LUISS

Dipartimento di Impresa e Management
Corso di Laurea Magistrale in Marketing
Cattedra di Machine Learning & Object Driven Marketing

La Privacy ed il Machine Learning: Sentiment Analysis sulla percezione della cessione di dati a terzi e come varia il comportamento d'acquisto con stimoli di Recommendation System

Chiar.mo Prof. Luigi Laura
RELATORE

Chiar.mo Prof. Giuseppe Francesco Italiano
CORRELATORE

Francesco Moscatello - Matricola 703471
CANDIDATO

Anno Accademico 2019/2020

Alla mia famiglia, ai miei cari ed a tutti coloro che mi hanno permesso di essere qui. A chi c'è, a chi c'è stato ed a chi ci sarà

Abstract:

Il *Machine Learning* sembra essere sempre più uno strumento utile in diversi ambiti tra cui indubbiamente anche il *Marketing*. Infatti questo tipo di applicazione di Intelligenza Artificiale lo vediamo sempre più spesso utilizzato da importanti aziende come, per citarne alcune, *Amazon* o *Netflix* che si avvalgono, ad esempio, del cosiddetto *Recommendation System* per orientarsi in maniera più conforme al need di ogni specifico consumatore. Inevitabilmente ogni impresa per essere maggiormente customer-oriented, soprattutto se vuole avvalersi di questi strumenti, necessita dei dati dei consumatori che, come si apprende dalla letteratura recente, rappresentano un bene molto delicato per gran parte della popolazione mondiale. Il diritto alla privacy, infatti, potrebbe costituire un limite alle applicazioni di *Machine Learning*. Pertanto questo studio si propone di verificare che sentiment hanno i consumatori circa la cessione di dati a terzi e se delle applicazioni del *Machine Learning*, in particolare riferibili al *Recommendation System*, influenzano il comportamento d'acquisto del consumatore verificando se, ed eventualmente in che misura, varia l'*intention to buy* quando vediamo inserzioni pubblicitarie di prodotti che sfruttano i nostri dati. Inoltre , in base a quanto suggerisce la letteratura, è stato ritenuto interessante valutare se varia l'atteggiamento circa la sicurezza ed il rischio percepito online tra i millenials e le categorie più adulte. I due studi principali, infine, mirano a verificare la compatibilità tra il diritto alla privacy e la possibilità, per le aziende, di soddisfare in maniera più precisa i bisogni dei consumatori. Per verificare i suddetti studi è stata effettuata una *Sentiment Analysis* per la prima parte, ed un *paired samples t-test* sulla seconda i cui dati sono stati raccolti attraverso un questionario svolto da 153 rispondenti. Per quanto riguarda, invece, l'approfondimento sulla eventuale differenza di atteggiamento online tra millenials e generazioni precedenti, è stato effettuato un *indipendent samples t-test* i cui dati sono stati collezionati sulla base dello stesso questionario con cui sono stati raccolti i dati per il *paired samples t-test*.

INDICE:

1. **INTRODUZIONE (pag. 6)**

2. IL MACHINE LEARNING NEL MARKETING (pag. 10)

2.1 Machine Learning (pag. 10)

2.2 Clustering (pag. 11)

2.3 Recommendation System (pag. 16)

3. SENTIMENT ANALYSIS (pag. 18)

3.1 Overview (pag. 18)

3.2 Twitter (pag. 19)

3.3 Normativa della Privacy (pag. 19)

3.4 R (pag. 20)

3.5 Natural Language Processing (pag. 21)

3.6 Metodologia (pag. 22)

3.7 Risultati (pag. 23)

3.8 Limitazioni (pag. 24)

4. VALUTAZIONE SPERIMENTALE: QUANTO IL RS INFLUENZA IL CUSTOMER (pag. 26)

4.1 Metodologia (pag. 26)

4.1 Data Collection (pag. 26)

4.2 Data cleaning (pag. 29)

4.3 Formulazione delle ipotesi (pag. 30)

4.4 Test delle ipotesi (pag. 31)

4.2 Risultati (pag. 35)

4.3 Implicazioni manageriali (pag. 39)

4.4 Limitazioni e spunti per ricerche future (pag. 40)

4.5 Micro-momenti (pag. 41)

5. CONCLUSIONI (pag. 42)

BIBLIOGRAFIA (pag. 48)

SITOGRAFIA (pag. 51)

APPENDICE (pag.51):

A. **SCRIPT R (pag. 51)**
B. **FLOW SURVEY (pag. 53)**
C. **OUTPUT SPSS (pag. 117)**

ELENCO DELLE FIGURE:

2.1 Distanza minima tra cluster

2.2 Proprietà distanza (fonte: ″Analisi dei dati e data mining per decisioni aziendali", Sergio Zani e Andrea Cerioli, 2007)

2.3 Formula distanza euclidea (fonte: ″Analisi dei dati e data mining per decisioni aziendali", Sergio Zani e Andrea Cerioli, 2007)

2.4 Distanza Euclidea e di Manhattan (fonte: ″Analisi dei dati e data mining per decisioni aziendali", Sergio Zani e Andrea Cerioli, 2007)

2.5 Dendrogramma

3.1 Prima raccolta BoW

3.2 Dataframe dopo stopwords

3.3 Output Sentiment Analysis lessico Bing (sx) e lessico NRC (dx)

4.1 Preferenze dei consumatori

4.2 Distribuzione t e z

4.3 Code test delle ipotesi

4.4 Acquisti dopo inserzione pubblicitaria

4.5 Manipulation Check libri

4.6 Manipulation Check videogiochi

4.7 Manipulation Check hotel

4.8 Grafico a torta genere rispondenti

4.9 Istogramma professioni rispondenti

4.10 Grafico a torta età rispondenti

CAPITOLO 1

INTRODUZIONE

Negli ultimi anni il *Marketing*, e non solo, è stato rivoluzionato in maniera importante grazie all'avvento della *Digital Transformation* che rappresenta oggi una grandissima risorsa ed opportunità per tutte le aziende. La *Digital Transformation* può essere considerata come "*una vera e propria leva capace di abbattere gli intermediari privi di valore strategico, e di massimizzare le relazioni alla base dei nuovi modelli di business, che necessita di una rinnovata ed ancora più profonda consapevolezza delle nuove logiche del mercato alla base della relazione tra domanda ed offerta, dei processi interni su cui poggiano le imprese e dell'impatto che il digitale comporta per le persone coinvolte in questi processi*" (Marco Ronchi & Marianna Ciancia, 2019). Partendo da questo assunto, indubbiamente tutte le innovazioni digitali costituiscono uno strumento imprescindibile per ogni impresa non solo perché queste possono sfruttare le nuove tecnologie per i processi interni ma anche e soprattutto perché 4,54 miliardi sono le persone connesse ad internet di cui 3,8 mld utilizzano regolarmente *social network*[1] e quindi sono più facilmente reperibili i dati di ogni consumatore. I *Big Data*[2], infatti, sono considerati una *commodity* che consente alle imprese di offrire ai consumatori servizi innovativi (Giovanni Pitruzzella, 2016). Tuttavia una forte limitazione alla possibilità di usufruire dei *Big Data*, per le imprese, e di beneficiare di offerte maggiormente *customer-oriented*[3] da parte di queste, per i consumatori, è indubbiamente la *Privacy* che come suggerisce gran parte della letteratura è un diritto a questi ultimi molto caro specie dopo alcuni avvenimenti come ad esempio lo scandalo di *Cambridge Analytica* che ha generato per *Facebook* la perdita di più di 3 milioni di utenti in Europa e circa 120 miliardi di dollari di valore di mercato (Isaak and Hanna, 2018). Tuttavia la letteratura accademica ha evidenziato il cosiddetto "paradosso della *Privacy*" (Norberg et al., 2007) secondo cui gli utenti online dichiarano, da una parte, di essere preoccupati dal raccoglimento ed utilizzo ambiguo da parte delle imprese (e.g. Bansal et al., 2010; Raiini et al., 2013; TRUSTe, 2014, Smith H.J. et al., 1996), dall'altra gli stessi utenti si mostrano disponibili a cedere i propri dati per compensi di poco conto come ha dimostrato Carrascal et al. (2013) che ha rilevato

[1] (Report Digital 2020 accessibile da:"https://wearesocial.com/it/blog/2020/01/report-digital-2020-i-dati-global" condotto da We Are Social in collaborazione da Hootsuite).

[2] all'inizio degli anni 2000 l'analista di mercato Doug Laney ha definito i big data con le tre V ovvero Volume (le organizzazioni raccolgono dati da diverse fonti come ad esempio dispositivi intelligenti nell'ambito IoT*, social network, transazioni commerciali ecc.), Velocità (grazie alla crescita dell'IoT i flussi di dati vengono dati alle imprese velocemente) e Varietà (i dati sono disponibili in tutti i tipi di formati).

*Per IoT si intende un sistema di dispositivi informatici correlati, macchine e digitali, oggetti, animali o persone dotati di identificatori univoci (uid) e la capacità di trasferire dati su una rete senza richiedere da uomo a uomo o interazione uomo-computer (V. Teigens, P. Skalfist, D. Mikelsten, 2019).

[3] L'orientamento al consumatore è la base dell'apprendimento organizzativo che produce come effetto una percezione del valore ed una soddisfazione del clienti maggiori dell'azienda che utilizza questo approccio rispetto ai competitors (Singula, J. M., Baker, W. E., e Noordewier, T., 1997; Slater, S. F. e Narver, J. C., 1995; Brady, M. K.; Cronin Jr., J. J., 2001).

che diversi utenti hanno accettato di valutare la propria cronologia di navigazione online 7 euro cioè quanto un *Big Mac*. In definitiva questo studio si propone di offrire alla letteratura un nuovo contributo circa la percezione degli utenti online della cessione di dati personali a terzi attraverso una *Sentiment Analysis* condotta considerando i *tweets* relativi al suddetto topic sul *social network Twitter* che vanta 336 milioni di utenti attivi mensili, 500 mln di *tweets* pubblicati ogni giorno e 1,3 miliardi di account *Twitter* di cui però 328 mln attivi (Matt Ahlgren, 2020). Inoltre la seconda parte della ricerca avrà come scopo quello di verificare come l'applicazione di algoritmi di *Machine Learning*, nello specifico il *Recommendation System* che per essere efficace necessita di una grande mole di dati, influisce sul marketing in termini di intention to buy. Questa seconda parte verrà analizzata sfruttando un esperimento condotto tramite un sondaggio lanciato attraverso il *software Qualtrics* dove verrà valutato come varia il comportamento d'acquisto dei consumatori sotto stimoli di *Recommendation System*. Il *Recommendation System* è quel sistema che sfruttando algoritmi di *Machine Learning*, in base ai dati in suo possesso, suggerisce agli utenti o consumatori, un determinato *output* (maggiori approfondimenti in questo ambito verranno forniti successivamente nel paragrafo dedicato 2.3). Questa seconda parte della ricerca parte dall'assunto che il *Machine Learning* (d'ora in poi ML) fornisce un importante supporto al *Marketing* (S. Zollo et al., 2006; Zhou, R., Khemmarat, S. e Gao, L, 2010). In particolare gli autori del secondo studio hanno analizzato che impatto ha il *Recommendation System* utilizzato da *YouTube* sulle visualizzazioni dei video all'interno della stessa piattaorma. All'interno di questa ricerca è stato rilevato che circa il 30% delle visualizzazioni complessive derivavano da una raccomandazione del relativo video per cui è possibile che i video analizzati, totalmente o in parte, siano stati visti grazie al fatto che siano stati suggeriti. Inoltre, un ulteriore finding di questo studio suggerisce che il *Recommendation System* incentiva a visualizzare video completamente diversi rispetto a quello alla fine del quale viene mostrato il video suggerito e questo dimostra che questa applicazione di ML permette di conoscere meglio i consumatori e di offrire loro video maggiormente di loro interesse rispetto a video *mainstream*. Partendo dai risultati di questo studio, questa parte della ricerca ha voluto estendere i risultati su azioni più dispendiose della semplice visualizzazione come l'*intention to buy* che è la variabile di riferimento.

Prima di procedere oltre, occorre chiarire ulteriormente cosa si intende per ML, per il quale si possono trovare diverse definizioni approfondite nel paragrafo dedicato (2.1). In generale si tratta di un termine coniato dal 1959 da Arthur Samuel che ha distinto due approcci distinti di questa branca dell'Intelligenza Artificiale (d'ora in poi IA). Anzitutto, quindi occorre prima chiarire cosa si intende per IA e per farlo, poiché è un argomento di recente rilevanza per cui trovare una definizione univoca appare abbastanza complicato, risulta opportuno rifarsi alle definizione di questa da parte di diversi autorevoli autori: "L'obiettivo del soggetto è identificare risolvibili ed interessanti problemi di elaborazione delle informazioni, e risolverli" (David Marr, 1976); "l'intelligente connessione della percezione all'azione" (Rodney Brooks, 1991); "le azioni che non sono

distinguibili da quelle umane" (Alan Turing, 1950). In definitiva, cercando di sfruttare tutte le precedenti definizioni, l'IA può essere definita come lo studio dei problemi di elaborazioni complesse di informazioni che spesso hanno le loro radici in alcuni aspetti di elaborazione di informazioni biologiche. Chiarito questo, è possibile proseguire nella spiegazione dei due metodi con cui opera il ML in cui il primo, identificato come rete neurale, porta a sviluppare delle macchine alle quali viene fornito apprendimento automatico per determinati impieghi generali (unsupervides learning) in cui il comportamento è appreso da una rete di commutazione connessa in modo random a seguito di una routine di apprendimento basata su ricompensa (in caso di successo) e punizione (in caso di insuccesso) (reinforcement learning). Il secondo metodo invece consiste nella programmazione specifica di determinate attività (supervised learning).

In conclusione l'obiettivo generale della ricerca è quello di comprendere che sentiment hanno gli utenti online circa la cessione di dati personali a terzi e se, ed eventualmente in che misura, il Recommendation System (che sfrutta i dati forniti dagli utenti stessi) riesce a dare alle imprese la possibilità di restituire ai propri clienti prodotti maggiormente customer-oriented e quindi più in linea con i needs di questi ultimi. Nell'elaborazione del secondo studio, ho ritenuto inoltre opportuno effettuare altre ricerche correlate che verranno meglio spiegati nel paragrafo 2.4. Perciò le *Research Questions* dell'intero studio sono le seguenti:

- Che *sentiment* hanno le persone circa la cessione di dati personali a terzi online?
- L'*intention to buy* di un prodotto la cui inserzione pubblicitaria viene mostrata ad un determinato consumatore in base ai suoi dati varia in maniera statisticamente significativa rispetto ad una non basata sui dati?
- C'è una differenza statisticamente significativa del tratto personale insicurezza online tra i *Millenials* e le generazioni precedenti?
- Lo stesso discorso dell'ultimo punto può essere fatto in relazione alla variabile "rischio percepito" in termini di sicurezza online?

CAPITOLO 2

IL MACHINE LEARNING NEL MARKETING

In questo capitolo è presente un approfondimento su cosa è il *ML*, su cosa è il *Recommendation System* e quindi su come i suoi algoritmi operano affinché assolvano ai compiti per cui sono programmati. Per questa ragione è anche presente una maggiore delucidazione circa il *Clustering* che rappresenta un aspetto fondamentale per spiegare quali tecniche utilizzano i suddetti algoritmi. Infine viene quindi chiarito come l'*Intelligenza Artificiale* possa costituire per le imprese uno strumento utile dal punto di vista strategico per ciò che concerne il *Marketing*.

2.1 Machine Learning Il *Machine Learning*, come l'*IA* essendo un suo sottoinsieme, è un ambito ancora abbastanza nuovo per cui racchiuderlo in una definizione univoca appare riduttivo. Per tale ragione, per spiegare più chiaramente a cosa ci si riferisce quando si parla di *ML*, ho raggruppato una serie di definizioni *di autorevoli autori:*

"Il Machine Learning è la programmazione di computer per ottimizzare un criterio di prestazione usando dati esemplari o esperienze passate" (Ethem Alpaydin, 2004)

"Il Machine Learning è una branca dell'Intelligenza Artificiale ed è focalizzata sul fare decisioni intelligenti e sul riconoscere di complessi modelli attraverso l'uso dell'esperienza passata e di dati storici" (Sahil Sharma and Vinod Sharma, 2016)

"L'obiettivo del Machine Learning è sviluppare metodi che possono automaticamente rilevare modelli nei dati, e poi usare i modelli scoperti per predire dati futuri o altri outcomes di interesse" (Kevin P. Murhy, 2012).

"Si dice che un programma apprende dall'esperienza E con riferimento a alcune classi di compiti T e con misurazione della performance P, se le sue performance nel compito T, come misurato da P, migliorano con l'esperienza" (Tom Michael Mitchell, 2012)

Chiarita ulteriormente la materia, si comprende facilmente la portata che può avere l'implementazione di algoritmi di *Machine Learning* in un contesto, come quello odierno, nel quale organizzazioni ed imprese si trovano a dover gestire una grandissima mole di dati (*Big Data*). Le applicazioni di questo campo possono essere diverse e sostanzialmente è possibile suddividere il ML in tre tipologie di apprendimento (*Supervised Learning, Unsupervised Learning* e *Reinforcement Learning*). Analizzandole nel dettaglio:

- *Supervised Learning* (Arthur Samuel, 1959): in questo caso la macchina si trova ad effettuare analisi su dati già etichettati dagli esperti per cui se deve gestire dati numerici la macchina procederà ad un problema di regressione mentre nel caso in cui questi siano categorici, ci troveremmo davanti ad un

problema di classificazione. Per chiarire la differenza tra i due punti: nel primo caso l'outcome sarà inevitabilmente una variabile continua mentre nel secondo una categoria categorica. Volendo fare un esempio: con una regressione ad esempio prediciamo quanto vino verrà prodotto in una determinata annata in base alle variabili indipendenti che influenzano questo dato, mentre nel caso di una classificazione si esprime se ad esempio è più probabile ottenere una buona cura da un determinato ospedale o meno (la variabile dipendente in questo caso si configura come una variabile dicotomica del tipo "0" o "1" per cui l'outcome sarà o un risultato o l'altro).

- Unsupervised Learning (Arthur Samuel, 1959): i dati non hanno attributi di destinazione, cioè in questo caso viene effettuata una ricerca esplorativa dei dati per trovare un'intrinseca struttura in essi per cui non ci sarà alcuna fase preliminare di un esperto ma sarà la macchina a cercare di trovare da sola una struttura degli input ricevuti.
- Reinforcement Learning (Leemon Baird, 1995): è una tecnica che presuppone l'utilizzo di input esterni in base alle scelte che effettua l'algoritmo per cui un outcome corretto genererà un premio per la macchina mentre uno sbagliato le restituirà una punizione. Sostanzialmente è un meccanismo trial and error.

Oltre all'approfondimento già fatto su questo paragrafo circa l'apprendimento supervisionato, merita una maggiore spiegazione anche quello relativo all'apprendimento non supervisionato poiché si riferisce a due aspetti molto rilevanti in questo studio quali la *Sentiment Analysis* ed il *Reccomendation System*. L'unsupervised learning infatti trova maggiormente riscontro nelle tecniche di **Clustering** e **Dimensionality Reduction**. Nel prossimo paragrafo (2.2) verrà fatto un ulteriore approfondimento circa la prima tecnica poiché maggiormente pertinente a questo studio siccome il *Recommendation System* la utilizza ma è assimilabile anche a quanto necessario fare per effettuare una *Sentiment Analysis* maggiormente efficace come spiegato nel paragrafo 1.8. Nel primo caso, invece, verrà fatto un approfondimento in seguito (2.3), per ora basti considerare che attraverso il Recommendation System, gli algoritmi di ML, sfruttando i dati in possesso del singolo utente/consumatore, restituiscono output a loro affini. Per fare questo viene effettuata una clusterizzazione dei singoli consumatori in base ai loro dati sensibili e soprattutto storici. La *Dimensionality Reduction* è, invece, il processo di riduzione del numero di variabili casuali in esame ottenendo un insieme di variabili principali (Roweis, S.T.; Saul, L.K, 2000). Gli approcci sono due: *feature selection* e *feature extraction* (Pudil, P.; Novovicova, J., 1998).

2.2 Clustering Risulta particolarmente rilevante per lo studio una spiegazione del Clustering. La *Cluster Analysis* si occupa di trovare gruppi di oggetti in dati in modo tale che gli oggetti in un gruppo saranno simili (o relazionati) tra di loro e differenti (o non relazionati) da quelli degli altri gruppi. Una buona clusterizzazione

è quella in cui le distanze interne al cluster sono minimizzate mentre quelle tra cluster differenti sono massimizzate (fig. 2.1).

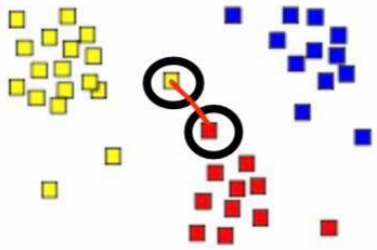


Figura 2.1 Distanza minima tra cluster

Per sua natura, quindi, questa tecnica risulta inevitabilmente interessante per il *Marketing* poiché permette all'aziende di definire con chiarezza i target di riferimento tra i consumatori ed anche perché viene applicata nell'ambito del *Recommandation System* come verrà spiegato nel capitolo 2.3. Questa tecnica ha tra i principali vantaggi quello di ridurre una quantità di dati in gruppi che appaiono simili per cui nell'ambito dei *Big Data* risulta in alcuni casi essenziale per ridurre il grado di complessità di analisi. Il *Clustering* quindi non è nient'altro che un set di cluster ma per operare alla suddivisione dei clusters va specificato che esistono due tecniche:

- Clustering partizionale: secondo questa tecnica, i dati vengono divisi in sottogruppi (che sono appunti i cluster) che non si sovrappongono tra loro in modo tale che ogni oggetto è esattamente e solamente in un sottogruppo (o cluster)

- Clustering gerarchico: è un insieme di clusters nidificati organizzati come un albero gerarchico

Prima di procedere alla spiegazione delle due tecniche va approfondito il calcolo di distanza dei punti. In statistica il concetto di distanza è mutuato dalla geometria ove si fa riferimento alla distanza tra due punti in uno spazio R_p. Procedendo ad una definizione: la distanza tra due punti corrisponde ai vettori x,y appartenenti allo spazio R_p una funzione d(x,y) che gode delle seguenti proprietà[4]:

[4] "Analisi dei dati e data mining per decisioni aziendali", Sergio Zani e Andrea Cerioli, 2007

1) *non negatività*:

$$d(\mathbf{x},\mathbf{y}) \geq 0 \qquad \forall \quad \mathbf{x},\mathbf{y} \in R^p$$

2) *identità*:

$$d(\mathbf{x},\mathbf{y}) = 0 \qquad \Leftrightarrow \quad \mathbf{x} = \mathbf{y}$$

3) *simmetria*:

$$d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x}) \qquad \forall \quad \mathbf{x},\mathbf{y} \in R^p$$

4) *disuguaglianza triangolare*:

$$d(\mathbf{x},\mathbf{y}) \leq d(\mathbf{x},\mathbf{z}) + d(\mathbf{y},\mathbf{z}) \qquad \forall \quad \mathbf{x},\mathbf{y},\mathbf{z} \in R^p$$

Figura 2.2 Proprietà distanza

Il metodo più famoso per calcolare la distanza è la Distanza Euclidea la cui formula è la seguente:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{ik} - x_{jk})^2}$$

Figura 2.3 Formula distanza euclidea

Dove d è la distanza, i e j sono due punti, x è una variabile indipendente e k è il numero di variabili indipendenti.

Altri metodi in questo ambito sono:

- Distanza di Manhattan: $_1dij=\Sigma^{p}_{s=1}|x_{is}-x_{js}|$. Per cui viene considerata la somma delle variabili in valori assoluti anziché dei loro quadrati. Questa distanza e quella euclidea si possono ottenere da una formula più generale ovvero la distanza di Minkowski. Si dice distanza di Mikowski di ordine k tra i punti i e j l'espressione seguente: $_kd_{ij}=[\Sigma^{p}_{s=1}|x_{is}-x_{js}|^k]^{1/k}$ con $k\geq1$ (per s si intende valore ennesimo generico). Da questa equazione appare evidente che la distanza euclidea corrisponde a quella di Mikowski con k=2 (perciò nella formula prima della d c'è 2) mentre quella di Manhattan con k=1 (perciò nella formula prima di d c'è 1). Questo tipo di distanza è anche chiamato "distanza della città a blocchi poiché poiché considera che tra due punti siano possibili solo movimenti con direzioni parallele agli assi come avviene in una città con griglie regolari di strade che si intersecano ad angolo retto come ad esempio Manhattan (fig.2.4).
- Distanza di Lagrange: $\lim_{k->\infty}{}_kd_{ij}=\max_s|x_{is}-x_{js}|=_\infty d_{ij}$. Pertanto considera solamente misure per cui i punti si distanziano maggiormente

- Distanza di Canberra: $_cd_{ij} = \Sigma^p_{s=1} [|x_{is}-x_{js}|/(x_{is}+x_{js})]$ proposta da Lance e Williams (1966, "A general theory of classificatory sorting strategies). E' una distanza utilizzabile solo nel caso in cui le variabili assumono solo valori positivi poiché in presenza di valori negativi non viene soddisfatta la disuguaglianza triangolare. Il vantaggio di questa distanza deriva dal fatto che in essa si considerano somme di rapporti che sono numeri puri. Perciò se le variabili sono espresse in diverso ordine di grandezze e unità di misura non è necessario standardizzare. Inoltre questa distanza è poco sensibile all'asimmetria della distribuzione delle variabili ed anche alla presenza di outlier.

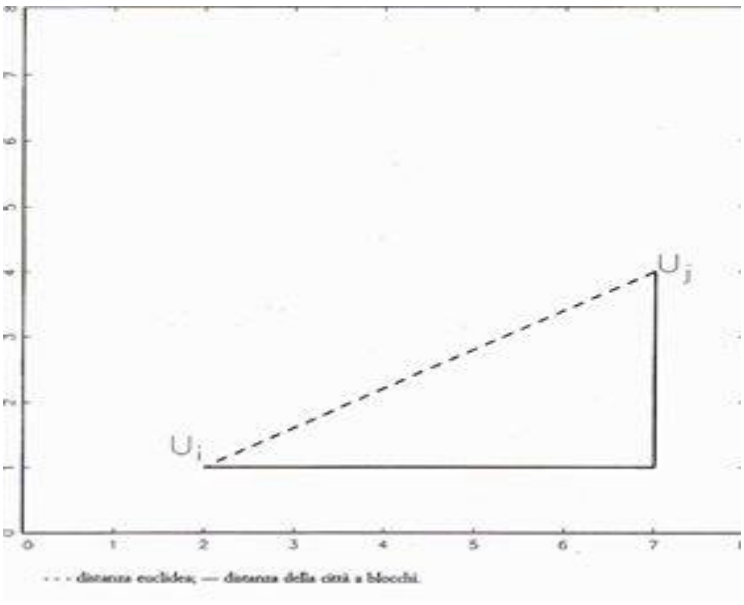


Figura 2.4 Distanza Euclidea e di Manhattan

Chiarito questo, è possibile calcolare diverse distanze tra i clusters:

- Massima distanza: la distanza tra i clusters corrisponde alla distanza tra i due punti più lontani dei rispettivi clusters
- Distanza dei centroidi: distanza tra i centroidi dei cluster. Il centroide è il punto che ha la media di tutti i punti in ogni componente
- Minima distanza: distanza tra i due punti più vicini dei cluster.

.Un problema che appare abbastanza comune nel processo di clustering è che la distanza è fortemente influenzata dalle misurazioni delle variabili che possono essere diverse per cui risulta opportuno effettuare una **normalizzazione dei dati**.

A questo punto è possibile procedere alla spiegazione della procedura dei due modelli di clustering citati prima. In quella gerarchica, nel primo step ci troviamo di fronte ad una situazione in cui ogni punto appartiene al proprio cluster, cioè ogni punto è separato e indipendente dagli altri. Nel secondo step si uniscono i due cluster

(che coincidono con i punti nel primo step) più vicini calcolati secondo la distanza euclidea o quella del centroide. Negli step successivi si ripete l'operazione fino a quando tutti i punti non si uniscono con quello più vicino, per in ogni passaggio si vede quali sono i due cluster più vicini e vengono entrambi raggruppati nello stesso cluster. Successivamente pe effettuare la partizione ottima occorrerà, quindi, il supporto grafico di un dendrogramma (grafo ad albero utilizzato per visualizzare la somiglianza nel processo di "raggruppamento") per cui in corrispondenza di un asse verticale vengono inseriti i numeri che corrispondono alle distanze tra i punti (o cluster), su quella orizzontale, invece, c'è la distanza logica dei clusters secondo la metrica definita. Il supporto grafico è utile poiché consente di vedere materialmente dove appare evidente effettuare il cut-off per cui si effettua la partizione ottima dei clusters. Tendenzialmente il taglio viene effettuato di solito in prossimità del salto poiché l'altezza rappresenta la devianza between-groups e quindi conviene farlo quando la varianza è massima. (fig. 2.5)

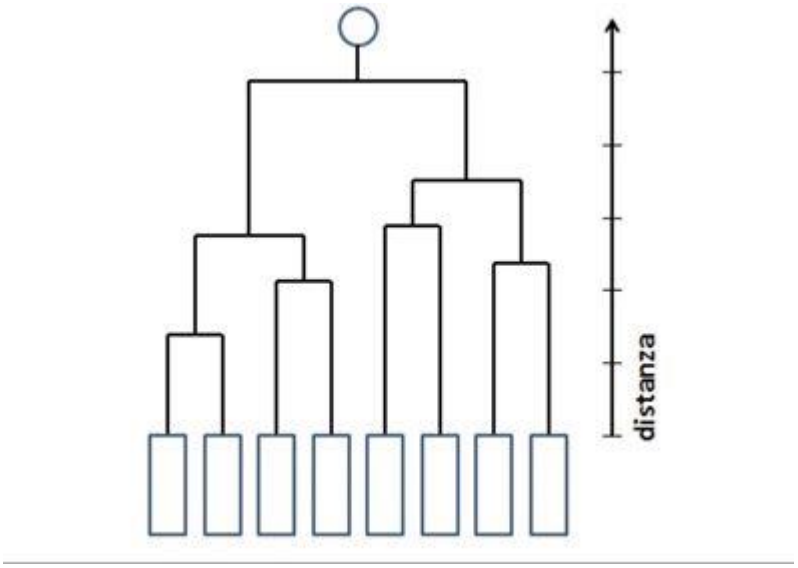


Figura 2.5 Dendrogramma

In alternativa a quello gerarchico c'è il criterio partizionale che trova la sua maggiore affermazione nel k-means clustering (l'altro noto metodo di questa categoria è lo spectral Clustering). Questa tecnica prevede 6 step:

1. Nel primo viene specificato il numero desiderato di cluster k (presupposto totalmente assente nell'altra metodologia)
2. Si assegna in maniera random ogni punto ad un cluster
3. Si calcolano i centroidi di ogni cluster
4. Si riassegna ogni punto al centroide del cluster più vicino
5. Si ricalcolano i centroidi
6. Infine si ripetono gli ultimi due step fino a quando il sistema non presenterà alcun miglioramento.

2.3 Recommendation System Si definisce *Recommendation System* un *software* di filtraggio che, in base ai dati in suo possesso su un determinato utente, crea per lui delle raccomandazioni personalizzate estremamente specifiche (Steven A. Shaya et al., 2010). Alcuni esempi di questa applicazione li troviamo quando vediamo i prodotti suggeriti da *Amazon*, le ricerche suggerite da *Google*, i *tweets* promossi (come già accennato nel capitolo 1.2), i film o le serie consigliate su *Netflix*. Sono già quindi diverse le importanti aziende che sfruttano questi algoritmi per uniformarsi maggiormente ai needs dei clienti e soprattutto per implementare una strategia comunicativa più efficace, confermando anche i risultati che emergono in questo studio (2.5). Le metriche che vengono utilizzate più spesso per verificare l'efficacia di questi algoritmi soprattutto in ambito manageriale sono:

- *customer satisfaction*: che cerca di calcolare la soddisfazione del consumatore rispetto all'output suggerito
- l'efficacia di previsione: che ha come obiettivo quello di stimare quanti suggerimenti hanno effettivamente portato i consumatori all'azione
- predizione dell'errore: che ha come obiettivo di predire, al contrario della metrica precedente, quanto il suggerimento non ha trovato riscontro effettivo. E' possibile calcolarlo per quelle coppie (u,i) per cui abbiamo sia la previsione che la valutazione effettiva. Si utilizzano principalmente due metriche a riguardo quali la distanza di Hamming[5] e la radice dell'errore quadratico medio (ci dà la misura della varianza interna ovvero del rapporto fra la devianza interna e la numerosità totale) la cui formula è la seguente: $\sqrt{\{\Sigma_{(u,i)}[(r_{ui}- \hat{r}_{ui})^2/C]\}}$ (C corrisponde alle coppie precedentemente descritte). In parole povere misura quanto i valori stimati si discostano da quelli effettivi calcolando la radice quadrata della media degli errori al quadrato. Cioè considera la media degli scostamenti tra valore stimato ed effettivo al quadrato (in modo da dare maggior peso ad ogni scostamento) dopodiché calcola la radice quadrata di questo valore. Ovviamente più basso l'errore quadratico medio, più risulta performante il Reccomendation System.

Gli algoritmi di *Reccomendation System* seguono tre tipologie di logiche e per chiarire la differenza tra questa, sono stato introdotti esempi considerando il caso di utilizzo di *Netflix*:

- *Collaborative filtering* (J.B. Schafer. D. Frankowski, J. Herlocker", S. Sen; 2007): se ad esempio i dati in mio possesso di due utenti (Tizio e Caio) su come hanno votato determinate serie li identifica come

[5] La distanza di Hamming tra due tringhe di uguale lunghezza corrisponde al numero di posizioni in cui i simboli corrispondenti sono diversi, cioè questa metrica misura quante sostituzioni sono necessarie affinché una stringa possa convertirsi nell'altra. Perciò possiamo definirla come calcolo del numero minimo di errori che possono esserci stati nel processo di trasformazione di una stringa nell'altra (Richard W. Hamming, 1950)

simili, se Tizio risulta gradire più della media la serie "*Games of thrones*", allora gli algoritmi suggeriranno a Caio questa serie

- *Content filtering* (Y. Koren, R. Bell, C. Volinsky, 2009): sostanzialmente questo sistema di filtraggio si focalizza maggiormente sulla similarità dei contenuti tra utenti per cui, facendo un esempio, se il sistema identifica che un utente, attraverso il rating che questo ha dato dei vari film che ha visto, gradisce più della media "*Pulp fiction*", otterrà come film suggerito "*Kill Bill*" che è un altro film di Quentin Tarantino (stesso regista di "Pulp Fiction)
- *Hydrid Reccomendation System* ("Hybrid web recommender systems", R. Burke, 2007): mischia i due sistemi di filtraggio precedenti. Se ad esempio dai dati in mio possesso su due utenti (Tizio e Caio) risulta che questi hanno preferenze simili, se a Tizio piace più della media la sitcom "*How I met your mother*", allora suggerirò a Caio "*Friends*" che appartiene allo stesso genere e lo stesso discorso può essere fatto al contrario (partendo cioè dai dati di Caio per generare un suggerimento a Tizio). *Netflix* utilizza questa tecnica.

Gli esempi sono basati sul caso specifico di *Netflix*, ma è lapalissiano che le stesse logiche possono essere considerate anche per prodotti di altre categorie merceologiche. Infine proprio siccome gli algoritmi di *Reccomendation System* seguono le logiche appena spiegate, sfrutta le tecniche di clusterizzazione (2.2) secondo algoritmi di ML.

CAPITOLO 3

SENTIMENT ANALYSIS

Questo capitolo è dedicato all'approfondimento dello studio di *Sentiment Analysis* su *Twitter* sul *topic* della *Privacy*. Per tale ragione, è anche presente una maggiore esplicazione della normativa del *topic* di riferimento su cui è possibile trovare alcuni approfondimenti (paradosso della *Privacy*) anche nell'Introduzione. Sono, inoltre, presenti chiarimenti sul tipo di analisi effettuata, sul programma utilizzato per compierla, sul social network utilizzato come campo d'analisi e, ovviamente, sulla metodologia, le limitazioni ed i risultati dello studio.

3.1 Overview Una buona parte dei dati di cui dispongono le aziende sono strutturati, numerici o categorici e le tecniche utilizzate per analizzare queste tipologie di dati sono tutte abbastanza avanzate. Tuttavia L'IA ha cercato di sviluppare ulteriori tecniche per analizzare anche dati qualitativi, poco strutturati, testuali o multilingue come ad esempio i contenuti che vengono pubblicati sui social networks. Sostanzialmente la differenza consiste nel fatto che i dati strutturati possono essere memorizzati in un database SQL (Structured Query Language) (Elmasri, 2010) mentre quelli non strutturati sono difficilmente assimilabili dagli algoritmi utilizzati dalle macchine. Un esempio in questo senso sono i tweets, ovvero i contenuti che gli utenti di Twitter pubblicano attraverso questo social network. Dati di questo tipo potrebbero fornire alle imprese, e non solo, informazioni molto interessanti (basti vedere quanto scritto nell'introduzione circa il numero di tweets pubblicati al giorno) ed è per questo che si sono sviluppate tecniche come ad esempio la Sentiment Analysis che si occupa proprio di analizzare questo tipo di dati che per loro natura richiedono un grado di complessità maggiore. Questo tipo di analisi rientra nel campo del Machine Learning (che verrà approfondito in seguito nel capitolo 2.1) per cui viene fornita ad una macchina una metodologia di apprendimento automatico che gli consente di imparare ad analizzare i dati che riceve senza input dal programmatore. Sostanzialmente alla macchina viene solamente insegnato come apprendere in maniera automatica in modo che questa sviluppi, attraverso i propri algoritmi, una logica con cui analizzare i dati senza che ci sia l'intervento umano che spieghi alla macchina cosa fare passaggio per passaggio. Il campo specifico che si riferisce all'apprendimento della macchina di comprendere dati testuali, è denominato tecnicamente come *Natural Language Processing*. Questo ambito si propone come obiettivo quello di comprendere e trarre significato dal linguaggio umano che risulta per sua natura ben più complesso di quello informatico. Gli albori di questi ambito sono identificabili nello studio condotto da Alan Turing nel 1950 ("Computing machinery and Intelligence-AM Turing") in cui l'illustre informatico ha proposto un settaggio dell'intelligenza artificiale affinché questa potesse prendere parte ad una conversazione in tempo reale con un umano in modo tale che la macchina interagisca come se fosse un altro uomo, cioè in modo tale che sia indistinguibile rispetto a questo. Applicazioni abbastanza note di questo campo le vediamo oggigiorno ad esempio con *Siri* o *Google Now*. Alla luce di tutto, lo scopo della

Sentiment Analysis di questa ricerca, è quello di comprendere se gli utenti di Twitter hanno sentiment positivo, negativo o neutro circa la privacy e quali emozioni vengono maggiormente evocate tra i vari tweets.

3.2 Twitter Ho selezionato *Twitter* per raccogliere le opinioni degli utenti circa il topic di mio interesse poiché, come già accennato nell'introduzione, è uno dei social network più noti ed utilizzati e poiché, grazie all'utilizzo degli hashtag che è una sua peculiarità, è possibile identificare più facilmente tutti i contenuti relativi ad un determinato argomento. Proprio per questa ragione, come verrà specificato meglio nel capitolo dedicato alla metodologia (1.6), la *data collection* è stata effettuata raccogliendo esclusivamente i tweets che contenessero l'hashtag relativo al topic di riferimento attraverso il tool "*twitteR*" che è un pacchetto del programma *R*. Su questo *social network*, infatti, è prassi comune tra gli utenti, porre alla fine dei propri contenuti delle parole che riguardano l'argomento che stanno trattando nel *tweet* seguiti dal cancelletto "#" (è questo procedimento a renderli *hashtag*, ndr) al fine di rendere possibile far comparire quel determinato contenuto a chi inserisce nella barra di ricerca quella stessa parola. Proprio per questa ragione su questo famoso *social network* è possibile visualizzare una seziona dedicata appositamente alla ricerca dove vengono anche suggeriti gli hashtag di tendenza per cui potremmo dire che *Twitter* identifica gli argomenti più discussi dai suoi utenti e li restituisce come suggerimento a chi visualizza la sezione dedicata alla ricerca. Una volta poi identificati gli hashtag più utilizzati riguardo un determinato topic, diventa prassi comune tra gli utenti utilizzare quello stesso hashtag proprio per dare più visibilità al proprio contenuto. Oltretutto i dati sono più facilmente accessibili grazie alle politiche, specificate nell'informativa della *Privacy* di *Twitter*, relative all'*API*[6], che sono ciò che permettono a due programmi altrimenti incompatibili di interagire, e sono anche più strutturati poiché ogni tweet ha un limite di 140 caratteri, per cui risulta meno complesso trasformarli in metadati che possono essere compresi dalla macchina poiché gli algoritmi di *Machine Learning* prediligono input a lunghezza fissa ben definiti. Proprio per queste dinamiche, *Twitter* rappresenta il *social network* che meglio si presta a questo tipo di analisi.

3.3 La normativa della Privacy Negli ultimi anni la *Privacy* ha acquisito sempre più rilevanza tra le persone tanto da diventare un argomento sempre più chiaramente disciplinato dal legislatore sia in ambito nazionale che internazionale. L'ultimo importante cambio della normativa riguardo la *Privacy* c'è stato nel 2018 con il decreto legislativo del 10 agosto 2018, n.101[7] in cui il legislatore italiano si è allineato al Regolamento UE 2016/679 a tutti noto con l'acronimo *GDPR* (*General Data Protection Regulation*) in vigore dal 25 maggio

[6] l'*Application Programming Interface* è un insieme di tecniche che fornisce un'interfaccia uniforme per l'esecuzione di calcoli della probabilità filogenetica su una varietà di piattaforme hardware di calcolo (Ayres et al. 2012). Sostanzialmente quindi grazie alle *API* un programmatore può far interagire due programmi o piattaforme altrimenti tra loro incompatibili.

[7] Gazzetta Ufficiale accessibile da: https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2018-09-04&atto.codiceRedazionale=18G00129&elenco30giorni=true&fbclid=IwAR3CrKps7QJ65zSQfAeIZU_pl_79otDyiu3Xf4O-aCUgFlLJwaJkzHDyCCg

2018. Tale rivoluzione, intesa a rinnovare quanto stabilito in materia con il precedente decreto legislativo 196/2003[8], ha avuto come scopo quello di modificare l'approccio della tutela della privacy con l'introduzione del principio della responsabilizzazione. Precedentemente, infatti, il codice della *Privacy* principalmente si preoccupava di regolamentare le sanzioni alla violazione di tale diritto ma grazie alla trasformazione digitale che ha portato ad una diffusione su vasta scala di *social network*, sistemi *cloud* e navigazione generica su internet, serviva disciplinare tale materia più secondo un'ottica preventiva piuttosto che reattiva poiché negli ultimi anni ha visto aumentare il proprio grado di complessità per le ragioni elencate. Sostanzialmente secondo la nuova normativa non è più sufficiente adottare le misure minime di sicurezza ma serve anche che i soggetti che dispongono di una grande mole di dati altrui, mostrino un atteggiamento proattivo mirato ad evitare possibili danni ai soggetti interessati. Tra i vari provvedimenti che deve prendere chi è intenzionato a raccogliere dati altrui, c'è quello di redigere un'informativa ben strutturata sulla *Privacy*. Tutti i soggetti interessati devono essere messi in condizione di sapere quali dati è interessata a raccogliere l'organizzazione che li richiede, e successivamente questi possono decidere se cedere o meno le proprie informazioni. Nell'informativa dovrà essere anche chiara la regolamentazione dei cookies (informazioni immesse sul browser personale quando si visita un sito web o quando si utilizza un *social network* con qualunque *smart object*[9]). Ogni *cookie* restituisce diversi dati su colui che sta navigando quali il nome del server di provenienza, un identificatore numerico, un *IP address* o altro (A. D. Miyazaki, 2008).

Tuttavia, nonostante la normativa sia diventata più stringente specie dopo lo scandalo di *Cambridge Analytica*, come rileva uno studio di Jensen et al. (2005), le persone non hanno una corretta consapevolezza della *Privacy* ed il loro atteggiamento online contrasta con la cura che effettivamente hanno nel proteggere i propri dati. Sul paradosso della *Privacy* già sono stati fatti alcuni cenni nell'Introduzione.

3.4 R Per effettuare *la Sentiment Analysis* mi sono servito di *R*, per spiegare il motivo di questa decisione occorre anzitutto effettuare un breve approfondimento circa il suddetto programma. *R* è nato come linguaggio di programmazione specifico per analisi statistiche dei dati, tuttavia oggi può essere definito come un ambiente contenente librerie ed oggetti utili ad analizzare dei dati e produrre grafici (Vito M.R Muggeo, 2002). In R è possibile strutturare i dati in vettori, matrici, *array* e liste oppure è possibile organizzarli secondo un dataframe per cui vengono organizzati in base a righe e colonne che si riferiscono alle variabili considerate. Per spiegare ad *R* cosa fare si utilizzano i comandi ma è possibile anche organizzare tutte le singole istruzioni in una funzione al fine di rendere più veloci le operazioni. Su *R* esiste un'innumerevole quantità di funzioni utili a

[8] Gazzetta Ufficiale accessibile da: https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2003-07-29&atto.codiceRedazionale=003G0218&fbclid=IwAR3Ofc5Gkk-YXYywDIT1B_jBVBNG2BHy_stYWjJLppTkmfnzVnah7WvR1k8

[9] Computer cookie control: transaction generated information and privacy regulation on the Internet, Joshua B. Sessler, 1996, Journal of Law and Policy)

calcolare i più comuni calcoli ed indici statistici come media, varianza, scarto quadratico medio ecc. per cui basta inserire su *R* l'argomento che si vuole calcolare tra le parentesi del comando utile ad effettuare il calcolo che ci interessa. Ad esempio considerando che su *R* esiste la funzione utile a calcolare la metà di un determinato argomento, basta inserire nel linguaggio di programmazione la funzione e tra parentesi l'argomento che ci interessa (ad esempio 10) ed il programma ci restituirà come risultato 5. Come già specificato, però, su *R* è possibile anche implementare pacchetti, librerie ed oggetti che permettono di compiere altre attività come ad esempio mostrare grafici o raccogliere dati da altri siti come verrà spiegato successivamente nella metodologia (3.6). Proprio grazie alla sua natura quantitativa arricchita però da altre funzioni che permettono di effettuare data scraping e visualizzazione dati attraverso grafici, R appare un software utile anche ad un'analisi complessa e soprattutto di carattere qualitativo come la Sentiment Analysis.

3.5 Natural Language Processing Prima di procedere ad una spiegazione più specifica su come è stata effettuata la Sentiment Analysis (3.6), occorre un approfondimento più generale su come un programma statistico come R (3.4), riesce ad identificare una grande mole di dati e ad associare a questi, il sentiment e addirittura le emozioni maggiormente evocate (nel paragrafo 3.7 i risultati specifici dell'analisi condotta nello specifico di questo studio). In questo tipo di analisi, il testo viene scomposto nelle sue singoli componenti ovvero le parole ed in basa ad un'analisi di queste viene stimato il sentiment emergente, cioè se c'è un atteggiamento positivo o negativo circa il topic di riferimento. Proprio per il fatto che in un'analisi come questa dove è richiesta una quantità di dati molto grande ed anche non semplice da analizzare, tecniche standardizzate di ML (approfondimenti nel paragrafo 2.1) risultano particolarmente utili. Come già scritto precedentemente (1.1), l'area del ML che si occupa di comprendere ed analizzare dati testuali relativi al linguaggio artificiale, è il *Natural Language Processing*. I suoi algoritmi, infatti, permettono alla macchina di comprendere ed analizzare un linguaggio diverso dal suo (che è quello artificiale) poiché processa informazioni di un linguaggio naturale che fondamentalmente, rispetto al linguaggio artificiale, è soggetto a cambiamenti che si presentano tra una generazione e l'altra e quindi segue regole molto ben definite (C.D. Manning & H. Schutze, 1999). Risulta evidente, quindi, come sia più difficilmente categorizzabile. La macchina nell'identificazione del linguaggio naturale fa un'analisi sia sintattica, quindi basata sul modo in cui le parole si combinano tra loro all'interno di una frase, che semantica, quindi basata sul significato[10]. Una macchina per poter assimilare il linguaggio naturale ha bisogno di una base ovvero di un pacchetto di informazioni contenenti il lessico del maggior numero di parole possibile e di un set di dati che gli permetta di comprendere le principali regole linguistiche e grammaticali. Anche questo aspetto verrà approfondito nella metodologia (3.6).

[10] Nel paragrafo 1.6 è presente la spiegazione nello specifico di come si è operato in questo senso attraverso una procedura che si è concentrata esclusivamente sulla funzione semantica delle parole

3.6 Metodologia Sentiment Analysis Per condurre *la Sentiment Analysis* mirata a comprendere l'opinione generale delle persone circa la Privacy, è stato utilizzato il programma statistico R per le ragioni spiegate nel paragrafo (3.4). Anzitutto su questo programma sono stati installati i pacchetti "*twitteR*", "*tidytext*", "*dplyr*" e "*ggplot2*", dopodiché prima di proseguire sullo stesso programma è stata mandata a Twitter la richiesta di poter utilizzare le sue API (1.1) per scopi di ricerca. Subito dopo averla ottenuta, è stato implementato il pacchetto "*twitteR*" da cui è partita la fase di *data scaping*[11] per cui su *R* è stato creato il collegamento con l'account *Twitter* che ha avuto l'autorizzazione di utilizzare le API del suddetto social network. Una volta che ad *R* è stato associato l'accesso alle *API del social network*, è iniziata la data collection attraverso la funzione "searchTwitter" per cui è stato richiesto al programma di raccogliere 3300 *tweets* contenenti l'*hashtag* "*Privacy*" e in lingua inglese. E' importante anche specificare la lingua poiché *R* riesce ad identificare i dati testuali in base al lessico che gli viene fornito e spesso questo si riferisce ad una specifica lingua per cui collezionare dati misti potrebbe risultare più complicato poiché richiederebbe un caricamento del lessico sul programma manuale. Una volta che *R* ha ottenuto il suo *dataset*, questo è stato trasformato in un *dataframe*[12] scomposto in vettori di uguale lunghezza. Successivamente è stata effettuata le *tokenizzazione*[13] per cui i dati testuali sono stati ricondotti a *token*. Nel grafico seguente (fig. 3.1) è rappresentato il *dataframe* delle parole con le relative frequenze assolute con cui si sono presentate.

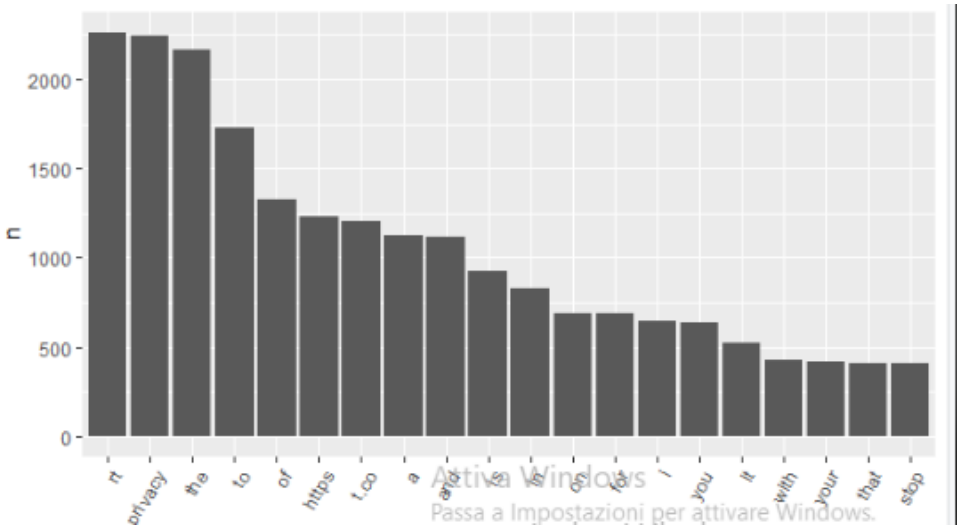


Figura 3.1 Prima raccolta BoW

[11] è una tecnica in cui un computer estrae dati dall'output leggibile dall'uomo proveniente da un altro programma (S. Munzert et al., 2014)

[12] è una lista di vettori (che si riferiscono alle variabili considerate, quindi nel nostro caso le frequenze assolute e le parole) che devono avere tutti la stessa lunghezza, ma possono essere di tipo diverso cioè sia qualitative che quantitative (K. Preiss, 1980)

[13] processo in cui i caratteri vengono rotti in parole chiamati token (J.E. Pentheroukadis, D. G. Bradlee, S. S. Knoll, 2006)

E' così che è stato generato un dataframe secondo il metodo "*Bag of Words*" per cui viene appunto conteggiato quante volte si ripete una determinata parola. Considerando proprio questa metodologia appare utile effettuare una fase di "*Stemming*" per cui vengono conteggiate allo stesso modo parole contenti la stessa radice per evitare eventuali eccezioni o errori di battitura da parte degli utenti. Prima ancora però appare utile evitare overloading di dati e per questo scopo è stato effettuato il cosiddetto procedimento di "*Stopwords*" per cui sono state eliminate tutte quelle parole non funzionali all'analisi (fig. 3.2).

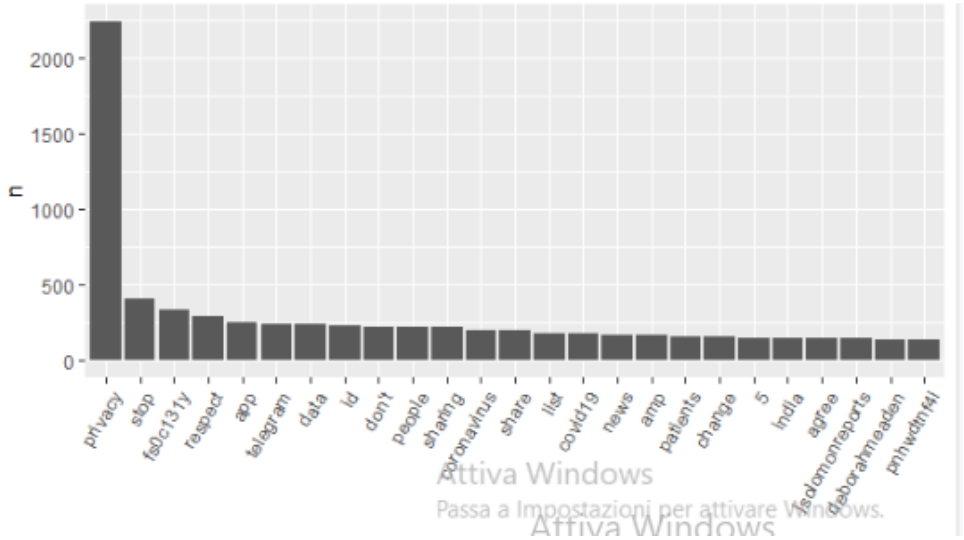
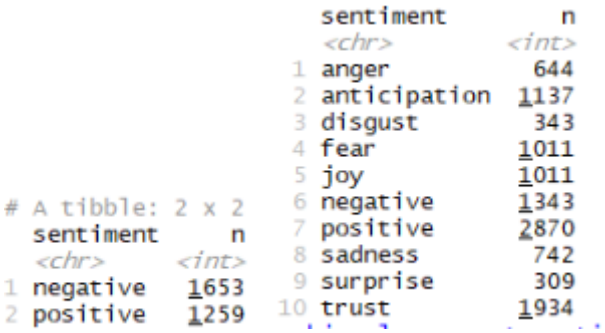


Figura 3.2 Dataframe dopo stopwords

E' evidente infatti che in un *tweet* sono contenute anche parole come ad esempio congiunzioni, preposizioni che servono a dare senso alla frase ma che, per permettere alla macchina di identificare il sentiment, non risultano utili. Questa infatti, in base al lessico fornito, associa ad esempio l'aggettivo "buono" come un qualcosa che può essere classificato come positivo pertanto tutte le parole all'interno della frase non utili in questo senso possono essere eliminate. Effettuata quindi questa fase di *data cleaning*, è stato fornito al programma il lessico "*Bing*" e quello "*NRC*" per approfondire anche circa le emozioni emerse dai tweets e con il comando "*get_sentiment*", è stata effettuata la clusterizzazione per cui ogni parola è stata associata ad una polarità (positiva, neutrale o negativa) e ad un sentimento. Entrambi i lessici sono stati creati grazie ad attività di *crowdsourcing* di diversi autori e sono stati convalidati utilizzando nuovamente nuove combinazioni di *crowdsourcing*, inoltre ovviamente non tutte le parole (che si rifanno alla lingua inglese) sono inserite nei lessici poiché molte considerate neutro. I risultati sono stati descritti nel successivo paragrafo (3.7).

3.7 Risultati Sentiment Analysis Il sentimento generato dalle analisi grazie al comando "*get_sentiments*" è stato trovato in base ai due lessici "*Bing*" ed "*NRC*" che si basano effettuano l'analisi su una sola parola. I risultati dei due lessici, come si nota dalla figura (fig. 3.3) sono discordanti e questa dimostra che le opinioni sono abbastanza bilanciate e che c'è un dibattito abbastanza accesso e nel complesso confusionario.



```
# A tibble: 2 x 2
  sentiment     n
  <chr>     <int>
1 negative   1653
2 positive   1259
```

```
   sentiment        n
   <chr>        <int>
 1 anger          644
 2 anticipation  1137
 3 disgust        343
 4 fear          1011
 5 joy           1011
 6 negative      1343
 7 positive      2870
 8 sadness        742
 9 surprise       309
10 trust         1934
```

Figura 3.3 Output Sentiment Analysis lessico Bing (sx) e lessico NRC (dx)

Un altro input interessante tra i risultati suggerisce che le emozioni maggiormente evocate sono a la fiducia e la gioia, tra quelle positive, e l'ansia e la paura tra quelle negative. Questo "secondo livello" di analisi sulle emozioni è reso possibile grazie al lessico "*NRC*" che categorizza le parole in base alle seguenti 8 emozioni: rabbia, ansia, disgusto, paura, gioia, tristezza, sorpresa e fiducia. Considerando il carattere molto esplorativo dell'analisi condotta, i findings principali si rifanno proprio al fatto che vengono espresse parole di felicità e fiducia circa la *Privacy* il ché ci restituisce come input che le persone considerano una normativa sulla protezione dei dati come qualcosa che rende felici le persone ed inoltre queste si sono mostrate fiduciose sull'efficacia attuale di tali normative. Dal lato, invece, delle parole negative, gli aspetti che andrebbero maggiormente curati sono l'ansia e la paura che fanno comprendere come una fuga dei dati personali costituisca effettiva preoccupazione per le persone.

3.8 Limitazioni La *Sentiment Analysis* condotta presenta indubbiamente alcune limitazioni che non vanno sottovalutate. Anzitutto l'analisi è stata effettuata su unigrammi per cui ogni parola contenuta tra i vari *tweets* è stata classificata come positiva, negativa o come espressione di una delle 8 emozioni come meglio descritto nel paragrafo 3.7. Pertanto non sono stati considerati i singoli tweets che si rifanno ad un determinato utente ma tutte le parole contenuti nei vari tweets sono state inserite in un unico "calderone" (*Bag of Words*) che ha poi classificato le singole parole. Tuttavia questo problema è parzialmente limitato dal fatto che i *tweets* hanno lunghezza massima di 140 caratteri quindi l'incidenza di diverse parole positive o negative da parte di un determinato utente ha potuto incidere in maniera limitata. Inoltre, su un *social network* potrebbe risultare utile anche effettuare un'analisi che categorizzi anche le emoticon[14]. I risultati delle analisi anche (3.7) hanno rilevato come nei due lessici ci sia stata una diversa categorizzazione per cui potrebbe risultare utile per studi futuri effettuare una analisi più robusta che riesca ad associare in maniera più adeguata il pensiero espresso nel singolo tweet al suo autore in modo da spostare l'analisi da quante parole positive o negative ci sono a quanti autori ci sono con sentiment positivo o negativo circa il topic di riferimento. A tale scopo, poiché ogni aggettivo viene classificato solamente come positivo o negativo, potrebbe essere utile una clusterizzazione per

[14] Riproduzioni stilizzate delle espressioni umane che esprimono una determinata emozione (Di Pasqua, M. 2009)

cui in maniera più adeguata viene associato un aggettivo ad una determinata caratteristica del *topic*. Ad esempio nel caso specifico, la *Privacy* può essere interpretata come adeguata o una cosa necessaria ma nell'analisi condotta in questo studio questa differenziazione non viene fatta. Se ad esempio grazie ad una attività di clusterizzazione (approfondimenti nel paragrafo 4.2) si associa la parola "adeguato" all'idea per cui si ritiene che la *Privacy* oggi è correttamente gestita e "buona" come un qualcosa riferibile alla sua condizione, sicuramente l'analisi appare più accurata. Un'altra limitazione dell'analisi, dovuta alla natura del programma utilizzato per effettuarla, è che non considera, ad esempio, qualificatori della parola come negazione (cioè ad esempio non considera "non" prima dell'aggettivo), ma si basa esclusivamente, come già spiegato, sugli unigrammi. Un ulteriore problema causato dalla "natura quantitativa" (3.4) di *R* è che risulta anche complicato qualificare correttamente *tweets* ironici o sarcastici.

Tuttavia il *focus* principale dell'analisi era quello di comprendere in che modo algoritmi del ML ed in particolare il *Recommendation System* potessero risultare utili in ambito marketing (i risultati sono stati mostrati nel paragrafo 2.5) mentre la *Sentiment Analysis* ha avuto semplicemente uno scopo esplorativo per cercare di comprendere quali emozioni fossero associate ad esse e se tendenzialmente in ambito online si abbia una attitudine positiva o negativa verso la cessione di dati a terzi. Come suggerito successivamente nelle Conclusioni, infatti, in base ai risultati dell'analisi spiegata nel presente capitolo e considerando anche quelli del capitolo successivo, questo studio potrebbe offrire uno spunto per le aziende comprendere se effettivamente possa essere conveniente una maggiore comunicazione ai consumatori circa la cessione di dati online data l'estrema confusione che anche i risultati di questa analisi, oltre a quanto già riscontrabile in letteratura, suggeriscono. L'input finale della ricerca complessiva infatti potrebbe essere racchiuso nella seguente domanda: "Se un corretto utilizzo dei dati dei consumatori genera per loro una maggior *intention to buy*, è lecito pensare di comunicare loro che la cessione dei dati, sempre effettuata in maniera responsabile e consapevole, può essere visto come un input che permette alle aziende di restituire loro un servizio migliore dando giovamenti sia lato consumatore che lato imprenditoriale?".

CAPITOLO 4

VALUTAZIONE SPERIMENTALE: QUANTO IL RS INFLUENZA IL CUSTOMER

In questo capitolo verrà analizzato nel dettaglio lo studio il cui obiettivo è principalmente quello di capire se ed eventualmente come gli algoritmi di ML possono risultare uno strumento utile alle imprese. Anzitutto viene fatta, quindi, un overview generale sul ML ed in seguito, più nel dettaglio, verrà spiegato dal punto di vista tecnico tutto il processo metodologico e di analisi includendo anche tutte le spiegazioni inerenti a tutte le tecniche adoperate in questo ambito, compresa la fase di raccolta dati.

4.1 Metodologia Considerando il potenziale di questi strumenti, ho voluto effettuare uno studio per sondare se, ed eventualmente in che modo, l'intention to buy di un'inserzione pubblicitaria che sfrutta i nostri dati (quindi su stimolo derivante da output generati da algoritmi di ML) differisce da una che rispetta maggiormente il diritto alla privacy individuale del consumatore (rifacendosi quindi alle inserzioni pubblicitarie che vediamo nei canali tradizionali come ad esempio la televisione). Per effettuare tale esperimento ho ideato un questionario su Qualtrics la cui metodologia viene chiarita maggiormente nel sottoparagrafo 2.4.1. Il numero di rispondenti è stato pari a 153 persone ma il campione considerato è stato minore come spiegato meglio nel sottoparagrafo 2.4.2 dove viene spiegato il processo di data cleaning. In sede di analisi, ho effettuato un paired samples t-test poiché entrambe le inserzioni pubblicitarie sono state visualizzate da tutti i dipendenti. Inoltre, per completezza, ho voluto approfondire l'analisi poiché, come si apprende dalla letteratura (N. Howe & W. Strauss 2007, 2009), i Millenials (o generazione Y), ovvero coloro che sono nati tra l'inizio degli anni '80 ed i primi anni 2000 (N. Howe & W. Strauss, 2000), hanno un rapporto verso l'ambiente online e delle caratteristiche differenti dalle generazioni precedenti (David C. Edelman, 2010, 2015). Perciò ho effettuato un indipendent samples t-test per verificare se il rischio percepito online e la sicurezza percepita online sono significativamente differenti tra i due gruppi. Ho cercato anche di "accorpare" le due scale precedentemente citate in un unico fattore quale "l'atteggiamento online" ma i risultati della factor analysis (approfondimenti nel paragrafo 2.5) non hanno permesso tale operazione.

4.1.1 Data collection Per raccogliere i dati sulle analisi proposte precedentemente (2.4), ho anzitutto preparato una Survey il cui flow è in Appendice (blocco B). Per la strutturazione di questo, dopo una breve intro di presentazione, ho cercato di raccogliere alcuni dati sui rispondenti chiedendo loro quale fosse l'acquisto o la ricerca più frequente fatti online tra tre categorie merceologiche o servizi quali libri, videogiochi e alloggio in hotel. La scelta di questi tipi di prodotti/servizi è stata fatta in seguito ad un focus group che ho svolto precedentemente per considerare su campioni molto piccoli delle variabili qualitative come i prodotti maggiormente cercati o acquistati online e sono emersi principalmente le tre categorie precedentemente elencate. Lo scopo della ricerca è molto esplorativo per cui non è stato effettuato un focus su una categoria merceologica anche se i risultati di questo studio possono offrire spunti per ricerche future maggiormente incentrati su singole categorie poiché potrebbero essere rilevate differenze rilevanti. Grazie a questa fase

preliminare è stato possibile limitare in maniera importante il numero di rispondenti che non ha acquistato o cercato quantomeno uno di questi tre prodotti. Come si apprende infatti dal grafico (figura 4.1), solo 5 (in termini percentuali il 4%) rispondenti su 142 (11 non hanno completato questa domanda) hanno risposto di non cercare nulla su Internet in generale o tra i prodotti selezionati.

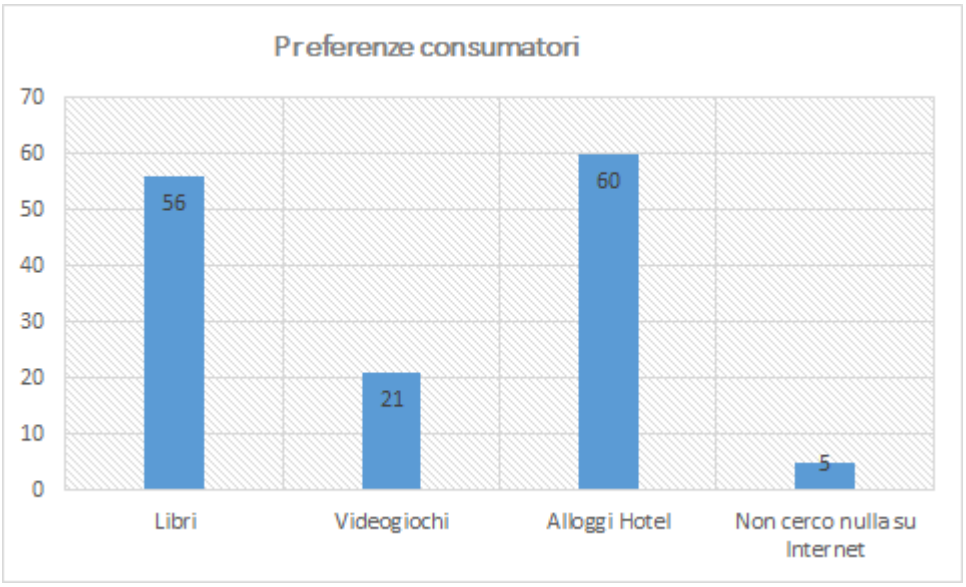


Figura 4.1 Preferenze dei consumatori

Successivamente in base al selezionamento delle tre categorie considerate, venivano visualizzate le domande successive: per coloro che hanno selezionato "Libri", sono apparse delle domande in cui veniva chiesto il genere di libri precedentemente più spesso cercati o acquistati online ed in base al genere selezionato, venivano proposti alcuni autori (anche questi determinati con focus group); a chi ha selezionato "Videogiochi", è stato chiesto il genere precedentemente più spesso cercati o acquistati ed in base a questo sono stati proposti alcuni videogiochi (sempre determinati su focus group); a chi ha selezionato "Alloggi per hotel per vacanze", invece, è stato chiesto dove ha cercato alloggi tra 5 possibilità di risposta (anch'esse determinate con focus group) quali: New York, Maldive, Parigi, Tokyo e nessuna tra queste. Inoltre a questi ultimi è stato chiesto se nelle ricerche/acquisti online recenti frequenti, hanno espresso preferenza per vicinanza al centro città o al mare (nel caso delle Maldive) oppure no. Successivamente ad ognuno di questi tre gruppi è stato chiesto su una scala Likert 1-7 di valutare l'intention to buy di un prodotto/servizio che rispettasse tutte le risposte selezionate precedentemente. La scala intention to buy è stata prevalidata da Mano, H., & Oliver, R. L. (1993) ed adattata in tutti i tre casi (libri, videogiochi ed hotel) a partire al lavoro di questi ultimi. Per rispondere alle domande relative a questa scala, è stato chiesto, quindi, di considerare un'inserzione pubblicitaria basata sulle risposte precedenti (ad esempio chi ha selezionato come categoria merceologica "Libri", come genere "Giallo" e come autore "Agatha Christie", è stata proposta l'inserzione pubblicitaria di un libro giallo di quest'ultima) Inoltre considerando l'importanza della leva prezzo (trova paper che dice che è molto importante) poiché non rilevante

ai fini di questo studio, nel caso del gruppo di rispondenti "Libri" è stato chiesto di considerare il prezzo dell'ultimo libro di quel genere acquistato, nel caso di "Videogiochi" è stata fatta una differenziazione a chi acquista abitualmente il gioco a prezzo pieno o chi preferisce aspettare tempo dalla sua uscita per trovare offerte mentre a chi ha selezionato "Alloggi ad hotel per vacanze" è stato precedentemente chiesto il prezzo (tra 4 fasce di prezzo) per cui l'inserzione in questo caso proponeva la fascia precedentemente selezionata. La procedura utilizzata per gli hotel sul prezzo non è stata utilizzata anche nel genere libri poiché in quel caso la variabilità dei prezzi è bassa (R. Creatini, 2001) come anche nel mercato dei videogiochi (T.A.Lorenzini, 2015). Successivamente indistintamente a tutti i rispondenti è stata sottoposta l'inserzione pubblicitaria di una FIAT Panda (pubblicità generica apparsa in televisione). Proprio per il fatto che tutti i rispondenti hanno visto questo prodotto, l'inserzione pubblicitaria non era basata sui dati dei consumatori e quindi, le domande mirate a comprendere l'intention to buy in scala likert 1-7 della FIAT Panda, hanno rappresentato un termine di confronto di un prodotto generico non basato sui dati dei consumatori rispetto ad uno invece basato su questi. Nella scelta di questo prodotto sono stati considerati solo due parametri per la sua natura generica ovvero:

A) che fosse anch'esso *hedonic* come i tre prodotti con viene confrontato per eliminare bias tra i due generi come suggerisce la letteratura (Terry L. Childers et al., 2001)

B) che fosse un'inserzione apparsa in televisione poiché come verrà approfondito nel capitolo sulle Conclusioni (5) per estensione lo studio potrebbe suggerire una differenziazione tra inserzioni pubblicitarie nell'ambito del *Social* (T.P. Liang et al., 2011) o *e-commerce* (Introduction to e-commerce, Jeffrey F. Rayport and Bernard J. Jaworski, 2002) e quelle trasmesse per canali tradizionali come ad esempio la TV che per loro natura sono meno targettizzate.

Nel questionario, inoltre, erano presenti domande sul rischio percepito online e domande sulla percezione di insicurezza online come tratto personale. Entrambe le scale erano *multi-items* e sono state prevalidate ed adattate a partire dagli studi di Mani, Z. & Chouk, I. (2018) nel primo caso e di Rojas-Méndez, J. I, Parasuraman, A., & Papadopoulos, N. (2017) nel secondo. Nella parte finale, invece, erano presenti domande demografiche su sesso, professione, età e titolo di studio del rispondente. In particolare i dati sull'età sono stati raccolti dando la possibilità ai rispondenti inserendosi in 8 differenti classi di età che poi in sede di analisi (*indipendent samples t-test* in questo caso) sono state divise in due parti tra *under* 30 ed *over* 30 per quanto spiegato precedentemente (4.1). Inoltre sono state effettuate alcune domande utili a testare con quanta frequenza gli utenti si sono trovati in passato ad inserzioni pubblicitarie che sfruttassero algoritmi di ML con particolare riferimento al *Recommendation System* e soprattutto una domanda per testare il *Manipulation Check* volta a verificare se i rispondenti hanno percepito o meno la violazione della *Privacy*. Data la natura dello stimolo, inoltre, non è stato ritenuto necessario effettuare alcun *pre-test* poiché i dati costituiscono esclusivamente uno strumento per generare un certo tipo di strategia comunicativa e non una certa caratteristica

di questa. Pertanto si è voluto verificare come algoritmi di ML che sfruttano i dati possono essere utili in termini di *intention to buy* e, quindi, non è stato visto come questa varia variabile in base a varianti di una strategia comunicativa per cui, invece, sarebbe risultato più opportuno effettuare un pre-test al fine di verificare se quella determinata variante di quella strategia comunicativa risultava "visibile" all'occhio dei rispondenti. Discorso analogo è stato fatto per *l'indipendent samples t-test*.

4.1.2 Data cleaning Dopo aver collezionato i dati attraverso il questionario sul *software Qualtrics*, c'è stata la fase di *data cleaning* utile a "pulire" il *dataset* in modo da eliminare tutti quei dati con scarso o assente carattere informativo per non incorrere in un overfitting di informazioni. In questa fase, quindi, dai 153 rispondenti iniziali, sono stati eliminati anzitutto tutti coloro i quali hanno completato meno del 79% del questionario. Sono stati considerati anche coloro che non l'hanno terminato poiché chi è giunto quantomeno a questa percentuale di completamento, ha risposto alle domande sulle due scale dell'*intention to buy* permettendomi quantomeno di effettuare il *paired samples t-test* su questa variabili. Inoltre sono stati eliminati anche coloro che hanno impiegato meno di 3 minuti a completare il questionario (che è il tempo di completamento stimato da *Qualtrics*) poiché potrebbero aver impiegato scarsa concentrazione nel test rischiando di creare *bias* nel campione.

Dopo queste due operazioni, il campione analizzato è stato di 123 rispondenti. In sede di analisi (questa parte è stata effettuata tutta sul programma, prima di procedere al test delle ipotesi, è stata effettuata un'analisi per testare l'affidabilità di tutte le scale presenti nello studio quindi: *intention to buy* (per libri, automobili, hotel ed automobile), rischio sicurezza online percepito, tratto individuale insicurezza ed i risultati sono stati tutti molto positivi poiché il *Cronbach Alpha* è stato nello specifico:

- Intention to buy libri: 0,94
- Intention to buy videogiochi: 0,967
- Intention to buy hotel: 0,917
- Intention to buy automobile: 0,911
- Intention to buy rischio percepito: 0,836
- Intention to buy insicurezza: 0,793

Non è stato necessario eliminare alcun item poiché il valore più basso del *Cronbach alpha*[15] è stato pari a 0,793, nonostante dalla tabella "Statistiche elemento-totale" (Appendice, sez. C), in alcuni casi il *Cronbach Alpha* subisse leggeri miglioramenti. Tuttavia i risultati sono risultati troppo positivi affinché si procedesse all'eliminazione di qualunque *item*. A confermare la bontà della scala, in Appendice è possibile trovare anche

[15] è un indice statistico utile a misurare l'attendibilità, ovvero per verificare la riproducibilità nel tempo a parità di condizioni, dei risultati da essi forniti (L.J. Cronbach, 1951)

la matrice di correlazione di elementi che ha rilevato in tutti i casi correlazioni molto alte tra tutti gli *item* all'interno della stessa scala. A questo punto, quindi, è stato possibile effettuare una media di tutti gli *items* che si riferivano ad una determinata scala, creando così una nuova variabile.

4.1.3 Formulazione delle ipotesi Rispetto a quanto scritto finora è stata, quindi, effettuata la formulazione delle ipotesi, poi testate con il *t-test*. La prima ipotesi formulata è stata:

H1: La media dell'*intention to buy* del prodotto "*no Privacy*" è maggiore a quella del prodotto "generico" non basato sui dati

In questo caso la supposizione che l'inserzione pubblicitaria del libro generi un'*intention to buy* più elevata rispetto a quella dell'automobile, è spinta dal fatto che nel primo caso l'inserzione ha sfruttato i dati precedentemente forniti dal consumatore e quindi era diversa in base a questi mentre nel secondo caso la pubblicità non si è basata su alcun dato ed infatti è stata generica. Infatti quest'ultima è stata indistintamente visualizzata da tutti i 123 rispondenti. Il supporto di questa ipotesi nasce da quanto suggerito dalla letteratura come già accennato in Introduzione (M.Ronchi & M.Ciancia, 2019; G. Pitruzzella, 2016) i *Big Data* forniscono un importante strumento per le aziende. In questo studio infatti si cerca di chiarire come questo asset possa trovare una sua più effettiva implementazione ed attraverso quali strumenti come ad esempio gli algoritmi di ML e più nello specifico attraverso il *Recommendation System* cui si cerca di fare una simulazione pur considerando tutte le limitazioni che vengono chiarite meglio nel paragfrato (2.7).

La seconda ipotesi ha il fine di verificare se ci fosse una differenza sul tratto individuale di insicurezza, intesa come caratteristica individuale nell'ambiente online tra due gruppi indipendenti quali gli *under* 30 (*Millenials*) e gli *over* 30 (generazioni precedenti). Pertanto:

H2: La media dell'insicurezza intesa come tratto individuale dei *Millenials* è minore di quella degli *over* 30

Il supporto della letteratura a riguardo è stata già accennata nel capitolo 2.4 per cui viene suggerito che gli under 30 possono essere categorizzati in maniera differente dalle generazioni precedenti in virtù di alcune differenza di atteggiamento riscontrabili anche e soprattutto nell'ambito digitale. Per questa ragione nel presente studio, è stato supposto che l'insicurezza nell'ambiente online intesa come caratteristica individuale e culturale fosse differente rispetto alle altre generazioni. In particolare, considerando i numeri forniti nell'Introduzione circa la diffusione al giorno d'oggi degli strumenti digitali, si suppone che questa sia minore negli under30

Discorso molto simile è stato fatto per testare un'altra variabile quale il rischio percepito online per cui l'ipotesi è:

H3: La media del rischio percepito online in termini di sicurezza dei *Millenials* è minore di quella degli *over* 30

Come già spiegato precedentemente (4.1), ho provato a riunire queste due variabili nel fattore "atteggiamento online" ma non è stato supportato dal risultato della *Factor Analysis* (Appendice cat. B) poiché solo un *item* ha presentato un *eigenvalue* maggiore di uno ma era in grado di spiegare solo il 55% della varianza, non raggiungendo quindi il valore soglia del 60%. I modelli, quindi, delle singoli analisi erano intesi a verificare in che modo la variabile x (le inserzioni pubbicitarie che sfruttassero i dati personali, quindi per estensione stimoli derivanti da algoritmi di ML, nei primi tre casi, l'età negli altri due). In entrambi i casi la variabile si è presentata come categorica poiché nelle prime tre formulazioni delle ipotesi, si suddivideva in inserzioni pubblicitarie "*si Privacy"* (quindi stimolo derivante da algoritmi di ML assente) o no privacy (stimolo presente) mentre anche nel caso dell'età, la variabile è stata configurata come categorica poiché, per quanto di per sé sia una variabile numerica, nel questionario è stata la possibilità ai rispondenti di rispondere a questa domanda identificandosi in una delle 8 classi presentate. Il programma statistico utilizzato (*SPSS*), infatti, ha restituito circa questo dato come ouput i numeri "1,2,3,4,5,6,7,8" in base a quale risposta è stata selezionata nell'ordine in cui è stata proposta nel questionario. SPSS ha infatti identificato la variabile come una categorica qualunque per cui ogni categoria viene assimilata ad un numero al fine di effettuare l'analisi. Le variabili dipendenti invece sono state tre: l'*intention to buy* del prodotto specifico nelle prime tre formulazioni di ipotesi, l'insicurezza come tratto individuale nel quarto e nell'ultimo il rischio percepito online in termini di sicurezza. In questi casi *SPSS* ha ricondotto queste variabili qualitative a numeri poiché gli *items* relativi a queste scale sono stati presentati nel questionario come domande in scala *Likert* 1-7 in cui veniva chiesto al rispondente di esprimere quanto fosse d'accordo con le affermazioni espresse nell'item. Le modalità di risposta sono state, quindi, espresse in ordine da "per niente d'accordo" a "completamente d'accordo" ed è stato quindi possibile per *SPSS* associare a queste variabili quantitative un numero ed effettuare tutte le statistiche descrittive e non presenti in Appendice (cat. B). I risultati sono stati discussi nella sezione successiva (2.4.3).

4.1.4 Test delle ipotesi Prima di procedere all'analisi ed il commento dei risultati, ho ritenuto opportuno approfondire brevemente la logica secondo cui vengono svolti i test delle ipotesi. Anzitutto questa tecnica viene utilizzata nell'ambito della statistica inferenziale (l'intera disciplina statistica è infatti possibile differenziarla in descrittiva, per cui in questo ambito viene esclusivamente descritto il dataset analizzato con indici come ad esempio la moda, la media, la mediana, la varianza ecc.; ed appunto in inferenziale) che è un insieme di metodi utili a permettere ai ricercatori di fare conclusioni o inferenze circa le caratteristiche dell'intera popolazione basandosi su dati del campione analizzato. Sostanzialmente ci permette di concludere se possiamo considerare di estendere i risultati riscontrati in un campione, per tutta la popolazione o meno. Nella formulazione delle ipotesi, come visto nel sottoparagrafo 4.1.2, c'è un'ipotesi, detta nulla (H0), ed

un'altra detta alternativa. L'ipotesi nulla esprime lo status quo mentre quella alternativa esprime esattamente l'opposto di quanto troviamo nella prima ipotesi. Gli indici che vengono utilizzati per esprimere quanto sono affidabili i risultati riscontrati nel campione, e quindi quanto è possibile estendere questi per tutta la popolazione sono essenzialmente due:

- il livello di confidenza: cioè la proporzione di volte che un intervallo stimato per un parametro della popolazione sarà corretto. Cioè se ho un livello di confidenza del 95%, significa che con questa probabilità l'intervallo di valori che ho stimato circa un determinato parametro risulta effettivamente **vero** per l'intera popolazione
- il livello di significatività: ovvero la possibilità con cui un'affermazione corretta viene accidentalmente (quindi erroneamente) rigettata. Se ad esempio ho un livello di significatività del 5%, questa è la possibilità che una affermazione giusta, estesa per tutta la popolazione, risulti sbagliata

Questi due indici risultano essere complementari, non a caso il livello di significatività viene espresso con la lettera greca "α" mentre il livello di confidenza è pari ad 1-α.

Tra le distribuzioni che utilizziamo in questo ambito, c'è la "t" che viene considerata quanto la varianza (è la media arimetica dei quadrati degli scostamenti dalla media, misura quale variabilità assume la variabile quindi più è alta più i valori di quella variabile possono assumere valori variabili) della popolazione è sconosciuta ed il campione che stiamo considerando è piccolo, ovvero inferiore a 30. Più cresce il numero di rispondenti, più la distribuzione tende ad essere normale ("z") (fig. 4.2).

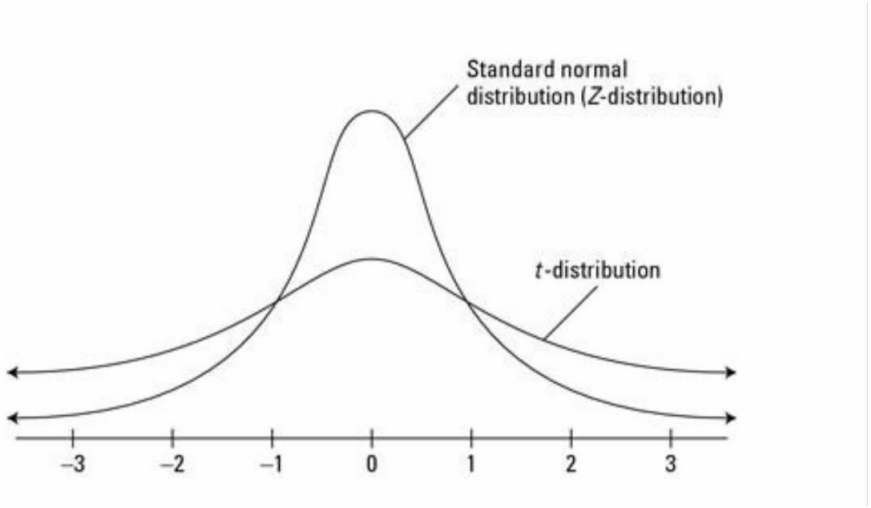


Figura 4.2 Distribuzione t e z

La forma di t deriva dai gradi di libertà (Ronald Fisher, 1922) (d'ora in poi GdL). Questi sono sempre funzione della grandezza del campione (GdL= dimensione campione "n"-1) per cui maggiore è la numerosità del campione, più la distribuzione tende ad essere nomale. Per spiegare maggiormente cosa sono i GdL: "sono il

numero massimo di valori logicamente indipendenti che hanno valori che hanno la libertà di variare nel campione considerato" (A.Ganti, 2019). Sostanzialmente quindi è il numero massimo di valori che possono assumere liberamente determinati valori, per cui ad esempio conoscendo i valori di tutti le variabili, la restante (o le restanti) non sono libere di variare ma dovranno per forza assumere un determinato valore, cioè conoscendo il valore di j variabili, la iesima dovrà per forza assumere un determinato valore.

Prima di spiegare gli ouput generati da *SPSS* circa ogni *t-test*, serve un ulteriore spiegazione anche del *p value*: mentre, come già accennato, α definisce quanto è probabile di rigettare un'ipotesi quando questa è vera, il p value calcola la probabilità in cui consideriamo vera un'ipotesi sbagliata (Wasserstein, R.L. e Lazar, N.A., 2016). Sostanzialmente, quindi, descrive la percentuale di popolazione o area sotto la curva che è al di fuori della nostra statistica. Mentre α è calcolato prima di aver collezionato i dati, il *p value* viene calcolato, quindi, dopo. Per chiarire con un esempio di un test a 2 code, il valore critico rappresenta quel valore che separa l'area blu dal centro (fig. 4.3)

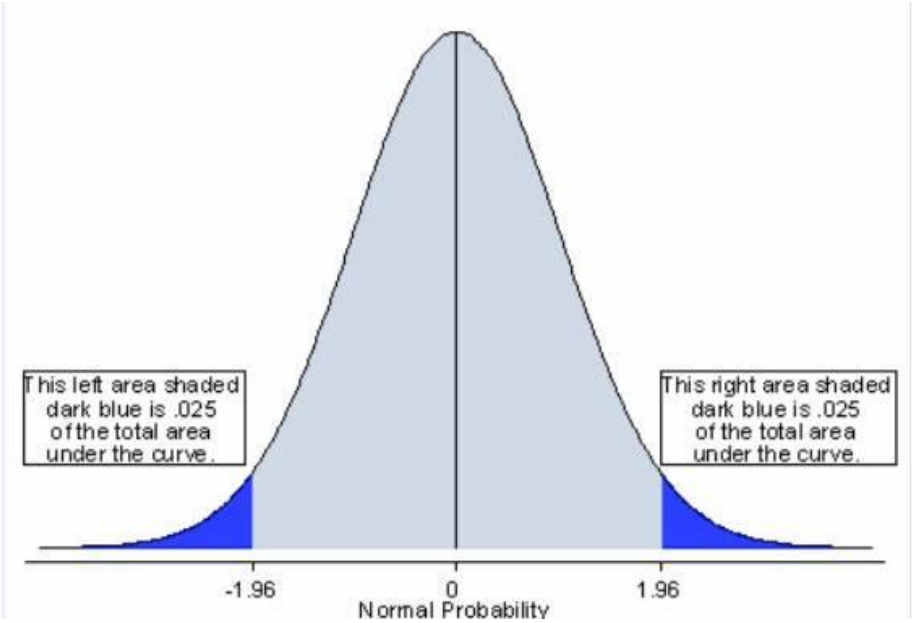


Figura 4.3 code test delle ipotesi

In un t-test per essere statisticamente significativo, il valore t deve essere all'interno della zona blu e perché ciò sia possibile, il p value ad esso associato deve essere inferiore ad α/2 (poiché nel caso di un test a due code, ad α nel caso di una coda. La differenza tra una coda e due code consiste nella differenza della formulazione delle ipotesi: in caso di uguaglianza tra primo e secondo membro, il test sarà a due code mentre in caso di maggiore/minore, il test è a una coda.

Perché venga rigetta l'ipotesi nulla, e venga quindi confermata l'ipotesi proposta dal ricercatore, il valore t deve essere nell'area blu, in caso contrario (*p value*>α), non rigettiamo l'ipotesi nulla.

In definitiva è possibile commettere due tipologie di errori in un test delle ipotesi:

- Errore di prima specie: quando rigettiamo l'ipotesi nulla quando in realtà è vera cioè i casi dei falsi positivi (probabilità corrisponde ad α)
- Errore di seconda specie: quando accettiamo l'ipotesi nulla quando è falsa, cioè i casi di falsi negativi.

Nello specifico di questo studio le ipotesi nulle (H0) ed alternative per i vari *t-test* (sia paired samples che indipendent samples) sono state le seguente:

1. Per quanto riguarda l'*intention to buy* del prodotto "*no Privacy*" dei libri:

H0: La media dell*'intention to buy* del prodotto "*no Privacy*" libri è uguale a quella del prodotto "generico" automobile

H1: La media dell'*intention to buy* del prodotto "*no Privacy*" libri è diversa da quella del prodotto "generico" automobile.

2. La seconda coppia di ipotesi invece è stata volta a verificare se lo stesso discorso è fattibile per un prodotto appartenente ad un'altra categoria merceologica come i videogiochi per cui:

H0: La media dell'*intention to buy* del prodotto "*no Privacy*" videogiochi è uguale a quella del prodotto "generico" automobile

H1: La media dell'*intention to buy* del prodotto "*no Privacy*" videogiochi è diversa da quella del prodotto "generico" automobile.

3. Stesso discorso è stato fatto con la terza coppia di ipotesi ma questa volta considerando un servizio quale l'alloggio in hotel per cui:

H0: La media dell'*intention to buy* del prodotto "*no Privacy*" alloggio in hotel per vacanze è uguale a quella del prodotto "generico" automobile

H1: La media dell'*intention to buy* del prodotto "*no Privacy*" alloggio in hotel per vacanze è diversa da quella del prodotto "generico" automobile.

4. Per quanto riguarda gli indipendent samples t-test nella prima coppia di ipotesi:

H0 = La media degli *under* 30 dell'insicurezza individuale intesa come tratto individuale è uguale a quella degli *over* 30

H0 = La media degli *under* 30 dell'insicurezza individuale intesa come tratto individuale è diversa da quella degli *over* 30

5. Il discorso in questo caso è simile a quello del punto 4 ma considerando un'altra variabile ovvero il rischio percepito online in termini di sicurezza. Perciò:

H0 = La media degli *under* 30 del rischio percepito online in termini di sicurezza è uguale a quella degli *over* 30

H0 = La media degli *under* 30 del rischio percepito online in termini di sicurezza è diversa da quella degli *over* 30

4.2 Risultati Terminata tutta la fase preliminare (4.1.2 e 4.2.3), e chiarito il funzionamento del test delle ipotesi (4.1.4), in questo paragrafo vengono resi noti i risultati del confronto medie con le due tipologie di *t-test* (*indipendent* e *paired samples*) considerate le ipotesi scritte nella fine del capitolo (2.4.4). La media dell'*intention to buy* dei libri (Minttbuy_libr= 5,4855, Appendice cat C.) è stata considerata statisticamente significativa poiché il *p value* calcolato dal campione (indicato come Sig. (2-code) nell'output) è stato indicato pari a 0.0000. Poiché, come già spiegato (2.4.3), questo valore indica l'area sotto la curva cioè per un valore t pari a 10,457 (che corrisponde al valore t nell'output in Appendice, cat. C) con GdL pari a 45 in una distribuzione t. Pertanto, poiché il *p value* (=0.0000)<α(=0,025 poiché il *test* è a due code ed il livello di significatività è del 5%), rigettiamo l'ipotesi nulla secondo cui le due medie sono considerate statisticamente uguali. Inoltre la differenza tra le due medie è di circa 2,62 (Mint_tobuy=5,4855; Minttbuy_automobile=2,8696) per cui, considerando che le domande erano in scala *Likert* 1-7, l'*intention to buy* del libro è risultata il 37,37% maggiore di quella dell'automobile. Per quanto riguarda l'intention to buy dei videogiochi: t(16)=0,2958; *p value*=0.009 <α(=0,025) per cui anche in questo caso rigettiamo H0 e consideriamo le differenze delle due medie come statisticamente significative. In questo caso l'*intention to buy* dei videogiochi (Minttbuy_videogiochi=4,7647; Minttbuy_automobile=2,9216) è risultata il 26,33% superiore. Nel caso, invece, dell'*intention to buy* degli alloggi in hotel per vacanze: t(55)=12,735; p value=0,000<α(0,025) per cui ancora una volta rigettiamo H0, con un *intention to buy* (Minttbuy_hotel=5,6369; Minttbuyautomobile=2,8869) del 40,14% più alta del prodotto "generico" dell'automobile.

Per quanto riguarda i due *indipendent samples t-test*, partendo dallo studio circa l'insicurezza come tratto individuale tra gli *under* 30 e gli *over* 30: prima di condurre l'analisi di confronto medie, bisogna controllare le statistiche de *Levene* che assumono che le varianze delle due medie sono uguali (H0= la varianza della

media degli *under* 30 è uguale a quella degli *over* 30; h1=la varianza della media degli under 30 è diversa da quella degli over 30). Poiché in questo caso il *p value*>*α* con *F* (116)=0,289(Appendice, cat.B), non rigettiamo l'ipotesi nulla ed assumiamo che le varianze sono uguali. Verificando quindi i risultati dell'output di *SPSS* che corrispondono con l'assunzione delle varianze uguali emerge che: t(116)=-4,888; p value=0.0000<α. Perciò rigettiamo l'ipotesi nulla ed assumiamo che la differenza tra la percezione di insicurezza come tratto personale è statisticamente significativa tra i gruppi under30 ed over30. In particolare gli under 30(Mmea_insicurezza_mill=4,2979;Mmea_insicurezza_nomill=5,5545) hanno mostrato insicurezza come tratto individuale il 17,95% in meno rispetto agli over30.

Nel caso del rischio percepito online, il test di *Levene* con F(116)=0,048 ha rilevato che le varianze sono uguali poiché *p value*>*α*. Osservando quindi le statistiche di questa assunzione: t116)=-4,967, p value(=0,000)<α(=0,025). Le due medie (Mmea_riskmill=4,3819;Mmea_risknomill=5,6667) hanno mostrato che gli under30 hanno una percezione del rischio, in termini di sicurezza, online inferiore del 18,37%.

Inoltre sono state effettuate alcune statistiche descrittive sul campione, con supporto dei grafici, su Excel:

- il 76% dei rispondenti ha rivelato di aver acquistato prodotto dopo aver visualizzato inserzioni pubblicitarie su internet (fig. 4.4)

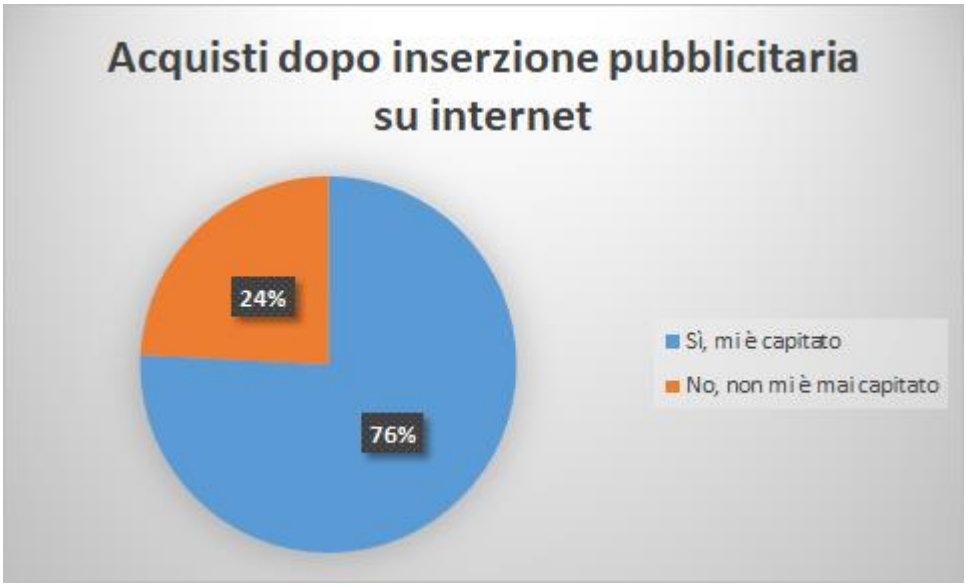


Figura 4.4 Acquisti dopo inserzione pubblicitaria

- è stata posta una domanda per testare il *Manipulation Check* quindi se i rispondenti hanno percepito una violazione della *Privacy* o meno nelle domande preliminari all'inserzione pubblicitaria del prodotto "*no Privacy*". I risultati hanno mostrato che: il 17% delle persone hanno sentito violata la propria *Privacy*, mentre l'83% no. (fig. 4.5)

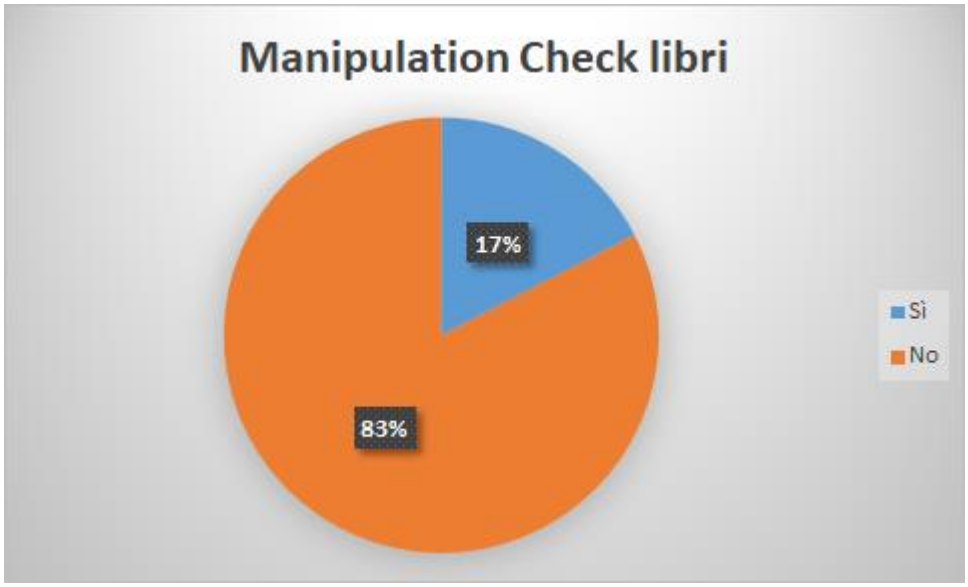


Figura 4.5 Manipulation Check libri

- nell'ambito dei videogiochi i numeri sono simili: (fig 4.6)

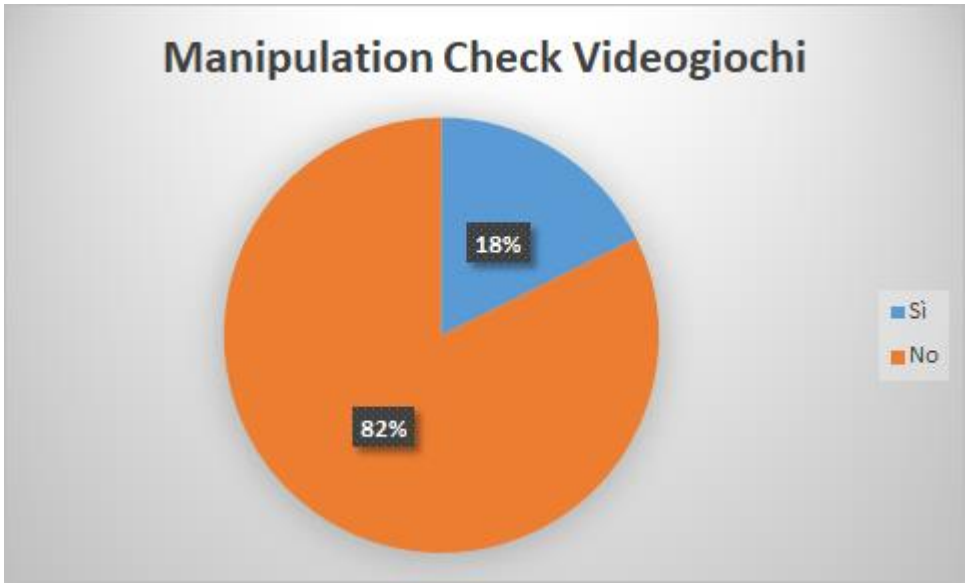


Figura 4.6 Manipulation Check videogiochi

- come anche in quello degli hotel: (fig. 4.7)

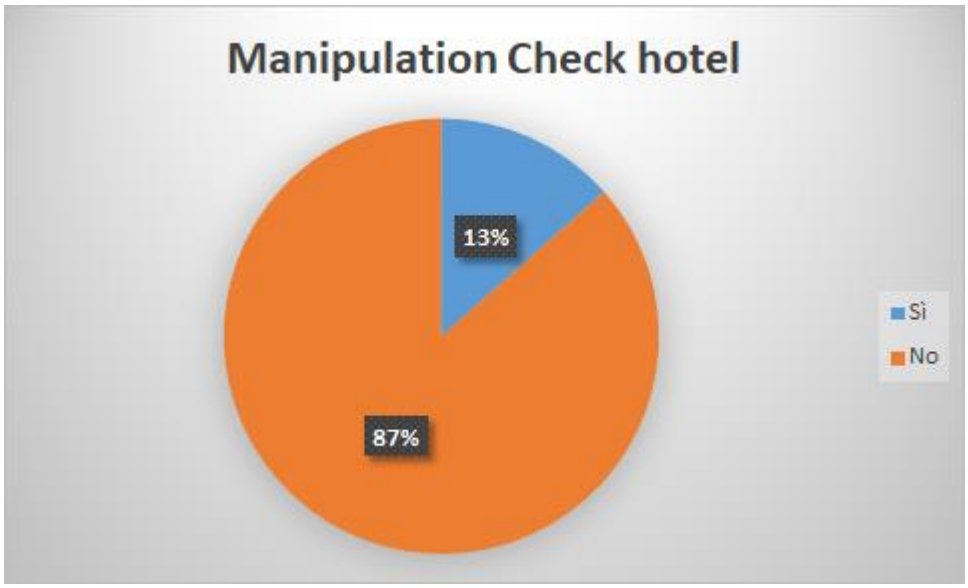


Figura 4.7 Manipulation Check hotel

- per quanto riguarda il genere: (fig. 4.8)

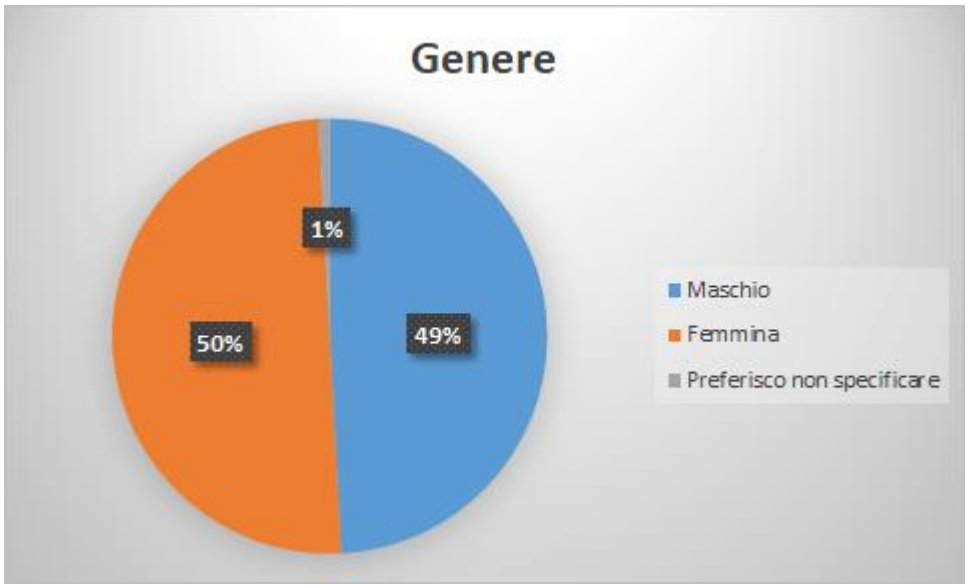


Figura 4.8 Grafico a torta genere rispondenti

- Per quel che concerne le professioni: (fig. 4.9)

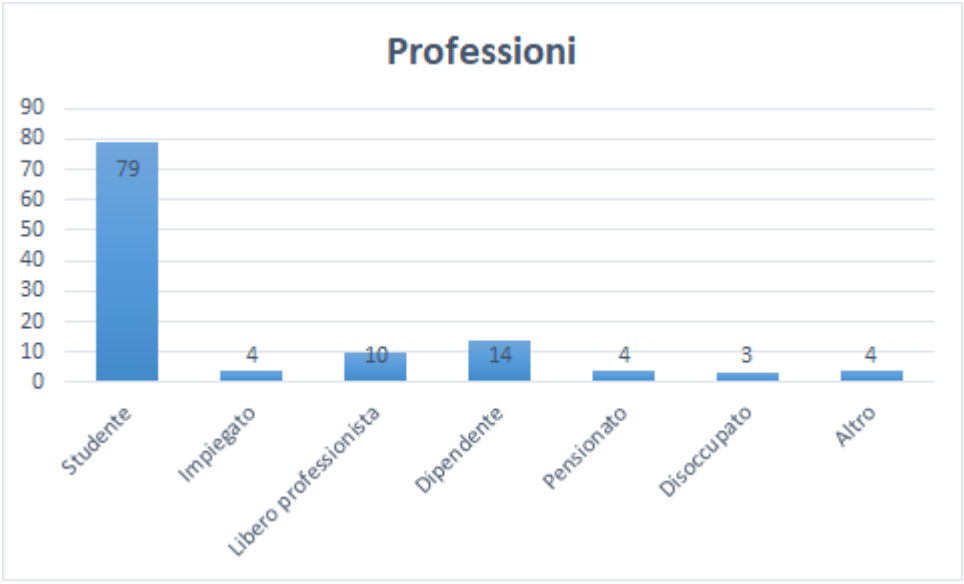


Figura 4.9 Istogramma professioni rispondenti

- sull'età invece i dati rivelano una distorsione del campione poiché l'81% del campione appartaneva ad una categoria (under 30) e solo il resto agli over 30. (fig. 4.10)

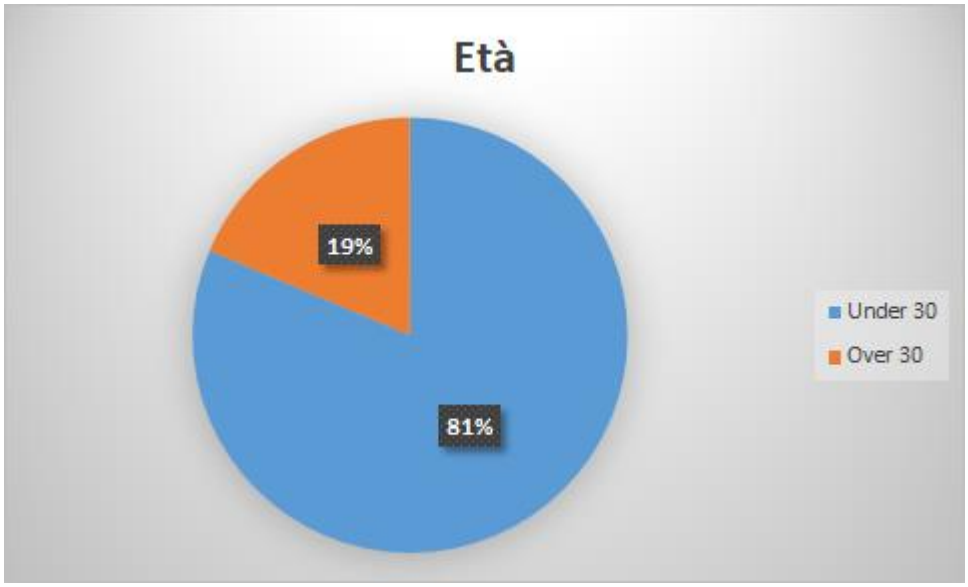


Figura 4.10 Grafico a torta età rispondenti

4.3 Implicazioni manageriali I *findings* di questo studio possono risultare interessanti in ambito manageriale poiché come già spiegato nell'Introduzione, grazie all'avvento della *Digital Transformation* oggi è possibile usufruire di una grande mole di dati (*Big Data*) dei consumatori il che implica da una parte che servono meccanismi che sappiano analizzarli in maniera facile e veloce (come ad esempio fanno gli algoritmi di ML), e da un'altra che costituiscono un'importante *commodity* per le aziende (come scritto nell'Introduzione). Inevitabilmente al giorno d'oggi, grazie all'avvento dell'*e-commerce* e del *Social commerce* (cenni a riguardo

nel capitolo 4.1.1) e alla digitalizzazione della società, esiste per le imprese l'esigenza di essere maggiormente incisiva nella strategia comunicativa poiché i consumatori sono esposti a molti più stimoli rispetto prima. Un approfondimento in merito viene fatto nel paragrafo 4.5 dove viene spiegato in maniera più precisa cosa sono i Micro-Momenti ed in che modo questi danno alle imprese come *feedback* che servono strategie pubblicitarie maggiormente incisive, rilevanti e customer-oriented. Ecco quindi che questo studio può risultare utile per far comprendere come l'implementazione di algoritmi di ML attraverso l'utilizzo dei dati dei consumatori, possono generare un'intenzione d'acquisto maggiore tra i consumatori rispetto a strategie comunicative che usufruiscono dei canali tradizionali come ad esempio la TV o gli eventi che per loro natura targettizzano in maniera meno precisa rispetto a quanto possa fare internet. Vero è che anche le strategie di eventi, televisioni o radio seguono logiche di targettizzazione per cui determinate pubblicità verranno mostrate in una determinata fascia oraria, prima o dopo un determinato programma televisivo per cercare di colpire con più probabilità il target di riferimento. Tuttavia come si apprende dai dati espressi nell'Introduzione, il numero di persone online è decisamente elevato ed inoltre in questo ambiente, proprio attraverso ad esempio algoritmi di ML, è possibile fare una "pubblicità personalizzata" che come si apprende da questo studio, inevitabilmente mostra un'intenzione d'acquisto maggiore degli utenti rispetto ad un'inserzione di un prodotto casuale.

4.4 Limiti della ricerca e spunti per ricerche future Questo studio presenta diversi limiti che possono fornire spunto per approfondimenti in ricerche future. Anzitutto il campione di 123 persone è stato ulteriormente frammentato in sede di analisi t-test poiché sono state confrontate le medie dell'intention to buy dei tre prodotti "*no Privacy*" separatamente rispetto a quello generico, generando addirittura nel caso dei videogiochi un campione di sole 17 persone. In statistica è sempre consigliato di utilizzare il campione più grande possibile poiché con maggiore precisione tende a rispecchiare il dato reale della popolazione. Per quanto riguarda, invece, i due *indipendent samples t-test*, il campione è apparso sicuramente maggiore poiché aggregato ma in questo caso c'è stato un altro *bias* poiché gran parte del campione (81% come si apprende nel paragrafo 2.5) apparteneva alla classe di età *under* 30. Potrebbe principalmente risultare utile, quindi, ripetere le analisi con campioni più grandi ottenendo quindi risultati più attendibili. Inoltre occorre anche specificare la logica seguita nell'ambito del *Manipulation Check* che in questo caso è stato effettuato principalmente per comprendere ulteriormente come i consumatori si approcciano circa la cessione dei propri dati personali ad aziende ma non è stato fatto nessun data cleaning in questo senso poiché, considerando la natura dello stimolo che non costituisce un elemento caratterizzante del prodotto o del suo aspettivo comunicativo quanto, invece, alla verifica di constatare l'efficacia di un input per rendere fruibile uno strumento quale il *Recommendation System*. Pertanto, proprio perché lo studio è stato effettuato considerando secondo la logica del *"what if"* per cui si è voluto verificare come varierebbe il comportamento d'acquisto dando come assunta la cessione dei dati personali.

Un'altra importante limitazione è dovuta dal fatto che lo studio ha esaminato principalmente quanto e se l'*intention to buy* cresce in presenza di maggiore concessione dei dati personali dei consumatori però, nel'elaborazione del questionario l'obiettivo è stato quello di avvicinarsi il più possibile alla realtà che troviamo con il Recommendation System cercando di seguire le stesse logiche che seguono gli algoritmi di ML in merito. Inevitabilmente, però, il fatto che le inserzioni pubblicitarie proposte nella *survey* sono stato frutto dell'intelligenza artificiale e non di quella informatica come avviene nel caso del *Recommendation System*, per verificare in maniera più attendibile quanto questi risultano davvero efficaci alle imprese in ambito *Marketing*, potrebbe risultare utile svolgere ulteriori studi che simulino o tentino di riprodurre la realtà in maniera più precisa. Questo studio, infatti, si limita solo ad offrire un contributo in questo senso.

4.5 Micro-momenti In un ambiente come quello attuale, l'automazione del *Marketing* attraverso l'ausilio dell'Intelligenza artificiale è fondamentale per sopportare ed analizzare i Big Data dei consumatori soprattutto nell'ambito dei *Social Media* (K. A. Grossberg, 2016). Oggigiorno capitano più spesso situazioni in cui ci troviamo a prendere decisioni mentre stiamo svolgendo diverse attività contemporaneamente, mentre guardiamo qualcosa per strada o prendiamo in generale una mole di informazioni molto maggiore. *Google* per esprimere questo concetto ha coniato il termine "Micro-Momenti". Per chiarire meglio il *mindset* odierno tipico delle persone, le persone sono spesso si trovano spesso in contesti in cui vengono sottoposte a stimoli di *Marketing* ed a prendere decisioni di qualunque tipo in un ambiente estremamente dinamico, complesso e che concede alle persone sempre meno tempo per completare il loro processo di decision-making (D.Ariely, 2016). Alla luce di questo, risultare rilevanti ed emergere in un ambiente iper-competitivo anche dal punto di vista del *Marketing* e della comunicazione, appare come un aspetto che deve essere curato da parte di ogni azienda.

CAPITOLO 5

CONCLUSIONI Nel suo complesso la ricerca si è basata su due studi esplicati nei capitoli 3 e 4 in cui il primo (la *Sentiment Analysis*) ha evidenziato come ci sia un dibattito acceso sulla *Privacy* e come le emozioni negative principalmente evidenziate relativamente ad essa sono l'ansia e la paura. Nel secondo studio, invece, è stato evidenziato come un'inserzione pubblicitaria che sfrutta i dati dei consumatori costituisca per questa una *intention to buy* più alta rispetto ad un prodotto non basato sui dati. I risultati del secondo studio, con le dovute cautele, per estensione potrebbero anche esprimere il paragone dell'efficacia che ha una pubblicità nell'ambito del *Social* e dell'*e-commerce* rispetto a quella più tradizionale visibile ad esempio in TV dove le aziende riescono a fare una targettizzazione dei clienti meno specifica basata esclusivamente sui programmi televisivi e sugli orari senza, quindi, raggiungere il grado di specificità delle inserzioni online. Occorre, inoltre, specificare che in questo studio è stata implementato esclusivamente un "modello base" di quanto riescono invece ad effettuare gli algoritmi di ML nell'ambito del *Recommendation System*. Basti considerare, infatti, che partendo dagli acquisti storici dei consumatori (come fatto nel caso specifico di questo studio) si possono delineare maggiormente i loro gusti, le loro tendenze e quindi, assumendo che in futuro le aziende riescano a possedere una grande mole di dati dei consumatori, potrebbe risultare anche più semplice predire prodotti che possano risultare loro utili anche nell'ambito di categorie merceologiche differenti. Osservando, cioè, in maniera estrememente precisa il comportamento dei consumatore e conoscendoli meglio, potrebbe aprirsi un importante solco per la ricerca per cui potrebbe risultare estremamente più facile targettizzare i consumatori in cluster maggiormente definiti che esprimono in maniera ulteriormente più precisa il comportamento d'acquisto di molti di loro arrivando quindi a proporgli prodotti in ogni categoria merceologica in base al cluster in cui vengono classificati. Per chiarire maggiormente questo aspetto potrebbe risultare utile un esempio: supponendo che un'azienda nel settore dell'automotive venga a conoscenza del fatto che un determinato consumatore tra i suoi acquisti recenti ha mostrato particolare tendenza ai prodotti "Bio" che privilegiano il rispetto per l'ambiente in diverse categorie merceologiche (come ad esempio quello della cura personale, alimentare, di abbigliamento ecc.), potrebbe risultare maggiormente interessante per questo un'inserzione pubblicitaria di un'automobile elettrica. A tal proposito basti vedere quanto scritto anche nell'introduzione dove lo studio di Zhou, R., Khemmarat, S. e Gao, L. suggerisce come, conoscendo meglio i gusti dei consumatori, è possibile offrire loro prodotti più conformi ai loro bisogno od interessi rispetto a prodotti popolari.

Considerando, inoltre, quanto emerso anche dai risultati degli *indipendent samples t-test* (paragrafo 4.2), gli *under* 30 mostrano anche un atteggiamento diverso rispetto alle precedenti generazioni in cui la preoccupazione online in termini di sicurezza percepita, ed anche l'insicurezza in ambito online come tratto personale, sono minori per cui potrebbe risultare meno difficoltoso comunicare, contrariamente a quanto si

pensi, che offrire i dati alle aziende, a patto che queste le utilizzino in maniera responsabile, non fa altro che dare a queste uno strumento che risulta utile a loro ma anche ai consumatori stessi come dimostrano i findings del secondo studio di questa ricerca che hanno mostrato un'*intention to buy* più elevata per prodotti che meglio rispecchiano i loro gusti espressi attraverso gli acquisti storici.

Infine occorre anche contestualizzare tipi di strategie comunicative che sfruttano algoritmi di ML secondo logiche di *Recommendation System*. Oggigiorno ci troviamo, infatti, in un clima dove il mercato risulta in quasi la totalità delle categorie merceologiche iper competitivo (come anche il contesto dello studio di partenza di Zhou, R., Khemmarat, S. e Gao, L. che si è basato su *YouTube*, dove la quantità di video è innumerevole, come campo d'analisi), i consumatori si ritrovano continuamente esposti ad una grandissima mole di informazioni per cui esprimere un messaggio rilevante per loro risulta particolarmente difficoltoso (Valdani E., 1997) e da qui poi deriva anche quanto già espresso nel paragrafo dedicato ai "Micro-momenti" (paragrafo 4.5) dove appunto viene evidenziato come il *time-pressure* per il consumatore lo porta sempre più spesso a fare scelte nell'ambito del *decision-making* che richiedano sempre meno tempo e sforzo. Un ulteriore spunto in questo senso potrebbe essere offerto proprio da quest'ultimo punto per cui l'implementazione di strategie comunicative che meglio riescono ad esaltare le euristiche ed il sistema 1 di Kanheman e Tversky (2011) possono risultare nell'attuale ambiente particolarmente efficaci. Infine l'ultima pandemia da Covid-19 risulta ancora troppo recente per essere già oggettom di ricerca ed analizzata dal punto di vista degli effetti che ha generato anche in ambito manageriale, ma considerando la portata dell'evento non è possibile escludere del tutto che ci sia un cambiamento del comportamento da parte dei consumatori che li ha portati a sperimentare maggiormente tutto ciò che riguarda l'*e-commerce*, il *social-commerce* ed in generale il mondo online per cui quanto già detto potrebbe risultare ulteriormente amplificato nei prossimi mesi.

Sostanzialmente, partendo dall'assunto che le aziende riescano in maniera responsabile ad occuparsi di come soddisfare i needs dei consumatori, queste potrebbero estendere un ulteriore servizio all'interno della loro offerta, una fase preliminare a quella d'acquisto (e non solo quindi di assistenza post-acquisto) in cui, conoscendo ed avendo mappato i consumatori, riescono a fornire a questi inserzioni pubblicitarie di prodotti maggiormente affini a loro e l'ambiente circostante potrebbe portare le aziende a fare di questo aspetto una competenza estremamente rilevante ed efficace se non necessaria. Conferme su quanto le aziende si stiano concentrando ad aumentare la propria presenza online e l'efficacia delle proprie strategie in questo ambito arrivano anche da quanto stiamo vedendo nell'ambito dell'Omnicanalità[16] dove le aziende stanno cercando di

[16] Oggi la capacità dell'uso simultaneo di diversi canali di interazione tra consumatoe e negozio (ad esempiol'uso dell'acceso a Internet mobile all'interno del negozio fisico al dettaglio per cercare informazioni sui prodotti e/o confrontare i prezzi dei prodotti) costituisce la caratteristica dominante del fenomeno del commercio al dettaglio "omnichannel" che si basa sulla ben consolidata infrastruttura al dettaglio multicanale sviluppata principalmente dallo sfruttamento commerciale del World Wide Web (Lazaris, C.,Vrechopoulos, C., 2014).

stimolare sempre maggiormente i consumatori a vivere esperienze d'acquisto online più appaganti, mettendo in pratica quanto suggerisce già parte della letteratura circa l'efficacia di questa strategia (Payne, E. M., Pelties, J.W., Barger V. A., 2017), affinché diventino più frequenti per adeguarsi all'ambiente circostante e sconfiggere le barriere geografiche che appaiono sempre più anacronostiche come ci suggerisce la letteratura (Levitt, T., 1993; Mittelman, J.H., 1993; Bandura, A., 2001) che da diversi anni ha definito come consolidato il processo di globalizzazione che l'umanità sta vivendo.

In definitiva, quindi, oltre agli input già forniti, questo studio rileva che da una parte le persone mostrano un dibattito bilanciato su *Twitter* circa la *Privacy* rilevando che proteggere i propri diritti dà gioia, fiducia ma anche paura ed ansia; e dall'altra che quando le aziende sfruttano i dati restituiscono agli stessi clienti prodotti che generano in questi una intention to buy più alta. C'è mancanza di comunicazione, quindi, per i consumatori di come lo sfruttamento dei dati personali restituisca a questi stessi un servizio migliore in termini di inserzioni pubblicitarie maggiormente conformi alle loro intenzioni d'acquisto?

BIBLIOGRAFIA:

(1) Alpaydin, E., "Introduction to Machine Learning", 2004, MIT press

(2) Ariely, D., "Time Pressure: Behavioral science considerations for mobile marketing", 2016, Moments that Matter, Think with Google

(3) Ayres et al., "BEAGLE: All Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics", 2012, Systematic Biology, Volume 61, Issue 1, pagine 170-173

(4) Baird, L.," Residual algorithms: Reinforcement learning with function approximation", 1995, U.S Air Force Academy

(5) Bandura, A., "The changinn face of psychology at the dawning of a globalization era", 2001, Canadian psychology/Psychologie canadienne 42 (1), 12

(6) Bansal, G. & Zahedi, F.M.; "Trading Trust for Discount: Does frugality moderate the impact of Privacy and Security Concerns?", 2010, AMCIS 2010 Proceedings

(7) Brady, M. K., "Customer orientation: Effecgts on customer service perceptions and outcome behaviors", 2001, Journal of Sevice Research 3 (3), 241-251

(8) Brooks, R.A., "Intelligence without representation", 1991, MIT Artificial Intelligence Laboratory

(9) Burke, R.; "Hybrid web recommender systems", 2007, School of Computer Science, Telecommunicationts and Informations Systems

(10) Carrascal et al.; "Your browsing behavior for a big mac: Economics of personal information online" ,2013, Proceedings of the 22end international conference on World Wide Web, 189-200

(11) Cascio Rizzo, G.L.; "The Impact of Privacy Factors on the Online Users' Wilingness to Leave Digital Footprints: An investigation of the effects of personalization in the perspective of Privacy Paradox Calculus, 2017, Luiss Guico Carli Univerity, Research Aproach for PhD Program in Management

(12) Cascio Rizzo, G.L.; "Sentiment analysis: algoritmi di clustering e metodi di classificazione supervisionata", 2018, LUISS Guido Carli

(13) Childers, T.L. et al., "Hedonic and utilitarian motivations for online retail shopping behavior", 2001, Journal of retailing 77 (4), 511-535

(14) Creatini, R., "Libri e prezzo fisso: non si gira mai pagina", 2001, Mercato Concorrenziale Regole

(15) Cronbach, L.J.; "Coefficient alpha and the internal structure of tests", 1951, Psychometrika 16 (3), 297-334

(16) Di Pasqua, M., "Emoticodex: la retorica delle emoticon, 2009, Università degli Studi Napoli Federico II

(17) Edelman, D. C.; "Branding in the Digital Age: You're spending your money in all the wrong places", 2010, Harvard Business review

(18) Edelman, D. C., "Competing on customer journeys", 2015, Harvard Business review

(19) Elmasri, "Fundumentals of Database Systems, 6th edition", 2010, Addison-Wesley, Indian Internationl Ed. In Softcover

(20) Fisher, R., "On the nathematical foundations of theoretical statistics", 1922, Royal Social Publishing

(21) Ganti, A., "Degrees of freedom", 2019, Accessibile da: https://www.investopedia.com/terms/d/degrees-of-freedom.asp

(22) Grossberg, K.A.; "The new marketing solutions that will drive strategy implementation", 2016, Strategic Leadership Forum

(23) Hamming, R.W.; "Error-detecting and error-correcting codes, 1950, The Bell system technica journal 29 (2), 147-160

(24) Howe, N. & Strauss, W.; “The next 20 years: how customer and workforce attitudes will evolve”, 2007, Harvard Business review
(25) Howe, N. & Strauss, W. ,”The fourth turning: What the cycles of history tell us about America's next rendezvous with destiny”, 2009, Three Rivers Press
(26) Howe, N. & Strauss, W. Millenials Rising 2000 Knopf Doubleday Publishing Group
(27) Isaak, J. and Hanna, M. J. ;”User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection, 2018, IEEE Computer Society
(28) Jensen, C. et al., “Privacy practises of Internet Users: Self-reports observed behavior”, 2005, International Journal of Human-Computern Science Studies 63 (1-2), 203-227
(29) Kanheman, D. & Tversky,A.; “Thinking fast and slow”, 2011, Farrar, Strauss and Giroux
(30) Koren, Y.; Bell, R. & Volinsky, C.; “Matrix factorization techniques for recommender systems, 2009, IEEE Computer Society
(31) Laney, D.; “3D Data Management Controlling Data Volume, Velocity and Variety”, 2001, META Group Inc.
(32) Lazaris, C. e Vrechopoulos, A., “From Multichannel to Omnichannel Retailing: Review of the Literature and Calls for Research”, 2014, 2nd International Conference on Contemporary Marketing Issues, (ICCMI) 6
(33) Levitt, T., “The globalization of markets”, 1993, Readings in international business: a decision approach 249
(34) Liang, T.P. et al.; “Introduction to the special issue social commerce: a research framework for social commerce”,2011, International Journal of electronic commerce 16 (2), 5-4
(35) Lorenzini, T.A.; “Il mondo dei videogames: mercato ed eventi, 2015, Università degli Studi di Padova
(36) Mani, Z. & Chouk, I.; “Drivers of consumer' resistance to smart products”, 2017, Journal of Marketing Management, 33 (1-2), 76-97
(37) Manning, C.D. e Schutze, H., “Foundations of statistical natural language processing”, 1999, MIT press
(38) Mano, H. & Oliver, R.L., “Assessing the Dimensionality and Structure of the Consumption Experience: Evaluation, Feeling, and Satisfaction”, 1993, Journal of Convention & Event Tourism
(39) Markun Hanjana, S.T. et al. “Understanding factors influencing consumers online purchase intention via mobile app: perceived ease of use, perceived usefulness, sytem quality, information quality and... “, 2019, Business Management Progran, Jakarta, Indonesia
(40) Marr, D.: “Early processing of visual information”, 1976, Royal Social Publishing
(41) Mitchell, T.M.; Carbonell, J.G. & Michalski, R.S.; “Machine Learning: a Guide to Current Research”, 2012, Kluwer Academic Publishers
(42) Mittelman, J. H., “Globalization: critical reflections”, 1996, Lynne Rienner International Political Economy Yearbook, vol. 9, pages/LC: 96-1319
(43) Miyazaki, A.D., “Online Privacy and the Disclosure of Cookie Use: effects on Consumer Trust and Anticipated Patronage”, 2008, Journal of Public Policy & Marketing
(44) Muggeo, V.M.R.; “Il linguaggio R: concetti introduttivi ed esempi”, 2002, Google Scholar accessibile da: http://www.cs.us.es/cursos/ci-2003/Programas/R/doc/contrib/nozioniR.pdf?fbclid=IwAR2OdLNOTYhXSra6fOUmr1cxwWhR2FHTDxLnSRdL1JXKabb_--v_pKAT-y4
(45) Munzert, S.; C. Rubba, C.; Meisner, P. & Nyhuis, D.; “Automated data collection with R: A practical guid to web scraping and text mining”, 2014, John Wiley & Sons
(46) Murphy, K.P.; “Machine Learning: a probabilistic perspective”, 2012, MIT press
(47) Norberg, P.A.et al., “The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors”, 2007, John Wiley & Sons
(48) Payne, E.M., Peltier, J.W. e Barger, V. A., “Omni-channel marketing, integrated marketing communications and consumer engagement”, 2017, Journal of Research in Interactive Marketing

(49) Pentheroudakis, J.E; Bradlee, D.G. & Knoll, S.S.,"Tokenizer for a natural language processing system", 2006, U.S. Patent
(50) Pitruzzella, G.; "Big data, competition and privacy: a look from antitrust perspective", 2016, Università degli Studi di Palermo
(51) Preiss, K.; "Data frame model for the engineering design process", 1980, Design Studies
(52) Pudil, P. & Novovicova, J. "Novel Methods for Featur Subset Selection with Respect to Problem Knowledge", 1998, The Springer International Series in Engineering and Computer Science, vol 453. Springer, Boston, MA
(53) Rainie et al.; "Anonimity Privacy and security online", 2013, Pew Research Center, 5
(54) Rayport, J.F. & Jaworski, B.J., "Introduction to e-commerce", 2002, Combining Busines and Information Technology, first edition
(55) Rojas-Méndez, J.I.; Parasuraman, A. & Papadopoulos, N.; "Demographics attitudes, and technology readiness", 2017, Marketing Intelligence & Planning
(56) Ronchi, M. & Ciancia, M.; "Digital Transformation. Metodi e strumenti per guidare l'evoluzione digitale delle imprese attraverso design, marketing e comunicazione", 2019, Transmedia Practice
(57) Roweis, S.T. & Saul, L.K.; "Nonlinear Dimensionality Reductionby Locally Linear Embedding", 2000, Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London
(58) Samuel, A.L.; "Some studies in machinelearning using the gameof checkers",1959, IBM journal of research and development
(59) Schafer, J.B.; Frankowski, D.; Herlocker; J. & S. Sen; "Collaborative filtering recommender systems", 2007, The adaptive web, 291-234
(60) Sessler, J.B.; "Computer cookie control: transaction generated information and privacy regulation on the Internet", 1996, Journal of Law and Policy
(61) Sharma,S.; Sharma, V. & Sharma, A.; "Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis", 2016, International Journal of Modern Computer Science, vol. 4 Issue3
(62) Shaya, S.A. et al.; "Intelligent performance-based product recommendation system", 2010, U.S. Patent
(63) Sinkula, J. M., Baker, W. E. e Noordewier, T., "A framework for market-based organizational learning: Linking values, Knowledge, and behavior", 1997, Journal of the academy of Marketing Science 25 (4), 305
(64) Slater, S. F. e Narver, J. C., "Market orientation and the learning organization", 1995, Journal of marketing 59 (3), 63-74
(65) Smith, H.J. et al., "Information Privacy: Measuring Individuals' Concerns about Organizational Practises", 1996, MIS Quarterly, vol. 20 No. 2
(66) Teigens, V.; Skalfist, P. & Mikelsten, D. "Intelligenza artificiale: la quarta rivoluzione industriale, 2019, Cambridge Standford Books
(67) Thaler, R.H., "Misbehaving: the making of behavioral economics", 2015, W.W. Norton & Co.
(68) Thaler, R.H. & Sunstein, C.R., "Nudge: Improving Decisions About Health, Wealth and Happiness", 2008, Council Library Main Collection
(69) TRUSTe, "T. R . U. S .T. E.Us consumerconfidence privacy report: consumer opinion and business impact", 2014, Research Report, TRUSTe Inc.
(70) Turing, A. M. ; "Computing machinery and Intelligence-AM Turing",1950, Mind 49: 433-460
(71) Valdani, E., "Dalla concorrenza all'ipercompetizione dall'evoluzione alla coevoluzione", 1997, Economia & management 3, 81-93
(72) Wasserstein, R.L. e Lazar, N.A.; "The ASA's Statement on p-values: Context, Process, andpurpose";,2016, in The American Statistician, vol. 70, n.2, pp. 129-133
(73) Zani, S. & Cerioli, A.; "Analisi dei dati e data mining per decisioni aziendali", 2007, IBS

(74) Zhou, R., Khemmarat, S. e Gao, L., "The Impact of YouTube recommendation System on Video Views", 2010, Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, 404-410

(75) Zollo, S. et al.; "Intelligent sales and marketing reccomendation system", 2006, US Patent App. 11/370, 526

SITOGRAFIA:

1. Degrees of freedom, A. Ganti, 2019, accessibile da:https://www.investopedia.com/terms/d/degrees-of-freedom.asp
2. Report di Matt Ahlgren 2020 accessibile da https://www.websitehostingrating.com/it/twitter-statistics/?fbclid=IwAR1wvss-gF38igfPDcvPrYB2XsxX3-sadfueBWQWo-7vbG55N_c0O48NU04
3. 3. Report Digital 2020 accessibile da:"https://wearesocial.com/it/blog/2020/01/report-digital-2020-i-dati-global" condotto da We Are Social in collaborazione da Hootsuite

APPENDICE:

A) SCRIPT R

COMANDI R:

```
> library(tidyverse)
-- Attaching packages --------------------------------------- tidyverse 1.3.0 --
v ggplot2 3.3.0     v purrr   0.3.4
v tibble  3.0.1     v dplyr   0.8.5
v tidyr   1.0.2     v stringr 1.4.0
v readr   1.3.1     v forcats 0.5.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter()   masks stats::filter()
x dplyr::id()       masks twitteR::id()
x dplyr::lag()      masks stats::lag()
x dplyr::location() masks twitteR::location()
> library(twitteR)


> library(twitteR)
> fn_twitter<-searchTwitter("Privacy",n=3300,lang="en")
> fn_twitter_df<-twListToDF(fn_twitter)
> tweet_words<-fn_twitter_df %>% select(id, text) %>% unnest_tokens(word,text)


> library("dplyr", lib.loc="~/R/win-library/4.0")
> library("ggplot2", lib.loc="~/R/win-library/4.0")
> tweet_words<-fn_twitter_df %>% select(id, text) %>% unnest_tokens(word,text)


  > library("tidyr", lib.loc="~/R/win-library/4.0")
  > library(tidyverse)
```

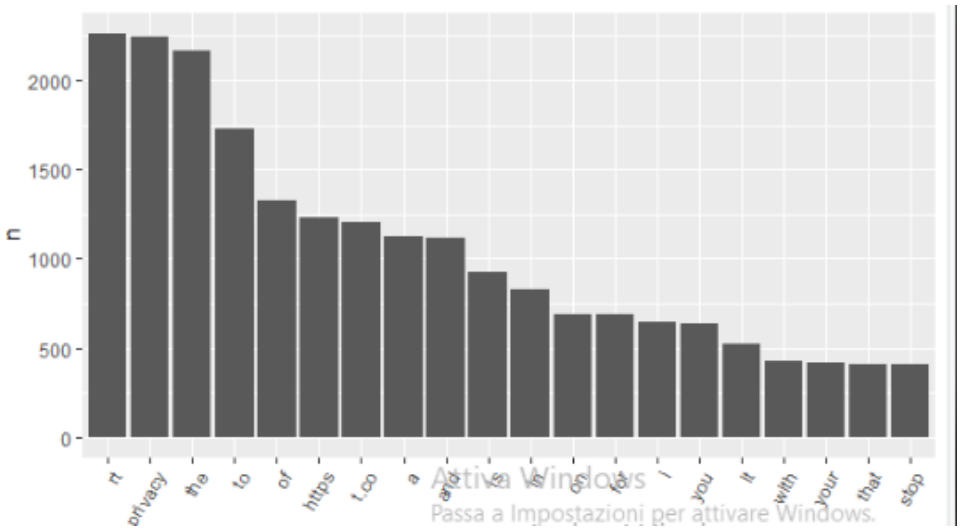
```
> library(tidytext)
> tweet_words<-fn_twitter_df %>% select(id, text) %>% unnest_tokens(word,text)
> tweet_words %>% count(word,sort=T) %>% slice (1:20) %>% ggplot(aes(x = reorder(word,
n, function(n) -n), y = n)) + geom_bar(stat="identity")+theme(axis.text.x = element_tex
t(angle=60,hjust = 1)) + xlab("")
>
> my_stop_words<-stop_words %>% select(-lexicon) %>% bind_rows(data.frame(word = c("htt
ps","rt","t.co","the","to","of","a")))
> tweet_words_interesting <- tweet_words %>% anti_join(my_stop_words)
Joining, by = "word"
> tweet_words_interesting %>% group_by(word) %>% tally(sort=TRUE) %>% slice(1:25) %>% g
gplot(aes(x=reorder(word,n, function(n) -n), y =n)) + geom_bar(stat="identity") + theme
(axis.text.x = element_text(angle = 60,hjust = 1)) xlab ("")
```

```
> tweet_words_interesting %>% group_by(word) %>% tally(sort=TRUE) %>% slice(1:25) %>% g
gplot(aes(x = reorder(word,n,function(n) -n), y = n)) + geom_bar(stat="identity") + the
me(axis.text.x = element_text(angle = 60, hjust=1)) + xlab("")
```

```
> tweet_words_interesting %>% group_by(word) %>% tally(sort=TRUE) %>% slice(1:25) %>% g
gplot(aes(x = reorder(word,n,function(n) -n), y = n)) + geom_bar(stat="identity") + the
me(axis.text.x = element_text(angle = 60, hjust=1)) + xlab("")
> bing_lex <- get_sentiments("nrc")
> fn_sentiment <- tweet_words_interesting %>% left_join(bing_lex)
Joining, by = "word"
> fn_sentiment %>% filter(!is.na(sentiment)) %>% group_by(sentiment) %>% summarise(n=n(
))
# A tibble: 10 x 2
   sentiment        n
   <chr>        <int>
 1 anger          644
 2 anticipation  1137
 3 disgust        343
 4 fear          1011
 5 joy           1011
 6 negative      1343
 7 positive      2870
 8 sadness        742
 9 surprise       309
10 trust         1934
> bing_lex <- get_sentiments("bing")
> fn_sentiment >_ tweet_words_interesting %>% left_join(bing_lex)
```

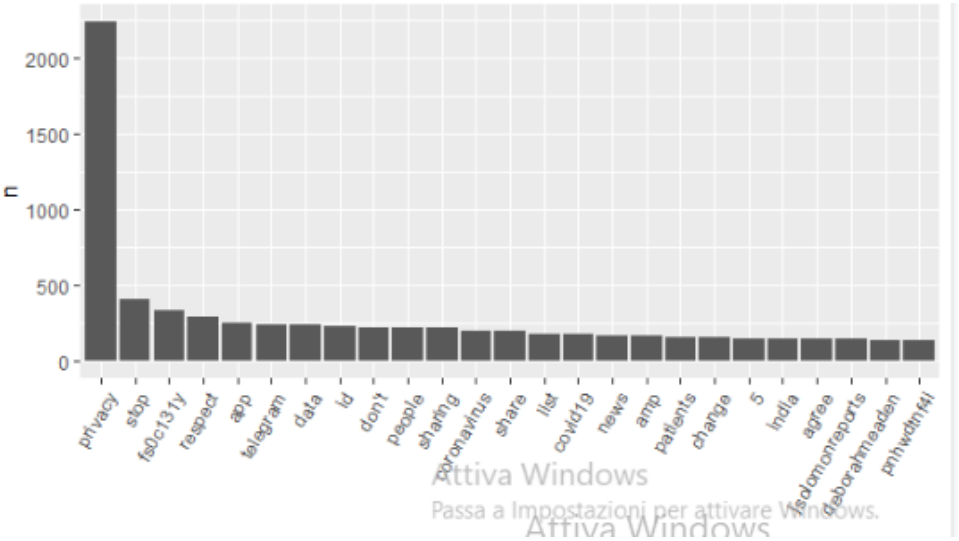
```
> fn_sentiment %>% filter(!is.na(sentiment)) %>% group_by(sentiment) %>% summarise(n=n(
))
# A tibble: 2 x 2
  sentiment     n
  <chr>     <int>
1 negative   1653
2 positive   1259
```

GRAFICO PRIMA STOPWORDS:



Attiva Windows
Passa a Impostazioni per attivare Windows.

GRAFICO POST STOPWORDS:



B) FLOW SURVEY

Start of Block: INTRO

Q1 Ciao! Sono uno studente LUISS e ti chiedo 5 minuti del tuo tempo per rispondere ad alcune semplici domande utili per un progetto di ricerca. Le risposte resteranno completamente anonime!

End of Block: INTRO

Start of Block: Categoria merceologica

Q2 Quale prodotto o servizio, tra questi, è quello che hai cercato o acquistato più spesso ultimamente su internet (motori di ricerca come ad esempio Google, Amazon o siti)?

- Libri (1)
- Videogiochi (2)
- Alloggi in hotel per vacanze (4)
- Non cerco nulla su Internet (5)

Q154 Con quanta frequenza fai questo acquisto nell'arco di un anno? (Indica il numero di volte)

_____

Q155 Mediamente quante ore spendi al giorno su internet in generale? (Indica il numero di ore)

_____

End of Block: Categoria merceologica

Start of Block: Libri

Q48 Quanto paghi mediamente un libro?

- Meno di 10 Euro (1)
- 10-30 Euro (2)
- Più di 30 euro (3)

Q3 Qual è il genere di libro che hai cercato/acquistato ultimamente? Alternativamente esprimi quale genere hai comprato più spesso in passato

- Giallo (1)
- Romanzo (2)
- Raccolta di poesie (3)
- Racconto (5)
- Libri universitari/scolastici (7)
- Nessuno tra questi (8)

Q4 Chi è tra questi l'autore del libro che hai cercato/acquistato ultimamente?

- Agatha Christie (1)
- Maurizio De Giovanni (2)
- Antonio Manzini (3)
- Andrea Camilleri (4)
- Stephen King (5)
- Altro (6)

Q5 Quale tra questi autori hai letto in passato maggiormente?

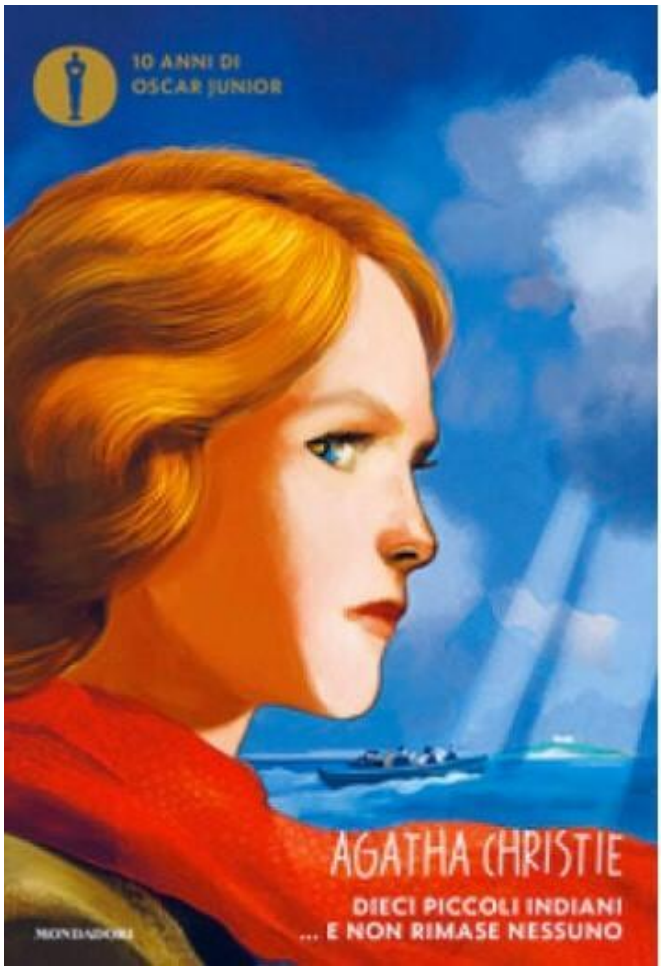
- Alessandro Baricco (1)
- Italo Calvino (2)
- Luigi Pirandello (3)
- Andrea Camilleri (4)
- George Orwell (5)
- Fedor Dostoevskij (6)
- Stephen King (7)
- Jane Austen (8)
- Elena Ferrante (10)
- Gabriel Garcìa Marquez (11)
- Umberto Eco (12)
- Lev Tolstoj (13)
- David Grossman (14)
- J.K. Rowling (15)
- Non ho mai letto nessuno tra questi (16)

Q6 Quale tra questi autori hai letto in passato maggiormente?

- Ernst Hemingway (1)
- Luigi Pirandello (2)
- Stephen King (3)
- Virginia Woolf (4)
- Italo Svevo (5)
- Andrea Camilleri (6)
- Fedor Dostoevskij (7)
- Non ho mai letto nessuno tra questi (8)

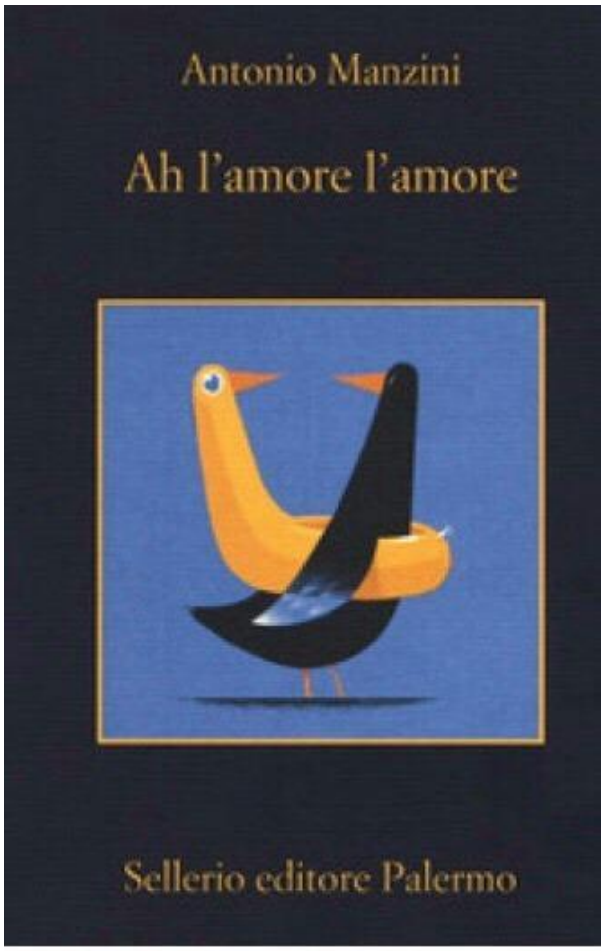
Q7 Quale tra questi autori hai letto in passato maggiormente?

- Eugenio Montale (1)
- Giacomo Leopardi (2)
- Umberto Saba (3)
- Alda Merini (4)
- Giuseppe Ungaretti (5)
- Charles Bukowski (6)
- Nicola Gardini (7)
- Erri De Luca (8)
- Pablo Neruda (9)
- Cesare Pavese (10)
- William Shakespeare (11)
- Nessuno tra questi (12)



Q8 Osserva questo giallo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

Page Break

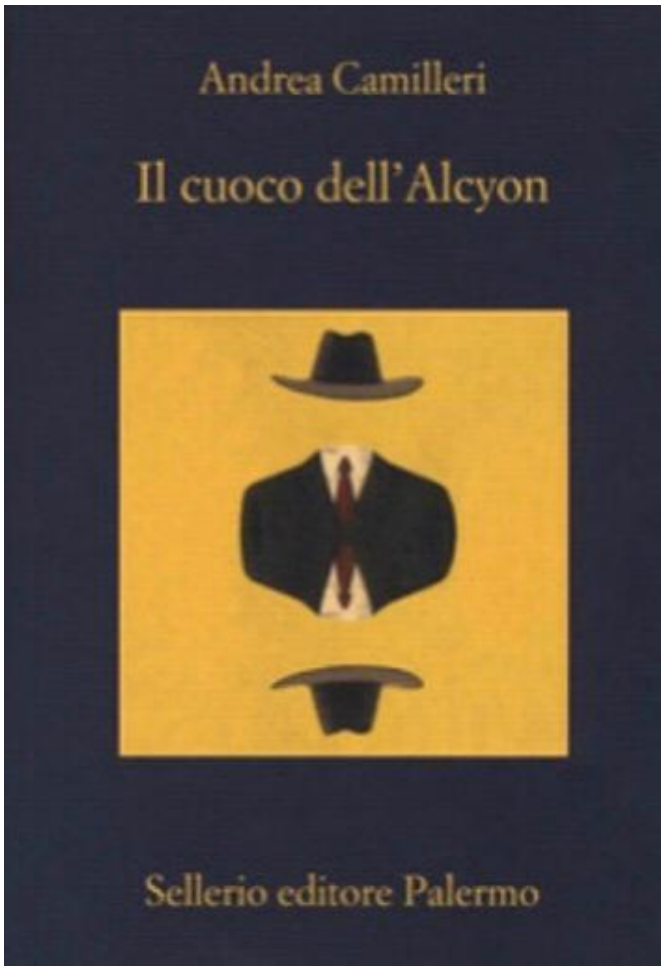


Q36 Osserva questo giallo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

Page Break

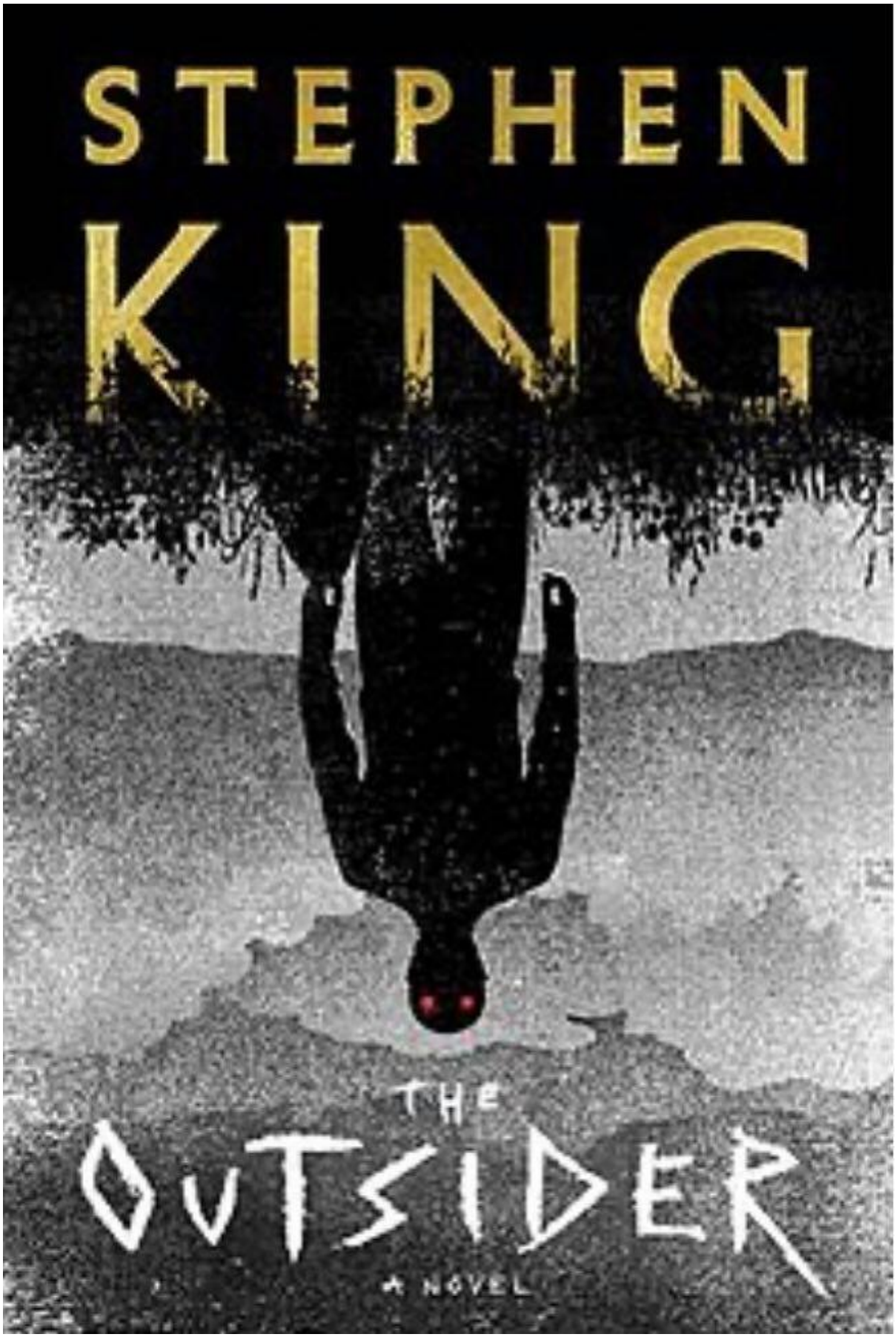
Q37 Osserva questo giallo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

Page Break



Q38 Osserva questo romanzo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

Page Break



Q39 Osserva questo romanzo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

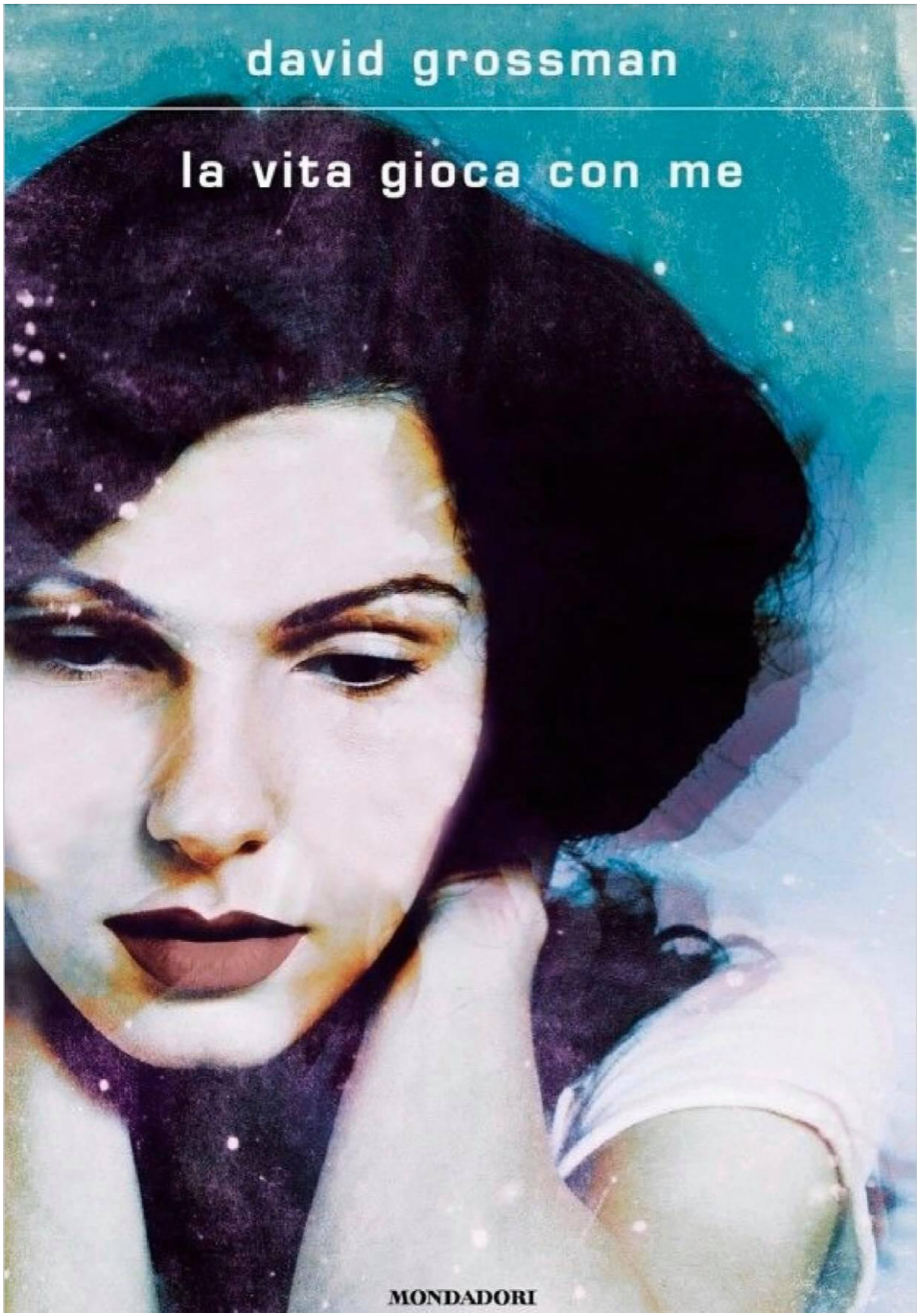
Page Break

Q57 Osserva questo romanzo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

Page Break

Q65 Osserva questo romanzo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

Page Break



Q69 Osserva questo romanzo ed immaginando che il suo prezzo sia simile all'ultimo romanzo che hai acquistato, rispondi alle prossime domande

Page Break

Q68 Immagina che J.K. Rowling abbia deciso di pubblicare il sequel di Harry Potter e che il suo prezzo sia simile a quanto hai pagato l'ultimo libro di questa serie, dopodiché rispondi alle seguenti domande

Page Break

Q58 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di Pirandello e rispondi alle domande successive

Page Break

Q77 Immagina che sia in vendita una raccolta di tutte le poesie di Eugenio Montale e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q78 Immagina che sia in vendita una raccolta di tutte le poesie di Giacomo Leopardi e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q79 Immagina che sia in vendita una raccolta di tutte le poesie di Umberto Saba e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q80 Immagina che sia in vendita una raccolta di tutte le poesie di Alda Merini e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q81 Immagina che sia in vendita una raccolta di tutte le poesie di Giuseppe Ungaretti e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q82 Immagina che sia in vendita una raccolta di tutte le poesie di Charles Buowski e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q83 Immagina che sia in vendita una raccolta di tutte le poesie di Nicola Gardini e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q84 Immagina che sia in vendita una raccolta di tutte le poesie di Erri De Luca e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q85 Immagina che sia in vendita una raccolta di tutte le poesie di Pablo Neruda e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q86 Immagina che sia in vendita una raccolta di tutte le poesie di Cesare Pavese e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q87 Immagina che sia in vendita una raccolta di tutte le poesie di William Shakespeare e, immaginando che il suo prezzo sia simile all'ultimo libro di poesie che hai acquistato, rispondi alle seguenti domande

Page Break

Q70 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo racconto che hai acquistato, puoi averne uno che non hai ancora letto di Pirandello e rispondi alle domande successive

Page Break

Q71 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo racconto che hai acquistato, puoi averne uno che non hai ancora letto di Ernst Hemingway e rispondi alle domande successive

Page Break

Q72 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo racconto che hai acquistato, puoi averne uno che non hai ancora letto di Stephen King e rispondi alle domande successive

Page Break

Q73 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo racconto che hai acquistato, puoi averne uno che non hai ancora letto di Virginia Woolf e rispondi alle domande successive

Page Break

Q74 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo racconto che hai acquistato, puoi averne uno che non hai ancora letto di Italo Svevo e rispondi alle domande successive

Page Break

Q75 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo racconto che hai acquistato, puoi averne uno che non hai ancora letto di Fedor Dostoevskij e rispondi alle domande successive

Page Break

Q76 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo racconto che hai acquistato, puoi averne uno che non hai ancora letto di Andrea Camilleri e rispondi alle domande successive

Page Break

Q67 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di Gabriel Garcìa Marquez e rispondi alle domande successive

Page Break

Q59 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di Italo Calvino e rispondi alle domande successive

Page Break

Q60 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di George Orwell e rispondi alle domande successive

Page Break

Q61 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di Fedor Dostoevskij e rispondi alle domande successive

Page Break

Q62 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di Jane Austen e rispondi alle domande successive

Page Break

Q63 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di Lev Tolstoj e rispondi alle domande successive

Page Break

Q64 Immagina di scoprire che per un prezzo simile a quello che hai pagato per l'ultimo romanzo che hai acquistato, puoi averne uno che non hai ancora letto di Umberto Eco e rispondi alle domande successive

Page Break

Q88 Immagina di trovare l'inserzione pubblicitaria del manuale del prossimo esame che devi sostenere (se sei uno studente liceale immagina si tratti di un libro richiesto da un tuo professore) e, immaginando il prezzo sia simile all'ultimo libro universitario/scolastico che hai acquistato, rispondi alle prossime domande

Page Break

Q47 Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni:

	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Completamente d'accordo) (7)
Sono disposto ad acquistare (1)	•	•	•	•	•	•	•
E' probabile che acquisti questo libro (2)	•	•	•	•	•	•	•
E' possibile che acquisti questo libro (3)	•	•	•	•	•	•	•

Q51 Già possedevi il libro suggerito nell'inserzione pubblicitaria?

- Sì (1)
- No (2)

End of Block: Libri

Start of Block: Atteggiamento

Insicurezza Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo"

	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Completamente d'accordo) (7)
Non consideri sicuro inserire il numero della carta di credito in un computer (1)	•	•	•	•	•	•	•
Il contatto umano è molto importante quando si fanno affari con un'azienda (2)	•	•	•	•	•	•	•
Non consideri sicuro	•	•	•	•	•	•	•

fare alcun tipo di affari finanziari online (3)							
Non ti senti sicuro nel fare affari con un posto che può essere raggiunto soltanto online (4)	•	•	•	•	•	•	•
Sei preoccupato che le informazioni che invii tramite internet possano essere viste da altre persone (5)	•	•	•	•	•	•	•

Rischio percepito Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo"

	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Completamente d'accordo) (7)
Il rischio che una terza parte non autorizzata supervisioni i processi di	•	•	•	•	•	•	•

pagamento che compio online è alto (1)							
Il rischio di abuso di utilizzo di informazioni è alto quando acquisto online (2)	•	•	•	•	•	•	•
Il rischio di abuso di informazioni di fatturazione è alto quando sono online (3)	•	•	•	•	•	•	•

End of Block: Atteggiamento

Start of Block: Videogiochi

Q9 Quale genere di videogiochi acquisti più spesso?

- Sportivi (ad es. FIFA, PES, ecc.) (1)
- Giochi di ruolo (ad es. The Witcher, Skyrim, Final Fantasy, ecc.) (2)
- Soul (ad esempio Nioh, Dark Souls, ecc.) (3)
- Picchiaduro (ad es. Tekken, Dragon Ball, ecc.) (4)
- Sparatutto (ad. es. Call of Duty, Battlefield, ecc.) (5)
- Strategia (ad es. Age of Empires, Civilization ecc.) (6)
- Open World (ad es. Assassin's creed, Red Dead Redemption, GTA, Metal Gear Solid V ecc.) (7)
- D'azione con supereroi (ad esempio Spiderman, Batman ecc.) (8)
- Di avventura/D'azione (ad es. Sekiro, Gof of War, Metal Gear Solid ecc.) (9)
- Nessuno tra questi (10)

Prezzo videogioco Preferisci aspettare che il prezzo di un giochi diminuisca, o di solito lo compri anche a prezzo pieno?

- Di solito aspetto qualche offerta (1)
- Di solito ho acquistato giochi a prezzo pieno (2)

Q10 Quale tra questi hai acquistato più spesso ultimamente? (Sono da considerare nella risposte tutte le versioni assimilabili a quel videogioco quindi per FIFA si intende, FIFAS 18, FIFA 19, FIFA 20 ecc.)

- FIFA (1)
- PES (2)
- NBA (3)
- Formula 1 (4)
- Madden NFL (5)
- EA Sports UFC (6)
- MotoGP (7)
- Bigben Interactive (8)
- Top Spin (9)
- Virtua Tennis (10)
- Nessuno tra questi (11)

Page Break

ruolo Quale tra questi hai acquistato più spesso ultimamente?

- The Witcher (1)
- Dark Souls (2)
- Nioh (3)
- Final Fantasy (4)
- Bloodborne (5)
- Skyrim (6)

- Sekiro (7)
- Fallout 76 (8)
- God of War (9)
- Altro (10)

Page Break

soul Quale tra questi hai acquistato più spesso ultimamente?

- Dark Souls (1)
- Nioh (2)
- Sekiro (3)
- Bloodborne (4)
- Hollow Knight (5)
- Altro (6)

Page Break

picchiad Quale tra questi hai acquistato più spesso ultimamente?

- Dragon Ball (1)
- WWE 2k (2)
- UFC (3)
- Tekken (4)
- Mortal Kombat (5)
- Street fighter (6)
- Naruto (7)
- Altro (8)

Page Break

sparat Quale tra questi titoli hai acquistato più spesso ultimamente?

- CoD (Call od Duty) (1)
- Battlefield (2)
- Destiny (3)
- Titanfall (4)
- Doom (5)
- Borderlands (6)
- Far Cry (7)
- Fortnite (8)
- Apex Legends (9)
- Resident Evil (10)
- Altro (11)

Page Break

strat Quale tra questi titoli hai acquistato più spesso ultimamente?

- Age of Empires (1)
- Civilization (2)
- Grepolis (3)
- Starcraft (4)
- Xcom (5)
- Rome: Total War (6)
- Altro (7)

Page Break

open world Quale tra questi titoli hai acquistato più spesso ultimamente?

- GTA (Grand Theft Auto) (1)
- Red Dead Redemption (2)
- Assassin's Creed (3)
- Mafia (4)
- Batman (5)
- Spiderman (6)
- Skyrim (7)
- The Witcher (8)
- Fallout (9)
- Metal Gear Solid V: The Phantom Pain (10)
- Altro (11)

Page Break

super Quale tra questi titoli hai acquistato più spesso ultimamente?

- Batman (1)
- Spiderman (2)
- Superman (3)
- Capitan America (4)
- X-men Legends (5)
- Altro (6)

Page Break

az./avv. Quale tra questi titoli hai acquistato più spesso ultimamente?

- Sekiro: die twice (1)
- Metal Gear Solid (2)
- God of War (3)
- Death Strandind (4)
- GTA (5)
- Red Dead Redemption (6)
- The last of us (7)
- Assassin's creed (8)
- Horizon (9)
- Resident Evil (10)
- Uncharted (11)
- Tomb Raider (12)
- Altro (13)

Page Break

Q89 A settembre uscirà FIFA 21 con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q95 A settembre uscirà FIFA 21 con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q92 A settembre uscirà PES2021 con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q96 A settembre uscirà PES2021 con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q94 Prossimamente uscirà una nuova edizione dell'ultimo videogioco sportivo che hai acquisto con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q97 Prossimamente uscirà una nuova edizione dell'ultimo videogioco sportivo che hai acquisto con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q100 In questi giorni è uscito Final Fantasy VII Remake. Il suo prezzo attualmente è di 69,99 Euro però immagina di trovarlo in offerta con uno sconto tra qualche settimana e rispondi alle prossime domande

Page Break

Q117 In questi giorni è uscito Final Fantasy VII Remake ed il suo prezzo è di 69,99 Euro. In riferimento a questo prodotto, rispondi alle prossime domande

Page Break

Q98 Ipotizza che a settembre esca Dark Souls IV con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q99 Ipotizza che a settembre esca Dark Souls IV con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q102 Ipotizza che a settembre esca Call of Duty: Black Ops V con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q103 Ipotizza che a settembre esca Call of Duty: Black Ops V con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q104 Ipotizza che a settembre esca uno dei più famosi giochi di strategia quale Age of Empires IV per ps4 e Xbox One con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q105 Ipotizza che a settembre esca uno dei più famosi giochi di strategia quale Age of Empires IV per ps4 e Xbox One con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q106 Ipotizza che a settembre esca Grand Theft (GTA) VI con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q107 Ipotizza che a settembre esca Grand Theft (GTA) VI con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q108 Ipotizza che a settembre esca un nuovo videogioco su Superman sullo stile del videogioco Marvel's Spider Man (eletto Game of the Year 2018) con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q111 Ipotizza che a settembre esca un nuovo videogioco su Superman sullo stile del videogioco Marvel's Spider Man (eletto Game of the Year 2018) con il prezzo di 69,99 Euro al day one.Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q109 Ipotizza che a settembre esca un nuovo videogioco su Superman con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q110 Ipotizza che a settembre esca un nuovo videogioco su Superman con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q112 Ipotizza che a settembre esca Metal Gear Solid VI con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q113 Ipotizza che a settembre esca Metal Gear Solid VI con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q114 Ipotizza che a settembre esca Mortal Kombat XII con il prezzo di 69,99 Euro al day one. Dopo qualche mese trovi un'offerta con uno sconto su questo prodotto. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q115 Ipotizza che a settembre esca Mortal Kombat XII con il prezzo di 69,99 Euro al day one. Pensando a questo videogioco rispondi alle prossime domande

Page Break

Q90 Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni:

	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Completamente d'accordo) (7)
Sono disposto ad acquistare questo videogioco (1)	•	•	•	•	•	•	•
E' probabile che acquisti questo videogioco (2)	•	•	•	•	•	•	•
E' possibile che acquisti questo videogioco (3)	•	•	•	•	•	•	•

End of Block: Videogiochi

Start of Block: Hotel

Q21 Ignorando il momento attuale in cui ci sono tantissime restrizione a causa del Coronavirus, dove vorresti fare un viaggio tra le città proposte?

- New York (1)
- Maldive (2)
- Parigi (3)
- Tokyo (4)
- Non mi piacerebbe visitare nessun posto tra questi (5)

Q23 Quando hai viaggiato in passato, quanto hai pagato in media un pernottamento con mezza pensione (prezzo a persona)?

- 40 Euro o meno (1)
- 40-70 Euro (2)
- 70-100 Euro (3)
- 100 Euro o più (4)

Q24 L'ultimo alloggio che hai prenotato, era vicino al centro-città o alle attrazioni del posto o vicino al mare (meno di 1 km)?

- Sì (1)
- No (2)

Q116 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo inferiore a 40 euro e distante meno di un km da New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q118 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo inferiore a 40 euro a New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q119 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra 40 ed i 70 euro a New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q120 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra 40 ed i 70 euro distante meno di un km dal centro di New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q121 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra 70 ed i 100 euro distante meno di un km dal centro di New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q123 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra 70 ed i 100 euro a New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q122 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 euro distante meno di un km dal centro di New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q124 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 euro a New York. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q125 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo inferiore a 40 euro alle Maldive. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q126 Immagina di trovare un'inserzione pubblicitaria di un ostello con prezzo inferiore a 40 euro alle Maldive distante meno di un km dal mare. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q127 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo tra i 40 ed i 70 euro alle Maldive distante meno di un km dal mare. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q128 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo tra i 40 ed i 70 euro alle Maldive. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q129 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo tra i 70 ed i 100 euro alle Maldive. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q130 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo tra i 70 ed i 100 euro alle Maldive distante meno di un km dal mare. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q131 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 euro alle Maldive. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q132 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 euro alle Maldive distante meno di un km dal mare. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q133 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 euro alle Maldive distante meno di un km dal mare. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q134 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo inferiore a 40 Euro a Parigi distante meno di un km dal centro città. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q135 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo inferiore a 40 Euro a Parigi. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q136 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 40 ed i 70 Euro a Parigi. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q137 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 40 ed i 70 Euro a Parigi distante meno di un km dal centro. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q138 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 70 ed i 100 Euro a Parigi distante meno di un km dal centro. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q139 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 70 ed i 100 Euro a Parigi. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q140 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 Euro a Parigi. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q141 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 Euro a Parigi distante meno di un km dal centro-città. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q142 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo inferiore a 40 Euro a Tokyo distante meno di un km dal centro-città. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q143 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo inferiore a 40 Euro a Tokyo. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q144 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 40 ed i 70 Euro a Tokyo. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q145 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 40 ed i 70 Euro a Tokyo distante meno di un km dal centro-città. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q146 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 70 ed i 100 Euro a Tokyo distante meno di un km dal centro-città. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q147 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo compreso tra i 70 ed i 100 Euro a Tokyo. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q148 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 Euro a Tokyo. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q149 Immagina di trovare un'inserzione pubblicitaria di un alloggio con prezzo maggiore a 100 Euro a Tokyo distante meno di un km dal centro-città. Considerando le caratteristiche dell'alloggio simili a quelle che hai trovato in un altro alloggio che hai pagato ad un prezzo simile, rispondi alle prossime domande

Page Break

Q91 Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni:

	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Estremamente d'accordo) (7)
Sono disposto ad acquistare (1)	•	•	•	•	•	•	•
E' probabile che pernotti in questo hotel (2)	•	•	•	•	•	•	•
E' possibile che pernotti in questo hotel (3)	•	•	•	•	•	•	•

End of Block: Hotel

Start of Block: Micro-moments pr no privacy

Q162 Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni:

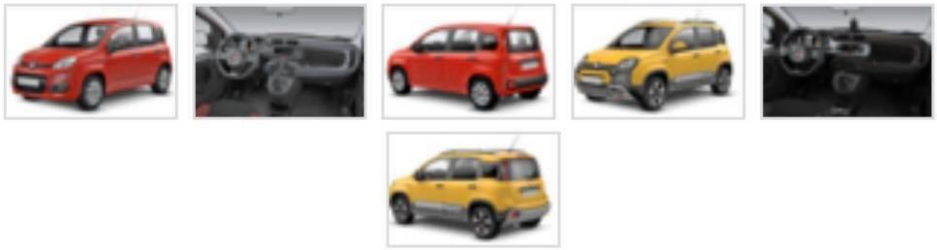
	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Estremamente d'accordo) (7)
Questa inserzione è stata utile (1)	•	•	•	•	•	•	•
Questa inserzione è stata rilevante (2)	•	•	•	•	•	•	•
Questa inserzione è stata appropriata (3)	•	•	•	•	•	•	•
Questa inserzione è stata pertinente (4)	•	•	•	•	•	•	•

End of Block: Micro-moments pr no privacy

Start of Block: Prodotto stnd

Q151 Adesso osserva questa automobile e rispondi alle domande seguenti

Fiat
Panda
da € 11.600
SCOPRI LA PROMOZIONE



Page Break

Q152 Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni:

	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Estremamente d'accordo) (7)
Sono disposto ad acquistare (1)	•	•	•	•	•	•	•
E' probabile che acquisti questa automobile (2)	•	•	•	•	•	•	•
E' possibile che acquisti questa automobile (3)	•	•	•	•	•	•	•

End of Block: Prodotto stnd

Start of Block: Micro-moments pr stnd

Q163 Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni:

	1 (Per niente d'accordo) (1)	2 (In disaccordo) (2)	3 (Parzialmente in disaccordo) (3)	4 (Né d'accordo né in in disaccordo) (4)	5 (Parzialmente d'accordo) (5)	6 (D'accordo) (6)	7 (Estremamente d'accordo) (7)
Questa inserzione è stata utile (1)	•	•	•	•	•	•	•
Questa inserzione è stata rilevante (2)	•	•	•	•	•	•	•
Questa inserzione è stata appropriata (3)	•	•	•	•	•	•	•
Questa inserzione è stata	•	•	•	•	•	•	•

pertinente (4)							

End of Block: Micro-moments pr stnd

Start of Block: Culturali

Q30 Ti è mai capitato di acquistare un prodotto dopo aver visto un'inserzione pubblicitaria su internet?

- Sì, mi è capitato (1)
- No, non mi è mai capitato (2)

Q157 Hai mai visto un film/una Serie TV su Netflix dopo che quest'ultimo te lo ha suggerito?

- Sì, mi è capitato (1)
- No, non mi è mai capitato (2)

Q31 Ti è mai capitato di trovare su internet l'inserzione pubblicitaria di un prodotto che avevi pensato nei giorni precedenti?

- Sì (1)
- No (2)

Q153 Ti è mai capitato di trovare inserzioni pubblicitarie su social network (ad es. Instagram, Youtube, Facebook, ecc.), su motori di ricerca (ad es. Google, bing, ecc.) o per mail?

- Sì (1)
- No (2)

Q32 Ti sei sentito violato/a della tua privacy oppure hai visto quell'inserzione pubblicitaria come un servizio aggiuntivo utile?

- Ho sentito violata la mia privacy (1)
- Ho trovato quell'inserzione utile (2)
- Nessuna delle due, l'ho vista come una coincidenza (3)

Q43 Le domande precedenti all'inserzione del libro le hai viste come una violazione della tua privacy?

- Sì (1)
- No (2)

Q44 Le domande precedenti all'inserzione del videogioco le hai viste come violazione della privacy?

- Sì (1)
- No (2)

Q45 Le domande precedenti all'inserzione dell'alloggio in hotel le hai viste come violazione della privacy?

- Sì (1)
- No (2)

End of Block: Culturali

Start of Block: Meme

Q159 No, non abbiamo ancora terminato ma mancano solo pochissime domande. Grazie per la pazienza!

End of Block: Meme

Start of Block: Demografiche

Q26 Qual è il tuo sesso?

- Maschio (1)
- Femmina (2)
- Preferisco non specificare (3)

Q27 Qual è la tua professione?

- Studente (1)
- Impiegato (2)
- Libero professionista (3)
- Dipendente (4)
- Pensionato (5)
- Disoccupato (6)
- Altro (7)

Q28 Quanti anni hai?

- 18 anni o meno (1)
- 18-21 (2)
- 21-25 (3)
- 25-30 (4)
- 30-40 (5)
- 40-50 (6)
- 50-60 (7)
- 60 o più (8)

Q29 Qual è il tuo titolo di studio?

- Nessuno (1)
- Licenza media inferiore (2)
- Licenza media superiore (3)
- Laurea triennale (4)
- Laurea magistrale o a ciclo unico (5)
- Dottorato (6)

End of Block: Demografiche

Start of Block: Outro

Q158 Ti ringrazio per il tempo dedicatomi! Il tuo supporto è stato fondamentale per la mia ricerca e spero di aver occupato un po' del tuo tempo durante questo periodo particolare e, anche se non sarà il migliore degli auguri, buona quarantena!

End of Block: Outro

C) OUTPUT SPSS:

AFFIDABILITA' DELLA SCALA INTENTION TO BUY LIBRI:

```
RELIABILITY
  /VARIABLES=Q47_1 Q47_2 Q47_3
  /SCALE('Intention to buy no privacy libri') ALL
  /MODEL=ALPHA.
```

Affidabilità

Scala: Intention to buy no privacy libri

Riepilogo elaborazione casi

		N	%
Casi	Valido	47	38,2
	Escluso[a]	76	61,8
	Totale	123	100,0

a. Eliminazione listwise basata su tutte le variabili nella procedura.

Statistiche di affidabilità

Alpha di Cronbach	N. di elementi
,940	3

AFFIDABILITA' DELLA SCALA INTENTION TO BUY HOTEL:

```
RELIABILITY
  /VARIABLES=Q91_1 Q91_2 Q91_3
  /SCALE('Intention to buy no privacy hotel') ALL
  /MODEL=ALPHA.
```

Affidabilità

Scala: Intention to buy no privacy hotel

Riepilogo elaborazione casi

		N	%
Casi	Valido	58	47,2
	Escluso[a]	65	52,8
	Totale	123	100,0

a. Eliminazione listwise basata su tutte le variabili nella procedura.

Statistiche di affidabilità

Alpha di Cronbach	N. di elementi
,917	3

AFFIDABILITA' DELLA SCALA INTENTION TO BUY VIDEOGIOCHI:

Affidabilità

Scala: Intention to buy videogiochi

Riepilogo elaborazione casi

		N	%
Casi	Valido	17	13,8
	Escluso[a]	106	86,2
	Totale	123	100,0

a. Eliminazione listwise basata su tutte le variabili nella procedura.

Statistiche di affidabilità

Alpha di Cronbach	Alpha di Cronbach basata su elementi standardizzati	N. di elementi
,967	,968	3

AFFIDABILITA' DELLA SCALA INTENTION TO BUY AUTOMOBILE:

Scala: Intention to buy si privacy automobile

Riepilogo elaborazione casi

		N	%
Casi	Valido	120	97,6
	Escluso[a]	3	2,4
	Totale	123	100,0

a. Eliminazione listwise basata su tutte le variabili nella procedura.

Statistiche di affidabilità

Alpha di Cronbach	Alpha di Cronbach basata su elementi standardizzati	N. di elementi
,911	,912	3

AFFIDABILITA' DELLA SCALA SICUREZZA:

Affidabilità

Scala: Sicurezza_atteggiamento

Riepilogo elaborazione casi

		N	%
Casi	Valido	122	99,2
	Escluso[a]	1	,8
	Totale	123	100,0

a. Eliminazione listwise basata su tutte le variabili nella procedura.

Statistiche di affidabilità

Alpha di Cronbach	Alpha di Cronbach basata su elementi standardizzati	N. di elementi
,793	,790	5

AFFIDABILITA' DELLA SCALA RISCHIO PERCEPITO:

Affidabilità

Scala: eiaschiopercepito_atteggiamento

Riepilogo elaborazione casi

		N	%
Casi	Valido	122	99,2
	Escluso[a]	1	,8
	Totale	123	100,0

a. Eliminazione listwise basata su tutte le variabili nella procedura.

Statistiche di affidabilità

Alpha di Cronbach	Alpha di Cronbach basata su elementi standardizzati	N. di elementi
,836	,837	3

STATISTICHE DESCRITTIVE LIBRI:

Statistiche degli elementi

	Media	Deviazione std.	N
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	5,70	1,366	47
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo libro	5,38	1,311	47
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questo libro	5,38	1,261	47

STATISTICHE DESCRITTIVE VIDEOGIOCHI:

Statistiche degli elementi

	Media	Deviazione std.	N
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare questo videogioco	4,82	2,298	17
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo videogioco	4,71	2,443	17
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questo videogioco	4,76	2,195	17

STATISTICHE DESCRITTIVE HOTEL:

Statistiche degli elementi

	Media	Deviazione std.	N
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	5,50	1,490	58
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che pernotti in questo hotel	5,64	1,252	58
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che pernotti in questo hotel	5,59	1,463	58

STATISTICHE DESCRITTIVE AUTOMOBILE:

Statistiche degli elementi

	Media	Deviazione std.	N
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	3,08	1,651	120
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questa automobile	2,68	1,506	120
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questa automobile	2,98	1,595	120

STATISTICHE DESCRITTIVE INSICUREZZA:

Statistiche degli elementi

	Media	Deviazione std.	N
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro inserire il numero della carta di credito in un computer	4,00	1,749	122
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il contatto umano è molto importante quando si fanno affari con un'azienda	5,45	1,355	122
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro fare alcun tipo di affari finanziari online	3,91	1,749	122
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non ti senti sicuro nel fare affari con un posto che può essere raggiunto soltanto online	4,31	1,632	122
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Sei preoccupato che le informazioni che invii tramite internet possano essere viste da altre persone	4,88	1,446	122

STATISTICHE DESCRITTIVE RISCHIO:

Statistiche degli elementi

	Media	Deviazione std.	N
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio che una terza parte non autorizzata supervisioni i processi di pagamento che compio online è alto	4,50	1,445	122
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di utilizzo di informazioni è alto quando acquisto online	4,85	1,389	122
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di informazioni di fatturazione è alto quando sono online	4,38	1,375	122

MATRICE DI CORRELAZIONE DEGLI ELEMENTI LIBRI:

Matrice di correlazione tra gli elementi

	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo libro	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questo libro
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	1,000	,793	,850
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo libro	,793	1,000	,882
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per	,850	,882	1,000

MATRICE DI CORRELAZIONE ELEMENTI VIDEOGIOCHI:

Matrice di correlazione tra gli elementi

	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare questo videogioco	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo videogioco	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questo videogioco
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare questo videogioco	1,000	,903	,859
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo videogioco	,903	1,000	,965
Indica su una scala da 1 a 7 (dove 1 sta per "per	,859	,965	1,000

MATRICE DI CORRELAZIONE ELEMENTI HOTEL:

Matrice di correlazione tra gli elementi

	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che pernotti in questo hotel	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che pernotti in questo hotel
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	1,000	,832	,764
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che pernotti in questo hotel	,832	1,000	,788
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente	,764	,788	1,000

MATRICE DI CORRELAZIONE ELEMENTI AUTOMOBILE:

Matrice di correlazione tra gli elementi

	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questa automobile	Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questa automobile
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	1,000	,750	,744
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questa automobile	,750	1,000	,832
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per	,744	,832	1,000

MATRICE CORRELAZIONE ELEMENTI INSICUREZZA:

Matrice di correlazione tra gli elementi

	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro inserire il numero della carta di credito in un computer	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il contatto umano è molto importante quando si fanno affari con un'azienda	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro fare alcun tipo di affari finanziari online	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non ti senti sicuro nel fare affari con un posto che può essere raggiunto soltanto online	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Sei preoccupato che le informazioni che invii tramite internet possano essere viste da altre persone
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro inserire il numero della carta di credito in un computer	1,000	,269	,465	,492	,474
Indica quanto sei	,269	1,000	,275	,377	,345

Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il contatto umano è molto importante quando si fanno affari con un'azienda	,269	1,000	,275	,377	,345
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro fare alcun tipo di affari finanziari online	,465	,275	1,000	,691	,394
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non ti senti sicuro nel fare affari con un posto che può essere raggiunto soltanto online	,492	,377	,691	1,000	,517
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente	,474	,345	,394	,517	1,000

MATRICE CORRELAZIONE ELEMENTI RISCHIO:

Matrice di correlazione tra gli elementi

	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio che una terza parte non autorizzata supervisioni i processi di pagamento che compio online è alto	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di utilizzo di informazioni è alto quando acquisto online	Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di informazioni di fatturazione è alto quando sono online
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio che una terza parte non autorizzata supervisioni i processi di pagamento che compio online è alto	1,000	,589	,578
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1	,589	1,000	,726
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di utilizzo di informazioni è alto quando acquisto online	,589	1,000	,726
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di informazioni di fatturazione è alto quando sono online	,578	,726	1,000

STATISTICHE ELEMENTO-TOTALE AFFIDABILITA' SE ITEM ELIMINATO LIBRI:

Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Correlazione multipla quadratica	Alpha di Cronbach se viene eliminato l'elemento
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	10,77	6,227	,846	,731	,937
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo libro	11,09	6,384	,869	,785	,918
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questo libro	11,09	6,427	,915	,840	,884

STATISTICHE ELEMENTO-TOTALE AFFIDABILITA' SE ITEM ELIMINATO VIDEOGIOCHI:

Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Correlazione multipla quadratica	Alpha di Cronbach se viene eliminato l'elemento
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare questo videogioco	9,47	21,140	,890	,818	,979
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questo videogioco	9,59	18,757	,968	,953	,924
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questo videogioco	9,53	21,390	,937	,933	,948

STATISTICHE ELEMENTO-TOTALE AFFIDABILITA' SE ITEM ELIMINATO HOTEL:

Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Correlazione multipla quadratica	Alpha di Cronbach se viene eliminato l'elemento
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	11,22	6,598	,841	,724	,875
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che pernotti in questo hotel	11,09	7,694	,863	,748	,866
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che pernotti in questo hotel	11,14	6,893	,810	,659	,901

STATISTICHE ELEMENTO-TOTALE AFFIIBILITA' SE ITEM ELIMINATO AUTOMOBILE:

Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Correlazione multipla quadratica	Alpha di Cronbach se viene eliminato l'elemento
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - Sono disposto ad acquistare	5,66	8,815	,780	,609	,908
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' probabile che acquisti questa automobile	6,05	9,191	,846	,731	,853
Indica su una scala da 1 a 7 (dove 1 sta per "per niente d'accordo" e 7 per "completamente d'accordo) quanto sei d'accordo con le seguenti affermazioni: - E' possibile che acquisti questa automobile	5,76	8,723	,840	,726	,855

STATISTICHE ELEMENTO-TOTALE AFFIDABILITA' SE ITEM ELIMINATO INSICUREZZA:

Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Correlazione multipla quadratica	Alpha di Cronbach se viene eliminato l'elemento
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro inserire il numero della carta di credito in un computer	18,55	22,349	,564	,334	,758
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il contatto umano è molto importante quando si fanno affari con un'azienda	17,10	27,296	,395	,175	,802
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro fare alcun tipo di affari finanziari online	18,64	21,588	,621	,497	,737
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non ti senti sicuro nel fare affari con un posto che può essere raggiunto soltanto online	18,24	21,207	,723	,573	,702
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Sei preoccupato che le informazioni che invii tramite internet possano essere viste da altre persone	17,67	24,454	,572	,349	,755

STATISTICHE ELEMENTO-TOTALE AFFIDABILITA' SE ITEM ELIMINATO RISCHIO:

Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Correlazione multipla quadratica	Alpha di Cronbach se viene eliminato l'elemento
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio che una terza parte non autorizzata supervisioni i processi di pagamento che compio online è alto	9,23	6,592	,628	,395	,841
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di utilizzo di informazioni è alto quando acquisto online	8,88	6,274	,738	,571	,732
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di informazioni di fatturazione è alto quando	9,35	6,379	,730	,562	,741

PAIRWED T-TEST INTENTION TO BUY:

Test t

Statistiche campioni accoppiati

		Media	N	Deviazione std.	Media errore standard
Coppia 1	Inttbuy_libri	5,4855	46	1,25454	,18497
	Inttbuy_automobile	2,8696	46	1,46478	,21597
Coppia 2	Inttbuy_videogiochi	4,7647	17	2,24154	,54365
	Inttbuy_automobile	2,9216	17	1,39209	,33763
Coppia 3	Inttbuy_hotel	5,6369	56	1,18965	,15897
	Inttbuy_automobile	2,8869	56	1,53119	,20461

Correlazioni campioni accoppiati

		N	Correlazione	Sign.
Coppia 1	Inttbuy_libri & Inttbuy_automobile	46	,229	,126
Coppia 2	Inttbuy_videogiochi & Inttbuy_automobile	17	,058	,824
Coppia 3	Inttbuy_hotel & Inttbuy_automobile	56	,315	,018

Test campioni accoppiati

		Differenze accoppiate					t	gl	Sign. (a due code)
		Media	Deviazione std.	Media errore standard	Intervallo di confidenza della differenza di 95% Inferiore	Superiore			
Coppia 1	Inttbuy_libri - Inttbuy_automobile	2,61594	1,69672	,25017	2,11208	3,11981	10,457	45	,000
Coppia 2	Inttbuy_videogiochi - Inttbuy_automobile	1,84314	2,56882	,62303	,52237	3,16390	2,958	16	,009
Coppia 3	Inttbuy_hotel - Inttbuy_automobile	2,75000	1,61589	,21593	2,31726	3,18274	12,735	55	,000

INDIPENDENT T-TEST ETA'-RISCHIO:

INDIPENDENT T-TEST ETA' INSICUREZZA:

Test t

Statistiche gruppo

	mill	N	Media	Deviazione std.	Media errore standard
Mea_insicurezza	1,00	96	4,2979	1,09583	,11184
	2,00	22	5,5545	1,04914	,22368

Test campioni indipendenti

		Test di Levene per l'eguaglianza delle varianze		Test t per l'eguaglianza delle medie						
		F	Sign.	t	gl	Sign. (a due code)	Differenza della media	Differenza errore standard	Intervallo di confidenza della differenza di 95% Inferiore	Superiore
Mea_insicurezza	Varianze uguali presunte	,289	,592	-4,888	116	,000	-1,25663	,25706	-1,76577	-,74749
	Varianze uguali non presunte			-5,025	32,366	,000	-1,25663	,25008	-1,76580	-,74746

INDIPENDENT T-TEST ETA'-RISCHIO:

Test t

Statistiche gruppo

	mill	N	Media	Deviazione std.	Media errore standard
Mea_risk	1,00	96	4,3819	1,10181	,11245
	2,00	22	5,6667	1,05910	,22580

Test campioni indipendenti

		Test di Levene per l'eguaglianza delle varianze		Test t per l'eguaglianza delle medie						
		F	Sign.	t	gl	Sign. (a due code)	Differenza della media	Differenza errore standard	Intervallo di confidenza della differenza di 95% Inferiore	Superiore
Mea_risk	Varianze uguali presunte	,048	,827	-4,967	116	,000	-1,28472	,25864	-1,79699	-,77246
	Varianze uguali non presunte			-5,093	32,270	,000	-1,28472	,25225	-1,79838	-,77107

FACTOR ANALYSIS ATTEGGIAMENTO:

Analisi fattoriale

Comunalità

	Iniziale	Estrazione
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro inserire il numero della carta di credito in un computer	1,000	,413
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il contatto umano è molto importante quando si fanno affari con un'azienda	1,000	,261
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro fare	1,000	,476
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non ti senti sicuro nel fare affari con un posto che può essere raggiunto soltanto online	1,000	,611
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Sei preoccupato che le informazioni che invii tramite internet possano essere viste da altre persone	1,000	,636
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente	1,000	,576

Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio che una terza parte non autorizzata supervisioni i processi di pagamento che compio online è alto	1,000	,576
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di utilizzo di informazioni è alto quando acquisto online	1,000	,634
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di informazioni di fatturazione è alto quando sono online	1,000	,601

Metodo di estrazione: Analisi dei componenti principali.

Varianza totale spiegata

	Autovalori iniziali			Caricamenti somme dei quadrati di estrazione		
Componente	Totale	% di varianza	% cumulativa	Totale	% di varianza	% cumulativa
1	4,208	52,599	52,599	4,208	52,599	52,599
2	,991	12,382	64,981			
3	,807	10,092	75,073			
4	,632	7,898	82,971			
5	,436	5,451	88,422			
6	,394	4,931	93,353			
7	,300	3,755	97,107			
8	,231	2,893	100,000			

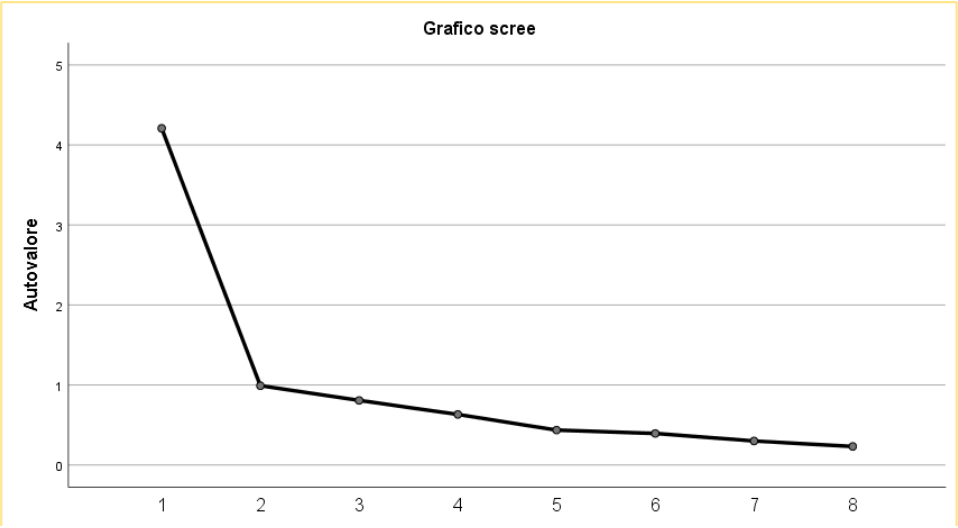
Metodo di estrazione: Analisi dei componenti principali.

Matrice dei componenti[a]

	Componente 1
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro inserire il numero della carta di credito in un computer	,642
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il contatto umano è molto importante quando si fanno affari con un'azienda	,511
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non consideri sicuro fare	,690
d'accordo" - Non consideri sicuro fare alcun tipo di affari finanziari online	
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Non ti senti sicuro nel fare affari con un posto che può essere raggiunto soltanto online	,781
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Sei preoccupato che le informazioni che invii tramite internet possano essere viste da altre persone	,798
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio che una terza parte non autorizzata supervisioni i processi di pagamento	,759

persone	
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio che una terza parte non autorizzata supervisioni i processi di pagamento che compio online è alto	,759
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di utilizzo di informazioni è alto quando acquisto online	,796
Indica quanto sei d'accordo con le seguenti affermazioni considerando che per 1 intendi "per niente d'accordo" mentre per 7 "completamente d'accordo" - Il rischio di abuso di informazioni di fatturazione è alto quando sono online	,775

Metodo di estrazione: Analisi dei componenti principali.



LUISS

Dipartimento di Impresa e Management
Corso di Laurea Magistrale in Marketing
Cattedra di Machine Learning & Object Driven Marketing

La Privacy ed il Machine Learning: Sentiment Analysis sulla percezione della cessione di dati a terzi e come varia il comportamento d'acquisto con stimoli di Recommendation System

Chiar.mo Prof. Luigi Laura
RELATORE

Chiar.mo Prof. Giuseppe Francesco Italiano
CORRELATORE

Francesco Moscatello - Matricola 703471
CANDIDATO

Anno Accademico 2019/2020

INDICE:

1. **INTRODUZIONE (pag. 135)**

2. **SENTIMENT ANALYSIS (pag. 137)**

   2.1 Overview (pag.137)

   2.2 Metodologia (pag. 138)

   2.3 Risultati (pag. 139)

   2.4 Limitazioni (pag. 139)

3. **VALUTAZIONE SPERIMENTALE: QUANTO IL RS INFLUENZA IL CUSTOMER (pag. 140)**

   3.1 Metodologia (pag. 140)

      3.1.1 Data Collection (pag. 141)

      3.1.2 Data cleaning (pag. 142)

      3.1.3 Formulazione delle ipotesi (pag. 143)

   3.2 Risultati (pag. 146)

   3.3 Limitazioni e spunti per ricerche future (pag. 147)

   CONCLUSIONI (pag. 148)

CAPITOLO 1

INTRODUZIONE

Negli ultimi anni il *Marketing*, e non solo, è stato rivoluzionato in maniera importante grazie all'avvento della *Digital Transformation* che rappresenta oggi una grandissima risorsa ed opportunità per tutte le aziende. La *Digital Transformation* può essere considerata come "*una vera e propria leva capace di abbattere gli intermediari privi di valore strategico, e di massimizzare le relazioni alla base dei nuovi modelli di business, che necessita di una rinnovata ed ancora più profonda consapevolezza delle nuove logiche del mercato alla base della relazione tra domanda ed offerta, dei processi interni su cui poggiano le imprese e dell'impatto che il digitale comporta per le persone coinvolte in questi processi*" (Ronchi, M. e Ciancia, M., 2019). Partendo da questo assunto, indubbiamente tutte le innovazioni digitali costituiscono uno strumento imprescindibile per ogni impresa non solo perché queste possono sfruttare le nuove tecnologie per i processi interni ma anche e soprattutto perché 4,54 miliardi sono le persone connesse ad internet di cui 3,8 mld utilizzano regolarmente *social network*[17] e quindi sono più facilmente reperibili i dati di ogni consumatore. I *Big Data*[18], infatti, sono considerati una *commodity* che consente alle imprese di offrire ai consumatori servizi innovativi (Pitruzzella, G., 2016). Tuttavia una forte limitazione alla possibilità di usufruire dei *Big Data*, per le imprese, e di beneficiare di offerte maggiormente customer-oriented[19] da parte di queste, per i consumatori, è indubbiamente la *Privacy* che come suggerisce gran parte della letteratura è un diritto a questi ultimi molto caro specie dopo alcuni avvenimenti come ad esempio lo scandalo di *Cambridge Analytica* che ha generato per Facebook la perdita di più di 3 milioni di utenti in Europa e circa 120 miliardi di dollari di valore di mercato (Isaak e Hanna, 2018). Tuttavia la letteratura accademica ha evidenziato il cosiddetto "paradosso della *Privacy*" (Norberg et al., 2007) secondo cui gli utenti online dichiarano, da una parte, di essere preoccupati dal raccoglimento ed utilizzo ambiguo da parte delle imprese (Bansal et al., 2010; Raiini et al., 2013; TRUSTe, 2014, Smith H.J. et al., 1996), dall'altra gli stessi utenti si mostrano disponibili a cedere i propri dati per compensi di poco conto come ha dimostrato Carrascal et al. (2013) che ha rilevato che diversi utenti hanno accettato di valutare la propria cronologia di navigazione online 7 euro cioè quanto un *Big Mac*. Questa contraddizione sembra seguire la scia degli studi di Kanheman e Tversky (2011), confermati poi in ambito economico dal premio Nobel R. Thaler (2008; 2015) che identificano un'irrazionalità dei consumatori in

[17] (Report Digital 2020 accessibile da:"https://wearesocial.com/it/blog/2020/01/report-digital-2020-i-dati-global" condotto da We Are Social in collaborazione da Hootsuite).

[18] all'inizio degli anni 2000 l'analista di mercato Doug Laney ha definito i big data con le tre V ovvero Volume (le organizzazioni raccolgono dati da diverse fonti come ad esempio dispositivi intelligenti nell'ambito IoT*, social network, transazioni commerciali ecc.), Velocità (grazie alla crescita dell'IoT i flussi di dati vengono dati alle imprese velocemente) e Varietà (i dati sono disponibili in tutti i tipi di formati).
*Per IoT si intende un sistema di dispositivi informatici correlati, macchine e digitali, oggetti, animali o persone dotati di identificatori univoci (uid) e la capacità di trasferire dati su una rete senza richiedere da uomo a uomo o interazione uomo-computer (V. Teigens, P. Skalfist, D. Mikelsten, 2019).

[19] L'orientamento al consumatore è la base dell'apprendimento organizzativo che produce come effetto una percezione del valore ed una soddisfazione del clienti maggiori dell'azienda che utilizza questo approccio rispetto ai competitors (Singula, J. M., Baker, W. E., e Noordewier, T., 1997; Slater, S. F. e Narver, J. C., 1995; Brady, M. K.; Cronin Jr., J. J., 2001).

diversi ambiti. In definitiva questo studio si propone di offrire alla letteratura un nuovo contributo circa la percezione degli utenti online della cessione di dati personali a terzi attraverso una *Sentiment Analysis* condotta considerando i *tweets* relativi al suddetto topic sul *social network Twitter* che vanta 336 milioni di utenti attivi mensili, 500 mln di *tweets* pubblicati ogni giorno e 1,3 miliardi account *Twitter* di cui però 328 mln attivi (Ahlgren, M., 2020). Inoltre la seconda parte della ricerca avrà come scopo quello di verificare come l'applicazione di algoritmi di *Machine Learning*, nello specifico il *Recommendation System* che per essere efficace necessita di una grande mole di dati, influiscono sul *Marketing* in termini di *intention to buy*. Questa seconda parte verrà analizzata sfruttando un esperimento condotto tramite un sondaggio lanciato attraverso il *software Qualtrics* dove verrà valutato come varia il comportamento d'acquisto dei consumatori sotto stimoli di *Recommendation System*. Il *Recommendation System* è quel sistema che sfruttando algoritmi di *Machine Learning,* in base ai dati in suo possesso, suggerisce agli utenti o consumatori, un determinato *output* (maggiori approfondimenti nell'intero studio). Questa seconda parte della ricerca parte dall'assunto che il *Machine Learning* (d'ora in poi ML) fornisce un importante supporto al *Marketing* (S. Zollo et al., 2006; Zhou, R., Khemmarat, S. e Gao, L, 2010). In particolare gli autori del secondo studio hanno analizzato che impatto ha il *Recommendation System* utilizzato da *YouTube* sulle visualizzazioni dei video all'interno della stessa piattaorma. All'interno di questa ricerca è stato rilevato che circa il 30% delle visualizzazioni complessive derivavano da una raccomandazione del relativo video per cui è possibile che i video analizzati, totalmente o in parte, siano stati visti grazie al fatto che siano stati suggeriti. Inoltre, un ulteriore finding di questo studio suggerisce che il *Recommendation System* incentiva a visualizzare video completamente diversi rispetto a quello alla fine del quale viene mostrato il video suggerito e questo dimostra che questa applicazione di ML permette di conoscere meglio i consumatori e di offrire loro video maggiormente di loro interesse rispetto a video *mainstream*. Partendo dai risultati di questo studio, questa parte della ricerca ha voluto estendere i risultati su azioni più dispendiose della semplice visualizzazione come l'*intention to buy* che è la variabile di riferimento. Prima di procedere oltre, occorre chiarire ulteriormente cosa si intende per ML, per il quale si possono trovare diverse definizioni approfondite nell'intero studio. In generale si tratta di un termine coniato dal 1959 da Arthur Samuel che ha distinto due approcci distinti di questa branca dell'*Intelligenza Artificiale* (d'ora in poi IA). Anzitutto, quindi occorre prima chiarire cosa si intende per IA e per farlo, poiché è un argomento di recente rilevanza per cui trovare una definizione univoca appare abbastanza complicato, risulta opportuno rifarsi alle definizione di questa da parte di diversi autorevoli autori: *"L'obiettivo del soggetto è identificare risolvibili ed interessanti problemi di elaborazione delle informazioni, e risolverli*" (David Marr, 1976); "*l'intelligente connessione della percezione all'azione*" (Rodney Brooks, 1991); *"le azioni che non sono distinguibili da quelle umane*" (Alan Turing, 1950). In definitiva, cercando di sfruttare tutte le precedenti definizioni, l'IA può essere definita come lo studio dei problemi di elaborazioni complesse di informazioni che spesso hanno le loro radici in alcuni aspetti di elaborazione di informazioni biologiche. Chiarito questo, è possibile proseguire nella spiegazione dei due metodi con cui opera il ML in cui il primo, identificato come rete neurale, porta a sviluppare delle macchine alle quali viene fornito apprendimento automatico per determinati impieghi generali (*unsupervides learning*) in cui il comportamento è appreso da una rete di

commutazione connessa in modo random a seguito di una routine di apprendimento basata su ricompensa (in caso di successo) e punizione (in caso di insuccesso) (*reinforcement learning*). Il secondo metodo invece consiste nella programmazione specifica di determinate attività (*supervised learning*). In conclusione l'obiettivo generale della ricerca è quello di comprendere che *sentiment* hanno gli utenti online circa la cessione di dati personali a terzi e se, ed eventualmente in che misura, il *Recommendation System* (che sfrutta i dati forniti dagli utenti stessi) riesce a dare alle imprese la possibilità di restituire ai propri clienti prodotti maggiormente *customer-oriented* e quindi più in linea con i *needs* di questi ultimi. Nell'elaborazione del secondo studio, ho ritenuto, inoltre, opportuno effettuare altre ricerche correlate che verranno meglio spiegati nel capitolo 3. Perciò le *Research Questions* dell'intero studio sono le seguenti:

- Che *sentiment* hanno le persone circa la *Privacy*?
- L'*intention to buy* di un prodotto la cui inserzione pubblicitaria viene mostrata ad un determinato consumatore in base ai suoi dati varia in maniera statisticamente significativa rispetto ad una non basata sui dati?
- C'è una differenza statisticamente significativa del tratto personale insicurezza online tra i *millenials* e le generazioni precedenti?
- Lo stesso discorso dell'ultimo punto può essere fatto in relazione alla variabile "rischio percepito" in termini di sicurezza online?

CAPITOLO 2

SENTIMENT ANALYSIS

2.1 Overview Una buona parte dei dati di cui dispongono le aziende sono strutturati, numerici o categorici e le tecniche utilizzate per analizzare queste tipologie di dati sono tutte abbastanza avanzate. Tuttavia L'IA ha cercato di sviluppare ulteriori tecniche per analizzare anche dati qualitativi, poco strutturati, testuali o multilingue come ad esempio i contenuti che vengono pubblicati sui *social networks*. Sostanzialmente la differenza consiste nel fatto che i dati strutturati possono essere memorizzati in un database *SQL* (*Structured Query Language*) (Elmasri, 2010) mentre quelli non strutturati sono difficilmente assimilabili dagli algoritmi utilizzati dalle macchine. Un esempio in questo senso sono i *tweets*, ovvero i contenuti che gli utenti di *Twitter* pubblicano attraverso questo *social network*. Dati di questo tipo potrebbero fornire alle imprese, e non solo, informazioni molto interessanti (basti vedere quanto scritto nell'Introduzione circa il numero di *tweets* pubblicati al giorno) ed è per questo che si sono sviluppate tecniche come ad esempio la *Sentiment Analysis* che si occupa proprio di analizzare questo tipo di dati che per loro natura richiedono un grado di complessità maggiore. Questo tipo di analisi rientra nel campo del *Machine Learning* (maggiormente approfondito nell'intero studio) per cui viene fornita ad una macchina una metodologia di apprendimento automatico che gli consente di imparare ad analizzare i dati che riceve senza input dal programmatore. Sostanzialmente alla macchina viene solamente insegnato come apprendere in maniera automatica in modo che questa sviluppi, attraverso i propri algoritmi, una logica con cui analizzare i dati senza che ci sia l'intervento umano che spieghi alla macchina cosa fare passaggio per passaggio. Il campo specifico che si riferisce all'apprendimento della macchina di comprendere dati testuali, è denominato tecnicamente come *Natural Language Processing*.

Questo ambito si propone come obiettivo quello di comprendere e trarre significato dal linguaggio umano che risulta per sua natura ben più complesso di quello informatico. Gli albori di questi ambito sono identificabili nello studio condotto da *Alan Turing* nel 1950 ("*Computing machinery and Intelligence-AM Turing*") in cui l'illustre informatico ha proposto un settaggio dell'*Intelligenza Artificiale* affinché questa potesse prendere parte ad una conversazione in tempo reale con un umano in modo tale che la macchina interagisca come se fosse un altro uomo, cioè in modo tale che sia indistinguibile rispetto a questo. Applicazioni abbastanza note di questo campo le vediamo oggigiorno ad esempio con *Siri* o *Google Now*. Alla luce di tutto, lo scopo della *Sentiment Analysis* di questa ricerca, è quello di comprendere se gli utenti di *Twitter* hanno sentiment positivo, negativo o neutro circa la *Privacy* e quali emozioni vengono maggiormente evocate tra i vari *tweets*.

2.2 Metodologia Sentiment Analysis Per condurre la *Sentiment Analysis* mirata a comprendere l'opinione generale delle persone circa la *Privacy*, è stato utilizzato il programma statistico *R* per le ragioni spiegate nel paragrafo 1.4 dell'intero studio. Anzitutto su questo programma sono stati installati i pacchetti "twitteR", "tidytext", "dplyr" e "ggplot2", dopodiché prima di proseguire sullo stesso programma è stata mandata a *Twitter* la richiesta di poter utilizzare le sue API (come spiegato nel paragrafo 1.1 dell'intero studio) per scopi di ricerca. Subito dopo averla ottenuta, è stato implementato il pacchetto "twitteR" da cui è partita la fase di data scaping[20] per cui su R è stato creato il collegamento con l'account *Twitter* che ha avuto l'autorizzazione di utilizzare le API del suddetto social network. Una volta che ad *R* è stato associato l'accesso alle *API* del *social network*, è iniziata la *data collection* attraverso la funzione "s*earchTwitter*" per cui è stato richiesto al programma di raccogliere 3300 *tweets* contenenti l'hashtag "*Privacy*" e in lingua inglese. E' importante anche specificare la lingua poiché *R* riesce ad identificare i dati testuali in base al lessico che gli viene fornito e spesso questo si riferisce ad una specifica lingua per cui collezionare dati misti potrebbe risultare più complicato in quanto richiederebbe un caricamento del lessico sul programma manuale. Una volta che *R* ha ottenuto il suo *dataset*, questo è stato trasformato in un *dataframe*[21] scomposto in vettori di uguale lunghezza. Successivamente è stata effettuata le *tokenizzazione*[22] per cui i dati testuali sono stati ricondotti a *token*. Nel grafico seguente (fig. 1) è rappresentato il *dataframe* delle parole con le relative frequenze assolute con cui si sono presentate. E' così che è stato generato un *dataframe* secondo il metodo "*Bag of Words*" per cui viene appunto conteggiato quante volte si ripete una determinata parola. Considerando proprio questa metodologia appare utile effettuare una fase di "*stemming*" per cui vengono conteggiate allo stesso modo parole contenti la stessa radice per evitare eventuali eccezioni o errori di battitura da parte degli utenti. Prima ancora però appare utile evitare *overloading* di dati e per questo scopo è stato effettuato il cosiddetto procedimento di "*stopwords*" per cui sono state eliminate tutte quelle parole non funzionali all'analisi. E' evidente infatti che in un *tweet* sono contenute anche parole come ad esempio congiunzioni, preposizioni che servono a dare senso alla frase ma che, per permettere alla macchina di identificare il *sentiment*, non risultano utili. Questa infatti, in base al

[20] è una tecnica in cui un computer estrae dati dall'output leggibile dall'uomo proveniente da un altro programma (S. Munzert et al., 2014)

[21] è una lista di vettori (che si riferiscono alle variabili considerate, quindi nel nostro caso le frequenze assolute e le parole) che devono avere tutti la stessa lunghezza, ma possono essere di tipo diverso cioè sia qualitative che quantitative (K. Preiss, 1980)

[22] processo in cui i caratteri vengono rotti in parole chiamati token (J.E. Pentheroukadis, D. G. Bradlee, S. S. Knoll, 2006)

lessico fornito, associa ad esempio l'aggettivo "buono" come un qualcosa che può essere classificato come positivo pertanto tutte le parole all'interno della frase non utili in questo senso possono essere eliminate. Effettuata quindi questa fase di *data cleaning*, è stato fornito al programma il lessico "*Bing*" e quello "*NRC*" per approfondire anche circa le emozioni emerse dai *tweets* e, con il comando "get_sentiment", è stata realizzata la clusterizzazione per cui ogni parola è stata associata ad una polarità (positiva, neutrale o negativa) e ad un sentimento. Entrambi i lessici sono stati creati grazie ad attività di *crowdsourcing* di diversi autori e sono stati convalidati utilizzando nuovamente nuove combinazioni di *crowdsourcing*, inoltre ovviamente non tutte le parole (che si rifanno alla lingua inglese) sono inserite nei lessici poiché molte considerate neutro. I risultati sono stati descritti nel paragrafo 1.7 dello studio completo.

2.3 Risultati Sentiment Analysis Il sentimento generato dalle analisi grazie al comando "get_sentiments" è stato trovato in base ai due lessici "*Bing*" ed "*NRC*" che si basano effettuano l'analisi su una sola parola. I risultati dei due lessici, come si nota dalla figura (fig. 1) sono discordanti e questa dimostra che le opinioni sono abbastanza bilanciate e che c'è un dibattito abbastanza accesso e nel complesso confusionario.

```
# A tibble: 2 x 2
  sentiment     n
  <chr>     <int>
1 negative   1653
2 positive   1259
```

```
   sentiment        n
   <chr>        <int>
 1 anger          644
 2 anticipation  1137
 3 disgust        343
 4 fear          1011
 5 joy           1011
 6 negative      1343
 7 positive      2870
 8 sadness        742
 9 surprise       309
10 trust         1934
```

Figura 1 Output Sentiment Analysis lessico Bing (sx) e lessico NRC (dx)

Un altro input interessante tra i risultati suggerisce che le emozioni maggiormente evocate sono a la fiducia e la gioia, tra quelle positive, e l'ansia e la paura tra quelle negative. Questo "secondo livello" di analisi sulle emozioni è reso possibile grazie al lessico "*NRC*" che categorizza le parole in base alle seguenti 8 emozioni: rabbia, ansia, disgusto, paura, gioia, tristezza, sorpresa e fiducia. Considerando il carattere molto esplorativo dell'analisi condotta, i findings principali si rifanno proprio al fatto che vengono espresse parole di felicità e fiducia circa la *Privacy* il ché ci restituisce come input che le persone considerano una normativa sulla protezione dei dati come qualcosa che rende felici le persone ed inoltre queste si sono mostrate fiduciose sull'efficacia attuale di tali normative. Dal lato, invece, delle parole negative, gli aspetti che andrebbero maggiormente curati sono l'ansia e la paura che fanno comprendere come una fuga dei dati personali costituisca effettiva preoccupazione per le persone.

2.4 Limitazioni La *Sentiment Analysis* condotta presenta indubbiamente alcune limitazioni che sono riportate di seguito. Anzitutto l'analisi è stata effettuata su unigrammi per cui ogni parola contenuta tra i vari *tweets* è stata classificata come positiva, negativa o come espressione di una delle 8 emozioni come meglio descritto nell'intero studio. Pertanto non sono stati considerati i singoli tweets che si rifanno ad un determinato utente ma tutte le parole contenuti nei vari *tweets* sono state inserite in un unico "calderone" (*Bag of Words*) che ha poi classificato le singole parole. Tuttavia questo problema è parzialmente limitato dal fatto che i *tweets* hanno lunghezza massima di 140 caratteri quindi l'incidenza di diverse parole positive o negative da parte di un determinato utente ha potuto incidere in maniera limitata. Inoltre, su un *Social Network* potrebbe risultare utile

anche effettuare un'analisi che categorizzi anche le *emoticon*[23]. I risultati delle analisi anche hanno rilevato come nei due lessici ci sia stata una diversa categorizzazione per cui potrebbe risultare utile per studi futuri effettuare una analisi più robusta che riesca ad associare in maniera più adeguata il pensiero espresso nel singolo *tweet* al suo autore in modo da spostare l'analisi da quante parole positive o negative ci sono a quanti autori ci sono con *sentiment* positivo o negativo circa il *topic* di riferimento. A tale scopo, poiché ogni aggettivo viene classificato solamente come positivo o negativo, potrebbe essere utile una clusterizzazione per cui in maniera più adeguata viene associato un aggettivo ad una determinata caratteristica del topic. Ad esempio nel caso specifico, la *Privacy* può essere interpretata come adeguata o come qualcosa di necessario ma nell'analisi condotta in questo studio questa differenziazione non viene fatta. Se ad esempio grazie ad una attività di clusterizzazione (approfondimenti nell'intero studio) si associa la parola "adeguato" all'idea per cui si ritiene che la *Privacy* oggi è correttamente gestita e "buona" come un qualcosa riferibile alla sua condizione, sicuramente l'analisi appare più accurata. Un'altra limitazione dell'analisi, dovuta alla natura del programma utilizzato per effettuarla, è che non considera, ad esempio, qualificatori della parola come negazione (cioè ad esempio non considera "non" prima dell'aggettivo), ma si basa esclusivamente, come già spiegato, sugli unigrammi. Un ulteriore problema causato dalla "natura quantitativa" di *R* è che risulta anche complicato qualificare correttamente *tweets* ironici o sarcastici. Tuttavia il focus principale dell'analisi era quello di comprendere in che modo algoritmi del ML ed in particolare il *Recommendation System* potessero risultare utili in ambito marketing mentre la *Sentiment Analysis* ha avuto semplicemente uno scopo esplorativo per cercare di comprendere quali emozioni fossero associate ad esse e se tendenzialmente in ambito online si abbia una attitudine positiva o negativa verso la cessione di dati a terzi. Come suggerito successivamente nelle Conclusioni, infatti, in base ai risultati, questo studio potrebbe offrire uno spunto per le aziende comprendere se effettivamente possa essere conveniente una maggiore comunicazione ai consumatori circa la cessione di dati online data l'estrema confusione che anche i risultati di questa analisi, oltre a quanto già riscontrabile in letteratura, suggeriscono.

CAPITOLO 3

L'UTILIZZO DI ALGORITMI DI MACHINE LEARNING NEL MARKETING

3.1 Metodologia Considerando il potenziale degli algorimti di ML nell'ambito *Marketing*, ho voluto effettuare uno studio per sondare se, ed eventualmente in che modo, l'intention to buy di un'inserzione pubblicitaria che sfrutta i nostri dati (quindi su stimolo derivante da output generati da algoritmi di ML) differisce da una che rispetta maggiormente il diritto alla *Privacy* individuale del consumatore (rifacendosi quindi alle inserzioni pubblicitarie che vediamo nei canali tradizionali come ad esempio la televisione). Per effettuare tale esperimento ho ideato un questionario su *Qualtrics* la cui metodologia viene chiarita maggiormente nell'intero studio. Il numero di rispondenti è stato pari a 153 persone ma il campione considerato è stato minore come spiegato meglio nel sottoparagrafo 3.1.2 dove viene spiegato il processo di *data cleaning*. In sede di analisi, ho effettuato un *paired samples t-test* poiché entrambe le inserzioni pubblicitarie sono state visualizzate da tutti i dipendenti. Inoltre, per completezza, ho voluto approfondire l'analisi poiché, come si apprende dalla

[23] Riproduzioni stilizzate delle espressioni umane che esprimono una determinata emozione (Di Pasqua, M. 2009)

letteratura (N. Howe & W. Strauss 2007, 2009), i *Millenials* (o generazione Y), ovvero coloro che sono nati tra l'inizio degli anni '80 ed i primi anni 2000 (N. Howe & W. Strauss, 2000), hanno un rapporto verso l'ambiente online e delle caratteristiche differenti dalle generazioni precedenti (David C. Edelman, 2010, 2015). Perciò ho effettuato un *indipendent samples t-test* per verificare se il rischio percepito online e la sicurezza percepita online sono significativamente differenti tra i due gruppi. Ho cercato anche di "accorpare" le due scale precedentemente citate in un unico fattore quale "l'atteggiamento online" ma i risultati della *Factor Analysis* non hanno permesso tale operazione.

3.1.1 Data collection Per raccogliere i dati sulle analisi proposte, ho anzitutto preparato una *Survey* il cui *flow* è in Appendice (blocco B). Per la strutturazione di questa, dopo una breve intro di presentazione, ho cercato di raccogliere alcuni dati sui rispondenti chiedendo loro quale fosse l'acquisto o la ricerca più frequente fatti online tra tre categorie merceologiche o servizi quali libri, videogiochi e alloggio in hotel. La scelta di questi tipi di prodotti/servizi è stata fatta in seguito ad un *focus group* che ho svolto precedentemente per considerare su campioni molto piccoli delle variabili qualitative come i prodotti maggiormente cercati o acquistati online e sono emersi principalmente le tre categorie precedentemente elencate. Lo scopo della ricerca è molto esplorativo per cui non è stato effettuato un focus su una categoria merceologica anche se i risultati di questo studio possono offrire spunti per ricerche future maggiormente incentrati su singole categorie poiché potrebbero essere rilevate differenze rilevanti. Grazie a questa fase preliminare è stato possibile limitare in maniera importante il numero di rispondenti che non ha acquistato o cercato quantomeno uno di questi tre prodotti. Dalle analisi è emerso, infatti, che solo 5 (in termini percentuali il 4%) rispondenti su 142 (11 non hanno completato questa domanda) hanno risposto di non cercare nulla su Internet in generale o tra i prodotti selezionati. Successivamente in base al selezionamento delle tre categorie considerate, venivano visualizzate le domande successive: per coloro che hanno selezionato "Libri", sono apparse delle domande in cui veniva chiesto il genere di libri precedentemente più spesso cercati o acquistati online ed in base al genere selezionato, venivano proposti alcuni autori (anche questi determinati con *focus group*); a chi ha selezionato "Videogiochi", è stato chiesto il genere precedentemente più spesso cercati o acquistati ed in base a questo sono stati proposti alcuni videogiochi (sempre determinati su *focus group*); a chi ha selezionato "Alloggi per hotel per vacanze", invece, è stato chiesto dove ha cercato alloggi tra 5 possibilità di risposta (anch'esse determinate con *focus group*) quali: New York, Maldive, Parigi, Tokyo e nessuna tra queste. Inoltre a questi ultimi è stato chiesto se nelle ricerche/acquisti online recenti frequenti, hanno espresso preferenza per vicinanza al centro città o al mare (nel caso delle Maldive) oppure no. Successivamente ad ognuno di questi tre gruppi è stato chiesto su una scala *Likert* 1-7 di valutare l'intention to buy di un prodotto/servizio che rispettasse tutte le risposte selezionate precedentemente. La scala intention to buy è stata prevalidata da Mano, H., & Oliver, R. L. (1993) ed adattata in tutti i tre casi (libri, videogiochi ed hotel) a partire al lavoro di questi ultimi. Per rispondere alle domande relative a questa scala, è stato chiesto, quindi, di considerare un'inserzione pubblicitaria basata sulle risposte precedenti (ad esempio chi ha selezionato come categoria merceologica "Libri", come genere "Giallo" e come autore "Agatha Christie", è stata proposta l'inserzione pubblicitaria di un libro giallo di quest'ultima) Inoltre considerando l'importanza della leva prezzo (trova paper che dice che è molto importante) poiché non

rilevante ai fini di questo studio, nel caso del gruppo di rispondenti "Libri" è stato chiesto di considerare il prezzo dell'ultimo libro di quel genere acquistato, nel caso di "Videogiochi" è stata fatta una differenziazione a chi acquista abitualmente il gioco a prezzo pieno o chi preferisce aspettare tempo dalla sua uscita per trovare offerte mentre a chi ha selezionato "Alloggi ad hotel per vacanze" è stato precedentemente chiesto il prezzo (tra 4 fasce di prezzo) per cui l'inserzione in questo caso proponeva la fascia precedentemente selezionata. La procedura utilizzata per gli hotel sul prezzo non è stata utilizzata anche nel genere libri poiché in quel caso la variabilità dei prezzi è bassa (R. Creatini, 2001) come anche nel mercato dei videogiochi (T.A.Lorenzini, 2015). Successivamente indistintamente a tutti i rispondenti è stata sottoposta l'inserzione pubblicitaria di una FIAT Panda (pubblicità generica apparsa in televisione). Proprio per il fatto che tutti i rispondenti hanno visto questo prodotto, l'inserzione pubblicitaria non era basata sui dati dei consumatori e quindi, le domande mirate a comprendere l'intention to buy in scala likert 1-7 della FIAT Panda, hanno rappresentato un termine di confronto di un prodotto generico non basato sui dati dei consumatori rispetto ad uno invece basato su questi. Nella scelta di questo prodotto sono stati considerati solo due parametri per la sua natura generica ovvero:

A) che fosse anch'esso *hedonic* come i tre prodotti con viene confrontato per eliminare *bias* tra i due generi come suggerisce la letteratura (Terry L. Childers et al., 2001)
B) che fosse un'inserzione apparsa in televisione poiché come verrà approfondito nel capitolo sulle implicazioni manageriali (3.6), per estensione lo studio potrebbe suggerire una differenziazione tra inserzioni pubblicitarie nell'ambito del social (T.P. Liang et al., 2011) o *e-commerce* (Introduction to e-commerce, Jeffrey F. Rayport and Bernard J. Jaworski, 2002) e quelle trasmesse per canali tradizionali come ad esempio la TV che per loro natura sono meno targettizzate.

Nel questionario, inoltre, erano presenti domande sul rischio percepito online e domande sulla percezione di insicurezza online come tratto personale. Entrambe le scale erano multi-items e sono state prevalidate ed adattate a partire dagli studi di Mani, Z. & Chouk, I. (2018) nel primo caso e di Rojas-Méndez et al. (2017) nel secondo. Nella parte finale, invece, erano presenti domande demografiche su sesso, professione, età e titolo di studio del rispondente. In particolare i dati sull'età sono stati raccolti dando la possibilità ai rispondenti inserendosi in 8 differenti classi di età che poi in sede di analisi (indipendent samples t-test in questo caso) sono state divise in due parti tra under 30 ed over 30 per quanto spiegato precedentemente (3.1). Inoltre sono state effettuate alcune domande utili a testare con quanta frequenza gli utenti si sono trovati in passato ad inserzioni pubblicitarie che sfruttassero algoritmi di ML con particolare riferimento al *Recommendation System* e soprattutto una domanda per testare il *Manipulation Check* volta a verificare se i rispondenti hanno percepito o meno la violazione della *Privacy*. Data la natura dello stimolo, inoltre, non è stato ritenuto necessario effettuare alcun *pre-test* poiché i dati costituiscono esclusivamente uno strumento per generare un certo tipo di strategia comunicativa e non una certa caratteristica di questa. Pertanto si è voluto verificare come algoritmi di ML che sfruttano i dati possono essere utili in termini di *intention to buy* e, quindi, non è stato considerato come questa varia variabile in base a varianti di una strategia comunicativa per cui, invece, sarebbe risultato più opportuno effettuare un *pre-test* al fine di verificare se quella determinata variante di quella

strategia comunicativa risultava "visibile" all'occhio dei rispondenti. Discorso analogo è stato fatto per l'*indipendent samples t-test*.

3.1.2 Data cleaning Dopo aver collezionato i dati attraverso il questionario su *Qualtrics*, c'è stata la fase di *data cleaning* utile a "pulire" il *dataset* in modo da eliminare tutti quei dati con scarso o assente carattere informativo per non incorrere in un *overfitting* di informazioni. In questa fase, quindi, dai 153 rispondenti iniziali, sono stati eliminati anzitutto tutti coloro i quali hanno completato meno del 79% del questionario. Sono stati considerati anche coloro che non l'hanno terminato poiché chi è giunto quantomeno a questa percentuale di completamento, ha risposto alle domande sulle due scale dell'*intention to buy* permettendomi quantomeno di effettuare il *paired samples t-test* su questa variabili. Inoltre sono stati eliminati anche coloro che hanno impiegato meno di 3 minuti a completare il questionario (che è il tempo di completamento stimato da *Qualtrics*) poiché potrebbero aver impiegato scarsa concentrazione nel *test* rischiando di creare *bias* nel campione. Dopo queste due operazioni, il campione analizzato è stato di 123 rispondenti. In sede di analisi (questa parte è stata effettuata tutta sul programma, prima di procedere al *test* delle ipotesi, è stata effettuata un'analisi per testare l'affidabilità di tutte le scale presenti nello studio quindi: *intention to buy* (per libri, automobili, hotel ed automobile), rischio sicurezza online percepito, tratto individuale insicurezza ed i risultati sono stati tutti molto positivi poiché il *Cronbach alpha* è stato nello specifico:

- Intention to buy libri: 0,94
- Intention to buy videogiochi: 0,967
- Intention to buy hotel: 0,917
- Intention to buy automobile: 0,911
- Intention to buy rischio percepito: 0,836
- Intention to buy insicurezza: 0,793

Non è stato necessario eliminare alcun *item* poiché valore più basso del *Cronbach alpha*[24] è stato pari a 0,793, nonostante dalla tabella "Statistiche elemento-totale" (Appendice, sez. C), in alcuni casi l'indice subisse leggeri miglioramenti. A confermare la bontà della scala, in Appendice è possibile trovare anche la matrice di correlazione di elementi che ha rilevato in tutti i casi correlazioni molto alte tra tutti gli item all'interno della stessa scala. A questo punto, quindi, è stato possibile effettuare una media di tutti gli items che si riferivano ad una determinata scala, creando così una nuova variabile.

3.1.3 Formulazione delle ipotesi Rispetto a quanto scritto finora è stata, quindi, effettuata la formulazione delle ipotesi, poi testate con il *t-test*. La prima ipotesi formulata è stata:

H1: La media dell'intention to buy del prodotto "*no Privacy*" è maggiore a quella del prodotto "generico" non basato sui dati.

In questo caso la supposizione che l'inserzione pubblicitaria del libro generi un'*intention to buy* più elevata rispetto a quella dell'automobile, è spinta dal fatto che nel primo caso l'inserzione ha sfruttato i dati precedentemente forniti dal consumatore e quindi era diversa in base a questi mentre nel secondo caso la

[24] è un indice statistico utile a misurare l'attendibilità, ovvero per verificare la riproducibilità nel tempo a parità di condizioni, dei risultati da essi forniti (L.J. Cronbach, 1951)

pubblicità non si è basata su alcun dato ed infatti è stata generica. Infatti quest'ultima è stata indistintamente visualizzata da tutti i 123 rispondenti. Il supporto di questa ipotesi nasce da quanto suggerito dalla letteratura come già accennato in Introduzione (M.Ronchi & M.Ciancia, 2019; G. Pitruzzella, 2016) i *Big Data* forniscono un importante strumento per le aziende. In questo studio infatti si cerca di chiarire come questo asset possa trovare una sua più effettiva implementazione ed attraverso quali strumenti come ad esempio gli algoritmi di ML e più nello specifico attraverso il *Recommendation System* cui si cerca di fare una simulazione pur considerando tutte le limitazioni che vengono chiarite meglio nel paragfrato (2.7).

La seconda ipotesi ha il fine di verificare se ci fosse una differenza sul tratto individuale di insicurezza, intesa come caratteristica individuale nell'ambiente online tra due gruppi indipendenti quali gli under 30 (millenials) e gli over 30 (generazioni precedenti). Pertanto:

H2: La media dell'insicurezza intesa come tratto individuale dei *Millenials* è minore di quella degli over 30

Il supporto della letteratura a riguardo è stata già accennata nel capitolo 3.1 per cui viene suggerito che gli under 30 possono essere categorizzati in maniera differente dalle generazioni precedenti in virtù di alcune differenza di atteggiamento riscontrabili anche e soprattutto nell'ambito digitale. Per questa ragione nel presente studio, è stato supposto che l'insicurezza nell'ambiente online intesa come caratteristica individuale e culturale fosse differente rispetto alle altre generazioni. In particolare, considerando i numeri forniti nell'Introduzione circa la diffusione al giorno d'oggi degli strumenti digitali, si suppone che questa sia minore negli under 30.

Discorso molto simile è stato fatto per testare un'altra variabile quale il rischio percepito online per cui l'ipotesi è:

H3: La media del rischio percepito online in termini di sicurezza dei millenials è minore di quella degli over 30

Come già spiegato precedentemente (3.1), ho provato a riunire queste due variabili nel fattore "atteggiamento online" ma non è stato supportato dal risultato della *Factor Analysis* (Appendice cat. B) poiché solo un *item* ha presentato un *eigenvalue* maggiore di uno ma era in grado di spiegare solo il 55% della varianza. I modelli, quindi, delle singoli analisi erano intesi a verificare in che modo la variabile x (le inserzioni pubbicitarie che sfruttassero i dati personali, quindi per estensione stimoli derivanti da algoritmi di ML, nei primi tre casi, l'età negli altri due). In entrambi i casi la variabile si è presentata come categorica poiché nelle prime tre formulazioni delle ipotesi, si suddivideva in inserzioni pubblicitarie si *Privacy* (quindi stimolo derivante da algoritmi di ML assente) o no *Privacy* (stimolo presente) mentre anche nel caso dell'età, la variabile è stata configurata come categorica poiché, per quanto di per sé sia una variabile numerica, nel questionario è stata la possibilità ai rispondenti di rispondere a questa domanda identificandosi in una delle 8 classi presentate. Il programma statistico utilizzato (*SPSS*), infatti, ha restituito circa questo dato come ouput i numeri "1,2,3,4,5,6,7,8" in base a quale risposta è stata selezionata nell'ordine in cui è stata proposta nel questionario. *SPSS* ha infatti identificato la variabile come una categorica qualunque per cui ogni categoria viene assimilata ad un numero al fine di effettuare l'analisi. Le variabili dipendenti invece sono state tre: l'*intention to buy* del

prodotto specifico nelle prime tre formulazioni di ipotesi, l'insicurezza come tratto individuale nel quarto e nell'ultimo il rischio percepito online in termini di sicurezza. In questi casi *SPSS* ha ricondotto queste variabili qualitative a numeri poiché gli items relativi a queste scale sono stati presentati nel questionario come domande in scala *Likert* 1-7 in cui veniva chiesto al rispondente di esprimere quanto fosse d'accordo con le affermazioni espresse nell'item. Le modalità di risposta sono state, quindi, espresse in ordine da "per niente d'accordo" a "completamente d'accordo" ed è stato quindi possibile per *SPSS* associare a queste variabili quantitative un numero ed effettuare tutte le statistiche descrittive e non presenti in Appendice (cat. B). I risultati sono stati discussi nella sezione successiva (3.2). All'interno dell'intero studio è presente un maggiore approfondimento circa le logiche che segue il *test* dell'ipotesi nell'ambito della statistica inferenziali. All'interno di questo riassunto prima di procedere alla spiegazione dei risultati occorre chiarire le varie ipotesi nulle ed alternative in corrispondenza delle varie ipotesi di ricerca (Introduzione) dei vari *paired samples* ed *indipentent samples t test* svolte in questo studio:

6. Per quanto riguarda l'intention to buy del prodotto "*no Privacy*" dei libri:

H0:La media dell'intention to buy del prodotto "no *Privacy*" libri è uguale a quella del prodotto "generico" automobile
H1:La media dell'intention to buy del prodotto "no *Privacy*" libri è diversa da quella del prodotto "generico" automobile.

7. La seconda coppia di ipotesi invece è stata volta a verificare se lo stesso discorso è fattibile per un prodotto appartenente ad un'altra categoria merceologica come i videogiochi per cui:

H0: La media dell'intention to buy del prodotto "no *Privacy*" videogiochi è uguale a quella del prodotto "generico" automobile
H1: La media dell'intention to buy del prodotto "no *Privacy*" videogiochi è diversa da quella del prodotto "generico" automobile.

8. Stesso discorso è stato fatto con la terza coppia di ipotesi ma questa volta considerando un servizio quale l'alloggio in hotel per cui:

H0: La media dell'intention to buy del prodotto "no *Privacy*" alloggio in hotel per vacanze è uguale a quella del prodotto "generico" automobile
H1: La media dell'intention to buy del prodotto "no *Privacy*" alloggio in hotel per vacanze è diversa da quella del prodotto "generico" automobile.

9. Per quanto riguarda gli *indipendent samples t-test* nella prima coppia di ipotesi:

H0:La media degli under 30 dell'insicurezza individuale intesa come tratto individuale è uguale a quella degli over 30
H1:La media degli under 30 dell'insicurezza individuale intesa come tratto individuale è diversa da quella degli over 30

10. Il discorso in questo caso è simile a quello del punto 4 ma considerando un'altra variabile ovvero il rischio percepito online in termini di sicurezza. Perciò:

H0:La media degli under 30 del rischio percepito online in termini di sicurezza è uguale a quella degli over 30

H1:La media degli under 30 del rischio percepito online in termini di sicurezza è diversa da quella degli over 30

3.2 Risultati Terminata tutta la fase preliminare (3.1.1 e 3.1.2), in questo paragrafo vengono resi noti i risultati del confronto medie con le due tipologie di t-test (indipendent e paired samples) considerate le ipotesi scritte nel sottoparagrafo precedente (3.1.3). La media dell'*intention to buy* dei libri (Minttbuy_libr= 5,4855, Appendice cat C.) è stata considerata statisticamente significativa poiché il p-value[25] calcolato dal campione (indicato come Sig. (2-code) nell'output) è stato indicato pari a 0.0000, per un valore t pari a 10,457 (che corrisponde al valore t nell'output in Appendice, cat. C) con GdL pari a 45 in una distribuzione t. Pertanto, poiché il *p value* (=0.0000)<α(=0,025 poiché il *test* è a due code ed il livello di significatività è del 5%), rigettiamo l'ipotesi nulla secondo cui le due medie sono considerate statisticamente uguali. Inoltre la differenza tra le due medie è di circa 2,62 (Mint_tobuy=5,4855; Minttbuy_automobile=2,8696) per cui, considerando che le domande erano in scala-*Likert* 1-7, l'*intention to buy* del libro è risultata il 37,37% maggiore di quella dell'automobile. Per quanto riguarda l'*intention to buy* dei videogiochi: t(16)=0,2958; p value=0.009 <α(=0,025) per cui anche in questo caso rigettiamo H0 e consideriamo le differenze delle due medie come statisticamente significative. In questo caso l'*intention to buy* dei videogiochi (Minttbuy_videogiochi=4,7647; Minttbuy_automobile=2,9216) è risultata il 26,33% superiore. Nel caso, invece, dell'intention to buy degli alloggi in hotel per vacanze: t(55)=12,735; p value=0,000<α(0,025) per cui ancora una volta rigettiamo H0, con un *intention to buy* (Minttbuy_hotel=5,6369; Minttbuyautomobile=2,8869) del 40,14% più alta del prodotto "generico" dell'automobile. Per quanto riguarda i due *indipendent samples t-test*, partendo dallo studio circa l'insicurezza come tratto individuale tra gli under30 e gli over30: prima di condurre l'analisi di confronto medie, bisogna controllare le statistiche de *Levene* che assumono che le varianze delle due medie sono uguali (H0= la varianza della media degli under 30 è uguale a quella degli over 30; h1=la varianza della media degli under 30 è diversa da quella degli over 30). Poiché in questo caso il *p value*>α con F (116)=0,289(Appendice, cat.B), non rigettiamo l'ipotesi nulla ed assumiamo che le varianze sono uguali. Verificando quindi i risultati dell'output di *SPSS* che corrispondono con l'assunzione delle varianze uguali emerge che: t(116)=-4,888; p value=0.0000<α. Perciò rigettiamo l'ipotesi nulla ed assumiamo che la differenza tra la percezione di insicurezza come tratto personale è statisticamente significativa tra i gruppi under 30 ed over 30. In particolare gli under 30(Mmea_insicurezza_mill=4,2979;Mmea_insicurezza_nomill=5,5545) hanno mostrato insicurezza come tratto individuale il 17,95% in meno rispetto agli over 30.

[25] questo valore indica la probabilità di ottenere risultati uguali o meno probabili di quello osservato durante il test assumendo per vera l'ipotesi nulla (Wasserstein, R.L. e Lazar, N.A., 2016)

Nel caso del rischio percepito online, il test di *Levene* con F(116)=0,048 ha rilevato che le varianze sono uguali poiché *p value*>α. Osservando quindi le statistiche di questa assunzione: t116)=-4,967, p value(=0,000)<α(=0,025). Le due medie (Mmea_riskmill=4,3819;Mmea_risknomill=5,6667) hanno mostrato che gli under30 hanno una percezione del rischio, in termini di sicurezza, online inferiore del 18,37%.

Inoltre sono state effettuate alcune statistiche descrittive sul campione:

- è stata posta una domanda per testare il manipulation check quindi se i rispondenti hanno percepito una violazione della *Privacy* o meno nelle domande preliminari all'inserzione pubblicitaria del prodotto "no *Privacy*". I risultati hanno mostrato che: il 17% delle persone hanno sentito violata la propria privacy, mentre l'83% no. (fig. 5). I numeri nell'ambito dei videogiochi e degli hotel a riguardo si sono mostrati simili (fig. 4.6 4 fig. 4.7 dell'intero studio)
- per quanto riguarda il genere: il 50% dei rispondenti è di sesso femminile, 49% maschile ed 1% ha preferito non rispondere
- sull'età invece i dati rivelano una distorsione del campione poiché l'81% del campione appartaneva ad una categoria (under 30) e solo il resto agli over 30.

3.3 Limitazioni della ricerca e spunti per ricerche future Questo studio presenta diversi limiti che possono fornire spunto per approfondimenti in ricerche future. Anzitutto il campione di 123 persone è stato ulteriormente frammentato in sede di analisi *t-test* poiché sono state confrontate le medie dell'*intention to buy* dei tre prodotti "no *Privacy*" separatamente rispetto a quello generico, generando addirittura nel caso dei videogiochi un campione di sole 17 persone. In statistica è sempre consigliato di utilizzare il campione più grande possibile poiché con maggiore precisione tende a rispecchiare il dato reale della popolazione. Per quanto riguarda, invece, i due *indipendent samples t-test*, il campione è apparso sicuramente maggiore poiché aggregato ma in questo caso c'è stato un altro bias poiché gran parte del campione (81% come si apprende nel paragrafo 2.5) apparteneva alla classe di età under 30. Potrebbe principalmente risultare utile, quindi, ripetere le analisi con campioni più grandi ottenendo quindi risultati più attendibili. Inoltre occorre anche specificare la logica seguita nell'ambito del *Manipulation Check* che in questo caso è stato effettuato principalmente per comprendere ulteriormente come i consumatori si approcciano circa la cessione dei propri dati personali ad aziende ma non è stato fatto nessun data cleaning in questo senso poiché, considerando la natura dello stimolo che non costituisce un elemento caratterizzante del prodotto o del suo aspettivo comunicativo quanto, invece, alla verifica di constatare l'efficacia di un input per rendere fruibile uno strumento quale il *Recommendation System*. Pertanto, proprio perché lo studio è stato effettuato considerando secondo la logica del *"what if"* per cui si è voluto verificare come varierebbe il comportamento d'acquisto dando come assunta la cessione dei dati personali, non sono state eliminate le risposte di coloro che hanno percepito violazione della *Privacy* nelle domande precedenti all'inserzione pubblicitaria della condizione *"no Privacy"*. Infine un'altra importante limitazione è dovuta dal fatto che lo studio ha esaminato principalmente quanto e se l'*intention to buy* cresce in presenza di maggiore concessione dei dati personali dei consumatori però, nell'elaborazione del questionario l'obiettivo è stato quello di avvicinarsi il più possibile alla realtà che troviamo con il

Recommendation System cercando di seguire le stesse logiche che seguono gli algoritmi di ML in merito. Inevitabilmente, però, il fatto che le inserzioni pubblicitarie proposte nella *survey* sono state frutto dell'*Intelligenza Artificiale* e non di quella informatica come avviene nel caso del *Recommendation System*, per verificare in maniera più attendibile quanto questi risultano davvero efficaci alle imprese in ambito *Marketing*, potrebbe risultare utile svolgere ulteriori studi che simulino o tentino di riprodurre la realtà in maniera più precisa. Questo studio, infatti, si limita solo ad offrire un contributo in questo senso.

CONCLUSIONI Nel suo complesso la ricerca si è basata su due studi esplicati nei capitoli 2 e 3 in cui il primo (la *Sentiment Analysis*) ha evidenziato come ci sia un dibattito acceso sulla *Privacy* e come le emozioni negative principalmente evidenziate relativamente ad essa sono l'ansia e la paura. Nel secondo studio, invece, è stato evidenziato come un'inserzione pubblicitaria che sfrutta i dati dei consumatori costituisca per questa una intention to buy più alta rispetto ad un prodotto non basato sui dati. I risultati del secondo studio, con le dovute cautele, per estensione potrebbero anche esprimere il paragone dell'efficacia che ha una pubblicità nell'ambito del *Social* e dell'*e-commerce* rispetto a quella più tradizionale visibile ad esempio in TV dove le aziende riescono a fare una targettizzazione dei clienti meno specifica basata esclusivamente sui programmi televisivi e sugli orari senza, quindi, raggiungere il grado di specificità delle inserzioni online. Occorre, inoltre, specificare che in questo studio è stata implementato esclusivamente un "modello base" di quanto riescono invece ad effettuare gli algoritmi di ML nell'ambito del *Recommendation System*. Basti considerare, infatti, che partendo dagli acquisti storici dei consumatori (come fatto nel caso specifico di questo studio) si possono delineare maggiormente i loro gusti, le loro tendenze e quindi, assumendo che in futuro le aziende riescano a possedere una grande mole di dati dei consumatori, potrebbe risultare anche più semplice predire prodotti che possano risultare loro utili anche nell'ambito di categorie merceologiche differenti. Osservando, cioè, in maniera estrememente precisa il comportamento dei consumatore e conoscendoli meglio, potrebbe aprirsi un importante solco per la ricerca per cui potrebbe risultare estremamente più facile targettizzare i consumatori in *cluster* maggiormente definiti che esprimono in maniera ulteriormente più precisa il comportamento d'acquisto di molti di loro arrivando quindi a proporgli prodotti in ogni categoria merceologica in base al cluster in cui vengono classificati. Per chiarire maggiormente questo aspetto potrebbe risultare utile un esempio: supponendo che un'azienda nel settore dell'*automotive* venga a conoscenza del fatto che un determinato consumatore tra i suoi acquisti recenti ha mostrato particolare tendenza ai prodotti "Bio" che privilegiano il rispetto per l'ambiente in diverse categorie merceologiche (come ad esempio quello della cura personale, alimentare, di abbigliamento ecc.), potrebbe risultare maggiormente interessante per questo un'inserzione pubblicitaria di un'automobile elettrica. Considerando, inoltre, quanto emerso anche dai risultati degli *indipendent samples t-test* (paragrafo 3.2), gli under 30 mostrano anche un atteggiamento diverso rispetto alle precedenti generazioni in cui la preoccupazione online in termini di sicurezza percepita, ed anche l'insicurezza in ambito online come tratto personale, sono minori per cui potrebbe risultare meno difficoltoso comunicare, contrariamente a quanto si pensi, che offrire i dati alle aziende, a patto che queste le utilizzino in maniera responsabile, non fa altro che dare a queste uno strumento che risulta utile a loro ma anche ai consumatori stessi come dimostrano i findings del secondo studio di questa ricerca che hanno mostrato

un'*intention to buy* più elevata per prodotti che meglio rispecchiano i loro gusti espressi attraverso gli acquisti storici. Infine occorre anche contestualizzare tipi di strategie comunicative che sfruttano algoritmi di ML secondo logiche di *Recommendation System*. Oggigiorno ci troviamo, infatti, in un clima dove il mercato risulta in quasi la totalità delle categorie merceologiche iper competitivo, i consumatori si ritrovano continuamente esposti ad una grandissima mole di informazioni per cui esprimere un messaggio rilevante per loro risulta particolarmente difficoltoso (Valdani E., 1997) e da questo deriva, poi, anche quanto già espresso nel paragrafo dedicato ai "Micro-momenti" (paragrafo 4.5 dell'intero studio) dove appunto viene evidenziato come il *time-pressure* per il consumatore lo porta sempre più spesso a fare scelte nell'ambito del decision-making che richiedano sempre meno tempo e sforzo. Un ulteriore spunto in questo senso potrebbe essere offerto proprio da quest'ultimo punto per cui l'implementazione di strategie comunicative che meglio riescono ad esaltare le euristiche ed il sistema 1 di Kanheman e Tversky (come già spiegato nell'Introduzione), possono risultare nell'attuale ambiente particolarmente efficaci. Infine l'ultima pandemia da Covid-19 risulta ancora troppo recente per essere già oggetto di ricerca ed analizzata dal punto di vista degli effetti che ha generato anche in ambito manageriale, ma considerando la portata dell'evento non è possibile escludere del tutto che ci sia un cambiamento del comportamento da parte dei consumatori che li ha portati a sperimentare maggiormente tutto ciò che riguarda l'*e-commerce*, il *social-commerce* ed in generale il mondo online per cui quanto già detto potrebbe risultare ulteriormente amplificato nei prossimi mesi. Sostanzialmente, partendo dall'assunto che le aziende riescano in maniera responsabile ad occuparsi di come soddisfare i *needs* dei consumatori, queste potrebbero estendere un ulteriore servizio all'interno della loro offerta, una fase preliminare a quella d'acquisto (e non solo quindi di assistenza post-acquisto) in cui, conoscendo ed avendo mappato i consumatori, riescono a fornire a questi inserzioni pubblicitarie di prodotti maggiormente affini a loro e l'ambiente circostante potrebbe portare le aziende a fare di questo aspetto una competenza estremamente rilevante ed efficace se non necessaria. Conferme su quanto le aziende si stiano concentrando ad aumentare la propria presenza online e l'efficacia delle proprie strategie in questo ambito arrivano anche da quanto stiamo vedendo nell'ambito dell'Omnicanalità[26] dove le aziende stanno cercando di stimolare sempre maggiormente i consumatori a vivere esperienze d'acquisto online più appaganti, mettendo in pratica quanto suggerisce già parte della letteratura circa l'efficacia di questa strategia (Payne, E. M. et al., 2017), affinché diventino più frequenti per adeguarsi all'ambiente circostante e sconfiggere le barriere geografiche che appaiono sempre più anacronostiche come ci suggerisce la letteratura (Levitt, T., 1993; Mittelman, J.H., 1993; Bandura, A., 2001) che da diversi anni ha definito come consolidato il processo di globalizzazione. In definitiva considerando gli input degli studi scritti ad inizio di questo capitolo, possiamo ritenere che c'è mancanza di comunicazione per i consumatori di come lo sfruttamento dei dati personali restituisca a questi stessi un servizio migliore in termini di inserzioni pubblicitarie maggiormente conformi alle loro intenzioni d'acquisto?

[26] Oggi la capacità dell'uso simultaneo di diversi canali di interazione tra consumatoe e negozio (ad esempiol'uso dell'acceso a Internet mobile all'interno del negozio fisico al dettaglio per cercare informazioni sui prodotti e/o confrontare i prezzi dei prodotti) costituisce la caratteristica dominante del fenomeno del commercio al dettaglio "omnichannel" che si basa sulla ben consolidata infrastruttura al dettaglio multicanale sviluppata principalmente dallo sfruttamento commerciale del World Wide Web (Lazaris, C.,Vrechopoulos, C., 2014).

RIASSUNTO