DOPPIOZERO

Chi ha paura dei data scientist? Numeri e pandemia

Roberta Rocca

5 Novembre 2020

Dopo un'inattesa *love story* (o una, ancor più inaspettata, *trust story*) tra Paese e istituzioni, durata appena il tempo d'una prima ondata, il prevedibile "ritorno" autunnale del coronavirus ha resuscitato i profondi dubbi dell'opinione pubblica italiana sulla capacità delle proprie istituzioni e classi dirigenti di gestire situazioni di crisi che richiedono soluzioni sistemiche. Tanto ammirevole è stata la gestione dell'*emergenza* in primavera – non era semplice, primi in Europa, decidere in favore di soluzioni tanto radicali e potenzialmente impopolari – quanto chiari sono adesso i limiti del governo nel pensare soluzioni che vadano oltre l'orizzonte temporale immediato. La natura dell'epidemia è infatti tale che tentare di limitare, oltre che la catastrofe sanitaria, quella economica, richiede la capacità di pensare a lungo termine e in potenza: per *previsioni* e per *ipotesi*.

Sono due, infatti, le caratteristiche fondamentali di questa (e non solo questa) pandemia.

La prima è la condizione di costante e radicale *incertezza* in cui il virus ci pone, dovuta alla *complessità* delle dinamiche di diffusione e alle scarse conoscenze finora accumulate. Quando e come un incremento lineare diventa esponenziale, e perché a Milano ma non a Catania? La seconda caratteristica fondamentale è l'inevitabile "sfasamento" tra l'effettiva diffusione del virus e la nostra percezione del suo stato di diffusione. I tamponi sono in grado di rivelare la presenza del virus solo diversi giorni dopo l'effettivo contagio; l'istantanea fornita dal tampone rappresenta uno stato di cose relativo a giorni (o settimane) addietro e il virus, nel frattempo, si è diffuso chissà quanto. Anche il numero di decessi, che avvengono anche mesi dopo il contagio, dà un'idea della gravità della situazione che è in ritardo di settimane rispetto alla realtà. L'incertezza e lo "sfasamento" fanno sì che misure efficaci debbano navigare su un orizzonte diverso da quello

della *reazione* ai dati sul contagio. Devono, piuttosto, provare a prevenire conseguenze *ipotetiche*, senza però, con restrizioni incondizionate, arrestare il Paese.

Come si fa a muoversi in un orizzonte così incerto, ipotetico e complesso? Il buon senso aiuta, ma non basta. Il senso di responsabilità politica (che avrebbe dettato, per esempio, di disporre del MES e potenziare efficacemente il sistema sanitario) è necessario: ma non basta. Persino pareri autorevoli, che pur giovano al dibattito pubblico, in questo caso non bastano. Per navigare l'incertezza e la complessità, esperienza individuale e intuizioni soggettive hanno un orizzonte limitato. Servono, piuttosto, strumenti che siano in grado non di semplificare, ma di *elaborare* e *interpretare* tale complessità.

La buona notizia è che questi strumenti esistono, e sono già alla base di una porzione immensa della nostra esperienza quotidiana. La nostra arma più efficace e più potente per contenere gli effetti della pandemia è, infatti, la disponibilità di dati e la capacità di usarli, attraverso modelli matematici, non solo per descrivere, ma anche per prevedere (leggi l'articolo). Si tratta di una lezione che altri paesi europei (come Germania e Danimarca) hanno fatto propria, giustificando pubblicamente le proprie decisioni sulla base delle predizioni di modelli di riferimento. La cattiva notizia, però, è che tali strumenti continuano a godere di una certa diffidenza nel dibattito pubblico del nostro paese, forse retaggio di un paradigma culturale in cui la cultura umanistica è quella nobile per eccellenza. Numeri, grafici, modelli ci respingono: ci terrorizzano e ci appaiono, a un tempo, troppo aridi per rientrare nel dominio (pragmatico e umanistico) della politica.

Ricordate la performance fantascientifica di Giulio Gallera sul fattore Rt, quando l'assessore al welfare lombardo dichiarò (testuali parole) che "l'indice Rt a 0,51 vuole dire che per infettare me bisogna trovare due persone nello stesso momento infette"? Più che un'eccezione, questo episodio è forse indicativo (sia pure in una forma iperbolica, quasi satirica) di una generale inadeguatezza della classe politica a fare i conti con strumenti di conoscenza quantitativi, e ad integrarli nei meccanismi di decisione politica.



Ma questa diffidenza va oltre la dimensione politica. Salvo alcune eccezioni, in Italia un giornalismo fortemente centrato su dati e loro visualizzazioni grafiche fatica ancora a decollare. In più, pur nell'era dei Big Data dove tutto è dati, gli esperti in materia sembrano relegato allo status di tecnici, più che aver riconosciuto il possesso di conoscenze politicamente rilevanti.

Banali considerazioni statistiche, che farebbero la differenza nella gestione della pandemia, sono rimaste largamente inascoltati (vedi su doppiozero l'articolo di Giovanni D'Alessio e Riccardo De Bonis "Covid: una caporetto statistica). Molti, ad esempio, hanno sottolineato come l'unico modo per avere un quadro adeguato degli *effettivi* livelli del contagio e di mertalità sia somministrare tamponi a campione sulla popolazione. Testare soltanto chi ha sintomi o ragioni per ipotizzare di essere stato contagiato, restituisce un quadro falsato di quale percentuale della popolazione abbia contratto il virus. Questa invocazione è caduta nel vuoto.

Di recente, uno splendido articolo su The Atlantic ha contribuito a divulgare altre conoscenze fondamentali sugli aspetti “statistici” del virus. Se nel dibattito pubblico si insiste sull’ormai noto fattore Rt – che cattura il numero di contagi innescato, in media, da ciascun individuo – bisognerebbe piuttosto prestare attenzione al cosiddetto fattore *K*, che descrive come il virus sembra diffondersi, più che in modo “ordinato” e costante, per “impulsi” ed eventi di “iperdiffusione” (ove un individuo contagia tanti contemporaneamente, attraverso grandi eventi, o per via di permanenze prolungate in luoghi chiusi). Queste considerazioni (del tutto assenti dal dibattito politico italiano) suggeriscono che la strategia ottimale sia quella di limitare gli eventi di iperdiffusione con interventi mirati (mantenere la chiusura delle discoteche, ad esempio; o limitare il più possibile l’uso mezzi pubblici come le metropolitane), piuttosto che optare *in extremis* per misure trasversali di impatto economico catastrofico.

La pandemia ci mostra che un cambio di rotta è necessario. Ci sono state e ci sono tuttora realtà virtuose che garantiscono solidi fondamenti. Già a febbraio, quando era quasi impossibile reperire dati centralizzati e realistici sul numero di tamponi e casi nel resto d’Europa, l’Italia diffondeva pubblicamente ed elegantemente i propri dati su Github. Esempi di informazione e giornalismo ispirati a *data mining* accurato e visualizzazioni efficaci sono stati forniti su Twitter da Matteo Villa dell’ISPI, già attivo su questo fronte prima della pandemia. Piuttosto che investire unicamente nell’istituzione di macchinosissimi comitati tecnico-scientifici, la strategia vincente sarebbe, forse, quella di investire massicciamente nel finanziare *task force* e gruppi di ricerca orientati alla raccolta di dati e alla formulazione di modelli predittivi che guidino le istituzioni e l’opinione pubblica.

Affidarsi a dati, modelli e tecnologie non sarebbe una sconfitta della politica, ma un potenziamento. La capacità *pratica* di decidere è prerogativa insostituibile di una buona politica. Ma può e dev’essere guidata da strumenti e conoscenze che permettano di visualizzare e comprendere gli scenari possibili. A beneficio tanto del bene collettivo, quanto della credibilità individuale di chi è tenuto a metterci la faccia.

Se continuiamo a tenere vivo questo spazio è grazie a te. Anche un solo euro per noi significa molto.

Torna presto a leggerci e SOSTIENI DOPPIOZERO

