

Roma, 24 luglio 2019

Prof. Giuseppe VALDITARA
Capo Dipartimento per la Formazione
Superiore e la Ricerca del MIUR
Via Michele Carcani, 61
00153 Roma

giuseppe.valditara@miur.it

SIS/345

Oggetto: Proposta di laurea magistrale in *Data Science*.

Caro Professore,

a seguito di quanto già comunicato al Presidente CUN, Prof. Antonio Vicino, in data 09.07.2019, Le segnalo le profonde perplessità mie e del Consiglio Direttivo della Società Italiana di Statistica in merito alla proposta di laurea magistrale in *Data Science* approvata dal CUN nella seduta del dicembre scorso.

Siamo consapevoli che i principi ispiratori dei contenuti della classe siano frutto del lavoro collegiale e prezioso del CUN durante lo scorso anno, basato anche sulle audizioni che furono effettuate nel mese di luglio 2018 e alle quali la SIS già portò le sue istanze.

Il tema è molto rilevante e la necessità di formare giovani in grado di padroneggiare le tecniche statistiche nel *Data Science* è sempre più sentita e urgente.

La statistica è uno dei linguaggi sovradisciplinari per eccellenza e oggi è di assoluta rilevanza, tanto che i laureati in questa materia trovano lavoro subito in multinazionali, istituti di ricerca, società di marketing e assicurazioni. Ma anche lontano dal mondo aziendale: energia, ambiente, scienze mediche, welfare e cultura sono tutti ambiti in cui servono figure in grado di estrarre le informazioni rilevanti da immensi giacimenti di dati, buttando via il "rumore".

Per questo ci permettiamo di tornare sul tema. Infatti, è nostra opinione che per cogliere la domanda di *data scientist* crescente in Italia e nel mondo, il ruolo di questa nuova laurea non debba essere solo quello di integrare le attuali lauree in Informatica con qualche insegnamento di Statistica, ma al contrario, quello di colmare un vuoto attualmente esistente tra le figure tecniche (statistici, informatici) e gli esperti di dominio. Le classi di laurea attuali mostrano una chiara prevalenza di uno solo degli ambiti richiesti, quello matematico statistico (per le lauree magistrali in Statistica), quello informatico (per le lauree magistrali in Informatica ed Ingegneria informatica) o quello economico-aziendale (Economia, Economia Aziendale, Sociologia, ecc.). Con la costruzione della nuova classe di laurea bisognerebbe contribuire a creare la nuova figura professionale del *data scientist* molto richiesta sul mercato del lavoro nazionale ed

internazionale, pertanto non si dovrebbero semplicemente giustapporre competenze tecniche diverse. Quello che si otterrebbe altrimenti sarebbe solo un'ulteriore figura tecnica, già presente tra gli sbocchi delle attuali lauree magistrali, ovvero ottenibile attuando una mera manutenzione delle classi delle lauree magistrali in Statistica (aggiungendo solo un po' di Informatica) o in Informatica.

Il tema dei Big Data pone importanti problemi di natura tecnologico-informatica, ma è improprio ascrivere alla *Data Science* solo l'immagazzinamento, la gestione e l'interrogazione di grossi database relegando ad un ruolo secondario la valutazione della qualità del dato in entrata, la sua pulizia, la sua analisi e la sua modellazione, tutte competenze che rientrano tra quelle proprie della figura professionale dello statistico. Inoltre, non è più concepibile il trattamento del dato statistico al di fuori di un suo utilizzo in chiave inferenziale-induttiva, allo scopo di ottenere non solo una descrizione dei fenomeni, ma anche una ricostruzione del plausibile meccanismo generatore del dato stesso, così da poter utilizzare i risultati delle analisi per capire la realtà, prevedere i comportamenti futuri e le conseguenze di interventi e azioni politiche. Tutti questi strumenti, fondamentali per un *data scientist*, rientrano nelle competenze professionali dello statistico.

Come dimostrano le lauree in *Data Science* già attivate nelle classi attualmente esistenti, al momento non mancano gli spazi progettuali per la definizione di profili di *data scientist* con un sofisticato bagaglio di tipo tecnico (con la prevalenza spesso di competenze informatiche). Quello che manca, invece, è una classe di laurea che permetta la formazione di una nuova figura, fortemente interdisciplinare, focalizzata sui temi della Statistica e dell'Informatica e con marcate competenze di tipo scientifico-tecnologico, biomedico, economico-aziendale, giuridico, sociopolitico e umanistico. I profili più richiesti sul mercato del lavoro delineano con chiarezza due tipi distinti di figure professionali di *data scientist*: la prima con un focus sulla raccolta, gestione e trattamento dei dati di tipo complesso ad elevata dimensionalità, la seconda, invece, con un focus marcato sull'analisi dei dati e sulla costruzione di modelli analitici, di tipo interpretativo e predittivo, per dati di tipo complesso, quale necessario e indispensabile supporto alle decisioni nei diversi domini applicativi.

Nella proposta attuale della nuova classe di laurea magistrale in *Data Science* le discipline statistiche sembrano ricoprire una mera funzione integrativa e di supporto. Così al futuro laureato magistrale in *Data Science* mancheranno le competenze per "far parlare" i dati, ovvero per intercettare e quantificare le relazioni tra variabili, per mettere in luce le variabili importanti, per comprendere i meccanismi che regolano i fenomeni.

Le discipline caratterizzanti la nuova proposta delineano una struttura facilmente incardinabile in dipartimenti di Ingegneria o di Informatica, dove solo in alcuni casi ci sono corsi e docenti di Statistica, ma molto meno proponibile per dipartimenti con una forte presenza delle discipline statistiche o che comprendano corsi di laurea in Statistica nella loro offerta formativa. Pur non volendo penalizzare gli statistici incardinati in dipartimenti di Ingegneria o di Informatica, anzi ritenendo fondamentale la valorizzazione del loro lavoro, vogliamo evidenziare la rigidità della

proposta attuale che contraddice l'idea di maggiore flessibilità che dovrebbe ispirare il lavoro di revisione delle classi per tenere il passo con una realtà culturale e lavorativa in continuo movimento.

Molti corsi di laurea magistrali in *Data Science* già istituiti in Italia e all'estero e già produttivi di laureati, nati quale tempestiva risposta alla crescente domanda globale e quasi sempre concertati con le realtà produttive locali, difficilmente o per nulla rientrerebbero nella classe delle lauree magistrali in *Data Science* così come attualmente proposta.

Riteniamo che quelle evidenziate costituiscano delle limitazioni importanti delle potenzialità della nuova classe di laurea, e rappresentino un'occasione persa nella costruzione della nuova figura professionale del *data scientist*.

Pertanto, è nostra convinzione che si debba intervenire sulla definizione dei CFU minimi delle attività formative nei diversi ambiti disciplinari, rafforzando lo spirito di elevata interdisciplinarietà proprio della *Data Science*, attribuendo uguale peso all'ambito matematico-statistico e a quello informatico. In tal modo si consentirebbe una maggiore flessibilità in sede di progettazione locale offrendo la possibilità di formulare due profili di *Data scientist*: il primo con prevalenza di materie informatiche nell'ambito della gestione del dato grezzo, il secondo più orientato alle discipline statistiche ed all'ambito della Business Analytics. Più specificatamente, a nostro parere, una struttura più equilibrata di quella attuale, e maggiormente confacente alle necessità del mercato del lavoro, richiederebbe un minimo di 21 CFU dell'ambito disciplinare matematico-statistico (in luogo degli attuali 15), portando a 48 il numero minimo di CFU riservati alle attività caratterizzanti nella classe (come, del resto, avviene già nella maggior parte delle lauree magistrali esistenti). Qualora, per offrire maggiore discrezionalità nella scelta delle materie affini, si intenda mantenere il numero di 42 CFU per il minimo delle attività caratterizzanti, riteniamo che dovrebbero essere attribuiti 18 CFU ad entrambi gli ambiti matematico-statistico ed informatico.

Infine vogliamo segnalare come ulteriori iniziative di revisione di classi di laurea quali la LM91, che le orientino verso applicazioni della *Data Science* in ambiti specifici propri dell'Area 13, come la Business Analytics, senza una presenza adeguata delle discipline statistiche e delle statistiche applicate (Statistica Sociale, Demografia e Statistica Economica) riproporrebbero lo stesso nefasto sbilanciamento culturale e sarebbero da evitare, nell'interesse di tutti ed in primis dei futuri studenti.

Certa dell'attenzione che vorrà porgere a questa nostra richiesta, l'occasione mi è gradita per inviare i miei più cordiali saluti.

Il Presidente della
Società Italiana di Statistica
Monica Pratesi
e il Consiglio Direttivo SIS

