

II RELAZIONE NEXT GEN.



Sala del Basolato, Fiesole, Martedì 22 novembre 2022

Il Dott. Lorenzo Testa, Dottorando in Statistica presso l'Università Bocconi di Milano e allievo ordinario in Economia alla Scuola Superiore Sant'Anna di Pisa, presenta il suo progetto – NextGenFDA - nell'ambito della *Data Science*, un concetto di statistica legato alla scienza di elaborazione dei dati.

Lo studio ha come oggetto lo sviluppo metodologico di metodi statistici utili per comprendere grandi quantità di *dati biomedici*, estraendo nuove informazioni che possono essere utilizzate per migliorare le nostre capacità di prevenzione e cura delle malattie.

In particolare, NextGenFDA si propone di studiare dati biomedici che possono essere osservati lungo domini continui, come il tempo e lo spazio. Questi dati, che prendono il nome di *dati funzionali*, sono estremamente versatili, perché sono in grado di catturare le forme e *pattern* di fenomeni complessi, rappresentandoli a un'incredibile risoluzione, e non sono sensibili a errori occasionali di misurazione. Allo stesso tempo, tuttavia, essi risentono di problematiche dovute alla loro complessità e alla loro instabilità, entrambe caratteristiche che derivano proprio dalla loro elevata risoluzione. NextGenFDA prova a rimediare a questi due problemi, sviluppando metodi statistici all'insegna della *scalabilità algoritmica* (cioè metodi che possono gestire efficientemente gigantesche quantità di dati) e della *stabilità statistica* (cioè metodi che misurano la robustezza e la riproducibilità dei loro risultati attraverso meccanismi di perturbazione).

NextGenFDA si propone di analizzare dati biomedici afferenti a tre ambiti applicativi principali. Innanzitutto, i *dati epidemiologici*, come quelli degli andamenti pandemici che abbiamo imparato a osservare negli ultimi anni, si prestano a una rappresentazione funzionale, che può aiutare ricercatrici e ricercatori a studiare i nessi causali tra diffusione di virus e batteri, pattern di mortalità e mobilità, fattori coadiuvanti e fattori limitanti del contagio. Inoltre, i *dati neurologici* generati dai dispositivi fMRI (macchinari di imaging attraverso risonanza magnetica) possono essere interpretati come dati funzionali e utilizzati per mappare quali aree del cervello si attivino in risposta a stimoli di varia natura. Ciò può contribuire in maniera decisiva ad un'attenta analisi delle differenze nel "funzionamento"

delle menti di soggetti sani e di soggetti affetti da patologie neurodegenerative. Infine, i *dati genomici* possono essere studiati per aumentare la nostra comprensione dei collegamenti tra struttura del DNA e probabilità di sviluppare determinate patologie.

Un progetto che oltre ad essere scientificamente ambizioso, ha la potenzialità di fornire importanti strumenti per intraprendere scelte su ogni livello, soprattutto in ambito politico e governativo.