

POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale

**Corso di Laurea Magistrale
in Ingegneria Gestionale**

Tesi di Laurea Magistrale

Il mondo dei dati



Relatore
Prof. Carlo Cambini

Candidato
Giovanni Peluso

Anno accademico 2022-2023

INDICE

INTRODUZIONE	4
CAPITOLO 1: I BIG DATA	5
<i>1.1 Generalità e caratteristiche</i>	5
<i>1.2 Caratteristiche dei Big Data: le 5 V</i>	8
<i>1.3 La Value Chain dei Big Data</i>	10
<i>1.3.1 Raccolta/Acquisizione dei dati</i>	11
<i>1.3.2 Archiviazione dati e cloud computing</i>	13
<i>1.3.3 Analisi e utilizzo dei dati</i>	13
<i>1.4 Le barriere all'ingresso</i>	15
<i>1.4.1 Ostacoli alla raccolta/acquisizione dei dati</i>	15
<i>1.4.2 Ostacoli all'archiviazione dei dati</i>	17
<i>1.4.3 Ostacoli all'analisi dei dati</i>	18
<i>1.4.4 Ostacoli all'utilizzo dei dati</i>	18
<i>1.5 Principali tecnologie</i>	19
<i>1.6 La classificazione dei dati</i>	21
<i>1.7 La condivisione dei dati</i>	24
<i>1.8 Introduzione alle piattaforme digitali</i>	26
<i>1.9 Implicazioni in termini di privacy</i>	28
CAPITOLO 2: PRIVACY E REGOLAMENTAZIONE	29
<i>2.1 Generalità e caratteristiche</i>	29
<i>2.1.1 Relazione tra target advertising e privacy</i>	33
<i>2.1.2 Il paradosso della privacy</i>	36
<i>2.1.3 Il “calcolo del valore della privacy”</i>	38
<i>2.2 Il GDPR</i>	41
<i>2.2.1 I soggetti del GDPR</i>	42
<i>2.2.2 Principali applicabili al trattamento dei dati personali</i>	44
<i>2.2.3 Sanzioni previste</i>	47
<i>2.2.4 Impatto del GDPR</i>	49
<i>2.3 Il Garante Privacy belga sanziona IAB Europe</i>	56

<i>2.3.1 Il Transparency & Consent Framework (TCF)</i>	57
<i>2.3.2 L'intervento dell'autorità belga e i punti di merito</i>	58
<i>2.3.3 Cosa seguirà alla decisione</i>	59
CAPITOLO 3: DATA BROKERS-MODELLO MATEMATICO DI MERCATO	61
<i>3.1 Generalità e caratteristiche</i>	61
<i>3.2 Modellazione del mercato dei DATA Brokers</i>	66
<i>3.3 Produzione dei DATA e commercializzazione</i>	75
CAPITOLO 4: MODELLI DI MERCATO	78
<i>4.1 Generalità e caratteristiche</i>	78
<i>4.2 Il modello di Evans e Schmalensee (2012)</i>	81
<i>4.3 Aspetti caratterizzanti il mercato</i>	89
CONCLUSIONI	92
BIBLIOGRAFIA	93
SITOGRAFIA	94
RINGRAZIAMENTI	96

INTRODUZIONE

La navigazione in rete, coinvolge la quotidianità di ognuno, dalla semplice introduzione su motore di ricerca, al mi piace su Facebook.

Siamo esposti ai data broker che raccolgono le tracce delle nostre visite per estrarre informazioni sui nostri consumi, che saranno venduti ad altri data broker, aziende e/o individui.

Le nostre tracce di navigazione rivelano informazioni, spesso personali, che vengono catturate per essere introdotte in un processo di vendita di informazioni che non tutela la privacy.

I dati vengono impiegati dai data broker per categorizzare i consumatori che non hanno la possibilità di interagire con gli stessi data broker che praticano attività di raccolta, analisi e condivisione.

L'industria dei data broker è fiorente ed anima un mercato in espansione caratterizzato da tre categorie di data broker:

1. data broker impegnati nella ricerca di informazioni personali;
2. data broker impegnati nel marketing per lo sviluppo di dossier per campagne di marketing;
3. data broker impegnati nella categorizzazione e modellazione dei consumatori attraverso interessi e attività offline per indirizzare sconti, offerte, coupon e altro ancora.

In sintesi l'elaborato:

- nel primo capitolo, esplora il mondo dei BIG DATA e le attività in cloud come argomentazioni principali;
- nel secondo capitolo affronta le problematiche di privacy e regolamentazione, soffermandosi sul regolamento europeo GDPR in materia di protezione dei dati personali che tende ad armonizzare le regole sulla raccolta e il trattamento dei dati personali;
- nel terzo capitolo introduce la figura del data broker, soffermandosi sulle attività prevalenti quali raccolta dei DATA e relativa commercializzazione;
- nel quarto capitolo riporta il tentativo, da intendersi come studio personale, di adattare un modello matematico di previsione esistente e sviluppato per un mercato differente ad un ipotetico scenario caratterizzato da data broker venditori ed acquirenti di data.

Capitolo 1

I BIG DATA

1.1 Generalità e caratteristiche

In che modo i dati influenzano le dinamiche aziendali?

Man mano che i mercati diventano più concentrati e dominati da aziende esperte di dati, è importante comprendere il ruolo macroeconomico dei dati.

Un quadro macro consente agli economisti di valutare i dati delle imprese e capirne gli effetti.

L'attuale società della conoscenza è prevalentemente digitale e caratterizzata da una crescente concentrazione di informazioni nelle mani di pochi soggetti, sia pubblici che privati.

Il ruolo svolto dalle aziende nella generazione dei flussi di dati (es. società di telecomunicazioni, grande distribuzione) e nella circolazione dei dati (es. motori di ricerca, data broker) è la ragione principale di questa concentrazione.

Le grandi aziende private raccolgono e analizzano enormi quantità di dati durante lo svolgimento delle loro attività quotidiane.

Questa mole di informazioni rappresenta un asset strategico ed economicamente rilevante, poiché la gestione di grandi database consente alle aziende di assumere il ruolo di gatekeeper per quanto riguarda le informazioni e le conoscenze che possono essere estratte dai dataset.

Questo sviluppo è stato coniato "ascesa dei big data" o "dataficazione" e si spiega con due recenti innovazioni tecnologiche simultanee (Mayer-Schönberger e Cukier, 2013):

-in primo luogo: la crescente disponibilità di dati;

-in secondo luogo: la crescente capacità delle imprese e dei governi di analizzare il nuovo big set di dati.

Nel mondo, in ogni minuto, vengono scambiate milioni di informazioni.

Solo per fare qualche esempio, ogni 60 secondi, su Facebook, vengono creati 3,3 milioni di post, pubblicati 510.000 commenti e aggiornati 293.000 stati; su Twitter vengono inviati 470.000 tweet; su WhatsApp vengono scambiati 29 milioni di messaggi; su Google vengono effettuate 3,8 milioni di ricerche.

I Big data sembrano entrare dappertutto, in ogni sfera della nostra vita, pubblica e privata. E il dibattito mondiale circa la loro estrazione e il loro impiego ha assunto ormai le caratteristiche di una vera e propria emergenza.

Secondo il rapporto “Digital in 2017 Global Overview” in Italia vengono spese, mediamente, 4 ore al giorno in rete davanti al desktop del nostro computer e due ore e otto minuti nella connessione da postazione da mobile. Un dato, quest’ultimo, che segna un primato europeo.

Come ha recentemente evidenziato l’Agcom: “i social network sono definitivamente divenuti parte integrante della dieta informativa quotidiana dei cittadini in Italia e nel mondo”.

Nel cercare di dare una definizione esaustiva al concetto di Big Data, senza tralasciare una accurata analisi della letteratura, non è possibile non partire dal semplice concetto di dato.

I dati sono descrizioni codificate di informazioni e costituiscono l’input principale del data market.

La loro importanza non dipende dal dato stesso ma dal suo uso da parte di aziende, individui e governi che ne usufruiscono con lo scopo di supportare i loro processi decisionali e gestionali.

Sono considerati un asset informativo fondamentale che necessita di nuove modalità di elaborazione mediante le quali è possibile migliorare le decisioni, intuire le scoperte ed ottimizzare processi. Godono della non rivalità nel consumo, ovvero l’utilizzo degli stessi da parte di un utente non preclude l’utilizzo ad altri soggetti e della non escludibilità, vale a dire non se ne può impedire la fruizione ad utenti che non hanno pagato per averli.

Non esiste una definizione univoca di Big Data ma ne esistono di diverse che comunque differiscono per alcuni elementi.

Ciò nonostante le definizioni che vengono maggiormente acclamate dalla letteratura individuano come fondamentale la grande dimensione del set di dati, la necessità di utilizzare potenza di calcolo su larga scala e software e metodi non standard per estrarre valore dai dati in un ragionevole lasso di tempo.

Nel 2010 Eric Schmidt sosteneva: “Nel giro di due giorni produciamo la stessa quantità di dati generati dagli albori della civiltà sino al 2003”.

Dunque ciò che maggiormente caratterizza i dati è il loro volume, giustificando la ragione per la quale tale fenomeno prende il nome di “Big Data”.

Nel 2011 la McKinsey and Company ha definito il fenomeno come la nuova frontiera dell'innovazione, della concorrenza e della produttività.

Non è di certo un segreto che negli ultimi anni i dati e la loro analisi vengano considerati come una risorsa fondamentale ai fini di una corretta gestione aziendale.

Secondo l'Unione Europea i big data sono “grandi quantità di tipi diversi di dati prodotti da varie fonti, fra cui persone, macchine e sensori”.

Nel lontano 2001, la società di consulenza strategica Gartner, definiva i big data come: “asset informativi ad alto volume, alta velocità e alta varietà che richiedono nuove forme di lavorazione ed elaborazione delle informazioni che permettono di migliorare le decisioni, intuire scoperte e ottimizzare i processi”.

Un'altra definizione può ricondursi a Manyika et al. (2011) secondo cui: “Con Big Data ci si riferisce a set di dati la cui dimensione va oltre la capacità dei tipici strumenti software di database di catturare, archiviare, gestire e analizzare”.

Tuttavia nel cercare una definizione che inglobasse tutte le caratteristiche del fenomeno riportiamo la definizione proposta da De Mauro (2016) secondo la quale i Big Data sono: “un insieme di risorse informative ad alto volume, ad alta velocità, alta varietà, che richiedono forme innovative e convenienti di elaborazione delle informazioni per una visione approfondita e un processo decisionale”.

L'OCSE riconosce che una definizione di Big Data con enfasi soltanto sul volume può risultare fuorviante, sebbene il volume sia la caratteristica più evidente che identifica i dati; dunque sottolinea l'importanza di includere concetti come: “velocità” e “varietà”.

1.2 Caratteristiche dei Big Data: le 5 V

Nel 2001, l'allora vicepresidente e Service Director dell'azienda Meta Group descrisse in un report il "Modello delle 3V", relativo alle 3 V dei Big Data: Volume, Velocità, Varietà.

Con il passare degli anni e con il conseguente sviluppo della letteratura in tal senso, ci si rese conto della necessità di mettere in risalto altre caratteristiche tali da definire come questi nuovi dati debbano essere utilizzati: Veridicità e Valore.



Analizziamo nel dettaglio le 5 V dei Big Data:

Volume: il volume dei dati rappresenta sicuramente la caratteristica che più facilmente si può accostare ai big data, visto che con questo termine si fa riferimento alla dimensione del fenomeno che genera la cosiddetta datasfera.

Numerosi studi e statistiche cercano di misurare tale caratteristica e l'unità di misura più idonea appare quella dei zettabyte: 1 zettabyte corrisponde a una capacità di archiviazione pari a oltre 36.000 anni (in termini di durata) di video in HD.

Una mole di dati sempre maggiore deriverà dal consumo di video online e dalla presenza di sensori legati all'IoT (Internet of Things), l'internet delle cose, dei quali ci si aspetta una crescita esponenziale, grazie all'avvento delle connessioni 5G.

Altro fattore che contribuisce all'aumento dei volumi di dati raccolti è sicuramente la riduzione dei costi di raccolta degli stessi.

Varietà: la varietà dei dati si riferisce all'eterogeneità nelle fonti sorgenti dei dati, nei formati con cui vengono acquisite le informazioni e nella rappresentazione e analisi dei dati immagazzinati.

Nel 2008 Chris Anderson sul Magazine Wired scriveva: “More isn’t just more. More is different” facendo riferimento alle differenti tipologie di dati oggi utilizzabili.

Una ricerca condotta da NVP (New Vantage Partners) sostiene che la varietà è una delle caratteristiche più distintive di molti modelli di business utilizzati dalle piattaforme online.

Il 40% delle aziende intervistate, appartenenti alla classifica Fortune 1000, fa riferimento alla varietà come il principale motore nelle scelte di investimento per la gestione dei Big Data, rispetto al 14,5% che cita il volume e al 3,5% che indica la velocità.

Velocità: la velocità dei dati risulta connessa con le tempistiche con cui le banche dati vengono alimentate e con la necessità di processare i dati in maniera rapida, per prendere decisioni ad un ritmo sempre più veloce e spesso in tempo reale.

Non è infatti un segreto come molte opportunità di business legate allo sfruttamento dei dati, l’analisi in tempo reale è sicuramente un prerequisito fondamentale.

Valore: cioè la capacità di estrarre valore economico dai big data, sia per il lancio di nuovi prodotti e servizi che per la loro commercializzazione e pubblicità indirizzate verso una domanda già profilata.

Non si tratta solo di generare valore per le imprese, ma anche di accrescere il benessere complessivo della società.

Affinché i dati acquisiscano davvero un valore economico essi devono essere funzionali allo svolgimento di analisi economico-statistiche, ed in tal senso sono necessari strumenti di Big Data Analytics. Lo sfruttamento dei dati permette non solo di rispondere a specifiche domande di ricerca (uso primario), ma anche di sfruttarne nel tempo, il loro “valore opzionale” (uso secondario). Il riutilizzo dei dati, infatti, è alla base dei numerosi progetti che Google e altre società della rete hanno in cantiere.

Veridicità e valenza: con veridicità ci si riferisce alla fiducia che in essi si può riporre ovvero alla qualità e significatività dei dati raccolti o elaborati. Si tratta di un requisito fondamentale in quanto solo se si lavora su dati veritieri si possono ottenere analisi utili e affidabili. La valenza dei dati cresce nel tempo rendendo di fatto le connessioni fra dati sempre più dense e complesse e determinando anche in questo caso nuove sfide da affrontare. Si potrebbe aggiungere un’altra “V”, cioè la visualizzazione dei dati. Con la giusta analisi e visualizzazione, infatti, le informazioni acquisiscono valore rispetto al livello di dati grezzi.

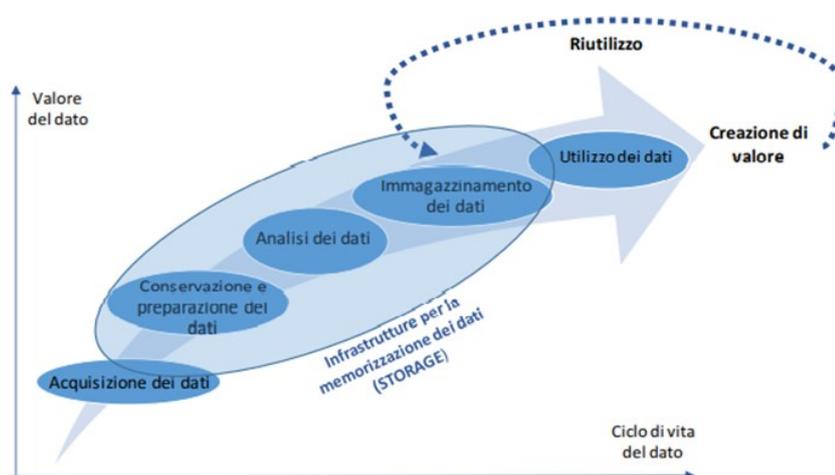
1.3 La Value Chain dei Big Data

La catena del valore è un modello che permette di descrivere la struttura di una organizzazione come un insieme limitato di processi. Questo modello è stato teorizzato da Michael Porter nel 1985 nel suo best seller “Competitive Advantage: Creating and Sustaining Superior Performance”.

Con particolare riferimento alla catena del valore in ambito Big Data la Commissione Europea, già nel 2013, ha ritenuto che la catena del valore dei dati fosse il fulcro dell’economia del futuro, con lo scopo di determinare opportunità di sviluppo digitale anche nei settori più tradizionali, quali ad esempio trasporti, servizi finanziari, sanità, manifatturiero, vendita al dettaglio.

La Value Chain dei Big Data si articola in una sequenza di attività principali e interdipendenti che ne aumentano gradualmente il valore. Come spiegato in OCSE (2015, cap.2 e 2016) così come in Mayer-Schonberger e Cukier (2013), i big data vengono raccolti, scambiati, archiviati e il valore viene estratto in un complesso ecosistema fatto di molti mercati correlati che sono spesso multiformi.

Potremmo rappresentare la Value Chain dei Big Data nel modo seguente:



Fonte: Agcom (2018)

Analizziamo più nel dettaglio le singole fasi.

1.3.1 Raccolta/Acquisizione dei dati

Un'impresa può raccogliere dati direttamente o indirettamente.

-Raccolta diretta

A sua volta la raccolta diretta dei dati può avvenire mediante due modalità: online e offline. In particolare:

- Alcune informazioni sono pubblicamente osservate tramite dispositivo, sistema operativo, indirizzo IP, ecc.
- Alcune informazioni sono fornite volontariamente dai consumatori al momento della registrazione ad un sito Web, come: nome, data di nascita, e-mail ecc.
- Alcune informazioni possono essere raccolte dal monitoraggio del consumatore in diversi modi, come:

-Tracking cookies, che sono un tipo specifico di cookie che viene distribuito, condiviso e letto attraverso due o più siti Web non correlati con lo scopo di raccogliere informazioni.

Ogni qual volta un utente visita un sito, il server fornisce all'utente un cookie che serve come forma di identificatore, qualora poi l'utente visiterà di nuovo il sito, le informazioni contenute nel cookie saranno ritrasmesse al server, che potrà quindi adattare il contenuto della pagina web alle info che aveva su quel particolare utente;

-History sniffing: annusare la storia, che è la pratica di tracciare quali siti un utente ha o non ha visitato (hackerando il suo elenco di cronologia del browser);

-Cross-device tracking: monitoraggio del dispositivo, offre la possibilità di interagire con lo stesso consumatore attraverso il suo desktop, laptop, tablet, smartphone, utilizzando informazioni sia online che offline;

attraverso l'uso di applicazioni da parte dell'utente, queste informazioni sono accessibili al proprietario del sistema operativo e allo sviluppatore dell'applicazione.

-Raccolta indiretta

Le aziende possono anche acquisire dati da terze parti, anche se questa rimane oggi una pratica marginale in Europa.

Sulla base di un'indagine sui modelli di business svolta da Deloitte, la Commissione Europea osserva che “nella stragrande maggioranza dei casi (78% delle aziende intervistate) i dati sono generati e analizzati internamente dall'azienda o da un subappaltatore. Questo è particolarmente vero nei settori con un'elevata presenza di grandi aziende tecnologicamente avanzate, come banche e fornitori di telecomunicazioni o produttori di automobili e macchinari.”

In questo caso, i dati possono essere ottenuti da broker di dati che sono stati definiti dalla US FTC come "società la cui attività principale è la raccolta di informazioni personali sui consumatori da una varietà di fonti e l'aggregazione, l'analisi e la condivisione di tali informazioni, o delle informazioni che ne derivano, per scopi quali marketing di prodotti, verifica dell'identità di un individuo o rilevamento di frodi". Dunque si osserva che i mercati dei dati sono negozi online in cui le aziende possono acquistare e vendere dati.

Il mercato secondario dei dati non è ancora ben compreso dalle autorità pubbliche, negli Stati Uniti la Federal Trade Commission ha condotto delle indagini concludendo che:

- il settore dei broker di dati negli Stati Uniti è complesso con più società che scambiano molti dati tra loro;
- i data broker dispongono di una grande quantità di dati su quasi tutte le transazioni domestiche e commerciali degli Stati Uniti;
- i broker di dati combinano dati offline e online da più fonti diverse ed è molto difficile per un consumatore conoscere e determinare come un broker di dati ha ottenuto i suoi dati;
- i broker di dati analizzano i dati e fanno inferenze sui consumatori, inserendoli in categorie che possono essere sensibili.

1.3.2 Archiviazione dati e cloud computing

La fase successiva alla raccolta riguarda l'archiviazione dei dati.

È fondamentale che la stessa venga fatta in modo da soddisfare le esigenze delle applicazioni che richiedono un accesso veloce ai dati.

L'archiviazione di enormi quantità di dati richiede grandi data center costituiti da grandi cluster di computer collegati da reti locali veloci.

Questi data center sono costosi da costruire e caratterizzati da economie di scala.

Tuttavia, lo sviluppo del cloud computing consente alle piccole imprese di affittare invece di possedere i data center, convertendo così i loro costi fissi in costi variabili (Greenstein et al., 2013).

Il cloud computing è un modello che consente un rapido accesso ai dati memorizzati, dati che sono condivisi con altri utenti e collocati in modo dinamico.

I grandi volumi di dati che tipizzano i Big Data hanno reso necessario il passaggio dai Relational Database Management Systems (RDBMS o modelli relazionali) verso altre tecnologie più moderne come NoSQL basate su modelli che non richiedono uno schema fisso.

Inoltre, la concorrenza tra i fornitori di servizi cloud può essere limitata dalla localizzazione dei dati, restrizioni che possono essere importanti per alcuni tipi di dati quali i dati sanitari, finanziari e di gioco/gioco d'azzardo, nonché per i dati di proprietà pubblica.

1.3.3 Analisi e utilizzo dei dati

Dopo la raccolta e l'archiviazione, le fasi successive della catena del valore riguardano l'analisi dei dati e le relative attività di esplorazione, trasformazione e modellazione, che permettono di portare alla luce informazioni e relazioni causali tra variabili che risultano spesso nascoste mediante l'utilizzo di modelli di correlazione.

Questo viene fatto da applicazioni ed algoritmi il cui sviluppo e miglioramento si basa su molti input come dati, forza lavoro qualificata e capitale.

Pertanto, i dati sono importanti ma probabilmente non sono l'input più importante come menzionato da Lerner (2014).

La particolarità di questi algoritmi di analisi consiste nell'apprendere e migliorare le loro performance man mano che vengono alimentati con un numero crescente di dati, dunque, è proprio la loro disponibilità alla base del progressivo miglioramento degli algoritmi impiegati e dello sviluppo di nuovi.

Tale fase si occupa di evidenziare i dati rilevanti dal punto di vista aziendale e garantire che gli stessi siano sicuri, completi e significativi.

Dunque, prevede l'organizzazione e l'elaborazione analitica dei dati grezzi al fine di dedurre informazioni sfruttabili per fini economici mediante l'utilizzo di tecniche analitiche sempre più innovative.

Lo stadio finale del dato è quello in cui da semplice informazione esso si trasforma in conoscenza e visione.

L'utilizzo dei dati nei processi decisionali può riguardare la riduzione dei costi di produzione, l'organizzazione del personale, l'invenzione di nuovi servizi e/o prodotti, così come qualsiasi apporto al miglioramento degli indicatori di performance nel privato come nel pubblico.

Le informazioni contenute nei dati possono migliorare i prodotti per tutti grazie ad una migliore comprensione delle esigenze dei consumatori, possono personalizzare al meglio i prezzi dei prodotti o le strategie di marketing, possono anche migliorare i processi, il marketing e l'organizzazione, aumentando così le efficienze produttive e dinamiche.

Pertanto, le informazioni dedotte dai big data possono aumentare il benessere dei consumatori e la crescita del PIL. L'OCSE (2015) suggerisce che le aziende di innovazione guidata dai dati beneficiano, in media, di una crescita della produttività più rapida dal 5% al 10% rispetto ad aziende simili che non utilizzano l'innovazione guidata dai dati.

Tutte queste fasi della catena del valore del dato consentono così di modellare il sistema dei big data e, di conseguenza, di identificare i vari passaggi attraverso cui generare valore, e, più in generale, conoscenza dai dati.

1.4 Le barriere all'ingresso

Le caratteristiche dei big data sopra citate, servono come base per il passaggio successivo della nostra analisi: esplorare le barriere all'ingresso nei mercati dei big data con particolare riferimento a ciascuna fase della catena del valore.

Fondamentale in tal senso, è stata l'analisi condotta da Rubinfeld e Gal (2017) che hanno individuato le barriere all'ingresso associate con le singole fasi della catena del valore.

Tradizionalmente, le barriere all'ingresso sono viste come provenienti dal lato dell'offerta. Tuttavia, possono sorgere barriere anche dal lato della domanda, come vedremo di seguito per la fase di raccolta/acquisizione dei dati.

1.4.1 Ostacoli alla raccolta/acquisizione dei dati

Identifichiamo tre tipi di barriere: tecnologiche, legali e comportamentali, che possono esistere in parallelo e rafforzarsi a vicenda.

Barriere tecnologiche:

Unicità dei dati o accesso ad essi: Si sostiene spesso che, a causa della sua natura non rivale e del fatto che le fonti di dati sono spesso abbondanti, i big data possono essere raccolti in modo semplice ed economico, ma questo è vero solo per quanto riguarda i dati pubblicamente disponibili e che sono liberamente disponibili a chiunque ad un costo basso.

Inoltre, quando i dati sono utili in forma aggregata, i dati possono spesso essere raccolti da diverse fonti, consentendo così una facile sostituzione tra fonti di dati alternative. Ad esempio, i dati utilizzati per determinare la velocità media dei conducenti su strade buie possono essere raccolti da una varietà di luoghi. Tuttavia, punti di accesso univoci a dati univoci possono portare a situazioni in cui i dati non possono essere facilmente replicati.

Infine, gateway unici per la raccolta dei dati potrebbero anche creare barriere tecnologiche all'ingresso. Ad esempio, in alcuni paesi del terzo mondo in cui i computer non sono comuni, i telefoni cellulari sono il principale dispositivo per l'utilizzo di Internet, creando così un vantaggio tecnologico per gli operatori di telefonia mobile nella raccolta dei dati.

Economia di scala, di scopo, learning by doing: Barriere tecnologiche dal lato dell'offerta si presentano se le imprese storiche hanno raggiunto sostanziali economie di scala o di scopo attraverso investimenti che sono parzialmente o totalmente affondati. La necessità di sostenere spese irrecuperabili può rendere proibitivo l'ingresso ad altre imprese.

Le economie di scale e di scopo possono derivare dai costi fissi di creazione, da dispositivi per la raccolta dei dati come antenne, crawler, strutture per interviste, cookie.

Spesso il costo della messa in atto dell'infrastruttura per la raccolta e l'analisi dei dati può generare costi fissi elevati, mentre i costi di estrazione dei dati possono essere bassi. La velocità di raccolta dei dati potrebbe creare quelle che possono essere definite "economie di velocità". Il Nowcasting consente alle aziende non solo di tenere traccia delle tendenze nella condotta degli utenti in tempo reale, ma anche delle tendenze nella condotta dei (potenziali) concorrenti e di rispondere più rapidamente.

Effetti di rete: Gli effetti di rete basati sui dati possono creare una barriera tecnologica all'ingresso sul lato della domanda.

Saranno necessarie spese sostanziali (affondate) per contrastare o addirittura superare gli effetti di rete esistenti. Ciò può accadere quando la qualità del prodotto dipende dalla qualità dei dati, che, a sua volta, è influenzata dal numero dei dati inseriti, dalla loro varietà e dalla loro freschezza.

L'ingresso di nuove imprese che non dispongono di tali dati potrebbe essere piuttosto difficile.

Barriere informative:

Gli ostacoli alla raccolta dei dati potrebbero derivare anche da informazioni limitate su chi possiede i dati rilevanti o sui costi di localizzazione e contratto con tali titolari di dati.

Barriere legali:

Leggi sulla protezione dei dati e sulla privacy, molte giurisdizioni hanno posto limitazioni alle attività di raccolta dei dati, più comunemente per quanto riguarda i dati personali, al fine di tutelare la privacy degli utenti ed altri obiettivi sociali.

Proprietà dei dati: che influiscono sulla facilità di accesso ai dati nonostante la loro natura non rivale.

Barriere comportamentali:

Esclusività: L' accesso esclusivo contrattuale ad un'unica fonte di dati può creare barriere all'ingresso sotto forma di preclusione all'ingresso o all'uscita. Ciò nonostante le preferenze dei consumatori possono talvolta essere osservate da fonti di dati alternative, dunque l'esclusività su una fonte di dati spesso potrebbe non creare una barriera significativa.

Prezzi e condizioni di accesso: prezzi di accesso e le condizioni stabilite dal titolare dei dati per la concessione dell'accesso ai propri dati.

Cosa collezionare: per un'azienda potrebbe non essere nel suo interesse raccogliere determinati tipi di dati che potrebbero dare ai suoi rivali un vantaggio comparativo.

1.4.2 Ostacoli all'archiviazione dei dati

I progressi, sia nell'hardware che nel software, hanno sostanzialmente ridotto questo problema.

I computer oggi hanno capacità di archiviazione molto migliori rispetto a prima. Ma soprattutto, una soluzione software ha in gran parte risolto il problema hardware: la creazione del cloud computing.

Barriere tecnologiche:

Limiti nell'hardware e nel software di archiviazione.

Costi di blocco e cambio: una volta che i dati vengono memorizzati in un ordine particolare, può risultare difficile il loro trasferimento.

Barriere legali:

Recentemente, l'Unione Europea ha eretto una barriera legale in merito all'archiviazione dei dati altrove.

1.4.3 Ostacoli all'analisi dei dati

Gran parte delle barriere che caratterizzano la raccolta dei dati, possono essere rilevanti anche per la sintesi e l'analisi dei dati.

Tuttavia è possibile identificare due ulteriori barriere che sono uniche di questa fase della catena del valore dei dati: compatibilità e interoperabilità dei dati.

Ogni titolare di dati organizza i propri dati in modo da soddisfare le proprie esigenze, tale organizzazione potrebbe creare una barriera per il suo utilizzo da parte di altri.

Strumenti analitici: La qualità degli algoritmi utilizzati per l'analisi dei big data può creare una barriera tecnologica.

1.4.4 Ostacoli all'utilizzo dei dati

Barriere tecnologiche:

Difficoltà di individuare e raggiungere il consumatore.

Il proprietario dei dati non può individuare facilmente i consumatori che potrebbero trarre vantaggio dal suo set di dati.

Barriere comportamentali:

Queste limitazioni di solito si applicano ai dati privati.

Tali restrizioni limitano la capacità dell'utente di esportare i dati personali da un'applicazione all'altra senza incontrare troppe difficoltà.

Tali limitazioni aumentano i costi.

Barriere legali:

Limitazioni legali volte a tutelare la privacy degli utenti o proteggere ulteriori obiettivi sociali, come leggi sulle pari opportunità che vietano la discriminazione basata su determinate caratteristiche.

1.5 Principali tecnologie

In questa sezione intendiamo analizzare le tecnologie principali utilizzate nelle varie fasi della catena del valore dei dati, entrando nel dettaglio delle stesse e fornendo una descrizione chiara e sintetica.

La fonte di riferimento sono le slides del corso di Strumenti ICT per la gestione dei servizi del professore Paolo Giaccone.

Nella fase di raccolta dei dati le tecnologie di riferimento sono: la Rete 5G e l'Internet of Things.

Il termine 5G indica l'insieme di tecnologie di telefonia mobile e cellulare, i cui standard definiscono la quinta generazione della telefonia mobile con una significativa evoluzione rispetto alla tecnologia 4G/IMT-Advanced (Wikipedia).

Il 5G è ampiamente superiore al 4G per quel che concerne la velocità di download: superano di molto i 300 Megabit al secondo e toccano, in alcuni casi, i 400 Megabit al secondo.

Le reti 4G hanno una latenza media di circa 30 ms, il 5G promette di arrivare a 2-3 ms. Il flusso di dati in tempo reale sarà di grandissima importanza anche per la mobilità e le auto a guida autonoma.

L'IoT si basa sull'idea di oggetti "intelligenti" tra loro interconnessi in modo da scambiare informazioni possedute, raccolte e/o elaborate.

L'Internet of Things è un paradigma che non conosce, potenzialmente, confini applicativi.

Nella fase di archiviazione dei dati le tecnologie di riferimento sono: il Cloud Computing e il Web Caching.

Il cloud computing è stato ampiamente descritto nei paragrafi precedenti.

La cache è un livello di storage dei dati ad alta velocità che memorizza un sottoinsieme di dati, in genere di natura temporanea, per rispondere alle richieste più rapidamente di quanto non sarebbe possibile accedendo ogni volta al percorso principale in cui si trovano i dati.

Nella fase di analisi dei dati le tecnologie di riferimento sono: il Machine Learning e i cookies.

Il Machine Learning (che include il deep learning) è la capacità di un computer di imparare, di modificare i propri processi sulla base delle nuove informazioni acquisite (Antonio Nicita e Marco Delmastro).

I cookies sono piccoli file di testo scaricati e salvati nel browser di un utente quando visita un sito web, per identificare e ricordare cose che lo riguardano, come ad esempio se è stato effettuato un login o se ha visitato determinate pagine.

I cookies sono classificati come cookie di prima parte, il cui dominio appartiene al fornitore di servizi online, quindi al sito stesso al quale si accede e cookies di terza parte, il cui dominio appartiene ad un'entità separata e differente dal fornitore dei servizi.

I cookies aiutano a creare il profilo dell'utente in modo da rendere i messaggi più mirati.

1.6 La classificazione dei dati

Esistono molti tipi diversi di dati, tuttavia per i nostri scopi, è opportuno distinguere tra quelli che vengono comunemente definiti dati "personali" e dati "non personali". Questo è importante perché, ai sensi della Direttiva generale sulla protezione dei dati (GDPR), i singoli utenti hanno determinati diritti sui dati personali che non si applicano ai dati non personali.

Dati personali sono definiti come «qualsiasi informazione relativa a una persona fisica identificata o identificabile» (Regolamento 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016).

È quindi comune suddividere tali dati personali in almeno quattro categorie:

- Dati forniti, che sono dati su un utente identificabile che sono stati forniti alla piattaforma dall'utente stesso, spesso nel momento in cui si registra per la prima volta. Tali dati possono avere un valore considerevole per le piattaforme digitali. È replicabile nel senso che se un utente può fornire dati ad una piattaforma digitale, allora potrebbe, in linea di principio, fornire gli stessi dati anche ad altre piattaforme. Se gli utenti hanno fornito molti dati per un lungo periodo o se utilizzano la piattaforma digitale stessa come mezzo per archiviare i dati, potrebbe essere più difficile replicare i dati forniti. Di conseguenza, il GDPR fornisce agli utenti il diritto di scaricare i dati che hanno fornito in precedenza alla piattaforma senza incorrere in costi.
- Dati osservati, ovvero i dati generati dalle interazioni tra un utente identificabile e la piattaforma. Osservando le risposte e le azioni dei singoli utenti, le piattaforme digitali sono spesso in grado di generare previsioni su come i singoli utenti interagiranno in futuro o ciò che potrebbe meglio soddisfare i loro bisogni.

Questi dati forniranno ad un motore di ricerca informazioni sia sulle probabili preferenze e necessità di quel singolo utente, ma anche preziose deduzioni, se combinati con i dati osservati da altri utenti.

Il GDPR ritiene che gli utenti debbano mantenere diritti simili sui dati osservati rispetto ai dati forniti, incluso il diritto di scaricarli gratuitamente.

- Dati dedotti, ovvero dati derivati dall'analisi di altri dati effettuata dalla piattaforma stessa.

La capacità di una piattaforma digitale di generare inferenze approfondite, può essere un'importante fonte di differenziazione e vantaggio competitivo nei mercati in cui tali

intuizioni possono essere tradotte in servizi differenziati.

- Dati acquisiti da terze parti, possono essere acquisiti per completare la profilazione degli utenti che hanno omesso di fornire alcuni dati personali, o per altre finalità.

È probabile che tali dati siano disponibili per la vendita da una varietà di fornitori di dati specializzati e facilmente ottenibili anche da piattaforme rivali.

Le stesse categorie possono essere applicate ai dati non personali, trattandosi di informazioni che non riguardano un individuo identificabile.

Pertanto, i dati forniti non personali possono essere informazioni originariamente fornite da un utente identificabile ma che sono state successivamente riconfigurate o rese anonime, in modo che l'identità dell'utente che le ha fornite non sia più evidente. I dati non personali sono informazioni che non sono protette dalla legge sulla protezione dei dati.

Si può distinguere tra dati anonimi e dati anonimizzati.

Dati anonimi, sono informazioni che non riguardano una persona fisica identificata o identificabile.

Un esempio di dati anonimi sono i dati macchina che si riferiscono e sono creati dall'attività di computer, telefoni cellulari e altri sistemi o dispositivi, ad esempio nel contesto dell'Internet of Things.

Dati anonimizzati, sono dati personali resi anonimi in modo tale che l'interessato non sia più identificabile. In altre parole, i dati anonimi sono informazioni che prima facevano riferimento ad una persona identificabile, ma dove tale identificazione non è più possibile.

Si evince che nel caso in cui i dati non possano essere attribuiti ad un individuo, è più probabile che questi siano in grado di fornire vantaggi economici senza incorrere in problemi in materia di privacy.

Tuttavia ci si pone delle domande sulla effettiva capacità di salvaguardare l'anonimato dei dati. Le norme dell'UE in materia di protezione dei consumatori si applicano alle relazioni B2C. Tali norme si applicano a beni o servizi scambiati dietro compenso. Tuttavia, in un ambiente online in cui sempre più servizi vengono pagati con dati e non con denaro, una questione cruciale è cosa costituisce una remunerazione e un pagamento.

Il diritto dell'UE si sta evolvendo su questo punto.

Una ulteriore classificazione dei dati in base alla loro struttura distingue i dati in dati strutturati, non strutturati e semi-strutturati.

- Dati strutturati, sono dati che sono stati predefiniti e formattati in una struttura impostata prima di essere inseriti nell'archivio dati.

Il miglior esempio di dati strutturati è il database relazionale.

Esempi comuni di dati strutturati generati dalla macchina sono le statistiche del weblog e i dati dei punti vendita, come i codici a barre e la quantità.

- Dati non strutturati, sono dati archiviati nel loro formato nativo e non elaborati fino a quando non vengono utilizzati.

È disponibile in una miriade di formati di file, tra cui e-mail, post sui social media, presentazioni, chat, dati dei sensori IoT e immagini satellitari.

- Dati semi-strutturati, si riferiscono a quelli che normalmente sarebbero considerati dati non strutturati, ma che hanno anche metadati che identificano determinate caratteristiche.

I metadati contengono informazioni sufficienti per consentire ai dati di essere catalogati, cercati e analizzati in modo più efficiente rispetto ai dati strettamente non strutturati. Sono il tramite tra dati strutturati e non strutturati.

Sebbene i dati strutturati siano quelli che contengono una densità di informazioni maggiore, la tendenza in atto è quella per la quale circa l'80% dei dati oggi disponibili ha natura non strutturata.

1.7 La condivisione dei dati

La condivisione dei dati è intrapresa per un'ampia gamma di scopi ed è stata attuata in base ad un'ampia gamma di accordi diversi, molti dei quali sono interamente volontari. A volte i beneficiari della condivisione dei dati sono individui particolari, a volte sono organizzazioni come aziende o organizzazioni pubbliche.

Diversi possono essere gli incentivi degli utenti alla condivisione:

- un singolo utente può richiedere l'accesso ai dati che sono stati accumulati dalla piattaforma su di lui per capire meglio di cosa si tratta.

L'OCSE si riferisce a questa motivazione come "autodeterminazione informativa".

- un singolo utente può richiedere l'accesso ai dati in suo possesso perché desidera condividerli con un altro fornitore di servizi.

Potrebbe essere il caso di un utente che desidera passare da un fornitore all'altro.

- i singoli utenti possono richiedere l'accesso ai dati che li riguardano perché la condivisione di tali dati può consentire loro di beneficiare dell' 'innovazione complementare' in altri mercati.

In tal caso, l'utente non cerca l'accesso ai dati per sostituire un servizio con un altro di tipo analogo ma per ottenere un nuovo servizio.

- I singoli utenti possono richiedere l'accesso ai propri dati nella speranza di poterli monetizzare e generare un reddito.

In realtà le motivazioni degli utenti per avviare un trasferimento di dati sono state più tosto contrastanti, alcuni sottolineano che gli utenti traggono vantaggio dal miglioramento della sicurezza e della protezione dei loro dati, altri dalla comodità che deriva dall'evitare di dover reinserire i dati su più piattaforme, altri sottolineano l'opportunità per gli utenti di monetizzare i propri dati "vendendoli" alla piattaforma.

Le prove mostrano che alcuni utenti hanno percepito benefici tangibili dalla possibilità di condividere i dati, tuttavia, il download di dati dalle piattaforme digitali da parte dei singoli utenti sembra essere limitato ad una minoranza e la monetizzazione dei dati personali da parte dei singoli rimane più un'aspirazione che una realtà.

Diversi possono essere gli incentivi delle imprese alla condivisione:

- le organizzazioni commerciali si scambiano dati quando vi è un reciproco vantaggio commerciale nel farlo.
- le organizzazioni condividono dati senza flussi di pagamento tra di loro, come quando i dati vengono condivisi per scopi di riferimento creditizio o per rilevare frodi.
- le organizzazioni commerciali possono fornire un accesso unilaterale ai dati per promuovere l'innovazione.

Questa strategia è stata ampiamente perseguita dalle piattaforme digitali come Facebook, Twitter, Google o Apple per facilitare l'innovazione che aumenta il valore della piattaforma stessa.

- le organizzazioni commerciali possono fornire l'accesso ai dati in cambio di "servizi ausiliari" oppure dietro pagamento.

I broker di dati acquisiscono e aggregano enormi volumi di dati da un'ampia gamma di fonti, e vendono set di dati ad altre società per scopi di marketing ed altri.

Le organizzazioni commerciali possono impegnarsi in "filantropia dei dati", qualora forniscano l'accesso ai dati a specifici enti di beneficenza o pubblici, per consentire loro di perseguire obiettivi sociali.

L'OCSE stima che il 65% di tutte le iniziative di condivisione dei dati siano state intraprese da organizzazioni del settore pubblico.

Gli esempi includono la condivisione di dati geospaziali, dati sanitari ed altri dati scientifici.

1.8 Introduzione alle piattaforme digitali

La Commissione europea ha affermato: “Le piattaforme online sono fattori chiave per il commercio digitale. Attualmente, più di un milione di imprese dell'UE commerciano attraverso piattaforme online per raggiungere i propri clienti e si stima che circa il 60% del consumo privato e il 30% del consumo pubblico di beni e servizi relativi all'economia digitale totale siano effettuati tramite intermediari online.”

Le piattaforme digitali connettono aziende, utenti e dati in un ecosistema, generando un modello di business in cui anche startup e PMI possono trovare il loro spazio per operare a livello globale. Si tratta di infrastrutture digitali in grado di connettere tra loro sistemi diversi ed esporli agli utenti attraverso interfacce semplificate ed integrate.

Le principali tipologie di piattaforme individuate da Evans e Gawer (2016) sono:

- Piattaforme di transazione: è una tecnologia, un prodotto o un servizio che funge da canale (o intermediario) facilitando lo scambio o le transazioni tra diversi utenti, acquirenti o fornitori.
- Piattaforme di innovazione: è una tecnologia, un prodotto o un servizio che funge da base su cui altre aziende (organizzate liberamente in un ecosistema innovativo) sviluppano tecnologie, prodotti o servizi complementari;
- Piattaforme integrate: è una tecnologia, un prodotto o un servizio che è sia una piattaforma di transazione che una piattaforma di innovazione.

Aziende come Apple ha entrambe, piattaforme corrispondenti come l'App Store e un grande ecosistema di sviluppatori di terze parti che supporta la creazione di contenuti sulla piattaforma.

Abbiamo anche piattaforme di investimento che sono costituite da società che hanno sviluppato una strategia di portafoglio di piattaforme e agiscono come una holding, un investitore attivo di piattaforme o entrambi.

Le piattaforme integrate sono quelle che contribuiscono maggiormente al valore di mercato globale. Soprattutto a partire dall'ultimo decennio le piattaforme online costituiscono un contributo fondamentale per le economie, per una serie di motivi.

In primis, perché rendono possibile un matching altamente efficiente (basti pensare ai mercati dell'e-commerce come eBay o Amazon) e forniscono un supporto ad un utilizzo più efficiente delle risorse (basti pensare ad Uber) con un notevole aumento in termini di produttività.

In secundis, sono una importante fonte di innovazione. Basti pensare che nel 2014 nove

piattaforme statunitensi hanno ottenuto 11.585 brevetti e non è un caso che molte piattaforme create da piccole start-up altamente innovative abbiano attratto investimenti significativi da parte di business angels e fondi di venture capital internazionali.

Evans e Gawer (2016) concentrandosi su piattaforme sia quotate in borsa che private e con una valutazione di mercato di almeno un miliardo di dollari, hanno stimato che la loro capitalizzazione di mercato globale superava i 4,3 trilioni di dollari.

Hanno inoltre notato che nonostante sia l'Asia ad avere il maggior numero di piattaforme sono però le piattaforme nord americane ad avere la maggiore capitalizzazione complessiva per un totale di 3,1 trilioni di dollari.

Altro elemento rilevante è che l'Europa è emersa come uno dei principali consumatori di servizi derivanti da piattaforme, ma ne ha generate relativamente poche, segno degli scarsi investimenti in tale settore.

1.9 Implicazioni in termini di privacy

“Se questa è l'era dell'informazione, allora la privacy è il problema dei nostri tempi”
(Alessandro Acquisti, Laura Brandimarte, George Loewenstein).

Comunichiamo tramite e-mail, messaggi e social media; impariamo tramite corsi online; cerchiamo risposte utilizzando i motori di ricerca; leggiamo notizie e libri nel cloud.

Attraverso queste ed altre attività, riveliamo informazioni, sia consapevolmente che inconsapevolmente, l'un l'altro, alle entità commerciali e ai nostri governi.

Il monitoraggio delle informazioni personali è onnipresente.

Sia le aziende che gli individui possono trarre vantaggio dalla condivisione di dati un tempo nascosti e dall'applicazione di analisi sempre più sofisticate a database più grandi e interconnessi.

Dove va posto il limite e, soprattutto, chi lo definisce e lo gestisce? Chi tutela il cittadino dalla conoscenza che di lui ha il mondo e che potrebbe essere utilizzata contro il suo interesse? Chi controlla i dati che riguardano la nostra sicurezza sociale, infrastrutturale, militare?

L'erosione della privacy può minacciare la nostra autonomia, non solo come consumatori ma come cittadini. Infatti condividere dati personali non si traduce necessariamente in più progresso, efficienza o uguaglianza.

La strada più promettente da percorrere è quella di uno sforzo serio e costruttivo, a livello di regolazione sovranazionale, che sappia rispondere a queste sfide e a queste paure.

Sia a livello europeo che a livello nazionale, sono state avviate iniziative volte ad indagare lo spazio possibile della regolazione nel campo dell'intelligenza artificiale, le opportunità economiche e i principali dilemmi etici.

I passi più importanti li ha fatti la vecchia Europa nel campo della tutela della privacy, imponendo uno standard al mondo: il GDPR (General Data Protection Regulation). Con esso viene rafforzata la responsabilizzazione dei titolari e dei responsabili del trattamento dei dati personali.

Approfondiremo ampiamente queste tematiche nel capitolo successivo.

Capitolo 2

Privacy e Regolamentazione

2.1 Generalità e caratteristiche

Nel cercare di fornire una definizione al concetto di privacy che risulti essere la più chiara possibile, iniziamo con l'individuare le caratteristiche principali mediante una rassegna della letteratura esistente su tale tematica.

Sebbene l'accesso alla privacy sia certamente influenzato da fattori socioeconomici (alcuni hanno definito la privacy un "bene di lusso"), e le norme sulla privacy differiscono notevolmente tra le culture, il bisogno di privacy sembra essere un tratto umano universale.

Dare una definizione di "privacy" non è semplice. È stata descritta come la protezione dello spazio personale di qualcuno e il suo diritto di essere lasciato solo (Warren e Brandeis, 1890); come controllo e salvaguardia delle informazioni personali (Westin, 1967); e come aspetto della dignità, dell'autonomia e della libertà umana (Schoeman, 1992). Tutte queste definizioni riguardano i confini tra il sé e gli altri, tra privato e condiviso, o pubblico (Altman, 1975).

Dunque, cercando di generalizzare, il termine privacy può essere inteso come: la protezione dello spazio personale dell'individuo, il controllo e la salvaguardia delle informazioni personali, un aspetto di dignità, autonomia e libertà dell'individuo.

Gli studiosi hanno scoperto prove di comportamenti di ricerca della privacy tra popoli e culture separati dal tempo e dallo spazio: dall'antica Roma e dalla Grecia alle società giavanesi, balinesi e tuareg preindustriali.

Spunti di una comune ricerca umana sulla privacy si trovano anche nei testi delle antiche religioni: Il Corano insegna a non spiarsi a vicenda; il Talmud consiglia ai costruttori di case di posizionare le finestre in modo che non siano direttamente di fronte a quelle dei propri vicini; la Bibbia (Genesi) racconta come Adamo ed Eva scoprirono la loro nudità dopo aver mangiato il frutto della conoscenza e si coprirono di vergogna dagli occhi indiscreti di Dio. Implicita, in questa eterogenea selezione di esempi storici, è l'osservazione che esistono molteplici nozioni di privacy.

Ogni volta che “clicchiamo” qualcosa sul web o effettuiamo una scelta su un dispositivo cediamo alla rete inconsapevolmente un pezzo di noi: un’opinione, un’emozione, un desiderio o un sogno.

La rete dunque colleziona i singoli frammenti e grazie ad opportuni algoritmi riesce a comporre una fotografia dell’utente in base alla quale vengono selezionati gli annunci da far visualizzare.

Questo massiccio utilizzo dei dati solleva importanti problemi in termini di privacy, in quanto gli utenti non hanno un modo semplice per impedire la raccolta dei propri dati e non hanno alcuna garanzia che gli stessi non vengano poi utilizzati in modo dannoso contro di loro.

In passato, la raccolta di dati personali era così costosa e difficile che solo le persone che godevano di qualche forma di celebrità erano vulnerabili all'intrusione della privacy. Tuttavia i recenti progressi tecnologici hanno reso la raccolta dei dati così scalabile che i dati di chiunque possono adesso essere raccolti e usati per scopi commerciali in quanto i costi di raccolta e archiviazione dei dati sono scesi notevolmente.

Particolarmente interessante è l’iniziativa portata avanti da Google che sta progettando per poi costruire un’intera città gestita in modo intelligente, Quayside.

Si tratta della prima città intelligente e completamente automatizzata, gestita dagli algoritmi che regoleranno il traffico, consentiranno di risparmiare energia, ridurranno l’inquinamento e avvertiranno anche quando è ora di svuotare i bidoni della spazzatura. Progetto che mai come ora, sta diventando realtà, vista l’acquisizione da parte di Sidewalk Labs (una consociata di Google) di un’area industriale dismessa di 800 acri sulla riva del lago Ontario a Toronto in Canada. Per la prima volta un progetto di innovazione smart che non si ferma ad un singolo edificio ma si estende ad un intero centro urbano. Quayside si prefigge lo scopo di diventare un modello per tutti gli insediamenti futuri canadesi e non, sfruttando al massimo le reti dati ultra veloci e tutte le nuove tecnologie messe a disposizione dal colosso californiano. Controlli in remoto di bus, auto, illuminazioni stradali, manutenzioni, luci dei semafori, controllo e gestione rifiuti, ma anche robot e automi che si spostano in tunnel sotterranei per consegnare e trasferire beni, insomma tutto quello che fino ad oggi è stato visto solo nei film di Steven Spielberg, Ridley Scott o George Lucas sembra che stia per diventare realtà.

Addirittura l'azienda di Mountain View ha annunciato un articolato progetto per il controllo dell'andamento meteorologico grazie a costruzioni e artifici in grado fermare la pioggia o controllare l'ombreggiamento nelle giornate estive.

Sembra per tanto evidente come questo nuovo ecosistema digitale imporrà un notevole uso di dati personali: gli spostamenti dei cittadini dovranno essere costantemente registrati, preferenze, usi e abitudini degli abitanti saranno incessantemente scambiati dai vari dispositivi.

Queste nuove dinamiche non fanno altro che aumentare la produzione di migliaia di informazioni mettendo in bella vista la nostra privacy.

Conseguenza è la possibilità da parte delle piattaforme di compiere degli abusi nei confronti degli utenti.

Si evince pertanto la necessità di importanti interventi normativi volti alla tutela dei diritti sulla privacy.

Le norme e i comportamenti relativi alla sfera pubblica e privata differiscono notevolmente da una cultura all'altra.

Gli americani, ad esempio, hanno fama di essere più aperti sulle questioni sessuali rispetto ai cinesi, mentre questi ultimi sono più aperti sulle questioni finanziarie.

Le preoccupazioni sulla privacy possono variare notevolmente per lo stesso individuo e per le società, nel tempo.

L'incertezza e la dipendenza dal contesto implicano che non si può sempre contare sulle persone per navigare i complessi compromessi che coinvolgono la privacy in modo egoistico.

Le persone spesso sono inconsapevoli delle informazioni che stanno condividendo.

La malleabilità, a sua volta, implica che le persone siano facilmente influenzabili su cosa e quanto rivelano. Inoltre, ciò che condividono può essere utilizzato per influenzare le loro emozioni, pensieri e comportamenti in molti aspetti della loro vita. Sebbene tale influenza non sia sempre o necessariamente malevola o pericolosa, la rinuncia al controllo sui propri dati personali e sulla propria privacy altera l'equilibrio di potere tra coloro che detengono i dati e coloro che ne sono i soggetti.

Gli approfondimenti della ricerca empirica sociale comportamentale sulla privacy suggeriscono che è improbabile che gli approcci politici forniscano una protezione adeguata contro i rischi posti dalle recenti tecnologie dell'informazione.

Consideriamo la trasparenza e il controllo, due principi concepiti come condizioni necessarie per la tutela della privacy.

Una ricerca mostra che possono fornire protezioni insufficienti e persino ritorcersi contro, se usati separatamente da altri principi di protezione della privacy.

In assenza di specifiche indicazioni normative, le aziende in grado di raccogliere e organizzare i dati esercitano un controllo più o meno assoluto ed esclusivo su di essi.

Il 20 giugno 2019 il Parlamento e il Consiglio Europeo hanno adottato il Regolamento 2019/1150 recante “Misure per promuovere l’equità e la trasparenza per gli utenti di servizi di intermediazione e dei motori di ricerca online” che è entrato in vigore il 12 luglio 2020.

Questo regolamento non obbliga direttamente la piattaforma a condividere i dati con gli utenti commerciali, ma impone trasparenza sulle pratiche di condivisione con tutte le terze parti. Tuttavia rimane intatta la posizione privilegiata della piattaforma sul mercato perché continua ad avere accesso indisturbato ai dati dei suoi partner commerciali, con conseguente asimmetria informativa a suo vantaggio.

Il dibattito su questi temi è ancora in pieno svolgimento, a febbraio 2020 la Commissione Europea ha indicato il 2030 come data limite entro il quale si mira a creare uno spazio unico europeo dei dati, uno spazio mondiale in cui i dati personali e non siano sicuri, di alta qualità e facilmente accessibili.

Un obiettivo della politica pubblica dovrebbe essere quello di raggiungere una maggiore equità di potere tra individui, consumatori e cittadini da un lato e, dall'altro, i titolari di dati come governi e società che attualmente hanno il sopravvento.

Per essere efficace, la politica sulla privacy dovrebbe proteggere le persone reali, che sono ingenui, incerti e vulnerabili, e dovrebbe essere sufficientemente flessibile per evolversi con le imprevedibili complessità emergenti dell'era dell'informazione.

2.1.1 Relazione tra target advertising e privacy

La pubblicità mirata è una forma di pubblicità online che si concentra su tratti, interessi e preferenze specifiche di un consumatore.

Gli inserzionisti scoprono queste informazioni monitorando l'attività su Internet.

La pubblicità mirata è un modo per i professionisti del marketing di presentare ai consumatori annunci che riflettono i loro tratti, interessi e comportamenti di acquisto specifici.

Ciò viene generalmente eseguito utilizzando i dati dei clienti per segmentare il pubblico in base a fattori quali dati demografici di base, interessi di acquisto o comportamenti di navigazione, quindi creando annunci pubblicitari unici e su misura per ciascun segmento di pubblico.

Sembra per tanto evidente come i dati degli utenti abilitano le piattaforme a creare profili molto dettagliati dei consumatori.

Tra i principali vantaggi della pubblicità mirata ritroviamo:

- Offrire un livello di personalizzazione più elevato;
- Stabilire la consapevolezza del marchio e aumentarne la percezione;
- Semplificare gli sforzi di marketing e mantenere le risorse focalizzate sulla crescita;
- Aumentare il ROI del marketing.

Internet e la rivoluzione delle tecnologie di comunicazione digitale hanno notevolmente aumentato la capacità delle imprese di indirizzare la pubblicità e assicurarsi che venga mostrata solo a determinati gruppi preselezionati di persone. Ciò significa, in teoria, che la pubblicità può essere più informativa per i consumatori di quanto non fosse prima.

Aspetto fondamentale del target advertising è che la pubblicità deve essere indirizzata solo a persone potenzialmente interessate a fare un acquisto. Ad esempio, un annuncio che spiega la tecnologia alla base di un pannolino potrebbe essere molto informativo per una neomamma.

Dunque l'idea generale è che se le imprese sono identiche e i consumatori sono poco informati, dal punto di vista del consumatore c'è utilità ad essere informati dell'esistenza e dei prezzi di un prodotto (Butters, 1977).

Quindi a prima vista sembrerebbe aumentare la potenziale utilità di un consumatore (Anand e Shachar). Tuttavia potrebbe non essere a costo zero infatti gli inserzionisti utilizzano i consumatori per costruire profili dettagliati acquisendo le loro impronte

click-by-click mentre navigano nel web in modo tale da rendere la pubblicità più accurata che mai.

I consumatori possono diffidare dall'essere seguiti troppo da vicino dalle aziende le quali utilizzano queste informazioni per personalizzare i prezzi (Acquisti e Varian, 2005).

Nell'esempio dei pannolini è improbabile che ci sia un effetto diretto sul prezzo pagato dalle madri a causa dell'esposizione all'annuncio, poiché i pannolini sono in gran parte acquistati offline.

Tuttavia il danno sembra provenire direttamente dalla disutilità che un cliente potrebbe sentire dall'invasione percepita della pubblicità.

Ad esempio, Turow et al. (2009) rilevano che l'86% dei giovani adulti afferma di non volere pubblicità su misura se essa è il risultato di seguire il loro comportamento su siti Web diversi da quello che stanno visitando.

In tal senso la letteratura di psicologia sociale descrive come i consumatori rispondono quando percepiscono qualcosa come eccessivamente invadente.

In questo caso la 'reattanza' li porta a resistere all'appello dell'annuncio (White et al., 2008).

La reattanza è quel meccanismo di difesa, una vera e propria resistenza psicologica intrinseca nell'uomo, nel non voler eseguire ordini che provengono da altri.

Pertanto, una caratteristica insolita dei mercati pubblicitari online è che sono caratterizzati da una tensione tra il desiderio di un'azienda di fornire informazioni al giusto gruppo di consumatori e l'apparente disgusto dei consumatori per come le aziende utilizzano i dati per cercare di migliorare l'informatività della pubblicità.

Data questa evidenza che i consumatori sembrano rispondere negativamente alla pubblicità eccessivamente invadente, ci si chiede come le aziende possono affrontare direttamente questo problema.

Un'opzione è semplicemente smettere di usare tecniche di pubblicità intrusiva. Tuttavia, un'opzione ovviamente preferibile dal punto di vista dell'azienda, è se ci fosse un modo mediante il quale potrebbero ottenere gli stessi benefici delle tecniche di targeting altamente sofisticate senza i costi in termini di disgusto dei consumatori. Pertanto, un modo potenziale e pratico per risolvere il compromesso tra informatività ed invasione è dare ai consumatori un controllo esplicito su come vengono utilizzate le loro informazioni nella speranza di ridurre la disutilità che deriva dall'invasione.

Secondo l'Australian Competition & Consumer Commission (2019) spesso il

consumatore è del tutto inconsapevole di quanti dati sono stati collezionati, conseguenza del fatto che le policy di privacy sono spesso ambigue o non permettono al consumatore un intervento adeguato.

Le piattaforme chiedono il consenso per i dati nella forma del tipo “prendere o lasciare” oppure sottopongono l’utente ad un bundle di consensi, per cui aderendo ad una richiesta, in automatico si esprime una preferenza anche per le richieste collegate. Dunque, il consumatore non possiede l’abilità di operare liberamente se intende usufruire del servizio e non è in grado di agire in modo informato.

Il governo del Regno Unito (HR Government, 2019), identifica il target advertising come responsabile di danni morali e psicologici ai consumatori quali manipolazioni messe in atto mediante propaganda e false informazioni.

Molti utenti online e gruppi di difesa sono preoccupati per i problemi di privacy relativi a questo tipo di targeting poiché la pubblicità mirata richiede l'aggregazione di grandi quantità di dati personali, inclusi quelli altamente sensibili che vengono poi scambiati tra centinaia di parti nel processo di offerte in tempo reale.

Questa è una controversia che l'industria del targeting comportamentale sta cercando di contenere attraverso l'istruzione, la difesa e i vincoli del prodotto al fine di mantenere tutte le informazioni non identificabili personalmente o per ottenere il permesso dagli utenti finali.

Gli accademici canadesi presso l'Università di Ottawa Canadian Internet Policy e Public Interest Clinic hanno recentemente chiesto al commissario federale per la privacy di indagare sulla profilazione online degli utenti di Internet per la pubblicità mirata. Nell'ottobre 2009 è stato riferito che un recente sondaggio condotto dall'Università della Pennsylvania e dal Berkeley Center for Law and Technology ha rilevato che la grande maggioranza degli utenti Internet statunitensi rifiutava l'uso della pubblicità comportamentale. Diversi sforzi di ricerca da parte di accademici ed altri, hanno dimostrato che i dati presumibilmente anonimi possono essere utilizzati per identificare individui reali. Tucker (2011) suggerisce che non è sempre vero che il compromesso tra protezione della privacy ed efficacia pubblicitaria debba essere negativo. Dando ai consumatori un controllo esplicito su come vengono utilizzati i loro dati e quindi potenzialmente aumentare le informazioni sulla raccolta dei dati, le aziende possono essere in grado di alleviare parte del compromesso tra quanto informativa può essere la loro pubblicità e quanto i consumatori lo trovano invadente.

2.1.2 Il paradosso della privacy

Nell'analizzare il tema del "paradosso della privacy" partiamo da una domanda provocatoria: quanti di noi sono realmente consapevoli delle informazioni personali che seminiamo in rete?

Il paradosso della privacy indica proprio la tendenza, tutta recente, al compromesso tra informazione e sicurezza. Siamo preoccupati della mancanza di privacy, soprattutto in rete, ma continuiamo a comportarci come se così non fosse. Spesso gli utenti dichiarano una forte preferenza per la privacy, ma sono disposti a rinunciare alle proprie informazioni personali per piccoli incentivi.

Westin in un sondaggio ha utilizzato delle domande non eccessivamente specifiche sulla privacy individuando i seguenti segmenti: fondamentalisti della privacy, pragmatici e indifferenti. Molte persone rientrano nel primo segmento: professano di preoccuparsi molto della privacy ed esprimono particolare preoccupazione per la perdita del controllo delle proprie informazioni personali o per l'accesso non autorizzato da parte di altri ad esse. È proprio la discrepanza tra atteggiamenti e comportamenti ad essere nota come il "paradosso della privacy". (Alessandro Acquisti, Laura Brandimarte, George Loewenstein).

L'evidenza empirica di una discrepanza tra le preoccupazioni che gli utenti mostrano per la loro privacy e il loro comportamento online mostra come questi molto spesso non siano abbastanza cauti nel divulgare informazioni private. Ad esempio, uno studio ha confrontato le risposte ad un sondaggio in merito al comportamento effettivo dei social media. Anche all'interno del sottogruppo di partecipanti che hanno espresso il più alto grado di preoccupazione per gli estranei che possono scoprire facilmente il loro orientamento sessuale, le opinioni politiche e i nomi dei partner, il 48% ha effettivamente rivelato pubblicamente il proprio orientamento sessuale online, il 47% ha rivelato il proprio orientamento politico e il 21% ha rivelato il nome del proprio partner attuale.

Dunque il processo decisionale sulla privacy è anche in parte il risultato di un razionale "calcolo" di costi e benefici; è anche influenzato da percezioni errate di tali costi e benefici, nonché da norme sociali ed emozioni.

Ognuno di questi fattori può influenzare il comportamento in modo diverso da come influenza gli atteggiamenti.

Indagini "esplicite" chiedono alle persone di fare compromessi diretti, in genere tra

privacy dei dati e denaro.

Ad esempio, in uno studio condotto sia a Singapore che negli Stati Uniti, gli studenti hanno fatto una serie di scelte ipotetiche sulla condivisione delle informazioni con siti Web che differivano per protezione delle informazioni personali e prezzi per l'accesso ai servizi.

Gli autori hanno concluso che i soggetti valutavano la protezione contro errori, accesso improprio e uso secondario delle informazioni personali tra \$ 30,49 e \$ 44,62. Acquisti, Brandimarte e Loewenstein hanno condotto un interessante esperimento online nel quale agli utenti venivano fatte delle domande abbastanza intrusive sul comportamento personale in determinate situazioni.

Ciò che gli autori hanno concluso è stato che nonostante il comportamento dei soggetti vari a seconda della cultura di origine, la scelta tra cosa condividere e cosa non condividere dipende fondamentalmente da tre elementi già citati in precedenza: dipendenza dal contesto, malleabilità e incertezza.

In tal senso si è soliti parlare di “miopia del consumatore” poiché condivide i propri dati in vista di un beneficio immediato senza riflettere sulle possibili implicazioni di lungo periodo sulla privacy.

Sembra per tanto evidente che le persone raramente hanno una chiara conoscenza di quali informazioni altre persone, aziende e governi hanno su di loro e come le stesse possano essere utilizzate e con quali conseguenze.

2.1.3 Il “calcolo del valore della privacy”

Una delle spiegazioni più consolidate per il paradosso si basa sulla Teoria del Calcolo della Privacy, secondo la quale gli individui eseguono un “calcolo” tra i potenziali costi - in termini di perdita della privacy - e i potenziali benefici - come sconti e servizi personalizzati - della divulgazione dei propri dati personali.

Il loro comportamento finale è determinato dall’esito di tale trade-off. Pertanto, si presume che gli individui decidano di divulgare le loro informazioni personali quando i potenziali benefici superano le perdite attese.

Una modello teorico, questo, fondato sull’assunto di un agente economico neoclassicamente razionale – il cosiddetto “Homo Economicus” – con preferenze stabili e sempre in grado di calcolare rischi e benefici, scegliendo l’alternativa che massimizza la propria utilità e che minimizza i costi (Posner., 1978; Stigler., 1980). Tuttavia, la ricerca in Economia Comportamentale ha dimostrato che il processo decisionale umano, lungi dal costituire il risultato di un calcolo perfettamente razionale, è in realtà influenzato da diversi fattori.

In primo luogo, l’incompletezza delle informazioni e le asimmetrie informative complicano il processo decisionale in materia di privacy. In secondo luogo, pur disponendo di tutte le informazioni necessarie, il soggetto in questione potrebbe non essere in grado di elaborare grandi quantità di dati per formulare una decisione informata, a causa della sua innata “razionalità limitata” (Herbert A. Simon, 1982).

Infine, quand’anche gli individui avessero informazioni complete e la capacità cognitiva illimitata di elaborarle, sarebbero soggetti ad alcune distorsioni psicologiche sistematiche, i cosiddetti bias cognitivi.

La letteratura propone numerosi studi che cercano di dare un valore monetario propriamente detto alla cessione di informazioni che possono limitare la privacy degli utenti.

Il Financial Times ha cercato di stabilire il valore di alcuni dati personali mediante l’impegno di un calcolatore online.

L’algoritmo alla base parte dall’ipotizzare che età, genere, livello di istruzione ed etnia degli utenti siano inizialmente disponibili.

In seguito vengono poste una serie di domande riguardante il particolare status dell’utente in questione quali status familiare, beni posseduti, hobby.

L'idea è che in base alle risposte fornite dall'utente l'algoritmo sia poi in grado di stimare il valore commerciale di ogni singolo dato fornito ed eventualmente incrementare o decrementare il valore dei dati precedentemente ceduti. Ad esempio, l'algoritmo stima che dichiarare di avere uno specifico hobby possa avere un valore di circa 0,08 dollari o dichiarare di possedere una barca un valore di circa 0,1 dollari.

Lo studio condotto da Chiou e Tucker rivela che per garantire la protezione della privacy online l'aspetto fondamentale sia il tempo in cui i dati vengono conservati. Ciò nonostante, da altri studi emerge come la riduzione dello storage di dati limita ma non contrasta del tutto la possibilità che le piattaforme restino in possesso di una quantità di dati sufficiente per poter attuare condotte abusive.

Molto interessante è anche lo studio condotto da Berasford sulla disponibilità a pagare per la privacy.

Dall'esperimento si evince una generale riluttanza degli utenti a pagare per la salvaguardia della privacy infatti soltanto la metà acquista DVD presso il rivenditore meno invasivo.

Uno studio condotto da Jentzsch, Preibusch e Harasser nel 2012 per conto della ENISA (Agenzia dell'Unione Europea per la cybersicurezza) analizza la monetizzazione della privacy. Con monetizzazione della privacy ci si riferisce alla decisione del consumatore di divulgare o non divulgare i dati personali in relazione ad una transazione di acquisto. Viene introdotto un modello teorico che tiene conto della concorrenza tra due fornitori di servizi.

I consumatori possono selezionare il fornitore di servizi di loro scelta, a seconda delle loro preoccupazioni sulla privacy e delle offerte fatte dai fornitori di servizi. In una variante del modello, i consumatori possono selezionare un fornitore di servizi, ma possono anche scegliere se desiderano che i propri servizi siano personalizzati in futuro. L'esperimento di laboratorio mostra che la maggior parte dei consumatori acquista da un fornitore più invasivo della privacy se il fornitore di servizi addebita un prezzo inferiore. Una percentuale non trascurabile dei partecipanti all'esperimento (13-83%), tuttavia, ha scelto di pagare un "premio" per la privacy.

Lo hanno fatto per evitare la divulgazione di ulteriori dati personali o perché il fornitore di servizi rispettoso della privacy ha promesso di non utilizzare i loro dati per scopi di marketing.

Viene poi effettuato un esperimento ibrido sul campo con oltre 2.300 partecipanti. L'esperimento sul campo ha confermato le tendenze osservate in laboratorio; l'unica

differenza rilevata è che in caso di nessuna differenza di prezzo i fornitori di servizi rispettosi della privacy che richiedono meno dati personali ottengono una quota di mercato maggiore. Le risposte fornite dai partecipanti al questionario hanno mostrato una preoccupazione piuttosto elevata per la privacy.

Analizzando nello specifico i risultati ottenuti:

- Quasi tutti i partecipanti all'esperimento di laboratorio (oltre il 90%) sono rimasti con il fornitore di servizi che avevano selezionato per primo in caso di due acquisti;
- Se il prezzo è lo stesso presso i due fornitori, la maggior parte degli acquisti in laboratorio viene effettuata presso il fornitore di servizi online rispettoso della privacy (circa l'83% di tutti i biglietti venduti);
- Nei casi in cui anche il prezzo differisca, la quota di mercato del fornitore di servizi rispettosi della privacy scende, inferiore o prossima a un terzo.

Confrontando i trattamenti in laboratorio e sul campo per tutti gli acquisti, si nota che il fornitore di servizi rispettoso della privacy ha una quota di mercato molto più ampia, se le differenze della raccolta dei dati è ovvia e i prezzi sono gli stessi. Tuttavia, una volta che i prezzi cambiano e un concorrente ostile alla privacy applica un prezzo inferiore, il fornitore di servizi rispettoso della privacy perde quote di mercato. Tuttavia, circa un terzo degli acquisti dei consumatori mostra di essere disposto a pagare un ricarico presso il fornitore di servizi rispettoso della privacy.

Dallo studio gli autori traggono le seguenti raccomandazioni:

- Se ci sono differenze minime o nulle nei prezzi offerti dai fornitori di servizi su beni omogenei, un concorrente che ha un fabbisogno di dati ridotto può ottenere un vantaggio competitivo purché questo tipo di differenziazione sia evidente per il consumatore;
- Il quadro normativo dovrebbe consentire una flessibilità sufficiente affinché i fornitori di servizi online possano offrire menu diversi per quanto riguarda i prezzi e i requisiti in materia di dati personali;
- Le differenze nei requisiti dei dati, nella protezione dei dati e nelle politiche sulla privacy devono essere rese più visibili ai consumatori. Più questi termini sono standardizzati e semplici, più facile sarà il confronto;
- Le autorità per la protezione dei dati dovrebbero applicare un quadro giuridico chiaro e coerente per la protezione dei dati.

2.2 Il GDPR

La tematica sulla protezione della privacy e in modo particolare sulla protezione dei dati personali è diventata soprattutto negli ultimi anni un tema di accesi dibattiti.

La presenza di poche piattaforme capaci di gestire enormi quantità di dati degli utenti rende palese il paradosso di internet, da un lato un processo di enorme produzione informativa ma dall'altro il concentrarsi di dati anche sensibili nelle mani dei big tech. Sembra per tanto evidente l'urgenza di regole che disciplinino la riservatezza dei dati degli utenti.

Il Regolamento generale per la protezione dei dati personali 2016/679 (General Data Protection Regulation o GDPR) è la principale normativa europea in materia di protezione dei dati personali.

Adottato il 27 aprile 2016, pubblicato sulla Gazzetta ufficiale dell'Unione Europea il 4 maggio 2016, entrato in vigore il 24 maggio dello stesso anno ed operativo a partire dal 25 maggio 2018.

Il presente regolamento stabilisce norme relative alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché norme relative alla libera circolazione di tali dati (art. 1 GDPR).

Dall'articolo 1 si evince come l'intero regolamento sia rivolto ai dati personali, definiti come: qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale (art. 4 GDPR).

Goldfarb e Tucker nel 2011 hanno osservato come l'applicazione di questa normativa abbia ridotto l'efficacia della pubblicità.

È noto come la pubblicità rappresenta una delle entrate principali delle aziende online, dunque, l'introduzione di una regolamentazione sulla privacy comporta cambiamenti economici.

Lo studio tuttavia sottolinea come è pur sempre possibile modificare la strategia pubblicitaria al fine di ridurre le perdite economiche dovute dall'imposizione delle nuove normative.

2.2.1 I soggetti del GDPR

Il GDPR distingue più soggetti in funzione del ruolo che svolgono nel trattamento dei dati. In particolare ai sensi dell'art. 4 troviamo:

- Titolare del trattamento: è la persona fisica o giuridica che decide finalità e mezzi del trattamento. Se ci sono più titolari, si parla di contitolarità.

In questo caso deve essere redatto un regolamento interno per suddividere le responsabilità fra contitolari, in particolare rispetto al trattamento dei dati degli interessati.

Con riferimento al titolare del trattamento, assume particolare rilevanza il principio di accountability stabilito dal regolamento europeo.

Con questo principio si stabilisce la piena responsabilità del titolare nell'organizzazione del proprio sistema di gestione della privacy.

- Responsabile del trattamento: è la persona fisica o giuridica che tratta i dati per conto del titolare.

La nomina a responsabile deve essere fatta tramite contratto o altro atto giuridico, che definisce quali dati vengono trattati, per quanto tempo e con quale finalità.

Attraverso l'atto sorgono gli obblighi in capo al responsabile del trattamento sia per quanto riguarda i dati dell'interessato sia rispetto al titolare dei dati, nei cui confronti il responsabile ha un obbligo di assistenza.

Il fatto che la nomina venga fatta tramite un atto non significa che, se un professionista tratta i dati degli interessati, non abbia responsabilità.

È la realtà dei fatti che determina la responsabilità.

- Autorizzato al trattamento: è colui che ha accesso ai dati e può trattarli solo se ha ricevuto specifiche istruzioni da parte del titolare o del responsabile del trattamento.

- Responsabile della protezione dati: è una persona altamente qualificata in ambito giuridico. I suoi compiti sono di assistenza e supporto alle attività del titolare e del responsabile del trattamento.

Può essere nominato Responsabile del Trattamento un soggetto interno o esterno all'azienda, l'importante è che ne sia garantita l'indipendenza e l'autonomia di spesa.

Il DPO è anche il referente per le autorità di controllo.

- Interessato al trattamento: è la persona fisica i cui dati vengono trattati.

Il GDPR definisce i suoi diritti:

1. Essere informato su quali dati vengono trattati dal titolare;
2. Avere l'accesso ai dati, richiesta a cui il titolare o il responsabile del trattamento deve rispondere in maniera celere;
3. Ottenere la rettifica o l'aggiornamento dei dati qualora siano sbagliati;
4. Chiedere e ottenere la cancellazione dei dati in possesso del titolare;
5. Revocare il consenso in qualsiasi momento;
6. Limitare l'elaborazione dei dati trattati;
7. Il diritto alla portabilità dei dati: l'interessato può richiedere al titolare del trattamento i dati in suo possesso in un formato che consenta il trasferimento ad altro titolare;
8. Diritti relativi al processo decisionale automatizzato e alla profilazione.

2.2.2 Principi applicabili al trattamento di dati personali

L'introduzione del GDPR ha comportato importanti cambiamenti nel modo in cui le aziende raccolgono, archiviano ed elaborano i dati dei consumatori, giustificando il periodo di tempo in cui conservano le informazioni.

Il tutto nell'ottica di una maggiore trasparenza. Elenchiamo di seguito quelli che sono i principi fondamentali su cui si basa il GDPR ai sensi dell'art. 5 e degli articoli chiarificatori ad esso collegati:

- **Liceità e correttezza:** Il GDPR elenca le basi giuridiche che determinano la liceità del trattamento. Queste possono essere suddivise in 3 categorie:

1. Consenso dell'interessato che deve essere esplicito, informato e specifico per le finalità del trattamento;

2. I casi di necessità espressamente previsti dal regolamento europeo;

3. Interesse legittimo del titolare del trattamento. L'interesse va valutato caso per caso;

- **Trasparenza:** Oltre ad essere un diritto dell'interessato, devono essere trasparenti le finalità del trattamento, accessibili i dati del titolare e tutte le info relative al trattamento stesso;

- **Limitazione dello scopo:** I dati devono essere trattati solo per lo scopo legittimo per cui sono stati raccolti;

- **Minimizzazione dei dati:** Le organizzazioni non dovrebbero raccogliere più informazioni personali di quelle di cui hanno bisogno;

- **Esattezza:** Devono essere trattati solo i dati esatti e aggiornati. Eventuali inesattezze devono essere corrette tempestivamente o cancellate;

- **Limitazione della conservazione:** I dati possono essere conservati solo per il tempo utile all'adempimento delle finalità per cui sono stati raccolti, dopodiché devono essere cancellati;

- **Integrità e riservatezza:** Il Titolare del trattamento deve utilizzare tecnologie in grado di garantire la protezione dei dati contro l'uso non autorizzato o illecito.

I dati personali devono essere protetti da "trattamenti non autorizzati o illeciti", nonché da perdita, distruzione o danni accidentali.

Il GDPR privacy non ha fatto altro che mettere per iscritto le pratiche migliori adottate nell'ultimo ventennio in materia di riservatezza del trattamento dei dati.

Il GDPR non impone l'adozione di specifiche misure di sicurezza, perché saranno diverse in funzione del tipo di organizzazione. Ad esempio, una banca dovrà proteggere

le informazioni in maniera più accurata di quanto dovrà fare il negozio di ferramenta che abbiamo sotto casa.

- Eccezioni: Il diritto degli interessati al controllo e alla gestione dei propri dati soccombe di fronte agli obiettivi di interesse pubblico generale dell'UE o degli Stati membri.

Si fa riferimento a ragioni di sicurezza nazionale, difesa, prevenzione o perseguimento di reati. Ma anche interessi economici e finanziari rilevanti.

Altri aspetti fondamentali introdotti dal GDPR sono:

- Applicazione territoriale: Regolato dall'art. 3 del GDPR;

il GDPR si applica:

1. Al trattamento di dati personali effettuati da un titolare del trattamento o responsabile del trattamento che risiede nell'Unione, indipendentemente dal fatto che il trattamento stesso sia effettuato o meno nell'UE;

2. Al trattamento di dati personali di persone fisiche (interessati) che si trovano nell'Unione, effettuato da un titolare o responsabile del trattamento al di fuori dell'Unione;

- Consenso dell'interessato: Regolato dall'art. 4 del GDPR, è qualsiasi manifestazione di volontà libera, specifica, informata e inequivocabile dell'interessato, con la quale lo stesso esprime il proprio assenso, mediante dichiarazione o azione positiva inequivocabile, al trattamento dei dati personali che lo riguardano.

Dunque, il consenso rappresenta sicuramente un mezzo molto semplice che l'utente può utilizzare per proteggere la sua privacy, infatti negandolo può impedire ad un sito di raccogliere informazioni sul suo conto;

- Pseudonimizzazione del dato: Regolato dall'art. 4 del GDPR, si tratta di una tecnica che consiste nel conservare i dati in una forma che impedisce l'identificazione del soggetto senza l'utilizzo di informazioni aggiuntive, il tutto volto a limitare il rischio di violazioni di privacy e furti d'identità;

- Responsabilità congiunta: Regolato dall'art. 13 del GDPR, i titolari del trattamento non possono esimersi dalla responsabilità per le violazioni della privacy che si verificano mentre collaborano con il fornitore di tecnologia che riveste il ruolo di responsabile del trattamento.

Il GDPR stabilisce che il titolare del trattamento è tenuto ad informare i propri utenti in merito a qualsiasi trasferimento di dati personali a fornitori terzi;

- Riduzione al minimo dell'asimmetria informativa: Regolato dall'art. 13 del GDPR,

prevede che l'utente debba essere informato sui dati collezionati dalle imprese, nonché sulle loro finalità e sui suoi diritti;

- Diritto alla cancellazione del dato o diritto all'oblio: Regolato dall'art. 17 del GDPR, è un diritto dell'interessato per cui la persona fisica – l'interessato - può ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano.

Il Regolamento prevede che la cancellazione avvenga senza ingiustificato ritardo e impone al titolare del trattamento l'obbligo di cancellare i dati personali se, per esempio, non sono più necessari rispetto alle finalità per le quali sono stati raccolti o trattati, se l'interessato ha revocato il consenso al trattamento o se i dati sono stati trattati illecitamente.

- Portabilità dei dati: Regolato dall'art. 20 del GDPR, consente a chiunque sappia che i suoi dati sono oggetto di trattamenti automatizzati compiuti dal titolare del trattamento, o col suo consenso o per contratto, di chiedere che i dati da lui forniti siano trasmessi, senza impedimenti, o a sé stesso o ad altro titolare da lui indicato, utilizzando un formato “strutturato, di uso comune e leggibile da dispositivo automatico”;

- Il principio di accountability: riguarda il sapere rispondere e rendere conto dei risultati ottenuti o di quanto sia stato fatto in merito al trattamento dei dati personali. La responsabilità implica la possibilità di verificare in che modo ci si è comportati per essere in compliance al regolamento.

L'Unione Europea ha intenzione di creare un unico spazio europeo per i dati, all'interno del quale è possibile accedere ad un infinito numero di dati personali e non di alta qualità.

Il tutto volto a promuovere crescita e valore nel rispetto della privacy degli utenti auspicando di diventare un modello da seguire (European Commission, 2020).

2.2.3 Sanzioni previste

Si può essere sanzionati:

- per il mancato adeguamento al regolamento anche se i dati non sono stati violati;
- per la mancata nomina del DPO;
- per violazione della sicurezza.

Tuttavia, la normativa europea è anche di difficile interpretazione, tant'è vero che per avere un quadro della disciplina esaustivo bisogna prendere in considerazione anche il d.lgs. 101/2018 e il regolamento privacy del 2003 nelle parti in cui non è stato abrogato. Il GDPR prevede sanzioni amministrative per le imprese che non si adeguano al regolamento. Il regolamento suddivide le sanzioni in due gruppi in base alla gravità dell'inadempimento.

- Per le sanzioni meno gravi l'importo della sanzione può arrivare a 10 milioni di Euro o fino al 2% del fatturato con riferimento all'anno precedente.
- Per le inadempienze più gravi si arriva fino a 20 milioni di Euro o 4% del fatturato, considerato sempre sull'anno precedente. Se oggetto della sanzione è una multinazionale si fa riferimento al fatturato del gruppo.

Oltre alla disciplina uniforme europea, il regolamento dà la facoltà agli Stati membri di prevedere ulteriori sanzioni.

Per quanto riguarda l'Italia, il codice Privacy ha ampliato le ipotesi di illeciti amministrativi, estendendoli notevolmente.

Il regolamento sulla privacy non prevede direttamente le sanzioni penali ma lascia liberi gli Stati membri di stabilire sanzioni penali per violazione del GDPR.

In Italia le autorità che si occupano di privacy hanno individuato delle linee guida e delle raccomandazioni da seguire in termini di gestione della privacy, in particolare ritengono fondamentale la cooperazione tra le autorità che potrebbe portare ad un coordinamento di carattere permanente.

Inoltre si evince la necessità di rafforzare il potere istruttorio ancor prima che venga aperto il procedimento e di un generico inasprimento delle sanzioni in caso di violazioni ("Big Data. Indagine conoscitiva congiunta. Luglio 2019").

Il Garante della Privacy è l'autorità preposta all'erogazione delle sanzioni per violazione del GDPR.

Fra le sue funzioni ci sono quelle di esame dei reclami e di irrogazione delle sanzioni disciplinari.

Il Garante ha inoltre il potere di sospendere, limitare e vietare il trattamento dei dati effettuato in violazione delle norme.

Con il GDPR la sua funzione di controllo si svolge soprattutto ex post, visto che le valutazioni preliminari di conformità alla normativa europea vengono fatte direttamente dal titolare in virtù del principio di accountability precedentemente definito.

2.2.4 Impatto del GDPR

La protezione della privacy in Europa è stata relativamente forte negli ultimi decenni a causa di fattori storici, culturali e ragioni sia politiche che giuridiche.

L'Unione Europea ha portato la protezione della privacy ad un nuovo livello con l'adozione del GDPR, a tal punto da far assomigliare la legge europea sulla privacy alla legge antitrust europea, dove la Commissione europea ha emesso diverse multe multimiliardarie nell'ultimo decennio.

Sempre più spesso le regole per il mondo digitale non sono adottate dai tradizionali meccanismi di coordinamento internazionale come le Nazioni Unite, ma sono avanzate da singoli paesi portando così ad una competizione tra paesi per diventare uno dei principali responsabili delle regole a livello mondiale. Esempi recenti includono l'Unione Europea con il GDPR e la California con il Consumer Privacy Act (2020). Ancora non è chiaro come progettare un'arena internazionale per la regolamentazione del mondo digitale.

L'intento di questo paragrafo consiste nell'analizzare gli effetti del GDPR con particolare riferimento allo studio condotto da Guy Aridor, Yeon-Koo Che, Tobias Salz nel 2020 per poi proseguire con una rassegna della letteratura su tale tematica.

Nello studio condotto da Aridor si analizzano gli effetti del GDPR sulla capacità delle aziende di raccogliere dati sui consumatori, identificare i consumatori nel tempo, accumulare entrate tramite la pubblicità online e prevederne il comportamento.

Si studia empiricamente gli effetti del GDPR, in particolare, la sua richiesta che i consumatori possano esprimere un consenso informato, specifico ed inequivocabile al trattamento dei loro dati.

L'opzione di consenso fornisce un mezzo semplice ma efficace per proteggere la privacy: negando il consenso, un consumatore può impedire ad un sito Web di raccogliere dati personali e condividerli con terze parti.

Si utilizziamo i dati forniti da un intermediario anonimo che opera in più di 40 paesi e stipula contratti con molte delle più grandi agenzie di viaggio online e motori di meta-ricerca di viaggi.

Il principale metodo di monitoraggio dei consumatori su cui ci concentriamo in questo studio sono i cookie web.

I cookie sono piccoli file di testo che vengono inseriti nei computer o nei telefoni cellulari degli utenti.

Ogni inserimento di un cookie fornisce ai siti web, in linea di principio, un identificatore persistente.

Tuttavia, i consumatori attenti alla privacy possono utilizzare vari mezzi per la tutela della loro privacy come la cancellazione manuale dei cookie, la modalità di "navigazione privata", o i cookie blocker tuttavia i dati che vengono generati sul sito web vengono comunque inviati ed archiviati.

Il GDPR offre ai consumatori un altro modo per proteggere la propria privacy poiché bloccano l'invio di tutte le informazioni non essenziali al sito Web di terzi: "Il requisito del consenso del GDPR va oltre e consente ai consumatori di negare "qualunque dato da inviare al sito".

Abbiamo ottenuto l'accesso ad un nuovo e completo set di dati da un intermediario anonimo che registra la totalità delle query di ricerca e degli acquisti dei consumatori nelle principali agenzie di viaggio online (OTA), negli Stati Uniti e in Europa dal 1° gennaio 2018 al 31 luglio 2018.

Osserviamo le ricerche dei consumatori, la pubblicità online, la previsione del comportamento dei consumatori da parte dell'intermediario e gli acquisti dei consumatori.

Il regolamento prevede che i siti web che effettuano transazioni con consumatori dell'UE fossero tenuti a chiedere ai consumatori il consenso esplicito per utilizzare i propri dati attraverso una procedura di opt-in, mentre coloro che trattavano dati di consumatori non UE non erano obbligati a farlo.

Viene effettuata un'analisi di regressione mediante il metodo difference in difference. Il metodo combina control group design e pre-post design, stimando l'effetto dell'intervento attraverso il confronto dell'evoluzione nel tempo di un gruppo esposto e un gruppo non esposto.

Nella sua formulazione di base, il metodo prevede l'osservazione del livello medio di Y nei due gruppi prima e dopo l'intervento.

All'interno del metodo lo shock esogeno riguarda l'introduzione in tutti gli Stati dell'Unione Europea del GDPR, il gruppo di trattamento siti di viaggio e piattaforme dei più importanti paesi europei mentre il gruppo di controllo siti di viaggio e piattaforme negli Stati Uniti, Canada e Russia poiché non influenzati direttamente dal GDPR.

La nostra specifica di regressione primaria è la seguente:

$$y_{tcjobp} = \alpha_t + \delta_{jc} + \kappa_c + \xi_j + \gamma_o + \zeta_b + \omega_p + \beta(EU_j \times after) + \epsilon_{tcjobp}$$

t = settimana dell'anno;

c = Stato;

j = sito web;

o = sistema operativo;

b = browser web;

p = tipo di prodotto (volo o hotel)

Il coefficiente che cattura l'effetto dell'introduzione del GDPR è associato alla variabile dummy che denota quando un sito web è soggetto alla regolazione, inoltre, è associato alla variabile che denota la settimana in cui si compie l'analisi dopo l'introduzione del GDPR.

Gli altri elementi presenti nella regressione corrispondono alle variabili che catturano gli effetti fissi dovuti al tempo, alle caratteristiche specifiche dei diversi paesi, dei siti web e dei siti web associati ai paesi, ai diversi sistemi operativi e browser web.

In una prima analisi misuriamo l'impatto dell'opt-out del GDPR (negare l'accesso alla raccolta dei cookies) sul numero totale di cookie e ricerche osservate dall'intermediario. Quindi esaminiamo se ci sono stati cambiamenti nella composizione dei restanti consumatori che hanno aderito.

L'outcome della regressione è il seguente:

Table 1: Difference-in-Differences Estimates for Cookies and Searches

	(1) log(Unique Cookies)	(2) Unique Cookies	(3) log(Recorded Searches)	(4) Recorded Searches
DiD Coefficient	-0.125** (-2.43)	-1378.1* (-1.71)	-0.107* (-1.87)	-9618.3** (-2.24)
Product Type Controls	✓	✓	✓	✓
OS + Browser Controls	✓	✓	✓	✓
Week FE	✓	✓	✓	✓
Website × Country FE	✓	✓	✓	✓
Observations	63840	63840	63840	63840

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: t-statistics are reported in parentheses. The standard errors for every regression are clustered at the website-country level. We aggregate every dependent variable to the website-country-product type-week level between weeks 16 and 29, including both weeks 16 and 29 (April 13th - July 20th). The dependent variables in the regression reported in the first and second column are the log and overall level of the number of unique cookies observed. The dependent variables in the regression reported in the third and fourth column are the log and overall level of the number of total recorded searches.

Le stime mostrano che, in aggregato, il GDPR ha ridotto il numero totale di cookie unici di circa il 12,5%. È importante notare che questo risultato non implica che il 12,5% dei consumatori ha utilizzato le funzionalità di opt-out. Ciò è dovuto al fatto che l'unità di osservazione è un cookie, piuttosto che un consumatore. Tuttavia, i risultati indicano un utilizzo relativamente ampio delle funzionalità di opt-out da parte dei consumatori. Un'altra misura della risposta del consumatore è il numero totale di ricerche registrate dall'intermediario. Eseguiamo nuovamente la stessa specifica con le ricerche registrate come variabile dipendente e riportiamo i risultati nelle colonne (3) e (4) di Tabella 1. Troviamo che c'è un calo del 10,7% nelle ricerche registrate complessive che è qualitativamente coerente con la dimensione dell'effetto della specifica utilizzando il numero di cookie unici.

In una seconda analisi studiamo l'effetto del GDPR sul mercato della pubblicità online. In particolare, esaminiamo la misura in cui l'opt-out del consumatore e l'aumento della tracciabilità media hanno influito sul valore medio dei consumatori per l'inserzionista e sulle entrate complessive sia per gli inserzionisti che per i siti web.

L'outcome della regressione è il seguente:

Table 3: Difference-in-Differences Estimates for Advertising Outcome Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	asinh(Total Clicks)	Total Clicks	asinh(Distinct Clicks)	Distinct Clicks	asinh(Revenue)	Revenue	Average Transfer	Average Bid
DiD Coefficient	-0.135** (-2.32)	-251.9* (-1.91)	-0.133** (-2.33)	-214.9* (-1.84)	-0.168 (-1.54)	-32972.3 (-0.75)	28.97** (2.12)	15.41*** (2.90)
OS + Browser Controls	✓	✓	✓	✓	✓	✓	✓	✓
Product Category Controls	✓	✓	✓	✓	✓	✓	✓	✓
Website × Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	62328	62328	62328	62328	62328	62328	62328	62328

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: t-statistics are reported in parentheses. The standard errors for every regression are clustered at the website-country level. We aggregate every dependent variable to the website-country-product type-week level between weeks 16 and 29, including both weeks 16 and 29 (April 13th - July 20th). The dependent variable in the regression reported in the first column is the total number of clicks associated with each observation and the second column is the inverse hyperbolic sine transform of this value. Likewise, the dependent variables in the third and fourth columns are the total number and inverse hyperbolic sine transform of the total number of unique cookies who interacted with advertisements. The dependent variables in the fifth and sixth column are the total number and inverse hyperbolic sine transform of the total revenue. The dependent variable in the seventh column is the average transfer between the intermediary and advertisers and in the eighth column it is the average bid by advertisers. Since some of the outcome variables can take zero values, in order to preserve these observations we utilize a common transformation in the applied microeconomics literature and use the inverse hyperbolic sine transform instead of the natural logarithm of the outcome variables (Bellemare and Wichman, 2019). The resulting transformed outcome variable, \bar{y} , is given by $\bar{y} = \text{arcsinh}(y) = \ln(y + \sqrt{y^2 + 1})$

Esaminiamo quindi l'effetto sui ricavi complessivi guadagnati da siti Web e inserzionisti. I ricavi dipendono sia dal numero di clic che dal prezzo per clic.

Le colonne (1) - (2) di Tabella 3 mostrano che c'è una diminuzione statisticamente significativa del 13,5% nel numero totale di clic, con una dimensione dell'effetto commisurata al calo dei cookie e delle ricerche totali.

Le colonne (3) - (4) mostrano che anche il numero di clic distinti diminuisce significativamente. Le colonne (5) e (6) di Tabella 3 forniscono la stima dell'impatto sui ricavi, negativo anche se non statisticamente significativo.

È importante sottolineare che le colonne (7) e (8) di Tabella 3 mostrano che le offerte e i trasferimenti medi dagli inserzionisti all'intermediario aumento.

Lo interpretiamo come un aumento del valore medio percepito, che gli inserzionisti attribuiscono ai consumatori rimanenti dopo il GDPR.

In una terza analisi studiamo le performance delle predizioni fatte con l'introduzione del GDPR, quindi se lo shock esogeno ha influenzato la capacità dell'intermediario di prevedere il comportamento dei consumatori.

L'outcome della regressione è il seguente:

Table 4: Difference-in-Differences Estimates for Prediction Outcome Variables

	(1) Class Proportion	(2) Average Predicted Probability	(3) MSE	(4) AUC	(5) Purchaser MSE	(6) Non-Purchaser MSE
DiD Coefficient	0.00915* (1.77)	0.00129 (0.17)	0.0130*** (3.74)	0.0124 (1.12)	-0.00579 (-0.43)	-0.00126 (-0.45)
Product Type Controls	✓	✓	✓	✓	✓	✓
OS + Browser Controls	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓
Website × Country FE	✓	✓	✓	✓	✓	✓
Observations	15470	15470	15470	15470	14298	15470

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: t-statistics are reported in parentheses. The standard errors for every regression are clustered at the website-country level. We aggregate every dependent variable to the website-country-product type-week level between weeks 16 and 29, including both weeks 16 and 29 (April 13th - July 20th). The dependent variable in the regression reported in the first column is the proportion of purchasers associated with each observation and the second column is the average predicted probability. The dependent variables in the third and fourth column are the MSE and AUC, respectively. Finally, in the fifth and sixth columns the dependent variables are the MSE conditional on the true class of the observation.

Innanzitutto, la colonna (1) mostra che il GDPR determina un piccolo ma significativo aumento della percentuale di acquirenti. Le colonne (3) e (4) mostrano l'impatto del GDPR sulle prestazioni di previsione dell'intermediario misurate rispettivamente in MSE e AUC. La colonna (3) mostra un aumento significativo di MSE dopo l'introduzione del GDPR.

Tuttavia, piuttosto che indicare il peggioramento delle prestazioni di previsione, è probabile che questo sia un artefatto del cambiamento nella proporzione della classe e la mancanza di adeguamento da parte del classificatore. Infatti, le colonne (5) e (6) mostrano che MSE condizionato alla vera classe non è aumentato; semmai sono diminuiti seppur statisticamente in modo insignificante.

Data la distribuzione distorta, un aumento della quota di acquirenti farà aumentare il MSE. Infatti, la colonna (4) mostra una stima positiva per l'effetto del trattamento sull'AUC che indica un miglioramento marginale della previsione, sebbene non sia statisticamente significativo.

Il miglioramento marginale dell'AUC indica che, nonostante la minore accuratezza delle probabilità stimate, è aumentata la capacità dell'intermediario di separare le due classi. Questa osservazione è coerente con quanto ci aspetteremmo dalla suddetta ipotesi di sostituzione di mezzi di privacy.

Nel complesso, i nostri risultati suggeriscono che il GDPR non ha avuto un impatto negativo sulla capacità di prevedere il comportamento dei consumatori e, semmai, il segno dell'effetto del trattamento suggerisce il contrario.

Concludiamo dicendo che le aziende operanti nei mercati dell'economia digitale, competono sempre più con le grandi aziende tecnologiche come Google.

Di conseguenza, mentre i nostri risultati evidenziano che maggiori requisiti di consenso potrebbero non essere del tutto negativi per le aziende, allora un'importante direzione per il lavoro futuro è capire fino a che punto tale regolamentazione metta le aziende in questi mercati in una posizione di svantaggio rispetto a quelle più grandi.

Analizzando la letteratura prodotta su tale tematica, al meglio delle nostre conoscenze Goldberg, Johnson e Shriver (2019) è il primo documento che esamina l'impatto economico del GDPR sui siti web europei. Allo stesso modo adottano un approccio di differenza nelle differenze per studiare l'effetto del GDPR sul traffico e sul volume degli acquisti dei siti Web europei. Tuttavia, la portata dei nostri dati ci consente di collegare le conseguenze delle decisioni di opt-out degli individui con l'identità del consumatore, i ricavi pubblicitari e l'efficacia delle tecnologie di previsione. Questo ci consente di andare oltre la misurazione diretta dell'impatto del GDPR e comprendere ulteriormente le esternalità associate alle decisioni sulla privacy individuali e il loro impatto economico indiretto.

Diversi altri documenti hanno studiato l'impatto del GDPR in altri domini. Jia, Jin e Wagman (2018) mostrano che il GDPR ha avuto effetti negativi sugli investimenti in capitale di rischio. Zhuo et al. (2019) studia l'impatto del GDPR sui contratti di connessione a Internet. Johnson e Shriver (2019) studia i cambiamenti nella concentrazione di mercato dei fornitori di tecnologia web a seguito del GDPR. Degeling et al. (2018) mostrano che un numero significativo di siti web ha risposto al GDPR aggiornando le proprie informative sulla privacy e aggiungendo il consenso ai

cookie, come richiesto dalla politica.

Utz et al. (2019) ha mostrato che c'era eterogeneità nell'attuazione del GDPR tra i siti web che ha portato a differenze significative nel fatto che i consumatori abbiano acconsentito alla raccolta dei dati.

Autorevole è anche lo studio condotto da Michail Batikas, Stefan Bechtold, Tobias Kretschmer e Christian Peukert nel 2020 in cui si nota una generica diminuzione nell'utilizzo di tecnologie web di terze parti da parte dei siti web nel breve periodo, a seguito dell'introduzione del GDPR.

Nel lungo periodo non si osserva un cambiamento nella tendenza all'aumento.

Altro effetto è una maggiore concentrazione dei mercati delle tecnologie web.

Questo secondo effetto può essere facilmente spiegato se si considera il fatto che il regolamento imponga come condizione per il trattamento dei dati personali il consenso dell'utente, che può essere più facilmente ottenibile da imprese grandi e note rispetto ad imprese piccole e poco conosciute.

Altri documenti studiano, invece, l'efficacia delle precedenti normative sulla privacy dei dati sulla pubblicità online. La prima è la Direttiva ePrivacy dell'UE del 2009, nota anche come Cookie Law, una precedente normativa europea volta ad aumentare la trasparenza e il controllo dei dati dei consumatori.

Goldfarb e Tucker (2011) utilizzano una metodologia basata su sondaggi per studiare l'efficacia della pubblicità online nell'UE dopo questa legge e scoprono che l'efficacia della pubblicità è diminuita. Il secondo è il programma Ad-Choices autoregolato che ha consentito ai consumatori di rinunciare alla pubblicità comportamentale online.

Gli individui che rinunciano a questo programma continuano a vedere annunci pubblicitari, ma non sono più presi di mira in base alle loro storie personali.

Johnson, Shriver e Du esaminano questa politica e trovano che i consumatori che si disattivano generano il 52% in meno di entrate rispetto ai consumatori che non lo fanno, ma che solo meno di un quarto di punto percentuale delle impressioni pubblicitarie proviene da consumatori che non lo fanno. Goldfarb e Tucker (2012) sostengono che le normative sulla privacy possono ostacolare l'innovazione delle banche dati in una serie di settori, compresa la pubblicità mirata.

Infine, Johnson (2013) stima un modello strutturale delle aste pubblicitarie ed esplora l'effetto di diverse politiche ipotetiche di consenso dei dati di opt-in e opt-out attraverso calcoli controfattuali. Trova che le entrate pubblicitarie diminuirebbero del 34,6% con una politica di opt-in e del 3,9% con una politica di opt-out.

2.3 Il Garante Privacy belga sanziona IAB Europe

Il Garante della privacy belga si è espresso circa la liceità dello standard TCF (Transparency & Consent Framework) adottato da IAB Europe e utilizzato da migliaia di inserzionisti online: la pronuncia definitiva del caso DOS-2019-01377 è giunta in data 2 febbraio, in territorio belga, da parte della Autorité de protection des données – Gegevensbeschermingsautoriteit.

Lo IAB Transparency & Consent Framework (TCF) è un framework tecnico a standard aperto che consente a siti Web, inserzionisti e agenzie pubblicitarie di ottenere, registrare e aggiornare il consenso dei consumatori per le pagine Web.

Il tutto a conferma dei molti aspetti che si lamentavano – da anni – non in regola con il GDPR e la Direttiva ePrivacy. Con effetti potenzialmente rivoluzionari per l'intero settore della promozione online, se pensiamo ad es. che tra i partecipanti al TCF si annoverano Google e Amazon.

Non è tanto l'importo della sanzione di 250.000 euro a dare il polso di questa decisione, lo è invece ciò che è stato accertato e dichiarato dall'autorità e cosa potrà avvenire di conseguenza, pensando agli operatori che si basano sull'attuale utilizzo di questo strumento.

Il framework dovrà trovare un nuovo equilibrio con la normativa europea, secondo le prescrizioni dell'autorità e così potenzialmente dover limitare, arrestare o comunque mutare sensibilmente molte delle sue prassi.

Con prevedibili effetti commerciali, di perdita dell'attuale efficacia di certe strategie promozionali basate sui cookie.

2.3.1 Il Transparency & Consent Framework (TCF)

Nel 2018 IAB (Interactive Advertising Bureau) Europe, una delle più importanti associazioni di categoria, tramite il suo IAB Tech Lab diffonde il proprio Transparency & Consent Framework v. 1.0, oggi arrivato alla versione 2.0, onde supportare il rispetto della normativa privacy appena entrata in vigore del GDPR e al contempo aiutare l'ecosistema del marketing digitale soprattutto nelle attività di retargeting. Questo nelle dichiarazioni di IAB.

L'obiettivo di favorire degli inserzionisti è parso da subito prevalente, pur affermando IAB di voler aiutare “tutte le parti della catena della pubblicità digitale a garantire che rispettino il GDPR dell'UE e la direttiva ePrivacy durante il trattamento dei dati personali o l'accesso e/o la memorizzazione di informazioni sul dispositivo di un utente, come cookie, identificatori pubblicitari, identificatori di dispositivi e altre tecnologie di tracciamento”.

La parte tecnica si è compiuta nella creazione di un ambiente condiviso in cui publisher e inserzionisti possono condividere dati, annunci e spazi pubblicitari digitali, applicando uno standard comune per la raccolta e gestione del consenso e “preferenze” degli utenti, a fini pubblicitari.

Di fatto il titolare di un sito web che voglia pubblicare nell'Unione annunci non solo propri ma di terzi, gestiti tramite network pubblicitari, è quasi costretto dal mercato a sposare il TCF di IAB. Uno standard comunque non rigido ma che offre diverse scelte e opzioni agli utilizzatori per il suo settaggio.

L'utente dei siti web partecipanti al TCF è scarsamente consapevole del suo uso: a parte la mancata trasparenza informativa, ciò che l'utente sperimenta è solo la comparsa di un banner o altra interfaccia grafica per i vari opt-in o opt-out richiesti per l'utilizzo dei dati raccolti (tramite cookie o altro).

Questa interfaccia rappresenta la Consent Management Platform (“CMP”) sfruttata da IAB per registrare le preferenze espresse dall'utente in un file, la “TC string”, condiviso tra i partecipanti al network TCF assieme a un cookie specifico e così contribuire al Real Time Bidding di IAB (denominato “OpenRTB”).

2.3.2 L'intervento dell'autorità belga e i punti di merito

L'RTB è un sistema di asta elettronica di piazzamento degli annunci digitali (display advertising) presso terzi, una forma di programmatic advertising.

Di fatto sono subito emerse notevoli criticità di compliance rispetto alla normativa dell'Unione, lamentate anche da associazioni di attivisti ed enti no profit a tutela della privacy: un'attea giunta a compimento con la prima pronuncia formale da parte di un'autorità di controllo. Il provvedimento dell'autorità è articolato e complesso, sfiorando le cento pagine di lunghezza: in questa sede presentiamo una prima panoramica dei principali punti critici emersi dal lavoro di indagine. A premessa va segnalato che la decisione è sì formalmente dell'autorità di controllo belga ma emessa in qualità di capofila (cioè come lead authority che guida e decide in casi che coinvolgono teoricamente più territori dell'Unione – c.d. “one-stop-shop” previsto dal GDPR per i casi transfrontalieri) in consultazione e condivisione con le autorità di tali territori, ben 27 in questa occasione. Quindi gli effetti di questa pronuncia vanno oltre il territorio belga e si ripercuotono di fatto sull'intera Unione e verso chi vuole trattare dati di persone presenti nell'Unione. Vedremo se e come le altre autorità coinvolte si esprimeranno nel proprio territorio di competenza, collegandosi alla decisione capofila. Sostanzialmente i punti censurati dall'autorità rispecchiano quelli già preannunciati nei mesi precedenti e che possiamo riassumere come segue:

- mancanza di liceità, viene fortemente criticata l'ammissibilità della base del legittimo interesse da parte degli utilizzatori, incompatibile con l'assetto combinato della Direttiva ePrivacy (basata sul consenso) e del GDPR, viziando la stessa conservazione e l'utilizzo dei dati personali confluiti nel TCF;
- mancanza di trasparenza, in assenza di un'informativa adeguata fornita tramite il CMP, troppo vaga sui complessi trattamenti attuati dal TCF, specie pensando all'uso di identificatori nella TCF string;
- carenze di accountability, security e privacy by default/by design, poiché IAB non ha comprovato all'autorità di poter garantire l'esercizio dei diritti da parte degli interessati e di poter monitorare la validità e integrità delle preferenze espresse dagli utenti;
- mancanza di corretto inquadramento di IAB Europe come titolare e, pertanto, dei correlati adempimenti quanto a nomina del DPO, di redazione dei registri di trattamento e di una DPIA afferente al TCF.

2.3.3 Cosa seguirà alla decisione

Come già detto sopra, il provvedimento commina una sanzione pecuniaria a IAB e intima alla stessa – entro un periodo massimo di sei mesi – di provvedere a “mettersi in regola”, ovvero:

- fornire una valida base giuridica per il trattamento dei dati personali, cioè l’elaborazione e la diffusione delle preferenze degli utenti nel contesto del TCF nonché vietare ai propri utilizzatori (attraverso una modifica ai propri termini contrattuali di utilizzo del TCF) l’uso della base del legittimo interesse nel TCF;
- assicurare efficaci misure di monitoraggio tecnico e organizzativo al fine di garantire l’integrità e la riservatezza del TC String;
- svolgere e perpetuare un rigoroso audit degli utilizzatori aderenti al TCF, sì da garantire che soddisfino i requisiti del GDPR;
- adottare misure tecniche e organizzative per impedire che il consenso sia pre-impostato “by default” nelle interfacce CMP;
- costringere le CMP degli utilizzatori ad adottare un approccio uniforme e conforme al GDPR circa le informative da rendere agli utenti;
- aggiornare i propri attuali registri dei trattamenti;
- effettuare una valutazione d’impatto sulla protezione dei dati (DPIA);
- nominare un proprio DPO, finora assente.

Il tutto dovrà avvenire in conformità a un piano di azione, da concordare con l’autorità belga.

Non solo: è stato anche ordinata la cancellazione definitiva dei dati personali già trattati nel proprio sistema TCF, aspetto che si estende a tutti gli utilizzatori del framework.

Precisato il ruolo di titolare di IAB Europe, i partecipanti dovranno configurarsi come contitolari, a seguito di apposito contratto di contitolarità ex art. 26 GDPR che dovrebbe essere realisticamente messo a disposizione da IAB (come accaduto in passato nel caso di Facebook e delle fan page).

A fronte di tutto questo, IAB da una parte sta valutando eventuali impugnazioni legali della decisione, dall’altro dichiara di voler lavorare con l’autorità per addivenire a un codice di condotta ex art. 40 GPDR basato sul TCF, come già aveva annunciato in precedenza.

D'altronde, puntualizza IAB, il provvedimento belga non ha comunque sancito un divieto d'utilizzo del TCF ma solo una sua "correzione".

Gli effetti dirompenti di questa decisione si potranno toccare con mano in futuro, in parallelo a ciò che sta accadendo circa la sempre maggiore chiusura da parte delle autorità europee di tutela della privacy verso i fornitori di servizi di marketing di diritto USA e alla crescente lotta ai dark pattern per acquisire consensi commerciali. Sommando sia la decisione di rilevanza europea appena descritta con la quasi-preclusione ai trasferimenti USA il mercato del digital marketing deve affrontare una sfida forse mai incontrata prima, una vera e propria rivoluzione più volte annunciata e che renderà il 2022 un punto di svolta.

CAPITOLO 3

DATA Brokers - Modello matematico di mercato

3.1 Generalità e caratteristiche

L'acquisizione di dati rappresenta l'impegno di attività che interessano i differenti settori dell'economia globale, come risultato della globalizzazione dove si attribuiscono priorità alle attività di analisi di mercato, per identificare l'orientamento alla produzione di beni materiali e l'erogazione di servizi.

L'acquisizione dei dati in tempo reale, secondo modalità differenti, rappresenta quanto oggi viene denominato come "Big Data".

Tra questi, per le imprese che dominano i mercati, sono considerati con valore i dati personali che è necessario catturare mediante lo sviluppo di algoritmi, specifici, per l'elaborazione. Pertanto, con la denominazione "Data Economy", stiamo ad indentificare i dati a valore economico che interessano, a modo di esempio, ambiti quali l'assistenza ospedaliera, la cura sanitaria a distanza, l'istruzione, i trasporti, l'informazione giornalistica, la fruizione di contenuti multimediali, i servizi finanziari e assicurativi, la ricerca del lavoro ed altro ancora.

Dall'impegno della cattura ed elaborazione dei dati emergono, come necessità indotta, nuove attività dal trattamento dei dati alla cyber security, senza trascurare le attività di monitoraggio e controllo del traffico, delle previsioni del tempo, delle attività sportive e tempo libero, dall'erogazione alla fruizione di servizi energetici ed altro ancora.

In breve, la cattura ed il trattamento dei dati, coinvolge la quotidianità. Pertanto, rappresenta oggi uno strumento, economico, per la valutazione della concorrenza e la crescita delle singole aziende, riconoscendo alle stesse la capacità di fornire indicazione di orientamento sul mercato specifico, su un segmento e su sotto mercato.

Si registrano studi che rivelano l'aumento di produttività ed una diminuzione degli sprechi come nuove opportunità per generare il miglioramento della qualità dei prodotti e servizi.

In altre parole, si attribuisce a tali attività il valore aggiunto che consente alle

aziende di orientare le proprie risorse verso le specifiche esigenze dei loro clienti, per far fronte alla forte innovazione indotta dagli sviluppi senza limiti della tecnologia Internet. Basti pensare alla competizione tra i motori di ricerca, dove solo chi possiede la tecnologia di gestione più avanzata, tendenzialmente, riesce a realizzare elaborazioni sempre più rapide e sofisticate orientando le scelte di mercato della vita collettiva.

È ben noto che la rete Internet continua a svilupparsi definendo un mondo che si trasforma da virtuale a reale e che coinvolge, sul piano socio-economico, la quasi globalità degli abitanti del pianeta senza confini ben definiti e regole giuridiche ed etiche.

È risaputo che i consumatori, quotidianamente, sono impegnati a praticare in rete attività online e offline, rivelando personali informazioni attraverso l'uso, soprattutto, di un dispositivo mobile.

In tali pratiche le entità, come oggi è possibile definire il consumatore, interagiscono con altre entità, i broker di dati, che vendono le informazioni al migliore offerente.

I Data Brokers sono aziende che raccolgono i dati direttamente, oppure tendono ad acquistarli da altre aziende come nel caso della vendita delle carte di credito e nel corso di attività offline, come la frequentazione dei social network.

L'obiettivo dei Data Brokers è la vendita e/o lo scambio delle informazioni con altre aziende, dopo attività di analisi ed aggregazione dei dati raccolti sui consumatori. Si mira ad orientare le scelte di quest'ultimi verso servizi e prodotti.

Le pratiche di intermediazione dei broker sollevano problemi di privacy poiché, i consumatori sono inconsapevoli delle attività che vengono condotti su di essi.

A tal proposito, per limitare le problematiche di privacy, a partire dal 1° gennaio 2010 la comunità europea ha deciso di limitare le loro pratiche mediante procedimenti che rendano note e trasparenti le proprie attività.

La pratica del data broker ha inizio con l'acquisizione dei dati sviluppando un processo economico che è modellato per comprenderne le dinamiche e per sviluppare delle valutazioni previsionali, per l'approccio ad un mercato sempre più sofisticato e complesso.

La trattazione è completata menzionando un rimando, breve, alle fonti di acquisizione dei dati, classificabili in:

- fonti governative;
- fonti pubblicamente disponibili;
- fonti commerciali;

Generalmente i broker aggregano tutti i dati per formare un composito dettagliato del consumatore.

La maggior parte recupera informazioni da blog e siti di social media, che sono monitorati da algoritmi, atti a catturare informazioni, con attività di scansione, sui visitatori che non hanno impostato alcuna limitazione sulla privacy, implicitamente e involontariamente.

Gli algoritmi sono sviluppati per consentire la cattura dei dati sensibili da più fonti, tracciando in rete l'entità individuata. Questo è quanto emerge secondo uno studio condotto, nel maggio del 2014, dalla Federal Trade Commission.

Nell'ambito della produzione dei dati, il processo di cattura è costantemente aggiornato mediante la stima della frequenza dell'entità, in tempo reale, su base giornaliera, settimanale, mensile e annuale.

A tal proposito gli stessi Data Brokers vengono classificati sulla base dei prodotti da loro offerti in:

- Data Brokers per il marketing e la pubblicità;
- Data Brokers per l'individuazione delle frodi;
- Data Brokers per la verifica di identità;
- Data Brokers per ricerca di persone;

Tra le metodologie di acquisizione, il broker impiega la modalità offline della gestione dei cookie per identificare e categorizzare i consumatori sulla base di una caratterizzazione particolare e specifica.

Tra le metodologie impiegate, si annoverano gli algoritmi sviluppati per modellare centinaia o migliaia di dati acquisiti da fonti differenti, con lo scopo di associare ai dati dei punteggi che consentano di categorizzare il consumatore secondo le sue particolari caratteristiche che vengono vendute a clienti interessati alla conoscenza

degli orientamenti dei consumatori, per gestirne le aspettative in termini di offerta di prodotti e servizi.

Dagli studi condotti emergono attività praticate dai broker, su richiesta di gruppi finanziari, per aiutare i loro clienti ad evitare le frodi tracciandone gli indirizzi e verificandone la veridicità della compilazione di un modulo, compilato dallo stesso consumatore del servizio all'atto dell'offerta, durante la transazione economica.

Si ritiene prioritario l'identificazione del consumatore, inteso come utilizzatore del servizio finanziario, per evitarne la violazione delle informazioni personali e tutelarlo dalle frodi.

La verifica delle identità rappresenta, per tanti data brokers, un fertile campo di business, soprattutto, per gli istituti bancari nel corso delle transazioni.

Nel caso specifico, offrono algoritmi esplicativi che associano il consumatore, opportunamente identificato, ad un punteggio stimato sulla base delle info condivise in modo ufficiale. Un consumatore con un punteggio classificato come punteggio a rischio elevato, viene ad essere interpretato come un consumatore, ad esempio, associato ad una persona deceduta. Pertanto, si associa ad esso l'attività di praticare un quiz, che include domande oscure ad un ladro di identità.

Per la maggiore, si impiegano quiz di tipo "match/no match". Ovvero, si chiede di confermare le informazioni da verificare con quanto conservato e disponibile nel file broker, identificato come verifica dello stato prodotto.

Altro campo di applicazione, molto interessante, è il servizio people search che identifica 'un'insieme di attività di ricerca basati su info personali quali nome, indirizzo, città/stato, numero di cellulare, indirizzo e-mail, nome utente o SSN'.

Trattasi di processi di ricerca che, sulla base di un solo parametro chiave, risalgono in maniera univoca al set di dati completo dell'entità univoca, in termini di informazioni quali l'età, la data di nascita, il genere, la formazione scolastica, le relazioni personali, gli interessi, eventuali matrimoni, parenti o informazioni su eventuali divorzi o precedenti penali ed altro ancora.

Il servizio people search offerto, in modalità gratuita, risulterà essere sicuramente meno affidabile di un servizio a pagamento. Sulla base della differenziazione di servizio indotta dalla tipologia di clientela, ad esempio, lo screening di monitoraggio richiesto da un istituto finanziario, sarà differente da quello richiesto da un rivenditore per verificare divieti sull'utilizzo dei prodotti per scopi illeciti, come nel caso di uso di software.

Le strategie di verifica di identità sono differenti, ci sono broker che pubblicano le loro condizioni d'uso sui siti web senza richiedere esplicitamente di accettare, altri, invece, richiedono di accettare i termini quando completano la transazione.

Si noti che nel processo di screening sono da considerarsi attività ad elevata frequenza, la verifica per accertare la reputazione di un'entità legittima mediante controlli di riferimento, implementando codici di codifica dei dati.

Pertanto, la qualità predittiva del file broker offerto viene esaltata nel corso della vendita soprattutto da brokers specializzati in diversi segmenti e servizi che impiegano processi ad apprendimento algoritmico per la personalizzazione del broker file da offrire.

Si vanno a definire settori dove i broker interagiscono e cooperano tra loro per rendere disponibili informazioni più accurate ed attendibili, opportunamente verificate rispetto a fonti dirette e combinate per offrirle come specialistiche.

Si creano scenari dove i broker condividono i dati per cooperare e altri dove gli stessi competono secondo un comportamento attribuibile, esclusivamente, alla domanda.

In linea di massima, sulla base di quanto introdotto, è possibile immaginare la creazione di un processo produttivo di dati che ha come obiettivo la necessità di fornire un prodotto “i dati” a clienti avidi di profitti.

Pertanto, si studiano modelli atti a rappresentare la gestione dei dati, dalla produzione fino alla messa a disposizione sul mercato.

Non manca da considerare la concorrenza tra i brokers, sempre più agguerriti, che spesso scelgono strategicamente la vendita diretta oppure la condivisione, a seconda della natura dei dati. Pertanto, decidono di sostenere direttamente un cliente in un sotto-mercato, mentre decidono di cooperare in un'altra parte del mercato.

A seguire si propone un modello di mercato su cui si affacciano i brokers, specialistici, che tendono di conquistare il cliente secondo la concorrenza diretta oppure la cooperazione per condivisione.

Il modello proposto, sulla base di ipotesi iniziali, modella il costo del file broker ed il prezzo che il cliente deve sostenere per acquisirlo.

3.2 Modellazione del mercato dei DATA Brokers

Si suggerisce di assumere a riferimento il modello matematico di Yiquan Gu, Leonardo Madio e Carlo Reggiani (febbraio 2019), che analizza il regime di concorrenza tra due DATA Brokers che praticano la propria strategia per acquisire uno stesso potenziale cliente, caratterizzante uno specifico segmento di mercato.

Si suppone che i due broker $k = 1,2$ possiedano, ciascuno, la sua lista broker sui consumatori da proporre al potenziale cliente da catturare.

Ciascuno ha raccolto i dati in modo indipendente.

Presentiamo il modello indicando:

con $N > 0$ il numero di consumatori, caratterizzati da un set di $M > 0$ di dati, supponendo che ogni broker ha accesso indipendente a un sottoinsieme di dati dei consumatori, fissata K la matrice logica di dimensione $M \times N$ che rappresenta le informazioni del broker k .

L'elemento $k_{ji} = 1$ della matrice rappresenta il broker k sul dato i -esimo del consumatore j -esimo, mentre con:

$$f() \text{ e } g()$$

viene indicata una funzione che misura il surplus extra da generare rispetto ad una situazione di partenza, dove i dati non sono disponibili all'impresa, obiettivo della cattura da parte del broker.

Il modello è di tipo incrementale considerando che i set di dati non sono uguali e non hanno lo stesso valore, per un cliente id. 1 modellato con la funzione $f()$ e per il cliente id. 2, modellato con $g()$.

Si noti che i dati di un broker differiscono tra settori e all'interno dello stesso.

In senso lato, il modello è stato generato per modellare una qualsiasi generazione di surplus da diversi tipi di attività sui dati dove il cliente id. 1 spera di ottenere:

$$h() > f()$$

analogamente per il cliente id. 2.

Considerando la funzione f , essendo possibile estendere la considerazione anche alla funzione g , si precisa che la struttura dei dati è di tipo:

- **Additivo**, se: $f(A_k | A_{-k}) = f(A_k) + f(A_{-k})$
- **Sub-additivo**, se: $f(A_k | A_{-k}) < f(A_k) + f(A_{-k})$
- **Super-additivo**, se: $f(A_k | A_{-k}) > f(A_k) + f(A_{-k})$

La struttura dei dati è additiva, quando il valore dell'insieme di dati unito è pari alla somma dei singoli valori dei set.

La struttura dei dati è sub-additiva, quando il valore dell'insieme di dati unito è inferiore alla somma dei valori dei set.

La struttura dei dati è super-additiva, quando il valore dell'insieme di dati unito è superiore alla somma dei valori dei set.

Combinando i set di dati su due segmenti di mercato, il set risultante, raggruppa tutti gli attributi rilevanti di tutti i consumatori su tutto il mercato; tanto da facilitarne la commercializzazione, indipendente, minimizzandone le sovrapposizioni.

Il modello si propone di prevedere i ricavi derivanti dalla vendita del loro set. Nell'impostazione del modello, secondo un primo approccio, i Data Brokers vendono il loro set in modo indipendente all'impresa, acquisendo gli utili k come:

$$\pi_k = \begin{cases} p_k & \text{se l'impresa compra i dati da } k \\ 0 & \text{se l'impresa non compra i dati da } k \end{cases}$$

dove p_k è il prezzo per i propri dati.

In secondo approccio, i brokers condividono il proprio set univoco, per una specifica azienda, acquisendo una quota s_k del profitto individuale rispetto al broker rivale.

Il modello stima il beneficio derivante dall'acquisto dei dati, ovvero, un profitto extra rispetto alla pratica normale con la possibilità di espansione sul mercato, praticando strategie di prezzo più sofisticate.

In particolare, quando i Data Brokers non condividono i dati, i profitti sono:

$$\Pi^r = \pi_0 + \begin{cases} f(A_k) - p_k & \text{se l'azienda acquista dati solo da } k \\ f(A_k | A_{-k}) - p_1 - p_2 & \text{se l'azienda acquista dati da entrambi i brokers} \end{cases}$$

L'azienda acquista i dati solo da un broker k , ottenendo $f(k)$ e pagando un prezzo p_k , mentre quando acquista indipendentemente i dati da entrambi i brokers li unisce, senza costi aggiuntivi, per ottenere un surplus aggiuntivo pagando entrambi i brokers.

Offrendo i set in combinazione, offrono al cliente la possibilità di ottenere più profitto:

$$\Pi^r = \pi_0 + f(A_k | A_{-k}) - P_{A_k | A_{-k}}$$

Secondo una regola di condivisione, che aumenta rigorosamente, anche i profitti di entrambi i brokers.

La strategia è preferita da entrambi i brokers, poiché il cliente decide se acquistare o meno il set offerto, valutando se il prezzo è tale da massimizzare i suoi profitti.

Ipotizzando una struttura dei dati additiva, si può definire una configurazione di equilibrio di Nash univoco tale che:

$$p_k^* = f(A_k)$$

per $k = 1$ e 2 .

Se, invece, la struttura dei dati è sub-additiva si può definire una configurazione di equilibrio di Nash tale che:

$$p_k^* = f(A_k | A_{-k}) - f(A_{-k})$$

per $k = 1$ e 2 .

Se, infine, la struttura dei dati è sovra-additiva si può definire una configurazione di equilibrio di Nash tale che:

$$p_1^{\mathcal{A}} + p_2^{\mathcal{A}} = f(\mathcal{A}_1 | \mathcal{A}_2) \quad e \quad p_k^{\mathcal{A}} \geq f(\mathcal{A}_k)$$

per $k = 1$ e 2 .

Il cliente finale mirante ad ottenere il massimo profitto, in accordo con il risultato predittivo di una struttura di dati additiva, acquista entrambi i set di dati e li usa congiuntamente, indipendentemente dal broker ed indipendentemente se uno oppure l'altro possieda i dati dell'altro.

In sintesi, il cliente finale preferisce acquistare entrambi i set di dati per usarli congiuntamente ad un prezzo finale pari a $p_k + p_{-k} = f(k) + f(-k)$.

I Data Broker, nella cooperazione, mirano ad ottenere il massimo dei profitti che coincide con il valore dei rispettivi set di dati.

Ipotizzando una struttura dei dati sub-additiva, i brokers fissano i prezzi come segue:

$$p_k^z = f(A_k | A_{-k}) - f(A_{-k})$$

per considerare il contributo marginale che è possibile estrarre dal valore dell'insieme di dati congiunto, utilizzato come surplus dal cliente finale, ottenendo profitti come:

$$\Pi^r = \pi^0 + f(A_k) + f(A_{-k}) - f(A_k | A_{-k})$$

Infine, ipotizziamo una struttura di dati sopra-additiva, caratterizzata da una qualsiasi coppia di (p_k, p_{-k}) in questo caso non è possibile definire una condizione di equilibrio univoco poiché abbiamo una molteplicità di equilibri dove, in corrispondenza di ciascuno di essi, il cliente finale acquista i dati da entrambi i brokers, ricavando profitti da tutta la generazione di surplus abilitata dai dati resi disponibili dai due data Brokers.

Concludendo, i profitti del cliente finale e i profitti combinati dei due brokers sono dati, considerando le seguenti ipotesi di struttura, da:

- *Additiva:*

$$\Pi^r = \pi^0 \quad e \quad \Pi_1 + \Pi_2 = f(A_1) + f(A_2) = f(A_1 | A_2)$$

- *Sub-additiva:*

$$\Pi^r = \pi^0 + f(A_1) + f(A_2) - f(A_1 | A_2) > \pi^0$$

$$\Pi_1 + \Pi_2 = 2f(A_1 | A_2) - f(A_1) - f(A_2) < f(A_1 | A_2)$$

- *Super-additiva:*

$$\Pi^r = \pi^0 \quad e \quad \Pi_1 + \Pi_2 = f(A_1 | A_2)$$

Gli autori ritengono che, quando la struttura dei dati è additiva o sovra-additiva, sia possibile da parte dei broker agire indipendentemente, per ottenere l'intero surplus dal cliente finale.

Nel caso in cui la struttura dei dati è sub-additiva, il cliente finale acquista il data list da entrambi i brokers ad un prezzo:

$$p_k^z = f(A_k | A_{-k}) - f(A_{-k}) < f(A_k)$$

che risulta inferiore al valore di ciascun set di dati poiché, quest'ultimo, è impostato considerando il contributo del singolo broker, fissato dallo stesso, considerando il proprio contributo marginale. Pertanto, in questa condizione, l'impresa si appropria di una parte dell'eccedenza generata dall'utilizzo del file broker.

Si consideri lo scenario in cui i brokers condividono i loro dati senza avere alcuna possibilità di vendere singolarmente il proprio set di dati, in tal caso, i profitti individuali sono:

$$\Pi_k = s_k(A_k, A_{-k}) \cdot f(A_k | A_{-k})$$

dove s_k è la regola di condivisione ed è data da:

$$s_k(A_k, A_{-k}) = \frac{f(A_k)}{f(A_k) + f(A_{-k})}$$

Si precisa che i broker condivideranno i dati se e solo se, per entrambi, la previsione dei profitti individuali è superiore a quelli ottenuti vendendo i dati in modo indipendente.

La regola di condivisione dovrebbe considerare il potere contrattuale. Pertanto, deve essere formulata modellando il valore relativo del set di dati.

Dall'analisi risulta che la condivisione dei dati esiste solo quando la struttura dei dati è sub-additiva. In tutti gli altri casi, come brevemente accennato, i dati vengono venduti in modo indipendente, poiché il surplus generato non comporta alcun miglioramento per ciascun broker.

Con riferimento alla struttura dei dati sub-additiva, la differenza nei profitti tra condivisione dei dati e vendita indipendente è:

$$\begin{aligned}\Pi_k - p_k^z &= s_k(A_k, A_{-k}) \cdot f(A_k | A_{-k}) - f(A_k | A_{-k}) + f(A_{-k}) = \\ &= f(A_{-k}) - s_{-k}(A_k, A_{-k}) \cdot f(A_k | A_{-k}) > 0\end{aligned}$$

dove il prezzo fissato congiuntamente è:

$$P^z = f(A_k | A_{-k})$$

Ciascun broker nella condivisione spera di migliorare la propria posizione, attenuando la concorrenza, a spese delle imprese e dei consumatori. Ma nello stesso tempo, mira ad evitare la guerra dei prezzi.

Qualunque sia l'oggetto di una commercializzazione, la guerra dei prezzi è causa di collasso, soprattutto, nel caso in cui non si riesce a sostenere l'adeguamento tecnologico ed innovativo. Quest'ultimo, è indispensabile nel processo di produzione della broker list, soprattutto, nella fase di elaborazione dei dati.

Indicando con c il costo per la fusione dei set di dati, il profitto è pari a:

$$\Pi^r = \pi^0 + f(A_k) + f(A_{-k}) - f(A_k | A_{-k}) + c$$

dove il termine:

$$f(A_k) + f(A_{-k}) - f(A_k | A_{-k})$$

indica il guadagno generato dalla competizione.

La c è modellata per soddisfare la condizione:

$$c \leq \min \{f(A_k | A_{-k}) - f(A_{-k}), f(A_k | A_{-k}) - f(A_k)\}$$

Considerando che il cliente finale paga un prezzo combinato:

$$p_k^z + p_{-k}^z = f(A_k) + f(A_{-k}) - 2c$$

Si vuole ipotizzare, completando la ns trattazione, che il prezzo fissato dal broker sia sufficientemente basso tale che:

$$P_{-k} \leq f(A_k | A_{-k}) - f(A_k) - c$$

Dove:

$$p_k \in [0, f(A_k | A_{-k}) - f(A_k) - c]$$

Di conseguenza:

$$p_k^z = f(A_k | A_{-k}) - f(A_k) - c$$

e quindi il prezzo combinato sostenuto dal cliente finale è:

$$p_k^z + p_{-k}^z = 2f(A_k | A_{-k}) - f(A_k) - f(A_{-k}) - 2c$$

a cui associo per i broker i seguenti profitti:

$$\Pi^r = \pi^0 + f(A_k) - f(A_{-k}) - f(A_k | A_{-k}) + c$$

il cui valore dipende, esclusivamente in termini di vantaggi, per ciascun broker unicamente dalla regola di condivisione.

Generalizzando, considerando la struttura di dati sopra-additiva descritta, si ottiene il seguente modello matematico:

$$\begin{cases} f(A_k) & \text{se } p_{-k} \geq f(A_k | A_{-k}) - f(A_k) - c \\ f(A_k | A_{-k}) - p_{-k} - c & \text{se } f(A_{-k}) \leq p_{-k} < f(A_k | A_{-k}) - f(A_k) - c \\ f(A_k | A_{-k}) - f(A_{-k}) - c & \text{se } p_{-k} < f(A_{-k}) \end{cases}$$

Dalla breve trattazione risulta evidente che lo scenario è complesso. Tanto che, il cliente finale, si trova a valutare: non acquistare nulla, acquistare solo da k , acquistare solo da $-k$ o acquistare da entrambi.

Pertanto, se:

$$P_{-k} \geq f(A_k | A_{-k}) - f(A_k) - c$$

il cliente finale acquista solo da k ;

se:

$$f(A_{-k}) \leq p_{-k} < f(A_k | A_{-k}) - f(A_k) - c$$

il cliente finale acquista da entrambi o non acquista niente se:

$$P_{-k} \leq f(A_{-k})$$

il cliente finale trova indifferente l'acquisto da $-k$ o l'acquisto da entrambi i brokers. Se ne deduce che, quando l'impresa deve sostenere un costo diverso da zero per unire i due set di dati, ogni broker gli offre uno sconto esattamente pari a tale costo per garantirsi che l'impresa acquisti anche il proprio set. Tale condizione, come precedentemente mostrato, ha come effetto ultimo l'intensificarsi della concorrenza tra i due brokers, che per tanto preferiranno condividere i loro set di dati indipendentemente dalla loro struttura. Tale condivisione punta alla massimizzazione dei profitti dei brokers e ad una completa estrazione del surplus generato.

Queste le condizioni che concludono la nostra breve trattazione sviluppata per descrivere lo scenario, complesso, della commercializzazione della data list.

3.3 Produzione dei DATA e commercializzazione

Si propone la classificazione dei Data, considerando le attività che vengono implementate sugli stessi.

- **Data Acquisition:** Processo di raccolta, filtraggio e selezione dei dati prima del loro inserimento in un data warehouse o in un qualsiasi altro archivio. Rappresentano i dati oggetto di query;
- **Data Analysis:** Esplorazione, trasformazione e configurazione dei dati con rilevanza maggiore. Dati grezzi utilizzabili nell'ambito di un processo decisionale, pertanto, rappresentano l'obiettivo di attività implementate sui data curation, data storage e data usage;
- **Data Curation:** Gestione dei dati per la creazione, la selezione, la classificazione, la trasformazione, la verifica, la conservazione di contenuti adattabili agli obiettivi prefissati;
- **Data Storage:** Modalità di gestione dei dati atta a soddisfare le esigenze di applicazioni che richiedono un accesso veloce a modelli strutturati e non;
- **Data Usage:** Attività atte ad integrare l'analisi dei dati all'interno dell'attività d'impresa;

Nell'ambito della Data Usage un aspetto su cui la letteratura si è spesa molto, sono i contratti stipulati tra Data Brokers e loro clienti. In fase contrattuale alcuni data brokers dichiarano che le loro informazioni si basano su determinate fonti e sulla loro accuratezza, senza assumersi alcuna responsabilità circa la loro qualità. Altri data brokers si concentrano sulla qualità predittiva dei loro dati.

A seguire, si vuole descrivere il ruolo dei data brokers attraverso una delineazione delle attività implementate.

Nello specifico, hanno la possibilità di accedere a determinate fonti, definite esclusive, acquisendo informazioni private e personali in settori quali: il finanziario e la comunicazione multimediale.

I dati vengono venduti per sviluppare previsioni e strategie economiche, attraverso diverse modalità, quali: il marketing e la pubblicità.

Offrono alle banche e agli operatori di telefonia mobile database di contatti, ovvero, identità per eliminare le frodi nel caso della concessione di prestito, nel caso di acquisti con carte di credito ed altro ancora.

L'operato è oggetto di critiche per le violazioni della privacy nel corso delle attività di raccolta e divulgazione delle informazioni, considerando che il valore dei dati viene sottoposto ad un processo di miglioramento dell'interpretabilità e funzionalità.

Il valore può essere incrementato mediante la trasformazione del formato di archiviazione di una risorsa informativa e la traduzione da una lingua all'altra, al fine di ottenere una piena corrispondenza tra domanda e offerta informativa.

Pertanto, la produzione dell'informazione dal dato grezzo al dato sofisticato richiede software ed applicazioni per la commercializzazione dei big data.

Essi collaborano con aziende specializzate in metodologie analitiche, implementate con elevate risorse tecnologiche, atte a sviluppare soluzioni customizzate. Ovvero, si adattano le informazioni al particolare cliente, che paga il valore per incrementare i propri profitti. Le informazioni più preziose, in termini di redditività, per i brokers sono le informazioni relative ad un tema specifico di settore.

In questa breve trattazione, è stato illustrato il commercio di dati eseguito tra i brokers. Approfondendo la modellazione della cooperazione e competizione tra i brokers in merito all'approccio strategico su un mercato, dove i clienti sono immersi in una realtà a carattere sempre più globale.

È stato presentato un modello relativamente semplice, per un'economia di scambio e commercializzazione dei dati, adattabile a mercati diversi. Dove i diversi brokers competono e/o cooperano per acquisire il favore del cliente finale, stimando il costo degli uni (i brokers), mentre il cliente finale, è interessato al prezzo da sostenere per acquisire il valore del set dati.

Il modello è da considerarsi uno strumento di stima dei profitti, indotto dalla gestione dei dati, per gli uni e gli altri.

Nel dettaglio, si riscontrano nella cooperazione, vantaggi per i broker che colgono l'occasione per aggiornare la propria data list e per il cliente che acquista la lista dei data broker offerta con informazioni più dettagliate sul consumatore.

Si evidenzia, che il mercato dei broker di dati, è caratterizzato da una mancanza di trasparenza verso consumatori, che vengono analizzati senza esserne informati. Pertanto, per quest'ultimi, si ritiene necessaria definire una legislazione che tuteli lo scambio, sia nell'ambito dello stesso settore che tra i diversi. Soprattutto, nello scambio di informazioni sensibili, per evitarne gli abusi.

La era è caratterizzata dall'economia digitale che si basa su transazioni di moneta virtuale tra gli utenti, su motori di ricerca, operatori telefonici, gestione di shopping online ed altro ancora.

Da numerosi studi emerge anche la necessità di istituire un antitrust che impedisca ai data brokers di praticare attività in regime anticoncorrenziale, producendo inefficienza come la violazione della privacy dei consumatori e perdita di profitti per le aziende clienti che acquistano la data list.

Si ritiene necessario suggerire una gestione a livello digitale, mediante l'applicazione specifica e sviluppata dai progettisti. Considerando la natura della struttura dei dati, richiamata durante la nostra trattazione in merito alla modellistica, è da considerarsi opinione puramente personale, la necessità di praticare l'aggiornamento dell'attuale normativa sulla privacy. Poiché si ritiene la stessa non adeguata all'attuale sviluppo tecnologico caratterizzante la globalizzazione della società.

Dobbiamo riflettere sulla realtà, oramai concreta, rappresentata dai data broker, caratterizzanti un contesto in cui non ci preoccupiamo di fornire le nostre informazioni a degli "sconosciuti" che tendono ad utilizzarli in maniera errata e per scopi illeciti. Senza avere la possibilità di proteggere la nostra identità in un mondo digitale sotto il controllo di aziende hi-tech, in forma di piattaforme sempre più sofisticate ed in concorrenza tra loro.

Pertanto, le prospettive di dominare il mercato sono vincolate al servizio competitivo offerto dalle piattaforme digitali che assumono un ruolo importante nella differenziazione della cattura dei dati. Specialmente se, per ottenerli, impiegano pratiche che violano illecitamente la privacy ed i diritti degli utenti.

CAPITOLO 4

Modelli di mercato

4.1 Generalità e caratteristiche

La commercializzazione del file data broker dipende dal valore e dalle caratteristiche tecnologiche del mercato di destinazione.

L'analisi dei mercati deve necessariamente considerare gli effetti che scaturiscono dalle attività tipiche dell'economia digitale caratterizzata dai big data e dai cloud.

Gli studi si concentrano principalmente su agenti economici e consumatori che determinano la definizione del prezzo.

Diversi settori di mercato sfruttano i dati acquisiti dai brokers, soprattutto, nel campo dell'e-commerce dominato dalle piattaforme digitali.

Le transazioni su piattaforma, rappresentano l'oggetto principale delle strategie di marketing miranti ad acquistare clienti.

La minimizzazione dei costi di gestione della piattaforma è necessaria per offrire al consumatore il minimo prezzo.

Generalmente, i consumatori preferiscono le piattaforme con numerosi partecipanti.

Nell'ambito della transazione digitale vengono preferite le piattaforme con l'interfaccia intuitiva e facilmente accessibile.

Il mercato multimediale è caratterizzato dai consumatori che partecipano attivamente alle offerte.

Il consumatore viene identificato secondo le direttive del file data broker.

Quest'ultimo viene configurato sulle caratteristiche della piattaforma per favorirne azioni di marketing e la gestione da parte delle imprese che collaborano, sia direttamente che indirettamente.

Obiettivo della piattaforma dedicata all'e-commerce, è la necessità di incrementare esponenzialmente i frequentatori.

Lo scenario è molto complesso. Il consumatore finale ha un ruolo importante sulla piattaforma.

Nell'ambito di un'economia digitalizzata, la sua frequenza e il suo numero esponenziale rappresenta la moneta di scambio delle transazioni tra le aziende.

Il comportamento del consumatore rappresenta l'interesse di studio di piattaforme specializzate e non.

Per cui, la piattaforma viene gestita sulla base dello sfruttamento delle informazioni acquisite con il file data broker.

Il file data broker è oggetto di analisi specifiche per la quantificazione del valore sulla base degli effetti di rete, in modo particolare, attraverso la simulazione di scenari.

Con il trascorrere del tempo, il valore riconosciuto al file data broker dipende dal continuo aggiornamento.

La piattaforma stessa richiede un continuo aggiornamento del software per aggiungere nuove funzionalità, scaturite dalle azioni sviluppate dagli utenti finali.

Teoricamente, la determinazione del valore del data file broker dipende dal comportamento del consumatore che influisce sull'accesso e partecipazione di nuovi consumatori.

Nel commercio su piattaforma, si cerca di estrarre il massimo profitto dal singolo consumatore.

Il mercato è fortemente mutevole nel tempo, in quanto, è caratterizzato da consumatori che acquisiscono una forte elasticità di prezzo rispetto ad altri.

Quest'ultimi, diventano più interessanti rispetto a coloro che possiedono un'elevata partecipazione, a causa della frequenza di acquisti.

Quanto introdotto, lascia intravedere la complessità della concorrenza tra piattaforme operanti in uno stesso settore.

Secondo una prima valutazione, il consumatore è catturato dalla riduzione del prezzo.

È necessario, però, evitare che la riduzione del prezzo diventi causa di degrado per la piattaforma.

Concludendo, si ritiene fortemente negativo la condotta dei data broker che si spostano da una piattaforma all'altra e che causano la migrazione dei consumatori stessi.

A tal proposito, non mancano esempi di piattaforme rese specializzate dalla presenza esclusiva di data broker specifici.

4.2 Il modello Evans e Schmalensee (2012)

La determinazione del prezzo di piattaforma è particolarmente complessa e la letteratura si è spesa molto su tale tematica.

Il modello proposto si riferisce ad un particolare mercato, il mercato a due versanti. Con mercati a due versanti si intendono mercati caratterizzati dalla presenza di una piattaforma gestita da un operatore, che funge da luogo di incontro e scambio fra due gruppi di utilizzatori tra loro interdipendenti. La piattaforma consente loro di realizzare transazioni minimizzando i costi.

La caratteristica fondamentale di questi mercati è la presenza di effetti di rete indiretti, cioè il gruppo di utilizzatori di un versante beneficia dell'aumentare del numero di utilizzatori dell'altro versante. Tale relazione è reciproca, dunque, entrambi i gruppi sono attratti dalla numerosità dell'altro gruppo.

Esempio tipico di mercato di questo tipo è il mercato delle carte di credito. In tale mercato i consumatori saranno incentivati ad entrare in un circuito compatibile con il più alto numero possibile di postazioni, allo stesso modo i commercianti saranno incentivati all'utilizzo di un determinato sistema di pagamento quanto più è alto il numero di clienti. Vero è che, in un simile mercato, è considerato importante anche il costo di utilizzo. Tuttavia nessuno sarebbe interessato ad una carta di credito economica che nessun commerciante accetterebbe.

In questi particolari mercati si palesano due strategie percorribili: il prezzo e l'apertura. Mentre la prima è stata oggetto di una ricerca molto rigorosa, la letteratura si è espressa poco in merito all'apertura. In effetti, appare evidente come la determinazione del prezzo in questi mercati è particolarmente complessa, dipende non solo da domanda e costi di un versante del mercato, ma anche da come la loro partecipazione influisce sulla partecipazione dell'altro versante. Il tutto tenendo in considerazione i profitti della piattaforma su entrambi i versanti del mercato. Si evince come i prezzi su entrambi i versanti dipendono dall'insieme congiunto di elasticità della domanda e costi marginali su ciascun lato.

Il modello proposto di seguito intende proprio analizzare tale problema, che noi cercheremo di adattare al mondo dei Data Brokers.

Consideriamo la presenza di due data broker che vendono i loro set di dati al prezzo p_1 e p_2 . Dato un certo numero di potenziali utilizzatori della piattaforma, il prezzo da sostenere per il file data broker deve essere tale da garantire la massima utilità:

$$U_L = k_L + \theta_{i,j}y_L - p_L$$

Prendendo in esame come prezzo della piattaforma p_L , si consideri la valutazione del bene k_L^* come segue:

$$\rightarrow k_1^z + \theta_{21}y_2 - p_1 = 0 \qquad k_1^z = p_1 - \theta_{21}y_2$$

$$\rightarrow k_2^z + \theta_{12}y_1 - p_2 = 0 \qquad k_2^z = p_2 - \theta_{12}y_1$$

Si concede l'accesso alla piattaforma solo se l'utilità di base è $\geq k_L^*$.

In particolare si considerano le funzioni:

$$\begin{cases} y_1 = 1 + \theta_{21}y_2 - p_1 \\ y_2 = 1 + \theta_{12}y_1 - p_2 \end{cases}$$

Da cui, risolvendo il sistema:

$$\begin{cases} y_1(p_1, p_2) = \frac{1 - p_1 + \theta_{21}(1 - p_2)}{1 - \theta_{21}\theta_{12}} \\ y_2(p_1, p_2) = \frac{1 - p_2 + \theta_{12}(1 - p_1)}{1 - \theta_{12}\theta_{21}} \end{cases}$$

Stiamo adattando un modello matematico alla commercializzazione del file data broker. Si vuole determinare il prezzo del servizio di piattaforma in modo da considerare nell'offerta un prodotto e/o bene.

Ipotizziamo di applicare le regole che governano l'offerta e la domanda.

Considerando che la domanda è una funzione decrescente di entrambi i prezzi:

$$\frac{\partial y_i}{\partial p_i} = - \frac{1}{1 - \theta_{21}\theta_{12}} < 0$$

la domanda del lato decresce all'aumentare del prezzo del lato "j":

$$\frac{\partial y_j}{\partial p_j} = - \frac{\theta_{ij}}{1 - \theta_{21}\theta_{12}} < 0$$

dove un incremento del prezzo sul lato j riduce la quantità domandata sul lato "i" al punto che "pj" aumenta, allora "yj" diminuisce e "yi" aumenta fino a determinare il prezzo che ciascun broker porterà sul mercato.

Date le funzioni di domanda precedentemente ricavate, è necessario massimizzare i possibili profitti. Pertanto, per l'impresa/broker 1 i profitti sono dati da prezzo per domanda:

$$Max_{p_1} \pi_1(p_1, p_2) = p_1 \cdot \frac{1 - p_1 + \theta_{21}(1 - p_2)}{1 - \theta_{12}\theta_{21}}$$

Ponendo la derivata uguale a zero:

$$\frac{\partial \pi_1}{\partial p_1} = 0$$

si ottiene:

$$p_1 = \frac{1}{2} + \frac{\theta_{21}(1 - p_2)}{2}$$

Per l'impresa/broker 2, applicando il *ragionamento analogo*:

$$\text{Max}_{p_2} \pi_2(p_1, p_2) = p_2 \cdot \frac{1 - p_2 + \theta_{12}(1 - p_1)}{1 - \theta_{12}\theta_{21}}$$

$$\frac{\partial \pi_2}{\partial p_2} = 0$$

da cui:

$$p_2 = \frac{1}{2} + \frac{\theta_{12}(1 - p_1)}{2}$$

Mettendo a sistema le due condizioni del primo ordine, si ottengono i prezzi di equilibrio:

$$p_1^{ind} = \frac{2 + \theta_{21}(1 - \theta_{12})}{4 - \theta_{12}\theta_{21}} \quad p_2^{ind} = \frac{2 + \theta_{12}(1 - \theta_{21})}{4 - \theta_{12}\theta_{21}}$$

e i profitti di equilibrio:

$$\pi_1 = \frac{[2 + \theta_{21}(1 - \theta_{12})]^2}{(1 - \theta_{12}\theta_{21})(4 - \theta_{12}\theta_{21})^2} \quad \pi_2 = \frac{[2 + \theta_{12}(1 - \theta_{21})]^2}{(1 - \theta_{12}\theta_{21})(4 - \theta_{12}\theta_{21})^2}$$

Consideriamo il caso di un solo acquirente del file data broker che cerca di massimizzare i propri profitti dalla cooperazione dei broker.

Si nota che ciascun offre il proprio file data broker in modo da massimizzare i propri profitti:

$$\text{Max}_{p_1 p_2} \Pi(p_1, p_2) = p_1 p_1(p_1, p_2) + p_2 p_2(p_1, p_2)$$

Considerando:

$$p_1(p_2) = \frac{1 + \theta_{21} - (\theta_{12} + \theta_{21})p_2}{2} \quad e \quad p_2(p_1) = \frac{1 + \theta_{12} - (\theta_{12} + \theta_{21})p_1}{2}$$

Mettendo a sistema queste due condizioni, si ricavano i prezzi di equilibrio:

$$p_1^* = \frac{1 - \theta_{12}}{2 - \theta_{12} - \theta_{21}} \quad e \quad p_2^* = \frac{1 - \theta_{21}}{2 - \theta_{12} - \theta_{21}}$$

e il profitto di equilibrio:

$$\Pi^* = \frac{1}{2 - \theta_{12} - \theta_{21}}$$

Confrontando quanto in breve introdotto, è possibile ritenere che:

- se i data brokers agiscono in modo indipendente, la piattaforma proietterà la sua scelta sul valore del file data broker offerto con il prezzo migliore, in relazione alla qualità delle informazioni;
- se i data brokers agiscono in cooperazione, la piattaforma proietterà il suo interesse sulla necessità di ottenere il massimo valore da entrambi i file offerti al prezzo più vantaggioso per i suoi profitti. I data brokers si preoccuperanno di minimizzare i costi e di contenere il rischio della concorrenza;
- se un data broker agisce in regime di monopolio su un segmento di mercato, offrendo il suo file come altamente specialistico per la qualità delle informazioni, la piattaforma proietterà il suo interesse sulla necessità di acquistarlo ad un costo da confrontare con le sue prospettive di mercato. La piattaforma confronterà il potenziale extra-surplus ottenibile con il costo del file broker. La decisione finale sarà il risultato di tale trade-off.

Il surplus, in un regime di cooperazione, è dato da:

$$SC = SC_1 + SC_2$$

Dove:

$$\Delta_{SC} = SC_M - SC_{CP}$$

rappresenta la differenza tra il surplus che la piattaforma estrae dal file broker in caso di monopolio SC_M e il surplus in caso di concorrenza perfetta SC_{CP} .

La piattaforma sceglie il monopolio perfetto, se e solo se, il valore SC_M estratto è superiore SC .

La difficoltà maggiore della modellazione è imputabile alla definizione rigorosa della differenza di valore SC_M e SC_{CP} .

Il modello di Evans è preso in considerazione per la sua applicabilità al mercato delle transizioni delle carte di credito. Non sono state considerate le ipotesi alla base della costruzione del modello. Si parte dall'aspetto matematico, la sola formulazione, con l'intento di costruire una funzione obiettivo che consenta di sviluppare un modello tecnico economico previsionale, specifico per il mercato nuovo e complesso dei data brokers. Per la creazione della variante si ritiene necessario l'affiancamento con l'esperto di economia in analisi dei mercati.

L'affiancamento nella fase di ricerca e sviluppo è importante per applicare progressive modifiche al modello di Evans, definendo nuovi parametri e nuovi vincoli. Oggi, la definizione dei modelli attraverso la funzione obiettivo rappresenta la tecnica della ricerca operativa preferita in numerosi settori per ottenere miglioramenti di performance.

Nella modellazione mostrata, gli utenti finali non sostengono il costo della gestione della piattaforma, ma solo il costo del prodotto e/o servizio della transazione in un mercato di acquirenti (B) e venditori (S)/brokers.

Gli acquirenti sono rappresentati dall'eccedenze b_B associate ad una specifica transazione, mentre l'eccedenza lorda b_S è attribuibile ai venditori/brokers.

La transazione/vendita del file broker ha un valore $c \geq 0$ associato ad un determinato prezzo p .

Considerando il numero di venditori N_S , la funzione di domanda degli acquirenti è data da:

$$N^B = Pr(b^B \geq p^B) = D^B(p^B)$$

ed è indipendente dal numero di venditori/brokers.

Mentre:

$$N^S = Pr(b^S \geq p^S) = D^S(p^S)$$

denota la domanda dei venditori su piattaforma.

Pertanto, il valore del file data broker è tale da massimizzare il profitto totale:

$$\pi = (p^B + p^S - c) D(p^B) D^S(p^S)$$

È intuitivo notare che il massimo è ricavato risolvendo le seguenti condizioni di primo ordine:

$$\frac{\partial(\log \pi)}{\partial p^B} = \frac{1}{p^B + p^S - c} + \frac{(D^B)'}{D^B} = 0$$

$$\frac{\partial(\log \pi)}{\partial p^S} = \frac{1}{p^B + p^S - c} + \frac{(D^S)'}{D^S} = 0$$

In particolare:

$$(D^B)' D^S = D^B (D^S)'$$

rappresenta la condizione che caratterizza i valori di p_B e p_S che massimizzano il valore per il prezzo totale p .

Ipotizzando che:

- p_B sia il prezzo degli acquirenti,
- p_S il prezzo dei venditori/brokers,

la relazione tra essi denota una opposizione di fase.

Pertanto, valutando una condizione di vincolo, il surplus derivante da ogni broker è dato, per una data transazione/file broker, da:

$$V^k(p^k) = \int_{p^k}^{+\infty} D^k(t) dt$$

per $k \in \{B, S\}$.

La quantità:

$$W = V^S(p^S)D^B(p^B) + V^B(p^B)D^S(p^S)$$

massimizza il vincolo:

$$p^B + p^S = c$$

Dalla condizione:

$$\frac{\partial W}{\partial p^B} = \frac{\partial W}{\partial p^S}$$

si ottiene:

$$V^S(D^B)' - D^B D^S = - D^S D^B + V^B(D^S)'$$

Stiamo cercando di adattare il modello generale di Rochet e Tirole (2006), sviluppato per il mercato delle carte di credito, al mercato del file broker.

Pertanto, si vuole adeguare la condizione:

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}$$

ed in particolare la sua variante:

$$\frac{p^B}{\eta^B} \left[\frac{V^B}{D^B} \right] = \frac{p^S}{\eta^S} \left[\frac{V^S}{D^S} \right]$$

considerando il surplus attribuibile a venditori/broker ed acquirenti.

Sarebbe opportuno approfondire ulteriormente tale aspetto.

4.3 Aspetti caratterizzanti il mercato

Nel mercato dominato dalle piattaforme multimediali, l'acquisto del file data broker determina le dinamiche tra i concorrenti, favorendone le strategie attraverso l'analisi di previsione dell'incompatibilità o compatibilità per l'integrazione e/o cooperazione. Nel contesto attuale, la tecnologia rappresenta una risorsa importante su cui investire per sviluppare il file data broker.

Il suo costo è determinato dal crescente interesse ad estrarre i profitti dal surplus di valore, indotto dal contesto di impiego, dove i dati vengono impiegati per la creazione di pubblicità. Quest'ultime sono rivolte a catturare specifici consumatori di prodotti e/o servizi.

Si mira, soprattutto, ad acquisire consumatori attraverso l'osservazione delle caratteristiche comportamentali considerandone la reattività alle offerte di prezzo.

Quanto in breve introdotto, è realizzabile mediante le risorse tecnologiche dei Big Data, che consentono di analizzare il file data broker in tempi ridottissimi con metodologie precise e mirate alla classificazione dei consumatori per gusti e per abitudini e preferenze di acquisto.

Le note introdotte sono il risultato della mia personale e quotidiana esperienza come frequentatore e consumatore di Amazon, Prime e Netflix.

I grandi dello shopping online e streaming sviluppano strategie miranti alla creazione di offerta personalizzate, evitando la tendenza a orientare i prezzi.

Infatti personalizzando le offerte riescono ad essere molto efficaci nei confronti del cliente, riuscendo inoltre a fidelizzarlo, aumentando fatturato, vendite e ad avere un impatto positivo anche sul valore percepito del cliente verso l'impresa. Personalizzare un'offerta, nel marketing, significa dare a quel tipo di cliente la soluzione il più adatta possibile alle sue necessità, sia da un punto di vista del prodotto o servizio offerto sia da un punto di vista comunicativo e di tempistiche richieste dal bisogno manifestato.

A tal proposito, si noti che la pratica dei prezzi personalizzati genera benessere solo nella condizione di prezzi tendenzialmente uniformi.

In tale configurazione di mercato, le piattaforme estraggono surplus progressivamente crescente dal venduto, sia in beni sia in servizi, minimizzando le vendite ai consumatori con una bassa disponibilità economica e controllando anche le vendite ai consumatori con una disponibilità più alta.

Nella strategia di definizione della discriminazione dei prezzi, è necessario evitare la transizione dei consumatori informati e tecnologicamente al passo, ad un bene e/o servizio alternativo.

Si ricordi che nella nuova economia del futuro, la gestione dei dati personali deve rispettare le norme sul trattamento e sulla diffusione.

I consumatori dovrebbero essere informati all'atto della raccolta dei dati sulla gestione, che deve avvenire nel rispetto del genere, della razza, della nazionalità, della religione e dell'età. In Europa, secondo la legislazione dell'UE.

Nella pratica della diffusione, la pubblicità rappresenta la risorsa principale per estrarre surplus di valore dai dati del comportamento.

Si praticano attività quali: social media marketing, e-mail marketing, display advertising e Search Engine Marketing, annunci audio o video che compaiono in diverse parti del display della piattaforma, sulla base delle caratteristiche ed interessi personali estratti nella fase di definizione del profilo comportamentale.

Il monitoraggio a seguire, diviene l'elemento di aggiornamento del profilo e del file data broker. Soprattutto, nel caso di prodotto e/o servizio specifico per gusti ed interessi.

Senza alcuna ombra di dubbio, si riconosce la necessità di informare al consenso i consumatori tecnologicamente più avanzati, coinvolti nel processo di proliferazione dei cookies.

Proponiamo di simulare uno scenario di mercato con i consumatori che vengono pagati per l'impiego della loro privacy, da imprese o terze parti, nei processi di diffusione.

In tale contesto, il singolo consumatore diviene un data broker da cui la piattaforma acquista le informazioni.

Il singolo broker impiega la propria privacy come una risorsa di scambio per ottenere una personalizzazione del prezzo.

In sintesi, stiamo ipotizzando uno spacchettamento dei dati del file data broker, dove il singolo consumatore vende la propria privacy riservandosi il diritto di scegliere quali dati cedere.

Uno scenario del tipo andrebbe a rivoluzionare un mercato digitale caratterizzato da:

- segmentazione dei consumatori in gruppi che ricevono pubblicità personalizzate;
- discretizzazione uniforme dei prezzi definita dalla quantità di cookies.

Sarebbe interessante modellare gli effetti derivanti dall'acquisto delle singole unità di consumatori che la piattaforma è disposta a pagare, escludendo le entità valutate di basso valore per le proprie attività commerciali. Alleviando le pratiche della pubblicità per orientarle verso l'alto valore, non ancora, coinvolto.

Una modellazione tipo, richiede la definizione di una funzione obiettivo simile a quanto riportato:

$$\pi(v, q) \triangleq vq - c * m(q)$$

Dove:

- q : rappresenta l'azione rivolta ai potenziali clienti;
- v rappresenta il *match value*;
- $m(q)$ rappresenta la matching cost function. Ovvero, il legame dello spazio pubblicitario con il costo "c" unitario di acquisto per tale spazio.

È buona prassi, nella modellazione, customizzare l'azione q per ogni consumatore con caratteristica i che possiede un certo valore v .

Dal punto di vista del data broker, è necessario modellare il match value del cookie. Per il data provider il match value rappresenta la richiesta di cookies, oppure il cookie stesso. Concludendo, si vendono le informazioni al prezzo unitario p associato a ciascun cookie notando che a ciascun utente possono appartenere anche più cookies.

Conclusioni

Si assiste al proliferare di compagnie di data broker che raccolgono dati ed elaborano informazioni per venderle ad altre aziende.

La crescente rilevanza dei servizi Internet ed il progresso degli strumenti di analisi dei big data, rende possibile ai data broker ottenere informazioni sempre più rilevanti, specifiche e sensibili.

Si evidenzia, dunque, la necessità di porre trasparenza etica ai servizi offerti da queste compagnie. Specialmente nel caso di dati generati quotidianamente da carte di credito, smartphone e dispositivi indossabili.

Oramai, le informazioni della nostra individualità alimentano un'industria multi-miliardaria.

In certi contesti, anche i dati comportamentali sotto forma di una cronologia di azioni che un utente, anonimo o identificato, ha compiuto durante la sua permanenza su un sito web, sono da considerarsi dati personali.

Tutti questi dati, una volta estratti, sono venduti, scambiati e integrati con quelli di altre aziende partner per accrescere il coinvolgimento degli utenti e la fedeltà verso un marchio attraverso campagne pubblicitarie atte ad influenzare la vita online dei consumatori.

In sintesi, le grandi moli di dati chiamati big data, possono essere ottenuti con facilità monitorando le tracce digitali degli individui.

Il modello di business più comune per la raccolta di dati è quello basato sull'offerta di servizi o prodotti gratuiti.

Concludendo, il mercato digitale è animato da una grande quantità di dati comportamentali, spesso, raccolti senza un consenso informato da parte degli utenti ma attraverso il tracciamento online di visite su portale governativo, commerciale e pubblico.

BIBLIOGRAFIA

- OECD (2017), Algorithms and Collusion: Competition Policy in the Digital Age
www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm
- Competition in Pricing Algorithms di Zach Y. Brown Alexander MacKay, University of Michigan Harvard University, novembre 2019
- J. Crémer, Y.-A. de Montjoye, H. Schweitzer (2019), Politica di concorrenza per l'era digitale: Relazione finale, pag. 83,84
- Furman, J., Coyle D., Fletcher A., Marsden P. & McAuley, D. (2019). Unlocking digital competition: Report of the digital competition expert panel. UK government publication, HM Treasury
- Lyskey, O. (2017). Regulating Platform Power, LSE Law, Society and Economy Working Papers 1/2017
- Lyskey, O. (2019). Grappling with ‘Data Power’: Normative Nudges from Data Protection and Privacy, *Theoretical Inquiries in Law*, 189–220 Marsden
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293– 310. <https://doi.org/10.1177/1461444816661553>
- Plantin, J.-C. & Punathambekar, A. (2019). Digital media infrastructures: pipes, platforms, and politics, 41(2) *Media, Culture & Society* 163-174
- Moody, J. (2019). “Why Colleges Look at Students’ Social Media”, U.S. News & World Report, <https://www.usnews.com/education/best-colleges/articles/2019-08-22/why-colleges-look-at-students-social-media-accounts>
- Scantamburlo, T., Charlesworth, A., Cristianini, N. (2019). “Machine Decisions and Human Consequences”, University of Bristol
- Oswald, M., Grace, J., Urwin, S., C. Barnes, G. (2018). “Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality”, *Information & Communications Technology Law*, 27, 223-250
- Rahman, M.M., Saha, A.R. (2019). “A Comparative Study and Performance Analysis of ATM Card Fraud Detection Techniques”, *Journal of Information Security*, 10, 188-197, <https://doi.org/10.4236/jis.2019.103011>.

SITOGRAFIA

- <https://managementcue.it/bigdata/9243/>
- <https://www.ilsole24ore.com/art/scandalizzati-facebook-vostri-dati-sono-vendita-anni-AEocgTXE>
- <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/3215070>
- https://blog.osservatori.net/it_it/big-data-cosa-sono
- https://blog.osservatori.net/it_it/big-data-e-tutela-dei-dati-personali-%C3%A8-possibile
- <https://www.specialistidellaprivacy.it/quanto-costano-nostri-dati-personali/>
- <https://www.infodata.ilsole24ore.com/2017/05/04/quanto-vale-mercato-europeo-dei-dati/>
- <http://bruegel.org/2016/11/search-engines-big-data-and-network-effects/>
- <https://www.idmerit.com/global-coverage/>
- <https://www.acxiom.com/what-we-do/data/>
- <https://developer.myacxiom.com/code/api/data-bundles/main>
- <https://developer.myacxiom.com/code/api/endpoints/portrait>
- <https://developer.myacxiom.com/code/api/data-bundles/level1bundle/demographics>
- <https://developer.myacxiom.com/code/api/data-bundles/bundle/personicxFinancial>
- <https://www.bloomberg.com/press-releases/2019-05-14/identity-verification-market-worth-12-8-billion-by-2024-exclusive-report-by-marketsandmarkets>
- <https://www.idmerit.com/behavior-monitoring/>
- <https://www.peakyou.com/>
- <https://www.tlo.com/social-media>
- <https://lists.nextmark.com/market?page=order/online/datacard&id=189752>
- <https://lists.nextmark.com/market?page=order/online/datacard&id=243005>
- <https://lists.nextmark.com/market?page=order/online/datacard&id=142394>
- <https://lists.nextmark.com/market?page=order/online/datacard&id=315873>
- <https://liveramp.com/blog/customer-data-onboarding/>
- https://www.braze.com/docs/help/best_practices/news_feed/
- <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-are-identifiers-and-related-factors/>
- <https://www.lotame.com/back-basics-data-onboarding/>
- <https://www.ibm.com/ibm/history/ibm100/us/en/icons/crimefighting/transform/>

RINGRAZIAMENTI

Ringrazio chi si è alzato presto la mattina per preparare qualcosa da portare il giorno dopo, chi tante volte ha ascoltato le mie lamentele, chi a mia insaputa ha infilato 50€ nei pantaloni, chi ha preferito uscire di casa pur di farmi studiare in silenzio, chi ha risparmiato pur di non farmi mancare niente, chi ha lavorato sodo per assecondare tutti i miei desideri, chi ha messo tutto se stesso per garantirmi un futuro, chi mi ha sempre detto: “piano piano” e chi “non rimandare mai quello che puoi fare oggi a domani”, chi mi ha detto: “alzati due ore prima altrimenti perdi il pullman” E chi mi ha comprato la macchina esasperato dei miei ritardi e desideroso dei miei divertimenti.

Ringrazio Mamma e Papà

Ringrazio chi ha pensato sempre che l’università è un posto dove fare belle conoscenze, andare a ballare e ubriacarsi.

Ringrazio mio Fratello

Ringrazio chi prima di un esame mi dedicava preghiere e rosari e chi mi dava uova fresche, pollo paesano, frutta e verdura di stagione.

Ringrazio Nonna e Nonno

Ringrazio chi ha capito i miei problemi e fatto di tutto pur di non farmeli pesare, chi ha posticipato i suoi impegni in base alle mie giornate libere e chi dopo un litigio prendeva il primo pullman e veniva sotto casa.

Ringrazio la mia ragazza

Ringrazio i miei amici di sempre che quando tornava a casa erano sempre pronti a “scroccare” e a fare serata insieme.

Ringrazio i miei compagni di università, tutti, per il bene che ci siamo voluti e per le serate passate a bere a cantare.

Ringrazio tutti voi per avermi reso la persona che sono oggi. Grazie di cuore.