

Master Universitario
"ANALISI DATI PER LA BUSINESS INTELLIGENCE"
A.A. 2018-2019

Titolo della tesi: DATA SCIENCE PER LA FILANTROPIA. *Agenda 2030: studio pilota per un classificatore dei progetti presentati alla Fondazione CRC*

Autore: Andrea Selva

Abstract

Il presente lavoro muove da un interrogativo: gli approcci, le metodologie e le architetture della Business Intelligence e della Data Science possono ricoprire anche all'interno di fondazioni e organizzazioni non profit un ruolo nevralgico nelle scelte strategiche, di governance, operative e nei processi di accountability e advocacy?

Tale interrogativo assume ulteriore importanza se collegato ad altre due grandi correnti che stanno catturando sempre di più l'attenzione degli addetti ai lavori: la prima, definita *Data for Good*, implica l'uso dei dati per finalità umanitarie e sociali, la seconda coincide con l'importanza crescente assunta dagli *Open Data*. Le profonde trasformazioni in atto a livello globale e nazionale non lasciano indifferenti neanche le Fondazioni di Origine bancaria, indirizzandole verso un processo di ridisegnamento della propria mission e delle modalità operative di intervento tuttora in corso. Da qui, l'idea di esplorare il collegamento tra le richieste di contributo degli Stakeholder della Fondazione CRC e l'Agenda 2030 attraverso un'analisi testuale e la realizzazione di uno studio pilota per un classificatore delle richieste di contributo potenzialmente rilevanti per il 17 SDGs.

Nel primo capitolo si presenterà brevemente la natura e la mission della Fondazione CRC, i 17 Sustainable Development Goals (SDGs - Obiettivi di Sviluppo Sostenibile) e le motivazioni per cui un allineamento al framework delineato dall'Agenda 2030 per lo Sviluppo Sostenibile possa essere produttivo per la mission delle organizzazioni non profit.

Nel secondo capitolo verrà presentata la prima parte del lavoro, con finalità prettamente esplorativa: un'analisi testuale, attraverso la ricerca di specifiche keyword, volta a selezionare le richieste di contributo presentate alla Fondazione potenzialmente rilevanti per il framework dell'Agenda 2030. I risultati suggeriscono che il numero di richieste potenzialmente rilevanti siano aumentate sensibilmente tra il 2016 e il 2018 e che gli obiettivi di sviluppo sostenibile più rappresentati siano l'*SDG4 Quality Education*, l'*SDG8 Decent Work And Economic Growth*, l'*SDG10 Reduced Inequalities* e l'*SDG11 Sustainable Cities And Communities*. Delle cinque aree in cui si articola l'Agenda 2030, due risultano principalmente importanti: l'area Persone e Prosperità con, rispettivamente, 301 e 468 progetti potenzialmente rilevanti deliberati nel periodo preso in considerazione.

Nel terzo capitolo si illustrerà la seconda parte del project work: in questo caso il task sperimentale consiste nella creazione di un modello che, dato un documento (richiesta di contributo) e k classi (i 17 SDGs), permetta di determinare quali SDGs siano più rilevanti per il testo in input. Più precisamente al sistema viene proposto un insieme di esempi di addestramento (training set), con l'obiettivo di fornire informazioni sufficienti per apprendere una funzione di classificazione, tale per cui sarà in grado di predire la classe del documento anche per elementi che non appartengono al training set. Verrà quindi illustrato il funzionamento del sistema, la creazione del dataset, la procedura di annotazione, la presentazione e discussione dei risultati. I risultati mostrano da un lato che per ottenere prestazioni utili è necessario disporre di quantità di dati annotati ben superiori rispetto a quelle disponibili oggi, e dall'altro che la soluzione prospettata, basata su architettura neurale, è adatta per il task di classificazione considerato: sarebbe quindi molto interessante estendere la sperimentazione in sinergia con altre Fondazioni di origine bancaria e/o enti e organizzazioni sensibili all'Agenda 2030.