

## **Audizione CUN sul tema "Data Science" --- 14 giugno 2018**

### **Contributo del Gruppo di Ingegneria Informatica (SSD ING-INF/05)**

Il Gruppo di Ingegneria Informatica sostiene la proposta di una nuova classe in "*Data Science*" (o "Scienza dei Dati", se si preferisce la denominazione in italiano, che può essere considerata equivalente), perché essa fa riferimento ad una nuova figura professionale, che è abbastanza specifica ma ha una prospettiva di ampio spettro. Le classi attuali non riescono a cogliere questa esigenza, in alcuni casi (in particolare la LM-91, nell'ambito della quale sono stati attivati corsi di *Data Science*) perché molto generiche, in altri perché mirate ad una specifica area. In quest'ultimo caso ricade la classe LM-32 (ingegneria informatica), cui il settore scientifico-disciplinare ING-INF/05 è particolarmente legato: essa può formare ingegneri informatici con specifiche competenze in *Data Science*, ma non veri specialisti in essa, perché non ha l'ampiezza e l'attenzione alla pluralità di discipline qui proposta.

#### **Gli obiettivi culturali di questa classe e i contenuti disciplinari e le competenze trasversali indispensabili che dovrebbero essere fornite da tutti i corsi della classe.**

Obiettivo della classe di "*Data Science*" (o "Scienza dei Dati") è formare il "data scientist" ("scienziato dei dati"), una figura professionale emergente motivata dalle opportunità fornite dalla crescita vertiginosa (in quantità e varietà) dei dati digitali accumulati in contesti di ogni tipo. I dati possono avere natura, forma e qualità molto diverse, debbono essere opportunamente acquisiti, gestiti, interpretati, analizzati e diffusi, il tutto con opportune metodologie e strumenti. Data l'ampiezza delle attività, sono possibili diverse definizioni della figura del data scientist, che condividono da una parte la presenza di competenze di base tipiche di varie aree disciplinari e dall'altra la capacità di operare in settori diversi, che possono richiedere a loro volta competenze mirate.

Le aree metodologiche di riferimento includono certamente quelle della matematica, della statistica e dell'informatica e ingegneria Informatica, ma potenzialmente anche altre. I settori in cui la scienza dei dati si applica sono molteplici, ad esempio economia, ingegneria nelle sue tante anime (civile, industriale e dell'informazione), sociologia, fisica, biologia, medicina, scienze dell'ambiente e diverse discipline umanistiche. Ciascuno di questi settori prevede specificità che non vanno considerate puramente applicative ma necessitano di una integrazione sostanziale delle competenze. Vi sono poi discipline, in un certo senso complementari, come per esempio le misure, l'elettronica o l'automatica, che si collocano "a monte" nel processo di rilevazione dei dati perché forniscono metodi e strumenti per l'acquisizione o estrazione dei dati.

In sostanza, possono esistere diverse "varianti" di *data scientist*, con caratterizzazioni relative ai domini di interesse, e la classe deve certamente tenerne conto, ma ha un nucleo che, pur aperto a queste caratterizzazioni oltre che alle applicazioni, non va vincolato a nessuna di esse. Tutto questo richiede una grande attenzione nel definire tanto gli obiettivi formativi quanto i contenuti caratterizzanti, per evitare da una parte di indicare vincoli troppo stringenti e dall'altra di fornire una specifica troppo generica. Probabilmente, è opportuno indicare un nucleo metodologico fondamentale, comunque flessibile, cui affiancare riferimenti ai domini applicativi e alle discipline complementari, in termini abbastanza modulari.

I laureati nei corsi di laurea magistrale della classe "*Data Science*" dovrebbero presentare buona parte (non necessariamente tutte, perché si potrebbe rischiare superficialità) delle seguenti caratteristiche:

- possedere gli strumenti cognitivi di base per pensare analiticamente, creativamente, criticamente e in modo indagatore, e avere le capacità di astrazione e soluzione di problemi necessarie per affrontare sistemi complessi;
- possedere solide conoscenze sui principi della modellazione dei dati e della rappresentazione e gestione della conoscenza, sia in senso induttivo che deduttivo;
- possedere solide conoscenze sui modelli e le metodologie statistiche e matematiche;
- possedere solide conoscenze in settori più specialistici, quali quelli del *data mining* e dell'apprendimento automatico (*machine learning*);
- possedere conoscenze sulle modalità di comunicazione dei risultati di processi di elaborazione dati, ricorrendo, ad esempio a strumenti di visualizzazione e di redazione di report orientati a tipi diversi di pubblico;
- avere capacità di gestione di sistemi *data-intensive*, in particolare devono essere in grado di operare su grandi collezioni di dati, anche di tipo eterogeneo e prodotti a grande velocità, al fine di integrarli, - in particolare sapendone gestire la provenienza e la qualità - e di eseguire analisi tematiche approfondite, ricavandone conoscenza utile a migliorare il processo decisionale;
- conoscere le problematiche delle infrastrutture tecnologiche abilitanti alla data science, quali, fra l'altro, calcolo ad alte prestazioni, *cloud* e *service-oriented computing*, reti di comunicazione, *Internet-of-things*, risorse pubbliche di dati, anche in formato aperto;
- saper operare in gruppi interdisciplinari costituiti da esperti con competenze in specifici ambiti delle tecnologie informatiche o in specifici domini applicativi, al fine di generare valore a partire dai dati;
- conoscere le problematiche relative alle normative connesse all'uso dei dati, con riferimento, tra gli altri, ai problemi di proprietà, restrizioni di licenza, *privacy* e sicurezza;
- possedere competenze relative all'utilizzo dei dati come strumento per l'innovazione e lo sviluppo economico, competenze quindi di natura economica ed economico-aziendali
- comprendere le conseguenze dell'uso delle tecniche di analisi dei dati e della conoscenza da essi estratta, in modo da rispettare il quadro normativo ed i principi etici che sono alla base della nostra società.

Per raggiungere questi obiettivi, è importante prevedere contenuti disciplinari indispensabili appartenenti agli ambiti della Matematica, della Statistica, dell'Informatica e dell'Ingegneria Informatica, essenziali per costruire una Scienza dei Dati, corredandoli con alcune importanti nozioni giuridiche ed economiche (anch'esse indispensabili) riguardanti l'uso appropriato di queste competenze. D'altra parte, i contenuti ai quali la Scienza dei Dati si applica, appartenenti a discipline pure fondamentali come l'economia, la sociologia, l'ingegneria, le varie scienze come la biologia, la fisica, la chimica, le discipline umanistiche, potrebbero essere declinati in maniera diversa nei vari corsi di laurea e non andrebbero inseriti tra le discipline caratterizzanti, altrimenti si rischia, da un lato, di creare una gabbia troppo costrittiva e, dall'altro, di favorire una eccessiva genericità.

In considerazione della valenza tanto scientifica quanto professionalizzante di questo percorso formativo, l'ammissione ai corsi di laurea magistrale della classe richiede il possesso di requisiti curriculari che prevedano un'adeguata padronanza di metodi e contenuti scientifici propedeutici a quelli di almeno uno degli ambiti disciplinari caratterizzanti l'ordinamento della classe di laurea magistrale in Data Science. È comunque opportuno prevedere la possibilità di proporre corsi cui possano accedere laureati di classi anche molto diverse fra loro, con conseguente organizzazione dei percorsi di studio diversificati al fine di compensare, almeno in parte, le diverse competenze.

**Quali potrebbero essere dei naturali sbocchi professionali, o sbocchi verso il proseguimento degli studi, coerenti con gli obiettivi della classe.**

Gli sbocchi occupazionali e professionali:

- nel settore terziario, con particolare riferimento ai servizi innovativi basati sui dati, eventualmente fruibili *on-line*, in modalità mobile, e/o basati su *social network*;
- nel settore industriale, con particolare riferimento ai processi innovativi nei quali le decisioni di livello operativo, tattico/manageriale e strategico/direzionale sono basate su informazioni estratte in modo tempestivo e sistematico a partire da dati raccolti internamente ed esternamente agli stessi processi industriali; particolare attenzione va rivolta all'introduzione di automazione massiccia derivante dal programma Industria 4.0
- nel settore primario, con particolare riferimento all'uso delle tecnologie digitali nello sfruttamento delle risorse naturali (agricoltura, estrazione di minerali etc.);

Gli sbocchi verso il proseguimento degli studi:

- nell'ambito scientifico, con particolare riferimento alle scienze come la fisica sperimentale, la chimica e la biologia, ma anche le "*X-informatics*", basate essenzialmente sul paradigma della scoperta scientifica guidata dai dati (a titolo di esempio, la bioinformatica, l'astroinformatica e la "*brain informatics*")
- nell'ambito economico-sociale, ad esempio in settori quali econometria, *marketing* quantitativo, *digital sociology*

**Se sia necessario introdurre altri elementi (per esempio presenza obbligatoria di tirocini o stage, attività laboratoriali, competenze linguistiche, eccetera) indispensabili per il raggiungimento degli obiettivi della classe.**

Come per molte altre classi di laurea, ai laureati di questa classe si richiede di essere capaci di comunicare efficacemente, in forma scritta e orale, in almeno una lingua dell'Unione Europea, oltre l'italiano, anche con riferimento ai lessici disciplinari.

Ai fini indicati, i curricula dei corsi di laurea magistrale della classe

- prevedono (fra i requisiti curriculari di accesso) almeno la conoscenza della lingua inglese;
- prevedono lezioni ed esercitazioni di laboratorio oltre ad attività progettuali autonome e attività individuali in laboratorio;

- possono prevedere attività esterne (ma è probabilmente opportuno non renderle obbligatorie), come tirocini formativi, presso aziende e industrie, enti pubblici o istituti di ricerca, laboratori, oltre a soggiorni di studio presso altre università italiane ed europee;
- culminano in un'attività di progettazione o di ricerca o di analisi di caso, che dimostri la padronanza degli argomenti e degli strumenti e la capacità di produrre in modo autonomo analisi organiche e approfondite di insiemi di dati di interesse.

### **Osservazioni aggiuntive**

Un requisito rilevante è quello della flessibilità, da due punti di vista.

In primo luogo, va ricordato che l'area è estremamente dinamica, con una definizione che potrebbe essere soggetta ad evoluzione, e quindi, se è importante cogliere l'occasione della corrente fase di manutenzione delle classi per introdurre una classe in *Data Science*, la sua struttura non deve essere troppo rigida, né negli obiettivi formativi né nella specifica degli ambiti e dei settori.

In secondo luogo, è importante tenere presenti altre possibili novità nelle aree multidisciplinari che coinvolgono l'informatica: segnaliamo l'importanza della classe LM-91, che fu definita vent'anni fa in modo forse visionario e venne inizialmente utilizzata ben poco, ma ha poi permesso l'istituzione di diversi corsi di *Data Science*, certamente non prevedibili all'epoca. Sembra quindi importante prevedere una classe con finalità simili a quelle della LM-91 (eventualmente insieme a quelle di altre, quali la LM-43), in grado di accomodare nuove necessità e tendenze che emergano nel prossimo futuro, senza avere collocazione nelle classi specifiche già esistenti.