Sofia D'Alessandro, Paola Ripa

Data is not Neutral

Dati, genere e potere: smascherare i bias nell'era della Data Science



FrancoAngeli

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati
possono consultare il nostro sito Internet:
www.francoangeli.it e iscriversi nella home page
al servizio “Informatemi” per ricevere via e.mail le segnalazioni delle novità.

Sofia D'Alessandro, Paola Ripa

Data is not Neutral

Dati, genere e potere: smascherare i bias nell'era della Data Science

FrancoAngeli

ISBN E-book:9788835199472

Copyright © 2026 by FrancoAngeli s.r.l., Milano, Italy.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore.
Sono riservati i diritti per Text and Data Mining (TDM), AI training e tutte le tecnologie simili.
L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito *www.francoangeli.it*.

Gli eventuali link attivi e QR code inseriti nel volume sono forniti dall'autore. L'editore non si assume alcuna responsabilità sui link attivi e QR code ivi contenuti che rimandano a siti non appartenenti a FrancoAngeli.

L'autore dichiara che, nella fase di stesura del presente volume, ha utilizzato strumenti di IA generativa esclusivamente per la revisione linguistica e l'ottimizzazione stilistica. Il controllo finale dei contenuti e la responsabilità scientifica dell'opera restano integralmente a carico dell'autore.

Indice

Prefazione
Alessia Cappello pag. 11

Introduzione
Sofia D'Alessandro » 17

1. Garbage In, Garbage Out
Sofia D'Alessandro » 27
1. La narrativa sui Big Data e sulla Data Science » 28
1.1. Introduzione: dati e immaginario contemporaneo » 30
1.2. La nascita della Data Science » 31
1.3. La promessa dell'oggettività dei Big Data » 33
1.4. Data Science e decisione automatizzata » 36
1.5. Big Data e disuguaglianze sociali » 39
1.6. Big Data e democrazia » 41
1.7. Critiche alla datafication della società » 44
2. La Data-Driven Decision » 47
2.1. Decisioni guidate dai dati nell'era digitale » 48
2.2. Principi chiave della Data-Driven Decision » 51
3. La matrice del dominio dei dati » 55
3.1. Rappresentazione matriciale dei dati » 57
3.2. Ruolo nella governance dei dati » 59

4. L'etica dei dati pag. 62
4.1. Bias e discriminazione algoritmica » 63
4.2. Trasparenza nell'ecosistema dei dati e dell'intelligenza artificiale » 66
4.3. Accountability » 70
4.4. Preprocessing dei dati: bilanciamento dei dataset e rimozione di variabili proxy per discriminazione » 74

2. Mind the Gap
Sofia D'Alessandro » 77
1. La nascita del Data Feminism » 79
1.1. Contesto storico: la società dei dati » 79
1.2. Le radici teoriche del Data Feminism » 81
1.3. Science and Technology Studies (STS) » 84
1.4. I sette principi del Data Feminism » 86
1.5. Data Feminism, intelligenza artificiale e potere decisionale » 89
1.6. AI governance e prospettiva di genere » 92
1.7. Limiti e critiche al Data Feminism in ambito IA » 95
2. I dati di genere » 98
2.1. Il concetto di genere » 99
2.2. Definizione di dati di genere » 101
2.3. Evoluzione storica dei dati di genere » 104
2.4. Il gender data gap » 107
2.5. Dati e dinamiche di genere nel mercato del lavoro » 109
2.6. Genere e istruzione: evidenze dai dati » 111
2.7. Data analysis per la salute in ottica di genere » 113
2.8. Violenza di genere e raccolta dei dati » 116
2.9. Genere e partecipazione politica: una lettura attraverso i dati » 118
3. Il Global Gender Gap Report » 122
3.1. Il World Economic Forum e il Global Gender Gap Report » 122

3.2. Metodologia dell'indice globale pag. 124
3.3. I quattro pilastri del Gender Gap Index » 125
3.4. Focus sull'Italia » 128
4. Il Gender Policy Report » 130
4.1. Il concetto di gender policy » 130
4.2. Evoluzione storica delle politiche di genere » 132
4.3. Gender budgeting » 135
4.4. Monitoraggio e valutazione » 138
4.5. Criticità e limiti delle gender policies » 140

3. Data Justice
Paola Ripa » 145
1. La raccolta dei dati » 147
2. Il Gender Equality Index » 150
3. L'Agenda 2030 delle Nazioni Unite » 153
3.1. La digitalizzazione della Giustizia italiana, qualche criticità » 157
4. Le modalità binarie di categorizzazione dei dati » 157

4. Mindset Data
Paola Ripa » 161
1. La nascita del "Femminismo Intersezionale" » 161
1.1. Un esempio pratico di intersezionalità » 163
2. I Gender Bias » 164
3. Il linguaggio di genere » 167
4. I dati per la co-liberazione » 170

Conclusioni
Paola Ripa » 175

Bibliografia » 179

Le Autrici » 189

«I dati sono migliori del petrolio,
perché sono una risorsa rinnovabile».
(Margrethe Vestager[1], intervista a *Euronews*, 2021)

Con questa affermazione Margrethe Vestager ha sintetizzato efficacemente il valore strategico assunto dai dati nel XXI secolo. Essi alimentano algoritmi, orientano decisioni pubbliche e private, modellano i sistemi di intelligenza artificiale e influenzano sempre più profondamente la vita delle persone. Tuttavia, come ogni risorsa che genera potere, anche i dati sollevano interrogativi fondamentali sulla distribuzione delle opportunità, sulla rappresentazione delle differenze e sulla giustizia sociale. È in questo spazio critico che si colloca il femminismo dei dati.

1. Politica danese, Commissaria europea per la concorrenza dal 2014 al 2024 e seconda Vicepresidente della Commissione europea dal 2019 al 2024.

Prefazione

Ho conosciuto personalmente una delle due autrici, Sofia D'Alessandro, attraverso "Mentorship Milano", il progetto che, come Assessorato alle politiche del lavoro e sviluppo economico del Comune di Milano, abbiamo costruito per mettere in relazione donne con esperienze professionali consolidate e giovani donne all'inizio del proprio percorso.

Di lei mi hanno colpito subito due cose: la competenza, certamente, ma soprattutto la disponibilità a metterla al servizio delle altre. Non è così scontato. Io credo molto nella capacità delle donne di fare squadra, di costruire alleanze, di aprire una porta a chi viene dopo. Perché la parità non si costruisce soltanto attraverso le battaglie individuali: si costruisce anche creando reti, condividendo opportunità, mettendo a disposizione tempo, esperienza e relazioni.

Quando Sofia mi ha chiesto di scrivere la prefazione di questo libro ne sono stata molto lusingata. E, lo confesso, anche un po' sorpresa.

Poi, leggendolo, ho capito.

Perché questo libro parla di dati, certo. Ma parla soprattutto di decisioni. E quindi parla profondamente di politica.

Chi governa una città, un Paese, un'organizzazione dovrebbe partire sempre dalla realtà. E per conoscere la realtà servono i dati: bisogna raccoglierli, leggerli, disaggregarli, confrontarli. Bisogna capire quali fenomeni nascondono e, qualche volta, quali fenomeni nascondiamo noi dietro una media apparentemente rassicurante.

Su questo, probabilmente, ho anche una deformazione professionale.

Prima di prestarmi alla politica ho lavorato per molti anni nel privato, nel marketing di grandi multinazionali, e lì ho imparato che senza numeri si rischia di affidare le decisioni alle impressioni, a "gut feeling" come

si dice in azienda quando testavamo le impressioni delle pubblicità tra i colleghi facendo un iper-scientifico test "di corridoio". Ogni scelta economica si basa sui numeri.

Ancora oggi, a ogni conferenza stampa della Milano Fashion Week, scherzo dicendo: «Io sono quella che dà i numeri. Ma qualcuno deve pur darli». E poi comincio: presenze attese, turisti, occupazione alberghiera, indotto economico.

Ma dietro la battuta c'è una convinzione seria: ciò che non conosciamo difficilmente possiamo governarlo. E ciò che non misuriamo difficilmente possiamo migliorarlo.

È la stessa ragione per cui, quando abbiamo lanciato "Mentorship Milano", ho voluto che non ci limitassimo a raccogliere impressioni o testimonianze positive. Dalla prima edizione il progetto è cresciuto fino a coinvolgere oltre 1.700 donne, tra mentor e mentee, unite dall'obiettivo di sostenere le più giovani nei loro percorsi professionali. Un progetto rivolto alle ragazze tra i 18 e i 35 anni, che permettesse loro di incontrare donne provenienti dall'impresa, dalle istituzioni, dall'università, dalla cultura e dalle professioni e confrontarsi, nell'arco di sei mesi, sulle scelte, sulle aspirazioni, sulle difficoltà e sui passaggi decisivi della propria vita personale e professionale. Volevo però capire se quegli incontri producessero davvero un cambiamento. Per questo abbiamo affidato alla società Lifeed una misurazione dell'impatto attraverso questionari rivolti alle partecipanti, per verificare l'evoluzione di alcune soft skills prima e dopo il percorso.

I risultati ci hanno raccontato qualcosa di molto più interessante del semplice numero delle partecipanti: la fiducia in sé stesse delle più giovani è aumentata del 33%, la capacità di gestire stress e pressioni sociali del 60%, mentre l'86% delle ragazze ha riconosciuto di possedere più competenze e risorse di quanto pensasse.

E dietro quei numeri sono arrivate decisioni vere: nuovi percorsi di formazione, opportunità professionali, esperienze e stage all'estero che alcune ragazze, prima del percorso, non avrebbero neppure immaginato di intraprendere.

È questo, per me, uno dei significati più concreti del dato applicato alla politica.

Perché un progetto pubblico non dovrebbe essere giudicato soltanto per la bontà delle sue intenzioni. Dovrebbe essere misurato per gli effetti che produce.

È anche la ragione per cui ho fortemente voluto che Milano lavorasse sul Bilancio di genere della città: non una dichiarazione di principio, ma una fotografia della città stessa attraverso dati su lavoro, redditi, conciliazione, rappresentanza, salute, istruzione.

Ed è proprio quando si comincia a guardare dentro i numeri che le cose diventano interessanti.

Milano ha un'occupazione femminile significativamente più alta della media italiana e, osservata da lontano, può sembrare una città vicina alla parità, in linea con la media europea. Potremmo fermarci qui. Ottimo risultato, una pacca sulla spalla e avanti così.

Ma proviamo a spacchettare il dato.

Il reddito medio dichiarato dalle donne milanesi è di circa 28 mila euro, quello degli uomini di circa 46 mila. Persino all'interno del Comune di Milano, dove le donne rappresentano circa due terzi del personale, il divario retributivo permane ai diversi livelli professionali. E se guardiamo al part-time, scopriamo che la quasi totalità dei rapporti a tempo parziale riguardano donne. Oltre il 90 per cento dei permessi è utilizzato da donne. Per non parlare dei dati impietosi delle donne che lasciano il lavoro dopo il primo figlio (riferimento ai dati Inps). Allora quella fotografia "positiva" cambia e racconta in realtà una storia ben diversa.

Ecco perché questo libro pone, a mio avviso, una questione essenziale per chiunque abbia responsabilità pubbliche: i dati sono indispensabili, ma i dati non sono neutrali.

Non lo sono perché qualcuno decide quali raccogliere e quali no. Quali domande fare. Quali categorie utilizzare. Quali fenomeni mettere in relazione. Se osservare soltanto il tasso di occupazione oppure anche la qualità di quell'occupazione. Se contare quante donne lavorano oppure chiedersi quante lavorano part-time, quante interrompono la carriera dopo la maternità, quanto guadagnano, quanto tempo dedicano al lavoro di cura, quante arrivano ai vertici.

Cambiare la domanda può cambiare radicalmente la realtà che il dato ci restituisce.

E se questo era vero ieri, lo è ancora di più oggi.

Perché i dati non alimentano più soltanto statistiche, ricerche e *report*. Alimentano algoritmi e sistemi di intelligenza artificiale che sempre più spesso contribuiscono a selezionare curricula, valutare profili, concedere credito, organizzare servizi, orientare decisioni.

Se una parte della realtà è sottorappresentata nei dati, rischia di es-

serlo anche nelle decisioni che quei dati contribuiscono a generare. E se i dati del passato contengono le disuguaglianze del passato, una tecnologia che semplicemente apprende da essi può finire non per correggerle, ma per replicarle, renderle più efficienti e persino amplificarle.

È uno dei passaggi più importanti di questo volume: dietro un numero, un dataset o un algoritmo apparentemente impersonale ci sono sempre scelte umane. Il libro lo mostra interrogandosi non soltanto sulla quantità dei dati disponibili, ma su chi venga rappresentato, chi rimanga invisibile e quali rapporti di potere possano riflettersi nella loro produzione e nel loro utilizzo.

Tuttavia credo che ci sia un passaggio ulteriore: i dati possono dirci dove siamo; non possono decidere dove vogliamo andare.

Possono dimostrarci, numeri alla mano, quanto siamo ancora lontani dalla parità sostanziale. Possono mostrarci dove il mercato del lavoro penalizza maggiormente le donne, quanto pesa la maternità sulle carriere, dove si concentrano le differenze retributive, quali politiche funzionano e quali no. Possono perfino aiutarci, grazie all'intelligenza artificiale, a immaginare soluzioni più sofisticate.

Ma poi viene il momento della scelta.

Possiamo sapere tutto sul gender gap e decidere di non cambiare nulla. Possiamo produrre il migliore Bilancio di genere possibile e continuare a considerare la cura una questione prevalentemente femminile. Possiamo misurare ogni anno la denatalità e non avere il coraggio di ripensare radicalmente il welfare, l'organizzazione del lavoro, i servizi alle famiglie.

Conoscere una disuguaglianza non significa automaticamente volerla rimuovere.

Ed è qui che torna la politica.

Servono dati migliori, certamente. Servono dati disaggregati, rappresentativi, capaci di rendere visibile chi troppo spesso non lo è stato. Serve una data governance consapevole. Ma servono anche politici e istituzioni disposti a utilizzare quella conoscenza per fare scelte.

Scelte anche difficili.

Significa, per esempio, chiedersi se un modello del lavoro ancora largamente costruito sulla presenza e sul tempo trascorso in ufficio sia compatibile con una società nella quale donne e uomini dovrebbero poter condividere davvero responsabilità professionali e familiari. Significa costruire davvero quel "villaggio" che oggi manca a troppe famiglie: servizi,

asili, scuole, tempi urbani, reti di prossimità e strumenti di conciliazione che rendano la libertà di lavorare e di avere figli una possibilità concreta e non un privilegio.

Significa decidere che la parità non è una politica per le donne, ma una politica di sviluppo.

E significa avere il coraggio di investirci davvero.

Perché possiamo misurare perfettamente la denatalità, ma se non interveniamo sulle condizioni che rendono difficile costruire una famiglia il grafico continuerà a scendere. Possiamo conoscere il gender pay gap fino all'ultimo decimale, ma se non cambiamo le condizioni che lo producono continueremo semplicemente a misurarlo meglio.

In altre parole: il dato rende visibile il problema, ma la politica deve scegliere se trasformarlo.

È forse questa la riflessione che porto con me dopo la lettura di queste pagine.

Viviamo in un tempo nel quale affidiamo ai dati e all'intelligenza artificiale una quantità crescente di decisioni e, proprio per questo, dobbiamo essere ancora più consapevoli del loro enorme potere ma anche dei loro limiti.

La tecnologia può elaborare quantità di informazioni che nessun essere umano potrebbe gestire. Può riconoscere correlazioni, costruire scenari, suggerire soluzioni. Può aiutarci a progettare politiche migliori. Ma non può stabilire quale società vogliamo diventare.

La visione appartiene ancora all'essere umano. E anche il coraggio.

Il coraggio di decidere che una disuguaglianza, pur essendo il risultato della storia, non debba necessariamente diventare il dato di partenza del futuro.

Alessia Cappello
Assessora alle Politiche del Lavoro e allo Sviluppo Economico
del Comune di Milano

Introduzione

Sofia D'Alessandro

Viviamo in un'epoca in cui i dati sono spesso descritti come il nuovo petrolio: una risorsa preziosa, apparentemente neutrale, capace di guidare decisioni, orientare politiche pubbliche e raccontare il mondo in modo oggettivo. La diffusione dei *Big Data* e della *Data Science* ha rafforzato questa narrazione, alimentando l'idea che i numeri possano parlare da soli e che le decisioni *Data-Driven* siano, per definizione, più razionali, efficienti e giuste. Ma è davvero così? Questo libro nasce da una domanda fondamentale: cosa succede quando guardiamo ai dati da una prospettiva di genere? E, soprattutto, chi viene incluso – o escluso – nei processi di raccolta, analisi e interpretazione dei dati?

A partire dai dati di genere e dai principali strumenti di misurazione delle disuguaglianze, come i *report* globali sul *gender gap* e gli indici europei sull'uguaglianza, emerge subito un primo elemento chiave: i dati non sono mai neutri. Essi riflettono le strutture sociali in cui vengono prodotti e, se non interrogati criticamente, rischiano di rafforzare le stesse disuguaglianze che dovrebbero contribuire a ridurre. Le politiche di genere e le iniziative internazionali, incluse quelle promosse nell'ambito dell'Agenda 2030, mostrano quanto sia importante disporre di dati accurati e inclusivi, ma anche quanto sia facile invisibilizzare intere categorie di persone attraverso scelte apparentemente tecniche. Uno dei nodi centrali affrontati in queste pagine riguarda infatti le modalità di categorizzazione dei dati. La tendenza a utilizzare schemi binari – come la distinzione rigida tra femminile e maschile – semplifica la complessità delle identità e contribuisce a escludere soggettività non conformi. Allo stesso tempo, i *Gender Bias* presenti nei *dataset* e negli algoritmi dimostrano come le tecnologie possano incorporare e amplificare pre-

giudizi già radicati nella società. Per comprendere a fondo queste dinamiche, il libro si intreccia con il pensiero del femminismo intersezionale, che invita a considerare come le diverse forme di oppressione – legate al genere, alla razza, alla classe, all'orientamento sessuale e ad altri fattori – si sovrappongano e si rafforzino a vicenda. In questo quadro, il *Data Feminism* emerge come un approccio capace di connettere teoria e pratica, proponendo una lettura critica dei dati e, allo stesso tempo, strumenti concreti per trasformarli in leve di cambiamento. L'etica dei dati diventa così una questione centrale. Non si tratta solo di proteggere la *privacy* o garantire la sicurezza delle informazioni, ma di interrogarsi su chi detiene il potere sui dati, su come questi vengono utilizzati e su quali conseguenze producono. La matrice del dominio dei dati aiuta a visualizzare queste relazioni, rendendo evidente come il controllo dei dati sia distribuito in modo diseguale e spesso concentrato in poche mani. In un contesto sempre più orientato verso decisioni *Data-Driven*, diventa quindi urgente sviluppare una consapevolezza critica che permetta di riconoscere i limiti e i rischi di un affidamento cieco agli algoritmi. La *Data Justice* si inserisce in questo scenario come un orizzonte teorico e pratico volto a promuovere equità, inclusione e responsabilità nell'uso dei dati, trasformandoli da strumenti di controllo a risorse per l'emancipazione. Infine, il libro riflette anche sul ruolo del linguaggio e della rappresentazione, sottolineando come le parole che utilizziamo – così come le categorie che adottiamo – contribuiscano a costruire la realtà che descriviamo. Ripensare il linguaggio di genere significa quindi intervenire non solo sul piano simbolico, ma anche su quello politico e sociale.

Questa introduzione apre a un percorso che non intende offrire risposte semplici, ma piuttosto strumenti per porsi domande migliori. Ripensare i dati in chiave femminista significa riconoscerne la dimensione profondamente umana, situata e politica. Significa, soprattutto, immaginare nuovi modi di produrre conoscenza e prendere decisioni, orientati non all'efficienza a ogni costo, ma alla giustizia, all'inclusione e alla coliberazione. *Data is Not Neutral* affronta tali tematiche in quattro specifiche sezioni. Il *Capitolo 1* ha una funzione meramente preparatoria. La crescente disponibilità di dati digitali e il rapido sviluppo delle tecnologie di analisi hanno profondamente trasformato il modo in cui organizzazioni, istituzioni e governi osservano la realtà e assumono decisioni. I dati rappresentano oggi una risorsa strategica, in grado di orientare politiche pubbliche, modelli di business, processi organizzativi e sistemi di

intelligenza artificiale. In questo scenario, la diffusione dei *Big Data* e della *Data Science* ha alimentato una narrativa secondo cui le decisioni basate sui dati sarebbero intrinsecamente più oggettive, razionali ed efficaci rispetto a quelle fondate esclusivamente sull'esperienza o sull'intuizione. Tuttavia, questa visione rischia di trascurare un elemento fondamentale: i dati non sono una rappresentazione neutrale della realtà, ma il risultato di processi di raccolta, selezione, classificazione e interpretazione influenzati da scelte metodologiche, organizzative, culturali e politiche. Come sintetizzato dall'espressione "*Garbage In, Garbage Out*", la qualità e l'affidabilità delle decisioni dipendono inevitabilmente dalla qualità dei dati che alimentano i sistemi informativi e gli algoritmi. Dati incompleti, distorti o non rappresentativi producono analisi inaccurate e possono contribuire a consolidare pregiudizi, stereotipi e disuguaglianze già presenti nella società. Per questa ragione, prima di affrontare il tema del *Data Feminism*, è necessario costruire un quadro teorico che permetta di comprendere come i dati vengano generati, governati e utilizzati all'interno delle organizzazioni. Il *Data Feminism*, infatti, non si limita a criticare gli effetti discriminatori degli algoritmi, ma propone una riflessione più ampia sulle relazioni di potere che influenzano l'intero ciclo di vita del dato: dalla sua produzione alla sua interpretazione, fino al suo impiego nei processi decisionali. Per cogliere pienamente la portata di questo approccio è quindi indispensabile comprendere preliminarmente i principi della *Data Governance*, della *Data Science* e del *Data-Driven Decision Making*, che costituiscono l'infrastruttura tecnica e organizzativa entro cui i dati vengono gestiti. L'analisi dei principi della *Data Governance* consente di approfondire temi quali la qualità del dato, la definizione di ruoli e responsabilità, la gestione del ciclo di vita delle informazioni, la sicurezza, la trasparenza e la conformità normativa. Questi aspetti rappresentano i presupposti essenziali per garantire che i dati siano affidabili, documentati e utilizzabili in modo responsabile. Solo a partire da una *governance* solida è infatti possibile interrogarsi criticamente sulla rappresentatività dei dati, sulla presenza di *bias* sistematici e sulle conseguenze sociali derivanti dal loro utilizzo. Questo primo capitolo ha pertanto l'obiettivo di fornire le basi teoriche necessarie per comprendere il rapporto tra dati, potere e processi decisionali. In particolare, verranno analizzate la nascita della *Data Science* e della cultura dei *Big Data*, l'evoluzione del paradigma delle decisioni *Data-Driven*, i principi della *Data Governance*, la matrice del dominio dei dati e le prin-