

CUN 14 giugno 2018

Audizione su una possibile nuova classe di laurea magistrale dedicata alla *Data Science*.

Intervento di Enrica Amato, Presidente AIS, Associazione Italiana di Sociologia.

La proposta di creare una classe di laurea magistrale in *Data Science* è a mio parere molto opportuna, in quanto essa risponde alle esigenze espresse da un mutato mercato del lavoro. L'analisi di grandi moli di dati permette infatti applicazioni in diversi domini di conoscenza: industriali, ambientali, economici, politici e sociali. La riconfigurazione delle nostre società come società *data intensive* richiede un apparato di tecniche di analisi e di trattamento dei dati automatiche (*machine learning, data mining, pattern recognition, visual analytics, etc.*), che pongono una serie di sfide metodologiche, che vanno dalla fase di preparazione dell'analisi dei dati fino a quella di visualizzazione dei risultati. Sarebbe un grave errore ritenere che queste sfide siano puramente tecniche; è necessario infatti formare laureati che siano consapevoli degli orizzonti teorici che sottendono l'era digitale. Solo una conoscenza epistemologicamente fondata del processo di costruzione dei dati può mettere al riparo da posizioni induttiviste di matrice vetero-positivista, che giustificano una nuova forma di "dataismo", secondo cui l'abbondanza di dati produce in se stessa risultati, senza la necessaria mediazione analitica di teorie, concetti, metodi.

Come è necessario che gli studenti sappiano usare metodi e strumenti propri della *Data Science* per acquisire nuove conoscenze nei vari domini della scienza, è altrettanto importante che siano in grado di valutare le ricadute degli approcci computazionali, che si riflettono sui contesti produttivi, organizzativi e sociali. Gli algoritmi di calcolo sono come è noto sempre più frequentemente utilizzati dai decisori politici per la programmazione, il monitoraggio e la valutazione delle politiche pubbliche. In questo contesto di *governance* riveste un'importanza fondamentale la *data awareness*, vale a dire la consapevolezza dei cittadini di essere generatori di dati aperti all'utilizzo da parte di terzi, anche a fini commerciali. I recenti cambiamenti dei regolamenti comunitari sulla protezione dei dati - con l'adozione del *General Data Protection Regulation (GDPR)* - manifestano chiaramente che la consapevolezza di quali siano i processi socio-tecnici che assemblano i dati è da considerarsi oggi un requisito indispensabile di cittadinanza.

Un progetto di laurea magistrale in *Data Science*, che si ponga all'incrocio tra la computer science, la statistica e i diversi domini di conoscenza delle scienze, ha l'obbligo di promuovere l'interdisciplinarietà, favorendo l'integrazione di dati e metodi. Per questo motivo la nuova classe deve essere necessariamente trasversale e prevedere la compresenza di scienze e discipline formali, al fine di temperare

aspetti teorici e approcci computazionali e di includere in questo progetto anche un sapere critico.

L'introduzione di settori scientifico disciplinari appartenenti all'area 14, in particolare a quelli sociologici, a cui si è fatto ampiamente ricorso in esperienze internazionali eccellenti (il *Master Degree in Social Data Science dell'Oxford Internet Institute*, o il *Master Degree in Web Science del Web Science Institute dell'Università di Southampton*), è determinante per incoraggiare l'analisi sistemica e critica degli usi, delle potenzialità e dei limiti degli approcci computazionali, e lo studio delle conseguenze (organizzative, metodologiche, sociali e politiche) che questi possono generare nei differenti contesti aziendali, amministrativi e altro in cui vengono inseriti.

Basti pensare a questo proposito all'enorme impatto che lo sviluppo dei *social network* ha avuto dal punto di vista politico, economico e relazionale, e alla necessità di sviluppare strumenti per la loro analisi capaci di tutelare anche la *privacy* dei cittadini. In aggiunta a questo nei nuovi ecosistemi delle *smart cities* e delle *smart communities* l'implementazione delle soluzioni di "internet delle cose" renderà imprescindibile l'adozione di competenze socio analitiche nella progettazione e utilizzo dei data base dinamici che raccoglieranno i nuovi massicci grafi determinati dalle relazioni tra gli oggetti.

a. Obiettivi culturali

La classe di laurea in *Data Science* dovrebbe formare figure sempre più richieste nel contesto della digitalizzazione della società e dell'industria 4.0, in grado di rispondere alle sfide sociali e tecnologiche legate all'emergere di piattaforme globali di memorizzazione ed elaborazione dei dati e di contribuire allo sviluppo della cultura del dato nelle istituzioni pubbliche, nelle aziende private e nelle amministrazioni locali. In particolare, tra gli obiettivi culturali segnalerei la valorizzazione degli *open data* e il loro uso per lo sviluppo di servizi più efficienti per i cittadini, e l'integrazione dei *big data* all'interno dei processi di analisi economica, politica e sociale.

b. Contenuti disciplinari

La classe di laurea magistrale in *Data Science* dovrebbe essere caratterizzata da un'offerta didattica interdisciplinare capace di coniugare contenuti più tecnici dell'ingegneria, dell'informatica e della statistica con conoscenze teoriche dei principali domini di applicazione in campo politico e sociale. Nello specifico, si rendono necessarie conoscenze nel campo dell'epistemologia delle scienze sociali, degli studi sociologici su scienza, tecnologia e società, e competenze nelle tecniche

di ricerca sociale più tradizionali - saperi incardinati nel settore della Sociologia Generale e della Metodologia della ricerca - che possono utilmente affiancare e integrare i metodi computazionali della *Data Science*.

c. Competenze trasversali

Autonomia di giudizio

Il laureato magistrale in *Data Science* deve avere la capacità di progettare, implementare e monitorare lo sviluppo di soluzioni ad uso intensivo di dati, valutando l'impatto di tali soluzioni nel contesto in cui sono calate, sia in merito agli aspetti tecnici e metodologici, che agli aspetti organizzativi. Il laureato magistrale in *Data Science* dovrebbe saper giudicare autonomamente le implicazioni economiche, sociali ed etiche che i processi di datificazione possono creare nei contesti in cui si realizzano.

Abilità comunicative

Il laureato magistrale in *Data Science* deve essere in grado di interagire alla pari con specialisti di diversi settori applicativi al fine di comprenderne le specifiche esigenze e di comunicare in modo chiaro e comprensibile soluzioni ed aspetti tecnici adottati. In particolare, deve saper addestrare collaboratori, coordinare e partecipare a progetti di ricerca, innovazione e sviluppo. Il laureato magistrale in *Data Science* deve poter utilizzare fluentemente, in forma scritta e orale, la lingua inglese, con riferimento anche ai lessici disciplinari.

Capacità di apprendimento

Il laureato magistrale in *Data Science* deve avere la capacità di acquisire in modo autonomo nuove conoscenze di carattere teorico e tecnico dalla letteratura scientifica e dall'esperienza, apprendendo anche metodi e cornici teoriche non previste nel curriculum. Per questo motivo andrebbe offerta una didattica che stimoli costantemente il confronto disciplinare.

d. Sbocchi professionali

Il laureato magistrale in *Data Science* può trovare occupazione in qualsiasi ecosistema lavorativo che preveda un uso estensivo dei dati. Si prevedono almeno i seguenti sbocchi occupazionali:

- Aziende e centri di ricerca che operano nel settore del *data providing and data brokerage*,
- Società ed enti pubblici coinvolti nella gestione di grandi moli di dati (per esempio: Agenzia delle entrate, Corte dei conti, Istat, Poste italiane, Trenitalia, agenzie ambientali)
- Laboratori e fondazioni di ricerca e sviluppo, pubblici e privati
- Assicurazioni e istituzioni finanziarie
- Società di consulenze nei settori IT e *Data processing*

Le professioni collegate sono:

- *Data scientist*
- *Web scientist*
- *Web analyst*
- *Data officer*

e. Presenza obbligatoria di tirocini e stage

Tirocini e stage in imprese e enti coinvolti in progetti di industria 4.0 o di digitalizzazione della pubblica amministrazione sono da considerarsi momenti di formazione importanti per la figura professionale che il corso dovrà formare.

f. Attività laboratoriali

La presenza di idonee attività laboratoriali è di certo indispensabile. Laboratori di software per la gestione e organizzazione dei big data, così come per la costruzione di tecniche e di modelli di analisi sono fondamentali, ma anche attività laboratoriali mirate alla formulazione di disegni di ricerca e di innovazione digitale che tengano in considerazione i dati come costrutti socio-tecnici.

g. Competenze linguistiche

Il laureato magistrale in *Data Science* deve saper utilizzare fluentemente, in forma scritta e orale, la lingua inglese, con riferimento anche ai lessici disciplinari. In particolare, una parziale didattica in lingua inglese dovrebbe preparare gli studenti all'interazione in ambito professionale in tale lingua.