

Calcio, Big Data e il sol dell'avvenire

Un computer a disposizione di un allenatore era un'utopia, nel 1973. La storia dei dati e del calcolo applicato al calcio non poteva che cominciare in un posto dove di utopia, nel 1973, se ne intendevano: l'Unione Sovietica. In quell'anno il colonnello Lobanovs'kyj iniziò ad allenare la Dinamo Kiev, e come prima richiesta nel suo staff volle uno statistico ed un computer. Di calcolatori se ne vedevano talmente pochi che il KGB lo mise subito sotto controllo. "Tutto è un numero" profetizzava il colonnello. Addestrò la sua squadra alla ripetizione di schemi elaborati al computer. Fu il primo ad introdurre il calcio totale e portò la Dinamo Kiev e la nazionale russa ai massimi livelli del calcio europeo.

Quando invece i computer erano ancora solo prototipi e i Big Data una visione fantascientifica, il ragionier Charles Reep riempiva taccuini su taccuini con i dati delle partite a cui assisteva. Quindi li elaborava, rigorosamente a mano. Cominciò nel 1950, quando durante un partita dello Swindon Town si era annoiato talmente tanto da cominciare ad annotare tutto ciò che succedeva in campo. La sua scoperta è ancora alla base di tanti sistemi di gioco: la probabilità di sbagliare un passaggio aumenta con il numero di passaggi consecutivi. Altro che Tiki-Taka, Reep dimostrò numericamente l'efficacia del contropiede: portando prima possibile la palla nell'area avversaria si massimizza il numero dei gol. Questa teoria, conosciuta come "Teoria della palla lunga", venne pubblicata sul prestigioso Journal of Royal Statistical Society e ha ispirato più o meno tutto il calcio inglese dagli anni '60 in poi.

Il frutto più recente della rivoluzione dei dati è il Midtjylland, squadra fondata nel 1999 e fresca vincitrice del

campionato danese. L'Ad della società, il trentaduenne Rasmus Ankersen, ha lanciato una provocazione fantascientifica: "l'algoritmo è più importante della classifica". Il padrone del Midtjylland è Matthew Benham, uno scommettitore incallito che con i proventi delle vincite acquista squadre di calcio. La scommessa di Benham e Ankersen si è rivelata poco azzardata: il loro scouting algebrico li ha portati a selezionare i giocatori giusti durante il calciomercato, facendoli poi rendere al meglio delle loro possibilità.



L'occhio (poco attento) di Ferguson su Stam

I dati possono anche portare a prendere una cantonata, come successe a Ferguson nel 2001. Le statistiche evidenziavano un calo dei tackle di Stam, suo difensore al Manchester United, perciò sir Alex decise di vendere il giocatore, considerandolo in calo. Stam restò su alti livelli di rendimento per altri sei anni, prima alla Lazio e poi al Milan. Ferguson non aveva considerato il "principio di Maldini": Paolo Maldini è stato uno dei più grandi difensori della storia pur facendo, in media, un tackle ogni due partite.

La ricerca dell'alchimia che si nasconde dietro una vittoria non poteva che affascinare anche il mondo della scienza. Undici giocatori che ne affrontano altrettanti, ognuno libero di muoversi nello spazio ma con un obiettivo comune a tutti e indipendente (o quasi) dal singolo. Una sfida, più che un problema da risolvere. Uno dei lavori più interessanti è quello di Taki e Hasegawa, con la loro definizione di 'regione

dominante'. Analizzando i dati di tracking dei giocatori, cioè la loro posizione in campo registrata ogni decimo di secondo, i giapponesi hanno elaborato un modello geometrico in grado di calcolare, ad ogni istante, l'area che un singolo giocatore può raggiungere prima di tutti gli altri. Sebbene onerosa in termini di calcolo, questa misura si è rivelata molto interessante: la strategia di attacco una squadra può essere valutata in base alla capacità di massimizzare le regioni dominanti dei suoi giocatori. Un team di ricercatori australiani guidati dal prof. Horton ha invece coinvolto dieci allenatori nella costruzione di un classificatore di passaggi: ogni allenatore ha visionato e valutato una serie di passaggi (intelligente, scontato, etc); sulla base di questi giudizi è stato sviluppato un algoritmo in grado di riprodurre il ragionamento degli allenatori, automatizzando tutto il processo.

Il santo Graal della Data Science applicata al calcio è proprio la valutazione dell'intelligenza di una mossa. Quanto incide la decisione sul dove e a chi passare la palla? Quanto è efficace uno scatto in profondità o una rincorsa dell'avversario a metà campo? Il fiuto nel rispondere a queste domande è ciò che fino ad oggi ha reso grandi gli allenatori. Lo scenario sta però cambiando, e i presidenti sanno già che il loro prossimo tecnico dovrà essere anche un bravo Data Scientist.

Mattarella e il complotto dei

dati

Accessi alla pagina del Presidente della Repubblica su wikipedia

A partire da dicembre un flusso di notizie, pettegolezzi, opinioni e retroscena ci ha sommerso, come per ogni elezione del presidente della Repubblica che si rispetti. Questa volta la portata della piena è stata maggiore, perché il periodo è tra i più contorti della storia contemporanea. L'ultima elezione del Presidente è avvenuta meno di due anni fa, anziché i sette convenzionali. Il Presidente uscente è dimissionario (raro) e al secondo mandato (mai successo prima): la riconferma di Napolitano arrivò al termine di votazioni in cui successe di tutto, dalla silurazione di Prodi alla rivolta contro Marini, passando per le quirinarie dei 5stelle dove arrivò prima una giornalista (Gabanelli) e secondo un medico (Gino Strada). Con questi presupposti, l'elezione del Presidente della Repubblica si preannunciava come una specie di thriller.

Eppure, nel mare magnum della rete la soluzione all'enigma era già lì, pronta due giorni prima dell'ultimo scrutinio. Il 29 gennaio Renzi sciolse la sua riserva. Durante la direzione PD comunicò chi aveva scelto come candidato da far votare al suo partito: Sergio Mattarella. I dati di accesso a Wikipedia erano chiari: il 29 gennaio, ben due giorni prima dell'elezione, la partita poteva già dirsi chiusa. Questi dati hanno una particolarità: indicano il numero di accessi alla pagina, ma comprendono anche gli accessi fatti quando il nome viene cercato su google e compare il box a destra con le informazioni principali tratte da wikipedia. Sono dunque una valutazione più che consona dell'interesse che c'è intorno a quel nome.

Oggi sappiamo già tutto di Mattarella, dalla storia personale

agli hobby, ma prima di quel giorno non era proprio così. Prima dell'annuncio, nessuno lo conosceva. I dati di Wikipedia sono inequivocabili: all'annuncio del premier, tutta la rete è andata alla ricerca della storia dell'ex giudice costituzionale-ex ministro-ex democristiano designato da Renzi come candidato unico alla presidenza.

Mattarella chi?

Nel frattempo, il secondo partito italiano ha organizzato le consuete Quirinarie, per chiedere ai suoi iscritti di scegliere il candidato da votare. L'annuncio dei 5stelle è arrivato lo stesso giorno, il 29 gennaio, ma l'interesse della rete è su scala decisamente minore. Il risultato si può interpretare in vari modi: Imposimato era già nome noto e popolare, oppure intorno al candidato dei 5stelle, famoso o meno che fosse, non c'era una grande curiosità. Lo stesso si può dire per il candidato dell'estrema destra Vittorio Feltri, sebbene partisse da una popolarità maggiore. Per meglio comprendere il fenomeno, vale la pena osservare anche l'interesse attorno al nome di Romano Prodi: ad un certo punto la minoranza del PD sembrava volesse votarlo, la rete sembra confermare che delle mosse della minoranza PD non gli importa poi chissà quanto. Anzi, "se i dati fossero confermati" - come si dice al termine delle elezioni - alla rete, delle intenzioni della minoranza PD, non gliene può fregare di meno.

Wikipedia anticipa i risultati delle elezioni. Il complotto dei dati.

Al termine della cavalcata trionfale di Mattarella, vale la pena dare uno sguardo a cosa è successo nelle retrovie. Come ad ogni elezione che si rispetti, a volte dalle urne sbucano fuori nomi poco consueti. Al quarto scrutinio, Walter Veltroni (PD), Roby Facchinetti (Pooh) e Lino Banfi (BA) hanno ottenuto lo stesso numero di voti: 1. Francesco Guccini li ha surclassati, ottenendone il doppio. Ancora prima, Giancarlo

Magalli aveva vinto a mani basse il sondaggio de Il Fatto Quotidiano. Quanta curiosità c'era intorno a questi nomi? In quanti li hanno cercati su Google? Vittoria schiacciante di Magalli, idolo di internet e trasformato da conduttore televisivo a opinionista politico. A gennaio ha inaugurato il suo blog all'interno de L'Espresso.

Il gennaio d'oro di Magalli

Il turbine di emozioni previsto per le elezioni del Presidente della Repubblica non si è rivelato quello previsto. Sergio Mattarella è stato votato senza affanni né tradimenti. Non c'era nessun altro in campo, ma soprattutto i 101 franchi tiratori del PD sono stati buoni questa volta. Bontà forse dettata dalla disciplina di partito - oggi il segretario è Renzi, nel 2013 c'era Bersani - o dalla consapevolezza che quando scende in campo la balena bianca, non c'è partita per nessuno.

Muovere il mouse sulla figura per scoprire cosa fa la DC ai franchi tiratori

Se il traffico è il problema, nei dati c'è la soluzione

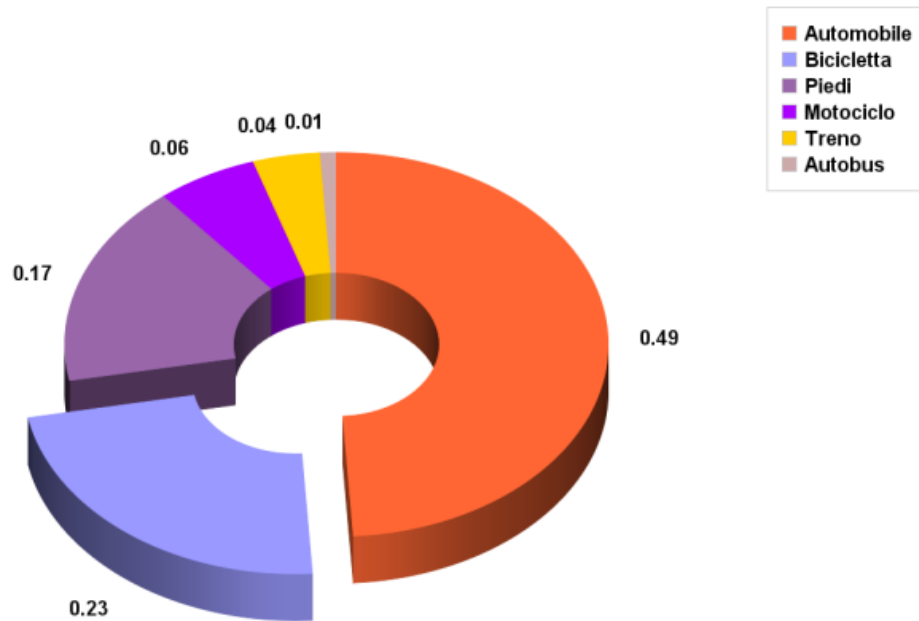
Il traffico genera dati, i dati forniscono risposte. Le centinaia di volontari del progetto [TagMyDay](#) hanno raccolto e donato i dati relativi alla loro mobilità: utilizzando il loro smartphone per ottenere dati GPS sui propri spostamenti, hanno poi annotato i loro viaggi in base a mezzo di trasporto e scopo. La prima fase di raccolta dati si è conclusa il mese

scorso, ora è il momento dell'analisi. Le prime risposte ai problemi del traffico che i dati ci suggeriscono sembrano quelle di un indovino, quelle che una volta lette il primo pensiero è: "era così facile?"

Come ci muoviamo?

Il dataset di TagMyday rappresenta, al momento, il sondaggio più preciso in circolazione sulla mobilità pisana. Pisa è una città fortunata sotto molti punti di vista e sfortunata per alcuni altri. L'orografia rende facile muoversi: a Pisa non esistono salite. La piccola dimensione favorisce la mobilità breve: i servizi sono tutti a disposizione nel raggio di pochi chilometri. Lo sviluppo urbano, invece, rappresenta un piccolo tallone d'achille. Nelle ore di punta, il traffico in città è lento, lentissimo. Le auto ingolfano i ponti che attraversano la città nella direttrice nord-sud, i bus lunghi hanno problemi a districarsi nelle strade strette, gli automobilisti girano alla ricerca di un parcheggio come anime perse. Questo mix di ingredienti decreta la vittoria di un mezzo di trasporto antico, povero, eppure ancora fondamentale: la bicicletta. L'automobile è il mezzo principale, ma non è in maggioranza. Il secondo mezzo più usato a Pisa è la bicicletta: quasi uno spostamento su quattro viene fatto pedalando.

Utilizzo mezzi di trasporto



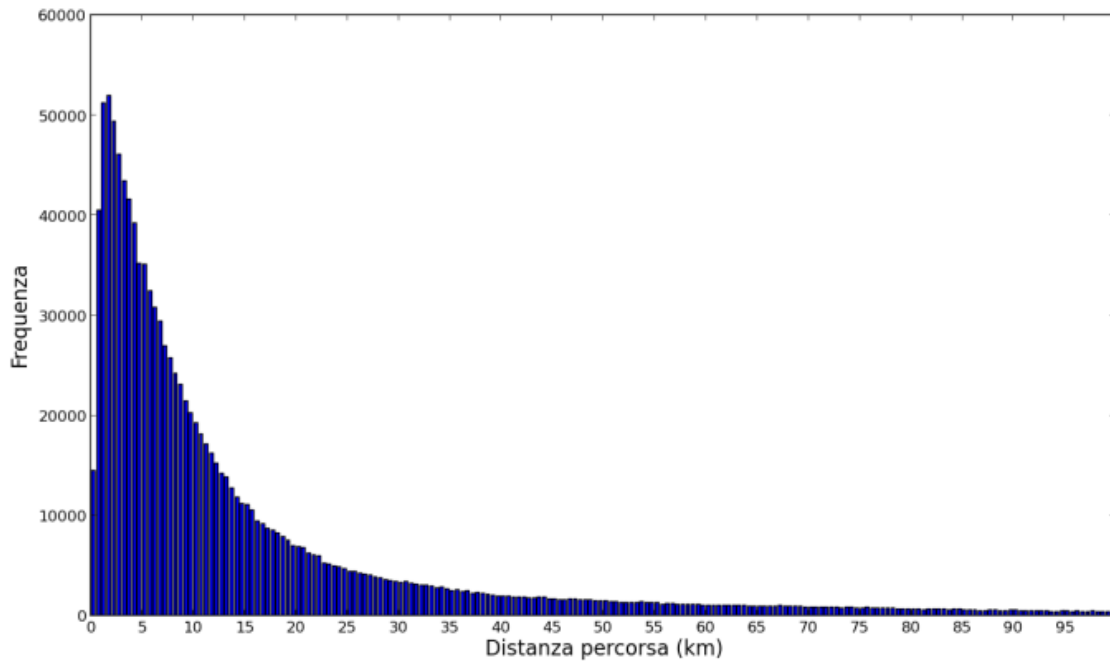
Vogliono la bicicletta e pedalano

Perché ci muoviamo in bicicletta?

La bicicletta è un'invenzione risalente a due secoli fa, ed è rimasta tale da allora. La modernità non ne ha intaccato l'utilizzo. I dati ci spiegano anche il perché di questo successo secolare. Questa volta non i dati di TagMyDay, ma un dataset ancora più grande, utilizzato dal [KDDLab](#) per la ricerca nel campo della mobilità.

I dispositivi installati dalle assicurazioni permettono non solo di pagare meno, ma anche di raccogliere dati preziosi per analizzare il traffico automobilistico a tutti i livelli, dal locale in su. L'analisi di un dataset simile fa emergere il primo indizio:

su circa un milione e mezzo di traiettorie GPS raccolte, i viaggi più frequenti sono quelli brevi, inferiori a cinque chilometri. I dati si riferiscono ad utenti residenti nell'area pisana e zone limitrofe (Lucca, Empoli, etc).



Quanto viaggiamo: poca strada, tante volte

I pisani fanno viaggi brevi, per questo usano la bici. Il grafico precedente si riferisce al traffico automobilistico e ispira nuove domande.

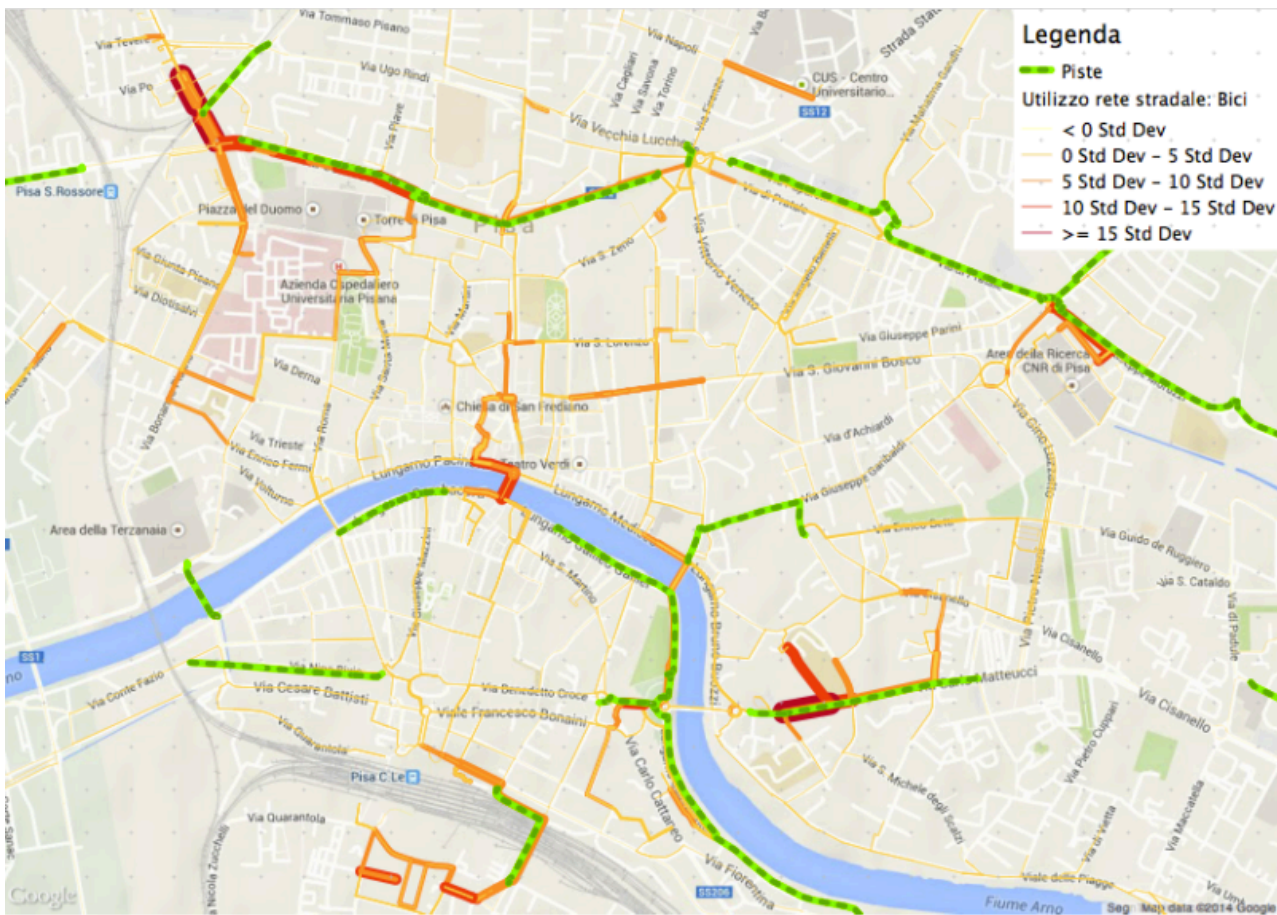
Quanto si guadagna ad andare in bicicletta?

In mezzo a quel milione di viaggi ce n'è un 55% costituito da viaggi brevi. Trasformiamoli tutti in viaggi in bicicletta. Questo significa un taglio della CO₂ di oltre il 20% (calcolo basato su dati emissioni presi da [Quattruote](#)). L'ambiente, come prevedibile, ne giova eccome. Il portafoglio del ciclista non è da meno. Prendendo in considerazione tutti gli automobilisti che effettuano almeno 20 viaggi brevi al mese (profilo tipico di un lavoratore), il risparmio mensile è compreso tra 24 e 76 euro al mese (costi al km ricavati da [tabelle ACI](#), considerando una vettura utilitaria). In un anno possono arrivare quasi a **1000 euro**. Chi sceglie la bici ha una tredicesima in più. La salute segue. A parte i vantaggi al sistema cardiovascolare, l'automobilista di cui sopra aggiunge al risparmio economico quello in termini di peso: andare al lavoro in bici significa bruciare una media di 200 calorie al

giorno in più (secondo il modello di [MapMyRide](#)). In un mese significa perdere 0.8 kg di peso corporeo (secondo il calcolatore fornito da [Livestrong](#)) Anche la palestra diventa superflua, insomma.

Sembra ormai chiaro come la bicicletta faccia guadagnare tutti. Ambiente, collettività, singolo utente, sono tanti i soggetti a poter approfittare di una migrazione di massa dall'automobile alla bici. I tanti nord europei che sfrecciano tra le strade di Amsterdam o Copenaghen in bici non sono dunque dei pazzi. Mettono il ketchup sugli spaghetti, al massimo, ma almeno dal loro usare intensivamente la bicicletta traggono un vantaggio notevole. Manca ancora qualcosa, però, alla risposta finale.

Dove usiamo la bici?



Dove si pedala di più a Pisa? (Non sempre sulle piste...)

Tra le lamentele più comuni tra gli utenti della strada, c'è

l'assenza di piste ciclabili. Pisa non è certo esente da queste critiche. Progettare e trovare spazio per le piste ciclabili in una città così piccola non è compito semplice. I dati di TagMyDay relativi al traffico ciclistico mostrano come le poche piste ciclabili di Pisa sono molto utilizzate, quando servono punti nevralgici. L'immagine sopra mostra l'utilizzo della rete stradale pisana da parte dei ciclisti. Da questi dati si può partire per una pianificazione della mobilità, andando incontro alle esigenze dei cittadini.

Era ovvio, no?

La facilità con cui si raccolgono i dati è un'opportunità da non gettare al vento. Da un'analisi partita da utenti dotati di smartphone – ormai lo strumento più comune del mondo – con una app passiva che ha registrato i loro movimenti, si può investigare e capire un fenomeno complesso come quello della mobilità urbana. Utilizzando i dati di automobili dotate di "scatola nera" si può capire quanto e come la mobilità incide sulla nostra vita anche in termini di salute e di soldi.

Un indovinello è semplice una volta che si conosce la risposta. La soluzione ai nostri problemi di mobilità sta nel mezzo più semplice di tutti, nonostante sia il mezzo più antiquato in circolazione. In un mondo in cui un'automobile dopo qualche anno diventa già da cambiare, non era immediato arrivare a questa risposta. La bicicletta non è più il mezzo dei poveri, **è invece il mezzo che ci rende più ricchi**. La bicicletta costa poco, è accessibile, è veloce in città e rende liberi di muoversi senza i vincoli di spazio e di tempo del trasporto pubblico.

Ci sono voluti dati, esperimenti, analisi e figure, ma ora, finalmente, la risposta ce l'abbiamo. E dire che era così facile.

