

Uno strumento per indagare i sistemi di decisione basati sui dati adottati nella Pubblica Amministrazione italiana

Suania Acampa (Università di Napoli Federico II); Roberto Artiaco (Università di Napoli Federico II)

Il lavoro che si intende presentare è parte integrante di un più vasto progetto PRIN che ha l'obiettivo di esplorare come la datificazione delle pubbliche amministrazioni stia trasformando il concetto di cittadinanza in Italia. La specifica fase del progetto a cui questo lavoro fa riferimento, si pone l'obiettivo di esaminare in che modo e in quale misura le amministrazioni pubbliche italiane stanno adottando sistemi data-driven che influenzano le decisioni, con l'obiettivo di creare uno strumento capace di esplorare e mappare in maniera automatica l'impiego di questi sistemi.

Nonostante il processo di digitalizzazione della PA italiana, preveda, tra le altre cose, l'utilizzo di tecnologie ad alta intensità di dati, ci sono ancora poche ricerche dettagliate su questo tema che riescano a mettere in luce in che modo la PA sta implementando sistemi di decisione basati sui dati. Se da una parte, l'uso di questi sistemi permetta di migliorare l'efficienza e l'efficacia dei servizi pubblici ottimizzando, ad esempio, l'allocazione delle risorse e la personalizzazione dei servizi per rispondere meglio alle esigenze dei cittadini; d'altra parte, l'implementazione di sistemi data-driven solleva preoccupazioni significative: la qualità dei dati e il modo in cui vengono raccolti, analizzati e utilizzati possono generare risultati parziali e non corretti che possono portare decisioni discriminatorie che potrebbero perpetuare o amplificare pregiudizi esistenti (O'Neil, 2016).

Attualmente, gli enti governativi italiani (sia a livello centrale che locale) non offrono un quadro chiaro di come i sistemi basati sui dati stiano influenzando le decisioni nei servizi pubblici. Per questi motivi, lo strumento che si intende creare sarà fondamentale per facilitare ulteriori studi e analisi su questo fenomeno, incrementando così la conoscenza, la comprensione e la consapevolezza di questi sistemi. Lo strumento che si intende realizzare funzionerà come un vero e proprio motore di ricerca utile anche a istituzioni, cittadini e altri attori sociali per effettuare ricerche ed entrare in contatto con su una serie di documenti e siti web della Pubblica Amministrazione italiana.

Per rispondere all'obiettivo di questa fase della ricerca è necessario raccogliere documenti e contenuti che popoleranno il database alla base del nostro strumento. Si è partiti dunque da un'attenta selezione di parole chiave relative alle macro aree legate ai temi della

datificazione, del data-intensive e dal processo decisionale automatizzato. In parte le keywords sono state riprese da un precedente lavoro della Northwestern University dal titolo “AlgorithmTips” ma arricchite e adattate al contesto italiano. Al fine di estrarre tutti i contenuti relativi ai temi citati, sarà utilizzata la tecnica del web scraping sui siti dei ministeri italiani che terminano con .gov.it, su altri siti ministeriali che non utilizzano il dominio .gov.it (es. beniculturali.it, ecc.) e infine su specifiche sezioni dei siti ministeriali che riguardano i servizi online ai cittadini (ad esempio lo “sportello unico digitale” del Ministero del Lavoro), uffici relazioni con il pubblico e l’amministrazione trasparente; spazi in cui si ipotizza siano contenute informazioni e documenti necessarie sull'utilizzo di sistemi di decisione basato sui dati.

Prima della costruzione del modello, sarà realizzato un assessment delle keywords finalizzato all’ottimizzazione dei contenuti del database. La procedura in questione richiederà un controllo di coerenza tra la parola chiave ricercata e l’output effettivo, evitando così la raccolta di dati eccessivamente rumorosi che potrebbero incidere negativamente sull’accuratezza dell’algoritmo da addestrare. A ogni contenuto raccolto, sarà assegnato manualmente un punteggio su una scala da 0 (non rilevante) a 1 (rilevante), ottenendo così una nuova variabile che ci fornisce informazioni sulla base della rilevanza dei temi d’interesse all’interno dei contenuti raccolti. Questo lavoro di etichettatura sarà la base di training da sottoporre a un modello di apprendimento supervisionato capace di assegnare in autonomia un punteggio di rilevanza a nuovi documenti sulla base di parole e frasi presenti nei dati di apprendimento. In questa fase sarà di fondamentale importanza il costante confronto tra classificazione automatica ed ermeneutica, basata sulla partecipazione attiva dei ricercatori nella lettura dei documenti. Solo successivamente il documento verrà caricato nel database. In ultima istanza, addestrando l’algoritmo BERT, sarà costruito un modello capace di classificare autonomamente ogni documento di cui è composto il database. A questo punto, funzionando come un motore di ricerca, lo strumento permetterà a chiunque sia interessato, di effettuare interrogazioni e accedere così a una varietà di documenti e risorse web sui sistemi decisionali presenti nelle amministrazioni pubbliche italiane.