

Giustizia e Machine Learning

Lorenzo Barbato Fornai, Pietro Bandecchi Ceccuti

18 febbraio 2023

Abstract

L'intelligenza artificiale negli ultimi anni ha trovato una vasta applicazione in gran parte delle attività umane, una fra queste è quella della giustizia. Nel presente articolo affrontiamo questo connubio attraverso una visione storica, evidenziandone quindi gli sviluppi, senza tralasciare le motivazioni per le quali il suo utilizzo si riveli così utile. L'articolo citerà, dapprima, 2 studi in campo sociale che ci pongono dinanzi ai limiti di noi esseri umani, per poi analizzare quali tecniche di previsione si siano rivelate le più efficaci in campo giudiziario ed, infine, illustrare le capacità di questi modelli a confronto con quelle umane.

Indice

1	Studi sulla coerenza e parzialità dei giudici	2
1.1	Disparità dei trattamenti e incoerenza	2
1.2	Le pene detentive seguono un andamento logaritmico	3
2	Modelli predittivi del rischio di recidiva	7
2.1	Studi pilota di Ernest W. Burgess	7
2.2	Alberi di decisione e foreste casuali	9
3	Decisioni umane e previsioni artificiali	12
3.1	Prevedere le decisioni della Corte Suprema Americana	12
3.2	Ridurre la popolazione carceraria con gli algoritmi	14

1 Studi sulla coerenza e parzialità dei giudici

Introduzione alla sezione

I giudici devono prendere decisioni sensate sulla base di approfondimenti che analizzano da soli: ciò rende il processo decisionale privo di un contesto approfondito e più vulnerabile ai pregiudizi. A conferma di ciò, ovvero che i magistrati, in quanto esseri umani, siano inclini a certe fallace cognitive, dagli anni settanta in poi sono stati condotti diversi esperimenti.

1.1 Disparità dei trattamenti e incoerenza

Se voi e il vostro gemello immaginario commettete lo stesso identico crimine vi aspettereste di ricevere la stessa condanna. Ma finirebbe davvero così?

Negli anni settanta, un gruppo di ricercatori americani provò a rispondere a una variante di questa domanda. Invece di ricorrere a criminali gemelli (una soluzione difficilmente realizzabile), imbastirono una serie di casi ipotetici e chiesero in maniera indipendente a 47 giudici distrettuali della Virginia che hanno partecipato a una conferenza giudiziaria statale nell'ottobre 1975, di leggere attentamente le istruzioni e le descrizioni dei casi, di raccomandare un verdetto e, se veniva emesso un verdetto di colpevolezza, di suggerire un'appropriata sentenza. Ecco il primo dei casi proposti:

L'imputata di 18 anni è stata arrestata per possesso di marijuana, insieme al suo ragazzo e a sette altri conoscenti. È stata trovata una quantità significativa di marijuana fumata e non fumata, ma non è stata trovata marijuana addosso all'imputata. L'imputata non ha precedenti penali, è una brava studentessa proveniente da una famiglia di classe media e non è né ribelle né si scusa per le sue azioni

Le differenze tra una sentenza e l'altra furono drammatiche. Dei 47 giudici, 29 giudicarono l'imputata non colpevole e 18 colpevole. Di questi, 8 raccomandarono la condizionale, 4 decisero che sarebbe bastata una multa, 3 optarono per multa e condizionale e 3 si dichiararono favorevoli a una pena detentiva.

In uno studio più recente, a 81 giudici del Regno Unito fu chiesto se avrebbero concesso la libertà su cauzione a un certo numero di imputati inventati per l'occasione. Ogni caso aveva una sua storia personale e una serie di precedenti. I giudici britannici si comportarono esattamente come i loro colleghi della Virginia: non uno dei 41 casi ricevette un giudizio unanime. Questa volta, però, tra i 41 casi ipotetici presentati a ogni giudice ce n'erano sette che si ripetevano: erano stati cambiati solo il nome e il sesso dell'imputato, per evitare che il giudice si accorgesse che si trattava di doppioni. La coerenza di ciascun magistrato nel prendere decisioni sulla custodia cautelare è stata misurata calcolando il va-

lore **Kappa di Cohen**¹, che corregge il caso. Le decisioni prese sull'insieme di sette casi duplicati sono state confrontate con quelle prese sulle loro controparti originali nel set di modellizzazione. Il valore di Kappa era inferiore a 0,40 per 12 magistrati, era "discreto" per 18 magistrati, "buono" per 20 magistrati e "eccellente" per 31 (perfetto per 29) magistrati. Alcuni giudici avrebbero quindi mostrato maggiore coerenza con le proprie decisioni precedenti se avessero concesso la cauzione tirando letteralmente a indovinare.

Molti altri studi sono giunti alla stessa conclusione: se si lascia ai giudici la libertà di valutare i casi da soli il risultato sarà completamente incoerente. Affidarsi al potere discrezionale di un giudice significa introdurre nel sistema una componente di fortuna.

1.2 Le pene detentive seguono un andamento logaritmico

Nel 2020 alcuni ricercatori guidati da Mandeep K. Dhimi analizzarono la distribuzione delle sentenze emesse in un anno in due diverse giurisdizioni (Inghilterra e Galles e Nuovo Galles del Sud, Australia). Il set di dati dell'Inghilterra e del Galles comprende 223.207 casi di adulti (di età pari o superiore a 21 anni) condannati presso una magistratura o la Crown Court dopo essere stati condannati o dichiarati colpevoli di uno dei 18 reati che possono essere classificati come reati contro la proprietà, reati violenti e reati di terrorismo, reati violenti, reati di guida o reati di droga.

Innanzitutto tracciarono le frequenze delle durate delle pene detentive e degli importi delle multe/compensazioni indipendentemente dal tipo di tribunale o dalla classificazione del reato. Ogni disposizione è stata esaminata utilizzando l'unità di misura registrata nel dataset. Le unità di misura erano i giorni per gli ordini di custodia mentre gli ordini di ammenda/compensazione sono stati misurati in sterline. I grafici della distribuzione di frequenza permisero di identificare le durate o gli importi delle sentenze comunemente utilizzati.

La Tabella 1 mostra che ci sono 21 "dosi comuni"² che rappresentano l'89% di tutte le pene detentive. Le dosi iniziano con multipli di sette giorni (cioè una settimana o due settimane) fino a 56 giorni (cioè otto settimane o due mesi). Un'eccezione degna di nota è stata quella di 30 giorni, forse perché rappresenta circa un mese. I multipli di 30 giorni sono stati poi utilizzati da 60 giorni fino a 180 giorni (cioè sei mesi). In seguito, sono stati utilizzati multipli di 90 giorni (cioè tre mesi) fino a 540 giorni (cioè un anno e nove mesi). Quindi, sono stati utilizzati multipli di 180 giorni (cioè sei mesi) fino a 1.440 giorni (cioè quattro anni). Infine, 1.800 giorni (cioè cinque anni) hanno rappresentato la

¹Il valore Kappa varia da 0 (che indica che l'accordo o la coerenza non è migliore del caso) a 1 (che indica un accordo o una coerenza perfetti). Un valore di Kappa compreso tra 0,40 e 0,60 è "discreto", tra 0,60 e 0,75 è "buono" e un valore superiore a 0,75 è "eccellente".

²Le "dosi" comuni sono state definite come durate della custodia cautelare o importi di multe/compensi che rappresentavano almeno l'1% della frequenza totale.

dose più alta di custodia comune. La tabella 1 mostra inoltre che l'89% delle multe/compensazioni è stato erogato in 16 dosi comuni. Si tratta di multipli di 10 e 25 sterline fino a 100 sterline, seguiti da multipli di 50 sterline fino a 300 sterline, sono state applicate anche dosi comuni di 400 e 500 sterline.

Genere di pena	Dosi Comuni	Percentuale dei casi
Custodia	7, 14, 28, 30, 42, 56, 60, 90, 120, 150, 180, 270, 360, 450, 540, 720, 900, 1,080, 1,260, 1,440, 1,800	89% (45,631)
Multa (£)	20, 25, 30, 40, 50, 60, 70, 75, 80, 100, 150, 200, 250, 300, 400, 500	89% (21,601)

Tabella 1: Inghilterra e Galles - Dosi di condanna identificate.

I ricercatori scoprirono così che i sentenziatori preferiscono determinati numeri quando emettono le sentenze (in carcere e ai servizi sociali) e gli importi (per le multe/compensazioni). Queste "dosi comuni" rappresentano oltre il 90% delle sentenze in ogni giurisdizione. In effetti la ricerca psicologica ha stabilito che le persone utilizzano alcuni numeri molto più frequentemente di altri quando stimano il livello di uno stimolo. Ad esempio, quando si chiede di generare numeri compresi tra 1 e 100, i multipli di 5 e 10 vengono scelti più frequentemente di altri numeri. In effetti, le serie di numeri preferiti possono essere previste con una certa precisione da alcune equazioni, come quella proposta da Noma e Baird (1975)³. Questi modelli predittivi generano serie di numeri con incrementi che aumentano in proporzione alla dimensione dei numeri. Ciò suggerisce che, oltre ad avere una preferenza per determinati numeri, le persone utilizzano scale numeriche con incrementi che aumentano in proporzione alla dimensione dei numeri coinvolti, seguendo, in particolare, un andamento già descritto dalla cosiddetta Legge di Weber (Figura 1). Weber (1846) osservò per la prima volta che la percezione è proporzionale (secondo un fattore k) al logaritmo dello stimolo. La legge di Fechner-Weber, fu pertanto enunciata come segue:

«Perché l'intensità di una sensazione cresca in progressione aritmetica, lo stimolo deve crescere in progressione geometrica»

³ $PN = kBn$, dove PN è un numero preferito, k è un numero intero compreso tra 1 e - 1, B è la base utilizzata e n è un esponente intero maggiore di 0, che predice la scelta dei numeri preferiti in un contesto di base 10 e può essere utilizzata per calcolare i numeri preferiti utilizzando altre basi. Albers (2001) ha presentato un'equazione simile (solo per la base 10) che produce una serie leggermente più fine: $PN = a \times 10^i : a \in 1, 1.5, 2, 3, 4, 5, 7, i \in \mathbb{N}$.

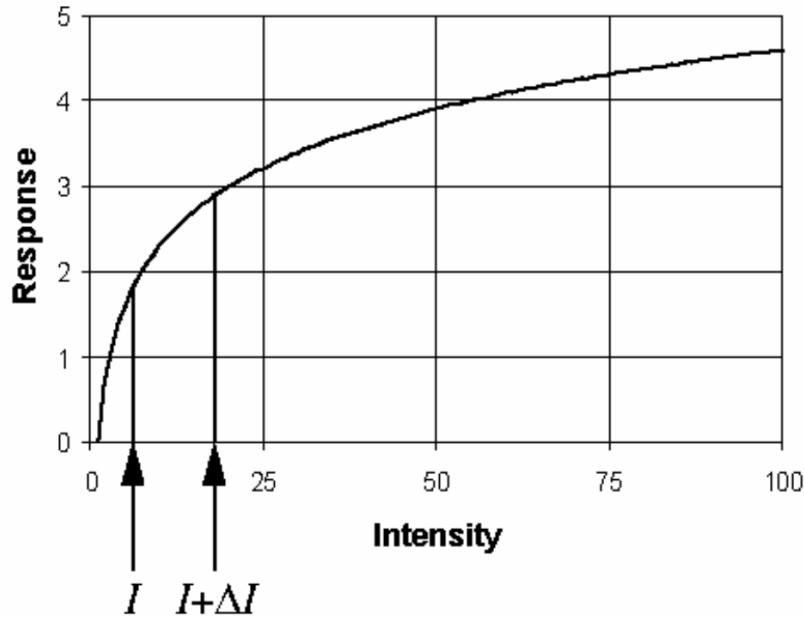


Figura 1: La legge di Weber può essere espressa dall'equazione $\Delta I = kI$, dove ΔI è una differenza appena percettibile nello stimolo I e k è una costante il cui valore dipende dalla dimensione in esame. La legge di Weber è una conseguenza della legge di Fechner: $R = \log(I)$.

Analizzando i dati in loro possesso, per esaminare la misura in cui ogni serie di dosi comuni è aumentata, hanno applicato prima un modello di regressione lineare e poi un modello di regressione log-lineare⁴ La variabile predittiva era il numero di dosi, che è su scala lineare e le variabili di esito erano le quantità di dosi comuni sulla scala lineare o sulla scala logaritmica. Sia per la custodia, che per le pene pecuniarie, le dosi comuni si adattavano meglio a un modello log-lineare(Figura2) che a un modello lineare.

Quindi, in definitiva, verificarono che gli intervalli tra le dosi fossero effettivamente cresciuti con l'aumentare del valore delle dosi, ovvero con l'aumentare della severità delle pene. I risultati sono contrari alle argomentazioni contro gli sforzi per ridurre la discrezionalità giudiziaria e potenzialmente minano la nozione di giustizia individualizzata, oltre a sollevare questioni sull'efficacia (in termini di costi) delle sentenze.

⁴Modello lineare, $y_i = \alpha + \beta i + \epsilon_i$ o modello log-lineare, $\log(y_i) = \alpha + \beta i + \epsilon_i$, dove y_i è la quantità di dosi e i è la dose per il caso i , α e β sono i parametri del modello da stimare, e ϵ_i è un termine di errore, che tra i casi si assume essere normalmente distribuito con una media di zero.

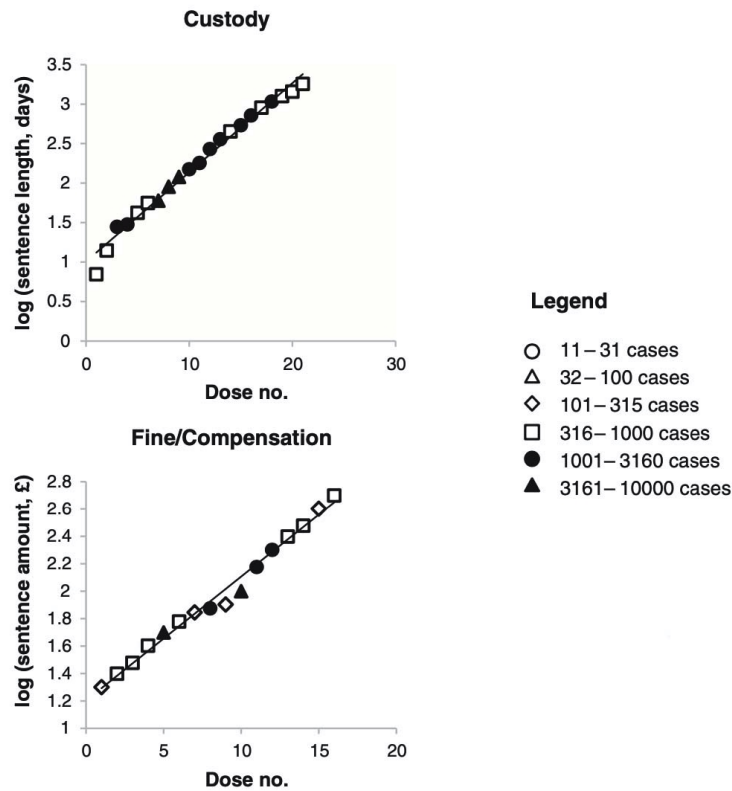


Figura 2: Modelli log-lineari condotti sui dati relativi alla custodia cautelare a all'ammontare delle pene pecuniarie.

All'aumentare delle pene previste si fanno più grandi anche le differenze tra quelle effettivamente inflitte. Per un reato appena più grave di uno che merita una condanna a venti anni, ad esempio, un'aggiunta di tre mesi non viene percepita come sufficiente. È come se le due pene fossero giudicate troppo simili. In realtà la differenza c'è: tre mesi in prigione sono pur sempre tre mesi in prigione, a prescindere da cosa li ha preceduti. Eppure, invece di aggiungere qualche mese, i giudici applicheranno la minima differenza apprezzabile, che in questo caso è di cinque anni, portando la sentenza definitiva a venticinque anni.

2 Modelli predittivi del rischio di recidiva

Introduzione alla sezione

Il sistema perfetto non esiste. Si finisce sempre per dover decidere tra scelte opposte ma altrettanto ingiuste. Una situazione così caotica e conflittuale, però, è proprio quella in cui un algoritmo può rivelarsi utile. Può sembrare incredibile, ma inserendo un algoritmo nel processo si possono garantire sentenze coerenti e mirate, evitando di dover scegliere tra le une e le altre.

2.1 Studi pilota di Ernest W. Burgess

Hornell Hart è stato tra i primi, se non il primo, a riconoscere la possibilità di costruire una tabella delle esperienze al fine di prevedere l'adeguamento alla libertà vigilata. In un articolo del 1923, egli sostenne che i candidati alla libertà vigilata dovessero ricevere un punteggio sulla base di elementi ritenuti probabili per il successo della libertà vigilata e che il rischio di violazione fosse stabilito per ogni punteggio o intervallo di punteggio. Non molto tempo dopo, questa idea fu applicata da Ernest W. Burgess che, in quello che oggi verrebbe definito uno studio pilota, analizzò i registri di 3.000 detenuti in libertà vigilata provenienti in egual misura da tre carceri dell'Illinois.

Partendo da tutti i dati disponibili, Burgess identificò 21 fattori che giudicò "potenzialmente significativi" per determinare la probabilità di violazione dei termini della libertà vigilata. Burgess attribuì a ogni detenuto un punteggio compreso tra 0 e 1 per ognuno dei 21 fattori, andando così a calcolare un punteggio di libertà vigilata per ogni persona. Chi aveva un punteggio alto (tra 16 e 21) era considerato a basso rischio di recidiva; i valori particolarmente bassi (minori o uguali a 4), invece, corrispondevano a un rischio elevato.

I fattori "che hanno superato più o meno bene la prova della correlazione con i precedenti dell'individuo in libertà vigilata sono:

- la natura del reato;
- il numero di complici nel commettere il reato per cui è stato condannato;
- la natura e la durata della pena inflitta;
- la durata del periodo trascorso prima della libertà condizionale;
- il fatto che la condanna sia stata o meno subordinata all'accettazione di un patteggiamento minore;
- la dichiarazione del giudice e del procuratore con riferimento alla raccomandazione a favore o contro la clemenza;
- i precedenti penali;

- i suoi precedenti penali in istituto;
- i suoi precedenti lavorativi;
- se era o meno datore di lavoro al momento della commissione del reato;
- tipo di criminale, come primo, occasionale o abituale delinquente, o criminale professionale;
- età al momento della libertà vigilata;
- nazionalità del padre;
- tipo sociale, come vagabondo, nullafacente, gangster;
- dimensione dell'area residenziale;
- residente o passeggero nella comunità al momento dell'arresto;
- tipo di quartiere in cui viveva;
- tipo di quartiere in cui è stato messo in libertà vigilata;
- il suo primo lavoro dopo la libertà vigilata;
- il suo ultimo lavoro in istituto;
- la sua età mentale secondo l'esame psicologico;
- il suo tipo di personalità secondo l'esame psichiatrico;
- la prognosi psichiatrica.

Infine, sono stati determinati i tassi di violazione per intervalli di punteggio selezionati, come nella Tabella 2. La progressione regolare dei tassi in base alla grandezza dei punteggi di previsione sembrava affermare la fattibilità della previsione da una tabella di esperienza.

La sua analisi, quindi, per quanto semplificata, si rivelò incredibilmente precisa. Il 98% dei detenuti a basso rischio arrivò senza intoppi al termine del periodo di libertà vigilata.

Numero di punti	numero di uomini per gruppo	Percentuale non violazione della libertà su cauzione
16-21	68	98,5%
14-15	140	97,8%
13	91	91,2%
12	106	84,9%
11	110	77,3%
10	88	65,9%
7-9	287	56,1%
5-6	85	32,9%
2-4	25	24,0%

Tabella 2: Tabella originaria sviluppata da Burgess.

2.2 Alberi di decisione e foreste casuali

Oggi gli algoritmi di valutazione del rischio più evoluti in uso nei tribunali sono molto più sofisticati degli strumenti rudimentali di Burgess. Non si limitano ad aiutare nella concessione della libertà vigilata ma vengono utilizzati anche per personalizzare i programmi di sostegno rivolti ai detenuti, per decidere chi ha diritto alla libertà su cauzione e, ultimamente, anche per aiutare i giudici a decidere la severità della condanna. Il principio fondamentale è sempre lo stesso: si inseriscono i dati relativi all'imputato - età, precedenti, gravità del reato e così via- e si ottiene una valutazione del rischio che si corre a lasciarlo in libertà.

Come funzionano? A grandi linee, gli algoritmi più efficaci attualmente in uso sfruttano la cosiddetta tecnica delle foreste casuali, che si basa su un'idea incredibilmente semplice: **l'albero di decisione**.

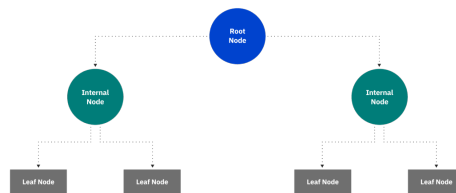


Figura 3: tipica struttura di un albero di decisione.

Si compone di una struttura ad albero gerarchica, che consiste di un nodo radice, di rami, nodi interni e nodi foglia. Come è possibile riscontrare dalla Figura 3, un albero decisionale inizia con un nodo radice, che non ha rami

in entrata. I rami in uscita dal nodo radice alimentano i nodi interni, noti anche come nodi decisionali. Sulla base delle funzionalità disponibili, entrambi i tipi di nodo conducono valutazioni per formare sottoinsiemi omogenei, che sono rappresentati da nodi foglia o nodi terminali. I nodi foglia rappresentano tutti i possibili risultati all'interno del set di dati. L'apprendimento dell'albero decisionale prevede l'utilizzo di una strategia **"dividi et impera"** applicata conducendo una ricerca per identificare i punti di divisione ottimali all'interno di un albero.

La Foresta Casuale invece è un algoritmo di apprendimento automatico che sfrutta la potenza di più alberi decisionali per prendere decisioni. In termini statistici più semplici, ecco come funziona: il computer esegue un algoritmo che seleziona i predittori in modo casuale e ripete e ripete questo processo per costruire diverse centinaia di alberi. - che poi permettono ai predittori selezionati a caso di fare la media per ottenere un unico risultato. Nel caso dello strumento di Philadelphia, questo risultato era l'assegnazione a una delle tre categorie di rischio (alto, moderato o basso) ai fini della sorveglianza della libertà vigilata.

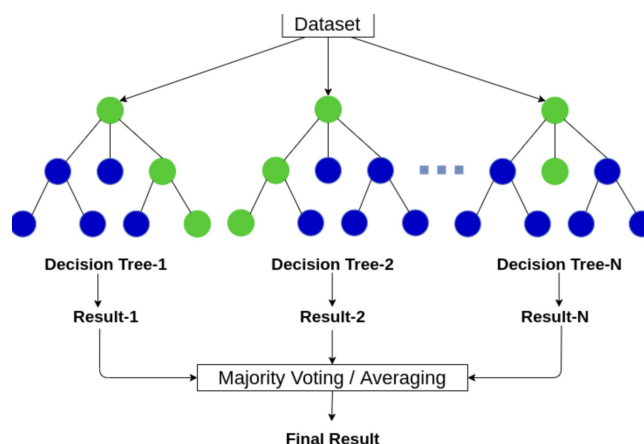


Figura 4: tipica struttura di una Foresta Casuale.

Nessuno strumento di previsione è perfetto, l'algoritmo difatti può commettere due tipi di errore: Il primo tipo di errore consiste nel rimettere in libertà un falso negativo (un falso negativo è una persona effettivamente ad alto rischio che è stata erroneamente identificata come a rischio moderato o basso). Il secondo tipo di errore è invece incarcerare un falso positivo, ovvero una persona effettivamente a basso o moderato rischio che è stata identificata, e quindi supervisionata, come ad alto rischio .

La chiave per costruire uno strumento di previsione a foresta casuale per qualsiasi aspetto del sistema giudiziario penale è bilanciare il rischio di sbagliare. Nel momento in cui gli operatori lavorano fianco a fianco con i ricercatori per stabilire questi parametri, incontreranno inevitabilmente la necessità di fare

dei compromessi con cui poter convivere. Questo processo comporta la determinazione, in anticipo, di un tasso di errore accettabile. Ecco come Hyatt ha spiegato il processo eseguito a Philadelphia⁵:

«Fondamentalmente, dovevamo determinare con precisione quanto sarebbe stato più costoso classificare erroneamente un soggetto in libertà vigilata in una categoria a basso rischio, che poi avrebbe commesso un reato grave, rispetto a quanto lo sarebbe stato sorvegliare intensamente qualcuno che in realtà è un vigilato a basso rischio perché lo strumento lo ha valutato ad alto rischio.»

La maggior parte delle giurisdizioni che intendono costruire uno strumento di previsione del rischio a foresta casuale probabilmente farebbero come a Filadelfia: fissare un costo relativo più alto per i falsi negativi che per i falsi positivi. L'APPD di Philadelphia ha deciso un rapporto di costo in cui i falsi negativi erano 2,6 volte più costosi dei falsi positivi. Ma qualsiasi giurisdizione che desideri progettare e implementare uno strumento simile dovrà determinare il proprio rapporto di costo o tasso di errore.

Come hanno notato Barnes e Hyatt, non esiste un'unica "risposta giusta" nella scelta dell'unità di previsione, dell'orizzonte temporale, della definizione degli esiti. Barnes ha aggiunto in merito:

«Non lo sottolineerò mai abbastanza, bilanciare questi diversi tipi di errori con il tasso di accuratezza complessivo del modello non è compito degli statistici del team. Poiché la dirigenza di un'agenzia deve convivere con le conseguenze di qualsiasi errore che si verifichi una volta che lo strumento di previsione diventa operativo, deve decidere con quale livello di accuratezza può convivere e quale equilibrio di potenziali errori preferisce.»

In definitiva, ciò richiede un'intensa collaborazione tra ricercatori e professionisti, in cui i funzionari delle agenzie - e non gli statistici - devono prendere decisioni politiche cruciali.

⁵Nel 2013 Geoffrey Barnes insieme al collega Jordan Hyatt del Jerry Lee Center of Criminology della Penn, ha creato e valutato lo strumento per aiutare il Dipartimento per la libertà vigilata e la libertà condizionale (APPD) di Philadelphia a creare un sistema computerizzato che preveda, con un alto grado di accuratezza, quali sono i probabili autori di reati violenti entro due anni dal rientro in comunità per garantire che i funzionari che sorvegliavano i soggetti in libertà vigilata ad alto rischio di recidiva avessero un carico di lavoro inferiore a quello dei funzionari che sorvegliavano i soggetti a rischio minore

3 Decisioni umane e previsioni artificiali

Introduzione alla sezione

L'intelligenza artificiale può accedere a un maggior numero di informazioni rilevanti per prevedere gli esiti delle controversie e dei procedimenti legali, supportando così il processo decisionale in tribunale e di previsione delle decisioni giudiziarie.

3.1 Prevedere le decisioni della Corte Suprema Americana

Gli scienziati e i giuristi hanno a lungo analizzato il lavoro della Corte Suprema⁶ degli Stati Uniti per capire quali sono le motivazioni che spingono i giudici nei loro verdetti. Nonostante le significative differenze metodologiche, entrambe le discipline cercano di spiegare le decisioni della Corte concentrandosi sull'esame dei casi passati. In altre aree del governo, per esempio, le elezioni presidenziali e i congressi decisionali del Congresso, gli scienziati politici si impegnano sistematicamente a prevedere gli esiti, ma pochi lo hanno fatto per le decisioni dei tribunali.

Utilizzando due metodi diversi, Andrew D. Martin insieme ad altri ricercatori, cercò di prevedere l'esito di ogni causa pendente davanti alla Corte Suprema durante il mandato dell'ottobre 2002 e confrontò tali previsioni con le decisioni effettive. Un metodo ha utilizzato un modello di previsione statistica basato su informazioni ricavate da precedenti decisioni della Corte Suprema. L'altro si è basato sui giudizi degli esperti in materia di accademici e professionisti. L'utilizzo di questi due metodi distinti permise di testare il loro potere predittivo non solo rispetto ai risultati effettivi della Corte, ma anche tra di loro.

Il modello statistico prendeva in considerazione solo una manciata di caratteristiche del caso, ognuna delle quali era facilmente osservabile senza una formazione specifica. Gli esperti legali, invece, potevano utilizzare conoscenze specifiche, come i fatti specifici del caso o le dichiarazioni dei singoli giudici in casi simili. Il modello statistico si differenziava dagli esperti anche perché prendeva esplicitamente in considerazione tutti i casi decisi da questo tribunale prima del mandato del 2002. Nessuno può avere una conoscenza così completa dei risultati della Corte negli ultimi otto mandati, e quindi gli esperti si sono necessariamente basati su un numero minore di osservazioni (anche se più dettagliate) del comportamento passato della Corte.

⁶La Corte Suprema è il più importante tribunale degli Stati Uniti per ciò che riguarda le leggi emanate nel paese e il loro rapporto con la Costituzione. È formata da nove giudici la cui carica è a vita: hanno il compito di garantire alla popolazione statunitense che ogni legge garantisca loro regole e diritti sanciti dalla Costituzione. Nel farlo, hanno ampio margine di manovra e possono farsi "interpreti" dei principi costituzionali.

Per ogni caso, i ricercatori scelsero alcune caratteristiche osservabili da utilizzare come variabili esplicative scelte per motivi esplicitamente teorici e in base alla loro disponibilità e alla relazione plausibile con il processo decisionale della Corte Suprema.

Per ogni caso che prevedettero per questo studio, utilizzarono gli alberi di classificazione stimati per generare una previsione prima della discussione della Giuria. Nella Figura 5 presentiamo gli alberi di classificazione stimata per il giudice O'Connor.

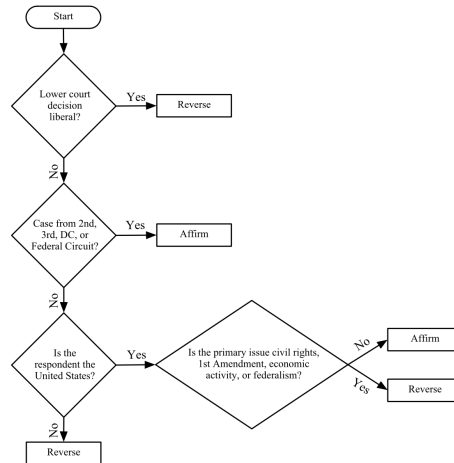


Figura 5: Albero di classificazione stimato per il giudice O'Connor.

L'altro metodo di previsione dello studio ha cercato di catturare i giudizi degli esperti di diritto. Gli esperti si distinguono dai non esperti per l'ampia formazione ed esperienza nel settore di riferimento; hanno inoltre la capacità di percepire schemi significativi che non possono essere facilmente codificati in un modello statistico e di strutturare le loro conoscenze attorno a schemi basati su principi; spesso i loro giudizi si basano su analisi qualitative. Il processo alla base dei giudizi dei nostri esperti legali differisce quindi notevolmente da quello alla base del modello.

Le 83 persone che hanno partecipato sono facilmente qualificabili come "esperti", avendo scritto e insegnato sulla Corte Suprema, esercitato la professione davanti alla Corte Suprema e/o fatto da cancelliere, e sviluppato conoscenze speciali in uno o più settori. Nel complesso, essi formarono un gruppo di 71 accademici e 12 avvocati d'appello. Di questo gruppo, 38 sono stati assistenti di un giudice della Corte Suprema, 33 sono titolari di cattedre e 5 sono attuali o ex presidi di facoltà di legge.

Agli esperti, liberi di considerare tutte le fonti di informazione o i fattori che

ritenevano rilevanti per fare la loro previsione, fu chiesto di prevedere uno o più casi all'interno delle loro aree di competenza. Le loro previsioni furono richieste prima della discussione orale, assicurando loro che le loro previsioni individuali e i casi a cui erano stati assegnati non sarebbero stati rivelati. Le loro previsioni hanno assunto la forma di una scelta "affermare o annullare" per la Corte nel suo complesso e per ciascun giudice: limitare gli esperti a una scelta binaria era necessario per un confronto diretto con i risultati del modello statistico.

Infine, su 68 dei casi utilizzati per analizzare le previsioni sull'esito del caso il modello ha previsto correttamente il 75,0% degli esiti finali dei casi, mentre le previsioni degli esperti hanno avuto un tasso di accuratezza solo del 59,1%. Quindi, già agli inizi dei primi anni Duemila questa nuova tecnologia, pur essendo stata sviluppata da poco, fu in grado di competere ad ottimi livelli con delle controparti umane, sorpendendo gli stessi autori di questo esperimento.

3.2 Ridurre la popolazione carceraria con gli algoritmi

Ogni giorno prendiamo decisioni basate sulle informazioni che abbiamo a disposizione. Allo stesso modo, i giudici devono decidere se trattenere in carcere le persone che stanno affrontando un processo penale o se permettere loro di tornare a casa fino al processo sulla base delle informazioni disponibili. Utilizzano le informazioni, come i precedenti penali e la natura del reato contestato per valutare il rischio che una persona non si presenti al suo futuro processo, noto come "rischio di fuga", o commetta un crimine prima del processo.

Nel 2017, un gruppo di ricercatori ha deciso di scoprire se le previsioni della macchina sono paragonabili, per accuratezza, a quelle dei giudici in carne e ossa. Per far ciò, gli scienziati hanno avuto accesso ai dati relativi a tutti gli arresti compiuti a New York nell'arco di quattro anni, dal 2008 al 2013. Le persone che in quel lasso di tempo avevano chiesto di essere messe in libertà su cauzione erano state 750 000, un numero di gran lunga sufficiente per poter confrontare sui singoli casi un algoritmo e un giudice umano.

Nei casi in questione il sistema giudiziario newyorchese non aveva fatto ricorso ad algoritmi; i ricercatori, quindi, hanno creato un gran numero di alberi di decisione per valutare con che precisione, all'epoca, si sarebbe potuto prevedere il rischio di una violazione della cauzione. Partendo dai dati personali - fedina penale, reato commesso e così via- si è calcolata la probabilità che l'accusato violasse i termini. Una cosa, tuttavia, è sicura: le valutazioni della macchina non hanno confermato quelle dei giudici. I ricercatori hanno dimostrato che molti degli imputati segnalati dall'algoritmo come soggetti davvero poco raccomandabili erano stati considerati dai giudici come elementi a basso rischio: quasi metà degli imputati segnalati dall'algoritmo come ad altissimo rischio erano stati effettivamente rilasciati su cauzione.

I dati hanno inoltre dimostrato che il gruppo identificato dall'algoritmo come ad alto rischio era stato davvero fonte di problemi. Poco più del 56% non si era presentato in tribunale e il 62,7% aveva approfittato della libertà per commettere altri reati, tra cui i più efferati di tutti: lo stupro e l'omicidio. Utilizzando il modello di apprendimento automatico del rischio di fuga invece di quelli dei giudici per selezionare le persone da trattenere prima del processo potrebbe avrebbero potuto diminuire le mancate comparizioni del 24,7% e tutti gli altri tipi di reato dell'11,1% senza alcun cambiamento nei tassi di detenzione preventiva, oppure ridurre la popolazione di detenuti in attesa di giudizio del 42,0% senza alcuna variazione del tasso di mancata comparizione.

Essi hanno riscontrato che il modello di apprendimento automatico supera l'attuale prassi dei giudici, prevedendo con maggiore precisione chi non si presenterà al processo e classificando correttamente le persone che dovrebbero essere detenute prima del processo più frequentemente di quanto non facciano i giudici, utilizzando le stesse informazioni sul caso di cui dispongono i magistrati.

L'uso dell'apprendimento automatico per valutare il rischio è sempre più comune nella ricerca sulla giustizia penale, e i risultati spesso dimostrano che l'apprendimento automatico è più accurato dei metodi statistici comunemente utilizzati nella pratica, superando anche le decisioni degli stessi giudici.

I vantaggi infatti non sono solo teorici. Lo stato del Rhode Island, i cui tribunali hanno adottato questo tipo di algoritmi otto anni fa, è riuscito a ridurre del 17% la popolazione carceraria con un calo del 6% delle recidive. Significa centinaia di imputati a basso rischio che non finiscono inutilmente in prigione e centinaia di reati in meno.

Conclusioni

Nonostante la crescente disponibilità di dati sull'esperienza delle persone in materia di giustizia, le istituzioni di tutto il mondo hanno ancora molta strada da fare per raggiungere la giustizia per tutti. L'IA potrebbe contribuire ai sistemi giudiziari assistendo l'amministrazione e, a lungo termine, a ridurre significativamente i costi per le autorità giudiziarie: tutti questi miglioramenti delle prestazioni possono infine tradursi in un miglioramento dell'accesso alla giustizia e in una riduzione dei tempi di pronuncia delle decisioni giudiziarie. Questi benefici, tuttavia, dipenderanno in larga misura dalla robustezza e dall'affidabilità degli strumenti impiegati.

Alcune delle applicazioni dell'IA nel campo della giustizia possono avere significative implicazioni legali, etiche e sui diritti fondamentali. In questi casi, lo sviluppo e la diffusione di soluzioni basate sull'IA dovrebbero essere valutati con particolare attenzione, ricordando che l'implementazione di tali soluzioni in ambito giudiziario non può essere considerata una sfida puramente tecnica. Tuttavia, come emerge da questo articolo, se i giudici potessero contare su un algoritmo - anche imperfetto - che rafforzi la loro capacità di giudizio, spesso così fallace, potremmo con sicurezza affermare di aver fatto un passo nella direzione giusta. Un algoritmo ben progettato e adeguatamente perfezionato potrebbe quantomeno eliminare la parzialità sistematica e gli errori casuali tipici di noi umani.

Riferimenti bibliografici

- [1] William Austin :*A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity.* 1977
- [2] Mandeep K. Dhani : *BAILING AND JAILING THE FAST AND FRUGAL WAY: AN APPLICATION OF SOCIAL JUDGEMENT THEORY AND SIMPLE HEURISTICS TO ENGLISH MAGISTRATES' REMAND DECISIONS.* 2001
- [3] Mandeep K. Dhani, Ian K. Belton, Elizabeth Merrall, Andrew McGrath, and Sheila M. Bird :*Criminal Sentencing by Preferred Numbers.* 2020
- [4] Karl F. Schuessler, :*Parole Prediction: Its History and Status.* 1955
- [5] Nancy Ritter :*Predicting Recidivism Risk: New Tool in Philadelphia Shows Great Promise.* 2013
- [6] Andrew D. Martin, Kevin M. Quinn, Theodore W. Ruger, and Pauline T. Kim :*Competing Approaches to Predicting Supreme Court Decision Making.* 2003
- [7] Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan :*Human Decisions and Machine Predictions.* 2017
- [8] Hannah Fry :*Hello World.* 2019
- [9] Pedro Domingos :*L'ALGORITMO DEFINITIVO.* 2016