Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale
A.a. 2020/2021
Sessione di Laurea Dicembre 2021

Progettazione ed implementazione di un modello di lead scoring

Il caso di un'impresa nel settore dei capital goods

Relatore:
Professor Fulvio Corno

Candidato:
Federica Curello

Sommario

Introduzione 4
Scopo della tesi 5
Prima Sezione: Letteratura 7
CRM- Customer Relationship Management 7
Targeted marketing 9
Il Regolamento Europeo sulla Protezione dei Dati (GDPR) 10
Marketing funnel 11
Lead scoring 14
Metodi di costruzione di un modello di lead scoring: approccio manuale versus metodi di AI 19
Albero decisionale 20
Random Forest 22
Gradient Boosted Trees 23
Regressione logistica 25
Indicatori di prestazione per la scelta del metodo di machine learning da adottare per il lead scoring 26
KPIs da misurare per migliorare la gestione dei lead 30
Prospettive di ricerche future 32
Seconda Sezione: Il caso applicativo 34
Presentazione del caso 34
Il contesto di CNH Industrial e la necessità di implementare il lead scoring 36
Aspetti metodologici 38
Identificazione qualitativa degli attributi adatti a costruire un modello di lead scoring 39
Reperibilità del lead 39
Interesse esplicitato dal lead 40
Attributi comportamentali che catturano l'interesse implicito 40
Profilo cliente target 41
Analisi quantitativa dei dati 43
Considerazioni preliminari 43
Attendibilità dei dati lato-utente 43
Affidabilità del sistema lato-utilizzatori 44
CNH Industrial Direct Marketing Consent 45
CNH Industrial Third Marketing Consent 47
Main Phone e Mobile Phone 48
E-mail 48
Contact Reason Detail 49
Normalization State 50

CreatedOn 51
Purchase Timeframe 53
Qualifying Opportunity 56
Channel 56
Activities Count 57
Open Deals e Won Deals 58
E-mail Clicked 59
Limite nelle analisi effettuate 59
Assegnazione dei pesi relativi per ciascun attributo 60
Attribuzione dei punteggi all'interno di ciascun campo 61
Suddivisione nelle soglie di rating 63
Implementazione in Microsoft Dynamics 365 Marketing 65
Monitoraggio e KPIs 70
Monitoraggio per rispondere ai cambiamenti 70
Verifica dell'adeguatezza dello strumento di lead scoring 70
KPIs per la misurazione delle prestazioni 73
Prospettive future all'interno dell'impresa 74
Conclusioni 77
Bibliografia 78
Sitografia 79

Introduzione

Negli ultimi decenni si è assistito ad una rapidissima evoluzione digitale, che si è manifestata in innovazioni di tipo tecnologico che hanno coinvolto praticamente ogni ambito della vita dell'uomo: dalla sfera privata e sociale, ai vari settori di mercato, come sanità, istruzione, industrie, aziende di servizi. Questa improvvisa e violenta variazione del contesto innescata da una diffusione sempre più pervasiva del digitale viene comunemente definita come "*digital disruption*", un'espressione fortemente rappresentativa di una vera e propria rottura con i vecchi paradigmi e cambiamento verso nuovi processi, nuove strutture, nuovi modelli di business nel panorama dei mercati di tutto il mondo. Si tratta di uno sconvolgimento sistemico, che ha portato alla necessità per le imprese di reagire al cambiamento con rapidità per non essere costrette ad uscire dal mercato a causa delle pressioni di aziende tecnologicamente più avanzate. D'altra parte, questo mutamento forzato ha comportato notevoli vantaggi nella creazione di valore aziendale, attraverso un efficientamento dell'uso delle risorse: si pensi ad esempio all'utilizzo di macchinari sempre più smart, algoritmi di intelligenza artificiale e strumenti sempre più sofisticati di calcolo, analisi statistica e predittiva. Non solo la trasformazione digitale ha migliorato le funzioni interne all'azienda, ma ha influenzato anche le modalità di commercializzazione dei prodotti, nonché, più ampiamente, il rapporto con il cliente.

Infatti, le imprese sono state indotte a reiventarsi per sfruttare l'utilizzo della rete, approdando online attraverso le piattaforme di e-commerce, che hanno creato un nuovo paradigma di acquisto. Nella prospettiva aziendale, questa nuova opportunità di commercializzazione dei prodotti introduce una serie di facilitazioni nelle procedure di acquisto di un utente: la possibilità di accedere da qualsiasi area geografica, in qualsiasi momento, acquistando con la semplicità di un click consentono, infatti, di attrarre clientela potenzialmente più ampia dei negozi fisici, a costi notevolmente inferiori. Tuttavia, l'altro lato della medaglia vede il consumatore acquisire potere contrattuale grazie all'ampia informazione disponibile sul web: cosicché diviene immediato per un utente poter confrontare prodotti di aziende diverse e riuscire a identificare quello più conveniente economicamente, quello di migliore qualità o semplicemente quello più adatto alle sue esigenze o gusti personali. Esplode quindi la competitività e diventa sempre più importante riuscire a catturare l'attenzione del potenziale cliente nel modo giusto. Pertanto, diventa cruciale per l'impresa investire nel marketing digitale, strumento di fortissimo impatto in un'era in cui le persone vivono costantemente connesse al proprio dispositivo tecnologico. Inoltre, il marketing digitale apre le porte a potenti metodi di raccolta, utilizzo e analisi dei "big data" attraverso algoritmi che permettono di processare una moltitudine di dati, che sarebbe stato impossibile (o molto time-consuming e money-consuming) gestire in maniera manuale, senza il supporto tecnologico.

Scopo della tesi

Il presente lavoro di tesi viene redatto allo scopo di analizzare le principali strategie di marketing "data-driven", focalizzandosi specificamente sulla costruzione ed implementazione di un modello di "lead scoring" attraverso l'applicazione aziendale in Cluster Reply srl. La ricerca si propone di presentare il caso applicativo per riuscire a proporre una risposta formalizzata alla domanda *"Come costruire un modello di lead scoring adatto alla propria impresa?"*, suggerendo quindi un processo esemplificativo da emulare. Infatti, sebbene il caso in esame riporti le specificità dell'impresa considerata, l'obiettivo ultimo è quello di arricchire la letteratura sperimentale di quest'ambito per portare le imprese a catturare gli aspetti procedurali della costruzione di un modello di lead scoring e traslarli nel proprio contesto aziendale.

In particolare, nella prima sezione verrà esaminata la letteratura, partendo da concetti di marketing legati al "funnel" di acquisto e andando poi a considerare la loro declinazione in ambito digitale attraverso tecniche di profilazione degli individui, che, grazie alle odierne tecnologie, sono andate ad affinarsi anche su larga scala. Tra queste, particolare considerazione verrà dedicata alla segmentazione dei potenziali clienti (lead) attraverso i modelli di "lead scoring", di cui verranno delineati gli elementi chiave per la progettazione e i benefici che apporta sia nella relazione con il cliente sia nell'efficientamento delle risorse e riduzione dei costi per l'azienda. Inoltre, verranno presentati gli indicatori di prestazione (KPIs) utili a monitorare l'andamento delle campagne di marketing, con attenzione particolare agli indicatori riguardanti il comportamento dei lead, per ottimizzarne la gestione e valutare in termini economici di MROI (Marketing Return On Investment) la redditività che deriva da tali investimenti.

Nella seconda sezione, i concetti approfonditi a partire dalla letteratura troveranno applicazione concreta grazie all'esperienza di collaborazione con l'impresa Cluster Reply srl, operante nel settore della consulenza informatica e, più specificamente, nell'implementazione di soluzioni tecnologiche per la gestione del CRM di imprese clienti. L'obiettivo perseguito dalla tesi sarà quello di progettare per la prima volta un sistema di lead scoring che si adatti alle esigenze dell'impresa cliente CNH Industrial, operante nel mercato B2B, in particolare nel settore dei capital goods. Pertanto, la costruzione del modello sarà avviata a partire da considerazioni specifiche del business in cui opera l'impresa e dei processi che coinvolgono un potenziale cliente verso l'acquisto. Si risponderà quindi alle seguenti domande:

1. *Come selezionare gli attributi del lead sui quali stabilire le condizioni di assegnazione dei punteggi per il modello di lead scoring?*

2. *Come procedere per l'assegnazione dei pesi relativi ad ogni attributo in modo da creare un sistema di premiazione efficace, basandosi su presupposti teorici e sullo studio di dati pregressi?*
3. *Come suddividere le soglie di classificazione dei lead per garantire un rating appropriato per creare un ordinamento di priorità di contatto?*

A questo scopo, verranno effettuate delle interviste alla responsabile CRM di CNH Industrial per comprendere, in una visione concettuale, quali specificità dell'azienda e del cliente target debbano essere prese in considerazione nella segmentazione dei lead. Parallelamente, per gli aspetti più tecnici, verranno esaminati i campi all'interno dello schema CRM per capire come questi vengano popolati e quindi quali siano adeguati all'attribuzione dei punteggi di scoring. Inoltre, per alcune tipologie di dati non immediatamente visualizzabili verrà studiato come estrarli, implementando soluzioni alternative.

Una volta completata la strutturazione teorica del modello di lead scoring, verrà fornita una breve descrizione dell'implementazione di questo nell'ambiente di Microsoft Dynamics 365 Marketing. Infine, sebbene a causa delle tempistiche necessarie non sarà possibile monitorare direttamente il sistema di lead scoring, verranno proposti indicatori specifici per misurare quantitativamente i benefici apportati in relazione agli obiettivi prefissati.

Prima Sezione: Letteratura

CRM- Customer Relationship Management

Agli inizi degli anni ’90 le imprese hanno avviato un massivo processo di computerizzazione delle funzioni aziendali per migliorare l’efficienza complessiva delle proprie operazioni. Ad esempio, sono stati sviluppati i sistemi ERP (Enterprise Resource Planning) per la gestione di tutti i processi di business più rilevanti dell’impresa, integrando quindi contabilità, vendite, acquisti, magazzini, produzione, logistica e risorse umane. Tecnologie come queste inizialmente apportano un vantaggio competitivo per le poche imprese che riescono a creare i propri sistemi tecnologici in modo da apportare notevoli vantaggi di efficacia ed efficienza, ma, dopo un breve periodo di assestamento del mercato, diventano un vero e proprio requisito di sopravvivenza. In questo contesto di competizione già pressante, il livello di concorrenza viene ulteriormente inasprito dalla diffusione di Internet nei primi anni 2000. Si rafforza, quindi, la necessità di differenziare i propri prodotti e servizi, puntando sulla soddisfazione delle esigenze del cliente. Pertanto, nasce il concetto di Customer Relationship Management (CRM) come strategia di business che si focalizza sulla relazione con il cliente durante tutto il suo ciclo di vita, diventando fulcro principale dell’impresa. Il CRM si basa quindi sugli obiettivi di incrementare il numero di clienti, la loro soddisfazione nei confronti dei prodotti/servizi forniti e quindi la loro fidelizzazione verso l’impresa. Gli obiettivi principali possono essere, infatti, riassunti in:

- *Customer Identification*: identificazione del segmento di clientela più profittevole;
- *Customer Attraction*: attuazione di strategie di marketing per attrarre il segmento identificato verso i prodotti dell’impresa;
- *Customer Retention*: costruire fedeltà del cliente verso il brand per mantenere una relazione a lungo termine;
- *Customer Development*: condurre i propri clienti ad espandere l’intensità delle transazioni, portando a crescere la profittabilità di ciascuno di essi.

In quest’ottica, risulta indispensabile un’integrazione coerente tra le funzioni aziendali. In particolare, la produzione deve sempre essere collegata a specifiche analisi di mercato, che possano rivelare i bisogni, le tendenze di acquisto in tempo reale, in modo da offrire un prodotto che rispecchi le aspettative della domanda di mercato. Inoltre, per migliorare la gestione dei potenziali acquirenti, il marketing deve essere in grado di guidare le vendite attraverso processi di segmentazione della

clientela, che sfruttino le informazioni di mercato per discernere i clienti più profittevoli. La soluzione a queste esigenze aziendali, ancora una volta, si è prospettata attraverso le infrastrutture tecnologiche. I sistemi CRM si configurano, quindi, come un potente strumento per facilitare e rafforzare una relazione duratura con il cliente. Il punto di forza di queste tecnologie si fonda sulla capacità di automatizzare i processi di gestione del cliente per le aree marketing, vendite e servizi post-vendita. Infatti, tutte le informazioni vengono raccolte in un database all'interno del sistema, che si definisce pertanto come "data warehouse" in quanto permette l'immagazzinamento dei dati del potenziale cliente e delle sue interazioni con l'impresa. Questi possono essere facilmente consultati oppure analizzati tramite tecniche di "OLAP" (On-Line Analytical Processing) e di "data mining", ovvero metodologie che permettono di effettuare analisi complesse su una grande mole di dati per guidare i processi decisionali all'interno dell'azienda. In particolare, esistono quattro diverse tipologie di CRM:

- **CRM operativo**: viene utilizzato per gestire i processi aziendali collegati al cliente, ovvero all'interno delle funzioni di marketing, vendite e post-vendite. Si occupa, quindi, di ottimizzare la gestione dei contatti, delle interazioni con essi, delle informazioni sulle transazioni effettuate, sugli ordini e sulle spedizioni e, infine, di registrare i feedback ed eventuali reclami del cliente, attraverso processi automatizzati, come l'*automated e-mail marketing*, il quale verrà approfondito nei prossimi paragrafi.
- **CRM analitico**: raccoglie i dati su diversi ambiti aziendali (finanza, marketing, vendite, produzione, customer care) e applica algoritmi, tecniche di machine learning e di business intelligence per compiere analisi sui dati storici, misurare le performance attuali e prevedere i comportamenti futuri, ad esempio le tendenze di acquisto. Il risultato sono grafici e altri strumenti di reportistica, che il team aziendale esaminerà per disegnare le prossime strategie da attuare in risposta alle considerazioni emerse.
- **CRM strategico**: aiuta a valutare l'adeguatezza degli obiettivi strategici proposti per l'acquisizione di nuovi clienti e il mantenimento sul lungo termine. Pertanto, raccoglie le informazioni individuali sul singolo cliente o per segmenti di clienti e ne calcola i potenziali ricavi e la profittabilità, in relazione ai costi di investimento. Infine, può confrontare i dati analizzati con i trend di mercato per elaborare nuove proposte che siano più vantaggiose delle imprese competitor.
- **CRM collaborativo**: facilita la comunicazione interna tra le varie funzioni aziendali ed esterna con i clienti e con le altre imprese con cui intrattiene una relazione. Ad esempio, è in grado di sfruttare canali di comunicazione più efficienti, come e-mail, social networks, sms.

Il CRM è quindi uno strumento in grado di concentrare tutte le informazioni dell'azienda, rendendole consultabili alle varie funzioni in tempo reale e consentendo di poterle sfruttare per integrare tra loro i processi in modo coerente e rispondente ai bisogni del cliente.

Targeted marketing

Grazie a questi sviluppi tecnologici, è diventato possibile, nonostante la grande mole di dati, riuscire a non trascurare l'obiettivo del marketing di catturare le esigenze del singolo cliente. I processi di marketing si sono quindi evoluti, passando da un modello "*bulk marketing*", che aggregava i potenziali clienti inviando campagne e-mail di massa senza alcuna distinzione personale, ad un modello "*targeted marketing*", che sfrutta l'intelligenza artificiale e le informazioni che gli utenti lasciano in rete per elaborare specifici modelli probabilistici dei comportamenti di acquisto, da utilizzare per targettizzare l'invio di posta pubblicitaria. Si parla quindi di trasformazione verso sistemi definibili come "*customer-centric marketing*" e "*data-driven marketing*". Questi due concetti risultano strettamente collegati: infatti, la raccolta e analisi delle informazioni che gli utenti lasciano (a volte inconsapevolmente) sul web è una potentissima risorsa per veicolare processi di marketing, guidati per l'appunto dai dati, cosicché sia possibile creare strategie focalizzate sull'utente, secondo i suoi interessi, le sue reazioni sui social networks, le sue interazioni con le pagine web. In questo contesto, i grandi colossi digitali e le piattaforme social, come Google e Facebook, acquistano sempre più potere grazie all'enorme quantità di dati raccolta e le aziende vedono in queste un'ottima opportunità di sponsorizzazione pubblicitaria targettizzata.

Un altro aspetto che rende particolarmente efficiente ed efficace il marketing basato sui dati è la possibilità di studiare gli analytics in tempo reale e, pertanto, di poter reagire in modo quasi immediato seguendo le tendenze del momento, estremamente mutevoli. Infatti, vengono studiati, in relazione ai dati raccolti, anche i comportamenti psicologici, per riuscire ad "entrare nella mente" del consumatore e determinare quali prezzi attribuire ai propri prodotti, quando proporre sconti esclusivi, come riuscire a catturare l'interesse verso il proprio brand. Naturalmente, più dati si riescono a raccogliere più l'analisi risulta dettagliata e suggerirà modalità di interazione specifiche, massimizzandone le probabilità di successo e riducendone i rischi. Tuttavia, gli utenti sono ormai in gran parte consapevoli dello sfruttamento dei dati personali e quindi cercano di tutelarsi bloccando il tracciamento delle informazioni sulle app, la geolocalizzazione, il consenso ai cookies. In questa direzione agiscono anche le normative vigenti per la tutela dei dati personali, come dettagliato nel prossimo paragrafo.

Il Regolamento Europeo sulla Protezione dei Dati (GDPR)

Come visto, le odierne tecnologie "data-driven" sfruttano costantemente un enorme mole di dati, ponendo numerose questioni sulla violazione della privacy. Pertanto, l'Unione Europea è intervenuta con una legislazione volta a garantire la protezione dei dati, definita come GDPR (General Data Protection Regulation).

Approvato nel 2016, ma con entrata in vigore il 25 maggio 2018, il regolamento GDPR si applica alle aziende (B2B e B2C, profit e no-profit senza distinzioni) e alle persone fisiche che gestiscano dati personali di cittadini europei.

La definizione di dati personali è sancita all'art.4 n.1 come *"qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale."* Pertanto, vengono incluse nella norma informazioni strettamente legate alla persona, come ad esempio il nome, l'indirizzo e-mail, l'indirizzo IP, pseudonimi, dati biometrici e dati clinici, mentre sono esclusi dalla definizione i dati relativi alle aziende o a gruppi di persone.

La legislazione GDPR impone obblighi informativi dell'uso che si intende fare dei dati raccolti (Privacy Policy), specificando, ove necessario, se si intende utilizzare i dati per campagne di e-mail marketing. Il testo dell'informativa deve essere presentato in un linguaggio semplice e comprensibile, senza che vi siano condizioni troppo lunghe e complesse.

Inoltre, prima della raccolta di dati personali, l'interessato deve fornire un consenso *informato ed esplicito* al trattamento dei dati personali. Per "informato" si intende che "*l'interessato dovrebbe essere posto a conoscenza almeno dell'identità del titolare del trattamento e delle finalità del trattamento cui sono destinati i dati personali"*; mentre il consenso "esplicito" prevede che l'individuo manifesti esplicitamente la sua volontà, ad esempio contrassegnando una casella. Quest'ultima, non è consentito venga presentata nel modulo di sottoscrizione come già contrassegnata perché andrebbe a compromettere il principio di "*scelta automaticamente libera*" dell'individuo, così come in tutti i casi di coercizione.

In questo processo di raccolta dei consensi, considerando l'e-mail marketing, il principale dei metodi di marketing, due sono i metodi di iscrizione:

- *Single opt-in*: molto semplice e immediato perché richiede la compilazione del solo modulo di iscrizione e quindi riduce i tassi di abbandono a metà processo, consentendo di raccogliere potenzialmente un bacino più ampio di subscribers.
- *Double opt-in*: richiede una conferma via e-mail all'utente per verificare la correttezza dell'indirizzo di posta elettronica fornito e per dare prova del consenso, con relativa data, orario e modalità di raccolta. Inoltre, il testo dell'e-mail inviata deve specificare per quali funzionalità si stia prestando consenso, ad esempio se per newsletter o per finalità di marketing. Sebbene sia stato evidenziato dalle analisi che il 20% degli utenti abbandoni il processo di iscrizione non completando la verifica attraverso l'e-mail, questo secondo metodo risulta comunque il più utilizzato per chi si occupa di e-mail marketing, in quanto la correttezza degli indirizzi e-mail si rileva come un'importante prerogativa per mantenere alti i tassi di recapito (*deliverability*) e perciò riducendo il numero di bounce e di segnalazioni spam, con impatto positivo sulla reputazione del brand.

Una volta completata l'iscrizione, l'utente può in qualsiasi momento richiedere una copia dei propri dati personali raccolti, in formato elettronico, oppure esercitare il suo diritto alla revoca, chiedendo la disiscrizione attraverso un link che deve essere obbligatoriamente incluso nella e-mail e ben visibile. Inoltre, il processo di disiscrizione deve essere immediato, senza richiesta di accesso con password o di compilazione di sondaggi.

La normativa impone pesanti sanzioni in caso di violazioni, che possono arrivare fino al massimo tra il 4% dei ricavi annui dell'azienda e 20 milioni di euro.

Marketing funnel

Il "marketing funnel" è un modello che rappresenta il "*path to purchase*", ovvero il percorso che il potenziale cliente segue per arrivare alla fase finale di acquisto. Tuttavia, la sua caratteristica forma ad imbuto (da cui il nome in inglese "funnel") rappresenta come la numerosità degli individui che entrino nella prima fase vada ad assottigliarsi progressivamente ad ogni step, portando solo una piccola fetta degli utenti entranti nell'imbuto alla conclusione del percorso, segnata dall'effettivo acquisto del prodotto.

Il marketing funnel è definibile secondo le quattro macro-fasi del modello AIDA:

1. **Awareness**: è il momento in cui un individuo generico acquisisce consapevolezza dell'esistenza di una soluzione ad un proprio problema oppure della presenza di un particolare

brand sul mercato, che cattura la sua attenzione attraverso pubblicità, contenuti nei social networks, ecc…

2. **Interest**: l'utente inizia ad interessarsi a quel prodotto/brand e a raccogliere informazioni; in questo modo passa da "semplice utente" a "lead", ovvero individuo i cui interessi possano farlo considerare un potenziale acquirente.
3. **Desire**: a questo punto, prende in considerazione l'acquisto e pertanto confronta i diversi fornitori per scegliere quello più adatto alle sue esigenze.
4. **Action**: viene concluso l'acquisto con il fornitore scelto e pertanto il lead viene qualificato a "cliente".

Dalla prospettiva dell'impresa, l'obiettivo è quindi innanzitutto quello di riuscire a coinvolgere un ampio bacino di utenti nella prima fase di Awareness, ma, ancor più importante, riuscire a condurre la più elevata percentuale di questi alla vendita.

A questo scopo, il primo passo fondamentale è quello di individuare le "buyer personas", ovvero i clienti target, delineati secondo analisi storiche (se il prodotto non è completamente nuovo) oppure in base alle caratteristiche di profili che potrebbero avere un interesse o bisogno nell'acquisto di quel prodotto. Risulterebbe invece eccessivamente costoso, nonché controproducente, effettuare pubblicità massiva non mirata, in quanto le persone che non hanno nessuna correlazione con il "cliente ideale" percepirebbero quel tipo di advertising come fastidioso e, su grandi numeri, questo causerebbe un danno all'immagine del brand.

La figura 1 mostra un marketing funnel espanso ad un livello di dettaglio maggiore.

Figura 1: Marketing funnel

La parte alta dell'imbuto (TOFU) comprende le sezioni in cui l'utente ha un potenziale interesse per il prodotto, ma non l'ha ancora espresso consapevolmente. In questa fase, è importante per l'impresa catturare gli individui con quest'interesse implicito tramite strategie di SEO[1], SMM[2], PPC (*pay-per-click*) oppure tramite Google AdWords o e-mail targettizzate. In genere sono da considerarsi utenti più interessati quelli che arrivano dalla ricerca di parole chiave piuttosto che coloro che cliccano su un banner, perché la ricerca attiva è una chiara manifestazione di interesse per quello specifico prodotto.

Statisticamente, si può supporre un 10% di CTR (numero di click sull'annuncio rispetto al numero di visualizzazioni), perciò un 10% degli utenti della fase precedente approda sulla cosiddetta "landing page" del sito web dell'azienda, pagina che deve risultare chiara, convincente,

[1] Il SEO (*Search Engine Optimization*), ovvero l'ottimizzazione per i motori di ricerca, consiste in una serie di tecniche che permettono di migliorare il posizionamento del proprio sito web sui motori di ricerca per essere più facilmente in grado di attrarre la visibilità degli utenti, che generalmente si concentrano esclusivamente sui primi risultati mostrati. Queste tecniche si focalizzano sull'attribuire delle *targeted keywords*, pensando a quali potrebbero essere i termini che un lead interessato al brand/prodotto inserirebbe nella barra di ricerca. Un'altra strategia di marketing, simile ed affiancabile al SEO, per ottenere visibilità più immediata, è quella denominata come SEM (*Search Engine Marketing*), che consiste nella pratica *pay-per-click marketing*, ovvero pagare il motore di ricerca per apparire tra i risultati sponsorizzati.

[2] L'acronimo si riferisce al *Social Media Marketing*, ovvero la strategia di creare annunci pubblicitari della propria impresa che compaiono nelle piattaforme di social networks. Un esempio molto efficace è il Facebook Advertising.

immediata e presentare contenuti interessanti, inseriti in un design accattivante, che veicoli una buona immagine del brand.

La parte intermedia del funnel (MOFU) va invece a scindere chi, presa visione della landing page, decide di lasciarla immediatamente, da chi invece mostra interesse, dedicando tempo alla lettura dei contenuti del sito. Per questi ultimi, il processo che segue si definisce come "Lead Generation". Infatti, circa il 20% di questi utenti aderiscono alla compilazione di form di acquisizione dei dati personali (nome, e-mail, recapito telefonico…,), convertendosi da semplici utenti non identificabili in effettivi potenziali clienti (lead), rintracciabili e assoggettabili a tecniche di customizzazione grazie alle informazioni raccolte. Per ciascun lead si avvia, quindi, un processo di "lead nurturing", che consiste in un insieme di strategie dell'inbound marketing volte a creare un rapporto diretto con il lead, personalizzando i contenuti delle campagne e-mail ai suoi interessi, in modo da farlo "affezionare" al brand.

Coloro che rispondono positivamente alla ricezione delle e-mail customizzate, ad esempio cliccando sul sito web linkato per più volte consecutivamente, entrano nella base dell'imbuto (BOFU), dapprima diventando "Prospects", poi, avvenuto l'acquisto, veri e propri "Customers" (circa il 20% dei Prospects), entrando nell'area *Vendite*.

Il processo però non deve concludersi a questo punto. Per essere una strategia a lungo termine non basta il completamento della vendita, ma l'impresa deve essere in grado di crearsi una rete di clienti fidelizzati, attraverso programmi di "*loyalty*" e "*retention*", offrendo ad esempio sconti sugli acquisti successivi o invitando ad acquistare un prodotto migliore del precedente (strategia di "up-selling") o prodotti abbinabili a quello acquistato ("cross-selling"). Inoltre, è di supporto a mantenere alta la reputazione del brand invitare gli utenti a recensire il prodotto online, per creare fiducia negli altri potenziali acquirenti.

Infine, una parte dei customers, coloro che più si sentono legati al brand e soddisfatti dei prodotti offerti, andrà a trasformarsi in "Customer Advocates": veri e propri ambassadors, che possono raccomandare il prodotto attraverso i social networks, blogs o passaparola, e valorizzare l'immagine dell'azienda, creando nuove opportunità di attrarre nell'imbuto nuovi utenti.

Lead scoring

Analizzate tutte le fasi del marketing funnel, emerge quanto sia rilevante la figura del lead per veicolare le opportune strategie di marketing inbound. In particolare, gli obiettivi di "lead generation" e "lead nurturing" hanno creato la necessità di tecniche di segmentazione e prioritizzazione dei lead.

Il cosiddetto “lead scoring” risponde a questo bisogno, strutturando un modello di punteggi, attribuibili in base a caratteristiche demografiche (dati espliciti dell’utente) e dati relativi al comportamento sul web (dati impliciti).

Le caratteristiche demografiche possono riferirsi alla località geografica, al genere, all’età, all’azienda in cui l’individuo opera e ad altre informazioni utili a filtrare i profili che vengono considerati in linea con il prototipo di cliente per uno specifico prodotto. Ad esempio, un’azienda che vende abbigliamento femminile teen andrà ad utilizzare i criteri del genere e dell’età per catturare l’attenzione delle “clienti target”; se il brand opera solo in Europa si aggiungerà anche ai fattori discriminanti la geolocalizzazione. Riferendosi invece a un’impresa che opera nel B2B, potrebbero diventare irrilevanti il genere, l’età e qualsiasi informazione personale, ad eccezione dell’azienda in cui presta attività lavorativa ed il ruolo ricoperto.

I dati comportamentali sul web, invece, sono legati alle azioni che l’utente compie in rete, che implicitamente vanno a designare i suoi interessi, il suo coinvolgimento verso un brand o prodotto e la sua propensione all’acquisto. Si possono considerare, ad esempio, la frequenza di apertura delle e-mail, di visita al sito web (o ad una specifica sezione di questo), di acquisto di prodotti (considerandone la categoria o il brand) e creare, quindi, analisi predittive sui futuri acquisti. Nel caso di un utente che ha acquistato uno smartphone, le campagne e-mail gli proporranno l’acquisto di prodotti correlati (es: cuffiette, cavo di alimentazione, cover...) e, all’uscita di un nuovo modello di quello stesso smartphone, verrà incluso nella campagna e-mail di promozione, magari proponendogli un pre-sale in quanto considerato in cima ai potenziali clienti. All’inverso, un utente che elimina molto frequentemente le e-mail pubblicitarie inviate da una specifica impresa, si rivela poco interessato al brand. Tuttavia, non bisogna eliminare quest’ultimo dalla lista di potenziali clienti, ma mantenerlo comunque presente ad un grado di priorità inferiore, a cui corrisponde un invio di campagne e-mail con minor frequenza per evitare che l’utente si disiscriva alla newsletter.

Alcuni degli strumenti tecnologici utilizzati per tracciare i comportamenti dell’utente sul web sono:

- ***Cookies***: sono piccoli file di testo, che vengono creati automaticamente al momento dell’accesso a un determinato sito, inviati dal server del sito web visitato al browser dell’utente, memorizzati in locale sul computer di quest’ultimo, per poi essere ritrasmessi, alla successiva visita, allo stesso sito. Questi hanno lo scopo di eseguire autenticazioni informatiche, monitoraggio di sessioni e memorizzazione di informazioni specifiche riguardanti l’utente e la sua visita. Più specificamente, esistono diverse tipologie di cookies, tra cui le più utilizzate per funzionalità di marketing sono:

1. *Cookies di profilazione o pubblicitari*: monitorano e profilano gli utenti nel corso della navigazione, studiando e memorizzando le loro opzioni e abitudini di navigazione sul sito e sul web in generale, nonché le abitudini di consumo, anche allo scopo di inviare pubblicità di servizi mirati e personalizzati. Possono quindi essere anche cookies da terze parti, ovvero provenienti da siti terzi rispetto a quello visitato, permettendo alla funzione marketing di avere una profilazione dell'utente molto più ampia rispetto ai soli dati estraibili dal sito proprietario.
2. *Cookies analitici*: sono utilizzati per raccogliere informazioni sull'utilizzo delle varie sezioni del sito e su come gli utenti interagiscano con esso, per analisi statistiche sulla visualizzazione delle pagine, sul numero di visitatori e sul tempo di visita. In questo caso, le informazioni sono raccolte in forma anonima e aggregata, senza conservazione di alcun dato personale degli utenti; pertanto, potrebbero essere utilizzate dal marketing aziendale per capire i prodotti più attrattivi e impostare le adeguate campagne di promozione via e-mail e via social networks.

L'accettazione e il consenso al trattamento delle informazioni a mezzo dei cookies avviene visionando il banner sul sito (in alto), e poi chiudendolo, scorrendo la pagina o cliccando qualunque elemento della pagina medesima. Qualora non avvenga alcuna interazione e l'utente esca dall'informativa chiudendola o proseguendo la navigazione nel sito, si considera prestato il consenso per tutti i cookies.

- ***Link traccianti***: sono dei link URL, nei quali parametri è contenuta l'informazione sul canale di provenienza e altri dettagli specifici, come la campagna marketing in cui è avvenuto il click. In questo modo, l'utente giunge alla "landing page" e i dati contenuti all'interno del link tracciante possono essere registrati a sistema, in modo che l'impresa possa studiare l'efficacia dei propri content media e dei canali digitali. Inoltre, se l'impresa ritiene che una campagna di marketing o un canale di provenienza debbano essere valorizzati maggiormente rispetto agli altri, potrà assegnare un punteggio positivo ai lead che provengono da questi.
- ***Analytic tools***: strumenti di *Analytics Intelligence*, come Google Analytics, che sono in grado di raccogliere dati e svolgere analisi (come, ad esempio, riguardanti il canale che porta ad un maggior grado di conversione), a rilevare anomalie e variazioni improvvise delle tendenze dei dati, a sfruttare algoritmi di machine learning per creare previsioni sui comportamenti degli utenti.

Il sistema di "lead scoring" si basa, quindi, sull'attribuzione di punteggi, opportunamente studiati sulle specificità dei clienti target dell'azienda, in modo che vengano considerati tutti i criteri rilevanti, sia demografici sia comportamentali. In quest'ottica, si può pensare di creare una matrice di lead scoring in grado di rappresentare la segmentazione per livelli dei lead (Figura 2). Le lettere rappresentano il punteggio legato ai dati demografici, considerando A come punteggio più alto e D come il più basso; mentre i numeri individuano il punteggio di "engagement" dell'utente, studiati i suoi comportamenti online (1 per gli utenti più attivi, 4 per quelli che hanno dimostrato minor interesse). Volendo studiare congiuntamente entrambi i fattori di lead scoring, si andranno ad individuare tipicamente tre livelli di segmentazione:

1. **Hot leads**: sono i "sales-ready leads", ovvero coloro che manifestano un interesse ad acquistare nel presente. Questi, infatti, presentano il miglior punteggio per entrambi i fattori (A1) oppure il miglior punteggio in uno dei due e il secondo miglior punteggio nell'altro (A2 e B1).
2. **Warm leads**: mostrano un interesse moderato, che potrebbe condurre ad un acquisto nel breve termine. Questi sono costituiti da punteggio ottimo in un fattore, ma discreto o pessimo nell'altro (C1, D1 e A3, A4) oppure da punteggi intermedi per entrambi i fattori (B2, B3, C2).
3. **Cold leads**: viene rilevato uno scarso interesse, è probabile che tendano ad acquistare nel lungo termine. Raggiungono un punteggio pessimo in uno dei due fattori (D2, D3, B4, C4), discreto in entrambi (C3) o pessimo in entrambi (D4).

A1	A2	A3	A4
B1	B2	B3	B4
C1	C2	C3	C4
D1	D2	D3	D4

Figura 2: Matrice di lead scoring. Fonte: Alvin, 2017

Il risultato di questo metodo di segmentazione (o simili) è quindi quello di riuscire a strutturare dei percorsi-cliente customizzati secondo il grado di lead scoring, guidando le azioni che il marketing deve intraprendere per essere efficace, senza spreco di risorse. Questo sistema di prioritizzazione, infatti, consente di capire su quali lead focalizzarsi, riducendo complessivamente il "costo per ciascun lead" e quindi risultando in maggiori ricavi e un più elevato ROMI (return on marketing investment). Inoltre, il processo migliora la relazione con il cliente, la sua soddisfazione e la sua fidelizzazione e consente di supportare il decision making, guidando anche le altre funzioni aziendali attraverso informazioni quantificabili e misurabili in tempo reale.

Di seguito, la tabella 1 suggerisce un modello di strategie di marketing basata sulla suddivisione dei lead come da matrice riportata sopra. Come accennato in precedenza, per i leads considerati "hot" le azioni proposte sono volte al processo di "conversione" che porta il lead ad essere qualificato come "pronto per le vendite", perciò si cerca un contatto più diretto telefonicamente o si avanza una proposta per incoraggiare all'acquisto. Per i lead a più basso scoring, invece, il processo che guida all'acquisto è più lungo ed è denominato "lead nurturing", in quanto è volto a creare nel cliente via via maggior engagement. In questo caso, la frequenza delle campagne di e-mail viene aumentata progressivamente nelle fasi in cui si vede crescere il punteggio del lead e quindi il suo interesse. Per riuscire in questo intento, l'intelligenza artificiale deve essere in grado di catturare i gusti dell'utente, in modo che le e-mail inviate presentino contenuti perfettamente customizzati, che perciò non infastidiscano il destinatario, ma creino una curiosità sempre crescente verso quel brand o prodotto. Presentando solo contenuti di interesse, il lead mostrerà un maggiore coinvolgimento e percepirà la comunicazione digitale in modo più stimolante. Questi risultati si traducono, quindi, in un miglior "tasso di conversione" e "tasso di mantenimento".

Tabella 1: Le strategie di marketing da seguire secondo il lead scoring. Fonte: Sperling,2017

Lead Score	Description	Marketing Action
A1	Target fit, high interest	Send email that encourages the lead to leave phone number or to call in directly, or propose to purchase the product directly.
A2	Target fit, medium interest	Send offer of free trial of the service or propose relevant material that is close to purchase of the product.
A3	Target fit, low engagement	Priority lead that need further nurturing and "why now"-message.
B1	Potential fit, high interest	Send email that encourages the lead to leave phone number or to call in directly, or propose to purchase the product directly.
B2	Potential fit, medium interest	Continue to nurture with marketing materials that can increase interest, send offer of free trial of the service. Pursue information to evaluate if it is a good fit.
B3	Potential fit, low engagement	Send nurturing content that can create a demand for the product. Pursue information to evaluate if it is a good fit.

Naturalmente, l'assegnazione di un utente ad una specifica cella della matrice viene effettuata in modo automatico e dinamico: in base alle azioni effettuate in risposta alle campagne e-mail (apertura, click link associato, cancellazione…) il lead si sposta in orizzontale in relazione all'incremento o decremento del suo punteggio. Non solo: anche le informazioni personali possono mutare nel tempo. Si pensi ad esempio all'avanzare dell'età che può determinare, a seconda dei casi, un avvicinamento

o un allontanamento dalla fascia d'età target. Allo stesso modo, anche il trasferimento da un paese ad un altro o un cambiamento della posizione lavorativa possono comportare spostamenti in verticale da una cella a un'altra. Nuovamente, si percepisce l'importanza di sistemi digitali automatizzati che siano responsivi ai cambiamenti in tempo reale. In quest'ottica, anche il sistema di lead scoring, una volta creato, non deve essere lasciato operare senza un continuo monitoraggio, bensì occorre provvedere agli opportuni cambiamenti per ottimizzarlo secondo quanto appreso o secondo i mutamenti di contesto.

Metodi di costruzione di un modello di lead scoring: approccio manuale versus metodi di AI

Come visto, il lead scoring è, quindi, uno strumento che guida la funzione marketing e sales a concentrare le proprie risorse sui lead che hanno maggior probabilità di portare a termine con successo l'acquisto. Pertanto, è fondamentale un'adeguata costruzione del modello per evitare di incorrere in valutazioni errate, che conducano i team a focalizzarsi su lead, che difficilmente porteranno a compimento l'acquisto, trascurando, invece, coloro che realmente avrebbero dovuto essere "Hot", portando a uno spreco di risorse, nonché un elevato costo opportunità.

Esistono diversi metodi per la costruzione di un modello di lead scoring all'interno di un'impresa, i quali possono consistere in approcci più tradizionali, come il metodo "manuale", o più innovativi, come procedure che sfruttano algoritmi di intelligenza artificiale.

Il metodo manuale può essere elaborato attraverso le seguenti fasi:

1. Analisi del business di riferimento dell'impresa e della sua pipeline di vendita: ogni settore di mercato prevede logiche e procedure di acquisto specifiche, a cui non si può prescindere per comprendere le caratteristiche del lead più significative nello stimare la sua propensione all'acquisto.
2. Preparazione dei dati: è necessario uniformare tra loro i dati, ad esempio riportando alla stessa dicitura dati dove un valore è scritto in maniera diversa pur riferendosi allo stesso significato (ad esempio quando ci sono abbreviazioni) ed esaminare attentamente se vi siano dei valori mancanti o sicuramente non corretti.
3. Analisi dei dati esplorativa: una volta selezionati gli attributi del lead dalla fase 1 e accertata la qualità dei dati dalla fase 2, si effettuano delle analisi sui dati storici registrati a sistema, in modo da catturare le tendenze dei clienti correnti e passati e poter comprendere quali valori assegnare ad ogni attributo.

4. Assegnazione dei punteggi: le considerazioni emerse dalle analisi esplorative e le esperienze dirette di business aziendali si coniugano nella creazione di una tabella di punteggi, che premia con punteggio positivo le caratteristiche del lead che possono farlo tendere all'acquisto con maggiore probabilità; mentre, penalizza con un punteggio negativo quelle con un pessimo impatto sulla tendenza ad acquistare. Se ne fornisce un esempio nella tabella 2. La somma algebrica dei punteggi costituisce lo score di quel lead.

Tabella 2: Tabella di punteggi esemplificativa. Fonte: Duncan and Elkan (2015)

Behavioral attributes	Value
Filled out a contact form	+10
Visited careers page	-5
Demographical attributes	**Value**
Job title is "student"	-10
Company is located in Northeast USA	+5

Sebbene l'approccio manuale sia il più semplice ed il più utilizzato nelle aziende, Duncan ed Elkan (2015) sostengono, tuttavia, che questo ha elevato rischio di errore umano, derivante dalle considerazioni che il team di marketing e di vendite formulano sulla base delle proprie esperienze aziendali o su sensazioni e intuizioni personali, che potrebbero rivelarsi errate nella realtà dei fatti. Utilizzare, invece, modelli predittivi basati su algoritmi di machine learning risulta essere una metodologia molto più oggettiva e data-driven. Infatti, proprio a partire dai dati del database del CRM, vengono elaborate analisi predittive in grado di individuare le caratteristiche comuni ai lead che si sono convertiti in customers e quelle che invece hanno accomunato i lead che non hanno mai portato a compimento l'acquisto. Il modello è, quindi, un lead scoring automatico, che dalle informazioni del lead in input riesce a prevedere la probabilità che si converta in cliente. Tra gli algoritmi di analisi predittiva i più utilizzati nella modellazione del lead scoring sono l'albero decisionale, il Random Forest, il Gradient Boosted Tree e la regressione logistica.

Albero decisionale

L'albero decisionale è l'algoritmo di più semplice costruzione e permette di usare dati di input sia di tipo qualitativo sia di tipo numerico nel discreto; tuttavia, può presentare maggiori problematiche nel caso di campi numerici continui o di attributi con dati mancanti oppure con un set di molteplici opzioni. Partendo dai dati di input, l'albero decisionale è in grado di creare ramificazioni if/else su ogni attributo che, seguendo i diversi percorsi, porteranno ai nodi terminali, che forniscono la

classificazione del lead. In particolare, per un numero N di dati di input è possibile creare N! possibili permutazioni di albero decisionale: per scegliere quello ottimale si procede settando nei nodi più alti dell'albero gli attributi che presentano una (presunta) relazione causa-effetto certa, ovvero tutti i dati che per quell'attributo X hanno registrato il valore *x* allora hanno ottenuto come output Y (nel caso del lead scoring il valore booleano "Yes" per la domanda "Ha portato a compimento l'acquisto?"). In caso nessuno dei dati presenti una relazione di causa-effetto "pura", si procede con una specifica "regola di splitting", ad esempio, si può misurare l'indice di impurità di Gini per ogni nodo figlio di ogni attributo, calcolato con la seguente formula:

$$I_G(t) = 1 - \sum_{i}^{m} (p_i)^2$$

dove *t* rappresenta il nodo figlio, *m* è il numero di classi possibili per *t*, mentre p_i è la frequenza relativa di dati della classe *i* per quel nodo figlio. Successivamente, si calcola l'indice di impurità di Gini per l'attributo X come media pesata degli indici di Gini dei suoi nodi figli. L'attributo che presenta una minore impurità viene scelto come nodo radice, ovvero quello di partenza da cui si ramificheranno le decisioni. Si procederà quindi in maniera ricorsiva, a partire da ogni ramo intermedio, sempre determinando l'impurità per ciascun nodo come precedentemente spiegato, ma tenendo conto del vincolo di provenienza del nodo da cui si sta generando. Ad esempio, se la condizione scelta come nodo radice è *"Età minore di 30 anni?"* e si sta analizzando la condizione *"È uomo?"* la frequenza relativa deve essere calcolata in base al nodo in cui ci si trova: se questo proviene dal ramo "Yes" in risposta alla prima condizione si considereranno solo *i=uomo che ha meno di 30 anni* oppure *i= donna che ha meno di 30 anni*; al contrario se ci si trova nel ramo "No" solo *uomo/donna con più di 30 anni*. In questo modo, nuovamente, dopo il calcolo della media pesata, si sceglie il nodo che ha indice di Gini inferiore. Si procede in questo modo, con l'accortezza di individuare correttamente i nodi foglia (o nodi terminali). Questi vengono scelti secondo una specifica "regola di stop" settata. Ad esempio, prendendo l'indice di Gini, si possono considerare come nodi terminali quelli con impurità minore rispetto ai nodi dello stesso livello: in questo modo si chiuderanno i nodi considerabili come "i più discriminanti", ottenendo un albero decisionale che non risulti troppo complesso a causa di eccessive ramificazioni.

Una volta costruito l'albero decisionale, viene impostata l'etichetta dei nodi finali come la classe che ha registrato la maggioranza dei casi. Se i dati hanno identica frequenza nelle classi, invece, allora l'etichetta del nodo non è assegnabile. Un esempio di albero decisionale è riportato in figura 3.

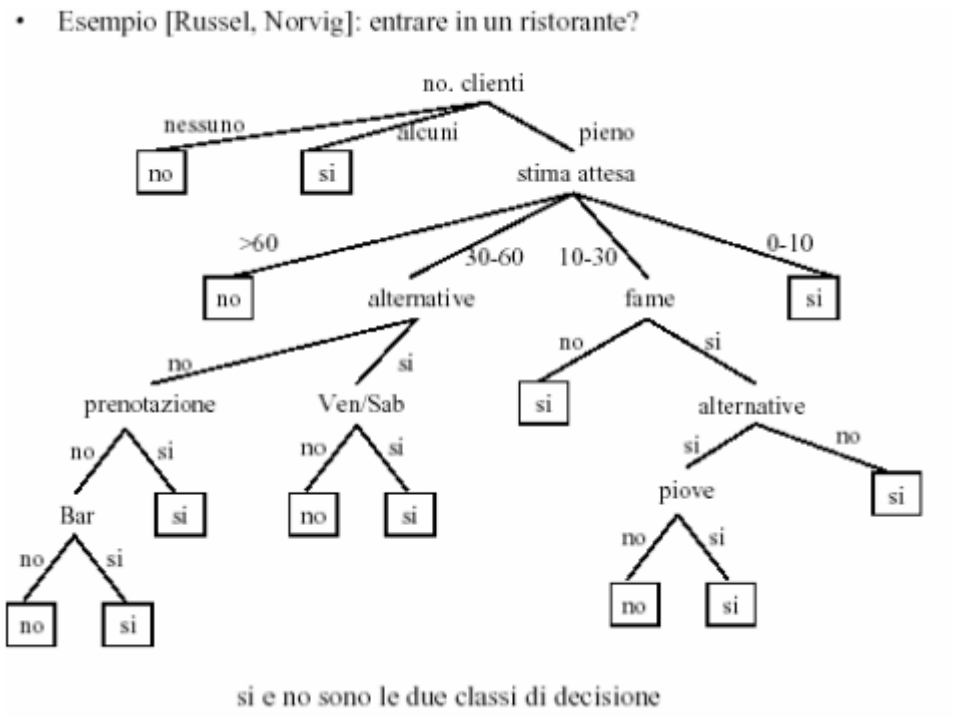

Figura 3: Esempio di albero decisionale

L'algoritmo di costruzione di un albero decisionale, anche detto algoritmo CART (*Classification And Regression Trees*), nel caso di alberi molto complessi, può presentare il problema dell'overfitting dei dati, ovvero riesce ad avere ottime performance per i dati usati per la sua creazione, ma non ha flessibilità e robustezza davanti a nuove osservazioni, portando a cattive predizioni. Per ridurre questo effetto, è necessario adottare l'algoritmo di Pruning, che consiste nella "potatura" dell'albero: permette di riconoscere i nodi meno significativi nella classificazione e di tagliarli, ottenendo un albero più snello.

Random Forest

Pur mantenendo la stessa semplicità degli alberi decisionali, l'algoritmo Random Forest ha il vantaggio di superare le problematiche di questi, presentando maggior flessibilità e accuratezza nel calcolo predittivo per i nuovi inserimenti di dati.

Come primo passo, l'algoritmo estrae dal dataset un sottoinsieme, definito come "bootstrap sample", che consiste in osservazioni scelte in modo randomico (possono comprendere anche record duplicati).

A partire da questo dataset ridotto, si costruiscono più alberi decisionali, nei quali ad ogni iterazione si sceglie di considerare un numero $m \leq p$ delle variabili di input[3] come candidate per lo split, che

[3] Per capire quale sia il numero adatto da scegliere per le variabili si può procedere per tentativi e selezionare in base alle performance migliori.

vengono scelte in modo casuale. In questo modo, da ogni albero risulterà una certa classe predittiva; pertanto, la classe con maggior frequenza tra tutti gli alberi decisionali diviene la classe predittiva del modello. Si veda un esempio nelle figure 4 e 5.

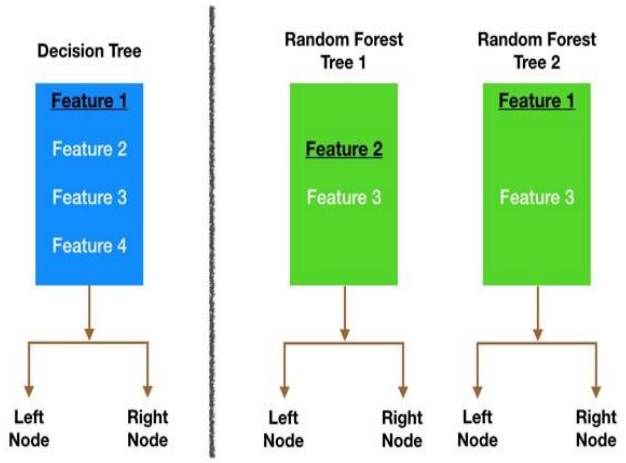

Figura 4: La differenza tra albero decisionale semplice e Random Forest

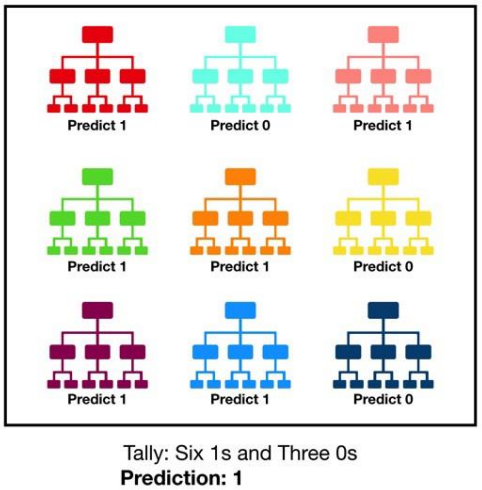

Figura 5: Trovare la classe predittiva del modello

Le ottime predizioni derivano proprio dal concetto "wisdom of crowds", ovvero dall'evidenza empirica che mettendo insieme alberi che non hanno nessuna correlazione tra loro si ottengano classificazioni predittive migliori di quelle che si svilupperebbero con i singoli alberi individuali.

Un ulteriore enorme vantaggio consiste nella capacità dell'algoritmo di riuscire a trattare con buone performance predittive anche i casi di osservazioni con dati mancanti. Infatti, è possibile stimare i valori dei dati mancanti attraverso una "matrice di prossimità", che permette di capire quali osservazioni sono più "simili" tra loro per poter completare i missing values di un'osservazione con i dati osservati nell'osservazione di maggiore prossimità.

Grazie a queste caratteristiche, l'algoritmo di Random Forest è forse la tecnica di *predictive analytics* più utilizzata in moltissimi altri ambiti oltre al lead scoring, quali, ad esempio, nel settore bancario, nelle diagnosi mediche, nel campo bioinformatico.

Gradient Boosted Trees

Anche l'algoritmo Gradient Boosted Trees, come il Random Forest, sfrutta il concetto che l'insieme di più *weak learners* (ovvero predittori deboli, estratti dagli alberi decisionali presi individualmente) può costituire uno *strong learner*, che ha accuratezza predittiva decisamente migliore rispetto ai singoli. L'algoritmo può essere applicato per predire una classificazione, come nel caso del lead

scoring, per cui l'obiettivo è prevedere se il lead si trasformerà in customer. Partendo da dati storici (training set), una prima predizione di classificazione è data da

$$L_0 = \log\left(\frac{\#\ lead\ convertiti}{\#\ lead\ non\ convertiti}\right)$$

in quanto, essendo l'output un valore booleano ("Yes" per avvenuta conversione, "No" in caso contrario), la funzione log(Odds) si presta bene alla stima. Per maggiore chiarezza, si può tradurre questo valore in probabilità di acquisto, utilizzando la trasformazione con la funzione logistica, ovvero $p = \frac{e^{L_0}}{1+e^{L_0}}$. A questo punto, per migliorare la predizione, si calcola per ognuno dei dati presenti nel training set il valore di residuo, ovvero la differenza tra il valore reale (1 per "Yes", 0 per "No") e il valore predetto *p,* e si costruisce l'albero decisionale che abbia come classe sui nodi foglia il valore $\frac{\sum Residuo}{\sum p(1-p)}$ (dove la sommatoria viene considerata solo nei casi in cui più osservazioni cadano all'interno dello stesso nodo foglia). Calcolati i valori di output per ogni nodo terminale, si può costruire una predizione migliorativa per ciascuna osservazione del training set, sommando a L_0 il valore del nodo foglia corrispondente moltiplicato per un "learning factor" (un numero da 0 a 1 che corregge l'effetto di overfitting). Questa più accurata previsione per ognuna delle osservazioni dovrebbe avvicinarsi maggiormente al valore di output effettivo; pertanto, i valori di residuo calcolati con la nuova funzione predittiva dovrebbero assottigliarsi. Lo stesso procedimento, a partire da questi nuovi residui, può essere iterato più volte, mediante la costruzione di alberi decisionali, ottenendo una funzione L_n (con n numero di iterazioni) che minimizza i residui di ciascuna delle osservazioni. In questo modo, per un nuovo lead entrante nel sistema si può prevedere la probabilità di acquisto calcolando, a partire dalle sue informazioni, L_n e trasformandolo attraverso la funzione logistica:

$$Probabilità\ di\ acquisto = \frac{e^{L_n}}{1+e^{L_n}}$$

Con la procedura di "bagging" introdotta da Breiman, l'algoritmo presenta una piccola variazione che incrementa notevolmente la sua accuratezza, eliminando il problema dell'overfitting: infatti, come per l'algoritmo di Random Forest, anche nel caso di Gradient Boosted Trees è possibile incorporare una componente di casualità, ovvero estraendo dal training set non tutte le osservazioni, ma un sottoinsieme ottenuto in modo randomico. In questo modo, l'algoritmo risulterà più robusto e meno collegato ai dati del training set. Per l'aggiunta di questa componente stocastica, il metodo con questa variante è definito come "Stochastic Gradient Boosted Trees".

Regressione logistica

La regressione logistica è un metodo di regressione applicato nei casi in cui la variabile dipendente è binaria, ovvero può assumere solamente i valori 1 o 0 ("Yes" o "No", nel caso del lead scoring riferito al lead che si è convertito in customer, come spiegato nel paragrafo precedente). In quest'ottica, i dati sono localizzati all'interno del dominio [0,1] e, in particolare, si presenteranno proprio in corrispondenza di questi due valori; pertanto, non potrebbero essere interpolati da una retta, che, invece, cresce in maniera lineare e può assumere un dominio (− ∞, + ∞). Infatti, la funzione che meglio interpola i dati ha la forma non lineare di una curva ad S, come mostrata in figura 6, dove le ordinate rappresentano le probabilità di acquisto del lead (ovvero la probabilità che la variabile indipendente assuma il valore 1) in relazione ad una variabile x indipendente, chiamata "variabile esplicativa".

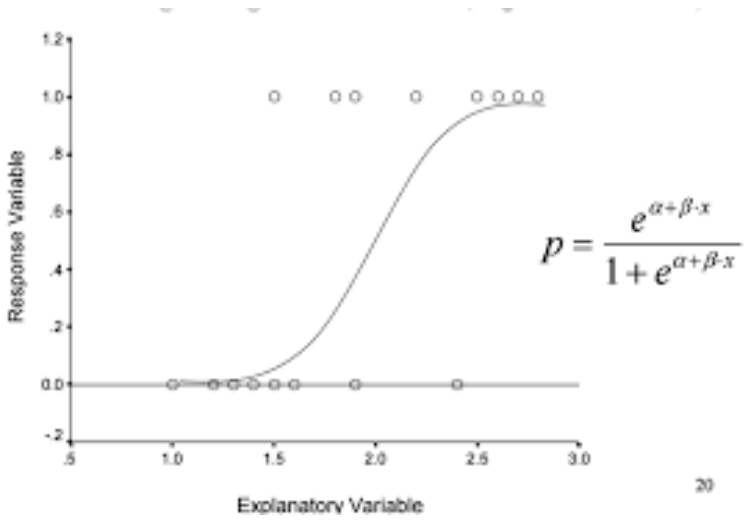

Figura 6: Curva ad S della funzione logistica

La formula riportata accanto al grafico esprime la probabilità di acquisto, calcolata attraverso la funzione di ripartizione logistica, valutata in $z = \alpha + \beta x$. È possibile ovviamente estendere l'analisi a più variabili esplicative, ognuna con il proprio coefficiente β_i. Attraverso software di analisi econometrica (ad esempio "STATA"), è possibile, a partire dai dati storici, effettuare una stima dei coefficienti delle variabili esplicative, attraverso il metodo della massima verosimiglianza, per poterne discernere l'effetto sulla variabile indipendente. Inoltre, viene effettuato un test del *p-value*, che rivela la significatività di ciascuno dei coefficienti stimati, in modo da escludere dalla funzione di predizione le variabili con coefficiente non significativamente diverso da zero. Pertanto, una volta ottenuta la funzione, completa di coefficienti stimati, è possibile inserire le informazioni di un nuovo

lead entrante nel sistema assegnandone il valore nelle variabili esplicative e, quindi, diviene immediato calcolare la sua probabilità di acquisto.

Indicatori di prestazione per la scelta del metodo di machine learning da adottare per il lead scoring

Per la costruzione di un sistema di lead scoring si possono impiegare tutti i metodi di machine learning sopra citati, ma si otterrebbero modelli predittivi diversi a seconda della tecnica impiegata; pertanto, per scegliere l'algoritmo con migliori performance predittive è necessario utilizzare un set di dati di test (diverso dal training set con cui sono stati costruiti i modelli) e misurare, attraverso, specifici indicatori, le performance di ciascuno nella predizione di quelle osservazioni. In particolare, gli indicatori più applicati in questa valutazione sono costruiti a partire dalla "Matrice di confusione", mostrata in figura 7.

Confusion matrix for binary classification			
Actual value	A	TP	FN
	B	FP	TN
		A	B
		Predicted value	

Figura 7: Matrice di confusione.

La matrice di confusione permette di visualizzare in modo immediato quanto le predizioni della classificazione (Predicted values) stiano fittando con le osservazioni effettive della realtà (Actual Values). Infatti, essendo la risposta in output una classificazione binaria, si costruisce una matrice 2x2, che riporta nelle colonne il numero di predizioni della risposta A (che corrisponde ad output positivo, quindi, nel nostro caso a "Yes"[4]) e della risposta B (output "No"); mentre nelle righe la misurazione reale di coloro che hanno registrato la classificazione A o la classificazione B. Pertanto, le predizioni corrette del modello sono contenute nella cella AA, in quanto il modello aveva predetto una risposta positiva e questa si è verificata realmente, quindi, è definibile come numero di *True Positive* (TP), e nella cella BB, che sarà rappresentativa del numero dei *True Negative* (TN), ovvero predizioni negative che sono poi state confermate. D'altra parte, le predizioni errate si rilevano dalla

[4] La domanda posta in esame è sempre se il lead si convertirà a customer, ovvero se la sua probabilità di acquisto è approssimabile ad 1.

cella BA, che rappresenta il numero di *False Positive* (FP), per i quali era risultata dal modello predittivo una risposta positiva, ma che non si è poi dimostrata tale, e dalla cella AB dei *False Negative* (FN), per i quali si è verificato output positivo, nonostante la predizione dell'algoritmo fosse negativa.

Da queste considerazioni, è quindi immediato calcolare l'indicatore di *Accuracy*, che rappresenta la percentuale di osservazioni correttamente predette dal modello rispetto al totale delle osservazioni, ovvero

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

e il complementare a questa frazione, ovvero il tasso di errore del modello:

$$Error\ Rate = \frac{FP + FN}{TP + TN + FP + FN}$$

Nonostante l'accuracy, data la sua facilità di calcolo e di interpretazione, sia la metrica più utilizzata per valutare le capacità predittive di un modello, essa può portare a considerazioni completamente errate se si tratta con un dataset sbilanciato *(imbalanced dataset*). In questi casi, infatti, i dati sono affetti da una forte sproporzione di distribuzione nelle classi: si supponga, ad esempio, in scala di 1 positivo: 100 negativi. Se si ipotizza che nei dati 100 lead abbiano portato a compimento l'acquisto, mentre 10.000 no, questa sproporzione può portare a calcolare valori di accuracy che considererebbero il modello come ottimo nella predizione, anche se questo non fa altro che restituire come previsione di output la classe che presenta la maggior frequenza. Perciò anche nel caso di un modello pessimo che non riesce a predire nessun positivo, ma restituisce sempre l'output negativo, l'accuracy paradossalmente ha il valore $\frac{10.000}{10.100}$= 0,99, che è interpretabile erroneamente con la valutazione di un modello con ottime performance predittive, che fornisce valutazioni corrette nel 99% dei casi. Per evitare di far emergere queste considerazioni fuorvianti è necessario bilanciare i dati prima del calcolo dell'accuracy, o, in alternativa, utilizzare altri indicatori, che studino separatamente le due classi.

La coppia di indicatori *sensitivity* e *specificity* restituisce una misura, rispettivamente del tasso dei positivi correttamente predetti e dei negativi correttamente predetti, come si vede dalle formule per il loro calcolo:

$Sensitivity = \frac{TP}{TP+FN}$ rappresenta la frazione di positivi che erano stati individuati dal modello predittivo rispetto ai positivi reali (che quindi includono anche coloro che sono stati classificati come negativi erroneamente).

$Specificity = \frac{TN}{TN+FP}$ rileva la percentuale di negativi rilevati dal modello rispetto ai negativi effettivi (ovvero compresi i falsi positivi).

In base all'obiettivo per il quale si sta effettuando la classificazione, può essere più rilevante considerare un modello con un elevato valore di sensitivity se è cruciale che venga riconosciuta una percentuale di positivi che si approssimi a quelli reali; piuttosto che, al contrario, in altri casi, si voglia puntare sulla specificity per catturare correttamente i negativi. Nel caso del lead scoring l'obiettivo principale è quello di individuare i lead più promettenti per la vendita; perciò, è essenziale riconoscere i positivi per poter sfruttare efficientemente le risorse aziendali. Infatti, un modello che non riconosce i lead che sono motivati all'acquisto può comportare che questi lead non vengano considerati e passino alla concorrenza, mentre un errore di classificazione per i lead non interessati può comportare uno spreco di risorse, ma ha comunque minor impatto rispetto al costo opportunità derivante da una mancata vendita.

Un'altra coppia di metriche spesso utilizzata è la *precision-recall*. La precision (anche denominata *Positive Predictive Value*, PPV) è la misura di precisione che il modello ha nel predire i positivi: $Precision = \frac{TP}{TP+FP}$, ovvero la frazione di positivi predetti correttamente rispetto al totale dei positivi predetti (comprendendo, quindi, i falsi positivi). La recall, invece, equivale alla sensitivity, ovvero la percentuale di reali positivi che è stata predetta dal modello. Pertanto, la differenza tra le due metriche consiste nel fattore a denominatore e, in base al contesto di riferimento, potrebbe essere preferibile avere un'alta precision, anche a costo di una bassa recall, o il viceversa. Questo dipende dalla considerazione se il modello debba evitare di riportare *falsi positivi*, pur tollerando i *falsi negativi* (come nel caso della precision) o il viceversa (recall). Nella classificazione del lead scoring, come detto, i *falsi positivi* sono accettabili perché comportano un piccolo costo in termini di risorse, mentre è di radicale importanza individuare i positivi, senza incorrere in errate valutazioni di *falsi negativi*, che porterebbero alla perdita di clienti, quindi è desiderabile una recall alta, a discapito di una bassa precision.

Se, invece, si vuole catturare l'effetto congiunto delle due metriche per verificare che entrambe siano elevate, un indicatore appositamente creato allo scopo è *F1-score*, calcolato come media armonica della precision e della recall. La media armonica, infatti, ha la proprietà di pesare maggiormente i valori piccoli: in questo modo il punteggio F1 sarà elevato solo se entrambi gli indicatori sono grandi.

Infine, un altro indicatore, derivato dalla sensitivity (*True Positive Rate*) e dal complementare della specificity (1- specificity, ovvero *False Positive Rate*) è la curva ROC (*Receiver Operative Characteristic*). La curva ROC, costruita come in figura 8, rappresenta il valore delle due metriche (1-specificity nell'asse x, sensitivity sull'asse y) al variare della soglia di classificazione. Infatti, se si prende in considerazione il metodo della regressione logistica, questo fornisce come output predittivo una probabilità, ovvero un numero compreso tra [0,1]; pertanto, è necessario stabilire una soglia al di sotto della quale la classificazione è negativa e al di sopra della quale è positiva. Per scegliere la soglia ottimale, è possibile guardare alla curva del ROC, che riassume i valori della matrice di confusione di ogni soglia, per vedere se il punto individuato da una certa soglia di classificazione è migliore di quello individuato dalle altre possibili soglie, in base a quale percentuale di *True Positive* si vuole raggiungere e quale di *False Positive* si può tollerare.

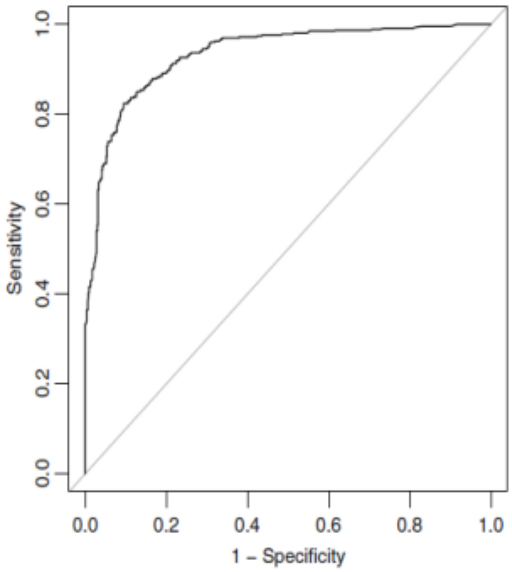

Figura 8: Esempio di curva ROC. Fonte: Kuhn & Johnson (2013)

Dalla costruzione della curva ROC è possibile determinare una misura di accuratezza del modello attraverso l'indicatore AUC (*Area Under The Curve*), che calcola l'area sottostante alla curva e quindi può assumere valori compresi tra 0,5 (caso estremo di ROC sulla bisettrice) e 1 (caso estremo opposto di ROC come curva che assume i valori (0;y) oppure (x;1)). Naturalmente, maggiore è l'area sottostante la curva, maggiore è il potere discriminatorio del modello: con AUC=0,5 il modello non sta fornendo alcuna informazione, con AUC=1, invece, discrimina perfettamente. Generalmente, tra i valori intermedi, si considera come moderatamente accurato un modello con AUC≥0,7.

Pertanto, utilizzando gli indicatori di performance riportati, può essere effettuato un confronto tra le varie metodologie di predictive analytics (Albero Decisionale, Random Forest, Gradient Boosted Trees, Regressione Logistica) per avere evidenza di quale tecnica rilevi le migliori prestazioni

predittive e, quindi, scegliere il modello associato per prevedere la classificazione di ciascun nuovo lead.

KPIs da misurare per migliorare la gestione dei lead

Successivamente alla strutturazione e implementazione del sistema di lead scoring, è indispensabile per un'impresa riuscire a misurare le proprie prestazioni per comprendere se essa stia operando nella giusta direzione per il raggiungimento degli obiettivi aziendali. A questo scopo, è utile scegliere i giusti KPIs (Key Performance Indicators), indicatori chiave per la misurazione delle performance aziendali. La costruzione dei KPIs idonei presuppone la definizione di un obiettivo strategico da raggiungere, il quale deve essere determinato soddisfacendo le caratteristiche del metodo SMART:

- **Specific:** deve inquadrare uno scopo specifico, puntuale, non generico. Ad esempio, l'aumento delle vendite in un arco temporale di sei mesi per il prodotto X.
- **Measurable:** deve essere misurabile in maniera oggettiva, quantitativa.
- **Achievable:** deve essere raggiungibile, non esageratamente pretenzioso da rimanere una mera utopia.
- **Realistic:** deve rimanere in linea con le realtà dell'impresa e del tipo di business.
- **Time-bound:** deve collocarsi in un arco temporale ben definito.

Avendo chiaro l'obiettivo a cui tendere, si passa quindi all'identificazione dei KPIs più adatti alla misurazione di questo. In genere, per ogni obiettivo si scelgono massimo 3 KPIs in quanto avere troppi indicatori potrebbe risultare confusionario; pertanto, la definizione di pochi, ma rilevanti, indicatori permette una visione essenziale ed immediata dei risultati in funzione dell'obiettivo da raggiungere. I KPIs dovranno, quindi, individuare i fattori critici di successo per l'obiettivo proposto e fornire rilevazione di dati che riflettano la situazione reale dell'impresa e siano in grado di guidare le decisioni. Di seguito sono riportati alcuni degli indicatori più utilizzati per un'ottimizzata gestione dei lead:

- ***Conversion rate***, ovvero il tasso di conversione: rappresenta la percentuale di lead che si sono convertiti rispetto al totale dei lead e traduce in dato effettivo quanto i team di marketing e di vendita siano efficaci nell'obiettivo di trasformare i lead in customer. Se il conversion rate risultasse al di sotto delle soglie stabilite, è necessario rivedere l'intero funnel di vendita per proporre strategie di miglioramento.
- ***Cost per lead***: fornisce una stima del costo che l'azienda sostiene per l'acquisizione di ciascun lead, calcolabile come $\frac{Costo\ per\ la\ campagna}{\#\ Leads\ acquisiti}$.

- ***Numero di lead***: indicatore fondamentale per monitorare l'efficacia della fase di Lead Generation.
- ***Ricavo medio per ogni lead***: rappresenta il ricavo medio di ogni lead in uno specifico arco temporale e viene calcolato come $\frac{Valore\ Totale\ delle\ Vendite}{\#\ Totale\ di\ lead}$. Bisogna assicurarsi che questo indicatore sia sempre maggiore rispetto al costo per ogni lead per poter garantire margini di guadagno.
- ***Return On Marketing Investment (ROMI):*** la frazione $\frac{Ricavi - costi\ del\ marketing}{Costi\ del\ marketing}$ esprime i ritorni netti sugli investimenti in marketing. Può essere formulata anche per una singola campagna, per misurarne il rendimento economico e confrontarlo con quello di altre campagne.
- ***Click-through rate***: esprime la percentuale di click rispetto al totale delle interazioni ad un contenuto presente in una e-mail, sul sito web dell'impresa o su un post sui social networks. Permette di valutare l'effetto che la campagna sta avendo sul pubblico che ne entra in contatto e quindi poter studiare quali strategie risultano più attrattive sul consumatore.
- ***Sales Cycle Time***: indica il tempo che mediamente intercorre tra la creazione del lead e la conclusione della fase di acquisto, permettendo quindi una stima delle tempistiche che occorrono per ogni opportunità di vendita, nell'ottica di portare a ridurre l'indicatore, ovvero abbreviare i tempi necessari per le vendite.
- ***Leads per Channel***: mostra il numero di lead acquisiti in ciascun canale (e-mail marketing, sito web, social networks…) per poter comprendere quale sia il canale che genera più lead e focalizzare su questo le strategie di investimento.
- ***Traffico organico sul sito web***: analizza il numero di visitatori del sito web dell'azienda o, più nello specifico, i visitatori di ciascuna pagina del sito web per capire su cosa si concentra l'interesse dei potenziali acquirenti e porsi obiettivi di miglioramento di questa metrica, adottando tecniche di SEO più avanzate o maggiore pubblicità sui social.
- ***Sessione media***: non basta considerare il numero di visitatori del sito per estrarne informazioni sulle preferenze dei consumatori, ma occorre anche considerare il tempo medio trascorso su ognuna delle pagine per poter valutare l'efficacia dei contenuti proposti e il livello di *user experience*.
- ***Tasso di ritenzione e di abbandono:*** il primo indica la percentuale di clienti che, soddisfatta dal primo acquisto, decide di rimanere cliente dell'impresa; mentre, al contrario, il tasso di abbandono indica la percentuale di clienti che, dopo un primo acquisto, hanno deciso di non

riacquistare prodotti dell'azienda. Per il calcolo di queste metriche viene stabilito un arco temporale ben definito.

- ***Net Promoter Score (NPS)***: attraverso sondaggi e recensioni è possibile catturare il grado di soddisfazione dei clienti e categorizzarli secondo questo come segue:

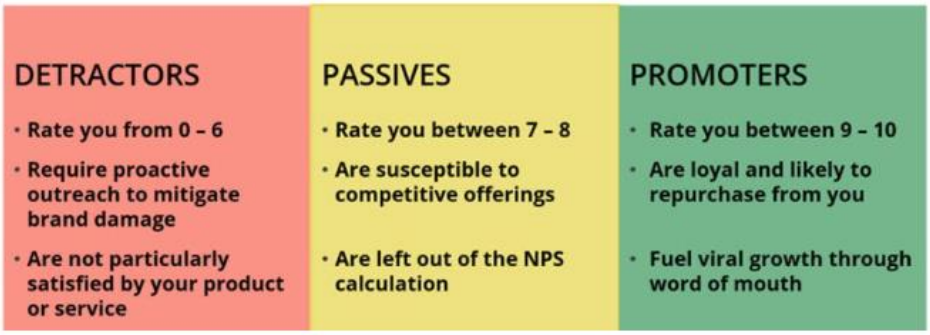

Figura 9: Suddivisione dei clienti secondo il rating che hanno attribuito nei sondaggi

In quest'ottica il *Net Promoter Score* è la probabilità che un cliente raccomandi il prodotto dell'azienda, ad esempio per passaparola a conoscenti oppure postando contenuti sui social networks. Il calcolo di questa metrica avviene con la seguente formula:

$$NPS = \frac{\#Promoters - \#Detractors}{\#\,Total\ Respondents}$$

Prospettive di ricerche future

Se fino ad ora la maggior parte delle ricerche si sono concentrate su una classificazione dei lead guidata dalla loro probabilità di acquisto, ancora pochi studiosi hanno provato ad esplorare metodi predittivi più sofisticati, che arrivino a prevedere il ricavo medio prospettabile di uno specifico lead. In questo modo, dopo una prima selezione qualitativa basata sul concetto di conversione dei lead, l'impresa può dettagliare maggiormente la sua classificazione sulla base di fasce di ricavo che questi possono generare. Naturalmente, per riuscire a questo scopo, è necessario riuscire ad estrarre informazioni sulla *willigness-to-pay* del potenziale cliente; ad esempio, nel caso di un mercato B2B, prendendo in considerazione le dimensioni dell'impresa. Gli studi potrebbero quindi concentrarsi sull'utilizzo di metodi di machine learning, non basati sulla classificazione (che mostra come output una classe di appartenenza), ma sulla regressione, che riporta una predizione numerica, in particolare stimando il valore monetario che il lead può apportare concludendo l'acquisto.

Inoltre, ad oggi, la maggior parte delle ricerche si basa sulla costruzione del modello predittivo tramite metodi di apprendimento supervisionato[5], ovvero, in cui l'algoritmo opera, a partire da un training

[5] Come quelli trattati nel paragrafo "Metodi di costruzione di un modello di lead scoring: approccio manuale versus metodi di AI"

dataset, conoscendo le variabili in input e le possibili classificazioni di output; pertanto, elabora una "regola" generale che associ i dati in input all'output e che possa essere utilizzata per la predizione di nuove osservazioni, diverse quindi dal training dataset. Tuttavia, occorrerebbe elaborare ricerche anche sull'utilizzo di metodi di apprendimento non supervisionato, i quali operano solamente a partire dai dati di input, senza conoscere a quale output devono arrivare. In questo modo, l'algoritmo viene lasciato libero di poter analizzare i dati, seguendo le loro analogie e differenze, e su queste definisce la regola di organizzazione degli output. In questo modo, si potrebbero scoprire tendenze dei dati differenti da quelle pensate e, quindi, anche predire classificazioni in maggior numero e più dettagliate.

Seconda Sezione: Il caso applicativo

Presentazione del caso

L'impresa Cluster Reply srl opera nella creazione di soluzioni CRM customizzate sulle imprese clienti, di cui la prevalenza è rappresentata dai vari brand raggruppati sotto il grande nome della multinazionale CNH Industrial. In particolare, nel contesto lavorativo affrontato nella mia esperienza in Cluster Reply srl, il focus si è concentrato sull'ambiente marketing, che, come affrontato nella letteratura esposta nella *Prima Sezione*, rappresenta uno dei cardini delle imprese in quanto fonte primaria di generazione di nuova clientela.

Recentemente, Cluster Reply srl ha adottato come sistema software per la gestione del marketing in CNH Industrial il nuovo rilascio di Microsoft, denominato Microsoft Dynamics 365 Marketing. È stato quindi necessario studiare le funzionalità di questo nuovo tool di lavoro e sperimentarle in ambienti creati appositamente come test. Infatti, oltre ad un mero adattamento al nuovo software e migrazione dei dati, il team si è trovato di fronte a nuovi tipi di entità e di processi, che non aveva mai gestito nei precedenti sistemi. Tra questi, l'entità del Lead, che nel precedente ambiente era denominato come Contact e non differiva dal punto di vista tecnico dal contatto, se non per la quantità di informazioni collezionate molto più scarna, essenziale.

In MS Dynamics 365 Marketing, invece, tre sono le entità principali:

- ***Contact***: rappresenta il singolo individuo, cliente o potenziale cliente, fornitore o un altro soggetto che intrattiene un qualche tipo di relazione con l'impresa. Ogni contatto possiede informazioni personali, che aiutano la funzione marketing a creare specifici segmenti di contatti, individuati filtrando secondo le query di interesse. In questo modo, è possibile creare i cosiddetti "*Percorsi cliente*" (*Customer journeys*), che, partendo dal segmento di contatti selezionato, costruisce un modello di processo che descrive l'interazione che si avrà con ogni singolo contatto del segmento. In genere i *percorsi cliente* includono delle attività di primo approccio dell'impresa verso il contatto (ad esempio l'invio di messaggio e-mail o la programmazione di una phone call), seguite da *condizioni if/then,* le quali innescano il trigger per azioni successive a seconda del tipo di risposta registrato per ogni contatto. Pertanto, i percorsi cliente si inseriscono in quella categoria di marketing automation, di cui trattato nella *Prima Sezione*, in quanto permettono la creazione di campagne di marketing massive, ma allo stesso tempo personalizzate in relazione al comportamento del singolo.
- ***Account***: identifica le aziende che entrano in relazione con l'impresa. Come per il *Contact*, queste aziende potrebbero essere imprese fornitrici, imprese partner o con legami

commerciali, ma anche imprese clienti nel caso l'impresa di riferimento operasse in un mercato B2B. In ogni caso, in genere l'account viene associato al contatto, cioè l'individuo che lavora in quell'azienda. In questo modo, si crea una relazione che rende immediatamente rintracciabile per ogni contatto l'azienda nella quale opera (se presente) e, all'inverso, chi sia il contatto di riferimento per l'azienda. Quest'ultimo è particolarmente importante per la creazione dei *percorsi cliente*, in quanto essi possono partire dai segmenti di contatti, mentre non è possibile creare dei segmenti di account aziendali.

- ***Lead***: si riferisce ad uno specifico interesse del contatto. Il lead rappresenta il vero e proprio potenziale cliente, ovvero colui che si registra nel modulo d'iscrizione di una campagna, nutrendo dunque un interesse specifico in una determinata categoria di prodotti. Pertanto, l'iscrizione a campagne differenti determina la creazione di diversi lead, ognuno che porta l'informazione circa la provenienza e quindi il tipo di interesse espresso implicitamente (o anche esplicitamente attraverso la compilazione di campi nel form d'iscrizione o in successivi di maggior dettaglio). Analogamente all'*Account*, anche per il *Lead* è consigliabile associare l'individuo cui si riferisce, ovvero il *Contatto padre*, e l'azienda (*Account padre*), ma in questo caso ad uno stesso contatto/account possono essere messi in relazione più record di lead. Sebbene al momento non sia ancora abilitata la creazione di segmenti lead per cui avviare un percorso cliente e quindi sia sempre necessario utilizzare i contatti associati; Microsoft ha annunciato che questa funzionalità verrà rilasciata nel 2022, pertanto le considerazioni elaborate in questo lavoro di tesi si prospetteranno anche sull'idea di uno sviluppo posteriore al rilascio della feature, quindi considerando possibile la creazione di *percorsi cliente* basati su *segmenti lead* (senza la necessità di un contatto associato).

  La particolarità dei lead è che possono essere classificati secondo una scala di priorità (*cold*=bassa, *warm*=media, *hot*=alta), creata attraverso un modello di *lead scoring* automatizzato. Quest'ultimo consente all'impresa di strutturare secondo le proprie esigenze uno schema di attribuzione di punteggi, in relazione ai valori assunti da fattori demografici, ovvero dai campi dell'entità *Lead* o delle entità padre del *Lead*, oppure da fattori comportamentali in risposta alle campagne e-mail (apertura con una determinata frequenza, ad esempio). La funzione del lead scoring è quindi quella di stabilire quali lead siano più promettenti per le vendite per poterli contattare in modo diretto prioritariamente rispetto agli altri. A questo scopo, può essere definito all'interno del modello di lead scoring un punteggio soglia che qualifica il *Lead* a "pronto per le vendite" in modo automatico (anche se è comunque possibile effettuare la conversione manualmente). Al momento, è consentito assegnare un punteggio da un modello di lead scoring solo ai lead associati ad un contatto o

ad un account, pertanto i lead senza contatto/account padre sono esclusi da ogni modello di scoring, selezionando già in parte i lead che, avendo più informazioni, risultano più attendibili.

Il contesto di CNH Industrial e la necessità di implementare il lead scoring

Il caso applicativo preso in esame si propone proprio la progettazione per la prima volta di un sistema di lead scoring per l'impresa cliente CNH Industrial, in particolare per il ramo Agriculture & Construction[6], che si occupa della vendita di macchine agricole (trattori, mietitrebbie, attrezzi per la lavorazione agricola) e di macchine per movimento terra e costruzione (escavatori cingolati e gommati di grandi o piccole dimensioni, pale gommate, motolivellatrici, carrelli elevatori). Per il tipo di prodotti venduti, il mercato di riferimento è principalmente un B2B (business-to-business) rivolto ad aziende agricole e aziende edili. Questo rende molto ampia la platea dei potenziali acquirenti e anche molto differenziata, in termini di categorie di prodotti d'interesse, ma anche di willingness-to-pay e frequenza di acquisto. Nasce così l'esigenza di segmentare per poter gestire in modo ottimale l'interazione con i potenziali clienti, focalizzando le proprie risorse su quelli che, con maggior probabilità, arriveranno a concludere l'acquisto. Il lead scoring permette proprio di poter classificare ognuno degli acquirenti potenziali (*Lead*) secondo determinate caratteristiche che l'azienda CNH Industrial definisce nell'identificare le "buyer personas". Questo modello di punteggio si configura quindi come uno strumento in grado di ottimizzare l'efficienza delle risorse impiegate nei processi di marketing e di vendita; pertanto, la sua implementazione attraverso MS Dynamics 365 Marketing può apportare un valore aggiunto per l'intero sistema.

Il processo di vendita in CNH Industrial non avviene in modo diretto, ma deve necessariamente passare attraverso i *Dealer*, ovvero i concessionari, che si occupano di trattare direttamente con il cliente fino all'effettivo acquisto. Perciò i Dealer rappresentano i veri utilizzatori del sistema di lead scoring, in quanto possono sfruttare le informazioni rilevate da questo per capire con quali precedenze contattare tutti i lead che sono stati loro assegnati.

I lead possono essere creati a partire da campagne digitali, ad esempio attraverso le piattaforme di social networks, oppure da eventi in presenza, come grandi fiere.

Nel primo caso, l'individuo interagisce con un contenuto digitale di interesse in uno dei social network e gli viene proposto di iscriversi inserendo informazioni essenziali in uno specifico form da sottomettere. Successivamente, viene mandata una survey per fornire informazioni più di dettaglio, quali, ad esempio, l'indirizzo, l'interesse, il ruolo nella propria azienda. Se il lead compila questi

[6] Da questo punto in poi ci si riferirà, per semplicità, a CNH Industrial, ma è sempre da intendersi come lo specifico ramo Agriculture & Construction

campi viene *qualificato*, cioè vengono creati l'*Account*, il *Contact* e l'*Opportunity*[7] associati e viene assegnato ad un dealer secondo criteri di prossimità geografica. Se, invece, il lead non compila la survey, si cerca comunque di assegnarlo ad un dealer (se possibile dalle informazioni raccolte) e quest'ultimo lo contatterà per informazioni aggiuntive: se fornite il lead viene qualificato e si creano *Account*, *Contact* e *Opportunity*; altrimenti, se dichiara espressamente un mancato interesse viene *disqualificato*.

Nelle fiere, invece, l'iscrizione avviene attraverso una landing page, nella quale già sono richieste, oltre alle informazioni obbligatorie, alcune facoltative, che se compilate portano direttamente all'*auto-qualifica*. In caso contrario, si procederà come spiegato per le campagne digital.

In quest'ottica, il modello di lead scoring, nel rispetto del vincolo imposto da MS Dynamics 365 Marketing, andrà a considerare solamente i lead qualificati[8], coloro i quali, grazie alla disponibilità di informazioni, hanno portato alla creazione di un Account e di un Contact. Tuttavia, a differenza della logica dell'ambiente MS che stabilisce un lead come "sales-ready" solamente dopo aver raggiunto il punteggio soglia del modello automatico di lead scoring (o in alternativa, averlo impostato manualmente); il modus operandi dell'impresa CNH Industrial prevede che ogni lead qualificato sia già potenzialmente pronto per le vendite: ad esempio, un lead potrebbe innescare la creazione di un'opportunità di vendita ed eventualmente concluderla immediatamente. Per questo motivo, non sarebbe corretto, per il processo in CNH Industrial, definire un momento di "conversione" del lead a cliente. L'obiettivo del lead scoring, che in letteratura viene delineato come incremento del tasso di conversione dei lead, si traduce, quindi, in questo caso, in aumento delle "Won Opportunities" generate dai lead (ovvero gli acquisti portati a compimento). Infatti, il sistema di lead scoring fornisce una classificazione dei lead secondo specifici parametri che collochino il lead in una delle fasi della pipeline, in modo che possa guidare il Dealer nel contattare in modo prioritario i lead che manifestano un interesse più spiccato verso l'acquisto. In questo modo, il lead con intenzione di acquisto immediato (*Hot*) viene catturato dal Dealer, che è in grado di soddisfare il bisogno nelle tempistiche desiderate, evitando che l'individuo/azienda passi ad un potenziale competitor. Allo stesso tempo, il Dealer ottimizza l'utilizzo delle proprie risorse nell'effettuare i contatti diretti con i lead, che generalmente avvengono per via telefonica, considerata molto time-consuming. Di fatto, i lead che vengono considerati poco promettenti per la vendita (*Cold*) vengono

[7] L'Opportunity è un'altra entità di MS Dynamics 365 Marketing che permette di tener traccia dell'opportunità di vendita creata dal singolo lead, registrando informazioni come il prodotto d'interesse, il periodo di apertura e di chiusura, lo stato (open, won, lost).

[8] La qualifica può essere avvenuta sul lead corrente o anche in lead precedenti creati dalla stessa persona fisica/giuridica, il requisito è che il lead sia associato ad un Account.

posti al fondo della classifica di priorità, in modo che, nonostante il Dealer contatterà ciascuno dei lead che gli è stato assegnato, verrà data precedenza ai lead che si ritiene abbiano maggior probabilità di successo nella vendita.

Aspetti metodologici

La progettazione del modello di lead scoring customizzato sull'impresa CNH Industrial è stata effettuata adottando il metodo manuale, che sebbene sia meno sofisticato delle procedure automatizzate di intelligenza artificiale, si adatta bene ad un primo approccio di costruzione. L'utilizzo della procedura manuale è anche giustificato dalla mancanza di dati storici raccolti dal CRM data la recente adozione dell'ambiente MS Dynamics 365 nell'impresa, motivo per il quale l'esiguità dei dati non avrebbe portato ad analisi consistenti nel campo statistico.

La metodologia seguita si compone quindi delle seguenti fasi, che verranno dettagliate nei paragrafi successivi:

- Identificazione qualitativa degli attributi adatti a costruire il modello di lead scoring: questa prima fase è attuata andando a scegliere tra tutti i campi attinenti al lead disponibili sul CRM quali siano i più rilevanti per la determinazione dei punteggi, attraverso un confronto diretto con l'impresa CNH Industrial.
- Analisi quantitativa dei dati: per ciascuno dei campi selezionati nella fase precedente vengono considerati i dati storici per stabilirne l'effettivo utilizzo e studiarne il potenziale effetto causale sulla probabilità di acquisto.
- Attribuzione dei pesi relativi e quindi dei punteggi per ciascuno dei singoli campi, derivandoli dalla combinazione dell'importanza percepita dal business CNH Industrial e dello studio dei trend storici eseguito dalla fase precedente.
- Determinazione delle soglie per la classificazione in "Hot", "Warm" e "Cold", attuata secondo i punteggi assegnati per ogni condizione.

La fase di progettazione verrà poi seguita dall'effettiva implementazione in MS Dynamics 365 Marketing, per la quale verranno illustrati gli aspetti tecnici più rilevanti, al fine di fornire un esempio per le imprese che, come CNH Industrial, si approcciano per la prima volta a questo strumento.

Infine, verranno proposti KPIs adeguati al monitoraggio dell'efficacia del modello, fotografando la situazione attuale con il calcolo di tali indicatori ed illustrando quali aspettative di miglioramento si dovrebbero scorgere grazie all'applicazione del modello di lead scoring.

Identificazione qualitativa degli attributi adatti a costruire un modello di lead scoring

La fase iniziale di identificazione degli attributi è stata affrontata scegliendo appropriati criteri di valutazione in risposta a specifiche domande:

- Reperibilità del lead: *quali informazioni il lead ha fornito per rendersi contattabile dall'impresa?*
- Interesse esplicitato dal lead: *da quali campi posso percepire l'interesse del lead?*
- Profilo cliente target: *quali attributi possono far individuare il cliente target?*

Per ciascuna di queste domande sono stati selezionati i campi più rispondenti, anche grazie ad un ciclo di incontri di confronto con la responsabile CRM di CNH Industrial, che ha fornito una visione più completa del significato degli attributi in relazione ai processi interni all'azienda.

Reperibilità del lead

La reperibilità del lead è un fattore essenziale per rendere possibile il contatto da parte del Dealer o l'invio di campagne marketing via e-mail; pertanto, si è deciso di attribuire un punteggio ai campi che permettono rintracciabilità e che afferiscono alla sfera di consensi GDPR di cui parlato nella *Prima Sezione*:

- *Main Phone* e *Mobile Phone*: risulta obbligatorio da sistema, al momento dell'iscrizione, che almeno uno dei due campi sia valorizzato per permettere al Dealer di contattare telefonicamente quel lead. Se tra i due attributi è presente il numero di un cellulare (*Mobile Phone*), quest'ultimo può essere utilizzato anche per effettuare comunicazioni via sms ed il lead è generalmente più attivo nel rispondere rispetto ad un telefono fisso.
- *E-mail:* anche l'e-mail rappresenta un importante punto di contatto con il potenziale cliente, ma in questo caso è raramente utilizzata per comunicazioni dirette; piuttosto, la compilazione di questo campo può rendere possibile l'invio di e-mail di marketing, al fine di promuovere i propri prodotti e di alimentare il processo di *lead nurturing*.
- *CNH Industrial Direct Marketing Consent*: questo campo, se impostato al valore "Yes", rappresenta il consenso fornito dal lead nel ricevere nel proprio indirizzo di posta elettronica contenuti di marketing diretto da parte dell'azienda CNH Industrial.
- *CNH Industrial Third Marketing Consent*: se il lead accetta di prestare il consenso, i suoi dati potranno essere mandati a terze parti, ovvero ai vari Dealer, e trattati a scopo di marketing.

- *Normalization state*: quest'attributo dichiara se l'indirizzo fisico del lead sia compilato in maniera corretta, ovvero attraverso una funzione di geolocalizzazione che permette di individuare il punto preciso su una cartina geografica. Se lo status di questo campo è "*Normalized*" significa che quel lead ha fornito informazioni corrette ed è assegnabile ad uno specifico Dealer. Al contrario, "*Not normalized*" indica che l'indirizzo compilato non è completo o è scritto in maniera errata; nonostante questo, a volte le informazioni disponibili sono comunque sufficienti al dispatchment. Pertanto, l'informazione del *Normalization state* è premiabile con un punteggio che attesti l'attendibilità del dato inserito dal lead e quindi anche la probabilità che il lead sia più motivato ad avere un contatto con l'azienda.

Interesse esplicitato dal lead

Altri attributi sono invece ricavati da informazioni direttamente esplicitate dal lead in campi facoltativi dei form e risultano fortemente rappresentativi dell'interesse verso l'acquisto:

- *Confirm Interest*: un campo booleano che se flaggato esprime l'interesse del lead verso le offerte proposte.
- *Contact Reason Detail*: nel caso in cui il lead voglia contattare direttamente l'azienda, ne fornirà le motivazioni in questo campo. Si tratta di un campo di tipo "Picklist", ovvero che comporta la scelta tra una delle opzioni proposte in un menù a tendina: *Receive technical information*, *Get a quotation, Demo*, *Other* oppure nessuna risposta. In questo caso, sarebbe opportuno dare la priorità ai lead che dimostrano un interesse in fase avanzata, ovvero coloro che richiedono un preventivo (*Get a quotation*) o una prova (*Demo*).
- *Purchase Timeframe*: questo campo ad opzioni predefinite (*0-3 mesi*, *3-6 mesi*, *6-12 mesi, oltre 12 mesi*) richiede al lead di inserire il range temporale in cui intende effettuare l'acquisto. Grazie alla compilazione di questo campo si è in grado di definire una scala di priorità che ponga in cima i lead con più vicina "due date" dichiarata per l'acquisto.
- *Type of interest*: il lead può compiere una selezione multipla tra le opzioni del menu (New, Rental, Used, Ex Demo Unit, Aftermarket) per scegliere se ha interesse ad acquistare un prodotto nuovo, usato, precedentemente utilizzato per la prova, accessori o prendere in noleggio. Il business di CNH Industrial ha dichiarato di avere una preferenza verso i potenziali acquirenti di prodotti nuovi; pertanto, grazie a questo campo è possibile individuare questi per porli ad un grado di attenzione maggiore grazie all'attribuzione dei punteggi.

Attributi comportamentali che catturano l'interesse implicito

Oltre agli attributi espliciti riportati, possono essere considerati altri tipi di attributi, definiti come comportamentali, che, come esposto nella *Sezione di Letteratura,* sono quelli che catturano,

implicitamente, il grado di interesse del lead nei confronti dell'impresa. Infatti, questi possono includere le interazioni alle e-mail (click, aperture, cancellazioni, bounces) oppure le visite al sito web dell'azienda, il check-in ad eventi, il download di file e molto altro ancora.

L'impresa CNH Industrial si configura principalmente in un mercato B2B, offrendo prodotti non standardizzati, ma customizzabili dal cliente, e quindi che necessitano di un approccio di vendita che esige l'interazione dealer-customer. Per questo motivo, il sito web dell'azienda presenta contenuti informativi sui vari brand e mercati, senza riportare direttamente i prodotti acquistabili. A differenza di un sito di e-commerce, quindi, la visita dell'utente al sito web presenta un contenuto informativo di scarsa rilevanza nel prevedere la propensione all'acquisto; pertanto, non viene presa in considerazione nella selezione dei campi per il lead scoring.

Per quanto riguarda gli eventi, l'impresa organizza periodicamente delle fiere, occasioni in cui promuovere i propri prodotti ai partecipanti, che, generalmente si configurano in individui/aziende fortemente interessati alle categorie di prodotti e motivati all'acquisto. Perciò, la presenza all'evento può risultare un ottimo indice per l'attribuzione di un punteggio positivo; tuttavia, non è ancora utilizzata alcuna tecnologia che registri il check-in e, attraverso un'integrazione tra sistemi, riporti il dato direttamente sul CRM. La prospettiva di utilizzare MS Dynamics 365 Marketing anche per la gestione degli eventi in presenza è comunque stata considerata dall'azienda, che ne prospetta un'implementazione nell'immediato futuro. A quel punto, anche il campo di check-in agli eventi potrà essere integrato nel modello di lead scoring.

Al momento, quindi, fattualmente è possibile proporre solo di considerare tra i campi del lead scoring le informazioni riguardanti la risposta comportamentale del lead rispetto alle e-mail di marketing ricevute. In particolare, si è scelto di valorizzare il campo *E-mail clicked,* che indica il numero di click che l'utente ha compiuto sulle e-mail, dato che è più restrittivo rispetto al campo *E-mail opened,* che mostra semplicemente il numero delle aperture. Inoltre, si sarebbero potuti conteggiare con un impatto negativo i campi *E-mail blocked* o *E-mail hard/soft bounced*, ma, per semplicità, in questa prima costruzione del modello di lead scoring, si è presupposto di adottare una logica premiante, ovvero con soli punteggi positivi, che incrementino il valore del lead.

Profilo cliente target

Partendo dall'idea di cliente target per CNH Industrial, sono stati considerati attributi che possano essere significativi nel determinare quanto un lead si avvicini al target, in modo da considerarlo, quindi, in quel caso, come più promettente per l'acquisto:

- *Industry*: il lead sceglie tra i settori di mercato preimpostati come opzioni quello di riferimento per l'azienda in cui opera. Naturalmente, il target per CNH Industrial si concentra su quei settori che necessitano per il proprio business di macchinari agricoli o di macchine movimentazione terra.
- *Created On*: è un campo che si compila in maniera automatica con la data di creazione di quel record di Lead. Può essere preso in considerazione, ad esempio, supponendo che i lead che si sono iscritti recentemente potrebbero averlo fatto per necessità di effettuare un acquisto; pertanto, il Dealer dovrebbe trattarli con priorità. Quest'ipotesi verrà testata attraverso l'utilizzo dei dati storici nel paragrafo seguente.
- *Qualifying Opportunity*: è un campo di tipo "Lookup", che fa riferimento al record dell'entità Opportunity creata nel momento della qualifica, sia essa avvenuta attraverso il processo digitale o attraverso qualifica nelle fiere in presenza. Se questo campo è popolato e lo status di questo record è settato in "Open" significa che nel processo di qualifica si è ravvisata una possibile opportunità di acquisto e quindi è necessario dedicare risorse per portarla a compimento.
- *Channel*: indica il canale di provenienza del lead ed è impostato automaticamente secondo valori predefiniti, tra cui i preferibili, secondo l'esperienza di CNH Industrial, sono "*Digital campaigns*", "*Events/ lead capture*", "*Website*" poiché giungono con un intenzioni di acquisto più consolidate rispetto a quelli provenienti da altri canali. Anche quest'ipotesi verrà studiata nelle analisi seguenti.
- *Decision Maker*: mettendo la spunta su questo campo booleano il lead dichiara di avere un potere decisionale all'interno dell'azienda in cui opera. La logica di classificazione, evidentemente, andrà a focalizzarsi anche su questo aspetto, in quanto ha maggiore attendibilità e possibilità di successo il contatto con uno dei capi dell'impresa, piuttosto che con un impiegato della stessa, il quale dovrebbe interfacciarsi ancora con i propri superiori.
- *Activites Count:* questo campo viene calcolato applicando la funzione di conteggio delle attività relazionate al Parent Account del lead, perciò si auto-aggiorna ogni qual volta si aggiunga un'Activity. In particolare, si definiscono come Activities le attività di interazione con il potenziale cliente, quali ad esempio una phone call o un appuntamento. L'ipotesi di base per lo scoring quindi si basa sul fatto che un lead che ha portato avanti frequenti momenti di contatto con il dealer possa avere un coinvolgimento maggiore e quindi una più accentuata propensione all'acquisto.
- *Open deals:* anche in questo caso, si tratta di un campo calcolato come *count* delle Opportunities che risultano in status "open" per l'Account Padre del lead. Lo stato "open",

infatti, indica che è stata aperta un'opportunità di acquisto per quel lead, ma non è ancora stata portata a completamento. Questo attributo è quindi fondamentale per fissare quei lead in cima alla classifica ed evitare che si perdano di vista opportunità aperte, che, se lasciate insoddisfatte per molto tempo, si trasformerebbero in mancati guadagni per l'impresa.

Da quest'ultimo attributo è poi nata l'idea di poter sfruttare un'informazione aggiuntiva, interrogando il database anche sul numero delle *Won Opportunities* del Parent Account di uno specifico lead. Per riuscire a compiere questo conteggio, dal punto di vista tecnico del sistema, sarà necessario creare un campo analogo al precedente, che però vada a conteggiare le Opportunities con status "won" (verrà definito per questo *Won Deals*). In questo modo, sarà possibile ricavare l'informazione se il lead abbia già acquistato in passato per l'impresa e conseguentemente derivarne l'ipotesi che si tratti di un cliente fidelizzato o comunque un soggetto con indubbio interesse verso l'acquisto, meritevole perciò di un punteggio maggiore.

Analisi quantitativa dei dati

Considerazioni preliminari

Successivamente alla selezione qualitativa dei campi concettualmente più rilevanti per poter definire punteggi del lead scoring rappresentativi della probabilità di acquisto, si è reso necessario passare ad un'analisi di tipo quantitativo per dimostrare che i suddetti campi presentassero valori in numero sufficiente ad essere considerati. In particolare, si è preso a riferimento lo storico di dati raccolti a partire dal 1° luglio 2020, in quanto la migrazione al sistema di MS Dynamics 365 è avvenuta nella seconda metà del giugno 2020, perciò i dati inseriti a sistema prima di tale data potrebbero avere caratteristiche differenti dovute al vecchio ambiente CRM.

Dalle più immediate rilevazioni è quindi emerso che i campi *Confirm Interest* e *Decision Maker* sono attributi standard che l'ambiente MS Dynamics 365 crea automaticamente, ma non sono utilizzati nel flusso di processi di CNH Industrial, ovvero non sono incluse in nessuno dei form proposti al lead alla sottoscrizione o nella procedura di auto-qualifica. Per questo motivo, la scelta obbligata è stata quella di eliminare tali campi dall'insieme prescelto. Inoltre, l'impresa CNH Industrial ha anche optato per l'eliminazione dei campi *Industry* e *Type of Interest*, posponendo l'ipotesi di eventuale inserimento futuro di questi campi.

Attendibilità dei dati lato-utente

Una questione importante da porre sotto analisi quando si tratta con i dati è quella di verificarne l'attendibilità. Spesso gli utenti per non fornire informazioni personali riempiono i campi di un form

con caratteri casuali o informazioni non veritiere. Pertanto, parallelamente alla selezione preliminare degli attributi da utilizzare per il lead scoring, le seguenti considerazioni sono state avanzate:

- I campi *Main Phone*, *Mobile Phone* ed *E-mail* sono gli unici di tipo String, che perciò l'utente inserisce manualmente da tastiera. Per questa ragione, sono effettuati da principio dei controlli formali, ad esempio riguardanti la presenza di soli caratteri numerici per il numero di telefono, non ripetuti (7777777777) e non sequenziali (1234567890), e l'inserimento dei caratteri "@" e "." per l'indirizzo e-mail. Tuttavia, passati i controlli formali, non esiste alcuna garanzia di autenticità del dato. È opportuno tener conto di quest'aspetto nella valutazione del punteggio da attribuire a questi campi.
- Alcuni attributi richiesti al lead alla compilazione dei form sono di tipo Boolean (*CNH Industrial Direct Marketing Consent, CNH Industrial Third Marketing Consent*) e Picklist (*Contact Reason Detail, Purchase Timeframe*): questo significa che il lead è obbligato a selezionare una delle opzioni predefinite. Sebbene la completa autenticità del dato non sia assicurata, è molto più raro che la selezione di un campo sia effettuata in maniera randomica o mendace, soprattutto perché si tratta di informazioni non troppo invasive. Pertanto, rispetto alle precedenti, risultano tendenzialmente più attendibili. Oltre a questo elemento, nel considerare questo tipo di campi, è stata osservata anche la possibilità di discriminare facilmente i valori dei campi in numero finito, attribuendo perciò un determinato punteggio per ogni opzione predefinita (con un campo di tipo String questo non sarebbe stato possibile perché ne sarebbero derivate un'alternativa potenzialmente diversa per ogni lead e quindi un numero spropositato di dati da analizzare uno ad uno).
- Infine, gli altri attributi, non richiesti direttamente al lead, ma calcolati in maniera automatica dal sistema risultano certamente attendibili lato-utente. Questi comprendono *Normalization State, CreatedOn, Qualifying Opportunity, Channel, Activities Count, Open Deals* (e il suo analogo di prossima creazione *Won Deals*).

Affidabilità del sistema lato-utilizzatori

Esaminata l'attendibilità delle informazioni dichiarate dall'utente, non si può prescindere dall'analizzare anche l'affidabilità dell'intero sistema di raccolta dei dati e di inserimento manuale da parte degli utilizzatori effettivi dell'impresa CNH Industrial e dei Dealer ad essa associati. Infatti, il sistema CRM di Dynamics 365, adottato solo da metà 2020, ha portato a dover effettuare diversi aggiornamenti di sistema, che potrebbero aver causato una parziale compromissione dei dati. L'ideale, in questo caso, sarebbe considerare solo i dati a partire dal luglio 2021 poiché oltre questa

data non sono state rilevate problematiche per cui siano stati richiesti aggiornamenti di sistema; tuttavia, al momento, non si disporrebbe di sufficienti dati per un'analisi consistente.

Un'ulteriore criticità rilevabile in questo periodo di assestamento al nuovo sistema è l'addestramento al suo utilizzo, che è ancora in corso nel caso dei Dealer, i quali talvolta non completano totalmente il processo all'interno del sistema e questo comporta una sottostima delle vendite effettuate.

In questo contesto, si rischia di incorrere in errori di valutazione, in quanto l'impatto che la qualità dei dati ha sulla strutturazione del modello potrebbe portare a considerare in modo limitativo (o addirittura fuorviante) alcuni attributi. Pertanto, le conclusioni a cui si giungerà devono essere considerate come affette da un elevato rischio di errore derivante dai dati di input. In questo lavoro di tesi, tuttavia, l'obiettivo è quello di suggerire un approccio valutativo, prendendo in esame le metodologie utilizzabili, piuttosto che i veri e propri risultati numerici. Lo studio, quindi, viene comunque condotto a scopo esemplificativo, pur consapevoli dell'inadeguatezza delle considerazioni quantitative.

In ottica aziendale, invece, per raggiungere un adeguato livello di affidabilità bisognerebbe utilizzare le analisi proposte, in futuro, dopo aver raccolto dati più attendibili. Per perseguire questo scopo, la progettazione dovrebbe partire ancor prima della fase in cui si considera il sistema e i dati raccolti, ovvero prendere le radici proprio nell'instaurare all'interno dell'azienda il corretto approccio all'utilizzo del sistema: da una parte, assicurandosi che gli utilizzatori effettivi, ovvero i Dealer, siano correttamente addestrati al suo utilizzo e incentivati da opportuni KPIs a sfruttare tutte le potenzialità del nuovo sistema; d'altra parte, è possibile anche, operando integrazioni tra eventuali ambienti operativi differenti, automatizzare una migrazione di dati istantanea da un sistema all'altro, evitando quindi che si creino gap informativi o dati incoerenti tra i diversi ambienti software. A quel punto, la premessa ad una qualsiasi progettazione all'interno del sistema deve essere effettuare una verifica della qualità dei dati, la quale necessita di monitoraggi anche successivi con frequenza prestabilita.

Nei successivi paragrafi si delineeranno le analisi relative a ciascuno degli attributi scelti nella selezione qualitativa.

CNH Industrial Direct Marketing Consent

Il campo *CNH Industrial Direct Marketing Consent* raccoglie le scelte dei lead sul prestamento del consenso ad essere contattati per finalità di marketing attraverso campagne effettuate direttamente dall'azienda CNH Industrial. Queste possono essere opportunità di sponsorizzazione indiretta anche per i singoli dealer; perciò, il flag del consenso dovrebbe far attribuire un punteggio positivo, anche in relazione ad un maggior interesse del lead verso l'azienda. La domanda posta in esame però è "*È*

vero che la quasi-totalità dei lead che hanno portato ad una won opportunity avevano prestato questo consenso?". Pertanto, partendo dalle Won Opportunities, è stata estratta l'informazione sul lead attraverso un'interrogazione SQL, che, seleziona questo e gli altri successivi attributi analizzati, sfruttando condizioni di join tra le tabelle Opportunity, Account, Lead. Questo processo è stato appositamente studiato per risalire agli attributi del lead attraverso la relazione con il proprio Parent Account, cui è collegata l'Opportunity. Pertanto, selezionando solo le opportunità vinte, l'informazione dell'account associato univocamente a quell'opportunità viene utilizzata per risalire ai lead che hanno relazione di Parent Account con questo ed estrapolarne i campi d'interesse. In questo modo è possibile studiare la distribuzione di valori degli attributi dei soli lead che hanno portato a termine l'acquisto.

Nel caso di *CNH Industrial Direct Marketing Consent*, la figura 10 riporta i dati rilevati.

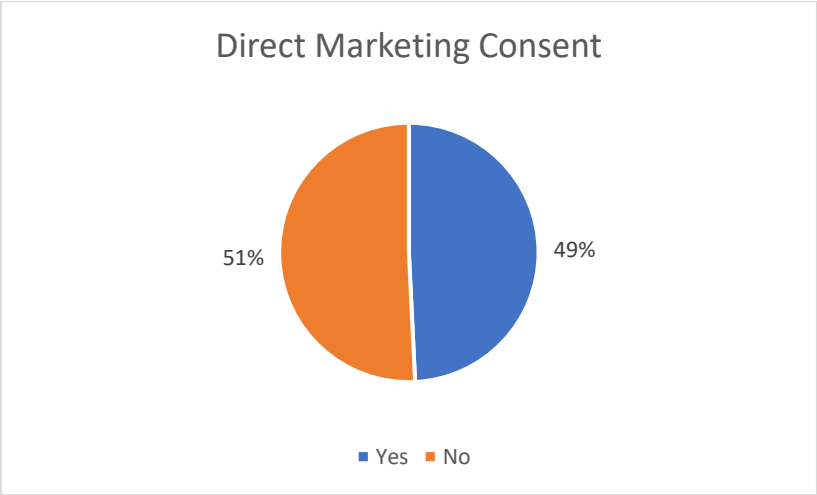

Figura 10: Distribuzione di CNHI Direct Marketing Consent considerando le Won Opportunities

Il grafico, spaccato quasi perfettamente a metà, mostra che i lead che hanno acquistato almeno una volta avevano, per metà prestato il consenso, ma per l'altra metà no. Questo ci porta a concludere che possa non esserci alcuna relazione di causa-effetto tra il consenso di direct marketing e la propensione all'acquisto. Si riassumono in tabella le considerazioni che serviranno all'attribuzione dei punteggi di lead scoring.

Tabella 3: Considerazioni sull'attributo CNHI Direct Marketing Consent

Attributo	Grado attendibilità	Relazione con l'acquisto
CNH Industrial Direct Marketing Consent	Alto	Debole

CNH Industrial Third Marketing Consent

Analoghe analisi rispetto al paragrafo precedente sono state svolte per l'attributo *CNH Industrial Third Marketing Consent*, che si riferisce al consenso di marketing effettuato da parti terze, ovvero i Dealer.

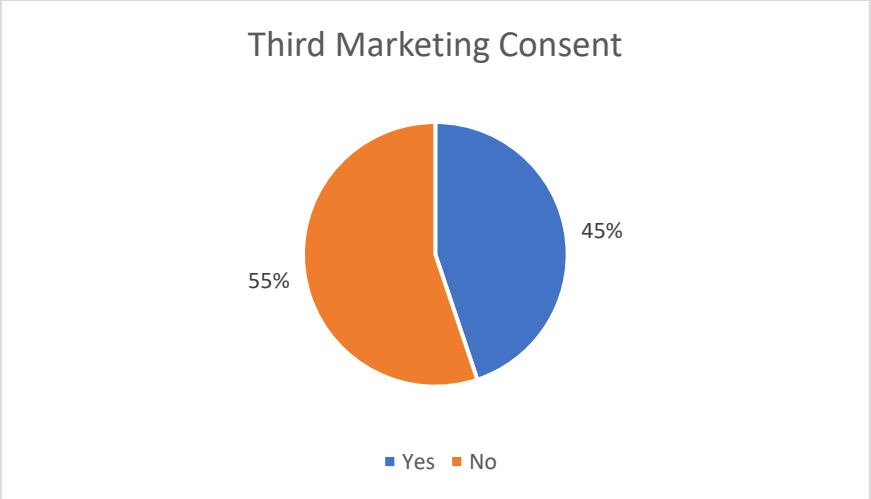

Figura 11: Distribuzione di CNHI Third Marketing Consent considerando le Won Opportunities

Anche in questo caso, le rilevazioni effettuate supportano l'idea di una scarsa relazione del consenso con la propensione all'acquisto.

Tabella 4: Considerazioni sull'attributo CNHI Third Marketing Consent

Attributo	Grado attendibilità	Relazione con l'acquisto
CNH Industrial Third Marketing Consent	Alto	Debole

Main Phone e Mobile Phone

I due campi *Main Phone* e *Mobile Phone* rappresentano il recapito essenziale per contattare in modo diretto il potenziale cliente e poter interagire in maniera molto più completa rispetto ad una semplice e-mail. Nelle ipotesi iniziali si è presupposto come obbligatorio possedere l'informazione di almeno uno dei due campi, in particolare, prediligendo il numero del cellulare perché più reperibile del telefono fisso. Le analisi effettuate su questi campi si sono quindi concentrate sulle Won Opportunities e sono andate a studiare quanti lead avessero entrambi i campi inseriti o solo uno dei due recapiti telefonici (Figura 12).

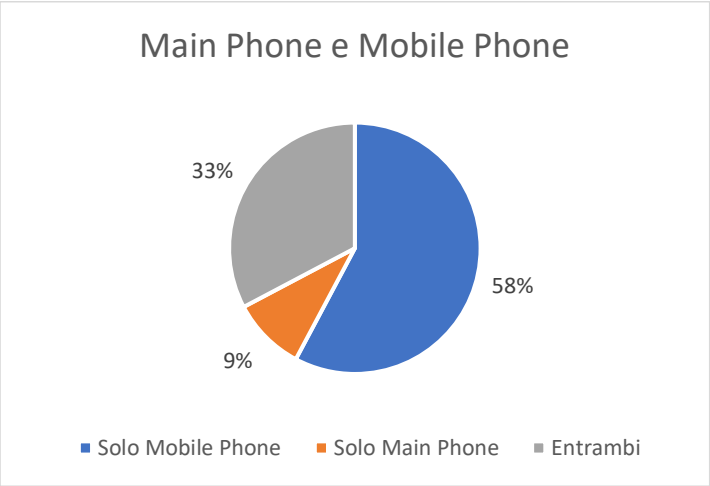

Figura 12: Distribuzione dei campi Main Phone e Mobile Phone considerando le Won Opportunities

Come ipotizzato, per i lead che hanno completato l'acquisto, l'ipotesi di maggior importanza relativa del *Mobile Phone* rispetto al *Main Phone* è verificata dalla maggioranza (58% contro il 9%). Un buon 33% ha invece inserito entrambi i recapiti. Di queste considerazioni (in Tabella 5) se ne terrà conto nell'attribuzione dei punteggi.

Tabella 5: Considerazioni sugli attributi Main Phone e Mobile Phone

Attributo	Grado attendibilità	Relazione con l'acquisto
Main Phone/Mobile Phone	Basso	Forte

E-mail

Anche l'e-mail rappresenta un'informazione di recapito importante, ma risulta più debole del recapito telefonico per colloqui diretti; viene utilizzata, piuttosto, per marketing più massivo.

Similmente, il database è stato interrogato con la stessa tipologia di query, considerando le Won Opportunities (Figura 13).

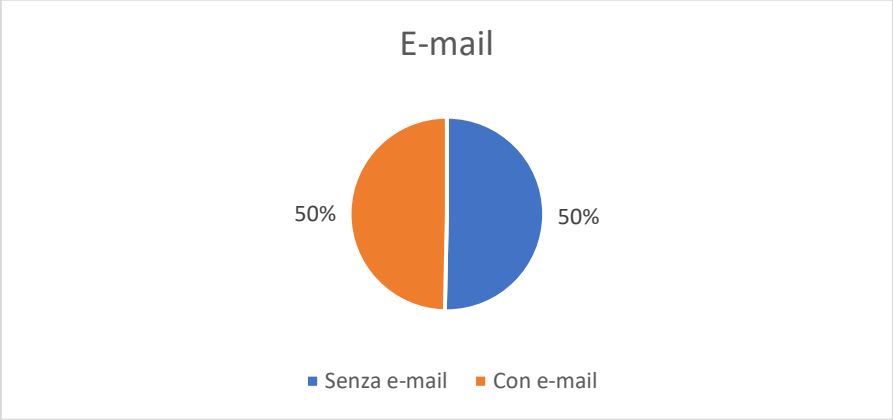

Figura 13: Distribuzione del campo e-mail considerando le Won Opportunities

Il grafico mostra una spaccatura esattamente a metà tra i lead che registrano l'informazione dell'e-mail e quelli con il campo nullo, evidenza che ci porta a concludere che non vi sia nesso causale tra disponibilità di quest'informazione e maggior probabilità di acquisto.

Tabella 6: Considerazioni sull'attributo E-mail

Attributo	Grado attendibilità	Relazione con l'acquisto
E-mail	Basso	Debole

Contact Reason Detail

Per quanto riguarda il campo "*Contact Reason Detail*", l'ipotesi da testare è che la maggioranza delle Won Opportunities sia generata da lead che hanno dichiarato come ragione del contatto "Demo", cioè poter effettuare una prova, oppure "Get a quotation", cioè ricevere un preventivo.

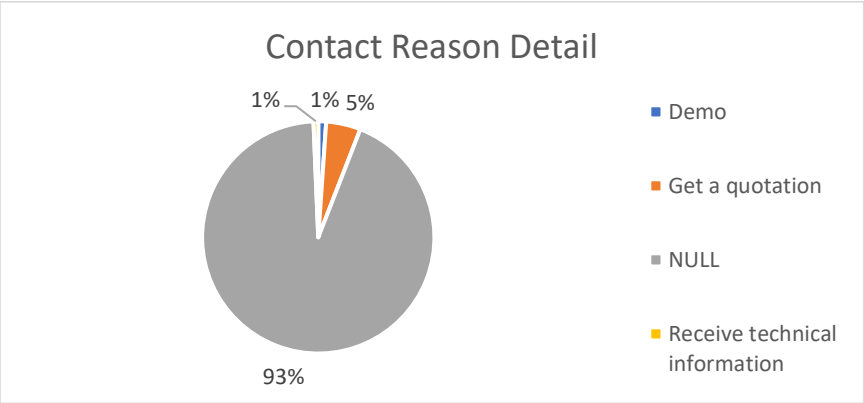

Figura 14: Distribuzione del campo Contact Reason Detail considerando le Won Opportunities

Purtroppo, trattandosi di un campo facoltativo alla compilazione del lead, sono stati rilevati insufficienti dati per poter trarre conclusioni adeguate. Pertanto, nella valutazione del dato ci si baserà sull'ipotesi di una media relazione con l'acquisto. Quest'analisi verrà comunque effettuata nuovamente a seguito del roll out di tutti i mercati (al momento solo alcuni hanno adottato il sistema CRM di MS Dynamics 365).

Tabella 7: Considerazioni sull'attributo Contact Reason Detail

Attributo	Grado attendibilità	Relazione con l'acquisto
Contact Reason Detail	Medio	Media

Normalization State

Per il campo *Normalization State*, si presume che un lead che abbia speso del tempo per scrivere correttamente il proprio indirizzo sia maggiormente motivato all'acquisto. Per testare quest'ipotesi, si è filtrata l'analisi alle Won Opportunities e ne è emerso che il 21% dei lead che hanno effettuato un acquisto risulta "Not Normalized", mentre ben il 79% ha lo stato "Normalized". Quest'evidenza empirica conferma l'idea iniziale di trattare l'attributo come indice dell'attendibilità del lead e quindi della sua maggiore propensione all'acquisto.

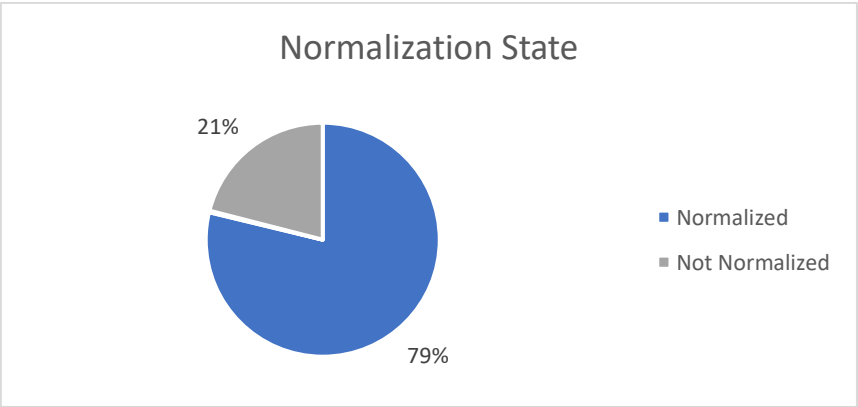

Figura 15: Distribuzione del campo Normalization State considerando le Won Opportunities

Le considerazioni che ne risultano sono espresse in tabella 8:

Tabella 8: Considerazioni sull'attributo Normalization State

Attributo	Grado attendibilità	Relazione con l'acquisto
Normalization State	Alto	Forte

CreatedOn

Nel considerare il campo *CreatedOn* la domanda posta è stata "*Il Dealer dovrebbe contattare con priorità i lead che si sono iscritti più recentemente? Un'interazione immediata può favorire l'acquisto?*". Per rispondere, si è cercata un'evidenza empirica nei dati. In particolare, per riuscire a catturare entrambe le informazioni di data di creazione del lead e data di acquisto, è stata pratica la creazione di un nuovo campo di tipo calcolato, denominato *TempoDiAcquisto*, che sfrutta la funzione "DIFFINDAYS", restituendo quindi la differenza in giorni tra la *CloseDate* di una Opportunity di stato "Won" (ovvero la conclusione dell'acquisto) e la *CreatedOn* dell'Account. Il motivo per cui si è considerata l'entità Account, invece che quella del Lead è la relazione univoca obbligatoria che c'è tra Opportunity e Account (ad ogni Opportunity è associato l'Account che l'ha generata), mentre la relazione tra Opportunity e Originating Lead risulta come opzionale, riportando quindi raramente il lead di riferimento. Questo, tuttavia, non comporta distorsioni dal punto di vista concettuale: si ricorda, infatti, che Lead e Account coincidono con la stessa persona fisica/giuridica; perciò, studiare il comportamento di acquisto di un Lead o quello del suo Parent Account è equivalente. Questo procedimento è stato applicato solo nel caso dell'attributo *CreatedOn*, e non per gli altri, per cui invece non era possibile riferire un attributo del Lead ad uno analogo dell'Account.

In questo modo, diventa immediato avere una percezione dell'arco temporale impiegato per l'acquisto. Il grafico di seguito, in figura 16, ne mostra l'andamento.

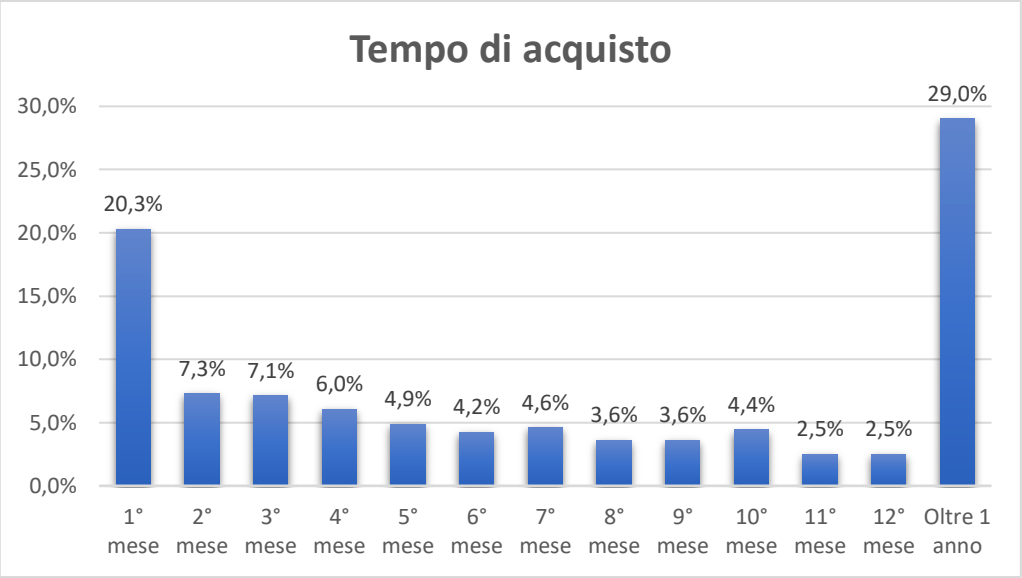

Figura 16: Distribuzione del tempo di acquisto secondo arco temporale mensile

Il grafico è stato creato a partire dalla tabella Opportunities di stato "Won", estraendo il dato, per ogni Account associato, del *TempoDiAcquisto*.

Le considerazioni che ne emergono rivelano che un buon 20% degli Account effettua un acquisto nel primo mese, il 7,3% conclude l'affare nel secondo mese dall'iscrizione e così via. Da notare che l'ultima barra dell'istogramma rappresenta un dato aggregato, che conteggia tutti gli acquisti dal tredicesimo mese in poi; pertanto, ne risulta che il 71% degli acquisti avviene entro il primo anno dalla registrazione, mentre il 29% oltre. Questi risultati verranno quindi considerati per l'attribuzione dei punteggi nelle prossime fasi di progettazione del modello di lead scoring.

Più dettagliatamente, è stato studiato come il 20% degli acquisti sia distribuito all'interno del primo mese, per poter passare ad un range temporale settimanale. Se ne riporta la distribuzione dei dati nella figura 17.

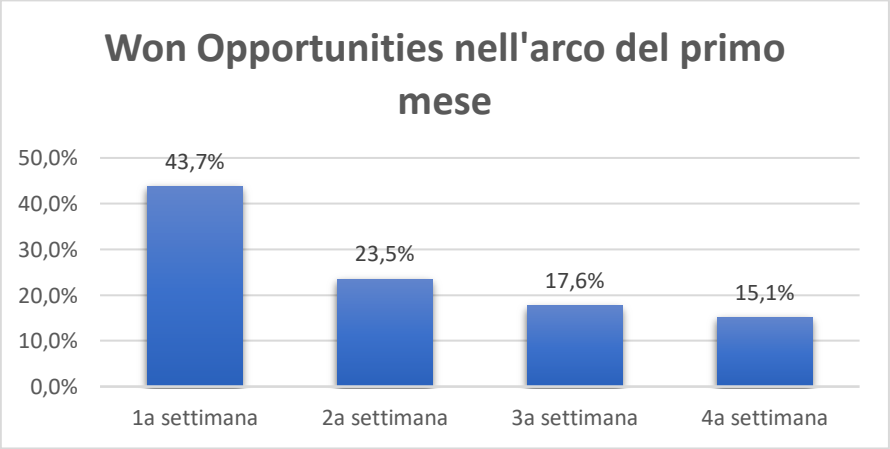

Figura 17: Distribuzione del tempo di acquisto secondo suddivisione settimanale

Il grafico mostra un trend decrescente al passare del tempo, che vede, quindi, il 43,7% degli acquisti concentrati nella prima settimana dalla creazione dell'Account. Sicuramente, questo dato sostiene la teoria secondo la quale un utente si registra perché ha un interesse all'acquisto generalmente immediato o nel breve periodo. Tuttavia, è da riconoscere una possibile distorsione dell'informazione: all'interno di quel 43,7%, il 35% circa di quegli Account ha registrato un tempo di acquisto pari a 0. È possibile, infatti, che il Dealer abbia registrato l'Opportunity vinta, creando solo in quel momento l'Account. Questo non permette di riuscire a risalire alle eventuali interazioni precedenti dell'Account; perciò, ne risulta che l'informazione potrebbe essere limitata ad un arco temporale più ristretto di quello effettivo. Si può concludere quindi che i dati mostrano un andamento degli acquisti decrescente allo scorrere del tempo, come ipotizzato in una prima selezione; ma, l'attendibilità del dato, considerata alta poiché inserto automaticamente dal sistema, è viziata da uno scorretto modus operandi dei Dealer, che sposta il grado di affidabilità a medio-alto.

Nella tabella 9, vengono riportate le valutazioni qualitative dell'attributo, in relazione a queste analisi.

Tabella 9: Considerazioni sull'attributo CreatedOn

Attributo	Grado attendibilità	Relazione con l'acquisto
CreatedOn	Medio-alto	Forte

Purchase Timeframe

Per il campo *Purchase Timeframe* si è tentato di capire la relazione con l'acquisto ponendosi la domanda: *"Coloro i quali manifestano un desiderio di acquistare nell'arco temporale x, effettivamente portano a conclusione l'acquisto entro i termini dichiarati?"*

Per questa analisi è stata effettuata la seguente query, riscritta per i vari range di valori dell'attributo, e, successivamente manipolate attraverso funzioni in Excel:

```
SELECT l.purchasetimeframename, o.tempodiacquisto
FROM Opportunity o, Account a, Lead l
WHERE o.accountid=a.accountid AND a.accountid=l.parentaccountid AND o.createdOn>='2020-07-01' AND
l.purchasetimeframename= '0-3 months'
```

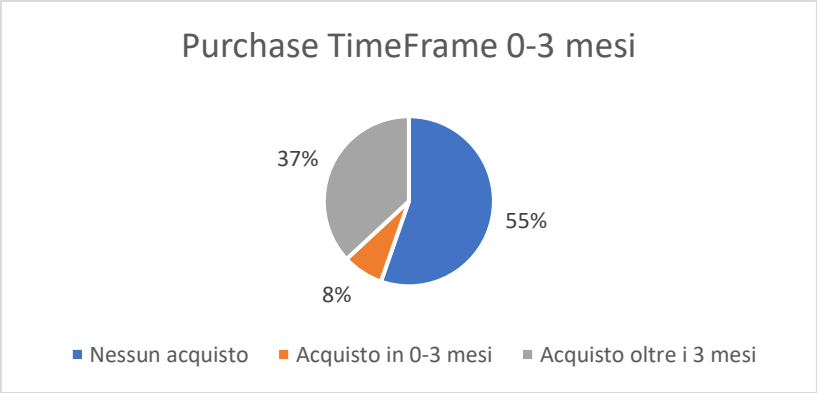

Figura 18: Distribuzione dei lead che hanno dichiarato Purchase Timeframe di 0-3 mesi

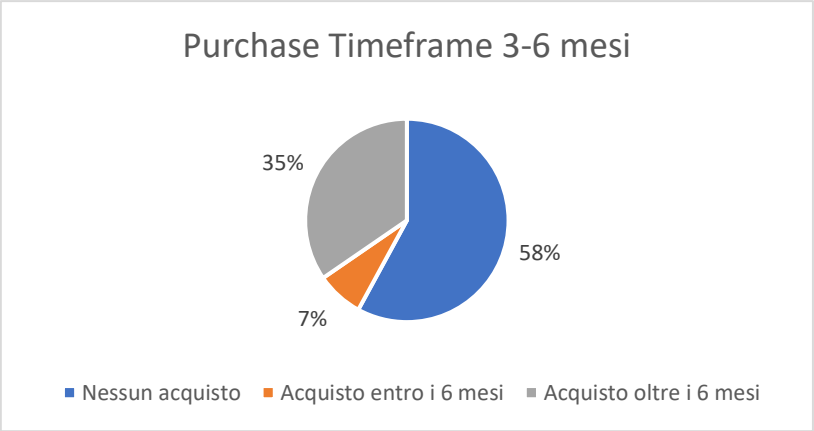

Figura 19: Distribuzione dei lead che hanno dichiarato Purchase Timeframe di 3-6 mesi

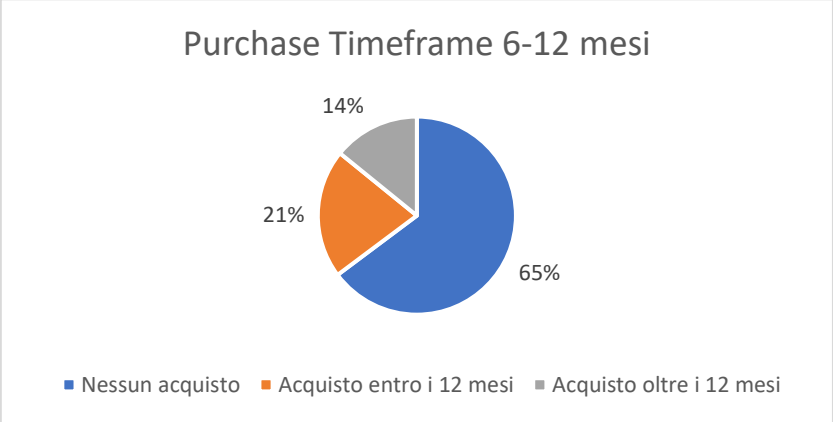

Figura 20: Distribuzione dei lead che hanno dichiarato Purchase Timeframe di 6-12 mesi

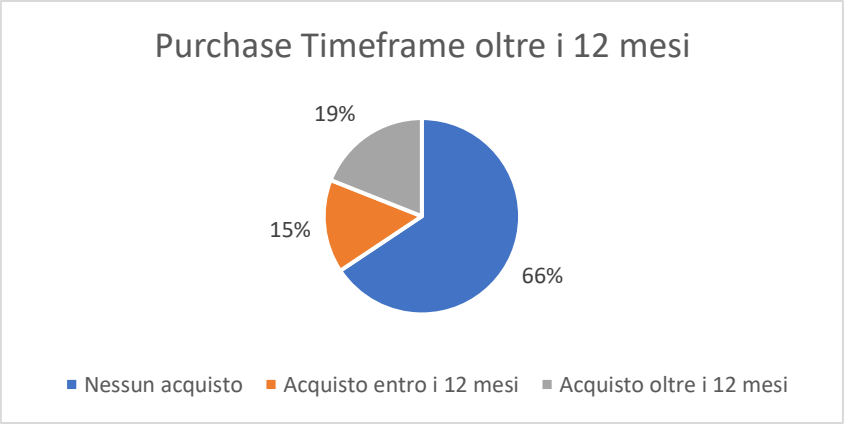

Figura 21: Distribuzione dei lead che hanno dichiarato Purchase Timeframe oltre i 12 mesi

Si nota dai grafici una forte discordanza tra dati effettivi e quanto dichiarato dal lead nel campo *Purchase Timeframe*, che nei peggiori dei casi si traduce in un mancato acquisto. Questa evidenza potrebbe far presupporre poca attendibilità per quest'attributo, ma in realtà il mancato acquisto potrebbe essere dovuto ad altre ragioni: ad esempio, il lead, informandosi più approfonditamente, può aver scelto di effettuare l'acquisto da un competitor, scelta magari anche dettata da tempistiche troppo lunghe per l'attesa di una chiamata dal Dealer. Quello che si vede, però, è che la tendenza a non concludere effettivamente l'acquisto è più contenuta nel caso di lead che manifestino il bisogno di acquistare entro i 3 mesi, mentre va via via aumentando all'allungarsi delle tempistiche. Questo significa che, nonostante solo l'8% abbia rispettato le tempistiche di acquisto dichiarate, comunque un restante 37% ha effettuato l'acquisto, anche se successivamente. Pertanto, nella logica di attribuzione di un punteggio, ha comunque un riscontro empirico premiare chi dichiara intenzione di acquisto 0-3 mesi rispetto a chi seleziona un arco temporale più lungo perché dimostra probabilità di acquisto più alte. Questo è confermato anche da un'analisi che, partendo dalle Won Opportunities, studia la suddivisione dei valori di *Purchase Timeframe* dei lead, come mostrato nella figura sottostante.

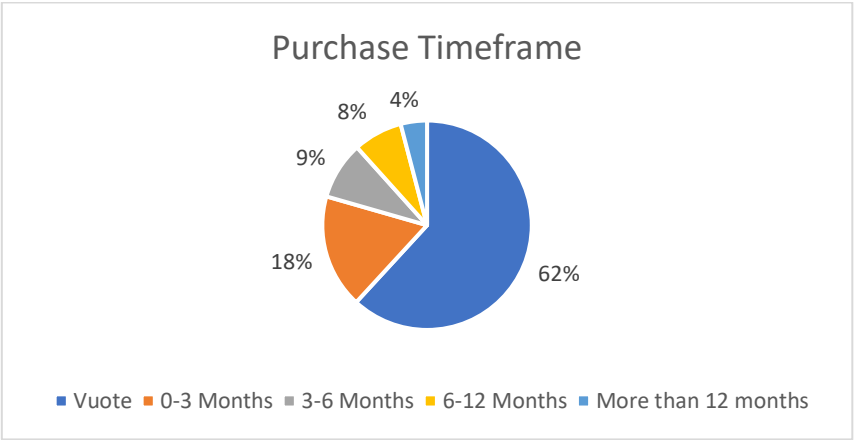

Figura 22: Distribuzione del campo Purchase Timeframe considerando le Won Opportunities

Tabella 10: Considerazioni sull'attributo Purchase Timeframe

Attributo	Grado attendibilità	Relazione con l'acquisto
Purchase Timeframe	Medio	Media

Qualifying Opportunity

Qualifying Opportunity è un campo di tipo Lookup che fa riferimento all'opportunità creata contestualmente alla qualifica del lead. Pertanto, quest'informazione indica se il lead sia già stato qualificato, comprendendo in questa condizione sia la possibilità di auto-qualifica attraverso le campagne digitali, sia la qualifica nelle fiere. Si conviene che ai lead che risultano qualificati debba essere attribuito maggior valore di quelli non qualificati, poiché hanno mostrato un maggior coinvolgimento ed interessamento verso l'azienda.

Guardando le statistiche riportate in figura si vede come ben l'88% dei lead che hanno effettuato l'acquisto arrivavano a fine pipeline con una Qualifying Opportunity, evidenza del fatto che questo campo ha forte impatto nel rilevare la probabilità di acquisto.

Figura 23: Distribuzione del campo Qualifying Opportunity considerando le Won Opportunities

Tabella 11: Considerazioni sull'attributo Qualifying Opportunity

Attributo	Grado attendibilità	Relazione con l'acquisto
Qualifying Opportunity	Alto	Forte

Channel

Anche per quanto riguarda le analisi effettuate sul campo *Channel* di provenienza del lead si è considerata la percentuale di suddivisione rispetto alle "Won Opportunities" (Figura 24).

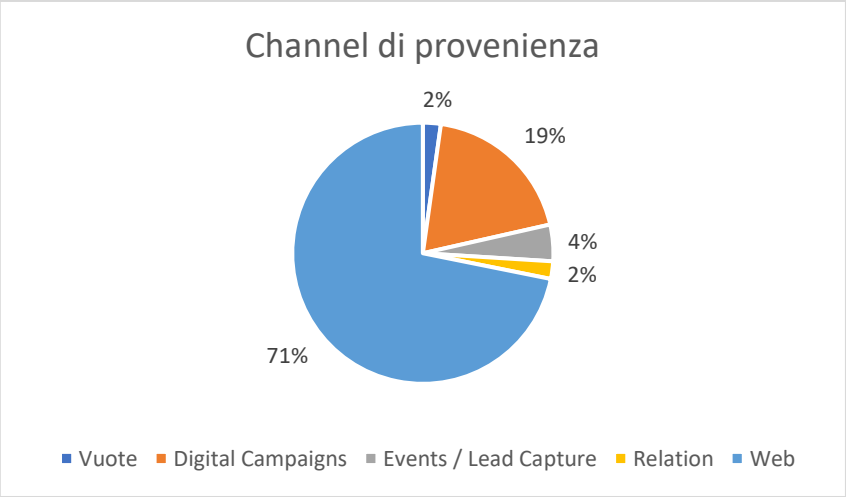

Figura 24: Distribuzione del campo Channel considerando le Won Opportunities

Dal grafico in figura 24 risulta che la grande maggioranza degli acquisti (71%) è costituita dai lead che provengono da una ricerca sul web. Questo dato conferma la teoria, riportata nella sezione dedicata alla letteratura, secondo cui i lead che arrivano ad iscriversi dopo aver effettuato una ricerca attiva tendono ad avere una maggiore propensione all'acquisto rispetto a coloro che sono portati all'iscrizione da moduli proposti sui social networks. Questi ultimi ("Digital Campaigns") sono proprio la seconda fonte di provenienza dei lead che diventano acquirenti, rappresentando il 19% della torta. Gli altri tipi di canale registrano invece scarsi valori. Questa drastica suddivisione dimostra, non solo che per i canali "Web" e "Digital Campaigns" dovrebbe essere assegnato un punteggio positivo, ma anche che può essere considerata una relazione forte tra il campo *Channel* e la probabilità di acquisto.

Tabella 12: Considerazioni sull'attributo Channel

Attributo	Grado attendibilità	Relazione con l'acquisto
Channel	Alto	Forte

Activities Count

L'attributo "Activities Count", associato agli Account, è un campo di tipo Rollup, che effettua il conteggio delle attività che l'Account ha registrato, ovvero le interazioni con il Dealer. In questo caso, l'analisi mira a rispondere alla domanda "*Un maggior numero di interazioni con il Dealer può tradursi in un più concreto interesse del lead verso l'acquisto?*". Perciò, lo studio si è concentrato sul verificare se all'aumentare delle Activities dell'Account crescessero, in media, anche i Won Deals (ovvero il campo che conteggia le Won Opportunities).

La query impiegata è stata così costruita:

```
SELECT a.cnhi_activitiescount, AVG(a.new_wondeals)
FROM Opportunity o, Lead l, Account a
WHERE o.originatingleadid=l.leadid  AND l.parentaccountid=a.accountid
AND o.cnhi_opendate>='2020-07-01'
GROUP BY a.cnhi_activitiescount
```

Ne deriva il seguente grafico (figura 25):

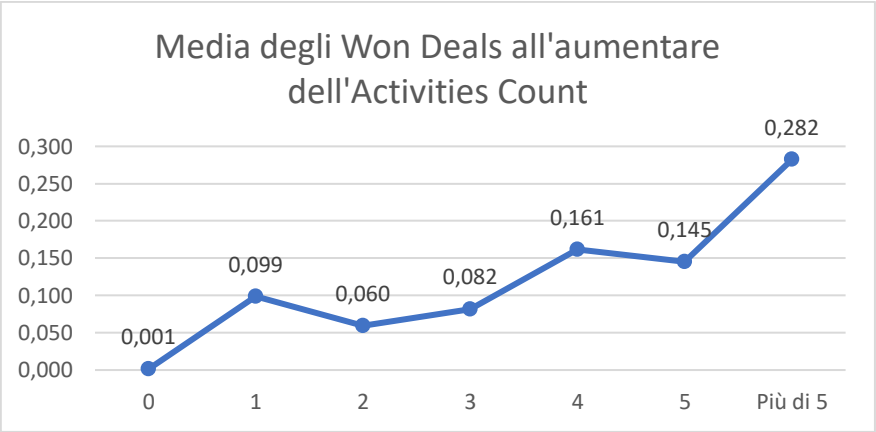

Figura 25: Andamento della media degli Won Deals all'aumentare dell'Activities Count

Sebbene la spezzata mostri un andamento un po' altalenante, si vede chiaramente come l'interpolazione dei dati è di tipo lineare, con una retta che tende a crescere al numero delle attività di interazione con il Dealer. Tuttavia, osservando l'incremento marginale all'aumentare di un'attività non sempre si riscontra un effettivo incremento nella media degli Won Deals, ragion per cui si dichiara una relazione con l'acquisto di entità media-forte.

Tabella 13: Considerazioni sull'attributo Activities Count

Attributo	Grado attendibilità	Relazione con l'acquisto
Activities Count	Alto	Media-forte

Open Deals e Won Deals

Per gli attributi *Open Deals* e *Won Deals* non è stata condotta alcuna analisi specifica in quanto si ritiene di poter immediatamente intuire l'effetto che hanno sulla probabilità di acquisto: un lead che ha almeno un'opportunità aperta (*Open Deals* >0) rileva già la necessità di contattare prioritariamente quel lead per portarla a compimento; mentre, un lead che è già customer (*Won Deals* >0) ha maggiore probabilità di riacquistare, perciò, se ha un'opportunity aperta, deve essere contattato preferenzialmente rispetto a chi non ha mai effettuato un acquisto.

Tabella 14: Considerazioni sugli attributi Open Deals e Won Deals

Attributo	Grado attendibilità	Relazione con l'acquisto
Open Deals+ Won Deals	Medio-alto[9]	Forte

E-mail Clicked

Anche per il campo *E-mail clicked* non sono state effettuate analisi quantitative in quanto il sistema di MS Dynamics 365 non permette al momento l'estrazione di questi dati sottoforma di attributi dell'entità. Si è comunque presupposta una relazione alta con l'acquisto, in quanto i click alle e-mail sono generalmente direttamente proporzionali all'interesse del lead.

Tabella 15: Considerazioni sull'attributo E-mail clicked

Attributo	Grado attendibilità	Relazione con l'acquisto
E-mail clicked	Alto	Forte

Limite nelle analisi effettuate

Le analisi riportate presentano un evidente limite: considerano la relazione di acquisto con ogni attributo in maniera indipendente, mentre il lead rappresenta un individuo a cui *congiuntamente* si riferiscono quegli attributi. Questo significa che l'effetto causale sull'acquisto non dovrebbe essere visto nel singolo attributo, ma nella combinazione di più campi, che, per l'appunto, *congiuntamente* influenzano la probabilità di acquisto. Per spiegare meglio questo concetto si fornisce quest'esempio: la maggior parte dei lead che hanno effettuato l'acquisto hanno attributi con i valori seguenti *Channel*= "Web", *Purchase Timeframe*="0-3 mesi" e *Activities Count*=0. Se si visualizzasse solo l'informazione dell'*Activities Count,* si concluderebbe (erroneamente) che coloro che non hanno avuto alcuna interazione con il Dealer hanno maggior probabilità di acquisto. In realtà, in questo caso, quel valore di *Activities Count* porta all'acquisto solo perché quei lead hanno determinati valori di *Channel* e *Purchase Timeframe*, che, molto più ragionevolmente, sono relazionabili all'acquisto. Ciò significa che considerare separatamente ogni campo potrebbe portare a considerazioni errate. Per migliorare l'attendibilità delle statistiche bisognerebbe essere in grado di condurre un'analisi multivariata, che prenda in considerazione contemporaneamente più fattori, ad esempio sfruttando la

[9] Il grado medio-alto è stato attributo poiché il campo non è gestito dall'utente, ma aggiornato automaticamente dal sistema; tuttavia, come descritto nel paragrafo "Affidabilità del sistema lato-utilizzatori", non sempre le Opportunity vengono chiuse in modo corretto dai Dealer.

regressione logistica. Questo procedimento non è stato percorribile per incompletezza ed inadeguatezza dei dati di input.

Assegnazione dei pesi relativi per ciascun attributo

Sfruttando le analisi effettuate nei paragrafi precedenti, riassunte in tabella, per arrivare ad attribuire dei pesi relativi, si è scelto di effettuare un ordinamento prendendo a riferimento, in primo luogo, la relazione di causa-effetto con l'acquisto; conseguentemente, a parità di questo, assegnare priorità all'attributo con maggior grado di attendibilità. Tenendo però conto dell'inadeguatezza dei dati che hanno portato a tali considerazioni, seppur rimanendo abbastanza fedeli alla classifica ottenuta come esplicitato, il peso relativo di alcuni campi è stato rivisto nell'ottica di coerenza con le ipotesi maturate grazie all'esperienza, suggerite da un confronto finale con la responsabile CRM di CNH Industrial.

Tabella 16:Riepilogo delle considerazioni per ogni campo e assegnazione dei pesi relativi

Campo	Grado di attendibilità	Relazione con l'acquisto	Peso relativo
CNH Industrial Direct Marketing Consent	Alto	Debole	2%
E-mail	Basso	Debole	5%
CNH Industrial Third Marketing Consent	Alto	Debole	5%
Main Phone e Mobile Phone	Basso	Forte	5%
Normalization State	Alto	Forte	5%
E-mail clicked	Alto	Forte	5%
Contact Reason Detail	Medio	Media	8%
Activities Count	Alto	Media-Forte	10%
Open Deals e Won Deals	Medio-Alto	Forte	10%
Channel	Alto	Forte	10%
Created On	Medio-Alto	Forte	10%
Qualifying Opportunity	Alto	Forte	10%
Purchase Timeframe	Medio	Media	15%

In particolare, si sono considerate percentuali molto basse (2% e 5%) per i campi che costituiscono un'informazione che rimane utile solo alla rintracciabilità del lead e poco indicative del suo stato di interesse. Per l'attributo comportamentale *E-mail clicked*, invece, si è scelto di effettuare una valutazione contenuta rispetto alle grosse potenzialità di indicazione qualitativa del coinvolgimento del lead perché questa funzionalità è ancora in sviluppo, il sistema non permette di estrarre gli analytics riguardanti le DEM; pertanto, questo campo è stato preimpostato per quando saranno disponibili aggiornamenti del sistema. Si è stabilito un punteggio massimo del 5% rispetto al totale, che verrà attributo secondo le logiche argomentate nel prossimo paragrafo.

Per quanto riguarda il campo *Contact Reason Detail,* si è optato per un peso relativo dell'8% in quanto l'informazione è inserita solo nel momento in cui il lead chiede in prima persona di avere un contatto con il dealer e quindi può tradursi in un lead motivato; tuttavia, per la maggior parte delle Won Opportunities è stato rilevato come nullo. Perciò, questo comporta un punteggio leggermente più alto rispetto agli attributi visti sopra, ma al di sotto di altri campi più incisivi nella relazione con l'acquisto.

Gli attributi valorizzati al 10% del peso relativo sono, invece, quelli che presentano una forte relazione con l'acquisto, accompagnata da un buon grado di attendibilità del dato; quindi, vengono identificati come molto rappresentativi e necessitano di essere premiati con un adeguato score.

Infine, per il campo *Purchase Timeframe*, si è scelto di andare oltre alle considerazioni emerse dalle analisi quantitative, focalizzandosi, invece, sul significato di quest'attributo, che risulta fondamentale nel costruire una classificazione di precedenza dei lead: è indispensabile per il dealer contattare con priorità coloro che manifestano un'esigenza di acquisto in tempistiche più vicine; perciò, è stato attribuito il peso relativo massimo, ovvero il 15%.

Attribuzione dei punteggi all'interno di ciascun campo

Per semplicità, è stato scelto come punteggio totale 100, in modo che le percentuali assegnate come peso relativo di ciascun campo determino direttamente il punteggio massimo associabile all'attributo. In quest'ottica, all'interno di ciascuno dei campi sono stati stabiliti punteggi per ogni valore alternativo, sfruttando coerentemente le analisi quantitative dei dati. La tabella 17 è da leggere, quindi, in questa chiave: *Per il campo X, se verificata la condizione X=valore, incrementa il punteggio del lead di Y punti.*

Tabella 17: Attribuzione dei punteggi per ogni campo secondo specifica condizione

Campo	**Condizione**	**Punteggio**
CNH Industrial Direct Marketing Consent	=Yes	+2
E-mail	NOT NULL	+5
CNH Industrial Third Marketing Consent	=Yes	+5
Main Phone e Mobile Phone	Entrambi NOT NULL e diversi l'uno dall'altro[10]	+5
Normalization State	=Normalized	+5
E-mail clicked	Ad ogni e-mail cliccata	+1 (fino ad un massimo di +5 in un mese)[11]
Contact Reason Detail	=Get a quotation	+8
	=Demo	+5
Activities Count	Almeno un'Activity con Response=In negotiation	+10
	Almeno tre Activities	+5
	Due Activities[12]	+3
Open Deals e Won Deals	Almeno una Open Opportunity e una Won Opportunity	+10
	Almeno una Open Opportunity	+5
Channel	=Web/Website	+10
	=Event/ Events-Lead Capture	+10 [13]

[10] Si è scelto di non attribuire un punteggio positivo anche a chi ha il solo *Mobile Phone*, in quanto si è visto che nei moduli attualmente utilizzati per l'iscrizione questo campo è obbligatorio.

[11] Incrementare di 1 punto ad ogni click ad una e-mail potrebbe portare a punteggi molto alti, occorrere stabilire un limite superiore. Si è quindi scelto come upper bound quello corrispondente al peso relativo dell'attributo. Per riportare dati sempre recenti, l'espressione di conteggio e-mail clicked verrà settata nel periodo di un mese, poi azzerata all'inizio del mese successivo.

[12] Si guarda a Due Activities perché una è generalmente creata contestualmente all'arrivo nel sistema di un lead che è stato qualificato; perciò, ha senso controllare che ve ne sia stata almeno un'altra.

[13] Il punteggio attribuito ai lead che provengono da eventi, fiere, è più elevato di quanto riscontrabile dai dati empirici riportati nel paragrafo di analisi quantitativa. Si suppone, infatti, che essendo in numero inferiore rispetto a quelli provenienti dal web o da campagne digitali, siano risultati solo in piccola proporzione nel considerare le Won

	=Digital Campaigns	+8
Created On	=Oggi[14]	+10
	=Ieri	+9
	=Due giorni fa o antecedentemente in questa settimana	+8
	=La scorsa settimana	+5
	=Due settimane fa	+2
Qualifying Opportunity	NOT NULL	+10
Purchase Timeframe	=0-3 Months	+15
	=3-6 Months	+10
	=6-12 Months	+5

Suddivisione nelle soglie di rating

La suddivisione nelle varie soglie di rating è stata derivata a partire dai punteggi assegnati secondo il paragrafo precedente.

In particolare, si è partiti con il considerare la categoria di "Hot Leads", ovvero quella più alta della classificazione, che vede il suo limite superiore a 100. Per stabilire, invece, la soglia inferiore si sono considerate le "caratteristiche minime" che un lead di questa fascia dovrebbe avere per essere considerato tra i più promettenti. Pertanto, si è delineato il profilo dei campi che più impattano nella determinazione del punteggio in questo modo:

Tabella 18: Esemplificazione dei punteggi ottenuti nei campi principali per un Hot Lead

Campo	**Punteggio**
Contact Reason Detail	+8
Activities Count	+10
Open Deals e Won Deals	+10
Created On	+10
Qualifying Opportunity	+10

Opportunities, ma, in realtà, si pensi a quanto partecipare a eventi in presenza possa essere indice di forte motivazione e interesse verso l'impresa e i suoi prodotti. Per questo motivo, è stato attribuito anche a questi lead ("pochi ma buoni") il punteggio massimo per l'attributo *Channel*.

[14] Si è considerato arco temporale giornaliero per valorizzare maggiormente gli ultimissimi arrivati, mentre si sono definiti poi punteggi settimanali per periodi antecedenti a due giorni.

Purchase Timeframe	+15

La somma di questi punteggi ha dato un totale di 63 punti, ma per prevedere un margine di tolleranza, la soglia è stata abbassata a 60 punti: questa definisce quindi il limite inferiore per gli "Hot Leads" e quello superiore per i "Warm Leads".

Per la classe "Warm", non riuscendo a definire dei "requisiti minimi", si sono calcolati i punteggi che mediamente potrebbe avere un lead "Warm", guardando sempre agli stessi campi:

Tabella 19: Esemplificazione dei punteggi ottenuti nei campi principali per un Warm Lead

Campo	**Punteggio**
Contact Reason Detail	+5
Activities Count	+5
Open Deals e Won Deals	+5
Created On	+8
Qualifying Opportunity	+10
Purchase Timeframe	+10

La somma di questi punteggi ha valore di 43 punti, che approssimativamente individua un punteggio medio per i "Warm leads". La soglia inferiore per la categoria "Warm" è stata quindi fissata a 30 punti, anche per avere dimensioni di classe abbastanza proporzionate (30-30-40). La fascia leggermente più ampia di "Hot leads" permette di prendere in considerazione un margine cautelativo sul numero di leads da contattare con priorità, per non incorrere nel rischio che venga attenzionata solo una ristretta élite di questi.

Per riassumere, le fasce di rating sono state così individuate:

Cold Leads: 0-29 punti

Warm Leads: 30-59 punti

Hot Leads: 60-100 punti

Implementazione in Microsoft Dynamics 365 Marketing

Microsoft Dynamics 365 Marketing è un ambiente software molto user-friendly per la creazione di un "automated lead scoring" [15]. Nell'apposita sezione "Scoring models", è possibile, infatti, definire diversi modelli di lead scoring, ognuno con i propri punteggi e le proprie soglie di rating. Nel caso applicativo in esame, si è scelto di adottare un unico modello per il momento, costruito secondo quanto descritto nei paragrafi precedenti. La possibilità di aggiungere nuovi modelli può comunque essere sfruttata in futuro per testare contemporaneamente condizioni differenti su alcuni campi e vedere come cambia la classificazione dei lead. Sfruttare questo metodo di confronto può fornire uno strumento che valuta l'adeguatezza stessa del lead scoring. In alternativa, si può anche optare per la costruzione di diversi modelli per catturare informazioni differenti; ad esempio, considerando più campi di tipo "comportamentali", si potrebbe creare un modello che, separatamente, vada ad esaminare solo questi aspetti di interazione del lead e quindi identifichi negli "Hot Leads" gli individui che effettivamente risultano i più attivi, a prescindere dai loro dati espliciti.

Ogni modello viene costruito imponendo delle condizioni, al verificarsi delle quali scatta l'assegnazione del punteggio, come si vede nella figura 26.

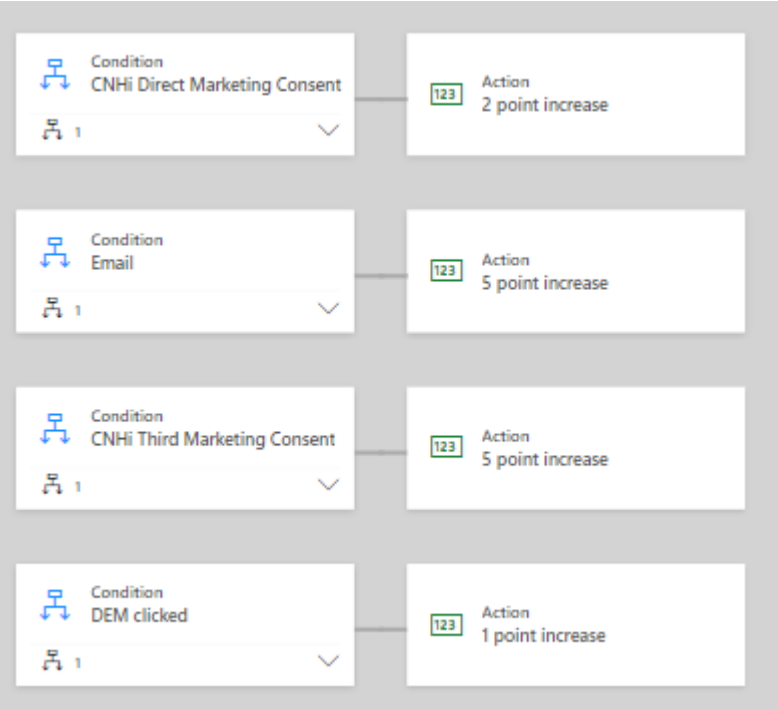

Figura 26: Layout di MS Dynamics 365 per la costruzione del modello di lead scoring

[15] La documentazione è disponibile alla pagina https://docs.microsoft.com/it-it/dynamics365/marketing/score-manage-leads e alle altre ad essa associate.

Per la maggior parte delle condizioni si è presa a riferimento l'entità Lead e si è impostata l'espressione su uno dei suoi attributi. Per i campi non direttamente all'interno del Lead, ma di un'entità ad esso referenziata, è stato necessario, invece, risalire all'entità collegata, purché la relazione sia univoca nella direzione dal Lead all'entità. Ad esempio, è stato possibile raggiungere i campi dell'Account padre, come nel caso del campo *Activities Count*.

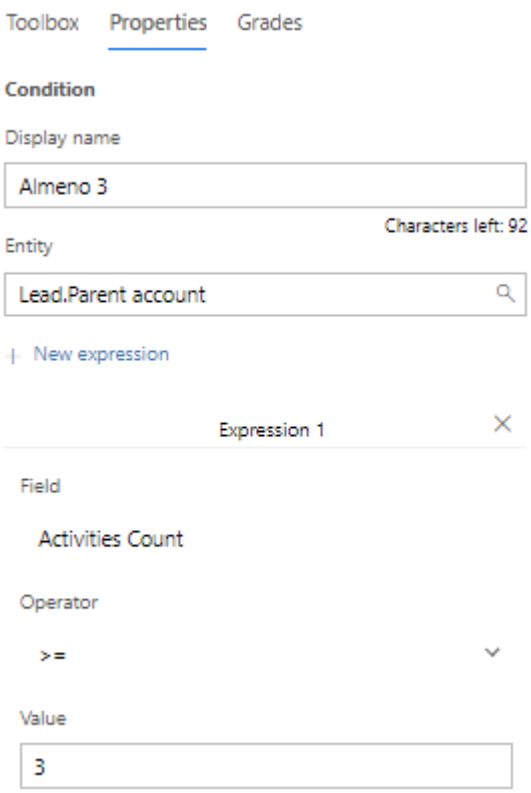

Figura 27: Impostazione di una condizione su un attributo del Parent Account

La scrittura "Lead." nell'*Entity* permette quindi di compiere un salto verso l'entità relazionata. È anche possibile, poi, dal Parent Account compiere un ulteriore hop verso un'entità relazionata. Nel caso in esame, ad esempio, vorremmo riferirci all'Activity per scrivere la condizione di almeno una Activity con Response= "In negotiation"; tuttavia, un Account può essere relazionato a N record di Activity, quindi, non è possibile scrivere un'*Entity* del tipo "Lead.ParentAccount.Activity". Una soluzione a questo problema è quella di creazione di un nuovo campo, di tipo Rollup, che sia costruito come l'attributo di Activities Count, ma con la condizione specifica di Response= "In negotiation". Analogamente, anche per il conteggio delle Won Opportunities dell'Account è stato creato un nuovo campo "Won Deals", così definito:

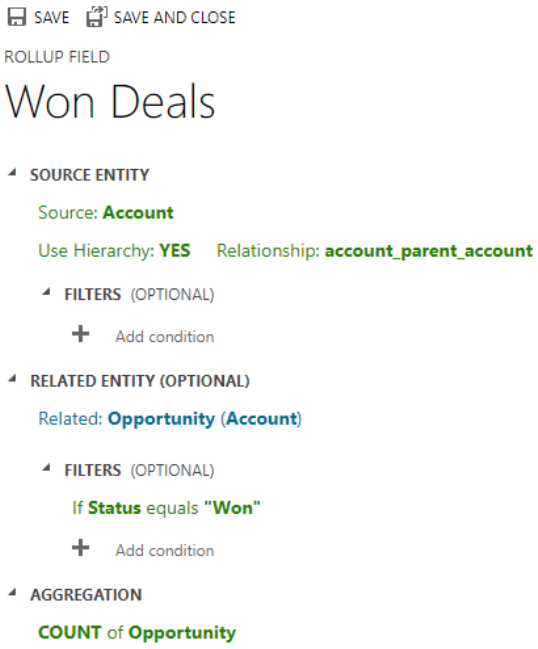

Figura 28: Creazione del campo Won Deals per il conteggio delle Won Opportunities di un Account

In questo modo, invece che risalire all'entità correlata (Activity o Opportunity), si crea un campo all'interno dell'Account, che si auto-aggiorna se avvengono dei cambiamenti, e che permette di calcolare l'espressione direttamente sul Parent Account:

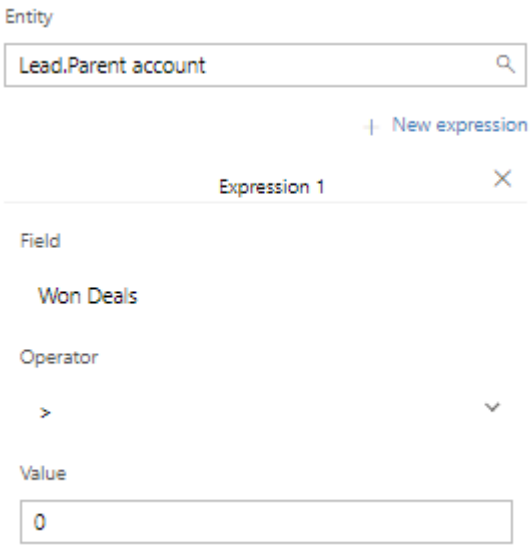

Figura 29: Utilizzo del campo Won Deals nell'impostazione della condizione per l'attribuzione del punteggio

Anche nel caso del campo *E-mail clicked*, che come precedentemente spiegato, può solo riferirsi al Contact, in quanto unica entità che può costituire un segmento di partenza per un percorso-cliente di e-mail marketing. Pertanto, la condizione viene così impostata:

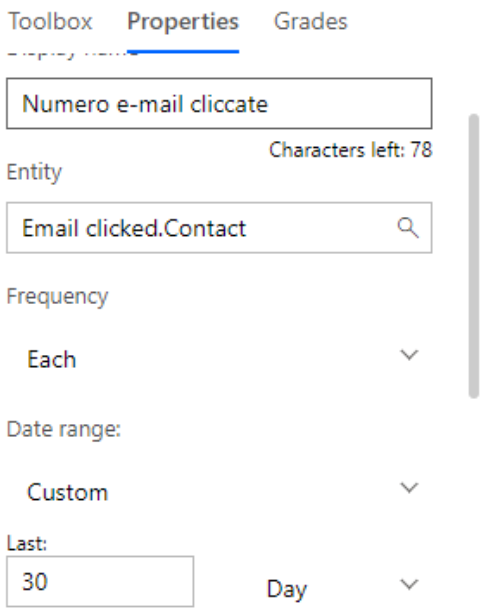

Figura 30: Impostazione della condizione sulle e-mail cliccate per l'attribuzione del punteggio

Infine, anche la condizione sul campo *CreatedOn* ha necessitato di una soluzione alternativa per essere implementato adeguatamente. In particolare, essendo un campo di tipo Date, l'espressione era impostata in modo vincolato a controllare con l'operatore "<" l'antecedenza rispetto ad una data definita dal calendario, come si vede nella figura sottostante.

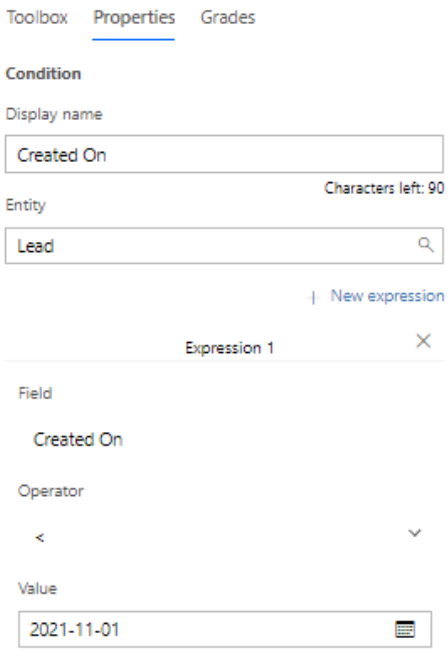

Figura 31: Vincolo per la creazione di una condizione su un campo di tipo Date

Quest'impostazione risulta molto sconveniente allo scopo di costruire la condizione su tempistiche relative, calcolate rispetto alla giornata odierna (oggi, ieri…) e che quindi dovrebbero far in modo di aggiornare la data imposta da calendario. Pertanto, anche in questo caso, la soluzione adottata è stata la creazione di un nuovo campo, che calcoli il tempo intercorso dalla data di creazione al "Time Now" (aggiornato in ogni istante dal sistema), restituendo un intero che rappresenta il numero di giorni. Nelle immagini sottostanti si riporta la costruzione del campo e la condizione settata nel modello di lead scoring.

Figura 32: Creazione di un nuovo campo per trasformare l'informazione in dato numerico

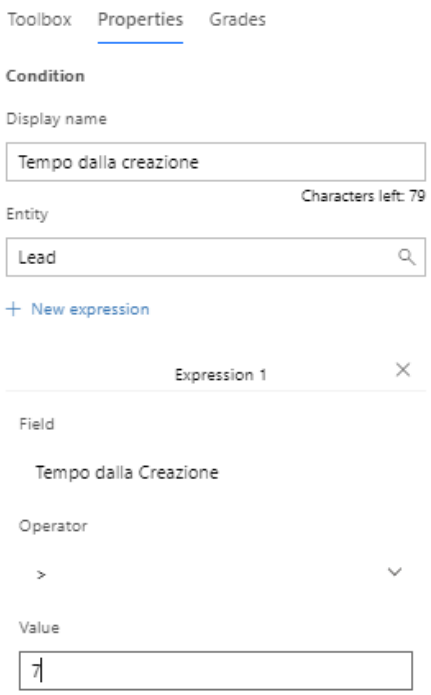

Figura 33: Impostazione della condizione sul Tempo dalla Creazione per evitare i vincoli del campo di tipo Date

Impostate tutte le condizioni, è possibile creare le fasce di classificazione nella sezione "Grades". L'ultimo passaggio da effettuare è la messa in Live, che andrà a calcolare in automatico i punteggi per ciascun lead. Questi sono consultabili nella sezione "Related"->"Lead scores" all'interno del modello e permettono una visualizzazione immediata dei lead, del loro score e del grade, in modo che siano anche ordinabili secondo uno score decrescente, creando direttamente una classifica di priorità.

Monitoraggio e KPIs

A conclusione del lavoro di progettazione ed implementazione del sistema di lead scoring, l'ultima fase da affrontare è la costruzione di un efficace strumento di monitoraggio, che possa misurare i risultati ottenuti grazie al modello in termini di miglioramento delle prestazioni; nonché rilevare eventuali cambiamenti delle tendenze degli attributi rispetto all'acquisto e adottare opportune tecniche di adattamento del sistema alle variazioni.

Monitoraggio per rispondere ai cambiamenti

In primo luogo, è necessario effettuare una costante rilevazione dei dati raccolti sui lead che hanno portato a completamento l'acquisto, così come presentato nel paragrafo "Analisi quantitativa dei dati", ovvero quegli stessi grafici devono essere periodicamente costruiti per osservare se le statistiche raccolte rispecchino le ipotesi adottate per la costruzione dei punteggi del modello di lead scoring e, in caso di discrepanze, portare il sistema ad allinearsi con i dati. Questo tipo di monitoraggio è utile non solo allo scopo di revisione continua dei punteggi, ma anche per guidare altre decisioni di investimento del business attraverso una sempre più dettagliata profilazione delle "Buyer Personas". Ad esempio, l'informazione sul *Channel* può rivelare su quale canale l'impresa ha più incentivo a creare campagne di marketing e quindi portarla a concentrare le proprie risorse nell'escogitare strategie specifiche per quel tipo di mercato di provenienza. Naturalmente, per questo tipo di supervisione prolungata nel tempo, è particolarmente indicato utilizzare dei software di reportistica, che permettano una visione immediata dei dati, aggiornati in tempo reale attraverso sincronizzazione automatica con il database. Nel caso di CNH Industrial è stata impostata la visualizzazione degli analytics sulla piattaforma "Qlik Sense".

Verifica dell'adeguatezza dello strumento di lead scoring

Il secondo punto da prendere in considerazione durante tutto il ciclo di vita del sistema è la misurazione dell'adeguatezza e dell'efficacia dello strumento di lead scoring. È quindi necessario, per costruire indicatori adatti, porsi ex-ante degli obiettivi concreti da raggiungere grazie al modello. In particolare, segue una descrizione dei principali con relativi indicatori di misurazione proposti.

- **Aumento delle vendite**: riuscire a dare priorità ai lead che hanno dimostrato un maggiore coinvolgimento ed interesse, secondo gli attributi scelti nel modello di lead scoring, dovrebbe servire ad accorciare le tempistiche di contatto del Dealer verso questi individui/aziende e quindi creare un'interazione più immediata, che può favorire la probabilità di acquisto. In aggregato quindi, a parità di Opportunity totali, ne dovrebbe risultare un aumento delle Won Opportunities. L'indicatore scelto, quindi, per misurare tale impatto è $\frac{\#di\ Won\ Opportunities}{\#\ di\ Total\ Opportunities}$. Ci si aspetta quindi che, considerando un intervallo di tempo (ad esempio trimestrale o semestrale) a partire dalla messa in Live del sistema di lead scoring, si inizi a registrare una percentuale maggiore rispetto alla soglia presa come riferimento da una fotografia (sullo stesso arco temporale) della situazione attuale, ovvero antecedente all'avvio del lead scoring.
- **Contrazione del tempo di vendita medio**: similmente, considerando tempistiche di contatto Dealer-Lead studiate secondo le esigenze del potenziale acquirente, ne deriva, nel complesso, una riduzione del tempo medio di acquisto dei lead. Di fatto, non cambierà per il lead l'arco temporale in cui si colloca il bisogno del prodotto, ma, la reazione più tempestiva nel contattarlo, ottimizza le tempistiche per portare la vendita a compimento effettivo, evitando quindi che il potenziale cliente passi ad un competitor a causa della prolungata attesa. L'indicatore da monitorare, pertanto, risulta dalla media del campo *TempoDiAcquisto* (in giorni) per ogni lead e la prospettiva di miglioramento delle prestazioni è ravvisabile dal suo decremento. Ovviamente, anche se non quantificabile, l'effetto consequenziale al miglioramento di questo indicatore ha un forte impatto anche reputazionale, in quanto il cliente che vede appagati i propri bisogni con rapidità percepisce un più alto grado di soddisfazione, che può creare un senso di fidelizzazione verso l'impresa e quindi anche una più spiccata tendenza a riacquistare, ricollegabile all'obiettivo precedente di aumento delle vendite.
- **Aumento dell'efficacia nella pipeline di vendita**: il sistema di segmentazione dei lead nei rating "Hot", "Warm" e "Cold" dovrebbe essere in grado di discriminare i lead secondo la propensione ad acquistare; pertanto, contattare in modo prioritario i lead "Hot" dovrebbe garantire una maggior probabilità di successo nella vendita. Si prevede quindi che, a parità di lead entranti nel sistema, quelli che sono classificati come "Hot" abbiano un più elevato numero di "Won Opportunities" rispetto a tutti gli altri lead. Questo concetto può essere tradotto in forma di disequazione come

$$\frac{\#\ Won\ Opportunities\ generate\ dagli\ Hot\ leads}{\#\ Hot\ leads} > \frac{\#\ Won\ Opportunities\ generate\ da\ Warm\ Leads+Cold\ Leads}{\#\ Warm\ Leads+Cold\ Leads}$$

Se la condizione è verificata, assicura una corretta attribuzione del punteggio per i lead "Hot", in quanto, secondo la definizione stessa di questi come più promettenti per la vendita, essi dovrebbero generare un numero di acquisti maggiore rispetto ai lead "Warm" e "Cold". Inoltre, si ricorda che non tutti i lead partecipano all'assegnazione dei punteggi di lead scoring in quanto il prerequisito del sistema di MS Dynamics 365 è quello di includere solo quelli che sono relazionati ad un Account/Contatto Padre. Per questi ultimi non si dispone di informazioni sufficienti per fornire un ordinamento relativo rispetto al rating, ma si può ragionevolmente presupporre che questi lead, non essendo stati qualificati, siano stati identificati come scarsamente interessati all'acquisto. Pertanto, si potrebbero collocare in una posizione di inferiorità rispetto ai "Cold Leads". In quest'ottica, la disequazione sopra riportata può comprendere nel lato destro, in aggiunta ai "Warm" e "Cold" leads, anche il conteggio dei lead senza score e delle relative "Won Opportunities".

Queste indicazioni devono fornire una guida per un adeguato controllo del sistema di scoring: infatti, se la condizione risultasse non soddisfatta, ne deriverebbe che l'attribuzione dei punteggi e delle soglie di classificazione è da rivedere. Naturalmente, ad un primo approccio questo potrebbe accadere, in quanto la costruzione del modello non è un metodo esatto, ma deriva da considerazioni ipotetiche o, come in questo caso, fondate su dati non completamente attendibili, che esigono di essere testate sperimentalmente, anche con molteplici tentativi, per riuscire a discernere gli effetti dei singoli fattori.

Nella stessa ottica, si può impostare la condizione di monitoraggio sui "Warm Leads":

$$\frac{\#\ Won\ Opportunities\ generate\ dagli\ Warm\ leads}{\#\ Warm\ leads} > \frac{\#\ Won\ Opportunities\ generate\ da\ Cold\ Leads}{\#\ Cold\ Leads}$$

anche in questo caso modificabile con l'aggiunta dei lead senza score.

Questo tipo di indicatore è altresì utilizzabile per effettuare un confronto tra la situazione attuale senza scoring e quella post-scoring. La considerazione che ne dovrebbe derivare è la seguente:

$$\frac{\#\ Won\ Opportunities\ generate\ dagli\ Hot +\ Warm\ leads}{\#\ Hot + Warm\ leads} > \frac{\#\ Won\ Opportunities\ generate\ dai\ Total\ Leads(pre - lead\ scoring)}{\#\ Total\ Leads(pre - leaad\ scoring)}$$

A sinistra della disequazione il tasso di acquisto per lead degli "Hot" e "Warm" nella situazione post-lead scoring, misurato in un determinato arco temporale, deve essere maggiore del tasso di acquisto dei lead totali nello stesso intervallo di tempo, considerato pre-lead scoring, ovvero non potendo distinguere i lead, ma equiparandoli allo stesso livello.

KPIs per la misurazione delle prestazioni

Verificata l'adeguatezza del sistema di lead scoring in relazione all'efficacia dello strumento e alla coerenza dei punteggi con i dati reali, si possono utilizzare degli indicatori chiave di performance (KPIs) derivanti dalle informazioni del lead scoring per poter comprendere al meglio con quali lead stia trattando l'impresa. In particolare, può essere indicato per la funzione marketing aziendale rilevare i seguenti aspetti:

- *A che punto della pipeline di vendita si trova la maggior parte dei lead dell'impresa?* Per rispondere a questa domanda, basti pensare che il rating del lead scoring fornisce un ordinamento di priorità legato alla probabilità di acquisto del lead; pertanto, esso può essere paragonabile alle fasi della pipeline di vendita del modello AIDA, visto nel paragrafo "Marketing funnel". Si può, infatti, associare il lead "Cold" ad un individuo/azienda che è ancora lontano dall'acquisto, ovvero nella fase di "Interest". Il lead "Warm" è, invece, nella fase di "Desire"; infine, il lead "Hot" è vicinissimo alla fase finale di "Action". Il modello di lead scoring si configura, dunque, come uno strumento che può tradurre in modo semplice e quantificabile queste informazioni qualitative sull'interesse del lead e sua prossimità all'acquisto, che altrimenti non sarebbe possibile riuscire a estrapolare in modo chiaro e definito. Riuscire a comprendere il posizionamento della maggioranza dei lead comporta che l'impresa è consapevole di quali potenziali clienti ha di fronte: *sono iscritti inattivi e con poco interesse all'acquisto oppure con un forte bisogno del prodotto e magari già clienti dell'impresa?* Se, ad esempio, prevale la prima tipologia, bisognerà investire su strategie di marketing per potenziare la Lead Generation e il Lead Nurturing; mentre, una percentuale maggiore della seconda categoria, presuppone di potenziare il team delle vendite per riuscire a prendere in carico tutta la domanda e, eventualmente, concentrarsi su strategie di fidelizzazione sempre più attrattive nei confronti del cliente. Per riuscire a catturare queste informazioni può essere funzionale utilizzare come indicatore la media dei punteggi di lead scoring (Average Lead Score). Questo score medio andrà immediatamente a riferirsi ad una categoria del rating.
- *Qual è la percentuale di Prospect che l'impresa deve gestire tempestivamente?* Questa percentuale può essere immediatamente calcolabile come frazione di "Hot leads" sul numero di lead totali. Analogamente a quanto spiegato nel punto precedente, se l'indicatore presenta una percentuale troppo bassa, questo comporterà decisioni di investimento sul marketing, mentre una alta sulla funzione sales. In quest'ultimo caso, è consigliabile anche rivedere il sistema di punteggi in quanto, se una grossa percentuale di lead si trovasse nella fase "Hot",

questo dato non sarebbe più molto discriminante nel contattare con priorità alcuni lead rispetto ad altri. Perciò, in quest'ottica, sarebbe opportuno o spostare la fase di "Hot" ad una soglia di punteggio più alta, o, in alternativa, creare una nuova categoria "Very hot", che fornisca un ordinamento ulteriore.

- *Quale valore possono generare i lead per l'impresa?* È possibile stimare il ricavo medio atteso dai lead registrati a sistema, secondo questa formula:

$$Ricavo\ medio\ di\ una\ Won\ Opportunity * \frac{\#Hot\ Leads}{\#Total\ Leads}$$

  dove il ricavo medio di una Won Oppurtunity è calcolato come media dei ricavi di tutte le Won Opportunities precedenti e rappresenta quindi il fatturato che si otterrebbe nel portare a compimento una vendita; mentre, la frazione rappresenta la probabilità che quella vendita si realizzi, considerata come la percentuale di "Hot Leads" presenti nel sistema. Quest'indicatore fornisce una stima non puntuale, ma approssimativa, però può essere comunque molto utile a "monetizzare" le informazioni sui lead, fornendo un'idea di quale valore economico possano apportare all'impesa.

Prospettive future all'interno dell'impresa

Volendo delineare le prospettive future per lo strumento di lead scoring all'interno dell'impresa CNH Industrial, si possono intravedere miglioramenti ed estensioni nel suo utilizzo, nonché ottimizzazione del modello di punteggi, attraverso progressivi raffinamenti.

In particolare, una volta costruito il sistema di lead scoring, anche superando positivamente i controlli di adeguatezza esplicitati nel paragrafo "Verifica dell'adeguatezza dello strumento di lead scoring", si può sempre ipotizzare di poter migliorare il modello. Infatti, se questa prima costruzione è stata effettuata procedendo con una metodologia manuale per l'esiguità di dati e la loro incompletezza, instaurato un corretto approccio alla qualità di dati, come suggerito nel paragrafo "Affidabilità del sistema lato-utilizzatori", risulta più accurato creare modelli predittivi maggiormente guidati dai dati, attraverso algoritmi di machine learning, come quelli presentati nella *Prima Sezione-Letteratura*. È ipotizzabile, quindi, creare classificazioni predittive indipendenti secondo il modello di Random Forest, Stochastic Grandient Boosted Trees e Regressione Logistica[16] (l'algoritmo dell'albero decisionale può non essere considerato perché poco robusto) utilizzando un training dataset considerato sufficientemente ampio ed attendibile. Ottenuti quindi i diversi metodi predittivi, per testarli, è possibile sfruttare la flessibilità offerta dal software MS Dynamics 365 Marketing, che

[16] Il riferimento a questi deriva dai paragrafi "Random Forest", "Gradient Boosted Trees", "Regressione Logistica", trattati nella *Prima Sezione- Letteratura*

permette la creazione e messa in live di più modelli di punteggio contemporaneamente. Pertanto, assegnando in ogni modello le regole di punteggio derivanti da ciascun algoritmo, per ogni lead automaticamente si aggiornerà il suo punteggio in ciascuno dei modelli. Questi potranno attribuirgli la stessa classe oppure una classe differente nei diversi modelli. Lasciando operare tutti i modelli di lead scoring per un periodo di tempo sufficientemente ampio da raccogliere dati da testare (minimo tre mesi), si procederà con la valutazione dei modelli (compreso quello creato manualmente) attraverso gli indicatori di prestazione e si sceglierà il migliore. Al momento MS Dynamics 365 Marketing è però molto rigido sul settaggio dei punteggi: permette solo la compilazione della condizione e del punteggio da un layout predefinito e non modificabile, mentre sarebbe opportuno avere la possibilità di includere l'algoritmo direttamente per la costruzione del modello. Si prospetta quindi quest'ipotesi di lavoro nell'ottica di un possibile aggiornamento del software che consenta, in alternativa al modello predefinito, l'inserimento di codice per predizione dell'acquisto. Solo nel caso si utilizzi un algoritmo di regressione logistica, il risultato, con la formulazione di una funzione multivariata, fornisce, attraverso la stima dei coefficienti delle variabili esplicative, i pesi di ciascun attributo, che è possibile tradurre in punteggi per il modello di lead scoring. Considerando poi che il risultato di una predizione di questo metodo è una probabilità di acquisto, si possono calcolare le soglie per la distinzione nelle diverse classi "Hot", "Warm", "Cold", guardando alla curva ROC. Analogamente anche per l'algoritmo Stochastic Gradient Boosted Trees, essendo il risultato una probabilità di acquisto, si può procedere con la stessa formulazione delle categorie. Al contrario, l'algoritmo di Random Forest, impostato come modello che studia la classificazione in "Yes" o "No" (riferiti all'evento di conversione del lead a customer), andrebbe a determinare una suddivisione in due sole classi, che possiamo identificare negli estremi del nostro rating, ovvero contrassegnando i lead con classe "Yes" come "Hot leads" e i lead con classe "No" come "Cold leads".

In alternativa all'utilizzo di algoritmi di machine learning, è anche possibile testare diversi modelli costruiti sempre con approccio manuale, ma con diverse ipotesi per l'assegnazione dei punteggi. Per un determinato periodo temporale, questi vengono lasciati operare simultaneamente e, al termine di questo test, se ne misurano i risultati per vedere quale dei modelli ha predetto nel miglior modo la classificazione. In particolare, si guarderà, in relazione alle sole "Won Opportunities" quale dei modelli ha correttamente predetto il maggior numero di "Hot leads".

In entrambi i casi, comunque, è opportuno prevedere una revisione periodica del modello di punteggi. Ad esempio, nel caso dell'adozione di un algoritmo di machine learning, semestralmente, si può ipotizzare di provare a ricostruire l'algoritmo dando come "training dataset" lo storico di dati degli ultimi sei mesi per essere sempre più preciso, anche in relazione ai cambiamenti. Nel caso, invece,

dei metodi di costruzione manuale, l'approccio da adottare è una continua supervisione dei grafici elaborati su Qlik Sense, come riportato nel paragrafo "Monitoraggio per rispondere ai cambiamenti". Nel caso di variazioni di uno specifico attributo, è opportuno prevedere un aggiustamento nella stessa direzione anche nel modello di punteggi.

Infine, il sistema di lead scoring può essere impiegato anche per scopi differenti dal creare una classificazione di priorità per contattare telefonicamente i lead. Ad esempio, come accennato nel paragrafo "Presentazione del caso", Microsoft ha dichiarato che nel 2022 (a pochi mesi, quindi, dalla data odierna) rilascerà per MS Dynamics 365 Marketing una nuova funzionalità che permetterà di creare segmenti di lead e utilizzarli come punto di partenza per i *customer journeys*. Questa introduzione permetterà di inviare campagne e-mail e strutturare un percorso di condizioni *if/else* direttamente sull'entità Lead, senza dover risalire al suo Parent Contact. In quest'ottica, potrebbe essere funzionale creare dei segmenti di lead proprio seguendo la classificazione risultata dal lead scoring, utilizzando specifici template di e-mail differenziati secondo la classe di riferimento, rispetto al tono, ai contenuti e alla frequenza di invio.

Conclusioni

In conclusione, la costruzione di un modello di lead scoring non può seguire un metodo esatto poiché basata su logiche predittive differenti a seconda dei casi e mutevoli nel tempo. Per questo motivo, ogni impresa deve individuare i fattori determinanti per la classificazione dei lead attraverso un'analisi del business di riferimento e delle logiche interne riguardanti i processi di marketing e di vendita. La creazione delle giuste logiche di attribuzione dei punteggi deve, quindi, seguire lo studio di dati raccolti internamente all'azienda circa i comportamenti di conversione dei lead e derivarne schemi predittivi, o interpretando autonomamente i dati studiati, o, in alternativa, affidandosi ad algoritmi di machine learning, che permettono un approccio più data-driven.

Il presente lavoro di tesi esplora alcune delle metodologie più frequentemente impiegate per la strutturazione di modelli di lead scoring, illustrandole da un punto di vista teorico, allo scopo di suggerire alle imprese che si approcciano a questo tipo di problema diverse soluzioni di progettazione.

Il caso applicativo preso in esame fornisce un'esemplificazione per quanto riguarda la costruzione manuale, poiché, l'esiguità dei dati raccolti dall'impresa e la loro scarsa affidabilità dal punto di vista della correttezza, hanno reso non percorribile l'ipotesi di affidarsi a metodi completamente fondati sui dati come i predictive analytics di cui citato sopra. Essendo la prima progettazione del sistema di lead scoring per l'impresa, l'adozione della procedura manuale risulta comunque una valida soluzione implementativa, che può coniugare i dati raccolti con le previsioni derivanti dall'esperienza del team di marketing. In relazione a quest'ultime si sono esplicitati quelli che possono essere i criteri di scelta per identificare gli attributi del lead scoring e per attribuire a questi un punteggio di valutazione, da cui è stata composta la suddivisione in classi. Con il passare del tempo e una sempre più ampia raccolta di dati, si potranno proporre raffinamenti del modello, anche in risposta a variazioni rilevate grazie ad un continuo monitoraggio sia del comportamento dei consumatori sia dell'efficacia dello strumento di lead scoring.

Nelle prospettive future per l'impresa, infine, si delineano ipotesi di lavoro che puntino ad un miglioramento continuo del sistema, tenendo in considerazione i vincoli imposti dall'ambiente software utilizzato e i suoi possibili aggiornamenti.

Bibliografia

- Đorđević, "Optimization of Digital Marketing Processes Through Modeling of Lead-Scoring," in Sinteza 2019 - International Scientific Conference on Information Technology and Data Related Research, Belgrade, Singidunum University, Serbia, 2019, pp. 32-37. doi:10.15308/Sinteza-2019-32-37
- Lindahl, Elin- "A qualitative examination of lead scoring in B2B marketing automation, with a recommendation for its practice."- KTH, School of Computer Science and Communication (CSC). 2017 (English)Independent thesis Advanced level
- Khadar, M. (2007). Influence of Artificial Intelligence on Digital Marketing. *College*, 25.
- Rosenbröijer, C. J. (2014). Customer Relationship Management and business analytics: a lead nurturing approach. *Proceedings of DYNAA*, *5*(1).
- Patrutiu-Baltes, L. (2016). Inbound Marketing-the most important digital marketing strategy. *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V*, *9*(2), 61.
- Nygård, R., & Mezei, J. (2020). Automating lead scoring with machine learning: An experimental study.
- Skog, D. A., Wimelius, H., & Sandberg, J. (2018). Digital disruption. *Business & Information Systems Engineering*, *60*(5), 431-437.
- Camilleri, M. A. (2020). The use of data-driven technologies for customer-centric marketing. *International Journal of Big Data Management*, *1*(1), 50-63.
- Ricardo Chalmeta, Methodology for customer relationship management, Journal of Systems and Software, Volume 79, Issue 7, 2006, Pages 1015-1024.
- Rodrigues, J. D. D. S. S. (2020). Automated lead scoring system: a case study of a Portuguese startup.
- Abdul Ghaaliq Lalkhen, MB ChB FRCA, Anthony McCluskey, BSc MB ChB FRCA, Clinical tests: sensitivity and specificity, *Continuing Education in Anaesthesia Critical Care & Pain*, Volume 8, Issue 6, December 2008, Pages 221-223, https://doi.org/10.1093/bjaceaccp/mkn041

Sitografia

https://www.selligent.com/it/blogs/pillole-di-marketing/real-time-marketing-tipologie-e-vantaggi

https://www.magnews.it/lead-nurturing-cose-strategie

https://www.netstrategy.it/seo/indicizzazione-seo-cose-e-come-funziona

https://www.netstrategy.it/sem-specialist

https://irp-cdn.multiscreensite.com/4df029b9/files/uploaded/Informativa%20cookies%20sito%20web%20Veronesi%20Srl-Italia%20On%20Line.pdf

https://www.italiaonline.it/risorse/marketing-funnel-cos-e-e-come-crearne-uno-efficace-503

https://www.digital-coach.it/blog/case-histories/marketing-funnel/

https://www.insidemarketing.it/glossario/definizione/lead-nurturing/

https://www.dmep.it/guida-definitiva-al-lead-scoring#pagina04

https://blog.advmedialab.com/lead-scoring

https://eur-lex.europa.eu/legal-content/it/TXT/?uri=CELEX:32016R0679

https://it.sendinblue.com/blog/gdpr-email/

https://blog.mailup.it/2016/06/single-opt-in-e-double-opt-in-due-metodi-a-confronto/

https://repositorio-aberto.up.pt/bitstream/10216/133060/2/433668.pdf

https://docs.microsoft.com/it-it/dynamics365/marketing/lead-lifecycle

https://protezionedatipersonali.it/cookie-law

https://it.sendinblue.com/blog/marketing-automation-kpi/

https://www.datalytics.it/10-kpis-da-monitorare/

https://www.affde.com/it/marketing-kpis-to-measure-for-better-lead-management.html

https://leadchampion.com/blog/5-metriche-di-lead-generation/

https://blog.digitalbuildingblocks.it/blog/i-kpi-da-tenere-docchio-per-le-tue-campagne-di-lead-generation

https://www.ictsviluppo.it/blog/che-cosa-crr-customer-retention-rate

https://www.tableau.com/it-it/learn/articles/kpi-key-performance-indicators-quickstart-guide

https://www.teamlewis.com/it/rivista/kpi-cosa-sono/

https://dashthis.com/blog/top-5-most-used-kpi-for-lead-generation/

https://www.artunlimitedusa.com/blog/calculate-lead-value

https://universeit.blog/crm-analitico-operativo-strategico-collaborativo/

http://www.ce.unipr.it/~medici/geometry/node104.html

https://www.researchgate.net/file.PostFileLoader.html?id=582dc4f9eeae392fc85afa91&assetKey=AS%3A429426035957760%401479394553209

https://towardsdatascience.com/understanding-random-forest-58381e0602d2

https://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting

https://jerryfriedman.su.domains/ftp/stobst.pdf

http://www.riani.it/DMM/DMM_W9e10.pdf

https://ichi.pro/it/matrice-di-confusione-spiegato-58157046543340

https://machinelearningmastery.com/failure-of-accuracy-for-imbalanced-class-distributions/

https://support.google.com/analytics/topic/10333392?hl=it&ref_topic=9143232

https://www.w3schools.com/php/php_sessions.asp

https://amslaurea.unibo.it/10415/1/Metodologie_di_apprendime.pdf

https://docs.microsoft.com/it-it/dynamics365/marketing/score-manage-leads

https://docs.microsoft.com/it-it/dynamics365/marketing/set-up-lead-scoring

https://docs.microsoft.com/it-it/dynamics365/marketing/mkt-settings-lead-score-options

https://docs.microsoft.com/it-it/dynamics365/marketing/lead-lifecycle