

UNIVERSITÀ DELLA CALABRIA



UNIVERSITA' DELLA CALABRIA

Dipartimento di FISICA

Dottorato di Ricerca in

Scienze ed Ingegneria dell'Ambiente, delle Costruzioni e dell'Energia

CICLO

XXXII

**APPLICAZIONE DEI BIG DATA NEL TURISMO, MARKETING
ED EDUCATION**

Settore Scientifico Disciplinare MAT/O7

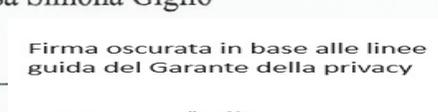
Coordinatore: Ch.mo Prof. Salvatore Critelli

Firma  Firma oscurata in base alle linee guida del Garante della privacy

Supervisore/Tutor: Ch.mo Prof. Pietro Pantano

Firma  Firma oscurata in base alle linee guida del Garante della privacy

Dottoranda: Dott.ssa Simona Giglio

Firma  Firma oscurata in base alle linee guida del Garante della privacy

INDICE

INTRODUZIONE	1
CAPITOLO 1	5
BIG DATA SCIENCE E MACHINE LEARNING	5
1.1	La nuova era dei dati 5
1.1.1	Le 5 V dei Big Data 7
1.1.2	Dal volume al valore 8
1.2	Big data e settori applicativi 11
1.3	Social media: risorsa illimitata di Big data 12
1.4	Cognitive Computing e Machine Learning 14
1.4.1	Apprendimento supervisionato e non supervisionato 16
1.4.2	Deep Learning e Reti neurali 18
1.5	Big data Analytics 19
1.5.1	Analisi dei Big data con Wolfram Mathematica 21
1.5.2	Image Analysis 22
1.5.3	Text e Content analysis 25
1.5.4	Sentiment analysis 26
1.5.5	Clustering analysis 28
CAPITOLO 2	31
BIG DATA ED INDUSTRIA DEL TURISMO	31
2.1	Il turismo ai tempi dei Big Data 31
2.2	Mobilità turistica e comportamento degli utenti 32
2.3	Flickr e Big data: uso dei social media per identificare l'attrattività turistica 35
2.3.1	Obiettivi della ricerca 35
2.3.2	Procedura e metodologia di analisi 36
2.3.3	Risultati 43
2.3.4	Discussioni 51
CAPITOLO 3	55
IMPATTO DEI BIG DATA SUL MARKETING E SULLA CUSTOMER EXPERIENCE	55
3.1	Digital marketing e Retail 55
3.2	Big data e Marketing Intelligence 58
3.3	Processi cognitivi e comportamento del consumatore 60
3.4	Twitter e Sentiment Analysis 62
3.4.1	Obiettivo della ricerca 62
3.4.2	Procedura e metodologia di analisi 62
3.4.3	Risultati 66
3.4.4	Discussioni 67

3.5	TripAdvisor e analisi delle immagini	69
3.5.1	Obiettivo della ricerca	70
3.5.2	Procedura e metodologia di analisi	70
3.5.3	Risultati	74
3.5.4	Discussioni	75
3.6	Cluster analysis e profilazione utenti	76
3.6.1	Obiettivo della ricerca	78
3.6.2	Procedura e metodologia di analisi	78
3.6.3	Risultati	80
3.6.4	Discussioni	84
CAPITOLO 4		86
BIG DATA ANALYTICS: NUOVE SKILLS NELLA FORMAZIONE		86
4.1	Big Data, IoT ed Industria 4.0	86
4.2	Valorizzazione della ricerca	90
4.3	Un esperimento didattico	91
4.3.1	Obiettivo	92
4.3.2	Procedura	93
4.3.3	Risultati	95
4.3.4	Discussioni	98
CAPITOLO 5		101
WEB SEMANTICO: PROGETTAZIONE DI UN SISTEMA INFORMATIVO		101
5.1	Web Semantico	101
5.2	Progettazione del Sistema informativo	103
CONCLUSIONI		113
INDICE DELLE FIGURE		132
INDICE DELLE TABELLE		135
APPENDICE A		136

INTRODUZIONE

Il mondo è attualmente inondato da dati e l'avanzare delle tecnologie digitali amplifica questo fenomeno in modo esponenziale, tale che l'uso massivo dei sistemi ICT (Information Communication Technologies) crea tracce digitali che riflettono le attività quotidiane di ciascun individuo, descrivendone con estrema precisione il comportamento, le abitudini e le preferenze. Queste tracce rispecchiano dettagliatamente i comportamenti collettivi e individuali delle persone, cosicché le diverse dimensioni della vita sociale in termini di opinioni, stili di vita, movimenti e relazioni, trovano un'immagine riflessa nello specchio digitale. Tale fenomeno è stato etichettato, in letteratura, con il termine Big data (Hwang & Chen, 2017).

Su questa base, è fiorita una vivace linea di ricerca che ha attratto scienziati di diverse discipline, non solo per la grande sfida intellettuale, ma anche per l'importanza in settori applicativi quali la pianificazione urbana, la mobilità sostenibile, l'ingegneria dei trasporti, la salute pubblica e la previsione economica. I Big data sono diventati il nuovo microscopio che rende "misurabile" la società. Il mutamento principale apportato dai Big Data, tuttavia, è già avvenuto, e risiede nel cambio di prospettiva evidenziatosi nella metodologia di analisi. Infatti, le tecniche utilizzate nell'era degli "small data" sono difficilmente applicabili ai nuovi dati a disposizione, a causa della loro estensione in termini di volume, velocità, varietà, ad esempio la regressione lineare non è più la tecnica migliore per decifrare le informazioni nascoste nei dati, in quanto si richiedono Dataset relativamente ridotti, omogenei e organizzati. A tal proposito, vari esperimenti stanno dimostrando come sia possibile rispondere a molte domande complesse sul comportamento umano sulla base dell'analisi dei Big data prodotti on-line (Gensler et al., 2015). Gli utenti pubblicano e condividono diversi tipi di dati sul Web, ad esempio le immagini sulle piattaforme social, la quali rappresentano sistemi socio-culturali ed economici utili alle amministrazioni locali, alle agenzie di viaggio e ad altri stakeholder; le recensioni, che riflettono la qualità e la popolarità di un determinato luogo, prodotto o servizio, e dunque la percezione e l'atteggiamento dell'utente. Le piattaforme di condivisione dei dati consentono dunque la creazione di un ponte tra mondi digitali e

mondi reali offrendo la possibilità di conoscere le preferenze ed il comportamento degli utenti.

Il *leitmotiv* che ha guidato il presente progetto di tesi è stato quello di studiare il comportamento dell'utente con un metodo indiretto, analizzando le tracce digitali che vengono lasciate e registrate dai sistemi tecnologici quotidianamente utilizzati per interagire con il mondo. Pertanto, la ricerca condotta si è focalizzata sull'analisi dei Big data estratti dai social media, da indagini online, da piattaforme di recensioni e da database applicando tecniche e strumenti sviluppati nell'ambito dell'Intelligenza artificiale. Algoritmi di Machine Learning sono stati utilizzati per estrarre, dai Dataset collezionati, informazione sotto forma di “conoscenza” e “valore”, dimostrando come dati di grandi dimensioni possano fungere da ricca fonte di informazione per comprendere il comportamento dell'utente (conoscenza), parte integrante di una società complessa, e sostenere i processi decisionali (valore).

Il *primo capitolo* consta di un excursus generale sui progressi tecnologici e sugli strumenti che hanno contribuito e contribuiscono alla diffusione e produzione dei dati. In particolare l'attenzione è stata posta sui Big data e sulle loro caratteristiche, sul valore che tali dati possono assumere e come questo valore può essere sfruttato per ricavare informazioni necessarie ai fini del processo decisionale. Si approfondisce l'argomento dei Big data e dei settori applicativi dunque dei social network che si qualificano come ricca fonte di Big Data. Inoltre, viene offerta una panoramica su argomenti quali Machine Learning, Deep Learning e Reti neurali. Il capitolo si conclude con la descrizione delle tecniche di Big data Analytics che sono state applicate nella ricerca.

Il *secondo capitolo* affronta le tematiche connesse all'industria del turismo e le innovazioni che i Big data hanno portato in questo settore. Si affronta il tema della mobilità e dello spostamento (fisico e virtuale) facendo riferimento ai dati GPS estratti dai social network. Nel capitolo viene presentata l'elaborazione e l'analisi di specifici dati digitali (immagini) al fine di indagare il rapporto che gli utenti hanno con la destinazione turistica. I dati analizzati sono stati estratti da Flickr, una piattaforma online attraverso la quale gli utenti condividono le loro foto. La raccolta dei dati è avvenuta grazie all'utilizzo di Application Programming Interface (API), un insieme di procedure utili ai programmatori per la realizzazione di applicazioni di vario genere. In base al numero di dati disponibili e immagazzinati, è stata determinata la relazione tra la posizione dell'utente ed il contenuto delle immagini al fine di studiare l'attrattività di varie località,

identificare i luoghi visitati ed i punti di interesse (POIs). Sono stati presentati i risultati che forniscono una metodologia utile ad analizzare e trasformare i Big data in informazioni significative per l'industria turistica. L'analisi ha permesso, inoltre, di elaborare una comprensione generale del movimento, degli interessi e delle esperienze dei turisti e ha confermato l'importanza dei dati ricavati dai social media come un vantaggio nella comprensione del comportamento degli utenti.

Il *terzo capitolo* è incentrato sull'analisi dei Big data nel marketing e nel Retail. In particolare, viene evidenziato come i Big data rappresentino per questi settori un valore aggiunto che contribuisce, non solo alla crescita di un'azienda, bensì al benessere del cliente/consumatore. Nel capitolo vengono riportate tre applicazioni: la prima riguarda tre aziende fast fashion e i dati estratti da Twitter. Si fa riferimento all'analisi eseguita su dati testuali (tweet) ovvero la *sentiment analysis* e se ne riportano i risultati, i quali illustrano le valutazioni dei consumatori sulle strategie di vendita per i tre rivenditori. I risultati hanno sottolineato come l'analisi dei Big Data, in particolare l'analisi del sentiment può essere uno strumento efficace per i marketer. La seconda applicazione riguarda il settore alberghiero, nello specifico i brand di lusso e l'analisi delle immagini condivise dai clienti sulla piattaforma TripAdvisor. La metodologia adottata ha consentito una comprensione più inclusiva della valutazione da parte dei consumatori degli attributi alberghieri più importanti nel settore del lusso, la cui importanza è in grado di influenzare l'immagine del marchio alberghiero. La terza applicazione concerne l'analisi dei cluster applicata per profilare classi di utenti nel settore dell'energia. L'analisi, condotta su dati estratti da un questionario online ideato *ad-hoc* e somministrato ad un campione rappresentativo, ha permesso di identificare i profili degli utenti sulla base del loro comportamento in relazione al consumo di energia. La ricerca ha permesso di raccogliere informazioni sul programma del consumatore, sulle sue preferenze e sulle sue abitudini e ha consentito di definire le classi di utenti al fine di adattare e promuovere specifiche attività di marketing, formulare efficaci strategie aziendali ed identificare punti di forza e di debolezza di un'azienda.

Nel *quarto capitolo* si parla di Industria 4.0, Big data ed Internet of Things e di come alcune tecnologie e strumenti all'avanguardia possano essere efficacemente integrati nei *curricula* per supportare l'attività di formazione attraverso l'applicazione in contesti reali. Il fine ultimo è stato dimostrare come l'acquisizione di competenze tecnologiche d'avanguardia possa contribuire a colmare il GAP tra contesto estremamente teorico, che

caratterizza la formazione accademica, e le esigenze del mercato del lavoro basato essenzialmente sulla pratica. Nel capitolo viene presentato un esperimento empirico che ha avuto come obiettivo quello di fornire agli studenti del corso di Laurea in Scienze Turistiche dell'Università della Calabria, un laboratorio sperimentale incentrato sul *framework* Industry 4.0 e i nove pillars (tra cui i Big Data) sottolineando come le tecnologie e i metodi innovativi possano contribuire allo studio del comportamento dei turisti e come i Big data possano essere sfruttati nella progettazione di tour accattivanti e personalizzati.

Nell'*ultimo capitolo* viene descritta e dettagliata la progettazione di un sistema informativo che, attraverso specifiche *query* su repository, associa variabili di varia natura (geografiche, bibliografiche) mappando e recuperando metadati a partire da una traccia digitale. Il fine ultimo è fornire all'utente che effettua una ricerca l'informazione semantica partendo da un'immagine.

CAPITOLO 1

BIG DATA SCIENCE E MACHINE LEARNING

1.1 La nuova era dei dati

L'introduzione della tecnologia Pentium nei primi anni '90, segna una svolta epocale. Per la prima volta i calcolatori registrano alte performance poiché coniugano in una macchina di modeste dimensioni, alta capacità di archiviazione dei dati, velocità di elaborazione e capacità di gestione di dispositivi software all'avanguardia (Kudyba & Diwan, 2002). Potremmo affermare che quello fu l'inizio della nuova era delle tecnologie dell'informazione. Ben presto, Big data, terabyte di dati, montagne di dati, avrebbero reso le modalità e gli strumenti di raccolta ed archiviazione precedenti semplicemente banali (Kudyba & Kwatinetz, 2014).

Oggi giorno, transazioni bancarie, ricerche web, registri telefonici e social network, rappresentano le principali attività che contribuiscono in modo esponenziale all'aumento della produzione dei dati. Attraverso motori di ricerca e piattaforme come, ad esempio, Google che processa 24 petabyte (10^{15} byte) di dati al giorno, Facebook che ottiene più di 10 milioni di foto per ora, con 2 miliardi di utenti in tutto il mondo (Clement, 2019), vengono prodotti ed offerti dati di milioni di individui con un livello di dettaglio senza precedenti. Si stima, infatti, che i dati aumentino ad un tasso di crescita annuale di quasi il 60% (Kashyap, 2017). L'estrema eterogeneità dei dati raccolti in un database descrive la varietà dei Big data che possono essere di diverso tipo: strutturati, semi-strutturati e non strutturati. I tipi di dati con mancanza di organizzazione strutturale, ovvero immagini, audio, video e testo, sono classificati come dati non strutturati, mentre solo una piccola parte, circa il 10% dei dati prodotti appartiene alla categoria di dati strutturati ed è

rappresentata dai dati presenti nei sistemi di gestione di database relazionali (RDBMS) tra cui fogli elettronici (fogli di calcolo). I dati semi-strutturati, invece, si estendono liberamente tra tipi di dati strutturati e non strutturati e raggruppano dati per i quali, nonostante la presenza di una qualche forma di struttura, essa non è prescrittiva, regolare e completa come richiesto dai tradizionali RDBMS. Un esempio di dato semi-strutturato è rappresentato in un file XML, oppure da dati eterogenei provenienti da sorgenti plurime. Allo stato attuale, la rapida crescita dei Big data è direttamente correlata alle azioni svolte dagli utenti durante le loro attività quotidiane e principalmente legate ad Internet, al Web e ai servizi cloud (Hwang & Chen, 2017).

All'inizio, il termine Big Data, era usato per definire set di dati prodigiosi e sofisticati. La definizione più comune parla di Big data come una grande quantità di dati difficile da gestire ed elaborare a causa della loro complessità e delle loro dimensioni (Hadi et al., 2015). In realtà, il concetto di Big data è rimasto comunque vago e non è mai stato chiaramente definito. Inoltre, a causa di interessi diversi da parte di imprese, scienziati, analisti dei dati e professionisti tecnici sono state elaborate definizioni diverse di Big data (Chen et al., 2014). Le origini del termine risalgono alla metà degli anni '90 nel periodo in cui John Mashey lavorava presso la Silicon Graphics (colosso della computer grafica negli USA degli anni '90) ed era coinvolto nell'elaborazione ed analisi di grandi set di dati (Diebold et al., 2012). Successivamente, nel 2001 Doug Laney, analista dell'attuale Gartner (multinazionale nel campo tecnologico) associò il termine Big data ad modello denominato "3Vs" attraverso cui venivano indicate le tre proprietà caratterizzanti i Big data ovvero, volume (aumento del volume dei dati), velocità (generazione rapida dei dati) e varietà (varie modalità di dati) (Laney, 2001). Gartner, IBM e altre aziende tra cui Microsoft continuarono ad utilizzare il modello "3Vs" per descrivere i Big data per altri dieci anni. La svolta sarebbe avvenuta nel momento in cui, oltre che a sviluppare una definizione corretta di Big Data, la ricerca si sarebbe concentrata su come utilizzare e trasformare un insieme di dati in "Big data" estraendo da essi il valore. A tal proposito, nel 2011, un rapporto della IDC (International Data Corporation - uno dei leader più influenti nei big data) che definì i Big data come *"una nuova generazione di tecnologie e architetture, progettate per estrarre economicamente valore da volumi molto grandi di dati, consentendo l'acquisizione, il rilevamento e/o l'analisi"*, diede le basi per ampliare il modello da "3Vs" a "4Vs" includendo un'altra proprietà, che rappresentava la sfida

maggiormente critica nei Big data, ovvero l'estrazione del "valore". I Big data, infatti, se efficacemente analizzati costituiscono delle informazioni che possono rispecchiare dinamiche di mercato, della ricerca e dello sviluppo; se opportunamente letti, rappresentano un valore per l'economia mondiale, poiché permettono di individuare ed estrarre conoscenze utili ed applicabili in vari campi.

1.1.1 Le 5 V dei Big Data

Come già discusso, la definizione di Big data non è riconducibile solo ed esclusivamente alla dimensione o alla quantità, ma ad altre proprietà che ne rappresentano i caratteri distintivi, ovvero: Volume, Velocità, Varietà, Veridicità, Valore.

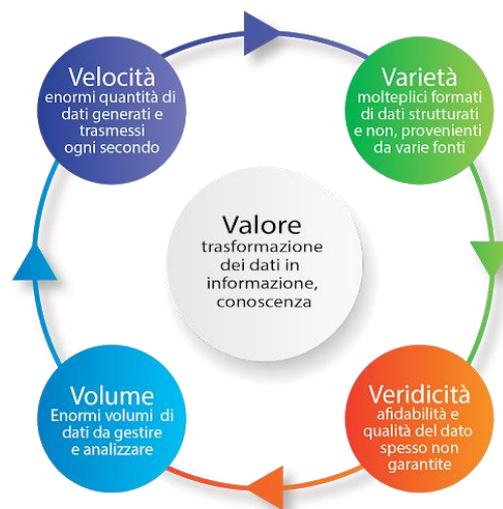


Figura 1.1. Le 5 caratteristiche dei Big Data

La prima caratteristica dei Big data ruota attorno alla grande quantità di dati derivanti dalle tecnologie digitali e dunque al *volume*. Ogni giorno, le aziende devono imparare a gestire il grande volume di dati che ricevono applicando nuovi processi. Il fatto che esista un numero così elevato di dati rende necessario suddividerli in gruppi, in quanto la loro gestione ed interpretazione può influenzarne o deviarne il risultato.

Un'altra caratteristica, riferita alla possibile inconsistenza dei dati analizzati, è la *varietà*. Essa si riferisce alle fonti eterogenee e alla natura dei dati, sia strutturati che non strutturati i quali presentano alcuni problemi per l'archiviazione, il *data mining* e l'analisi. L'uso dei dispositivi elettronici e di internet non solo aumenta il volume e la varietà dei dati ma la *velocità*, ovvero l'alta frequenza con cui i dati vengono prodotti. In relazione a questo parametro i Big data richiedono l'uso di strumenti per garantirne la corretta memorizzazione. La velocità è relativa, non solo ai Big Data, ma anche a tutti i processi, tra cui l'analisi. Infatti, una buona analisi può richiedere ore o addirittura giorni (Hashem et al., 2015). Una delle sfide dell'analisi dei Big data è determinare se, tra tutti i dati elaborati, ve ne siano alcuni falsi. È proprio quel volume di dati che genera dubbi sulla loro *veridicità*, un'altra caratteristica dei Big Data. A causa della grande quantità prodotta, molti dati potrebbero essere errati oppure incompleti. Questo comporta una maggiore attenzione da parte degli analisti che devono appurarne l'affidabilità e la validità (Sagiroglu & Sinanc, 2013). È infatti l'affidabilità dei dati che permette di raggiungere determinati obiettivi e di estrarre *valore* da essi (Khan et al., 2018).

1.1.2 Dal volume al valore

Le innovazioni tecnologiche hanno spostato l'attenzione dalla quantità alla qualità dei dati evidenziando l'importanza degli stessi (Bradlow et al., 2017). Per qualità dei dati si intende il valore assegnato a questi ultimi e, quando si considera il valore, probabilmente il più significativo valore aggiunto apportato dai Big data è il miglioramento nel processo decisionale. Man mano che la quantità di dati continua a crescere, la necessità di sfruttarla diventa più importante. A tal fine, Verhoef e colleghi (2016) hanno elaborato un modello per spiegare come avviene il processo di creazione di valore dai Big Data, evidenziando le strategie e le soluzioni per generare un valore bidirezionale, sia per i consumatori (value to consumer -V2C) che per l'azienda (value to firm-V2F).

Il modello si basa su quattro componenti principali:

- i. *risorse*
- ii. *capacità*
- iii. *analisi*

iv. *valore*

Le *risorse*, generalmente archiviate dalle società, possono essere suddivise in due tipologie: tangibili (edifici) e immateriali (clienti, prodotti). Ad oggi, la produzione di dati ha portato dei cambiamenti nella gestione aziendale, inducendo i manager a concepire anche i Big data come risorse preziose da utilizzare a supporto della gestione commerciale. Tuttavia, per ben sfruttare i dati è necessario evidenziare gli elementi rilevanti, rispetto al tipo di attività svolta, attraverso i quali le imprese saranno in grado di coordinare le proprie attività e le proprie risorse. Investire in dati appropriati contribuirebbe, secondo Gandomi e Haider (2015) allo sviluppo di dinamiche di successo nella gestione dei processi aziendali e nella creazione di solide relazioni con i consumatori. Le *capacità* riguardano invece la parte di gestione delle risorse, in quanto assicurano un controllo migliore e allo stesso tempo consentono una distribuzione dei dati più soddisfacente (Day, 1994). Gli elementi principali che permettono alle aziende di sviluppare competenze analitiche di successo sono: (a) le persone, l'individuazione di un *team* che abbia le giuste competenze per l'analisi dei Big data consente di disporre di conoscenze sufficienti sui dati e di sviluppare strategie di successo; (b) i sistemi, la selezione del sistema giusto per l'analisi è l'elemento chiave per garantire risultati di alta qualità, in cui coesistono nuove soluzioni, intese come *match* tra nuove tecniche di analisi e soluzioni aziendali tradizionali; (c) i processi, l'insieme di procedure che definiscono le linee guida per l'accesso, la gestione dei dati e la comunicazione all'interno delle aziende; (d) la struttura organizzativa, la quale prevede l'inserimento della funzione analitica all'interno delle aziende. Per quanto concerne l'*analisi* dei Big Data, essa si basa su due forme: (a) un'analisi finalizzata ad ottenere approfondimenti e (b) un'analisi finalizzata a favorire lo sviluppo di modelli. Sia il raggiungimento di approfondimenti che lo sviluppo di modelli può generare nel marketing valore per le imprese e miglioramento nelle campagne di vendita (Marr, 2015) nonché sviluppo di soluzioni basate sulle esigenze degli utenti (Thaler & Tucker, 2013). Infine, attraverso l'analisi viene generato *valore* dai dati. Il processo di analisi prevede, infatti, la creazione di valore nel momento in cui i dati vengono convertiti in conoscenza: conoscenza ed informazione per le aziende e conoscenza dei prodotti e dei servizi offerti per i consumatori. L'analisi dei Big data può fornire dunque una conoscenza sistematica delle valutazioni degli utenti sugli elementi

considerati più o meno interessanti di un'azienda in termini di prodotti, beni e/o servizi (Fan et al., 2015). In particolare, la creazione di valore è la chiave del successo per le aziende in quanto permette di proporre agli utenti prodotti o servizi tenendo anche conto del vantaggio rispetto ai concorrenti. Inoltre, permette alle imprese di esaminare eventuali difficoltà all'interno delle attività aziendali in vista dell'ottimizzazione dei processi (Cossío-Silva et al., 2016).

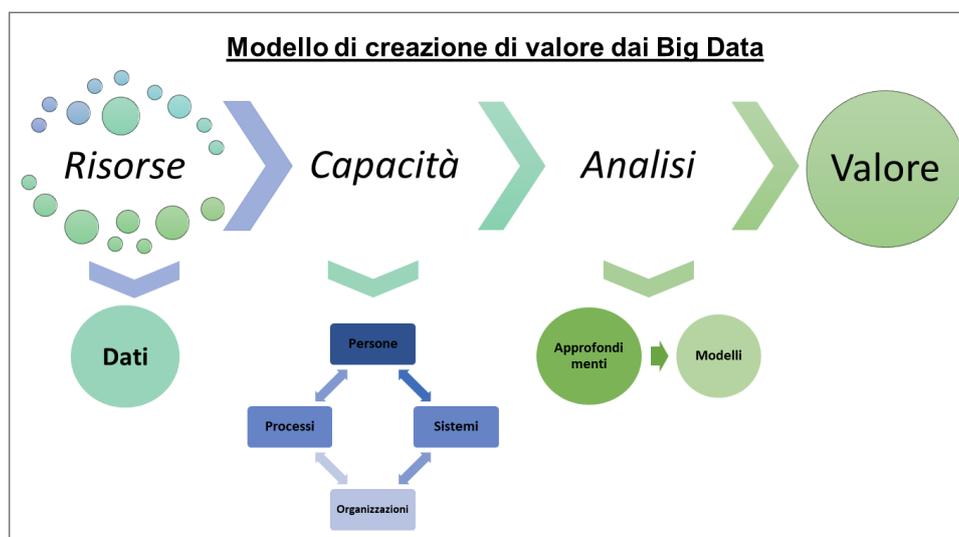


Figura 1.2. Modello di creazione di valore dai Big data proposto da Verhoef et al., 2016

Attualmente, per eseguire operazioni sui dati in modo efficace, efficiente e accurato, le aziende stanno espandendo i loro set di dati tradizionali integrandoli con i dati provenienti per esempio dai social media o dai sensori di sistemi tecnologici ed applicando altre tipologie di analisi come quella relativa all'analisi del testo. Infatti, la creazione di valore per i consumatori, e di conseguenza per l'azienda, in termini di valore economico presuppone una conoscenza approfondita del consumatore e delle variabili cognitive e comportamentali. Pertanto, la capacità di analizzare l'enorme quantità di dati emergenti dal comportamento degli utenti è divenuta una sfida promettente ed un'opportunità interessante sotto diversi punti di vista e per diversi fini. A tal fine, l'analisi dei Big data potrebbe essere definita come la ricerca di indizi e leggi che implicitamente supportano e aiutano a percepire meglio il presente (Jin et al., 2015).

1.2 Big data e settori applicativi

I Big data stanno suscitando attenzione da parte del mondo accademico, governativo e industriale (Jin et al., 2015). Essi infatti, non interessano solo il settore dell'Information Technology (IT), la quale rappresenta lo starter da cui partire, ma sono prodotti nei settori più disparati. Nessun settore in cui esiste un marketing e dei dati da analizzare può ritenersi indenne dalla rivoluzione dei Big Data.

Le principali fonti da cui vengono estratti i Big data sono:

- i. *Black Box Data*, generati dagli aerei. Si tratta dei dati della scatola nera che contengono voci del personale di volo, registrazioni microfoniche ed informazioni relative alle prestazioni dell'aeromobile;
- ii. *Social Media Data*, generati su siti di social media come Twitter, Facebook, Instagram;
- iii. *Stock Exchange Data*, dati della Borsa. In genere contengono informazioni sulle vendite e acquisti sia di commercianti che di clienti privati;
- iv. *Marketing Data*, relativi ad attività svolte online, messaggi di testo, annunci;
- v. *Health Care Data*, generati dalle macchine che forniscono cure ai pazienti, cartelle cliniche elettroniche (EHR), immagini digitali, dispositivi medici wireless;
- vi. *Retail Data*, contengono informazioni sul volume di affari dei centri commerciali, dati demografici, ecc.
- vii. *Sensor Data*, provenienti da sensori integrati in prodotti atti a monitorarne l'utilizzo. Dati integrati in sistemi intelligenti e collegati tramite Internet of Things (IoT);
- viii. *Power Grid Data*, relativi al consumo e all'utilizzo energetico pubblico (commerciale) e privato (residenziale);
- ix. *Transport Data*, contengono informazioni su capacità, modello, disponibilità e distanza coperta da un veicolo;
- x. *Search Engine Data*, prodotti dai motori di ricerca, che si classificano come una delle principali cause e fonti di Big Data, poiché i motori di ricerca dispongono di enormi database e repository dove vengono archiviati questi tipi di dati.

Quindi, quando si parla di Big Data, bisogna tenere presente la fonte dalla quale i dati provengono. A questa è infatti correlata la tipologia di dati che è possibile estrarre; conseguentemente, la scelta e l'eventuale adattamento delle tecnologie utili alla loro elaborazione. Le sfide che si presentano riguardano quindi l'acquisizione, l'archiviazione, la ricerca, la condivisione, l'analisi, il trasferimento e la visualizzazione dei dati.

1.3 Social media: risorsa illimitata di Big data

Con la rapida espansione delle tecnologie digitali ed in particolare dei social media, il tradizionale concetto di gruppo è cambiato. Oggigiorno, far parte di un social network significa far parte di un gruppo, di una comunità. È in questo contesto che viene esteso alla comunità virtuale il concetto di “appartenenza” come definito da Maslow (2010). Questo concetto spiega il grande successo dei social network, dove la partecipazione virtuale ha sostituito quella reale e dove i membri coinvolti sono collegati tra loro attraverso una piattaforma *social*.

Le ragioni della popolarità delle comunità virtuali o *community* risalgono al diciannovesimo secolo quando, il sociologo tedesco Tönnies (1887) propose la distinzione tra *società*, dove vengono stabilite relazioni formali, impersonali e strumentali, caratterizzate da grandi città, stato e grandi organizzazioni, e *comunità privata*, dove i legami sono informali, organici o istintivi e tipizzati dalla famiglia o dal vicinato. L'appartenenza ad una comunità è auto-soddisfacente (motivazione intrinseca), mentre essere membro di una società è un mezzo per raggiungere ulteriori obiettivi individuali (motivazione estrinseca). Nicholas Negroponte (1995) parlò invece dei “quartieri digitali dediti alla socializzazione” anticipando, appunto, quella che ad oggi, viene definita comunità virtuale. Nell'ultimo ventennio si è discusso molto sul concetto e sull'uso dei media digitali per la creazione di *community*. Nel 1997, Hagel e Armstrong presentarono un lavoro nel quale veniva sottolineato il valore delle comunità virtuali, e ne veniva evidenziato il “potere” nel generare “circoli virtuosi” capaci di contribuire alla quantità e alla qualità delle conoscenze. Una comunità tradizionale è formata da un gruppo di individui che condividono lo stesso ambiente fisico formando un gruppo riconoscibile ed uniti da vincoli organizzativi, interessi linguistici, religiosi ed economici (Kim et al., 2011). Allo stesso modo, una comunità virtuale è caratterizzata dagli stessi

elementi chiave di una reale e consiste in uno scambio di informazioni personali (anagrafica, luogo di provenienza, studi effettuati) oppure informazioni generali (registrazioni ad eventi, curiosità) che avviene tra i partecipanti (Füller et al., 2006). Le attività di una comunità virtuale sono abilitate dalla tecnologia dell'informazione, i suoi partecipanti guidano i contenuti o gli argomenti e la relazione si evolve attraverso la comunicazione (McLoughlin et al., 2018). Pertanto, ogni social network rappresenta una struttura sociale costituita da persone come amici, familiari o semplicemente da persone che condividono interessi, obiettivi e valori comuni. Le competenze individuali sono disponibili all'intero gruppo come la creazione, la manipolazione di spazi condivisi e la comunicazione, intesa come espressione di idee o controllo costante sulla qualità dei servizi offerti (Toivonen et al., 2009). In effetti, ad oggi, le comunità virtuali costituiscono un luogo dove condividere ed acquisire informazioni e conoscenze su prodotti e servizi. Rappresentano degli spazi in cui gli utenti stabiliscono connessioni sociali accedendo al cyberspazio (Mueller et al. 2011) e si propongono come un potenziale vantaggio per le aziende in quanto si formano naturalmente attorno a temi ed interessi particolari che possono riguardare prodotti, beni e servizi. Pertanto, l'utilizzo delle piattaforme social facilita l'interazione tra cliente e azienda ed è diventato una pratica consolidata nelle campagne di marketing in quanto permette all'azienda sia di trovare nuovi clienti (sfruttando appunto il passaparola elettronico - eWOM), sia di mantenere quelli esistenti (Lim et al., 2015). Inoltre, dal punto di vista della gestione del marchio, lo sfruttamento delle *fanpage* su social media come Facebook, consente alle aziende e nello specifico al *brand* di aumentare la reputazione (Lee et al., 2012), la consapevolezza (Chen et al., 2014) e l'identità aziendale (Devereux et al., 2017). Queste piattaforme quindi sono tanto a disposizione e a beneficio delle aziende, quanto dell'utente e rappresentano un vero e proprio valore cooperativo nel mondo del web.

Dunque, le piattaforme social come Instagram, Twitter e Facebook, si propongono come canali attraverso cui gli utenti possono condividere liberamente la propria opinione in merito ad esperienze passate ed aspettative su un determinato prodotto, marchio o servizio e sono attivamente incoraggiati a condividere valutazioni e recensioni online (Bronner & de Hoog 2010). Altro esempio sono le piattaforme dedicate: TripAdvisor o Booking, le quali offrono ai viaggiatori l'accesso ad una grande quantità di recensioni online estremamente utili per la pianificazione delle vacanze o dei pacchetti soggiorno. Diverse

ricerche hanno riscontrato come gli utenti, potenziali consumatori, sono più interessati alle raccomandazioni di altri utenti piuttosto che semplicemente alle informazioni sui prodotti offerte dal fornitore (Ridings & Gefen, 2004).

Riepilogando, i social media da un lato, rappresentano una comunità in cui i membri si supportano sia a livello informativo che emotivo (Ballantine & Stephenson, 2011), dall'altro consentono alle aziende, attraverso le analisi delle valutazioni e recensioni, lasciate dagli utenti, di aumentare il livello di fiducia nel *brand* e ridurre i rischi commerciali (Liang & Scammon, 2011). Infatti, la conoscenza dell'aspetto emotivo e informativo, in un contesto online, supporta l'aspetto sociale dunque sia i venditori che gli acquirenti ed è in grado di influenzare il successo e l'insuccesso del commercio elettronico. Questi aspetti sono stati oggetti di studio di molti ricercatori i quali hanno utilizzato l'analisi digitale per capire come sfruttare la presenza dei social media per generare fiducia positiva nel cliente oppure come comprendere il loro comportamento (Sun et al., 2015; Jung et al. 2018; Pantano et al., 2019).

1.4 Cognitive Computing e Machine Learning

Il riconoscimento del valore dell'enorme quantità di dati quotidianamente prodotta dagli individui è un fenomeno che sta cambiando il panorama dell'Information Technology. Questa immensa mole di dati contiene un'incredibile quantità di informazioni che oggi possiamo estrarre e conoscere grazie a tecniche e metodi capaci di coniugare Intelligenza Artificiale (AI) ed Elaborazione dei segnali, note anche come *cognitive computing*. Il *cognitive computing* connette le scienze cognitive e l'informatica, con l'obiettivo di simulare processi di pensiero umani attraverso modelli computerizzati, utilizzando algoritmi di auto-apprendimento che sfruttano sistemi di *data mining*, *pattern recognition* e l'elaborazione del linguaggio naturale. In questa prospettiva, la necessità di sfruttare i Big data per ottenere valore ha spinto i ricercatori a sviluppare strumenti sempre più sofisticati di Intelligenza Artificiale tra cui gli algoritmi di apprendimento automatico (machine learning) (Alpaydin, 2016). Attualmente, l'applicazione di nuovi modelli di elaborazione come il Machine Learning (ML) per l'analisi dei Big data è uno degli argomenti più affrontato in diversi contesti e sono le tecniche maggiormente adottate in numerosi e complessi campi poiché forniscono possibili soluzioni per estrarre le

informazioni nascoste nei dati (Qiu et al., 2016).

«*Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed*». Con questa espressione Arthur Samuel nel 1959 definì l'apprendimento automatico. Tuttavia, la sua definizione lasciava un velo di incertezza sulla modalità di apprendimento del computer.

Gli algoritmi basati su ML sono un campo di indagine teorica, principalmente legato all'Informatica e all'Intelligenza Artificiale e sono strettamente correlati alla matematica e alla statistica applicata. Tuttavia, l'apprendimento automatico è noto per la sua applicazione in settori multidisciplinari, trova infatti applicazione in filosofia, sociologia, biologia, ecc. (Tarca et al., 2007). L'apprendimento automatico prevede lo sviluppo di algoritmi che possono essere utilizzati su grandi classi di problemi, necessitando di un adattamento nel caso di applicazione alla risoluzione di uno specifico compito. Più dati o "esperienza" acquisiscono i computer, migliore diventa la performance, alla pari di un essere umano. Infatti, un apprendimento attraverso l'esperienza è analogo alle funzioni che sono alla base dello sviluppo della cognizione umana. Nello specifico, grazie al progresso tecnologico è possibile avvicinare la capacità analitica della macchina alle capacità umane. Gli algoritmi di ML utilizzano metodi computazionali per estrarre informazioni dai dati creando, ad esempio, delle categorie concettuali. Essi possono essere sviluppati con stili diversi per modellare un problema. Lo stile è dettato dall'interazione con l'ambiente di dati espresso come input per il modello. Lo stile di interazione dei dati determina il modello di apprendimento che un algoritmo di ML può produrre. L'utente deve comprendere i ruoli dei dati di input e il processo di costruzione del modello. L'obiettivo è selezionare il modello in grado di risolvere il problema ottenendo il miglior risultato previsto.

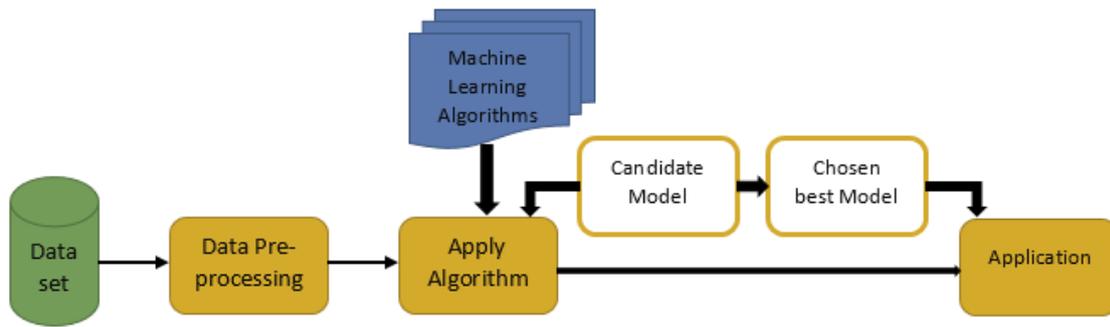


Figura 1.3. Schema funzionale dell’algoritmo di Machine Learning nell’analisi dei dati

1.4.1 Apprendimento supervisionato e non supervisionato

Non esiste una sola forma di intelligenza artificiale in quanto ogni applicazione si basa su modelli di Machine Learning specifici e su tecniche differenti in base all’analisi dei dati, anche in funzione dell’obiettivo che si cerca di raggiungere.

Non esiste nemmeno una forma unica di “apprendimento” ma esistono vari tipi con diverse caratteristiche. Si parla infatti di:

- i. apprendimento supervisionato (*supervised learning*);
- ii. apprendimento non supervisionato (*unsupervised learning*).

Nell’apprendimento supervisionato vengono forniti esempi sotto forma di possibili input e dei rispettivi output richiesti. Lo scopo è quello di estrarre una regola generale che associ l’input all’output corretto creando un modello. L’apprendimento supervisionato è il processo di apprendimento in cui è nota la variabile di output. In questo tipo di apprendimento ai dati viene attribuita una categoria di riferimento, in altre parole si è consapevoli del risultato previsto. In realtà, questo algoritmo contiene una variabile target o *outcome* che deve essere prevista da un determinato set di predittori (variabili indipendenti). Usando questo insieme di variabili, verrebbe generata una funzione che mappa gli input ai risultati previsti. Il processo di addestramento prosegue fino a quando il modello raggiunge un livello di correttezza previsto sui dati di addestramento. Le principali tecniche utilizzate nell’apprendimento supervisionato sono:

- i. *Regressione*, dato un numero di variabili predittive ed una variabile target

continua, si tenta di trovare una relazione tra le variabili per prevedere un risultato;

- ii. *Classificazione*, in base a dati precedentemente etichettati, lo scopo è di prevedere l'etichettatura di future classi di dati.

Questi sistemi sono chiamati classificatori (maggiore è la complessità degli oggetti da classificare e il numero di categorie, più grande sarà il set di addestramento).

Nell'apprendimento non supervisionato, la macchina unisce gli oggetti in gruppi, apprendendo dai dati senza un modello di riferimento. In particolare, durante il processo, mira a trovare una struttura negli input forniti, senza che gli input siano etichettati in alcun modo (Love, 2002). L'apprendimento non supervisionato viene in genere utilizzato per trovare specifiche relazioni all'interno del set di dati. Non ci sono esempi di formazione utilizzati in questo processo, ma al sistema viene fornito un insieme di dati in input dai quali trovare modelli, correlazioni nascoste o strutture intrinseche. Le principali tecniche utilizzate nell'apprendimento non supervisionato sono:

- i. *Clustering*, una tecnica esplorativa che permette la creazione di gruppi, non avendo precedente conoscenza sui dati raggruppati. È una tecnica che consente di scoprire le relazioni tra i dati, in quanto nei gruppi vi saranno dati aventi caratteristiche simili tra loro;
- ii. *Riduzione della dimensionalità*, una tecnica utilizzata nelle pre-elaborazione delle caratteristiche, relative ai dati, con lo scopo di combinare le informazioni correlate e ridondanti ed eliminare il "rumore" dai dati.

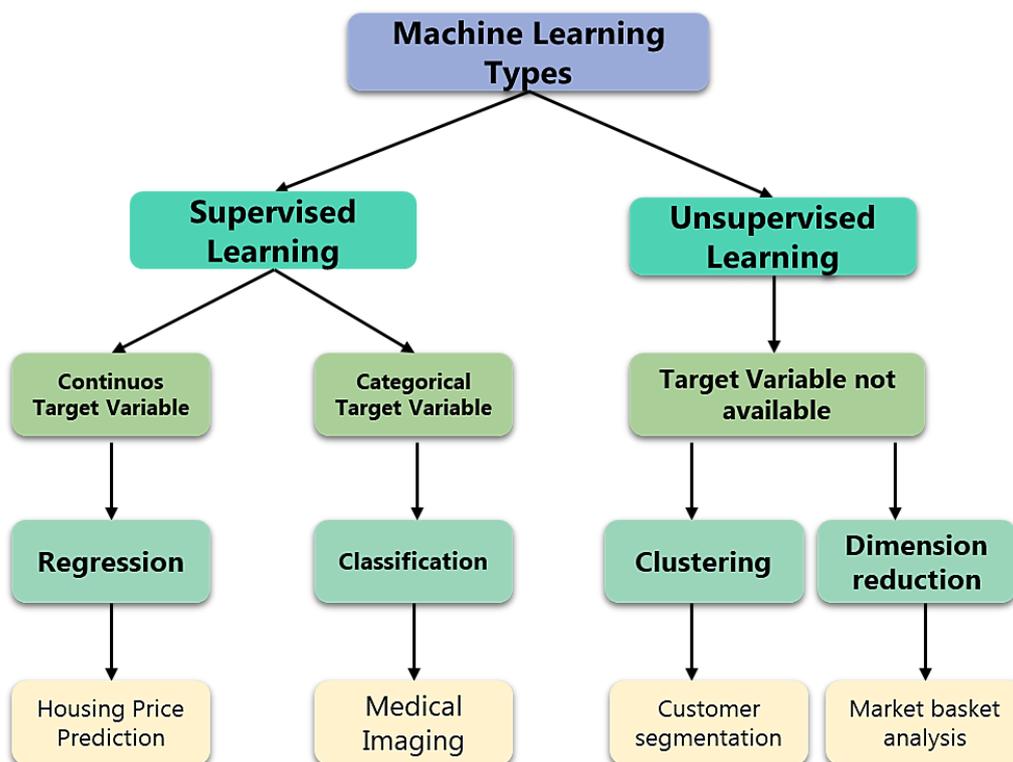


Figura 1.4. Tipi di Machine Learning

In definitiva, l'apprendimento automatico consiste nello studio e nella modellizzazione di algoritmi che possono apprendere dai dati, estrapolando da essi nuove conoscenze.

1.4.2 Deep Learning e Reti neurali

Lo strumento più classico per costruire il Machine Learning è la rete neurale. Fonda le sue radici nella ricerca degli anni '70 del secolo scorso, quando furono sviluppati software in grado di conservare, organizzare e gestire conoscenze specifiche.

Le reti neurali si ispirano al funzionamento biologico del cervello umano, ovvero a modelli costituiti da interconnessioni di informazioni. Le reti neurali artificiali (Artificial Neural Network - ANN) sono modelli matematici composti da neuroni artificiali e rappresentano la principale implementazione del *Deep Learning*, oggetto di studio di numerosi scienziati sin dal 1979. Il Deep Learning è una branca del Machine Learning che trova applicazione anche nella risoluzione di problemi complessi derivanti dai Big data (Deng et al., 2014). Esso concerne una serie di tecniche di apprendimento automatico

multilivello; infatti il termine *deep* fa riferimento alla profondità degli strati (Ahmad et al. 2018). Questo tipo di apprendimento ha avuto successo in molti settori applicativi, tra cui l'elaborazione delle immagini, l'assistenza sanitaria, l'agricoltura e i trasporti e permette di sfruttare sempre più dati per generare informazione. Recentemente, questo settore ha richiamato una grande attenzione nell'elaborazione del linguaggio naturale e nella classificazione delle immagini, intesa sia come rilevamento e localizzazione degli oggetti che come creazione delle immagini grazie alle dimensioni e alle connessioni in strutture complesse delle ANN che possono facilmente contenere fino a centinaia di strati. Grandi società come Google, Microsoft ed Amazon usano quotidianamente tecniche di Deep Learning per analizzare enormi quantità di dati. In particolare, gli assistenti personali computerizzati, come Siri di Apple, Alexa di Amazon o Cortana di Microsoft, fanno un uso intensivo di Deep Learning per riconoscere, comprendere e rispondere a domande umane (Kraus et al., 2019). Netflix, invece, utilizza le *Deep Neural Network* (reti neurali profonde) come sistema di raccomandazione per offrire consigli personalizzati sui prodotti ai clienti.

1.5 Big data Analytics

L'estrema variazione tra le fonti di dati e i tipi di dati ha ulteriormente aggravato i colossali compiti di gestione dei dati. Quindi, la nozione di Big data è stata estesa a Big data Analytics (BDA) che mira ad estrapolare da questi dati intelligenza sotto forma di informazione. In particolare, la letteratura suggerisce che l'analisi dei Big data consiste nella "raccolta, analisi, visualizzazione, uso e interpretazione dei dati per varie divisioni funzionali al fine di ottenere approfondimenti attuabili, creare valore aziendale e stabilire un vantaggio competitivo" (Motamarri et al., 2017, pp. 625). Pertanto, come già sottolineato, progettare un *framework* con il piano e gli strumenti giusti per gestire i Big data è fondamentale per valorizzare ed orientare al meglio i risultati dell'analisi dei dati (Blazquez & Domenech, 2018). Negli ultimi anni, lo sfruttamento di tali dati ha favorito lo sviluppo di applicazioni in grado di portare opportunità interessanti e benefici tangibili alla collettività (Gil & Song, 2016). Tecniche avanzate consentono di eseguire analisi più complesse ed efficaci fornendo a professionisti e ricercatori risultati più validi e precisi (Sheng et al., 2019). In particolare, l'analisi dei Big data comprende diverse tecniche

quali: l'analisi descrittiva, l'analisi prescrittiva e l'analisi predittiva (Wang et al., 2018).

- i. L'analisi *descrittiva* consente alle organizzazioni di riassumere ciò che è accaduto e adottare soluzioni ottimali che potrebbero influire sulle prestazioni future. Gli strumenti adottati da questo tipo di analisi per accedere ed analizzare i dati sono utili per descrivere la situazione attuale o passata dei processi aziendali. L'analisi descrittiva comprende tecniche come la regressione e la visualizzazione.
- ii. L'analisi *prescrittiva* cerca di quantificare l'effetto delle decisioni future per consigliare possibili risultati prima che vengano effettivamente implementate. L'analisi prescrittiva fornisce ciò che accadrà e spiega il perché accadrà fornendo raccomandazioni a riguardo. Questa analisi comprende tecniche come l'ottimizzazione, la simulazione e la modellazione numerica.
- iii. L'analisi *predittiva* tenta di scoprire modelli e mostrare relazioni nei dati. Questa analisi consente di capire cosa è successo in passato e prevedere probabili tendenze o eventi futuri. Gli strumenti adottati sono caratterizzati da tecniche matematiche come i modelli predittivi.

L'ambiente emergente dei Big data stimola gli studiosi verso nuove ricerche in grado di esplorare “nuovi argomenti al limite delle nostre attuali conoscenze” (Merendino et al., 2018). Inoltre, poiché la quantità di dati si evolve continuamente e i formati e le qualità dei dati continueranno a crescere e a essere digitalizzati, la chiave per un'innovazione di successo è investire in infrastrutture per Big Data, inclusa la figura del Data Scientist. La crescita della scienza dei dati o *data science* (termine da cui il prende il nome il Data Scientist) è vista come uno dei fattori più influenti, soprattutto nel settore del marketing contemporaneo (Balducci & Marinova, 2018) in quanto permette di analizzare i dati con successo per raggiungere un vantaggio competitivo trasformando i risultati ottenuti in strategie di successo.

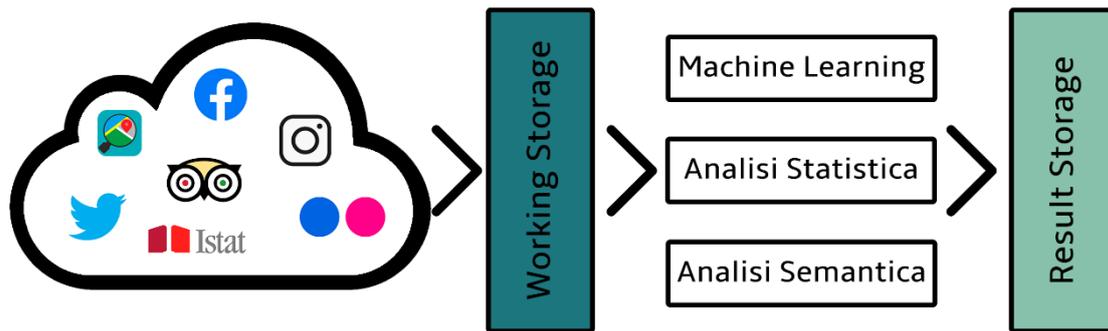


Figura 1.5. Framework di analisi dei Big data

1.5.1 Analisi dei Big data con Wolfram Mathematica

Wolfram Mathematica® o *Mathematica* è un moderno sistema di elaborazione tecnica a supporto di tutte le aree dell'informatica. Rispetto ad altri software, ad esempio R o Matlab, *Wolfram Mathematica* offre differenti funzioni utili per eseguire analisi in una vasta gamma di settori (Chonacky & Winch, 2005). Inizialmente, l'impatto di questo software è stato avvertito principalmente nel campo della scienza, dell'ingegneria e della matematica, diventando, solo successivamente, importante ambiente di calcolo anche in altri settori. Nel campo ingegneristico *Wolfram Mathematica* è diventato uno strumento standard sia per lo sviluppo che per la produzione. In campo economico è stato significativo nella crescita di sofisticati modelli finanziari, oltre ad essere ampiamente utilizzato per la pianificazione ed analisi generali, mentre per quanto riguarda l'informatica la sua componente linguistica è ampiamente utilizzata come ambiente di ricerca, prototipazione ed interfaccia. Da un punto di vista prettamente tecnico, è un ambiente di programmazione che utilizza il "linguaggio Wolfram". È basato sulla riscrittura di espressioni (term-rewriting) e supporta svariati paradigmi di programmazione, tra cui la programmazione funzionale, la programmazione logica, la programmazione basata sul riconoscimento di schemi (pattern-matching) e sulle regole di sostituzione (rule-based), nonché la più tradizionale programmazione procedurale (Wolfram, 2002). *Wolfram Mathematica* è realizzato principalmente in C e in C++, ma gran parte delle numerose librerie fornite con il programma sono scritte nel linguaggio

proprietario di *Mathematica*, che può essere utilizzato per espandere ulteriormente le funzionalità del sistema. In questo ambiente il linguaggio di base viene interpretato da un kernel (ovvero il software avente il compito di fornire ai processi in esecuzione sul computer un accesso sicuro e controllato all'hardware) che esegue l'elaborazione vera e propria, i risultati vengono quindi comunicati ad una specifica interfaccia tra quelle disponibili. La comunicazione tra il kernel e questi ultimi utilizza il protocollo di MathLink, spesso attraverso una rete. Il kernel stesso è composto da oltre 1800 funzioni disponibili dopo l'inizializzazione di *Mathematica* e gestisce calcoli come: differenziazioni simboliche, integrazioni simboliche, rappresentazioni grafiche, valutazioni di serie, somme e così via (Baumann, 2006).

Ad oggi il software *Mathematica* viene applicato nella biologia per l'elaborazione di immagini, nella scienza dei dati, nella geometria e nella visualizzazione. Il linguaggio *Mathematica* impiega diversi metodi e strumenti per ridurre la complessità dei dati e fornisce una serie di metodi all'avanguardia che consentono di eseguire automaticamente diverse operazioni come l'analisi delle recensioni dei turisti (Pantano et al., 2017), l'identificazione dell'interazione cliente-robot (Bertacchini et al., 2017) oppure supporta procedure applicate su dati reali e relativi a sequenze di scosse di assestamento (Del Pezzo & Bianco, 2010).

Wolfram Mathematica è stato scelto come software per condurre le analisi in questa ricerca. In generale, i dati sono stati esaminati attraverso tecniche innovative quali algoritmi di *machine learning*, analisi *semantica* ed analisi *statistica*. Gran parte del lavoro è stato basato su tecniche di Machine Learning utilizzate per il riconoscimento delle immagini, per la creazione di cluster, per l'analisi del testo (*sentiment analysis*) e per la profilazione di classi di utenti. Le tecniche di analisi applicate e di seguito presentate hanno permesso di elaborare modelli di analisi e scoprire realmente il valore intrinseco dei Big Data, processo probabilmente "inattuabile" applicando algoritmi tradizionali.

1.5.2 Image Analysis

Il concetto di riconoscimento delle immagini è strettamente correlato al concetto di *computer vision*, il quale indica un processo di addestramento durante il quale la macchina

impara a “riconoscere” così come gli esseri umani. In letteratura, il riconoscimento delle immagini viene definito come un processo che consiste nell’identificazione di oggetti, luoghi e persone o altri elementi con l’obiettivo di estrarre da essi informazioni significative (Pantano et al., 2019). Questo processo viene utilizzato per eseguire attività come l’etichettatura delle immagini, la quale ne definisce la categoria. Il riconoscimento delle immagini è legato alle tecnologie informatiche ed è dunque relativo al grado di riconoscimento di elementi attraverso l’uso e l’applicazione dell’Intelligenza Artificiale (AI). Viene eseguito in modi diversi, le migliori tecniche prevedono l’uso di reti neurali convoluzionali (CNN), un tipo di Artificial Neural Network (ANN) feed-forward dove i neuroni sono organizzati a strati, dunque l’informazione si muove solo in una direzione, e attraverso le quali le immagini vengono filtrate tramite una serie di strati di neuroni artificiali (LeCun et al., 1998). Una tipica architettura CNN utilizza una struttura tridimensionale, in cui ogni gruppo di neuroni analizza una regione o “caratteristica” particolare dell’immagine. Nello specifico, un’immagine (composta da una matrice di pixel) viene inviata alla rete; l’immagine di input passa attraverso un numero infinito di passaggi ovvero la parte convoluzionale durante la quale, la rete mantiene le caratteristiche essenziali dell’immagine ed esclude il rumore irrilevante. Ad esempio, il modello sta imparando a riconoscere da un’immagine un elemento un’automobile con un albero sullo sfondo. Se durante il processo venisse applicata una rete neurale tradizionale, il modello assegnerà un peso a tutti i pixel, compresi quelli dell’albero, mentre una rete neurale convoluzionale utilizzerà una tecnica matematica per estrarre solo i pixel più rilevanti (solo relativi all’automobile). Questa tecnica consente alla rete di apprendere funzionalità sempre più complesse su ogni livello. Lo scopo della convoluzione è quello di estrarre localmente le caratteristiche dell’oggetto nell’immagine. Questo dimostra che la rete apprende schemi specifici all’interno dell’immagine i quali saranno immagazzinati come informazione. L’output finale è un vettore che contiene, per ogni caratteristica nell’immagine, la probabilità che questi appartenga ad una specifica classe o categoria.

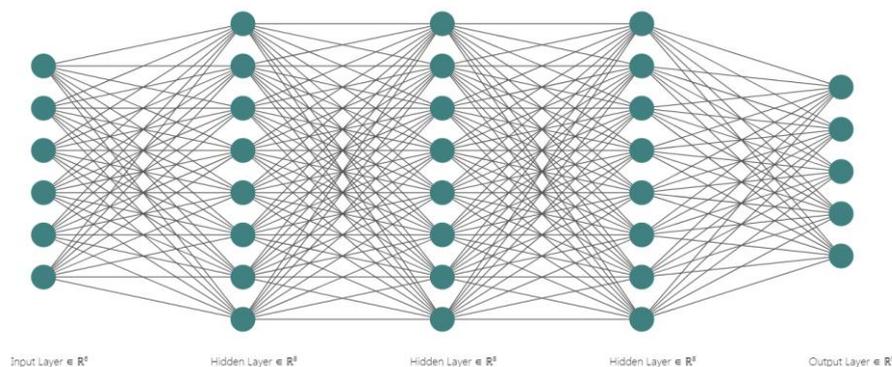


Figura 1.6. Rappresentazione di una rete neurale

Come anticipato, oltre alle CNN, altre tecniche vengono applicate per il riconoscimento delle immagini tra cui l'algoritmo di riconoscimento delle immagini che rappresenta un classificatore. Il processo di riconoscimento avviene con applicazione dell'algoritmo che seleziona un'immagine in input e fornisce in output l'informazione sul suo contenuto. L'output è un'etichetta di classe. Tuttavia, al fine di apprendere e distinguere tra le classi l'algoritmo deve essere addestrato. Ad esempio, per creare un algoritmo di classificazione in grado di identificare le immagini raffiguranti automobili, verrà creata una rete neurale con migliaia di immagini di automobili e migliaia di immagini di sfondi senza automobili. L'algoritmo imparerà ad estrarre le caratteristiche che identificano l'oggetto "automobile" e a classificare correttamente le immagini che contengono tale elemento. Il riconoscimento delle immagini è entrato nel *mainstream* dell'analisi dei dati non strutturati tale che le implementazioni sviluppate per eseguire il riconoscimento delle immagini includono anche il riconoscimento dei volti ed oggetti, la geo-localizzazione visiva, il riconoscimento dei gesti, l'analisi delle immagini mediche, assistenza alla guida, sicurezza e sorveglianza.

Il riconoscimento di volti, foto e fotogrammi video viene utilizzato anche da piattaforme social come Facebook, Google e da molte altre applicazioni di alto profilo. Infatti, i social media come Instagram, che lavorano principalmente con le immagini, sono utilizzati anche per "enfaticizzare" luoghi e brand, il che non implica la generazione di nuove immagini, ma la scelta di immagini particolari e di alta qualità, in base alle funzionalità del mezzo utilizzato, ovvero la scelta di filtri particolari per aggiungere effetti digitali accattivanti. Inoltre, grazie alla capacità delle piattaforme online di fornire un accesso

sempre più ampio agli utenti, le immagini sono tra i contenuti preferiti inclusi nei post online e sono in grado di migliorare l'attrattività di location ed ambienti (Minazzi & Lagrosen, 2013). La letteratura considera le immagini scattate dagli utenti come ricche fonti di dati per la ricerca (Balomenou & Garrod, 2019) in quanto la loro analisi permette di comprendere la percezione dell'immagine del marchio (O'Connor, 2010), l'attrattiva di una destinazione turistica (Perez-Vega et al., 2018) oppure consente di determinare la qualità dell'esperienza vissuta (Banrjee & Chua, 2016).

1.5.3 Text e Content analysis

L'analisi del testo è una popolare tecnica di analisi basata su una serie di procedure di classificazione e di tecniche analitiche in grado di estrarre informazioni da fonti testuali per approfondimenti, tendenze o modelli (Krippendorff, 2018). Sebbene l'analisi del testo viene utilizzata per codificare il testo di un documento, essa offre ai ricercatori anche l'opportunità di studiare i valori, i sentimenti e le preferenze degli utenti, in quanto affrontare l'analisi dei contenuti significa studiare le attitudini dei consumatori e i loro cambiamenti nel tempo (Morris, 1994). Infatti, la letteratura suggerisce che (Titscher et al., 2000) il metodo di analisi del testo più consolidato tra l'insieme dei metodi empirici di indagine è l'analisi del contenuto. In generale, gli elementi dell'analisi dei contenuti sono i testi tra cui gli articoli di giornale, le trascrizioni delle interviste e così via. Il loro contenuto, cioè il significato dei simboli, è organizzato secondo regole di elaborazione delle informazioni intersoggettivamente comprensibili (Kromrey, 1998). Il principio alla base dell'analisi del contenuto è che le forme di espressione culturale possono essere espresse nei testi, questo comporta che l'analisi del contenuto dei testi concerne la realtà sociale e tutti i risultati dell'analisi e la loro interpretazione dipendono in modo corrispondente. Inizialmente, l'analisi del contenuto mirava alla "descrizione obiettiva, sistematica e quantitativa del contenuto manifesto della comunicazione" (Berelson, 1952, p. 18). Nel tempo, l'analisi è stata estesa anche all'interpretazione del contenuto.

Attualmente, le tecniche principali utilizzate per eseguire questo tipo di analisi sono: i metodi qualitativi di analisi del contenuto e le analisi di quantificazione del contenuto. Il primo consente esplicitamente una quantificazione in termini di valutazione statistica dei dati raccolti (Mayring, 2010), mentre il secondo integra nella loro concezione, gli aspetti

di una teoria orientata alla formazione delle categorie e relativa valutazione. Quando si esegue l'analisi del contenuto qualitativo, si affronta una problematica legata al tipo di contenuto da analizzare. L'analisi di ciò che dice il testo, tratta dell'aspetto del contenuto e descrive le componenti visibili e ovvie, indicate come contenuto manifesto. Al contrario, l'analisi di ciò di cui parla il testo tratta dell'aspetto relazionale e implica un'interpretazione del significato sottostante il testo, indicato come contenuto latente. L'approccio quantitativo è spesso utilizzato nella ricerca sui media.

Labrinidis e Jagadish (2012) identificano cinque fasi per l'estrazione e la trasformazione di elevati volumi di dati da un testo. Il processo proposto dagli studiosi inizia con la gestione dei dati: acquisizione, estrazione e interazione e procede con l'analisi dei dati: modellizzazione, analisi ed interpretazione. Nella prima fase dell'analisi del testo uno specifico strumento di estrazione del testo acquisisce documenti, estrae e raccoglie dati. Successivamente, i dati raccolti vengono analizzati fino a quando non vengono estratte le informazioni. Lo scopo dell'analisi è ridurre la complessità dei dati e sviluppare uno schema utile al dominio di varie applicazioni. L'analisi del testo comprende l'analisi statistica, la linguistica computazionale e l'apprendimento automatico. In effetti, esiste una relazione tra analisi del testo e la linguistica computazionale che rispecchia l'aspetto emozionale degli utenti in specifici contesti. Nell'ambito della ricerca sui Big Data, gli studiosi hanno applicato diversi metodi automatizzati su differenti tipologie di dati per eseguire analisi in relazione al contenuto del testo. Questo tipo di analisi consente di studiare il comportamento degli utenti senza dover prima avviare il processo di osservazione e quindi permette ai ricercatori di effettuare l'analisi basandosi non su ciò che le persone hanno fatto ma sul perché le persone lo fanno (Verma et al., 2016).

1.5.4 Sentiment analysis

L'analisi del sentiment, nota anche come *opinion mining*, è un processo di analisi che riguarda l'estrazione del significato delle parole, in particolare dell'emozione “dietro” le parole. L'analisi del sentiment è un campo di elaborazione del linguaggio naturale, della linguistica computazionale e del text mining. In linea generale questo processo di analisi mira ad identificare l'atteggiamento, in termini di giudizio o valutazione, di un oratore o uno scrittore in relazione ad un argomento. L'origine dell'analisi del sentiment risale agli

anni '50, quando venne utilizzata principalmente su documenti cartacei scritti. Attualmente, questo processo sta diventando un metodo popolare ed utile al fine di estrarre informazioni soggettive dai contenuti condivisi online, inclusi testi, tweet, recensioni e commenti, attraverso procedure automatizzate (Xu et al., 2011). Alcuni dei primi lavori, in questo contesto furono, condotti da Turney (2002) che applicò diversi metodi per rilevare la polarità di recensioni relative ad automobili, banche, film e destinazioni di viaggio. Anche Pang e colleghi (2002) e Snyder e colleghi (2007) applicarono l'analisi del sentiment utilizzando però un metodo differente basato sulla classificazione della polarità di un documento tramite una scala di valori. Nello specifico, Pang e colleghi esaminarono un filmato classificandolo come positivo o negativo, in base ad una scala a 3 o 4 stelle. Snyder e colleghi effettuarono, invece, un'analisi approfondita su recensioni di ristoranti, analizzando i giudizi sui vari attributi del ristorante come cibo e atmosfera avvalendosi di una scala a 5 stelle. Pertanto, in entrambe le ricerche il compito fondamentale è stato classificare la polarità dei sentimenti espressi in un testo ovvero il livello positivo, negativo o neutro.

Esistono tre livelli principali di classificazione nella sentiment analysis: (i) livello del documento, (ii) livello della frase e (iii) livello dell'aspetto. Il primo livello mira a classificare un documento come espressione di un giudizio o sentimento positivo o negativo; il secondo mira a classificare il sentimento espresso in ciascuna frase (oggettivo o soggettivo); l'ultimo livello mira a classificare il sentimento in relazione agli aspetti specifici delle entità (Medhat et al., 2014). Durante il processo di identificazione del sentiment l'aspetto oggettivo e soggettivo è molto importante in quanto la soggettività di parole e frasi può dipendere dal loro contesto e dunque anche un documento oggettivo può contenere frasi soggettive, basti pensare ad un testo di un articolo che riporta le opinioni della gente. Tuttavia, Pang e Lee (2008) dimostrarono che la rimozione di frasi oggettive da un documento prima di classificare la sua polarità può contribuire a migliorare le prestazioni.

Per quanto riguarda le applicazioni utilizzate per eseguire l'analisi del sentiment possono essere citate le seguenti: *Latent Semantic Analysis*, *Support Vector Machines (SVM)* e *Bag-of-Words* che corrispondono a tecniche di Machine Learning. Inoltre, software open source come GATE e RapidMiner forniscono Machine Learning e tecniche di elaborazione del linguaggio naturale che permettono di automatizzare l'attività di

sentiment analysis su grandi collezioni di dati e testi come ad esempio le pagine web, gruppi di discussioni su blog, web e social media quali Facebook. Per fare qualche esempio, a partire da una raccolta di dati online, un ciclo di analisi del sentiment consente di scoprire informazioni sulle percezioni positive o negative dei consumatori; le differenze tra esperti e consumatori in merito alla valutazione dei prodotti (Lee & Bradlow, 2011); il contenuto emotivo dei blog (Weng & Lee 2011); il flusso del linguaggio emotivo e i loro effetti persuasivi (Van Laer et al., 2018).

1.5.5 Clustering analysis

L'analisi dei cluster o *clustering* è una tecnica di analisi dei dati multivariata che consente la selezione e il raggruppamento di elementi omogenei in un gruppo (Jain et al., 1999). I gruppi (o i loro centroidi) vengono utilizzati al posto dei dati come dimensioni dello spazio vettoriale. Tale tecnica, può essere adottata per identificare uno stato adatto tra i dati considerando la variabile desiderata. Lo scopo dell'analisi dei cluster è identificare il minor numero di gruppi in modo che gli elementi appartenenti a un gruppo siano più simili tra loro, nello specifico, dividere un set di dati in sottoinsiemi "sufficientemente distinti". Il punto di partenza dell'analisi dei cluster è la definizione di una misura di somiglianza o distanza tra le righe della matrice di dati. L'altro punto fondamentale è la regola con cui vengono creati i gruppi. In linea di principio, le tecniche di *clustering* possono essere applicate a qualsiasi set di dati.

È una tecnica di *data mining* e risulta particolarmente utile quando si dispone di Big data e si prova a trasformarli in semplici oggetti rappresentativi. In generale, ogni volta che risulta necessario classificare dati di grandi dimensioni in informazioni significative e gestibili, l'analisi dei cluster è molto utile (Rubtzov et al., 2016). Si distinguono diversi tipi di raggruppamenti in *cluster*, i più noti sono:

- i. *Partizionale*
- ii. *Gerarchico*
- iii. *Esclusivo*
- iv. *Sovrapposto*
- v. *Probabilistico*

Un cluster (i) partizionale è una divisione dell'insieme di elementi in sottoinsiemi (cluster) non sovrapposti in modo tale che ogni elemento appartenga esattamente ad un sottoinsieme. Se permettiamo ai cluster di avere dei cluster secondari, otteniamo un cluster (ii) gerarchico, che è un insieme di cluster nidificati organizzati ad albero; nel cluster (iii) esclusivo un determinato dato appartiene a un cluster definito e non potrebbe essere incluso in un altro cluster; il cluster (iv) sovrapposto, viene utilizzato per raggruppare i dati, in modo che ciascun punto possa appartenere a due o più cluster con diversi gradi di appartenenza. In sostanza un oggetto può appartenere a più di un gruppo (classe) contemporaneamente. L'ultimo tipo, (v) cluster probabilistico raggruppa gli elementi in modo casuale (Tan et al., 2006). Generalmente, ad ogni tipo di *clustering* possono essere associati alcuni algoritmi applicati nell'analisi come il *K-mean* che si classifica tra gli algoritmi di clustering esclusivo; *Fuzzy C-mean* che è un algoritmo di clustering sovrapposto; *Mixture of Gaussians* è un algoritmo di cluster probabilistico.

Le attuali sfide nel *clustering* includono l'introduzione di una nuova generazione di algoritmi distribuiti o *distributed algorithms* (algoritmo che viene eseguito su un sistema distribuito, ovvero un insieme di computer indipendenti che non condividono la loro memoria, ma in cui ogni processore ha la sua memoria e comunica tramite reti), ovvero algoritmi in grado di operare facilmente con le recenti infrastrutture di elaborazione distribuita, come il *cloud computing*, abbinandole alla complessità dei dati (Mahmood, 2016). Infatti, l'analisi dei cluster, basata su un processo di apprendimento senza supervisione, è una delle tecniche più comunemente applicate e utile per formare il partizionamento di un set di dati in sottogruppi che sono simili tra loro e sono abbastanza diversi dagli altri gruppi (Antoniadis, 2013). Attualmente, esiste un numero infinito di esempi in cui il *clustering* svolge un ruolo importante ad esempio nel campo della geologia (Pappadà et al., 2018) dove il *clustering* può essere applicato per analizzare le catastrofi alluvionali; in archeologia per stabilire tassonomie di oggetti in pietra (Caspari e Jendryke, 2017); nel turismo, per supportare lo sviluppo del turismo rurale e attuare piani di gestione, promozione e valorizzazione delle destinazioni turistiche (Dona & Popa, 2013).

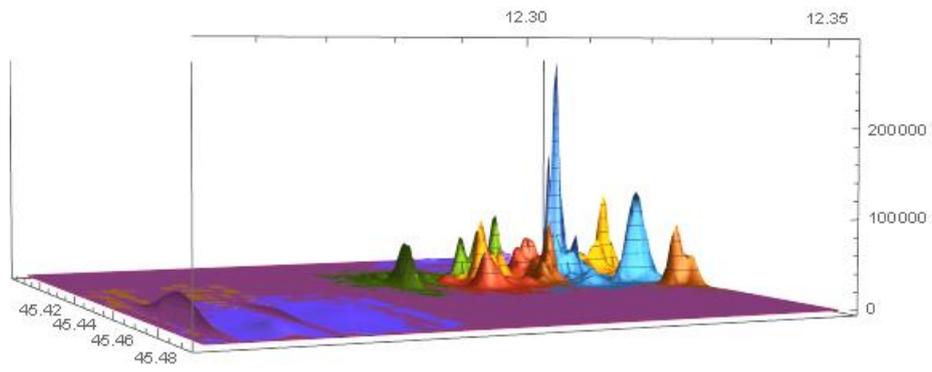


Figura 1.7. Istogramma 3D rappresentativo in cui il set di dati è stato diviso in 19 gruppi distinti.

CAPITOLO 2

BIG DATA ED INDUSTRIA DEL TURISMO

2.1 Il turismo ai tempi dei Big Data

Oggigiorno il settore turistico riveste un'importanza fondamentale come fattore di benessere economico e di sviluppo sociale rappresentando una delle maggiori attività economiche mondiali grazie, soprattutto, allo sviluppo delle comunicazioni, dei trasporti e delle tecnologie informatiche. Il turismo si classifica, infatti, come uno dei settori industriali in continua evoluzione ed in cui le nuove tecnologie hanno trovato e trovano una vasta applicazione, rapidamente adottata da tutti gli attori coinvolti ed in grado di generare enormi effetti sinergici attraverso l'uso di Internet. Nell'ultimo ventennio, Internet ha trasformato il mondo delle comunicazioni, affermandosi come canale privilegiato per la cooperazione e l'interazione tra le persone e i loro dispositivi digitali. L'invenzione di Internet è considerata, infatti, uno dei risultati di maggior successo del serio investimento in ricerca e lavoro tra università, istituzioni ed industria. Infatti, con l'avvento della rivoluzione informatica, la società globale è stata radicalmente ridisegnata ed il mondo è stato spinto sempre più verso un'economia basata sull'informazione. La tecnologia dell'informazione si è affermata come potenziale strumento per attuare cambiamenti nella struttura e nei processi dell'intero settore turistico (Merten, 2007) e viene utilizzata per migliorare i servizi turistici come: le prenotazioni di viaggi (Xiang et al., 2015), la pianificazione di itinerari, il marketing territoriale e della destinazione (Di Pietro et al., 2012), contribuendo in modo radicale alla condivisione delle informazioni. In particolare, il fenomeno della digitalizzazione ha favorito la nascita di un bacino di servizi e di un mercato totalmente nuovo originato, principalmente, dall'enorme produzione e flusso di dati, di informazioni e di conoscenze a disposizione di chiunque

volesse accedervi, senza la necessità di particolari abilità. Attualmente, esistono varie tecnologie che permettono di condividere ed usufruire facilmente e rapidamente di informazioni digitali. Questo processo è facilitato dalla varietà, dall'adattabilità e dell'uso incontrollato di dispositivi elettronici come PC, Smartphone, Tablet, console di gioco ed elettrodomestici a livello hardware da un lato e da applicazioni, servizi online e media a livello software dall'altro. Si tratta quindi di dispositivi e applicazioni che generano una grande quantità di dati e attraverso cui gli utenti lasciano continuamente "tracce digitali" di qualsiasi tipo: da semplici foto o video a documenti. La crescente diffusione dei dati digitali ha portato alla nascita di un fenomeno chiamato Big Data, di cui si è largamente parlato nel capitolo precedente. I Big data rappresentano, senza dubbio, una nuova sfida per il turismo, in cui le preferenze dei turisti, le loro abitudini e il loro comportamento sono diventati una componente chiave per gestire gli interessi economici e culturali di un luogo al fine di capire come dirigere gli investimenti degli stakeholder.

2.2 Mobilità turistica e comportamento degli utenti

Svariate sono le attività quotidiane che implicano il movimento "spaziale": recarsi a lavoro, andare a scuola, fare shopping, condurre tutte quelle attività il cui comune denominatore è lo spostamento nello spazio (Ai et al., 2019). Tuttavia, oggi, con lo sviluppo dell'informatica e delle tecnologie dell'informazione e della comunicazione (ICT) un numero crescente di attività umane viene svolto, non più solamente nello spazio fisico, bensì in quello virtuale. Le attività e le interazioni condotte nello spazio virtuale non interferiscono o modificano lo spazio fisico-geografico, ma lo influenzano, lo arricchiscono ed interagiscono con esso. Il comportamento spaziale ha acquisito maggiore rilevanza con l'avvento di Internet che ha permesso, al contrario dei tradizionali metodi basati su questionari (Barbosa et al., 2018), di comunicare con precisione i comportamenti dell'utente ad esempio relativi all'utilizzo di dispositivi elettronici (quali il telefono) (Chittaranjan et al., 2013) o relativi al movimento.

In questo contesto la produzione dei Big data diventa un elemento importante e di supporto per comprendere le dinamiche in entrambi gli spazi - fisici e virtuali - (Shaw et al., 2016) consentendo di analizzare gli schemi spazio-temporali dei turisti e contribuendo a sviluppare conoscenze relative ai luoghi più comunemente visitati e ai comportamenti assunti dai viaggiatori. In particolare, il desiderio di comprendere le preferenze turistiche

di tutto il mondo ha portato all'emergere di molti servizi basati sulla localizzazione. Gran parte della attività "social" condotte dagli utenti, generate e condivise online, che implicano spostamenti, sono geo-localizzate. Queste informazioni, volontariamente fornite dall'individuo sono denominate Volunteered Geographic Information (VGI) (Devkota et al., 2019) e includono strumenti per creare, assemblare e divulgare i dati geografici forniti. Il Web 4.0 con le sue applicazioni fornisce una distribuzione continua e volontaria delle informazioni attraverso piattaforme social come Flickr, Facebook, Twitter, Instagram, che possono essere considerate "produttrici" di VGI. Queste piattaforme estremamente popolari consentono agli utenti di pubblicare contenuti, messaggi brevi, immagini e video (Salas-Olmedo et al., 2018) e sono utili per pianificare e gestire i percorsi turistici, per ottenere la cronologia dei viaggi degli utenti e per identificare la presenza turistica in una città (Memon et al., 2015). Inoltre, consentono all'utente di visualizzare e caricare immagini attraverso siti e applicazioni nonché di rendere pubblica la propria esperienza di viaggio ad altre persone, che possono visualizzare le foto, condividerle o scaricarle. Queste informazioni sono importanti per identificare i luoghi più visitati, ad esempio i punti di interesse (POI) della città e per suggerire come gestire i flussi turistici. Con il termine POI viene indicato un punto specifico che può essere utile o interessante per i visitatori (Yuan et al., 2013). Esso viene spesso mostrato su una mappa per determinare un'attrazione quale un luogo storico, un edificio, un monumento, un museo, o qualsiasi altro elemento attrattivo (Guo et al., 2018). I dati digitali, prodotti dagli utenti, forniscono prove di mobilità e spostamento supportando i ricercatori nello sviluppo di nuovi metodi di osservazione e analisi di un luogo che riguardano, sia l'individuazione delle dinamiche turistiche che l'identificazione delle preferenze e del comportamento dell'utente. Infatti, i dati, oltre ad essere collegati ad una specifica area geografica, consentono di estrarre altre informazioni legate all'utente e alla sua sfera emotiva. Un indicatore importante per comprendere il comportamento del turista è attraverso il modo in cui si muove all'interno di una destinazione (Md Khairi et al., 2019). Diversi aspetti della mobilità turistica sono stati studiati nel corso degli anni utilizzando le informazioni spazio-temporali fornite dai social media quali YouTube, Flickr, Twitter e da fonti come società di telefonia mobile quali TIM (Jankowski et al., 2010). Ad esempio, Zheng e colleghi (2010) hanno progettato un sistema di social networking in grado di identificare le persone e generare una cooperazione tra i vari utenti nella scelta dei luoghi da visitare attraverso l'acquisizione

di informazioni relative alla mobilità e alle attività di ciascun utente tramite un sistema chiamato *GeoLife*. Le informazioni fornite sono visualizzate da altri utenti e amici “social” i quali saranno probabilmente influenzati nella scelta del luogo da visitare. Altri dati digitali come le immagini scattate dagli utenti ed il testo dei messaggi condivisi sul web sono stati utili per comprendere la percezione che le persone hanno dell’ambiente circostante, per delineare i punti di interesse della città oppure per determinare le differenze tra le foto scattate da turisti stranieri e le foto scattate dalle persone del luogo (Girardin et al., 2008; Mirkovic et al. 2011). Attraverso l’analisi di immagini geo-referenziate di Flickr, Sun e colleghi (2015) hanno identificato le principali destinazioni di viaggio in una città e hanno raccomandato i migliori itinerari di viaggio tra le famose destinazioni. Il loro lavoro, condotto utilizzando un metodo di raggruppamento spaziale, in cui classificano i punti di riferimento, ha dimostrato come sia possibile, attraverso questo approccio, pianificare in modo ottimale un viaggio. Anche Arain e colleghi (2017) acquisendo immagini geo-taggate dal database di Flickr hanno eseguito alcuni esperimenti utili per presentare un sistema di raccomandazione su possibili destinazioni turistiche. Pfaender e Fen-Chong (2011), invece, attraverso le tracce digitali dei telefoni cellulari, hanno studiato la mobilità dei visitatori stranieri a Parigi e hanno svolto le loro indagini utilizzando vari strumenti e software. Alcuni ricercatori hanno inoltre proposto un metodo previsionale relativo alle location preferite degli utenti in una città utilizzando le informazioni temporali disponibili su Flickr (Clements et al., 2010) mentre altri, associando le informazioni temporali delle immagini ricavate da siti turistici, hanno determinato il tempo di permanenza dei turisti in una specifica area attrattiva (Popescu & Grefenstette 2009). Pertanto, le informazioni fornite tramite le immagini dei social media possono essere utilizzate per individuare i movimenti dei turisti ed offrire loro consigli di viaggio. Oltre a Flickr, altre piattaforme come Panoramio e Twitter sono state utilizzate per monitorare il comportamento dei turisti. Studi condotti utilizzando post geo-tagcati estratti da Twitter hanno permesso di tracciare la posizione dell’utente e sviluppare dei modelli di previsione (Zhao et al., 2016).

Come si evince dalla letteratura scientifica sopra menzionata, i social media hanno acquisito, un’importanza strategica e fondamentale nell’ambiente turistico, facilitando le interazioni tra i vari utenti online e contribuendo a cambiare l’approccio dell’utente ai servizi e ai prodotti turistici (Xiang & Gretzel, 2010). Ad oggi, anche il singolo individuo attraverso i social media ha la possibilità di reperire facilmente informazioni su prodotti,

servizi e destinazioni turistiche; inoltre interagendo con altri utenti può accedere ad altre esperienze e raccomandazioni utili a supporto della scelta decisionale.

2.3 Flickr e Big data: uso dei social media per identificare l'attrattività turistica

Come emerge dalla letteratura, negli ultimi anni il settore turistico ha assunto un ruolo determinante nella valutazione del livello di sviluppo economico di un paese (Cárdenas-García & Pulido-Fernández 2019). Dai risultati delle statistiche presenti nel rapporto sul Turismo internazionale e sulla base di uno studio effettuato dall'Organizzazione Mondiale del Turismo (United Nations World Tourism Organization - UNWTO) si denota come i numeri in questo settore siano in costante aumento a livello mondiale. Tuttavia, si evince, una perdita di competitività dell'Italia rispetto ad altri paesi europei tra cui Francia e Spagna. Attualmente, il sistema turistico imprenditoriale italiano, che ha rappresentato un punto di forza per decenni, risulta essere molto debole nella commercializzazione e nel marketing tale da far rallentare l'attrattività dell'Italia a vantaggio di altri paesi. In generale, web e social network rappresentano ad oggi il nesso per eccellenza in grado di condizionare le scelte dell'utente da diversi punti di vista, incluso quello turistico, relativamente alla scelta della prenotazione sia di viaggi, che di vacanze. Il Politecnico di Milano ha evidenziato come tale capacità negli ultimi anni, rispetto al passato, sia quadruplicata contribuendo in modo sensibile alla nascita e alla diffusione di una nuova figura, ossia quella del turista-consumatore che sceglie e crea una richiesta precisa partendo dai suoi interessi e desideri, mentre diminuisce il numero di turisti abbastanza standardizzati e in un certo modo "adattabili" che acquistano prodotti preconfezionati. Eppure, nonostante i trend evidenziati dalle ricerche statistiche, da un'analisi generale sembra che il sistema turistico italiano non abbia sviluppato forti politiche commerciali e azioni sinergiche, da consentire un progresso dal punto di vista della commercializzazione e del web-marketing.

2.3.1 Obiettivi della ricerca

Partendo da queste premesse, l'Italia è stata selezionata come oggetto di studio. In

particolare la scelta è ricaduta su sei città identificate come le destinazioni turistiche con maggiore attività di viaggio svolta dai visitatori. Sono state selezionate le città con maggiore impatto culturale secondo la classifica ufficiale pubblicata nel 2016 dall'Istituto Nazionale di Statistica (ISTAT). Le sei città scelte sono ricche di edifici storici, chiese, castelli, musei e sono le destinazioni turistiche ideali. Tutte sono città d'arte, conservano le tracce del passato e sono contrassegnate dall'attività di grandi artisti e delle loro opere d'arte. La ricerca condotta ha fornito la possibilità di utilizzare l'apprendimento automatico (Machine Learning) per estrarre informazioni sui dati estratti dai social media (Flickr), nello specifico le immagini scattate da turisti stranieri e turisti locali. Tramite l'analisi dei dati ottenuti da Flickr è stata definita la tendenza annuale dell'attività fotografica in ogni città ed identificata la provenienza degli utenti (reperibile sul profilo pubblico di Flickr di ogni singolo utente) per determinare quali luoghi sono da considerarsi attrattivi per i turisti italiani e quali luoghi sono di interesse per i turisti stranieri. Di rilievo per la comprensione delle tendenze e preferenze dei turisti è stata l'analisi effettuata sulle immagini che ha permesso di conoscere quali elementi il turista preferisce fotografare. L'analisi delle coordinate geografiche ha permesso, invece, di indagare quali sono i POI nelle città che generano benefici per il territorio sottolineando l'importanza della condivisione di foto online per studiare il comportamento turistico.

2.3.2 Procedura e metodologia di analisi

Sono state prese in esame 26.392 immagini geo-referenziate condivise da 4.205 utenti da gennaio 2014 a dicembre 2016 su Flickr e relative a sei città d'arte italiane: Milano, Venezia, Firenze, Roma, Napoli e Palermo, città contrassegnate dall'attività di grandi artisti e con il maggior numero di turisti. I dati sono stati collezionati da Flickr, una piattaforma online in cui gli utenti condividono fotografie personali, con un totale di 87 milioni di membri registrati e oltre 3,5 milioni di nuove immagini caricate quotidianamente (www.flickr.com). Questa piattaforma è utile per catturare eventi di vario genere con alta precisione, ad esempio eventi atmosferici oppure eventi riguardanti la politica, presentandosi come un database altamente utile ed affidabile per analizzare le immagini scattate in tutto il mondo (Preis et al., 2013; Stylianou-Lambert, 2012). I dati sono stati immagazzinati utilizzando pubbliche Application Programming Interface (API) di Flickr (<http://www.Flickr.com/services/api/>), un insieme di procedure che consentono

l'espletamento di operazioni di modellazione per una varietà di applicazioni informatiche (Banta et al., 2006); spesso con tale termine vengono indicate le librerie software di un linguaggio di programmazione. In particolare, la raccolta è avvenuta tramite l'implementazione di una piattaforma che ha permesso, attraverso l'esecuzione di una *query*, di ottenere i dati. Nello specifico la piattaforma ha permesso, in un primo momento, la connessione al sito Web di Flickr e di seguito la possibilità di acquisire i dati delle città selezionando anche il periodo (giorno, mese ed anno) (Figura 2.1).

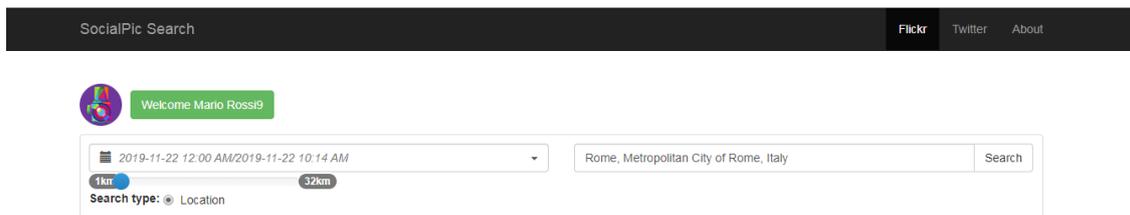


Figura 2.1. Interfaccia input della Piattaforma

L'output ricavato dalla piattaforma fornisce immagini e metadati descrittivi, i quali costituiscono un corredo di informazioni necessarie per descrivere l'oggetto, in questo caso l'immagine, in maniera accurata e precisa, ovvero:

- i. *Posizione,*
- ii. *ID Posizione,*
- iii. *Latitudine e Longitudine,*
- iv. *Nome utente,*
- v. *ID Utente,*
- vi. *Data e Ora*
- vii. *Link dell'immagine*

Per ogni città campione è stato condotto lo stesso processo e la ricerca è stata effettuata su un raggio di 5 km dal centro della città.

SocialPic Search Flickr Twitter About

Results:
25

[Export To CSV](#) [Save Images](#)

Location	Location id	Latitude	Longitude	Username	User id	Datetime	Link	Image
Colonna - Roma RM, Italia		41.901047	12.483247	Marcial Bernabeu	128784882@N04	Fri, 22 Nov 2019 09:33:59 +0100	https://farm66.staticflickr.com/65535/49103864997_f8ff15cdc7.jpg	
Centro Storico - Roma RM, Italia		41.898779	12.482206	raffaele pagani	94165526@N04	Fri, 22 Nov 2019 09:06:06 +0100	https://farm66.staticflickr.com/65535/48998516228_f180a9657d.jpg	
Pigna - Roma RM, Italia		41.897079	12.478580	john etvidge	132318869@N08	Thu, 21 Nov 2019 17:57:46 +0100	https://farm66.staticflickr.com/65535/49100170948_bb5dc908b7.jpg	
Roma, Subura (Ancient Urabn Quarter) -		41.894604	12.493321	Eric.Parker	13484951@N00	Thu, 21 Nov 2019 14:15:30 +0100	https://farm66.staticflickr.com/65535/49099776246_99ed42a301.jpg	

Figura 2.2. Interfaccia output Piattaforma

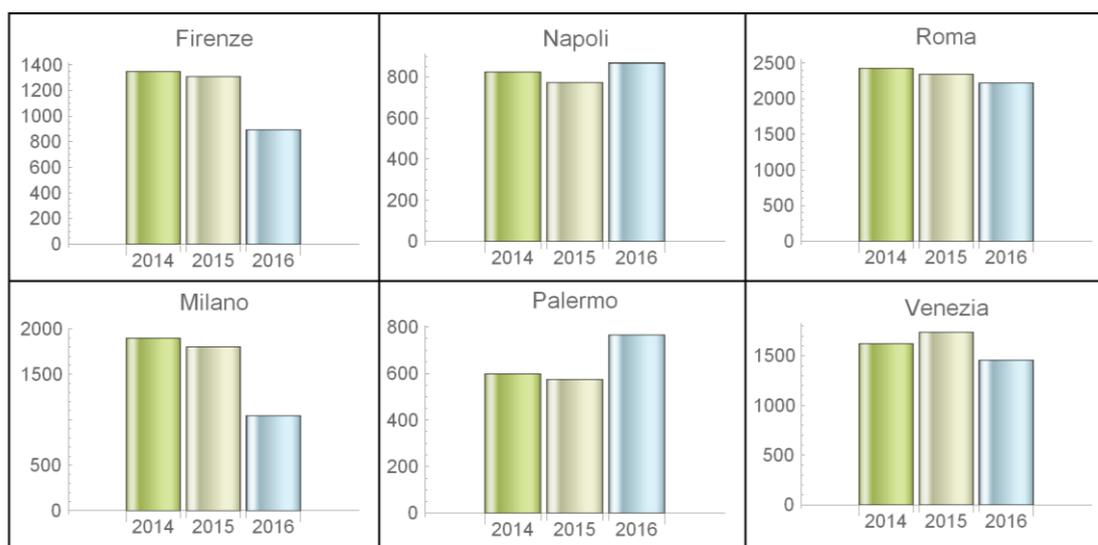
I metadati sopra elencati sono stati utilizzati nella ricerca per condurre i diversi tipi di analisi che saranno presentate di seguito.

I dati raccolti sono stati analizzati utilizzando, *Wolfram Mathematica*, un software ad alte prestazioni di cui si è parlato nel capitolo precedente. In particolare, sono state eseguite diversi tipi di analisi ed applicati diversi metodi, tra cui specifici algoritmi di Machine Learning, che hanno permesso di:

- i. determinare la città ed il paese di origine degli utenti;
- ii. comprendere quanti turisti stranieri e quanti turisti locali arrivano in una zona;
- iii. identificare i luoghi visitati;
- iv. mostrare luoghi e POI attraverso l'analisi spazio-temporale.

Dopo aver immagazzinato i dati relativi alle sei città è stata condotta una prima analisi descrittiva per determinare il numero di foto scattate da turisti. La Tabella 1 riassume il numero di foto scattate in ciascuna città dal 2014 al 2016.

Tabella 1. Numero di fotografie scattate dagli utenti in ciascuna città dal 2014 al 2016



Il set di dati raccolto include l'ID utente, utile per ricavare alcune informazioni quali la città e il paese di origine degli utenti. Tramite il software *Mathematica*, è stato possibile creare una connessione all'API di Flickr attraverso una query e adottando il modulo *ServiceConnect* ["Flickr"] che ha permesso di rintracciare il luogo indicato dagli utenti come luogo di origine.

Per raggiungere questo risultato sono stati richiamati gli ID degli utenti (UserID) in un array ed è stata estratta la "Location" tramite un ciclo "for" (Figura 2.3). In particolare, ponendo la condizione $i \leq 10$, finché i è minore o uguale a 10, la variabile i viene incrementa di 1. Inoltre, tenendo conto delle eventuali problematiche dovute ad un *overflow* del sistema e quindi per evitare l'interruzione della connessione con il server di Flickr, è stato deciso di ottimizzare il lavoro estraendo i dati in più sessioni.

L'obiettivo è stato quello di ricavare la provenienza degli utenti per identificare i luoghi maggiormente attrattivi per i turisti italiani e i luoghi di interesse per i turisti internazionali oltre che individuare quanti turisti stranieri e quanti turisti locali arrivano in una zona.

```
ax = For[i = 2, i ≤ 10, i++,
n = ServiceExecute["Flickr", "UserData", "User" → {"UserID", cc[[i]]}];
Print[n["Location"]]]
```

Figura 2.3. Codice sviluppato in Mathematica per ricavare la Location

Dopo aver ricavato il paese di origine, gli utenti sono stati raggruppati per nazionalità ed etichettati in base alla provenienza. Ad esempio un utente di Madrid ed un utente di Barcellona sono stati entrambi riuniti sotto l'etichetta "Spagna". Di seguito il codice realizzato in *Mathematica* per classificare gli utenti.

```
Table[Union[Table[Flatten[lc2[[m]]][[1+2(n-2)]], {n, Length[lc2[[1]]}], {m, 1}],  
{1, Australia;6, Austria;1, Belarus;1, Belgium;4, Brazil;5, Bulgaria;1,  
Canada;6, China;1, Colombia;1, Congo;1, Croatia;1, Denmark;1, France;5,  
Germany;9, Greece;2, Hungary;2, Ireland;2, Israele;1, Italia;97,
```

Figura 2.4. Codice sviluppato in Mathematica per classificare gli utenti

Dopo aver calcolato il numero di immagini scattate dagli utenti per ogni città, determinato il luogo di provenienza degli utenti ed il numero di turisti italiani e stranieri, l'attenzione è stata rivolta verso i luoghi che essi preferiscono fotografare.

Infatti, ai fini della ricerca, che ha come scopo anche la comprensione del comportamento del turista, di fondamentale importanza è stato comprendere cosa spinge le persone a fotografare un oggetto in una specifica zona. Al software è stato fornito l'ID relativo alla posizione (Location ID -metadato integrato nel set di dati iniziale) di ogni immagine ed avviata un'analisi descrittiva al fine di ricavare i luoghi con maggiore attività fotografica.

```
luoghi =  
Table[Union[Table[Flatten[lcx[[m]]][[1+2(n-1)]], {n, Length[lcx[[1]]}], {m, 6}],  
{(Firenze, Giardino di Boboli - Firenze, San Niccolo - Firenze,  
Santo Spirito - Firenze), (Boffalora - Milano, Milano, Moncucco - Milano,  
Oggiono - Milano, Segnanino - Milano, Vigentino - Milano),  
(Napoli, Porto di Santa Lucia - Napoli, Rione I.n.c.i.s. - Napoli,
```

Figura 2.5. Codice sviluppato in Mathematica per ricavare i luoghi con maggiore attività fotografica

Al fine di identificare cosa gli utenti preferissero fotografare in un determinato luogo è stata eseguita l'analisi delle immagini, ricavate dalla piattaforma Flickr, attraverso l'applicazione di un algoritmo di Machine Learning, basato sul riconoscimento degli

oggetti presenti nelle immagini. Nello specifico, nel software sono state importate le immagini raccolte e tramite la funzione *ImageIdentify* è stato possibile identificare l'elemento principale in esse rappresentato. La funzione applicata è una Machine Learning ed il metodo di identificazione si riferisce alle reti neurali artificiali. La macchina è stata addestrata usando milioni di immagini fornite come esempi da cui ha imparato progressivamente a fare distinzioni. Questa funzione opera in modo probabilistico, generando molteplici risultati con diverse probabilità stimate. In particolare, *Wolfram ImageIdentify* identifica l'oggetto principale in un'immagine. Il training set interno è composto da oltre tre milioni di immagini di training e oltre 4.000 classi di oggetti (non disponibili pubblicamente). In particolare, la *Neural Net Repository* è costruita per essere una risorsa globale per i modelli di reti neurale ed ospita una raccolta, in espansione, di modelli di reti neurali addestrati e non addestrati, adatti per valutazione immediata, formazione, visualizzazione e apprendimento. Di seguito un esempio di come opera la rete e della probabilità di riconoscimento di un oggetto.

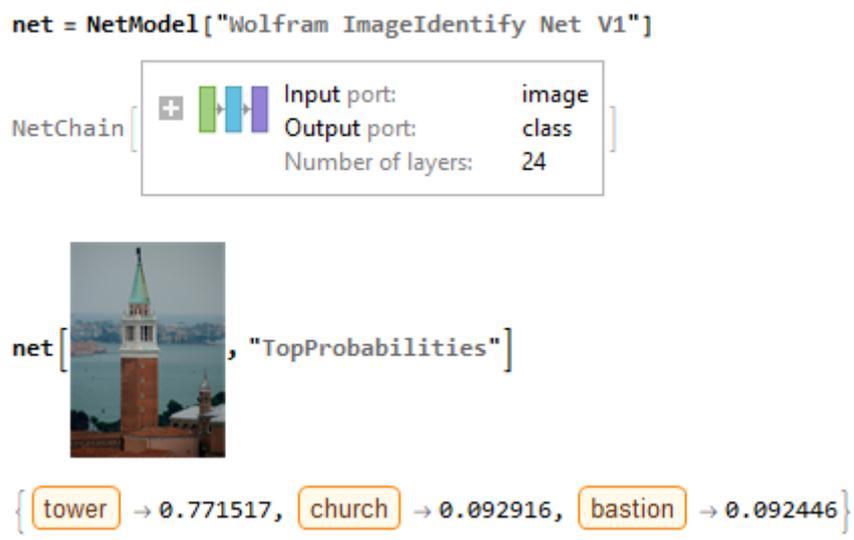


Figura 2.6. Wolfram Neural Net Repository e probabilità di riconoscimento dell'oggetto rappresentato nell'immagine

Il riconoscimento avviene attraverso (i) la rappresentazione: l'algoritmo di apprendimento crea un modello che produce i risultati desiderati dai dati in input; (ii) la valutazione, l'algoritmo di apprendimento può creare più di un modello, un'ulteriore funzione di valutazione determina quale dei modelli funziona meglio per fornire il

risultato desiderato da specifici dati in input (la funzione di valutazione ha il compito di valutare i modelli creati, poiché più di un modello potrebbe fornire risultati utili per risolvere un problema specifico); (iii) l'ottimizzazione, il processo di apprendimento produce una serie di modelli, che di solito possono generare il giusto risultato per un determinato set di input: attraverso questi modelli il processo di apprendimento cerca di determinare quale funziona meglio; infine viene selezionato il modello migliore.

Il processo di riconoscimento delle immagini è stato ripetuto automaticamente per tutte le immagini raccolte per ogni città. Nell'analisi le immagini sono state classificate in categorie in base al contenuto.

Altri metadati, ovvero le coordinate geografiche correlate alle immagini raccolte, sono state impiegate per identificare la distribuzione spaziale degli utenti, per studiare le attrattività di ogni città e per indagare sul cambiamento o la staticità dell'attività fotografica nel corso degli anni. Per analizzare i dati geo-localizzati è stata applicata una funzione di *clustering*. Come già discusso nel primo capitolo, quando si decompone un insieme di oggetti in cluster, l'obiettivo principale è trovare gruppi in cui gli oggetti mostrano molte somiglianze tra loro. L'analisi dei cluster viene utilizzata come metodo non supervisionato per cercare di riconoscere la struttura sottostante nei dati, per la classificazione, il riconoscimento dei modelli e per la riduzione e l'ottimizzazione. Esistono molti metodi che cercano di raggiungere questo obiettivo. Nella presente ricerca, è stata applicata la funzione *FindClusters* - $f(e)$ implementata in *Wolfram Mathematica*. Questa funzione utilizza l'analisi dei cluster come metodo per partizionare oggetti funzionanti per una varietà di tipi di dati, inclusi dati numerici, testuali ed immagini. Essa classifica elementi simili applicando l'apprendimento automatico senza supervisione, attraverso il quale il software confronta i dati senza ricevere alcuna informazione a riguardo. Dato un gruppo di elementi e , la funzione $f(e)$ lo suddivide in gruppi di elementi simili:

$$f(e) = [\{e1, e2, \dots\}]$$

Per impostazione predefinita, *FindClusters* pre-elabora i dati automaticamente a meno che non vi sia un valore di distanza da assumere tra due punti specificati. Nell'analisi corrente, il software, in base alla qualità dei dati di input, ha applicato automaticamente un algoritmo considerato migliore in base ai modelli implementati disponibili.

Tramite la funzione *ClassifierInformation* è stato possibile ricavare informazioni in

merito all'algorithmo scelto dalla macchina (Figura 2.7).

Classifier information	
Input type	NumericalVectorSequence
Classes	1, 2, 3, 4, 5
Method	DBSCAN
Single evaluation time	7.35 ms/example
Batch evaluation speed	18.7 examples/ms
Classifier memory	55.7 kB
Training examples used	22 examples
Training time	223. ms

Figura 2.7. Informazioni sul metodo applicato per il raggruppamento dei dati

Per raggruppare i dati (latitudine e longitudine) in cluster il software ha applicato il metodo DBSCAN un popolare metodo di apprendimento non supervisionato che identifica e separa i cluster ad alta densità dai cluster a bassa densità. Nello specifico, per ogni punto nel set di dati, determina una forma n dimensionale attorno a quel punto, quindi conta il numero di punti che rientrano in quella forma. Successivamente, conta questa forma come un cluster ed espande in modo iterativo il cluster, esaminando ogni singolo punto all'interno del cluster e contando il numero di altri punti dati nelle vicinanze. Per quanto riguarda la distanza tra punti di interesse (POI) e utenti, essa viene calcolata in base alla metrica euclidea.

La funzione applicata restituisce il numero di cluster rilevati che varia ampiamente di anno in anno per le sei città italiane.

2.3.3 Risultati

I risultati della prima analisi, come si evince dalla Tabella 1, mostrano che la città più fotografata è Roma con 6.988 fotografie, seguita da Milano con 5.596 foto. Nello specifico, il numero di fotografie scattate in ciascuna città differisce nel corso dei tre anni. Alcune città hanno visto un aumento dell'attività fotografica in termini di attività turistica (Napoli 11% e Palermo 25%), mentre è stata registrata una diminuzione dell'attività fotografica per le città di Roma, Venezia, Firenze e Milano.

Per quanto concerne l'analisi degli utenti e del luogo di provenienza, i risultati dimostrano come nel 75% dei casi la città natale degli utenti sia identificabile, mentre per il resto degli utenti (25%) il luogo è stato definito come non disponibile. In effetti, alcuni utenti hanno indicato come città di origine una città falsa o un luogo di origine immaginario, ad esempio: “*Fantasy*”, “*On a quest to discover the world*”; tali dati sono stati assemblati come non disponibili. I risultati rivelano che il numero dei visitatori italiani è maggiore dei visitatori stranieri in tutte le città. A tal proposito, al fine di rappresentare al meglio i dati ed ottenere da essi maggiore significatività, soprattutto per quanto riguarda la visualizzazione, gli *outliers* (picchi) sono stati esclusi. Dai risultati si evince che un numero significativo è rappresentato dai visitatori provenienti dal Regno Unito, dagli Stati Uniti e dalla Germania, in particolare in città come Roma e Venezia. Di seguito un grafico rappresentativo del numero di visitatori nella città di Palermo nel corso dei tre anni (2014/2015/2016) con l'indicazione del paese di provenienza.

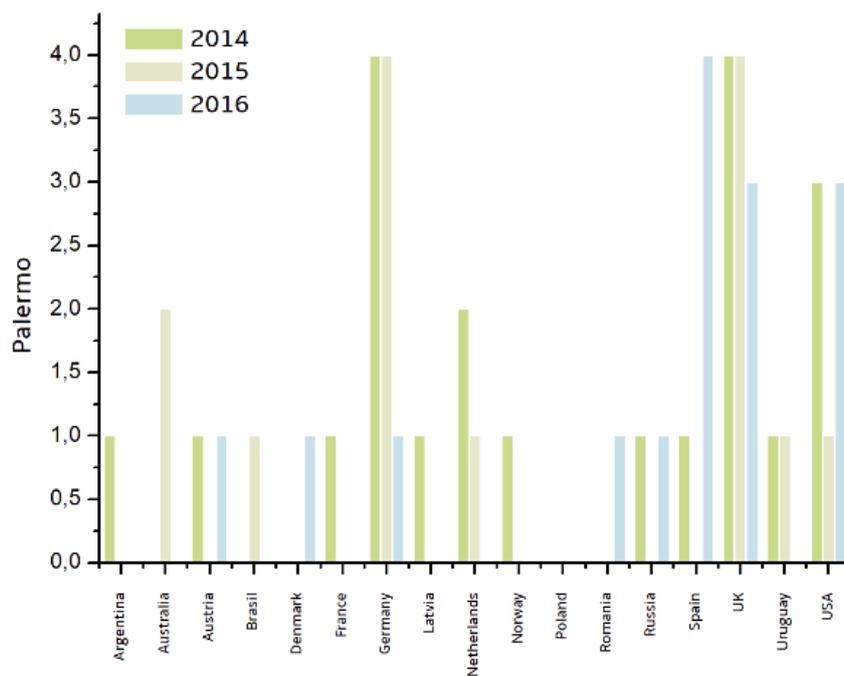


Figura 2.8. Numero di visitatori nella città di Palermo nel corso dei tre anni e paese di provenienza

Invece l'analisi condotta sui dati relativi alla *Location* ha fornito informazioni sui luoghi visitati dagli utenti e ha permesso di identificare le principali attrazioni turistiche in cui si concentra maggiore attività fotografica.

Di seguito la tabella indicante il nome dei luoghi maggiormente fotografati dagli utenti

nelle sei città italiane nel corso dei tre anni.

Tabella 2. Luoghi visitati dagli utenti

CITTÀ	LUOGHI MAGGIORMENTE FOTOGRAFATI
Firenze	Giardino di Boboli
	San Niccolò
	Santo Spirito
Milano	Galleria Vittorio Emanuele
	Moncucco
	Oggiaro
	Vigentino
Napoli	Porto Santa Lucia
	Rione I.N.C.I.S.
	S. Ferdinando
	Vasto
Palermo	Vomero
	Fiera del Mediterraneo
	Guadagna
	Palazzo dei Normanni
Roma	Perpignano
	Campo Marzio
	Flaminio
	Gianicolo
	Rione di Trastevere
Venezia	San Paolo
	Castello (sestiere di Venezia)
	San Marco
	San Polo

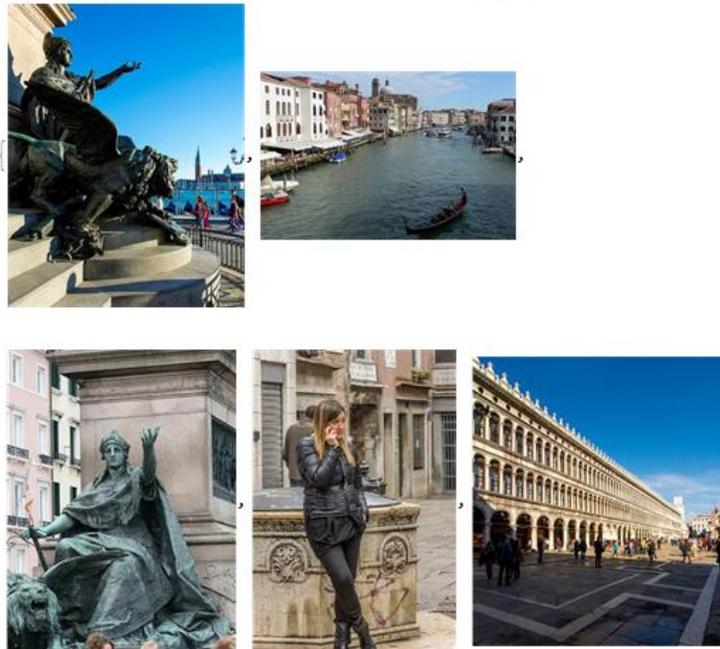
La tabella evidenzia come per ogni città la più alta concentrazione si trovi in luoghi storici come: San Niccolò, Santo Spirito e Giardino di Boboli a Firenze. Quest'ultimo luogo

rappresenta uno storico parco cittadino considerato un museo all'aperto; Galleria Vittorio Emanuele a Milano; Porto di Santa Lucia a Napoli; Campo Marzo, celebre quartiere della capitale dove si possono trovare molti edifici storici, a Roma; Palazzo dei Normanni a Palermo, Ponte dell'Accademia e Piazza San Marco, un centro sociale, religioso e politico, una delle piazze italiane più importanti, conosciuta in tutto il mondo per la sua bellezza architettonica a Venezia. Nella città di Napoli, uno dei luoghi dove vi è maggiore attività fotografica è Vasto, un quartiere in cui vi è un gran numero di chiese erette nei primi anni del '900. Altri luoghi sono invece connessi alle aree commerciali, zone in cui le persone trascorrono molto tempo. A Palermo, ad esempio, il secondo luogo più fotografato è Guadagna, dove vi è un centro commerciale.

Lo scenario mostrato sottolinea che le persone fotografano e amano trascorrere il loro tempo in luoghi in cui vi sono attrazioni famose ma anche dove possono incontrare altre persone ed interagire con esse.

Dopo aver analizzato i luoghi sono state prese in esame le immagini.

```
yy = Table[Import[xx[[a]], "Image"], {a, Length[xx]}]
```



```
yy2 = ImageIdentify[yy]
```

```
{statue, canal, statue, person, architecture}
```

Figura 2.9. Applicazione della funzione ImageIdentify che permette il riconoscimento degli elementi di alcune immagini scattate nella città di Venezia ed estratte da Flickr.

Capitolo 2

I risultati dell'analisi relativa al riconoscimento dell'oggetto nell'immagine, eseguita con *ImageIdentify*, mostrano gli elementi più fotografati in ogni città. Tutte le immagini sono state classificate in categorie in base al contenuto. Gli elementi più fotografati sono le persone, seguite dalle strutture inclusi gli edifici e beni culturali (torri e archi). Di seguito delle tabelle rappresentative degli output ottenuti dal software applicato. La figura mostra i dieci elementi più fotografati nel corso dei tre anni in tutte le sei città campione.

palermo			
person	867	person	223
house	43	house	34
tower	42	arch	34
arch	40	tower	27
church	38	road	27
road	34	building	20
building	30	store	19
ballroom	26	coast	19
store	21	mount	13
altarpiece	21	correctional institution	13
		correctional institution	8
		correctional institution	8

firenze			
person	256	person	867
tower	90	house	43
church	76	tower	42
arch	63	arch	40
house	54	church	38
building	52	road	34
bridge	41	building	30
altarpiece	34	ballroom	26
road	28	store	21
hotel	23	altarpiece	21
		correctional institution	13
		correctional institution	13

roma			
person	783	person	265
tower	56	arch	29
arch	50	house	27
house	42	road	24
building	33	altarpiece	22
ballroom	32	coast	15
church	30	store	12
road	25	hotel	12
store	21	building	12
Tesla coil	17	tower	9
		woody plant	10
		coniferous tree	8
		coniferous tree	8

milano			
person	277	person	783
arch	70	tower	56
tower	67	arch	50
building	60	house	42
house	57	building	33
church	48	ballroom	32
altarpiece	39	church	30
bridge	33	road	25
road	30	store	21
statue	19	Tesla coil	17
		altarpiece	22
		coast	15
		hotel	12
		building	12
		tower	9

venezia			
person	805	person	188
tower	62	house	45
house	57	arch	41
arch	50	altarpiece	27
building	47	road	25
church	43	tower	20
store	25	burial chamber	19
ballroom	23	mount	18
factory	18	coast	16
road	16	church	16
		burial chamber	27
		entablature	18
		church	17
		altarpiece	13
		hotel	12

napoli			
person	355	person	805
building	99	tower	62
arch	95	house	57
tower	94	arch	50
house	93	building	47
altarpiece	87	church	43
church	73	store	25
bridge	48	ballroom	23
road	42	factory	18
hotel	38	road	16
		church	16
		burial chamber	19
		mount	18
		coast	16
		church	16

Figura 2.10. I dieci elementi più fotografati nel corso dei tre anni in tutte le sei città

Per quanto riguarda l'analisi dei cluster i risultati sottolineano la varianza dei dati e di conseguenza dell'attività fotografica nelle città nel corso dei tre anni (Figura 2.11).

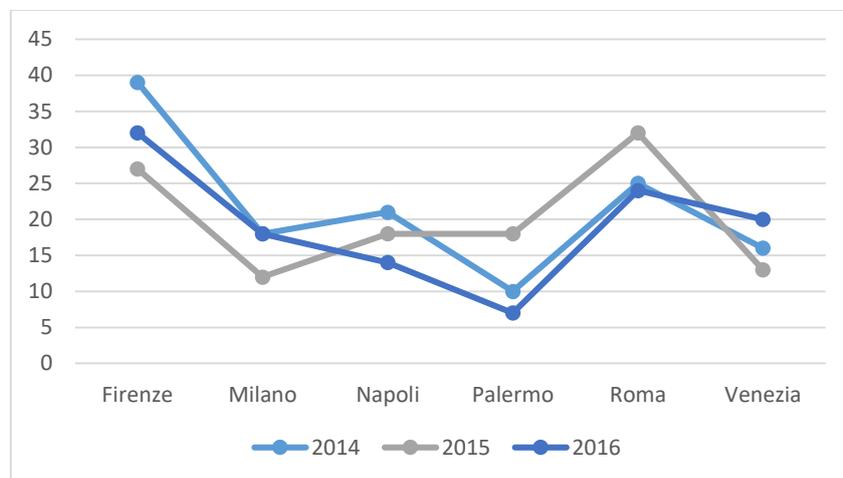


Figura 2.11. Tendenza annuale dei cluster dal 2014 al 2016 per le sei città italiane

La figura mostra la tendenza annuale dei cluster e illustra la variazione del numero di cluster in ciascuna città nel corso di tre anni. Per alcune città il software, attraverso l'applicazione della funzione *FindClusters*, ha identificato una diminuzione del numero di cluster (Firenze, Napoli, Palermo e Roma) nel 2016, mentre in altre città ha determinato un aumento (Milano e Venezia). La varianza del numero di cluster per città fornisce una misura dell'attività fotografica turistica, definendo l'indice di popolarità delle città su Flickr.

Di seguito si riportano le mappe dei cluster per ogni città analizzata e per ogni anno. Nella figura i cluster mappati mostrano la densità dei dati all'interno delle città. Ogni cluster è costituito da unità ed è rappresentato da colori diversi. I punti dello stesso colore sono assegnati allo stesso cluster.

In definitiva si può affermare che i cluster rappresentano l'attenzione che i turisti hanno prestato alle attrattività di un luogo.

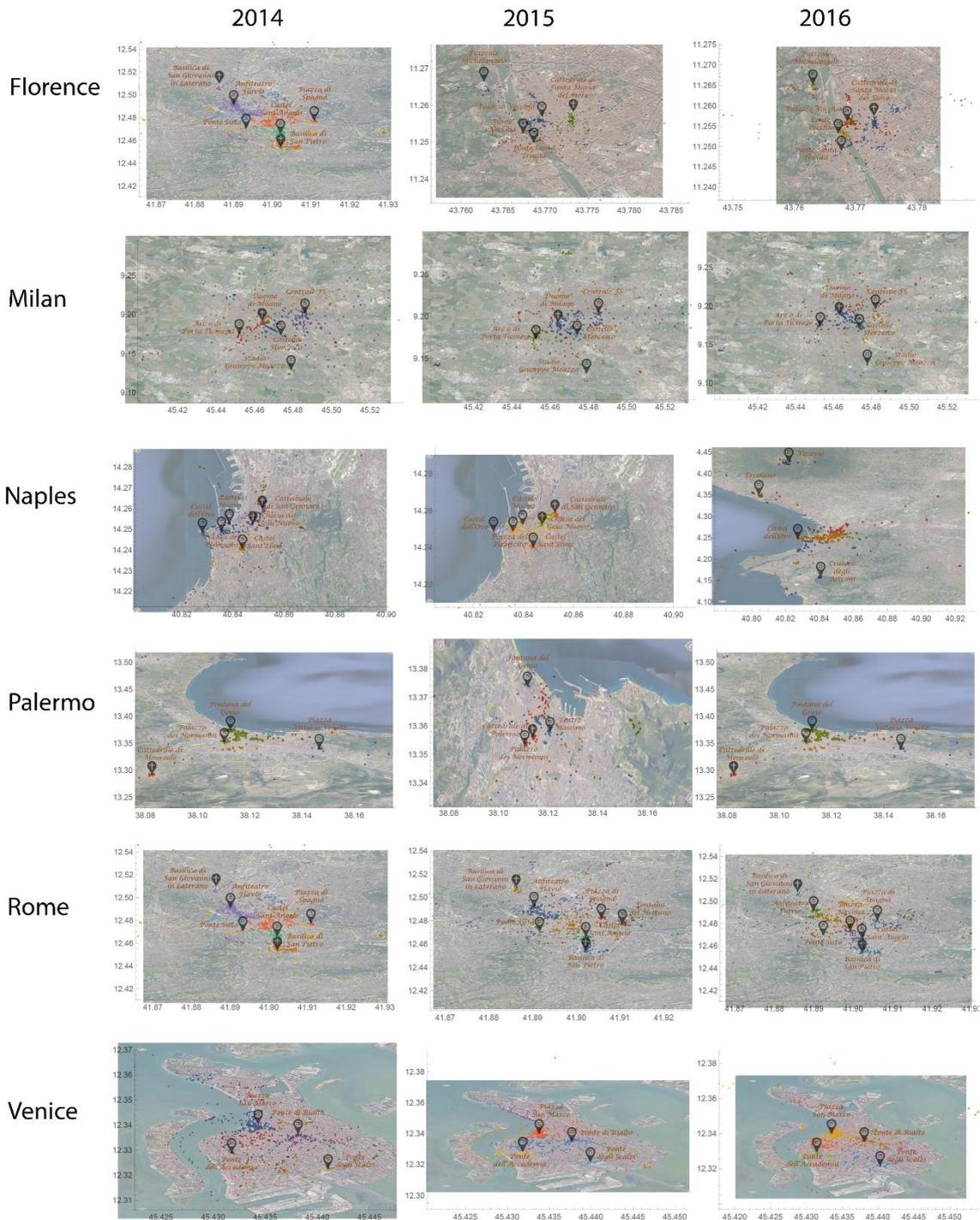


Figura 2.12. Distribuzione dei cluster nelle sei città italiane e confronto geo-referenziato con mappe delle città

Un accurato codice sviluppato nel *notebook* di *Mathematica* ha permesso di tracciare i cluster su mappe insieme ai POI turistico-culturali in diverse aree delle sei città italiane. Le informazioni geografiche e territoriali sul set di dati relativo ai punti di interesse sono state estratte da OSM (<http://www.Openstreetmap.org/>).

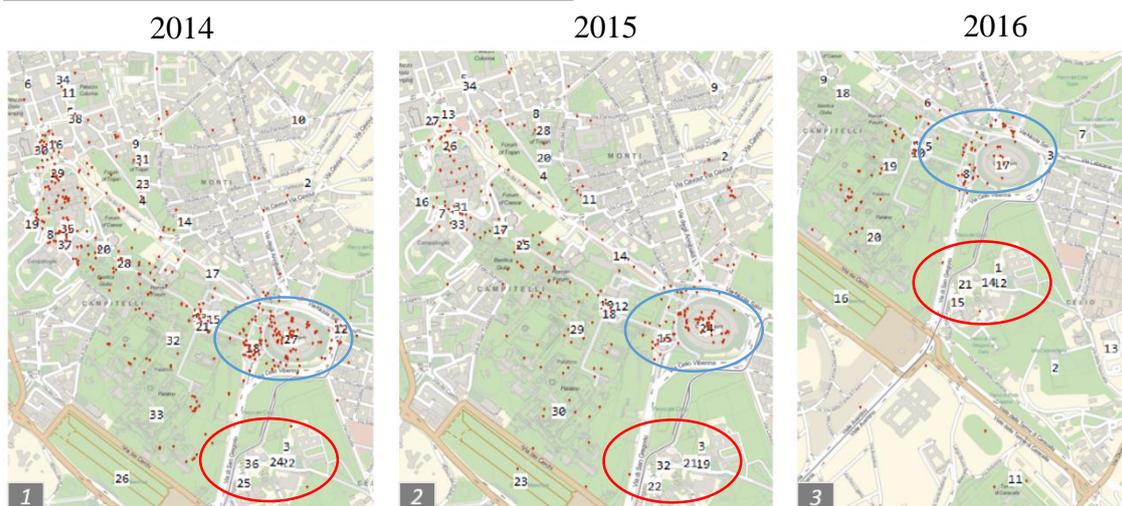


Figura 2.14. Punti di interesse e cluster di un'area di Roma dal 2014 al 2016

La figura mostra come l'area più attiva, in termini di attività fotografica, sia vicino al Colosseo (27a-24b-17c) e all'Arco di Tito (18a-15b-8c) area archeologica, mentre le aree meno attrattive risultano nei pressi della *Domus Augustana* (36a-32b-21c) e del *Circo Massimo* (24a-21b-14c). In questi punti non si registra attività fotografica nel corso dei tre anni. Il colore del cerchio mostra il rapporto tra gli *hotspot* – punti caldi (azzurro) e luoghi meno visitati (rossi). Sebbene vi sia una diminuzione del numero di utenti nella città di Roma, come riportato dalle statistiche mostrate nella Tabella 1, e dunque un cambiamento evidente dell'attività fotografica, la situazione attorno ai punti di interesse rimane costante nel corso degli anni.

Il set di dati relativi alle altre cinque città sono stati raggruppati allo stesso modo.

Un risultato significativo è che viene mostrata la distribuzione degli utenti attorno ai POI sia in aree famose, cosiddette *hotspot* che in aree impopolari, ricche comunque di POI ma con una bassa densità. Questo risultato riflette l'ipotesi che la distribuzione del turismo all'interno delle città sia irregolare.

2.3.4 Discussioni

Diversi ricercatori hanno dimostrato che ogni aspetto dell'esperienza turistica è pieno di meccanismi psicologici (Larsen, 2007). Durante un viaggio, gli utenti vanno alla ricerca di sentimenti ed emozioni, per apprezzare e provare nuove esperienze, creano relazioni sociali, le quali rappresentano il prodotto dell'identità e dell'atteggiamento del turista immerso in una società. Riprendendo un concetto espresso da Pearce (1982), un turista si

trasforma in un ambasciatore “naturale” di esperienze vissute dall’uomo che possono essere amplificate e rese note attraverso i social media. Come emerge dalla letteratura, è importante condurre studi sui modelli comportamentali turistici negli ambienti urbani (Miah et al., 2017) e l’uso di strumenti innovativi, utili nella comprensione del comportamento dei turisti in una determinata città, rappresenta un vantaggio per la ricerca in questo campo. Jansen-Verbeke (1986), nella sua ricerca parla di elementi costitutivi di una città classificandoli in elementi primari, ovvero luogo di attività e tempo libero; secondari, ossia strutture e servizi forniti per supportare i turisti; ed elementi aggiuntivi, affermando come questi tre elementi siano presenti soprattutto nella maggior parte delle destinazioni culturali. Esaminare la destinazione culturale urbana è importante per spiegare il consumo turistico in termini di spazio, tempo e denaro e la comprensione dell’atteggiamento rafforza la conoscenza del turista inteso come un consumatore. Sulla base di queste affermazioni, lo studio sul comportamento degli utenti acquisisce, nella ricerca contemporanea, una posizione privilegiata in quanto utile nel proporre adeguate strategie di marketing al fine di promuovere una località. Pertanto, gli stakeholder potrebbero sfruttare quest’analisi e le analisi effettuate sui dati estratti dai social media per conoscere lo stato attuale di una particolare destinazione turistica ed attuare politiche per migliorarla a livello di efficienza e di sviluppo. Le piattaforme di condivisione dei dati rappresentano sistemi socio-culturali ed economici che si adattano ai modelli di business adeguandosi di conseguenza alla diversità tecnologica e alle differenti tipologie di utenti (Scott & Orlikowski, 2012). Attraverso appropriati modelli di business, l’industria del turismo può influenzare le scelte dei potenziali turisti durante la ricerca di informazioni su un viaggio e l’analisi dei dati prodotti può rappresentare una componente significativa nelle interazioni tra utenti e fornitori di servizi. Dunque, l’uso delle tecnologie e di Internet può consentire alle destinazioni turistiche di migliorare la propria competitività. Alcuni turisti durante i loro viaggi, vogliono rilassarsi, mentre altri vogliono divertirsi e godere di un’esperienza senza precedenti. Immagini significative e attrattive, linguaggio persuasivo e contenuti affascinanti, possono aiutare ad attirare l’attenzione degli utenti ed ottenere visitatori.

Un contributo importante fornito dalla presente ricerca è che attraverso questa analisi è possibile identificare le dinamiche associate ad un sito turistico utilizzando i dati estratti dai social media applicando analisi basate su tecniche di apprendimento automatico, di valida natura (si veda percentuale di probabilità dei risultati).

Inoltre, questo studio non solo promuove, l'approccio all'apprendimento automatico come supporto nello studio sul comportamento dei turisti/consumatori, ma estende gli attuali approcci metodologici per comprendere importanti caratteristiche basate sui luoghi e sulle destinazioni turistiche e fornisce informazioni su un potente software (*Wolfram Mathematica*) il cui ambiente di sviluppo può essere utilizzato con o senza il potenziamento di strumenti e competenze specifiche.

Inoltre, lo studio presentato evidenzia come il turismo sia collegato alla cultura e suggerisce di porre maggiormente l'accento sulla promozione delle attrazioni culturali in Italia. Attualmente, la domanda di cultura è alta ed ha un peso considerevole per l'Italia, che rappresenta una destinazione scelta per ragioni culturali nel 53% dei casi. Infatti, musei, monumenti e siti archeologici hanno un numero maggiore di visitatori. Questo atteggiamento consente alla cultura di essere a disposizione di tutti, generare turismo e di conseguenza prolungare i soggiorni. Nella ricerca condotta, non sono state analizzate le cause, ma è stato verificato, attraverso l'analisi dei dati estratti dai social, quali sono i siti turistici in cui vi è una maggiore concentrazione di utenti e quali sono le aree meno visitate utilizzando tecniche innovative come l'apprendimento automatico. Come già evidenziato, la ricerca, proponendo sei *case studies*, non prende in esame tutte le foto scattate da tutti i turisti che nel corso dei tre anni hanno visitato le sei città italiane, ma analizza le immagini che turisti/fotografi hanno scattato e successivamente condiviso sulla piattaforma Flickr. Tuttavia, lo studio mostra come il turista scattando le varie foto, le quali sono considerate sempre più come opportuni strumenti di marketing (Konijn et al., 2016), effettua una selezione per condividere con gli altri utenti alcuni luoghi preferiti di una determinata città. Infatti, le foto condivise su Flickr mostrano esattamente come i turisti si muovono all'interno di una città e forniscono informazioni su spazi pubblici, come piazze e quartieri. Sulla base di queste informazioni e partendo dalla letteratura precedente (Girandin et al., 2008; Konijn, et al., 2016; Zhang et al., 2019) i risultati suggeriscono quanto sia importante esplorare determinati modelli di comportamento, in questo caso l'attività fotografica del turista al fine di supportare concretamente le parti interessate nella gestione, ad esempio, dei flussi turistici che influenzano in modo significativo l'economia della città ed il benessere dei cittadini. La ricerca condotta potrebbe influenzare iniziative quali la proposta di nuovi itinerari per incoraggiare il viaggiatore a visitare altri luoghi, l'organizzazione di eventi e la predisposizione di trasporti efficienti per raggiungere luoghi lontani dal centro città al fine di migliorare

l'efficienza di una destinazione ed il suo sviluppo economico e culturale.

Riassumendo, è stato fornito uno strumento in grado di mappare i cluster all'interno di una città o di un gruppo di città e di conseguenza consentire la costruzione di modelli probabilistici in grado di produrre previsioni sufficientemente accurate sui dati. Il sistema di analisi adottato rivela chiari vantaggi nel riuscire a identificare rapidamente le dinamiche turistiche legate all'ambiente economico e culturale di una città, esaminando i dati in tempo reale estratti dai social media. Inoltre, l'analisi dei dati geo-referenziati e l'estrazione di informazioni rilevanti dai dati attraverso gli algoritmi di Machine Learning offre informazioni approfondite e significative che possono contribuire a prevedere le attività dell'utente e i suoi cambiamenti nel tempo.

Un futuro lavoro potrebbe considerare non solo i dati estratti da un'unica piattaforma, ma presentare un confronto con altre sorgenti, ad esempio TripAdvisor, tale da evidenziare le potenzialità di sviluppo di un'offerta turistica vasta e che coinvolga più segmenti, più settori (non solo i POI turistico-culturali) per una gestione sostenibile dei flussi turistici.

CAPITOLO 3

IMPATTO DEI BIG DATA SUL MARKETING E SULLA CUSTOMER EXPERIENCE

3.1 Digital marketing e Retail

Con oltre 3 miliardi di persone in tutto il mondo che utilizzano regolarmente il Web per intrattenimento, per comunicare con amici o per lo *scouting* di prodotti di vario genere, si sta assistendo ad un cambiamento nell'atteggiamento sia dei consumatori che delle imprese e del modo in cui queste ultime commercializzano. Quindi, la capacità di attuare con successo i processi di innovazione si basa sulle abilità delle aziende di cambiare forme, pratiche e strategie organizzative al fine di sfruttare il cambiamento tecnologico (Volderba et al., 2013). Infatti, la rivoluzione digitale e lo sviluppo delle innovazioni tecnologiche sono considerate, ad oggi, le discriminanti che favoriscono la crescita organizzativa delle aziende. La sfida per gli esperti di marketing è sviluppare politiche innovative tramite le quali ottenere vantaggi significativi, integrando strategie di marketing digitale con strategie di marketing tradizionale, tenendo conto anche delle specifiche che caratterizzano ciascuna azienda.

Il marketing digitale o *digital marketing*, precedentemente noto come e-marketing o marketing di rete, è al centro del business digitale e viene definito da Chaffey & Ellis-Chadwick (2019) come il raggiungimento degli obiettivi di marketing attraverso l'applicazione di tecnologie e media digitali. Questo dovrebbe determinare la spinta verso investimenti delle aziende nel web e non solo verso l'adozione delle tecnologie. Aziende come Zalando, risultano un'eccellenza nel marketing digitale in quanto sfruttano il potere

dei media e della distribuzione digitale per far crescere il proprio business. In effetti, grazie alle politiche adottate, Zalando è diventata una società da 6 miliardi di euro affermandosi come la migliore piattaforma di vendita al dettaglio di moda in Europa che gestisce le vendite di oltre 1.500 marchi internazionali. L'azienda ha sviluppato, per ciascuno dei 15 siti web, delle misure dedicate in base al paese da cui viene fatto l'accesso. Queste misure includono strategie personalizzate di marketing, dettate dal contesto socio-culturale del paese.

Il marketing digitale, se correttamente implementato, pienamente collima con la definizione di marketing, inteso come funzione organizzativa ed insieme di processi utili a creare, comunicare e fornire valore ai clienti e gestire le relazioni con essi in modo vantaggioso sia per le organizzazioni che per le parti interessate (American Marketing Association). Dunque, l'obiettivo del digital marketing è quello di avvicinarsi ai bisogni reali dei clienti per comprenderli meglio, aggiungere valore ai prodotti ampliando i canali di distribuzione ed aumentare le vendite attraverso l'esecuzione di campagne promosse tramite i media e pubblicità online. Inoltre, lo scopo è anche quello di avvalersi dei siti web per facilitare i clienti, guidare al meglio le vendite e gestire il servizio post-vendita. Come per il marketing tradizionale, il marketing digitale è un modo di pensare, un modo di mettere il cliente al centro di tutte le attività online. Per questo motivo il *brand* deve puntare a generare una *customer experience* senza soluzione di continuità, ovvero una soluzione di continuità tra i vari punti di contatto che permette l'interazione dei clienti con i marchi, i servizi e i prodotti.

La digitalizzazione del commercio si afferma quindi come una risorsa strategica per il business, una delle trasformazioni più importanti che attualmente caratterizza le opportunità e i modelli commerciali, i processi di acquisto e le attività di vendita tradizionali, innescando, ulteriormente, una drammatica trasformazione dei rivenditori e dello scenario competitivo in cui essi si esibiscono (Willems et al., 2017). Un'impostazione di vendita al dettaglio *smart* può rappresentare un modo vantaggioso per un'azienda di generare maggior valore per clienti e impresa stessa (Pantano & Timmermans, 2014). Le innovazioni introdotte nelle impostazioni di vendita al dettaglio sono molteplici, e vanno dalle APP mobili sviluppate *ad-hoc*, ai chip per pagamenti *contact-less* inseriti, ad esempio, nelle carte di credito, fino ai tavoli *multitouch* e vetrine interattive. L'intersezione tra vendita al dettaglio ed innovazione rappresenta un argomento di ricerca fruttuoso poiché la possibilità di integrare nuove e più efficienti

tecnologie digitali è un fattore che implica lo sviluppo di nuovi modelli di vendita dovuti al differente comportamento del consumatore intrinseco a questa modalità. I fattori che hanno determinato questo cambiamento includono:

- i. la continua spinta tecnologica, che ha portato ad una nuova tecnologia efficiente per supportare la gestione della vendita al dettaglio a diversi livelli;
- ii. la domanda dei consumatori per nuove esperienze di acquisto;
- iii. le strategie di innovazione pionieristiche della concorrenza.

Studi precedenti (Evanschitzky et al., 2015; Inamn & Nikolova, 2017) concordano sul fatto che l'introduzione delle tecnologie digitali migliora la qualità del servizio al dettaglio sia dal punto di vista dei consumatori che da quello dei rivenditori. Lo scopo è quello di innovare gli ambienti di vendita al dettaglio fornendo ai consumatori l'accesso ad informazioni aggiuntive sui prodotti per supportarli nella ricerca, la localizzazione e per fornire informazioni legate all'acquisto di beni, migliorando al contempo le loro esperienze di acquisto con servizi nuovi ed efficienti.

Di conseguenza, sono sempre più frequenti i casi in cui viene offerta al cliente la possibilità di vivere esperienze *in-store* altamente coinvolgenti ed entusiasmanti, implementate facendo particolare attenzione sia all'atmosfera, sia ai vari elementi multisensoriali che possono influire positivamente sull'aspetto percettivo ed emotivo dell'esperienza stessa. Particolari tecnologie digitali *in-store* includono la segnaletica digitale (chioschi e totem), app mobili create *ad-hoc*, commesse umanoidi e così via.

Nello scenario emergente, le nuove tecnologie stanno spingendo i rivenditori verso un mercato al dettaglio omnicanale, che si traduce nella giusta combinazione di canali online, offline e mobili (Bell et al., 2018).

Nel marketing, comprendere la situazione del mercato di riferimento è molto importante per indirizzare i *driver* del business. Nei mercati di consumo, i clienti stanno acquisendo maggiore sintonia con la spesa online e le tecnologie digitali consentono loro di essere ben informati nelle decisioni di acquisto. Secondo Doherty ed Ellis-Chadwick (2010) in futuro quando il consumatore diventerà più informato ed esigente potrebbe innescarsi una significativa lotta per il potere tra il rivenditore ed il consumatore.

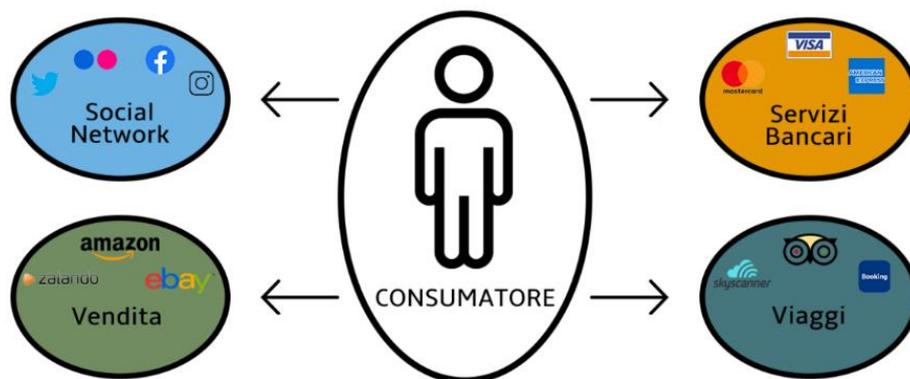


Figura 3.1. Produzione di Big data da parte del consumatore e settori di interesse

3.2 Big data e Marketing Intelligence

L'automazione sta diventando di fondamentale importanza nel marketing e, più in generale, nel settore dei servizi attraverso l'adozione di tecnologie digitali, in particolare quelle legate all'intelligenza artificiale (AI), le quali permettono alle aziende di interagire in modo adattivo con i clienti. Inoltre, la digitalizzazione sta spingendo sempre più verso i canali elettronici (Trampe et al., 2014) che si presentano come un nucleo di comunicazione che mette a disposizione una grande quantità di informazioni di diversa natura. Consideriamo, ad esempio, tutte le informazioni a cui è possibile accedere, che riguardano il tempo speso sul sito web da parte degli utenti, la visualizzazione di ciascun prodotto, i successivi acquisti e pensiamo a come i dati prodotti da queste azioni possano trasformarsi in informazioni utili alle aziende per migliorare le decisioni operative, prevedere la domanda ed indirizzare i consumatori. Infatti, i Big data generati dai consumatori, se gestiti ed analizzati attraverso l'uso delle tecnologie, costituirebbero un valido supporto per i responsabili di marketing e per i rivenditori i quali adeguerebbero le strategie di gestione e di vendita alla "conoscenza" ricavata dalle analisi.

Tuttavia, sebbene i Big data rappresentino un simbolo di avanguardia in termini di gestione delle informazioni, la loro esplosione si sta rivelando sia un problema che un'opportunità. Attualmente, le aziende applicano l'analisi dei Big data per ottenere un vantaggio sulla qualità del prodotto e sulla soddisfazione dei clienti (Zhao et al., 2019), ma sono ancora "nuove" nello sfruttamento del vero potenziale dei Big data. Le principali sfide nelle quali incorrono sono correlate all'incompletezza, alla tempestività e alla

scalabilità dei Big data (Labrinidis & Jagadish, 2012). Ad esempio, problemi con software, database, personale e competenze insufficienti sono i principali ostacoli a questo tipo di analisi e, in un mercato competitivo, tali negligenze potrebbero danneggiare un'azienda. Come già anticipato nel Capitolo 1, le aziende potrebbero risolvere questi problemi sviluppando e organizzando la funzione analitica all'interno della propria organizzazione o accedendo a servizi di terze parti, come i servizi offerti da Google Cloud. Inoltre, l'adozione e l'uso di sofisticate tecniche di analisi, come gli algoritmi di apprendimento automatico, consentirebbe di comprendere fenomeni complessi relativi ai Big data e offrire risultati innovativi per migliorare il processo decisionale in relazione a tutti gli aspetti del *marketing intelligence*. Quest'ultimo concetto venne definito da Tan e Ahmed (1999) come una struttura di persone, attrezzature e procedure utili a collezionare, ordinare, analizzare e distribuire informazioni pertinenti, tempestive e accurate, che i responsabili di marketing possono utilizzare per migliorare la pianificazione, l'implementazione ed il controllo del marketing. Dunque, nel contesto del *marketing intelligence* con nuovi strumenti analitici gli esperti potrebbero estrarre conoscenza dai Big data e personalizzare i programmi di marketing. Si sta assistendo ad un nuovo approccio di marketing che considera in tempo reale tutte le potenziali offerte per ogni singolo cliente e ne determina quella ottimale. La futura migliore offerta di azione è determinata dagli interessi e dalle esigenze del cliente. Ciò è in netto contrasto con gli approcci di marketing tradizionali, i quali prevedevano di creare prima un'offerta per un prodotto o servizio e poi tentare di trovare prospettive interessanti e ammissibili. Pertanto, ad oggi, gli esperti di marketing devono:

- i. estrarre i contenuti generati dai consumatori (ascoltando ciò che dicono i consumatori);
- ii. utilizzare questi contenuti per sviluppare delle previsioni.

In questo contesto, i Big data possono essere adoperati come ricco set di dati sull'esperienza dei consumatori in relazione ad un determinato marchio, prodotto o servizio e le loro analisi possono fornire una conoscenza sistematica delle valutazioni dei consumatori sugli elementi considerati più o meno interessanti concernenti un determinato rivenditore (Gensler et al., 2015). Inoltre, la capacità di analizzare l'enorme quantità di dati emergenti dal comportamento dei consumatori online è una sfida promettente ed un'opportunità interessante per i responsabili di marketing che possono

utilizzare le informazioni ricavate per un'ulteriore massimizzazione del profitto (Bradlow et al., 2017). Indubbiamente, questi grandi sforzi di raccolta dei dati generati da piattaforme digitali interattive forniscono un modo per comprendere costrutti e processi psicologici che sono stati poco pratici, se non impossibili, fino a pochi anni fa (Jaffe, 2014).

3.3 Processi cognitivi e comportamento del consumatore

Le scienze cognitive rappresentano un sistema di riferimento importante per quanto riguarda il settore del marketing. Permettono di comprendere i processi cognitivi legati alle scelte di acquisto e di consumo e consentono di definire i fattori che orientano i comportamenti individuali e la motivazione dei consumatori. Inoltre, permettono di esaminare le componenti delle organizzazioni e i meccanismi collettivi e sociali ad esse interni, che ne condizionano le dinamiche (Williams, 2001).

Tra i modelli teorici su cui si fonda l'analisi del comportamento si annoverano i "modelli comportamentali" (Kotler, 1991) che prevedono l'attivazione di una "risposta" da parte del consumatore ogni qual volta viene sottoposto ad uno "stimolo".

Solomon (2004) definì il comportamento del consumatore come lo studio dei processi coinvolti quando individui o gruppi selezionano, acquistano, usano o eliminano prodotti, e servizi per soddisfare bisogni e desideri. Inoltre, evidenzia come il consumo di prodotti e servizi è sempre più connesso ai "significati" degli stessi piuttosto che ai benefici funzionali, che nelle economie moderne sono assunti come un dato. In questa prospettiva, particolarmente rilevanti risultano, sia le ricerche che analizzano il comportamento del consumatore (Dalli & Romano, 2003), sia lo studio del marchio come strumento cognitivo e fiduciario che ispira ed influenza i comportamenti di acquisto e di consumo (Busacca, 2000).

A tal proposito, i fattori critici nelle decisioni di marketing sono rappresentati dalla relazione tra il consumatore e i processi di acquisto, in particolar modo nei contesti ipercompetitivi. Il consumo dei prodotti e servizi da parte dei consumatori, è inteso come:

- i. *esperienza*, in termini di emozioni che essi sono in grado di fornire;
- ii. *integrazione* intesa come valore sociale;
- iii. *legame* in riferimento a qualcuno o qualcosa;

- iv. *commedia* in termini di partecipazione e condivisione di un'esperienza con altri.

L'analisi del comportamento relativo al consumo assume, dunque, nuovi significati alla luce dei cambiamenti del consumatore e del proprio comportamento di acquisto. Generalmente, la linea di ricerca che si sviluppa attorno al concetto di consumatore può essere suddivisa in tre aree:

- i. *elaborazione delle informazioni;*
- ii. *teoria della cultura del consumatore;*
- iii. *ricerca sulle decisioni comportamentali.*

Per *elaborazione delle informazioni* si intende l'interazione dei processi motivazionali e affettivi, tipici in psicologia, della cognizione sociale (Simonson et al., 2001) e utili nella comprensione di temi come la persuasione e le influenze implicite sul comportamento del consumatore. La *teoria della cultura del consumatore* riguarda invece la dimensione esperienziale e socio-culturale del consumo utilizzando dati qualitativi. L'ultima area *ricerca sulle decisioni comportamentali* concerne le analisi normative, descrittive e prescrittive che riguardano rispettivamente come i consumatori dovrebbero prendere decisioni, come migliorare le decisioni e come effettivamente essi decidono.

Di seguito vengono riportati tre ricerche relative all'analisi dei Big data eseguite con lo scopo di comprendere il comportamento del consumatore.

1. La prima ricerca riguarda l'applicazione della sentiment analysis sui Big data estratti da Twitter e relativi a tre aziende *fast fashion*.
2. La seconda ricerca concerne l'analisi di Big data non strutturati, ovvero immagini condivise dagli utenti sulla piattaforma TripAdvisor.
3. La terza ricerca presenta l'analisi dei cluster applicata per profilare classi di utenti nel settore dell'energia attraverso dati estratti da un questionario online ideato *ad-hoc* e somministrato ad un campione rappresentativo.

3.4 Twitter e Sentiment Analysis

L'industria della moda affronta continuamente sfide per cercare di soddisfare più rapidamente il desiderio dei consumatori riguardo nuovi prodotti e servizi. Questo andamento comporta la necessità di indagare continuamente sul mercato per comprendere e rispondere in modo rapido ed efficace alla domanda.

Diversi ricercatori tra cui Homburg e colleghi (2015) e Liang e colleghi (2015) hanno adottato la metodologia della sentiment analysis al fine di indagare sui dati condivisi sulle piattaforme social. Come già riportato nel Capitolo 1 (sezione 1.4.4) questa metodologia permette di raccogliere informazioni sui consumatori quasi in tempo reale e risulta meno costosa rispetto alle tecniche tradizionali basate su questionari strutturati (Shayass et al., 2017). Tuttavia, nella maggior parte degli studi, l'analisi del sentiment è stata condotta attraverso un approccio semi-manuale che ha richiesto la segmentazione e la codifica del testo da parte del ricercatore, allo scopo di estrarre parole trasmettenti il sentimento, ulteriormente codificate ed analizzate attraverso modelli di regressione tradizionali (Mostafa, 2013).

3.4.1 Obiettivo della ricerca

L'obiettivo preposto in questo studio è stato quello di esplorare i contenuti generati online da parte dei consumatori in termini di commenti positivi o negativi al fine di aumentare le informazioni sul marketing, scegliendo uno studio a più casi ed applicando una procedura sistematica ed automatizzata.

Generalmente, gli studi a più casi, che si basano sulla logica della replica, prevedono lo studio di ogni caso come un esperimento indipendente che si erge a sé stante come unità analitica offrendo risultati più affidabili e generalizzabili rispetto agli studi a caso singolo (Ellonen et al., 2009).

3.4.2 Procedura e metodologia di analisi

Seguendo la strategia suggerita da Yin (2003), i casi scelti sono stati tre e riguardano rivenditori con caratteristiche simili in quanto:

- i. provenienti dallo stesso settore: *fast fashion* (specializzato in accessori);
- ii. aventi prestazioni nello stesso mercato: Regno Unito;
- iii. aventi dimensioni simili: grandi aziende.

Pertanto, i casi scelti sono concorrenti l'uno con l'altro. Le aziende selezionate sono state: *Skinnydip London*, *Missguided* e *Accessorize*. Il campione è costituito da 9.652 tweet (2.552 per il rivenditore A, 4.900 per il rivenditore B e 2.200 per il rivenditore C) pubblicati a febbraio 2018. Per l'estrazione dei dati è stato interrogato il database di *Mathematica* che, attraverso una connessione all'API di Twitter, ha permesso di ricavare i dati. La presenza di un *hashtag* specifico ha permesso di ottenere i tweet recenti aventi le caratteristiche scelte.

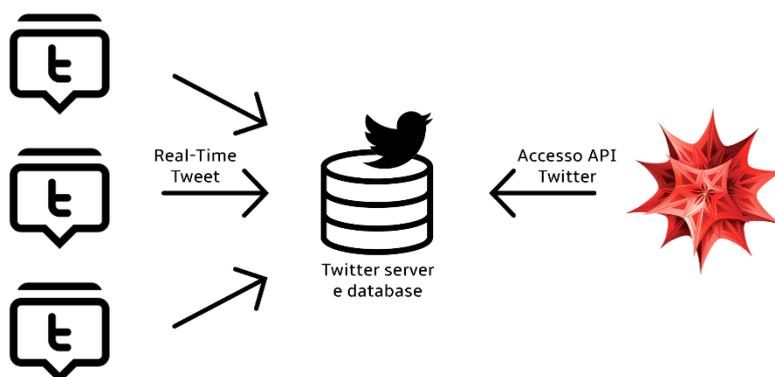


Figura 3.2. Interrogazione delle API di Twitter tramite Wolfram Mathematica

In particolare, i dati sono stati ricavati effettuando la ricerca tramite il comando *TweetSearch* che invia una richiesta all'API di Twitter, utilizzando specifici parametri

[TweetSearch, "request", params]

dove *request* è la query (search string) e *params* è la parola chiave (hashtag).

Come parametro per la ricerca dei tweet sono stati utilizzati i nomi, nello specifico gli *hashtag* (tipo di etichetta) delle tre aziende scelte. Il processo è stato avviato in egual modo per ogni azienda.

Il software ha generato un database con il testo completo del tweet ed altri metadati, tra cui la data della pubblicazione, la lingua ed il retweet (condivisone da parti di terzi). Lo

studio ha limitato le raccolte di tweet a quelle in sola lingua inglese per evitare possibili problemi emergenti dall'analisi dei tweet multilingue (Mostafa, 2013).

ID	Text	CreationDate	Entities	Language	FavoriteCount	RetweetCount
968444114097328128	Absolute brunch goals 🍷 tag your brunch bestie in the comments! ☺ Shop our lol 🍷	Tue 27 Feb 2018 11:14:04	hashtags symbols user_mentions	en	9	0
968250449890545664	May or may not have just ordered a lush new backpack and cute choker from @Skinn	Mon 26 Feb 2018 22:24:31	hashtags symbols user_mentions	en	0	0
968081729155227648	Monday hustle photo: @topgirl.studio #skinnydiplondon #mondaymotivation #hustle	Mon 26 Feb 2018 11:14:04	hashtags symbols user_mentions	en	11	1

Figura 3.3. Output dei tweet e dei relativi metadati ricavati con Mathematica

Sia per l'estrazione dei dati che per l'analisi il software adottato è stato *Wolfram Mathematica*. Per procedere all'analisi dall'output (Figura 3.3) è stato estratto il testo su cui avviare la sentiment.

Di seguito viene riportato il codice realizzato per l'estrazione del testo.

```

text = Table[StringDrop[StringSplit[ToString[ll[[n, 2]]], ","][[1]],
  7], {n, Length[ll]}]

{ for you 🍷🍷🍷 @imlaurenblack looking 🍷🍷🍷 with our
  chilli manoa shoulder bag and red lola sunglasses.
  Shop the bag here https://t.co/QQN3ci6HP3 and the
  sunglasses here https://t.co/O4Wm4v9M2a #skinnydiplondon
  #newin #sunglasses #bag https://t.co/96RJAqQ5Xx,

```

Figura 3.4. Codice sviluppato in Mathematica per l'estrazione del testo dalla tabella di output

In questo studio è stato applicato un modello già implementato, creato e addestrato tramite la “*Wolfram Neural Net Repository of Neural Network Models*”, costituita da 70 modelli di reti neurali. Nello specifico è stata utilizzata una Machine Learning supervisionata ovvero pre-addestrata. Durante il processo di addestramento, alla macchina viene fornito un *dataset* costituito da un'enorme quantità di dati, dal quale essa apprende. La fase di apprendimento si interrompe quando l'algoritmo raggiunge un buon livello di prestazione.

In particolare, in questo caso è stata adottata la funzione *Classify* che genera l'operazione di classificazione basata su esempi e classi già contenuti nel software e raggruppa i dati in categorie. Basandosi sul framework di reti neurali *Wolfram Mathematica* fornisce un

sistema uniforme per la memorizzazione dei modelli di rete neurale. Il processo di classificazione del sentimento si riferisce a reti neurali artificiali.

Quando viene applicato la funzione *Classify* il software seleziona il classificatore incorporato “Sentiment” per dedurre il sentimento di un messaggio, in questo caso estratto da un social network.



Figura 3.5. Modulo Classify e classificatore incorporato “Sentiment”

Il software, basandosi sui dati in input, sceglie automaticamente il miglior algoritmo da adottare in base ai modelli disponibili creati e formati attraverso la *repository* tentando di trovare il modello con la più alta probabilità di dati invisibili (ovvero su set di test). Innanzitutto, vengono selezionati i possibili candidati (in base alle caratteristiche dei dati); in secondo luogo, i modelli competono l’uno contro l’altro utilizzando tecniche di validazione incrociata (Kohavi, 1995); infine, viene selezionato il modello migliore per identificare il sentimento trasmesso da uno *snippet* (frammento) di testo.

Dopo aver importato una lista di esempi il software, adoperato per l’analisi, permette inoltre di ottenere le probabilità sull’individuazione delle possibili classi: positive, negative e neutre e quindi di avere nel dettaglio, un visione sulle performance del classificatore impiegato.

```

Classify["Sentiment", {"I love this movie", "so depressed today",
  "My phone broke again"}, "Probabilities"]
{<|Positive → 0.709808, Neutral → 0.055217, Negative → 0.23467|>,
 <|Positive → 0.499522, Neutral → 0.28636, Negative → 0.214118|>,
 <|Positive → 0.0498463, Neutral → 0.0615645, Negative → 0.888589|>}

```

Figura 3.6. Percentuale di probabilità sull’individuazione delle categorie: positivo, negativo e neutro

Per ottenere una descrizione del metodo adottato dal software è stato applicato il modulo *ClassifierInformation* che fornisce informazioni sul classificatore. L’immagine seguente,

mostra il metodo scelto, in questo caso il metodo Markov generalmente utilizzato per l'analisi del testo in *Wolfram Mathematica*.

Classifier information	
Method	Markov
Number of classes	3
Number of features	1
Number of training examples	4
Number of tokens	12
Order	0

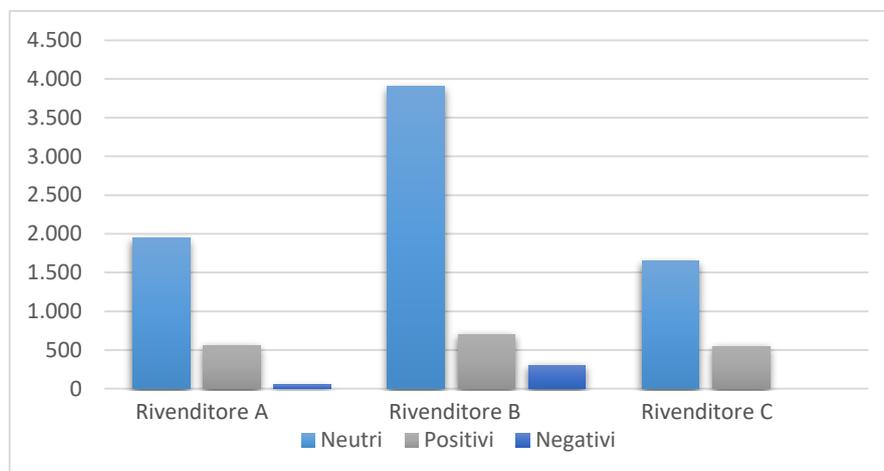
Figura 3.7. Informazioni sul metodo applicato per la classificazione del sentiment

Data una determinata categoria o classe (nel presente caso positiva, neutra o negativa), il classificatore di Markov dell'ordine 0 presuppone che i token siano generati in modo indipendente e adotta il teorema di Bayes (che fornisce la probabilità di un evento, attingendo dallo stato dell'evento immediatamente precedente) per prevedere la categoria. Questa procedura è anche chiamata Unigram model o naïve Bayes model.

3.4.3 Risultati

Sebbene esista un numero massimo predefinito di tweet che possono essere analizzati contemporaneamente in base alla capacità delle prestazioni del computer, in questo caso 5.000 richieste, nessuno dei rivenditori ha raggiunto questo numero. In particolare, sono stati raccolti 2.552 tweet per il rivenditore A, 4.900 per il rivenditore B e 2.200 per il rivenditore C. Il numero di tweet per rivenditore fornisce una misura preliminare della quantità di comunicazioni e-WOM (ovvero il passaparola elettronico) per ciascun rivenditore, definendo anche l'indice di popolarità su Twitter. In effetti, per il rivenditore A, sono stati identificati 1.944 tweet neutri, 559 tweet positivi (22% dei tweet totali) e 49 negativi. Per il rivenditore B, l'analisi del sentiment ha identificato 3.900 tweet neutri, 700 positivi (14% del totale) e 300 negativi (6% del totale). Per il rivenditore C, l'analisi del sentiment ha identificato 1.650 tweet neutri, 550 positivi e nessun tweet negativo, quindi i tweet positivi sono il 25% del totale (Tabella 3).

Tabella 3. Confronto dei risultati dell'analisi del sentiment



Dall'analisi emerge come il rivenditore C sia mancante di tweet negativi, risultando come l' esercente che ha sollecitato nei consumatori solo "sentimenti" positivi. Pertanto, l'analisi fornisce una sorta di indice del sentimento, composto da una parte positiva e negativa, consistente nella percezione nei confronti di un determinato rivenditore da parte del cliente. I prodotti, i servizi e le campagne di marketing del rivenditore C sono percepiti dai consumatori, utenti di Twitter, come di maggior successo se confrontati con i principali concorrenti (rivenditore A e rivenditore B). D'altra parte, il rivenditore B mostra il numero più alto di tweet negativi (14%), il che implica che un numero abbastanza consistente di consumatori non risulta soddisfatto, esprimendo, di conseguenza, sentimenti negativi. Ciò potrebbe avere sfavorevoli conseguenze nelle comunicazioni e-WOM, sollecitando altri consumatori a condividere commenti simili. Dalla letteratura emerge, infatti, come leggendo uno *status* o vedendo un'immagine fortemente caratterizzata dal contenuto emotivo, gli utenti possano essere influenzati e cambiare la propria inclinazione inconsciamente pensando o provando lo stesso sentimento (Lee & Lim, 2015).

3.4.4 Discussioni

L'analisi del sentimento fornisce misure delle comunicazioni online effettuate dai consumatori effettivi e relative a tre diversi rivenditori di fast fashion determinando la

percezione (sentimento) che i consumatori hanno dello stesso rivenditore (indice dei sentimenti) e ciò che i consumatori esprimono in termini di commenti positivi, negativi e neutri. Poiché i contenuti online generati dai consumatori riflettono la qualità del prodotto e la popolarità di un determinato marchio o servizio (Tsekouras, 2017), questo studio evidenzia l'importanza dell'analisi dei Big data, in quanto presenta una conoscenza sistematica della valutazione dei consumatori in merito agli elementi più o meno interessanti relativi ad un determinato rivenditore fornendo un approccio di supporto per farlo concretamente. Dalla letteratura emerge che mentre altri autori hanno cercato di identificare gli effetti di commenti, post e recensioni online positivi o negativi (Hornik et al., 2015; Pang & Qiu, 2016), il presente studio è il primo a mostrare come rilevare sistematicamente i sentimenti positivi o negativi dei tweet condivisi per migliorare le informazioni di marketing dei rivenditori di *fast fashion* fornendo, allo stesso tempo, una misura della relazione marchio-cliente. Ricerche precedenti attraverso l'analisi del *sentiment* hanno misurato l'effetto dei commenti positivi, negativi o neutri sulle vendite giornaliere online di un'azienda e l'effetto influente che le recensioni negative hanno avuto sulle vendite (Sonnier et al. 2011). A tal proposito, la ricerca condotta presenta implicazioni pratiche per l'industria della moda *fast*, caratterizzata dalla necessità di rispondere rapidamente alle richieste dei consumatori su nuovi prodotti avvalendosi di sistemi di supporto utili per accedere alle informazioni sulle attitudini dei clienti in tempo reale (Cook & Yurchisin, 2017) e mostra anche fino a che punto è possibile raccogliere questo tipo di dati (tweet) su diversi rivenditori, citando solo il nome dell'azienda scelta. In particolare, l'analisi effettuata dimostra fino a che punto un sistema informativo come quello presentato può rendere sistematica la procedura di rilevazione dei commenti negativi (oltre a quelli positivi), che richiede tempo e costi per essere eseguita manualmente. Ottenere una rapida panoramica delle opinioni dei consumatori consente di valutare i punti di forza e di debolezza dei prodotti, nonché di raccogliere informazioni aggiuntive sulle prestazioni dei concorrenti. Dunque, i rivenditori di *fast fashion* possono beneficiare di un rapido accesso alle opinioni dei consumatori in termini di commenti positivi e negativi e al contempo confrontarli con quelli riguardanti la concorrenza. Ricerche precedenti consideravano il *data mining* dei contenuti online dei consumatori come un'attività non banale a causa della quantità di contenuti e dello stile informale (Xu et al., 2011). Lo studio presentato risponde a queste domande ed estende altri studi fornendo un approccio efficace e sistemico per accedere ed indagare sul ricco set di dati

per ottenere approfondimenti in grado avere un impatto sul *marketing intelligence*.

3.5 TripAdvisor e analisi delle immagini

Il successo nella gestione del marchio deriva dalla giusta comprensione della percezione del *brand* da parte dei consumatori e dalla capacità dei manager di rispondere correttamente alla domanda per generare redditività. In particolare, la crescente richiesta di marchi di lusso sta aggiungendo complessità al mercato in questo settore, ponendo nuove sfide per i manager e per la gestione di hotel di lusso che sta acquisendo l'interesse di studiosi e professionisti sia per quanto riguarda il *brand* che il turismo in questo campo. Per tali motivi, alcuni studi si sono concentrati sulla soddisfazione dei consumatori in relazione ai marchi alberghieri attraverso questionari, valutazioni online, analisi del sentiment ed interazioni sui social media (Lai & Hitchcock, 2017; Kim & Lee, 2019). Invece, per quanto riguarda il *luxury brand*, in relazione al settore alberghiero, la ricerca si concentra principalmente sulla valutazione della qualità dei servizi percepita dai clienti, su promozioni e prezzi, sui fattori che favoriscono la fidelizzazione dei clienti e sulle prestazioni dei siti web (Khan & Rahman, 2017; Liu et al., 2017). Tuttavia, a causa della continua crescita nel settore della ricettività turistica, in linea con una crescente richiesta di esperienze di lusso, l'indagine sui marchi di hotel di lusso sta emergendo come argomento caldo nella ricerca nel settore turistico. A differenza degli altri settori del lusso, l'industria alberghiera si basa sulla "vendita" di esperienze piuttosto che sulla vendita di beni materiali, quindi mira a fornire un servizio superiore atto a soddisfare le esigenze dei clienti. Gli hotel di lusso mostrano, inoltre, caratteristiche di prestigio, prezzo premium e conoscenza intima degli ospiti (Sherman, 2005). Ad esempio, i viaggiatori potrebbero preferire un hotel di lusso per servizi quali le prenotazioni assicurate, il parcheggio gratuito, alte prestazioni del personale, servizi in camera e così via (Ariffin et al., 2018). Gli ospiti si sono, dunque, trasformati in consumatori nel momento in cui sono diventati più consapevoli della qualità e del valore dei servizi offerti dagli hotel di lusso e le loro valutazioni, in termini di recensioni e/o immagini scattate negli hotel, diventano fondamentali in quanto influiscono sulla reputazione del *brand*. Estrarre informazioni dai dati condivisi dai clienti online, in particolare sulle piattaforme quali TripAdvisor, sosterebbe in primo luogo la raccolta di approfondimenti sulle preferenze e sulle scelte

dei consumatori e di conseguenza permetterebbe all'azienda di soddisfare appieno il cliente. L'applicazione di potenti strumenti di analisi quali gli algoritmi di Machine Learning (ML) permette di estrarre, con successo, informazioni da dati non strutturati come immagini per esplorare le carenze e le risorse dei servizi degli hotel (Valdivia et al., 2017) con particolare attenzione alle fonti quali i social media, che risultano ancora poco "sfruttati" negli studi relativi al *luxury branding*.

3.5.1 Obiettivo della ricerca

La presente ricerca è stata condotta con lo scopo di identificare le caratteristiche di un hotel di lusso che acquistano maggiore impatto sulla valutazione da parte dei clienti adottando una nuova metodologia in grado di supportare la raccolta e l'analisi di "grandi dati visivi", in termini di immagini. In particolare, sono stati adottati algoritmi di Machine Learning per analizzare un set di dati estratti da TripAdvisor che hanno permesso di acquisire e misurare gli attributi degli hotel di lusso come principali *driver* per la valutazione dell'esperienza in alcuni hotel da parte dei clienti. Pertanto, mentre gli studi precedenti si sono concentrati sull'aspetto sociale della fotografia (Lo & McKercher, 2015), questa ricerca riflette sulla comprensione del comportamento del consumatore e sulle sue preferenze e suggerisce come sia necessario che la ricerca relativa al *luxury brand*, in particolare al settore alberghiero si concentri sull'elaborazione e sull'analisi delle immagini considerando i progressi degli algoritmi di apprendimento automatico.

3.5.2 Procedura e metodologia di analisi

Sono state prese in esame alcune celebri catene operanti nel settore del lusso presenti sul territorio londinese. In particolare, sono stati esaminati i casi di sei Hotel di lusso aventi caratteristiche comuni quali:

- i. essere nella stessa area geografica;
- ii. disporre di un numero elevato di camere (tra 85 e 100 - comprese le suite) con tariffe comparabili (a partire da 1.000 sterline a notte), servizi (es. piscina, spa, trasporti aeroportuali, ecc.);
- iii. avere più di 500 recensioni su TripAdvisor.

TripAdvisor è una piattaforma web di viaggi, che pubblica informazioni riguardo hotel, ristoranti ed attrazioni turistiche. Essa, è considerata la più grande piattaforma di viaggio con in media oltre 455 milioni di visitatori mensili e oltre 630 milioni di recensioni relative ad hotel, ristoranti ed attività commerciali correlate (TripAdvisor, 2018). In particolare, le valutazioni vengono scritte dagli utenti, fruitori delle strutture, ed ogni recensione viene valutata dallo staff che ne giudica la coerenza con le linee guida del sito. La piattaforma, permette anche la condivisione di foto scattate dai clienti che si presenta come la possibilità di condividere con altri l'esperienza vissuta in un determinato luogo ed ottenere l'apprezzamento di altri migliorando l'esperienza in una sorta di "circolo ermeneutico" (Balomenou & Garrod, 2019; Nikjoo & Bankhshi, 2019). Ai fini della ricerca, sono state raccolte 7.395 immagini pubblicate unicamente dagli utenti di TripAdvisor nella sezione "Viaggiatori" (Figura 3.8).

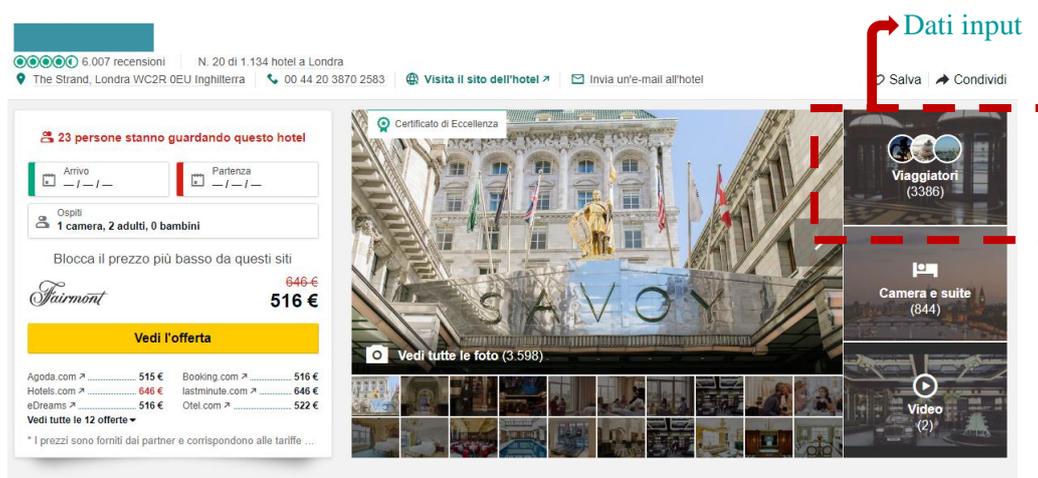


Figura 3.8. Dati riassuntivi relativi ad un hotel su TripAdvisor. Fonte https://www.tripadvisor.it/Hotel_Review-g186338-d187686-Reviews-The_Savoy-London_England.html

La raccolta dei dati è avvenuta utilizzando un web-crawler, *JDownloader*, un gestore di download di file con oltre 20 milioni di utenti che esplora il Web e attraverso collegamenti ipertestuali archivia le pagine memorizzate in un ampio database. Attraverso la creazione di uno script (figura 3.9), è stato possibile ricavare anche i commenti relativi alle foto ed ottenerli in file .csv.

```

var bWriteFile = false;
var sInfoFilePath = "";
var sText = "";
var sInfoFileType = ".csv"
var iContent = 0;
var myPackage = package;
var aParts = myPackage.getDownloadLinks();

if (myPackage.isFinished() == true) {
    bWriteFile = true;
}

if (bWriteFile == true) {
    sText += "Page,Instagram URL,Image,Comment" + "\r\n"
    for (var i = 0; i < aParts.length; i++) {
        sText += myPackage.getName() + ",";
        if (aParts[i].getUrl() != undefined) {
            sText += aParts[i].getUrl() + ",";
        }
        sText += myPackage.getDownloadFolder() + aParts[i].getName() + ","
        if (aParts[i].getComment() != undefined) {
            sText += "'" + aParts[i].getComment() + "'" + ",";
            iContent += 2;
        }
        sText += "\r\n";
    }
}

if (bWriteFile == true && iContent > 0) {
    sInfoFileType = "." + iContent + sInfoFileType
    sInfoFilePath = myPackage.getDownloadFolder() + "/" + myPackage.getName() + sInfoFileType;
    if (sInfoFilePath.length > 255) {
        sInfoFilePath = myPackage.getDownloadFolder() + "/" + "jd" + sInfoFileType;
    }
    try {
        writeFile(sInfoFilePath, sText, true);
    } catch (e) {
    }
}
}
!

```

Figura 3.9. Script per ottenere i commenti

I dati acquisiti sono stati, dunque immagini e commenti. Tuttavia, ai fini dell'analisi sono state prese in esame solo le immagini, dunque i "contenuti visivi".

Per l'analisi sono stati applicati algoritmi di Machine Learning, nello specifico il software utilizzato fornisce la funzione *ImageIdentify* già addestrata e implementata che consente di riconoscere l'oggetto presente nell'immagine e assegnare a ciascuno di essi la categoria di riferimento, creando una classificazione.

Al software sono state fornite le immagini scattate dai viaggiatori nei sei hotel ed è stata applicata la funzione *ImageIdentify* in grado di distinguere l'elemento principale presente in ogni immagine restituendo come risultato una classificazione espressa in categorie.

La funzione adottata si basa su un particolare tipo di intelligenza artificiale denominata *deep learning*, volta a simulare il processo di apprendimento umano al fine di rendere il sistema in grado di migliorarsi qualora siano disponibili nuove informazioni. Oltre che fornire un'etichetta di classificazione, dalla funzione si può ricavare la probabilità e l'accuratezza con cui opera nel processo di riconoscimento. La probabilità descrive il grado di identificazione dell'aspetto dell'oggetto utilizzando un intervallo di valori compreso tra 0 e 1 mentre l'accuratezza indica la quantificazione della differenza tra il valore effettivo e quello osservato.

Dal punto di vista matematico, il software applica la funzione che ha la più alta probabilità

di risposta sul set di dati (Wolfram, 2017). Il metodo applicato come classificatore è una regressione logistica (Figura 3.10), un metodo di interpolazione tra dati. I modelli di regressione logistica vengono applicati nel momento in cui si vuole studiare la relazione causale tra una variabile dipendente dicotomica e una o più variabili indipendenti quantitative (Hosmer et al., 2013).

La figura seguente riassume le informazioni sul classificatore e sul metodo applicato.



Figura 3.10. Metodo applicato per la classificazione

Tuttavia, per testare il software e ricavare la probabilità e l'accuratezza sulla corretta identificazione dell'oggetto nell'immagine è stato praticato un "addestramento". In particolare, è stato addestrato un classificatore in grado di riconoscere le varie stanze di un hotel. Alla macchina è stato fornito un set di dati iniziale (training set) costituito da foto scattate dai consumatori negli hotel di lusso che suggeriscono una categoria specifica (camera da letto, ristorante, bagno) dal quale la macchina ha appreso. Successivamente, è stata applicata la funzione *Classify*, attraverso la quale il software ha categorizzato gli elementi forniti durante la fase di training restituendo le classi assegnate a ciascuna immagine. Infine è stato eseguito il test set al fine di verificare la correttezza dell'algorithmo e l'eventuale *overfitting* con l'insieme di addestramento.

La figura seguente (Figura 3.11) mostra il codice adottato per addestrare e testare le prestazioni ed il funzionamento del software.

```

trainingset =
{
   -> "bathroom",  -> "bathroom",  -> "bathroom",
   -> "bathroom",  -> "bathroom",
   -> "bedroom",  -> "bedroom",  -> "bedroom",
   -> "bedroom",  -> "bedroom",
   -> "restaurant",  -> "restaurant",  -> "restaurant",
   -> "restaurant",  -> "restaurant"};

c = Classify(trainingset)

ClassifierFunction [  Input type: Image
Classes: bathroom, bedroom, restaurant ]

c [  ]
bedroom

c [  , "Probabilities" ]
( | bathroom -> 8.83925 x 10-161, bedroom -> 1., restaurant -> 2.48288 x 10-198 | )

testset = {  -> "bedroom",  -> "restaurant",  -> "bathroom" };

ClassifierMeasurements [c, testset, "Accuracy"]
1.

```

Figura 3.11. Training set, test set, probabilità e accuratezza

L'output mostra il grado di precisione con cui il software raggiunge gli obiettivi richiesti.

3.5.3 Risultati

Sebbene i consumatori abbiano scattato diverse foto negli hotel, prestando attenzione a diversi oggetti, ci sono alcuni elementi ricorrenti nelle immagini dei sei hotel. La camera da letto e il tavolo da pranzo sono i più fotografati in ogni immagine, apparendo come il primo o il secondo elemento rappresentato nelle foto scattate dagli ospiti in ogni hotel.

Le principali attrazioni dell'hotel sono il bagno, in particolare il lavandino, il letto matrimoniale ed il soggiorno per quanto riguarda la camera; le persone ed il ristorante per quanto riguarda il resto degli elementi fotografati. Dai risultati si evince come i consumatori siano molto attenti alla qualità dei servizi delle aree private come camera da letto e bagno, con enfasi in particolari come il lavandino o il letto. Questo risultato potrebbe indicare una particolare attenzione, da parte del cliente, sia alla pulizia che alla cura di questi elementi. Inoltre, fotografare questi elementi o luoghi indicherebbe ulteriore attenzione al *design* e all'architettura, al comfort e agli spazi in cui i consumatori si rilassano. In altre parole, questi risultati sintetizzano i *driver* della valutazione dei consumatori sull'esperienza di consumo degli hotel di lusso.

Nella tabella di seguito vengono mostrati i cinque elementi più fotografati in ogni hotel, con il relativo numero di volte in cui appaiono nelle immagini.

Tabella 4. Output di Mathematica con elenco dei cinque elementi maggiormente fotografati in ogni hotel e numero di volte in cui appaiono

hotel	elemento	numero
haymarkethotel	living room	51
	bedroom	48
	bathroom	31
	seat	11
	washbasin	11
intercontinentalhotel	bedroom	67
	washbasin	47
	airport terminal	47
	bathroom	41
	shower	36
milestonehotel	living room	42
	bedroom	37
	dinner table	34
	buffet car	20
	double bed	17
bulgarihotel	home theater	10
	restaurant	9
	double bed	9
	bedroom	8
	airport terminal	8
savoyhotel	bedroom	171
	bathroom	95
	dinner table	91
	antechamber	88
	living room	74
ritzhotel	dinner table	152
	bedroom	128
	restaurant	106
	antechamber	91
	living room	63

3.5.4 Discussioni

Sinora, nella maggior parte delle ricerche sono state adottate metodologie tradizionali per raccogliere dati sui consumatori impiegando non più di 1.000 soggetti e avvalendosi di metodologie tradizionali per la raccolta e l'analisi dei dati (ad esempio interviste, sondaggi, questionari, ecc.). La ricerca presentata mostra invece la possibilità di indagare su una grande quantità di dati avvalendosi di tecniche di intelligenza artificiale. In particolare, i risultati dell'analisi mostrano la parte significativa di elementi non testuali ovvero di immagini che non possono essere esplorate con metodi tradizionali come l'analisi del contenuto, ma attraverso algoritmi di Machine Learning che ne hanno fornito la valutazione. La metodologia adottata ha consentito una comprensione più inclusiva delle opinioni dei consumatori sugli attributi alberghieri più importanti nel settore del

lusso, la cui importanza è in grado di influenzare la crescita e dare maggior risalto all'immagine del marchio alberghiero. Di conseguenza, lo studio estende il lavoro di Kim e Stepchenkova (2015) fornendo una pratica per riassumere quantitativamente e sistematicamente i contenuti manifesti delle immagini visive.

I consumatori, ad oggi, valutano più la qualità generale della stanza piuttosto che le strutture e le caratteristiche di intrattenimento o svago (ad esempio le piscine). Inoltre, sembra che i consumatori tendano a scattare un numero limitato di foto (compresi i *selfie*) negli hotel di lusso. La ricerca mostra inoltre che gli algoritmi di analisi dei Big Data, per valutare le foto generate dai consumatori, contribuiscono al miglioramento del monitoraggio dei social media nel contesto della gestione di hotel e del brand di lusso in quanto supportano la comprensione e l'importanza rivestita dai social media per scopi di *marketing* e *branding*. Le immagini rappresentano, la conservazione di un inventario visivo composto dagli elementi più attraenti per il visitatore (MacKay & Couldwell, 2004). Questo inventario può essere capitalizzato e potrebbe fungere da input per nuove strategie tra cui quella di spingere i manager a monitorare in modo più frequente e appropriato i siti web e le pagine disponibili sul web concedendo loro particolare attenzione per la gestione del marchio online. Includere contenuti visivamente più accattivanti oppure coinvolgere i consumatori attraverso i post pubblicati sui social media potrebbe rafforzare l'azienda in termini di fidelizzazione dei clienti e attrazione di nuovi consumatori.

Infine, la metodologia adottata potrebbe essere considerata come una tecnica per una "riduzione delle dimensioni" in termini di volume dei "Big data visivi", aprendo al contempo la strada verso l'analisi di questo tipo di dati, come mai fatto prima.

3.6 Cluster analysis e profilazione utenti

L'adozione di tecnologie altamente avanzate come le *Smart Grid* (SG) ha causato dei profondi cambiamenti in una vasta gamma di comportamenti relativi al settore energetico da parte dei clienti, compresa l'adozione di fonti energetiche sostenibili, di tecnologie ad alta efficienza energetica, di investimenti nei processi di efficienza energetica degli edifici e, soprattutto, di cambiamenti nel comportamento diretto e indiretto da parte dei clienti nel consumo di energia. Le reti intelligenti (SG) per il risparmio energetico sono ormai una realtà in quasi tutti i paesi del mondo. Queste infrastrutture generano un gran numero

di dati eterogenei che possono essere utilizzati anche per valutare il comportamento dei consumatori in merito al consumo energetico e fornire ai clienti finali un feedback più dettagliato sull'uso dell'elettricità. Tuttavia, come per tutti i dati di grandi dimensioni, anche i dati legati a questo settore, posso presentare dei problemi con l'archiviazione, l'elaborazione e l'analisi. Ecco perché sono stati sviluppati diversi metodi automatici di raccolta ed analisi e formulati algoritmi in grado di raggruppare i dati e di conseguenza i clienti in macro-categorie, in base ai loro modelli comportamentali. Negli ultimi anni, la ricerca scientifica ha considerato gli algoritmi di Machine Learning un metodo utile per analizzare i dati di consumo del carico elettrico (Tanwar et al, 2015), in particolare utilizzando l'analisi dei cluster, un valido metodo per esplorare le strutture all'interno dei dati. Nella letteratura, i metodi utilizzati per effettuare una buona analisi di cluster prevedono le seguenti fasi:

- i. Raccolta ed elaborazione dei dati che consentono di progredire con l'associazione di ciascun cliente al suo modello di carico rappresentativo, da prendere in considerazione ai fini della categorizzazione. I dati devono essere indicativi del comportamento di consumo del cliente raccolti su un modello di carico giornaliero. La durata deve essere abbastanza lunga da garantire la comparsa di schemi comportamentali. Pertanto, sono richieste non meno di due o tre settimane di osservazione nelle stesse condizioni di carico. Al termine della raccolta, i dati vengono elaborati ed eliminati quelli che presentano anomalie;
- ii. Definizione delle principali funzioni utilizzate come input per consentire l'analisi degli algoritmi utili al raggruppamento dei clienti, facendo una selezione delle caratteristiche più rappresentative di ciascun utente;
- iii. Valutazione delle procedure di *clustering* e adozione di algoritmi per eseguire l'analisi;
- iv. Classificazione dei clienti (fase post-clustering).

Nella presente ricerca, sono state applicate tecniche di *clustering* per profilare classi di utenti.

3.6.1 Obiettivo della ricerca

L'obiettivo della ricerca è stato quello di profilare classi di utenti. In particolare, l'approccio dominante è stato basato sull'applicazione di algoritmi di Machine Learning che hanno permesso di generare automaticamente gruppi di cluster al fine di visualizzare una rappresentazione concisa del comportamento dei clienti nei confronti del consumo e dell'uso di energia. In effetti, l'analisi dei cluster, basata su un processo di apprendimento non supervisionato, è una delle tecniche di *data mining* più comunemente applicate e utile per il partizionamento di un set di dati in sottogruppi che sono simili tra loro. Le tecniche di *data mining* si dividono in metodi di previsione e metodi di descrizione. I primi usano variabili conosciute per predirne altre sconosciute, mentre i secondi interpretano pattern assunti dai dati, che in questo caso, sono rappresentati dai clienti. L'analisi dei cluster si categorizza come un metodo di descrizione, poiché analizza le informazioni contenute nei dati e permette di individuare dei segmenti al loro interno. Individuare le informazioni relative al target a cui un prodotto si rivolge è decisivo per conoscere i desideri/bisogni/problemi dei clienti e soddisfarli. Così, attraverso lo studio della clientela si riescono, infatti, a valutare le caratteristiche e i comportamenti dei consumatori, personalizzando l'offerta, in un'ottica di fidelizzazione ed aumento delle quantità dei prodotti venduti. I clienti, infatti, non sono tutti uguali, ma presentano preferenze, aspettative, e comportamenti diversi. Un'efficace segmentazione che raggruppa clienti attraverso variabili che li rendono simili, permette di sviluppare strategie di marketing mirate e diversificate.

L'idea è stata quella di indagare quali sono gli utenti disponibili a far parte di una comunità energetica avvicinandosi al mercato delle energie rinnovabili e all'automazione.

3.6.2 Procedura e metodologia di analisi

Per raggiungere questo scopo, è stato implementato e somministrato ad un campione rappresentativo di utenti un sondaggio online al fine di ricavare informazioni sul consumo di energia delle famiglie. Il questionario online sviluppato con il format di Google Moduli è composto da 2 sezioni e da 22 domande. La prima sezione è stata dedicata alla raccolta di informazioni demografiche, la seconda sezione alla raccolta di dati relativi alle preferenze dell'utente in merito al consumo di energia. Dopo aver ottenuto le risposte i

dati sono stati riconfigurati al fine di renderli interoperabili e per una buona ed efficace costruzione del set di dati. Le risposte con input mancanti sono state eliminate, in questo caso, 4 questionari. Di conseguenza, sono stati predisposti ed analizzati 257 vettori (risposte). L'analisi dei dati ha previsto due fasi: (i) un'analisi descrittiva, per esaminare le caratteristiche demografiche, atteggiamenti e motivazioni dei consumatori correlate a variabili tradizionali come ad esempio la classe energetica adoperata; (ii) un'analisi dei cluster per identificare gruppi distinti aventi caratteristiche simili. Su ciascuna delle variabili demografiche: età, nucleo familiare e servizi online utilizzati è stata calcolata la media e la deviazione standard.

Per la misurazione dell'atteggiamento è stata adottata la scala Likert a cinque punti che ha permesso di rilevare il livello di accordo degli utenti rispetto tre fattori, corrispondenti le tre macro-aree costitutive del questionario:

- i. Propensione al risparmio energetico;
- ii. Disponibilità ad adottare energie rinnovabili;
- iii. Conoscenza dei sistemi domotici e di automazione.

La cluster analisi è stata eseguita sulle tre variabili demografiche: età, unità familiare e servizi online utilizzati, messe in relazione con le tre macro-aree di cui sopra.

Tecnicamente l'analisi dei cluster determina una riduzione della quantità totale di informazioni, tale da pervenire a gruppi omogenei. Ciascun cliente analizzato possiede n variabili, demografiche, comportamentali e/o affettive, che possono essere quantitative e qualitative e che rappresentano le informazioni che consentono il raggruppamento dei dati in cluster. Per raggruppare i dati in cluster è stata adottata la funzione *FindClusters* (Capitolo 2) che applica algoritmi di Machine Learning per selezionare e raggruppare elementi omogenei in un insieme di dati per ridurre lo spazio di ricerca e trovare soluzioni ottimali (Wolfram, 2017). Per impostazione predefinita, questa funzione identifica i cluster sulla base della distanza più breve, confrontando i dati senza acquisire alcuna informazione al riguardo. Il metodo adottato per la clusterizzazione del set di dati in input è Gaussian Mixture (Figura 3.12).



Figura 3.12. Informazione sul metodo adottato per la clusterizzazione ottenuta con l'applicazione di ClassifierFunction

Inoltre, per quanto riguarda la distanza, Mathematica ha raggruppato i dati secondo una funzione di somiglianza. La funzione di somiglianza applicata automaticamente da Mathematica è la distanza euclidea, che raggruppa automaticamente gli elementi in gruppi in base alla loro vicinanza nello spazio.

Come precedentemente eseguito da Frank e colleghi (2019) è stata condotta un'ulteriore analisi statistica sui risultati relativi ai cluster per comprendere la relazione dei gruppi di consumatori (allocati nei diversi cluster) con le differenti tre macro-aree (propensione al risparmio energetico; disponibilità ad adottare energie rinnovabili e conoscenza della domotica e dei sistemi di automazione) e fornire infine le informazioni riguardo i profili.

3.6.3 Risultati

I risultati della prima fase, ovvero dell'analisi statistica, mostrano i dati di un campione composto da 257 risposte. I soggetti, ugualmente composti da uomini e donne, hanno prevalentemente un'età compresa tra 18 e 34 anni. Il 61% vive in un appartamento, mentre il 39% in una casa indipendente. Il 42% di essi ha una classe energetica A. Usano il web per svolgere numerose attività, dall'acquisto di pacchetti di viaggio e vacanze all'utilizzo dei servizi online. Tuttavia, nella maggior parte dei casi, più della metà (59%) utilizza i social network. La casa è per lo più riscaldata da una caldaia autonoma, infatti emerge una percentuale molto alta (84%) ed utilizza gas per generare acqua calda (79%). Le caratteristiche demografiche del campione sono riassunte nella tabella seguente.

Tabella 5. Caratteristiche demografiche del campione

Caratteristiche	Campione (n = 257)
Genere	
Maschi	43%
Femmine	57%
Età	
18 – 24	38%
25 – 34	39%
35 – 44	10%
Superiore 45	13%
Composizione del Nucleo Familiare	
1 componente	20%
2 componenti	11%
3 componenti	28%
4 componenti	26%
5 componenti	13%
Abitazione	
Abitazione indipendente	39%
Appartamento	61%
Riscaldamento abitazione	
Caldaia condominiale	8%
Pannelli solari	5%
Caldaia autonoma	84%
Teleriscaldamento	3%
Sistema di raffreddamento estivo	
Termoconvettore	7%
Condizionatore fisso	38%
Condizionatore Mobile	14%
Altro/Nessuno	40%

Caratteristiche	Campione (n = 257)
Riscaldamento acqua	
Scaldabagno elettrico	16%
Gas	79%
Teleriscaldamento condominiale	5%
Sistema di distribuzione del riscaldamento	
Unità di riscaldamento	82%
Riscaldamento a pavimento	5%
Termoconvettore	3%
Altro/Nessuno	10%
Classe energetica dell'abitazione	
A+	29%
A	42%
B	15%
Altro	14%
Servizi Online utilizzati	
Acquisti Online	26%
Social network	59%
Pagamento utenze	4%
Ricerca di pacchetti vacanza, voli aerei, viaggi in treno	10%

Per quanto riguarda l'analisi dei cluster i risultati ottenuti sono omogenei. Infatti, sono emersi 2 o 3 cluster principali e solo in un caso il raggruppamento dei cluster è uguale ad 11. Tuttavia, anche in quest'ultimo caso il maggior numero di elementi è raggruppato nei primi 3 segmenti.

I risultati relativi alla Media e alla Deviazione Standard (SD) in relazione all'analisi dei cluster sono mostrati nella Tabella seguente.

Tabella 6. Statistiche descrittive, Media e SD effettuate sul campione

	Statistiche descrittive	
	Media sul valore della scala Likert	Deviazione Standard
Propensione al risparmio energetico e opinione sull'attuale fornitore di energia		
<i>Sono disponibile a utilizzare solo elettrodomestici di classe A ++</i>	3.91	1.29
<i>Il mio attuale fornitore di energia è affidabile</i>	3.39	1.06
<i>Penso che sia importante avere trasparenza e chiarezza nella fattura da parte del fornitore di energia elettrica</i>	4.33	1.14
<i>Penso che le offerte dei fornitori di energia attualmente sul mercato siano economiche</i>	2.86	1.01
Disponibilità ad adottare energie rinnovabili		
<i>Sono disponibile a utilizzare fonti di energia rinnovabili e / o prodotti / tecnologie in grado di risparmiare energia</i>	4.19	1.18
<i>Sono d'accordo se nella mia città verrà prodotta più energia da fonti rinnovabili</i>	4.32	1.16
<i>Accetto di pagare un extra per utilizzare l'energia prodotta da fonti rinnovabili</i>	3.00	1.26
Conoscenze relative ai sistemi domotici e di automazione		
<i>La domotica è un valido aiuto per anziani o disabili</i>	3.70	1.15
<i>I sistemi di automazione domestica consentono il risparmio energetico e quindi riducono consumi e sprechi</i>	3.65	1.08
<i>I sistemi di automazione domestica sono in grado di rendere la casa più sicura</i>	3.73	1.10

Un segmento (133 intervistati) appartiene a una fascia di età compresa tra 18 e 24 anni, consapevole che la domotica non sia un valido aiuto solo per gli anziani o i disabili, ma può contribuire al risparmio energetico e quindi ridurre consumi e sprechi e rendere la casa più sicura. Una forte consapevolezza per quanto riguarda la conoscenza dei sistemi domotici e di automazione è stata osservata nella fascia d'età tra i 25 e 34 anni, mentre gli utenti appartenenti a una fascia di età superiore ai 45 anni non mostrano una chiara conoscenza di questi sistemi. La maggior parte degli intervistati ha risposto di essere d'accordo o fortemente d'accordo (il punteggio medio è 4,32) con l'adozione nella propria città di tecnologie atte alla produzione di energia da fonti rinnovabili.

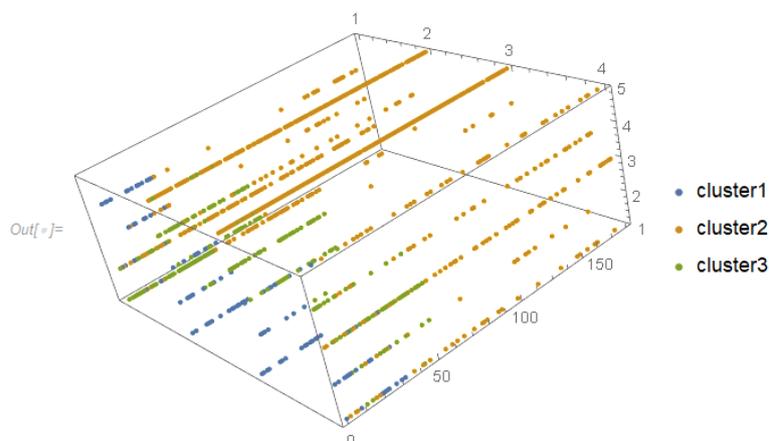


Figura 3.13. ListPlot del cluster generato correlando la variabile “età” alla macro-area “disponibilità ad adottare energie rinnovabili”

Infine sulla base dell’analisi dei cluster e delle caratteristiche demografiche del campione, sono emersi i seguenti profili:

1. *Profilo 1* con età compresa tra 18 e 24 anni, unità familiare composta da 3 persone, utilizza sia i servizi di e-commerce che i social network. Questi soggetti sono propensi ad adottare tecnologie utili al risparmio energetico e all’impiego di fonti rinnovabili.
2. *Profilo 2* con fascia d’età compresa tra 25 e 34 anni, unità familiare composta da 3 persone, concorda fortemente sul risparmio energetico e sull’adozione di tecnologie per l’energia rinnovabile. Ritengono che sia importante avere chiarezza e trasparenza sul tema fatturare e sono disposti a utilizzare gli elettrodomestici A++. Tuttavia, considerano il mercato dell’energia di oggi a buon mercato.
3. *Profilo 3* con età compresa tra 35 e 44 anni, unità familiare composta da 3 persone, non è d’accordo con la produzione e l’uso di fonti di energia rinnovabile e considera il loro fornitore di energia inaffidabile.

3.6.4 Discussioni

Dall’analisi dei dati emerge che la maggior parte degli intervistati quotidianamente mette in pratica comportamenti “virtuosi” per risparmiare, risultando sensibile al costo

dell'energia. Molti sono gli utenti disposti a modificare le proprie abitudini ed adottare uno stile di vita sostenibile per ridurre gli sprechi e abbattere i costi, per esempio attraverso l'acquisto di prodotti efficienti dal punto di vista energetico. Ciò fa emergere quanto i consumatori siano aperti nei confronti di uno stile di vita sostenibile che si proietti verso sistemi *smart*. Bisognerebbe, dunque, migliorare le conoscenze sul mercato dell'energia attraverso più chiare informazioni e migliorare la sensibilizzazione degli operatori economici riguardo gli interessi dei consumatori. Inoltre, bisognerebbe promuovere in modo efficiente i diritti degli utenti attraverso il rafforzamento della collaborazione tra gli organismi nazionali competenti ed i servizi di consulenza offerti ai consumatori. Questo permetterebbe la possibilità di contribuire ad un migliore funzionamento del mercato, aumentando la fiducia del consumatore e riducendo la sua avversione nei confronti dei rapporti commerciali. Una comunicazione fondata sull'utilità delle soluzioni di efficienza energetica che, ad oggi, sono disponibili sul mercato può essere un valido elemento sul quale si potrebbe giocare la futura competitività delle offerte sul mercato riguardo l'energia, i sistemi domotici e le sue componenti.

Quindi, dividere in categorie significa considerare, in un gruppo di individui, caratteristiche comuni con il gruppo a cui appartiene.

In questa ricerca l'applicazione dell'analisi dei cluster ed in particolare, l'utilizzo di algoritmi di apprendimento automatico hanno consentito di rilevare somiglianze tra i set di dati raccolti dal questionario al fine di identificare modelli distinti in grado di descrivere il profilo di comportamento dei clienti. Questa analisi si presenta dunque utile per definire il profilo per ciascuna classe di clienti e per il calcolo delle informazioni globali sulla potenza e sull'energia per le classi di clienti al fine di adattare le tariffe di fatturazione. Le informazioni ottenute possono rivelarsi utili per promuovere specifiche attività di marketing, gestire al meglio le esigenze dei clienti e formulare efficaci strategie aziendali identificando i punti di forza e di debolezza di un'azienda. Ad esempio, la pubblicità potrebbe essere mirata a convincere i consumatori che non sono disposti a pagare un extra per adottare energia rinnovabile. Sarà utile evidenziare che l'esborso iniziale consentirà di ottenere vantaggi economici in un secondo momento.

CAPITOLO 4

BIG DATA ANALYTICS: NUOVE SKILLS NELLA FORMAZIONE

4.1 Big Data, IoT ed Industria 4.0

Il processo di globalizzazione ha prodotto profondi cambiamenti in molti campi valorizzando l'imprenditorialità e l'innovazione. Parallelamente, nelle aziende manifatturiere è aumentata la necessità di incorporare nell'organico lavorativo nuove figure in grado di gestire l'intera catena del valore¹ in modo agile e reattivo. In generale, le aziende necessitano di figure virtuali e fisiche che consentano una stretta cooperazione ed un rapido adattamento lungo il ciclo di vita del prodotto, dall'innovazione alla produzione e distribuzione.

Le recenti iniziative governative ed industriali hanno delineato uno scenario in cui i progressi nel campo dell'informatica e della tecnologia saranno sfruttati per migliorare le attività commerciali e produttive. Le trasformazioni generali causate dall'integrazione digitale e dall'ingegneria intelligente, fondono il mondo virtuale e quello reale con enfasi su applicazioni ingegneristiche come robotica, digitalizzazione ed automazione. Questo *framework* definisce la quarta rivoluzione industriale, comunemente denominata Industry 4.0 (I4.0). Industria 4.0 è il termine usato per etichettare la Quarta Rivoluzione industriale che riassume una serie di innovazioni e cambiamenti tra cui l'emergere di nuovi ruoli nell'intero processo di lavoro, miglioramenti tecnologici, la necessità di figure lavorative con nuove competenze, diventando così volano di sviluppo nell'attuale panorama

¹ Modello *value chain* teorizzato da Porter (1985)

industriale (Pereira & Romero, 2017). Il concetto è apparso per la prima volta durante la fiera di Hannover in Germania nel 2011, al fine di promuovere un approccio strategico alla digitalizzazione nel campo produttivo. In effetti, una delle caratteristiche chiave dell'I4.0 è la creazione di industrie altamente automatizzate basata su una forte interazione uomo-macchina (Xu et al., 2018). L'I4.0 rappresenta uno stato industriale caratterizzato da un'accurata digitalizzazione dei flussi economici e produttivi. Conseguentemente, la necessità dell'Industria 4.0 è quella di convertire le macchine ordinarie in macchine autocoscienti, capaci di autoapprendimento per migliorare le loro prestazioni complessive e sviluppare una piattaforma di produzione aperta ed intelligente. I progressi nella tecnologia moderna prodotti nell'ambito del *framework* dell'I4.0 sono raggruppabili nei seguenti nove pilastri: Big data e Analytics, Robot autonomi, Simulazione 2D e 3D, Integrazione di sistemi orizzontali e verticali, Internet of Things, Cyber Security e Cyber Physical Systems (CPS), Cloud, Additive Manufacturing e Realtà Aumentata, di seguito descritti in modo dettagliato:

- i. *Big data e Analytics*, indicano la generazione di nuovi dati e l'analisi approfondita di dati provenienti da macchine e processi differenti utili per aumentare la produttività e la competitività delle imprese;
- ii. *Robot autonomi*, impiegati per eseguire metodi di produzione autonomi; inoltre se programmati in modo preciso e *smart* possono occuparsi di sicurezza, flessibilità, versatilità e collaborazione;
- iii. *Simulazione 2D e 3D*, utilizzate per sfruttare i dati in tempo reale e per rappresentare il mondo fisico in un modello virtuale di cui fanno parte persone, macchine e prodotti;
- iv. *Integrazione di sistemi orizzontali e verticali*, delineato da tre dimensioni: (i) integrazione orizzontale che implica comunicazione con soggetti esterni all'azienda, ad esempio i fornitori o i distributori, (ii) integrazione verticale, che implica la comunicazione e condivisione di informazioni all'interno dell'azienda in maniera trasversale rispetto alla struttura gerarchica della stessa e (iii) sistemi di produzione in rete basati su progettazione end-to-end per l'intero ciclo di vita del prodotto;
- v. *Internet of Things*, ovvero una rete di oggetti interconnessi che comunicano attraverso protocolli standard. Nell'Internet of Things sono integrati oggetti fisici, fattori umani, macchine intelligenti, sensori intelligenti, processo di

- produzione e linee di produzione attraverso i confini dell'organizzazione;
- vi. *Cyber Security e Cyber Physical Systems (CPS)*, definito come un sistema in cui i sistemi naturali e umani (spazio fisico) sono strettamente integrati con i sistemi di calcolo, comunicazione e controllo (spazio cibernetico). La forte connessione del mondo fisico col mondo digitale può migliorare la qualità delle informazioni richieste per la pianificazione, l'ottimizzazione ed il funzionamento dei sistemi di produzione. A causa dell'enorme quantità di tracce che i consumatori lasciano nei sistemi informatici fisici, la sicurezza dei dati in tutti i tipi di transazioni diventa un elemento chiave che consente ai consumatori ad accedere in sicurezza ai prodotti e ai servizi di rete;
 - vii. *Cloud*, fa riferimento alla necessità delle organizzazioni di una maggiore condivisione dei dati tra i siti e le aziende, vale a dire raggiungere i tempi di reazione in millisecondi o anche più velocemente. Il continuo scambio di dati viene effettuato collegando i sistemi cyber fisici in modo intelligente con l'aiuto dei sistemi cloud in tempo reale;
 - viii. *Additive Manufacturing*, ovvero i metodi di fabbricazione e produzione additiva saranno ampiamente utilizzati per produrre piccoli lotti di prodotti personalizzati che offrono vantaggi di costruzione, come ad esempio design complessi e leggeri;
 - ix. *Realtà Aumentata*, intesa come insieme di sistemi che supportano una varietà di servizi e che possono essere utilizzati dall'industria per fornire ai lavoratori informazioni in tempo reale al fine di migliorare il processo decisionale e le procedure di lavoro. I lavoratori possono ricevere istruzioni di riparazione su come sostituire una particolare parte mentre stanno guardando il sistema reale che necessita di riparazione (Vaidya et al., 2018).

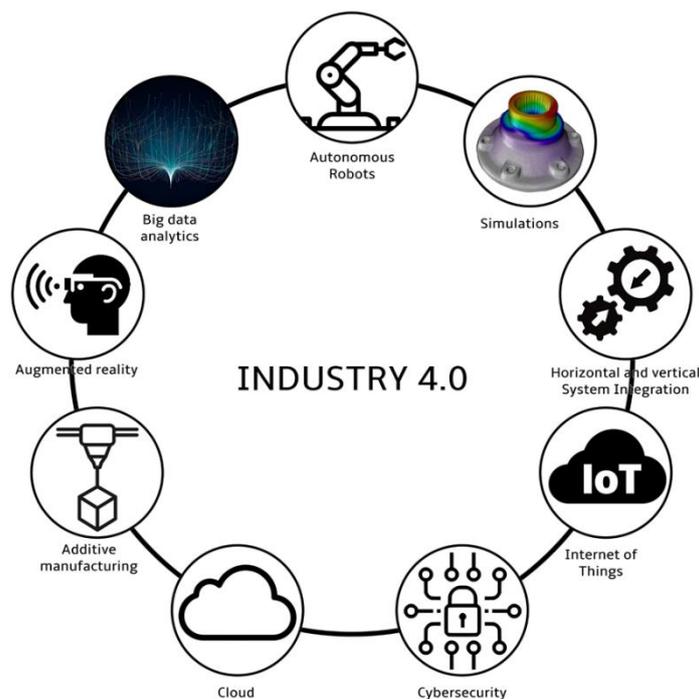


Figura 4.1. Le nove tecnologie che hanno trasformato la produzione industriale

Il paradigma I4.0 si basa sulla comunicazione e sulla cooperazione con i sistemi cyberfisici (CPS) e con le persone in tempo reale attraverso l'Internet of Things (IoT), al fine di migliorare le prestazioni in termini di produttività, efficienza delle risorse, costi e sicurezza. In altre parole, l'obiettivo del CPS è fornire in tempo reale l'acquisizione e la gestione di dati di consistente volume (Big Data) dal mondo fisico ed informazioni dallo spazio informatico in modo da consentire la gestione e l'analisi "intelligente" dei dati (Xu & Duan, 2019). L'effettivo trattamento dei Big data in informazioni utili è la chiave dell'innovazione sostenibile all'interno di I4.0. Molte ricerche discutono degli aspetti tecnici dell'I4.0 non prestando attenzione agli approcci manageriali e alla cultura organizzativa, aspetti significativi, che possono influenzare e condizionare il successo dell'attivazione di questo concetto.

Come già discusso nei capitoli precedenti, nell'attuale contesto economico e competitivo, per migliorare la produttività, le aziende si trovano ad affrontare sfide importanti che riguardano l'analisi dei Big Data. Tuttavia, uno dei problemi ancora aperti riguarda proprio la formazione degli specialisti per l'analisi di tali dati che in alcuni casi si rivela non soddisfacente (Deng & Cao, 2018). La sfida è avere linee guida pratiche, quadri teorici di riferimento ed approfondimenti su come i Big data possano davvero creare

valore per un'azienda e indicazioni su come integrare le risorse umane in questo scenario. A tal proposito, lo sviluppo di applicazioni IoT ha posto le basi per l'implementazione di un ambiente *smart* utile al miglioramento delle attività di formazione ed apprendimento. Attualmente, l'istruzione e l'industria stanno investendo molto nello sviluppo di sistemi ed infrastrutture IoT. Poiché, l'IoT è considerato sempre più una priorità nelle strategie nazionali in materia di ICT (Tecnologie dell'informazione e della Comunicazione), i nuovi piani per lo sviluppo di ambienti informatici avanzati per la gestione dei Big data creati dai dispositivi IoT sono più importanti che mai per il futuro del processo decisionale (Gubbi et al., 2013). Big data, Internet of Things ed Industry 4.0 rappresentano i pilastri delle aziende del futuro, della loro trasformazione ed innovazione.

Al fine di tradurre questi concetti in supporto concreto, per migliorare le attività aziendali, è necessario attuare un processo di formazione che miri all'acquisizione di conoscenze e competenze direttamente collegate al settore dell'Industria 4.0.

4.2 Valorizzazione della ricerca

Le tecnologie dell'I4.0 sono comunemente integrate nel nostro mondo accompagnandoci nella nostra quotidianità, sia in ambito lavorativo che in ambito ricreativo. Esse sono facilitatori, spesso permettono risparmio di tempo e denaro; ci supportano, cambiando la nostra percezione del mondo ed il nostro comportamento. La quarta rivoluzione industriale sta enormemente estendendo il potere della digitalizzazione creando così un nuovo mondo. In quanto esseri umani viviamo in un mondo fisico, ma trascorriamo parte del nostro tempo in mondi tecnologici virtuali, ossia su sistemi cibernetici (CPS), inconsapevoli del potenziale di trasformazione che queste tecnologie stanno portando con sé e i cambiamenti socio-economici che stanno prefigurando. Gli effetti di questa rivoluzione sono molto più invasivi di quelli precedenti: la tecnologia si incarna negli strumenti che usiamo e persino nel nostro corpo. Oggigiorno, possiamo leggere libri, guardare film su richiesta, ascoltare musica, prenotare servizi facilmente e con tempistiche ridotte, sempre e ovunque nel mondo. I vantaggi di queste tecnologie sono innumerevoli. Tuttavia, le disparità diventeranno sempre più evidenti, per le persone, per i prodotti, per le industrie e per i sistemi di formazione, consentendo l'emergere di piattaforme che diventeranno dominanti in questa era tecnologica. Conseguentemente, la mancanza di lavoro aumenterà, a causa dell'automazione e dell'uso di robot industriali.

Questo fenomeno ridurrà il bisogno della presenza di soggetti umani nei settori industriali, più inclini al progresso tecnologico. Inoltre, la specializzazione del lavoro crescerà, poiché saranno richiesti livelli di istruzione superiore e studi specializzati. Tuttavia, sarà anche necessario aumentare competenze e flessibilità dei lavoratori: le capacità di *problem solving*, le capacità sociali e tecnologiche saranno i processi più influenti richiesti in questi nuovi luoghi di lavoro (The Future of Job Report). Pertanto, nella prospettiva I4.0, i metodi vecchi e obsoleti necessitano di una profonda revisione e aggiornamento in termini di conoscenze, esperienze ed abilità. Questa sfida coinvolge molti settori, dai processi di produzione, ai modelli di business, ai sistemi educativi.

In questo scenario, il settore dell'*education* gioca un ruolo fondamentale nella formazione delle nuove generazioni verso l'uso delle tecnologie I4.0.

In una prospettiva futura, in ambito formativo sarà necessario dapprima effettuare un'attenta analisi delle dinamiche lavorative e delle funzionalità che stanno cambiando; conseguentemente sarà necessario ridefinire i profili lavorativi con funzioni e competenze specifiche da conseguire, appunto, durante la formazione. Di conseguenza, le istituzioni accademiche dovrebbero svolgere un ruolo chiave in quanto poli tecnologici responsabili del collegamento tra accademia, industria, società e territorio (Baena et al., 2017).

4.3 Un esperimento didattico

L'evoluzione di nuove e sofisticate tecnologie sta cambiando il quadro tradizionale anche nell'industria del turismo. Secondo Sigala (2018) i gestori turistici possono utilizzare le tecnologie come strumenti utili per creare nuove opportunità sia per i fornitori che per i consumatori. Infatti, seppur l'adozione di strumenti innovativi di gestione ed analisi permette di comprendere meglio il mercato turistico, le aziende spesso applicano piani inappropriati di analisi dei dati causando, in alcuni casi, un ritorno negativo sugli investimenti (Verhoef et al., 2016). L'aggiunta di esperti professionisti nei vari settori della conoscenza tecnologica per l'industria turistica è necessaria nei processi di erogazione di servizi (Van Doorn et al., 2017). Come suggerisce la letteratura, il turismo è riconosciuto come un "settore pilastro", ovvero un settore strategico che può realmente contribuire alla crescita della ricchezza generale di un paese.

Nel *framework* I4.0 i nuovi sistemi organizzativi stanno crescendo definendo quello che alcuni autori chiamano "turismo intelligente" (Gretzel et al., 2015) o "destinazioni del

turismo intelligente” (Lamsfus et al., 2015). Una destinazione intelligente, accessibile a tutti, favorisce lo sviluppo sostenibile delle aree turistiche attraverso l’adozione di tecnologie ICT nell’ambiente. Questo ecosistema tecnologico e funzionale supporta l’interazione e l’integrazione dei clienti con, e nell’ambiente, migliorando così anche la qualità della vita dei residenti. Tali sistemi non sono stati sviluppati solo in diverse parti del mondo, ma è stato creato un movimento sociale di condivisione delle esperienze turistiche *smart* sui social media.

Le sfide più rilevanti potrebbero essere correlate sia all’analisi delle principali problematiche affrontate dall’industria del turismo, sia allo studio di come formare professionisti che lavoreranno nel settore del turismo e dell’ospitalità (Balasubramanian & Ragavan, *in press*) i quali dovranno essere formati su tutti gli aspetti delle tecnologie proprie dell’I4.0. Gli operatori turistici dovranno adottare un approccio proattivo con il consumatore e dovranno essere consapevoli delle esigenze di ogni utente. L’adozione di nuovi ed innovativi strumenti e la creazione di partenariati tra accademia e aziende potrebbe essere fondamentale per favorire il miglioramento delle competenze trasversali da parte dell’industria turistica fornendo conoscenze adeguate in materia di tecnologia ed innovazione (Chen et al., 2012). Pertanto, le organizzazioni, in particolare i centri di formazione, dovrebbero diffondere la cultura I4.0 e contribuire allo sviluppo delle competenze I4.0 attraverso corsi universitari dedicati.

4.3.1 Obiettivo

La ricerca educativa evidenzia la crescente necessità di preparare i laureati nel settore del turismo a numerose attività in questo contesto, comprese le capacità imprenditoriali e tecniche. È sulla base di questa premessa che è nata l’idea di condurre un esperimento didattico con l’obiettivo di fornire ai laureati che opereranno nel settore del turismo specifiche *skills* imprenditoriali e tecniche proprie del *framework* Industry 4.0.

I punti chiave del programma progettato sono stati:

- i. promuovere lo sviluppo delle competenze trasversali degli studenti;
- ii. consentire agli studenti di affrontare problemi di lavoro reali;
- iii. colmare il divario tra università e impresa;
- iv. creare un *curriculum* formativo-professionale ritenuto coerente con la

tendenza tecnologica dell'I4.0.

4.3.2 Procedura

Il programma di lavoro è stato avviato nell'anno accademico 2017/2018, durante l'insegnamento di Psicologia generale tenuto per gli studenti del corso di Laurea in Scienze Turistiche presso l'Università della Calabria. Il corso ha coinvolto 30 studenti (27% maschi e 73% femmine) ed il programma è stato concepito per favorire l'acquisizione di competenze "hard", in termini di conoscenze tecniche ed abilità di programmazione, e "soft", in termini di empatia ed organizzazione, da applicare nell'attività lavorativa al fine di aumentare la gamma di opportunità nel mercato del lavoro. Il corso ha previsto due fasi di lavoro: una prima fase teorica ed una seconda fase pratica. Nella prima fase sono state organizzate delle lezioni preliminari su argomenti propri dell'I4.0 quali Big data e analisi dei Big data, Machine Learning, Realtà Virtuale e Aumentata, Simulazione, Internet of Things e Robotica, tali da consentire agli studenti di avere un quadro multidimensionale basato su nuove conoscenze utili alle imprese turistiche. In particolare, gli studenti sono stati orientati su due degli argomenti sopra menzionati: analisi dei Big data e uso dell'Intelligenza Artificiale.

La seconda fase ha previsto invece l'esecuzione di "esercizi pratici", nello specifico agli studenti sono stati forniti strumenti di elaborazione ed analisi dei dati al fine di aumentare le loro *hard* e *soft* skills utili e spendibili nel mercato del lavoro.

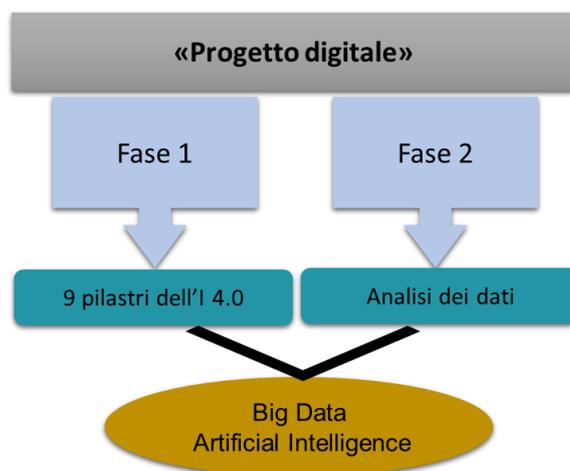


Figura 4.2. Framework relativo all'organizzazione del corso

In particolare, in questa fase, gli studenti hanno scelto una località turistica, localizzata nella regione Calabria ed hanno ricavato informazioni utili per la realizzazione di un “progetto digitale”. Gli studenti hanno seguito una procedura specifica e dettagliata per la raccolta e l’analisi dei dati. Innanzitutto, hanno ricavato informazioni specifiche sulla località scelta, ovvero informazioni geografiche, storico-culturali, territoriali e demografiche identificando sia i punti di interesse naturali e culturali del luogo che le offerte turistiche in quel momento disponibili. Successivamente hanno imparato ad acquisire dati da piattaforme di social network al fine di effettuare un’analisi su testi (*sentiment analysis*) e su immagini (*object recognition*). Le piattaforme consultate sono state Flickr e Twitter e le analisi condotte sui dati raccolti hanno consentito di estrarre informazioni relative alle attrattività della località scelta (in base alle foto scattate) e hanno permesso di indagare sul comportamento dei consumatori e sulle loro emozioni (in base ai tweet).

Per l’estrazione e l’analisi dei dati è stato applicato il software *Wolfram Mathematica*. Per la raccolta delle immagini è stato necessario lo sviluppo di un codice che ha permesso dapprima la connessione all’API di Flickr e conseguentemente la possibilità di ricavare informazioni relative alla destinazione scelta tramite l’indicazione delle coordinate geografiche.

```
i = ServiceExecute["Flickr", "ImageSearch", {"Keywords" -> {█}, MaxItems -> 50]
```

Thumbnail	Keys				Owner	Title
	ID	Farm	Server	Secret		
	29501266525	9	8482	821a1166fc	75924821@N03	█
	28665101584	9	8099	8afd7d4864	120524337@N03	waterfall
	27167985370	8	7246	906007eb6a	96275607@N02	█

Figura 4.3. Metadati ricavati da Flickr tramite Wolfram Mathematica

Per quanto riguarda il testo, il software ha permesso di raccogliere tweet tramite una *query* e la connessione alla piattaforma Twitter. Nello specifico ogni studente applicando un determinato *hashtag*, relativo alla location selezionata, ha ottenuto il testo dei tweet e i

corrispondenti metadati. Dopo aver ottenuto testo e immagini è stata eseguita l'analisi dei dati.

Infine, le risorse ottenute sono state raggruppate in tre categorie principali: variabili fisiche, variabili geografiche e variabili comportamentali. Queste procedure sono state necessarie per sviluppare il lavoro. Infatti, alla fine del corso, ogni studente ha sviluppato un progetto sfruttando attributi propri dell'I4.0. Agli studenti è stata data la possibilità di scegliere il progetto digitale da realizzare: brochure digitali, un sito Web, uno spot, un blog, progetti attraverso i quali la località selezionata avrebbe ottenuto valorizzazione e pubblicità.

Al fine di misurare il livello di soddisfazione in relazione alle nuove direttive del corso, è stato ideato e somministrato ai partecipanti, un breve questionario. Il questionario è stato sviluppato utilizzando Google Moduli ed inviato via email ai 30 studenti che hanno partecipato al corso. Il test, composto da 12 item a risposta chiusa, costruite su una scala Likert a 5 punti, dove 1 equivale a *disaccordo totale* mentre 5 ad *accordo totale*, è stato suddiviso in due sezioni, nella prima sezione le domande erano inerenti l'organizzazione del corso e i risultati attesi dalla sperimentazione, mentre la seconda sezione riguardava la valutazione del docente. Le 7 domande della prima sezione prevedevano per esempio di capire se *“il corso fornisce utili strumenti tecnologici per l'analisi del comportamento del turista”*, oppure se *“gli obiettivi del corso sono stati raggiunti in breve tempo”*. La seconda sezione era composta da 5 domande tra cui *“il docente ha assunto un atteggiamento motivante utile a far apprendere agli studenti la materia”*. Ai partecipanti è stato chiesto di rispondere alle domande, scegliendo una delle risposte che più rispecchiava la propria opinione. In tal modo è stato possibile calcolare il livello di gradimento in riferimento a ciascun item. I dati raccolti sono stati testati statisticamente utilizzando SPSS.

4.3.3 Risultati

Lavorando in gruppi di due, gli studenti hanno sviluppato quindici progetti digitali relativi a quindici località calabresi. Nello specifico, essi hanno realizzato spot, siti web, blog, brochure digitali e video seguendo la procedura suggerita durante il corso.

Grazie alle immagini raccolte, gli studenti hanno ricavato informazioni sulla posizione dei turisti identificando i punti di interesse e grazie alla posizione hanno definito i percorsi

dei turisti nella località (Figura 4.4); sono state realizzate mappe ed in base alla data di pubblicazione dell'immagine anche dei diagrammi con immagini raffiguranti eventi religiosi o culturali, eventi quotidiani, monumenti, piazze, chiese e cibo.

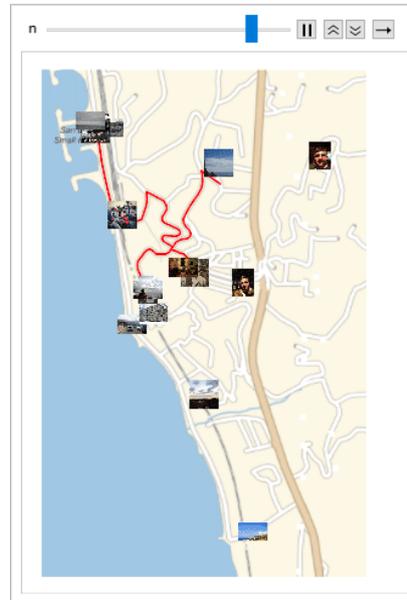


Figura 4.4. Percorso di un turista ottenuto combinando le coordinate geografiche e le immagini ottenute da Flickr

Il recupero e l'analisi di queste informazioni possono quindi aiutare in modo significativo l'auto-miglioramento degli esperti. Inoltre, gli approcci all'analisi del sentimento riguardo al testo dei social media si concentrano principalmente su un punto di vista psicologico necessario per riconoscere le emozioni e i fenomeni sociali che potrebbero essere indicatori rilevanti nella valutazione o nella previsione degli atteggiamenti e dei comportamenti delle comunità online (Miháلتz et al., 2015).

Inoltre, il questionario ha permesso di calcolare il livello di soddisfazione degli studenti in relazione alla qualità delle attività e agli strumenti tecnologici forniti durante il corso. I dati raccolti dai 30 questionari sono stati analizzati con il software SPSS 22.0. In particolare sono state calcolate la media, la deviazione standard (Tabella 7) e la frequenza (Figura 4.5).

Tabella 7. Analisi descrittiva delle risposte alle domande del questionario ideato per misurare la soddisfazione dei partecipanti al corso.

	Media	Deviazione standard
Attività Face-to-face e strumenti tecnologici forniti		
<i>Gli esami riflettono il materiale del corso</i>	4,30	,877
<i>Gli obiettivi specifici di questo corso sono stati raggiunti in breve tempo</i>	4,40	,621
<i>L'atmosfera in aula era positiva</i>	4,40	,621
<i>Sono contento di aver seguito questo corso</i>	3,93	,640
<i>Le aspettative per questo corso non sono state raggiunte</i>	1,77	,774
<i>Il corso fornisce utili strumenti tecnologici per analizzare il comportamento turistico</i>	4,43	,728
<i>Le tecnologie e i software appresi sono rilevanti per studiare le dinamiche turistiche</i>	4,50	,682
Soddisfazione delle aspettative degli studenti		
<i>Le mie aspettative durante il completamento di questo corso non sono state soddisfatte</i>	1,77	,774
<i>Molto ben informato sull'argomento</i>	4,50	,572
<i>Ben preparato per questo corso</i>	1,33	,606
<i>Ha assunto un atteggiamento motivante utile a far apprendere agli studenti la materia</i>	4,50	,777
<i>Nel complesso preparato</i>	4,43	,626

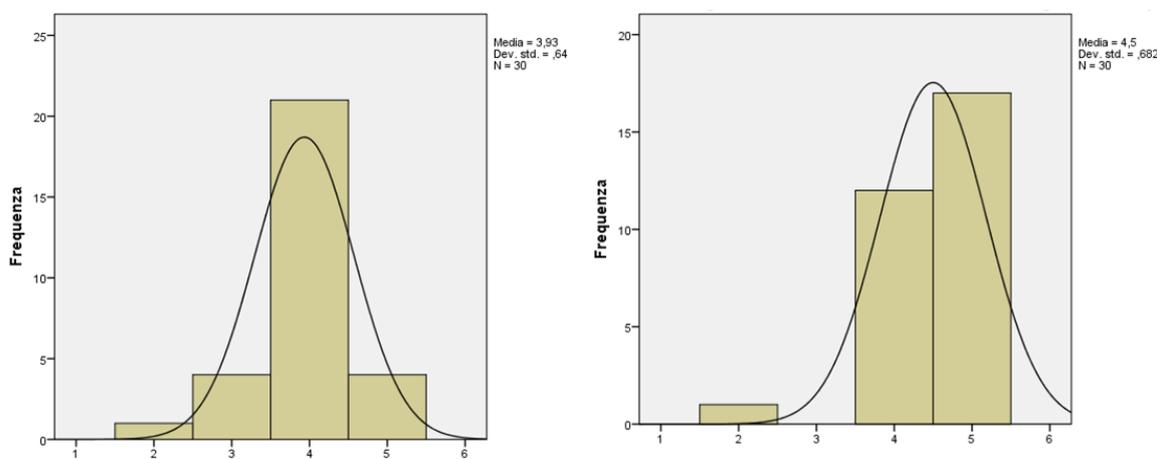


Figura 4.5. Analisi delle frequenze delle risposte alle domande del questionario: sono contento di aver seguito questo corso (a) e le tecnologie e i software appresi sono rilevanti per studiare le dinamiche turistiche (b)

I risultati dell'analisi mostrano che il 97% degli studenti ha apprezzato i contenuti educativi relativi all'applicazione delle nuove tecnologie nel turismo, soprattutto per le potenzialità nel modellare al meglio le attività professionali dei futuri operatori turistici, adattando i contenuti alle nuove esigenze tecnologiche e ai consumatori contemporanei.

Inoltre, secondo il 93% degli studenti, un ruolo chiave ha avuto il clima della classe, ovvero gli aspetti emotivi e fisici della classe e la capacità del docente di motivare gli studenti ad avvicinarsi a nuovi strumenti scientifici. L'83% degli studenti si è ritenuto contento e soddisfatto della partecipazione al corso.

Tabella 8. Percentuali dei dati raccolti

Attività Face-to-face e strumenti tecnologici forniti	Negativo	Positivo
<i>Gli esami riflettono il materiale del corso</i>		93%
<i>Gli obiettivi specifici di questo corso sono stati raggiunti in breve tempo</i>		90%
<i>L'atmosfera in aula era positiva</i>		93%
<i>Sono contento di aver seguito questo corso</i>	-3%	83%
<i>Le aspettative per questo corso non sono state raggiunte</i>	-80%	
<i>Il corso fornisce utili strumenti tecnologici per analizzare il comportamento turistico</i>		87%
<i>Le tecnologie e i software appresi sono rilevanti per studiare le dinamiche turistiche</i>	-3%	97%
Soddisfazione delle aspettative degli studenti		
<i>Le mie aspettative durante il completamento di questo corso non sono state soddisfatte</i>	- 93%	
<i>Molto ben informato sull'argomento</i>		97%
<i>Il docente non era ben preparato per questo corso</i>	- 93%	
<i>Il docente ha assunto un atteggiamento motivante utile a far apprendere agli studenti la materia</i>	-3%	90%
<i>Nel complesso buono</i>		93%

4.3.4 Discussioni

L'esperimento ha previsto l'acquisizione pratica di metodi e strumenti per ottenere e analizzare i Big data attraverso l'utilizzo di tecniche di intelligenza artificiale quali il Machine Learning. Nello specifico, gli studenti hanno analizzato testi ed immagini estratti dai social media ed applicato *sentiment analysis* e *object recognition* per identificare il comportamento del turista misurando la sua soddisfazione, in termini di emozioni, e le sue preferenze in termini di luoghi di interesse visitati. Dunque, l'esperimento ha consentito agli studenti di interfacciarsi con nuovi e innovativi metodi di analisi e ha permesso loro di prendere consapevolezza sull'utilità di lavorare a stretto contatto con la ricerca scientifica. Inoltre, tutti o parte di questi metodi sono stati poi incorporati nel progetto digitale. Per questo motivo, si può parlare di *Design Thinking*, uno strumento eccellente per lo sviluppo della creatività. Ispirato alle tecniche e agli strumenti utilizzati dal designer nello sviluppo di progetti innovativi, il Design Thinking è il processo

creativo con cui il designer parte da un problema in cui la soluzione è sconosciuta e arriva allo sviluppo di un'idea creativa ed originale che risponde perfettamente al bisogno identificato. Gli studenti hanno, infatti, creato dei “sistemi di comunicazione” utili per facilitare concretamente la realizzazione di progetti innovativi per località turistiche, siti archeologici, ecc.

I risultati di questo esperimento mostrano che lo studio ha avuto successo in quanto l'accettabilità complessiva del lavoro proposto si è dimostrata elevata. Gli studenti hanno riconosciuto l'importanza del lavoro svolto e hanno riportato un alto livello di motivazione e soddisfazione, anche durante la prova d'esame.

Si è innescata negli studenti la consapevolezza di poter diventare “intermediari” tra la ricerca scientifica ed il mercato del lavoro dove questo processo di diffusione della cultura tecnologica potrebbe portare a benefici significativi ed economie sostenibili nel settore turistico.

I risultati contribuiscono a sostenere l'importanza di trasferire la ricerca “industriale” nell'insegnamento, facendo partecipare attivamente gli studenti ai progressi nei campi tecnologici e di abituarli a leggere articoli scientifici per attingere alla ricerca come fonte di informazione, che può quindi essere trasformata in conoscenza pratica del loro lavoro. L'esperimento è tuttavia completamente empirico in quanto resta ancora molto lavoro da un punto di vista organizzativo e sperimentale. Tuttavia, è stato un contributo importante in quanto ha fornito un quadro di come gli studenti possano essere motivati a diventare promotori di affari e conoscenza nelle situazioni della vita reale. Spinti dal pragmatismo dell'approccio e dalla facilità d'uso dei metodi proposti, consapevoli di trattare con materiali che conoscono, come dati raccolti dai social media, gli studenti sono stati in grado di realizzare innovativi progetti digitali, lavorando al meglio ed implementando sistemi di comunicazione con grande abilità valorizzando il territorio nazionale. A tal proposito, secondo l'osservatorio nazionale del turismo (<http://www.ontit.it>), l'Italia è uno dei pochi paesi con un patrimonio culturale molto ricco, con i suoi 53 siti UNESCO, è stata riconosciuta nel 2017 una delle cinque principali destinazioni turistiche al mondo, con 55 milioni di arrivi annuali. Tuttavia, solo 8,1 milioni di turisti trascorrono il loro tempo in Calabria (regione dell'Italia meridionale). In particolare, il sud Italia con i suoi splendidi paesaggi montani e marini, con i suoi monumenti storici, con il suo patrimonio culturale rappresenta una “destinazione di riferimento” in grado di soddisfare anche il turista più pretenzioso. Il basso flusso turistico, il 14,7%, registrato dalla Calabria nel

2017 suggerisce di indagare meglio le reali motivazioni che influenzano questo scenario. In questo *framework*, l'esperimento non solo ha consentito agli studenti di acquisire conoscenze tecnologiche connesse al nuovo scenario dell'I4.0 ma ha permesso loro di indagare anche dinamiche del turismo calabrese, offrendo l'opportunità, per alcuni di loro di interfacciarsi con gli stakeholder locali, di divulgare un innovativo approccio al settore turistico e di immagazzinare conoscenze utili per un futuro inserimento professionale.

CAPITOLO 5

WEB SEMANTICO: PROGETTAZIONE DI UN SISTEMA INFORMATIVO

5.1 Web Semantico

Come già largamente discusso nei capitoli precedenti, l'enorme flusso di dati, strutturati e non strutturati, prodotti su Internet sta ponendo delle sfide in relazione al recupero delle informazioni coinvolgendo tutti i settori scientifici tra cui il campo della Linguistica Computazionale (LC). Le tematiche scientifiche legate a questa nuova ondata di sviluppo tecnologico, vedono le ricerche nel settore della LC improntate sulla necessità di recuperare la sfera semantica e quella pragmatica dal web.

Attualmente gli utenti del web non forniscono a pieno ed in modo chiaro e codificato le informazioni riguardo i loro contenuti pubblicati in rete, pertanto i *search engine*, così come altri servizi e applicazioni, faticano a creare collegamenti ontologici tra essi (Abdou, et al., 2018). La ricerca computazionale, con la sua capacità di gestire grandi database di strutture linguistiche ed esaminare grandi corpora di prove linguistiche, porta intuizioni uniche nella complessità della semantica lessicale.

In questo contesto di sviluppo, nasce il web semantico con il cui termine si intende la trasformazione del mondo del web in un ambiente in cui tutti i dati disponibili sono associati ad informazioni ed a metadati i quali ne illustrano il significato semantico. L'ideatore di tale termine fu Tim Berners-Lee che, agli inizi del decennio, introdusse nella comunità informatica il web semantico definendolo come *“un'estensione del Web attuale, in cui le informazioni hanno un significato ben definito, consentendo ai computer e alle persone di lavorare meglio in cooperazione”* (Berners-Lee et al., 2001). Quindi, il web semantico è basato sul concetto del collegamento dei dati, ovvero sulla connessione di informazioni nel web sfruttandone la potenza, sia per connettere informazioni non

attualmente collegate sia per unificarle ad altre informazioni che attualmente sono connesse.

Di conseguenza, le macchine non avrebbero accesso solo alla visualizzazione, bensì anche all'elaborazione e alla comprensione dei dati. Le macchine sarebbero quindi in grado di analizzare tutti i dati sul web, di interpretare il loro contenuto, di gestire i link e le interazioni tra persone e computer imitando i flussi cerebrali. In questo contesto, l'enfasi del Web semantico si sposta dalla mera semplificazione della conoscenza alla compilazione della conoscenza. Quest'ultima continuerà ad essere creata, ma i computer saranno più coinvolti nel processo. Nella letteratura, il web è ritenuto un ambiente che si affida a tecnologie semplici, pervasive e fortemente abilitanti come HTTP, URL e HTML. Nel contesto del web semantico i dati collegati utilizzeranno le stesse tecnologie del web tradizionale o tecnologie molto simili (Keller, 2011). Ad oggi, i collegamenti di cui si avvale il web semantico avvengono attraverso le cosiddette Resource Description Frameworks (RDF) che sono strutture che esprimono semplici dichiarazioni di relazioni leggibili dai computer. Gli RDF si servono degli Uniform Resource Indicators (URI) che collegano le informazioni agli RDF in modo molto simile agli URL.

Inoltre, le tecniche innovative di *data mining* e gli algoritmi di Machine Learning sono parte integrante dei problemi del web semantico. Una buona gestione ed analisi dei dati crea conoscenza e quindi offre la possibilità di esplorare i dati e i documenti web da un punto di vista semantico. Sebbene l'architettura del web semantico sia ancora oggetto di dibattito nel campo dell'informatica, la caratteristica principale rimane la creazione di un web nuovo che si trasforma da "contenitore di informazioni" ad "ecosistema" in cui tutte le parti che lo costituiscono sono correlate ed interpretabili da una macchina.

Nella prossima sezione verrà mostrata la progettazione di un sistema informativo basato sul *framework* del web semantico. Il sistema presentato, non si propone come soluzione ai problemi aperti in questo settore, inoltre è un *work in progress*, bensì offre un modello che permette ad un utente di acquisire informazioni approfondite durante una ricerca, attraverso un sistema formato da una rete di relazioni e connessioni tra documenti.

Partendo da dati digitali, nello specifico immagini, il sistema codifica tre diversi aspetti del significato relativo al dato in input: strutture geografiche, concettuali e letterarie, servendosi di tecniche informatiche e di algoritmi di apprendimento automatico che mostrano la plausibilità e la naturalezza delle relazioni tra i dati ed il loro significato semantico.

5.2 Progettazione del Sistema informativo

Al fine di sviluppare un'architettura tipica del web semantico, che rappresenta una rete composta da svariati "elementi", è necessario convertire i documenti esistenti del web in contenuti semantici. Nella presente ricerca, viene proposto un nuovo approccio basato sulla struttura semantica dei dati digitali. Lo scopo è quello di creare un sistema informativo combinando diverse tecniche ed applicando varie funzionalità utili per la mappatura ed il recupero di informazioni dai dati.

Il compito principale del sistema è analizzare il documento web per estrarre etichette che ne identifichino e ne definiscano il contenuto. Il processo è così composto: Database, costituito da immagini estratte dai social media; Applicazione di Machine Learning, per il riconoscimento degli elementi nelle immagini; Interrogazione di tre canali, per ricevere l'informazione a più strati.

Le fasi del processo informativo sono le seguenti:

- i. *Integrazione dei dati*
- ii. *Estrazione dell'informazione*
- iii. *Elaborazione dell'informazione*
- iv. *Presentazione dell'informazione*

Durante la prima fase *Integrazione dei dati*, al sistema vengono forniti dati in input sotto forma di immagini ricavate dal web. In particolare immagini estratte da piattaforme di social media (Flickr, Facebook, Instagram e/o TripAdvisor) di cui si è largamente parlato nei capitoli precedenti. Le immagini fornite sono strettamente correlate alle attività dei turisti o visitatori e rappresentano le "tracce digitali" lasciate e condivise sul web, manifesto delle loro preferenze e del loro comportamento all'interno di una città.

Nella seconda fase *Estrazione dell'informazione*, dopo aver fornito i dati digitali al software, *Wolfram Mathematica*, viene applicata una Machine Learning, nello specifico la funzione *Image Identify* che, ottenendo le immagini in input, fornisce in output delle etichette di classe indicanti il contenuto dell'immagine, ovvero l'oggetto/elemento in essa rappresentato. Un'etichetta di classe viene generalmente utilizzata nel contesto dell'apprendimento automatico supervisionato, ed in particolare nella classificazione. L'algoritmo impiegato dalla funzione, come definito nel paragrafo 2.3.2, crea una classificazione degli elementi, i quali vengono suddivisi in categorie. Le categorie integrate nel software per classificare le immagini sono più di 10.000. Il riconoscimento

dell'oggetto nell'immagine si basa su un particolare tipo di Intelligenza Artificiale chiamato *deep learning*, volto a simulare il processo di apprendimento umano.

Una volta ottenute le etichette di classe, si procede con l'*Elaborazione dell'informazione* ricevuta sotto forma di categorie differenti classificabili in: oggetti, monumenti, edifici, ecc.

Data un'etichetta in input vengono eseguite delle richieste al fine di elaborare a più *layer* l'informazione e dunque recuperare contenuti e metadati dalle categorie intese come entità di *Wolfram Language*. Un'entità è un oggetto esistente. Solitamente si parla di classi di oggetti, ovvero insiemi di oggetti che hanno caratteristiche comuni. Essa può rappresentare una cosa, una persona, un luogo o un oggetto ed è indicato con un nome significativo all'interno di un contesto applicativo. Un'entità differisce da un attributo in quanto quest'ultimo definisce le informazioni sull'entità che deve essere memorizzata. Un'entità può possedere attributi nulli o più attributi e ciascuno di questi attributi si applica solo a quella entità. A tal proposito, sono stati istruiti tre canali attraverso cui elaborare le richieste per ricavare i metadati sulle entità. I canali sono i seguenti:

- i. *POIs*
- ii. *Wikipedia*
- iii. *Google Book*

- i. *Punti di interesse (POIs)*

Per punto di interesse si intende un punto specifico che può essere utile o interessante per gli utenti. Generalmente, viene mostrato su una mappa per determinare un'attrazione quale un luogo storico, un edificio, un monumento, un museo, o qualsiasi altro elemento attrattivo.

In questo canale l'etichetta istituisce un collegamento ad altri due sistemi. Il primo sistema, tramite l'applicazione del software *Mathematica*, permette l'interrogazione del pacchetto *GraphStore* che offre funzionalità per la rappresentazione, l'interrogazione e l'aggiornamento di grafici e set di dati RDF.

```

In[ ]:= Needs["GraphStore`"]

In[ ]:= rdfs[s_] := IRI["http://www.w3.org/2000/01/rdf-schema#" <> s];
skos[s_] := IRI["http://www.w3.org/2004/02/skos/core#" <> s];
wikibase[s_] := IRI["http://wikiba.se/ontology#" <> s];
bd[s_] := IRI["http://www.bigdata.com/rdf#" <> s];
wdt[s_] := IRI["http://www.wikidata.org/prop/direct/" <> s];

```

Figura 5.1. Codice sviluppato in Mathematica per l'interrogazione del pacchetto GraphStore

Il pacchetto definisce un livello simbolico che può essere utilizzato per eseguire query SPARQL su archivi RDF locali o endpoint SPARQL remoti. RDF utilizza gli IRI (Internationalized Resource Identifier) per identificare risorse, proprietà e valori per consentire una rappresentazione e l'integrazione di dati provenienti da fonti diverse. Dopo aver definito i pacchetti, al sistema viene fornita un'etichetta di output e il nome della città di cui interessa estrarre i punti di interesse.

```

SPARQLSelect[
  {typeQuery["building"],
   aroundQuery[GeoPosition[Entity["City", {"Florence", "Toscana", "Italy"}]],
   Quantity[1, "Kilometers"]], labelQuery[]} -> {"placeLabel", "location"}] /.
RDFString[l_, _] -> l

```

Figura 5.2. Codice sviluppato in Mathematica per ricavare tutti i “building” relativi alla città di Firenze

Mathematica fornisce come risultato mappe e coordinate geografiche. Nello specifico tutti i “building” (etichetta scelta), riconosciuti come tali della città di Firenze.

```

<| location -> Point[GeoPosition[{43.7751, 11.2496}]],
  placeLabel -> Cappella Bardi in Santa Maria Novella|>,
<| location -> Point[GeoPosition[{43.7751, 11.2494}]], placeLabel -> Tornabuoni Chapel|>,
<| location -> Point[GeoPosition[{43.7748, 11.2489}]], placeLabel -> Spanish chapelle|>,

```

Figura 5.3. Coordinate ed etichette ottenute in output

Le coordinate a loro volta possono essere mostrate su una mappa.



Figura 5.4. Mappa dei punti di interesse relativi all'etichetta "building" della città di Firenze

Il secondo sistema riguarda l'interrogazione di uno specifico Database, costituito da fogli di calcolo che registrano dati sui punti di interesse turistico-culturale che si trovano nel territorio italiano. L'interrogazione avviene tramite *Mathematica* ed alcuni codici implementati *ad hoc* che consentono di ricavare sia le coordinate geografiche che i *Label*.

```
xx1 = Cases[Table[If[xx[[n, 4]] == "Firenze", n, x], {n, 2, Length[xx]}], Except[x]]
{2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455,
2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475,
2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495,
2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515,
2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534,
2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553,
```

Figura 5.5. Interrogazione di un Database tramite Mathematica

```
poi = Table[Grid[Table[xx[xx1[[sefi[[s, n]]]], 2 ;; 3]], {n, Length[sefi[[s]]}], Frame -> All], {s, 7}]
```

Raccolte d'arte Arciconfraternita della misericordia	Museo
Battistero di San Giovanni a Firenze	Edificio di culto
Cattedrale di Santa Maria del Fiore (Duomo di Firenze)	Edificio di culto
Piazza della Repubblica	Punto di Interesse
Loggia del Bigallo	Monumento
Collezioni dell'osservatorio ximeniano	Museo
Museo del Bigallo	Museo
Museo dell'Opera di Santa Maria del Fiore	Museo
Museo di Casa Martelli	Museo
Cupola di Brunelleschi	Monumento
Campanile di Giotto	Monumento
Santa Reparata	Area Archeologica

Figura 5.6. Output ottenuto a seguito dell'interrogazione del Database

ii. *Wikipedia*

Il secondo canale riguarda l'interrogazione di Wikipedia attraverso il software *Mathematica*. In particolare la richiesta avviene attraverso *WikipediaData* che utilizza l'API di *MediaWiki* per recuperare contenuti e metadati di articoli e categorie da Wikipedia.

```
In[ ]:= SystemOpen["https://www.wikipedia.org/"]
TextSentences[WikipediaData["Colosseum"]][[ ;; 1]]
```

Figura 5.7. Interrogazione di Wikipedia tramite Mathematica

Il sistema si collega attraverso *SystemOpen*, successivamente attraverso *TextSentences* rileva i metadati relativi ad una determinata entità e le importa sotto forma di informazioni.

iii. *Google Book*

Il terzo canale è rappresentato da Google Books che consente di effettuare ricerche all'interno di oltre 100M di testi digitalizzati tramite riconoscimento OCR (Optical Character Recognition) e consente di scaricare la versione digitale del libro o della rivista ove possibile.

In questo sistema vengono impiegate le RESTful API di Google Books che permettono di accedere a molte funzionalità tra cui librerie e volumi digitali. Un volume è un oggetto read only contiene tutte le informazioni di un libro o una rivista digitalizzata e memorizzata sui server di Google, mentre una libreria (*bookshelf*) può essere modificata

tramite autenticazione.

Per il sistema è stato sviluppato uno specifico codice, in linguaggio Javascript (approccio JSON-P), per effettuare le richieste necessarie ad ottenere i dati e i metadati. In particolare, utilizzando un approccio AJAX (Asynchronous JavaScript and XML) tramite una chiamata XMLHttpRequest (XHR) è stato possibile effettuare delle richieste HTTP ad un web server direttamente dal web browser. Il web server a sua volta risponderà alla chiamata con un file JSON contenente le informazioni desiderate.

Le API di Google Books richiedono delle credenziali per poter effettuare la richiesta. Nello scopo della ricerca le richieste riguardano un dataset pubblico pertanto è sufficiente creare ed utilizzare una API Key che identificherà univocamente il progetto, assegnandone le risorse necessarie ad effettuare le query richieste.

```
<html>
]
  <head>
    <title>Google Books API</title>
    <meta charset="UTF-8">
    <meta name="author" content="" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
  ]
  <style>
    table, td, th {
      border: 1px solid #ddd;
    }
    table {
      border-collapse: collapse;
      width: 100%;
    }
    th, td {
      text-align: left;
      padding: 8px;
    }
    th {
      background-color: #4CB0CF;
      color: white;
    }
    tr:nth-child(even) {background-color: #f2f2f2}
  </style>
  <script>
    var gk = "AIzaSyC-KTfsOoGmMme8mOHbilISz5LVb1tD0uA";
    var url = "https://www.googleapis.com/books/v1/";
```

Figura 5.8. Codice in linguaggio Javascript

Il progetto dispone di un'interfaccia GUI (HTML web page) che permette l'immissione della richiesta nell'apposito spazio dedicato "search" e l'ottenimento di un output

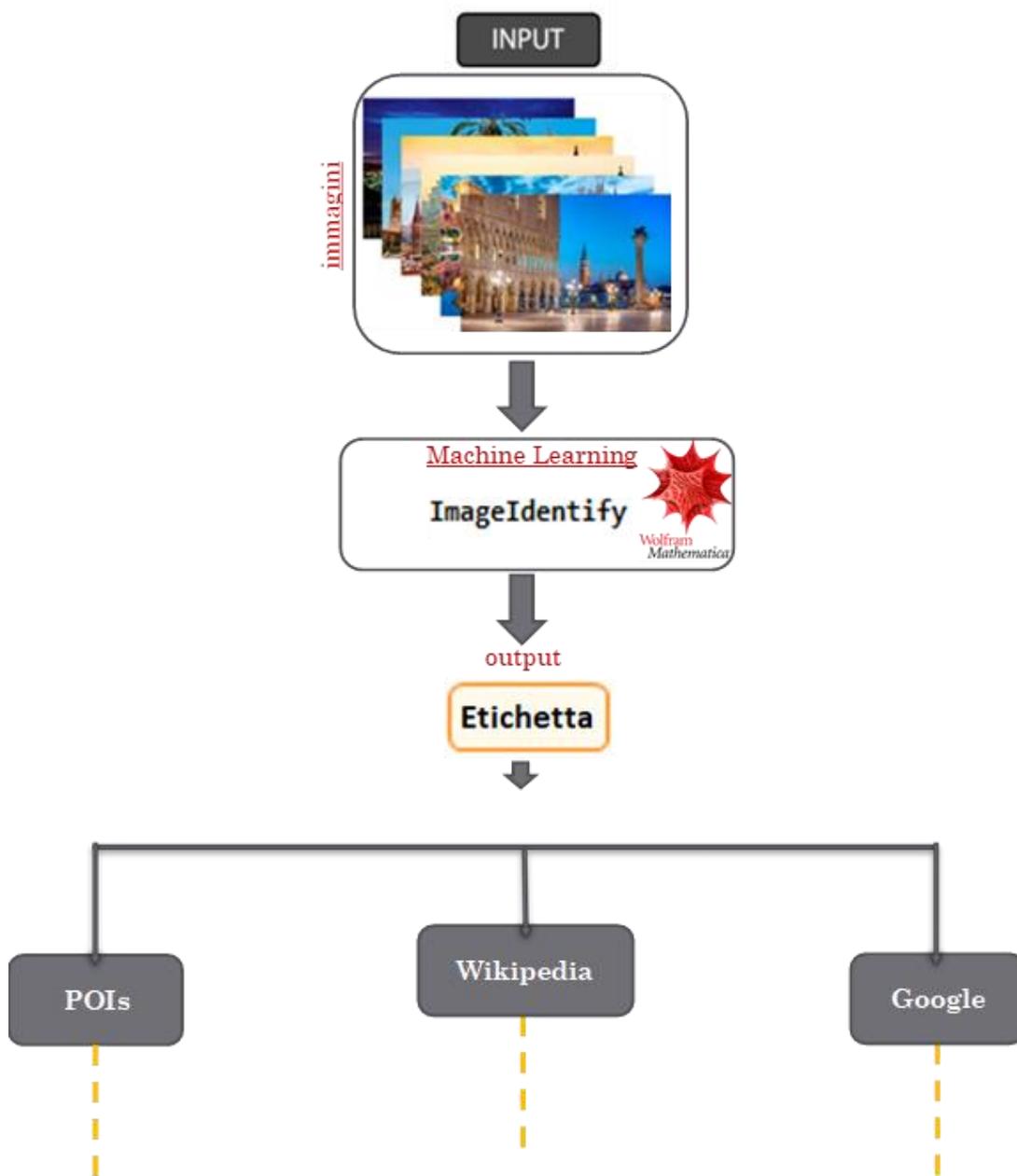
informativo sui libri scritti sull'argomento/etichetta fornita in input.

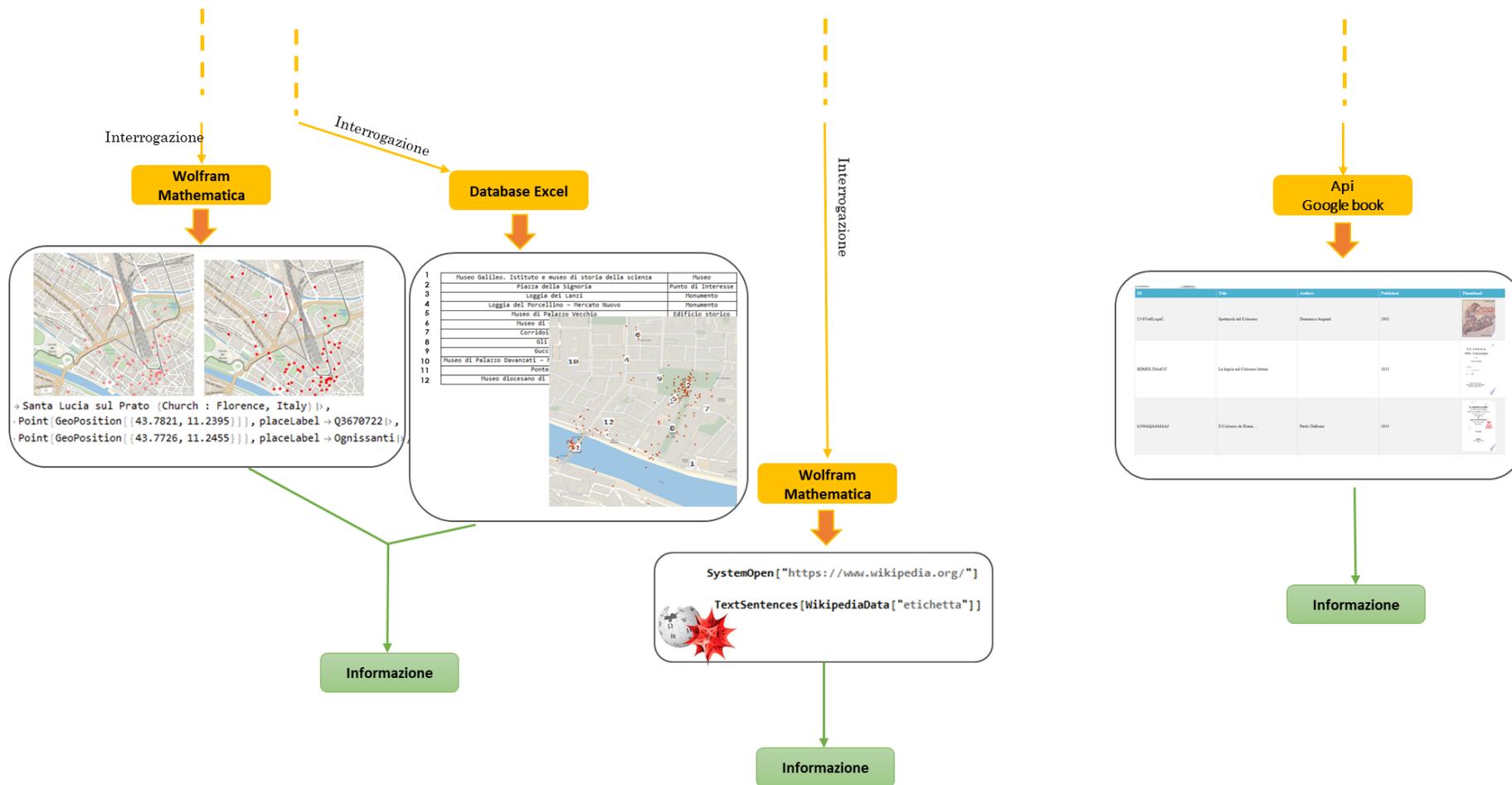
Il risultato è caratterizzato da una serie di metadati che raggruppano informazioni quali: ID del libro; Titolo; Autore; Anno di pubblicazione; Miniatura della copertina; URL (che fornisce la possibilità di vedere l'anteprima del libro).

ID	Title	Authors	Published	Thumbnail	URL
ugMVAAYAAJ	Le chiese d'Italia		1864		JSON View
sWIFAAAQAJ	Le chiese di Napoli		1845		JSON View

Figura 5.9. Interfaccia contenente metadati

L'ultima fase, *Presentazione dell'informazione*, prevede la trasformazione dell'informazione ricevuta in conoscenza, derivata considerando sia i metadati ottenuti dal sistema, sia l'esperienza acquisita dall'utente in passato, la quale rappresenta l'insieme di formazioni rilevanti memorizzate dall'uomo. Il fine ultimo è fornire all'utente che sta effettuando una ricerca un'informazione semantica completa.





Al momento, ci sono enormi quantità di media digitali disponibili sul web. Se associati ai metadati, comprensibili automaticamente, il web potrebbe fungere da fornitore potenzialmente illimitato di servizi web multimediali.

Il sistema presentato si propone come potenziale strumento di recupero di metadati partendo da un'immagine. Permette di associare alla "domanda" dell'utente la relativa semantica, così da poter estrarre l'informazione.

Nello specifico, partendo da un dato digitale, in questo caso un'immagine, l'utente può inserire, nelle apposite interfacce dedicate, l'etichetta (recuperata tramite funzionalità basate su tecnologie di Intelligenza Artificiale) per la quale richiede assistenza. Inoltre, il software utilizzato si avvale di altre fonti esterne permettendo così che le richieste in arrivo vengano processate velocemente.

Mettendo insieme diversi sistemi, come mostrato, e garantendone l'interoperabilità il sistema proposto potrebbe rappresentare per gli utenti un valido strumento informativo in grado di coniugare disponibilità con utilità.

CONCLUSIONI

La ricerca ha fornito un approccio efficace e sistemico all'accesso al ricco set di dati sulle esperienze degli utenti, basato sull'enorme numero di contenuti che essi generano e condividono online ed ha indagato sull'enorme mole di dati al fine di ottenere approfondimenti in grado di offrire l'opportunità di identificare tendenze e modelli che sono importanti da affrontare in vari settori. Infatti, lo studio propone dei vantaggi come l'approccio multidisciplinare tra settori differenti quali le scienze sociali, statistiche e l'informatica. Tale approccio ha permesso di approfondire la ricerca sui Big data nella teoria e fornire un efficace recupero ed analisi dei dati nella pratica. A tal proposito, l'analisi, effettuata su grandi set di dati estratti da Database e da piattaforme come Instagram, Twitter, TripAdvisor ed altri siti di social media, ha fornito una visione psicologica sugli atteggiamenti e sul comportamento di un ampio spettro di utenti.

I dati sono stati esaminati attraverso tecniche quali algoritmi di *Machine Learning*, *analisi semantica* ed *analisi statistica*. Gran parte del lavoro è stato basato sull'adozione di algoritmi di apprendimento automatico applicati per (i) il riconoscimento delle immagini, (ii) per la creazione di cluster, (iii) per l'analisi del testo (*sentiment analysis*) e (iv) per la profilazione di classi di utenti. Per il riconoscimento delle immagini l'approccio ha richiamato le reti neurali artificiali (*deep artificial neural networks*), algoritmi e sistemi computazionali ispirati al cervello umano. Nello specifico, il riconoscimento dei contenuti delle immagini ha permesso di comprendere alcuni fattori quali: l'attrattiva della destinazione turistica e la qualità dell'esperienza dell'utente, nonché dal punto di vista del marketing, l'identificazione della percezione che il cliente ha nei confronti del *brand*. L'adozione del *clustering*, altra tecnica di Machine Learning, ha permesso, invece, di selezionare e raggruppare elementi omogenei in un insieme di dati determinando i punti di interesse in una città. Tale risultato può essere utile all'industria turistica per lo sviluppo di modelli di business appropriati e per influenzare le scelte degli utenti durante la ricerca di informazioni relative ad un viaggio. L'analisi dei cluster è stata applicata, a seguito di analisi statistiche, anche su dati relativi al settore energetico, al fine di identificare e profilare classi di utenti. I risultati ottenuti hanno fornito sia un quadro della situazione energetica nelle famiglie sia informazioni utili per il *provider* (fornitore di servizi) in quanto riflettono maggiore capacità di rappresentare la realtà e dunque maggiore

opportunità di agire sulle leve che permettono di attuare efficienze, di ottimizzare processi interni e di sviluppare nuovi prodotti e servizi.

L'analisi del testo ha permesso, invece, di identificare il significato delle parole nei testi (post e commenti) e definirne il "sentimento" nascosto. In base alle conoscenze sulla letteratura precedente, per quanto riguarda l'analisi del testo, mentre altri autori hanno cercato di identificare l'effetto di commenti, post e recensioni online positivi o negativi, la ricerca condotta è la prima a mostrare come rilevare sistematicamente i sentimenti positivi o negativi dei tweet (messaggi di testo pubblicati su Twitter) condivisi per migliorare le informazioni nel settore del marketing e del Retail. Per quanto concerne il settore turistico, lo studio della destinazione culturale urbana si è rivelato importante al fine di indagare il consumo turistico in termini di spazio e tempo e la comprensione dell'atteggiamento ha rappresentato un punto di forza nella conoscenza del turista inteso come consumatore. A tal proposito, l'aggiunta di esperti professionisti nei vari settori della conoscenza tecnologica per l'industria turistica è necessaria nei processi di erogazione di servizi. La proposta di un percorso di formazione, verificato nella presente ricerca attraverso un esperimento didattico, si è rivelata una sfida interessante che ha fornito indicazioni su come integrare le risorse umane nel *framework* dell'Industria 4.0 e ha dimostrato come i Big data possano realmente creare valore per un'azienda.

Attingendo da studi precedenti, i risultati raggiunti dimostrano che gli algoritmi di analisi dei Big data quali l'apprendimento automatico, per valutare i contenuti generati dagli utenti contribuiscono, da un lato alla comprensione sull'esperienza dell'utente verso un luogo, una destinazione; d'altra parte, la loro analisi fornisce una conoscenza sistematica delle valutazioni dei consumatori su un determinato prodotto o servizio e verso lo sviluppo di una sorta di "intelligenza sociale". Pertanto, la ricerca risponde alla richiesta di sviluppare nuove analisi per comprendere come estrarre valore dai dati strutturati e non strutturati (Dindar & Yaman, 2018; Balducci & Marinova, 2018; Hartman et al., 2019). Estrarre informazioni dall'analisi del comportamento del consumatore come individuo significa indagare sulle modalità con cui egli raccoglie le informazioni e le condivide. Inoltre, considerare l'utente/consumatore come parte integrante dell'ambiente significa prestare attenzione anche agli stimoli ai quali è soggetto che sono in gran parte responsabili delle decisioni prese in relazione a destinazioni, prodotti, servizi in quanto influenzano il processo decisionale. Tali presupposti sono stati determinanti per comprendere gli elementi psicosociali e motivazionali propri dell'individuo.

Inoltre, i risultati della ricerca propongono come, un approccio più sofisticato al monitoraggio dei social media nel contesto turistico e nel marketing, nonché nel settore dell'education, possa contribuire a migliorare le decisioni strategiche e le politiche operative degli stakeholder.

Per quanto concerne la progettazione del sistema informativo esso potrebbe per facilitare le comunicazioni tra gli utenti e permette di fare, non solo ricerche basate su parole chiave, ma anche sui contenuti semantici. A tal proposito, partendo dalle ricerche di Ogden e Richards (1923) sullo studio del significato e di Jakobson (1966) che studiò i processi comunicativi, si è cercato di strutturare e sistematizzare un processo che riflette un atto comunicativo ed informativo tale che un simbolo (immagine) attraverso l'applicazione di un significante (machine learning che si sostituisce al processo mentale proprio dell'uomo) permettesse l'esplicitazione di un referente (oggetto\etichetta) che opportunamente porta alla trasmissione di un messaggio sotto forma di conoscenza. Il tutto coordinato da un Sistema in grado di coniugare fattori differenti in un'ottica interdisciplinare dove l'analisi dei dati combacia perfettamente con la linguistica.

Bibliografia consultata

- Abdou, M., AbdelGaber, S., & Farhan, M. (2018). A semi-automated framework for semantically annotating web content. *Future Generation Computer Systems*, 81, 94-102.
- Ahmad, J., Muhammad, K., Lloret, J., & Baik, S. W. (2018). Efficient conversion of deep features to compact binary codes using Fourier decomposition for multimedia big data. *IEEE Transactions on Industrial Informatics*, 14(7), 3205-3215.
- Ai, P., Liu, Y., & Zhao, X. (2019). Big Five personality traits predict daily spatial behavior: Evidence from smartphone data. *Personality and Individual Differences*, 147, 285-291.
- Alpaydin, E. (2016). *Machine learning: the new AI*. MIT press.
- Antoniadis, A., Brossat, X., Cugliari, J., & Poggi, J. M. (2013). Clustering functional data using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(01), 1350003.
- Arain, Q. A., Memon, H., Memon, I., Memon, M. H., Shaikh, R. A., & Mangi, F. A. (2017). Intelligent travel information platform based on location base services to predict user travel behavior from user-generated GPS traces. *International Journal of Computers and Applications*, 39(3), 155–168.
- Ariffin, A. A. M., Maghzi, A., Soon, J. L. M., & Alam, S. S. (2018). Exploring the Influence of Hospitality on Guest Satisfaction in Luxury Hotel Services. *e-Review of Tourism Research*, 15(1), 1-20.
- Baena, F., Guarin, A., Mora, J., Sauza, J., & Retat, S. (2017). Learning factory: The path to industry 4.0. *Procedia Manufacturing*, 9, 73-80.
- Balasubramanian, K., & Ragavan, N. A. (*in press*). What are the key challenges faced by the Malaysian hospitality and tourism industry in the context of industrial revolution 4.0?. *Worldwide Hospitality and Tourism Themes*.
- Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557-590.
- Ballantine, P. W., & Stephenson, R. J. (2011). Help me, I'm fat! Social support in

- online weight loss networks. *Journal of Consumer Behaviour*, 10(6), 332-337.
- Balomenou, N., & Garrod, B. (2019). Photographs in tourism research: prejudice, power, performance and participant-generated images. *Tourism Management*, 70, 201-217.
 - Banta, E. R., Poeter, E. P., Doherty, J. E., & Hill, M. C. (2006). JUPITER: Joint universal parameter identification and evaluation of reliability-an application programming interface (API) for model analysis (Techniques and Methods 6-E1). Denver, CO:Author.
 - Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., ... & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734, 1-74.
 - Baumann, G. (2006). *Mathematica for theoretical physics: classical mechanics and nonlinear dynamics* (Vol. 1). Springer Science & Business Media.
 - Bell, D. R., Gallino, S., & Moreno, A. (2018). Offline Experiences and Value Creation in Omnichannel Retail. *Available at SSRN*.
 - Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: Free Press.
 - Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, 284(5), 28-37.
 - Bertacchini, F., Bilotta, E., & Pantano, P. (2017). Shopping with a robotic companion. *Computers in Human Behavior*, 77, 382-395.
 - Blazquez, D., & Domenech, J. (2018). Big Data Big data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, 130, 99-113.
 - Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data Big data and predictive analytics in retailing. *Journal of Retailing*, 93(1), 79-95.
 - Bronner, F., & De Hoog, R. (2010). Consumer-generated versus marketer-generated websites in consumer decision making. *International Journal of Market Research*, 52(2), 231-248.
 - Busacca, B. G. (2000). *Il valore della marca tra postfordismo ed economia digitale*. Egea.
 - Cárdenas-García, P. J., & Pulido-Fernández, J. I. (2019). Tourism as an economic

- development tool. Key factors. *Current Issues in Tourism*, 22(17), 2082-2108.
- Caspari, G., & Jendryke, M. (2017). Archsphere—A cluster algorithm for archaeological applications. *Journal of Archaeological Science: Reports*, 14, 181-188.
 - Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital marketing*. Pearson UK.
 - Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data Big data to big impact. *MIS quarterly*, 36(4).
 - Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, 19(2), 171-209.
 - Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, 17(3), 433-450.
 - Chonacky, N., & Winch, D. (2005). 3Ms for instruction: Reviews of Maple, Mathematica, and Matlab. *Computing in Science & Engineering*, 7(3), 7.
 - Clement, J. (2019). Number of monthly active Facebook users worldwide as of 1st quarter 2019 (in millions). Statista Retrieved from <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide>.
 - Clements, M., Serdyukov, P., De Vries, A. P., & Reinders, M. J. (2010, July). Using flickr geotags to predict user travel behaviour. *Proceedings of the 33rd international ACM SIGIR conference on Research and Development in Information Retrieval*, Geneva, Switzerland (pp. 851–852). New York, NY: ACM.
 - Cook, S. C., & Yurchisin, J. (2017). Fast fashion environments: consumer's heaven or retailer's nightmare?. *International Journal of Retail & Distribution Management*, 45(2), 143-157.
 - Cossío-Silva, F. J., Revilla-Camacho, M. Á., Vega-Vázquez, M., & Palacios-Florencio, B. (2016). Value co-creation and customer loyalty. *Journal of Business Research*, 69(5), 1621-1625.
 - Dalli, D., & Romani, S. (2004). Il comportamento del consumatore. Acquisti e consumi in una prospettiva di marketing (Vol. 30). FrancoAngeli.
 - Day, G. S. (1994). The capabilities of market-driven organizations. *Journal of marketing*, 58(4), 37-52.
 - Del Pezzo, E., & Bianco, F. (2010). MathLTWA: Multiple lapse time window

- analysis using Wolfram Mathematica 7. *Computers & Geosciences*, 36(10), 1388-1392.
- Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4), 197-387.
 - Deng, M., & Cao, Y. (2018). Innovation and Effect Evaluation Model of Education and Training Outsourcing of State-owned Enterprises under Big Data. *Kuram ve Uygulamada Egitim Bilimleri*, 18(6), 3017-3027.
 - Devereux, L., Melewar, T.C., & Foroudi, P. (2017). Corporate Identity and Social Media: Existence and Extension of the Organization. *International Studies of Management and Organization*, 47(2), 110-134.
 - Devkota, B., Miyazaki, H., Witayangkurn, A., & Kim, S. M. (2019). Using Volunteered Geographic Information and Nighttime Light Remote Sensing Data to Identify Tourism Areas of Interest. *Sustainability*, 11(17), 4718.
 - Dindar, M., & Dulkadir Yaman, N. (2018). # IUseTwitterBecause: content analytic study of a trending topic in Twitter. *Information Technology & People*, 31(1), 256-277.
 - Di Pietro, L., & Pantano, E. (2012). An empirical investigation of social network influence on consumer purchasing decision: The case of Facebook. *Journal of Direct, Data and Digital Marketing Practice*, 14(1), 18-29.
 - Diebold, F. X., Cheng, X., Diebold, S., Foster, D., Halperin, M., Lohr, S., ... & Schorfheide, F. (2012). *A Personal Perspective on the Origin (s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline*. University of Pennsylvania, USA.
 - Doherty, N. F., & Ellis-Chadwick, F. (2010). Internet retailing: the past, the present and the future. *International Journal of Retail & Distribution Management*, 38(11/12), 943-965.
 - Dona, I., & Popa, D. (2013). Tourism destination mapping through cluster analysis. *Management, Economic Engineering in Agriculture and Rural Development*, 13(2), 133-138.
 - Ellonen, H. K., Wikström, P., & Jantunen, A. (2009). Linking dynamic-capability portfolios and innovation outcomes. *Technovation*, 29(11), 753-762.
 - Evanschitzky, H., Iyer, G. R., Pillai, K. G., Kenning, P., & Schütte, R. (2015). Consumer trial, continuous use, and economic benefits of a retail service

- innovation: The case of the personal shopping assistant. *Journal of Product Innovation Management*, 32(3), 459-475.
- Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data Big data analytics for business intelligence through the lens of marketing mix. *Big Data Big data Research*, 2(1), 28-32.
 - Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210, 15-26.
 - Füller, J., Bartl, M., Ernst, H., & Mühlbacher, H. (2006). Community based innovation: How to integrate members of virtual communities into new product development. *Electronic Commerce Research*, 6(1), 57-73.
 - Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
 - Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2015). Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, 20(1), 112-141.
 - Gil, D., & Song, I. Y. (2016). Modeling and management of big data: challenges and opportunities.
 - Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing*, 7(4), 36–43.
 - Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets*, 25(3), 179-188.
 - Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7), 1645-1660.
 - Guo, J., Zhang, W., Fan, W., & Li, W. (2018). Combining geographical and social influences with deep learning for personalized point-of-interest recommendation. *Journal of Management Information Systems*, 35(4), 1121–1153.
 - Hadi, H. J., Shnain, A. H, Hadishaheed, S., Ahmad, A., H. (2015). Big data Big data and five v's characteristics. *International Journal of Advances in Electronics and Computer Science*, 2(1), 2393-2835.

- Hagel, J., & Armstrong, A. G. (1997). *Net. Gain, Harvard Business School Press, Boston.*
- Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research, 56(3), 387-404.*
- Hartmann, N. N., Wieland, H., & Vargo, S. L. (2018). Converging on a new theoretical foundation for selling. *Journal of Marketing, 82(2), 1-18.*
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information systems, 47, 98-115.*
- Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research, 52(5), 629-641.*
- Hornik, J., Satchi, R. S., Cesareo, L., & Pastore, A. (2015). Information dissemination via electronic word-of-mouth: Good news travels fast, bad news travels faster!. *Computers in Human Behavior, 45, 273-280.*
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression (Vol. 398).* John Wiley & Sons.
- Hwang, K., & Chen, M. (2017). *Big-data analytics for cloud, IoT and cognitive computing.* John Wiley & Sons.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR), 31(3), 264–323.*
- Jankowski, P., Andrienko, N., Andrienko, G., & Kisilevich, S. (2010). Discovering landmark preferences and movement patterns from photo postings. *Transactions in GIS, 14(6), 833–852.*
- Jansen-Verbeke, M. (1986). Inner-city tourism: resources, tourists and promoters. *Annals of Tourism Research, 13(1), 79-100.*
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data Big data research. *Big Data Big data Research, 2(2), 59-64.*
- Jakobson, R. (1966). *Saggi di linguistica generale, a cura di L. Heilmann, Feltrinelli, Milano.*
- Jung, T. H., Tom Dieck, M. C., & Chung, N. (2018). Determinants of hotel social media continued usage. *International Journal of Contemporary Hospitality Management, 30(2), 1152-1171.*

- Kashyap, P. (2017). Let's Integrate with Machine Learning. In *Machine Learning for Decision Makers* (pp. 1-34). Apress, Berkeley, CA.
- Keller, M. A. (2011). Linked data: A way out of the information chaos and toward the semantic web. *Educause Review*, 46(4), 10-11.
- Khan, I., & Rahman, Z. (2017). Development of a scale to measure hotel brand experiences. *International Journal of Contemporary Hospitality Management*, 29(1), 268-287.
- Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A. A., & Salehian, S. (2018, March). The 10 Vs, Issues and Challenges of Big Data. In *Proceedings of the 2018 International Conference on Big Data Big data and Education* (pp. 52-56). ACM.
- Kim, H., & Stepchenkova, S. (2015). Effect of tourist photographs on attitudes towards destination: manifest and latent content. *Tourism Management*, 49, 29-41.
- Kim, J., & Lee, K. H. (2019). Influence of integration on interactivity in social media luxury brand communities. *Journal of Business Research*, 99, 422-429.
- Kim, J., Song, J., & Jones, D. R. (2011). The cognitive selection framework for knowledge acquisition strategies in virtual communities. *International Journal of Information Management*, 31(2), 111-120.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 14(2), 1137-1145.
- Konijn, E., Sluimer, N., & Mitas, O. (2016). Click to share: Patterns in tourist photography and sharing. *International Journal of Tourism Research*, 18(6), 525-535.
- Kraus, M., Feuerriegel, S., & Oztekin, A. (2019). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.
- Kromrey, H. (1998). *Modelle und Methoden der Datenerhebung und Datenauswertung*, überarbeitete und erweiterte Auflage, Opladen
- Kudyba, S., & Diwan, R. K. (2002). *Information technology, corporate productivity, and the new economy*. Greenwood Publishing Group.
- Kudyba, S., & Kwatinetz, M. (2014). *Introduction to the big data Big data era*.

Big Data, Mining, and Analytics, 1-17.

- Lai, I. K. W., & Hitchcock, M. (2017). Sources of satisfaction with luxury hotels for new, repeat, and frequent travelers: A PLS impact-asymmetry analysis. *Tourism Management*, 60, 107-129.
- Lamsfus, C., Martín, D., Alzua-Sorzabal, A., & Torres-Manzanera, E. (2015). Smart tourism destinations: An extended conception of smart cities focusing on human mobility. In *Information and communication technologies in tourism 2015* (pp. 363-375). Springer, Cham.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.
- Larsen, S. (2007). Aspects of a psychology of the tourist experience. *Scandinavian Journal of Hospitality and Tourism*, 7(1), 7-18.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Lee, D., Kim, H. S., & Kim, J. K. (2012). The role of self-construal in consumers' electronic word of mouth (eWOM) in social networking sites: A social cognitive approach. *Computers in Human Behavior*, 28(3), 1054-1062.
- Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5), 881-894.
- Liang, B., & Scammon, D. L. (2011). E-Word-of-Mouth on health social networking sites: An opportunity for tailored health communication. *Journal of Consumer Behaviour*, 10(6), 322-331.
- Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach. *International Journal of Electronic Commerce*, 20(2), 236-260.
- Lim, J., Hwang, Y., Kim, S., & Biocca, F. (2015). How social media engagement leads to sports channel loyalty: mediating roles of social presence and channel commitment. *Computers in Human Behavior*, 46, 158-167.
- Liu, M. T., Wong, I. A., Tseng, T. H., Chang, A. W. Y., & Phau, I. (2017). Applying consumer based brand equity in luxury hotel branding. *Journal of Business Research*, 81, 192-202.
- Lo, I. S., & McKercher, B. (2015). Ideal image in process: Online tourist photography and impression management. *Annals of Tourism Research*, 52, 104-

- 116.
- Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic bulletin & review*, 9(4), 829-835.
 - MacKay, K. J., & Couldwell, C. M. (2004). Using visitor-employed photography to investigate destination image. *Journal of Travel Research*, 42(4), 390-396.
 - Mahmood, Z. (Ed.). (2016). Data science and big data Big data computing: frameworks and methodologies. Springer.
 - Marr, B. (2015). Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance. John Wiley & Sons.
 - Maslow, A. H. (2010). *Motivazione e personalità*. Armando editore.
 - Mayring, P. (2010). Qualitative inhaltsanalyse. In *Handbuch qualitative Forschung in der Psychologie* (pp. 601-613). VS Verlag für Sozialwissenschaften.
 - McLoughlin, C., Patel, K. D., O'Callaghan, T., & Reeves, S. (2018). The use of virtual communities of practice to improve interprofessional collaboration and education: findings from an integrated review. *Journal of interprofessional care*, 32(2), 136-142.
 - Md Khairi, N. D., Ismail, H. N., & Syed Jaafar, S. M. R. (2019). Tourist behaviour through consumption in Melaka World Heritage Site. *Current Issues in Tourism*, 22(5), 582-600.
 - Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
 - Memon, I., Chen, L., Majid, A., Lv, M., Hussain, I., & Chen, G. (2015). Travel recommendation using geo-tagged photos in social media for tourist. *Wireless Personal Communications*, 80(4), 1347-1362.
 - Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data Big data on board level decision-making. *Journal of Business Research*, 93, 67-78.
 - Merten, P. S. (2007). The transformation of the distribution process in the airline industry empowered by information and communication technology. In *Information and communication technologies in support of the tourism industry* (pp. 76-113). IGI Global.
 - Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A big data Big data

- analytics method for tourist behaviour analysis. *Information & Management*, 54(6), 771-785.
- Miháltz, M., Váradi, T., Csertő, I., Fülöp, É., Pólya, T., & Kővágó, P. (2015, September). Beyond Sentiment: Social Psychological Analysis of Political Facebook Comments in Hungary. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 127-133).
 - Minazzi, R., & Lagrosen, S. (2013). Investigating social media marketing in the hospitality industry: Facebook and European hotels. In *Information and Communication Technologies in Tourism 2014* (pp. 145-157). Springer, Cham.
 - Mirkovic, M., Culibrk, D., Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., McArdle, G., & Crnojevic, V. (2011, October). A comparative study of spatial, temporal and content-based patterns emerging in YouTube and Flickr. *2011 international conference on Computational Aspects of Social Networks (CASoN)*, Salamanca, Spain (pp. 189–194). IEEE.
 - Morris, R. (1994). Computerized content analysis in management research: A demonstration of advantages & limitations. *Journal of management*, 20(4), 903-931.
 - Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.
 - Motamarri, S., Akter, S., & Yanamandram, V. (2017). Does big data Big data analytics influence frontline employees in services marketing?. *Business Process Management Journal*, 23(3), 623-644.
 - Mueller, J., Hutter, K., Fueller, J., & Matzler, K. (2011). Virtual worlds as knowledge management platform—a practice-perspective. *Information Systems Journal*, 21(6), 479-501.
 - Negroponte, N. (1995). The digital revolution: Reasons for optimism. *The Futurist*, 29(6), 68.
 - Nikjoo, A., & Bakhshi, H. (2019). The presence of tourists and residents in shared travel photos. *Tourism Management*, 70, 89-98.
 - O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management*, 19(7), 754-772.

- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1-135.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.
- Pang, J., & Qiu, L. (2016). Effect of online review chunking on product attitude: The moderating role of motivation to think. *International Journal of Electronic Commerce*, 20(3), 355-383.
- Pantano, E., & Dennis, C. (2019). Toward a Smart Store. In *Smart Retailing* (pp. 65-77). Palgrave Pivot, Cham.
- Pantano, E., & Di Pietro, L. (2013). From e-tourism to f-tourism: emerging issues from negative tourists' online reviews. *Journal of Hospitality and Tourism Technology*, 4(3), 211-227.
- Pantano, E., & Timmermans, H. (2014). What is smart for retailing?. *Procedia Environmental Sciences*, 22, 101-107.
- Pantano, E., Giglio, S., & Dennis, C. (2020). Integrating Big data Analytics Into Retail Services Marketing Management: The Case of a Large Shopping Center in London, UK. In *Handbook of Research on Innovations in Technology and Marketing for the Connected Consumer* (pp. 205-222). IGI Global.
- Pantano, E., Giglio, S., & Dennis, C. tweets: Sentiment outcomes for fast fashion retailers through Big Data analytics. *International Journal of Retail and Distribution Management*, 47 (9), 915-927.
- Pantano, E., Priporas, C. V., & Stylos, N. (2017). ‘You will like it!’ using open data to predict tourists' response to a tourist attraction. *Tourism Management*, 60, 430-438.
- Pappadà, R., Durante, F., Salvadori, G., & De Michele, C. (2018). Clustering of concurrent flood risks via hazard scenarios. *Spatial Statistics*, 23, 124–142.
- Pearce, P. L. (1982). *The social psychology of tourist behaviour*. Oxford: Pergamon.
- Pereira, A. C., & Romero, F. (2017). A review of the meanings and the implications of the Industry 4.0 concept. *Procedia Manufacturing*, 13, 1206-1214.
- Perez-Vega, R., Taheri, B., Farrington, T., & O'Gorman, K. (2018). On being

- attractive, social and visually appealing in social media: The effects of anthropomorphic tourism brands on Facebook fan pages. *Tourism management*, 66, 339-347.
- Pfaender, F., & Fen-Chong, J. (2011, July 6–8). Exploratory analysis of urban mobility: From mobile phone usage data to tourist behaviour. *IET International Conference on Smart and Sustainable City (ICSSC)*, Shanghai, China.
 - Popescu, A., & Grefenstette, G. (2009, April 20–24). Deducing trip related information from Flickr. *Proceedings of the 18th international conference on World Wide Web*, Madrid, Spain (pp. 1183–1184). New York, NY: ACM.
 - Preis, T., Moat, H. S., Bishop, S. R., Treleaven, P., & Stanley, H. E. (2013). Quantifying the digital traces of Hurricane Sandy on Flickr. *Scientific reports*, 3, 3141.
 - Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data Big data processing. *EURASIP Journal on Advances in Signal Processing*, 2016(1), 67.
 - Ridings, C. M., & Gefen, D. (2004). Virtual community attraction: Why people hang out online. *Journal of Computer-mediated communication*, 10(1), JCMC10110.
 - Rubtzov, V., Rozhko, M., & Gabdrakhmanov, N. (2016). Cluster analysis in tourism development. International Business survey. *Enterprise Information Systems*, 13(2), 148-169.
 - Sagioglu, S., & Sinanc, D. (2013, May). Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 42-47). IEEE.
 - Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data Big data sources. *Tourism Management*, 66, 13–25.
 - Samuel, A.rthur L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*. Management, 10(22), 5291–5294.
 - Scott, S. V., & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, organizations and society*, 37(1), 26-40.

- Shaw, S. L., Tsou, M. H., & Ye, X. (2016). Human dynamics in the mobile and big data Big data era. *International Journal of Geographical Information Science*, 30(9), 1687-1693.
- Sheng, J., Amankwah-Amoah, J., & Wang, X. (2018) (*in press*). Technology in the 21st century: New challenges and opportunities. *Technological Forecasting and Social Change*.
- Sheng, J., Amankwah-Amoah, J., & Wang, X. (2019). Technology in the 21st century: new challenges and opportunities. *Technological Forecasting and Social Change*, 143, 321-335.
- Sherman, R. (2005). Producing the superior self: Strategic comparison and symbolic boundaries among luxury hotel workers. *Ethnography*, 6(2), 131-158.
- Sigala, M. (2018). New technologies in tourism: From multi-disciplinary to anti-disciplinary advances and trajectories. *Tourism management perspectives*, 25, 151-155.
- Snyder, B., & Barzilay, R. (2007, April). Multiple aspect ranking using the good grief algorithm. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, (pp. 300-307) ACL.).
- Solomon M. R. (2004). *Consumer behavior, buying, having, and being*. Pearson, Prentice, Hall, New Jersey.
- Sonnier, G. P., McAlister, L., & Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, 30(4), 702-716.
- Stylianou-Lambert, T. (2012). Tourists with cameras: Reproducing or producing. *Annals of Tourism Research*, 39(4), 1817-1838.
- Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2015). Road-based travel recommendation using geotagged images. *Computers, Environment and Urban Systems*, 53, 110–122.
- Tan, P. N., Steinbach, M., & Kumar, V. (2006). Cluster analysis: basic concepts and algorithms. *Introduction to data mining*, 8, 487-568.
- Tan, T. T. W., & Ahmed, Z. U. (1999). Managing market intelligence: an Asian marketing research perspective. *Marketing Intelligence & Planning*.
- Tanwar, A. K., Crisostomi, E., Ferraro, P., Raugi, M., Tucci, M., & Giunta, G. (2015, July). Clustering analysis of the electrical load in European countries. In

- 2015 *International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
- Tarca, A. L., Carey, V. J., Chen, X. W., Romero, R., & Drăghici, S. (2007). Machine learning and its applications to biology. *PLoS computational biology*, 3(6), e116.
 - Thaler, R. H., & Tucker, W. (2013). Smarter information, smarter consumers. *Harvard Business Review*, 91(1), 45–54.
 - Titscher, S., Meyer, M., Wodak, R., & Vetter, E. (2000). Content analysis. *Methods of text and discourse analysis*, 55-73.
 - Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J. P., Saramäki, J., & Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. *Social networks*, 31(4), 240-254.
 - Tönnies, F. (1887). Community and society. *The urban sociology reader*, 13.
 - Trampe, D., Konuş, U., & Verhoef, P. C. (2014). Customer responses to channel migration strategies toward the e-channel. *Journal of Interactive Marketing*, 28(4), 257-270.
 - Tsekouras, D. (2017). The effect of rating scale design on extreme response tendency in consumer product ratings. *International Journal of Electronic Commerce*, 21(2), 270-296.
 - Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.
 - Vaidya, S., Ambad, P., & Bhosle, S. (2018). Industry 4.0—a glimpse. *Procedia Manufacturing*, 20, 233-238.
 - Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*, 32(4), 72-77.
 - Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research*, 20(1), 43-58.
 - Van Laer, T., Edson Escalas, J., Ludwig, S., & van den Hende, E. (2018). What happens in Vegas stays on TripAdvisor? A theory and technique to understand

- narrativity in consumer reviews. *Journal of Consumer Research, Forthcoming*.
- Verhoef, P. C., Kooge, E., & Walk, N. (2016). *Creating value with big data Big data analytics: Making smarter marketing decisions*. Routledge.
 - Verma, V., Sharma, D., & Sheth, J. (2016). Does relationship marketing matter in online retailing? A meta-analytic approach. *Journal of the Academy of Marketing Science, 44*(2), 206-217.
 - Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change, 126*, 3-13.
 - Weed, K. (2015). Unilever CMO Keith Weed's five tech trends for 2015. <https://www.prweek.com/article/1330175/unilever-cmo-keith-weeds-five-tech-trends-2015>. Last access 06/11/2019
 - Weng, J., & Lee, B. S. (2011, July). Event detection in twitter. In *Fifth international AAAI conference on weblogs and social media*.
 - Willems, K., Smolders, A., Brengman, M., Luyten, K., & Schöning, J. (2017). The path-to-purchase is paved with digital opportunities: An inventory of shopper-oriented retail technologies. *Technological Forecasting and Social Change, 124*, 228-242.
 - Williams, K. C., & Trani, A. (2001). *Psicologia per il marketing*. Il mulino.
 - Wolfram, S. (2002). *A new kind of science* (Vol. 5, p. 130). Champaign, IL: Wolfram media.
 - Wolfram, S. (2017) *An elementary introduction to the Wolfram Language*. Wolfram Media, Incorporated.
 - Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179–188.
 - Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services, 22*, 244-249.
 - Xu, K., Liao, S. S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for competitive intelligence. *Decision support systems, 50*(4), 743-754.
 - Xu, L. D., & Duan, L. (2019). Big data Big data for cyber physical systems in industry 4.0: A survey. *Enterprise Information Systems, 13*(2), 148-169.

- Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: state of the art and future trends. *International Journal of Production Research*, 56(8), 2941-2962.
- Yin, R.K. (2003), *Case Study Research: Design and Methods*, Sage Publications, Newbury Park, CA.
- Yuan, Q., Cong, G., Ma, Z., Sun, A., & Thalmann, N. M. (2013, July). Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 363-372). ACM.
- Zhao, K., Cong, G., & Sun, A. (2016, October). Annotating points of interest with geo-tagged tweets. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 417-426). ACM.
- Zhao, N., & Li, H. (2019). How can social commerce be boosted? The impact of consumer behaviors on the information dissemination mechanism in a social commerce network. *Electronic Commerce Research*, 1-24.
- Zhang, K., Chen, Y., & Li, C. (2019). Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tourism Management*, 75, 595-608.
- Zheng, Y., Xie, X., & Ma, W. Y. (2010). Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Engineering Bulletin*, 33(2), 32–39.

INDICE DELLE FIGURE

Figura 1.1. Le 5 caratteristiche dei Big Data.....	7
Figura 1.2. Modello di creazione di valore dai Big data proposto da Verhoef et al., 2016	10
Figura 1.3. Schema funzionale dell’algoritmo di Machine Learning nell’analisi dei dati	16
Figura 1.4. Schema dei tipi di Machine Learning	18
Figura 1.5. Framework di analisi dei Big Data.....	21
Figura 1.6. Rappresentazione di una rete neurale.....	24
Figura 1.7. Istogramma 3D rappresentativo in cui il set di dati è stato diviso in 19 gruppi distinti.....	30
Figura 2.1. Interfaccia input della Piattaforma	37
Figura 2.2. Interfaccia output Piattaforma.....	38
Figura 2.3. Codice sviluppato in Mathematica per ricavare la Location.....	39
Figura 2.4. Codice sviluppato in Mathematica per classificare gli utenti	40
Figura 2.5. Codice sviluppato in Mathematica per ricavare i luoghi con maggiore attività fotografica	40
Figura 2.6. Wolfram Neural Net Repository e probabilità di riconoscimento dell’oggetto rappresentato nell’immagine	41
Figura 2.7. Informazioni sul metodo applicato per il raggruppamento dei dati	43
Figura 2.8. Numero di visitatori nella città di Palermo nel corso dei tre anni e paese di provenienza	44

Figura 2.9. ImageIdentify e riconoscimento degli elementi di alcune immagini della città di Venezia estratte da Flickr.....	46
Figura 2.10. I dieci elementi più fotografati nel corso dei tre anni in tutte le sei città...	47
Figura 2.11. Tendenza annuale dei cluster dal 2014 al 2016 per le sei città italiane	48
Figura 2.12. Distribuzione dei cluster nelle sei città italiane.....	49
Figura 2.13.....	50
Figura 2.14. Punti di interesse e cluster di un'area di Roma dal 2014 al 2016	51
Figura 3.1. Produzione di Big data da parte del consumatore e settori di interesse	58
Figura 3.2. Interrogazione delle API di Twitter tramite Wolfram Mathematica.....	63
Figura 3.3. Output dei tweet e dei relativi metadati ricavati con Mathematica.....	64
Figura 3.4. Codice sviluppato in Mathematica per l'estrazione del testo dalla tabella di output.....	64
Figura 3.5. Modulo Classify e classificatore incorporato "Sentiment"	65
Figura 3.6. Percentuale di probabilità sull'individuazione delle categorie: positivo, negativo e neutrale	65
Figura 3.7. Informazioni sul metodo applicato per la classificazione del sentiment.....	66
Figura 3.8. Dati riassuntivi relativi ad un hotel su TripAdvisor. Fonte https://www.tripadvisor.it/Hotel_Review-g186338-d187686-Reviews-The_Savoy-London_England.html	71
Figura 3.9. Script per ottenere i commenti	72
Figura 3.10. Metodo applicato per la classificazione	73
Figura 3.11. Training set, test set, probabilità ed accuratezza.....	74
Figura 3.12. Informazione sul metodo adottato per la clusterizzazione ottenuta con l'applicazione di ClassifierFunction.....	80

Figura 3.13. ListPlot del cluster generato correlando la variabile “età” al fattore “disponibilità ad adottare energie rinnovabili”	84
Figura 4.1. Le nove tecnologie che hanno trasformato la produzione industriale.....	89
Figura 4.2. Framework relativo all’organizzazione del corso	93
Figura 4.3. Metadati ricavati da Flickr tramite Wolfram Mathematica.....	94
Figura 4.4. Percorso di un turista ottenuto combinando le coordinate geografiche e le immagini ottenute da Flickr	96
Figura 4.5. Analisi delle frequenze delle risposte alle domande del questionario: sono contento di aver seguito questo corso (a) e le tecnologie e i software appresi sono rilevanti per studiare le dinamiche turistiche (b).....	97
Figura 5.1. Codice sviluppato in Mathematica per l’interrogazione del pacchetto GraphStore	105
Figura 5.2. Codice sviluppato in Mathematica per ricavare tutti i “building” relativi alla città di Firenze.....	105
Figura 5.3. Coordinate ed etichette ottenute in output	105
Figura 5.4. Mappa dei punti di interesse relativi all’etichetta “building” della città di Firenze.....	106
Figura 5.5. Interrogazione di un Database tramite Mathematica	106
Figura 5.6. Output ottenuto dopo l’interrogazione del Database	107
Figura 5.7. Interrogazione di Wikipedia tramite Mathematica	107
Figura 5.8. Codice in linguaggio Javascript	108
Figura 5.9. Interfaccia contenente metadati.....	109

INDICE DELLE TABELLE

Tabella 1. Numero di fotografie scattate dagli utenti in ciascuna città dal 2014 al 2016	39
Tabella 2. Luoghi visitati dagli utenti.....	45
Tabella 3. Confronto dei risultati dell'analisi del sentiment	67
Tabella 4. Output di Mathematica con elenco dei cinque elementi maggiormente fotografati in ogni hotel e numero di volte in cui appaiono	75
Tabella 5. Caratteristiche demografiche del campione	81
Tabella 6. Statistiche descrittive, Media e SD effettuate sul campione	83
Tabella 7. Analisi descrittiva delle risposte alle domande del questionario ideato per misurare la soddisfazione dei partecipanti al corso.....	97
Tabella 8. Percentuali dei dati raccolti	98

APPENDICE A

Data Mining: un insieme di tecniche e metodologie che si occupano dell'estrazione di informazioni significative da grandi quantità di dati.

Facebook: un social media attraverso il quale è possibile svolgere molte attività tra cui pubblicare dei post personali, inviare dei messaggi o chattare in privato con agli amici, condividere foto e video.

Google Moduli: un App inclusa nella suite per ufficio di Google Drive; si presenta come uno strumento che consente di raccogliere informazioni dagli utenti tramite un sondaggio o un quiz personalizzato.

Hashtag: un particolare tipo di tag utilizzato nei social network (ad esempio su Twitter e Instagram) e microblogging che consente agli utenti di applicare un determinato tag e consente ad altri utenti di trovare facilmente i messaggi con un tema o contenuto specifico (identificato da questo tag).

Instagram: un social media attraverso il quale gli utenti possono scattare foto, applicarvi filtri e condividerle sul web.

Marketing Intelligence: un insieme di informazioni accurate sul comportamento dei consumatori (intelligence dei consumatori) e sul comportamento dei concorrenti (informazioni sulla concorrenza) che supportano le aziende nel processo decisionale per identificare le opportunità di mercato e le strategie di successo.

Snippet: un ritaglio, un frammento di testo.

TripAdvisor: la più grande piattaforma di viaggio con in media oltre 455 milioni di visitatori mensili e oltre 630 milioni di recensioni relative ad hotel, ristoranti ed attività commerciali correlate che pubblica informazioni riguardo hotel, ristoranti ed attrazioni.

Twitter: è un servizio di notizie e microblogging, utile ad amici, familiari e colleghi in quanto permette di comunicare e rimanere in contatto attraverso lo scambio di messaggi in modo rapido e frequente.