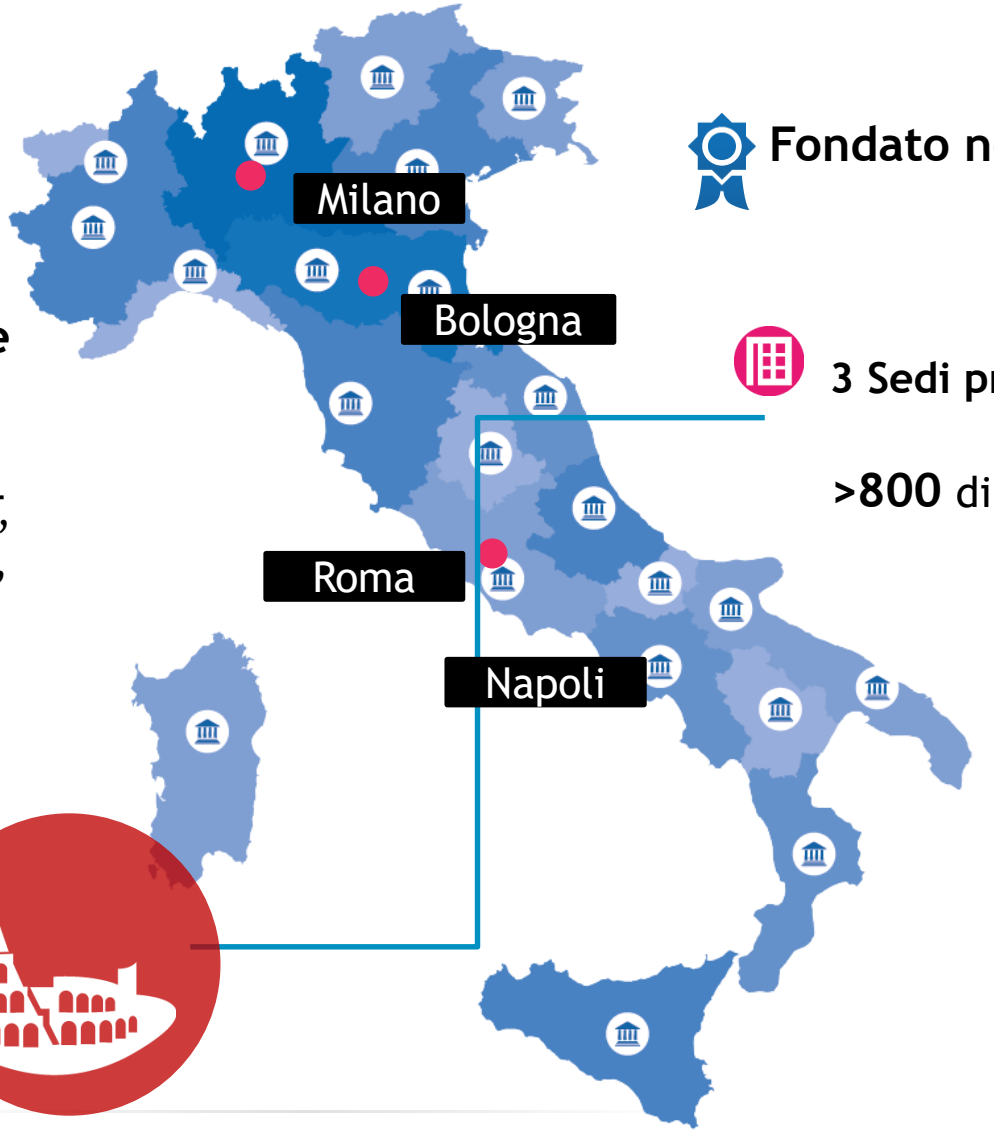
CINECA

Cineca: il principale consorzio universitario italiano

MIUR

66 Università Italiane

10 Enti di Ricerca
CNR, CREA, INDIRE, INAF, INFN, OGS, INVALSI, SZN, INRiM, GSSI

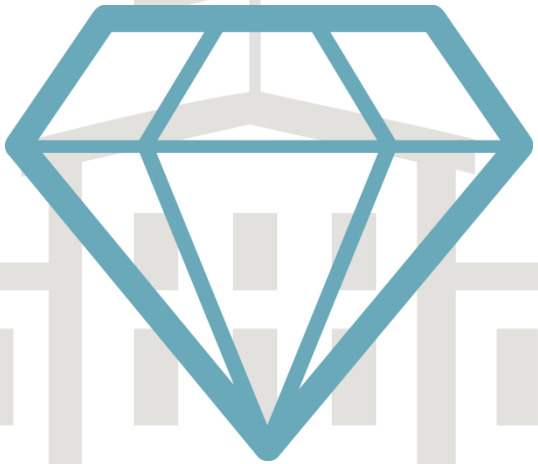


Fondato nel 1969

3 Sedi principali

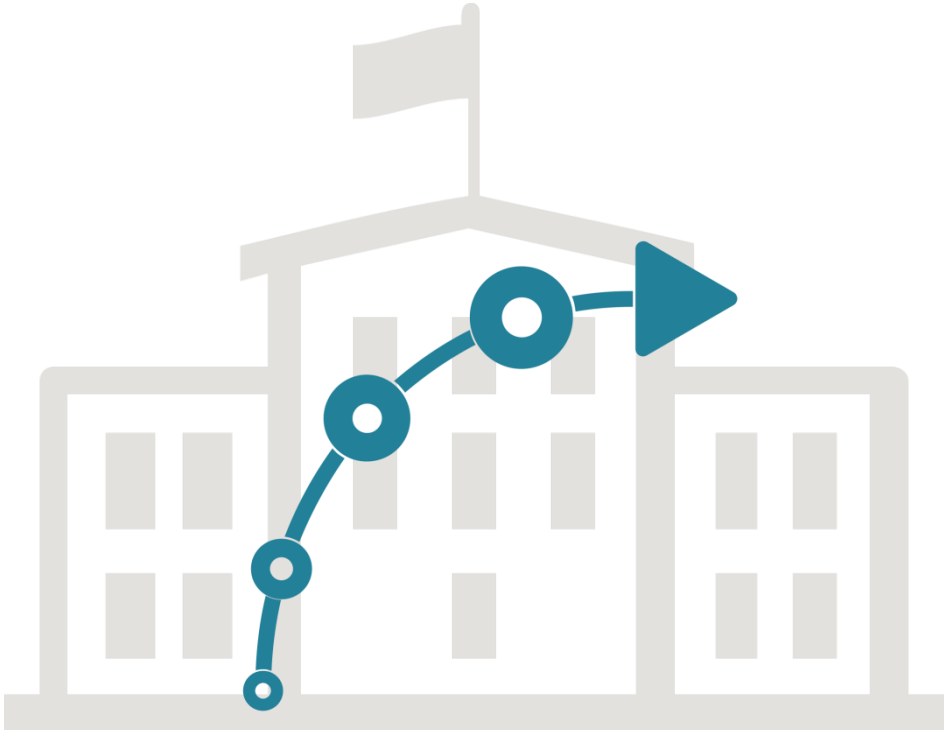
>800 dipendenti

I dati sono ricchezza



- Definiti il «petrolio del terzo millennio»
- La vendita di servizi digitali collegati all'utilizzo dei dati è diventato un fattore strategico decisivo per molti settori industriali: sono diventati un asset aziendale
- Anche per la pubblica amministrazione: il piano triennale va in questa direzione
- i dati sono una grandissima risorsa che ogni organizzazione si trova oggi ad avere, in maniera più o meno consapevole
- maggiore è la consapevolezza maggiore è la pulizia con cui il dato è stato raccolto e quindi inferiore è lo sforzo di raffinamento che è necessario fare per trasformare davvero la materia prima in valore: maggiore è il loro valore

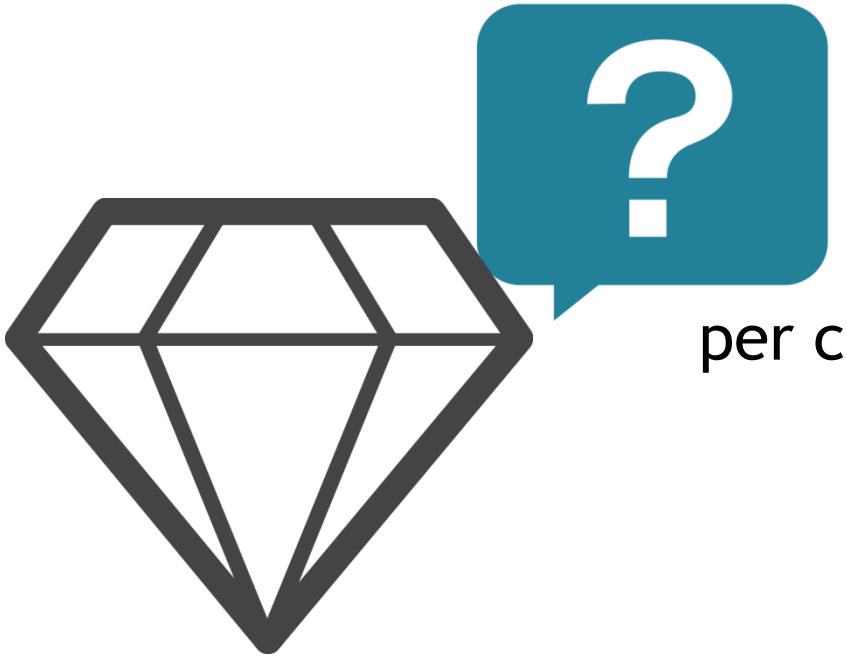
I dati sono storia



- I dati rappresentano il passato di una organizzazione, la sua storia, ma raccontano anche il suo presente e possono aiutarci ad anticiparne il futuro

I dati sono ricchezza…

Che caratteristiche devono avere?



per chi?

Caratteristiche per generare valore

- VOLUME – il valore è legato alla quantità di dati a disposizione: più è grande più è grande il loro valore (infatti spesso si parla di big data)
- VELOCITA' – i dati devono essere a disposizione e accessibili in real time
- NON CONOSCIUTI – le elaborazioni che vengono effettuate devono portare ad informazioni sconosciute
- PROTETTI
- VERI/VERIDICITA' – il dato ha valore se è accurato
- UTILIZZABILI – il dato di sintesi deve essere accessibile quando serve con l'ausilio di strumenti tecnologici
- VARIETA' – dati strutturati e dati non strutturati
- VARIABILITA' – devono poter assumere diverso significato a seconda dell'utilizzo che se ne fa

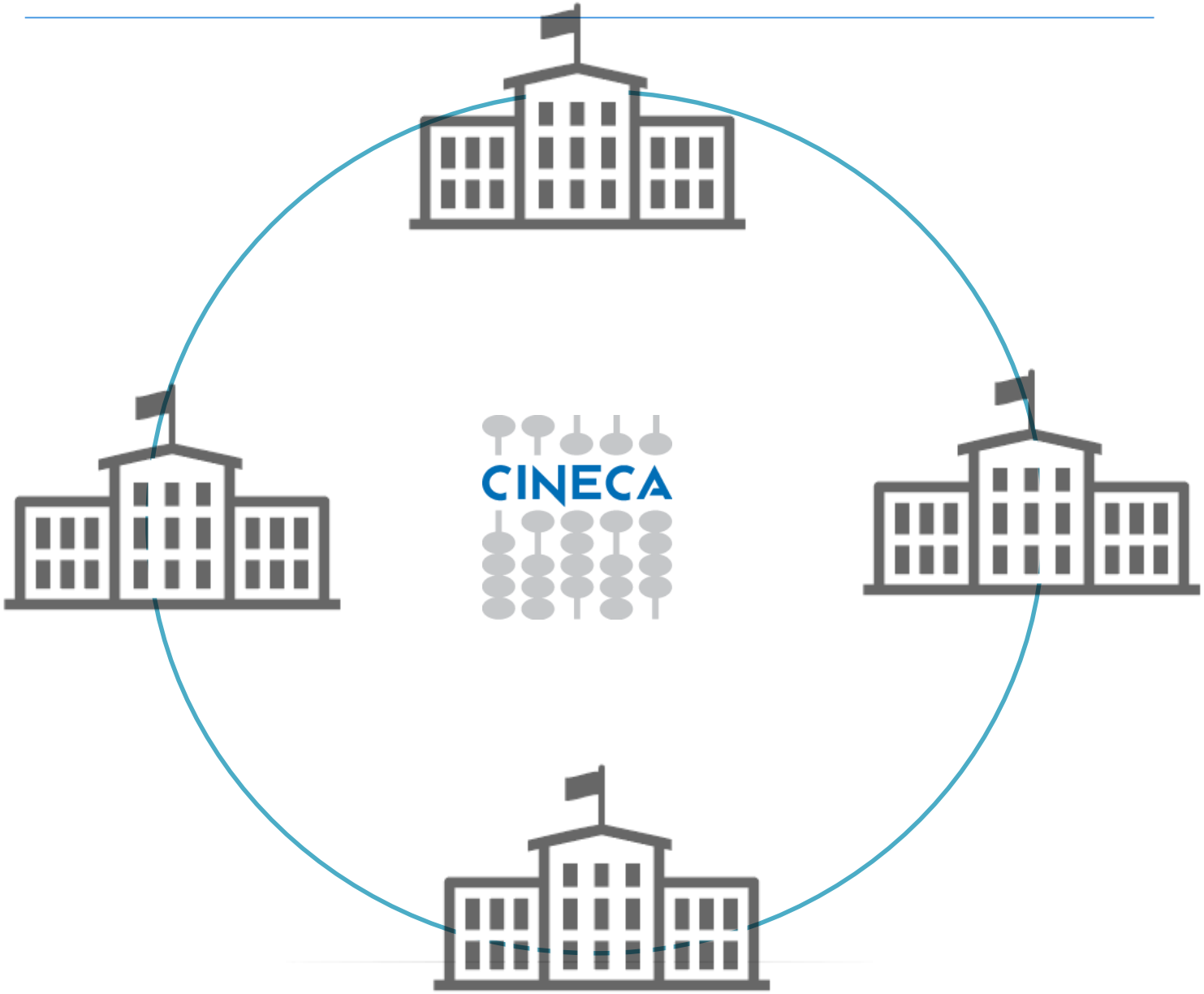
Le attività di Cineca

Supercalcolo per la comunità scientifica

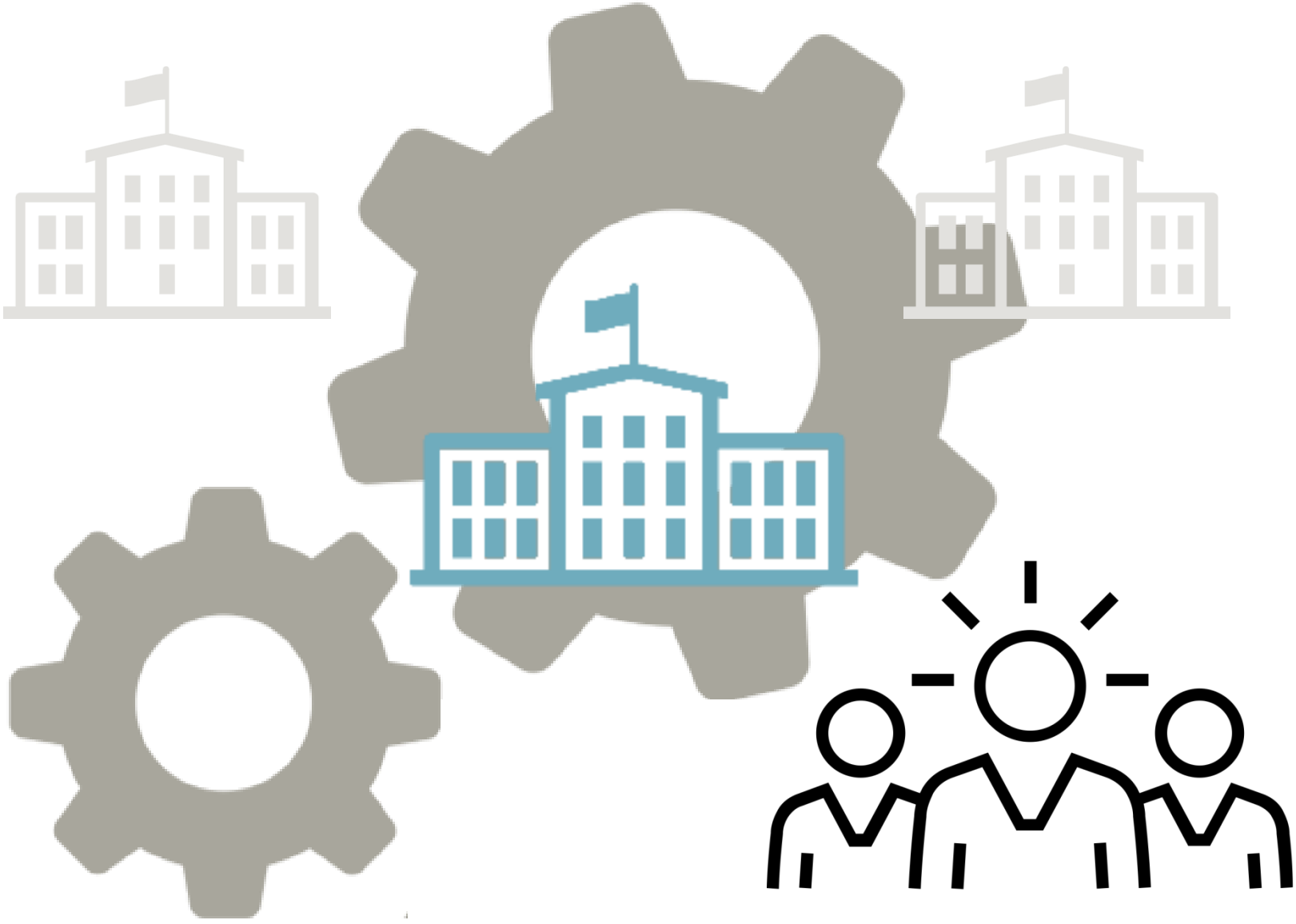
Sistemi gestionali e servizi a sostegno delle Università e del MIUR

Trasferimento tecnologico alle imprese e alla pubblica amministrazione

Le università italiane: una comunità di pratica



Atenei, Studenti, Paese



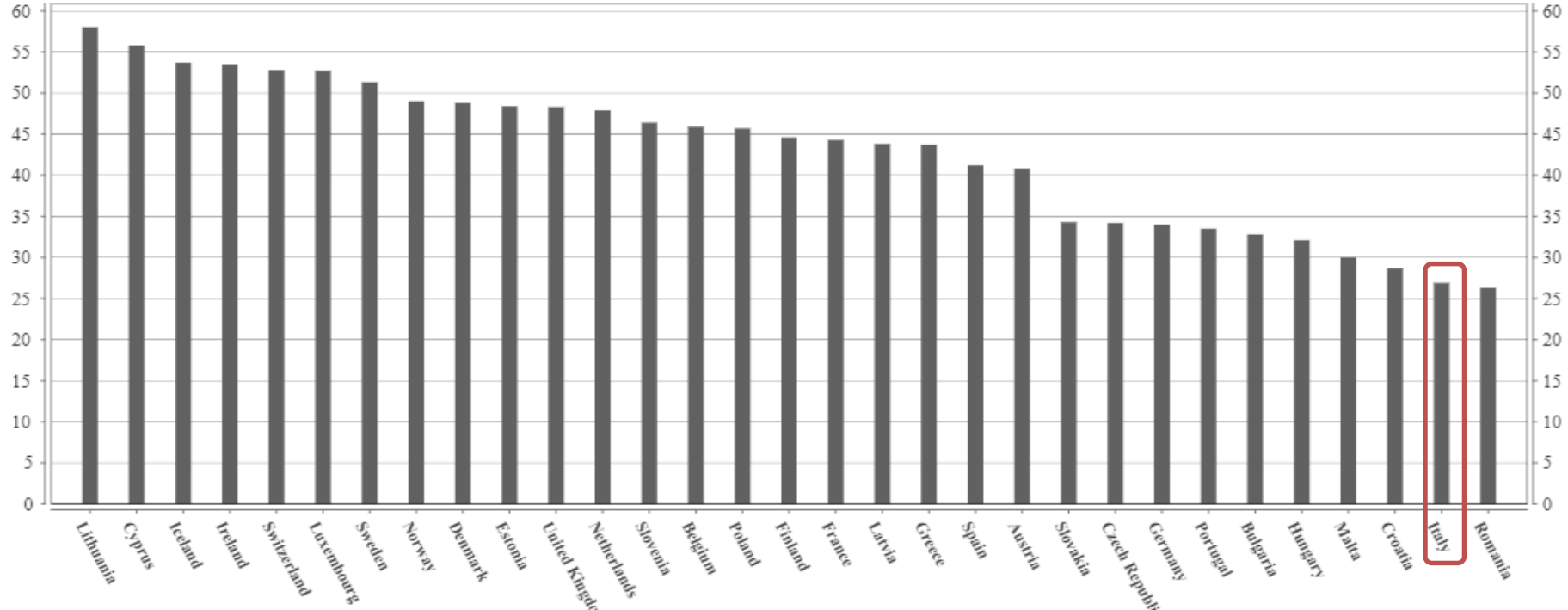
Osservatorio abbandoni

Osservatorio abbandoni: il valore

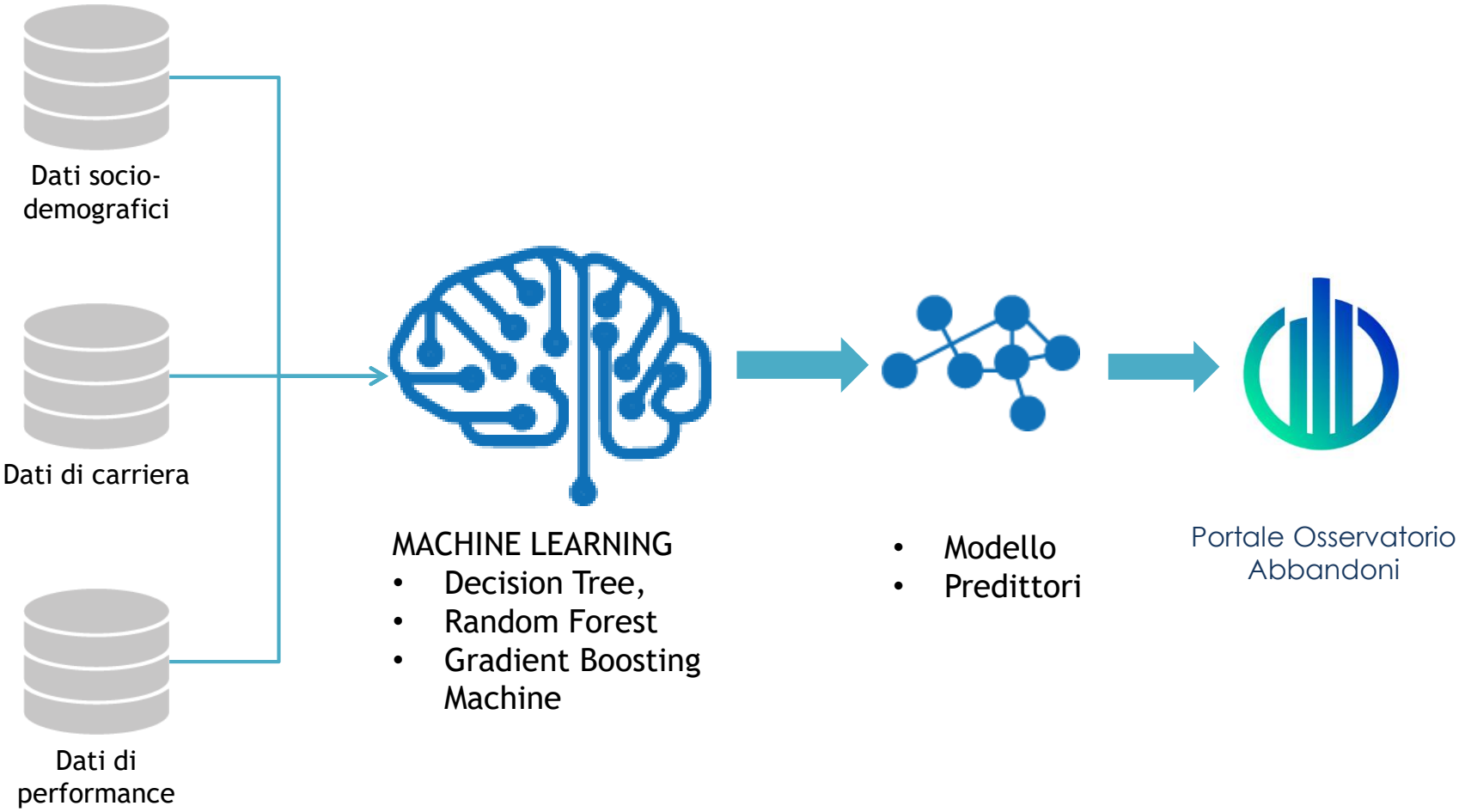
Un abbandono è valore perso

- Per l'Ateneo: investimenti perduti ed elemento di demerito nella valutazione ministeriale
- Per il Paese: risorse pubbliche adoperate senza successo
- Per lo Studente: una scelta sbagliata, investimento in tempo e denaro

Dall'indagine Eurostat sui laureati fra i 30 e 34 anni



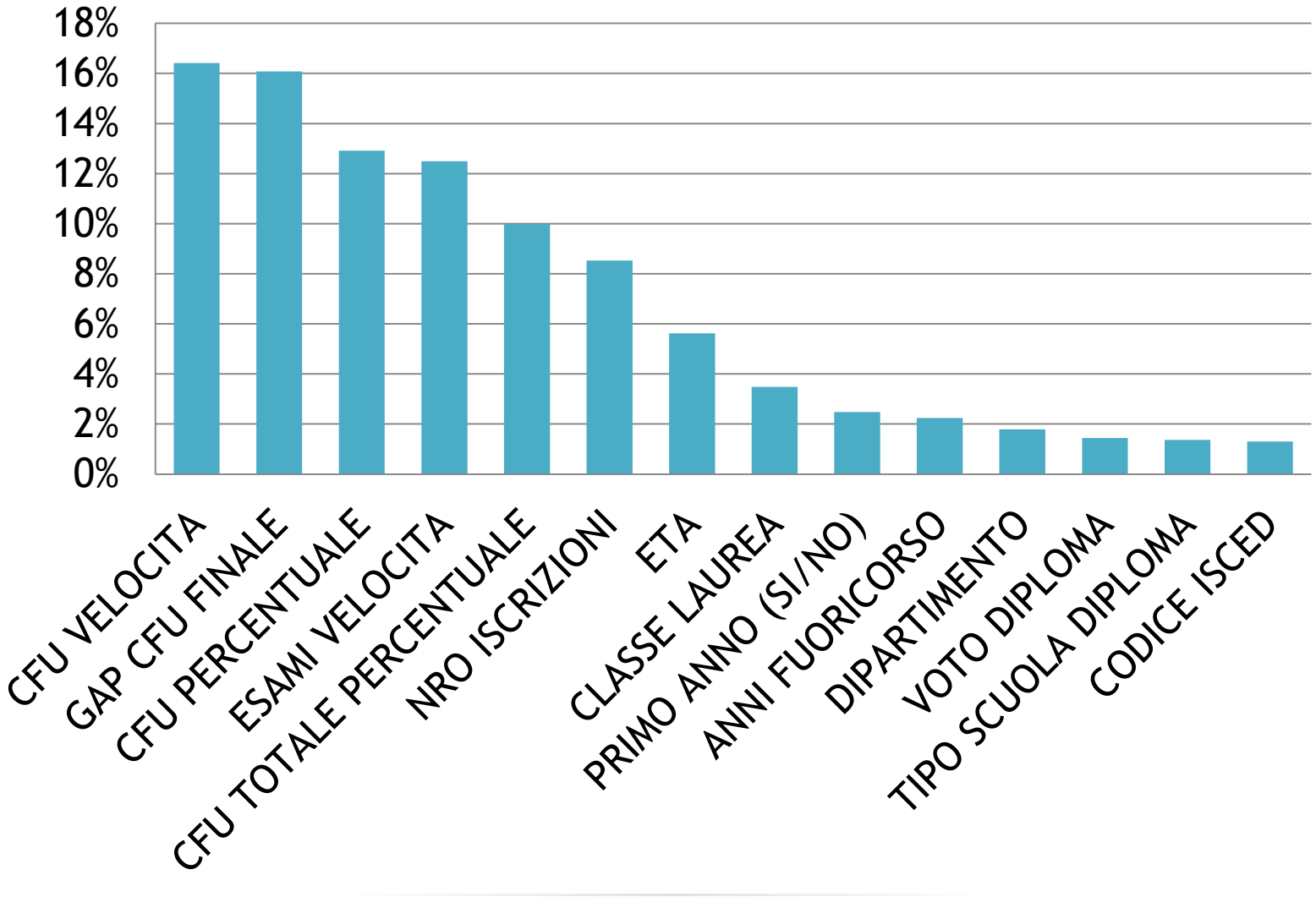
Dati e Tecnologia



Machine Learning: verifica e incremento

- Il modello estratto da un set di dati viene verificato su un dataset omogeneo ma differente
  - Altri anni accademici per cui si ha già l'informazione degli eventuali abbandoni
- Questa verifica porta nuova conoscenza
  - La maggior parte degli abbandoni si verifica tra il primo e il secondo anno
- La nuova conoscenza viene integrata
  - Analisi più approfondita sui dati del primo anno

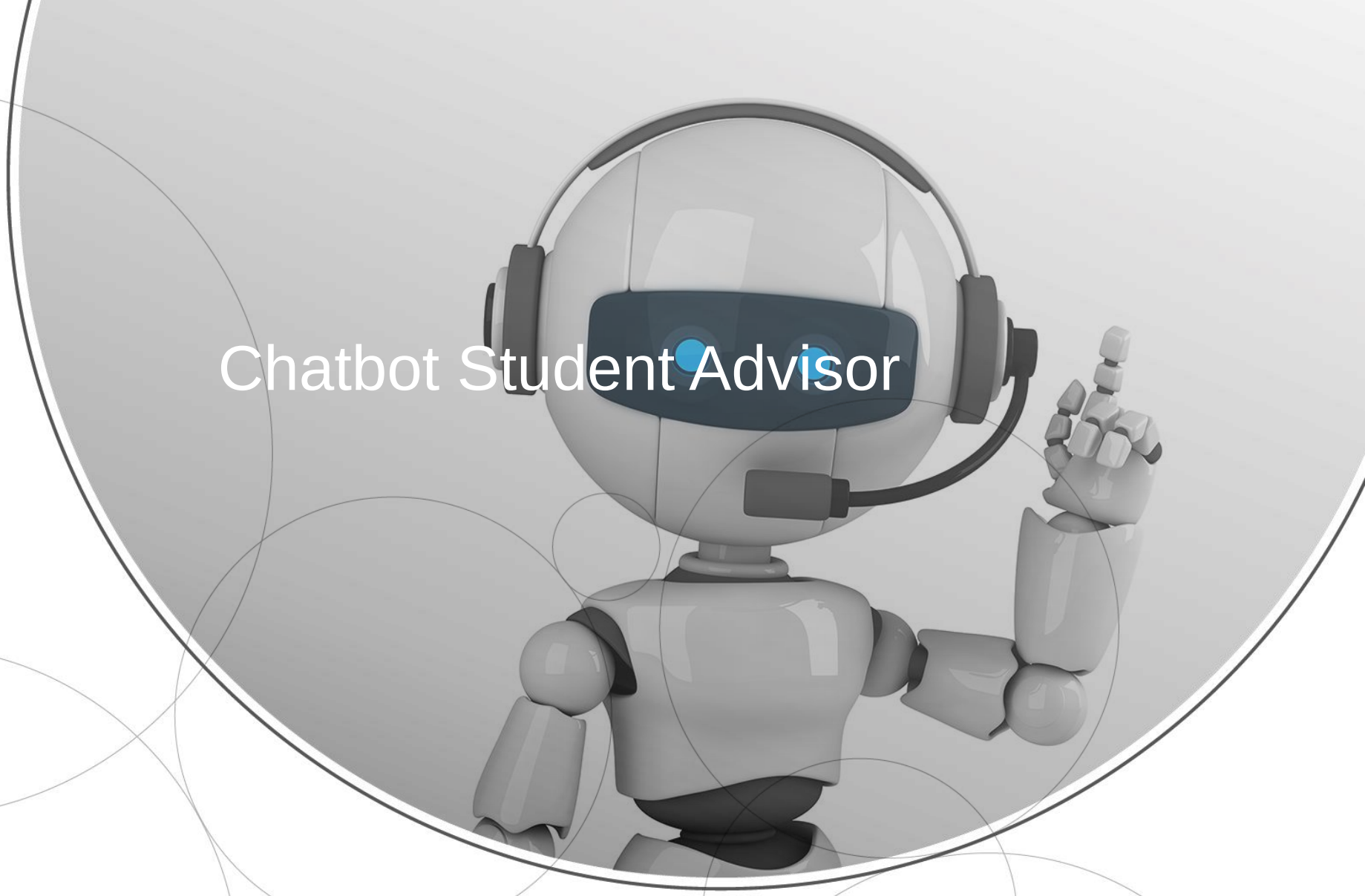
Le variabili più influenti



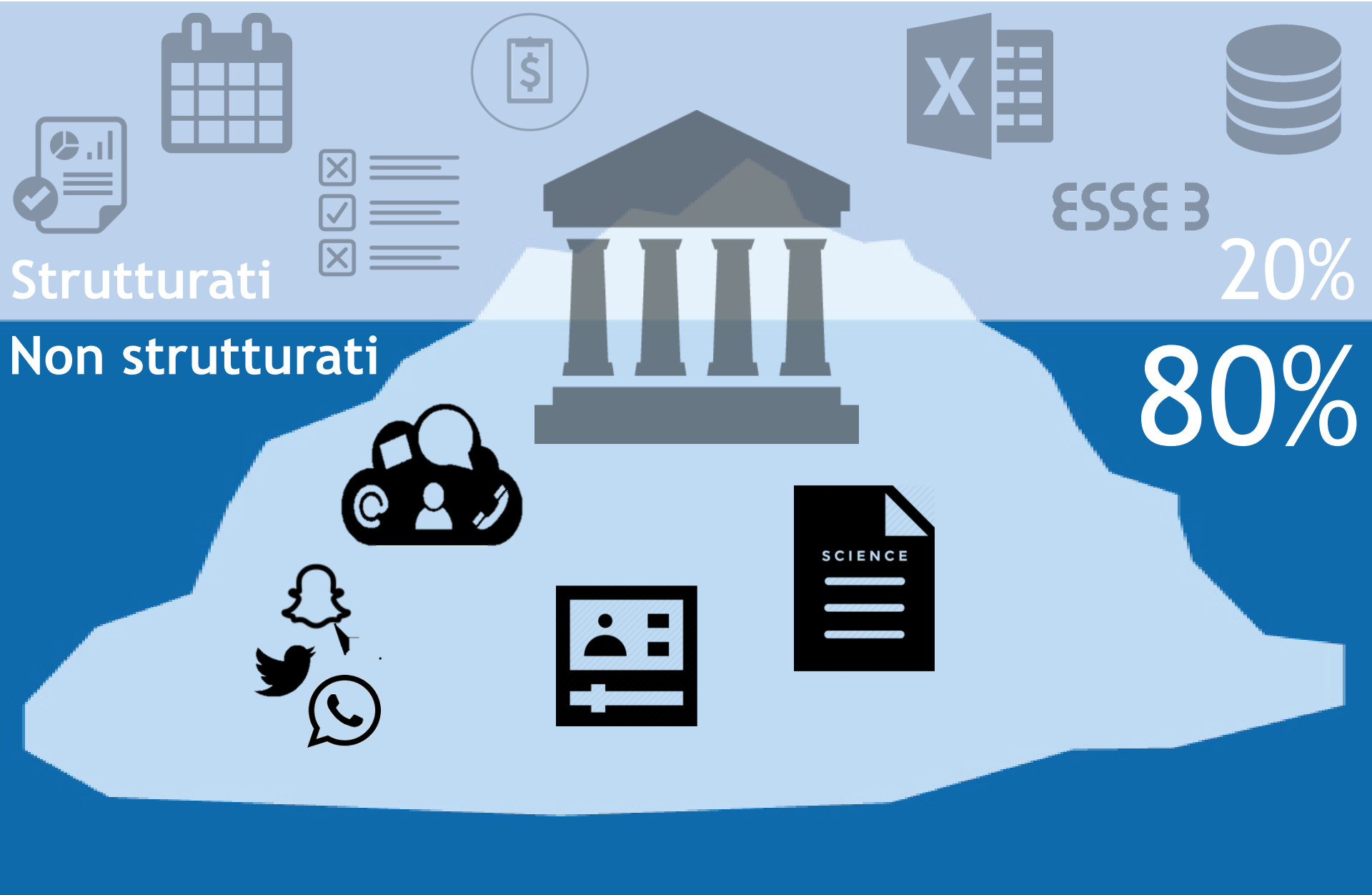
Dal modello alla persona

Obiettivo: date le caratteristiche di uno studente, qual è la probabilità che tale studente abbandonerà nell'anno corrente?

Chatbot Student Advisor



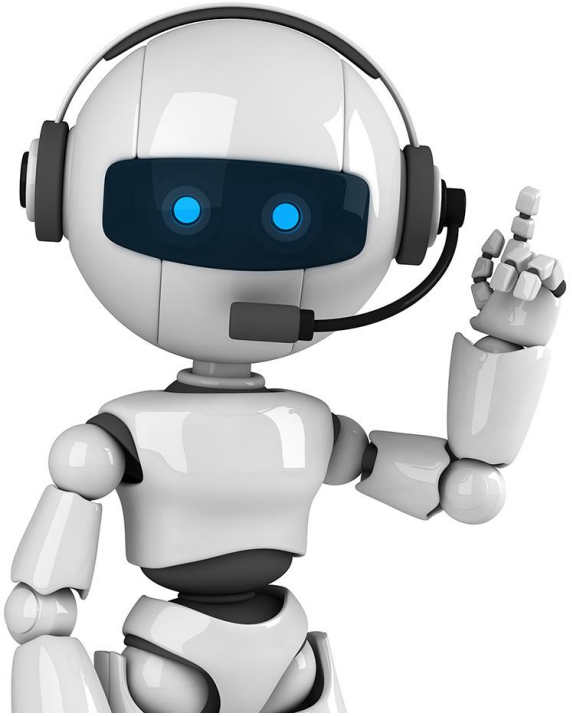
Il valore dei dati non strutturati

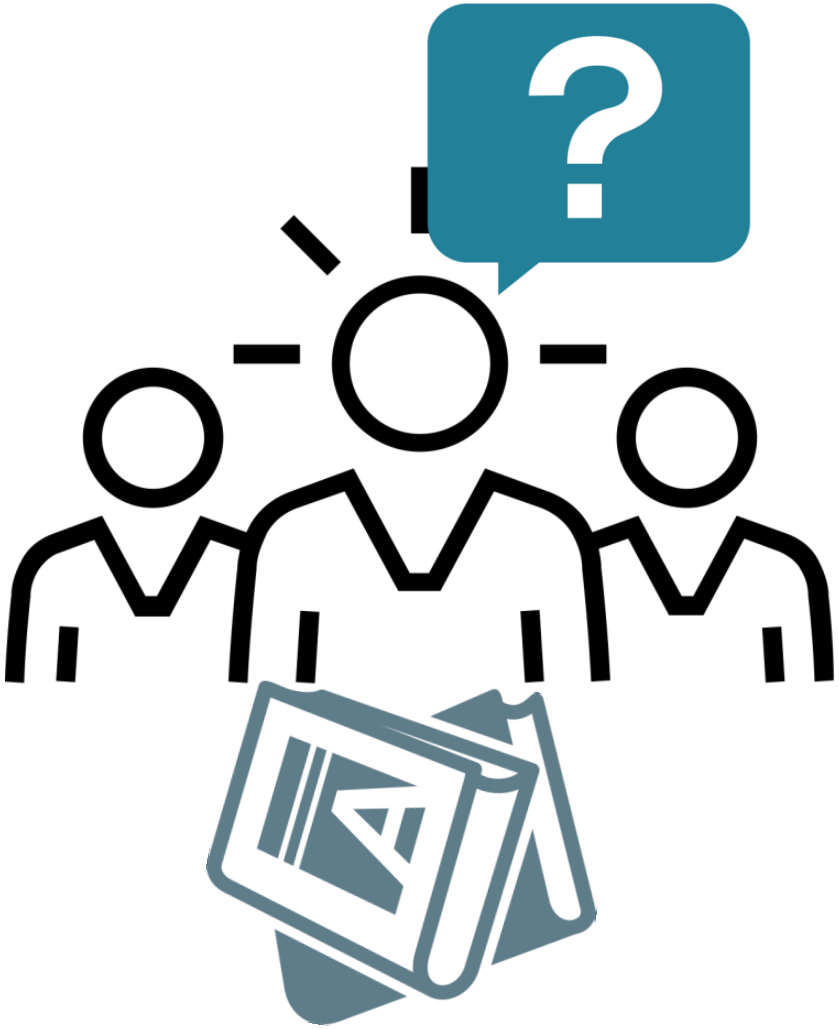


Cognitive Computing

Il Cognitive Computing permette di individuare ed estrarre informazioni e relazioni da **grandi moli di dati non strutturati** usando le più innovative ed efficaci tecniche di **Intelligenza Artificiale**. Sono soluzioni basate su servizi di tipo probabilistico.

Student Advisor





Cognitive Computing e Natural Language Processing

Dopo una prima elaborazione della Knowledge Base e delle domande in linguaggio naturale più frequenti (il 20% di domande copre l'80% dei casi), l'algoritmo viene allenato («*taming*», domato) con successive iterazioni su varianti delle domande e correzione delle risposte.

Il Machine Learning permette al sistema di raffinare la sua capacità di individuare l'*intento* di una domanda e rispondere in maniera pertinente dopo averla collegata ad un caso noto.

Individuare l'intento

Student Advisor: un’interfaccia conversazionale

Conversare per semplificare

- Sempre disponibile
- Familiare allo studente
- Attivabile su diversi canali (web, app studenti, FB Messenger…)
- Solleva l'help desk da un'ampia parte di domande
- Può guidare la conversazione verso la risposta più adatta
- Sa individuare i casi in cui è necessario rimandare all'operatore

Verso il riutilizzo delle conversazioni e dei processi

Possibilità di **riutilizzare il training** per gli intent più comuni:

- Come faccio a…
- Come richiedo…
- Dove trovo…

Personalizzare le risposte a seconda del **tono di voce** dei singoli chatbot

- **Quando posso depositare Il titolo della tesi?**
  - Come deliberato dagli organi direttivi, gli studenti sono tenuti a presentare il titolo entro...
  - Puoi presentare la tua tesi dopo l'iscrizione al penultimo anno.
  - Puoi già farlo, coraggio, è scritto in evidenza nella pagina My Univ!

una ultima considerazione ….

- I dati per avere valore devono essere protetti, non conosciuti, nuovi
- Alcune situazioni che sono accadute nell'ultimo periodo ci pongono degli interrogativi:
  - Cambridge analytica
  - Google/Mastercard
  - Mail aziendali

Grazie

d.vannozzi@cineca.it

www.cineca.it