

Programmazione di Applicazioni Data Intensive

Introduzione al Corso

<https://tinyurl.com/gianlucamoro>



Materiale didattico

<https://virtuale.unibo.it/course/view.php?id=52948>

Proff. [Gianluca Moro](#) e tutor Roberto Pasolini (lab.)
Dip. di Informatica (DISI) - Università di Bologna, Cesena
nome.cognome@unibo.it

Secondo Semestre



Rivoluzione in Corso: Cos'hanno in Comune queste Applicazioni Straordinarie ?

- **Assistenti virtuali** con capacità umana nel conversare al telefono per fare una prenotazione <https://bit.ly/2Kyjbgw>
- *chatGPT, Bard, Falcon, Llama, Alpaca ...* Large Language Models
- **Negozi al dettaglio interamente automatizzati** senza nemmeno cassieri e casse per pagare <https://bit.ly/2FuQgbi>
- **Diagnosi** di melanoma con precisione superumana (pub. Nature)
 - [Dermatologist-level classification of skin cancer with deep neural networks](#),
- **Profilazione utenti**: capacità di prevedere da pochi post e like caratteristiche personali altamente sensibili <https://bit.ly/1mG7Go6>
 - age, gender & sexual orientation, ethnicity, religious and political views, use of illegal substances, happiness, intelligence, personality traits, parental separation
- **Robot che imparano da zero e da soli** a svolgere nuovi compiti: camminare, correre, evitare ostacoli <https://bit.ly/2sQmx5z>
- **Auto a guida autonoma**: Tesla, Google e tanti altri ...



Cos'hanno in Comune queste Applicazioni?



Ottenere Intelligenza dai Dati: Rivoluzione Con Impatto Equiparato alla Scoperta dell'Elettricità

- L'intelligenza di queste applicazioni **non è cablata** nel codice, ma **è appresa da insiemi di dati** con algoritmi standard di **dominio pubblico** e indipendenti dal tipo di problema
 - **"Artificial Intelligence is the New Electricity"** Prof. Andrew Ng (Stanford) tra i massimi esperti mondiali di machine learning
 - *l'intelligenza artificiale applicata cambierà radicalmente interi settori della vita economica e sociale* dalla salute, all'industria, all'agroalimentare, ai trasporti. L'impatto sulla società è equiparato a quello dell'invenzione del motore a vapore o dell'elettricità ([Ministero dello Sviluppo Economico](#)).
- E.g. applicazioni che sfruttano l'**intelligenza ricavata dai dati**
 - prevedere l'**andamento delle vendite** di prodotti/servizi
 - scoprire le **propensioni di acquisto** di ogni utente
 - **riconoscere condizioni di rischio** di pazienti da sensori indossabili (IoT)
 - prevedere **consumi energetici**, **andamento della borsa** e tanto altro ...

Ottenere Intelligenza dai Dati: Rivoluzione Con Impatto Equiparato alla Scoperta dell'Elettricità

- L'intelligenza di queste applicazioni non è innata ma **è appresa dai dati**.
- Repubblica: L'economia 4.0 cambia il lavoro: i più richiesti saranno i **DATA SCIENTIST**
- SOLE 24 ORE: Data Scientist, il lavoro da 100mila euro l'anno
- IBM Predicts Demand For **DATA SCIENTIST** Will Soar 28% By
- LinkedIn's Fastest-Growing Jobs Today Are In **DATA SCIENCE**
- Gartner Group: **Applied Artificial Intelligence** a top strategic technology
- E.g. applicazioni che sfruttano l'**intelligenza ricavata dai dati**
 - prevedere **DATA SCIENCE ??**
 - scoprire le *Principi, tecniche e algoritmi per ricavare*
 - riconoscere *intelligenza dai dati con metodo scientifico*
 - prevedere **consumi energetici, andamento della borsa** e tanto altro ...

G. Moro, R. Pasolini - DIA - Dipartimento di Informatica, Università di Bologna, Cesena

△

Programmazione di Applicazioni Data Intensive: Obiettivo del Corso

- I problemi citati richiedono applicazioni capaci di estrarre intelligenza dai dati → DATA INTENSIVE
- **Obiettivo del corso:**
 - Fornire contenuti teorici e pratici di base di **Data Science** dai fondamentali al **Machine Learning**, per lo sviluppo in Python di modelli e **applicazioni di Intelligenza Artificiale**
- Pochi rudimenti di base della matematica, non richiesti all'esame e già implementati nelle librerie, sono la base di queste applicazioni
 - *vettori, regressione, min/max di funzioni*
- L'informatica fondata su di essi rende competitivi nell'area in grande crescita del **data science** che include anche *machine learning*
 - in piccole e grandi aziende sono in forte aumento le posizioni di **data scientist e machine learning engineer** (stage in Google, Amazon, Microsoft, IBM, Apple ecc.)

G. Moro, R. Pasolini - DIA - Dipartimento di Informatica, Università di Bologna, Cesena

C



Programma del Corso e Applicazioni in Lab

- Sviluppo di **applicazioni predittive**
 - predire il prezzo di abitazioni dalle vendite precedenti, predire i consumi energetici, predire le **quotazioni di titoli di borsa**
 - Predire **se un prestito verrà restituito**, predire le **spese di manutenzione di auto**, *riconoscere attività umane mediante smartphone*
- Applicazioni di **recommendation** di prodotti/servizi/news
 - Scoprire **interessi e propensioni di acquisto** di utenti con algebra lineare, algoritmi di collaborative filtering e SVD, recommendation impersonali e collocazione di prodotti in magazzino per minimizzare il pick-up time
- Applicazioni di base per il **Natural Language Processing**
 - elaborazione del linguaggio con machine learning e **language model**: *motori di ricerca, riassumere doc, question answering, **intro a chatbot** ...*
 - determinare la customer satisfaction su **dati reali da Amazon, Twitter, Facebook**, distinguere commenti in positivi/negativi

G. Moro, R. Pasolini - DIA - Dipartimento di Informatica, Università di Bologna, Cesena

6



Recommendation: Una Storia Vera (i)

- Minneapolis, Supermercati TARGET
 - Un padre un po' arrabbiato si presenta in una filiale della catena chiedendo di parlare con un manager
 - *"mia figlia riceve da settimane vostri coupon sconto su **prodotti per la maternità come vestiti, culle e pannolini per neonati ... ma sta facendo ancora la scuola superiore, la state incoraggiando a rimanere incinta ?**"*
 - Il manager rileva che il materiale indirizzato alla figlia dell'uomo conteneva coupon sconto e pubblicità per abbigliamento e articoli premaman, ma anche foto di neonati sorridenti ...
 - Il manager: *"ci scusi deve esserci stato un errore"* e alcuni giorni dopo richiama il padre per scusarsi nuovamente
 - Al telefono, però, il padre era piuttosto imbarazzato. *"Ho parlato con mia figlia ... ci sono state alcune attività in casa mia di cui non ero consapevole. **Lei partorirà in Agosto. Le devo io delle scuse**"*



fonte: Forbes - Charles Duigg of New York Times

Data Intensive Applications - G. Moro, R. Pasolini - DISI, Università di Bologna, Cesena

7



Recommendation: Una Storia Vera (ii)

- Come il supermercato ha previsto che la ragazza era incinta e quali prodotti sarebbero stati più appropriati per lei ?
- il supermercato associa ad ogni scontrino un codice cliente che dipende dalla carta di credito, dall'email o dall'indirizzo
 - perciò ha lo **storico degli acquisti** di ogni cliente
 - ed anche **dati demografici** acquisiti direttamente o acquistati
- Le spese per neonati e bambini sono un grande business
 - il supermercato avviò anni prima un progetto mirato ad “agganciare” i genitori imminenti, prima che diventino clienti della concorrenza
- Cos'è stato scoperto dall'analisi dei dati ?
 - le **donne incinta** (*stato desumibile anche a posteriori dall'età del figlio*) acquistano di **più lozioni inodore all'inizio del 2° trimestre di gravidanza**
 - nelle **prime 20 settimane anche integratori** di calcio, magnesio e zinco
 - l'aumento del consumo di **battufoli di cotone, disinfettanti per le mani e salviette indicano che sono prossime al parto**

Data Intensive Applications - G. Moro, R. Pasolini - DISI, Università di Bologna, Cesena

8



Recommendation: Una Storia Vera (iii)

- **C'è di più ...**
 - sono stati **identificati circa 25 prodotti** che insieme permettono di **stimare la data di nascita** in un piccolo intervallo
 - ciò consente di proporre coupons diversificati in base alla fase della gravidanza
 - uno score predice lo stato di gravidanza in base agli acquisti:
 - e.g. donna di 23 anni, a **Marzo** raddoppia il consumo di lozioni inodore, magnesio, zinco, acquista un tappeto blu brillante -> **parto in Agosto 87%**
- il supermercato ha rilevato anche un **comportamento inatteso**
 - un alto numero di futuri genitori che ricevevano solo materiale pubblicitario per la maternità non diventavano clienti
 - distribuendo invece articoli di maternità in mezzo ad altri, ha procurato maggiori vendite
 - Perché ? la conclusione fu che a **nessuno piace essere troppo “spiato”**
- Aumentato il fatturato del 50%

Data Intensive Applications - G. Moro, R. Pasolini - DISI, Università di Bologna, Cesena

9



Programma del Corso: Tecnologie Python

- **Introduzione da zero al linguaggio Python 3**
 - **primo linguaggio più richiesto** a livello internazionale dopo SQL e Java
 - **Standard di fatto** in Data Science, Machine & Deep Learning
 - facile da imparare, utilizzabile in modo interattivo
 - usando l'ambiente **Jupyter**, descrizione dell'attività, dati da usare e codice eseguibile interattivamente sono forniti in un unico documento
- **Introduzione alle librerie di Data Science e Machine Learning**
 - **NumPy** per operazioni tra array N-dimensionali e algebra lineare
 - **pandas** per l'elaborazione di dati tabulari
 - **scikit-learn** algoritmi di machine learning per l'intelligenza artificiale
 - **NLTK** per il trattamento di dati testuali
 - **KERAS/Pytorch** per lo sviluppo di reti neurali
 - **Hugging Face** Language Model
 - **Flask** per realizzare web application per le applicazioni sopracitate

G. Moro, R. Pasolini - DIA - Dipartimento di Informatica, Università di Bologna, Cesena

10



Organizzazione del Corso

Lezioni: *Nessun pre-requisito richiesto*

- **lezioni** divise tra aula e **laboratorio assistito** (con seminari, esclusi dall'esame, su **deep neural network in Python** su casi di studio reali allo stato dell'arte)

Modalità di esame (6 CFU): *sviluppo di un progetto e colloquio*

- Discussione individuale di un **progetto** di laboratorio di gruppo (1-3 studenti) concordato col docente, su uno o più argomenti del corso

Materiale didattico – testo suggerito non obbligatorio

- **[Data Science con Python, Dai Fondamenti al Machine Learning](#)** EGEA, Joel Grus, Edizione italiana curata da Gianluca Moro, 2021
- **Slide e codice** disponibili nel sito del corso
 - <https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2023/385100>
 - **Riferimenti** bibliografici per approfondimenti opzionali (sito web corso)
 - Applicazioni recenti di intelligenza artificiale sviluppati anche con tesi di questo corso <https://disi-unibo-nlp.github.io/projects/>

G. Moro, R. Pasolini - DIA - Dipartimento di Informatica, Università di Bologna, Cesena

11

