



ASSOCIAZIONE ITALIANA DI RICERCA OPERATIVA

# **Parere dell'Associazione Italiana di Ricerca Operativa sulla costituzione di una nuova classe di Laurea Magistrale in Data Science**

Il Consiglio Direttivo dell'AIRO (Associazione Italiana di Ricerca Operativa), che dal 2016 ha assunto la rappresentanza del SSD MAT/09, Ricerca Operativa, ha nominato un gruppo di lavoro che ha elaborato il presente documento in preparazione all'audizione presso il Consiglio Universitario Nazionale del 14 giugno 2018.

Il documento propone alcuni spunti di riflessione funzionali alla costituzione di una possibile classe di laurea magistrale in Data Science, fornendo il punto di vista maturato in seno al settore scientifico-disciplinare MAT/09. In particolare, il documento si articola in due parti, la prima illustra la situazione internazionale e nazionale dei corsi in Data Science, la seconda fornisce una risposta ad alcune domande poste dal CUN in merito alla creazione di una nuova classe di laurea Magistrale.

## **La situazione internazionale e nazionale**

Come osservato in un recente rapporto del McKinsey Global Institute<sup>1</sup> la disponibilità in volume e varietà dei dati è cresciuta esponenzialmente negli ultimi anni rendendo i Big Data, e di conseguenza la loro gestione, analisi e valorizzazione, un fenomeno pervasivo dell'economia mondiale. Parimenti, l'aumento della capacità computazionale e lo sviluppo di algoritmi sempre più efficaci alimenta rapidi progressi tecnologici che stanno portando ad una nuova rivoluzione industriale. In maniera del tutto analoga, il Rapporto del Gruppo di Lavoro Big Data@MIUR<sup>2</sup> sottolinea che "nei prossimi anni crescerà ancora esponenzialmente il numero e la diversità delle possibili sorgenti di dati, in una complessità alimentata da un processo di coalescenza per cui insiemi di dati un tempo fisicamente o logicamente isolati sono ora accessibili insieme".

La disponibilità di strumenti di analisi ed estrazione di informazioni dai dati sarà cruciale per consentire non solo di migliorare le prestazioni operative ma di creare nuove opportunità di mercato. Servono quindi competenze per estrarre informazione e valore dai dati analizzandoli secondo schemi di indagine che possono anche essere diversi da quelli per i quali sono stati raccolti, rispondendo alle domande attraverso il riconoscimento di modelli emergenti dal comportamento dei dati e usando strumenti e algoritmi in pieno sviluppo.

---

<sup>1</sup> McKinsey Global Institute -THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD in collaboration with McKinsey analytics – December 2016

<sup>2</sup> disponibile a <http://www.istruzione.it/allegati/2016/bigdata.pdf>

Questa forma di innovazione guidata dai dati sta introducendo modelli di business completamente nuovi e può avere un forte impatto sulla qualità della vita. I sistemi di supporto alle decisioni evidence-based di tipo data-driven diventano quindi sempre più fondamentali e con essi lo sviluppo di nuovi paradigmi di decisione basati sull'utilizzo massivo di dati.

Anche secondo i dati dell'Osservatorio Big Data Analytics & Business Intelligence della School of Management del Politecnico di Milano (Rapporto 2017), per quanto riguarda le grandi imprese in Italia, nonostante un generale utilizzo dei *descriptive analytics*, strumenti che descrivono la situazione attuale e passata dei processi aziendali, con una crescita dell'11% rispetto all'anno precedente, l'area di maggiore interesse è quella dei *predictive analytics*, che consentono di effettuare previsioni sull'evoluzione del mercato e sulle strategie, già diffusi nel 73% dei casi (contro il 59% del 2016). Sono ancora indietro, invece, ma assumeranno sempre più importanza i *prescriptive analytics*, strumenti avanzati e basati su modelli di ottimizzazione capaci di proporre soluzioni sulla base delle analisi svolte, presenti solo nel 33% delle grandi imprese (rispetto al 23% del 2016) e ancora di più gli *automated analytics*, capaci di avviare autonomamente l'azione proposta secondo il risultato delle analisi, diffusi solo nel'11% delle organizzazioni, prevalentemente a livello pilota, pressoché in linea con quello dello scorso anno (10%).

In questo contesto si individuano due categorie di figure professionali: il Data Analyst o Data Expert competente per lo sviluppo e/o l'utilizzo di modelli ed algoritmi specifici per l'analisi e la valorizzazione dei dati da un lato, e il Data Scientist esperto di dominio di applicazione, quali le scienze sociali, il settore aziendale-organizzativo e quello economico e giuridico dall'altro. La domanda di tali figure professionali è cresciuta vorticosamente in questi ultimi anni e sembra destinata a crescere ulteriormente in futuro dato che la carenza di esperti in Data Science è uno dei fattori principali di limite alla crescita delle nuove opportunità economiche legate allo sfruttamento dei Big Data.

### **Il contesto internazionale**

In conseguenza della rilevanza del fenomeno e della crescente domanda di Data Scientists, la Data Science spesso inserita nel contesto più ampio della Data e Business Analytics ha assunto un ruolo fondamentale nell'offerta formativa dei maggiori atenei del panorama internazionale. Tra questi, a titolo di esempio, si ricordano:

- 1) Stanford Univ., USA (<https://statistics.stanford.edu/academics/ms-statistics-data-science>)
- 2) MIT, USA (<http://mitsloan.mit.edu/master-of-business-analytics/>)
- 3) New York Univ., USA (<https://cds.nyu.edu/academics/ms-in-data-science/>)
- 4) Imperial College, UK  
(<https://www.imperial.ac.uk/business-school/programmes/msc-business-analytics/>)
- 5) Edinburgh Univ., UK (<http://www.drps.ed.ac.uk/18-19/dpt/ptmscdatsc1f.htm>)



ASSOCIAZIONE ITALIANA DI RICERCA OPERATIVA

- 6) ETH Zurich, CH  
(<https://www.math.ethz.ch/studies/master-programmes/master-data-science.html>)
- 7) Amsterdam Univ., NL  
(<http://gss.uva.nl/content/masters/information-studies-data-science/data-science.html>)
- 8) TU Berlin, DE ([https://www.analytics.tu-berlin.de/data\\_analytics\\_lab/data\\_analytics\\_track/](https://www.analytics.tu-berlin.de/data_analytics_lab/data_analytics_track/))

Analizzando la struttura dei corsi in Data Science offerti dai più prestigiosi atenei internazionali si evince chiaramente il ruolo delle discipline fondanti che sono invariabilmente rappresentate dall'informatica, dalla statistica, dall'ottimizzazione matematica e dalla ricerca operativa. Tali discipline forniscono i principali strumenti metodologici a disposizione del data scientist e costituiscono una parte molto consistente dei crediti dei programmi di studio sia quando questi sono offerti a studenti di ambito STEM (*Science, Technology, Engineering and Mathematics*) sia non-STEM. A questo si affiancano contenuti provenienti dai settori applicativi specifici (ad esempio nel campo delle scienze sociali o naturali), contenuti trasversali quali quelli legati agli aspetti giuridici del trattamento del dato e contenuti specialistici legati ai dispositivi ed alle reti per l'acquisizione e l'elaborazione dei dati.

### **Il contesto nazionale**

Nel panorama italiano sono stati attivati negli ultimi anni diversi corsi in ambito Data Science ed in assenza di una classe di laurea magistrale specifica sono stati inquadrati principalmente nella classe LM-91 (TECNICHE E METODI PER LA SOCIETÀ DELL'INFORMAZIONE) quali, ad esempio,

- Roma La Sapienza, L.M. Data Science (in inglese)
- Cagliari, L.M. Data Science, Business Analytics e Innovazione
- Padova, L.M. Data Science (in inglese)
- Milano Bicocca, L.M. Data Science (in inglese)
- L'Aquila, L.M. Applied Data Science (in inglese)

e nelle classi LM-18 (INFORMATICA) quali il corso in Data Science and Business Informatics dell'Università di Pisa e LM-44 (MODELLISTICA MATEMATICO-FISICA PER L'INGEGNERIA) Data Science & Scientific Computing dell'Università di Trieste. Inoltre, in diversi casi sono stati attivati curricula specifici in Data Science in Lauree Magistrali di varie classi. Per una panoramica dell'offerta didattica sia in Italia che all'estero si veda anche il rapporto Big data@MIUR precedentemente citato.

In quasi tutte le esperienze presenti in Italia sono stati attivati quali corsi a scelta od obbligatori insegnamenti del Settore Scientifico Disciplinare MAT/09 oltre ad insegnamenti delle aree statistiche e di informatica ma, a differenza di quanto avviene nei corsi Data Science esteri, l'offerta formativa è molto più disomogenea e spesso presenta una forte accentuazione o verso contenuti applicativi o di tipo tecnologico che faticano a coprire le competenze di tipo più



ASSOCIAZIONE ITALIANA DI RICERCA OPERATIVA

metodologico in grado sia di utilizzare in maniera consapevole i modelli e gli algoritmi allo stato dell'arte che di svilupparne di nuovi al fine di proporre soluzioni innovative sulla base delle analisi svolte.

## **Obiettivi culturali della classe**

La Laurea Magistrale in Data Science deve fornire avanzate competenze tecnico-scientifiche nelle tecnologie dell'informazione, della statistica, della ottimizzazione matematica e dell'uso di modelli data-driven per la formazione di competenze sia per lo sviluppo di nuove tecnologie innovative per l'analisi e l'uso di dati, sia per lo sviluppo di metodologie di apprendimento dai dati, sia per l'applicazione di tali tecnologie a diversi possibili domini di applicazione. In sintesi, la laurea in Data Science deve formare una figura professionale multidisciplinare ed innovativa che, oltre a conoscere le modalità di formazione, di raccolta e di trattamento dei dati, conosce i meccanismi di funzionamento degli strumenti matematico-modellistici dell'analisi statistica e dell'ottimizzazione che gli permettano di progettare e condurre esperimenti, interpretare i risultati e, ove necessario, sviluppare soluzioni e metodologie originali per migliorare i processi di apprendimento, e per individuare soluzioni innovative per l'analisi e l'ottimizzazione dei processi di aziende, istituzioni e pubbliche amministrazioni.

Il Data Scientist deve avere inoltre la capacità di comprendere i problemi relativi ad affidabilità e sicurezza dei dati e dei sistemi software e dei sistemi di acquisizione e trasmissione dei dati e deve inoltre integrare le capacità metodologiche con la conoscenza della normativa giuridica relativa alla raccolta, trattamento e uso dei dati nonché degli aspetti etici relativi ai dati nei diversi ambiti applicativi.

## **Contenuti disciplinari e competenze trasversali indispensabili nella classe**

La classe Data Science deve permettere di articolare un percorso formativo coerente ed efficace ed in grado di accogliere studenti provenienti sia da un ambito STEM che non-STEM. A tal fine non solo deve offrire una solida base metodologica delle materie fondative della Data Science, con particolare riferimento alle discipline informatiche, statistiche ed alla ricerca operativa, ma deve anche prevedere un'ampia possibilità di discipline specifiche e trasversali su cui articolare indirizzi di specializzazione ed approfondimento nei diversi settori applicativi, dalle scienze sociali, all'economia ed alla gestione di impresa, all'industria 4.0, alle applicazioni sanitarie, alla biologia, alle scienze naturali ed alla fisica.

Si propone di articolare le discipline caratterizzanti su due livelli. Il primo gruppo è rappresentato dal nucleo fondamentale di discipline che forniscono le metodologie fondative della Data Science

come riconosciuto in tutti i corsi internazionali in modo definire un laureato completo e facilmente riconoscibile sul mercato interno ed internazionale. Il secondo gruppo prevede tre diversi percorsi di approfondimento, che possono essere articolati in modo tra loro alternativi ed in grado di coprire le diverse figure professionali di Data Scientist.

### **Discipline fondative la Data Science**

Le capacità suddette si acquisiscono attraverso insegnamenti avanzati di natura metodologica e attraverso insegnamenti dedicati a mettere in pratica le conoscenze acquisite in specifici settori applicativi. In particolare, si ritiene indispensabile che la classe preveda un nucleo fondamentale di discipline caratterizzanti tra le quali debba essere necessariamente selezionato un numero consistente di crediti (da 24 a 30 ed almeno 6 per ciascuna macro area) comprendente le discipline fondative della Data Science:

- Informatica (INF/01 Informatica e ING-INF/05 Ingegneria informatica)
- Ricerca Operativa ed Ottimizzazione (MAT/09 Ricerca Operativa)
- Statistica (SECS-S/01 Statistica, MAT/06 Probabilità e Statistica Matematica)

miranti a fornire le conoscenze statistiche, informatiche e della programmazione matematica di base necessarie per lo sviluppo degli strumenti software e delle infrastrutture necessarie per la raccolta, l'elaborazione, e l'organizzazione delle grandi moli di dati e dei modelli matematico-statistici utili per la loro analisi. Tale raccordo ed integrazione è fondamentale sia nel caso di corsi di laurea a prevalente accesso di studenti non-STEM in cui è fondamentale dare la necessaria base metodologica per la successiva applicazione a settori specifici, sia nel caso di corsi a prevalente accesso di studenti di area STEM, in cui possono essere utilizzati per fornire competenze maggiormente avanzate e specifiche al contesto Big-Data ed utili a formare Data Experts capaci sia di utilizzare in modo critico e consapevole gli strumenti esistenti sia di progettarne di nuovi per il trattamento, l'analisi e la valorizzazione dei dati.

### **Percorsi/Aree di approfondimento**

In aggiunta alle discipline fondamentali potrebbe essere opportuno prevedere vari gruppi di materie caratterizzanti finalizzate a completare la formazione dei Data Scientists in diversi settori applicativi ed a cui attribuire i crediti residui (fino ad un raggiungimento del massimo di 48 cfu caratterizzanti). In particolare si potrebbero ipotizzare le seguenti aree:

#### **Area Gestionale-organizzativa**

Tale area comprende settori disciplinari in ambito economico e gestionale, nonché discipline di matematica e statistica applicata, quali ad esempio

- Economia e Finanza
- Organizzazione e gestione aziendale
- Statistica e matematica per l'economia e la gestione
- Ottimizzazione ed informatica per l'economia e la gestione

### Area delle Scienze sociali, umane e naturali

Tale area comprende settori applicativi di estrema rilevanza per la Data Science, quali ad esempio

- Sociologia generale ed applicata
- Medicina e Biologia
- Discipline giuridiche
- Scienze naturali
- Scienze agrarie

### Area Tecnologico-Industriale

Tale area comprende le discipline relative all'acquisizione, al trattamento ed alla trasmissione dei dati nonché il settore applicativo industriale, quali ad esempio

- Tecnologie dell'informazione (Elettronica e Telecomunicazioni)
- Sistemi per l'elaborazione delle informazioni e
- Progettazione e controllo di impianti industriali e macchine
- Matematica industriale, analisi numerica

Con riferimento allo specifico contributo della Ricerca Operativa si evidenziano le metodologie di Ricerca Operativa che possono essere utilizzate in quanto di estrema rilevanza per Data Science sia sul piano metodologico sia su quello applicativo:

<b>Tema</b>	<b>Metodologie MAT/09</b>
<i>Algoritmi di apprendimento</i>	Ottimizzazione non lineare non vincolata e vincolata convessa e non convessa Tecniche di ottimizzazione globale Metodi basati sul rilassamento lagrangiano e del subgradiente Algoritmi di ottimizzazione su reti
<i>Apprendimento con rinforzo</i>	Programmazione dinamica Markov decision models
<i>Classificazione supervisionata e non supervisionata</i>	Algoritmi di decomposizione a blocchi Algoritmi di ottimizzazione online (o con errore) Algoritmi e modelli di graph clustering e partitioning Set covering e set partitioning Metodi a generazione di colonna e dualità
<i>Problemi decisionali su larga scala</i>	Metodi di decomposizione Metodi a generazione di righe e di colonne Tecniche metaeuristiche
<i>Analisi dei dati ed ottimizzazione</i>	Metodi di ottimizzazione stocastica e di ottimizzazione robusta

<i>Simulazione ed ottimizzazione black/grey box</i>	Tecniche di simulazione e generazione di scenari Metodi senza derivate Utilizzo integrato di modelli surrogati e ottimizzazione
<i>Teoria delle decisioni</i>	Alberi di decisione Indici di stocasticità e valore dell'informazione

## Principali sbocchi occupazionali del laureato magistrale in Data Science

L'evoluzione dei sistemi di produzione di beni e servizi rende il Data Scientist una figura professionale cruciale la cui domanda da parte del mercato e della pubblica amministrazione è attualmente molto forte ed è destinata ad aumentare e restare alta anche nel futuro. Gli sbocchi "naturali" si presentano in tutti i settori nei quali è necessario identificare problemi di sistema o processo che possono essere meglio risolti con una attenta analisi dei dati e successivamente utilizzare tecniche descrittive, predittive e prescrittive nei problemi decisionali.

Il percorso formativo consente quindi al laureato magistrale in Data Science di trovare occupazione presso piccole e medie imprese, grandi aziende, pubblica amministrazione, amministrazioni locali, Istituti di ricerca pubblici e privati e libera professione. Il dottorato di ricerca e i Master sono altre possibilità a cui lo studente in Data Science potrebbe rivolgersi al termine del suo percorso di studio.

## Elementi indispensabili per il raggiungimento degli obiettivi

I corsi di laurea della classe Data Science dovrebbero:

- prevedere che i laureati dispongano di buone competenze di lingua inglese (ad esempio, pari ad un livello B2)
- favorire un approccio formativo fortemente orientato al lavoro "sul campo" introducendo un numero minimo di CFU dedicati ad attività laboratoriali che permettano agli studenti di sperimentare le competenze metodologie acquisite su casi di studio concreti e di sviluppare la capacità di partecipare ed organizzare gruppi di lavoro
- prevedere e favorire tirocini formativi presso aziende ed enti di ricerca che permettano di applicare le conoscenze acquisite in contesti multidisciplinari.