



Università
Ca' Foscari
Venezia

Corso di Laurea Magistrale
in Marketing e Comunicazione

Tesi di Laurea

Il marketing data-driven: come migliorare l'esperienza del cliente online e aumentare le vendite sull'e-commerce grazie a data science e marketing automation. Applicazioni al caso Tognana Porcellane.

Relatore

Ch.ma Prof.ssa Daniela Favaretto

Laureanda

Martina Peruzzo

Matricola 868442

Anno Accademico

2021 / 2022

Ringraziamenti

Al termine di un percorso ci si ferma a pensare a tutto quello che è stato e, una volta giunti alla fine, ci si volta indietro ricordando i momenti trascorsi. Conservo molti ricordi del mio percorso di studi, alcuni leggeri e spensierati, altri più turbolenti. Ho riso, pianto, sono caduta e mi sono rialzata e tutto questo mi ha fatto crescere e mi ha permesso di arrivare fino a questo punto. A rendere speciali questi momenti sono state le persone con cui li ho condivisi, che mi hanno accompagnato in questo breve ma intenso percorso di vita. Persone che ci tengo a ringraziare perché mi hanno aiutato a raggiungere la vetta, contribuendo a rendere la salita un po' meno faticosa.

Ringrazio la mia relatrice, la professoressa Daniela Favaretto, che mi ha permesso di concludere il mio percorso di studi con serenità, seguendomi durante la tesi ma lasciandomi libertà di scelta e decisione nella stesura dell'elaborato, dimostrandosi esigente ma allo stesso tempo comprensiva.

Ringrazio Valentina, Ilenia e Lara, cuore e mente di Madam Agency, che mi hanno seguita, istruita ed accompagnata nel mio percorso di stage permettendomi di crescere sia professionalmente che umanamente, comprendendo le mie necessità e guidandomi verso la definizione del mio progetto di tesi magistrale.

Ringrazio Luca Peron, E-commerce Manager di Tognana Porcellane, ed i suoi collaboratori che si sono sempre dimostrati gentili e disponibili nel fornirmi tutte le informazioni di cui avevo bisogno per la stesura della tesi.

Un grazie anche ai miei compagni di studi: Aurora, Edoardo, Francesco, Nicolò, Paola, che hanno reso due anni di didattica e lavori di gruppo a distanza più leggeri e mi hanno fatto respirare un po' di quotidianità in un momento in cui nulla sembrava essere normale.

Grazie a Davide, che in tutti questi anni è stato per me un punto di riferimento ed ha saputo darmi la forza di non mollare e la giusta dose di lucidità anche quando tutto sembrava andare storto. Grazie perché il tuo starmi accanto è stato essenziale per diventare la persona che sono e per raggiungere tutti i miei traguardi, che considero in parte anche tuoi.

Il ringraziamento più doveroso e importante va alla mia famiglia che mi ha sempre sostenuta lasciandomi la libertà di cercare e intraprendere la mia strada, aiutandomi e appoggiandomi in tutte le mie scelte. Un grazie particolare a mamma e papà, che oltre ad avermi permesso di proseguire il mio percorso di studi, sono sempre stati per me un pilastro ed un esempio da seguire.

Indice

1. Introduzione	1
Capitolo 1 - Il marketing digitale: l'evoluzione della disciplina fino al giorno d'oggi	3
2. Dal marketing tradizionale al marketing digitale	3
2.1. Marketing 1.0	4
2.2. Marketing 2.0	5
2.3. Marketing 3.0	6
2.4. Marketing 4.0	7
3. Il nuovo consumatore digitale: l'evoluzione del marketing post covid-19	9
3.1. Nuove tendenze e comportamenti	9
3.2. Le sfide del marketing digitale	12
3.3. Marketing 5.0	14
Capitolo 2 – Il marketing data-driven	17
4. I dati	17
4.1. Il ruolo del dato nel marketing nella digital economy	17
4.2. Vari tipi di dati	18
4.3. Big data	20
4.3.1. La definizione di big data	20
4.3.2. Le Vs dei big data	20
4.3.2.1. Il modello a 3V	21
4.3.2.2. Il modello a 5V	22
4.3.2.3. Il modello ITMI	23
5. Data science per il marketing	26
5.1. Cos'è la data science?	26
5.2. I data analytics	28
5.2.1. Descriptive analytics	28
5.2.2. Diagnostic analytics	28
5.2.3. Predictive analytics	29
5.2.4. Prescriptive analytics	30
5.3. Data mining e machine learning	32
5.3.1. Data mining	32
5.3.1.1. Definizione e caratteristiche	32

5.3.1.2.	Sviluppo di un progetto di data mining	34
5.3.1.2.1.	Knowledge Discovery in Databases	34
5.3.1.2.2.	CRISP-DM	36
5.3.2.	Tecniche di data mining	38
5.3.3.	Applicazioni di data mining per l'e-commerce	43
5.4.	Machine learning	45
5.4.1.	Tipologie di apprendimento automatico	48
Capitolo 3 - Il caso Tognana Porcellane		51
6.	Il settore dei prodotti per la tavola, la cucina ed il regalo	52
6.1.	La situazione attuale	52
6.1.1.	Le imprese della filiera	53
6.1.2.	I consumatori	54
6.2.	Trend del settore	55
6.2.1.	Riscoperta della casa	55
6.2.2.	E-commerce	56
6.2.3.	Sostenibilità	57
7.	Tognana Porcellane	58
7.1.	L'azienda	58
7.2.	La storia	59
7.3.	Portafoglio prodotti	59
7.3.1.	La tavola	60
7.3.2.	La cottura	60
7.3.3.	Il tessile e l'arredo casa	61
7.4.	E-commerce e strategia digitale	61
7.4.1.	Lo shop online: struttura e modalità di navigazione	61
8.	Marketing automation per lo shop IT	64
8.1.	Cos'è la marketing automation?	64
8.2.	Il software: SALESmanago	68
8.3.	Campagne di marketing automation	69
8.3.1.	Segmentazione utenti	69
8.3.2.	Automation Workflow	71
8.3.3.	Web personalization e Lead Generation	74
9.	Approccio data-driven e analisi del database clienti	76

9.1. Analisi RFM	77
9.2. Predictive analytics	85
9.3. Cluster analysis	88
9.3.1. Il software KNIME	88
9.3.2. Lavorare in KNIME	89
9.3.2.1. Nodi	90
9.3.2.2. Workflow	91
9.3.2.3. Interfaccia utente	92
9.3.3. Cluster analysis: applicazione pratica	93
9.3.3.1. Fase 1	93
9.3.3.2. Fase 2	97
9.3.3.3. Fase 3	103
9.3.3.4. Fase 4	105
10. Conclusioni	108
11. Elenco delle figure	110
12. Elenco delle tabelle	111
13. Bibliografia e sitografia	112

1. Introduzione

Il mio percorso di laurea magistrale è iniziato quando il mondo si trovava nel pieno di una pandemia globale. Affrontare un nuovo corso di studi in ambito economico nel bel mezzo di un'emergenza sanitaria mi ha permesso di aprire gli occhi su come, non solo il modo di vivere e rapportarsi con gli altri, ma tutti i comportamenti di un'intera società tendano a cambiare in circostanze anomale, e di come sia facile, soprattutto nel mondo del business, che vengano in un attimo ribaltate tutte le regole del gioco e della concorrenza. Vivere una situazione di questo tipo porta a capire come, nell'arco di poco tempo, tutto possa mutare, le priorità personali e collettive vengano rivalutate ed anche i comportamenti dei consumatori si evolvano alla velocità della luce. E le aziende, in primis, devono farsi trovare pronte ad affrontare il cambiamento. Un cambiamento che si è visto in questo caso soprattutto attraverso la digitalizzazione delle attività delle imprese e la crescita delle vendite online. In questo contesto si è più che mai rafforzata l'idea che l'unica cosa che permette alle organizzazioni di gestire presente e futuro, cercando di attenuare i rischi, soprattutto nel mondo digitale, sia la conoscenza. La conoscenza del mercato, dei propri clienti, delle nuove tecnologie, che consente di mettersi al riparo da quello che potrebbe riservare il futuro. Tutto ciò può avvenire solo ed esclusivamente attraverso la raccolta e l'utilizzo dei dati. Avere informazioni a propria disposizione permette di acquisire un vantaggio competitivo, di anticipare e prevedere cosa potrebbe accadere in futuro, creando piani di riserva per situazioni inaspettate. Ma non solo, tenere le redini del proprio business e conoscere i propri utenti permette di migliorare sempre più la qualità dei prodotti e servizi offerti, di far fronte alla crescente concorrenza e a consumatori più informati ed esigenti e di sviluppare una relazione di fedeltà di lunga durata con i propri clienti. Ne ho avuto la prova durante il mio stage curriculare, presso l'agenzia di web marketing Madam Agency Srl. Da questa esperienza lavorativa ho imparato non solo ad avvicinarmi a livello pratico alle realtà aziendali legate al marketing delle performance, ma ho avuto anche la possibilità di capire come i dati siano al giorno d'oggi alla base di tutte le attività che svolge un'azienda, soprattutto online. In seguito alla pandemia gran parte dei business hanno dovuto accelerare la digitalizzazione in atto già da qualche anno. Per sopravvivere alle chiusure forzate, l'evoluzione verso l'online è stata necessaria e ci troviamo oggi in un mondo in cui l'utilizzo del digitale si è moltiplicato, permettendo alle aziende di tenere sempre più traccia delle azioni dei propri clienti, facendone tesoro per lo sviluppo del business. È qui che la gestione delle informazioni diventa fondamentale per migliorare l'esperienza del cliente con i brand e ricercare stabilità in un mercato in continuo mutamento.

L'idea che sta alla base dell'elaborato nasce di conseguenza da un'esperienza vissuta "sul campo", in particolare relativamente all'importanza della gestione e della raccolta del dato nel marketing per sviluppare la relazione con i clienti ed aumentare le vendite in ambito digitale. Nel primo capitolo

viene presentata un'introduzione al mondo del marketing al giorno d'oggi, ripercorrendo l'evoluzione a partire dalla nascita delle pratiche di marketing fino all'arrivo della pandemia, con un focus sui radicali cambiamenti avvenuti negli ultimi anni e sulle principali sfide da affrontare. La trattazione prosegue con un secondo capitolo in cui viene introdotto il tema dei dati, in particolare dei *big data*, e che si focalizza su come sia possibile ricavare informazioni dagli stessi, attraverso gli strumenti della *data science*: a partire dagli *analytics*, fino all'utilizzo di più avanzate tecniche di *data mining* e di algoritmi di *machine learning*. Il cuore della trattazione si sviluppa nel terzo capitolo dove viene introdotto il caso aziendale, Tognana Porcellane. Ho avuto la possibilità di conoscere l'azienda durante il periodo di *stage* curriculare. Il reparto marketing di Tognana collabora con l'agenzia in cui ho lavorato per un progetto in ambito di *marketing automation*, nella gestione dei flussi automatici di comunicazione con i clienti e di personalizzazione del sito web B2C, oltre che per il tracciamento dei dati raccolti attraverso Google Analytics. Il capitolo presenta, in primo luogo, il settore di riferimento, introducendo la situazione attuale del mercato ed i trend che lo caratterizzano. In secondo luogo, vengono introdotti l'azienda e la sua storia, i prodotti che offre e la strategia digitale, con un focus sulle caratteristiche della navigazione all'interno dell'e-commerce B2C. Viene presentato poi un approfondimento sulle principali attività di *marketing automation* sviluppate dall'agenzia ed un'analisi del database clienti. Quest'ultima rappresenta la parte più personale dell'elaborato, poiché si struttura principalmente in un commento ad un'analisi RFM, seguito da un'analisi di *clustering* effettuata attraverso il software di *data mining* e *machine learning* KNIME. L'obiettivo di questa trattazione è quello di mostrare come strumenti diversi di *data science* e *marketing automation*, guidati dall'intelligenza artificiale, possano integrarsi tra loro ed offrire da un lato uno snellimento nei processi di CRM all'interno delle aziende, e dall'altro lato delle opportunità per sviluppare dei *customer journey* digitali personalizzati per migliorare la *customer experience* nei canali digitali, aumentando, di conseguenza, ROI e vendite attraverso gli e-commerce.

Capitolo 1

Il marketing digitale: l'evoluzione della disciplina fino al giorno d'oggi

2. Dal marketing tradizionale al marketing digitale

Il mondo in cui viviamo è in continuo mutamento. Da un punto di vista economico, lo possiamo notare sia a livello macro, per cui stiamo assistendo a continui spostamenti di potere dalle nazioni occidentali a quelle asiatiche emergenti come la Cina, che a livello micro, con un cambiamento rilevante nel profilo dei consumatori nel tempo che sono oggi sempre più giovani, produttivi e con un livello di reddito più elevato. Stiamo assistendo a cambiamenti in molti settori; negli anni siamo passati da una produzione ad alto volume ad una produzione di nicchia, con aziende rinnovate e che si avvicinano a nuove attività e nuovi campi di business. L'estensione dei processi di mondializzazione e globalizzazione ha livellato l'area di "battaglia" in cui si confrontano le aziende per l'occupazione del mercato. Poter accedere a prodotti e servizi da tutto il mondo ha permesso anche alle realtà più piccole di competere con i leader del mercato per l'acquisizione degli stessi clienti e per la leadership di settore, permettendo, alle volte, di avere anche la meglio. Oggi le PMI sono costrette a pensare fuori dagli schemi e trovare ispirazioni da fonti esterne dal mercato, ivi compresi i consumatori stessi, per cui si sta assistendo ad un passaggio da processi di sviluppo verticale ad orizzontale. Mentre fino ad alcuni anni fa le aziende convincevano i consumatori ad acquistare attraverso efficaci campagne di marketing, ora devono necessariamente tenere conto di altri fattori esterni all'impresa su cui i consumatori fanno riferimento, che sono specialmente: amici, famiglia, colleghi, post sui social media. I principali fattori di influenza non sono più preferenza, motivazione, esperienza e opinione altrui, ma sono fattori esogeni, tra i più diversi e personali. Alcuni importanti cambiamenti sono stati causati e spinti dallo sviluppo e dall'uso estremo di Internet degli ultimi anni. Questo strumento di comunicazione ci ha portato ad una maggiore connettività, ci ha avvicinato sia come persone che come aziende, rendendoci allo stesso tempo più consapevoli ma anche più vulnerabili. Con lo sviluppo delle piattaforme informatiche e dei social media, il modo in cui gli individui comunicano è cambiato drasticamente. Queste piattaforme di comunicazione online hanno unito le persone, infranto barriere geografiche e demografiche e hanno aiutato a creare forti relazioni personali tra le persone. È stato in gran parte questo forte sviluppo dell'era di Internet dell'ultimo decennio a spostare il mondo degli affari: con le nuove tecnologie e metodi di produzione si creano oggi prodotti tecnologicamente avanzati, adatti a tutti i tipi di mercato e a soddisfare tutte le richieste del consumatore.

In questo contesto di evoluzione del mondo del 21esimo secolo, il marketing, come altre discipline, si è adattato al progresso dell'umanità, delle tecnologie, delle conoscenze e dei bisogni [32].

Nel secolo scorso, tra gli anni Cinquanta e Sessanta del Novecento, l'attenzione delle aziende era completamente focalizzata sulla produzione e le attività commerciali e pubblicitarie non avevano nulla a che vedere con il marketing come lo intendiamo oggi. Tutto ciò che era prodotto veniva venduto e agli imprenditori non interessava delle richieste dei propri clienti perché si focalizzavano sulla qualità del prodotto e, al massimo, sul packaging, causando una concentrazione verso l'innovazione di prodotto a scapito delle tecniche di vendita. Solo con il tempo il marketing si è evoluto in un'attività strategica, con le aziende che hanno piano piano iniziato a prestare attenzione alle esigenze del cliente e ad utilizzare tecniche di segmentazione e *targeting*, per dare maggiore impatto a queste attività. Dall'approccio produttivo siamo arrivati ad un marketing, come lo conosciamo oggi, dove il cliente viene prima e sta alla base del processo di produzione perché viene consultato sui suoi bisogni e soddisfatto nei suoi desideri [66]. Così si è sviluppata la grande rivoluzione del marketing, che ha condotto da una visione *product-oriented* delle attività ad una visione *consumer-oriented* in cui si mettono in risalto i bisogni del cliente, dal valore razionale, ed i suoi desideri, che hanno invece un valore irrazionale.

Lo sviluppatore di questo concetto di marketing evolutivo è il famoso professore Philip Kotler, considerato il padre del marketing moderno, che, attraverso le sue pubblicazioni e con l'aiuto di numerosi altri autori, ha definito una serie di fasi che il marketing ha attraversato nel corso del secolo scorso e che ancora tutt'oggi sta attraversando. Con i suoi scritti ha definito l'evoluzione dal marketing 1.0 al marketing 4.0, fino ad arrivare ad una nuova recente fase che si è sviluppata in seguito alla pandemia da covid-19, il marketing 5.0 [32]. In questa prima parte della trattazione, ci focalizzeremo su come si sia evoluto il concetto di marketing a partire dal secolo scorso e su come i cambiamenti avvenuti in questa disciplina vengano applicati al mondo in cui ci ritroviamo oggi.

2.1. Marketing 1.0

Il concetto di Marketing 1.0 fa riferimento all'evoluzione della teoria e della pratica del marketing nella prima metà del secolo scorso, specialmente a partire dagli anni 50 negli Stati Uniti. Il marketing 1.0 nasce infatti in un momento in cui la produzione di massa era in pieno svolgimento e non c'erano criteri di consumo, veniva semplicemente consumato ciò che il mercato offriva, indipendentemente dalle diverse caratteristiche di prodotti o servizi. Le aziende in questo periodo investono solamente nello sviluppo del prodotto e la comunicazione con il cliente è unidirezionale, ovvero sono solo le aziende a fornire informazioni relativamente al prodotto ed alle offerte [66]. Possiamo definire quindi il marketing 1.0 come marketing centrato sul prodotto, con obiettivo principale quello di creare

prodotti e servizi perfetti, che generassero il massimo valore per il cliente. La concorrenza, infatti, si basava sull'offerta di prodotti dotati di maggiori funzionalità e che offrivano il massimo dei benefici ai consumatori, richiedendo prezzi anche molto elevati.

Il concetto di marketing in quest'epoca era quindi focalizzato sullo sviluppo del prodotto e sul suo ciclo di vita, oltre che sulla creazione e definizione delle 4P più efficaci. L'obiettivo primario era la soddisfazione del consumatore, ma questo ha portato anche a grossi limiti: le aziende così facendo inducevano spesso i clienti ad acquistare anche beni di cui non avevano realmente bisogno e questo ha portato, a sua volta, allo sviluppo della cultura del consumismo [40].

2.2. Marketing 2.0

L'evoluzione al concetto di marketing 2.0 è arrivata nel momento in cui ha iniziato a farsi spazio l'era dell'informazione contemporanea, in cui iniziano a svilupparsi le tecnologie della comunicazione e dell'informazione. Le maggiori sfide che le aziende si trovano a dover affrontare in questo periodo attraverso le azioni di marketing sono principalmente legate al fatto che il consumatore è diventato più consapevole ed informato. Con lo sviluppo delle tecnologie dell'informazione è stato possibile iniziare a confrontare prodotti e servizi della concorrenza e ricercare facilmente informazioni a riguardo. Il valore del prodotto o servizio venduto da un'azienda diventa quindi ciò che il consumatore ritiene che valga, di conseguenza da parte delle aziende diventa fondamentale ascoltare e cercare di soddisfare i bisogni ed i desideri dei consumatori sempre più esigenti ed informati [32].

Il marketing 2.0 si concentra quindi su un approccio più *customer-oriented*, in un contesto in cui il cliente presta attenzione e confronta offerte diverse, prendendo decisioni. L'obiettivo delle aziende diventa di conseguenza non solo la soddisfazione del cliente, ma anche la fidelizzazione dello stesso. Nascono a questo scopo i primi centri assistenza clienti e, con il web 2.0, diventano rilevanti i media digitali e i social media, che consentono di ricevere feedback dalle esperienze degli utenti. È in questo contesto che nasce il social media marketing [66].

In quest'era, il marketing si è concentrato principalmente sulla comprensione della segmentazione, del *targeting* e del posizionamento. Le imprese non cercavano più di creare prodotti e servizi perfetti per tutti, ma si concentravano sulla conoscenza approfondita del proprio mercato di riferimento, per definire con precisione il proprio posizionamento. L'attenzione non era più focalizzata su fronzoli e abbellimenti d'effetto, ma sulle caratteristiche specifiche dei prodotti, definite in relazione ai bisogni e desideri dei consumatori. Il tutto ebbe un riflesso anche sui prezzi di vendita, che venivano fissati su un livello corretto per i segmenti individuati come *target*. Inoltre, le imprese iniziarono anche a sforzarsi nel costruire una relazione durevole con i consumatori, attraverso l'utilizzo dei primi

approcci di CRM per fidelizzare i clienti e non perderli a scapito della concorrenza. L'obiettivo di questo marketing 2.0 passò rapidamente dalla soddisfazione alla fidelizzazione del cliente [40].

2.3. Marketing 3.0

Con il marketing 3.0 si passa dall'era del marketing orientato al consumatore all'era del marketing orientato al valore. Kotler, nel suo libro Marketing 3.0, spiega come anche in questa fase ci si proponga di soddisfare il consumatore ma, a differenza della precedente, *“le imprese che lo applicano possiedono anche missioni, visioni e valori volti a dare un contributo al mondo”* [39]. Sostanzialmente, ciò significa che le imprese non si limitano a fornire beni e servizi ma si impegnano e danno il proprio contributo anche nel risolvere problemi legati alla società. *“Il marketing 3.0 eleva il concetto di marketing alla sfera delle aspirazioni, dei valori e dello spirito (l'”anima”) dell'umanità”* [39]. In questa fase il consumatore viene considerato come vero e proprio essere umano, nel senso proprio del termine, con esigenze e speranze che vanno oltre al semplice bisogno o desiderio di acquisto di beni e servizi. Kotler, infatti, definisce il marketing 3.0 come un'integrazione tra il marketing emozionale ed il marketing “umano” [39].

Questo passaggio raggiunge l'apice con la fine degli anni 2000 con l'ascesa della generazione y e le turbolenze causate dalla crisi finanziaria mondiale. In questo periodo, vista la grande quantità di informazioni disponibili e gli scandali che stavano travolgendo il settore finanziario, la stima verso le imprese che dimostravano di avere obiettivi legati esclusivamente al profitto stava calando e la generazione y ha iniziato a chiedere alle imprese di produrre beni e servizi non solo che rispondessero ai loro bisogni, ma che impattassero positivamente sulla società, la cultura e l'ambiente circostante [40]. In tempi di crisi economica globale, il marketing 3.0 diventa fondamentale per la vita di consumatori costretti sempre più frequentemente a fare i conti con cambiamenti e inquietudini sociali, economiche e ambientali rapide e drastiche. Le imprese che praticano marketing 3.0 forniscono risposte e ragioni di speranza a chi si deve confrontare con questi problemi e, di conseguenza, impatta sui consumatori a un livello più elevato. Nel marketing 3.0 le imprese si differenziano in base ai valori di cui sono portatrici. E, in tempi turbolenti, è una differenziazione che conta. Siamo in un periodo in cui le principali forze che plasmano l'ambiente di business sono: la partecipazione, il paradosso della globalizzazione, la creatività sociale. La centralità di questi tre aspetti aprono un'era in cui il consumatore si trasforma, diventando più collaborativo e più attento agli aspetti culturali del consumo, facendosi guidare, come mai finora, dagli aspetti più spirituali dell'animo umano. Questo spiega come il marketing 3.0 rappresenti l'apice del marketing tradizionale ed un'integrazione equilibrata di marketing collaborativo, culturale e spirituale [39].

È da qui che nasce il marketing centrato sull'essere umano, nel momento in cui le imprese hanno iniziato a farsi carico anche di questioni non prettamente aziendali, integrando nel proprio modello di business anche attività di marketing etico, sostenibile e socialmente responsabile. I consumatori nell'era del marketing 3.0 non scelgono i brand solo per funzionalità o soddisfazione emotiva, ma danno molta importanza alla propria realizzazione spirituale. Di conseguenza, la differenziazione tra le imprese non risiede più nelle caratteristiche intrinseche dei prodotti, ma nei valori di cui si fanno portavoce, nei prodotti che possono fornire soluzioni ai maggiori problemi ambientali e sociali del mondo [40].

Di seguito vediamo come Kotler in Marketing 3.0 faccia un confronto tra le tre fasi del marketing tradizionale.

	Marketing 1.0 <i>Marketing focalizzato al prodotto</i>	Marketing 2.0 <i>Marketing focalizzato al consumatore</i>	Marketing 3.0 <i>Marketing guidato dai valori</i>
<i>Obiettivo</i>	Vendere prodotti	Soddisfare e fidelizzare i consumatori	Rendere il mondo un posto migliore in cui vivere
<i>Fattori abilitanti</i>	Rivoluzione industriale	Tecnologia dell'informazione	Nuova ondata tecnologica
<i>Come le imprese guardano al mercato</i>	Mercato di massa con esigenze di beni fisici	Consumatore intelligente con esigenze anche immateriali	Essere umano nel pieno senso del termine, con esigenze materiali e spirituali
<i>Concetto chiave di marketing</i>	Sviluppo prodotti	Differenziazione	Valori
<i>Linee guida dell'impresa</i>	Caratteristiche dei prodotti	Posizionamento dell'impresa e dei prodotti	Missione, visione e valori dell'impresa
<i>Proposte di valore</i>	Funzionali	Funzionali ed emotive	Funzionali, emotive e spirituali
<i>Interazione con i consumatori</i>	Transazione da uno a molti	Rapporto da uno a uno	Collaborazione da molti a molti

Figura 1: Confronto tra gli aspetti del marketing 1.0, 2.0, 3.0 (Fonte: Marketing 3.0, 2011)

2.4. Marketing 4.0

Si inizia a parlare di marketing 4.0 con l'ingresso delle aziende nell'era digitale, con lo sviluppo dell'intelligenza artificiale, degli algoritmi predittivi e del *machine learning*, che rendono più semplice prevedere comportamenti e aspettative dei consumatori [51]. Philip Kotler lo definisce come “*un approccio di marketing che combina l'interazione online e offline tra aziende e consumatori*” (Kotler, 2017). Rappresenta l'evoluzione naturale del marketing 3.0, che unisce le potenzialità della macchina e della nuova intelligenza artificiale ad altre tecnologie ITC per aumentare

la produttività, sfruttando allo stesso tempo la connettività tra uomo e uomo per migliorare il processo di interazione e di relazione con il cliente [32].

Con il marketing 4.0 l'obiettivo si sposta sull'acquisizione della fiducia del cliente e sulla sua fidelizzazione, utilizzando l'interazione tra la parte migliore del marketing tradizionale e l'online del marketing digitale. Marketing 4.0 e innovazione digitale vanno di pari passo, in un mondo in cui il consumatore ha meno tempo, vuole le cose subito e cerca varie possibilità di scelta. Il compito del marketing in questa era diventa quello di raggiungere il cliente con un messaggio che vuole davvero sentire. A fronte di una concorrenza sempre più globale, diventa fondamentale la ricerca di una comunicazione a 360 gradi che sia il più possibile diretta, prendendo in considerazione sia i canali offline che quelli online, come i social network, strumento essenziale per prevedere tendenze e individuare nuovi possibili scenari attraverso i contenuti generati dagli utenti stessi.

In questa nuova fase di evoluzione del marketing, assistiamo anche alla comparsa dei *big data* ed alla nascita della cultura aziendale del dato. Nuove opportunità vengono offerte dall'integrazione dell'analisi dei dati anche nei reparti di marketing, con la possibilità di ottenere preziose informazioni sui clienti attuali a potenziali su larga scala, definendo in maniera più chiara i loro gusti, le preferenze e gli interessi. L'obiettivo diventa capire il comportamento del proprio *target* di consumatori, prevederlo e anticiparlo grazie alle potenzialità dell'analisi dei *big data*. Per sfruttare le possibilità offerte da questi nuovi flussi di dati e informazioni, le aziende si vedono costrette in questa fase a cambiare ed evolvere la loro cultura decisionale, educando dipendenti e collaboratori ad un nuovo modo di fare business, per il quale le decisioni basate sui dati tendono ad essere le decisioni migliori [66].

Il marketing 4.0 apre le aziende ad un nuovo modo di relazionarsi, ai brand che si "umanizzano" per stabilire connessioni emotive con il cliente ed allo sviluppo di una comunicazione che si basa sempre più su un approccio multicanale. L'importanza si sposta nel trasmettere correttamente i valori del marchio, instaurare legami con il consumatore creando contenuti di valore. Il lavoro delle aziende passa dalla gestione dell'acquisto di prodotti e servizi offline alla gestione ed al monitoraggio degli acquisti online. Il digital marketing diventa il motore del marketing 4.0, con le imprese che investono sempre di più sulle attività online per promuovere il proprio brand ed i propri prodotti. Da elemento di supporto, il marketing digitale diventa in questa fase la strategia principale per raggiungere il cliente e aumentare le vendite a livello internazionale, permettendo di fornire contenuti in maniera sempre più semplice e rapida. Il marketing richiede ora un approccio necessariamente omnicanale, da implementare sia online che offline [51].

3. Il nuovo consumatore digitale: l'evoluzione del marketing post covid-19

Ogni fase di evoluzione del concetto di marketing è stata segnata da eventi naturali, sociali, culturali, politici che ne hanno influenzato lo sviluppo. Dalla nascita delle tecnologie dell'informazione, all'era del consumismo e della regressione, dalla crisi finanziaria, fino all'arrivo di moderne tecnologie e dell'intelligenza artificiale. Tutti questi avvenimenti che hanno fatto la storia sono stati alla base dello sviluppo del comportamento dell'uomo e di tutte le attività che ne studiano l'evoluzione, come il marketing stesso.

Nel 2020, è stata la pandemia da covid-19 a stravolgere il mondo che ci circonda, costringendoci a nuove abitudini e stili di vita e portando sia aziende che consumatori ad adottare nuovi comportamenti digitali. Le abitudini di acquisto ormai radicate sono state abbandonate e l'accelerazione nell'adozione digitale ha portato in pochi mesi a sviluppi che si sarebbero attesi nell'arco di un decennio. Le aziende stesse si sono viste costrette ad innovare per trovare nuove modalità per connettersi ai clienti e, le nuove tecnologie, hanno permesso di elaborare enormi quantità di dati granulari sui consumatori, che hanno nuove esigenze da soddisfare e si aspettano un'offerta di esperienze sempre più personalizzate [48].

3.1. Nuove tendenze e comportamenti

I comportamenti mutevoli dei consumatori causati dal covid-19 hanno portato all'accelerazione delle tendenze, all'emergere di nuove preferenze e ad un completo capovolgimento di una serie di abitudini radicate. L'ambiente e le preferenze dei clienti continueranno ad evolversi, quindi le aziende devono essere preparate ad adattarsi, iterare e mettere in pratica i cambiamenti all'interno delle proprie organizzazioni [47].

I dati raccolti da Google Trends mostrano in primo luogo che durante il periodo pandemico i consumatori hanno assunto dei comportamenti sempre più digitali. Questa evoluzione è stata una necessaria conseguenza delle misure di distanziamento sociale imposte dai governi di molti Paesi, che hanno reso i canali digitali l'unico modo per poter rimanere in contatto con altre persone [50]. Come conseguenza a questa nuova situazione, in tutto il mondo le aziende si sono mosse per cercare di accogliere il repentino passaggio al digitale, orientando verso questa direzione i loro modelli di business. Oggi ogni attività, dall'acquisto di pasti o generi alimentari, alla finanza e all'istruzione, fino al fitness e benessere, ha sviluppato un equivalente digitale o online, con un aumento di utenti negli ultimi due anni ed è molto probabile che, terminata la crisi dovuta dalla pandemia, i consumatori preferiscano continuare a sfruttare queste nuove opportunità digitali [47].

Ciò che sorprende è stata la velocità con cui l'adozione digitale è avvenuta. I consumatori si sono abituati rapidamente al nuovo stile di vita digitale, affidandosi al canale e-commerce ed alle app di consegna di generi alimentari e della ristorazione per soddisfare le necessità quotidiane. Questo ha portato anche ad un aumento rilevante dell'utilizzo del *digital banking* e delle modalità di pagamento senza contanti per le operazioni di acquisto. Le piattaforme di videoconferenza come Zoom e Google Meet hanno permesso alle persone di parlare, lavorare e di usufruire della didattica a distanza direttamente da casa e, non potendo uscire, anche il numero di streaming su YouTube e Netflix è aumentato vertiginosamente [40]. L'online è diventato, durante il *lockdown*, il primo canale di acquisto per molti consumatori. Google ha registrato, solo in Italia, un aumento di nuovi acquirenti online triplicato nei primi 5 mesi del 2020, rispetto all'anno precedente. L'accelerazione digitale che ha spinto all'adozione dell'e-commerce come canale di acquisto primario non ha digitalizzato solo il settore della vendita al dettaglio: i consumatori hanno iniziato ad utilizzare Internet per attività per cui prima non avrebbero probabilmente nemmeno utilizzato i canali digitali, come lezioni online, chat dal vivo o palestre virtuali [50]. Secondo i dati raccolti da McKinsey, da gennaio 2020 si è registrato un aumento quasi del 20% della spesa online. Il 92% dei consumatori che nel 2019 ha provato ad acquistare online si è convertito, consolidando quella che sembrava una risposta all'emergenza pandemica in un'abitudine indispensabile, che ha ridefinito le comodità [48].

La pandemia ha messo a rischio anche la fidelizzazione del consumatore ai brand. Nel periodo pandemico è stato registrato un livello di cambio canale e interruzione della fedeltà al marchio senza precedenti. Il 75% dei consumatori ha provato nuovi comportamenti d'acquisto, guidati dalla convenienza e dalla ricerca di esperienze con maggior valore. Il 39% degli stessi, appartenenti principalmente alle generazioni y e z, ha abbandonato i marchi di fiducia per provarne dei nuovi e riuscire a trovare i brand che riflettano i propri valori [48].

Attraverso il Report 2021 sull'e-commerce nell'area EMEA, anche Adobe ha identificato i più frequenti comportamenti di acquisto online dei consumatori nel corso dello scorso anno. Il primo anno di pandemia ha segnato un drastico cambiamento per gli acquisti sia online che in negozio, ma comportamenti e preferenze dei consumatori si sono dimostrati in continua evoluzione nel corso dell'anno, mantenendo come costante la crescita dello *shopping* online. Ad esempio, rispetto al 2020, le preferenze di dispositivo sono cambiate; mentre prima gli utenti preferivano acquistare tramite computer portatile (45%) invece che attraverso lo smartphone (36%), ora grazie alle innovazioni come ad esempio Apple Pay, l'equilibrio si è ribaltato ed il 78% tra gli intervistati da Adobe, afferma di aver effettuato un acquisto con lo smartphone nell'ultimo anno a fronte di un 64% che ha preferito utilizzare il pc. Con la riapertura dei negozi fisici nel 2021, la frequenza online non è diminuita con

quasi un quarto degli acquirenti che si possono definire oggi abituali (13%), con una frequenza media di acquisto che va dalle 11 alle 20 volte al mese, e frequenti (11%), con una media di acquisti superiori alle 20 volte al mese. Anche con la riapertura, la metà degli intervistati prevede di mantenere lo stesso volume di acquisti online, mentre un 44% prevede un aumento delle proprie spese. L'online si è rivelato un canale preferenziale per i consumatori che prediligono la praticità, la comodità e che hanno mancanza di tempo per effettuare gli acquisti in negozio fisico, nonché per coloro che ricercano un'esperienza di *shopping* più semplice. Il periodo pandemico che ha forzato molti acquirenti ad utilizzare i canali digitali, ha portato anche ad una maggiore fiducia nello *shopping* online, con utenti di diverse fasce demografiche che si dimostrano più a loro agio con l'e-commerce e maggiormente aperti a nuove esperienze in tutte le fasi del loro *customer journey*. Lo dimostra il fatto che negli ultimi 12 mesi il 44% degli utenti intervistati afferma di aver acquistato degli articoli online da rivenditori che non conoscevano prima, confermando il trend della sempre più crescente difficoltà da parte dei brand di creare una relazione di fedeltà con i clienti. La crescente fiducia verso lo *shopping* online ha aumentato anche la tranquillità nell'utilizzo delle tecnologie e dei pagamenti elettronici, portando ad una maggiore sicurezza anche nell'acquisto di prodotti più costosi. [3]

Il cambiamento che sta avvenendo alle preferenze ed alle aspettative dei consumatori non si limita solo allo *shopping* online. In questo periodo si è rafforzato anche il legame dei clienti con i propri valori e la propria etica, che si riflette sui loro comportamenti di acquisto. Tra gli utenti intervistati per la ricerca di Adobe, il 58% degli acquirenti (6 su 10) afferma che potrebbe smettere di utilizzare prodotti di un determinato marchio se non lo ritenesse compatibile con i propri valori. Essendo sempre più coinvolto in importanti discussioni sui problemi del mondo che li circonda, come cambiamento climatico e diseguaglianze sociali, il consumatore oggi è più consapevole e vuole avere un maggiore controllo nella scelta su come e dove effettuare gli acquisti. Valori e identità individuale stanno diventando aspetti fondamentali per la differenziazione del brand e si richiede di conseguenza sempre più alle aziende di comprendere questi valori ed abbracciarli, lavorando su empatia e sensibilità maggiori per rispondere alle aspettative crescenti del cliente. [3]

I nuovi comportamenti dei consumatori che si sono sviluppati durante la pandemia hanno portato ad importanti novità che sono destinate a restare. Le aspettative dei clienti non si sono solamente evolute, ma stanno diventando variegata, allargata e spesso complesse da soddisfare. Gli acquirenti non sempre hanno interesse nelle esperienze innovative, ma la portata e la frequenza con cui hanno accolto e adottato le nuove tecnologie negli ultimi tempi sono la prova di come non appena esse vengano accettate, la domanda cresca in maniera esponenziale. In questo periodo post-pandemico, i consumatori sono aperti alle nuove esperienze come mai prima d'ora. Ricercano un percorso di

acquisto sempre più chiaro e semplificato, vogliono avere la libertà di fare una scelta bilanciata che sia il più possibile semplice ma anche personalizzata, affidandosi ad un marchio che sia guidato dagli stessi valori etici e che dia il proprio contributo nella risoluzione dei problemi sociali e ambientali [3].

3.2. Le sfide del marketing digitale

Dopo aver approfondito come si sono evoluti i comportamenti di acquisto nel corso degli ultimi due anni, è necessario capire come le aziende devono affrontare il nuovo contesto in cui si stanno sviluppando i mercati e come sfruttare le attività di marketing come leva per spingere le vendite in un mondo sempre più digitale.

A livello sociale, sono tre le principali sfide che i *marketers* devono affrontare oggi: il gap generazionale, la polarizzazione delle aspettative ed il tema del *digital divide*. Oggi convivono sulla terra 5 differenti generazioni (*baby boomers*, generazione x, y, z, alfa) che hanno comportamenti e preferenze in conflitto e detengono posizioni differenti all'interno delle aziende e della società stessa. Questo porta molte volte alla disconnessione tra le idee e le preferenze dei più anziani che prendono le decisioni ed i manager o i clienti stessi, creando enormi ostacoli nelle decisioni aziendali e nella scelta dei prodotti da offrire sul mercato. Le cinque generazioni presentano anche diversi livelli di competenza tecnologica e sono plasmate da un diverso ambiente socioculturale e da esperienze di vita differenti, richiedendo ai mercati offerte, *customer experience* e modelli di business a volte anche molto distanti tra loro. Inoltre, la maggior parte delle imprese non ha le risorse necessarie per poter soddisfare le richieste di tutte le generazioni di utenti: il ciclo di vita del prodotto è sempre più corto, con esigenze e desideri dei giovani consumatori in continua evoluzione, il portafoglio prodotti è spesso rigido ed i servizi offerti non permettono una customizzazione per ogni generazione, costringendo le aziende a scegliere di servire un massimo di due o tre generazioni contemporaneamente.

Anche la polarizzazione del mercato rappresenta un'enorme sfida; i *marketers* devono affrontare disuguaglianze e squilibri nella distribuzione della ricchezza che portano il mercato intermedio a contrarsi, costringendo lo spostamento degli operatori del settore verso l'alto o verso il basso. Nonostante ciò che ha portato la pandemia ed i passi avanti che sono stati fatti, bisogna ancora fare il conto con il divario che esiste tra coloro che sono fiduciosi del potenziale delle nuove tecnologie e della digitalizzazione e coloro che ne sono ancora scettici. La digitalizzazione, accelerata dalla pandemia, ha portato infatti con sé una sorta di paura dell'ignoto, di minaccia per la perdita dei posti di lavoro e di preoccupazione per le violazioni della privacy sul web, accompagnata anche da una promessa di crescita esponenziale e di una vita migliore per l'umanità [40].

Dalle interviste effettuate a manager aziendali dal Sole24Ore per valutare l'impatto dell'accelerazione digitale post-pandemica sulle strategie digitali, sembra che oggi poco sia rimasto dei modelli e delle azioni di marketing impiegati prima dello scoppio della pandemia, dai canali utilizzati per attirare i clienti, fino a quelli dedicati alla gestione dei flussi interni. Per creare valore con le attività di marketing nella società di oggi sono diventati di fondamentale importanza il coinvolgimento dell'audience in tempo reale, l'innovazione digitale e la coesione e relazione tra i diversi canali in ottica di omnicanalità. Durante la pandemia, gli annunci sui canali digitali sono stati lo strumento in maggiore aumento e gran parte dei dipartimenti di marketing nelle aziende, ha ridefinito metriche e KPI per valutare le proprie performance in modo più accurato e puntuale, considerando maggiormente gli indici relativi alla soddisfazione del cliente, i costi di acquisizione ed il *content engagement*. Siamo giunti in una fase in cui l'esperienza dell'utente sul proprio sito o e-commerce vale quanto la qualità dei prodotti e servizi offerti e rappresenta un punto fondamentale per acquisire un vantaggio competitivo sui competitors [38].

Anche la creazione di un'ottima *customer experience* per i propri clienti è diventata oggi una delle maggiori sfide per chi lavora nel marketing. Con il cambiamento radicale delle esigenze, delle aspettative e dei comportamenti dei consumatori a cui abbiamo assistito nell'arco di alcuni mesi, le aziende hanno dovuto ripensare e adattare non solo le strategie di comunicazione e di marketing, ma l'intera esperienza di acquisto del cliente. Secondo l'Head of Marketing di Fineco, in un'intervista al Sole24Ore, le aree fondamentali su cui concentrarsi dopo questo periodo sono principalmente due: l'innovazione ed il *meaning and purpose*. Ciò significa indirizzare la propria attenzione sulla creazione di *marketing journey* sempre più customizzati, con la definizione di contenuti mirati e interazione continua con gli utenti attraverso un utilizzo massivo dell'automazione e mettendo al centro l'analisi e l'elaborazione dei dati nelle fasi strategiche. Il fatto che la pandemia abbia stravolto anche il nostro sistema valoriale ha reso necessario riportare il focus del marketing sulla *human experience* e sulla centralità e partecipazione dell'utente nello sviluppo della relazione con il brand, portando l'azienda a focalizzarsi di conseguenza sul proprio *meaning e purpose* [38].

Questo rapido spostamento del processo di acquisto di beni e servizi verso il canale online, ha costretto le aziende a focalizzare le strategie di acquisizione del consumatore sul digitale. Inoltre, è aumentata anche la consapevolezza che fosse necessario dotarsi di strumenti adatti a comprendere e soddisfare le nuove esigenze: è per questo che i responsabili del marketing hanno iniziato ad orientarsi all'innovazione, all'implementazione di tecnologie *martech*, all'utilizzo e gestione dei dati ed alla personalizzazione in tempo reale per migliorare l'esperienza di acquisto dell'utente e renderla omogenea, sia a livello di canale che di dispositivo utilizzato [38].

Con l'aumento del digitale, le aziende hanno ora a portata di mano anche una grande quantità di dati dinamici. Ecco perché, per offrire esperienze di valore in un ambiente in continua evoluzione, è fondamentale investire nei dati e nella tecnologia, per anticipare e prevedere il *sentiment* del cliente ed il suo valore. Questo significa per le aziende sfruttare i dati che hanno a disposizione e servirsi di strumenti in grado di estrarre informazioni da questi dati adottando un approccio proattivo all'evoluzione dei comportamenti e rispondendo in tempo reale alle esigenze del cliente, andando oltre all'approccio reattivo caratterizzato dall'utilizzo di sondaggi e interviste. Acquisire tutti i dati dei clienti, non solo di un campione di intervistati, può permettere di ottenere una visione più accurata delle esigenze e delle aspettative dei consumatori e, grazie alle informazioni raccolte, è possibile intervenire in maniera più significativa e focalizzarsi sulla fidelizzazione del cliente, aumentando il valore dello stesso e riducendo i costi di servizio. Per questo motivo gli investimenti su sistemi di raccolta ed elaborazione dati completi e predittivi possono consentire alle imprese e organizzazioni di rispondere rapidamente alle esigenze dei clienti anche durante i periodi di crisi come quello appena trascorso [47].

3.3. Marketing 5.0

Finora ci siamo focalizzati su come la pandemia da covid-19 abbia di fatto accelerato la digitalizzazione delle imprese e su come, con il *lockdown* e le politiche di distanziamento fisico, sia i *marketers* che mercati e consumatori si siano dovuti adattare rapidamente ad una nuova realtà rinnovata, *touchless* e digitale. L'interesse e le preoccupazioni delle generazioni dei più giovani che stanno vivendo un mondo in rapido cambiamento, sono oggi focalizzate sul desiderio di apportare uno sviluppo positivo all'umanità e sul miglioramento della qualità della vita, promuovendo allo stesso tempo il progresso tecnologico in tutti gli aspetti dell'esistenza umana. In seguito allo sviluppo del marketing digitale e all'approccio omnicanale che hanno caratterizzato il marketing 4.0, con l'ascesa della generazione z e della generazione alfa, è arrivato per il marketing il tempo di evolversi nuovamente per adattarsi ai recenti mutamenti del mercato. È in questo contesto che si sviluppa il marketing 5.0.

Nel suo ultimo libro Philip Kotler definisce il Marketing 5.0 come *“l'applicazione di tecnologie che imitano l'umano per creare, comunicare, offrire e aumentare il valore lungo il customer journey”* (Kotler 2021). Uno tra gli aspetti critici di questa nuova fase è l'utilizzo di quell'insieme di tecnologie avanzate che mirano ad eguagliare le competenze ed abilità dei marketer umani. Sono ad esempio l'AI, NPL, i sensori, VR, AR, IoT, la robotica, il *blockchain*, la cui combinazione costituisce il fattore abilitante del marketing 5.0. Questo tipo di tecnologia va ben oltre la distribuzione di contenuti sui social media o la costruzione di un'esperienza omnicanale per il cliente. Le nuove tecnologie come

l'intelligenza artificiale, l'elaborazione del linguaggio naturale, i sensori e l'internet delle cose stanno cambiando le regole del gioco in ambito di marketing all'interno delle imprese. Nonostante gran parte delle tecnologie avanzate fossero già presenti nella seconda metà del secolo scorso, il fatto che si siano affermate solo negli ultimi anni è dovuto alla mancanza in precedenza di tecnologie abilitanti adatte al loro sviluppo. Ciò che rende possibile l'ascesa di queste tecnologie avanzate è infatti il raggiungimento di un grado di maturità nella potenza di calcolo, nei software *open source*, in Internet, nel *cloud computing*, nei dispositivi mobili e nella presenza di *big data*, che è stata raggiunta solo da qualche anno. Queste tecnologie avanzate saranno la base del marketing 5.0 nel prossimo decennio e la loro diffusione permetterà alle imprese di automatizzare mansioni noiose e ripetitive, abbattere gli ostacoli geografici, migliorare la sicurezza digitale nei settori che utilizzano dati sensibili, fino a ridurre la necessità di risorse umane in contesti ad alto rischio [40].

Ciò che spaventa oggi è l'impatto che queste tecnologie potrebbero avere sull'operato dell'essere umano ed il problema centrale di questa fase del marketing 5.0 risiede quindi nell'identificare il modo in cui macchine e persone possono integrarsi le une con le altre, al fine di creare il massimo valore per il cliente lungo il suo percorso di acquisto. L'obiettivo del marketing 5.0 è quello di sfruttare le potenzialità della tecnologia avanzata mantenendo al centro l'umanità. La *next tech* viene utilizzata per supportare i *marketers* nella creazione, nella comunicazione e nell'offerta, al fine di aumentare il valore lungo il *customer journey* del cliente, creando una *customer experience* che sia allo stesso tempo fluida e coinvolgente. Questa fase richiede quindi di realizzare una simbiosi bilanciata tra l'intelligenza umana e quella della macchina. L'estrazione di modelli di comportamento del consumatore dai dati grazie all'utilizzo dell'AI non basta. È necessario l'apporto dei *marketers* per filtrare ed interpretare le motivazioni sottese ai pattern di comportamento individuati dalla tecnologia, poiché nessuna macchina è in grado di replicare ciò che l'intelligenza umana è in grado di fare. Di conseguenza, dal momento in cui la macchina non riesce a costruire una relazione con i consumatori analoga a quella umana e che non possiamo insegnare al computer ciò che nemmeno noi umani sappiamo come imparare, anche nel marketing 5.0 il ruolo del marketer umano resta fondamentale [40].

Il compito delle imprese in questa fase è dimostrare ai propri consumatori che la corretta applicazione di queste tecnologie non minaccia l'essere umano ma può, al contrario, aumentarne la felicità. Questo perché le tecnologie permettono di risolvere i problemi attraverso l'utilizzo di approcci personalizzati e customizzati, non andando ad eliminare le relazioni sociali ma fornendo, al contrario, piattaforme per rendere le connessioni più intime tra loro e le loro *community*. L'obiettivo di offrire una *customer experience* migliore passa necessariamente per l'integrazione di interazioni *hi-tech* e *hi-touch*.

Oggi lo *human touch* rappresenta quel fattore essenziale nella nostra esistenza che, secondo gli autori di Marketing 5.0 Kotel, Kartajaya e Setiwan, è al servizio del quale le *martech* devono operare. Il valore sottostante alle soluzioni tecnologiche che i *marketers* andranno a sviluppare avrà come base un'immane etica delle relazioni e un'importanza fondamentale nel rapporto tra l'uomo ed il resto del mondo [40].

Capitolo 2

Il marketing data-driven

4. I dati

4.1. Il ruolo del dato nel marketing della digital economy

La spinta alla digitalizzazione e lo sviluppo dell'economia digitale hanno rivoluzionato il modello di business tradizionale, costringendo da qualche anno le aziende ad innovare per restare al passo con i tempi e non rischiare di perdere il proprio vantaggio competitivo. Come diretta conseguenza di questa nuova realtà, anche i reparti marketing hanno dovuto reinventarsi. Il fondamentale cambiamento che è avvenuto nel passaggio dal marketing tradizionale al marketing digitale ha coinvolto anche persone, processi e tecnologie, i fattori attraverso i quali si sviluppa il marketing in azienda, ed ha permesso di focalizzare l'attenzione su una quarta dimensione divenuta ormai indispensabile: i dati [43].

Il marketing è stato caratterizzato fin dalle sue origini da un orientamento *data-friendly*, con la centralità dell'analisi e delle attività di esplorazione del mercato, in particolare per ricavarne informazioni relative ai consumatori. L'uso massiccio dei dati, prima di natura esclusivamente quantitativa e poi qualitativa, è sempre stato rilevante anche nelle più ordinarie attività di vendita e distribuzione come sostegno al management e come verifica dei risultati a posteriori, rimpiazzando presto anche nella pianificazione aziendale i metodi empirici e la semplice esperienza. Nella prima metà del secolo scorso, con l'avvento dei supporti di archiviazione magnetici, degli strumenti di elaborazione e con lo sviluppo della disciplina statistica, i dati hanno assunto particolare importanza in azienda, soprattutto nella funzione del marketing. Si iniziano a produrre dati in proprio, ad archivarli in database ed utilizzarli sistematicamente in ogni processo di pianificazione e controllo. Inoltre, con lo sviluppo dell'informatica negli anni '70 e l'imposizione delle correnti scientifico-disciplinari legate al *data mining* ed alla più recente *digital transformation*, la cultura del dato in azienda ha iniziato a farsi sempre più consolidata [13]. Ad oggi, in un periodo in cui elaborazione culturale e strumentazione tecnica si intersecano, l'enorme mole di dati ottenibili attraverso la digitalizzazione dei business e, più in generale, da tutte le attività compiute dall'uomo, sposta l'attenzione delle aziende verso lo sviluppo dell'analisi dei dati per acquisire informazioni significative ed utilizzarle per aumentare le performance aziendali [43], portando ad un approccio di marketing che non può prescindere dalla sua dimensione *data-driven* [13].

Tutte le azioni che gli utenti compiono sul web vengono archiviate sotto forma di dati, che possono essere successivamente analizzati per estrarne tendenze e modelli. I dati raccolti grazie al digitale sono caratterizzati da elevata granularità e da un'ampia varietà di fonti. È possibile ricavare informazioni socio-demografiche e contrattuali, tracciare i percorsi di acquisto degli utenti, le loro interazioni con il brand, il comportamento su siti web e app. Incrociando questi dati con le informazioni relative all'anagrafica dei prodotti, è possibile costruire modelli descrittivi o predittivi sui futuri comportamenti di utenti e clienti. Grazie al progresso nelle scienze dell'informazione, questi dati possono oggi essere gestiti attraverso sofisticate strategie, analizzati applicando l'intelligenza artificiale e utilizzati nelle attività di automazione e di marketing o di *business intelligence* [58].

Ma i soli dati non bastano. Devono essere integrati e guidati dalla strategia competitiva e di marketing, essere calati nel contesto ambientale ed utilizzati per fare *benchmarking* con il settore di riferimento e per ricavarne informazioni strategiche che offrano una più ampia prospettiva sull'intero business. Diventano quindi sempre più richieste nel marketing competenze di tipo analitico e tecnologico, specialmente con riferimento alla gestione di piattaforme digitali e analisi dei dati, in combinazione con competenze strategiche. Ai giorni nostri, in un mondo in cui tutto accade velocemente e gli *analytics* sono diventati fondamentali nella catena del valore del marketing, diventa imprescindibile per i *marketers* saper gestire *tool* e processi per l'analisi dei dati, i quali rappresentano la nuova era delle misurazioni in ambito aziendale [23].

4.2. Vari tipi di dati

In un periodo storico in cui l'era dei dati ha raggiunto il suo apice, è necessario capire come sfruttarli e come generare conoscenza per risolvere i problemi ed ottimizzare le decisioni aziendali. I dati vengono utilizzati per prendere decisioni strategiche operative, migliorare l'esperienza offerta agli utenti, suggerire acquisti di prodotti affini a quelli già acquistati, proporre pubblicità o contenuti sui social media in linea con i gusti del cliente e per molte altre attività. Per essere utilizzati vengono raggruppati in dataset, ovvero in insiemi di dati organizzati in forma relazionale, strutturati in tabelle o matrici, in cui ciascuna colonna corrisponde ad una variabile e ciascuna riga ad un'osservazione [76]. Per costruirli, è necessario capire con che tipologie di dati ci si trova ad avere a che fare e come questi si comportano. In prima istanza è possibile suddividere i dati in due categorie principali:

- *Dati interni*: dati che si possono già trovare nei sistemi operazionali dell'azienda. Sono relativi agli acquisti, alla produzione, alle vendite e spesso sono contenuti in *Data Warehouse* e già opportunamente ripuliti.

- *Dati esterni*: dati non prodotti direttamente in azienda, provenienti da fonti esterne e utilizzabili per rendere le analisi più complete. Ne sono un esempio i dati provenienti da fonti quali ISTAT o dai social network.

I dati vengono classificati e distinti anche per il loro tipo di struttura. La presenza o meno di quest'ultima permette facilmente di identificare gli attributi che caratterizzano i dati stessi. Si possono suddividere in:

- *Dati strutturati*: sono un tipo di dati che può essere immagazzinato in database di tipo relazionale, sotto forma di tabelle costituite da righe e colonne. Vengono identificati attraverso una chiave primaria. Sono altamente organizzati e possono essere facilmente elaborati utilizzando strumenti di *data mining*.
- *Dati semistrutturati*: si caratterizzano per essere organizzati e strutturati, ma non possono essere archiviati in database relazionali. Un esempio di dati semi-strutturati possono essere i JSON e gli XML.
- *Dati non strutturati*: sono dati grezzi e non organizzati e non possono essere immagazzinati in alcun sistema di database relazionale. Rappresentano l'80% dei dati raccolti e includono video, audio, immagini, email, file di testo, commenti e post sui social media e risiedono su file di testo o su file binari.

Ultima ma non meno importante, la classificazione dei dati in quantitativi, qualitativi, *big data*.

- *Dati quantitativi*: sono quella tipologia di dati che può essere espressa sotto forma di numeri. Un classico esempio è la quantità di un determinato prodotto acquistata da un cliente.
- *Dati qualitativi*: sono dati categorici, ovvero che possono essere classificati in gruppi ed hanno la funzione di descrivere una qualità dell'osservazione.
- *Big data*: sono dati dalle tipologie più varie, raccolti in dataset molto grandi, difficilmente memorizzabili all'interno di una singola macchina. Oggi rappresentano una risorsa preziosa soprattutto per svolgere attività di marketing personalizzate e più profittevoli. Verrà approfondito il tema dei *big data* nella sezione seguente [29] [9].

4.3. Big data

4.3.1 La definizione di big data

Ancora oggi non esiste una definizione completa e totalmente condivisibile sul significato di *big data*. Il termine è stato utilizzato per la prima volta già verso la fine del secolo scorso, nel 1997, in un documento redatto dalla NASA relativo ad un problema nella visualizzazione di grandi set di dati e che rappresentava all'epoca una grande sfida per i *data scientists*. I *big data* hanno iniziato però ad acquisire popolarità attorno al 2008, quando degli informatici americani iniziarono a prevedere l'impatto che il *big data computing* avrebbe potuto avere sulle attività delle aziende, dei ricercatori e degli informatici [35]. In particolare, è a partire dal 2011 che il termine ha spopolato, anche se è spesso stato utilizzato in modo improprio, anche a causa della banalità della traduzione del termine dalla lingua inglese ("grande mole di dati"). Specialmente in questi anni, molti ricercatori hanno cercato di darne definizione. Per trarne una migliore interpretazione, possiamo definire i *big data* come "una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore" [25]. Si parla sostanzialmente di particolari tipi di "dati complessi che richiedono metodologie, strumenti e competenze atte a gestirli, processarli, estrarli ed analizzarli." [35]. Il fatto che siano un'enorme mole di dati non è una condizione sufficiente perché venga attribuita loro l'etichetta di "*big data*". Essi devono essere originati da strumenti, tecnologie e macchinari di ultima generazione, con la capacità di estrarli, organizzarli e analizzarli in modo da poterne ricavare informazioni di business anche in tempo reale sul comportamento dei consumatori, sulle loro caratteristiche o sulle performance dell'azienda [35].

Per approfondire il significato di *big data* è necessario introdurre il concetto delle V dei *big data* ed il modello ITMI.

4.3.2. Le Vs dei big data

Nonostante non esista ancora una definizione universalmente riconosciuta di *big data*, è possibile spiegarne puntualmente le caratteristiche e gli aspetti che li definiscono [35]. Il primo ad introdurre i principi base che li descrivono fu Doug Laney nel 2001, che definì il modello delle 3v dei *big data*: volume, velocità e varietà [28].

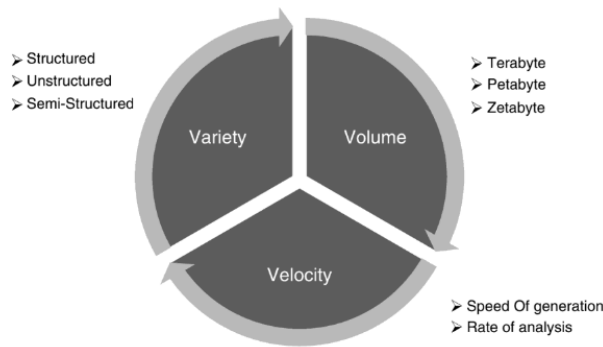


Figura 2: Le 3V dei *Big data* (Fonte: *Big data: Concepts, Technology, and Architecture*, 2021)

4.3.2.1. Il modello a 3V

Volume. I dati generati e processati dai *big data* sono in continuo aumento ad un ritmo sempre crescente. Il loro volume cresce esponenzialmente perché le aziende continuano a raccogliere dati per migliorare le loro soluzioni di business e questo richiede modalità di archiviazione sempre più scalabili e affidabili. I dati provengono principalmente dai social media, dalle transazioni effettuate tramite POS e online banking, dai sensori GPS, dai sensori nei veicoli [9]. Oggi si parla di *zetabyte* o addirittura di *brontobyte* di dati, quantità che superano di gran lunga il *terabyte* e che richiedono di essere gestiti attraverso tecnologie ad hoc [37]. Inoltre, il 90% dei dati mondiali sono stati generati negli ultimi anni e questo prospetta un continuo aumento nella loro proliferazione. Il problema risiede nell'acquisizione, conservazione e contenimento di questo elevato tipo di informazioni, sempre più varie, che devono a loro volta essere verificate, organizzate ed analizzate [28].

Velocità. Rappresenta una caratteristica essenziale dei *big data* che li distingue da semplici dataset di grandi dimensioni. Fa riferimento sia alla velocità di generazione e diffusione dei dati attraverso la rete che alla necessità di analizzare sempre più rapidamente i dati raccolti in tempo reale, per identificare trend ed opportunità prima dei concorrenti [28]. Pensiamo ai contenuti che diventano virali in un attimo, ai post sui social network che si condividono con un clic o alle più delicate transazioni attraverso carte di credito che richiedono controlli di sicurezza con spostamenti di dati che avvengono in millesimi di secondo [37].

Varietà. I *big data* si caratterizzano per essere un'ampia gamma di tipologie di dati, provenienti dalle fonti più disparate e gli strumenti di raccolta e analisi tradizionali non sono in grado di gestire in modo efficiente e contemporaneamente tutti i diversi tipi di dati [28]. Sia i dati generati dalle persone che quelli generati dalle macchine possono essere strutturati, semi strutturati o non strutturati, come

definiti precedentemente. Questa varietà di dati comporta la necessità di nuove soluzioni specifiche per raccogliere, immagazzinare ed analizzare i *big data*.

4.3.2.2. Il modello a 5V

Dal 2001, come affermato in precedenza, il concetto di *big data* si è evoluto e con esso anche il modello delle 3V di Laney. Si è passati da un modello a 3V ad un modello a 5V, che si è espanso ulteriormente fino a comprendere 7 dimensioni, poi 9 e così via. Alle 3V originali, che costituiscono la base per i *big data*, sono state aggiunte altre caratteristiche che arricchiscono il modello originale carente di aspetti come la scoperta dei dati, l'identificazione di nuovi modelli, la qualità, l'affidabilità ed il loro processo di sviluppo [41]. Da come si evince da alcuni articoli scientifici, sembra che oggi gli studi contino addirittura 42V. Nell'elaborato verranno approfondite quelle maggiormente citate dagli studi.

Il modello a 5V è quello che viene più comunemente utilizzato nella definizione dei *big data*. Le due dimensioni ulteriori rispetto a quelle descritte in precedenza sono: Veridicità e Valore.

Veridicità. Fa riferimento all'accuratezza ed all'affidabilità dei dati, alla loro significatività per risolvere i problemi di *data mining*. I dati non interessanti rappresentano rumore, anomalie nel sistema ed è necessario che vengano separati dai dati di interesse [41]. Essi devono essere utili e creare valore per poter supportare i processi decisionali, di conseguenza è necessario che siano dati di qualità e, anche se è più difficile da garantire, che siano affidabili ed accurati [37]. Questa dimensione rappresenta oggi una delle maggiori sfide nell'analisi dei dati [41].

Valore. I dati assumono un'importanza rilevante in azienda solo nel momento in cui si riesce a trasformarli in valore. Ciò significa sfruttarli per poter fare delle previsioni e prendere decisioni ottimizzate. Un classico esempio è rappresentato dalle aziende che analizzano i dati che raccolgono sui consumatori e sulle loro attività per prevederne il comportamento di acquisto e modificare i propri prodotti, servizi, campagne di marketing sulla base dei risultati di queste analisi [37]. Ricavare un significato dai dati permette di lavorarci per obiettivi differenti come il *targeting* per le pubblicità, ma anche effettuare analisi del rischio, intercettare frodi, ecc. [41].

Studi più recenti hanno definito altre quattro dimensioni: Validità, Volatilità, Visualizzazione, Variabilità.

Validità. Fa riferimento al corretto utilizzo dei dati. È una dimensione correlata alla veridicità, ma prende in considerazione anche l'integrità dei dati [57].

Volatilità. Riguarda il sistema di conservazione dei dati e le tempistiche di archiviazione che ne consentano un utilizzo futuro [57].

Visualizzazione. Questa dimensione riguarda la capacità di rendere comprensibili più dimensioni dei *big data*. È necessaria una modalità di visualizzazione che renda leggibile ed interpretabile l'enorme quantità di dati che vengono analizzati [41] [57].

Variabilità. Si riferisce ai mutamenti che può subire il significato o la connotazione di un dato nel tempo ed in base ai diversi contesti in cui viene analizzato. Il problema che ne deriva è quello di riuscire a decodificarne il senso correttamente nel momento in cui i dati devono essere interpretati [28].

4.3.2.3. Il modello ITMI

A partire dalla definizione di *big data*, Andrea De Mauro, nel 2019, elabora in una delle sue pubblicazioni un modello concettuale chiamato modello ITMI che definisce le quattro componenti essenziali del concetto di *big data*: l'informazione, la tecnologia, i metodi e l'impatto.

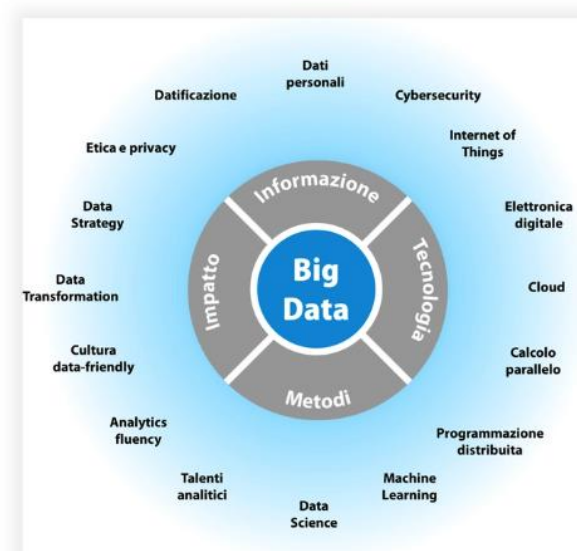


Figura 3: Il modello ITMI dei *Big data* con gli argomenti più popolari connessi a ciascuna componente (Fonte: *Big data per il Business*, 2020)

Informazione. Una delle componenti essenziali dei *big data* è l'informazione. Oggi viviamo in un'era di sovraccarico informativo in cui i dati non vengono generati solamente dalle persone (come quelli di Facebook, Whatsapp, ecc.), ma anche dalle macchine stesse, dai dispositivi dotati di sensori che generano e distribuiscono autonomamente dati attraverso la connessione internet. Si tratta dei dispositivi intelligenti della rete *Internet of Things*. Li ritroviamo negli ambiti più vari, dal controllo della produzione nelle aziende, fino alla tecnologia IoT presente all'interno delle nostre stesse case, che ci permette ad esempio di ottimizzare l'energia controllando elettrodomestici, luci e caldaie. Le macchine intelligenti creano continuamente dati, che sommati alla già elevata quantità prodotta dalle persone, permettono di raccoglierne un'enorme mole, quella a cui la definizione di *big data* fa riferimento.

Tecnologia. È il fattore scatenante e che maggiormente rappresenta il successo dei *big data* negli ultimi anni. Grazie al progresso nei campi dell'elettronica, delle telecomunicazioni e dell'informatica, si è assistito ad una crescita esponenziale delle capacità di calcolo, di archiviazione e trasferimento dei dati. De Mauro nel suo libro spiega come questa crescita tecnologica possa essere approfondita attraverso la legge di Moore, proposta dal cofondatore di Intel Gordon Moore nel 1965 e che descrive lo sviluppo della microelettronica in quegli anni con una progressione esponenziale.

La legge enuncia che «*La complessità di un microcircuito, misurata ad esempio tramite il numero di transistor per chip, raddoppia ogni 18 mesi (e quadruplica quindi ogni 3 anni).*» [79]

In effetti è possibile notare come, negli ultimi anni, siano aumentati costantemente la velocità di calcolo relativamente ai microprocessori dei computer, il volume della memoria digitale a parità di costo, la velocità della connessione internet e dello scambio dei dati tramite dispositivi mobili. Tutte queste tecnologie non hanno quindi seguito un aumento lineare ma si sono evolute ad un ritmo esponenziale. Nonostante dati e tecnologie, anche quelle intelligenti, siano conosciute da svariato tempo, non è da molto che sono diventate protagoniste delle nostre vite e che si sono fatte spazio nelle nostre aziende, ripetendo la progressione tipica delle tecnologie cosiddette esponenziali. Questo progresso ha permesso ai *big data* di esplodere e di vivere oggi uno sviluppo all'ennesima potenza. Inoltre, fondamentali sono stati anche i progressi ottenuti in alcune specifiche aree tecnologiche, come il calcolo distribuito, che hanno permesso di creare nuove infrastrutture informatiche adatte alla gestione dei *big data*. La più popolare è Hadoop, un *open source framework software*, composto da vari moduli in grado di elaborare una grande mole di dati in parallelo su macchine distribuite con l'obiettivo di velocità nel calcolo ed elevata affidabilità. I due moduli principali sono Hadoop Distributed File System, «*un file system distribuito che gestisce dataset di grandi dimensioni in*

esecuzione sull'hardware commerciale.” [36] e MapReduce, che rappresenta il cuore di Hadoop ed è “*un paradigma di programmazione che permette una scalabilità massiccia su centinaia o migliaia di server in un cluster Hadoop.*” [36]. Anche l’evoluzione della modalità di scambio risorse in *cloud* è stata fondamentale nello sviluppo dei *big data*. Questa modalità permette anche alle aziende più piccole di farne utilizzo senza esporsi ad investimenti a lungo termine per creare dei cluster Hadoop, affidandosi a dei provider che mettono a disposizione temporaneamente parte dei propri data center connessi online in cambio di un canone mensile. Tutto ciò conferma quanto sia essenziale lo sviluppo tecnologico per la crescita esponenziale dei *big data*.

Metodi. Con l’avvento dei *big data* ci si è presto accorti che le tecniche e metodologie analitiche utilizzate tradizionalmente non sarebbero state efficaci nello sfruttare al massimo le potenzialità dei nuovi strumenti tecnologici e di questi dati caratterizzati da maggiore velocità, varietà e volume. Per questo motivo si sono affermati quelli che oggi chiamiamo *data analytics*, metodi che sono stati sviluppati a partire da conoscenze derivanti da ambiti scientifici diversi, dalla matematica, all’informatica, alla statistica, nel contesto multidisciplinare della *data science* e utilizzati per interpretare ed estrarre conoscenza dai dati.

Impatto. I *big data* hanno avuto ed hanno tuttora un impatto ad ampio spettro non solo sul business ma anche sulle nostre vite perché sono entrati nella quotidianità in tutti i settori e le discipline, siano esse umanistiche o scientifiche. L’utilizzo massivo dei dati sta impattando in maniera contrastante anche a livello emozionale, dividendo coloro che vedono nei *big data* un futuro efficiente, equo, sostenibile e prospero, da coloro che sono spaventati dalle implicazioni relative alla privacy, al controllo totalitario dei dati e alla presa di potere delle macchine. Per questo motivo è fondamentale definire l’effettiva portata di questo tipo di dati, per poterne sfruttare le potenzialità senza perdere il controllo. Dotarsi di tecnologie e *big data* può portare le aziende a ridurre i costi, migliorare il processo di *decision making* e la definizione dei nuovi prodotti da lanciare sul mercato. Sfruttare i dati permette quindi di anticipare le opportunità e trovare il modo per impattarle anche grazie alla tecnologia ed alle giuste figure professionali come i *data scientists*. Resta comunque sempre vivo il problema legato alla privacy ed alle limitazioni delle libertà personali, considerando che, anche se l’utente accetta di condividere le proprie informazioni personali, potrebbe non essere consapevole del fatto che dai suoi dati si ricavano schemi di comportamento che l’azienda potrebbe utilizzare a proprio vantaggio. È necessario rimanere all’erta e sempre aggiornati sulle normative europee in tema di privacy per poter sfruttare i *big data* senza ledere i diritti personali degli individui [24].

5. Data science per il marketing

5.1. Cos'è la data science?

Negli ultimi anni, grazie alla maggiore disponibilità in rete di dati sui comportamenti degli utenti e allo sviluppo degli strumenti informatici, sono aumentate considerevolmente anche le possibilità che le campagne di *digital marketing* possono offrire alle aziende per migliorare le performance. *L'email marketing*, il *search engine marketing*, il *social media marketing* hanno potenziato le attività di acquisizione, segmentazione e profilazione dei clienti per le aziende, creando maggiori opportunità per il business. Tutti questi dati grezzi che oggi riusciamo a raccogliere sul comportamento degli utenti rappresentano il petrolio del nostro secolo e la moltiplicazione dei dati relativi alle campagne, agli acquisti ed agli stessi percorsi che compiono i potenziali clienti prima di effettuare una spesa, rappresentano un'enorme opportunità per i *marketers* di oggi. Per sfruttare al massimo questa grande quantità di dati e ricavarne informazioni interessanti, sono necessari strumenti, tecniche e modelli statistici che possiamo ritrovare all'interno di quella branca della scienza definita *data science*, che ci permette di osservare tendenze e cause di questi comportamenti su larga scala, con applicazioni dirette sulle attività di *digital marketing*. In questo contesto, nel suo libro *Data science Marketing: applicazioni data driven nel digital marketing*, Fabio Celli dà una particolare definizione di *data science*, affermando che: “*se i big data rappresentano una potenziale miniera di conoscenze, la data science è la cassetta degli attrezzi per estrarne il prezioso valore, perché permette di trattare automaticamente i dati quantificando e visualizzando relazioni, individuando gruppi simili tra loro, riconoscendo regolarità, estraendo modelli predittivi di andamenti futuri e confrontando gli attributi tra loro.*” [16].

Dando una definizione più scientifica, è possibile affermare che la *data science* è una materia multidisciplinare che combina diversi campi tra cui l'informatica, la matematica, la statistica e l'analisi dei dati e che necessita anche di una buona conoscenza del dominio di applicazione dei dati per poter contestualizzare e dare un senso alle informazioni estratte dagli stessi [16]. È una scienza che si occupa, attraverso l'utilizzo di tecniche statistiche e algoritmi di *machine learning*, di analizzare i dati ed estrapolarne *insights* e previsioni e di trasformarli per renderli praticamente utilizzabili attraverso un insieme di regole e processi. La figura che si occupa di queste attività è il *data scientist*, un ruolo relativamente nuovo in ambito aziendale ma anche molto ricercato e che richiede competenze trasversali per analizzare dati raccolti da varie fonti, tra cui il web, le piattaforme mobile, o gli stessi clienti. Partendo dalle richieste del business, si occupa di produrre analisi, previsioni, regressioni, grafici per ricavare conoscenza dai dati a disposizione, cercando di prevedere

correlazioni non visibili dall'uomo in prima istanza. Permette di ottenere informazioni che il business possa utilizzare per risolvere problemi specifici [29] [53]. Le attività necessarie per implementare un progetto di *data science* comprendono anche la preparazione dei dati, la loro pulizia, l'aggregazione e la manipolazione, per poter poi eseguire le analisi avanzate, estrarre modelli da dati complessi e utilizzarli nei problemi di vita reale [53] [14].

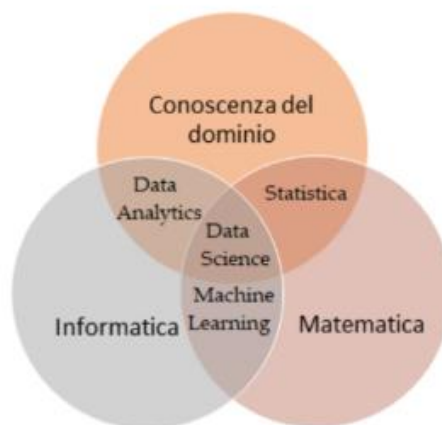


Figura 4: Le skills del data scientist (Fonte: *Data science e Machine learning: Dai Dati alla Conoscenza*, 2021)

Molto spesso quando si parla di *data science* si fa riferimento in maniera confusionaria anche a termini come *machine learning* ed intelligenza artificiale. In realtà, sono tre discipline differenti, strettamente collegate tra loro e spesso utilizzate contemporaneamente. La *data science*, infatti, sfrutta algoritmi di *machine learning* per ricavare “modelli dai dati”, mentre l’intelligenza artificiale utilizza l’apprendimento automatico per far “imparare” la macchina dai dati ed ottimizzare le azioni da intraprendere. Il territorio della *data science* si estende dal *data mining*, che si avvale degli algoritmi per processare i dati, alla business intelligence, che invece si appoggia prevalentemente sugli strumenti per la visualizzazione dei dati [29].

La contaminazione dell’ambito delle performance aziendali con discipline di tipo scientifico, consente oggi di gestire in maniera più efficiente ed efficace le informazioni rispetto a quanto avveniva attraverso l’utilizzo di tradizionali metodi di analisi dei dati, specialmente nel campo del marketing. Nella seguente parte del capitolo verranno analizzate alcune delle discipline della *data science* maggiormente utilizzate nel business e come si combinano tra loro per permettere alle aziende di acquisire vantaggio competitivo.

5.2. I data analytics

Il termine *data analytics* fa riferimento a quell'insieme di metodi, tecniche e strumenti utilizzati per estrarre informazioni di valore dai dati. Rappresenta sostanzialmente il modo per trasformare i dati grezzi in qualcosa di più utile e prezioso. Quando si parla di *data analytics* si fa riferimento ad un concetto molto ampio. Al suo interno è possibile riconoscere quattro tipi di analisi dei dati, caratterizzate da proprie peculiarità e applicazioni pratiche. Possiamo distinguere la *data analytics* in: *descriptive analytics*, *diagnostic analytics*, *predictive analytics* e *prescriptive analytics* [26].

5.2.1. Descriptive analytics

Le *descriptive analytics* rappresentano la base dell'analisi dei dati all'interno dell'azienda. Sono metodologie di analisi che si focalizzano sull'esaminare i dati storici per descrivere ciò che è avvenuto in passato e riportare un'immagine chiara ed interpretabile sullo stato del business. Rispondono in pratica alla domanda "Cos'è successo?". Rappresentano la forma più tradizionale di capacità analitiche, con un output che prende tendenzialmente la forma di un report statico distribuito con regolarità ai vertici aziendali. Il documento raccoglie e spiega i KPI di interesse attraverso tabelle, grafici, statistiche riassuntive che mostrano semplici trasformazioni o aggregazioni di valori. Oltre ai report statici, attraverso tecniche di analitica descrittiva è possibile creare *dashboard* interattive, che vengono condivise via web e mostrano grafici e tabelle che possono essere personalizzati attraverso filtri, selettori e pulsanti e permettono all'utente di approfondire gli aspetti di maggiore interesse. Anche i cosiddetti *management cockpit* sono uno strumento di *descriptive analytics*. Sono una sorta di cruscotti decisionali indirizzati al top management che evidenziano i KPI necessari per facilitare il *decision making* [26]. Fanno parte di questo gruppo i servizi di *analytics* forniti da Google o dalle piattaforme Meta, che mostrano dati relativi a traffico, obiettivi, costi e ricavi delle campagne permettendo di ricavare metriche come il ROI, *clickthrough rate*, costo per clic e così via [16].

5.2.2. Diagnostic analytics

Le tecniche di analisi diagnostica vengono utilizzate per studiare le relazioni di causa delle tendenze e le correlazioni tra le variabili [20]. È un tipo di analisi che si pone tra il livello descrittivo e predittivo ed ha come obiettivo quello di ricercare in maniera sistematica i motivi che causano dinamiche rilevanti nell'evoluzione del business [25]. Le *diagnostic analytics* comprendono tecniche statistiche come i metodi di confronto (ad esempio il calcolo della t-stat), i test di verifica delle ipotesi (chi

quadrato) o i metodi per analizzare gli attributi come la correlazione e le tecniche di trasformazione dei dati (*Principal Component Analysis* o selezione degli attributi) [16].

5.2.3. Predictive analytics

I metodi di *predictive analytics* sono tecniche analitiche più sofisticate, spesso basate sull'intelligenza artificiale per andare oltre la semplice descrizione dei fatti avvenuti nel passato, utilizzando i dati per prevedere quello che potrebbe accadere in futuro. Infatti, questo tipo di analisi consente rispondere alle domande “Perché è successo?” e “Che cosa succederà?” [26]. Si basa sull'utilizzo di dati storici per prevedere potenziali scenari che possono aiutare a guidare le decisioni strategiche. Le previsioni possono riguardare un futuro a breve termine, come ad esempio nel caso in cui si voglia prevedere il mal funzionamento di un macchinario nell'arco di una giornata, o un futuro più lontano, come nella previsione dei flussi di cassa della propria azienda per l'anno seguente [21]. Fare previsioni permette di acquisire la capacità di anticipare il comportamento di clienti e concorrenti e questo può produrre un importante vantaggio competitivo e migliorare il ritorno sugli investimenti in futuro. Esistono vari strumenti di *predictive analytics*, più o meno sofisticati:

- *Strumenti diagnostici*: accrescono il valore dei report statici attraverso indicatori che spiegano le cause di quello che viene osservato nei dati evidenziandone schemi rilevanti. Utilizzano metodi semplici come l'analisi di correlazione, le carte di controllo ed i test statistici.
- *Business alert*: permettono di inviare delle notifiche quando i controlli diagnostici che vengono effettuati automaticamente rilevano situazioni notevoli per il business che necessitano di essere approfondite.
- *Anomaly detection*: sono strumenti in grado di ricercare incongruenze e fenomeni anomali nei dati che possono essere gestiti se opportunamente segnalati.
- *Propensity model*: fanno parte di quei modelli che cercano di anticipare il futuro facendo delle previsioni. In questo caso vengono utilizzati algoritmi di intelligenza artificiale per prevedere il livello di soddisfazione di un cliente rispetto ad un'offerta commerciale o il tasso di *churn*, ovvero di abbandono, della clientela.
- *Strumenti per la segmentazione*: si utilizzano algoritmi per creare dei gruppi omogenei di clienti, negozi o prodotti che mostrano similarità tra loro. Trattare con gruppi al posto che con i singoli permette di rendere i processi più snelli e migliorare l'esperienza del cliente [26].

Questo livello di analitiche comprende, più in generale, tutti gli algoritmi di *machine learning*, da quelli per risolvere problemi di classificazione, alle reti neurali, fino agli alberi decisionali e agli

algoritmi per la regressione. Gli algoritmi di *forecasting* rappresentano l'espressione più avanzata delle analitiche predittive, che fanno previsioni sul futuro basandosi su dati del passato attraverso regressioni su serie temporali [16].

5.2.4. Prescriptive analytics

Le *prescriptive analytics* utilizzano i dati per fornire raccomandazioni e determinare una linea d'azione ottimale nelle decisioni da prendere [22]. Rappresentano un gruppo di analitiche che indicano direttamente quali azioni effettuare per il raggiungimento di un particolare obiettivo, rispondendo alla domanda che spesso si pone il management su "Che cosa fare?" [26]. Attraverso gli algoritmi di *machine learning*, analizzano in maniera veloce ed efficiente una grande quantità di dati e, utilizzando le condizioni "if"-*"else"*, formulano raccomandazioni basate su specifiche combinazioni di requisiti [22]. Sono strumenti che simulano un gran numero di scenari alternativi e li filtrano di volta in volta per restituire la soluzione che massimizza i profitti o minimizza i costi. Nel marketing permettono ad esempio di definire livelli di sconto personalizzati per spingere gli utenti del nostro e-commerce ad acquistare. Il massimo degli strumenti prescrittivi è rappresentato dalla macchina che è in grado di lavorare in modo autonomo senza il continuo intervento dell'uomo [25].

Un esempio di *prescriptive analytics* sono i *recommendation systems*, sempre più presenti nella nostra esperienza digitale di tutti i giorni, attraverso i quali gli algoritmi raccomandano all'utente dei prodotti a seconda delle preferenze, degli acquisti precedenti e del comportamento sul sito web. Nel *programmatic advertising*, le analitiche prescrittive raggiungono un livello ancora più sofisticato. Gli algoritmi sono progettati per applicare un apprendimento automatico continuo attraverso un processo iterativo e prendere così decisioni anche in *real time*. Cercano di massimizzare i rendimenti cumulativi prodotti dalle decisioni attraverso un continuo processo di test della loro strategia nel mondo reale ed apportando le opportune correzioni volta per volta [26]. L'*email automation* è un chiaro esempio di analisi prescrittiva all'opera. I *marketers* la utilizzano per inviare contenuti personalizzati agli utenti, attraverso la definizione di alcune categorie per cui differenziare i *lead* in base a motivazioni, mentalità e intenzioni. Qualsiasi interazione che gli utenti hanno con le email possono automaticamente inserirli all'interno delle categorie, attivando l'invio di un diverso insieme di messaggi. È da sottolineare che anche le analitiche prescrittive sono uno strumento per prendere decisioni e creare strategie, di conseguenza il giudizio umano alle raccomandazioni fornite dagli algoritmi è sempre prezioso e necessario per raggiungere gli obiettivi prefissati [22].

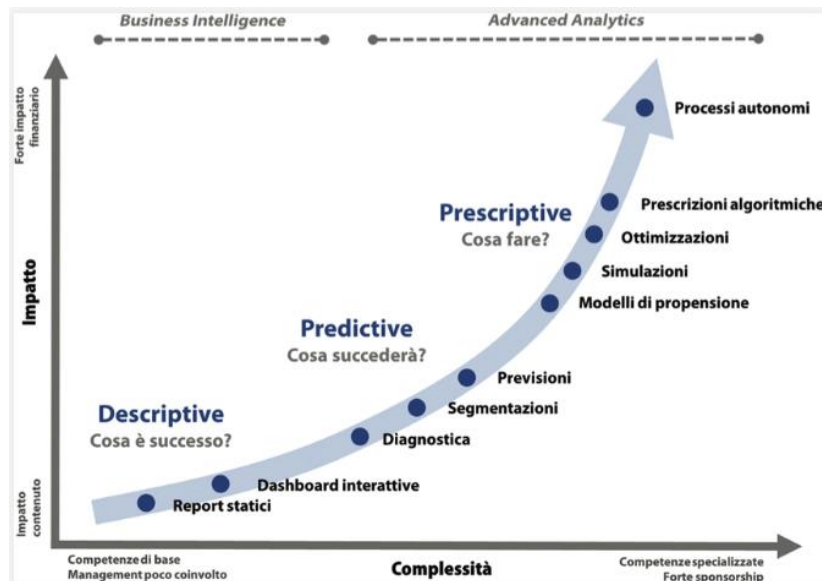


Figura 5: I tre livelli di data analytics (Fonte: *Big data per il Business*, 2020)

La figura 5 mostra come, passando da un livello di *data analytics* all'altro, si generi un impatto economico sul business sempre più rilevante, a discapito però di un aumento della complessità degli strumenti utilizzati e, di conseguenza, di un bisogno crescente di risorse sempre più specializzate e di un coinvolgimento del management sempre più forte. Più ci si vuole spingere nella creazione di valore attraverso i dati, più è naturale incontrare ostacoli come la resistenza culturale dell'azienda ed i limiti tecnologici e di *know-how*. È comunque possibile affermare che non esiste una netta demarcazione fra i diversi livelli di *analytics* ed essi non vanno visti come alternative l'uno all'altro. È molto frequente infatti veder coesistere diversi strumenti analitici all'interno della stessa organizzazione o nello sviluppo di uno specifico progetto. Infatti, la *business intelligence*, che riguarda la parte più descrittiva dell'analisi, sarà sempre utile per il controllo gestionale e servirà da base per spingere le analisi predittive e prescrittive a cui ci si riferisce oggi con il termine di *advanced analytics*. L'utilizzo degli strumenti di *data analytics* è essenziale all'interno delle aziende perché consente di supportare l'azienda nella riduzione dei costi, nell'aumento alle vendite e, ad un livello più alto, porta ad un miglioramento della presa di decisioni e della definizione di prodotti e servizi da lanciare sul mercato [25].

5.3. Data mining e Machine learning

La convergenza tra informatica e comunicazione ha portato allo sviluppo di una società che si nutre di informazioni, che raccogliamo ancora per la maggior parte nella loro forma grezza di dato. I database delle aziende racchiudono un'enorme quantità di informazioni potenzialmente importanti per il business ma ancora non scoperte o non abbastanza articolate [34]. Con lo sviluppo dei *big data* e dell'*information technology*, le potenzialità assunte dalla crescente quantità di dati diventati disponibili in azienda sono in crescita esponenziale. Le continue ricerche in ambito database hanno dato origine ad un approccio per archiviare, manipolare ed estrarre informazioni da questi dati preziosi per raggiungere gli obiettivi di business, specialmente in termini di crescita organizzativa e *customer satisfaction* [2] [10]. In particolare, sono due gli ambiti della *data science* che si sono sviluppati con l'avvento dei *big data*: il *data mining* ed il *machine learning*.

5.3.1. Data mining

5.3.1.1. Definizione e caratteristiche

Il *data mining* è un processo di estrazione di informazioni implicite, precedentemente sconosciute, e potenzialmente utili da set di dati [34]. Utilizza una serie di tecniche per analizzare i database e ricercare degli schemi, pattern ricorrenti e correlazioni tra i dati definendo modelli computazionali che delineano gli sviluppi dei processi in corso e ne predicono il futuro andamento [19]. È un ambito della *data science* nato negli anni '90 ed è caratterizzato dal suo approccio multidisciplinare che comprende campi come la statistica, l'intelligenza artificiale, il *machine learning* e la gestione dei database. Il *data mining* permette di sviluppare modelli in grado di riconoscere delle correlazioni tra milioni di record presenti nei database aziendali e per questo costituisce un elemento fondamentale dei *data analytics* [62].

Un processo di *data mining* si articola nella raccolta e analisi dei dati, nella creazione di modelli matematici e nell'utilizzo della conoscenza acquisita per prendere decisioni. A partire da un campione di osservazioni passate, queste tecniche hanno l'obiettivo di trarre delle conclusioni che sia possibile generalizzare all'intera popolazione, così da predirne il comportamento nel modo più accurato possibile. Il processo prevede una fase di *training*, in cui vengono utilizzati i dati storici per far apprendere il modello, ed una fase successiva, in cui il modello generato viene utilizzato per costruire un legame funzionale e fare delle previsioni [31]. Da questa operazione è possibile scovare schemi e modelli forti, che possono essere generalizzati e utilizzati per fare previsioni accurate sui dati futuri.

Nonostante ciò, capita spesso di individuare pattern banali, poco interessanti o spuri, ovvero dovuti a coincidenze accidentali e non a vere e proprie correlazioni tra i dati. I dati reali non sono perfetti, anzi, e questo può portare a trovare modelli che siano in parte confusi o con dati mancati. Sarà quindi sempre possibile trovare eccezioni a ogni regola prodotta o casi a sé stanti non compresi in nessuna regola. Ciò che fa il *data mining* è applicare algoritmi che devono essere sufficientemente robusti per gestire questi tipi di dati imperfetti ed estrarre pattern e regolarità che siano generalizzabili e utili per chi le utilizza [34]. Grazie a questo tipo di processo, i dati vengono trasformati in conoscenza che viene poi sfruttata per prendere decisioni migliori. I modelli e le regolarità che vengono scovate tra i dati possono essere espresse sotto forma di espressioni lineari, sequenze di regole *if-then-else*, *cluster*, grafici o alberi decisionali [31].

L'utilizzo delle tecniche di *data mining* permette *in primis* alle aziende di assumere un atteggiamento proattivo dando la possibilità di intervenire in anticipo sui comportamenti dei clienti. Consente, inoltre, di reagire in maniera tempestiva ai cambiamenti nelle condizioni competitive e di ottimizzare l'impatto delle campagne e attività di marketing [31]. La possibilità di analizzare dati di diverse tipologie (quantitativi, qualitativi, testuali, immagini, suoni, ecc.), il fatto che non richieda necessariamente la formulazione di ipotesi a priori, la capacità di elaborare un numero di osservazioni e variabili elevato, la semplicità nell'interpretazione dei risultati e la loro visualizzazione, sono tutti vantaggi che hanno portato allo sviluppo di questa disciplina e gli hanno permesso di farsi strada all'interno dei reparti marketing e produzione delle aziende [30]. La crescita irrefrenabile dei database negli ultimi anni ha portato il *data mining* al centro delle nuove tecnologie aziendali [34]. Questo perché la quantità di dati da elaborare è sempre maggiore ed il mercato in rapida evoluzione richiede necessariamente di anticipare i concorrenti per guadagnare vantaggio competitivo. Per raggiungere questo tipo di obiettivi, gli strumenti tradizionali non sono più adeguati al processo di *knowledge discovery* e ci si affida alle più accurate tecniche di *data mining* [30], che diventano fondamentali per scoprire schemi nascosti nei dati, rendendoli una risorsa inestimabile per le aziende e fonte di crescita per il business [34]. Inoltre, per stare al passo con il potenziale illimitato dei *big data* e lo sviluppo della potenza di calcolo e della velocità di elaborazione, anche la tecnologia di *data mining* continua costantemente ad evolversi, permettendoci di effettuare analisi dei dati sempre più rapide, semplici e soprattutto, automatizzate. Le aziende di vari settori si servono spesso di questi metodi per ottimizzare i prezzi e le promozioni o per scoprire come diversi fattori (es. dati demografici, rischio, concorrenza, social media) possono influenzare i loro modelli di business, i profitti, le attività o le relazioni con i clienti [62].

5.3.1.2. Sviluppo di un progetto di data mining

Con lo sviluppo delle tecniche di *data mining*, sia accademici che industriali hanno cercato fin dagli anni '90 di definire degli standard per quest'area di conoscenza. I primi puntavano a creare un linguaggio per il *data mining* che diventasse un canone (come è avvenuto per il linguaggio SQL relativamente ai database relazionali), i secondi a strutturare processi e metodologie che potessero guidare l'implementazione delle applicazioni di *data mining*. A partire dal tradizionale processo di *Knowledge Discovery for Databases* utilizzato nell'ampio campo di ricerca della conoscenza dai dati, sono stati definiti negli anni una serie di metodi e processi per lo sviluppo di un processo standard di *data mining*. Il *CRISP-DM*, che verrà approfondito nei prossimi paragrafi, è uno tra questi [8].

5.3.1.2.1. Knowledge Discovery in Databases

Nonostante vengano spesso utilizzati come sinonimi, *il data mining* si colloca all'interno di un più ampio processo di ricerca di nuova conoscenza dai dati, definito *Knowledge Discovery in Databases* (KDD), e ne costituisce lo stadio fondamentale, pur non coincidendo completamente con il processo stesso. Definito nel 1996 da Usama Fayyad, Piatetsky-Shapiro e Smyth, questa metodologia coinvolge l'applicazione degli algoritmi di *data mining* in maniera ripetuta per interpretarne i pattern generati e utilizzare la conoscenza a supporto delle decisioni aziendali.

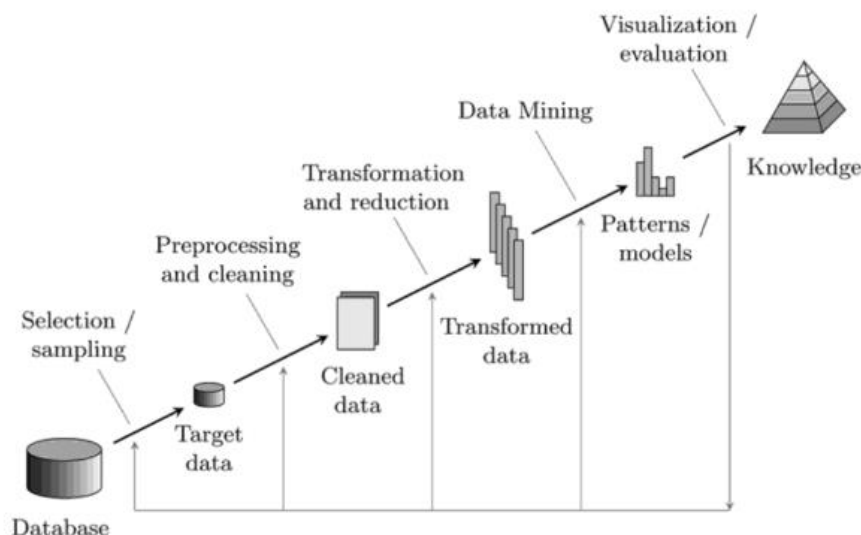


Figura 6: Processo di KDD (Fonte: Introduzione al KDD e al *Data mining*)

La figura mostra le diverse fasi del processo di KDD in cui il *data mining* rappresenta la fase più importante.

1. *Selection and sampling*: in questa prima fase viene effettuata una segmentazione dei dati grezzi che vengono poi selezionati secondo alcuni criteri per definire un set di dati *target* utili a risolvere il problema sotto studio, eliminando i dati inutili.
2. *Preprocessing and data cleaning*: per la maggior parte delle analisi, anche dopo aver individuato un *target data* non si ritiene necessario analizzarne l'intero contenuto. Nella fase di pre-elaborazione vengono campionate le tabelle ed esplorate in modo da effettuare un'analisi su base campionaria. Si procede poi alla pulizia dei dati che prevede l'eliminazione di errori e la decisione di come comportarsi in caso di dati mancanti.
3. *Transformation*: i dati per essere utilizzabili devono essere trasformati. Questa fase consiste nel convertire gli stessi in altri tipi di dati o utilizzare formule matematiche e logiche sulle variabili per definirne di nuovi. Durante questa fase è importante anche effettuare la riconfigurazione dei dati, al fine di garantirne la consistenza, soprattutto se provengono da fonti diverse.
4. *Data mining*: è la fase centrale del processo di KDD che comprende le fasi di scelta del ruolo dei sistemi di *data mining* per l'analisi, la selezione degli algoritmi e la loro effettiva applicazione, al fine di ricercare modelli interessanti agli scopi del progetto [30]. In questo stadio ai dati trasformati vengono applicate delle tecniche di analisi per ricavarne delle informazioni interessanti, utili e non banali. Gli algoritmi vengono scelti a seconda dei dati a disposizione e degli obiettivi stabiliti per il progetto stesso. Considerando che l'intero processo di *Knowledge Discovery* è interattivo tra utente, software e obiettivi e allo stesso tempo iterativo, questa fase può prevedere un ritorno per più volte anche alle fasi precedenti per trasformare o ripulire ulteriormente i dati a seconda delle esigenze degli algoritmi di analisi.
5. *Visualization and evaluation*: dopo aver scovato pattern e modelli dai dati attraverso l'utilizzo degli algoritmi di *data mining*, è necessario interpretare e visualizzare l'output dell'analisi, valutare i modelli e capire in che modo possono essere utili e di supporto alle decisioni. Se i risultati non sono soddisfacenti, è possibile rivedere e/o modificare alcune fasi dell'intero processo [44].

La metodologia di *Knowledge Discovery in Databases* comprende numerosi passaggi e molte decisioni da prendere in cui è coinvolto anche il *data scientist*. Alle fasi sopra descritte è necessario aggiungere uno stadio precedente relativo allo sviluppo di una comprensione del dominio dell'applicazione, delle conoscenze pregresse pertinenti e degli obiettivi dell'utente finale ed uno stadio successivo che prevede il consolidamento della conoscenza, incorporandola nel sistema aziendale [8].

5.3.1.2.2. CRISP-DM

Un modello sviluppato in seguito per definire un approccio standard strettamente relativo ai progetti di *data mining* è il *Cross Industry Standard Process for Data mining* (CRISP-DM). È stato delineato da un consorzio di aziende inizialmente composto da Daimler Chrysler, SPSS e NCR e finanziato dalla Commissione Europea [8]. Lo scopo di questo progetto era quello di definire e convalidare un sistema di approccio al *data mining* indipendente dalla tipologia di business in cui avrebbe potuto essere applicato.

La figura seguente mostra il ciclo di vita di un progetto di *data mining* che si sviluppa in sei fasi con sequenza non rigida. Questo perché molto spesso il processo richiede di ritornare alle fasi precedenti prima di procedere con le successive a seconda della bontà dei risultati prodotti in ogni fase. Il processo è raffigurato con delle frecce che rappresentano le principali e più frequenti dipendenze tra le fasi e circondato da un'ellisse che pone l'accento sulla ciclicità del processo di *data mining*, che continua anche dopo l'individuazione e la sperimentazione dei risultati ottenuti, utilizzando le nuove conoscenze come informazioni per processi consecutivi.

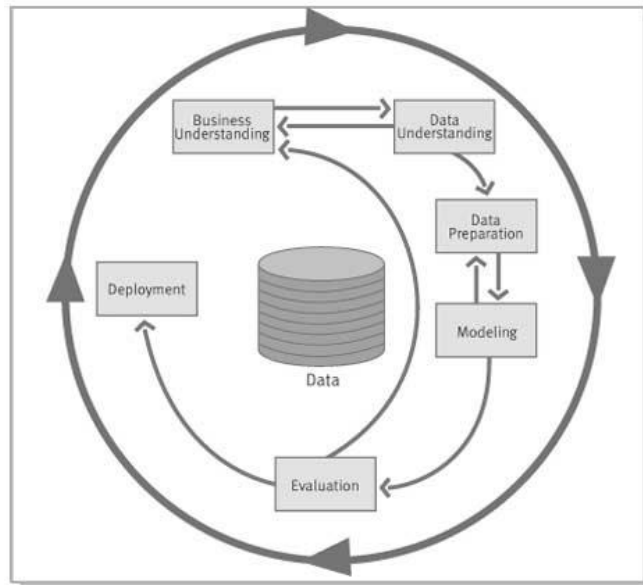


Figura 7: Il ciclo di vita del processo CRISP-DM (Fonte: Researchgate.net)

Vengono di seguito approfondite le fasi dell'approccio CRISP-DM.

1. *Business Understanding*: prima di implementare un qualsiasi progetto di *data mining* è necessario conoscere il settore in cui si opera. Una prima fase consiste quindi nella definizione degli obiettivi da parte dei manager aziendali che devono essere tradotti in un problema di *data mining*. Viene effettuata la stesura preliminare di un piano per raggiungere gli obiettivi.
2. *Data Understanding*: questa fase prevede la raccolta dei dati necessari all'analisi ed una loro prima elaborazione in modo da acquisirne familiarità, identificare problemi nella qualità dei dati raccolti e scovare le prime informazioni attraverso il calcolo delle statistiche di base.
3. *Data Preparation*: comprende tutte le attività necessarie a costruire l'insieme di dati finali a cui applicare le tecniche di *data mining* a partire dai dati grezzi. In questa fase vengono selezionate tabelle, *records* e attributi, e vengono trasformati e puliti i dati.
4. *Modelling*: rappresenta lo stadio di scelta delle tecniche di *data mining* da applicare, in modo da ricavare modelli dai dati. Considerato che alcune tecniche per essere applicate necessitano di accogliere in input i dati sotto una particolare forma, questa fase, come mostrato dallo schema, spesso richiede di tornare allo stadio di *data preparation* per modificare il dataset e adattarlo alle specifiche esigenze.

5. *Evaluation*: prima di procedere alla fase di distribuzione del modello è necessario valutarlo in modo approfondito e rivedere i passaggi effettuati per costruirlo in modo da assicurarsi che sia adatto per il raggiungimento degli obiettivi aziendali. Questo stadio termina con la decisione di come utilizzare i risultati del progetto di *data mining*.
6. *Deployment*: è l'ultima fase del processo e consiste nell'utilizzo del modello che è stato costruito e valutato per raggiungere gli obiettivi stabiliti. Per questo motivo la conoscenza acquisita dovrà essere organizzata e presentata in modo da poter essere utilizzata dal cliente [44] [45].

5.3.2. Tecniche di data mining

Il *data mining* è utile alle organizzazioni perché le assiste nello sviluppare un'efficiente analisi dei *big data* e permette di risolvere un'ampia serie di problemi, a seconda delle tecniche che vengono applicate e degli algoritmi scelti nella fase di *modelling*. I *data scientists* utilizzano molte tecniche differenti che permettono di utilizzare questo approccio all'analisi dei dati in vari settori e campi di applicazione. Ogni campione utilizzato e ogni tecnica possono portare a risultati differenti, considerando che utilizzano proprie regole e metodi di analisi. Per questo motivo è necessario dare molta importanza all'identificazione del problema di business, per scegliere ed eseguire la giusta tecnica e ricavarne il miglior risultato [2].

I principali metodi di *data mining* utilizzati nel campo aziendale e del marketing sono: associazione, classificazione, alberi decisionali, *clustering*, predizione e reti neurali.

Associazione. Le regole di associazione sono una delle tecniche più conosciute e utilizzate. Permettono di scoprire pattern e correlazioni nascoste tra i dati basandosi sulla relazione tra alcune variabili memorizzate generalmente a partire dagli acquisti effettuati dai clienti. Utilizzano le relazioni tra gli oggetti appartenenti ad una stessa transazione per scovare ricorrenze frequenti di diversi *items* che si ripetono all'interno del dataset. Si basano sulla regola *if-then*, attraverso la quale viene definita la probabilità delle relazioni scoperte tra gli oggetti o le variabili della transazione [2]. Molto spesso il numero di possibili regole di associazione per un dataset risulta elevato e una grande proporzione di esse sono spesso di basso valore [10]. Si tratta di determinare quali di queste sono regole forti che possono essere generalizzate e utilizzate. Questo viene fatto sulla base di due indicatori, il supporto e la confidenza, che permettono di identificare una soglia minima per cui una

regola associativa può essere considerata forte. Queste tecniche sono molto utili nel marketing perché permettono di spiegare le correlazioni esistenti tra alcuni tipi di prodotti acquistati condizionatamente l'uno all'altro, valutano a priori che effetto potrebbero avere delle attività promozionali di *cross-selling* su insiemi di prodotti correlati o cosa comporterebbe l'introduzione di nuovi prodotti rispetto agli esistenti correlati negativamente. Inoltre, permettono di ottimizzare le attività di merchandising attraverso una miglior collocazione dei prodotti sugli scaffali e facilitano le operazioni di riassortimento nei punti vendita [31]. È una tecnica molto utilizzata dai retailers attraverso la *Market Basket Analysis* che permette di capire il comportamento d'acquisto del cliente analizzando la composizione del suo carrello della spesa [2]. Le tipologie di regole associative maggiormente conosciute sono: *multilevel association rule*, *multidimensional association rule*, *quantitative association rule* [10].

Classificazione. Le tecniche di classificazione hanno l'obiettivo di sviluppare un modello in grado di classificare una collezione di dati in differenti gruppi o classi e fare previsioni accurate su grandi dataset [2]. In un problema di classificazione si prende in considerazione un dataset contenente n osservazioni costituite da m attributi esplicativi e 1 attributo target categorico che identifica la classe di appartenenza di ciascun record in relazione agli obiettivi dell'analisi di *data mining*. “*I modelli di classificazione hanno l'obiettivo di individuare dei legami ricorrenti tra le variabili esplicative corrispondenti ad osservazioni appartenenti ad una medesima classe. Questi legami vengono tradotti come regole di classificazione che vengono impiegate per predire la classi di osservazioni di cui è noto solo il valore degli attributi esplicativi*” [31]. Questo tipo di tecnica di *data mining* può essere utilizzato per avere un'idea di quanti consumatori o osservazioni appartengono ad una determinata categoria piuttosto che ad un'altra, utilizzando attributi multipli per identificare una classe in particolare [2]. Per questo motivo trova spesso applicazione negli ambiti di rilevamento delle frodi e per valutare il rischio di credito o la probabilità di disattivazione di un servizio da parte di un cliente [10]. La classificazione può essere utilizzata anche in combinazione ad altre tecniche per migliorarne il risultato, come negli alberi decisionali o nelle operazioni di *clustering* [2].

Un processo di classificazione dei dati prevede una fase di *training*, una fase di *test* e una fase di predizione. Nella prima fase viene identificato un *training set* che viene utilizzato per “allenare” l'algoritmo di classificazione. Nella fase di *test*, le regole vengono impiegate sul restante set di dati, utilizzato in tal modo per stimare l'accuratezza delle regole di classificazione. Se l'accuratezza è accettabile, si passa alla fase di predizione in cui le regole prodotte possono essere applicate a nuovi set di dati. Questo processo consente all'algoritmo di determinare l'insieme di parametri richiesti per

effettuare una corretta discriminazione, permettendo così di codificarli in un modello chiamato classificatore. I modelli di classificazione si distinguono in: *classificazione per induzione dell'albero decisionale*, *classificazione bayesiana*, *reti neurali*, *support vector machine*, *classificazione basata sull'associazione* [10].

Alberi decisionali. Rappresentano una delle tecniche di classificazione più utilizzate di tipo gerarchico. Vengono utilizzati specialmente per la loro semplicità concettuale e d'uso, di interpretazione dei risultati, per la loro velocità nell'elaborazione e per la robustezza rispetto agli *outlier* o a dati mancanti. Gli alberi decisionali si basano su un processo che ripartisce in gruppi la popolazione attraverso spaccature di tipo gerarchico. Per farlo vengono selezionate tra le variabili esplicative quelle che spiegano al meglio la variabilità della variabile target di cui si vuole prevedere il valore. I tre elementi principali che costituiscono gli alberi decisionali sono:

- *Nodi interni:* rappresentano i raggruppamenti degli elementi in diversi stadi della segmentazione.
- *Rami:* sono le condizioni che determinano la divisione.
- *Nodi foglia:* sono i nodi da cui non parte più alcun ramo, per cui non sono ritenute utili altre suddivisioni.

Gli alberi decisionali possono essere prodotti attraverso molti algoritmi. L'algoritmo AID è il più conosciuto ed accetta solo variabili target qualitative effettuando delle partizioni binarie, che massimizzano la devianza tra i gruppi. A partire da esso si sono sviluppati altri due algoritmi molto conosciuti: l'algoritmo CHAID, che effettua segmentazioni multiple e l'algoritmo CART, che effettua segmentazioni binarie ma accetta sia variabili di tipo qualitativo che quantitativo e adotta la tecnica di *pruning* per determinare degli alberi di taglia ottimale [31].

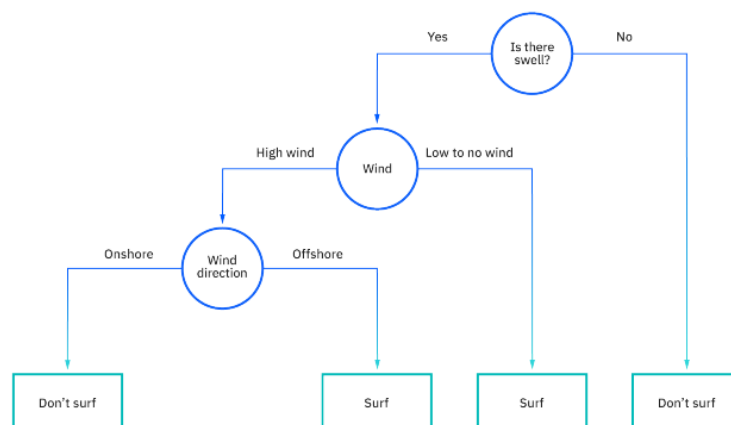


Figura 8: Funzionamento di un albero decisionale (Fonte: IBM)

Clustering. Le tecniche di *clustering* consistono in un processo di raggruppamento di un insieme di elementi in classi che comprendano gli *items* più simili tra loro. La *cluster analysis* viene molto spesso utilizzata nelle attività di *data mining* perché permette di classificare in maniera automatica una raccolta di modelli in gruppi, chiamati *cluster*, basandosi sulla somiglianza. Il problema di *clustering*, infatti, consiste nel suddividere un set di dati in insiemi più piccoli, così che gli *items* raggruppati all'interno di un *cluster* siano più simili tra loro rispetto a quelli di altri gruppi, sulla base di determinati criteri [56]. Analizzando uno o più attributi, questo processo permette di identificare le istanze simili tra loro così da comprendere differenze e somiglianze tra i dati. Spesso questo processo di raggruppamento viene chiamato anche segmentazione perché suddivide i dati in categorie con l'obiettivo di identificare un *cluster* di risultati correlati [2]. In base alla strategia che viene utilizzata per distinguere gli elementi del dataset, le tecniche di *clustering* si suddividono in tecniche di tipo gerarchico e di tipo non gerarchico. Oltre a queste due modalità più conosciute, esistono anche metodi basati sulla densità, sulla griglia o metodi basati su modelli.

Gli algoritmi gerarchici effettuano delle scomposizioni gerarchiche di un dato insieme di oggetti. In base al processo di *clustering* possono essere classificati in processi agglomerativi o divisivi. Nell'approccio agglomerativo, chiamato anche *bottom up*, si presenta una situazione iniziale in cui ciascun oggetto costituisce un gruppo separato e l'algoritmo lavora per aggregare successivamente oggetti e gruppi vicini. Il processo prosegue fino al livello più in alto della gerarchia in cui il cluster comprende tutti gli oggetti in analisi o fino al momento in cui non si verifica una condizione per cui si ritiene necessario terminare l'aggregazione. Nell'approccio divisivo, chiamato anche *top-down*, si verifica la situazione esattamente contraria, per cui inizialmente gli oggetti appartengono tutti ad un unico cluster. In seguito ad ogni iterazione successiva, il cluster viene diviso fino a quando ogni oggetto costituisce un cluster differente o se si verifica una condizione di terminazione [46].

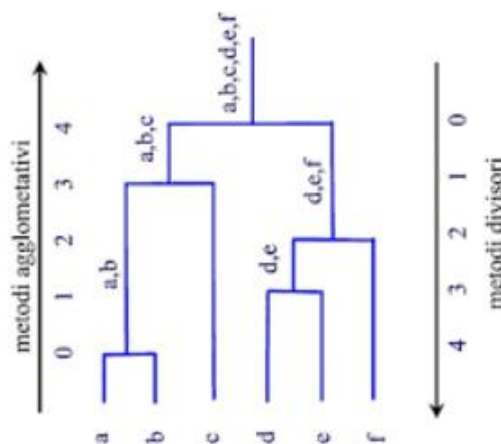


Figura 9: Metodo di *clustering* gerarchico rappresentato attraverso un dendrogramma (Fonte: developersmaggioli.it)

I metodi di partizionamento non gerarchici, invece, assegnano gli elementi a k cluster in modo che ciascun gruppo contenga almeno un oggetto e ciascun oggetto appartenga esattamente ad un gruppo e li riallocano attraverso un processo iterativo per migliorare la qualità del *clustering*. Questo permette agli oggetti appartenenti allo stesso cluster di essere “vicini” tra loro e correlati ed agli oggetti appartenenti a cluster differenti di essere molto distanti tra loro [46]. L’algoritmo più utilizzato per questi tipi di analisi è il *k-means*.

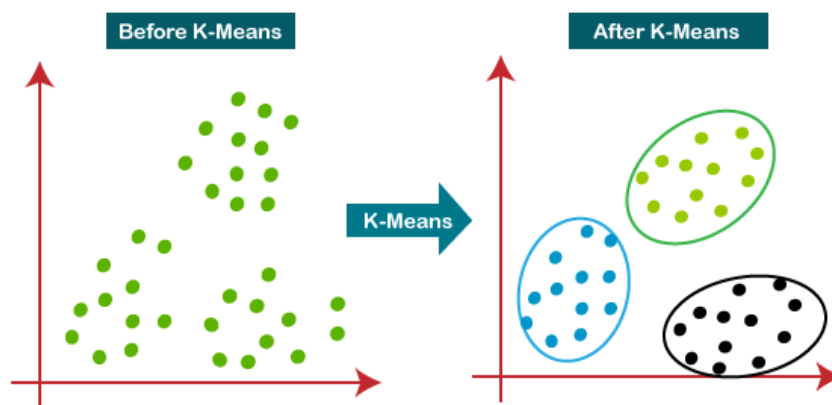


Figura 10: Modello di *clustering* non gerarchico, attraverso l’utilizzo dell’algoritmo *k-means*. (Fonte: medium.com)

Il *clustering* fa parte dei processi di classificazione non supervisionati. Molto spesso, quando l’approccio di classificazione diventa costoso, viene utilizzata questa tecnica come approccio di pre-elaborazione per la selezione e la classificazione di sottoinsiemi di attributi, ad esempio, per formare gruppi di clienti in base a modelli di acquisto [10].

Predizione. L’argomento relativo alle tecniche di predizione è molto ampio. Sia la classificazione che il *pattern matching*, ad esempio, possono essere utilizzate per questo tipo di scopo. Una tecnica che viene adottata molto spesso e che può essere adattata per la previsione è la regressione. Questo tipo di analisi può essere utilizzato per studiare la relazione tra una o più variabili indipendenti, che rappresentano gli attributi già noti, e variabili dipendenti o di risposta, che sono quelle che si vuole prevedere. Opera in modo simile ai classificatori, ma in questo caso i risultati non sono variabili categoriche ma valori numerici. Infatti, molto spesso, gli stessi tipi di modelli possono essere utilizzati sia per la regressione che per la classificazione, come l’algoritmo CART (albero decisionale) o le reti neurali. Nel marketing questo tipo di analisi può essere utilizzata per studiare il ROAS e

l'associazione tra le spese di pubblicità e le vendite. La regressione può essere lineare, lineare multivariata, non lineare o multivariata non lineare [10].

Reti neurali. “Le reti neurali sono sistemi di calcolo con nodi interconnessi, che funzionano in modo molto simile ai neuroni del cervello umano.” [63]. Utilizzano gli algoritmi per ricercare correlazioni e modelli nei dati non strutturati, li raggruppano e li classificano, in un continuo processo di apprendimento nel tempo. [63] Sono costituite da due elementi principali: il nodo e il collegamento. Il nodo rappresenta i neuroni e il collegamento le connessioni tra essi. Sono una forte tecnica di modellazione predittiva, ma spesso molto complessa da capire anche dagli esperti [2]. Si caratterizzano per la loro grande capacità di estrarre significati dai dati più complicati e imprecisi, non gestibili da una semplice analisi effettuata dall'uomo o da altre tecniche informatiche [10]. Spesso nel business viene utilizzata per rilevare le frodi [2].

5.3.3. Applicazioni di data mining per l'e-commerce

Lo sviluppo dell'e-commerce negli ultimi anni sta portando le aziende ad affidarsi sempre di più alle potenzialità delle tecniche di *data mining*. Gli *shop* online rappresentano infatti un campo di applicazione naturale per gli *advanced analytics*. Grazie al commercio elettronico le aziende oggi più che mai sono in possesso di *big data*. Ogni utente quando visita uno *shop* online lascia dietro di sé dei fatti, compie azioni che le aziende possono tracciare e archiviare in database. Tutte queste tracce rappresentano dei dati strutturati o non strutturati che possono essere utilizzati dalle aziende per acquisire vantaggio competitivo. L'unico modo per sfruttarli al massimo è estrarre le informazioni da questa grande mole di dati per definire un miglior processo decisionale ed effettuare un'efficace *data analysis*. In particolare i database relazionali e di marketing sono un ottimo campo di applicazione per i modelli di *data mining*, che permettono attraverso l'estrazione di risultati da un mercato campione, di studiarne le caratteristiche e fornire una base per identificare le potenzialità del mercato generale. Possono fornire un supporto tecnico e informativo più accurato nella definizione di modelli di business, strategie di marketing e decisioni d'impresa e rappresentano uno strumento fondamentale per supportare il business con informazioni relative all'attività di commercio elettronico [1].

Sono molte le aree nel campo dell'e-commerce in cui il *data mining* può dare un contributo fondamentale per migliorare il business. Di seguito vengono approfonditi gli ambiti di applicazione più frequenti.

Profilazione del cliente. L'obiettivo finale di qualsiasi analisi di *data mining* in un e-commerce è quello di migliorare i processi che contribuiscono a fornire valore ai clienti, in modo da acquisirne di nuovi e fidelizzare gli esistenti. Nella parte di *back-end* dei siti web operano dei sistemi tecnologici che comprendono sofisticati strumenti di *data mining* in grado di definire i profili dei clienti e scenari di modellazione predittiva delle interazioni con la piattaforma di e-commerce [60]. L'estrazione della conoscenza dai dati di navigazione degli utenti permette di identificarne il loro comportamento, se acquistano articoli che già conoscevano, se navigano e acquistano prodotti che non conoscevano o prodotti nuovi, e tutto questo aiuta le aziende a migliorare la propria infrastruttura e ad ottimizzare i processi di pianificazione. La possibilità di raccogliere informazioni di profilazione del cliente permette alle aziende di pianificare le proprie attività e operazioni commerciali e sviluppare continue ricerche per migliorare prodotti e servizi offerti con l'e-commerce. Inoltre, classificare i clienti potenzialmente più redditizi a partire dai dati delle visite al sito web può aiutare le aziende a ridurre i costi di vendita e di acquisizione della clientela [1]. Sistemi basati sul *data mining* vengono utilizzati per proporre prodotti alternativi ai clienti o per tenerli informati su eventi importanti per cui hanno dimostrato interesse [60].

Personalizzazione del servizio. Le tecnologie di *data mining* applicate all'e-commerce permettono di fornire contenuti e servizi orientati ai singoli utenti sulla base delle informazioni raccolte riguardo i loro bisogni e comportamenti. Ciò su cui si è maggiormente concentrata la ricerca della comunità del *data mining* sono i sistemi di raccomandazione, sempre più frequenti, e su alcuni strumenti correlati come il filtraggio collaborativo [1]. Attraverso i *recommendation systems*, i consumatori ottengono suggerimenti personalizzati su prodotti e servizi di loro interesse e, dal canto loro, le aziende sviluppano una migliore relazione con il cliente. Esistono diversi tipi di approcci su cui si basano i sistemi di raccomandazione. I modelli possono essere guidati dai contenuti, offrendo suggerimenti sulla base di ciò che il cliente ha osservato o acquistato in passato, o dalla community, ovvero sulla base delle azioni compiute da utenti con caratteristiche simili. Inoltre si dividono in sistemi che utilizzano classificazioni o *tag* assegnati ai clienti, che ricevono feedback diretto, o che si basano sugli articoli acquistati insieme nel tempo, che ricevono un feedback indiretto. Molto spesso chi gestisce gli e-commerce utilizza anche i meccanismi di suggerimento in *real time* o pubblicità targetizzate che sfruttano i contenuti che l'utente sta navigando [77].

Analisi del paniere. La *Market Basket Analysis* è una tecnica molto utilizzata dai retailers, soprattutto online, per conoscere meglio i propri clienti attraverso i dati raccolti dall'e-commerce. Con l'analisi del paniere è possibile identificare le affinità tra i vari articoli acquistati così da pianificare campagne

di *cross-sell* o *up-sell*, combinazioni e *bundle* di prodotti e raccogliere informazioni per costruire il profilo del cliente con i dati demografici, le abitudini di acquisto, le preferenze, e per offrire una migliore esperienza di navigazione e acquisto online.

Previsioni di vendita. Con le applicazioni di *data mining* per l'e-commerce è anche possibile, attraverso lo studio del tempo che il cliente impiega per acquistare un articolo, prevedere se e quando comprerà di nuovo. Questo può servire alle imprese per determinare strategie di obsolescenza pianificata o per individuare prodotti complementari da pubblicizzare sul mercato. Prevedere il flusso di cassa derivante dalle vendite dell'e-commerce permette alle aziende di non trovarsi impreparate e di creare un piano finanziario per supportare il caso in cui la previsione delle vendite non venga rispettata.

Pianificazione della merce. Le tecniche di *data mining* aiutano in ambito di pianificazione della merce nella determinazione delle opzioni di stoccaggio e nella gestione di magazzino e inventario. Permettono di determinare il miglior prezzo per prodotti e servizi, di capire quali articoli maggiormente desiderano i clienti e di bilanciare le scorte, utilizzando come supporto anche strumenti di *business intelligence* sui dati della concorrenza.

Segmentazione del mercato. Rappresenta uno dei migliori utilizzi delle tecniche di *data mining* per l'e-commerce. I *big data* forniti dalla rete permettono di suddividere gli utenti sulla base di caratteristiche significative come: reddito, età, sesso, occupazione e molte altre informazioni che possono essere utilizzate per sviluppare delle campagne personalizzate di *email marketing* o nelle strategie SEO. Inoltre, segmentare il mercato può essere utile nell'analisi della concorrenza per capire quali concorrenti sono coinvolti in ogni singolo segmento e definire strategie personalizzate sul target di riferimento. La segmentazione del database di una società di vendita al dettaglio permette così di migliorare i tassi di conversione poiché l'azienda può concentrare i propri sforzi su un mercato aderente e altamente ricercato [1].

5.4. Machine learning

Il *machine learning*, o apprendimento automatico, può essere definito come “la capacità di un sistema intelligente di imparare a svolgere una determinata funzione direttamente sul campo, senza la necessità di una programmazione a monte e in un regime di miglioramento continuo” [42].

Attraverso l'utilizzo di metodi statistici, logiche induttive ed altre tecniche scientifico-informatiche, l'apprendimento automatico ha l'obiettivo di affinare in maniera progressiva l'attendibilità di un algoritmo nell'individuare pattern e modelli nei dati [42]. Deriva dal campo dell'intelligenza artificiale e rappresenta un ramo molto ampio di essa. Si occupa di riprodurre l'abilità di imparare, che caratterizza l'intelligenza umana, attraverso le macchine [25] e la utilizza per elaborare regole di comportamento, che mette poi in pratica per risolvere problemi relativi all'analisi dei dati [64].

La tecnologia di *machine learning* rappresenta una sorta di variante al tradizionale approccio di programmazione informatica, in cui il programmatore codifica, ovvero descrive in linguaggio di programmazione, i passi che il computer deve eseguire per passare dai dati di input al risultato, dovendo necessariamente conoscere il legame che li unisce. Questo approccio viene sorpassato con il *machine learning*, nel quale è la macchina che deve capire che tipo di legame unisce dati e risultati [25].

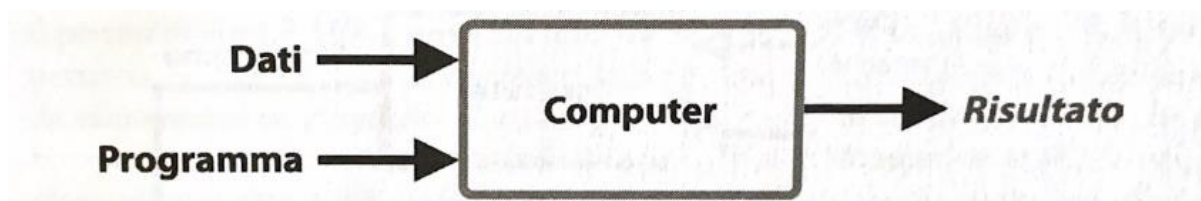


Figura 11: Approccio tradizionale alla programmazione informatica (Fonte: *Big data per il Business*, 2020)

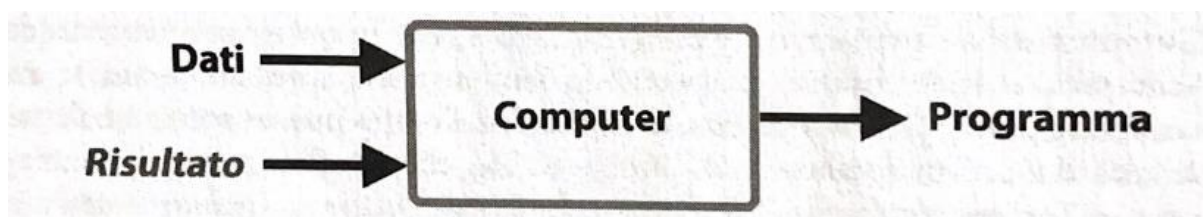


Figura 12: Approccio di *machine learning* (Fonte: *Big data per il Business*, 2020)

Come mostrano le figure 11 e 12, gli elementi che compongono i due metodi sono esattamente gli stessi: i dati, un programma, il computer ed i risultati. Quello che cambia è l'ordine in cui appaiono nel processo. Il programma utilizzato per passare dai dati al risultato non rappresenta più un input in ingresso ma diventa l'informazione in uscita al computer. Questo ci permette di capire come

l'algoritmo di *machine learning* cerchi sostanzialmente di imparare verosimilmente come passare dai dati al risultato e restituisca il processo come output finale. Uno dei vantaggi principali del *machine learning* è quello di poter riutilizzare il programma creato dal computer, per applicarlo su altri casi e scoprire informazioni da nuovi set di dati, sfruttando ciò che la macchina ha imparato ed accrescere al tempo stesso il valore del programma [25].

Il *machine learning* è uno tra i temi più studiati e discussi del momento. Le ricerche e le sperimentazioni iniziarono già dalla metà del secolo scorso grazie ad alcuni matematici che si cimentarono nell'applicazione di metodi probabilistici per cercare di automatizzare il processo di *decision making* attraverso le macchine. Ciò ha permesso di arrivare oggi all'utilizzo dell'apprendimento automatico in un ampio raggio di applicazioni, dal riconoscimento vocale, alla guida autonoma delle auto, fino ai suggerimenti di ricerca su Amazon, Netflix o Google, apportando particolare valore soprattutto alle analisi di marketing [42].

Oggi le aziende, per poter estrarre informazioni dai *big data* devono necessariamente servirsi dell'intelligenza artificiale. Il *machine learning* diventa quindi fondamentale nel campo dell'analisi dei dati perché le applicazioni di *data analytics* più interessanti per il business si basano proprio su questo tipo di tecnologia [25]. L'apprendimento automatico costituisce la base tecnica del *data mining*, che utilizza molti dei suoi algoritmi per estrarre informazioni dai dati grezzi presenti nei database [34]. Le tecniche di *data mining* possono essere infatti utilizzate come una sorta di fonte di informazioni per le applicazioni di *machine learning*, perché permettono di definire insiemi di dati più organizzati come input per gli algoritmi di apprendimento automatico [61]. Questa tecnologia diventa quindi fondamentale nell'automatizzare il processo di ricerca schemi e correlazioni tra i dati [18].

Nel campo del *digital marketing*, in particolare, come per quanto visto nelle tecniche di *data mining*, il *machine learning* permette di sfruttare le informazioni relative ai contatti e provenienti da una varietà di fonti (CRM, social media, email, sito web, ecc.), per costruire un *customer journey* personalizzato a livello di singolo utente. L'apprendimento automatico rende possibile automatizzare le attività di segmentazione e profilazione dei clienti, basandosi su svariati fattori e variabili, per poi applicare i giusti strumenti di marketing e sviluppare le attività di engagement per ogni segmento. Inoltre, permette di ottimizzare tempi e risultati delle campagne promozionali mirate, di distribuire contenuti che siano rilevanti per l'utente offrendogli un'esperienza di valore e coerente in tutti i canali. Permette anche di intervenire anticipatamente in caso di segnali di abbandono da parte dei clienti e di gestire di conseguenza il *churn rate*. Secondo uno studio effettuato da *Forbes* nel 2018, tra le

aziende che hanno introdotto il *machine learning* e tecnologie di intelligenza artificiale, il 75% ha accresciuto la *customer satisfaction* del 10%.

Le applicazioni più utilizzate di *machine learning* nel *digital marketing* possiamo riconoscerle nelle attività di *email marketing* e *web personalization*, specialmente con i *recommendation systems*, nel *social media marketing* e attraverso la *marketing automation*, che verrà approfondita nel prossimo capitolo [42].

5.4.1. Tipologie di apprendimento automatico

Gli algoritmi di apprendimento automatico non elaborano i dati tutti nello stesso modo. Esistono differenti modalità di *machine learning* che si caratterizzano per insegnare modelli di comportamento diversi alla macchina che porta di conseguenza a risultati differenti a seconda della modalità utilizzata. Ad oggi si riconoscono tre modi di apprendere da parte della macchina: apprendimento supervisionato (*supervised learning*), apprendimento non supervisionato (*unsupervised learning*) e apprendimento con rinforzo (*reinforcement learning*) [64].

Apprendimento supervisionato. L'obiettivo di questa tipologia di apprendimento è quello di definire un modello, partendo dai dati etichettati di un *training set*, che permette di effettuare previsioni su dati nuovi. L'aggettivo "supervisionato" fa riferimento al fatto che questo tipo di apprendimento si basa su una serie di esempi precedentemente "addestrati" come dati di input, dove i segnali di output (etichette) sono già noti alla macchina [59]. Infatti, in questo tipo di apprendimento automatico, la macchina impara attraverso l'osservazione di esempi passati, in cui viene associato un risultato ai dati [25].

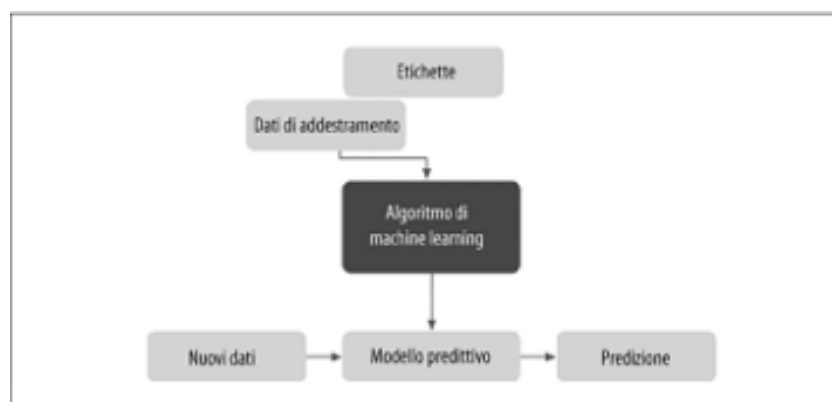


Figura 13: Processo di supervised learning (Fonte: apogeonline.com)

Nel processo in figura vediamo come i dati di addestramento provenienti dal training set vengano utilizzati come input per l’algoritmo per definire un modello in grado di fare predizioni su dati nuovi, non etichettati. Appartengono a questa categoria di apprendimento i modelli di classificazione e di regressione. Il filtraggio dei messaggi email in spam e non spam rappresenta un classico esempio di applicazione di *supervised learning*, per cui l’azione di filtraggio avviene grazie ad una precedente etichettatura di prototipi di messaggi in base a delle classificazioni che vengono indicate alla macchina e a cui resta solo il compito di individuare le stesse caratteristiche nei nuovi messaggi in entrata [64] [59].

Apprendimento con rinforzo. Nel *reinforcement learning* l’obiettivo è quello di creare un sistema, chiamato agente, che in base alle proprie interazioni con l’ambiente migliori le proprie prestazioni. [42]. Questo permette alla macchina di imparare dai suoi stessi errori. Essa mette quindi in pratica ciò che ha imparato, vede se funziona attraverso un sistema di ricompense e, attraverso questa nuova esperienza, rinforza la sua conoscenza [25]. I segnali di ricompensa, chiamati *reward*, non sono né etichette, né valori di verità, ma misurano sostanzialmente la qualità dell’azione effettuata dalla macchina. In tal modo il sistema utilizza questa forma di apprendimento automatico per imparare un insieme di azioni che possano massimizzare la ricompensa attesa. È un approccio del tipo “prova-sbaglia-correggi, impara e riprova”.

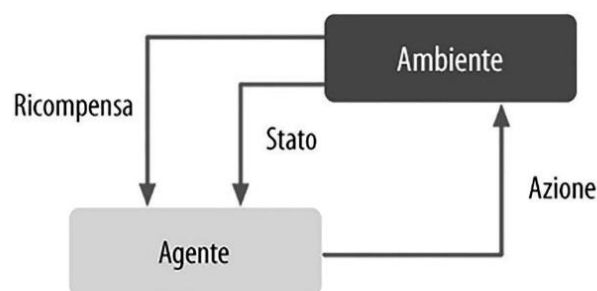


Figura 14: Processo di reinforcement learning (Fonte: apogeeonline.com)

Apprendimento non supervisionato. La macchina in questo tipo di apprendimento deve imparare a riconoscere gli aspetti della struttura dei dati di input senza avere come base esempi e risultati passati [25]. L’apprendimento non supervisionato prevede di conseguenza la costruzione di modelli per l’estrazione di informazioni dai dati senza etichette o variabili note, né meccanismi di rinforzo. Questo permette di individuare strutture nascoste nei dati e le loro conseguenti caratteristiche. L’obiettivo è quello di generare una rappresentazione propria dei dati in input producendo

un'organizzazione in grado di catturare schemi e probabilità [64]. Alcune tra le tipologie di analisi di approccio non supervisionato più utilizzate sono la *cluster analysis* e la riduzione della dimensionalità. Il *clustering*, come visto precedentemente, è infatti una tecnica esplorativa dei dati che permette di organizzarli in sottogruppi significativi, senza alcun tipo di conoscenza sull'appartenenza a precedenti gruppi. Viene alle volte chiamato anche classificazione senza supervisione [59].

Capitolo 3

Il caso Tognana Porcellane

Da quanto esposto nei capitoli precedenti possiamo dire che, principalmente in seguito alla pandemia, il consumatore sta cercando oggi di essere più attento ed informato nelle sue spese sia online che offline e sta effettuando acquisti sempre più razionali e soppesati. I nuovi trend di comportamento stanno portando le aziende ad una digitalizzazione forzata ed alla necessità di definire strategie di vendita personalizzate e incentrate sulla gestione della relazione con il cliente che diventa sempre più *one to one*. Con l'aumento dei dati a disposizione e lo sviluppo degli strumenti di analisi, inoltre, il business non può far altro che basare le proprie strategie di marketing sulle informazioni ricavate dai dati, per raggiungere gli obiettivi sia di breve che di lungo periodo. In questa parte della trattazione, l'obiettivo è quello di mettere in pratica quanto precedentemente approfondito riguardo l'importanza dei dati e l'efficacia dell'utilizzo di strumenti di *data science* nella gestione delle relazioni con i clienti. Verrà presentato in questo capitolo il caso aziendale Tognana Porcellane ed introdotta la *marketing automation*, che combina analisi dei dati ad intelligenza artificiale per spingere le azioni di CRM ad un alto livello. Si fa riferimento ad una serie di strategie e strumenti sempre più frequentemente utilizzati da agenzie di performance e reparti marketing, che permettono non solo di ricavare i dati degli utenti sul web, ma anche di gestire la relazione ed effettuare attività di marketing mirate. L'elaborato mira di conseguenza a dimostrare in maniera pratica l'importanza dei dati nella strategia di marketing e come l'analisi permetta di conoscere al meglio il proprio cliente permettendo di sviluppare un rapporto redditizio.

Il capitolo introduce una panoramica sul settore di riferimento in cui opera Tognana Porcellane ed una presentazione dell'azienda, per poi descrivere le attività di *marketing automation* che vengono svolte internamente attraverso l'utilizzo del software *all-in-one* SALESmanago. Una seconda parte del capitolo sarà dedicata al commento di una serie di analisi effettuate sui dati raccolti sul comportamento dei clienti sull'e-commerce e all'utilizzo del software KNIME per l'implementazione di un'analisi sul database delle transazioni attraverso la tecnica del *clustering*.

6. Il settore dei prodotti per la tavola, la cucina ed il regalo

6.1. La situazione attuale

Secondo i dati raccolti dall'associazione ART¹ (Arti della Tavola e del Regalo) attraverso l'indagine di mercato effettuata in collaborazione con *Format Research*, il settore degli articoli per la tavola, la cucina e il regalo in Italia vale oggi quasi 7 miliardi di euro. La filiera è composta da circa 10 mila imprese: 609 produttori, 1.097 distributori intermedi e 8.448 rivenditori al dettaglio per un totale di circa 38mila lavoratori operanti nel settore. Nel 2020 si è assistito ad una perdita di circa 900 milioni di euro, dovuta alla crisi sanitaria ed economica ed alla conseguente riduzione del potere di acquisto degli italiani che è imputabile per un 45% alla chiusura forzata di ristoranti, strutture alberghiere e, più in generale, alle difficoltà che hanno colpito l'intero settore Ho.Re.Ca.

Nonostante la crisi, il settore è riuscito a contenere in parte le perdite già nel secondo semestre del 2020 con miglioramenti significativi a partire dai primi mesi del 2021. Questo scenario di ripresa è stato possibile grazie al cambiamento nelle abitudini dei consumatori dovuti al periodo trascorso in *lockdown* che ha registrato un ritrovato interesse verso la casa in tutte le sue sfaccettature, anche in termini di maggiore attenzione verso l'oggettistica, gli articoli per la tavola e per la cucina. In seguito alla perdita di ricavi avvenuta nel 2020, già dai primi mesi del 2021 si è assistito ad un aumento sia del clima di fiducia e di ottimismo da parte delle imprese che dei principali indicatori economici rispetto alle performance registrate nell'anno precedente, trend che si è consolidato, portando ad un trimestre di chiusura dell'anno 2021 decisamente positivo.

Nonostante la forte eterogeneità che caratterizza il mercato, sono state rilevate dinamiche positive in tutte le categorie di spesa anche se la maggior parte della crescita nel fatturato è attribuibile al segmento "preparazione del cibo" che nel periodo gennaio-settembre 2021 ha avuto una crescita dell'11,5%. Il veloce recupero del livello di consumo pre-covid è stato sostenuto principalmente dai nuovi bisogni dei consumatori emersi durante la pandemia, nati come conseguenza all'aumento del tempo trascorso in casa anche grazie all'adozione sempre più diffusa della modalità di lavoro in *smart working*, permettendo un allineamento ad una crescita media annua dei consumi interni del + 4,2%.

Ruolo centrale in questa ripresa è stato rivestito dalle vendite online di prodotti per la casa. Nel 2021 il giro di affari generato dall'e-commerce è stato stimato attorno ai 39,4 miliardi di euro, con una crescita del 21% rispetto al 2020 [7]. Nel 2022, invece, il mercato si presenta in fase di rallentamento, con le nuove crisi aperte in ambito politico, sociale ed economico che hanno interrotto il trend positivo

¹ Associazione imprenditoriale, aderente al Sistema Confcommercio, che riunisce produttori, importatori, distributori, agenti di case estere e rivenditori specializzati di articoli per la tavola, la cucina, il regalo e la decorazione della casa, in collaborazione con *Format Research*.

ed impedito il consolidamento delle performance della ripresa post pandemica. Circa il 70% delle imprese, infatti, afferma che il mercato non è in miglioramento rispetto al 2021. I mercati mondiali sono attualmente agitati, l'emergenza sanitaria non ha ancora trovato una fine e la guerra in Ucraina ha decisamente influito sulle tensioni nei prezzi di trasporto, dell'energia e delle materie prime. Tutto questo risulta determinante nella perdita del potere di acquisto dei consumatori, che vedono anche in questo settore aumentare l'inflazione e diminuire le proprie disponibilità economiche per l'acquisto di prodotti non di prima necessità [5].

6.1.1. Le imprese della filiera

Di seguito verrà approfondito nel dettaglio come hanno performato nel periodo di crisi i diversi segmenti di imprese appartenenti alla filiera del settore di prodotti per la tavola, la cucina e il regalo: le imprese manifatturiere, le imprese dell'ingrosso e le imprese al dettaglio.

Lo studio è stato realizzato dall'associazione di categoria ART, attraverso la combinazione di diversi tipi di ricerche di mercato tra cui un'indagine quantitativa su un campione delle imprese che costituiscono la filiera del commercio nel settore ed un'indagine quantitativa effettuata su un campione di consumatori italiani. Dalle ricerche è emerso che il segmento delle imprese manifatturiere rappresenta il gruppo più in salute, con un livello di fiducia prospettico da parte del campione superiore a 50 (in una scala da 0 a 100), che identifica un mercato non in contrazione. Il cambiamento più significativo che hanno affrontato le imprese manifatturiere del settore è stata la trasformazione dei propri modelli di business, per adattarsi ad un nuovo orientamento spinto dalla crisi sanitaria verso il servizio al cliente ed a nuove modalità di gestione dello sviluppo commerciale. Secondo l'indagine di mercato, le scelte di acquisto dei consumatori sono ricadute per il 40% su imprese che offrono migliori prezzi, per il 34,5% verso aziende che rispettano requisiti e specifiche tecniche stringenti e, per il 23,1%, verso le aziende manifatturiere che hanno saputo identificare soluzioni tecnologiche adeguate nella gestione della relazione. Il consumatore, di conseguenza, sta attribuendo sempre più rilevanza non solo al prezzo ma anche alla maggiore personalizzazione dei prodotti offerti dalle imprese (68,6%) ed alla co-creazione con l'impresa degli stessi (65,7%). Il mercato sembra essere sempre più guidato dalla richiesta di prodotti innovativi, dalla digitalizzazione della relazione tra cliente e fornitore e dall'attenzione verso la sostenibilità.

Le imprese grossiste sono il segmento che è stato più colpito dalla pandemia, a causa dell'aumento dei costi e delle difficoltà di trasporto e logistica ed agli impedimenti burocratici, registrando la chiusura di molte imprese rispetto al 2019. Le importazioni hanno costituito un grosso ostacolo per questa tipologia di imprese. Tra le intervistate, il 49% importa dall'estero e circa il 55% ha risentito del periodo pandemico in termini di maggiori difficoltà nelle importazioni dovute all'aumento dei

costi, al trasporto ed alla logistica, alla difficoltà legata ai quantitativi ordinati ed agli ostacoli burocratici.

Nonostante la tendenza alla riscoperta della casa da parte degli italiani durante il *lockdown*, le imprese al dettaglio hanno subito un duro colpo durante il periodo pandemico principalmente a causa della generale diminuzione dei consumi (perdita di 130 miliardi di consumi in Italia registrata solo nel 2020) che ha interessato anche il settore degli articoli per la casa. Con le riaperture nel 2021, però, si è assistito ad un aumento del livello di fiducia anche da parte di questo segmento sia per quanto riguarda il settore nel suo complesso, sia per le performance di vendita delle singole imprese. Le principali novità presentate dalle imprese del commercio al dettaglio per far fronte agli ostacoli dell'emergenza sanitaria sono relative all'introduzione di innovazioni di processo e di servizio come il potenziamento o lo sviluppo della propria piattaforma di e-commerce, il miglioramento dei servizi a supporto del cliente sul web e sui social, la spinta verso una migliore gestione della relazione commerciale grazie all'invio di email e comunicazioni. Anche la sempre più strategica presenza delle aziende sui social network ed il potenziamento dei sistemi di pagamento digitali sono stati elementi fondamentali per la ripresa. L'introduzione di queste innovazioni ha permesso quasi ad un 35% delle imprese del settore di sopravvivere alla crisi, ad un 42% di minimizzare le perdite sui ricavi rispetto al 2019, al 13% di mantenere lo stesso livello di ricavi ed a quasi un 10% di aumentarli rispetto al 2019.

6.1.2. I consumatori

Nel 2020, l'87,1% dei consumatori finali intervistati ha acquistato almeno un articolo da tavola, da cucina o da regalo, registrando una diminuzione del 2,4% rispetto al 2018. Questi dati confermano che la pandemia non ha inciso significativamente sul numero di consumatori che ha fatto almeno un acquisto nel periodo ma la contrazione principale si è verificata nel valore dei consumi stessi, a causa della diminuzione del potere di acquisto dovuta dalla crisi. Durante la pandemia, infatti, l'interesse dei consumatori verso gli articoli della casa è aumentato, ma è diminuita la quantità nella spesa totale, specialmente nel 2020, con un miglioramento a partire dal 2021. La contrazione della spesa media nel periodo dell'emergenza sanitaria non può essere attribuibile ad un'unica ragione. La minor spesa è dipesa infatti dall'aumento della propensione al risparmio da parte dei consumatori in un momento di elevata incertezza economica (41,3%), dalla chiusura forzata dei negozi fisici che ha impedito l'acquisto ad alcuni consumatori (40,6%), dalla paura dei contagi (33,6%) e dalla riduzione stessa della propria capacità di spesa (27,8%).

Anche il comportamento del consumatore in fase di acquisto ha subito dei mutamenti nell'ultimo anno. In particolare, si ricerca sempre di più un'esperienza multi-canale, sfruttando sia lo *shop* fisico

che quello online. Secondo le ricerche dell'associazione ART, il 45% dei consumatori che hanno acquistato nel 2020 articoli per la casa, la tavola o la cucina in negozi fisici, hanno prima cercato informazioni online sui prodotti di interesse, principalmente per quanto riguarda prezzi e caratteristiche del prodotto. Al contrario, il 29% ha preferito cercare le informazioni sui prodotti prima in negozio. Il 51% di coloro che si sono recati in un negozio fisico con intenzione di acquistare prodotti di cui avevano bisogno ha poi preferito effettuare l'acquisto online per la convenienza nel prezzo di vendita, mentre quasi il 37% ha deciso di acquistare online perché in negozio non ha trovato il prodotto che cercava [6].

La crisi portata dallo scoppio della pandemia da covid-19 ha avuto un impatto considerevole sia sull'economia che sul comportamento stesso dei consumatori. L'utilizzo "obbligato" di canali diversi dal classico negozio fisico ha portato alla definizione di nuove abitudini per il consumatore di tutte le età ed ha spinto le imprese a sviluppare nuovi modelli di business in cui il cliente è posto al centro. L'impossibilità di muoversi dalle proprie case ha stimolato le aziende a focalizzarsi su una strategia di omnicanalità che permanesse anche in seguito alla riapertura dei negozi fisici e che ha accelerato la trasformazione e l'evoluzione del processo d'acquisto, con la moltiplicazione dei *touchpoint* in tutte le fasi. Questa situazione ha portato anche allo sviluppo di trend nel settore degli articoli per la casa come il massiccio uso dell'e-commerce e l'attenzione alla sostenibilità ed agli acquisti *green* che verranno approfonditi di seguito.

6.2. Trend del settore

L'emergenza sanitaria ha portato al cambiamento del comportamento di acquisto di alcune fasce di consumatori, portando un'accelerazione a livello esponenziale di una serie di trend di mercato che si sarebbero altrimenti sviluppati nell'arco di diversi anni. Vediamo di seguito i principali trend di mercato che si sono affermati nel settore degli articoli per la tavola, la cucina ed il regalo nell'ultimo biennio ed i prodotti più richiesti.

6.2.1. Riscoperta della casa

La casa e, più in generale, l'ambiente domestico, hanno acquisito un ruolo di fondamentale importanza durante i periodi di *lockdown*. Sono diventati una sorta di contenitore, di scudo per fornire sicurezza ed incolumità e, per alcuni, anche un nuovo luogo di lavoro. Per questo motivo, anche tutti gli oggetti che la compongono, dall'arredamento, agli utensili, fino all'attrezzatura più pratica, hanno attirato maggiore attenzione e cure, portando i consumatori ad effettuare anche acquisti mirati in ambito casalingo. È aumentato il tempo per vivere intensamente la propria abitazione e per dedicarsi

alla cucina, con una riscoperta dei cibi tradizionali e per i lievitati fatti in casa, che hanno portato a riservare un'attenzione particolare verso gli articoli per la cottura come padelle e pentole.

Il trend che più si è diffuso è stata la riscoperta del piacere di cucinare. Lo stop alla vita frenetica ha portato le famiglie a riunirsi ed il calo degli impegni fuori casa ha permesso a molti di ricavarsi del tempo libero, spesso utilizzato per sperimentare in cucina e provare ricette tradizionali per cui nella vita "normale" non avrebbero avuto il tempo. Ecco che la conseguente preparazione di pane, pizza e dolci ha permesso lo sviluppo di nuovi comportamenti e stili di vita che hanno avuto un impatto significativo sulla macrocategoria dei casalinghi, principalmente nella seconda parte della crisi. Questo ha permesso alla maggior parte delle aziende appartenenti a questo settore di registrare buone performance commerciali. I produttori di *cookware* hanno sottolineato come gli utensili per la cottura, soprattutto per le preparazioni da forno, siano stati venduti come i *bestseller* del momento.

Questa evoluzione nel comportamento del consumatore tra le mura domestiche ha portato le aziende del settore casalinghi ad ampliare e diversificare l'offerta, individuando delle soluzioni merceologiche che rispondessero alle nuove esigenze per la casa degli italiani come i prodotti per la profumazione, la pulizia ed i piccoli elettrodomestici [54].

6.2.2. E-commerce

Nel mese di dicembre 2020, in Italia, la diffusione dell'online tra la popolazione ha raggiunto il 74,7%, aumentando del 4,7% rispetto al 2019. Lo annuncia il gruppo Casaleggio e Associati che, nel report annuale sull'andamento dell'e-commerce in Italia, ha registrato 44,7 milioni di utenti unici mensili nel 2020, con un incremento di 3,2 milioni rispetto al 2019. Questo rapido incremento di utenti online, con una conseguente evoluzione delle abitudini di consumo, ha costretto i brand ad una rapida evoluzione, per affrontare temi come la digitalizzazione dei pagamenti, la gestione di magazzini e logistica e lo sviluppo di un nuovo tipo di relazione con il cliente, tutte questioni diventate fondamentali a partire dal 2020. La richiesta di digitalizzazione dei brand da parte dei consumatori continua a salire ed il cambiamento nelle abitudini che ha caratterizzato questi due anni resterà irreversibile per la maggior parte degli italiani [15]. *"La crescita esponenziale dell'utilizzo del canale e-commerce per gli acquisti non solo nel settore degli articoli per la casa, è probabilmente una della novità più rilevanti dovuta alla pandemia da covid-9"* [6]. Nel 2020 il nuovo comportamento d'acquisto dei consumatori ha portato nel settore dei casalinghi ad un'esplosione del canale e-commerce dovuta anche conseguentemente ai blocchi prolungati alle aperture dei punti vendita fisici. Gli esperti ritengono che, nonostante non sia semplice prevedere l'evoluzione del fenomeno, è ragionevole pensare che almeno una parte della crescita in questo canale perdurerà anche dopo la fine dell'emergenza sanitaria [55]. Il settore della casa e dell'arredamento pesa ancora poco all'interno del

totale delle vendite online (circa un 1%). Nonostante ciò, nel corso del 2020 è stato caratterizzato da una forte crescita che si è aggirata attorno al 24%.

L'associazione ART rileva che oggi il 24% delle imprese di produzione, il 39% dei grossisti ed il 32% delle imprese al dettaglio del settore svolgono attività di commercio elettronico. Le imprese che svolgono attività di commercio all'ingrosso, inoltre, si caratterizzano per la stragrande maggioranza per l'utilizzo di piattaforme messe a disposizione dai *marketplace* come *Amazon* e *eBay*. Il canale di vendita online ha visto una crescita del 50% rispetto al 2019, con una quota di consumatori che hanno acquistato online articoli per la casa, la tavola e la cucina che è passata dal 17% a circa 35%, portando un'importante accelerazione nel processo di digitalizzazione delle imprese. Il vantaggio di questo incremento è stato sia degli operatori nativi digitali che delle imprese tradizionali. Si rileva che un 25% dei consumatori ha acquistato prodotti per la casa sui siti web di catene di negozi come *Kasanova*, *Zara Home* e *Coin Casa*, il 15% ha preferito acquistare direttamente dalle imprese produttrici, mentre il 13% su siti di negozi di prossimità [6].

6.2.3. Sostenibilità

Un altro trend particolarmente in crescita nel settore è quello delle produzioni *green*, che continuano a crescere confermando un comparto in buona salute. Secondo il rapporto annuale di Assobioplastiche, nel giugno 2020 la produzione italiana di manufatti compostabili è stata di oltre 10mila tonnellate, per un fatturato complessivo di 745 milioni di euro (+8,8% rispetto al 2019). Una delle grandi sfide che l'industria europea deve affrontare, infatti, è rappresentato dallo sviluppo di un'economia circolare: la qualità ambientale sta diventando un elemento competitivo di fondamentale importanza da sfruttare come strategia vincente soprattutto nei mercati internazionali. Il tema *green* si dimostra oggi quanto mai centrale anche nel settore dei casalinghi. Il consumatore è sempre più attento in fatto di imballaggi utilizzati nel confezionamento degli alimenti e di materiali utilizzati negli utensili e nei recipienti per la cottura e la conservazione dei cibi. L'attenzione si sposta verso il benessere proprio e della famiglia e cresce la consapevolezza dell'impatto ambientale di imballaggi e materiali non sostenibili [54]. I consumatori tendono ad acquistare sempre più prodotti casalinghi e utensili destinati a durare nel tempo e danno rilevanza a quegli articoli che vengono prodotti in stabilimenti che adottano scelte produttive attente all'ambiente e all'utilizzo consapevole delle risorse.

Il trend è stato accolto e condiviso dalla stragrande maggioranza dei produttori di articoli per la tavola e la cucina, che puntano sempre più sull'eco-sostenibilità sia nei processi di manifattura che nelle linee di prodotto. Questo viene tradotto in una serie di pratiche come la riduzione degli sprechi delle risorse, la limitazione delle emissioni nocive, l'utilizzo di materiali *green* e la propensione verso un

approccio sempre più circolare con le logiche del riciclo e del riutilizzo dei materiali. Lo dimostrano ad esempio la *Zero Waste*, capsule collection di Tognana per la conservazione ed il trasporto dei cibi per evitare gli sprechi e *Recy*, la nuova linea di pentole di Alluflon Spa, che unisce ad un processo industriale basato sulla trasformazione delle lattine riciclate in dischi di alluminio, una filiera corta, il 100% della qualità Made in Italy ed una conseguente riduzione dell'impronta ecologica della produzione. In controtendenza rispetto agli ultimi scenari globali in cui l'importanza della qualità è stata progressivamente soppiantata dalle strategie di prezzo, questa nuova realtà ha portato ad un interesse sempre maggiore anche verso la provenienza dei prodotti acquistati, spostando le preferenze del consumatore verso il meglio delle manifatture nazionali. Un esempio è l'utilizzo sempre più frequente del vetro, totalmente compatibile con tale approccio, essendo un materiale "healthy", versatile e allo stesso tempo molto affidabile. Pyrex®, in quanto leader di mercato, sta spingendo per la valorizzazione del concetto "all glass", in linea anche con le scelte di consumo sostenibile. La sostenibilità sta caratterizzando anche il campo dei casalinghi monouso. In seguito alla direttiva europea SUP (*Single Use Plastic*), tutte le classiche linee di stoviglie monouso in plastica sono state sostituite con nuovi prodotti realizzati in materiali eco-sostenibili. Il comparto registra risultati eccezionali, con un trend in significativa crescita e che ha registrato un aumento fino a tre punti percentuali [55].

7. Tognana Porcellane

Dopo aver definito il panorama di mercato in cui Tognana opera, è necessario conoscere da vicino la realtà aziendale, per scoprire la storia, le linee di prodotto che l'azienda offre e le strategie di business, per contestualizzare una successiva analisi sul database clienti.

7.1. L'azienda

Tognana è un'azienda italiana leader nella produzione e commercializzazione, sia al dettaglio che per il canale Ho.Re.Ca., nel settore degli articoli in porcellana e complementi per la tavola. Negli ultimi anni l'azienda si è trasformata da importante realtà manifatturiera ad affermata azienda internazionale con un volume di vendita che supera i 25 milioni di pezzi all'anno ed un fatturato in continua crescita. L'Italia rappresenta il mercato principale per Tognana (70% del fatturato totale), ma l'azienda è presente anche all'estero, principalmente nell'area balcanica dell'Europa e negli Stati Uniti. Con i suoi 75 anni di storia, Tognana è riuscita nel tempo a sviluppare un brand forte e riconosciuto, che fa leva sulla tradizione e che, allo stesso tempo, sta dimostrando negli ultimi anni uno spirito innovativo che ha permesso di ampliare l'offerta anche a dei segmenti di mercato più giovani [54]. Per questo motivo Tognana riesce ad offrire una gamma di prodotti con alte performance di qualità dove forme e decori eleganti, tipici di un gusto classico e della tradizione della manifattura in porcellana, si

fondono con le richieste del mercato attuale di leadership tecnologica e distributiva universalmente riconosciuta. Tognana Porcellane si dimostra anche un'azienda capace di stare al passo con i trend dell'attualità, con una sempre maggiore attenzione verso l'eco-sostenibilità e lo sviluppo e l'attenzione costante verso il canale e-commerce sempre più essenziale e performante nella strategia distributiva. Il brand è presente a livello internazionale attraverso diversi canali. Conta infatti con 1000 punti vendita nel canale retail, 600 punti vendita nella GDO/GDS e distribuzione nel canale Ho.Re.Ca., con un portafoglio prodotti che comprende più di 1600 referenze [65].

7.2. La storia

L'inizio della storia dell'azienda risale al 1775, con la famiglia Tognana impegnata nella produzione industriale nel settore dei laterizi. È necessario attendere fino al 1946 per assistere all'inaugurazione dell'azienda ceramica come la conosciamo oggi, con un breve passaggio dalla produzione di maioliche alla porcellana, che seguiva tradizionali sistemi di realizzazione: dalla cottura alla sinterizzazione fino alla vetrificazione. Tognana si instaurò subito nel mercato come uno tra i maggiori produttori italiani di stoviglie in porcellana di alta qualità ed acquisì ulteriore popolarità grazie anche alle collaborazioni con designer di fama internazionale e all'incorporazione nel 1989 dell'azienda Nuova Fontebasso Spa, storica realtà trevigiana operante anch'essa nel settore della ceramica e delle porcellane. Uno sviluppo importante del brand si ebbe nel 1998 con la partnership con la Famiglia Morosi che nel 2000, dopo l'acquisizione totale dell'azienda, si occupò del rilancio della stessa, apportando cambiamenti importanti e concentrando gli investimenti sul fronte della modernizzazione degli impianti e dei cicli produttivi. Grazie a questo rilancio del brand, il Gruppo Tognana iniziò ad acquisire popolarità anche a livello internazionale. Negli anni seguenti l'azienda continuò a crescere arrivando nel 2009 ad introdurre le prime linee di pentole in fusione di alluminio con interno in ceramica, nuovo inizio per il brand, e nel 2014 ad evolversi ancora nel segmento cottura introducendo *Evoluta*, la linea di pentole creata in seguito ad un importante studio sul risparmio energetico. Un'altra novità di questi anni fu *Ubiqua*, la gamma completa di pentole utilizzabili su tutti i tipi di piano cottura che permette una modalità di cottura naturale e salutare, accompagnata dalla linea *Roq*, costituita da una serie di articoli coordinati alla linea *Ubiqua*, con le medesime caratteristiche per i prodotti per la tavola. Oggi Tognana, grazie alle evoluzioni che hanno visto il brand protagonista fin dalla nascita, occupa in maniera indiscutibile una posizione di leader di mercato nel settore dei prodotti per la tavola, la cucina e la casa [65].

7.3. Portafoglio prodotti

L'azienda si propone sul mercato con un ampio assortimento di prodotti che si differenziano in primo luogo per l'ambito di destinazione. Le referenze si suddividono in proposte per il settore *professional*

da un lato, e per il consumatore finale dall'altro, puntando in entrambi i casi su un prodotto di qualità, sulla tradizione e la creatività del design italiano [65]. Concentrandosi sul segmento B2C, che sarà oggetto di studio, l'azienda offre al consumatore quattro tipologie di prodotti differenti ma complementari, raggruppati nelle categorie tavola, cottura, tessile e arredo casa.

7.3.1. La tavola

Il segmento tavola identifica l'azienda fin dai suoi primi anni di attività. Passione e cura nei particolari sono le qualità che contraddistinguono i prodotti per la tavola di Tognana, assieme al materiale di produzione: la porcellana. L'elevata qualità delle ceramiche Tognana è testata e controllata in laboratorio: il brand propone prodotti resistenti al test di lavaggio in lavastoviglie, di abrasione e resistenza meccanica, materiali con la capacità di assorbire calore, nonché vernici certificate da parte della *food and drug administration* per la sicurezza della salute. Le linee di prodotto di questa categoria comprendono tutto il necessario per la tavola: dai servizi piatti, ai bicchieri, posate, coltelli e taglieri. Propone inoltre tazze e tazzine, barattoli per caffè, zucchero, biscotti, spezie e accessori per donare alla tavola un tocco di stile ed innovazione [65]. Rappresentano la categoria di punta del brand, che conta con il maggior volume di prodotti venduti tra gli articoli dell'assortimento in tutti i canali di vendita.

7.3.2. La cottura

Il segmento cottura rappresenta la categoria in maggior sviluppo per il brand, soprattutto nell'ultimo periodo, in seguito all'evoluzione delle abitudini del consumatore in casa dovute alla pandemia. Come afferma l'amministratore delegato dell'azienda Morosi: *“Con le limitazioni causate dal Covid-19, le persone hanno trascorso sempre più tempo all'interno delle mura domestiche, riscoprendo o enfatizzando il piacere di cucinare a casa con tutti gli strumenti a disposizione”,[...]* *“questo è il motivo per cui abbiamo continuato ad investire in questo campo, creando prodotti che coniugano in maniera sempre più magistrale la funzionalità e l'estetica, e puntando, in parallelo sulla produzione italiana”* [54].

Tognana Porcellane offre una vasta scelta di pentole e padelle, ma anche bistecchiere, grill e pirofile per soddisfare sempre più i desideri dei consumatori, con design esclusivi e qualità certificata [65]. La qualità dei materiali è un punto di forza anche in questa categoria di prodotti, che garantiscono un'ottima cottura dei cibi grazie ai materiali antiaderenti e all'acciaio inox 18/10, nonché alle ceramiche di alta qualità.

7.3.3. Il tessile e l'arredo casa

La categoria tessile comprende tovaglie, runner, strofinacci, set all'americana, grembiuli, guanti, presine e cuscini realizzati in correlazione alle fantasie dei prodotti per la tavola. Molte proposte vengono fabbricate in cotone riciclato e attraverso processi produttivi che pongono particolare attenzione all'eliminazione degli sprechi e al riutilizzo degli scarti, sotto un'ottica *eco-friendly*. Inoltre, Tognana offre a completamento dell'offerta anche una serie di oggetti di arredamento per la casa come alcuni vasi, statue, portafoto. Con i brand Tognana e Andrea Fontebasso 1760, l'obiettivo è quello di unire il mondo della tavola con quelli del tessile e dell'arredo e presidiare il mercato con linee funzionali all'arredamento *total look* della casa [54].

7.4. E-commerce e strategia digitale

Tognana ha sviluppato con il tempo una strategia distributiva che le permette di presidiare la maggior parte dei canali distributivi. L'azienda al momento è molto presente nella distribuzione organizzata GD/GDO, come canale fisico ed altri canali come il dettaglio qualificato, il negozio, o il canale alberghiero per circa un 40% del fatturato. Secondo quanto dichiarato dall'E-commerce Manager di Tognana Luca Peron, il valore dei ricavi provenienti dall'online non è ancora rilevante nei confronti dei canali fisici sopra descritti, contando solo con un 10% sul totale del fatturato, ma l'azienda sta investendo molto sull'e-commerce perché si dice certa che rappresenti lo sviluppo del futuro delle vendite nel settore.

Considerato lo sviluppo che il canale e-commerce dell'azienda sta avendo negli ultimi anni e l'importanza attribuitagli nella crescita del brand in un prossimo futuro, l'elaborato presenta un'analisi approfondita sul database clienti che acquistano online, esponendo le tecnologie di *marketing automation* che vengono utilizzate per acquisire e gestire nuovi contatti. Prima di addentrarsi nella vera e propria analisi dei dati, viene presentata una panoramica sulla struttura dello *shop* italiano online, così da contestualizzare le modalità di raccolta dei dati sugli utenti, il loro percorso di acquisto e la successiva analisi dei dati comportamentali raccolti.

7.4.1. Lo shop online: struttura e modalità di navigazione

Tognana è presente sul web in Italia sia attraverso un sito istituzionale, che con lo *shop* online ed un *magazine* digitale. Nel primo è possibile scoprire maggiori informazioni riguardo l'azienda, relativamente alla sua storia, alle caratteristiche del brand ed ai suoi prodotti. Contiene anche una sezione dedicata agli eventi ed ai cataloghi sia per il target B2B che B2C. Dalla home page del sito

istituzionale è possibile accedere al *magazine*, una raccolta di articoli dedicata al mondo Tognana e Andrea Fontebasso 1760 che trattano dei prodotti dell'assortimento ma anche di curiosità sul mondo del *food, beverage* e, più in generale, di tutto ciò che fa riferimento all'ambito alimentazione e cura della casa.



Figura 15: Home page del sito istituzionale di Tognana Porcellane (Fonte: www.tognana.com/it)

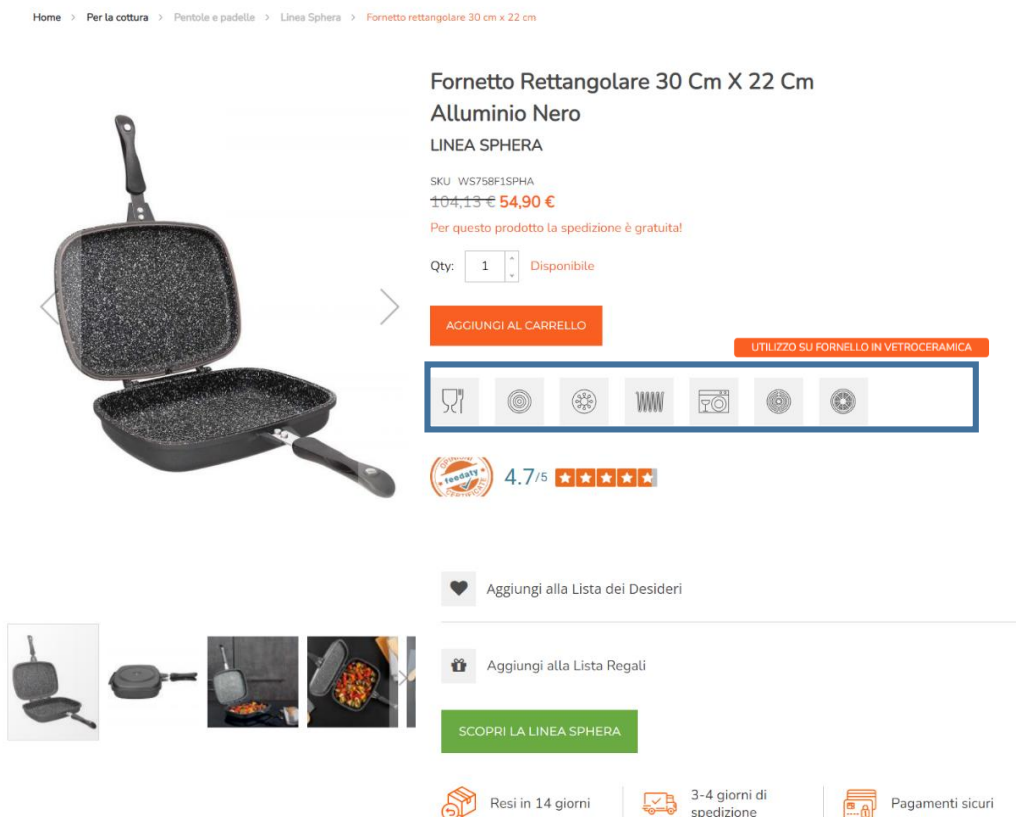
Lo *shop* online di Tognana è suddiviso per ambito di destinazione. Non appena si accede alla home page del sito istituzionale, infatti, è possibile scegliere tra due tipi di negozi: una sezione riservata al canale Ho.Re.Ca, utilizzata dal *target* B2B (*professional*), ed una sezione specifica per il consumatore finale, che costituisce il *target* B2C (*shop*). Nell'elaborato si farà riferimento allo *shop* dedicato alla vendita non professionale. Nella home page dell'e-commerce B2C è possibile navigare il sito per categoria (tavola, cottura, tessile, arredo casa) e sono presenti le sezioni "novità" e "outlet" che raccolgono rispettivamente i prodotti appena lanciati sul mercato ed i prodotti in saldo. La pagina raccoglie anche delle sezioni con le linee di prodotto in evidenza, i prodotti raccomandati ed una selezione di articoli raggruppati per momenti speciali, eventi o festività. Inoltre dalla home page è possibile accedere facilmente al proprio carrello, all'account personale ed alla *wishlist* dei prodotti preferiti.



Figura 16: Home page dello shop IT online di Tognana Porcellane (Fonte: www.tognana.com/it)

Una volta selezionata la categoria di interesse, si viene indirizzati allo shop online vero e proprio, con la possibilità di scegliere i prodotti ed inserirli a carrello. La pagina categoria dà la possibilità di filtrare i prodotti per tipologia, prezzo, ma anche per modalità di utilizzo, materiale, colore e marchio. Inoltre, in ogni pagina del sito è possibile utilizzare una la barra di ricerca che permette di individuare dei prodotti specifici o una tipologia di articoli già conosciuta.

La scheda prodotto di ogni prodotto si presenta ricca di informazioni e strutturata come segue.



Il fornello rettangolare 30 cm x 22 cm con **doppia induzione** della linea Sphera funge da fornello, piastra, griglia, girafrittata e saltiera: un unico immancabile prodotto per esprimere tutta la tua creatività in cucina. Il corpo è in alluminio pressofuso con uno strato interno in **microsfere di porcellana** che garantisce un'**antiaderenza superiore**; il fondo è in acciaio inox thermoradiante antideformazione che consente performance eccellenti ad alte temperature anche su piano ad induzione. Il fornello rettangolare Sphera è adatto quindi a tutti i piani di cottura, e permette inoltre una cucina sana e salutare in quanto è Nickel Free e PFOA Free. Il manico in bakelite Soft Touch ha un design elegante, raffinato e brevettato. È possibile lavare il fornello rettangolare della linea Sphera in lavastoviglie. EVERGREEN

Figura 17: Scheda prodotto dell'e-commerce di Tognana Porcellane (Fonte:www.tognana.com/it)

Oltre alla possibilità di inserire il prodotto a carrello e scegliere la quantità da acquistare, l'utente nelle schede prodotto dello *shop* di Tognana ha accesso ad una numerosa serie di informazioni molto utili per la scelta dei prodotti da acquistare. La schermata mostra, ad esempio, se la spedizione è gratuita o a pagamento, il tempo di spedizione, una serie di foto che presentano al meglio l'articolo ed alcune caratteristiche o i punti di forza dello stesso. Passando con il puntatore del mouse sopra alle icone evidenziate in figura 17, è possibile avere tutte le informazioni riguardo le funzionalità del prodotto come ad esempio la possibilità di utilizzare i prodotti cottura in un forno in vetroceramica piuttosto che nei piani ad induzione. Anche le recensioni sono importanti per migliorare l'esperienza di acquisto e spingere il cliente a mettere il prodotto a carrello e, in questo caso, sono semplici da consultare ed immediate. Scorrendo verso fondo pagina sono presenti anche una serie di prodotti correlati all'articolo visualizzato, una sezione prodotti raccomandati ed un ricettario per cui è possibile effettuare il download.

Le schede prodotto e lo *shop* online più in generale, si presentano ricchi di contenuto, e, oltre che attrarre e spingere l'utente all'acquisto, permettono ai visitatori di effettuare una serie di azioni sulla piattaforma, che possono essere rilevate, tracciate ed analizzate, al fine di utilizzarle per migliorare l'esperienza utente online ed aumentarne il tasso di fidelizzazione e la redditività.

8. Marketing automation per lo shop IT

Per immagazzinare i dati raccolti attraverso il sito web e gestire i contatti provenienti dallo *shop* online IT, Tognana si affida ad un'agenzia specializzata che cura questi delicati aspetti della relazione con il cliente attraverso strumenti di *marketing automation*.

8.1. Cos'è la marketing automation?

Quando si sente parlare di *marketing automation*, si fa riferimento ad un ramo del marketing che ha come scopo la personalizzazione dell'esperienza dell'utente, principalmente nei canali digitali, e la massimizzazione del profitto per ogni acquirente. È una funzione del marketing relativamente recente, o meglio, esiste da molti anni, ma è sempre stata utilizzata da pochi innovatori o da grandi e

strutturate società del digitale. Tra la maggior parte delle aziende che dovrebbero utilizzarla nei loro canali di vendita, soprattutto in quelle di piccole dimensioni, a volte risulta ancora poco conosciuta. Per questo il tema è ancora un po' confusionario, spesso associato solamente alle automazioni relative all'*email marketing*, alle attività di *remarketing* o, ancora, alla sola creazione di *workflow* o all'automazione dei processi aziendali. Altre volte la pratica di *marketing automation* viene identificata semplicemente con la scelta di alcuni software che permettono di inserire funzionalità aggiuntive sul proprio e-commerce o sito web. Ma, in realtà, rappresenta più una strategia che un semplice strumento tecnologico [17].

L'utilizzo della *marketing automation* consente alle aziende di rendere rilevante il proprio e-commerce per ogni visitatore o per gruppi di essi, permettendo di personalizzare il sito web a seconda del tipo di utente e del comportamento di navigazione e acquisto che assume una volta che è atterrato sulla pagina. Questo è ciò che differenzia e che dà un vantaggio competitivo agli *shop* online, rispetto ai tradizionali negozi di retail. La *marketing automation* permette di utilizzare e sfruttare il traffico esistente sul sito ed aumentare i ricavi sui contatti già presenti nel database dell'e-commerce. Di conseguenza, più visite si hanno nel sito e più email utenti sono presenti nella piattaforma, più le attività di automazione saranno efficaci. Permette di inviare due tipi di comunicazioni automatiche sulla base del comportamento di acquisto o delle preferenze degli utenti. Si distinguono tra attività di comunicazione *on-site*, che avvengono in *real time* (pop up, banner, ecc.) e *off-site*, che comprendono tutte quelle comunicazioni inviate quando l'utente non è sul sito e che vengono personalizzate a seconda dei dati raccolti dalle precedenti visite, relativamente alla navigazione o agli acquisti effettuati. Le attività *off-site* sono per esempio l'invio di *email one to one*, di messaggi, notifiche push, ecc.. Le automazioni funzionano secondo una serie di regole definite dai *marketers*, per cui se si verifica l'evento x, scatta in automatico l'automazione y, sia essa *on-site* o *off-site* [17].

La *marketing automation* non funziona da sola. Per aumentare il profitto sui clienti online è necessario strutturare alla base delle attività di ottimizzazione dell'e-commerce, in particolare relativamente all'usabilità ed al disegno dei percorsi di navigazione, che permettano di raccogliere nuovi contatti e di perfezionare le attività di automazione. Un percorso di navigazione ben strutturato diventa fondamentale al giorno d'oggi, soprattutto da mobile, dove l'esperienza online è costituita da una serie di micromomenti e di connessione 5G, per cui la velocità di risposta di un e-commerce durante gli acquisti è essenziale [17].

Raccogliere dati sugli utenti durante la navigazione permette di creare dei segmenti nella piattaforma di *marketing automation*, necessari per poter attivare le automazioni impostate. La segmentazione sta infatti alla base delle attività non solo di *automation* ma di marketing più in generale per suddividere

i clienti in base al loro stato nel *funnel* di acquisto o alle loro caratteristiche demografiche e comportamentali. L'utilizzo delle *analytics* diventa anche qui fondamentale nella definizione del ciclo di vita del cliente medio e per attribuire i giusti valori ad ogni segmento. Come vedremo in seguito, queste piattaforme propongono spesso anche delle matrici RFM, che vanno accuratamente settate a seconda del pubblico di riferimento.

Gli strumenti di *marketing automation* consentono di potenziare siti e piattaforme di e-commerce attraverso una serie di funzionalità avanzate, per la maggior parte supportate dall'intelligenza artificiale e dal *machine learning*.

Recommendation systems dinamici. Sono un elemento molto importante nella *marketing automation*. In media solo il 5% dei prodotti presenti a catalogo vengono visualizzati dagli utenti su un e-commerce, che trascorrono un tempo in pagina generalmente limitato a pochi secondi. I prodotti raccomandati dinamici permettono ai visitatori di trovare fin da subito i prodotti per loro rilevanti e che hanno una più alta probabilità di essere acquistati. Considerando che la selezione di prodotti mostrati sul sito potrebbe non interessare tutti gli utenti, l'obiettivo dei sistemi di raccomandazione dinamici diventa quello di restituire al cliente i prodotti che effettivamente gli interessano durante la navigazione. Attraverso questa funzionalità si riesce ad automatizzare un'operazione che fatta manualmente richiederebbe molte ore di lavoro. La *marketing automation* in questo caso permette di fare ciò che *Amazon* fa ormai da molto tempo. Grazie ai software di automazione, le raccomandazioni possono essere gestite e sostituite in maniera dinamica, sulla base di complessi algoritmi che restituiscono i prodotti correlati con la probabilità più alta di essere venduti. È possibile scegliere tra diversi algoritmi di selezione a partire dai più generici come: visti di recente, acquistati insieme a questo prodotto o i più popolari oggi. Questa funzionalità permette anche di impostare degli A/B test, in modo da far ruotare sulla stessa pagina due diversi algoritmi e misurare quale risulta più performante. Ciò che caratterizza i sistemi di raccomandazione è che più dati gli algoritmi possono processare, meglio funzionano. Diventa quindi fondamentale costruire alla base una strategia per portare traffico al sito e permettere così a *big data* e *machine learning* di migliorare via via le proprie previsioni [17].

Web personalization in real time. Gli strumenti di *marketing automation* permettono di personalizzare i contenuti del sito web, come testi e grafiche, in tempo reale in base al comportamento dell'utente durante la navigazione. L'utilizzo delle micro personalizzazioni, ad esempio, può spingere i nuovi utenti ad iscriversi alla newsletter o a creare un account sul sito, può mostrare la spedizione gratuita solo quando viene effettivamente applicata, o creare urgenza comunicando il numero di prodotti ancora disponibili [17].

Coupon e offerte personalizzate. Attraverso l'utilizzo di pop up, message bar, message box, replace html e email, è possibile automatizzare la gestione delle offerte in base ai differenti segmenti di utenti identificati e massimizzare così i margini sulle vendite [17].

Cross-selling e up-selling dinamico. Oltre ad effettuare questo tipo di attività attraverso i sistemi di raccomandazione, è possibile integrare ulteriori automazioni per spingere l'utente verso l'acquisto di prodotti addizionali o prodotti più costosi rispetto a quelli di partenza. Un esempio può essere l'offerta di pacchetti risparmio o kit in promozione prima che l'utente effettui acquisti sul sito, oppure, una volta che il prodotto è stato inserito a carrello, mostrare articoli più rilevanti o correlati con il prodotto visto, spingendo verso attività di *cross-selling* [17].

Email marketing one to one. Rappresenta la funzionalità più potente, efficace e anche maggiormente sfruttata da chi utilizza piattaforme di *marketing automation* per il proprio e-commerce. Attraverso la profilazione avanzata e la segmentazione degli utenti del database è possibile inviare comunicazioni mirate agli utenti, che risulteranno molto più efficaci delle email di massa a contenuto standard [17].

Recupero carrelli abbandonati. Questa funzionalità permette di concentrarsi sulla fase finale del processo di acquisto, quando questo per qualsiasi motivo non viene concluso. La fase di pagamento e *checkout* è generalmente la più delicata del *sales funnel* e per questo diventa fondamentale avere degli strumenti che permettano di provare a recuperare i carrelli che non si sono ancora tramutati in acquisti [17].

Geolocalizzazione e marketing di prossimità. Permette di riconoscere continente, nazione, città, lingua dell'utente, informazioni che possono essere utilizzate per migliorare le comunicazioni sulla spedizione, informare della presenza di un punto vendita in prossimità di utenti che navigano da paesi esteri o da determinate regioni o città italiane, mostrare contenuti diversi per paese [17].

Notifiche push, sms, chat e bot. Sono funzionalità che aiutano a mantenere un contatto costante con il proprio utente, sia a fini commerciali che di servizio [17].

Lead generation. Con *lead generation* si fa riferimento a tutte quelle attività che mirano all'acquisizione di nuovi contatti da inserire nel database. Aumentare il numero di *lead* (email) permette alla *marketing automation* di essere più efficace e per fare ciò una delle tecniche più utilizzate consiste nell'impostazione di pop-up comportamentali all'interno del sito web. È possibile crearli all'interno dei software di *automation* o attraverso servizi esterni; l'importante è realizzare dei pop-up personalizzati ed utilizzarli su specifici segmenti potenzialmente interessati, come potrebbero essere ad esempio i nuovi visitatori [17].

A/B test. I software di *marketing automation* permettono anche di fare *A/B test* sulle azioni e sugli algoritmi, consentendo di confrontare le performance tra versioni diverse e migliorandole continuamente. [17]

Dopo aver presentato le principali caratteristiche e potenzialità degli strumenti di *marketing automation*, viene presentato SALESmanago, il software scelto dall'azienda per implementare alcune attività di automazione sullo *shop IT*. Verranno presentate anche alcune tra le principali funzionalità sopra descritte che Tognana attualmente utilizza per la *marketing automation* sul proprio e-commerce.

8.2. Il software: SALESmanago

SALESmanago è una piattaforma *all-in-one* di *Customer Data* e *Marketing Automation* in *cloud* basata sull'intelligenza artificiale. Il software sfrutta tutta la potenza dei dati proprietari e dei *zero-party data*, combinando analisi avanzate, apprendimento automatico e automazione dell'AI per offrire esperienze personalizzate altamente configurabili, attraverso una gamma completa di canali di esecuzione del marketing integrati. Rappresenta la nuova generazione dei software di *marketing automation*, poiché attraverso *machine learning* e intelligenza artificiale permette l'aggregazione e l'analisi dei dati dei clienti, consentendo la personalizzazione dei canali anche in tempo reale. I reparti di marketing hanno in questo modo tutti gli strumenti e le informazioni necessarie per la creazione di campagne con un tasso di conversione elevato ed un miglioramento continuo nei servizi offerti ai clienti. L'azienda ha sede in Polonia, a Cracovia, e collabora con imprese di medie dimensioni ma anche con realtà di alto calibro come Starbucks, Victoria's Secret, Vodafone e molte altre. È una tra le piattaforme di automazione classificate da *Forbes* come in più rapida crescita in Europa [12].

In primo luogo, SALESmanago identifica gli utenti del sito web di riferimento, analizzando il comportamento online e tenendo traccia della cronologia delle transazioni. Una volta effettuata l'analisi, permette di creare comunicazioni personalizzate durante il ciclo di vita del cliente attraverso svariati canali di marketing come l'email, le notifiche web push, i social media, le live chat, le app da mobile, le reti pubblicitarie (Google, Facebook, ecc.) ed anche attraverso il sito web stesso, con la possibilità di personalizzazione della schermata a seconda delle caratteristiche del visitatore. La piattaforma consente di analizzare in profondità il database clienti per scovare le informazioni necessarie per la gestione dei contatti. Permette di creare anagrafiche e profili sui singoli clienti con informazioni sia comportamentali e personali che transazionali, monitora le abitudini di consumo, registra come reagiscono gli utenti alle campagne di marketing dirette, gestisce il consenso degli iscritti per garantire la conformità al GDPR sulla privacy, effettua analisi predittive sui clienti e permette di sincronizzare i dati con piattaforme DMP o *data warehouse*. Questo consente al software

di effettuare anche una segmentazione dinamica del database clienti e di inviare contenuti personalizzati che soddisfino le esigenze di ognuno e ne aumentino il coinvolgimento. Ogni flusso di automazione viene attivato da alcune azioni, o *trigger*, che l'utente compie sul sito web, permettendo di far entrare i clienti di interesse nel processo desiderato e assegnando anche delle etichette, chiamate *tag*, che aiutino nella segmentazione della base utenti. Ai *lead* acquisiti viene attribuito anche un punteggio, lo *score*, che identifica chi di loro ha maggiore interesse o ha avuto maggiori interazioni con il brand [33].

SALESmanago è una piattaforma completa e multifunzionale che crea canalizzazioni di vendita, ne misura le prestazioni e le ottimizza per rendere i risultati sempre migliori [33]. Inoltre, è un sistema in continuo aggiornamento e che permette di comprendere quali tra le attività di marketing sono le più efficienti, individuare le aree in cui è più conveniente investire e creare una *user experience online* adatta ad ogni tipo di cliente [4].

8.3. Campagne di marketing automation

Tognana Porcellane sfrutta una serie di funzionalità del software SALESmanago nelle sue campagne di *marketing automation* per l'e-commerce. Le principali sono approfondite di seguito.

8.3.1. Segmentazione utenti

SALESmanago permette di segmentare in maniera automatica gli utenti che entrano in contatto con l'azienda tramite il sito web. Attraverso i comportamenti e le interazioni online, raccoglie dati sui clienti creando dei profili personalizzati per singolo contatto in cui vengono immagazzinate tutte le informazioni relative alle azioni svolte sul sito. Permette di effettuare una iper-segmentazione del database contatti senza necessità di utilizzare il codice, attraverso l'intelligenza artificiale, che consente di creare segmenti rilevanti in pochi clic [70].

Il modo più semplice per segmentare gli utenti in entrata sono i *tag*, che rappresentano delle etichette assegnate ai contatti per consentire la loro separazione in base ad interessi, attività sul sito web, dati demografici o provenienza del contatto. È possibile assegnarli attraverso modalità e tempistiche differenti. La definizione dei *tag* può avvenire per particolari gruppi di utenti durante l'importazione del database, in seguito alla registrazione volontaria di moduli da parte dell'utente, attraverso le matrici di segmentazione, in base a ciò che un contatto ha trovato interessante sul sito web (ad esempio attraverso il tracciamento delle visite agli url). Inoltre è possibile impostarli attraverso le *automation rules* o i *workflow*, definendo, a seconda delle interazioni svolte, che etichetta assegnare anche in base a delle condizioni predefinite. Il software assegna automaticamente i *tag* di sistema che

raccogliono informazioni relativamente al browser utilizzato dall'utente ed al sistema operativo. Possono essere aggiunti anche manualmente, direttamente all'interno delle schede contatto [69].



Figura 18: Automation rule per l'assegnazione del tag SLEEPY (Fonte:SALESmanago)

In figura 18 viene presentato un esempio di *automation rule* (che verrà approfondita in seguito) dedicata all'assegnazione del tag "SLEEPY" ai quei clienti che non acquistano da più di 180 giorni e vengono definiti di conseguenza "dormienti". In questo caso l'automazione si attiva per quei contatti che hanno effettuato almeno un acquisto sul sito web IT (contrassegnati con il tag "CLIENTE_IT"). Dopo 181 giorni il flusso richiede di verificare se per il contatto si verifica un acquisto, ovvero se avviene un *external event* chiamato *purchase*. Se sì, non succede nulla e il contatto esce dal flusso, se invece la condizione non si verifica, al contatto viene assegnato il tag "SLEEPY". Questo è un esempio di come una regola di automazione permetta di attribuire delle etichette a gruppi di clienti e segmentarli.

Oltre a suddividere i contatti attraverso i tag, è possibile, nel *Customer Segmentation Center*, creare dei segmenti rilevanti per le necessità dell'azienda in maniera automatica. Questa sezione rappresenta un centro unificato in cui è possibile creare, gestire, modificare e analizzare tutti i segmenti che il reparto marketing dell'azienda ritiene interessanti. Un esempio di segmenti efficaci soprattutto in ambito e-commerce sono i segmenti dinamici, o *Dynamic Segments*. La segmentazione dinamica permette di creare e aggiornare in tempo reale tutti i gruppi di contatti, in base a dei criteri definiti tra i dati CRM, comportamentali e transazionali, in un momento determinato. È possibile in questo modo aggiornare in maniera automatica dei gruppi *target* senza dover generare report di dati manualmente

per selezionare i contatti appropriati, gestire dati multivariati senza supporto di sistemi esterni, IT o analisi, determinare precisamente e in maniera immediata la dimensione del gruppo per le campagne automatiche e generare analisi dettagliate che mostrino come cambiano composizione e dimensione dei segmenti nel tempo [71].

8.3.2. Automation Workflow

Il cuore delle attività di SALESmanago sono i flussi di *automation* che permettono di automatizzare le comunicazioni con i clienti attraverso vari canali. Accelerano i processi di vendita, in modo da rendere i contatti pronti ad effettuare un acquisto e poter rispondere individualmente al comportamento degli utenti sul sito web.

Regole di automazione. All'interno della piattaforma si possono creare processi avanzati grazie all'utilizzo delle regole di automazione che vengono costruite sulla base di tre specifici elementi: gli eventi, le condizioni e le azioni che permettono di creare regole su misura a seconda delle esigenze del business.

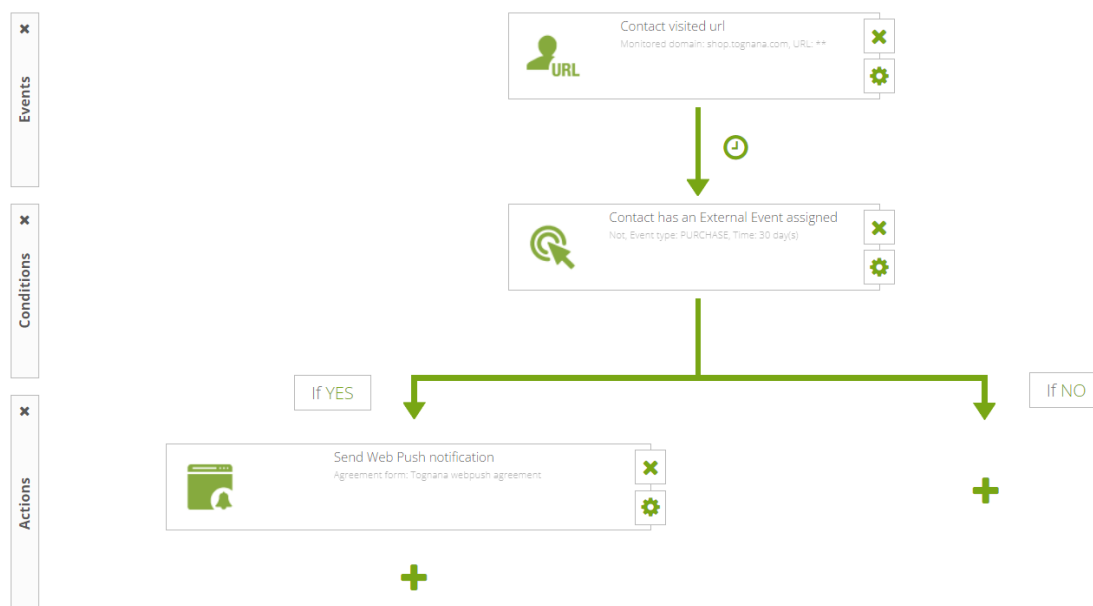


Figura 19: Struttura di una automation rule per l'invio del modulo di agreement per le web push (Fonte:SALESmanago)

Le *automation rules* permettono di rendere automatica la comunicazione con i clienti, ma anche di velocizzare il processo di vendita, preparare i *lead* da passare al *sales team* e rispondere in maniera mirata ai comportamenti degli utenti sul sito web. Vengono create, come in figura 19, attraverso la configurazione di:

- **Eventi:** sono i *trigger* che attivano le regole per il contatto. A seconda del comportamento di ciascun utente permettono di adattare le strategie di marketing e le azioni di comunicazione o di *web personalization* da compiere. Nell'esempio in figura l'evento assegnato al contatto è la visita alla home page dello *shop* di Tognana.
- **Condizioni:** vengono impostate come elementi necessari che devono essere soddisfatti per permettere al contatto di proseguire all'interno della regola di automazione. La condizione da soddisfare nel caso in figura è che l'utente non abbia effettuato un acquisto negli ultimi 30 giorni.
- **Azioni:** se i contatti soddisfano tutte le condizioni precedentemente verificate, vengono eseguite le azioni definite nella regola di automazione. Nel caso in figura, se tutte le condizioni sono state soddisfatte viene inviata una notifica *web push* con la richiesta di *agreement* per le raccomandazioni prodotto [67].

Flussi di automazione. Le regole di automazione possono costituire una piccola parte dei più complessi *workflow*, che permettono di costruire campagne di marketing strutturate su più livelli ed includono processi più avanzati. Rappresentano uno strumento di pianificazione e automazione delle campagne più articolate e varie, anch'esse attivate sulla base dei comportamenti dei singoli clienti sul sito web. Tutte le attività effettuate dal contatto e che vengono tracciate possono essere utilizzate come eventi e/o condizioni per attivare le azioni del flusso [68]. In figura 20 è possibile notare come si struttura un *automation workflow* all'interno della piattaforma.

Di seguito vengono approfonditi alcuni tra i flussi più comunemente utilizzati dai *marketers* negli e-commerce, come definiti per Tognana Porcellane.²

² Tutti i flussi descritti sono stati ideati dall'agenzia di riferimento Madam Agency Srl e studiati durante il periodo di *stage* curriculare.

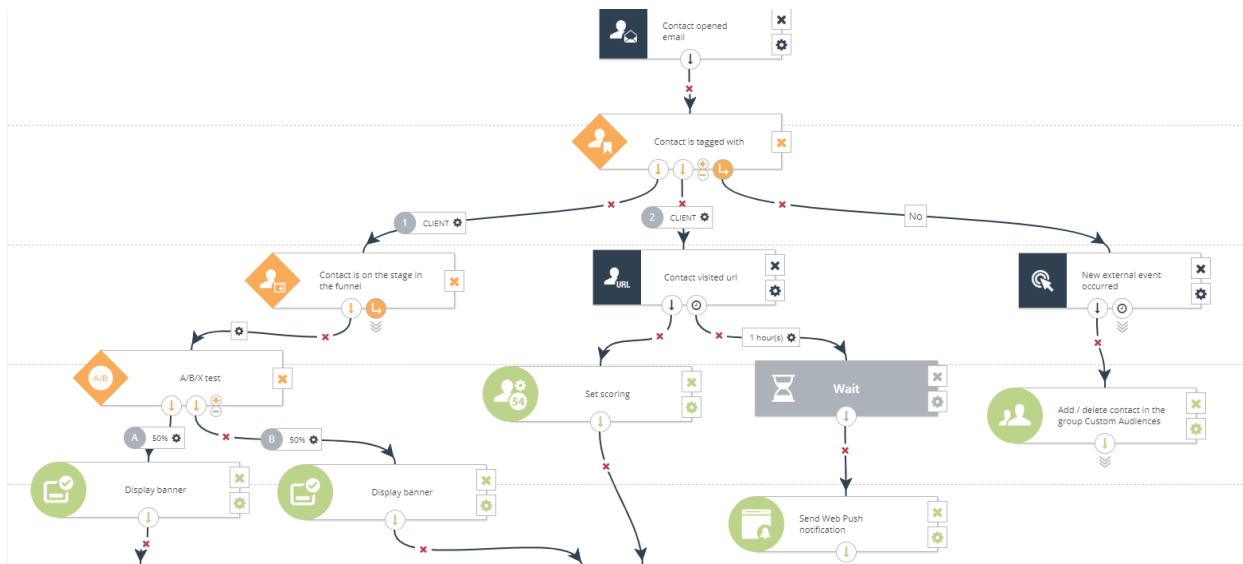


Figura 20: Esempio di struttura di un workflow (Fonte:SALESmanago)

Welcome campaign. Come suggerisce il nome stesso, è una campagna di benvenuto per i nuovi utenti. Viene attivata per tutti i visitatori che si iscrivono alla newsletter da qualsiasi sezione del sito web ed ai quali viene automaticamente assegnato il *tag* NEWSLETTER_IT. Una volta taggati, gli utenti entrano all'interno di questo flusso per cui gli viene inviato un ciclo di email che dà il benvenuto nella community di Tognana Porcellane ed assegna uno sconto del 10% da utilizzare nello *shop* online. Allo stesso tempo, il *workflow* attiva per questi utenti dei banner dinamici e li aggiunge alla lista di *remarketing* per le pubblicità su Facebook e su Google. Se l'utente utilizza lo sconto effettuando un acquisto, gli viene automaticamente assegnato il *tag* WELCOME_PURCHASE, altrimenti si attiva un flusso di email che avvisa l'utente sulla scadenza dello sconto e che suggerisce alcune idee per l'acquisto. La condizione *purchase* viene controllata secondo gli intervalli: 1gg, 5gg, 9gg, 14gg, 1gg. A conclusione del flusso viene inviata una email in cui è presente un form di profilazione avanzata per conoscere le preferenze degli utenti che non hanno compilato il form durante la fase di iscrizione alla newsletter. Il flusso dura 30 gg.

Carrello abbandonato. Le automazioni di carrello abbandonato sono le più utilizzate e anche quelle che generalmente risultano più redditizie per l'azienda. Vengono attivate per recuperare gli acquisti che non sono stati conclusi, spingendo l'utente a rivedere i prodotti che ha lasciato a carrello. Il *workflow* ha inizio nel momento in cui il sistema registra che alcuni articoli sono stati messi a carrello. Dopo 1 ora, se non si registra l'evento *purchase*, il flusso invia una email con i prodotti che l'utente ha "dimenticato" a carrello e lo inserisce nel pubblico di *remarketing* per Facebook e Google. Se anche in seguito a questa prima email non avviene l'acquisto, dopo 24h e dopo 5 giorni viene inviato

un ulteriore flusso di comunicazioni che si conclude con l'assegnazione al cliente di uno sconto del 5% per completare l'ordine. Se l'utente procede all'acquisto, esce dal flusso in qualsiasi fase si trovi.

Post acquisto. Il *trigger* per attivare questo flusso consiste nella conclusione di un acquisto. Dopo 10 giorni che la piattaforma registra un evento *purchase*, viene inviata un'email di ringraziamento dal sistema che presenta dei prodotti raccomandati, un invito a lasciare una recensione, dei banner dinamici relativi a promozioni e un link che riporta ad alcune ricette del *magazine*. Dopo 12 giorni, al contatto viene inviata un'ulteriore email con idee su come completare la *mise en place* della tavola o con altre ricette da provare con i prodotti della categoria cottura Tognana. All'interno delle email di questo flusso sono sempre presenti dei banner di promo dinamiche, che cambiano a seconda del periodo di riferimento.

8.3.3. Web personalization e Lead Generation

Un'altra funzionalità che rende gli strumenti di *marketing automation* sempre più popolari e attrattivi consiste nella possibilità di espandere le comunicazioni con i contatti, creando dei contenuti dinamici su misura sul sito web, che rispondano ai loro specifici interessi. I moduli di *Web personalization* e di *Lead Generation* presenti su SALESmanago, consentono di sviluppare nuove modalità di comunicazione, attraverso la visualizzazione di pop-up dinamici, banner e frame di raccomandazione personalizzati, che rendono unica l'esperienza di acquisto di ogni utente e permettono allo stesso tempo di acquisire informazioni su di essi.

Social proof widget. Vengono utilizzati principalmente per mostrare ai nuovi clienti l'interessamento degli utenti ai prodotti del catalogo e creare urgenza all'acquisto sul sito web attraverso la creazione di un *widget* personalizzato di testo e immagine. Possono essere utilizzati anche per mettere in evidenza alcuni prodotti che potrebbero interessare al cliente ma a cui spesso non si rivolge attenzione. Sullo *shop* di Tognana, ad esempio, compare un *widget* che invita ad aggiungere a carrello una pratica *tote bag* con il logo del brand, che può essere acquistata per un costo molto basso [74].

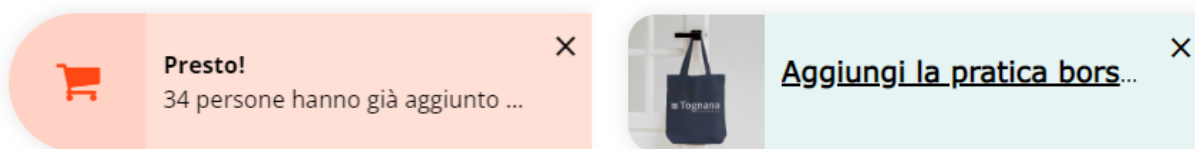


Figura 21: Esempi di social proof widget (Fonte:SALESmanago)

Pop-up dinamici. I pop-up dinamici rappresentano un modo efficace per generare *lead* all'interno del sito web. La piattaforma permette di selezionare le immagini appropriate e definirne le dimensioni,

per non disturbare la navigazione dell'utente, di definire dopo quanto tempo verrà visualizzato il messaggio, che tipo di messaggio far recapitare all'utente e le modalità di visualizzazione. Può essere impostato diversamente a seconda del dispositivo, permette di scegliere la posizione dello schermo in cui potrà essere visualizzato ed una serie di impostazioni per la grafica e per il tracciamento dei dati.

Ciò che è fondamentale è che i pop-up dinamici permettono di scegliere a che utenti mostrare il messaggio. È possibile, ad esempio, farlo apparire ai visitatori che sono iscritti alla newsletter o che non lo sono, ai contatti che sono identificati attraverso un particolare *tag*, su che pagine visualizzarlo o ogni quante visite. Sul sito di Tognana vengono impostati diversi pop-up a seconda dei periodi e delle promozioni in corso. In figura 22, l'esempio di un pop-up che si attiva per gli utenti non iscritti alla newsletter e viene visualizzato pochi secondi dopo che il contatto è entrato nella pagina categoria cottura.



Figura 22: Esempio di pop-up dinamico (Fonte:SALESmanago)

Web push. Sono delle notifiche che vengono visualizzate sul dispositivo degli utenti e costituiscono una modalità conveniente e veloce per comunicare con i clienti. Sono composte da un'icona quadrata, un titolo ed un breve testo, contenenti un link che rimanda ad una specifica pagina del sito web. La particolarità delle *web push* è che vengono visualizzate nel browser anche quando l'utente non si trova nel sito web di riferimento. Non sono notifiche che necessitano di dati personali degli utenti ma solamente dell'accettazione per la ricezione attraverso un modulo di *agreement*. Se l'utente accetta di ricevere notifiche, possono essere utilizzate per indirizzare comunicazioni significative, per rimanere in contatto anche con gli utenti non iscritti alla newsletter e non monitorati e permettono così al momento dell'accettazione di convertire utenti anonimi in contatti identificati. Inoltre, sono

comunicazioni in *real time* e possono essere impostate anche come azioni in seguito a delle condizioni verificate attraverso le *automation rules* [75].

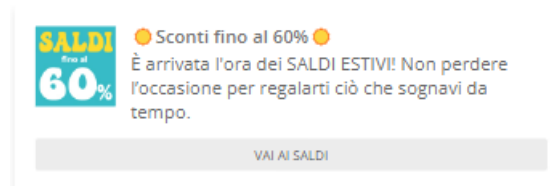


Figura 23: Esempio di notifica web push per i saldi (Fonte:SALESmanago)

Banner personalizzati. Grazie alla possibilità di personalizzare i banner presenti sul sito web o all'interno delle email diventa più semplice adattare i contenuti agli interessi degli utenti. È possibile modificare le comunicazioni attraverso lo strumento di segmentazione o in base alla fase del *sales funnel* in cui il visitatore si trova. I banner personalizzati possono anche essere adattati all'esperienza dell'utente in base ad alcune regole di automazione predeterminate [72].



Figura 24: Esempio di banner personalizzato (Fonte:SALESmanago)

9. Approccio data-driven e analisi del database clienti

In Tognana la cultura del dato in azienda, assume importanza fondamentale. Viene considerata, infatti, una dimensione necessaria per poter pianificare il futuro ed aumentare le performance aziendali nel tempo. Anche gli strumenti e le attività di *marketing automation* stesse, come già anticipato, si basano su una forte cultura del dato. Tutto ciò che viene automatizzato e personalizzato necessita alla base di informazioni analitiche sui comportamenti e le caratteristiche degli utenti, per massimizzare il ritorno sugli investimenti e sulle attività di *advertising*. Le piattaforme spesso

permettono anche l'integrazione con strumenti di analisi come *Google Analytics* per poter gestire all'interno di un unico luogo tutte le informazioni ricavate dalle interazioni degli utenti con il sito web e per trarre vantaggio dalle attività di *analytics* nella gestione delle relazioni con i clienti.

Di seguito vengono presentati due tipi di analisi dei dati raccolti attraverso le piattaforme di *marketing automation* e di e-commerce, che permettono di approfondire le caratteristiche degli utenti che acquistano sullo *shop* online IT di Tognana Porcellane e possono essere d'aiuto nello sviluppo di attività di gestione e/o fidelizzazione del cliente.

9.1. Analisi RFM

L'analisi RFM viene spesso utilizzata nel marketing per effettuare la segmentazione della clientela secondo i comportamenti d'acquisto in maniera immediata e semplice. Permette di individuare i segmenti di clienti più redditizi e fedeli e di programmare azioni di marketing mirate e personalizzate per aggiungere valore alla relazione con la clientela e spingere alla fidelizzazione. Spesso all'interno delle aziende la continua ricerca nell'acquisizione di nuovi clienti non lascia spazio al mantenimento dei clienti attuali, che rappresentano in realtà una grande fonte di valore per l'impresa. Secondo il principio paretiano, infatti, l'80% del fatturato di un'azienda viene generato dal 20% dei suoi clienti. È su questa idea che si basa l'analisi RFM che cerca di individuare i clienti migliori per mantenerli ed aumentare la *brand loyalty*. Il modello RFM segmenta la clientela sulla base di tre metriche principali:

- *Recency*: identifica quanto tempo è trascorso dall'ultima volta che un cliente ha effettuato un acquisto. Coloro che hanno effettuato acquisti di recente, secondo il modello RFM, sono maggiormente sensibili a promozioni e comunicazioni di marketing rispetto ai clienti che non effettuano un acquisto da molto tempo.
- *Frequency*: registra la frequenza di acquisto. I clienti con un numero di acquisti più elevato si presentano più ricettivi rispetto ai clienti occasionali.
- *Monetary*: calcola la spesa totale per cliente in un determinato periodo di tempo. Anche in questo caso, secondo il modello RFM, i clienti che spendono di più tendono ad essere più reattivi rispetto a coloro che spendono poco.

Individuate le soglie di RFM, il reparto marketing si occuperà di configurare e pianificare delle attività differenziate a seconda delle caratteristiche del segmento. Ad esempio, per i "clienti migliori" verranno implementate attività e campagne di fidelizzazione, mentre per coloro che rientrano nel gruppo di "clienti a rischio", verranno pianificate campagne di *re-engagement*. Questo tipo di analisi permette di individuare i clienti più qualificati da contattare, così da ottenere maggiori profitti e

ridurre i costi e gli sforzi di marketing, che può in questo modo focalizzarsi in maniera puntuale su specifici segmenti, offrendo ad ogni gruppo ciò che vuole ed ottimizzando così le performance delle proprie attività [11].

SALESmanago, all'interno del modulo *Segmentation*, permette di creare delle analisi RFM automatizzate sulla base di alcuni criteri stabiliti dall'utente. Nel caso di Tognana, la necessità era quella di segmentare i clienti *Fidelity*, ovvero coloro che hanno effettuato almeno un acquisto sullo *shop* online e che sono iscritti alla newsletter ed hanno quindi accettato di ricevere comunicazioni di marketing. L'arco di tempo considerato per l'analisi sono gli ultimi 12 mesi.

Di seguito i parametri utilizzati per la costruzione del modello RFM.

RECENCY	
<i>SHORT</i>	Il cliente ha effettuato l'ultimo acquisto meno di 90 giorni fa.
<i>MEDIUM</i>	Il cliente ha effettuato l'ultimo acquisto 90 o più giorni fa, ma meno di 290 giorni fa.
<i>LONG</i>	Il cliente ha effettuato l'ultimo acquisto 290 o più giorni fa.
FREQUENCY	
<i>CASUAL</i>	Il cliente ha effettuato meno di 2 acquisti negli ultimi 12 mesi.
<i>REGULAR</i>	Il cliente ha effettuato 2 o più acquisti, ma meno di 4 acquisti negli ultimi 12 mesi.
<i>COMMON</i>	Il cliente ha effettuato 4 o più acquisti negli ultimi 12 mesi.
MONETARY	
<i>SAVER</i>	Il cliente ha effettuato acquisti per meno di 100€ negli ultimi 12 mesi.
<i>MEDIUM</i>	Il cliente ha effettuato acquisti per un valore uguale o maggiore a 100€ ma minore di 266€.
<i>SPENDER</i>	Il cliente ha effettuato acquisti per un valore uguale o maggiore a 266€.

Tabella 1: Criteri per la costruzione del modello RFM (elaborazione personale)

Il vantaggio di eseguire questo tipo di analisi attraverso una piattaforma di *marketing automation* come SALESmanago, consiste nel fatto che il software, sulla base dei criteri definiti, segmenta automaticamente la base utenti e perfeziona giornalmente i risultati dell'analisi e le matrici combinate RFM, permettendo ai *marketers* ed alle aziende di aver sempre a disposizione delle analisi aggiornate. Viene presentata di seguito l'analisi RFM sui dati dei *Fidelity* aggiornati al 31 agosto 2022.

Recency

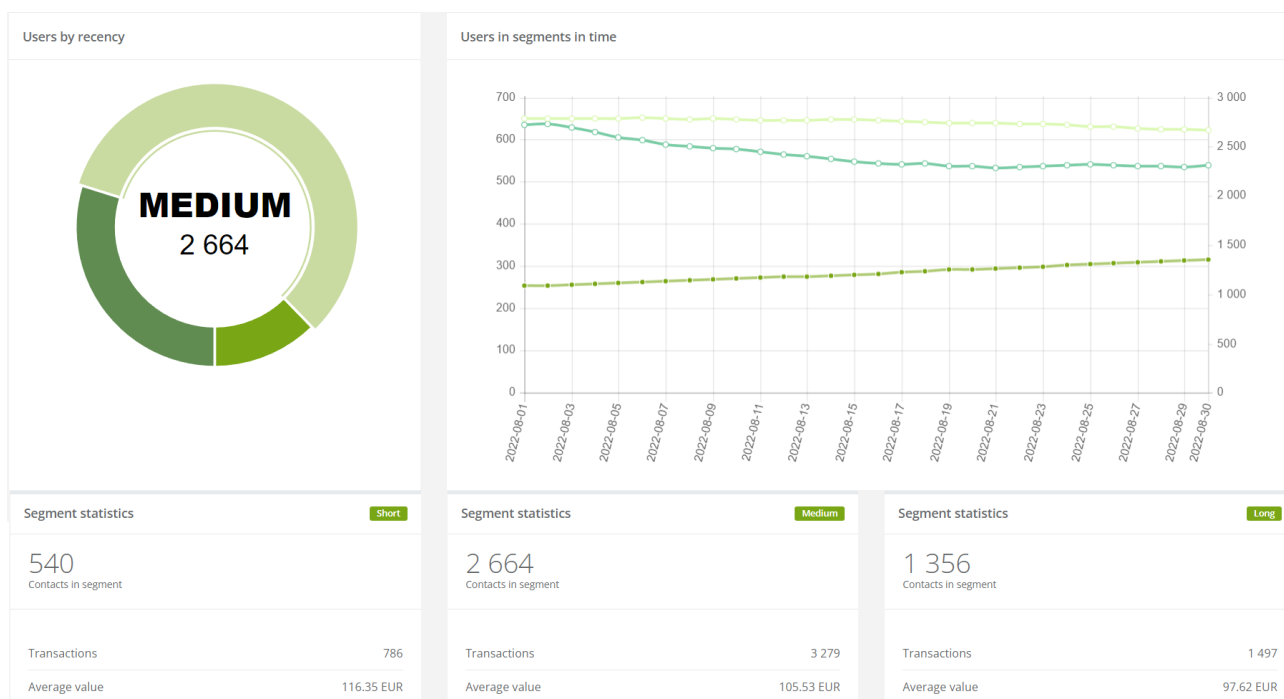


Figura 25: Analisi della recency (Fonte:SALESmanago)

La figura 25 mostra la *recency* del segmento di clienti selezionato attraverso una serie di grafici e statistiche. Nel primo riquadro in alto a sinistra viene riportato un diagramma a torta che mostra la composizione del segmento secondo i parametri scelti. Seguono un grafico a serie temporali relativo all'evoluzione della distribuzione degli elementi del gruppo nell'arco del mese di riferimento e delle statistiche numeriche più approfondite riguardanti i dati sulle transazioni, il valore medio di acquisto ed il fatturato totale per segmento (opportunamente ommesso in figura per mantenere la privacy sui dati sensibili dell'azienda).

Secondo l'analisi effettuata dal software SALESmanago, il 58% dei clienti è caratterizzato da un livello di *recency medium*. La maggior parte dei clienti del segmento ha quindi effettuato l'ultimo acquisto più di 90 giorni fa, ma meno di 290 giorni fa. Oltre a contenere il maggior numero di contatti,

questo gruppo conta anche con il più alto numero di transazioni totali. Ciò significa che la maggior parte degli acquisti è stato effettuato nei primi tre mesi del 2022. Seguono il segmento *long*, con 1356 clienti e 1496 transazioni ed il segmento *short* che contiene il numero più basso di clienti e di transazioni, rispettivamente 540 e 786, ma registra un valore medio per transazione più elevato. Da questa analisi si può notare anche come, al passaggio dal segmento *long* fino al segmento *short*, il valore medio per transazione sia aumentato, con gli acquisti più recenti che registrano il maggiore valore medio a carrello.

Frequency

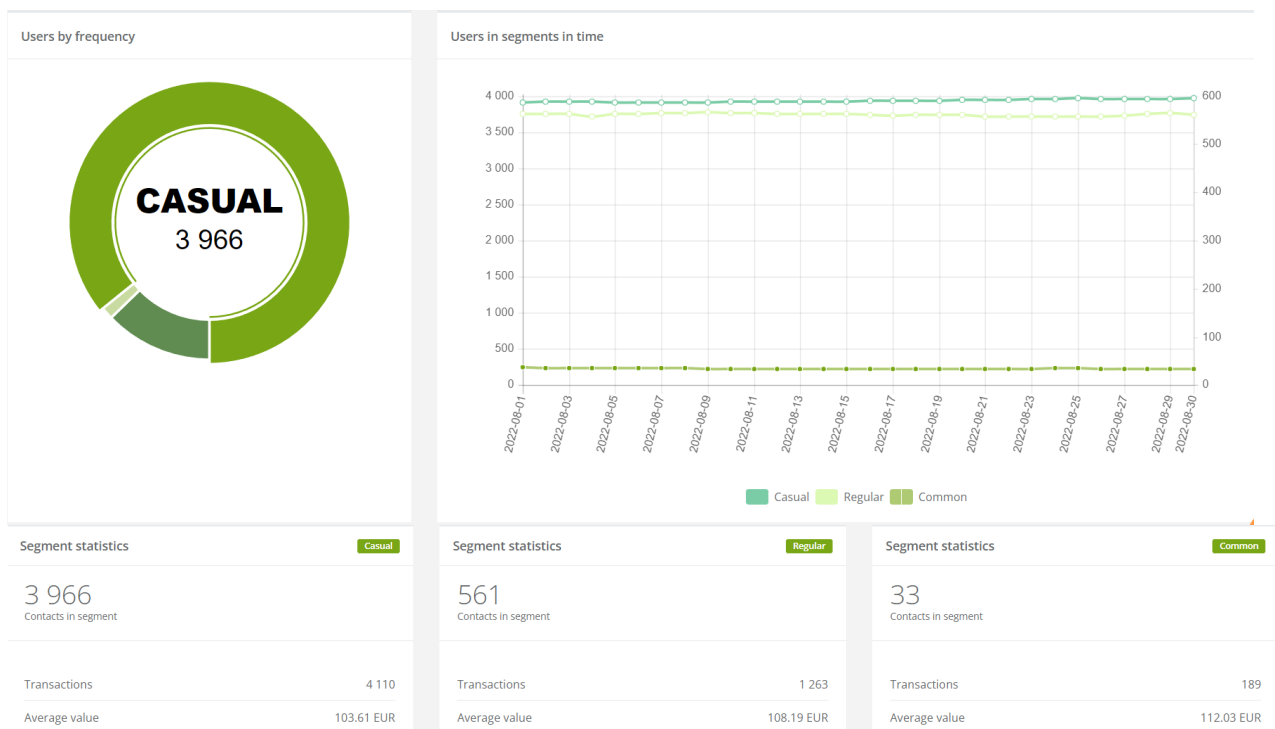


Figura 26: Analisi della Frequency (Fonte:SALESmanago)

Analizzando la frequenza di acquisto, emerge che più dell'85% dei *Fidelity* appartengono al segmento *casual*, ovvero che hanno effettuato meno di 2 acquisti nell'arco di 12 mesi. Segue il segmento *regular* con 561 clienti. I contatti *common*, che hanno completato il più elevato numero di acquisti, sono 33 e sono il segmento che registra un maggior valore medio per carrello. Dall'analisi della *frequency* si può ricavare che l'utente tipo di Tognana effettua principalmente acquisti unici sul sito web B2C. Questo tipo di segmentazione suggerisce che è necessario programmare una serie di attività di marketing che puntino non solo ad acquisire nuovi utenti, ma soprattutto a riportare gli utenti *casual*

a riacquistare, ad esempio attraverso l'implementazione di un programma fedeltà o dei premi/sconti per i successivi acquisti.

Monetary

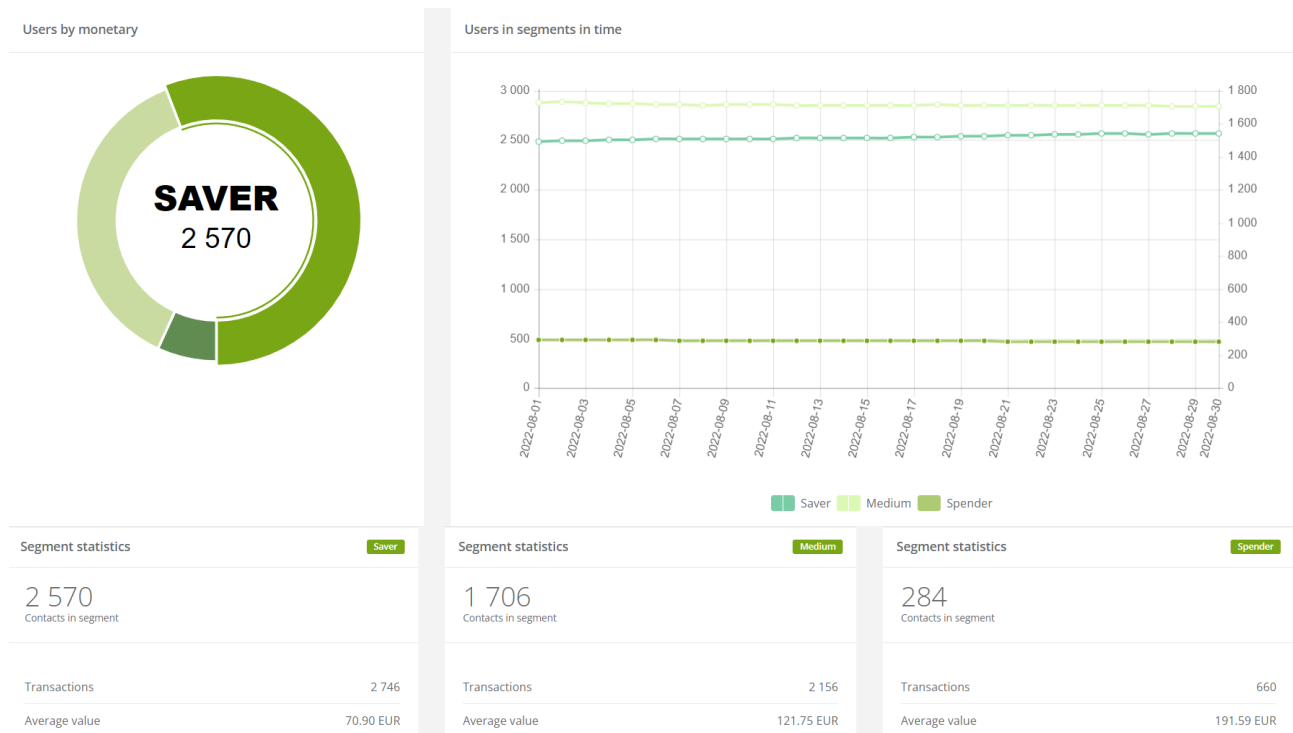


Figura 27: Analisi della Monetary (Fonte SALESmanago)

Nonostante la maggior parte del segmento clienti analizzato appartenga alla fascia *saver*, una buona parte di essi (37% circa) si presenta come clienti con spesa *medium*, compresa tra i 100€ ed i 266€. Anche il numero di transazioni per questo segmento si registra elevato (2156). Considerando che i prezzi medi dei prodotti venduti da Tognana non sono molto alti, il numero considerevole di clienti appartenenti al segmento *medium* rappresenta un buon segno per quanto riguarda il livello di carrello medio e fa presupporre anche un numero di prodotti acquistati per buona parte delle transazioni maggiore di uno.

Al fine di identificare i clienti migliori e quelli più a rischio di abbandono, SALESmanago fornisce all'utente delle matrici in cui vengono combinate due alla volta le tre metriche considerate nell'analisi RFM, come presentate di seguito.

Frequency-Monetary



Figura 28: Analisi incrociata frequency-monetary (Fonte:SALESmanago)

La combinazione delle metriche *frequency* e *monetary* mostra come una grossa fetta degli utenti considerati siano clienti che effettuano pochi acquisti e di importo inferiore a 100€ o di importo medio tra 100€ e 266€. Un buon 8% di essi si trova invece in una fascia di spesa *medium* e frequenza *regular*, con un numero di acquisti effettuati nell'ultimo anno compreso tra 2 e 4. Questo gruppo di utenti potrebbe essere interessante per l'azienda perché si presenta in una situazione favorevole per poter stimolare i contatti attraverso attività di *engagement* ed aumentare il numero di acquisti e/o il valore medio del carrello. A partire dai dati di questa matrice è stato creato anche un segmento *RFM / VIP* all'interno della piattaforma di *marketing automation*, che raggruppa tutti gli utenti che sono compresi nel riquadro verde scuro in alto a sinistra e rappresentano quindi i clienti più redditizi: hanno frequenza elevata (più di 4 acquisti negli ultimi 12 mesi) e livello di carrello sopra alla media (*monetary = spender*).

Monetary-Recency



Figura 29: Analisi incrociata monetary-recency (Fonte:SALESmanago)

Confrontando *recency* e *monetary*, emerge che i riquadri più popolati sono quelli con *recency* compresa tra il 90 ed i 290 giorni, e *monetary* minore o uguale a 266€. Un segmento interessante è rappresentato dal riquadro verde chiaro centrale, molto numeroso, che presenta sia *monetary* che *recency* media. Questi utenti, che hanno acquistato per l'ultima volta nel primo trimestre 2022 registrando un valore di carrello medio significativo, sono in una posizione intermedia per cui possono essere spinti, attraverso attività di comunicazione e marketing mirate, ad effettuare un nuovo acquisto e ad aumentare la *monetary*. Seguono gli utenti meno redditizi, che hanno effettuato l'ultimo acquisto, di valore minore di 100€, più di 290 giorni fa, per cui si potrebbero implementare delle attività di *re-engagement*.

Frequency-Recency



Figura 30: Analisi incrociata frequency-recency (Fonte:SALESmanago)

La maggior parte degli utenti raggruppati nella matrice ricadono nel segmento con *recency medium* e *frequency casual*. Ciò significa che il più elevato numero di clienti ha effettuato un unico acquisto durante l'ultimo anno, che risale tra 90 e 290 giorni fa. Questo gruppo di utenti, considerando che sono iscritti alla newsletter, potrebbero essere stimolati nell'aumentare entrambe le metriche attraverso promozioni nei mesi di settembre e ottobre e nei prossimi mesi invernali o con sconti dedicati ad un secondo acquisto.

9.2. Predictive analytics

Oltre agli strumenti utilizzati per la segmentazione dei contatti, SALESmanago offre anche dei moduli relativi alle analisi di predizione sugli acquisti e sul tasso di abbandono dei clienti. Sono presenti sia all'interno di una sezione dedicata nel menù (*Artificial Intelligence* → *Predictive Analytics*), che nelle singole schede di contatto alla finestra *Prediction Analytics*. La piattaforma permette, grazie all'utilizzo dei *big data*, di fare un passo in avanti nel mondo delle *analytics*, passando dall'analisi retrospettiva, all'analisi in *real time*, fino all'analisi predittiva per singolo contatto o per l'intero database. L'obiettivo è quello di prevedere le attività future del cliente sulla base dei dati transazionali e comportamentali considerando le azioni compiute nel presente e nel passato. Questo tipo di analisi può indirizzare coloro che si occupano di marketing nell'adattamento delle attività di comunicazione e gestione del contatto al fine di ridurre il rischio di perdita del cliente ed aumentare le performance dell'azienda [73].

Si prenda in considerazione la finestra di *Prediction Analytics* presente in ogni scheda contatto. Essa è costituita da 3 sezioni principali, ognuna contenente un tipo di analisi differente.

1. Prediction value of purchase occurrence

Prediction value of purchase occurrence

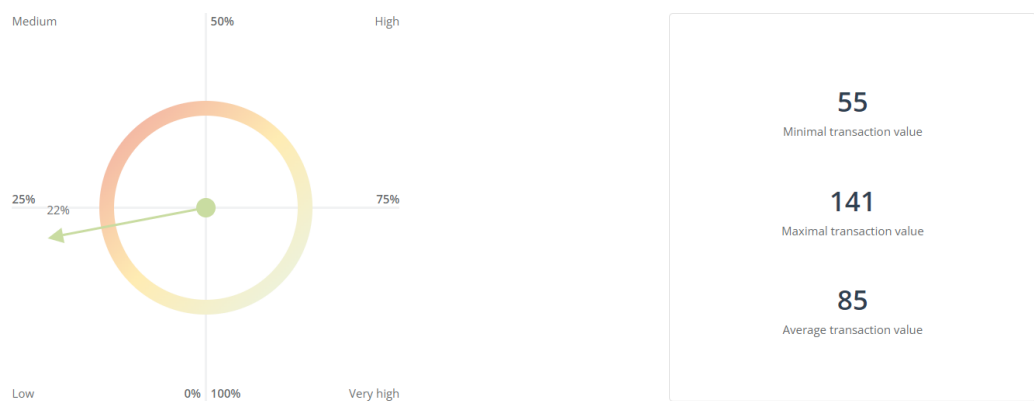


Figura 31: Analisi di predizione di acquisti futuri (Fonte:SALESmanago)

Questo modulo fornisce una stima sulla probabilità che un determinato contatto effettui nuovamente un acquisto. La previsione viene calcolata sulla base dei dati transazionali del contatto e delle visite effettuate sul sito web. Il grafico si compone di 4 quadranti, che indicano da 0 a 100 il valore della probabilità di riacquisto del singolo cliente. Nella parte più a destra della sezione vengono riportati i

dati che fanno riferimento al valore minimo tra le transazioni effettuate, al valore massimo ed al valore medio, che vengono anche utilizzati per effettuare il calcolo della probabilità. Nell'esempio presentato in figura, la probabilità di riacquisto misurata per il cliente è del 22%, che consente alla freccia verde di collocarsi all'interno del grafico solamente nel quadrante *Low* in basso a sinistra.

2. Churn prediction value and CLV analysis

Churn prediction value and CLV analysis

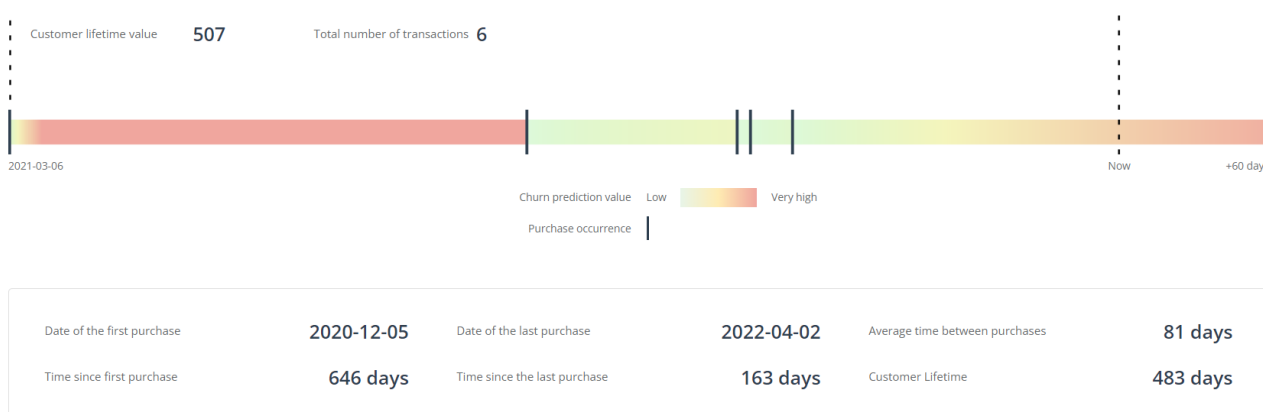


Figura 32: Analisi della probabilità di abbandono e CLV (Fonte:SALESmanago)

Nella linea temporale raffigurata, vengono rappresentati la previsione del tasso di abbandono del contatto e gli acquisti effettuati in uno specifico arco di tempo (l'evento *purchase* viene identificato nella linea del tempo con una barretta blu verticale). In particolare, questa sezione mostra i dati storici delle azioni di acquisto del cliente e lo sviluppo della probabilità di abbandono nel tempo, la probabilità di *churn* attuale e quella prevista per i prossimi 60 giorni. Questo tipo di analisi di abbandono permette di valutare il rischio che il cliente non torni più sullo *shop* per non aver effettuato altre transazioni. Inoltre, la sezione presenta una serie di informazioni numeriche sul contatto calcolate in base alle operazioni di acquisto effettuate dal cliente. Le metriche misurate sono:

- Il *Customer Lifetime Value*, calcolato semplicemente come il valore totale degli acquisti effettuati da un contatto.
- Il numero totale delle transazioni.
- La data della prima e dell'ultima transazione.
- Il numero di giorni passati dal momento della prima e dell'ultima transazione fino ad ora.
- Il tempo medio tra le transazioni, che permette di fare una previsione su quando il contatto potrebbe effettuare un prossimo acquisto.

- Il *Customer Lifetime*, calcolato come il numero di giorni trascorsi tra la prima e l'ultima transazione.

Nel caso dell'analisi del contatto in questione, dalla figura 32 è possibile notare come la barra di previsione di abbandono si colori di diverse sfumature in base alla lontananza o vicinanza all'avvenimento di un evento *purchase*. Inoltre questa sezione permette di ricostruire la storia dello stesso cliente all'interno del nostro sito. In questo caso, ad esempio, si sta analizzando un contatto che è attivo da più di un anno ed ha effettuato un numero elevato di acquisti, con l'ultimo risalente ad aprile 2022, per un totale di 507€.

3. Prediction value of purchase occurrence from source

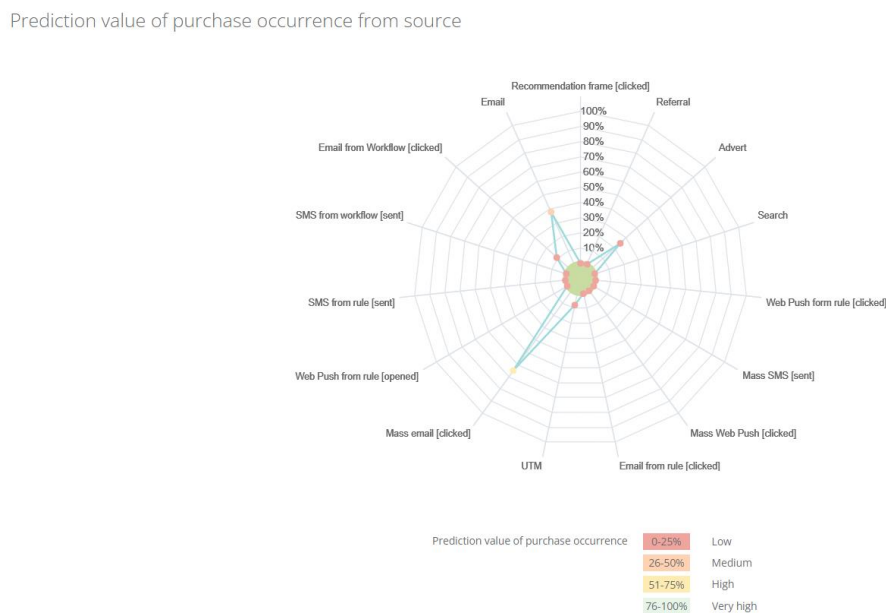


Figura 33: Analisi di predizione di acquisti futuri in base alla fonte (Fonte:SALESmanago)

Questa sezione permette di fare un focus sulla provenienza degli acquisti per fonte. Presenta un grafico a ragnatela con rappresentate tutte le fonti utilizzate sul sito web per attrarre traffico e clic e da cui può provenire un acquisto. Per ogni fonte specifica la probabilità di attribuzione diretta dell'acquisto. Attraverso questo tipo di analisi è possibile valutare quale canale di comunicazione è stato il più redditizio in termini di transazioni concluse e di conseguenza qual è quello che presenta la probabilità di acquisto più alta ed il miglior modo per comunicare con il cliente. Nel caso in figura si nota subito come il canale che più ha portato il cliente all'acquisto siano le newsletter (*mass email*),

con probabilità di acquisto alta, seguite da altre tipologie di email, con probabilità di acquisto media e dagli *advert* [73].

Nel modulo *Artificial Intelligence* → *Predictive Analytics*, sono contenute sostanzialmente le stesse sezioni ma in forma più sintetica poiché l'analisi predittiva viene effettuata sull'intero database contatti e restituisce dei risultati da cui è possibile ricavare delle informazioni in linea generale sul portafoglio clienti completo.

9.3. Cluster analysis

Un altro tipo di analisi relativo alla segmentazione del database clienti sulla base dei loro comportamenti di acquisto può essere effettuata attraverso il *clustering*.

Le tecniche di *data mining* ci permettono di segmentare ed effettuare delle analisi più approfondite sulla clientela. La *cluster analysis*, in particolare, può essere utilizzata in diverse applicazioni: come strumento *stand-alone* per esaminare la distribuzione dei dati all'interno di un dataset, come step di *pre-processing* per l'applicazione di ulteriori algoritmi come gli alberi decisionali, o per scoprire pattern di comportamento degli utenti sul web [27]. Inoltre, la clusterizzazione può essere utile nel business per facilitare il *Customer Relationship Management*; creare piccoli gruppi omogenei di clienti dalle caratteristiche simili permette di differenziare messaggi e offerte, massimizzando il ROI delle campagne di comunicazione, incrementando la *loyalty* e aumentando di conseguenza il fatturato generato. Attraverso l'utilizzo del software KNIME è possibile suddividere il database in gruppi contenenti clienti simili nel comportamento di acquisto e così individuarli e gestirli in maniera differente e mirata. Dopo aver presentato una *overview* sul funzionamento e le caratteristiche della piattaforma KNIME, si procederà con l'applicazione della *cluster analysis* sul database clienti di Tognana Porcellane.

9.3.1. Il software KNIME

KNIME è una piattaforma *open source* di analisi dei dati, reportistica e integrazione [78]. Nasce nel 2006 in Svizzera, nei laboratori dell'Università di Costanza, da alcuni sviluppatori che avevano l'obiettivo di creare una piattaforma facile da utilizzare, con interfaccia grafica accattivante, gratuita, espandibile e modulare. Grazie alle sue caratteristiche KNIME è stato nominato da Gartner come una tra le cinque piattaforme analitiche leader in ambito di *machine learning* e *data science*.

Tra le caratteristiche della piattaforma, le modalità di *Data Input* e *ETL (Extract, Load, Transform)* sono un punto di forza di KNIME. Fanno riferimento alle attività di import dei dati ed alla loro trasformazione ed esportazione che nella piattaforma possono essere implementate in maniera

immediata e facile. A differenza di quanto avviene con R, Python o Excel, in cui queste fasi richiedono più passaggi manuali, in KNIME è possibile effettuarle semplicemente scegliendo una serie di nodi e collegandoli in maniera adeguata tra di loro. Una volta elaborati i dati in entrata, KNIME permette di utilizzare oltre 200 algoritmi di *machine learning* che è possibile configurare con semplicità anche da utenti non esperti. Gli algoritmi esistenti permettono di effettuare delle analisi che possano soddisfare vari settori di business. Se non fossero abbastanza, la piattaforma permette anche l'integrazione con R e Python, così da poter sfruttare le loro funzionalità più avanzate.

Altro punto di forza di KNIME, che permette il suo utilizzo anche ad utenti non esperti di *data science*, è la *user friendliness* della sua interfaccia. La piattaforma si presenta infatti intuitiva e semplice da navigare, nonché attraente anche da un punto di vista visivo. Essa permette di far vedere ciò che fa, consentendo di scegliere l'ordine e quali attività analitiche effettuare attraverso la creazione di un flusso di lavoro contenente le trasformazioni e gli algoritmi che si vogliono applicare. Inoltre, KNIME permette di lavorare efficacemente in team. Grazie alla sua implementazione grafica è possibile commentare e spiegare i flussi di lavoro direttamente nell'interfaccia principale, consentendo così ad altri di utilizzarli in maniera immediata. È possibile anche modularizzare, esportare e commentare i dati con semplicità. Un punto a sfavore di KNIME è rappresentato dalle modalità di *data visualization*. Alcune funzionalità per rappresentare i dati (come si vedrà di seguito) sono presenti su questa piattaforma, ma per rendere i risultati il più chiari possibile è consigliabile farsi supportare da altri software di *data visualization* più specifici. Per chi utilizza KNIME, è stata creata anche una *community* principalmente popolata da *data scientists* e programmatori che mantengono attivo un forum su cui si può fare riferimento per risolvere i problemi nell'utilizzo della piattaforma o per chiedere aiuto ai più esperti sulle analisi da effettuare.

Anche se non rappresenta il miglior strumento per la *data analytics* esistente, KNIME può essere molto utile per gestire in maniera efficace ed in autonomia gran parte delle analisi che necessitano l'utilizzo del *machine learning*. Inoltre, il fatto che possa essere integrato con funzionalità aggiuntive più avanzate, gli permettono di essere uno strumento utile non solo per iniziare a fare *analytics*, ma anche per effettuare delle analisi più avanzate e complete. Rappresenta uno strumento da utilizzare come base e da arricchire attraverso l'utilizzo di altri *tool* complementari [26].

9.3.2. Lavorare in KNIME

Per iniziare a lavorare in KNIME è necessario conoscere gli elementi attraverso cui vengono effettuate le analisi dei dati: i nodi ed i *workflow*.

9.3.2.1. Nodi

I nodi sono le unità computazionali di base utilizzate su KNIME. Ogni nodo all'interno della piattaforma rappresenta un processo analitico che, unito agli altri, costituisce l'intero flusso di *analytics*. La finestra *repository* contiene nodi utili a leggere i dati dei file (siano essi in formato .xls, .csv, e così via), nodi per trasformare tabelle aggiungendo, eliminando, aggregando righe e colonne, nodi per elaborare i dati ed applicare gli algoritmi di *machine learning*. Ciò che li caratterizza è il fatto che ogni nodo svolge uno specifico compito ed ha caratteristiche uniche, che li differenzia l'uno dall'altro. A livello grafico, vengono rappresentati invece in maniera uniforme e sono costituiti da elementi standard che permettono di impostarli facilmente.

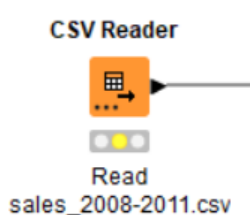


Figura 34: Esempio di nodo su KNIME (Fonte:KNIME)

Tutti i tipi di nodi si presentano come dei piccoli quadrati con un semaforo sottostante e una serie di triangoli e/o quadratini in entrata e uscita chiamati connettori, che servono a collegare i nodi tra loro, a seconda della tipologia e delle attività analitiche da svolgere. In figura, ad esempio, è rappresentato un nodo di input che viene utilizzato per caricare dati dall'esterno. Considerando che svolge un'attività che segna l'inizio di un flusso di lavoro, esso comprende solo un connettore in output (il triangolo nero), mentre in altri nodi è possibile

avere connettori sia in entrata che in uscita. Esistono anche nodi che presentano solo connettori in entrata, come quelli che salvano file sul computer o visualizzano i risultati dell'analisi. In base alla loro forma e colore, i connettori indicano il tipo di dato che viene trasferito: se sono triangoli si occupano di trasferire tabelle contenenti numeri o stringhe di testo, se sono quadrati si occupano di rappresentare modelli, immagini, documenti o altri tipi di dati. La differenza di forma e colore deve essere rispettata anche nelle connessioni tra nodi, per cui triangoli neri in uscita devono essere collegati a nodi che presentano triangoli neri in entrata e, allo stesso modo, a quadrati blu devono corrispondere connettori della stessa forma e colore. A livello grafico, i nodi si collegano tra di loro in un ordine determinato attraverso delle linee che partono dai connettori situati nel bordo destro o sinistro del quadrato, a seconda che la connessione avvenga in entrata o in uscita al nodo stesso. Ogni nodo trova pieno compimento solo se adeguatamente collegato ad una serie di altri nodi appartenenti ad un *workflow* che si sviluppa da sinistra verso destra. I nodi più a destra verranno eseguiti solo nel momento in cui sono stati implementati ed eseguiti adeguatamente i precedenti collocati a sinistra. Lo stato in cui il nodo si trova viene indicato da un semaforo, attraverso tre puntini che si colorano di rosso, giallo o verde in base alla configurazione ed alla buona riuscita o meno dell'operazione. A seconda del colore mostrato possiamo distinguere diverse situazioni:

- *Semaforo rosso*: il nodo non è stato ancora adeguatamente configurato o connesso correttamente. In questa situazione il nodo non può essere eseguito ed è necessario fare doppio clic per configurare le impostazioni correttamente.
- *Semaforo giallo*: indica la corretta configurazione e dà il permesso di effettuare l'azione di *Execute* che mette il nodo in azione. Una volta eseguito, apparirà al posto del semaforo giallo una barra che mostrerà lo stato di avanzamento di elaborazione dei dati da parte del nodo.
- *Semaforo verde*: nel momento in cui il semaforo diventa verde significa che il nodo ha terminato l'elaborazione dei dati ed è stato quindi correttamente eseguito. È possibile a questo punto anche visualizzare gli output del lavoro, a seconda della tipologia di analisi effettuata.
- *Triangolo giallo*: se qualcosa non ha funzionato o potrebbe mettere a rischio il corretto funzionamento del nodo, appare nel mezzo del semaforo un'icona raffigurante un triangolo giallo con un punto esclamativo al suo interno. Rappresenta una sorta di *warning* che spinge l'utente a rivalutare le impostazioni di configurazione o i dati di input.
- *X rossa*: se l'esecuzione del nodo è stata interrotta perché qualcosa non ha funzionato, viene visualizzata una "x" rossa all'interno del semaforo. Sia nel caso del triangolo giallo, che nel caso della "x" rossa, KNIME ci permette di sapere di più riguardo l'errore o il *warning* mantenendo il puntatore del mouse sull'icona in modo da far apparire una finestra o un commento che spieghi il motivo della notifica [26].

9.3.2.2. Workflow

I nodi utilizzati singolarmente non permettono all'utente di effettuare delle analisi complete, ma rappresentano solamente gli step di un *workflow*, ovvero di un flusso analitico costituito dall'insieme dei nodi collegati tra loro. KNIME permette di combinare e collegare i nodi ordinandoli e configurandoli in maniera adeguata, così da ottenere un *workflow* in grado di eseguire un processo di analisi o trasformazione dei dati completo. All'interno della piattaforma è possibile utilizzare e modificare flussi di esempio già impostati, o crearne di completamente nuovi partendo da un foglio vuoto su cui trascinare i nodi a disposizione. Questa modalità di creazione dei flussi di lavoro permette all'utente di mantenere sempre il controllo sugli step di analisi e di poter seguire ogni passo dell'esecuzione del *workflow*, con la possibilità di valutare anche i risultati parziali in ogni nodo del flusso [26].

9.3.2.3. Interfaccia utente

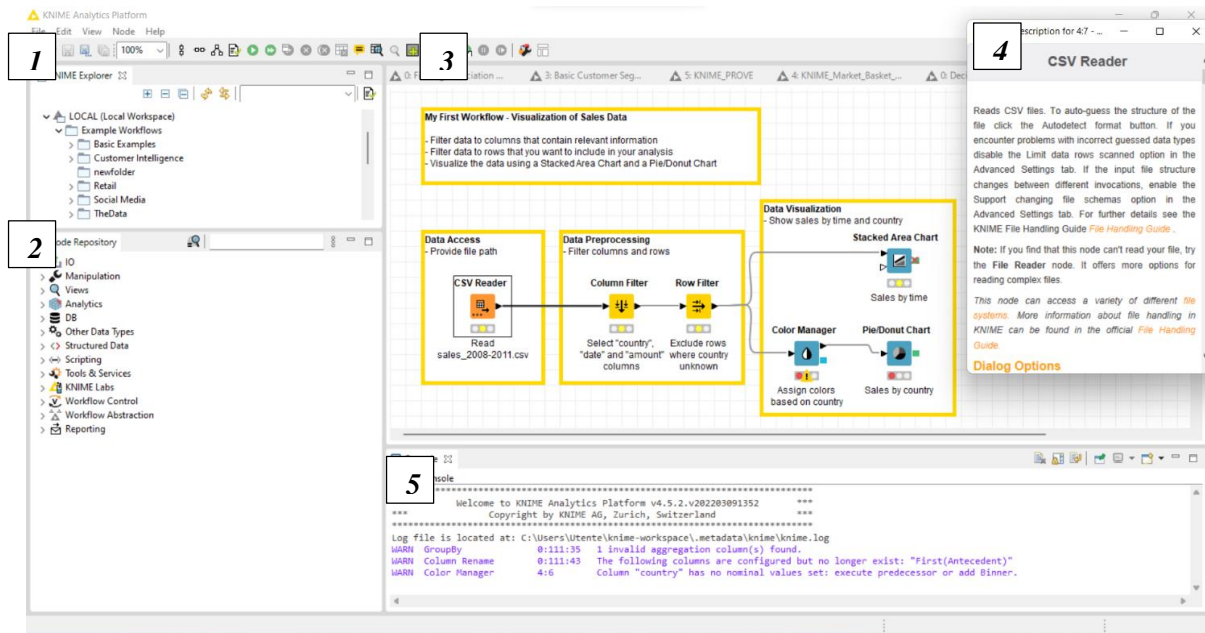


Figura 35: Interfaccia utente su KNIME (Fonte:KNIME)

L'interfaccia utente di KNIME è immediata e di facile utilizzo. Una volta effettuato il download ed installato il programma, esso si apre in una schermata che contiene 5 finestre principali come da figura 35.

1. *Explorer*: è il luogo in cui vengono archiviati i *workflow* e mantenuti in ordine. È organizzata attraverso delle cartelle e presenta anche dei flussi di lavoro di esempio già impostati e che è possibile utilizzare e dei file di dati di default di dominio pubblico. Contiene anche una cartella *Local*, in cui sono presenti i *workflow* salvati sul computer. All'interno di questa sezione è possibile anche connettersi con dei server esterni per accedere a flussi di lavoro remoti.
2. *Node Repository*: è una finestra in cui sono raccolti tutti i nodi disponibili all'utente. Sono organizzati per tipologia e possono essere facilmente trovati attraverso una barra di ricerca situata in alto a destra attraverso la digitazione delle parole chiave.
3. *Workflow Editor*: è la sezione principale in cui vengono creati i flussi di lavoro. Permette di rilasciare i nodi trascinandoli dal *Node Repository* e collegarli tra loro attraverso delle frecce con pochi clic del mouse. Inoltre, in questa finestra è possibile inserire dei commenti per spiegare i passi compiuti nella creazione del flusso analitico, costruire dei box che raggruppano concettualmente delle parti di *workflow* e aggiungere del testo per descrivere il funzionamento dei singoli nodi.

4. *Node description*: è una finestra che si apre cliccando su qualsiasi nodo e permette di ottenere una spiegazione su cosa fa e come si utilizza il nodo stesso. Vengono date anche indicazioni su quali input necessita e che tipologia di output produce.
5. *Console*: registra tutto ciò che avviene nel *workflow* e permette di consultare le informazioni aggiornate ogni qual volta viene effettuata una modifica. Tutti gli errori che si incontrano nell'esecuzione dei nodi o *warning* che allertano su possibili intoppi nel flusso possono essere consultati in questa finestra e vengono allo stesso tempo salvati su un file log [26].

9.3.3. Cluster analysis: applicazione pratica

Prendendo spunto da un flusso di analisi cluster definito da Andrea De Mauro nel suo libro *Big data Analytics*, è stata implementata un'analisi sul database clienti di Tognana Porcellane, con l'obiettivo di identificare tra gli utenti che hanno effettuato delle transazioni sul sito dei comportamenti di acquisto simili in modo da poterli suddividere in gruppi distinti e omogenei.

9.3.3.1. Fase 1

La prima fase del lavoro consiste nella selezione, *sampling* e *pre-processing* dei dati. Sono stati analizzati due file, uno proveniente dall'e-commerce di Tognana denominato *Estrazione ordini Tognana.xls*, l'altro ricavato attraverso un export di dati dal portale *CDP* di SALESmanago di tutti gli eventi *purchase* registrati dalla piattaforma e rinominato *Purchase SM.xls*. Sono stati utilizzati due file provenienti da fonti differenti per poter identificare il maggior numero di attributi dalle tabelle di transazioni in possesso, in modo da ricavarne caratteristiche significative relativamente agli acquisti effettuati dagli utenti sullo *shop* online. Di seguito viene presentato il processo di selezione del campione di dati da utilizzare per il *clustering* con la corrispondente descrizione dei nodi utilizzati.

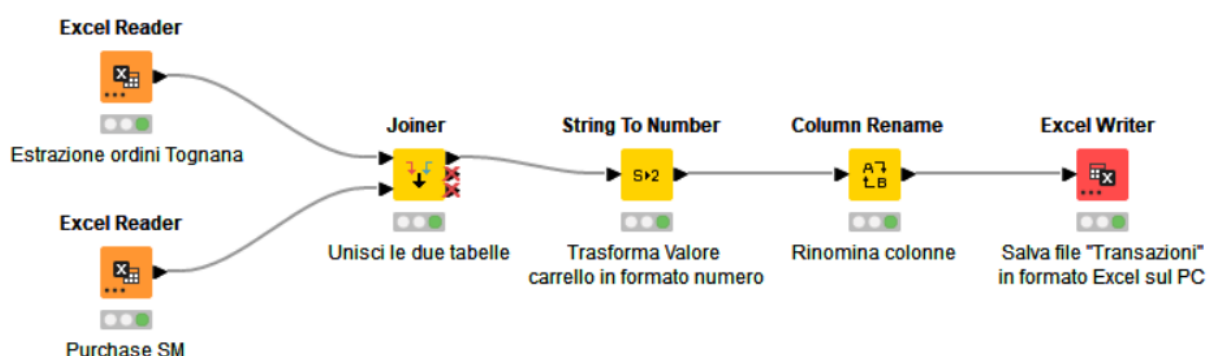


Figura 36: Workflow di elaborazione dei file di dati a disposizione (Fonte: elaborazione personale su KNIME)

Il primo nodo utilizzato per visualizzare i file di dati su KNIME è *Excel Reader*, che permette di navigare tra la raccolta del computer e selezionare i file *.xls* contenenti dati di tipo stringa, numero, di tipo booleano, o data e ora necessari per effettuare l'analisi dei dati. I fogli di calcolo caricati si presentano come segue:

Estrazione ordini Tognana

Il file è costituito da 9 colonne:

- *Ordine numero*: identifica il numero di ordine progressivo degli acquisti effettuati online.
- *Riferimento ordine cliente*: contiene dei codici cliente associati al numero di ordine.
- *Data ordine*: fa riferimento alla data in cui l'ordine è stato registrato dal software di e-commerce.
- *SKU*: numero identificativo degli articoli acquistati.
- *Articolo*: descrizione degli articoli acquistati.
- *Forma decoro*: fa riferimento allo stile/collezione di cui fanno parte gli articoli acquistati.
- *Valore netto*: identifica il valore totale al netto dell'IVA e di eventuali sconti dell'acquisto di ogni singolo articolo.
- *Quantità*: numero di pezzi acquistati per prodotto.
- *Prezzo unitario*: prezzo unitario per articolo al netto del calcolo dell'IVA.

Le righe della tabella si riferiscono ad ogni articolo acquistato, di conseguenza un singolo ordine potrebbe risultare suddiviso su più righe. La struttura del file si presenta quindi con una chiave primaria composta da *ordine numero*, *riferimento ordine cliente* e *SKU*.

Purchase SM

Il file contenente i dati ricavati da SALESmanago è invece costituito da 6 colonne:

- *Email*: contiene gli indirizzi email attraverso cui i contatti vengono identificati all'interno di SALESmanago. Viene utilizzato come codice identificativo cliente.
- *Data ordine*: data e ora in cui il cliente completa la transazione per l'acquisto.
- *SKU*: lista dei codici identificativi di tutti gli articoli acquistati.
- *Valore carrello*: valore del carrello totale compreso di IVA e ridotto da eventuali sconti per cui il cliente ha effettuato la transazione.
- *Detail*: raccoglie in una lista la descrizione degli articoli acquistati.
- *ExternalId*: comprende un identificativo dell'ordine cliente che corrisponde alla colonna *Riferimento ordine cliente* del primo file.

A differenza della precedente, all'interno di questa tabella gli SKU degli articoli acquistati sono raggruppati attraverso una lista all'interno di ogni riga, che registra i singoli ordini effettuati dagli utenti, identificati attraverso la loro email.

Dopo aver caricato il file di dati all'interno della piattaforma, per combinare le due tabelle e creare un nuovo documento contenente maggiori informazioni sui comportamenti di acquisto dei clienti sono stati utilizzati i nodi *Joiner*, *String to Number*, *Column Rename*, *Excel Writer*. Si vedano in dettaglio di seguito.

Joiner. Questo nodo permette di combinare le righe delle due tabelle date in input che presentino gli stessi valori per le colonne selezionate. L'operazione viene effettuata collegando le due tabelle caricate con *Excel Reader*, attraverso le due porte a forma di triangolo presenti nel lato sinistro del nodo *Joiner*. Ogni riga della tabella data in input dalla porta superiore viene combinata con ogni riga data in input dalla porta inferiore che contenga valori identici nelle colonne selezionate.³

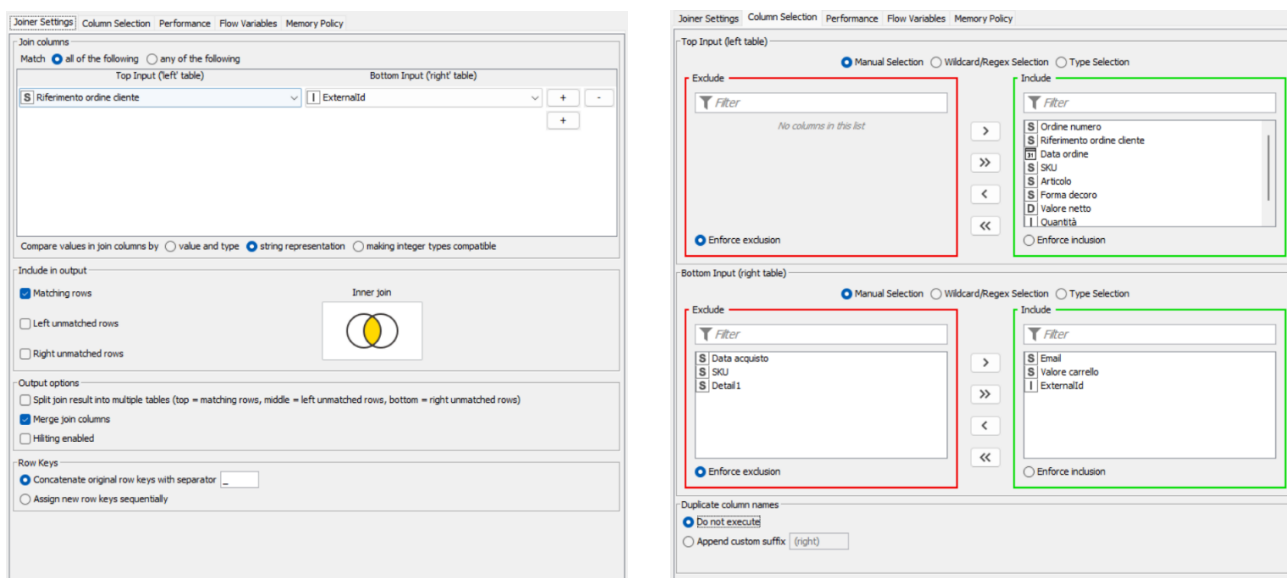


Figura 37: Configurazione nodo Joiner (Fonte:KNIME)

La figura 37 mostra i *settings* definiti durante la configurazione del nodo. Inizialmente viene richiesto di scegliere una colonna da entrambe le tabelle da utilizzare per l'unione. In questo caso sono state selezionate le colonne *Riferimento ordine cliente* e *Externalid*, che sono gli unici attributi che contengono valori coincidenti in entrambe le tabelle. Si è voluto in questo modo creare un database che comprenda, oltre alle informazioni relative agli acquisti sull'e-commerce, anche il campo *email* che, fungendo da codice identificativo, permette di individuare in maniera univoca i clienti e quali

³ KNIME – Node description – Joiner

ordini appartengono allo stesso cliente. È necessario, in seguito, definire quali righe si vogliono includere nell'output. In questo caso è stata selezionata l'opzione *Inner Join* per cui si chiede al nodo di includere nella tabella di output solo le righe la cui chiave è presente in entrambe le tabelle. Selezionando *Left unmatched rows* o *Right unmatched rows*, è possibile mantenere anche le righe di una delle due tabelle (o di entrambe nel caso di *Full outer join*) contenenti valori che non coincidono con quelle dell'altra. Come opzioni di output è stato scelto *Merge join columns* che permette di unire i valori creando un'unica colonna denominata *Riferimento ordine cliente=ExternalId*. Nella finestra *Column Selection*, viene inoltre richiesto di selezionare le colonne di entrambe le tabelle che si vuole mantenere nell'output. In questo caso si è deciso di mantenere tutte le colonne, escludendo quelle presenti in entrambe le tabelle.

String to Number. Ad ogni colonna presente nelle tabelle caricate, a seconda di come vengono trattati i dati nel file originale, la piattaforma KNIME assegna un formato di default tra cui: stringa, numero, formato data e ora, ecc. Il tipo di formato assegnato incide sulle operazioni che possono essere effettuate sulle colonne. Se, ad esempio, la colonna si presenta in formato stringa, non potrà essere selezionata per effettuare operazioni matematiche come può avvenire nel caso di una colonna numerica. Non sempre però il formato assegnato dalla piattaforma corrisponde a quello effettivo dei dati. In questo caso alla colonna *Valore carrello* è stato assegnato il formato stringa di default, nonostante si tratti di valori numerici. Il nodo *String to Number* ha permesso di convertire i suddetti valori stringa dell'attributo in numeri per cui è possibile effettuare operazioni matematiche.

Column Rename. Una volta eseguite una serie di operazioni sui dati e sulla combinazione delle colonne, le etichette nominative delle stesse subiscono delle modifiche. Attraverso il nodo *Column Rename* è possibile correggere i nomi delle colonne ed il formato. Da questa operazione risulta una tabella finale strutturata come segue.

Ordine numero | Riferimento Ordine Cliente | Data ordine | SKU | Articolo | Forma decoro | Valore netto | Quantità | Prezzo Unitario | Email | Valore carrello totale

Excel Writer. Attraverso questo nodo è possibile esportare i dati della nuova tabella creata nel computer con la creazione di un foglio di calcolo excel che può essere nuovo o esistente. Dopo avere elaborato i file di dati a disposizione, l'output finale risulta in una nuova tabella che viene esportata e rinominata *Transazioni*, contenente dati più dettagliati sugli acquisti effettuati sull'e-commerce dal 23/11/2020 al 30/08/2022.

9.3.3.2. Fase 2

Dopo aver selezionato e campionato i dati necessari per l'analisi, si è passati ad una seconda fase di lavoro, che costituisce il cuore del processo di *clustering*. Tutti gli step del *workflow* sono rappresentati in figura 38.

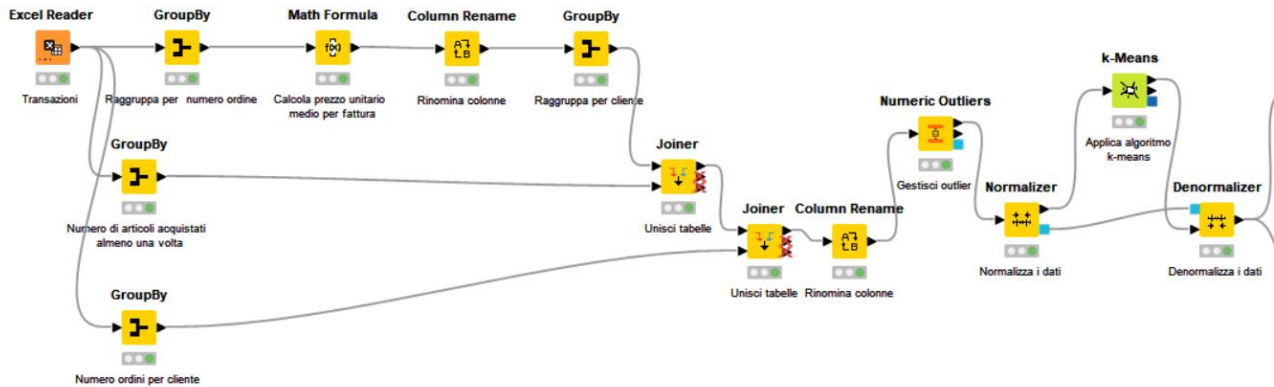


Figura 38: Workflow di *clustering* (Fonte: elaborazione personale su KNIME)

Come prima operazione, attraverso il nodo *Excel Reader* sono stati caricati nella piattaforma i dati della nuova tabella *Transazioni*. Si è proseguito con una serie di ulteriori step di *pre-processing* dei dati, prima di applicare i nodi per la *cluster analysis* effettiva, partendo dall'utilizzo del nodo *GroupBy*.

GroupBy. Questo nodo permette di raggruppare le righe di una tabella sulla base dei valori univoci rilevati nelle colonne del gruppo selezionato, fornendo come output una riga per ogni set di valori identificato. Le colonne che non sono selezionate, vengono poi aggregate in base alle impostazioni definite nella finestra *Manual aggregation* della schermata di configurazione del nodo. È possibile selezionare diverse modalità di aggregazione delle colonne a seconda del formato dei dati che contengono.⁴ Il nodo *GroupBy* è stato utilizzato in tre diverse modalità: per raggruppare le righe per numero di ordine, per contare il numero di articoli diversi acquistati almeno una volta dal cliente (e definirne quindi la varietà del carrello media) e per calcolare il numero di ordini totali effettuati da ogni contatto. In figura 39 vengono presentate le finestre di configurazione utilizzate per il raggruppamento del numero ordine.

⁴ KNIME – Node description - GroupBy

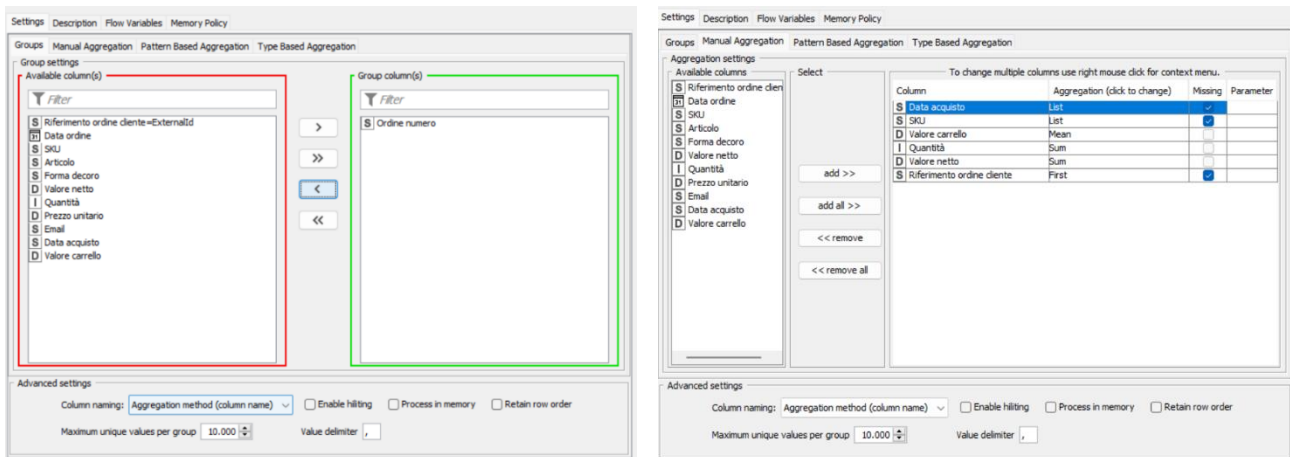


Figura 39: Configurazione nodo GroupBy (Fonte: KNIME)

Come è possibile notare dalla figura, per questo tipo di aggregazione è stata utilizzata l'opzione *List* per le colonne *Data di acquisto* e *SKU*, in modo da mantenere tutti i dati presenti nelle righe e creare una lista per ogni riga. Per *Quantità* e *Valore netto* è stato invece utilizzato il metodo di aggregazione *Sum*, per poter visualizzare la somma totale delle quantità e del carrello per ogni ordine. L'attributo *Valore carrello* è stato raggruppato attraverso la media dei valori, mentre per *Riferimento ordine cliente*, considerando che la colonna comprende valori che si ripetono, è stata effettuata un'aggregazione che mantenesse il primo valore riconosciuto per gruppo.

Math Formula. Una volta effettuato il raggruppamento per numero di ordine, è stato calcolato il *Prezzo Unitario Medio* per fattura attraverso il nodo *Math Formula*. Nonostante fosse già presente una colonna relativa al prezzo nella tabella iniziale, non poteva essere aggregata per media perché ogni prezzo faceva riferimento ad articoli acquistati in quantità differente e ne sarebbe quindi risultato un prezzo medio non ponderato. Attraverso l'utilizzo del nodo *Math Formula* è stato invece possibile calcolare un prezzo ragionevole dividendo il *Valore netto* per la *Quantità* acquistata per ogni fattura. Dopo aver effettuato questa operazione, si è nuovamente passati per il nodo *Column Rename* per rinominare le etichette delle nuove colonne. In figura 40 viene presentata la configurazione del nodo *Math Formula*.

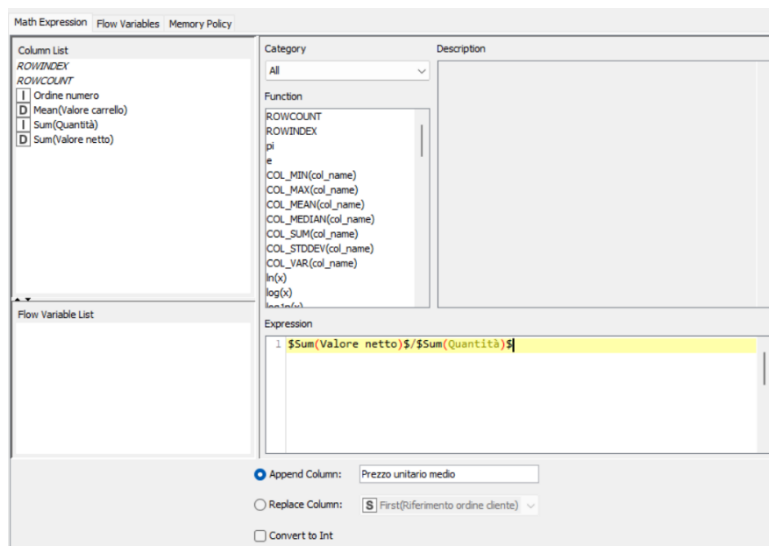


Figura 40: Configurazione del nodo Math Formula (Fonte: KNIME)

Il nodo *GroupBy* è stato utilizzato anche per calcolare il numero di articoli acquistati almeno una volta dai clienti. La tabella è stata raggruppata sulla base della colonna *Email* ed è stata aggregata la colonna *SKU* attraverso la modalità *Unique count*, che permette di contare ogni membro appartenente al gruppo a cui si fa riferimento solo una volta. Ciò ha permesso di creare una colonna nuova in cui venisse associato ad ogni cliente il numero di prodotti diversi che ha acquistato nell'arco di tempo di riferimento, che permette di dare un'idea sulla varietà del carrello media di ogni contatto. Lo stesso procedimento è stato effettuato per calcolare il numero di ordini totali per singolo cliente.

L'utilizzo del nodo *GroupBy* in questa modalità ha permesso di definire delle nuove colonne rispetto alla tabella iniziale che sono poi state nuovamente aggregate attraverso il nodo *Joiner* sulla base dell'attributo *Email*. Attraverso queste operazioni è stato possibile ricavare una serie di informazioni sugli acquisti effettuati dai clienti che possono essere utilizzate per creare dei *cluster*.

Si è proceduto nuovamente alla rinomina delle colonne attraverso il nodo *Column Rename* e l'output che se ne è ricavato è risultato in una tabella strutturata come segue.

Email | Dimensione media carrello | Valore carrello medio netto | Prezzo medio | Valore carrello medio totale | Articoli acquistati almeno una volta | Numero ordini per cliente

Le operazioni presentate finora hanno permesso di ricavare una serie di variabili descrittive che ci consentono di identificare alcune caratteristiche nel comportamento di acquisto degli utenti sullo *shop* online di Tognana Porcellane. Per procedere nell'analisi dei clienti più simili tra loro e creare dei *cluster* il più possibile omogenei, una volta ricavate questo tipo di informazioni attraverso semplici interventi nella tabella, è necessario implementare una serie di nodi di *data mining* che permettano di

elaborare i dati in maniera automatica e di fornirci risultati più complessi. In questa fase sono stati utilizzati quattro nodi principali, come rappresentato in figura 41.

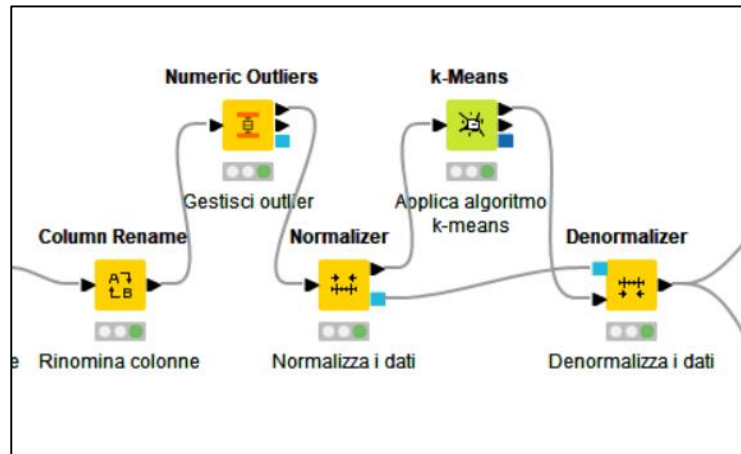


Figura 41: Implementazione dei nodi di *data mining* per l'analisi dei cluster (Fonte: elaborazione personale su KNIME)

Numeric Outliers. Lo scopo del *clustering* è quello di creare gruppi che contengano osservazioni il più simili tra loro e di fornire all'utente una suddivisione dei dati di buona qualità che consenta di ottimizzare le attività di business. Lavorando con i dati reali, però, ricavare analisi significative non è sempre possibile perché ci si deve spesso confrontare anche con dei valori anomali, chiamati *outliers*, che ostacolano la creazione di cluster coerenti e separabili [52]. Per poter superare questo problema, è necessario considerare l'esistenza di questi valori e gestirli in maniera adeguata prima di suddividere la popolazione di riferimento in gruppi. Il nodo *Numeric Outliers* permette di rilevare e trattare i valori anomali per tutte le colonne selezionate individualmente attraverso il calcolo dello scarto interquartile (IQR).

Per ogni colonna vengono calcolati il primo ed il terzo quartile (Q_1 , Q_3) e un'osservazione viene considerata anomala se si trova al di fuori dell'intervallo $R = [Q_1 - k(IQR), Q_3 + k(IQR)]$ con $IQR = Q_3 - Q_1$ e $k \geq 0$. Il valore di default è impostato in $k=1,5$; ciò significa che vengono contrassegnati come *outliers* tutti i valori minori del valore più piccolo in R e maggiori del valore più grande in R, ovvero dell'estremità inferiore e superiore dei baffi di un boxplot. Nel caso in analisi è stato configurato (dopo una serie di iterazioni e confronti sui risultati finali) un valore $k=3,0$, concentrandosi solamente sul taglio dei valori anomali più estremi.⁵

⁵ KNIME – Node description – Numeric Outliers

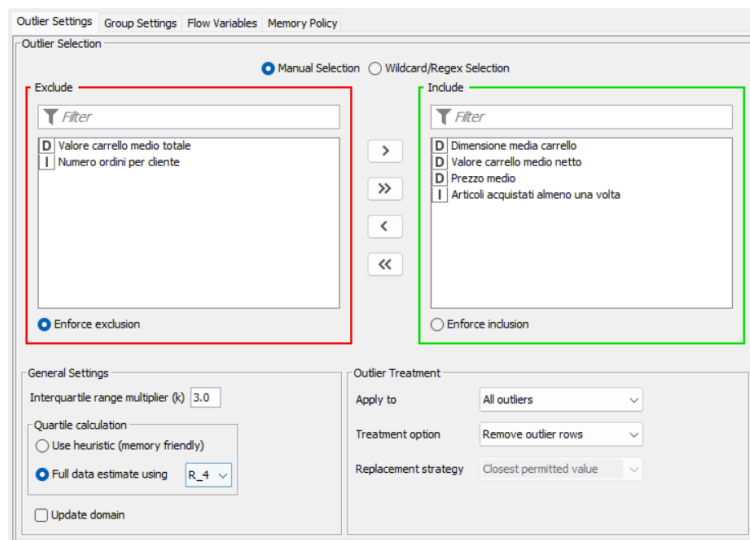


Figura 42: Configurazione nodo *Numeric Outliers* (Fonte: KNIME)

Nei *settings* del nodo, oltre ad impostare il range dello scarto interquartile, è possibile indicare anche come devono essere trattati i valori che vengono contrassegnati come *outliers*. Il trattamento scelto in questo caso è la rimozione delle righe contenenti i valori anomali. Questa configurazione ha permesso di eliminare solo una piccola parte dei dati presenti nel database, corrispondente a circa un 7% delle righe totali (12.326 righe originali, contro 11.467 righe dopo aver scartato i valori anomali).

Normalizer. Il nodo in questione viene utilizzato per effettuare la normalizzazione dei valori di tutte le colonne numeriche del database. La tecnica di normalizzazione viene spesso applicata nella fase di preparazione dei dati da dare in pasto agli algoritmi di *machine learning* ed ha l'obiettivo di modificare i dati utilizzando una scala comune che eviti distorsioni negli intervalli di valori o una perdita di informazioni dovuta alle scale numeriche utilizzate nelle diverse colonne. Permette così agli algoritmi di modellare in maniera corretta i dati [49]. La finestra di dialogo del *Normalizer* consente di selezionare le colonne su cui si vuole lavorare e di scegliere che metodo utilizzare per la normalizzazione. Nella configurazione è stata scelta la normalizzazione *Min-Max* che prevede una trasformazione lineare dei valori in modo che il minimo ed il massimo di ciascuna colonna siano rispettivamente 0 e 1. È possibile utilizzare anche le normalizzazioni *Z-score* o su scala decimale.

k-means. Una volta normalizzati i dati è stato possibile applicare il nodo che implementa l'algoritmo *k-means* di *machine learning* per effettuare la clusterizzazione. L'algoritmo permette di suddividere un set di dati in un numero predefinito di *k cluster*. Per la formazione dei cluster, assegna ad ogni record delle coordinate in un determinato "spazio" dei record, che ha un numero di dimensioni pari al numero dei campi considerati. Ogni coordinata del record è rappresentata quindi dal valore di ciascun campo. Per effettuare questo tipo di operazione l'algoritmo di *clustering* utilizza la distanza

euclidea sugli attributi selezionati. Al fine di trarre utilità da questa interpretazione geometrica, è necessario far precedere il nodo dalla normalizzazione dei dati, come fatto precedentemente, per permettere la comparazione delle variazioni tra le diverse dimensioni. L'assegnazione dei record ai *cluster* avviene attraverso un processo iterativo partendo da gruppi centrati in posizioni casuali nello spazio dei record e spostando i centroidi [27]. Il nodo, in particolare, permette di restituire i centri del *cluster* per un numero predefinito k ed esegue un *crispclustering*, assegnando esattamente un vettore di dati ad ogni gruppo di elementi definito. Il lavoro dell'algoritmo termina quando le assegnazioni dei dati ai *cluster* non cambiano più. L'output che il nodo presenta dopo aver effettuato le operazioni di *clustering* consiste in una tabella con gli attributi di partenza a cui viene aggiunta una colonna che etichetta ogni record (ancora in scala 0-1) a seconda dell'appartenenza al *cluster*.

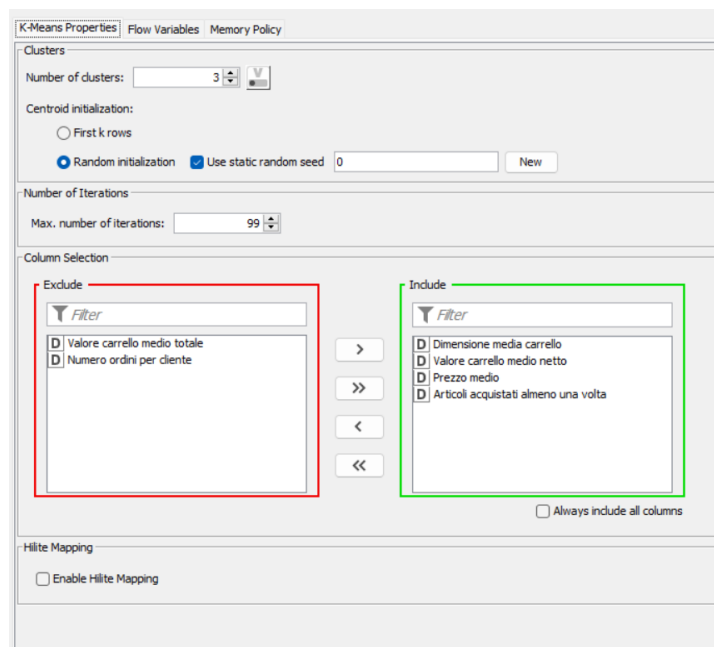


Figura 43: Configurazione del nodo *k-means* (Fonte: KNIME)

Per il *clustering*, sono state scelte le quattro dimensioni più significative ricavate dalle precedenti elaborazioni dei dati: *dimensione media carrello*, *valore carrello medio netto*, *prezzo medio*, *articoli acquistati almeno una volta*. Si è preferito mantenere il valore del carrello netto per poterlo comparare con il prezzo medio, anch'esso netto, ed è stata scartata la variabile *numero ordini per cliente* perché, come precedentemente visto con l'analisi RFM, si ha avuto conferma del fatto che il numero di ordini effettuati dai clienti varia in media da 1 a 3 (solo pochi record contano con un numero di ordini maggiore) e la dimensione non risulta quindi significativa nella differenziazione del comportamento dei clienti. Si è preferito utilizzare la quantità e la varietà dei prodotti acquistati, rispettivamente con la *dimensione media carrello* e *articoli acquistati almeno una volta*. Il numero k di *cluster* ottimale

viene generalmente riconosciuto tra i 3 ed i 5 gruppi. Prima di scegliere il numero definitivo di *cluster*, sono state effettuate delle prove con 3 e 4 gruppi. Da queste operazioni, i risultati più significativi sono stati raggiunti con l'utilizzo del parametro $k=3$, che è stato scelto come definitivo. Come conferma è stato utilizzato il nodo *Statistics* che, come si vedrà in seguito, ha permesso di valutare la distribuzione dei record e le statistiche di base dei *cluster* a seconda del k scelto.

Denormalizer. Questo nodo permette di denormalizzare i dati di input forniti dal nodo *Normalizer*. La trasformazione avvenuta precedentemente viene in questo modo convertita ed i valori ritornano ed essere visualizzati nella loro scala originale, permettendo all'utente una più chiara interpretazione delle variabili. Viene sempre utilizzato dopo la normalizzazione ed in seguito all'utilizzo di uno dei nodi di *mining*.

9.3.3.3. Fase 3

La terza fase del flusso di lavoro consiste nella visualizzazione dei risultati derivati dall'analisi, attraverso l'utilizzo dei nodi visibili in figura 44.

Color Manager. Il nodo viene utilizzato per assegnare dei colori ai *cluster* creati e per poterli identificare facilmente una volta implementati i nodi per la visualizzazione dei risultati.

Statistics. Calcola delle statistiche descrittive di base come i valori di minimo e massimo, la media, la deviazione standard, la varianza, la mediana, la somma complessiva, il numero di valori mancanti e il conteggio delle righe su tutte le colonne numeriche. Permette anche di contare tutti i valori nominali assieme alle loro occorrenze. Come anticipato precedentemente, questo ha permesso di valutare la composizione dei *cluster* e scegliere il parametro k più significativo per effettuare l'analisi.

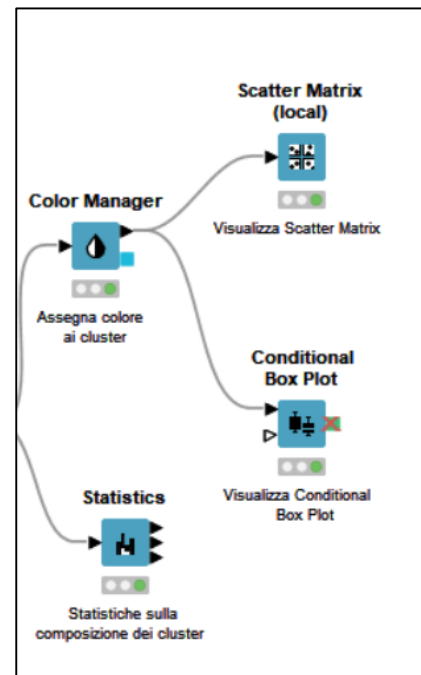


Figura 44: Nodi per la visualizzazione dei risultati (Fonte: KNIME)

Scatter Matrix. La matrice a dispersione permette di osservare la posizione dei *cluster* a seconda delle varie dimensioni. Ogni elemento della matrice è un grafico a dispersione che visualizza i valori i della i -esima colonna sull'asse delle ascisse ed i valori della j -esima colonna sull'asse delle ordinate, mentre le coordinate vengono visualizzate in maniera alternata su tutti i lati della trama.⁶ Grazie alla

⁶ KNIME – Node description – Scatter Matrix

Scatter Matrix è possibile identificare le relazioni che esistono tra le variabili ed i *cluster*, a seconda di dove tendono ad incontrarsi i punti dello stesso colore.

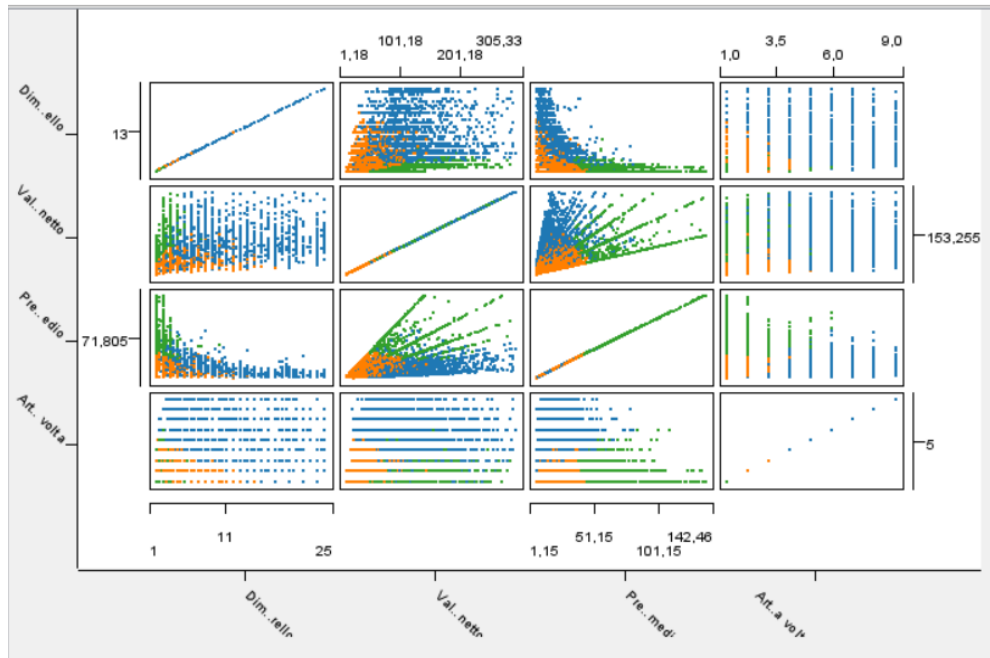


Figura 45: Rappresentazione della Scatter Matrix (Fonte: KNIME)

Conditional Box Plot. Il nodo *Conditional Box Plot* permette di visualizzare graficamente la distribuzione dei dati per variabili continue. Un box plot mostra i parametri statistici robusti di una distribuzione, ovvero quelli non sensibili agli *outliers* estremi, come il minimo ed il massimo, il quartile inferiore e superiore e la mediana. Il diagramma suddivide i dati di una colonna numerica in classi in base ad un'altra colonna nominale e crea una “scatola” per ciascuna delle classi. Un box plot per una variabile numerica si compone di un numero di scatole quanti sono i *cluster* individuati, che vanno dal quartile inferiore Q_1 al quartile superiore Q_3 . Nel mezzo del box è indicata la mediana, rappresentata da una barra orizzontale. La distanza tra primo e terzo quartile è definita scarto interquartile (IQR). Nella parte superiore e inferiore della figura sono presenti i cosiddetti “baffi”, collocati al valore minimo e massimo e rappresentati attraverso due barre orizzontali collegate al riquadro da una linea tratteggiata. Essi non superano mai il valore $1,5 \cdot IQR$, di conseguenza tutti i punti che oltrepassano $Q_1 - (1,5 \cdot IQR)$ o $Q_3 + (1,5 \cdot IQR)$ vengono disegnati separatamente come *outliers*. Questi valori anomali si distinguono ulteriormente in lievi o estremi. Gli *outliers* lievi sono quei punti che si trovano tra $1,5 \cdot IQR$ e $3 \cdot IQR$, mentre gli *outliers* estremi sono quei punti che presentano valori minori di $Q_1 - (3 \cdot IQR)$ o maggiori di $Q_3 + (3 \cdot IQR)$. I primi vengono visualizzati

nel box plot come cerchi, i secondi, invece, sono rappresentati attraverso delle croci.⁷ In figura 46 vengono riportati i risultati rappresentati nel box plot per ogni dimensione considerata nell'analisi.

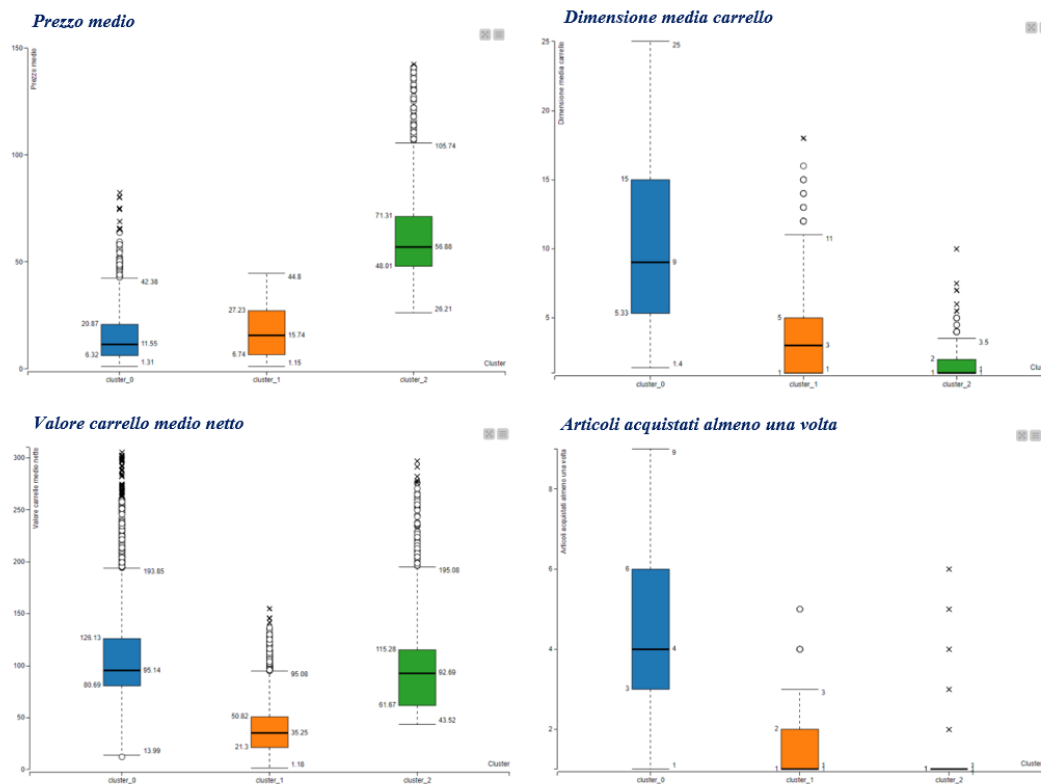


Figura 46: Visualizzazione dei risultati nel box plot divisi per variabile (Fonte: KNIME)

9.3.3.4. Fase 4

Dopo aver visualizzato i risultati ottenuti, è necessario procedere alla loro interpretazione per contestualizzare e dare un senso pratico all'analisi di *clustering* effettuata.

Dal set di dati a disposizione sono stati ricavati 3 *cluster*, come è possibile notare dal box plot sopra riportato.

Cluster_0. Il cluster_0, rappresentato con il colore blu, è il meno numeroso e comprende 2797 clienti. Ciò che lo differenzia principalmente dagli altri due *cluster* sono valori elevati sia nella dimensione media del carrello (con mediana uguale a 9) che nella varietà dello stesso, con un box plot relativo al numero di articoli diversi acquistati almeno una volta molto ampio (da 3 a 6). Possiamo vederlo dal fatto che le scatole blu per queste due dimensioni sono molto più alte delle altre e non c'è

⁷ KNIME – Node description – Conditional Box Plot

sovrapposizione con box di colori diversi. Anche il valore del carrello è elevato rispetto alla media registrata tra *cluster*, ma il prezzo unitario per fattura risulta tra i più bassi. Facendo un'analisi più approfondita, anche prendendo in considerazione la tabella di output in cui vengono classificati i record, è possibile notare che appartengono a questo *cluster* la maggior parte dei clienti che hanno un numero di ordini leggermente più elevato degli altri. Ne ricaviamo che i contatti appartenenti a questo *cluster* potrebbero essere verosimilmente coloro che non acquistano generalmente prodotti di alto valore, ma provano spesso prodotti diversi, comperano in maggiore quantità e ritornano ad acquistare in media un numero di volte più elevato rispetto agli utenti appartenenti agli altri *cluster*. I clienti che si differenziano per questo tipo di caratteristiche potrebbero rispondere in maniera positiva a comunicazioni frequenti, che propongano buoni sconto o campioni gratuiti, vista la propensione ad acquistare prodotti diversi nel tempo. L'obiettivo potrebbe essere quello di mantenerli e fidelizzarli, cercando di aumentare il prezzo medio degli articoli che vengono messi a carrello.

Cluster_1. Il *cluster_1*, rappresentato dal colore arancione, comprende 5510 clienti. È caratterizzato da una dimensione carrello media, che va da 1 a 5 prodotti e da una diversità di acquisto nei prodotti bassa, generalmente 1 o 2 prodotti riacquistati più di una volta. Il valore del carrello medio ed il prezzo medio sono bassi. Questo gruppo comprende i clienti presumibilmente più occasionali, con carrelli di basso valore e scarsa ampiezza di prodotti del catalogo provati. Sono contatti non fidelizzati al brand. È il *cluster* più numeroso e, ricollegandoci a quanto visto nell'analisi RFM, possiamo trovare una correlazione con i risultati precedentemente ricavati per cui la maggior parte di clienti è caratterizzata da una frequenza di acquisto casuale e carrelli medi bassi. Gli utenti di questo tipo, meno redditizi ma più numerosi, possono essere stimolati con campagne che fanno informazione su prodotti nuovi e con promozioni o sconti sui prodotti simili o complementari a quelli acquistati in precedenza (es. prodotti tavola appartenenti alla stessa collezione). È possibile che si tratti di utenti che hanno acquistato sul sito sapendo già di cosa avevano bisogno e che non hanno avuto modo di navigare tra l'assortimento prodotti. In questo caso potrebbe essere rilevante utilizzare i *recommendation systems* presenti nella piattaforma di *marketing automation* per stimolare la curiosità su prodotti nuovi e spingere l'utente ad effettuare altri acquisti.

Cluster_2. Il *cluster_2*, rappresentato con il colore verde, comprende 3160 clienti. Si caratterizza per una dimensione e varietà di carrello media, sensibilmente inferiori rispetto agli altri gruppi, ma conta con un valore medio del carrello alto ed il prezzo medio più elevato tra i *cluster*. Possono essere classificati come clienti che acquistano generalmente pochi prodotti (in media 1 o 2), ma che tendono a scegliere tra gli articoli più costosi del catalogo. Ci immaginiamo che gli acquisti siano principalmente legati alle batterie di pentole o ai servizi piatti completi. Questo comportamento

permette di indirizzare verso questo *cluster* di clienti comunicazioni e promozioni che spingono a dare visibilità alla sezione “premium” dell’assortimento, focalizzandosi, più che su buoni sconto e offerte, sugli aspetti qualitativi degli articoli top di gamma.

L’utilizzo della tecnica di *clustering* ci ha permesso di studiare delle tendenze nei comportamenti di acquisto dei clienti sul sito web, di confrontarne le caratteristiche e di segmentare attraverso un approccio di *data mining* l’enorme mole di contatti presenti nel database delle transazioni. Grazie alle informazioni ricavate è possibile andare a creare dei nuovi segmenti all’interno della stessa piattaforma di *marketing automation* sulla base delle caratteristiche che definiscono i diversi *cluster* e creare dei contenuti dinamici per il sito web o, per gli iscritti alla newsletter, di gestire in maniera più personalizzata l’invio delle comunicazioni periodiche. Questo tipo di tecnica avanzata, attraverso l’utilizzo di software *open source* come KNIME, può essere di supporto e complementare alle attività di *data analytics* già effettuate dall’azienda e può permettere, a seconda degli obiettivi, di ricavare informazioni più nascoste tra i dati, non immediatamente visibili, ed approcciarsi alla relazione con il cliente ed alle attività di marketing con un occhio più analitico e basato su ciò che mostrano i dati.

10. Conclusioni

Negli ultimi anni i *marketers* si sono trovati ad affrontare una serie di sfide complesse. La saturazione del mercato e le richieste sempre più esigenti dei consumatori stanno spingendo oggi le aziende a focalizzarsi non solo sull'acquisizione di nuovi clienti ma, soprattutto, sullo sviluppo di relazioni stabili e durature con quelli esistenti, cercando di alimentare e stimolare il più possibile i contatti per fidelizzarli al brand. Più aumenta la concorrenza, più diventa necessario per le imprese sapersi distinguere offrendo al cliente il prodotto giusto, al momento giusto, attraverso il canale migliore. Prendendo anche in considerazione l'evoluzione che sta attraversando il mondo del business negli ultimi anni, diventa imprescindibile per chi si occupa di marketing sviluppare competenze analitiche di analisi dei dati per permettere alle aziende di sfruttare le informazioni che possiedono o che potrebbero raccogliere, per acquisire vantaggio competitivo sui concorrenti, soprattutto online. Oggi la cultura del dato diventa fondamentale all'interno dell'impresa e si sta acquisendo sempre più consapevolezza sull'importanza di adottare un approccio *data-driven* alla base di tutte le attività di *decision making*. Il marketing rappresenta una tra le funzioni aziendali che può ricavare enormi vantaggi da questo tipo di orientamento ai dati.

Ripercorrendo i temi trattati nell'elaborato, risulta subito chiaro come la contaminazione tra attività legate alla *data science* ed alle nuove tecnologie stia oggi sempre più caratterizzando la realtà imprenditoriale in ogni settore. Il marketing delle performance diventa fondamentale nell'era del digitale per raggiungere in maniera più efficiente ed efficace gli obiettivi di business riducendo gli sprechi ed ottimizzando le risorse a disposizione, consentendo all'uomo di ridurre tempi e sforzi, facendosi affiancare da dispositivi e software veloci, automatici ed intelligenti. Nonostante i dati siano spesso sporchi e difficili da interpretare, gli strumenti di *data mining* e *analytics* permettono di lavorare ad una loro trasformazione ed interpretazione per ricavarne informazioni utili al business a seconda delle esigenze. Sulla base di queste premesse, toccare con mano la realtà delle automazioni per il marketing mi ha permesso di combinare gli studi di *marketing analytics* con le conoscenze pratiche acquisite in ambito lavorativo, consentendomi di studiare un approccio ibrido tra strumenti di *data mining* e di *marketing automation*, che potesse essere utile a livello aziendale per ampliare le possibilità delle imprese o delle agenzie partner relativamente alle modalità di segmentazione e analisi della clientela. In seguito alla presentazione dell'elaborato nel suo complesso, è possibile concludere che i risultati ottenuti attraverso gli strumenti di *data mining* e *machine learning* possano integrarsi e potenziare le attività svolte dai software di *marketing automation* e intelligenza artificiale per ricavare informazioni e similarità tra gli utenti al fine di indirizzare specifiche campagne di marketing e comunicazione ad una serie di contatti che si assomigliano per comportamenti e caratteristiche. Tutto

ciò permette alle aziende di utilizzare al massimo i dati raccolti attraverso le piattaforme digitali acquisendo ed elaborando informazioni sui propri clienti, prevedendone le azioni future e, allo stesso tempo, assicurando ad ogni utente un *customer journey* personalizzato ed un'esperienza di acquisto online rilevante, che spinga alla conversione.

11. Elenco delle figure

Figura 1: Confronto tra gli aspetti del marketing 1.0, 2.0, 3.0 (Fonte: Marketing 3.0)	7
Figura 2: Le 3V dei <i>Big data</i> (Fonte: <i>Big data: Concepts, Technology, and Architecture</i> , 2021)	21
Figura 3: Il modello ITMI dei <i>Big data</i> con gli argomenti più popolari connessi a ciascuna componente (Fonte: <i>Big data per il Business</i> , 2020)	23
Figura 4: Le skills del data scientist (Fonte: <i>Data science e Machine learning: Dai Dati alla Conoscenza</i> , 2021)	27
Figura 5: I tre livelli di data analytics (Fonte: <i>Big data per il Business</i> , 2020)	31
Figura 6: Processo di KDD (Fonte: Introduzione al KDD e al <i>Data mining</i>)	34
Figura 7: Il ciclo di vita del processo CRISP-DM (Fonte: Researchgate.net)	37
Figura 8: Funzionamento di un albero decisionale (Fonte: IBM)	40
Figura 9: Metodo di <i>clustering</i> gerarchico rappresentato attraverso un dendogramma (Fonte: developersmaggioli.it)	41
Figura 10: Modello di <i>clustering</i> non gerarchico, attraverso l'utilizzo dell'algoritmo <i>k-means</i> . (Fonte: medium.com)	42
Figura 11: Approccio tradizionale alla programmazione informatica (Fonte: <i>Big data per il Business</i> , 2020)	46
Figura 12: Approccio di <i>machine learning</i> (Fonte: <i>Big data per il Business</i> , 2020)	46
Figura 13: Processo di supervised learning (Fonte: apogonline.com)	48
Figura 14: Processo di reinforcement learning (Fonte: apogonline.com)	49
Figura 15: Home page del sito istituzionale di Tognana Porcellane (Fonte: www.tognana.com/it)	62
Figura 16: Home page dello <i>shop</i> IT online di Tognana Porcellane (Fonte: www.tognana.com/it)	63
Figura 17: Scheda prodotto dell'e-commerce di Tognana Porcellane (Fonte:www.tognana.com/it)	64
Figura 18: Automation rule per l'assegnazione del <i>tag</i> SLEEPY (Fonte:SALESmanago)	70
Figura 19: Struttura di una automation rule per l'invio del modulo di agreement per le web push (Fonte:SALESmanago)	71
Figura 20: Esempio di struttura di un workflow (Fonte:SALESmanago)	73
Figura 21: Esempi di social proof widget (Fonte:SALESmanago)	74
Figura 22: Esempio di pop-up dinamico (Fonte:SALESmanago)	75
Figura 23: Esempio di notifica web push per i saldi (Fonte:SALESmanago)	76
Figura 24: Esempio di banner personalizzato (Fonte:SALESmanago)	76
Figura 25: Analisi della recency (Fonte:SALESmanago)	79
Figura 26: Analisi della Frequency (Fonte:SALESmanago)	80
Figura 27: Analisi della Monetary (Fonte SALESmanago)	81

Figura 28: Analisi incrociata frequency-monetary (Fonte:SALESmanago)	82
Figura 29: Analisi incrociata monetary-recency (Fonte:SALESmanago)	83
Figura 30: Analisi incrociata frequency-recency (Fonte:SALESmanago)	84
Figura 31: Analisi di predizione di acquisti futuri (Fonte:SALESmanago)	85
Figura 32: Analisi della probabilità di abbandono e CLV (Fonte:SALESmanago)	86
Figura 33: Analisi di predizione di acquisti futuri in base alla fonte (Fonte:SALESmanago)	87
Figura 34: Esempio di nodo su KNIME (Fonte:KNIME)	90
Figura 35: Interfaccia utente su KNIME (Fonte:KNIME)	92
Figura 36: Workflow di elaborazione dei file di dati a disposizione (Fonte: elaborazione personale su KNIME)	93
Figura 37: Configurazione nodo Joiner (Fonte:KNIME)	95
Figura 38: Workflow di <i>clustering</i> (Fonte: elaborazione personale su KNIME)	97
Figura 39: Configurazione nodo GroupBy (Fonte: KNIME)	98
Figura 40: Configurazione del nodo Math Formula (Fonte: KNIME)	99
Figura 41: Implementazione dei nodi di <i>data mining</i> per l'analisi dei cluster (Fonte: elaborazione personale su KNIME)	100
Figura 42: Configurazione nodo Numeric Outliers (Fonte: KNIME)	101
Figura 43: Configurazione del nodo <i>k-means</i> (Fonte: KNIME)	102
Figura 44: Nodi per la visualizzazione dei risultati (Fonte: KNIME)	103
Figura 45: Rappresentazione della Scatter Matrix (Fonte: KNIME)	104
Figura 46: Visualizzazione dei risultati nel box plot divisi per variabile (Fonte: KNIME)	105

12. Elenco delle tabelle

Tabella 1: Criteri per la costruzione del modello RFM (elaborazione personale)

13. Bibliografia e sitografia

- [1] Abdugofur Temirov, Ren Dongxiao (2018), *Data mining Techniques in E-Commerce*, International Journal of Science and Research (IJSR).
<https://www.ijsr.net/archive/v8i9/ART2020929.pdf>
- [2] Abdullahi Sidow Osman (2019), *Data mining Techniques: Review*, International Journal of Data science Research. https://www.academia.edu/77542877/Data_Mining_Techniques_Review
- [3] Adobe, *Report 2021 Area EMEA: Il modo giusto di fare e-commerce negli anni 2020*.
https://info2.magento.com/rs/585-GGD-959/images/Consumer-Research-2022-Adobe-EMEA-Report-V1.4_it.pdf
- [4] Arcuri, L. (2021), *Sales Manago: Cos'è E Come Funziona*, Digitools – Data Driven Marketing.
<https://www.digitools.it/sales-manago-cose-come-funziona/>
- [5] ART – Arti della Tavola e del Regalo, *Living - Osservatorio ART: Dati Mercato Italia 2° Trim 2022*, Italia News, 5 Luglio 2022. <https://art-tavolaregalo.it/living-art-living-osservatorio-art-dati-mercato-italia-2-trim-2022/>
- [6] ART – Arti della Tavola e del Regalo, *Osservatorio ART: Nuovi comportamenti di acquisto e futuro del retail nell'epoca del Covid-19*, Format Research, Milano, 23.04.2021. <https://art-tavolaregalo.it/wp-content/uploads/2021/04/Sintesi-Osservatorio-ART-2021.pdf>
- [7] ART – Arti della Tavola e del Regalo, *Living – Osservatorio ART: Rilevazione dati trimestrali e sentiment mercato Italia settore casa, tavola e cucina – 4Q 2021*, MMC PARTNERS per ART 01/2022. <https://art-tavolaregalo.it/wp-content/uploads/2022/01/LIVING-ART-4Q-2021-SENTIMENT.pdf>
- [8] Azevedo, A., Filipe Santos, M. (2008), *KDD, SEMMA and CRISP-DM: a parallel overview*. IADIS European Conference on *Data mining* 2008, Amsterdam.
https://www.researchgate.net/publication/220969845_KDD_semma_and_CRISP-DM_A_parallel_overview
- [9] Balusamy, B., Kadry, S., Abirami, N., Gandomi, A. H. (2021). *Big data: Concepts, Technology, and Architecture*. Wiley.
- [10] Bharati Ramageri, M. (2010), *Data mining Techniques and Applications*, Indian Journal of Computer Science and Engineering.
https://www.researchgate.net/publication/49616224_Data_mining_techniques_and_applications

- [11] Blog di ADABRA, *Analisi RFM per la profilazione utenti: cos'è e come funziona*.
<https://blog.adabra.com/analisi-rfm-per-la-profilazione-utenti/>
- [12] Blog SALESmanago (2018), *FT1000: SALESmanago fastest growing European marketing automation platform*. <https://blog.salesmanago.com/salesmanago-2/ft1000-salesmanago-fastest-growing-european-marketing-automation-platform/>
- [13] Brognara, R. (2021). *Come e cosa pensano gli algoritmi di marketing*. Franco Angeli Edizioni.
- [14] Carvalho, A., Moreira, J., Carvalho, A. C. P. d. L. F., Horvath, T. (2018), *A General Introduction to Data Analytics*. Wiley.
- [15] Casaleggio Associati, *Pensare digitale, E-commerce in Italia 2021 – L'e-commerce si consolida: le grandi manovre*, XV edizione del rapporto, Maggio 2021.
https://www.casaleggio.it/wp-content/uploads/2020/12/CA-E-commerce-2021-report-ITA_WEB.pdf
- [16] Celli, F. (2020). *Data science marketing. Applicazioni data driven nel digital marketing*. Maggioli Editore.
- [17] Chiappini, F. (2018), *Marketing automation per e-commerce*, Dario Flaccovio Editore.
- [18] Computerworld Italia, CIO, *Che cos'è il data mining? Scoprire modelli e tendenze nei dati*.
<https://www.cwi.it/big-data/big-data-analytics/che-cose-il-data-mining-scoprire-modelli-e-tendenze-nei-dati-140130>
- [19] Condemi, J. (2022), *Data mining: cos'è, come funziona, esempi e software*, Ai4business.it.
<https://www.ai4business.it/intelligenza-artificiale/machine-learning/data-mining-cose-come-funziona-esempi-e-software/>
- [20] Cote, C. (2021), *What is Diagnostic Analytics: 4 examples*, Harvard Business School Online.
<https://online.hbs.edu/blog/post/diagnostic-analytics>
- [21] Cote, C. (2021), *What is Predictive Analytics: 5 examples*, Harvard Business School Online.
<https://online.hbs.edu/blog/post/predictive-analytics>
- [22] Cote, C. (2021), *What is Prescriptive Analytics: 6 examples*, Harvard Business School Online.
<https://online.hbs.edu/blog/post/prescriptive-analytics>

- [23] Della Bella, F. (2021), *I KPI del Marketing: le metriche fondamentali per misurazione delle performance*, Digital4. <https://www.digital4.biz/marketing/big-data-e-analytics/kpi-marketing-misurazione-delle-performance/>
- [24] De Mauro, A. (2019). *Big data Analytics: Analizzare e interpretare dati con il machine learning*. Feltrinelli Editore.
- [25] De Mauro, A. (2020). *Big data per il Business: Guida strategica per manager alle prese con la trasformazione digitale*. Feltrinelli Editore.
- [26] De Mauro, A. (2022). *Data analytics per tutti. Imparare ad analizzare, visualizzare e raccontare i dati*. Apogeo.
- [27] Dispense di didattica online, Dipartimento di Informatica, Università degli Studi di Torino, *Cluster analysis*. <http://www.di.unito.it/~cena/materiale/GestioneDB/Cluster.pdf>
- [28] Ducange, P., Pecori, R., Mezzina, P. (2016), *A Glimpse on Big data Analytics in the Framework of Marketing Strategies*. <https://arpi.unipi.it/bitstream/11568/988867/3/postPrintSOCOBigData.pdf>
- [29] Di Nuzzo, M. (2021), *Data science e Machine learning: Dai Dati alla Conoscenza*.
- [30] Dulli, S., *Il Data mining*, Dispense per la didattica, Università degli Studi di Padova. <https://www.math.unipd.it/~dulli/corso06/DMteoria.pdf>
- [31] Favaretto, D., *Dispense per la didattica, Corso di Marketing Analytics*.
- [32] Fuciu, M. Dumitrescu, L. (2018). *From Marketing 1.0 To Marketing 4.0 – The Evolution of the Marketing Concept in the Context of the 21ST Century* - International conference Knowledge-Based Organization, vol.24, no.2, pp.43-48. <https://doi.org/10.1515/kbo-2018-0064>
- [33] Greaves R., *Una revisione completa di SALESmanago: tutto ciò che devi sapere*, E-commerce Platforms. <https://ecommerce-platforms.com/it/articles/salesmanago-review>
- [34] Hall, M. A., Witten, I. H., Pal, C. J., Frank, E. (2016). *Data mining: Practical Machine learning Tools and Techniques*, Elsevier Science.
- [35] Iandiorio, E. (2021). *Big data: cosa sono, come analizzarli e utilizzarli per fare marketing*. Dario Flaccovio Editore.
- [36] IBM, *Cos'è Hadoop?, Cos'è Map Reduce?*, www.ibm.com

- [37] InsideMarketing.it, *Big data cosa sono e perché sono importanti per le aziende*.
<https://www.insidemarketing.it/glossario/definizione/big-data/>
- [38] IlSole24Ore (2021), *La relazione con i clienti è sempre più digitale, fra nuove metriche e smart data. Con la pandemia, il marketing ha fatto un nuovo passo in avanti*.
https://www.ilsole24ore.com/art/la-relazione-i-clienti-e-sempre-piu-digitale-nuove-metriche-e-smart-data-la-pandemia-marketing-ha-fatto-nuovo-passo-avanti-AE6fBUt?refresh_ce=1
- [39] Kartajaya, H., Kotler, P., Setiawan, I. (2011). *Marketing 3.0: Dal prodotto al cliente, all'anima*. Gruppo 24 ore.
- [40] Kartajaya, H., Kotler, P., Setiawan, I. (2021). *Marketing 5.0: Tecnologie per l'umanità*. Hoepli.
- [41] Khurshid, K., Khan, A. A, Siddiqi, H., Rashid, I. (2018), *Big data-9Vs, Challenges and Solutions*, Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan.
https://www.researchgate.net/publication/329359973_Big_data-9Vs_Challenges_and_Solutions
- [42] Leonardi, A. (2019), *Il valore del Machine learning nel Marketing*, Ai4business.com.
<https://www.ai4business.it/intelligenza-artificiale/machine-learning/machine-learning-nel-marketing/>
- [43] Lim, J., Miller, J. D. (2020). *Data-First Marketing: How To Compete and Win In the Age of Analytics*. Wiley.
- [44] Manganaro, V.A., *Introduzione al KDD e al Data mining*.
<https://digilander.libero.it/statistici/articoli/manganaro1.pdf>
- [45] Marbán, Ó., Mariscal, G., Segovia, J. (2009) , *A Data mining & Knowledge Discovery Process Model*, Facultad de Informática (Universidad Politécnica de Madrid), Universidad Europea de Madrid.
https://www.researchgate.net/publication/221787522_A_Data_Mining_Knowledge_Discovery_Process_Model
- [46] Materiale Didattico, Università degli Studi Mediterranea di Reggio Calabria, *Il Clustering*.
https://www.unirc.it/documentazione/materiale_didattico/599_2008_93_1623.pdf
- [47] McKinsey & Company (2020), *Elevating customer experience excellence in the next normal*.
<https://www.mckinsey.com/business-functions/operations/our-insights/elevating-customer-experience-excellence-in-the-next-normal>

- [48] McKinsey & Company, *Emerging consumer trends in a post COVID 19 world*.
<https://www.mckinsey.com/business-functions/growth-marketing-and-sales/our-insights/emerging-consumer-trends-in-a-post-covid-19-world>
- [49] Microsoft, *Normalizzare il componente dati*, 27/07/2022. <https://docs.microsoft.com/it-it/azure/machine-learning/component-reference/normalize-data>
- [50] Moneta, J., Sinclair, L. (2020), *Il COVID-19 ha accelerato l'adozione del digitale: è il momento giusto per la trasformazione*, Think with Google.
<https://www.thinkwithgoogle.com/intl/it-it/futuro-del-marketing/digital-transformation/covid-ha-accelerato-adozione-digitale/>
- [51] Mura, C., *Marketing 4.0: evoluzione nell'era digitale*, Blog Digital Coach. <https://www.digital-coach.com/it/blog/case-histories/marketing-4-0/>
- [52] Nowak-Brzezińska A, Gaibei I. (2022), *How the Outliers Influence the Quality of Clustering?*, Entropy (Basel). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9324173/>
- [53] Oracle.com, *Cosa sono i Big data?* <https://www.oracle.com/it/big-data/what-is-big-data/>
- [54] Pagani, P. (2021), *La casa al centro dell'universo consumer*, GDOWEEK.
<https://www.gdoweeek.it/la-casa-al-centro-delluniverso-consumer/>
- [55] Pagani, P. (2021), *Lockdown e smart working riaccendono i riflettori sui casalinghi*, Mark-up.
<https://www.mark-up.it/lockdown-e-smart-working-riaccendono-i-riflettori-sui-casalinghi/>
- [56] Pande, S.R., Sambare, S.S., Thakre, V.M. (2012), *Data Clustering Using Data mining Techniques*, International Journal of Advanced Research in Computer and Communication Engineering.
https://www.academia.edu/41021790/Data_Clustering_Using_Data_Mining_Techniques
- [57] Pepping, J. (2017), *The individual and organizational factors influencing the implementation of Data-driven Marketing*, Master Thesis, University of Twente.
https://essay.utwente.nl/72492/1/Pepping_MA_BMS.pdf
- [58] Ramon Saura, J., Ribeiro-Soriano, D., Palacios-Marqués, D. (2021) *From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets*, International Journal of Information Management, vol. 60.
<https://doi.org/10.1016/j.ijinfomgt.2021.102331>

- [59] Raschka, S., Mirjalili, V. (2020), *I tre diversi tipi di machine learning e a che cosa servono*, Apogeeonline.com. <https://www.apogeeonline.com/articoli/i-tre-diversi-tipi-di-machine-learning-e-a-che-cosa-servono-sebastian-raschka-vahid-mirjalili/>
- [60] Restegari, H., *Data mining and E-Commerce: Methods, Applications and Challenges*, Faculty of Computer Science and Information Systems - University Technology Malaysia. <https://core.ac.uk/download/pdf/11784915.pdf>
- [61] Sap Insights, *Che cos'è il machine learning?*. <https://www.sap.com/italy/insights/what-is-machine-learning.html>
- [62] SAS Italy, *Che cos'è il Data mining?* https://www.sas.com/it_it/insights/analytics/data-mining.html
- [63] SAS Italy, *Le Reti Neurali Artificiali: cosa sono e perché sono importanti*. https://www.sas.com/it_it/insights/analytics/neural-networks.html
- [64] Sistelli, C. (2022), *Machine learning: 3 tipi di apprendimento per l'intelligenza artificiale*, Investech.blog. <https://investech.blog/2022/06/22/machine-learning-3-tipi-di-apprendimento-per-lintelligenza-artificiale/>
- [65] Sito web di Tognana Porcellane. <https://www.tognana.com/it/>
- [66] Suárez, T. (2018). Evolución del marketing 1.0 al 4.0, Redmarka. Revista de Marketing Aplicado, vol 01, núm. 022, 209-227. <https://doi.org/10.17979/redma.2018.01.022.4943>
- [67] Support SALESmanago, *Adding New Automation Rule*. <https://support.salesmanago.com/adding-new-automation-rule/>
- [68] Support SALESmanago, *Automation Processes*. <https://support.salesmanago.com/automation-processes/>
- [69] Support SALESmanago, *Contact Segmentation Tags*. <https://support.salesmanago.com/contact-segmentation-tags/>
- [70] Support SALESmanago, *Customer Segmentation Center*. <https://support.salesmanago.com/customer-segmentation-center/>
- [71] Support SALESmanago, *Dynamic Segments*. <https://support.salesmanago.com/dynamic-segments-2/>

[72] Support SALESmanago, *Personalized banners*. <https://support.salesmanago.com/dynamic-banners/>

[73] Support SALESmanago, *Prediction Analytics on the contact card*.
<https://support.salesmanago.com/prediction-analytics-on-the-contact-card/>

[74] Support SALESmanago, *Website Personalization*. <https://support.salesmanago.com/dynamic-content/>

[75] Support SALESmanago, *Web Push*. <https://support.salesmanago.com/web-push/>

[76] Treccani Enciclopedia Online. <https://www.treccani.it/enciclopedia/dato>

[77] Thompson, W., *Come fanno i sistemi di raccomandazione a sapere cosa potrebbe piacerti?*, SAS Italy. https://www.sas.com/it_it/insights/articles/big-data/recommendation-systems.html

[78] Wikipedia, *KNIME*. <https://it.wikipedia.org/wiki/KNIME>

[79] Wikipedia, *Legge di Moore*. https://it.wikipedia.org/wiki/Legge_di_Moore