

# Percorso Formativo “Predictive modeling and data science”

## Manifesto

### Docente di riferimento

Prof. Dario Malchiodi

### Sede di erogazione

Milano

### Obiettivi generali

Una delle conseguenze della diffusione delle tecnologie informatiche è la produzione di grandi quantità di dati in formato digitale. Questa presenza massiccia di dati, che caratterizza ormai molti settori della ricerca e dell'industria, sta stimolando la crescita di nuove figure professionali come quella del *data scientist*. Questi è una persona in grado di estrarre, integrare, organizzare, analizzare e presentare informazioni provenienti da numerose sorgenti di dati. Le iniziative in genomica, fisica delle particelle, astrofisica e neuroscienze nel settore della ricerca, così come la crescente rilevanza delle applicazioni basate sull'analisi di reti sociali sono un fattore trainante per la formazione di professionisti e ricercatori che sappiano sviluppare algoritmi in grado di filtrare, raggruppare e categorizzare massicci flussi di dati in rapida e costante evoluzione.

L'attività del *data scientist* richiede un insieme di competenze ampio e articolato. È necessario saper organizzare l'acquisizione di differenti tipologie di dati (testi, immagini, relazioni, geni, ...) strutturati in modalità diverse (flussi, gerarchie, reti, ...) e provenienti da sorgenti di vario tipo (Web, sensori, terminali mobili, ...), nonché la loro organizzazione in basi di dati opportunamente strutturate. Questi dati devono poter essere analizzati combinando tecniche statistiche, algoritmiche e di apprendimento automatico al fine di sintetizzare dei modelli predittivi.

### Sbocchi professionali

Il percorso forma esperti con un'approfondita conoscenza della disciplina informatica e delle sue applicazioni nei campi caratterizzati dalla necessità di estrarre informazioni in contesti in cui i dati provengono da fonti eterogenee o sono affetti da errori o incertezza, con una particolare sensibilità a valutare gli approcci da utilizzare in funzione della scala dei *data set* a disposizione. Specificamente, i laureati avranno competenze legate alla gestione dei dati su piattaforme distribuite, all'elaborazione di dati su larga scala e all'applicazione di tecniche intelligenti di analisi dei dati. Ciò permetterà loro di inserirsi proficuamente in contesti lavorativi focalizzati su tematiche quali l'analisi del rischio, la gestione avanzata dei dati (*data modeling, data warehousing e knowledge management*), l'analisi intelligente dei dati (*data mining, text mining, data quality*), la prevenzione antifrode, l'ottimizzazione in ambienti Web (SEO, *ad placement, marketing optimization*). I settori più promettenti in tal senso sono quelli legati a commercio elettronico, banche e finanza, telecomunicazioni, assicurazioni, sanità, trasporti e logistica, farmaceutica e pubbliche amministrazioni. Il profilo del laureato permetterà anche una prosecuzione degli studi verso il terzo livello della formazione, nell'ottica di un percorso professionale all'interno di laboratori di ricerca pubblici e privati.

## Piano formativo

Coerentemente con gli obiettivi generali, il piano formativo prevede un nucleo di insegnamenti relativi ai settori dei sistemi informativi, dei sistemi intelligenti e dei sistemi distribuiti, affiancati a insegnamenti che curino gli aspetti metodologici relativi alla statistica e all'ottimizzazione. Nella fattispecie: i) gli insegnamenti di *Metodi statistici per l'apprendimento* e di *Complementi di ricerca operativa* forniscono gli strumenti matematici necessari per comprendere e applicare gli strumenti e le metodologie introdotte nei rimanenti corsi; ii) l'insegnamento di *Sistemi distribuiti e pervasivi* permette di sviluppare le competenze necessarie per implementare e gestire le infrastrutture necessarie allo sviluppo di applicazioni distribuite; iii) l'insegnamento di *Gestione dell'informazione* rafforza le capacità degli studenti di trattare i dati nei moderni sistemi di supporto alle decisioni; iv) l'insegnamento di *Sistemi intelligenti* offre una panoramica sui differenti approcci informatici all'analisi dei dati. A partire da questo nucleo il percorso si sviluppa in due direzioni: una prima relativa alla gestione dei dati e una seconda relativa alla loro elaborazione e analisi.

Insegnamenti di indirizzo generale e metodologico:

- *Complementi di ricerca operativa (anni alterni, prossima attivazione nell'a.a. 2017/18)*
- *Gestione dell'informazione*
- *Metodi statistici per l'apprendimento* (propedeuticità: *Sistemi intelligenti*)
- *Sistemi distribuiti e pervasivi*
- *Sistemi intelligenti*

Modellazione, gestione e manipolazione dei dati:

- *Algoritmica per il Web*
- *Bioinformatica*
- *Programmazione Avanzata*
- *Sviluppo di applicazioni per dispositivi mobili* (propedeuticità: *Sistemi distribuiti e pervasivi*)

Data analytics:

- *Analisi di dati su larga scala* (propedeuticità: *Metodi statistici per l'apprendimento, Gestione dell'Informazione*)
- *Information retrieval*

Sono inoltre di interesse per il percorso una serie di insegnamenti relativi a specifiche metodologie di analisi dei dati, nonché alla loro modellazione e alla privacy nel loro rilascio:

- *Metodi e modelli per le decisioni (anni alterni, prossima attivazione nell'a.a. 2017/18)*
- *Modelli dei dati e DBMS di nuova generazione*
- *Privacy e protezione dei dati*
- *Simulazione (anni alterni, prossima attivazione nell'a.a. 2016/17)*
- *Visione artificiale*
- *Algoritmi euristici (anni alterni, prossima attivazione nell'a.a. 2016/17)*

## Informazioni aggiuntive

Le tesi di laurea possono essere svolte sotto la supervisione dei numerosi docenti del Dipartimento di Informatica che svolgono attivamente ricerca nell'ambito delle discipline coinvolte nel percorso, individualmente oppure all'interno di laboratori di ricerca.

## COMPATIBILITA' E SOSTENIBILITA'

### Docenti proponenti

*Prof. Paolo Boldi, Prof. Nicolò Cesa Bianchi, Prof. Dario Malchiodi, Dr. Massimo Santini, Prof. Giorgio Valentini*

### Sede di erogazione

Milano

### Insegnamenti e coperture

#### **Ipotesi di ripartizione degli insegnamenti su anni e semestri**

	<b>Primo semestre</b>	<b>Secondo semestre</b>	<b>CFU</b>
<b>I anno</b>	<i>Algoritmica per il Web</i> <i>Analisi dei dati su larga scala</i> <i>Programmazione avanzata</i> <i>Sistemi intelligenti</i> Competenze linguistiche	<i>Gestione dell'informazione</i> <i>Metodi statistici per l'apprendimento</i> <i>Information retrieval</i> <i>Sistemi distribuiti e pervasivi</i> Insegnamento a scelta	<b>57</b>
<b>II anno</b>	<i>Sviluppo di appl. per disp. Mobili</i> <i>Complementi di ricerca operativa*</i> Insegnamento a scelta Prova finale	<i>Bioinformatica</i> Prova finale	<b>63</b>
<b>Totale CFU</b>			<b>120</b>

\* Insegnamento erogato ad anni alterni: prossima erogazione a.a. 2017/18.