



UNIVERSITÀ DI PISA

Corso di Laurea Magistrale in Informatica Umanistica

TESI DI LAUREA MAGISTRALE

**Shapley attribution per il search advertising: un
modello basato sullo stemming di termini di ricerca**

Candidato: *Mariasilvia Vanni*

Relatore: *prof. Nicola Ciaramella*

Anno Accademico 2020 - 2021

Indice

1. Introduzione	5
2. SEO e SEM: tecniche per aumentare la visibilità di un sito web	10
2.1 La visibilità nei risultati dei motori di ricerca	10
2.2 Cos'è la SEO e quali sono i passaggi che la costituiscono	12
2.3 On-Page SEO	13
2.3.1 Scelta del nome del dominio	13
2.3.2 Scelta delle keyword	14
2.3.3 Tag HTML	15
2.3.4 Heading tag	17
2.3.5 Attributo ALT nelle immagini	18
2.4 Off-Page SEO	18
2.4.1 Link Building e Link Earning	18
2.4.2 Menzione del Brand	19
2.4.3 Blog e Social Network	20
2.4.5 Altre strategie per aumentare la visibilità online	21
2.4.6 Google My Business	22
2.5 Cos'è la SEM e quali sono i passaggi che la costituiscono	23
2.6 Elementi costituenti della SEM	24
2.7 Google Ads	24
2.8 Oltre a Google Ads, altri tool per la pubblicità online	27
2.9 Differenze tra attività SEO e SEM	29
3. L'importanza delle keywords	33
3.1 L'algoritmo di ricerca di Google	34
3.2 Pianificazione delle parole chiave	35
3.3 Keyword per annunci a pagamento	39
3.4 Alcuni strumenti per l'analisi delle parole chiave	45
4. Calcolo delle performance di ogni componente di tutte le parole chiave	51
4.1 Dataset utilizzato	51
4.2 Pulizia dei dati	55

4.3	Processo di attribuzione	56
4.3.1	Marketing di attribuzione	57
4.3.2	Attribuzione applicata al seguente progetto	58
4.4	Stemming	61
4.5	Lemmatizzazione	65
4.6	Stemming o Lemmatizzazione?	67
4.7	Calcolo dei risultati	68
5.	Il valore di Shapley	76
5.1	Cos'è il valore di Shapley	76
5.2	La formula	81
5.3	Le proprietà	84
5.4	Come viene utilizzato il valore di Shapley nel seguente progetto	84
5.5	Rivisitazione del calcolo di Shapley	88
5.6	Applicazione del calcolo	88
5.7	Funzione caratteristica	90
5.8	Risultati ottenuti	91
6.	Conclusione	94
7.	Tabelle complete dei risultati finali	96
8.	Bibliografia	119
9.	Sitografia	119
9.1	Articoli	121

1. Introduzione

L'obiettivo del seguente progetto di tesi è di offrire una tecnica differente per coloro che si occupano del search advertising, ambito dedicato alla pubblicità online attraverso i motori di ricerca, in particolare per la fase di pianificazione delle parole chiave di un annuncio pubblicitario promosso dall'azienda per cui lavorano. Il metodo proposto permette loro di comprendere quali dei termini utilizzabili per la generazione delle parole chiave, risultano più promettenti di altri poiché rispetto alle scelte prese in passato per la stessa azienda, hanno contribuito maggiormente all'ottenimento dei risultati finali.

Il progetto è stato sviluppato a partire da informazioni fornite dall'azienda in cui ho svolto il percorso di tirocinio curriculare: WebSoup SRL¹. Situata a Lucca, è un'agenzia SEO e Software House, la quale si occupa principalmente di realizzare siti web, la parte e-commerce e della loro promozione sul web. Pianificano campagne pubblicitarie online e strategie SEO per l'ottimizzazione dei siti web dei loro clienti nei motori di ricerca.

Intraprendendo l'esperienza di tirocinio presso di loro, si è deciso di sviluppare un progetto di tesi basato sui dati di uno dei loro clienti: nello specifico, sono stati presi in considerazione i dati relativi alle campagne pubblicitarie che WebSoup ha generato per un cliente proprietario di un e-commerce di prodotti tipici toscani da poco online, chiamato Quintovizio². WebSoup SRL, utilizza per la gestione delle campagne pubblicitarie dei loro clienti la piattaforma Google Ads, strumento che Google offre per l'inserimento e la gestione di spazi pubblicitari nei risultati di ricerca: l'utente digita una particolare richiesta ed il motore di ricerca mostra una serie di annunci pertinenti che sono stati creati tramite piattaforme come Google Ads. Per ogni cliente dell'agenzia corrisponde un account specifico sulla piattaforma, tramite il quale è possibile verificare gli andamenti delle campagne attive, i risultati di quelle passate e la creazione di nuove. Tra le informazioni reperibili attraverso l'account Google Ads di ogni cliente è possibile accedere ad una sezione interamente dedicata alle parole chiave utilizzate per gli annunci. Ogni campagna pubblicitaria, oltre ad avere un titolo ed una descrizione, è caratterizzata da una serie di parole chiave, scelte con molta cura,

¹ WebSoup SRL <https://www.websoup.it/>

² Quintovizio <https://quintovizio.com/>

che permettono a Google di comprendere rispetto a quali termini di ricerca pubblicare l'annuncio creato. L'utente che effettua la propria ricerca utilizza una serie di parole che riassumono il concetto basilare delle informazioni a cui è interessato e il motore di ricerca osserva le parole chiave che sono state abbinate ad annunci o ai contenuti delle pagine web, dai loro creatori.

Nello specifico ambito degli annunci a pagamento, i risultati che l'utente visualizza sono frutto dell'abbinamento tra le parole chiave che caratterizzano l'annuncio ed i termini di ricerca selezionati dall'utente, i quali hanno attivato la pubblicazione dello stesso. Tramite l'account Google Ads è possibile dunque verificare i risultati che le varie parole chiave di ogni campagna hanno ottenuto, permettendo così a coloro che si occupano dell'ambito pubblicitario di comprendere quali di queste hanno raggiunto i migliori risultati.

Concentrandosi su queste informazioni si è pensato di ideare un metodo che potesse permettere a coloro che si occupano di questo ambito, di poter valutare non solo i risultati ottenuti dalle parole chiave predefinite, ma poter effettuare uno studio più approfondito che vada ad osservare il comportamento e di conseguenza i risultati raggiunti dai singoli termini che compongono insieme ad altri una parola chiave. L'idea del progetto prevede quindi di offrire uno strumento che possa aiutare a comprendere quali abbinamenti tra termini che compongono una parola chiave risultano più vincenti e quindi osservare quali tra i vocaboli utilizzati per la costruzione di una o più parole chiave hanno portato al raggiungimento di risultati più elevati, per poter valutare le possibili scelte future.

La parola chiave nel mondo del web marketing costituisce un elemento fondamentale in quanto permette all'azienda che promuove un certo prodotto o servizio di rendersi visibile nell'intero mercato online. Coloro che lavorano nel mondo del search advertising devono capire quali parole chiave porteranno loro un maggiore numero di clic da parte degli utenti, attirando la loro attenzione più di quanto riescono a fare gli altri annunci. Il concetto che è stato preso in considerazione e adattato al tipo di studio svolto in questa prima parte, è quello alla base dell'attribuzione di marketing, processo che permette di valutare nelle comunicazioni multicanale quali sono i mezzi tra quelli utilizzati, che riescono ad attirare maggiormente l'attenzione degli utenti.

I passaggi che si sono svolti per il raggiungimento finale dell'intero lavoro sono diversi, sono stati inizialmente scaricati e puliti i dati di interesse ed una volta ottenuto il dataset su cui sviluppare l'intero lavoro, si procede con la scrittura di un programma

in linguaggio Python, attraverso il quale sono stati svolti tutti gli studi del caso in esame. Il primo algoritmo ad essere implementato consiste nell'applicazione ad ogni vocabolo che formava una parola chiave del dataset del concetto di stemming.

Lo stemming è un processo che riduce la forma flessa di una parola alla sua forma radice detta tema, il quale non sempre corrisponde alla radice del termine, ma può anche essere costituito dalla radice accompagnata da altre lettere che precedono la desinenza. La scelta di procedere in questa direzione è stata dettata dal voler evitare di considerare tutte le possibili varianti di uno stesso argomento che potrebbe essere espresso tramite l'utilizzo di differenti termini che variano per genere (maschile e femminile) e/o per numero (singolare e plurale), ma raggruppare i risultati ottenuti da ciascuno senza distinguere tra le possibili flessioni dello stesso vocabolo, considerando quindi solo il significato del termine.

L'esecuzione del processo di stemming prevede l'utilizzo di un algoritmo specifico scritto in Python e reperibile sul web: negli anni ne sono stati implementati diversi, quello scelto nel progetto è l'algoritmo Snowball. A seguito di tale applicazione, si è potuto ottenere una prima analisi sulle parole esaminate visualizzando i risultati che i singoli stem di ciascun termine, equivalente inglese del concetto di tema di una parola, le coppie e le triple che formano hanno ottenuto nel corso delle varie campagne studiate.

Nella seconda parte del lavoro presentato, il focus si è rivolto verso un'ulteriore analisi da effettuare sulle parole chiave del cliente, ottenuti dunque i valori delle performance di ogni singolo componente delle varie parole chiave si è pensato di applicare il principio alla base del calcolo del valore di Shapley.

Il valore di Shapley, pensato e ideato dal matematico ed economista statunitense Lloyd Stowell Shapley dal quale riprende il nome, consiste in un calcolo sfruttato nell'ambito della teoria dei giochi cooperativi che permette di calcolare a quanto corrisponde il contributo di ciascun giocatore in una collaborazione tra più membri. La teoria dei giochi è una disciplina che studia modelli matematici che descrivono come interagiscono più agenti ed aiuta a definire e prevedere l'andamento di un possibile scenario reale. Viene applicata in molti dei campi in cui è necessario uno studio sulle previsioni di certe decisioni da prendere e sulle situazioni che possono verificarsi di conseguenza. Il valore di Shapley nello specifico considera più elementi che partecipano ad uno stesso gioco e valuta, a partire dai risultati che i giocatori in squadra hanno ottenuto, in che modo ognuno di questi ha contribuito al

raggiungimento del risultato finale. Prende in considerazione tutte le combinazioni tra giocatori, anche quelle probabili, per poter fornire ad ognuno un valore a partire da tutte le possibili combinazioni che possono essere generate tramite i partecipanti del gioco in esame.

Il significato basilare di tale calcolo è stato applicato alle parole chiave utilizzate per le pubblicità online del cliente, con lo scopo di offrire ai gestori di tale ambito un metodo per poter valutare quali tra i vari termini utilizzati hanno contribuito maggiormente al raggiungimento dei risultati finali e dunque su quali puntare per la generazione di eventuali future parole chiave. La formula su cui si fonda tale valore è stata però rivista per poterla meglio applicare al tipo di giocatori su cui si sta lavorando. L'idea è stata quella di considerare i termini che compongono le varie keywords come i giocatori per cui si è interessati a valutare il rispettivo valore di Shapley, mentre per quanto riguarda le possibili combinazioni, si è proceduto considerando solo quelle realmente avvenute in quanto nel caso esaminato risulta essenziale tenere presente anche del significato dei vari giocatori poiché non tutti hanno modo di giocare con ogni altro per mancanza di coerenza tra i concetti che esprimono, per questo non sono effettivamente possibili tutte le combinazioni. Inoltre, sempre per lo specifico ambito studiato, ovvero le parole chiave utilizzate per gli annunci pubblicitari, risulta difficile a livello computazionale applicare l'algoritmo che permette il calcolo del valore di Shapley, poiché i giocatori in partita sono molti e le combinazioni generabili crescono in modo esponenziale.

La formula è stata rivisitata, si è evitato di considerare ogni possibile combinazione tra giocatori, ma solo quelle realmente avvenute ed in base a queste è stato calcolato il valore di Shapley di ogni vocabolo che ha giocato la sua partita insieme ad altri. Il tipo di cliente scelto non ha importanza ai fini di tale lavoro in quanto il seguente progetto di tesi si propone di offrire un metodo che possa affiancare i molti strumenti già esistenti, per poter analizzare le campagne pubblicitarie online di qualsiasi tipologia di prodotto o servizio offerti da un'azienda. Gli strumenti già utilizzati permettono tipicamente di studiare come nel web vengono ricercate le informazioni, quali parole chiave sono più efficienti e quali vengono principalmente utilizzate, mentre l'idea di questo lavoro è di proporre una strategia applicabile direttamente ai soli risultati del cliente, senza dover verificare l'andamento di una certa keyword nell'intero spazio internet, ma potendola esaminare solo per quello che

riguarda il cliente studiato e poter comprendere meglio come abbinare i vari termini tra loro per garantire delle parole chiave vincenti.

2. SEO e SEM: tecniche per aumentare la visibilità di un sito web

Le persone possono reperire le informazioni che cercano sul web tramite diversi canali: digitando l'indirizzo web, se conosciuto, tramite i motori di ricerca, oppure tramite i Social Network. Trovare le informazioni in modo rapido ed efficace è un requisito fondamentale e sempre più richiesto dai visitatori del web.

Oggi internet ha rivoluzionato i comportamenti d'acquisto dei vari clienti, prima di comprare un qualsiasi prodotto o servizio gli utenti si informano sul web, cercano opinioni, consigli, confrontano prezzi e funzionalità. La fiducia dei consumatori rappresenta un elemento decisivo e per le imprese diventa fondamentale oltre alla visibilità sul web, il saper gestire con efficacia la propria reputazione online. Per coloro che possiedono un sito web, è dunque importante riuscire a conquistare le prime posizioni della classifica di risposte che offre un motore di ricerca, essere quindi molto abili nel raggiungere i loro potenziali clienti, capendo come farsi notare tra i molti altri presenti in rete e come differenziarsi rendendo i loro contenuti più interessanti.

Il termine SEO racchiude tutte quelle attività che una volta messe in atto, permettono a coloro che sono visibili online di ottenere posizioni migliori nei risultati di ricerca, di comparire dunque tra i primi risultati che il motore di ricerca offre all'utente ed avere una maggiore probabilità di essere raggiunti. Il termine SEM descrive invece un processo che contiene anche le tecniche di tipo SEO, in quanto consiste in una serie di attività in ambito marketing con l'obiettivo di ottenere maggiore visibilità nel web, tramite tecniche sia gratuite che a pagamento. Le informazioni che circolano sulla rete si espandono notevolmente e assumono forme e contenuti diversi (documenti, video, link, streaming, flussi audio), per questo motivo i motori di ricerca rivestono un ruolo cruciale.

Nel seguente capitolo si riportano le principali tecniche per ottenere posizioni migliori tra i risultati di ricerca dei vari motori utilizzati.

2.1 La visibilità nei risultati dei motori di ricerca

I dati presenti su Internet sono in costante crescita, siti, blog e pagine di Social Network sono in aumento continuo e questo implica che per aziende, società e professionisti sia fondamentale ma allo stesso tempo sempre più complicato rendersi distinguibili nel

web: la cosa importante non è più essere presenti nello spazio Internet, ma rendersi visibili al suo interno.

Il posizionamento delle varie pagine web nei risultati dei differenti motori di ricerca si ottiene tramite un insieme di tecniche utilizzate con lo scopo di migliorare la visibilità della pagina stessa e di conquistare il meccanismo alla base dei vari algoritmi che determinano il loro funzionamento. Un motore di ricerca è un sistema automatizzato che riesce ad analizzare un'enorme quantità di dati, restituendoli come un elenco ordinato di contenuti e queste informazioni vengono classificate in base a complessi algoritmi che ne indicano l'importanza per una determinata combinazione di parole chiave di ricerca. È un sistema, dunque, che in maniera del tutto automatica reperisce le informazioni presenti nel web e le organizza attraverso opportuni indici. I meccanismi che regolano il funzionamento di un motore di ricerca permettono agli utenti di ricevere in base alle loro ricerche, risultati utili e coerenti, l'utente struttura una query tramite un elenco di parole chiave da ricercare, successivamente un software denominato *crawler o spider*, esegue la scansione del maggior numero possibile di pagine presenti nel web, le quali vengono poi archiviate e catalogate nel database del motore di ricerca interrogato. Opportuni algoritmi recuperano le più pertinenti ed alla fine forniscono all'utente un elenco di pagine che soddisfano la sua richiesta. L'insieme di tutte le pagine restituite come risultato di una determinata ricerca si identifica con il nome di *SERP (Search Engine Results Page)*, il successo di un motore di ricerca dipende principalmente dalla rilevanza dei risultati forniti, l'utente che lo interrogherà fornendogli un input in ingresso riceverà come output l'elenco delle informazioni ritenute più rilevanti dal motore di ricerca. Un motore di ricerca funziona dunque da "scanner" delle pagine web presenti nella rete: in modo automatico le categorizza e le restituisce all'utente come risultato di una determinata richiesta.

Esistono vari motori di ricerca, per citarne alcuni *Bing, Yahoo!, Baidu, YANDEX, DuckDuckGo*, ma quello che si identifica come il più utilizzato al mondo resta attualmente *Google*³. Caratteristica di Google che gli ha sicuramente permesso di innalzarsi in pochi anni a leader mondiale, consiste nel modo differente rispetto a quanto fanno gli altri motori di ricerca di ottenere i risultati della SERP. L'idea alla base è stata quella di inserire negli algoritmi di ricerca i link come variabili di calcolo per la rilevanza, considerando i collegamenti tra siti come elementi matematici. Nasce

³ Fonte: StatCounter Global Stats, aggiornato a dicembre 2021 <https://gs.statcounter.com/search-engine-market-share> (visitato il 28 dicembre 2021)

quindi l'algoritmo Pagerank, adoperato per catalogare ed indicizzare i contenuti presenti nel web: le query ricercate dai vari utenti ottengono come risultato un sottoinsieme dell'intero insieme di pagine, ordinato in base al valore del PR. Molteplici sono gli altri aspetti che hanno determinato il successo di Google come ad esempio una progettazione user-friendly, molto semplice per qualsiasi tipo di utente ed un continuo aggiornamento per ciò che riguarda l'algoritmo di ricerca e le parti che lo compongono.

2.2 Cos'è la SEO e quali sono i passaggi che la costituiscono

SEO è l'acronimo di *Search Engine Optimization* (in italiano Ottimizzazione per i motori di ricerca) e definisce l'attività di ottimizzazione delle pagine web con lo scopo di renderle maggiormente *search engine friendly*, ovvero mettere in atto tutte le tecniche tendenti a migliorare la posizione del sito stesso nei risultati di ricerca e rafforzarne la presenza nel web. Chi si occupa di SEO sono dunque professionisti che studiano a fondo l'ottimizzazione, per riuscire a comprendere quali siano i nuovi parametri che i motori di ricerca tengono in considerazione per valutare i link ai siti da mostrare. Il posizionamento di un sito web tra le prime risposte dei motori di ricerca è comunque un processo guidato da tanti fattori, limitarsi ad ottimizzare un sito non risulta per questo sufficiente a renderlo visibile poiché gli elementi che insieme collaborano per l'ottenimento di tale risultato sono differenti.

Il primo importante passo da affrontare per ottenere ottimi risultati consiste nel pianificare la propria strategia per poter capire su quali aspetti puntare e soprattutto individuare nel tempo eventuali modifiche da apportare. I motori di ricerca evolvono, i piani prestabiliti e le scelte d'azione necessitano quindi di essere pensate in modo flessibile e adattabile ad eventuali cambiamenti. I parametri presi in considerazione dai motori di ricerca sono molti e le principali strategie e pratiche di ottimizzazione SEO si possono suddividere in due macroaree, definibili come:

- *On-Page SEO*;
- *Off-Page SEO*.

Con il termine On-Page SEO, si indicano tutte quelle pratiche che hanno lo scopo di modificare gli elementi che vanno a costituire la struttura globale di un sito web o di una pagina, in modo tale da renderla più appetibile nei confronti del motore di

ricerca. La Off-Page SEO, consiste invece in tutte quelle strategie SEO che si applicano al di fuori del sito, ma che risultano comunque molto efficaci per il suo posizionamento.

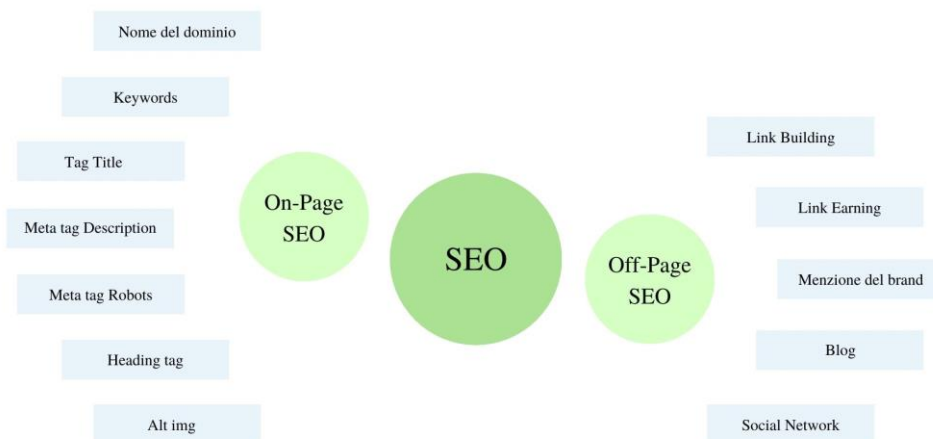


Figura 1. Elenco di alcuni dei principali fattori della On-Page SEO e Off-Page SEO⁴.

2.3 On-Page SEO

Gli elementi che compongono la On-Page SEO, sono definiti *On-Page factor*, ovvero tutti quei fattori interni alla pagina che possono influenzare il posizionamento di un sito web e per questo necessitano di essere curati ed affrontati nel migliore dei modi. Di seguito si riportano le spiegazioni di alcuni tra i principali fattori On-Page, elencati in Figura 1.

2.3.1 Scelta del nome del dominio

La scelta del nome del dominio è uno tra i primi passaggi da affrontare nell'ambito della SEO e i fattori di cui tener conto affinché la scelta presa risulti vincente, sono diversi. I principali si possono identificare in tre caratteristiche che il nome deve possedere: corto, non confondibile e facile da pronunciare.

La scelta di un nome troppo lungo può risultare limitante per vari motivi, può essere difficoltoso l'inserimento in cartelloni pubblicitari o in annunci a pagamento, inoltre per il cliente può essere più complesso da ricordare. Per quest'ultima ragione

⁴ Fonte: Giorgio Taverniti. *SEO POWER, Strategie e strumenti per essere visibili sui motori di ricerca*. Milano, Ulrico Hoepli Editore, 2013

non deve nemmeno essere facilmente confondibile con altri domini, con siti di altre aziende o in particolar modo con eventuali competitor e soprattutto semplice da pronunciare e da ricercare, è preferibile che contenga una keyword e soprattutto che rifletta la realtà aziendale.

I motori di ricerca tendono a dare peso alle parole chiave presenti all'interno del nome di dominio, per questo è consigliabile avere una lista di pochi vocaboli tra cui scegliere che descrivano al meglio il dominio, per far sì che per gli utenti che ricercano sul web risulti facilmente rintracciabile. Oltre a questo, con il nome scelto si deve enfatizzare il tipo di servizio o prodotto proposti alle persone per rendere la loro ricerca più semplice ed immediata. In rete sono presenti molti tool che aiutano nel processo di scelta, alcuni dei quali permettono anche di verificare in tempo reale la disponibilità del dominio selezionato.

2.3.2 Scelta delle keyword

Archiviato il primo step, ovvero la scelta del nome del dominio, si procede con le altre strategie SEO per rendere la pagina più visibile di altre.

Individuare le keyword migliori è un punto cruciale, deve essere dedicato a questa fase il giusto tempo in quanto le parole chiave selezionate si utilizzeranno poi all'interno del testo, nei meta tag opportuni e nelle campagne di promozione. Molte analisi svolte e dati raccolti dimostrano che gli utenti che interrogano i motori di ricerca pongono maggiore attenzione sui primi risultati, difficilmente procedono oltre la seconda o terza pagina di ricerca e tendono a creare frasi talvolta anche abbastanza lunghe nella costruzione della loro query.

In questa fase in cui l'obiettivo è quello di scegliere le keyword migliori è necessaria una pianificazione adeguata, è importante comprendere i bisogni dei nostri utenti, le diverse tipologie di parole chiave che abbiamo a disposizione e pensare a possibili parole o combinazioni di esse con cui un utente potrebbe cercare la nostra pagina. Creare un elenco di sinonimi associati al nostro brand prendendo in considerazione anche aziende leader in determinati settori, per prendere spunto da loro oppure avvalersi di strumenti di supporto per la ricerca delle keyword, tenendo sempre ben presente che l'utente può rintracciare la pagina sul web tramite la combinazione di parole differenti. Le keyword scelte avranno poi molteplici funzioni: serviranno per

creare contenuti ad hoc, strutturare URL, creare titoli accattivanti nelle pagine e molto altro.

2.3.3 Tag HTML

In fase di costruzione del sito, nel codice HTML, ovvero nella parte che permette di definire la struttura delle varie pagine ed i contenuti finali che l'utente andrà a visualizzare, vengono inseriti dei tag e meta tag, porzioni di codice invisibili all'utente ma non al crawler che definiscono il layout della pagina e di maggiore importanza per l'ambito SEO, permettono l'accesso dei motori di ricerca ai contenuti stessi del sito.

I meta tag sono elementi nascosti all'interno del codice HTML studiati appositamente per fornire informazioni aggiuntive riguardo a un documento ipertestuale, documento formato da testo e da codice. Esistono vari tipi di meta tag, alcuni nel corso degli anni hanno perso importanza per quanto riguarda il loro peso ai fini di posizionamento, per altri invece il loro contributo resta ancora significativo. Va dedicata loro, per questo motivo, la giusta attenzione affinché la loro funzionalità sia sfruttata al meglio: in rete sono presenti molti tool che guidano i principianti o anche gli esperti di SEO nella scelta dei tag migliori per rendere la loro pagina più cliccabile delle altre.

Il tag Title è di fondamentale importanza in quanto consiste nel titolo che i motori di ricerca mostrano nell'elenco dei risultati. Come suggerisce Google, il tag Title indica a utenti e motori di ricerca l'argomento di una determinata pagina, va posizionato all'interno dell'elemento <head> del documento HTML e creato per ogni pagina del sito⁵. Il titolo verrà utilizzato dai motori di ricerca all'interno della SERP, per questo motivo è importante che il contenuto della pagina web sia coerente con il titolo scelto, che il titolo sia unico per ogni pagina e che esso risulti descrittivo utilizzando le parole chiave più significative per il sito, ma allo stesso tempo breve, nelle raccomandazioni di W3Schools viene specificato che non dovrebbe superare i 50-60 caratteri⁶. Sempre Google spiega in modo chiaro il motivo per cui a questo tag vada fornita la giusta attenzione, specificando che i titoli sono estremamente

⁵ Google Search Central, voce *Guida introduttiva all'ottimizzazione per i motori di ricerca (SEO)*
<https://developers.google.com/search/docs/beginner/seo-starter-guide?hl=it> (visitato il 20 settembre 2021)

⁶ W3Schools, voce *HTML <title> tag*
[https://www.w3schools.com/tags/tag_title.asp#:~:text=The%20tag%20defines%20the,search%20engine%20optimization%20\(SEO\)](https://www.w3schools.com/tags/tag_title.asp#:~:text=The%20tag%20defines%20the,search%20engine%20optimization%20(SEO)) (visitato il 20 settembre 2021)

importanti poiché permettono agli utenti di avere informazioni immediate riguardo il contenuto di un risultato e sul perché questo sia pertinente alla loro richiesta, inoltre sono spesso l'informazione principale utilizzata dagli utenti stessi per decidere su quali risultati fare clic⁷. Un meta tag Title ben scritto migliora quindi sia il posizionamento della pagina, che il numero di clic e visualizzazioni.

Il meta tag description è un altro elemento utile in quanto può comparire tra i risultati prodotti dai motori di ricerca per una specifica keyword. I motori di ricerca, per visualizzare i risultati nella SERP, utilizzano porzioni di contenuti prelevati dalle pagine chiamati *snippet*, la generazione di tale parte avviene in modo automatico e il tag description risulta utile per suggerire al motore di ricerca quali informazioni inserire. La sua funzione principale consiste nel fornire una descrizione della pagina, per questo motivo è necessario che rispetti alcune caratteristiche affinché il suo utilizzo risulti più utile ai fini del posizionamento della pagina stessa. Deve dunque essere unico per ogni pagina, coerente con il contenuto che l'utente si aspetta di trovare nel sito e non troppo generico, nella sua brevità deve colpire l'attenzione dell'utente. Non sempre, comunque, i motori di ricerca mostrano come snippet quello che è stato specificato nel tag description, ad esempio Google è uno di questi, in quanto talvolta sceglie di mostrare dei contenuti interni alla pagina piuttosto che la descrizione presente nel tag. In ogni modo, utilizzare opportunamente questo preciso elemento del codice HTML, permette di avere la possibilità di influenzare in modo positivo i motori di ricerca e allo stesso tempo migliorare il ranking della pagina. Questo tipo di tag, se scritto nel modo giusto fornendo una spiegazione chiara e convincente di ciò che gli utenti troveranno nella pagina, non permette di migliorare il posizionamento, ma la sua presenza risulta comunque significativa poiché può migliorare il numero finale di click ottenuti.

Il meta tag robots, ha il compito di comunicare con i programmi che si occupano della scansione del sito web, permettendo infatti all'esperto di SEO di specificare come comportarsi con certe risorse presenti al suo interno. Ha il compito di indicare al crawler quali delle sezioni della pagina web che sta scandendo, siano accessibili e quali non indicizzabili, per poter dunque specificare al motore di ricerca quali parti non prendere in considerazione. Può risultare utile, ad esempio, per tutte quelle

⁷ Google Search Central, voce *Creare titoli e snippet utili per i risultati della Ricerca*
<https://developers.google.com/search/docs/advanced/appearance/good-titles-snippets?hl=it> (visitato il 20 settembre 2021)

informazioni che devono necessariamente essere inserite per una questione di servizio all'interno del sito, ma che poi per gli utenti non costituiscono un valore reale, come le pagine dedicate alla privacy o tutte quelle informazioni per cui è importante che il motore di ricerca sappia della loro esistenza, ma che non hanno un vero e proprio valore in termini di SEO. Usare il meta tag robots permette quindi di dare specifiche indicazioni ed evitare possibili penalizzazioni. Tale meta tag si inserisce nella sezione header del documento HTML ed è possibile al suo interno definire il tipo di comando⁸. I predefiniti sono *follow* e *index*, i quali non devono essere specificati, mentre in caso sia necessario l'uso di altri comandi disponibili, devono essere indicati. Index, indica che i contenuti della pagina devono essere inclusi nell'archivio del motore di ricerca, mentre follow richiede di seguire tutti i link presenti nella pagina al fine di individuare altre pagine del sito web in modo automatico. Se nessun meta tag robots viene inserito, i crawler indicizzeranno automaticamente tutte le pagine e seguiranno tutti i link presenti all'interno del sito.

2.3.4 Heading tag

Da non confondere con i meta tag o i tag Title, gli heading tag o tag di intestazione, vengono utilizzati per organizzare i testi di una pagina, permettono di specificare la gerarchia di titoli e sottotitoli presenti e fornire maggiore rilevanza a parole o frasi racchiuse all'interno di tali tag.

Le tipologie di tag di intestazione utilizzabili sono 6, vanno dalla <H1> alla <H6>, l'ordine di scansione da parte dei principali crawler dei motori di ricerca è da sinistra a destra, partendo dall'alto verso il basso. Quindi è buona norma utilizzare il tag <H1> nella posizione più alta della pagina in alto a destra, in quanto con questo tag viene tipicamente indicato l'argomento principale della pagina, sarà dunque quello a cui verrà assegnata maggiore rilevanza dai motori di ricerca. Utilizzare gli heading tag non migliora il posizionamento della pagina, ma se sfruttati nel modo giusto, risultano importanti ai fini di indicizzazione, poiché è possibile indicare ai crawler quali siano le frasi più importanti all'interno del testo ed agli utenti durante la navigazione del sito.

⁸ Per visualizzare altri comandi disponibili per il Meta tag Robots di Google, consultare il seguente link https://developers.google.com/search/docs/advanced/robots/robots_meta_tag?hl=it#directives

2.3.5 Attributo ALT nelle immagini

Le immagini sono parte dell'estetica di un sito web, ma oltre ad abbellire l'aspetto visivo devono essere anch'esse considerate per ottimizzare il sito in un'ottica SEO. L'attributo alt è un elemento interno al tag HTML ed ha lo scopo di fornire all'utente un testo alternativo nel caso in cui stia navigando sul sito tramite un browser che non supporta le immagini. Tale attributo fornirà le informazioni relative all'immagine cui corrisponde e rispetto all'ambito SEO potrebbe risultare una buona scelta quella di utilizzare nell'alt una delle keyword basilari per i contenuti del sito.

2.4 Off-Page SEO

Il corretto posizionamento e la corretta indicizzazione di un sito web dipendono da svariati fattori in quanto i motori di ricerca devono dare risultati sempre più appropriati e attinenti alle query richieste dagli utenti. I fattori esterni che garantiscono una maggiore popolarità nel web consistono nella gestione dei link che puntano al sito, ad esempio sui social media o nei podcast sono le recensioni, i commenti, tutto ciò che rende la reputazione della pagina migliore di quella di altre e che permette ai motori di ricerca di capire come una certa pagina è considerata dagli utenti del web e di conseguenza anche il brand o l'azienda che vi sono dietro. Popolarità, reputazione e autorevolezza, sono i tre fattori cruciali della Off-Page SEO.

Di seguito si riportano le spiegazioni di alcuni tra i principali fattori Off-Page, elencati in Figura 1.

2.4.1 Link Building e Link Earning

Il processo di Link Building è uno dei pilastri su cui si fonda la SEO e consiste nel far pubblicare dei *backlink* verso il sito per il quale si vuole ottenere una maggiore visibilità, in maniera gratuita o a pagamento, su pagine di altri siti che trattano di argomenti simili e che godono di un'ottima reputazione nel settore cui fanno riferimento. I backlink possono essere artificiali o naturali: i primi, si ottengono nel momento in cui i backlink verso il sito di interesse vengono acquistati, scambiati o richiesti gratuitamente con lo scopo di migliorare i posizionamenti. I naturali invece, sono quelli che si ottengono quando un webmaster inserisce spontaneamente un backlink verso il sito o la pagina poiché ritenuti utili per i suoi utenti.

La Link Building si basa su backlink artificiali, per questo motivo è molto importante verificare la qualità dei siti sui cui si pubblica, scegliere di pubblicare backlink su siti di scarsa qualità penalizza lo scopo di tale strategia nei confronti del sito a cui si è interessati.

La Link Earning consiste nel produrre e promuovere contenuti strategici capaci di stimolare l'acquisizione di backlink in maniera spontanea, senza dunque che l'inserimento sia forzato e senza investire per l'acquisizione diretta. Tale tecnica può produrre ottimi risultati, purtroppo però la sua efficacia non sempre è garantita in quanto trattandosi di backlink naturali, la pubblicazione è spontanea e quindi non si ha nessun controllo sui siti dai quali si ricevono tali backlink e sul numero di quelli ottenuti. Buona pratica risulta dunque la scelta di una modalità mista, ovvero fare link building, ma allo stesso tempo produrre anche buoni contenuti che garantiscano la possibilità di ottenere collegamenti in modo spontaneo. L'obiettivo della Link Building o della Link Earning è quello di ottenere molti backlink di qualità, ovvero da domini autorevoli verso il proprio sito: è più importante capire da dove provengono i link, piuttosto che sapere il numero di quanti puntano al sito, questo perché un link da una pagina più autorevole con traffico attivo ed in salita, eleva notevolmente la reputazione del sito a differenza di quanto possano fare tanti link di pagine di fama minore. Esistono su internet tool specifici che aiutano in questa fase del lavoro, ad esempio alcuni permettono di verificare il numero di link acquisiti, altri invece permettono di verificare l'autorità di un dominio, ma è comunque buona norma valutare anche personalmente l'autorevolezza di un dominio, analizzandolo manualmente.

2.4.2 Menzione del Brand

Una menzione si verifica quando un sito nomina un brand, ma non inserisce un link verso la pagina web cui fa riferimento. A differenza delle citazioni che risultano cliccabili per l'utente, la menzione, non mostrando nessun link da poter visitare, non permette di generare traffico referral, ovvero tutto il traffico verso un sito proveniente da fonti esterne ad un motore di ricerca poiché ottenuto tramite altri siti web. Le menzioni del brand caratterizzano comunque la popolarità del sito, motori di ricerca come Google tengono conto anche di queste informazioni per valutare l'autorevolezza del brand poiché considera che spesso le persone menzionano siti o marchi senza però

riportare il link diretto al sito, come succede invece nel caso dei backlink che consistono in vere e proprie citazioni.

Nonostante questo, le citazioni generano traffico referral cosa che con le menzioni non si può ottenere, dunque può risultare una buona pratica giunti a questa fase di lavoro contattare coloro che menzionano il sito ed invitare ad inserire il link diretto, trasformando la menzione in citazione. Anche in questo caso esistono modi per vedere quante volte il brand sia stato menzionato, utilizzando particolari operatori di ricerca di Google.

2.4.3 Blog e Social Network

I blog sono siti web in cui è possibile pubblicare contenuti di vario genere come immagini, video, opinioni sugli argomenti trattati. Promuovere il sito su un blog o crearne direttamente uno al suo interno, può consentire all'azienda di farsi notare nel mondo di internet.

L'azione di promozione tramite altri blog non deve risultare eccessiva, il modo migliore è commentare in maniera opportuna contenuti di interesse personale o che siano attinenti a ciò che si offre al cliente. Un'ottima strategia è anche quella di scegliere di creare un blog all'interno del sito di un'azienda, offrendo periodicamente articoli tramite i quali i clienti possano conoscere la storia di un prodotto, i servizi offerti e curiosità legate a tutto ciò che possono rintracciare navigando il sito. Produrre contenuti di qualità, garantisce una loro facile indicizzazione da parte dei motori di ricerca, favorendo di conseguenza la possibilità di nuove visite provenienti dal web e quindi una sempre maggiore visibilità e popolarità.

I social network, siti che forniscono agli utenti della rete un punto d'incontro virtuale per scambiarsi messaggi, condividere foto e video, hanno ottenuto uno spazio sensibile all'interno del web, per tale motivo per promuovere attività o servizi non si possono non prendere in considerazione. L'enorme diffusione di questi nuovi canali di comunicazione ha rivoluzionato il modo in cui gli utenti si avvicinano al mondo del web, hanno notevolmente ampliato il bacino di persone che oggi navigano in internet, offrendo nuovi possibili visitatori per le pagine presenti in rete.

Il mondo del web marketing e nel dettaglio dunque anche l'ambito SEO, hanno dovuto necessariamente adattarsi a questa novità, tanto che è nata una nuova disciplina la Social Media Optimization, definita tipicamente tramite l'acronimo SMO, che si

occupa nello specifico di promuovere un marchio, un'azienda o dei servizi all'interno dei principali canali di social network.

2.4.5 Altre strategie per aumentare la visibilità online

La visibilità in rete si ottiene a seguito di un'accurata catena di molteplici fattori che necessitano di essere studiati e trattati in maniera attenta ed efficace per il raggiungimento dello scopo finale. Di seguito si illustrano alcuni degli altri elementi che contribuiscono alla realizzazione di tale processo.

La gestione della landing page, che nel campo del web marketing corrisponde alla specifica pagina che il visitatore raggiunge dopo aver cliccato un link apposito o una pubblicità. Tale pagina è solitamente sviluppata per trattare argomenti specifici e mostra contenuti che corrispondono ad un'estensione dei link o della pubblicità, ed è ottimizzata per una specifica parola chiave o frase per attirare l'attenzione dei motori di ricerca⁹. Le landing page sono importanti in quanto servono a catturare l'attenzione del visitatore ed invitarlo a compiere una specifica azione all'interno del sito, è la prima pagina che l'utente si trova davanti dopo aver cliccato su un banner pubblicitario o su un link, il suo ruolo è di fondamentale importanza, in modo particolare per quanto riguarda l'obiettivo di ottenere la conversione da semplice visitatore a cliente.

Le e-mail marketing sono un altro ottimo strumento per aumentare di popolarità in rete ed ottenere clienti. Consistono in uno dei veicoli principali attraverso il quale si possa raggiungere l'utente, i loro punti di forza sono un maggior rispetto per i clienti e la loro privacy, informazioni e pubblicità sempre più mirate e una comunicazione sempre più personalizzata. Questo canale, così come gli altri, per risultare efficace deve essere ben organizzato, è importante dunque comprendere gli utenti a cui siamo interessati, in base alle fasce d'età, argomenti che maggiormente ricercano, ed inviare di conseguenza e-mail adeguate e pertinenti. Messaggi brevi, coinvolgenti e semplici, aggiungere qualche immagine e rispettare le caratteristiche che distinguono il sito web a cui la mail fa riferimento.

⁹ Wikipedia, voce *Landing Page* https://it.wikipedia.org/wiki/Landing_page (visitato il 20 settembre 2021)

2.4.6 Google My Business

Google tra i canali di ambito Off-Page SEO, offre Google My Business¹⁰, uno strumento gratuito che permette a chi possiede un'attività, di gestirne la presenza online, ovvero avere la possibilità di mostrarsi in rete ed attirare clienti della zona. Infatti, se viene specificata l'area coperta dal servizio offerto dall'attività in questione, con un account My Business si ha la possibilità di raggiungere con più facilità gli utenti che ricercano informazioni in una specifica area geografica. Alla persona che in una determinata località ricerca un particolare servizio su internet, verranno mostrate tutte le attività vicine alla zona e attinenti al tipo di ricerca effettuata.

I profili di Google My Business, sono quelli che si posizionano nella parte superiore dei risultati di ricerca per le query con intento di ricerca locale, risultano una buona soluzione per coloro che possiedono un negozio piccolo e che puntano principalmente ad una vendita locale. Oltre alle informazioni sull'attività viene mostrato il collegamento diretto a Maps per il raggiungimento immediato ed anche le varie recensioni ottenute. Quest'ultime consistono in un altro elemento molto efficace, migliori sono quelle ottenute nella scheda di Google My Business, più probabilità vi è di venire posizionati nello "snack pack", definito anche come "Local 3 Pack", che corrisponde ad uno spazio presente nella lista di risultati offerti da Google che mostra tra i primi risultati di ricerca tre profili di attività locali che possiedono un account My Business. L'ottenimento di tale posizione si raggiunge con una scheda ben strutturata, completa e attiva, rispondendo sempre alle recensioni e pubblicando spesso nuovi contenuti.

In generale un'ottimizzazione SEO di tipo Off-Page svolta in modo corretto genera molteplici vantaggi per la popolarità di un sito nel web, alcuni di questi sono l'aumento dei posizionamenti nelle SERP dei motori di ricerca, una maggiore autorevolezza ed una maggiore visibilità che portano di conseguenza all'ottenimento di più presenze nelle prime posizioni rispetto ad altri siti e quindi più backlink, più visite e più menzioni.

¹⁰ Guida di Google My Business <https://support.google.com/business/answer/6300665?hl=it> (visitato il 22 settembre 2021)

2.5 Cos'è la SEM e quali sono i passaggi che la costituiscono

L'acronimo SEM indica il *Search Engine Marketing*¹¹, ossia un ramo del web marketing che tramite l'uso dei motori di ricerca permette di aumentare la visibilità e rintracciabilità di un sito web. Così come il marketing tradizionale si incentra sull'attività di promozione di prodotti o servizi di un'azienda, nonché sullo sviluppo, monitoraggio e fidelizzazione della clientela, allo stesso modo il web marketing tramite il Search Engine Marketing, fornisce la possibilità al gestore di un sito web di generare un traffico di utenti interessati a scoprire quel determinato sito. Con la sigla SEM, si identifica un ambito più ampio della SEO, che include la SEO stessa e la SEA, acronimo di *Search Engine Advertising*, ovvero la disciplina che comprende le tecniche finalizzate al miglioramento del posizionamento di un annuncio tra i risultati a pagamento di un motore di ricerca. Attraverso le tecniche SEO l'obiettivo è quello di migliorare il posizionamento del sito web nelle SERP dei vari motori di ricerca, con le tecniche di SEM, lo scopo è differente in quanto mirano a generare una convergenza di utenti realmente interessati al sito su cui stanno navigando, cercando dunque di condurre gli individui realmente motivati verso un determinato sito. L'obiettivo principale non consiste solo nel condurre il sito alle prime posizioni, ma anche generare maggior traffico possibile, attirare un numero cospicuo di utenti che navigano sui motori di ricerca e che sono realmente interessati al prodotto o ai servizi offerti da uno specifico sito.

Una strategia SEM completa prevede dunque l'uso sia di pubblicità a pagamento che di tecniche di tipo SEO, garantendo per la pagina un ottimo posizionamento da una parte, ed un miglioramento significativo della visibilità nei motori di ricerca dall'altra. La collaborazione tra le due risulta vincente, in quanto una campagna advertising, se accompagnata da una buona pagina web in cui l'utente è reindirizzato, può portare all'ottenimento di buoni risultati. Negli anni l'acronimo SEM è stato sempre più utilizzato per riferirsi esclusivamente alle attività a pagamento (SEA), per questo motivo si proseguirà con tale termine anche nella seguente relazione.

¹¹ Wikipedia, voce *Search Engine Marketing* https://it.wikipedia.org/wiki/Search_engine_marketing (visitato il 22 settembre 2021)

2.6 Elementi costituenti della SEM

Diversi sono gli elementi di cui tener conto per ottenere una buona strategia Search Engine Marketing. Come già anticipato nel paragrafo precedente, sicuramente è importante che si ponga l'attenzione sia sulle strategie SEM che su quelle SEO, le quali se integrate possono condurre ad ottimi risultati.

Volendosi concentrare maggiormente su quelle che sono le caratteristiche fondamentali di una buona strategia SEM, uno degli aspetti significativi è la scelta del budget, ovvero del prezzo che si è disposti a pagare nel momento in cui un utente effettua un clic sulla pagina sponsorizzata. Il budget può essere modificato nel tempo in base ai risultati che sono stati ottenuti, l'attenzione si deve quindi focalizzare sul monitoraggio delle campagne per tenere traccia del loro andamento. Un altro elemento consiste nello stabilire un limite temporale, poiché questo permette di valutare in quale periodo sia utile alzare il budget e in quale invece diminuirlo, oltre a stabilire anche in quanto tempo vogliono essere raggiunti determinati risultati. Infine, si deve pensare alla creazione dell'annuncio vero e proprio, alle parole da scegliere per la descrizione di ciò che offre per poter attirare l'attenzione dei possibili clienti, ma anche dei motori di ricerca che devono decidere quali contenuti mostrare.

Per mettere in pratica queste tecniche esistono strumenti nati con tale scopo, Google offre ad esempio la piattaforma Google Ads, ma anche gli altri motori di ricerca offrono le proprie piattaforme di advertising.

2.7 Google Ads

Google Ads, tra le varie piattaforme è la più apprezzata dagli specialisti SEM, è uno strumento che Google offre per la pubblicità online per dare la possibilità di promuovere i propri servizi o prodotti su Ricerca Google, YouTube e altri siti presenti sul Web. Google Ads consente inoltre agli inserzionisti di scegliere obiettivi specifici per i propri annunci, i quali possono essere di vario genere, dall'ottenere l'acquisto da parte dei clienti di uno specifico prodotto, l'iscrizione alla newsletter, oppure generare un aumento di chiamate telefoniche verso la propria attività o visite al sito web, visualizzazione di una specifica promozione e molto altro.

Con un account Google Ads gli inserzionisti dopo aver scelto uno specifico obiettivo possono personalizzare il budget impostando un limite massimo mensile, il targeting ovvero un pubblico che rispetti certe caratteristiche, selezionare l'area

geografica in cui mostrare l'annuncio o avviare e interrompere la pubblicazione degli annunci in qualsiasi momento. La scelta dell'area in cui mostrare l'annuncio può essere a corto raggio, ovvero vicina all'attività oppure anche più ampia selezionando dunque città, regioni o interi paesi.

Nel momento in cui un utente cerca online un particolare prodotto o servizio, Google Ads mostra gli annunci promossi da coloro che possono soddisfare tali richieste utilizzando tecnologie intelligenti che aiutano a mostrare gli annunci ai potenziali clienti. Una volta creato l'annuncio e personalizzato secondo le proprie esigenze, si attende l'approvazione da parte di Google, ed ottenuta questa, l'annuncio verrà mostrato tutte le volte che un utente dell'area scelta come target ricerca un prodotto o un servizio affini a quelli proposti dall'annuncio stesso. L'inserzionista pagherà solo nel momento in cui l'utente andrà ad interagire con l'annuncio, solo quando farà quindi clic su di esso o svolgerà in generale una delle attività impostate come obiettivo.

Google Ads offre diverse tipologie di campagne, tra le più comuni:

- Campagne sulla rete di ricerca, gli annunci sono solitamente in formato di testo e possono essere pubblicati sulle pagine dei risultati di Ricerca Google quando un utente cerca un prodotto o servizio simile a quelli offerti. Sono campagne basate sulla ricerca di parole chiave da parte degli utenti ed è una tipologia di advertising che permette di comparire nelle posizioni più alte della SERP.
- Campagne sulla Rete Display, gli annunci sono in formato illustrato e vengono pubblicati su altri siti web o app visitate dal proprio target di clienti. L'annuncio comparirà sotto forma di banner o display su siti che trattano di servizi e prodotti simili a quelli offerti.
- Campagne video, ovvero annunci video della durata di pochi secondi che vengono mostrati immediatamente prima o durante la riproduzione di contenuti su YouTube.

Tra le più specifiche:

- Campagne Shopping, annunci in cui vengono mostrati direttamente i prodotti offerti ed indirizzano il cliente verso il sito web.
- Campagne per App, le quali consentono di promuovere applicazioni per iOS o Android su Ricerca Google, YouTube, Google Play e altro ancora.

La creazione dell'annuncio su Google Ads si compone di elementi diversi che verranno visualizzati dagli utenti nella SERP. Tra questi il primo su cui si pone l'attenzione è sicuramente il titolo, la parte più in vista dell'annuncio che rappresenta il primo modo di approcciarsi con i potenziali clienti e deve quindi indicare di cosa si occupa l'attività o quali prodotti e servizi offre. Immediatamente sotto al Titolo, si trova la descrizione, la quale dovrebbe fornire maggiori dettagli sui prodotti e servizi offerti o evidenziare quali sono i vantaggi per l'utente che visita il sito. Altri due elementi sono l'URL di visualizzazione e l'URL di destinazione. Il primo non corrisponde alla pagina effettiva sulla quale gli utenti saranno reindirizzati quando fanno clic sulla pubblicità, ma si riferisce generalmente all'home page del sito. L'URL di destinazione è invece la pagina esatta in cui chi fa clic sull'annuncio in questione, viene reindirizzato, è importante per questo motivo che sia il più pertinente possibile ai termini di ricerca ad essa associati. Per ognuno di questi elementi esiste un numero massimo di caratteri utilizzabili, nella guida di Google Ads su come creare un annuncio¹² è riportato che per il titolo è possibile utilizzare massimo 25 caratteri, mentre per gli altri elementi 35, quindi il messaggio è contenuto in poche righe, per questo è molto importante utilizzare le giuste parole per ottenere un buon punteggio da parte di Google ed attirare immediatamente l'attenzione degli utenti.

Attraverso la piattaforma di advertising Google Ads si possono creare gli annunci tramite campagne di tipo Pay per click che, come suggerisce il nome, consistono nel far pagare l'inserzionista nel momento in cui l'utente effettua un click sull'annuncio da lui pubblicato. Il procedimento prevede di definire le parole chiave sulle quali fare un'offerta, Google Ads mostra gli annunci da lui ritenuti idonei agli utenti che cercano tramite le parole chiave che sono state acquistate ed infine vengono portati in una landing page studiata per ottenere una conversione, ovvero trasformare l'utente in cliente.

Google Ads funziona come un'asta, quando un utente effettua una ricerca su Google parte un'asta automatica che in base a diversi fattori, stabilisce quali annunci pubblicare e in quale posizione. Gli inserzionisti decidono il loro budget massimo, il bid, ovvero quanto sono disposti a spendere per un click. Più il valore del bid è elevato, maggiore sarà la probabilità di essere visti, ma non sempre l'offerta più alta vince. Google, infatti, combina il fattore monetario con un fattore di qualità, non vuole

¹² Google Ads, voce *Come creare un annuncio* https://ads.google.com/intl/it_it/home/resources/come-creare-un-annuncio-online-con-google/ (visitato il 22 settembre 2021)

mostrare agli utenti solo gli annunci del miglior offerente, poiché potrebbero essere non in linea con la ricerca effettuata e quindi non avere un'adeguata percentuale di clic rispetto al numero di impressioni, rispetto cioè al numero di volte in cui quel particolare annuncio è stato visualizzato. Preferisce piuttosto mostrare annunci pertinenti e migliori, anche se con un costo per click massimo offerto inferiore, perché in questo modo gli utenti sono soddisfatti e spinti ad utilizzare Google anche per le ricerche future. Il costo effettivo per il click ricevuto potrebbe essere inferiore rispetto all'offerta massima, soprattutto se l'annuncio ottiene un punteggio di qualità alto. Il punteggio di qualità è una metrica che permette di valutare quanto risultano rilevanti i propri annunci Google Ads, basandosi sulla pertinenza delle parole chiave scelte, sulla pagina di destinazione per l'utente, la landing page, e sul valore di CTR previsto.

Nel momento in cui l'annuncio viene considerato idoneo per la pubblicazione, Google Ads ne calcolerà il ranking, ovvero il posizionamento nella SERP, in base all'offerta e alla qualità dell'annuncio al momento dell'asta. Google confronta il ranking dell'annuncio con quelli degli altri inserzionisti e determina in questo modo il costo effettivo ed il posizionamento.

Quando l'inserzionista vince l'asta, il suo annuncio viene visualizzato dall'utente, il quale può decidere di andarlo a visitare. Questo per l'inserzionista risulta essere già un buon risultato, se poi l'utente procede anche con l'effettuare una conversione, cioè completa una certa azione prestabilita, come l'acquisto di un prodotto o l'iscrizione alla newsletter, sicuramente il suo obiettivo iniziale viene completamente raggiunto. L'inserzionista quindi volendo attirare più traffico cerca di elaborare strategie SEM per creare annunci che possano ottenere nel complesso dei punteggi di qualità elevata, per questo deve cercare di capire quali parole o frasi attirano più utenti e comportino nello stesso tempo un miglior punteggio di qualità, rispetto alle metriche di cui tengono conto i motori di ricerca.

2.8 Oltre a Google Ads, altri tool per la pubblicità online

In questa relazione ci si è principalmente soffermati sullo studio di Google Ads in quanto consiste nella piattaforma utilizzata per lo svolgimento del progetto, ma i tool presenti per la gestione della pubblicità online sono diversi.

Il social network Facebook, è ad esempio un canale potentissimo per la pubblicità online, grazie ai dati che riesce a conservare come like, preferenze che le

persone esprimono, post che esse condividono, contenuti che commentano e molto altro. Fornisce per coloro che sono amministratori di una pagina lo strumento Facebook for Business¹³, tramite il quale si possono generare inserzioni pubblicitarie a pagamento, annunci personalizzati che vengono sponsorizzati tramite le proprie pagine Facebook. Viene creato l'annuncio con immagine, titolo e descrizione, definito il target di pubblico a cui farlo visualizzare ed impostato un budget giornaliero di spesa.

Microsoft Advertising¹⁴ è una piattaforma pubblicitaria che consente di pubblicare annunci sponsorizzati nelle pagine dei risultati del motore di ricerca di Microsoft, nello stesso modo in cui avviene con gli altri strumenti di advertising, l'inserzionista quindi offre un prezzo per il posizionamento dell'annuncio il quale viene visualizzato accanto ai risultati organici quando gli utenti effettuano una ricerca su un argomento correlato. Il pagamento avviene nel momento in cui l'utente effettua il clic sull'annuncio.

Quora di base consiste in un sito web in cui gli utenti possono pubblicare domande e risposte su qualsiasi argomento, oltre che votare e aggiungere dei commenti andando a generare la nascita di una vera e propria community di persone che condividono la loro conoscenza. La community di Quora è composta da persone che tramite e-mail e password possono iscriversi gratuitamente alla piattaforma ed intervenire riguardo ad argomenti per cui altri utenti chiedono informazioni o chiedere direttamente delucidazioni sui loro dubbi. Grazie alla sua struttura, la piattaforma Quora può fornire anche la possibilità di far aumentare la brand awareness, grado di conoscenza da parte del pubblico di riferimento del proprio marchio, e generare contenuti di valore al fine di attirare più persone sul proprio sito. Con Quora for Business¹⁵ è possibile creare delle campagne per ottenere maggiore traffico verso il proprio sito ed aumentare la propria autorità in un determinato settore, risultando dunque anch'essa una piattaforma utile per la promozione online.

Anche Amazon, famosa azienda di e-commerce, permette di fare pubblicità online in alternativa alle classiche piattaforme offerte da Google o Facebook. Nello specifico, Amazon ha un suo sistema di pubblicità chiamato Amazon Marketing

¹³ Facebook for Business <https://www.facebook.com/business>

¹⁴ Microsoft Advertising <https://about.ads.microsoft.com/it-it/introduzione/microsoft-advertising-iscriviti>

¹⁵ Quora <https://q.quoraforbusiness.com/business>

Services (AMS)¹⁶, che consente ai venditori che lo desiderano di sponsorizzare i loro prodotti in varie modalità, offrendo la possibilità di portare traffico verso la pagina descrittiva del prodotto ed aumentare di conseguenza le vendite.

La pubblicità online può essere generata anche attraverso altri social network come Reddit, LinkedIn, Pinterest, Snapchat, TikTok, Youtube, Spotify, Instagram, consentendo tutti quanti oltre alle loro funzionalità classiche, la possibilità di inserire contenuti pubblicitari da parte degli inserzionisti, offrendo loro la possibilità di dare agli utenti una via alternativa per raggiungere direttamente i siti da loro sponsorizzati.

2.9 Differenze tra attività SEO e SEM

La sostanziale differenza tra le due discipline consiste nel modo in cui permettono di ottenere visibilità nel web, la SEO consente da una parte di posizionare il sito web in modo naturale, tutte le azioni che richiede non comportano alcun tipo di pagamento ma un investimento in termini di tempo. Questo non implica una minore efficacia di tale strategia, se attuata nel modo giusto garantisce infatti l'ottenimento di una corretta indicizzazione e quindi degli ottimi risultati. Dall'altra parte però, per rafforzare la posizione sul web di una certa pagina o sito, è possibile anche optare per attività di tipo SEM come per esempio l'acquisto di link, l'apertura di spazi pubblicitari su siti rilevanti o l'adesione a campagne Pay per Click come quelle gestite da Google Ads.

Entrambe portano dunque dei risultati se eseguite nel modo giusto, sicuramente una collaborazione tra le due implica una migliore prestazione poiché l'inserzionista ha maggiore possibilità di ottenere ottimi punteggi di qualità per i motori di ricerca, concentrandosi sull'aspetto SEO e rendendo dunque il contenuto della pagina pertinente rispetto a quanto realmente offre all'utente, ma anche concentrandosi sulla parte relativa alle strategie di tipo SEM, partecipando all'asta per l'ottenimento delle prime posizioni ed offrendo un proprio prezzo.

Osservando i risultati ottenuti a seguito di una richiesta effettuata al motore di ricerca Google, è possibile capire come si distinguono i risultati delle due attività svolte, comprendere quindi quali siti sono presenti perché occupano uno degli spazi degli annunci a pagamento e quali invece costituiscono i risultati della cosiddetta ricerca organica. I risultati definiti organici sono quelli che tra i vari risultati nella

¹⁶ Amazon Marketing Services <https://advertising.amazon.com/it-it>

SERP di Google non sono stati creati per mezzo di campagne a pagamento, vengono visualizzati perché ritenuti i più pertinenti dagli algoritmi rispetto alla ricerca che è stata effettuata dal cliente e sono frutto di sole tecniche di tipo SEO. I risultati non organici invece sono quelli a pagamento effettuati in questo caso per mezzo di Google Ads e che compaiono solitamente in testa, prima dei risultati organici o ai lati. Si distinguono dagli altri risultati in quanto sono evidenziati con la scritta “Annuncio”.

Nella Figura 2 si riporta un esempio di SERP Google in cui è possibile distinguere tra le due tipologie di risultato.

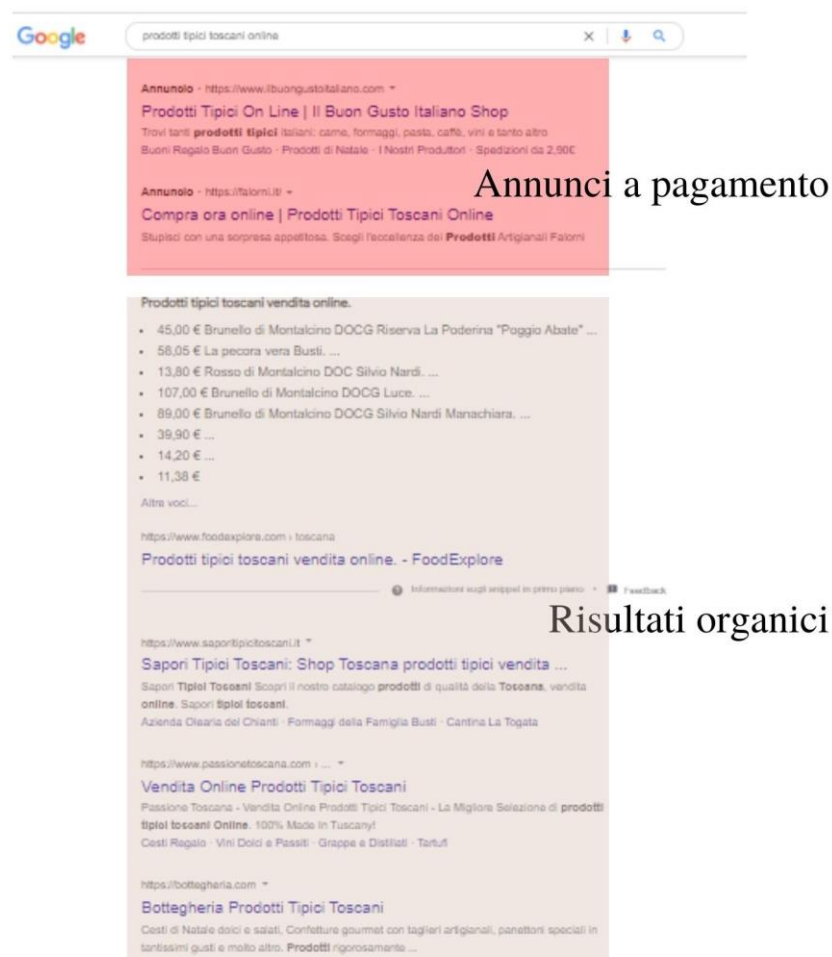


Figura 2. Esempio di risultati organici e annunci a pagamento.

Le azioni di Search Engine Marketing possono risultare molto efficaci, ma talvolta anche molto costose, inoltre è bene tener di conto del fatto che spesso l'attività a pagamento ottiene risultati brillanti nel periodo in cui si intraprende ed una volta

terminata la visibilità lentamente scompare, per questo motivo è fondamentale che vi sia un'ottima sinergia tra visibilità a pagamento e naturale, tra l'attività di SEM e SEO dunque.

Dopo aver individuato nei paragrafi precedenti le principali caratteristiche di una strategia di tipo SEM e di tipo SEO, si vedono rappresentate in Figura 3 le fasi principali da percorrere per avere sotto controllo la propria campagna di posizionamento.

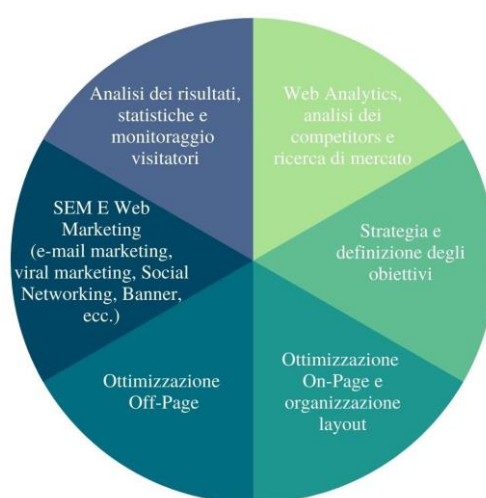


Figura 3. Esempio di modello SEM e SEO¹⁷.

La prima fase si identifica nell'analizzare il mercato, conoscere eventuali competitor e capire le possibili soluzioni da adottare. Già in questa prima parte del lavoro è opportuno effettuare il monitoraggio delle azioni intraprese utilizzando strumenti utili per la Web Analytics, per la raccolta cioè dei dati sui visitatori del sito per poterne analizzare i comportamenti e studiare una strategia di marketing più mirata. L'analisi dei dati aiuta a capire quali delle azioni svolte e delle decisioni prese hanno avuto successo e su quali settori risulta più vantaggioso puntare per cercare di migliorare il business nel web. Vengono consultati per questo i KPI, acronimo di Key Performance Indicators, ovvero indicatori chiave utili a misurare le performance di un'azienda che permettono di valutare il successo nel raggiungimento di certi obiettivi prefissati.

¹⁷ Fonte: Marco Maltraversi. *SEO e SEM, Guida avanzata al web marketing*. Milano, Edizioni FAG, 2011

L'analisi del mercato online deve essere affiancata alle strategie tradizionali, è importante infatti capire come i competitor si muovono nel web, quali attività intraprendono, che mercati toccano. Altri criteri di cui tener conto, come si è visto nei precedenti paragrafi sono strategia e pianificazione che, come in tutte le attività, ricoprono un ruolo cruciale per cercare di ottenere i risultati desiderati. Gli obiettivi vengono prefissati ed insieme ad essi le possibili strategie d'azione, analizzando anche come il mercato si muove ed i risultati ottenuti. Le analisi interne ed esterne alla pagina che racchiudono tutte le caratteristiche basilari dei principali processi in ambito SEO ed infine, le campagne SEM, che permettono di ampliare il raggio d'azione, andando a lavorare non solo sulla pagina stessa, ma anche sulle attività che promuovono il servizio offerto tramite campagne a pagamento, annunci sui Social network e molto altro. Tra le tecniche più utilizzate in questo ambito vi è la keyword advertising, metodo di pubblicità che viene attuato sui motori di ricerca e permette di determinare le ricerche per le parole chiave più adatte all'offerta della propria attività, posizionando i propri annunci nei risultati della SERP. Per concludere, l'ultimo step da affrontare consiste nell'analisi dei risultati, è importante studiare i dati ottenuti, monitorare le azioni messe in atto e comprendere quali di queste hanno raggiunto l'obiettivo e quali no.

Nel corso del capitolo sono state viste le principali fasi che si susseguono nel percorso di promozione di un sito web o di un'attività nel mercato online. I punti su cui porre l'attenzione sono molti, tutti di importanza rilevante in quanto ognuno di essi nel loro campo d'azione specifico, giocano un ruolo decisivo per l'ottenimento dell'obiettivo finale. Nel prossimo capitolo si pone l'attenzione su uno di questi elementi, ovvero lo studio e l'importanza che ricoprono le keywords nelle strategie SEO e SEM, in quanto tutto il progetto presentato è stato svolto proprio sulle parole chiave utilizzate per la generazione di annunci a pagamento.

3. L'importanza delle keywords

Tra le varie fasi della Search engine marketing vi è la ricerca delle parole chiave da abbinare al proprio annuncio o sito web. Si definisce in inglese come Keyword research e corrisponde ad una delle attività SEM più importanti e ad alto rendimento. In questo capitolo si descrivono gli strumenti, le tecniche e le caratteristiche della fase di pianificazione di parole chiave per il Search advertising.

Tramite lo studio sulla domanda di parole chiave del mercato di riferimento, si conoscono non solo quali termini e frasi indicizzare con le strategie SEO, ma anche molte più informazioni sui propri clienti nel loro insieme. La keyword research permette di prevedere i cambiamenti della domanda, rispondere alle mutevoli condizioni del mercato e produrre i contenuti relativi ai prodotti ed ai servizi che gli utenti stanno già attivamente cercando nel web. Ogni frase o termine di ricerca digitata in un motore viene registrata e gli strumenti di ricerca delle parole chiave esistenti, consentono di recuperare tutte queste informazioni. Le parole chiave costituiscono per questo un elemento molto importante nell'ambito SEM in quanto permettono di esprimere le tematiche principali affrontate dal sito che si vuole promuovere o dal servizio e prodotti che si offrono ai possibili clienti. Gli utenti nel momento in cui si avvicinano al mondo di Internet per effettuare un acquisto o cercare informazioni, utilizzano proprio le parole chiave come termini di ricerca, poiché permettono loro di verbalizzare i problemi o gli argomenti su cui desiderano informazioni. Le aziende creano contenuti su argomenti riguardanti il loro business che le aiutano a essere classificate e rintracciate, mentre dall'altro lato ci sono gli utenti che tramite un motore di ricerca, effettuano le loro richieste utilizzando delle parole che richiamano l'argomento cui sono interessati. Le keyword per questo sono alla base del contenuto di un sito web, il tema trattato in ogni pagina dovrebbe ricondursi direttamente ad una specifica parola o frase chiave, in quanto aiutano in questo modo i visitatori ed i potenziali clienti a capirne l'argomento principale, oltre ad essere utili per i motori di ricerca che durante la fase di indicizzazione, possono comprendere tramite l'analisi delle keywords, se il sito può essere associato o meno alla richiesta effettuata dall'utente.

Nel primo capitolo, in parte è già stato affrontato questo tema, in quanto lo studio delle parole chiave rappresenta uno degli elementi di cui trattare nella fase di

pianificazione SEM e SEO, devono essere scelte in modo da rendere i contenuti delle proprie pagine accattivanti, facili da rintracciare per l'utente e pertinenti rispetto a quanto riportato, così che, un motore di ricerca come ad esempio Google possa reputare gli argomenti proposti di qualità elevata e far balzare il sito tra le prime posizioni dei risultati.

La scelta delle giuste keyword prevede così come gli altri elementi SEO e SEM, una vera e propria pianificazione ed è importante che in questa fase si tenga conto dei bisogni dell'utente, di come questo possa ricercare i contenuti che gli si offrono e di come è solito comportarsi nelle ricerche sul Web. La pianificazione e la selezione delle parole chiave non prevede però solo l'ambito SEO, è infatti necessario scegliere keywords da poter utilizzare anche nelle campagne a pagamento come, ad esempio, negli annunci promossi tramite lo strumento Google Ads.

Google Ads, come anticipato, è una piattaforma offerta da Google per la gestione delle campagne a pagamento. Permette di creare annunci che l'utente visualizza tra le prime pagine dei risultati, segnalati dal termine "Annuncio", se Google li ritiene idonei e pertinenti al servizio o alla pagina che viene offerta. L'inserzionista offre un valore monetario per i click che riceve, Google non assegna però le prime posizioni solo in base a quanto gli viene offerto ma anche rispetto alla qualità dei contenuti. Questo perché se si concentrasse solo sul premiare il miglior offerente, potrebbe capitare che i risultati visualizzati dall'utente non siano realmente in linea con la ricerca effettuata e quindi non otterrebbe un'adeguata percentuale di clic rispetto al numero di impressioni. Preferisce dunque mostrare annunci con un costo per clic massimo offerto inferiore, ma più pertinenti rispetto a quanto descrivono, così da rendere l'esperienza dell'utente soddisfacente e garantire il suo ritorno.

3.1 L'algoritmo di ricerca di Google

Nel 2013 Google ha rilasciato un nuovo algoritmo di ricerca, *Hummingbird*, che ancora oggi influenza una buona percentuale dei risultati nella SERP.

In generale, un algoritmo di ricerca, si avvia nel momento in cui una persona effettua una ricerca tramite uno dei vari motori esistenti. L'algoritmo non appena l'utente effettua la propria richiesta, analizza rapidamente milioni di pagine con l'obiettivo di far visualizzare i risultati migliori per offrire all'utente un'esperienza di ricerca positiva. Studiare tutto ciò che riguarda la SEO, significa reperire anche le

informazioni che vengono fornite riguardo al funzionamento degli algoritmi di ricerca usate dai motori con cui principalmente ci si appropria, così da poterne prevedere il comportamento ed eventuali evoluzioni.

Hummingbird, è l'algoritmo di ricerca specifico di Google, il cuore del suo funzionamento consiste nel fatto che agisce sul significato delle parole chiave, sulle frasi di ricerca e sulla qualità dei testi inseriti all'interno di una pagina web. Per fornire risultati soddisfacenti ai propri utenti, Google deve saper comprendere il vero significato di un testo e Hummingbird si fonda proprio su tale principio. È stato creato per riuscire a gestire le tipiche frasi di ricerca della nostra epoca, lunghe e dettagliate, spesso sottoforma di domanda, che gli utenti pongono quotidianamente al motore di ricerca. L'algoritmo di Google deve interpretare tali domande e soprattutto individuare i contenuti giusti per rispondere, riesce a farlo poiché è stato costruito in modo da poter riconoscere il significato delle parole chiave e può quindi anche con lunghe e articolate richieste indicizzare le pagine che meglio le soddisfano. Lo scopo di Hummingbird è quello di capire pienamente un testo, al fine di comprenderne il grado di qualità. Per questo come si è già sottolineato, più i contenuti sono di qualità maggiore sarà la possibilità di ottenere un posto nella prima pagina della SERP di Google ed aumentare le vendite di conseguenza.

3.2 Pianificazione delle parole chiave

La decisione delle keywords da utilizzare è un processo estremamente importante e complesso poiché permettono di mantenere o posizionare il sito web tra le prime risposte che i motori di ricerca offrono. Gli elementi di cui tener conto anche in questo caso sono differenti, per prima cosa può risultare utile osservare come agiscono i propri concorrenti, analizzare i loro siti e capire quali parole chiave utilizzano principalmente. Si procede quindi con l'analisi dei loro siti web, si osservano le parole chiave o le frasi da loro utilizzate nella descrizione dei prodotti e servizi che possono competere con i propri. Definire inoltre una lista di alternative e sinonimi dei vocaboli da loro utilizzati, verificare i contenuti pubblicati, studiare i primi risultati della SERP di ricerca all'interno dei principali motori di ricerca con contenuti affini a quanto si vuole pubblicizzare.

Non solo la concorrenza va studiata, ma anche e soprattutto il target di clienti su cui si vuole puntare, i visitatori sono alla ricerca di informazioni e per questo

interrogano i vari motori di ricerca con keyword differenti. Importante quindi è selezionare le parole chiave più popolari tra i visitatori di un determinato sito web, studiando a fondo la tipologia di utenti su cui il sito vuole puntare e scegliere di conseguenza in modo adeguato e consapevole le keyword per ogni singola pagina.

Quando gli utenti accedono ai motori di ricerca e digitano la loro specifica richiesta possono utilizzare un linguaggio completamente diverso da quello previsto e, anche se il prodotto o servizio fornito può soddisfare le loro esigenze, potrebbe accadere che l'annuncio o il sito non vengano visualizzati poiché non perfettamente in linea con i termini di ricerca da loro composti: dedicare del tempo alla comprensione delle tipiche sequenze di ricerca del proprio pubblico è un aspetto che influisce notevolmente sulla decisione finale. Altri elementi che influenzano il comportamento di ricerca e che vanno esaminati includono tutto ciò che fa riferimento ai dati demografici dell'utente, le indicazioni fornite dai vari strumenti riguardanti il sesso delle persone, l'età, il reddito, la posizione geografica.

Esaminato il proprio pubblico, i competitors e nel caso di siti già avviati, anche i dati statistici relativi a quanto accaduto fino a quel momento con le attività e le strategie intraprese, si procede con la stesura di un elenco di possibili parole. Nel manuale *The art of SEO: Mastering Search Engine Optimization*, si può leggere una lista dettagliata di punti da considerare durante questa fase. Riportandone alcuni, i tre autori suggeriscono di produrre una lista di chiavi, composte da una a tre parole, che descrivono il prodotto o servizio. Considerare eventuali sinonimi che i clienti potrebbero digitare per la ricerca di quanto offerto consultando appositi dizionari. Creare una classificazione di termini di maggiore livello e di argomenti di cui possono fare parte i prodotti in oggetto. Utilizzare gli strumenti esistenti che offrono tra le varie funzionalità, la possibilità di vedere quali termini di ricerca hanno già utilizzato le persone per raggiungere il proprio sito, nel caso in cui se ne possieda uno.

L'elenco generato dovrà contenere possibili parole chiave che differiscono sia per i termini che contengono sia per il loro grado di specificità. Esistono in generale due principali tipologie di keywords: le generiche o le specifiche. Le generiche, definite anche a coda corta o *Short Tail Keyword*, sono parole chiave brevi composte al massimo da uno o due termini e per questo si definiscono generiche, proprio perché descrivono il contenuto in modo generale. Sono quelle che più facilmente vengono utilizzate anche da tutte le altre aziende che offrono prodotti o servizi simili, quelle a cui è molto probabile che corrisponda un numero di ricerche effettuato molto elevato,

poiché sono le più scontate ed immediate. Le specifiche o *Long Tail Keyword*, sono invece quelle parole chiave costituite da più termini, andando a descrivere in modo più dettagliato ciò che si sta offrendo. Corrisponde loro un numero di ricerche più basso, poiché essendo composte da più termini risulta meno frequente la possibilità che molte persone che effettuano richieste dettagliate, vadano a digitare esattamente la stessa identica sequenza di termini.

Il dataset su cui si è svolto il lavoro presentato fa riferimento ad un e-commerce di prodotti tipici toscani, per fare dunque qualche esempio di parole generiche e specifiche rimanendo in tema con quanto analizzato, per la prima tipologia di keyword, se si considera un utente interessato a ricercare siti di prodotti tipici toscani online si possono riportare casi come “prodotti tipici”, “prodotti toscani”, “prodotti”. Per l’altra tipologia la ricerca sarebbe stata effettuata invece tramite termini di ricerca quali “prodotti tipici toscani online”, “vendita di prodotti tipici toscani online”. Immediatamente si nota un’informazione più dettagliata nella seconda tipologia di esempi.

Rispetto a questa distinzione è stata ideata la Teoria della coda lunga, proposta da Chris Anderson, un saggista e giornalista statunitense, che ha descritto inizialmente il concetto alla base in un articolo pubblicato nel 2004 sulla rivista *Wired*¹⁸, poi nel suo libro *The Long Tail: Why the Future of Business Is Selling Less of More*¹⁹, pubblicato nel 2006. La teoria della coda lunga, afferma che le ricerche effettuate in Internet tendono ad essere molto lunghe e focalizza l’attenzione sulla possibilità che molte persone possano essere disposte ad acquistare oggetti poco conosciuti all’interno della rete, a scapito di quella branchia di persone che tende a comprare solo i prodotti commerciali di marche conosciute. Vuole quindi sottolineare il concetto secondo cui i prodotti meno popolari e non di nicchia, possono produrre, se presi in maniera collettiva, un numero più elevato di vendite rispetto a prodotti con maggiore popolarità e si riferisce quindi a tutte quelle chiavi che di solito sono caratterizzate dal portare poco traffico se prese singolarmente in quanto composte da più termini, dando vita a frasi lunghe e dettagliate e non contenenti nomi specifici dei brand interessati alla ricerca. Più la chiave è lunga e dettagliata, più possibilità si hanno di dare risposte precise agli utenti che entrano nel sito, anche se il loro numero non è elevato. Per

¹⁸ Anderson, Chris. 2004. *The Long Tail*. In "Wired" <https://www.wired.com/2004/10/tail/>

¹⁹ Chris Anderson. *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion, 2006

questo motivo, concentrarsi sulla coda lunga può portare a risultati sorprendenti, poiché l'insieme di tanti piccoli risultati, appare solitamente migliore dei singoli maggiori risultati. In Figura 4 è possibile visualizzare tale concetto, dove si può notare che i pochi utenti che approdano sul sito utilizzando keyword specifiche, corrispondono alla parte costituita dalla coda lunga la quale rappresenta una varietà di parole chiave formate a partire dalla combinazione di più vocaboli, che possono assumere una rilevanza sempre maggiore. Proprio su queste si dovrebbe quindi prestare maggiore attenzione, in quanto generano solitamente bassi volumi di accessi verso il sito perché hanno una competitività minore, ma se prese complessivamente possono fruttare un numero di accessi molto importanti, mirano infatti proprio a raggiungere un target specifico di clienti.

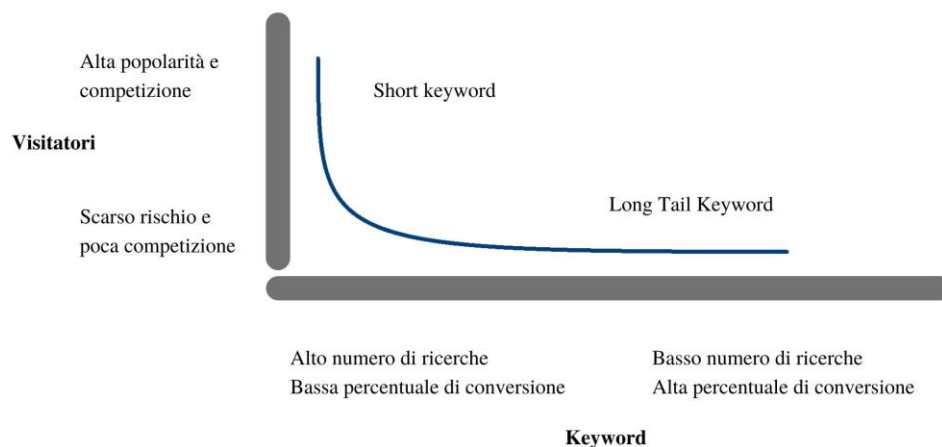


Figura 4. Grafico che rappresenta l'andamento delle keyword secondo la teoria della coda lunga²⁰.

Gli utenti che svolgono ricerche tanto dettagliate interrogando il motore di ricerca con parole chiave formate da tre o quattro termini, sono persone che già sanno ciò di cui hanno bisogno, consapevoli di quello a cui sono interessati e dunque probabilmente già pronti ad acquistare, per questo motivo è una tipologia di ricerche con un tasso di conversione alto. Generano quindi un traffico inferiore in termini quantitativi, ma decisamente migliore dal punto di vista qualitativo.

Il grafico permette di comprendere come le keyword a coda lunga possiedono una percentuale di conversione maggiore rispetto ai termini più generici. Una persona

²⁰ Fonte: Eric Enge, Stephan Spencer, Jessie C. Stricchiola. *The art of SEO, Mastering Search Engine Optimization*. Sebastopol, O'Really Media, 2015

che cerca una parola generica come ad esempio “Prodotti tipici” avrà meno possibilità di raggiungere l’annuncio relativo al sito del negozio che si vuole pubblicizzare rispetto a chi utilizza termini di ricerca più specifici come “Negozio di prodotti tipici toscani a Lucca”. Va comunque specificato che non sempre vale la regola secondo cui la coda corta sia sinonimo di basso tasso di conversione, soprattutto nel caso di campagne PPC strutturate o quando si ha a che fare con brand molto forti sul mercato.

3.3 Keyword per annunci a pagamento

Le parole chiave sono importanti in ambito SEO, come si è visto, ma lo sono anche per le tecniche che riguardano la pubblicità online. Il termine Keyword Advertising indica un metodo per fare pubblicità che si basa proprio sulle parole chiave, ovvero l’inserzionista che offre un prezzo per ottenere visualizzazioni, sceglie ogni singola parola per cui desidera che il proprio annuncio venga pubblicato e dall’altro lato l’utente che digita la propria richiesta ottiene tra i risultati anche messaggi promozionali attinenti alle stesse parole chiave da lui ricercate.

Nella fase di pubblicazione di un annuncio a pagamento su Google, ad esempio, si identificano una serie di parole chiave, oltre agli altri elementi richiesti, ed il motore di ricerca se considera l’annuncio pertinente a quanto l’utente sta ricercando, lo farà comparire tra le prime posizioni. Le parole chiave vengono utilizzate dunque per abbinare gli annunci ai termini utilizzati dagli utenti per le ricerche, permettono all’inserzionista di essere visti dal cliente e di avere la possibilità di ottenere una conversione. La fase di keyword advertising in questo modo, proprio perché mirata e inerente all’argomento d’interesse per l’utente, ha molte possibilità di risultare efficace. La creazione dell’annuncio, oltre alla scelta dei termini da utilizzare come parole chiave, offre la possibilità di scelta del tipo di corrispondenza. Il tipo di corrispondenza delle parole chiave permette di determinare quali termini di ricerca degli utenti attiveranno la pubblicazione dell’annuncio, ovvero, come si legge dalla Guida di Google Ads²¹, in che misura la parola chiave deve corrispondere alla query di ricerca dell’utente in modo che l’annuncio possa essere preso in considerazione per l’asta. Questa scelta permette di aggiungere un dettaglio alla keyword, in quanto dà la possibilità di decidere se l’annuncio cui fa riferimento la parola chiave, possa essere

²¹ Guida di Google Ads, voce *Informazioni sulle opzioni di corrispondenza delle parole chiave*
<https://support.google.com/google-ads/answer/7478529> (visitato il 30 settembre 2021)

preso in considerazione dal motore di ricerca solo se la richiesta dell'utente contiene una serie di termini specifici, o se ne contiene alcuni di quelli presenti nella keyword. Google Ads, ad esempio, dà la possibilità di scegliere tra “corrispondenza generica”, “corrispondenza a frase” o “corrispondenza esatta”.

La tipologia di corrispondenza generica è quella di default ed è pensata principalmente per mostrare l'annuncio ad un'ampia gamma di persone con l'obiettivo di portare quanto più traffico possibile sulla pagina di destinazione. Essendo predefinita, non va specificata, ma si attiva in automatico per ogni tipo di parola per cui non si è specificata la tipologia di corrispondenza, la sintassi prevede l'inserimento dei termini scelti senza nessun'altra indicazione. Il concetto alla base di tale tipologia è che gli annunci possono essere pubblicati per ricerche correlate alla parola chiave scelta, ovvero per quelle richieste dell'utente che possono anche non includere i termini corrispondenti alla specifica keyword. In questo modo si ha la possibilità di attirare un numero di visitatori più elevato sul proprio sito web e dedicare meno tempo alla creazione di elenchi di possibili alternative, ma allo stesso modo il controllo sulla campagna da parte dell'advertiser è meno incisivo, poiché Google ha maggiore libertà di scelta su quando far visualizzare l'annuncio e soprattutto lo considera pertinente anche rispetto a ricerche che possono essere collegate al prodotto offerto ma che non sono poi così affini. Per esempio, l'annuncio con la parola chiave *prodotti tipici toscani* può essere abbinato a richieste effettuate tramite termini correlati come *ristorante di prodotti tipici toscani*, *migliori prodotti tipici toscani* o *vendita di prodotti tipici toscani*, l'utente si troverà tra i risultati di ricerca varie possibili opzioni che richiamano quanto da lui digitato.

La tipologia di corrispondenza a frase, attiva l'annuncio quando un utente utilizza come termine di ricerca una frase che corrisponde alla parola chiave dello stesso e con le parole nell'esatto ordine scelto, oppure quando è presente la frase scelta seguita o preceduta da altre parole. Questa tipologia prevede che gli annunci possano essere pubblicati per ricerche che includono il significato della parola chiave, risulta molto utile per meglio definire il target dei propri annunci e allo stesso tempo per diminuire la probabilità che la pubblicazione venga attivata anche nei casi in cui i clienti utilizzano termini non correlati al prodotto o servizio offerti. Un esempio di questo tipo di corrispondenza si verifica quando l'annuncio con una parola chiave come “*prodotti tipici toscani*” viene attivato rispetto a ricerche di utente quali *prodotti tipici toscani*, *prodotti tipici toscani online*, *vendita online prodotti tipici toscani*. La

sintassi che caratterizza questa tipologia si forma con l'inserimento della parola chiave tra virgolette, come riportato sopra.

Nel caso della tipologia di corrispondenza esatta, gli annunci vengono pubblicati con ricerche che hanno lo stesso significato o lo stesso intento della parola chiave, la query di ricerca dell'utente deve corrispondere esattamente alla frase che costituisce la parola chiave. A differenza della precedente, non attiva la pubblicazione se l'utente inserisce o prima o dopo anche altre parole, genera sicuramente meno impressioni ma al contempo permette di indirizzare i termini scelti su ciò che l'utente sta realmente cercando, offre quindi il massimo controllo su chi vede l'annuncio, raggiungendo però meno persone rispetto alle altre due. In Google Ads, la sintassi per questa tipologia prevede l'inserimento della parola chiave tra parentesi quadre. Se per l'annuncio è stata scelta la parola chiave a corrispondenza esatta [*vendita di prodotti tipici toscani*], l'attivazione avverrà con ricerche come *vendita di prodotti tipici toscani* oppure *vendita prodotti tipici toscani*, ma non con *vendita di prodotti tipici toscani online*.

Google Ads oltre alle tipologie di corrispondenza appena viste, offre la possibilità di definire anche delle parole da escludere, ovvero parole chiave per cui non è prevista la pubblicazione dell'annuncio nel caso in cui siano utilizzate come termini di ricerca. L'applicazione di parole chiave escluse può aiutare ad evitare costi inutili a causa di clic accidentali o non pertinenti. Oltre a creare un elenco di parole chiave rilevanti per il proprio annuncio, è utile scegliere anche parole irrilevanti che si vogliono escludere, così da poter impedire la pubblicazione dei propri annunci quando vengono effettuate ricerche con tali termini. Se un negozio vende prodotti tipici toscani, ma non dispone di un particolare alimento, può questo essere aggiunto alla lista delle parole chiave escluse così da evitare visite poco utili, sia per l'inserzionista che dovrà pagare il clic, sia per l'utente che si trova tra i risultati di ricerca informazioni poco affini a quanto richiesto. Per determinare le parole chiave da escludere è possibile individuare termini di ricerca simili alle proprie keywords, ma potenzialmente interessanti per utenti che cercano un prodotto diverso da quello offerto. Per rendere una campagna fortemente mirata, è necessario stabilire anche quale target di utenti può risultare poco interessante per il proprio prodotto o servizio. Con l'aiuto delle parole escluse si può definire un targeting più preciso e di conseguenza raggiungere più facilmente solo i clienti realmente interessati all'annuncio, coloro dunque che possono effettivamente incrementare il ritorno sull'investimento (ROI).

Selezionando parole da escludere per le campagne sulla rete di ricerca, è possibile specificare anche in questo caso la tipologia di corrispondenza tra generica, di frase o esatta. Tuttavia, questi tipi di corrispondenza nel caso delle parole escluse funzionano in modo diverso rispetto a quanto accade con le keywords scelte per la pubblicazione dell'annuncio. Una keyword generica viene associata a query che possono contenere varianti simili, errori di grammatica, sinonimi, ricerche correlate e molto altro. Se non viene specificata nessuna preferenza, anche per le parole chiave escluse la tipologia predefinita sarà quella generica. In questo caso, gli annunci verranno evitati per il solo fatto di contenere tutti i termini che costituiscono la parola chiave esclusa, senza bisogno di doversi preoccupare per il loro ordine.

Nel caso della corrispondenza generica esclusa, è però necessario escludere le forme singolare, plurale e maschile, femminile, poiché non saranno automaticamente dedotte dalla piattaforma, per questo se come parola chiave esclusa si indica *birre artigianali*, l'annuncio verrà comunque associato a ricerche come *birra artigianale*.

Le parole chiave con corrispondenza a frase esclusa associate ad un certo annuncio, fanno in modo che questo non venga visualizzato se la parola chiave appare nella ricerca con lo stesso preciso ordine anche se fossero inclusi altri termini. La sequenza deve essere cercata esattamente nel modo specificato, se presenti ad esempio delle lettere in più, l'annuncio potrebbe comunque essere visualizzato.

Le parole chiave a corrispondenza esatta esclusa permettono di fare in modo che l'annuncio non venga pubblicato se la ricerca dell'utente contiene i termini esatti della parola chiave esclusa, nello stesso ordine e senza parole aggiuntive. Eventuali parole in più potrebbero causare la pubblicazione dell'annuncio.

La scelta delle parole chiave da escludere consiste in un passaggio delicato in quanto può risultare da una parte molto utile, ma allo stesso tempo anche poco efficace poiché utilizzarne troppe potrebbe comportare che l'annuncio venga raggiunto solo da un numero molto ristretto di clienti. In generale, quanto più generica è la corrispondenza scelta tanto più sarà il traffico che la parola chiave può generare, mentre al contrario più la corrispondenza è specifica, maggiore sarà il controllo e la pertinenza sulle ricerche che gli utenti effettuano. Coloro che sono in possesso di un account Google Ads, possono consultare la tipologia di corrispondenza per ogni parola chiave tramite la tabella delle statistiche che riporta per ognuna le rispettive informazioni.

Gli elementi che vengono mostrati per ogni parola chiave sono diversi e si affronteranno più nel dettaglio nel prossimo capitolo, nel quale viene spiegato il lavoro e lo studio svolti per il compimento del progetto presentato. In questo paragrafo, si riporta un riferimento, oltre alla tipologia di corrispondenza anche allo stato delle parole chiave, altra caratteristica che può essere visualizzata.

Lo stato di una parola chiave indica se questa è attiva o non attiva. Nel caso in cui non lo sia, significa che non compare negli annunci pubblicati, Google quindi provvede in tali casi ad indicarne la ragione. Permette di capire se la parola chiave cui fa riferimento determina la visualizzazione degli annunci da parte del target di clienti selezionato. In generale, può indicare che gli annunci vengono pubblicati, non vengono pubblicati perché si è deciso di mettere in pausa o di rimuovere una parola chiave, un gruppo di annunci o una campagna, o che non vengono pubblicati perché si è verificato un problema relativo all'account o alla parola chiave stessa.

Nella Guida di Google Ads²², relativa alle informazioni sullo stato delle parole chiave, è possibile conoscere il significato dei valori che la colonna “Stato” può assumere rispetto alle diverse parole. Tali valori indicano le differenti situazioni in cui si trova la parola chiave e di conseguenza l’annuncio cui fa riferimento, le quali sono frutto di decisioni prese da chi gestisce il profilo Google Ads dell’azienda e dunque le campagne pubblicitarie.

Quando viene riportata per una parola chiave l’indicazione *Non idonea*, può corrispondere ad uno dei seguenti casi:

- Campagna in pausa, le parole chiave non sono attive perché la campagna è stata messa in pausa e quindi anche le parole chiave che le appartengono cessano di pubblicare annunci.
- Campagna in attesa, le parole chiave non sono attive perché la pubblicazione della campagna è prevista per una data futura e i relativi annunci non vengono ancora pubblicati.
- Campagna rimossa, le parole chiave non sono attive poiché si è deciso di rimuovere la campagna che le contiene.

²² Guida Google Ads, voce *Informazioni sullo stato delle parole chiave*
<https://support.google.com/google-ads/answer/2453978?hl=it> (visitato il 30 settembre 2021)

- Campagna terminata, indica che le parole chiave non sono attive perché la campagna ha superato la data che era stata definita inizialmente come data di fine e per questo non è più attiva.
- Gruppo di annunci in pausa, si verifica quando le parole chiave non sono attive perché il gruppo di annunci è stato messo in pausa e quindi le parole chiave di quest'ultimo interrompono la pubblicazione degli annunci.
- Gruppo di annunci rimosso, le parole chiave non sono attive perché il gruppo di annunci è stato in questo caso rimosso e le relative parole chiave interrompono la pubblicazione degli annunci.

Lo stato della parola chiave quando corrisponde a *In pausa*, specifica che è inattiva, in quanto messa in pausa e per tale ragione non pubblica annunci e non accumula di conseguenza nuovi costi fino al momento della riattivazione. Se invece indica *Rimossa*, allora significa che la parola chiave è stata rimossa, per questo non è attiva e quindi non pubblica più gli annunci che le corrispondono.

Parola chiave *Idonea*, indica che è idonea alla pubblicazione di annunci, mentre *Non approvata*, significa che non è attiva e non pubblica annunci. In quest'ultimo caso, solitamente il valore dello stato indica che è presente un problema con una o più norme pubblicitarie. Deve essere rivista la parola chiave stessa in quanto viola una o più delle regole vigenti. Infine, lo stato può indicare *Idonea (limitata)*, questo implica che la parola chiave sia idonea ed attiva per la pubblicazione degli annunci, ma riscontra qualche problema e quindi è probabile che non riesca a raggiungere ottimi risultati. L'essere limitata, rappresenta principalmente le seguenti situazioni:

- L'offerta è inferiore all'offerta stimata per la prima pagina, la parola chiave è attiva e può pubblicare annunci, ma non è attualmente in grado di pubblicarne nella prima pagina dei risultati di ricerca di Google.
- Basso volume di ricerca, la parola chiave non è idonea a pubblicare annunci poiché associata ad un volume di traffico di ricerca molto scarso, non risulta dunque pertinente alle ricerche degli utenti o è troppo specifica o insolita.
- Gli annunci vengono pubblicati raramente in quanto il punteggio di qualità che corrisponde loro è basso.
- Poche corrispondenze di pagina, si verifica quando l'annuncio è stato creato per la Rete Display, per essere quindi pubblicato su altre pagine o siti web, ma

la parola chiave è temporaneamente sospesa perché corrisponde a pochissime pagine, ossia la frase non è comune nei contenuti presenti nella Rete Display.

Consultando la guida di Google Ads, è possibile inoltre conoscere modi per gestire i casi appena presentati, fornendo ottimi suggerimenti per il raggiungimento dell'obiettivo finale.

3.4 Alcuni strumenti per l'analisi delle parole chiave

La keyword research, ovvero la ricerca delle parole chiave, corrisponde ad una fase fondamentale: più la qualità è alta, maggiore è la possibilità di attirare in modo efficiente traffico verso il sito. Per rendere questa ricerca completa, oltre ai passaggi descritti, esistono molti tool specifici che aiutano gli esperti nell'ottimizzazione delle pagine da loro gestite ed offrono funzionalità molteplici per la comprensione delle ricerche effettuate rispetto ad una specifica parola chiave, ed aiuta a scoprire possibili alternative adatte alla tematica su cui si sta lavorando.

Tra i vari, due che personalmente ho avuto la possibilità di vedere durante il percorso di tirocinio sono Google Trends e SEMrush. Google Trends²³, strumento offerto da Google che offre la possibilità di visualizzare l'andamento della popolarità di una parola chiave nel tempo e rispetto ad un'area geografica di interesse. Digitando un particolare termine di ricerca, mostra l'andamento delle richieste su Google nel tempo, le zone geografiche maggiormente interessate all'argomento, eventuali altre tematiche che possono essere correlate a quanto studiato e altre richieste effettuate dagli utenti che vi si possono associare e dalle quali si può prendere spunto. Permette inoltre di vedere la stima di Google di quali città, regioni e lingue hanno eseguito il maggior numero di ricerche per una determinata parola chiave. I risultati possono essere filtrati rispetto alle proprie esigenze, si può selezionare una specifica area geografica oppure un periodo particolare dell'anno. Google Trends è completamente gratuito e per questo si è potuto testare con un esempio per mostrare il tipo di informazioni che riporta ed in che modo le rappresenta. In figura 5 e 6, si osservano i risultati offerti da Google Trends rispetto al termine di ricerca *prodotti tipici toscani*. La figura 5 corrisponde alla prima sezione dei risultati nella quale risulta immediatamente visibile il grafico dell'andamento nel tempo. In questo caso si

²³ Google Trends <https://trends.google.it/trends/?geo=IT>

possono verificare il numero di ricerche effettuate con la parola chiave selezionata, in Italia negli ultimi 12 mesi, ma come si vede sopra al grafico può essere ristretto il campo, selezionando aree geografiche più ristrette o periodi di tempo differenti.

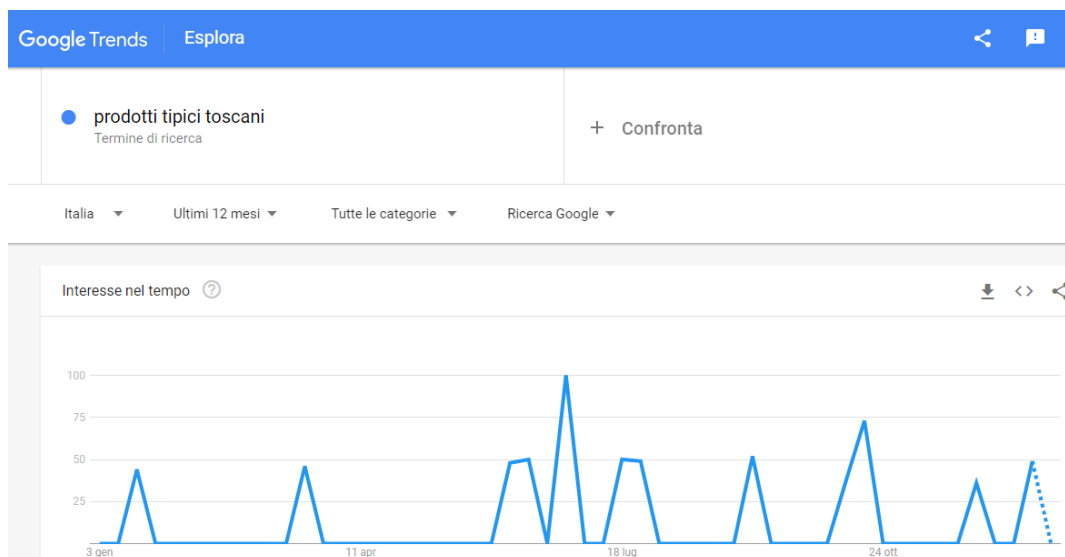


Figura 5. Andamento nell’anno 2021 delle ricerche su Google con il termine selezionato, preso da Google Trends.

La figura 6, riporta invece la seconda sezione dei risultati di Google Trends, nella quale si pone l’attenzione sulle regioni, in questo caso d’Italia, che più sono interessate all’argomento e che hanno effettuato ricerche su Google con la parola chiave.

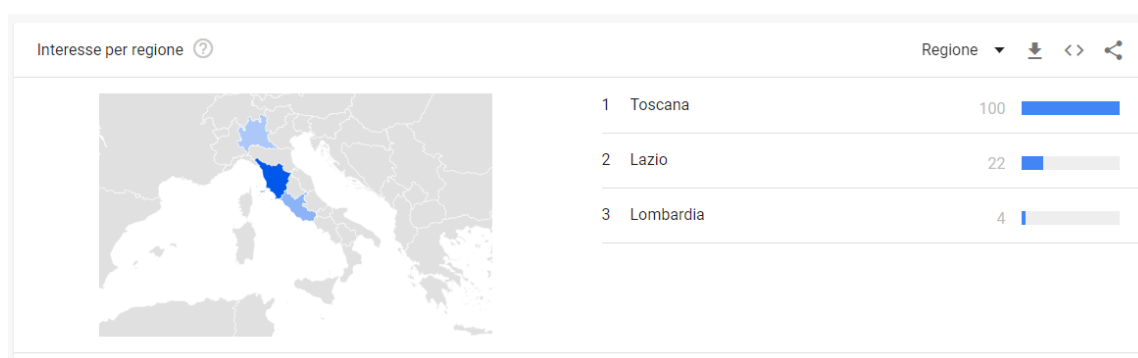


Figura 6. Dati relativi all’interesse per regione della parola chiave studiata, presi da Google Trends.

Google Trends offre anche la possibilità di confrontare due o più termini di ricerca tra loro per comprendere quale, rispetto al tempo o alla zona, è risultato più

popolare. Inserendo i termini nella ricerca si vedrà la cronologia delle tendenze dei termini richiesti rappresentata in diversi colori su un grafico distribuito in un determinato periodo di tempo, anche in questo caso è possibile modificare i risultati cambiando il periodo di tempo e la regione.

SEMrush²⁴, è una piattaforma che fornisce dati relativi alle ricerche effettuate sul web sia organiche che a pagamento. Raccoglie i dati relativi alle varie keyword facendo riferimento ai risultati di ricerca Google e le organizza in database, al momento sono più di 20 miliardi²⁵. Per utilizzare i servizi offerti, è necessario inserire un dominio, una URL o una keyword nella barra di ricerca ed i report che SEMrush riporta riguardano i risultati relativi alle ricerche organiche ed a quelle a pagamento, che sono tra loro separati, ma per entrambi fa riferimento alle parole chiave che hanno portato gli utenti alla pagina web o ad un determinato annuncio, osservando i primi venti risultati di Google. Per ogni keyword si riscontrano varie informazioni come la posizione raggiunta nei risultati di ricerca di Google, la quantità di richieste effettuate, il valore di CPC, l'URL della pagina di destinazione corrispondente e molto altro. Viene visualizzato anche un grafico che rappresenta il numero di parole chiave per cui un dominio è stato posizionato tra i primi venti risultati, rispetto ad un periodo di tempo osservato. Le parole chiave vengono inoltre raggruppate rispetto alle posizioni a cui hanno portato i siti web, per comprendere immediatamente quali di queste può garantire una posizione alta o bassa nei risultati di ricerca di Google. SEMrush dà la possibilità di analizzare anche i propri competitor, tramite una sezione del report delle statistiche che fornisce i domini dei siti concorrenti e le parole chiave comuni per dare la possibilità di vedere i risultati a cui certe chiavi hanno loro portato.

Molti sono gli altri strumenti esistenti che offrono funzionalità simili a quelli descritti, volendone presentare alcuni, si citano Bing Keyword Research²⁶, SEOZoom²⁷, Ahrefs Keywords Explorer²⁸ e Google Keyword Planner²⁹.

Bing Keyword Research, è uno strumento che genera il suggerimento di possibili keyword basato su un termine di ricerca o un sito web selezionati. Recupera i dati dalle richieste che gli utenti effettuano tramite il motore di ricerca Microsoft Bing di

²⁴ SEMrush <https://www.semrush.com/>

²⁵ Statistiche SEMrush aggiornate al 30 dicembre 2021 <https://it.semrush.com/stats/> (visitato il 30 dicembre 2021)

²⁶ Bing Keyword Research <https://www.bing.com/webmasters/help/keyword-research-628070b6>

²⁷ SEOZoom <https://www.seozoom.it/keyword-research/>

²⁸ Ahrefs Keywords Explorer <https://ahrefs.com/it/keywords-explorer>

²⁹ Google Keyword Planner https://ads.google.com/intl/it_it/home/tools/keyword-planner/

proprietà di Microsoft e permette di esportare i risultati in file Excel, così da poterli visionare su fogli di calcolo. L'inserimento di una parola chiave nella barra di ricerca restituisce dati che includono frasi di ricerca contenenti la parola chiave digitata durante il mese precedente, le percentuali di clic ottenute rispetto al numero di volte in cui l'annuncio è stato visto ed il costo medio per clic. Il tool Bing Keyword Research è gratuito, ma richiede la creazione di un profilo su Bing Ads e di fornire quindi i dati della propria carta di credito per eventuali pubblicità nella rete di Microsoft Bing.

SEOZoom vanta di un database di miliardi di keywords e garantisce una panoramica completa del proprio settore di riferimento, permettendo a coloro che lo utilizzano di effettuare una ricerca delle parole chiave migliori per il proprio sito, controllare il posizionamento giornaliero, osservare i competitor e le loro strategie, analizzare il proprio profilo backlink, analizzare gli investimenti PPC dei competitor e molto altro. Il servizio che offre non è gratuito, dispone di diversi piani che cambiano in base al prezzo e alle funzionalità offerte.

Ahrefs Keywords Explorer, piattaforma che fornisce all'utente un ampio rapporto sulle parole chiave, si basa sui dati clickstream, ovvero su una tecnica di Web Analytics che consiste nell'analizzare e monitorare i click degli utenti eseguiti all'interno di una pagina internet per rilevare le principali attività dei visitatori sul sito. Mostra il numero stimato di clic per le parole chiave scelte ed elabora dati relativi a differenti motori di ricerca, fornisce inoltre dati per 171 paesi diversi.

Google keyword planner, il pianificatore di parole chiave fornito da Google, altro strumento che l'azienda fornisce per la ricerca di parole chiave, indica con quale frequenza vengono cercate determinate parole su Google, come sono cambiate le loro ricerche nel tempo, fornisce termini correlati, stime del volume di ricerca, stime dei costi degli annunci per qualsiasi parola chiave o URL inserite. Offre diversi modi per effettuare la ricerca, in base a parole o frasi, a siti Web o pagine di destinazione ed in base a categorie. L'inserimento di una parola chiave o di una frase nella barra di ricerca restituisce un elenco di idee per le parole chiave correlate a quella inserita, il volume della query di ricerca per la parola chiave in base alle località, alle lingue ed alle proprietà di Google definite e permette di visualizzare la competitività relativa alla parola chiave nella ricerca a pagamento. Nel momento in cui viene utilizzato il Google keyword planner è possibile scegliere tra una serie di impostazioni per meglio definire la ricerca. Ad esempio, si può scegliere tra l'opzione di inclusione di certi termini, per la quale il pianificatore di parole chiave mostrerà solo i suggerimenti che includono i

termini scelti, oppure l'opzione di esclusione, che consente di aggiungere una parola chiave da escludere per qualsiasi termine di ricerca che non riguarda la propria attività. Vi è inoltre la possibilità di impostare il paese cui fare riferimento, la lingua predefinita da utilizzare e se mostrare, come da impostazione predefinita, i volumi di ricerca degli ultimi 12 mesi o specificare un intervallo di mesi personalizzato e confrontare i volumi per due diversi intervalli di tempo. Specificare se si desidera che le stime di traffico vengano restituite solo per le reti di ricerca Google o includano anche i partner di ricerca di Google, ovvero quei siti web che collaborano con Google al fine di estendere lo spazio pubblicitario disponibile. Google keyword planner, permette inoltre di ottenere previsioni di traffico su differenti parole chiave, restituendo il numero previsto di clic, visualizzazioni e costo giornalieri e la posizione media raggiungibile dall'annuncio, inoltre si ottengono stime più precise se si selezionano il budget o i costi per dei clic specifici. In base ai clic stimati al giorno, si può avere un'idea relativa di quali delle parole chiave a disposizione, sono le più popolari e possono potenzialmente generare più traffico.

La principale differenza tra Google Keyword Planner e Google Trends è il modo in cui calcolano e visualizzano la popolarità di un determinato termine di ricerca. Google Trends fornisce solo statistiche relative che tengono conto del numero totale di ricerche su Google effettuate in un determinato periodo di tempo, Google Keyword Planner, mostra il volume di ricerca totale stimato. I due strumenti funzionano dunque in modo diverso, se l'obiettivo delle proprie ricerche è quello di voler sapere quanto traffico sta ottenendo una parola chiave per valutarne il posizionamento raggiunto, allora può risultare più soddisfacente l'uso di Google Keyword Planner. Se si vuole capire invece la popolarità di un certo tema, allora Google Trends aiuta maggiormente a capire se l'interesse nel pubblico sta aumentando, diminuendo o mantenendosi stabile. Questo perché Google Trends non si concentra su quante centinaia, migliaia o milioni di volte viene cercata una parola chiave in un determinato intervallo di date, basa i suoi grafici sulla percentuale di ricerche fatte su Google in quel periodo, composte da quella specifica parola chiave. Entrambi possono risultare estremamente importanti ed utili, l'unico aspetto di cui si può tenere conto per decidere quale dei due usare, è capire se l'obiettivo delle proprie ricerche riguarda lo studio di un nuovo argomento, categoria o nicchia da trattare per Google Trends oppure Keyword Planner quando arriva il momento di trovare parole chiave specifiche in modo da poter ottenere dati sul volume di ricerca effettivo.

Gli strumenti visti, sono solo alcuni di quelli esistenti, le alternative possibili sono molte, cambiano i prezzi e le funzionalità offerte, ma in ogni caso il loro utilizzo aiuta e migliora notevolmente la fase di pianificazione delle parole chiave. Una volta ottenuti i dati dalla ricerca di parole chiave, si devono analizzare per capire quale tra le varie keywords permettono di ottenere risultati migliori, sia in termini di punteggio di qualità che di costi. Si procede dunque con una categorizzazione delle stesse in base ai risultati ottenuti ed una scelta finale delle parole da utilizzare.

Gli esperti di SEM possono quindi agire tramite le varie possibili strategie viste in questo capitolo e nel precedente, lo scopo per tutti è chiaramente quello di comprendere quali siano le migliori per il tipo di mercato che si trovano ad affrontare, tenendo presente dei molteplici aspetti che possono influenzare il risultato finale. La scelta delle parole chiave più attinenti a quanto viene pubblicato nel web è, come si è più volte specificato, uno step di fondamentale importanza poiché permette di attirare maggiore traffico verso il proprio annuncio o sito web e quindi di conseguenza raggiungere facilmente gli obiettivi prefissati. Su tale concetto si sviluppa il progetto di questa tesi di laurea, il quale viene raccontato nel dettaglio nei prossimi due capitoli.

Il titolo di questo capitolo, *L'importanza delle keywords*, vuole sottolineare la principale caratteristica che appartiene loro, ovvero il fatto di costituire il fulcro di uno dei passaggi più delicati nell'ambito della Search Engine Marketing, ma anche il rappresentare il concetto alla base del progetto descritto.

4. Calcolo delle performance di ogni componente di tutte le parole chiave

In questo capitolo viene esposta la prima parte del progetto svolto, tramite la quale è stato osservato il dataset relativo alle performance ottenute dalle varie parole chiave utilizzate per poter valutare i risultati ottenuti dai singoli termini che le costituiscono. Il concetto basilare che viene applicato in questa prima fase di lavoro è quello del marketing di attribuzione, processo usato tipicamente nel mondo del web marketing per valutare, a partire da una comunicazione che sfrutta l'uso di più canali, i risultati che ognuno di questi ha ottenuto, permettendo di capire quale dei mezzi adoperati è risultato il più efficace. Tale processo è stato in questo caso specifico, applicato ai termini che formano insieme ad altri le parole chiave utilizzate per le campagne pubblicitarie online del cliente dopo essere state sottoposte all'applicazione dell'algoritmo di stemming.

4.1 Dataset utilizzato

Il dataset sul quale si sviluppa il lavoro è stato recuperato dall'account Google Ads di uno dei clienti dell'agenzia presso la quale ho svolto il tirocinio. Il cliente in questione è il titolare di un e-commerce di prodotti tipici toscani da poco online.

Il periodo di tempo cui fanno riferimento i dati, parte dal 23 settembre 2020 ed arriva al 30 giugno 2021 e rappresenta la panoramica dei risultati ottenuti in questo lasso di tempo dalle parole chiave selezionate per gli annunci pubblicitari a pagamento. Accedendo all'account Google Ads, nella sezione dedicata alle parole chiave si ha la possibilità di osservare l'andamento delle performance relative a tutte le keywords che sono state scelte per la promozione tramite annunci pubblicitari.

La pagina si apre mostrando un primo grafico, nel quale vengono indicati il numero di clic ottenuti quotidianamente, nel periodo di tempo osservato, come illustrato in Figura 7.

Parole chiave per la rete di ricerca



Figura 7. Rappresentazione grafica dei clic giornalieri ottenuti nel lasso di tempo selezionato.

Al di sotto di questa rappresentazione grafica troviamo la tabella di dati, nella quale sono inserite tutte le parole chiave e le rispettive caratteristiche. La visualizzazione di tutte le parole può essere personalizzata tramite la possibilità di tre opzioni: possono essere visualizzate tutte le keywords, tutte quelle attivate o tutte tranne quelle rimosse. L'opzione scelta per questo lavoro è stata quella di scaricare tutte le parole poiché l'interesse è orientato verso le performance ottenute da ognuna di esse e la cosa importante è che le parole analizzate siano state attivate almeno una volta per campagne pubblicitarie.

La tabella dati cui si è fatto riferimento, è formata da 490 parole e 15 colonne, quest'ultime riportate in ordine sono: Parola chiave, Tipo di corrispondenza, Campagna, Gruppo di annunci, Stato, CPC max, URL finale, Impressioni, Interazioni, Tasso di interazione, Costo medio, Costo, Conversioni, Costo / Conversioni, Tasso di conversione. Volendo dare una breve spiegazione per ognuna di queste, la prima colonna "Parola chiave", come suggerisce il nome riporta direttamente la keyword scelta per uno degli annunci; a seguire "Tipo di corrispondenza", riporta il tipo di corrispondenza scelta per la parola chiave cui fa riferimento, che come già visto nel capitolo 3 (v. 3.3), possono essere generiche, a frase o esatte.

La colonna "Campagna", riporta il nome della campagna per cui la parola chiave è stata utilizzata ed ogni campagna è costituita da uno o più gruppi di annunci, i quali sono riportati nella colonna denominata proprio "Gruppo di annunci". Questa contiene il nome del gruppo a cui appartiene la parola, ogni gruppo contiene uno o più annunci che condividono obiettivi simili ed è importante tale suddivisione poiché permette di

organizzare i vari annunci attorno a un tema comune. Nella colonna "Stato" sono riportate le informazioni sullo stato attuale di ogni campagna presente nell'account nel momento in cui si accede ad esso e come risolvere eventuali problemi non previsti. I tipi di valore che tale colonna può assumere sono diversi, si distinguono tra "idonea", "non idonea", "rimossa", "in pausa", ecc., già spiegati nel capitolo 3 relativo alle keywords (v. 3.3). In "CPC max", viene indicato il costo per clic massimo, ovvero l'importo massimo che si è disposti a pagare per un clic sull'annuncio; la colonna "URL finale", riporta l'URL raggiunto dagli utenti dopo aver fatto clic sull'annuncio.

Le colonne rimanenti sono quelle in cui si visualizzano effettivamente tutti i risultati legati alle singole parole chiave e sono quelle su cui ci si è maggiormente focalizzati per lo svolgimento di tale lavoro. I valori riportati come "Impressioni", rappresentano il numero di volte in cui l'annuncio viene pubblicato sulle reti pubblicitarie di Google, come Google.com, YouTube o altri siti web ed applicazioni, permettono di stabilire la frequenza di visualizzazione dell'annuncio ed in questo caso riportano nello specifico il numero di visualizzazioni dell'annuncio in cui compare la parola chiave osservata. I numeri presenti in "Interazioni" rappresentano invece i clic che l'annuncio in cui è presente la parola ha ottenuto: monitorare le interazioni consente di stabilire se gli annunci stiano raggiungendo gli obiettivi di rendimento. "Tasso di interazione", è sempre collegato ai clic ottenuti, ma indica la frequenza con la quale gli utenti interagiscono con un annuncio dopo averlo visualizzato. Tale valore aiuta a stabilire l'efficacia di un annuncio, è rappresentato solitamente in percentuale ed è maggiormente conosciuto nel mondo del Web advertising come CTR³⁰, acronimo che sta per Click-through rate. Corrisponde alla percentuale di persone che hanno effettivamente cliccato sull'annuncio dopo averlo visto.

Il CTR in percentuale si calcola con la formula seguente:

$$CTR = \frac{Clic}{Impressioni} * 100$$

³⁰ Wikipedia, voce *Click Trough rate* https://it.wikipedia.org/wiki/Click-through_rate (visitato il 21 settembre 2021)

La colonna “Costo medio” corrisponde al CPC³¹ effettivo, acronimo che sta per Costo per Click o nella versione inglese PPC, Pay per click. Come suggerisce il nome, tale metrica rappresenta l'importo medio pagato per la singola interazione, ovvero il costo generato da ogni click effettuato da un utente su un'inserzione a pagamento.

Il calcolo di tale metrica si ottiene dividendo il budget speso in advertising con il numero di click effettuati dagli utenti su uno specifico annuncio, quindi con la seguente formula:

$$CPC = \frac{Costo}{Clic}$$

La colonna denominata “Costo” rappresenta la somma dei costi totali nel periodo di riferimento, “Conversioni”, invece mostra il numero di conversioni ricevute dopo le interazioni con gli annunci, possono rappresentare ad esempio i clic sugli annunci di testo, le visualizzazioni degli annunci video, o altro. A partire dal numero di conversioni si ottengono le altre due metriche che Google Ads fornisce, ovvero “Costo/Conversioni” che corrisponde al Costo per Acquisizione, CPA³², e “Tasso di conversione”, denominato nel mondo del web marketing come Conversion Rate, CVR. “Costo / Conversioni”, indica il costo medio di una conversione, ovvero il costo che l'inserzionista sostiene per ogni conversione da parte degli utenti. Attraverso il CPA si misura dunque il costo unitario necessario per ottenere una determinata azione da parte del cliente che può consistere nell'acquisizione di un click, di una vendita o altro. Si ottiene dividendo il costo totale con il numero di conversioni generate dalla campagna, in formula:

$$CPA = \frac{Costo}{Conversioni}$$

Nell'ultima colonna sono presenti invece i valori del “Tasso di conversione” o Conversion Rate³³, il quale indica la frequenza media con cui un'interazione con un annuncio genera una conversione, la percentuale di visitatori unici che hanno effettuato

³¹ Wikipedia, voce *Pay per click* <https://en.wikipedia.org/wiki/Pay-per-click> (visitato il 21 settembre 2021)

³² Wikipedia, voce *Cost per action* https://it.wikipedia.org/wiki/Cost_per_action (visitato il 21 settembre 2021)

³³ Wikipedia, voce *Conversion Rate* https://it.wikipedia.org/wiki/Conversion_rate (visitato il 21 settembre 2021)

quella specifica azione che l'inserzionista ha definito come obiettivo della campagna. Si calcola dividendo il numero di persone che compiono la determinata azione per il numero totale di clic ottenuti.

La formula per calcolare il CVR in percentuale:

$$CVR = \frac{Conversioni}{Clic} * 100$$

In Figura 8, è rappresentata una parte della tabella visualizzabile tramite l'account di Google Ads del cliente, con i valori corrispondenti per ogni parola chiave dei quali sono mostrati solo quelli di maggior interesse per il lavoro effettuato. Le parti cancellate corrispondono ai costi effettuati dal cliente esaminato, il quale ha richiesto che vengano omessi. Nelle tabelle che seguono, se presenti i dati reali relativi ai costi, si procede allo stesso modo.

<input type="checkbox"/> <input checked="" type="radio"/> Parola chiave	Impr.	Interazioni	↓ Tasso di interazione	Costo medio	Costo	Conversioni	Costo/conv.	Tasso conv.
Totale: parole chiave	365.716	17.623 clic	4,82%	€	€	189,00	€	1,07%
<input type="checkbox"/> <input checked="" type="radio"/> box prodotti toscani	97	14 clic	14,43%	€	€	0,00	€	0,00%
<input type="checkbox"/> <input checked="" type="radio"/> shop online prodotti toscani	952	118 clic	12,39%	€	€	1,00	€	0,85%
<input type="checkbox"/> <input checked="" type="radio"/> prodotti toscani online	4.289	522 clic	12,17%	€	€	4,00	€	0,77%
<input type="checkbox"/> <input checked="" type="radio"/> sapori tipici toscani	140	17 clic	12,14%	€	€	0,00	€	0,00%
<input type="checkbox"/> <input checked="" type="radio"/> prodotti tipici toscani da regalare	1.064	128 clic	12,03%	€	€	3,00	€	2,34%
<input type="checkbox"/> <input checked="" type="radio"/> vendita online prodotti tipici toscani	817	95 clic	11,63%	€	€	4,00	€	4,21%
<input type="checkbox"/> <input checked="" type="radio"/> prodotti tipici toscani online	1.948	226 clic	11,60%	€	€	4,00	€	1,77%
<input type="checkbox"/> <input checked="" type="radio"/> "e-commerce prodotti tipici"	339	37 clic	10,91%	€	€	0,00	€	0,00%

Figura 8. Tabella delle parole chiave visualizzabile sull'account Google Ads.

4.2 Pulizia dei dati

Il dataset è stato scaricato in formato .csv ed è stato pulito tramite OpenRefine³⁴, tool che permette di lavorare con i dati, di pulirli, o trasformarli da un formato ad un altro. Durante la fase di data cleaning sono state effettuate trasformazioni quali il passaggio da testo a valori numerici dove necessario, la correzione di alcuni errori di battitura presenti nell'elenco delle parole chiave, la modifica del nome di alcune

³⁴ OpenRefine <https://openrefine.org/>

colonne per evitare spazi bianchi tra i termini che ne compongono il nome ed infine l'eliminazione di quelle colonne che non risultano utili ai fini di tale progetto.

Le colonne eliminate sono “Tipo di corrispondenza”, “Campagna”, “Gruppo di annunci”, “Stato”, “CPC max”, “URL finale”. “Tipo di corrispondenza” è stata eliminata perché la maggior parte delle parole aveva una corrispondenza generica, solo poche di tipo a frase e quindi risultava difficile poter improntare un possibile studio basandosi su questa caratteristica. Stessa scelta è stata presa per “Campagna”, perché anche in questo caso facevano tutte riferimento ad un'unica campagna.

Le colonne rinominate sono invece Interazioni, Tasso di interazione, Costo medio, Costo / Conversioni, Tasso di conversione, per le quali si è preferito usare il nome maggiormente utilizzato nell'ambito del web marketing, in ordine sono state dunque modificate in Click, CTR, CPC, CPA, CVR. A queste poi è stata aggiunta un'ulteriore metrica alla quale si fa spesso riferimento nella fase di valutazione delle performance di un annuncio a pagamento: il CPM. L'acronimo CPM³⁵ sta per Cost per mille ed è una metrica utilizzata per stimare il costo di una campagna pubblicitaria rispetto a mille visualizzazioni della stessa, quanto costa quindi mostrare una certa pubblicità per mille visualizzazioni. Questo indicatore si determina moltiplicando per mille il rapporto fra costo della campagna e numero di contatti raggiunti. Di seguito la formula:

$$CPM = \frac{Costo}{Impressioni} * 1000$$

4.3 Processo di attribuzione

Una volta eseguite le varie fasi di pulizia, si ottiene il dataset finale sul quale sono stati svolti i vari studi. Le parole chiave a disposizione sono 490 per le quali si conoscono le rispettive performance in termini di Impressioni, Conversioni generate, Clic ottenuti, Costo, CVR, CTR, CPC, CPA, CPM. Lo scopo iniziale consiste nell'andare ad individuare i risultati ottenuti da ciascun termine che costituisce ognuna delle parole chiave, questo studio sulle parole è stato ottenuto mettendo in pratica il processo di Marketing di attribuzione.

³⁵ Wikipedia, voce *Cost per mille* https://it.wikipedia.org/wiki/Cost_per_mille (visitato il 21 settembre 2021)

4.3.1 Marketing di attribuzione

Il termine Marketing di attribuzione indica tutti quei passaggi che sfruttano l'uso dei modelli di attribuzione, i quali come riporta la guida di Google Analytics, corrispondono alla regola o l'insieme di regole che determina il modo in cui il credito per vendite e conversioni viene assegnato ai punti di contatto nei percorsi di conversione³⁶.

Solitamente questo tipo di processi viene utilizzato da aziende che seguono strategie di marketing multicanale, sfruttando dunque l'uso di più mezzi per le loro attività di marketing, come per esempio Facebook, Google Ads, campagne e-mail, e mirano ad andare a verificare i risultati ottenuti da ciascuno strumento per valutare in base ad una serie di obiettivi raggiunti, come per esempio la vendita di un prodotto, quali dei vari canali utilizzati detiene il maggior merito per poi eventualmente comprendere come modificare la propria strategia per renderla il più efficace possibile. Se un utente vede un annuncio su Google, clicca e converte, il modello di attribuzione risulta piuttosto semplice, se però prima di effettuare la conversione, l'utente segue un percorso differente che passa da varie visite attraverso canali diversi e in giorni lontani tra loro, allora la questione diviene più complessa da valutare. I modelli di attribuzione aiutano ad assegnare un valore a questi touchpoint, attribuendo il merito della conversione o ad uno dei canali utilizzati o distribuendolo in modo differente. Esistono vari tipi di modelli di attribuzione, i quali si differenziano in base alla strategia che seguono nella fase di assegnazione del merito ad un canale piuttosto che ad un altro.

I modelli di attribuzione esistenti sono:

- Ultima interazione;
- Ultimo click non diretto;
- Ultimo click Google Ads;
- Prima interazione;
- Decadimento temporale;
- In base alla posizione.

³⁶ Guida di Analytics, voce *Panoramica dei modelli di attribuzione*
<https://support.google.com/analytics/answer/1662518?hl=it> (visitato il 4 ottobre 2021)

Per comprendere al meglio le varie tipologie, viene proposto di seguito un esempio preso dalla Guida di Google Analytics: un cliente trova il sito osservato facendo clic su uno degli annunci Google Ads; una settimana dopo torna sul sito facendo clic da un social network e nello stesso giorno torna una terza volta attraverso una delle campagne e-mail e, dopo qualche ora, torna di nuovo, raggiungendo il sito in maniera diretta per effettuare un acquisto e dunque generare una conversione.

Nel modello di attribuzione "Ultima interazione", l'ultimo punto di contatto riceve il 100% del credito della vendita, in questo caso il canale Diretto. Nel modello "Ultimo clic non diretto", tutto il traffico diretto viene ignorato e il 100% del merito viene attribuito all'ultimo canale su cui il cliente ha fatto clic prima di una conversione, nel caso riportato in esempio, il canale E-mail. "Ultimo clic Google Ads", è un modello secondo il quale il primo e unico clic sul canale Ricerca a pagamento riceve il 100% del credito della vendita. Per il modello di attribuzione "Prima interazione", il 100% del merito della conversione spetta al primo punto di contatto, in questo caso il canale Ricerca a pagamento. Nel modello di attribuzione "Lineare" ciascun punto di contatto nel percorso di conversione riceve la stessa percentuale di credito per la vendita, quindi essendo Ricerca a pagamento, Social Network, E-mail e Diretto i canali coinvolti, spetta il 25% ciascuno. "Decadimento temporale", prevede che i punti di contatto più vicini nel tempo alla vendita o alla conversione ottengano la maggior parte del credito. In questa particolare vendita, i canali Diretto ed E-mail riceverebbero la maggior parte del credito, in quanto il cliente ha interagito con essi poche ore prima della conversione. Il canale Social Network riceverebbe meno credito rispetto ai canali Direct o E-mail e dato che l'interazione con Ricerca a pagamento è avvenuta una settimana prima, questo canale riceverebbe un credito notevolmente inferiore. "In base alla posizione", il 40% del credito viene assegnato alla prima e all'ultima interazione, mentre il restante 20% del credito viene distribuito uniformemente tra le interazioni centrali. In questo esempio, i canali Ricerca a pagamento e Diretto riceverebbero ciascuno il 40% del credito, mentre i canali Social Network ed E-mail riceverebbero il 10% del credito.

4.3.2 Attribuzione applicata al seguente progetto

Il concetto di attribuzione cui si è appena fatto riferimento, è un tipo di processo utilizzato nel mondo del web marketing e partendo da questo principio, si è deciso di sfruttare l'idea alla base ed applicarla all'insieme di parole chiave appartenenti al

dataset scaricato. Chi si occupa di elaborare strategie SEM per creare annunci efficaci, che riescano ad ottenere dei punteggi di qualità elevata ed attirare l'attenzione di molti utenti, deve cercare di capire quali siano le parole o le frasi che più attirano gli utenti e che nello stesso tempo generino un miglior punteggio di qualità, per questo è importante porre l'attenzione sui singoli termini che compongono le keywords e riuscire a comprendere quali tra i tanti risultano essere i più vincenti.

Per questa ragione, come è stato già anticipato, la fase iniziale ha come obiettivo quello di andare a valutare i punteggi per ciascun componente di ogni keyword rispetto alle performance ottenute, ovvero capire basandosi sui risultati ottenuti da ciascuna parola chiave, chi tra i vari elementi che compongono la keyword ha ottenuto il risultato migliore. Di seguito si riportano le fasi che si sono susseguite per l'ottenimento di tale scopo.

La prima azione messa in pratica consiste nell'aver escluso dal set di dati tutte quelle parole chiave che non hanno ottenuto nessuna impressione in quanto questo implica che non siano mai state visualizzate su annunci Google. Rispetto al tipo di analisi che si vuole svolgere dunque, si considerano queste come combinazioni tra parole mai avvenute. Dopo questa rimozione, da 490 parole chiave ne restano a disposizione 277, nella figura esposta sotto è possibile vedere una parte del dataset dopo tutte le varie modifiche apportate.

	Parola_chiave	Gruppo_di_annunci	Impressioni	Click	Costo	Conversioni	CVR	CTR	CPA	CPC	CPM
1	aleatico	Dolci_liquori	5	2		0	0.00	40.00			
6	biscotti cantucci	Dolci_liquori	270	8		2	25.00	2.96			
7	pan puccini	Dolci_liquori	26	1		0	0.00	3.85			
8	confettura di fichi	Miele_confetture	409	6		0	0.00	1.47			
9	uva e noci	Miele_confetture	9	0		0	0.00	0.00			
...			
478	ragù di cinghiale	Pasta_sughi	1026	27		0	0.00	2.63			
481	prodotti della toscana	food_box	2974	196	1	0	0.00	6.59			
485	sapori tipici toscani	food_box	140	17		0	0.00	12.14			
486	cesti pasquali	food_box	4324	256		3	1.17	5.92			
487	sugo ai funghi porcini	Pasta_sughi	1148	9		0	0.00	0.78			

277 rows x 11 columns

Figura 9. Parte del dataset sui cui si è svolto lo studio.

Per ridurre ulteriormente la complessità delle parole che compongono il dataset, si è proceduto con l'eliminazione dalle varie keywords delle stopwords, ovvero tutte quelle parole che non hanno un significato particolare in quanto sono molto frequenti in qualsiasi tipo di linguaggio e dunque non si ritengono interessanti. Per l'italiano, ad esempio, con stopwords si considerano le congiunzioni, preposizioni, articoli, ecc.

Questo ulteriore grado di pulizia dei dati è stato messo in pratica facendo riferimento a NLTK³⁷ (=Natural Language Toolkit), un insieme di librerie scritte in linguaggio Python, specifiche per l'elaborazione del linguaggio naturale (in inglese: Natural Language processing, acronimo: NLP), termine che si riferisce ad un campo dell'Intelligenza artificiale che consiste in tutti quei processi che permettono ai computer di "comprendere" il linguaggio naturale, cioè il linguaggio che utilizzano comunemente gli esseri umani quando comunicano tra loro, per poterlo analizzare e studiare. NLTK fornisce una lista di parole predefinita ed identificata con il termine stopwords, che corrispondono ad un elenco di parole molto comuni, ma che non forniscono informazioni utili per la maggior parte delle procedure di analisi del testo, come ad esempio preposizioni, articoli, congiunzioni o altri caratteri specifici che possono essere aggiunti alla lista già esistente. Data la loro poca rilevanza spesso si eliminano dai testi da analizzare e proprio per questo motivo, anche nel caso di tale studio, ci si è mossi in questa direzione, facendo riferimento nello specifico all'elenco di stopwords per l'italiano con l'aggiunta di altri pochi caratteri come il simbolo &, virgolette alte ed apici, gli ultimi due erano presenti nel dataset in quanto appartenenti a parole chiave che riportavano come tipo di corrispondenza, quella a frase (v. 3.3).

Il dataset ha subito tutti i vari step di data cleaning, si procede quindi a questo punto con la fase di tokenizzazione. Tutte le parole chiave presenti nel dataset vengono sottoposte ad una tokenizzazione poiché alla base di tutto il lavoro vi è la volontà di effettuare uno studio sulle radici delle singole parole, che combinate insieme vanno a formare le differenti parole chiave.

La divisione in token è stata effettuata tramite l'uso del pacchetto `nltk.tokenizer`³⁸ fornito dalla libreria NLTK, il quale procede con la separazione in token di un testo, ovvero con la separazione di tutte le singole frasi o le singole parole di cui è formato. In base a come viene effettuata la suddivisione si hanno a

³⁷ Natural Language Toolkit <https://www.nltk.org/>

³⁸ NLTK Tokenizer Package, `nltk.tokenize` package <https://www.nltk.org/api/nltk.tokenize.html>

disposizione diversi tipi di tokenizer: quello per frasi, che usa le frasi come token, dato quindi un testo composto da 5 frasi verrà suddiviso in 5 token ciascuno contenente una delle frasi del testo iniziale. Altrimenti esiste il tokenizer per parole, il quale usa le parole come token ed in questo caso se un testo è composto da 253 parole, verrà suddiviso in 253 token. NLTK consente inoltre di definire i tokenizer usando le cosiddette regular expressions, o regex, sequenza di simboli che identifica l'insieme di stringhe, ovvero di caratteri da considerare come token, permettendo dunque di personalizzare il tipo di tokenizer. Nel caso descritto nella seguente relazione si effettua una tokenizzazione su tutte le sequenze di vocaboli che formano una parola chiave, si utilizza la funzione `word_tokenize()` del pacchetto NLTK e si ottengono quindi tanti token quanti sono gli elementi che formano ogni keyword, evitando come specificato sopra tutte le parole presenti per la lingua italiana nell'elenco stopwords.

4.4 Stemming

La tokenizzazione è stata fatta, si procede quindi con lo stemming.

Lo stemming³⁹ è un processo di Text Normalization, che significa ridurre il più possibile le variazioni presenti all'interno di un testo sostituendo le parole che lo compongono con la loro forma canonica. Lo scopo dello stemming è proprio quello: ricondurre quindi una parola dalla sua forma flessa alla radice o tema, permettendo a differenti forme di vocaboli simili di convergere verso un unico tema comune. Il tema non corrisponde necessariamente alla radice morfologica, infatti in alcuni vocaboli la desinenza si unisce direttamente alla radice, ma in altri fra radice e desinenza si inserisce una lettera o un gruppo di lettere che insieme alla radice formano il tema.

Il pacchetto NLTK offre tre diversi algoritmi per il calcolo dello stemming: Snowball⁴⁰, Porter e Lancaster. In generale tutti e tre agiscono allo stesso modo, vanno a rimuovere le desinenze dalle parole conducendole verso una forma comune. Ciò che li differenzia è la qualità dei risultati ed il fatto che Porter e Lancaster, a differenza dello Snowball, sono specifici per la lingua inglese, mentre l'altro funziona anche con altre lingue, motivo questo per cui la scelta dell'algoritmo per tale progetto, si sia affidata immediatamente allo stemmer Snowball. L'algoritmo di stemming Porter è il più vecchio dei tre perché nato negli anni '80; lo stemmer Snowball, è un algoritmo

³⁹ Stemming <https://www.nltk.org/api/nltk.stem.html>

⁴⁰ Algoritmo Snowball <https://www.nltk.org/api/nltk.stem.snowball.html>

conosciuto anche come Porter2, poiché considerato migliore del precedente: è stato infatti costruito basandosi sui problemi riscontrati con lo stemmer Porter. Infine, lo stemmer Lancaster, tende talvolta a ridurre le parole in maniera eccessiva, trasformandole in radici strane, ma nel complesso risulta efficiente quanto gli altri.

Come anticipato si è utilizzato lo stemmer Snowball, questo è una classe e non una funzione quindi come prima cosa deve essere creata una specifica istanza sulla base della classe: questo passaggio lo si mette in pratica chiamando la classe e passando tra parentesi i parametri che si desiderano per la particolare istanza della classe. Nel progetto si procede con l'importazione dell'algoritmo da NLTK, successivamente è stata creata l'istanza ed impostata come parametro la lingua italiana ed infine si è applicato lo stemmer ad ogni parola come riportato di seguito:

```
from nltk.stem import SnowballStemmer
st = SnowballStemmer("italian")
st.stem(parola)
```

Lo stemming è una procedura utilizzata spesso durante la fase di preparazione di un testo ad analisi successive, come ad esempio la classificazione, dove l'intento è quello di capire quale sia l'argomento di cui tratta un certo testo. Dunque, in casi come questo, se nello stesso testo sono presenti le parole “correre”, “corro”, “corriamo” e “correremo”, è importante che il computer capisca che non vengono affrontati quattro argomenti diversi, ma uno solo, quello relativo alla corsa, che è infatti associabile al tema “corr”. Un altro esempio si può riportare con le forme singolari, plurali, maschili o femminili di una stessa parola, “bambino”, “bambina”, “bambini”, “bambine”, verranno collegate tutte al tema “bambin”.

La scelta di utilizzare uno stemmer per tutti i token che formano le keywords studiate, è stata dettata dal fatto che in molti casi capita di incontrare il nome di uno stesso prodotto declinato nelle varie forme flesse, in questo specifico dataset si riscontrano principalmente casi di versioni plurali e singolari della stessa parola. Concentrandosi sul fatto che lo scopo principale del lavoro è quello di capire quali siano i termini che ottengono fra tutti i risultati migliori, riferendosi in modo particolare ai significati delle parole stesse, si è assunto di considerare per ogni

keyword solo l'unità semantica e dunque si è proceduto con un algoritmo di stemming su tutte le componenti di ogni parola chiave.

Lo stemming sui token del dataset esaminato permette quindi di poter valutare complessivamente le performance ottenute dalle differenti parole senza andare a distinguere tra le varie forme che la stessa parola può assumere, soffermandosi in questo modo solo sul significato alla base del termine stesso.

Riportando qualche esempio preso proprio dal dataset scaricato da Google Ads, si osservano tra le varie le seguenti parole chiave, con il rispettivo numero di impressioni⁴¹:

“cesto regalo san valentino”, 27

“cesti gastronomici regalo”, 307

“ceste regalo aziendali”, 22

“cesti gastronomici”, 13.332

“cesti di natale”, 7.171

I token “cesto”, “cesti”, “ceste”, se non fosse stato applicato lo stemming, sarebbero stati considerati come parole distinte e di conseguenza anche le loro performance, ma applicando tale processo si riconducono tutte e tre al tema “cest” e questo permette di andare a sommare tutti i risultati ottenuti dalle singole nelle varie combinazioni in cui sono state inserite, ed ottenere dunque un unico valore finale che garantisce l'attribuzione del risultato relativo a quel particolare termine. Si effettua la tokenizzazione e si isola il termine cesto dalle altre componenti, si applica lo stemming e si ottiene la radice “cest”. Ad ogni “cest” corrisponde il numero di impressioni ottenute dalla keyword in cui è inserito e volendo calcolare il numero di impressioni ottenute solo dal termine “cest”, si sommano tutti i valori ottenuti per ogni tipologia di metrica considerata. Riprendendo l'esempio, otteniamo per la radice “cest”, 20.859 impressioni, valore che indica quante impressioni sono state ottenute in totale da “cest”.

È necessario tenere presente che risalire al tema di una parola partendo dalla sua forma flessa non è sempre semplice: per quanto si possano definire delle regole

⁴¹ Nell'esempio si riporta solo questa metrica per renderlo più semplice ai fini della spiegazione, nel caso reale però si sono considerate tutte le metriche elencate nel paragrafo 4.3 di questo capitolo.

precise, il linguaggio naturale è ricco di eccezioni. Per questo motivo, si verificano anche casi come il seguente, illustrato a partire da alcune keywords appartenenti al dataset studiato:

“liquori toscani”

“birra artigianale toscana”

“vino toscano”

“salumi toscani”

“vendita specialità toscane”

Con le parole chiave appena elencate, il risultato che lo stemmer restituisce per i termini “toscani”, “toscana”, “toscani” corrisponde al tema “toscan”, mentre per “toscano” a “tosca”, lasciando quest’ultimo separato dagli altri. Nel momento in cui si sommano le performance ottenute dal singolo termine, questi due restano distinti l’uno dall’altro, non rispettano quindi l’idea alla base del progetto.

Questo risultato sottolinea uno dei problemi dello stemming: esistono due tipologie di errore negli algoritmi di stemming: Over-stemming e Under-stemming. Nell’esempio sopra si verifica il fenomeno dell’Under-stemming, ovvero quando due parole dovrebbero essere ricondotte allo stesso tema perché effettivamente derivano dal medesimo, ma invece vengono ricondotte a due differenti, “toscan” e “tosca”. I termini sono più specifici ma si rischia di perdere qualche relazione tra le parole che hanno lo stesso significato.

L’Over-stemming, al contrario, è un errore che si verifica quando due parole vengono ricondotte alla stessa radice, ma in realtà derivano da due temi completamente differenti. Succede che vengono individuati più legami tra le parole con lo stesso significato, rischiando però di includere anche quelle con significato diverso. Con le parole appartenenti al dataset non si verifica mai il caso di Over-stemming, ma volendo per chiarezza riportare un esempio inventato, se lo stemmer con le parole “università” e “universale” restituisce per entrambe lo stesso risultato ovvero “univers”, si verifica un caso di Over-stemming poiché i due termini non appartengono alla stessa famiglia di significati, ma vengono trattati come tali.

L’algoritmo Snowball, applicato al dataset esaminato, rientra nel fenomeno dell’Under-stemming con pochi vocaboli, tra quelli rintracciati vi sono:

“italiano”, “italiana”, “italiani”, “italiane”, restituendo per “italiano”, “ital”, mentre per le altre tre “italian”;

“sughi”, “sugo”, per i quali restituisce rispettivamente “sugh” e “sug”;

“uva”, “uve”, che vengono considerate due parole differenti e vengono restituite dallo stemmer così come sono;

“toscano”, “toscana”, “toscani”, “toscani”, già citati sopra.

In questo caso essendo un dataset non troppo ampio è facile rendersi conto di eventuali errori, ma se si lavora con una quantità di parole molto più elevata, risulta complicato controllarle tutte. Le parole che rientrano in questo problema tipico dello stemming sono poche, si è proceduto dunque con una correzione manuale, ma quando il numero diviene più alto diventa complicato procedere in tale maniera.

Un’alternativa allo stemming potrebbe essere l’uso della lemmatizzazione, la quale comunque non è detto garantisca risultati migliori perché è un tipo di procedimento che lavora bene quando va a lemmatizzare le parole che compongono delle frasi di un testo, quindi parole inserite in un contesto, ma dato che nel caso in esame si tratta di gruppi di parole scollegate l’una dall’altra è probabile che il risultato sia simile o comunque poco differente da quello dello stemming.

4.5 Lemmatizzazione

La lemmatizzazione è il processo di riduzione di una forma flessa di una parola alla sua forma canonica detta lemma. Nell’elaborazione del linguaggio naturale, la lemmatizzazione è il processo algoritmico che determina automaticamente il lemma di una data parola⁴².

Il processo di lemmatizzazione è molto simile a quello di stemming, in quanto entrambi hanno come scopo quello di ricondurre diverse forme flesse di una parola allo stesso tema. I due si differenziano per il fatto che la lemmatizzazione cerca di ricondurre la parola ad un termine di senso compiuto, il lemma: restituirà infatti la forma della parola reperibile in un qualsiasi dizionario, quindi un termine che deve essere necessariamente valido. Lo stemming invece si limita ad eliminare la parte finale della parola, restituendo dunque la parte iniziale che non necessariamente corrisponderà alla sua radice morfologica. In genere, è sufficiente che le parole

⁴² Wikipedia, voce *Lemmatizzazione* <https://it.wikipedia.org/wiki/Lemmatizzazione> (visitato il 20 settembre 2021)

correlate siano associate alla stessa sequenza iniziale di lettere, quella parte che in inglese viene definita con il termine stem e da cui deriva appunto il nome stemming, anche se non sempre corrisponde ad una radice valida.

Questa differenza tra i due processi di normalizzazione del testo avviene poiché uno stemmer opera su una singola parola senza avere conoscenza del contesto in cui è inserita, non riesce quindi a distinguere quelle parole che assumono significati diversi in base alla parte del discorso in cui si trovano. La Lemmatizzazione invece fa in genere riferimento ad un'analisi delle parole, rimuovendo solo le terminazioni flessive e restituendo la corrispondente forma base. Molte parole possono corrispondere a differenti lemmi in base al contesto in cui vengono utilizzate, per questa ragione la lemmatizzazione garantisce risultati migliori se prima di procedere viene effettuato il Part of Speech tagging o PoS tagging, ovvero il processo di etichettatura di una parola in un testo, detto corpus, con lo scopo di identificare la parte del discorso a cui corrisponde, in base sia alla sua definizione che al contesto. Partendo dunque da un corpus, ad ogni token si associa l'etichetta che rappresenta il suo ruolo all'interno del testo. Tra le possibili categorie, quelle comuni alla maggior parte delle lingue sono: sostantivo, verbo, articolo, aggettivo, preposizione, pronome, avverbio, congiunzione; a queste si aggiungono altre sottocategorie in base al tipo di linguaggio utilizzato.

La lemmatizzazione in Python, si può implementare grazie a diversi pacchetti offerti dalle varie librerie esistenti. Nltk, offre ad esempio *WordNet Lemmatizer*⁴³, modulo che effettua la lemmatizzazione facendo riferimento al *WordNet*⁴⁴, un ampio database lessicale della lingua inglese che mira a stabilire relazioni semantiche tra le parole. Per procedere è necessario creare un'istanza di `WordNetLemmatizer()` e chiamare la funzione `lemmatize()` su ogni singola parola. Tra i parametri della funzione, oltre al termine da esaminare è possibile fornire il POS tag corrispondente per ottenere un miglior risultato. Non sempre risulta possibile inserire manualmente il relativo POS tag per ogni token soprattutto quando si lavora con testi di grandi dimensioni, per questo motivo nltk offre il metodo `nltk.pos_tag()`⁴⁵ che restituisce in modo automatico il tag.

⁴³ NLTK, WordNet Lemmatizer https://www.nltk.org/_modules/nltk/stem/wordnet.html

⁴⁴ WordNet <https://wordnet.princeton.edu/>

⁴⁵ NLTK, nltk.tag package https://www.nltk.org/api/nltk.tag.html?highlight=nltk%20pos_tag

Volendo effettuare una lemmatizzazione delle parole presenti nel dataset di tale progetto, non si potrebbe però utilizzare il tipo di funzione appena spiegata, in quanto è specifica per la lingua inglese. È possibile però utilizzare la libreria SpaCy⁴⁶, la quale si basa su modelli predefiniti in grado di analizzare il testo e calcolare varie funzionalità relative al NLP, tra le quali fornisce anche la possibilità di ottenere il lemma di una parola. Il POS tag in questo caso viene assegnato di default e le lingue per cui è stata scritta sono molte, quella di cui si è interessati va specificata in fase di download del pacchetto durante la quale è richiesto di specificare il modello corrispondente alla lingua⁴⁷.

4.6 Stemming o Lemmatizzazione?

La scelta di usare l'uno o l'altro processo, è guidata dal tipo di studio che ci si trova ad affrontare, se l'applicazione di tali algoritmi dipende molto dalla comprensione corretta del significato delle parole nel contesto in cui sono inserite, allora la migliore opzione sarà quella della lemmatizzazione, se però risulta necessario che le parole correlate vengano ricondotte allo stesso tema anche se questo non si identifica poi nella radice valida, allora si può optare per lo stemming, considerando anche che quest'ultimo è più facile e veloce da implementare e quindi può essere più ottimale in termini di tempo.

Il progetto si basa su una serie di parole chiave che possono essere formate da più unità, ma è molto difficile che corrispondano a delle vere e proprie frasi, piuttosto ad un insieme di termini che richiamano la descrizione del prodotto. La scelta dello stemming, quindi, è stata dettata proprio da tale caratteristica del dataset, non trattandosi di frasi inserite in un testo, per le quali sarebbe risultato importante comprendere il contesto, è stato possibile procedere tramite l'uso dello stemming ottenendo comunque degli ottimi risultati. La fase di normalizzazione del testo in questo caso era essenzialmente guidata dal voler evitare troppe forme di una stessa parola e garantire un raggruppamento delle varie possibili versioni.

Il concetto basilare di questo lavoro è comprendere se una determinata parola chiave possa ottenere risultati migliori se arricchita con l'aggiunta di un termine che

⁴⁶ SpaCy <https://spacy.io/>

⁴⁷ Altre librerie esistenti, sono elencate al seguente link
<https://www.machinelearningplus.com/nlp/lemmatization-examples-python/>

ne descrive una qualità in più. Per fare un esempio, se la keyword osservata fosse *prosciutto crudo*, l'idea sarebbe quella di comprendere se l'aggiunta della provenienza del prodotto possa garantire risultati migliori. Non è quindi importante sapere se la parola da aggiungere sia *toscano*, *toscani* o *della Toscana*, ad esempio, ma semplicemente se risulti utile in termine di performance ottenute, aggiungere un riferimento relativo alla provenienza del prodotto sotto forma di un qualsiasi vocabolo che, come parte iniziale, abbia la radice *tosc*.

4.7 Calcolo dei risultati

Una volta ottenuta la radice di tutti i termini presenti nelle varie keywords, si procede affinché si riescano ad ottenere i risultati raggiunti da ciascuno.

L'obiettivo è proprio quello di capire quali siano le performance dei singoli vocaboli che combinati in vario modo vanno a comporre le parole chiave e quindi partendo da una lista formata da ogni singola unità di ciascuna parola chiave, sia essa formata da un solo termine che da molti, si ottiene un elenco di tutte le radici di ogni vocabolo con i rispettivi valori derivanti dalla parola chiave a cui appartengono.

La tabella riporta un piccolo esempio per la comprensione di quanto svolto, facendo riferimento ad una piccola parte dei dati studiati, dove i valori di Parola_chiave, Impressioni, Click, Conversioni, CVR e CTR corrispondono a quelli reali, mentre gli altri sono stati modificati come richiesto dal proprietario. La stessa scelta è stata presa per tutte le altre tabelle che si presentano a seguire ed in generale in tutti i casi in cui si sono riportati dei valori per le metriche elencate.

Parola_chiave	Impressioni	Click	Costo	Conversioni	CVR	CTR	CPA	CPC	CPM
biroldo	603	40	9,50 €	0	0%	6,63%	0 €	0,24 €	15,75 €
biscotti cantucci	270	8	3,12 €	2	25%	2,96%	2 €	0,39 €	11,56 €
confettura di fichi	409	6	1,70 €	0	0%	1,47%	0 €	0,28 €	4,16 €
miele di castagno	1048	15	3,86 €	0	0%	1,43%	0 €	0,26 €	3,68 €
miele e tartufo	16	1	0,30 €	0	0%	6,25%	0 €	0,30 €	18,75 €
vendita miele	5442	163	65,48 €	0	0%	3%	0 €	0,40 €	12,03 €
dolci toscani	828	38	4,35 €	0	0%	4,59%	0 €	0,11 €	5,25 €
birra artigianale toscana	89	7	2,26 €	0	0%	7,87%	0 €	0,32 €	25,39 €

Tabella 1. Keywords estratte dal dataset esaminato.

Prendendo come esempio le keywords rappresentate in Tabella 1, dopo la tokenizzazione e l'applicazione dello stemming per ogni singolo componente si ottiene

la seguente lista di parole, di cui i valori numerici corrispondono allo stesso ordine presente in tabella ovvero Impressioni, Click, Costo, Conversioni, CVR, CTR, CPA, CPC, CPM:

['birold', 603, 40, 9.50, 0, 0, 6.63, 0, 0.24, 15.75]
['biscott', 270, 8, 3.12, 2, 25.0, 2.96, 2, 0.39, 11.56],
['cantucc', 270, 8, 3.12, 2, 25.0, 2.96, 2, 0.39, 11.56],
['confettur', 409, 6, 1.70, 0, 0, 1.47, 0, 0.28, 4.16],
['fich', 409, 6, 1.70, 0, 0, 1.47, 0, 0.28, 4.16],
['miel', 1048, 15, 3.86, 0, 0, 1.43, 0, 0.26, 3.68],
['castagn', 1048, 15, 3.86, 0, 0, 1.43, 0, 0.26, 3.68],
['miel', 16, 1, 0.30, 0, 0, 6.25, 0, 0.30, 18.75],
['tartuf', 16, 1, 0.30, 0, 0, 6.25, 0, 0.30, 18.75],
['vend', 5442, 163, 65.48, 0, 0, 3.0, 0, 0.40, 12.03],
['miel', 5442, 163, 65.48, 0, 0, 3.0, 0, 0.40, 12.03],
['dolc', 828, 38, 4.35, 0, 0, 4.59, 0, 0.11, 5.25],
['tosc', 828, 38, 4.35, 0, 0, 4.59, 0, 0.11, 5.25],
['birr', 89, 7, 2.26, 0, 0, 7.87, 0, 0.32, 25.39],
['artigianal', 89, 7, 2.26, 0, 0, 7.87, 0, 0.32, 25.39],
['tosc', 89, 7, 2.26, 0, 0, 7.87, 0, 0.32, 25.39]

Come si può notare ogni componente viene riportato con la radice a cui lo stemmer lo ha ridotto e con tutti i risultati ottenuti dalle parole chiave cui fanno parte. Lo scopo ancora non è stato raggiunto poiché alcuni stem si ripetono, quindi svolto questo passaggio si vanno a prendere tutte le radici identiche che derivano dall'uso di uno stesso termine in più keywords e si sommano i rispettivi valori di Conversioni, Impressioni, Click e Costo, mentre le altre metriche CVR, CTR, CPA, CPC e CPM, dato che si ottengono a partire da divisioni tra le precedenti, vengono ricalcolate sulla base dei nuovi valori così da ottenere i risultati finali di ciascun vocabolo che insieme ad altri costituisce una parola chiave.

Di seguito l'elenco in ordine alfabetico dei risultati di ogni singolo stem, dove per quelli che hanno partecipato solo all'interno di un'unica keyword, i valori numerici restano identici, mentre per coloro che sono stati utilizzati in diverse combinazioni vengono sommati. I valori riportati sotto rappresentano quindi il risultato della singola

radice rispetto a tutte le keyword utilizzate, sempre rispetto all'ordine dei valori numerici di Impressioni, Click, Costo, Conversioni, CVR, CTR, CPA, CPC, CPM:

['artigianal', 89, 7, 2.26, 0, 0, 7.87, 0, 0.32, 25.39],
['birold', 603, 40, 9.50, 0, 0, 6.63, 0, 0.24, 15.75],
['birr', 89, 7, 2.26, 0, 0, 7.87, 0, 0.32, 25.39],
['biscott', 270, 8, 3.12, 2, 25.0, 2.96, 2, 0.39, 11.56],
['cantucc', 270, 8, 3.12, 2, 25.0, 2.96, 2, 0.39, 11.56],
['castagn', 1048, 15, 3.86, 0, 0, 1.43, 0, 0.26, 3.68],
['confettur', 409, 6, 1.70, 0, 0, 1.47, 0, 0.28, 4.16],
['dolc', 828, 38, 4.35, 0, 0, 4.59, 0, 0.11, 5.25],
['fich', 409, 6, 1.70, 0, 0, 1.47, 0, 0.28, 4.16],
['miel', 6506, 179, 69.64, 0, 0, 2.75, 0, 0.39, 10.70],
['tartuf', 16, 1, 0.30, 0, 0, 6.25, 0, 0.30, 18.75],
['tosc', 917, 45, 6.61, 0, 0, 4.90, 0, 0.14, 7.20],
['vend', 5442, 163, 65.48, 0, 0, 3.0, 0, 0.40, 12.03].

La stessa procedura si attua per valutare i risultati ottenuti dalla combinazione di due o tre parole che compongono le varie keywords. L'obiettivo in questo caso è quello di valutare i risultati per ciascuna coppia o tripla che hanno, insieme ad altri vocaboli, contribuito alla costruzione di una parola chiave. Quindi si sono prese solo le keyword per cui il numero di componenti sia maggiore di uno o due termini e come per il caso precedente si dividono tali keyword in singoli token sui quali si effettua poi lo stemming.

Una volta ottenuta la radice di ogni componente di tutte le parole chiave, si utilizza una delle funzioni di itertools⁴⁸, ovvero combinations()⁴⁹. La funzione combinations() prende come parametri la sequenza di parole per cui vogliamo ottenere tutte le combinazioni ed un valore numerico che indica il numero di termini di cui vogliamo siano formate le sotto sequenze generate. Combinations() restituisce infatti tutte le sotto sequenze derivabili dalla parola data come input, le fornisce considerando l'ordine originale e senza ripetizioni.

⁴⁸ itertools <https://docs.python.org/3/library/itertools.html>

⁴⁹ Combinations <https://docs.python.org/3/library/itertools.html#itertools.combinations>

Si consideri la parola chiave “*confezioni regalo prodotti alimentari*”, presa dal dataset. La funzione combinations(), se ha come parametro numerico il valore 2, restituirà le seguenti coppie di radici:

(*confezion, regal*),
 (*confezion, prodott*),
 (*confezion, alimentar*),
 (*regal, prodott*),
 (*regal, alimentar*),
 (*prodott, alimentar*).

Se invece il valore impostato è uguale a 3, il risultato sarà in questo caso:

(*confezion, regal, prodott*),
 (*confezion, regal, alimentar*),
 (*confezion, prodott, alimentar*),
 (*regal, prodott, alimentar*).

Ogni sotto sequenza è associata alle performance della sequenza iniziale, questo permette di valutare il risultato di ogni coppia o tripla di parole nell’intero insieme di keywords. Una volta calcolate le varie combinazioni, anche in questo caso si cercano eventuali duplicati e si sommano le rispettive performance.

Di seguito si riporta la rappresentazione di tale passaggio a partire da poche keywords riprese dal dataset.

Parola chiave	Impressioni	Click	Costo	Conversioni	CVR	CTR	CPA	CPC	CPM
confezioni regalo prodotti alimentari	752	48	23,62 €	4	8,33%	6,38%	5,91 €	0,49 €	31,41 €
biscotti cantucci	270	8	1,23 €	2	25%	2,96%	0,62 €	0,15 €	4,56 €
prodotti tipici toscani	13748	1380	460,68 €	17	1,23%	10,04%	27,10 €	0,33 €	33,51 €
prodotti tipici toscani da regalare	1064	128	32,10 €	3	2,34%	12,03%	10,70 €	0,25 €	30,17 €

Tabella 2. Keywords estratte dal dataset esaminato.

Considerando le quattro parole chiave in Tabella 2, si ottiene inizialmente la lista delle radici combinate a coppie e triple della stessa keyword con le rispettive performance.

Lista di coppie:

[‘confezion regal’, 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],

[‘confezion prodott’, 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],

['alimentar confezion', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['prodott regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar prodott', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['biscott cantucc', 270, 8, 1.23, 2, 25.0, 2.96, 0.62, 0.15, 4.56],
['prodott tipic', 13748, 1380, 460.68, 17, 1.23, 10.04, 27.10, 0.33, 33.51],
['prodott tosc', 13748, 1380, 460.68, 17, 1.23, 10.04, 27.10, 0.33, 33.51],
['tipic tosc', 13748, 1380, 460.68, 17, 1.23, 10.04, 27.10, 0.33, 33.51],
['prodott tipic', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],
['prodott tosc', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],
['prodott regal', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],
['tipic tosc', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17]

Si procede verificando eventuali coppie che si ripetono per ottenere infine i risultati finali ottenuti da ciascuna. La lista finale prevede le seguenti coppie in ordine alfabetico:

['alimentar confezion', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar prodott', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['biscott cantucc', 270, 8, 1.23, 2, 25.0, 2.96, 0.62, 0.15, 4.56],
['confezion prodott', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['confezion regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['prodott regal', 1816, 176, 55.72, 7, 3.98, 9.69, 7.96, 0.31, 30.68],
['prodott tipic', 14812, 1508, 492.78, 20, 1.33, 10.18, 24.64, 0.32, 33.27],
['prodott tosc', 14812, 1508, 492.78, 20, 1.33, 10.18, 24.64, 0.32, 33.27],
['tipic tosc', 14812, 1508, 492.78, 20, 1.33, 10.18, 24.64, 0.32, 33.27].

Lista di triple:

['confezion prodott regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar confezion regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar confezion prodott', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar prodott regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['prodott tipic tosc', 13748, 1380, 460.68, 17, 1.23, 10.04, 27.10, 0.33, 33.51],
['prodott tipic tosc', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],

['prodott regal tipic', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],
['prodott regal tosc', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],
['regal tipic tosc', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],

La keyword “biscotti cantucci” non è presente in lista perché per ottenere l'elenco di tutte le triple vengono considerate solo le parole con più di due termini al loro interno.

Si procede poi verificando eventuali triple che si ripetono per ottenere infine i risultati finali ottenuti da ciascuna. La lista finale prevede le seguenti triple in ordine alfabetico:

['alimentar confezion prodott', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar confezion regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['alimentar prodott regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['confezion prodott regal', 752, 48, 23.62, 4, 8.33, 6.38, 5.91, 0.49, 31.41],
['prodott tipic tosc', 14812, 1508, 492.78, 20, 1.33, 10.18, 24.64, 0.32, 33.27],
['prodott regal tipic', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],
['prodott regal tosc', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17],
['regal tipic tosc', 1064, 128, 32.10, 3, 2.34, 12.03, 10.70, 0.25, 30.17].

Ogni coppia ed ogni tripla presente nell'elenco è stata ordinata alfabeticamente anche al suo interno, questo perché si evita che due o tre radici identiche che formano due coppie o triple distinte in quanto provenienti da diverse parole chiave, non vengano poi considerate uguali solo perché il loro ordine originale è opposto. Se nel dataset ho due parole chiave del tipo:

“*confezioni natalizie alimentari*” e “*confezioni alimentari natalizie*”, dopo la generazione di ogni coppia si avranno per la prima:

(*confezion, nataliz*)

(*confezion, alimentar*)

(*nataliz, alimentar*)

per la seconda:

(confezion, alimentar)

(confezion, nataliz)

(alimentar, nataliz)

Senza un ordinamento interno per ciascuna coppia o tripla, succedrebbe che in casi come questo, *(nataliz, alimentar)* e *(alimentar, nataliz)* verrebbero considerati come due coppie differenti. Ma dato che lo scopo del lavoro è quello di studiare i risultati ottenuti dalle radici di ogni componente della keyword per valutare quanto la loro presenza risulti fondamentale all'interno della parola chiave, risulterebbe errato considerare loro distintamente poiché non è interessante in questo specifico caso il loro ordine.

Nella parte finale, quando viene generato il nuovo dataset e ricalcolate le metriche di CVR, CTR, CPA, CPC, CPM, alcuni risultati corrispondono a valori del tipo *inf* e *NaN*. Il primo dei due si ottiene quando la divisione ha un valore 0 come divisore e quindi il risultato è impossibile, mentre l'altro quando sia dividendo che divisore corrispondono a 0 e dunque la divisione risulta indeterminata. Questi valori sono stati trasformati in 0 poiché ai fini di tale lavoro non è importante che venga specificato il fatto che la divisione sia risultata indeterminata o impossibile, ma semplicemente che la parola chiave corrispondente non abbia ottenuto risultati; evitando così problemi di calcolo soprattutto nella seconda parte di tale progetto, visto che il dataset utilizzato è proprio quello generato durante la prima fase. Tale risultato, infatti, indica che la specifica parola chiave in base alla metrica cui viene fatto riferimento, non ha generato o nessun clic o nessuna conversione, dato che ognuna delle formule viste (v. 4.1 e 4.2) ha come denominatore o una di queste due metriche o il valore corrispondente alle impressioni, per le quali però sono già state eliminate in fase iniziale quelle con valore uguale a 0.

Una volta che tutte queste procedure sono state effettuate, tutti gli elenchi di singole radici, coppie e triple con rispettive performance vengono inseriti in un nuovo dataset raffigurato nell'immagine seguente, il quale in definitiva contiene 748 record di cui 211 corrispondono alle singole radici di ogni componente di tutte le parole chiave, 368 le coppie e 169 le triple. Prima della gestione dei duplicati il numero di elementi era 1393, di cui 657 singole radici, 536 coppie e 200 triple.

	Radice_parola	Impressioni	Click	Costo	Conversioni	CVR	CTR	CPA	CPC	CPM
0	acac	1785.0	70.0		0.0	0.00	3.92			
1	acet	341.0	6.0		0.0	0.00	1.76			
2	affumic	2014.0	67.0		0.0	0.00	3.33			
3	aleat	5.0	2.0		0.0	0.00	40.00			
4	alimentar	3656.0	198.0		6.0	3.03	5.42			
...			
743	regal tipic tosc	1064.0	128.0		3.0	2.34	12.03			
744	ros tosc vin	23.0	1.0		0.0	0.00	4.35			
745	ross tosc vin	371.0	22.0		0.0	0.00	5.93			
746	sapor tipic tosc	140.0	17.0		0.0	0.00	12.14			
747	tipic tosc vend	817.0	95.0		4.0	4.21	11.63			

748 rows × 10 columns

Figura 10. Dataset finale.

5. Il valore di Shapley

Nel seguente capitolo viene spiegato in che cosa consiste il valore di Shapley, come si calcola, ma soprattutto in che modo è stato applicato. Partendo dalle stesse parole chiave reperite tramite l'account Google Ads del proprietario dell'e-commerce di prodotti tipici toscani, Quintovizio, ci si è posti il problema di poter offrire a coloro che si occupano dello studio di possibili nuove keywords per le future campagne, uno strumento che permettesse loro di poter verificare non solo il risultato ottenuto dalle parole chiave utilizzate, ma anche quanto le singole componenti di ciascuna abbiano contribuito per l'ottenimento di tale performance, per poter comprendere quindi quanto la presenza di un certo termine all'interno di una keyword possa migliorare i risultati finali.

5.1 Cos'è il valore di Shapley

Il valore di Shapley⁵⁰ è un concetto di soluzione della teoria dei giochi cooperativi. La teoria dei giochi, nella versione inglese Game Theory, si può descrivere come lo studio matematico del processo decisionale che porta i giocatori analizzati a prendere certe decisioni, mira a comprendere come i decisori interagiscono tra di loro considerando che per ogni giocatore il risultato ottenuto dipende dalle azioni di tutti i coinvolti. Quando il giocatore sceglie la propria strategia deve comunque tenere di conto anche delle scelte degli altri, che a loro volta stanno ragionando sulle proprie strategie considerando i possibili pensieri di ogni altro partecipante.

La teoria dei giochi cooperativi si basa sullo studio dei giochi di coalizione o cooperativi, giochi in cui la competizione avviene tra gruppi di giocatori e si concentra sulla previsione di quali coalizioni si formeranno, le azioni congiunte che i gruppi intraprendono e i risultati collettivi raggiunti. Si oppone alla tradizionale teoria dei giochi non cooperativi, la quale si concentra sulla previsione delle azioni e dei guadagni dei singoli giocatori⁵¹.

⁵⁰ Wikipedia, voce *Valore di Shapley* https://it.wikipedia.org/wiki/Valore_di_Shapley (visitato il 7 ottobre 2021)

⁵¹ Wikipedia, voce *Game Theory* https://en.wikipedia.org/wiki/Game_theory (visitato il 7 ottobre 2021)

Il termine concetto di soluzione consiste invece nel definire una regola formale per prevedere come verrà giocato un gioco. Queste previsioni sono chiamate soluzioni, descrivono le strategie che verranno adottate dai giocatori ed il risultato del gioco.

Il valore di Shapley prende il nome da Lloyd Stowell Shapley, matematico ed economista americano che lo ha teorizzato. Il concetto alla base è il seguente: una coalizione di giocatori coopera e ottiene un certo guadagno complessivo da tale cooperazione. Dal momento che alcuni giocatori possono contribuire alla coalizione più di altri, lo scopo del valore di Shapley è quello di comprendere quanto è importante ogni giocatore per la cooperazione complessiva e quale ricompensa può aspettarsi in base al contributo dato.

Nel capitolo precedente è stato affrontato il tema dell'attribuzione che viene ripreso in quanto il calcolo del valore di Shapley è proprio un tipo di modello di attribuzione data-driven, guidato cioè dai dati. I modelli di attribuzione classici vengono molto utilizzati e svolgono un ruolo importante nell'ambito del web marketing, ma hanno anche alcuni aspetti che li limitano, in particolare il fatto che si basano su regole che devono essere scelte a priori, l'utente deve quindi decidere in anticipo come dividere il credito per le conversioni di vendita tra i canali. Si è visto infatti che possono essere scelti modelli che accreditano la quota del payoff nello stesso modo tra tutti i punti di contatto, oppure può essere accreditata una percentuale decrescente di payoff in base a quanto è più lontano nel tempo un punto di contatto dalla data di conversione e molti altri modi.

I vari responsabili dell'attività di web marketing all'interno di un'azienda utilizzano i risultati ottenuti tramite il modello di attribuzione scelto per poter valutare il ROI⁵², acronimo di Return on investment o ritorno sull'investimento, indice che indica il tasso di rendimento sul totale degli investimenti di un'impresa. Questo indice permette loro di poter prendere decisioni su dove investire in futuro a partire da informazioni relative ai ritorni ottenuti dagli investimenti fatti. I modelli di attribuzione permettono dunque di compiere questo tipo di studi sull'andamento delle scelte prese, ma esistono anche approcci basati sui dati che permettono di superare i loro limiti. L'attribuzione basata sui dati è una soluzione che permette di studiare la complessità

⁵² Enciclopedia Treccani, voce *ROI (Return On Investment)*
[https://www.treccani.it/enciclopedia/roi_\(Dizionario-di-Economia-e-Finanza\)](https://www.treccani.it/enciclopedia/roi_(Dizionario-di-Economia-e-Finanza)) (visitato il 7 ottobre 2021)

dei percorsi degli acquirenti osservando il modo in cui i canali ed in particolar modo le loro diverse combinazioni interagiscono con i possibili clienti per influenzare il risultato di vendita desiderato. Un modello basato sui dati fornisce la visualizzazione più accurata di quali siano i canali che stanno ottenendo i risultati migliori.

Il valore di Shapley consiste proprio in un modello di attribuzione basato sui dati che permette di trovare un modo per dividere equamente la vincita tra più giocatori, a partire da una partita in cui i giocatori stessi possono formare tra loro delle coalizioni, collaborando per aumentare la probabilità di un risultato desiderato. Può essere definito dunque come una misura del contributo marginale medio di un giocatore in ciascuna coalizione, basandosi sul fatto che i giocatori possono unirsi alle coalizioni in diversi momenti e quindi coprire all'interno di esse *ordini* differenti ed avere vari gradi di influenza su di esse assumendo differenti *valori*. Si basa sul presupposto che ogni ordine ha la stessa probabilità di verificarsi e quindi i giocatori vengono premiati per il loro contributo dato a tutte le permutazioni.

Nell'ambito del marketing i giocatori corrispondono ai canali della campagna, mentre le coalizioni, corrispondono ai vari modi in cui i canali interagiscono con gli account durante il percorso dell'acquirente. La teoria dei giochi cooperativi e più nello specifico il valore di Shapley, forniscono un modo concreto per misurare l'influenza del canale e dividere equamente il credito per le conversioni di vendita tra i canali in base al loro contributo individuale al payoff totale.

Il valore di Shapley è una soluzione ampiamente utilizzata, anche Google Analytics la sfrutta per l'attribuzione del canale. Tra i benefici che ne derivano, si possono elencare la capacità di fornire informazioni più approfondite sulle prestazioni di un canale, un'equa ripartizione del credito basata sul contributo ed una capacità di ottimizzare gli investimenti di marketing e influenzare i risultati di vendita. Si riporta di seguito un esempio per capire meglio il concetto del valore di Shapley. Si prendono in considerazione i percorsi di alcuni utenti che durante il loro tragitto verso la conversione, vengono raggiunti da uno o più canali. Questo significa che i canali hanno lavorato insieme formando coalizioni per aumentare la probabilità di una possibile conversione. I tre canali in questione sono Facebook, E-mail, Instagram, i quali vengono utilizzati per promuovere delle campagne pubblicitarie e spingere l'utente ad effettuare un acquisto, l'obiettivo è calcolare il guadagno ottenuto dai tre canali in base al contributo che hanno dato alla coalizione per poter verificare in che modo le

ricompense andrebbero divise e di conseguenza capire chi dei tre risulta effettivamente il più efficace.

I dati a disposizione sono i seguenti e rappresentano come i tre canali raggiungono i vari utenti presi in esame, il risultato ottenuto quando agiscono individualmente oppure quello ottenuto quando collaborano con altri partecipanti. Il valore che corrisponde a ciascuna coalizione riporta il rispettivo Conversion Rate, ovvero la percentuale di utenti che effettivamente ha effettuato la conversione.

Coalizioni	CVR
Facebook	0,1
E-mail	0,3
Instagram	0,15
Facebook + E-mail	0,2
Facebook + Instagram	0,2
E-mail + Instagram	0,4
Facebook + E-mail + Instagram	0,9

Tabella 3. Dataset di esempio per spiegare il concetto alla base di Shapley.

L'obiettivo del calcolo è comprendere a quanto corrisponde il reale risultato di ciascun canale, tenendo in considerazione il contributo che ha apportato alla coalizione in cui è comparso.

Per prima cosa si procede calcolando il valore di ciascuna coalizione tramite la funzione caratteristica. Considerando S una coalizione di giocatori, la funzione caratteristica assegna un valore $v(S)$ ad ogni coalizione S che rappresenta il profitto che può generare quando i suoi giocatori lavorano insieme. Esistono varie opzioni per definire la funzione caratteristica, si possono infatti considerare le entrate totali generate da ciascuna coalizione, il numero totale di conversioni di vendita generate da ciascuna coalizione, il Conversion rate di ciascuna coalizione oppure la probabilità condizionata di conversione, ovvero la probabilità di conversione dati una serie di canali. Nel seguente progetto il valore è stato calcolato in due versioni, sia considerando la somma del CVR di ciascun partecipante in una coalizione che considerando il rispettivo CTR, per questo motivo procediamo nell'esempio considerando il CVR. In questo caso, facendo riferimento direttamente al relativo CVR

di ciascuna coalizione il valore di Facebook + E-mail corrisponde al CVR che i due giocatori hanno ottenuto durante la loro collaborazione:

$$v(\text{Facebook, E-mail}) = \text{CVR}(\text{Facebook} + \text{E-mail}) = 0,20$$

Applicato poi lo stesso procedimento anche con le altre coalizioni avvenute, si ottengono i valori per ciascuna come riportato in tabella 4.

Coalizioni (S)	v(S)
Facebook	0,1
E-mail	0,3
Instagram	0,15
Facebook + E-mail	0,2
Facebook + Instagram	0,2
E-mail + Instagram	0,4
Facebook + E-mail + Instagram	0,9

Tabella 4. Rappresentazione delle coalizioni d'esempio ed i rispettivi valori.

Ottenuto il valore di ogni coalizione, si procede considerando tutte le possibili permutazioni dei tre partecipanti in sequenza, calcolando dunque l'incremento che ognuno di loro comporta: lo scopo è quello di valutare il contributo marginale di ciascun partecipante alla coalizione.

Iniziando con la sequenza Facebook + E-mail + Instagram, in questo caso il primo canale a partecipare alla coalizione è Facebook che quando agisce da solo assume un valore pari a 0,10. Il secondo canale a raggiungere la coalizione è E-mail, si procede osservando il valore che ottengono i due canali Facebook ed E-mail quando collaborano ovvero 0,20. Il fatto che E-mail e Facebook insieme raggiungono un valore pari a 0,20 e Facebook da solo 0,10, implica che l'arrivo di E-mail ha portato ad un incremento pari a 0,10 poiché $v(\text{E-mail} + \text{Facebook}) - v(\text{Facebook}) = 0,10$. Tale valore corrisponde al contributo marginale di E-mail in questa particolare sequenza dei tre canali in gioco. Terzo elemento della sequenza è Instagram, il valore dei tre giocatori insieme è pari a 0,90 mentre il valore dei due che lo precedono, $v(\text{Facebook} + \text{Email})$ corrisponde a 0,20, per questo il contributo marginale di Instagram in questa

specifica sequenza corrisponde a $v(\text{Facebook} + \text{Email} + \text{Instagram}) - v(\text{Facebook} + \text{Email})$, ovvero 0,70.

Lo stesso calcolo viene ripetuto per ogni possibile sequenza in cui i tre giocatori possono essere combinati, ottenendo i rispettivi contributi marginali in base alla posizione che ciascun canale ricopre, in base dunque all'ordine con cui hanno raggiunto la coalizione come riportato di seguito.

(Facebook, E-mail, Instagram) \rightarrow (0.10, 0.10, 0.70)

(Facebook, Instagram, E-mail) \rightarrow (0.10, 0.10, 0.70)

(E-mail, Facebook, Instagram) \rightarrow (0.3, -0.10, 0.70)

(E-mail, Instagram, Facebook) \rightarrow (0.3, 0.10, 0.50)

(Instagram, Facebook, E-mail) \rightarrow (0.15, 0.05, 0.70)

(Instagram, E-mail, Facebook) \rightarrow (0.15, 0.25, 0.50)

A questo punto si procede con il calcolo del valore di Shapley per ogni canale, che corrisponde alla media dei contributi marginali di ciascuno. Il valore di Shapley di Facebook è $(0.10 + 0.10 + (-0.10) + 0.50 + 0.05 + 0.50) / 6 = 0.19$; per E-mail è $(0.10 + 0.70 + 0.30 + 0.30 + 0.70 + 0.25) / 6 = 0.39$; per Instagram risulta essere $(0.70 + 0.10 + 0.70 + 0.10 + 0.15 + 0.15) / 6 = 0.32$.

La somma dei tre risultati corrisponde a 0.90, lo stesso valore che assumono i canali nel momento in cui collaborano tutti e tre insieme. Questo accade poiché il concetto alla base del calcolo del valore di Shapley consiste nel dividere in modo opportuno il guadagno della grande coalizione, ovvero quella formata da tutti i membri, per ogni giocatore coinvolto, rispettando il contributo che ciascuno ha dato. Per ciascuno si ottiene dunque il rispettivo guadagno che ognuno dei tre otterrebbe se partecipassero insieme alla stessa coalizione.

5.2 La formula

Grazie all'esempio riportato sopra è possibile comprendere che due concetti fondamentali di questo calcolo sono proprio l'ordine dei partecipanti e tutte le possibili permutazioni. L'ordine è importante perché in base a questo cambia il tipo di contributo che il giocatore apporta, si è visto sopra che quando il canale entra nella sequenza per primo assume un certo valore, mentre se si unisce per secondo o terzo ne

assume un altro. Le permutazioni vanno tenute in considerazione perché permettono di valutare Shapley per ciascun giocatore, assumendo che tutte le sequenze possibili si possano verificare. Quindi, anche se nel dataset di partenza ne è presente solo una, il valore di Shapley è calcolato considerando tutte le sequenze possibili e quindi anche quelle che effettivamente non si sono verificate ma che possono farlo.

La formula su cui si basa tale calcolo e che tiene dunque conto di tutti questi elementi è la seguente:

$$\phi(i, v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S))$$

N corrisponde all'insieme di giocatori ed $|N|$ è la cardinalità di N , il numero dei giocatori.

S è la coalizione tra giocatori appartenenti a N , si forma quando un sottoinsieme di giocatori collabora ed $|S|$ è la cardinalità di S , il numero dei giocatori che la costituiscono.

$v(S)$ è il valore assegnato a ciascuna coalizione S dalla funzione caratteristica.

La sommatoria è su tutti i sottoinsiemi S di N che non contengono il giocatore i , ovvero il giocatore per cui si sta calcolando il valore di Shapley.

La formula consiste in una media pesata del guadagno che il giocatore i fornisce nel momento in cui si aggiunge alle coalizioni in cui non è presente. La media pesata, si usa quando ogni numero ha una determinata importanza, detta peso, che influisce sul calcolo finale. Per calcolare la media ponderata o pesata tra più valori numerici si procede con la somma dei prodotti di ogni numero per il rispettivo peso e la divisione del risultato ottenuto per la somma dei pesi. Nel caso del calcolo del valore di Shapley, si procede calcolando la media rispetto a tutte le permutazioni possibili, poiché il contributo del giocatore dipende dall'ordine in cui esso ha partecipato alla coalizione.

$v(S \cup \{i\}) - v(S)$ rappresenta il contributo marginale del giocatore i se si unisce alla coalizione dopo che tutti i giocatori appartenenti ad S si sono aggiunti. Tramite tale differenza si ottiene il contributo marginale di i , ovvero l'aumento di guadagno

ricollegabile all'aggiunta di una nuova unità, lasciando tutti gli altri invariati. Se $v(S \cup \{i\})$ ha un certo valore, quindi permette di ottenere un certo risultato e $v(S)$ da solo un altro, la differenza tra questi due valori permette di calcolare il contributo marginale di i , ovvero quanto la comparsa di i ha aggiunto alla collaborazione tra le componenti di S .

La prima parte della formula invece permette di calcolare il peso di ciascuna coalizione, tenendo conto delle permutazioni possibili tramite l'uso del fattoriale. Il fattoriale⁵³ di un numero n corrisponde al prodotto dei numeri interi positivi, minori o uguali ad n e rappresenta il numero delle possibili permutazioni di un insieme contenente n elementi non ripetuti. Se ho tre elementi, questi possono essere combinati tra loro in sei possibili sequenze. Il numeratore $|S|!(|N|-|S|-1)!$ analizzato nel dettaglio, è formato da $|S|!$ che rappresenta il numero di ordini in cui i membri $|S|$ di S possono aggregarsi alla coalizione prima del giocatore i . Prendendo di nuovo l'esempio fatto sopra, se $S = \text{Facebook, Instagram}$ ed $i = \text{E-mail}$, con $|S|!$ posso considerare tutte le possibili permutazioni tra i due elementi di S prima che si unisca a loro E-mail. In questo caso $|S| = 2$, il fattoriale $2! = 2$ e per questo i possibili ordini che i giocatori Facebook ed Instagram possono ricoprire sono esattamente due:

Facebook, Instagram

Instagram, Facebook.

$(|N|-|S|-1)!$ rappresenta invece gli $|N|-|S|-1$ membri di $N \setminus (S \cup \{i\})$, ovvero coloro che appartengono ad N ma che non fanno parte della coalizione in cui è presente anche i e che possono aggregarsi alla coalizione dopo il giocatore i . Se $S = \text{Facebook}$, $i = \text{E-mail}$ e $N = \text{Facebook, Instagram, E-mail}$, significa che i possibili giocatori che possono raggiungere la coalizione dopo i sono $(|N|-|S|-1) = 3 - 1 - 1 = 1$, questo valore corrisponde ai 3 giocatori in gioco, dai quali sottraggo il numero dei giocatori presenti nella coalizione S ed il giocatore i . Anche in questo caso si calcola il fattoriale per considerare ogni possibile ordine secondo il quale gli altri partecipanti possono aggregarsi alla coalizione dopo i . Invece che sommare individualmente su tutte le permutazioni, si raccolgono tutte quelle in cui un certo insieme di giocatori si unisce prima del giocatore i , considerando qualsiasi ordine, e quelle che si uniscono dopo, sempre in qualsiasi ordine. Il denominatore $|N|!$ indica tutte le possibili sequenze generabili dalla quantità di giocatori che appartengono ad N e permette di calcolare la

⁵³ Wikipedia, voce *Fattoriale* <https://it.wikipedia.org/wiki/Fattoriale> (visitato il 7 ottobre 2021)

media rispetto alle $|N|!$ permutazioni. Il contributo del giocatore i rispetto alle permutazioni di S è identico, che sia $S = \text{Facebook, Instagram}$ o $S = \text{Instagram, Facebook}$, il contributo marginale di E-mail che si unisce alla coalizione dopo gli altri due corrisponde allo stesso valore e per questo motivo si moltiplica per ognuna il risultato di $v(S \cup \{i\}) - v(S)$.

5.3 Le proprietà

Il valore di Shapley è caratterizzato da una serie di proprietà, le quali si elencano di seguito:

Efficienza. La somma dei valori di Shapley di tutti i partecipanti è uguale al valore della coalizione completa, ovvero quella in cui tutti collaborano, per garantire che l'intero guadagno sia distribuito tra tutti.

Simmetria. Due giocatori si definiscono simmetrici se il valore della coalizione S unito ad uno è identico al valore di S unito all'altro. Quando due giocatori sono simmetrici significa che il loro valore di Shapley sarà identico e che quindi contribuiscono in modo equivalente al guadagno finale.

Linearità. Se si combinano due giochi descritti da due funzioni distinte, la ricompensa distribuita corrisponderà alla ricompensa che deriva da una funzione, sommata a quella che deriva dall'altra.

Giocatore nullo. Se un giocatore ha un contributo marginale pari a zero, avrà un valore di Shapley uguale a zero, non riceverà quindi nessuna ricompensa.

Monotonicità. Se un giocatore contribuisce maggiormente al guadagno finale rispetto ad un altro giocatore, il valore di Shapley del primo giocatore sarà superiore di quello dell'altro.

5.4 Come viene utilizzato il valore di Shapley nel seguente progetto

Nella seconda fase di questo lavoro, l'obiettivo è stato quello di applicare il concetto base del calcolo di Shapley al dataset esaminato. Nel dettaglio si è partiti dalla tabella dati ottenuta dalla prima parte del programma, rappresentata in figura 10 nel capitolo 4, ed è stato definito il tipo di studio.

La tabella dati risultante, come già spiegato nel capitolo precedente, è composta da tutte le radici delle varie componenti di ciascuna parola chiave suddivise per singoli termini, coppie e triple. Questa rappresentazione dei dati è stata interpretata rispetto al

valore di Shapley come una tabella di giocatori e coalizioni, dove le radici delle parole riportate individualmente corrispondono al giocatore per cui si vuole calcolare il valore di Shapley, mentre le radici raggruppate per due o tre elementi corrispondono alle coalizioni in cui i vari giocatori hanno effettivamente partecipato.

Il concetto base è stato rivisitato per due principali motivi: il valore di Shapley si calcola considerando tutte le possibili coalizioni, anche quelle dunque che non si sono verificate realmente, ma che sono comunque probabili. Questo implica nella scrittura del programma l'inserimento di una funzione che genera tutte le possibili combinazioni tra i giocatori in questione. Il numero di possibili combinazioni a partire da un insieme n di canali osservati, corrisponde a (2^n) combinazioni, questo ci permette di comprendere che se il numero di membri per cui è interessante calcolare il valore di Shapley è elevato, diventa automaticamente un calcolo molto complesso: se i giocatori fossero ad esempio 20, il numero delle possibili sotto sequenze generabili corrisponderebbe a 1.048.576. Nel dataset studiato le radici riportate singolarmente sono 211, il numero di tutte le combinazioni che si possono ottenere è enorme e crea dunque problemi soprattutto a livello computazionale.

L'altro motivo per cui si è pensato di rivedere la formula originale è guidato dal fatto che i giocatori di questo progetto sono degli elementi che costituiscono parole chiave per cui non risulta effettivamente interessante calcolare tutte le possibili combinazioni tra tutte le componenti poiché alcuni termini non verranno mai associati ad altri: per esempio il vocabolo *prosciutto* può essere associato a *crudo* o a *toscano*, ma rispetto al nostro studio, sarà poco probabile che venga utilizzato insieme alla parola *birra* o alla parola *cantucci*.

Queste due principali questioni hanno portato a prendere in considerazione il concetto alla base del calcolo di Shapley, la sua formula originale e provare a sfruttarlo in modo differente. Il focus si è proiettato esclusivamente sulle coalizioni reali, ovvero solo su quelle presenti nel dataset e che dunque sono state realmente utilizzate: non si considerano le probabili coalizioni, ma solo quelle per cui si hanno dati di riferimento. Questa scelta è stata presa come già anticipato per ovviare al problema della complessità dell'algoritmo in termini di tempo e di spazio, ma anche perché risultava poco interessante ai fini di tale studio.

Quando si ottiene il valore di Shapley per ciascuna radice, comprendiamo quale delle varie in gioco ha dato un migliore contributo alle performance finali: se dai risultati prevale che il termine *toscan* è quello che ha raggiunto un valore di Shapley

più elevato, nel momento in cui si vogliono generare nuove parole chiave posso fare riferimento a questi risultati e capire che tale vocabolo può portare ad importanti performance per le campagne future. Lo scopo è quello di fornire a chi si occupa della gestione delle nuove keywords di poter partire da una base di valori che fanno riferimento allo stesso cliente ma analizzando le performance di campagne passate. Come abbiamo visto nel capitolo 3, esistono molti strumenti per poter conoscere quanto una parola chiave è buona o quanto è ricercata dagli utenti, ma lo si può conoscere rispetto all'intero mondo del Web. Con questo progetto l'obiettivo è proprio quello di fornire un elemento in più, si dà la possibilità di poter conoscere non solo le performance di certe parole rispetto al pianeta internet, ma anche di potersi focalizzare su ciò che è successo in passato con le pubblicità dello stesso cliente per cui si stanno cercando nuove parole chiave per la generazione delle future campagne.

Lo scopo ultimo di questo studio è quindi quello di generare una guida per coloro che hanno il compito di programmare le future keywords, dando loro la possibilità di focalizzarsi anche su dati più specifici che fanno riferimento alla situazione passata dello stesso cliente per cui stanno lavorando.

In generale i risultati esprimono a quanto corrisponde il guadagno che ogni giocatore ottiene se si considera il contributo che ha effettivamente dato alla collaborazione con altri membri. Può capitare che per alcuni giocatori il valore di Shapley corrisponda a 0 e questo indica che non ha contribuito al raggiungimento dei risultati finali ottenuti dalla coalizione, la sua presenza nella squadra non ha aggiunto un valore a quanto già apportato dagli altri partecipanti, oppure può succedere che il valore di Shapley per un giocatore risulti essere negativo. Questo risultato, se si considera il concetto generale alla base del calcolo di Shapley può indicare che due giocatori a cui corrisponde un valore negativo non sono tra loro in sinergia quindi scegliere di metterli nella stessa squadra può risultare deleterio ai fini del raggiungimento dello scopo finale. Esaminando però lo stesso risultato rispetto al tipo di dati a cui si fa riferimento in questo studio, si potrebbe dare la stessa interpretazione a quanto accaduto poiché può effettivamente capitare che due termini non siano in sinergia tra loro e dunque se utilizzati insieme nella stessa parola chiave non riescono a garantire ottime performance, ma è altrettanto necessario sottolineare che proprio facendo riferimento a questa tipologia di dati non sempre tale spiegazione risulta troppo convincente in quanto i vari termini che compongono le parole chiave non interagiscono tra loro. Per comprendere il motivo di possibili valori negativi è

importante quindi considerare che i dati analizzati fanno riferimento a campagne search e per questo motivo sono risultati che per loro natura non possono essere completamente veritieri in quanto vengono influenzati da vari fattori necessari in tale ambito. Non è possibile pensare che tutto dipenda esclusivamente dal comportamento dell'utente poiché i fattori che entrano in gioco e che influenzano in parte i risultati ottenuti dalle campagne sono diversi: può influire ad esempio il tempo, sia se si considerano i giorni della settimana o un particolare periodo dell'anno, una certa campagna può essere stata visualizzata in un particolare giorno della settimana che risulta più proficuo del giorno in cui è stata vista invece un'altra campagna, oppure se si osservano i risultati ottenuti ad esempio da certi prodotti tipici del Natale, è normale che le performance che raggiungono nei mesi di novembre e dicembre saranno ben diverse da quelle ottenute nei mesi di marzo o aprile.

Altri fattori che influiscono sono i due attori principali che in simultanea gestiscono la campagna che sono il manager della stessa e Google: se da una parte vi è la componente naturale ovvero quella composta dagli utenti e dai loro comportamenti, dall'altra entrano in gioco i due attori che si possono definire come artificiali in quanto agiscono perché guidati da un particolare scopo, perché vogliono ottenere qualcosa. Il gestore della campagna deve prendere delle decisioni come quella del bid o del tipo di target, ma nello stesso tempo anche Google occupandosi della fase di ottimizzazione per la pubblicazione dell'annuncio prende le proprie decisioni. Il dataset utilizzato è un altro elemento importante, quando i numeri diventano più grandi chiaramente il disturbo dovuto a tutti i vari elementi in gioco tende a scomparire e questo si riscontra poi anche nei risultati, è importante quindi essere consapevoli della presenza di tali fattori che creano rumore ma anche del fatto che questa caratteristica tipica del settore non può influire sul concetto alla base del metodo proposto.

Tutti i fattori elencati devono far capire che i valori di Shapley ottenuti sono statisticamente soggetti a rumore ma perché calcolati sulla base di dati empirici disturbati, dati che per loro natura possono essere influenzati da vari elementi così come accade in generale in tale settore. Per qualsiasi applicazione si voglia fare delle informazioni prese da piattaforme come Google Ads è necessario tenere presente che non raggiungeranno mai un grado di affidabilità assoluta poiché tale settore dipende per sua natura dall'influenza dei vari elementi che ne rappresentano però una sua caratteristica fondante.

5.5 Rivisitazione del calcolo di Shapley

Quello che viene rivisto rispetto al calcolo ufficiale, non è tanto il concetto alla base, ma due caratteristiche che ne fanno parte.

Per prima cosa ciò che viene ignorato corrisponde all'ordine delle parole, in questo studio si assume infatti che l'ordine in cui due vocaboli interagiscono per formare una parola chiave, non sia interessante. Lavorando con keywords e non con canali attraverso i quali l'utente ha intrapreso un percorso particolare verso la conversione, risulta di maggiore importanza capire se le combinazioni generate tra parole risultano vincenti e quali dei due termini abbia influito maggiormente sul raggiungimento del risultato finale, non è importante conoscere quindi se l'ordine prescelto della coalizione sia stato *prosciutto toscano* o *toscano prosciutto*, ma capire se la presenza dell'uno o dell'altro vocabolo comporti dei cambiamenti significativi in termini di performance.

Altro elemento ignorato sono le permutazioni, se nella versione originale si considerano tutte le possibili permutazioni generabili da ogni coalizione S , in questo caso la concentrazione è orientata appunto solo su quelle realmente avute e quindi non si tiene di conto della parte di formula che va a considerare le permutazioni di ogni sequenza. Si esclude quindi la parte che permette di calcolare il peso di ciascuna coalizione, dato che non si è interessati a calcolare Shapley tenendo conto di tutti i modi in cui i vari giocatori possono combinarsi tra di loro nel formare una coalizione, ma lo si vuole solo ottenere occupandosi esclusivamente di quelle sequenze di parole che effettivamente hanno generato nel caso reale delle keywords ottenendo dei risultati.

La formula di base resta la stessa, si evita il calcolo dei fattoriali in quanto permettono di tenere di conto di tutte le possibili permutazioni a cui in questo caso non si è interessati in quanto non si hanno dati a disposizione, ma resta comunque il concetto del calcolo medio del contributo marginale e la funzione caratteristica che assegna un valore a ciascuna coalizione esistente.

5.6 Applicazione del calcolo

Il programma scritto in Python per il calcolo del valore di Shapley fa riferimento al dataset risultante dalla prima parte del lavoro e si suddivide nelle seguenti fasi: per prima cosa viene generata una lista delle coalizioni, nella versione originale si sarebbe

dovuto costruire una lista contenente tutte le possibili coalizioni tra giocatori, in questo caso generiamo solo quelle che realmente si sono avverate, ovvero quelle presenti nel dataset cui si fa riferimento.

Per il calcolo effettivo di Shapley si necessita di altre due funzioni, una che permette di generare tutti i sottoinsiemi esistenti dell'insieme di giocatori, l'altra invece per il calcolo del valore di ciascuna coalizione. Infine si procede con il calcolo vero e proprio di Shapley, quindi per ogni singolo giocatore si escludono quelli che non hanno mai partecipato a nessuna coalizione, ovvero coloro che formano individualmente una parola chiave, mentre per i rimanenti si prendono in considerazione tutte le coalizioni in cui sono realmente comparsi e senza considerare l'ordine se ne calcola il contributo marginale e la media di tutti i contributi dati rispetto al numero di coalizioni cui hanno realmente partecipato.

I giocatori nel seguente progetto corrispondono alle singole radici presenti nel dataset iniziale, alcune delle quali vengono riportate in tabella 5. Le coppie e le triple di radici costituiscono le coalizioni a cui hanno realmente partecipato i giocatori.

Parola_chaive	Impressioni	Click	Costo	Conversioni	CVR	CTR	CPA	CPC	CPM
acac	1785	70	22,35	0	0	3,92	0	0,32	12,52
acet	341	6	2,25	0	0	2,7	0	0,38	6,60
affumic	2014	67	15,36	0	0	10,39	0	0,23	7,63
alimentar	3656	198	99,8	6	12	22,01	16,63	0,50	27,30
amar	46	0	0	0	0	0	0	0	0
befan	48	3	0,25	0	0	6,25	0	0,08	5,21
biadin	6	0	0	0	0	0	0	0	0
bianc	532	9	5,76	0	0	8,37	0	0,64	10,83
biold	603	40	11,48	0	0	6,63	0	0,29	19,04
biscott	1294	40	7,52	2	25	15,27	3,76	0,19	5,81
bosc	100	2	0,9	0	0	2	0	0,45	9,00
bresaol	5868	199	41,54	0	0	3,4	0	0,21	7,08
caciocavall	256	11	3,89	0	0	4,3	0	0,35	15,20
caff	1585	38	20,95	0	0	4,45	0	0,55	13,22

Tabella 5. Esempio di alcuni dei giocatori.

Per comprendere il modo in cui viene calcolato il valore di Shapley di ciascun giocatore nel lavoro presentato, si riporta un esempio con dati inventati.

In Tabella 6 è presente un esempio di giocatori, le coalizioni a cui partecipano ed il rispettivo valore calcolato sulla base del CVR che ciascuna coalizione ha ottenuto durante le campagne pubblicitarie.

Coalizioni (S)	v(S)
prosciutt	0,11
pecorin	0,15
tosc	0,12
prosciutt + tosc	0,24
pecorin + tosc	0,33

Tabella 6. Dataset di esempio per il calcolo di Shapley.

Si procede a questo punto con il calcolo. Dato che non vi è interesse per tutte le possibili permutazioni e per l'ordine secondo cui le parole giocano insieme, si calcola il valore di Shapley tramite la media del contributo marginale del giocatore rispetto alle coalizioni in cui effettivamente è comparso.

Partendo dal termine *prosciutt*, è comparso in *prosciutt + tosc*, se dunque il valore di *prosciutt + tosc* è pari a 0,24 e quello di *tosc* a 0,12, il contributo marginale di *prosciutt* sarà in questo caso $(0,24 - 0,12) = 0,12$. Non compare in altre coalizioni quindi la media si calcola con la seguente divisione $0,12/1$ ed il risultato 0,12 corrisponde al valore di Shapley di *prosciutt*.

Passando al vocabolo *tosc*, in questo caso la presenza si rileva in due coalizioni e si calcola quindi il contributo marginale per ciascuna delle due. Nel caso di *prosciutt + tosc*, il contributo marginale di *tosc* è pari a $(0,24 - 0,11) = 0,13$ e nel caso invece di *pecorin + tosc*, si ottiene $(0,33 - 0,15) = 0,18$. La media tra i valori calcolati corrisponde a $(0,18 + 0,13) / 2 = 0,155$, valore di Shapley per *tosc*.

Il giocatore *pecorin* è comparso in una sola coalizione ed il suo valore di Shapley corrisponde quindi a $(0,33 - 0,12) = 0,21$.

L'esempio appena riportato è stato costruito con lo scopo di comprendere i calcoli alla base del metodo presentato, su un campione più ampio di dati si possono ottenere dei risultati più significativi.

5.7 Funzione caratteristica

Il valore di ciascuna coalizione è stato calcolato facendo riferimento al valore del CTR e del CVR, la funzione caratteristica quindi considera in un primo momento tutti i Conversion Rate di ogni giocatore appartenente alla coalizione e poi ripete la stessa

cosa con i CTR, calcolando in definitiva due versioni distinte del valore di Shapley per gli stessi giocatori.

La scelta si è orientata verso queste due metriche in quanto sembrano le più interessanti per il tipo di calcolo. Potevano anche essere utilizzate altre metriche come impressioni o conversioni che permettono ad esempio di valutare quanto può migliorare la loro performance rispetto alla presenza o meno di un termine in una parola chiave: ad esempio, in termini di impressioni è possibile valutare cosa comporta l'aggiunta della parola *toscano* alla keyword *prosciutto crudo*. In realtà però anche in quest'occasione si deve prendere atto del fatto che conversioni ed impressioni non dipendono solo dall'utente che interagisce con l'annuncio, ma sono molto influenzate anche dal valore del bid scelto, ovvero dalla cifra offerta per ogni click nelle campagne Pay Per Click. Come descritto nel capitolo 2 (v. 2.7) una volta decise le parole chiave su cui puntare per la campagna a pagamento è necessario definire quale cifra si è disposti a pagare per ognuna, ed il valore del bid risulta essere dunque determinante per la visibilità della pagina: maggiore è l'offerta, maggiore sarà la possibilità di essere visti.

Valutare Shapley per ogni componente di tutte le parole chiave basandosi sul valore delle impressioni, delle conversioni o dei click può risultare molto interessante, ma solo se studiato su un insieme di parole per cui l'offerta per ogni clic è uguale per tutte, altrimenti il confronto non risulta equo. Anche nel CTR e nel CVR si fa comunque riferimento a conversioni ed impressioni, ma la scelta iniziale del bid ha un impatto differente su queste metriche poiché meno sensibili a tale valore.

5.8 Risultati ottenuti

Il risultato finale è composto da un dataset formato da due colonne, nella prima delle quali è riportato il giocatore ovvero la radice dei vocaboli che hanno partecipato insieme ad altre alla composizione di una o più parole chiave, nell'altra colonna invece è riportato il relativo valore di Shapley. Come anticipato è presente sia la versione che fa riferimento al CVR che quella che fa riferimento al CTR e per ognuna si riportano tutti i valori in ordine decrescente, così da poter immediatamente capire quali siano i termini che hanno ottenuto rispetto allo studio affrontato il miglior risultato.

Di seguito una parte delle tabelle dei valori ottenuti, riportate poi per intero in appendice. In totale i giocatori per cui viene riportato un valore sono 196, non più 211

quanto erano inizialmente, questo perché non viene calcolato il valore di Shapley per coloro che non hanno mai partecipato a nessuna competizione.

	Player	Shapley_value
0	decortic	62.11
1	caseif	47.05
2	maiones	11.02
3	castagn	9.80
4	bazzon	4.39
...
191	prezz	-6.45
192	birrific	-7.31
193	biadin	-7.89
194	tradizion	-9.10
195	cec	-17.52

196 rows × 2 columns

Figura 11. Valore di Shapley calcolato in base al CTR.

	Player	Shapley_value
0	verdur	98.15
1	cinghial	5.76
2	bazzon	4.07
3	pat	3.43
4	cantucc	3.14
...
191	mandorl	-3.77
192	morbid	-3.77
193	fettuccin	-3.85
194	befan	-5.00
195	nocciol	-5.00

196 rows × 2 columns

Figura 12. Valore di Shapley calcolato in base al CVR.

Alcune parole hanno ottenuto un valore di Shapley pari a 0 e questo implica che il rispettivo valore di CVR o CTR fosse uguale a 0 e che il contributo dato da queste alle keywords cui hanno partecipato è nullo. Sono inoltre presenti anche valori negativi tra i risultati che come si è visto nel paragrafo 5.2, possono indicare delle reali sinergie negative, che due termini quindi non riescono effettivamente a produrre buoni risultati se utilizzati insieme, oppure che il campione osservato è piccolo o dipendere da un'interferenza tra le azioni dei vari soggetti in gioco.

I risultati finali dipendono dal dataset su cui lo studio si svolge, lo scopo principale è stato quello di voler descrivere il tipo di studio effettuato partendo dal concetto del valore di Shapley e adattando tale definizione ad un tipo di dati relativi a parole chiave in generale. Non conta molto la tipologia del cliente scelto o i valori effettivamente ottenuti, l'interesse si focalizza piuttosto sul metodo proposto e sul fatto che possa essere applicato alle keywords di qualsiasi tipologia di prodotto o azienda: l'idea è stata quella di creare uno strumento aggiuntivo, utile per la fase di generazione di keywords future basandosi sui risultati ottenuti dalle scelte passate.

6. Conclusione

I risultati finali raggiunti a seguito dello svolgimento del progetto appena presentato si possono visualizzare in appendice, dove vengono riportate tre tabelle con i rispettivi valori per ciascun componente di ogni parola chiave. Nel dettaglio è possibile osservare nella prima tabella i risultati della prima parte del progetto, ovvero quello dedicato al concetto di marketing di attribuzione, mentre nelle altre due si possono osservare i valori ottenuti dall'applicazione del calcolo di Shapley, considerando le due metriche prese in esame: CVR e CTR.

I risultati effettivi vengono riportati per completezza del lavoro svolto, ma il vero scopo di tale progetto è quello di fornire un metodo che possa affiancare gli strumenti già esistenti per affrontare la fase di pianificazione delle parole chiave nelle campagne pubblicitarie a pagamento. L'obiettivo inizialmente prefissato è stato dunque raggiunto, in quanto si è riusciti dopo una prima fase di stemming su tutti i termini che costituiscono una parola chiave, ad applicare loro la formula del calcolo del valore di Shapley, riuscendo quindi ad utilizzare questo calcolo tipico di altri ambiti del web marketing, per la fase di generazione e pianificazione delle parole chiave per la rete di ricerca.

Un altro elemento di cui tener conto è che la scelta del cliente è stata concordata con l'agenzia WebSoup SRL per poter avere a disposizione dei dati su cui svolgere le analisi e le operazioni messe in atto, ma non limita l'usabilità di tale progetto, il quale può essere applicato a qualsiasi altra tipologia di azienda. Il programma alla base del progetto è stato scritto in Python, sono state utilizzate varie librerie predefinite, come riportato nei capitoli 4 e 5, e mi ha permesso di mettere in pratica alcune delle conoscenze apprese durante gli anni attraverso i vari corsi seguiti. La parte più complessa è stata quella relativa all'applicazione del calcolo di Shapley, soprattutto perché le informazioni reperibili principalmente fanno riferimento all'applicazione standard di tale principio. Si sono dunque dovute considerare alcune caratteristiche che si distaccano dalla versione originale (v. 5.4 e 5.5) senza però alterare il concetto base di tale calcolo.

Volendo suggerire possibili sviluppi futuri o eventuali modifiche a quanto svolto, si propone per esempio la possibilità di applicare ai termini presi in esame non solo il processo di stemming, ma anche quello di lemmatizzazione, per garantire un

livello di precisione più elevato dei termini, dovendo comunque gestire le possibili problematiche riportate nel capitolo 4, che i due algoritmi presi autonomamente possono comportare. Altra possibilità futura, consiste nell'applicazione del calcolo di Shapley considerando anche altre metriche, non solo quelle viste, così da poter sfruttare tale metodo osservando i risultati raggiunti attraverso uno studio più ampio.

7. Tabelle complete dei risultati finali

La seguente tabella riporta i valori ottenuti a seguito dell'applicazione del concetto di marketing dell'attribuzione. Le colonne lasciate in bianco rappresentano i costi effettuati dal cliente esaminato quindi come da lui richiesto sono stati omessi.

Radice_parola	Impressioni	Click	Costo	Conversioni	CVR	CTR	CPA	CPC	CPM
acac	1785	70		0	0	3,92			
acet	341	6		0	0	1,76			
affumic	2014	67		0	0	3,33			
aleat	5	2		0	0	40			
alimentar	3656	198		6	3,03	5,42			
amar	46	0		0	0	0			
ambrat	63	0		0	0	0			
antic	71	0		0	0	0			
antipast	746	34		1	2,94	4,56			
aranc	46	0		0	0	0			
arroto	1595	42		0	0	2,63			
artigianal	16696	1084		1	0,09	6,49			
azi	1	0		0	0	0			
aziendal	22	2		0	0	9,09			
balsam	228	6		0	0	2,63			
bazzon	492	40		2	5	8,13			
befan	48	3		0	0	6,25			
biadin	6	0		0	0	0			
biane	532	9		0	0	1,69			
bio	2239	109		1	0,92	4,87			
biodinam	28	0		0	0	0			
biolog	8695	548		0	0	6,3			
biold	603	40		0	0	6,63			
birr	4450	193		0	0	4,34			
birrif	21	2		0	0	9,52			
birrific	3	0		0	0	0			
biscott	1294	40		2	5	3,09			
bosc	100	2		0	0	2			
box	322	33		0	0	10,25			
bresaol	5868	199		0	0	3,39			
buccell	398	15		0	0	3,77			
cac	3316	85		1	1,18	2,56			
caciocavall	256	11		0	0	4,3			
caff	1585	38		0	0	2,4			
cannellin	899	24		0	0	2,67			
cantucc	1503	53		2	3,77	3,53			
cantuccin	895	33		0	0	3,69			
cappell	2051	56		0	0	2,73			
carn	3928	106		0	0	2,7			
cas	1675	99		0	0	5,91			
caseif	195	107		0	0	54,87			

castagn	1177	47		0	0	3,99			
cec	674	14		0	0	2,08			
cen	1626	97		0	0	5,97			
cest	41061	3237		75	2,32	7,88			
chianin	6	0		0	0	0			
chiar	95	3		0	0	3,16			
cib	175	11		0	0	6,29			
cinghial	3095	79		1	1,27	2,55			
ciocol	199	6		0	0	3,02			
collin	1	0		0	0	0			
colomb	6416	260		0	0	4,05			
compr	445	26		0	0	5,84			
cond	58	0		0	0	0			
confettur	859	17		0	0	1,98			
confezion	2115	134		5	3,73	6,34			
copp	186	4		0	0	2,15			
cost	285	10		0	0	3,51			
crem	1351	71		0	0	5,26			
crostin	1874	88		0	0	4,7			
crud	12701	426		3	0,7	3,35			
cucin	6658	248		1	0,4	3,72			
culinar	648	41		0	0	6,33			
cuor	15	0		0	0	0			
decortic	182	130		0	0	71,43			
dolc	828	38		0	0	4,59			
dur	2484	42		0	0	1,69			
ecommerc	339	37		0	0	10,91			
eucalipt	293	4		0	0	1,37			
extra	215	11		0	0	5,12			
extravergin	491	15		0	0	3,05			
fagiol	1452	43		1	2,33	2,96			
farin	8055	304		1	0,33	3,77			
farr	1044	26		1	3,85	2,49			
fest	206	2		0	0	0,97			
fettuccin	242	5		0	0	2,07			
fich	409	6		0	0	1,47			
filett	11	0		0	0	0			
fondent	169	4		0	0	2,37			
food	466	41		0	0	8,8			
formagg	18443	796		8	1,01	4,32			
fresc	1453	59		1	1,69	4,06			
frollin	257	14		0	0	5,45			
frutt	100	2		0	0	2			
fung	1173	10		0	0	0,85			
garfagnan	40	2		0	0	5			
gastronom	19688	1568		47	3	7,96			
genoves	4029	137		0	0	3,4			
giall	336	7		0	0	2,08			
gioves	444	18		0	0	4,05			
girasol	369	5		0	0	1,36			
glass	6	0		0	0	0			
gourmet	267	14		0	0	5,24			

gran	7129	124		0	0	1,74			
ide	890	48		1	2,08	5,39			
inter	627	102		0	0	16,27			
ital	7583	258		0	0	3,4			
lard	8171	421		8	1,9	5,15			
legum	2543	237		0	0	9,32			
lepr	234	9		0	0	3,85			
limon	4	0		0	0	0			
linchett	125	6		0	0	4,8			
linguin	777	7		0	0	0,9			
liquor	152	12		0	0	7,89			
lucc	4	0		0	0	0			
lucches	10	1		0	0	10			
lup	1	0		0	0	0			
macin	711	37		0	0	5,2			
maiones	798	25		0	0	3,13			
maltagl	646	20		0	0	3,1			
mamm	206	2		0	0	0,97			
mandorl	21	0		0	0	0			
manz	72	4		0	0	5,56			
marmell	4353	218		1	0,46	5,01			
menu	49	2		0	0	4,08			
merlot	613	25		0	0	4,08			
miel	31693	906		0	0	2,86			
millefior	449	11		0	0	2,45			
mondiol	24	1		0	0	4,17			
mor	350	9		0	0	2,57			
morbid	136	4		0	0	2,94			
natal	7420	719		13	1,81	9,69			
nataliz	13784	965		13	1,35	7			
ner	235	10		0	0	4,26			
noc	9	0		0	0	0			
nocciol	129	0		0	0	0			
oli	15143	327		0	0	2,16			
oliv	2789	109		0	0	3,91			
onlin	51777	2984		29	0,97	5,76			
ortic	134	1		0	0	0,75			
pacc	821	38		0	0	4,63			
paccher	1702	26		0	0	1,53			
pan	26	1		0	0	3,85			
pancett	5561	173		1	0,58	3,11			
panfort	6981	259		4	1,54	3,71			
papp	884	17		0	0	1,92			
particular	602	39		1	2,56	6,48			
pasqu	4642	173		1	0,58	3,73			
pasqual	4456	256		3	1,17	5,75			
past	2542	542		1	0,18	21,32			
pat	1105	26		1	3,85	2,35			
pecor	4552	181		2	1,1	3,98			
pecorin	16741	828		3	0,36	4,95			
penitent	19	1		0	0	5,26			
pep	3319	85		1	1,18	2,56			

peperoncin	280	4		0	0	1,43			
pest	4029	137		0	0	3,4			
pic	6101	222		1	0,45	3,64			
piccant	1392	31		0	0	2,23			
pienz	4463	392		2	0,51	8,78			
pietr	114	28		0	0	24,56			
pistacc	147	6		0	0	4,08			
pomarol	300	5		0	0	1,67			
pomodor	884	17		0	0	1,92			
porcin	1173	10		0	0	0,85			
pozz	72	4		0	0	5,56			
preg	941	57		0	0	6,06			
prezz	559	8		0	0	1,43			
prodott	36504	3540		45	1,27	9,7			
produzion	2	0		0	0	0			
prosciutt	25888	969		9	0,93	3,74			
puccin	26	1		0	0	3,85			
ragù	2097	54		1	1,85	2,58			
regal	11640	687		13	1,89	5,9			
riboll	553	31		0	0	5,61			
ros	23	1		0	0	4,35			
ross	919	41		1	2,44	4,46			
rustic	9	0		0	0	0			
sal	253	14		0	0	5,53			
salam	9111	246		5	2,03	2,7			
sals	4196	117		0	0	2,79			
salum	19370	1069		9	0,84	5,52			
san	2655	147		0	0	5,54			
sant	79	4		0	0	5,06			
sapor	140	17		0	0	12,14			
scoppol	159	13		0	0	8,18			
secc	1678	68		0	0	4,05			
selvaggin	171	6		0	0	3,51			
senator	2051	56		0	0	2,73			
sepp	90	3		0	0	3,33			
shop	952	118		1	0,85	12,39			
slow	241	22		0	0	9,13			
soppress	2948	97		0	0	3,29			
souven	10	1		0	0	10			
spaghet	130	3		0	0	2,31			
special	3688	270		5	1,85	7,32			
stagion	291	14		0	0	4,81			
stes	1839	62		1	1,61	3,37			
sug	4134	113		2	1,77	2,73			
syrah	472	8		0	0	1,69			
tagliatell	234	4		0	0	1,71			
tartin	1015	56		0	0	5,52			
tartuf	2883	98		2	2,04	3,4			
tener	3187	43		0	0	1,35			
tipic	18289	1894		28	1,48	10,36			
torron	2641	77		0	0	2,92			
tosc	69593	5441		60	1,1	7,82			

tradizion	10	0			0	0	0			
tranc	1022	27			0	0	2,64			
tuad	14	1			3	300	7,14			
uov	2213	72			0	0	3,25			
urlo	1	0			0	0	0			
uva	80	3			0	0	3,75			
valentin	2211	129			0	0	5,83			
veloc	1003	55			0	0	5,48			
vend	10332	445			6	1,35	4,31			
verdur	151	1			1	100	0,66			
vergin	215	11			0	0	5,12			
vin	4160	223			0	0	5,36			
zenzer	1	0			0	0	0			
zupp	1620	74			0	0	4,57			
acac miel	1785	70			0	0	3,92			
acet antic	71	0			0	0	0			
acet balsam	228	6			0	0	2,63			
acet glass	6	0			0	0	0			
acet vin	42	0			0	0	0			
affumic copp	186	4			0	0	2,15			
affumic lard	62	3			0	0	4,84			
affumic pancett	1766	60			0	0	3,4			
alimentar cest	1154	50			1	2	4,33			
alimentar confezion	1681	110			5	4,55	6,54			
alimentar pace	821	38			0	0	4,63			
alimentar prodott	752	48			4	8,33	6,38			
alimentar regal	3656	198			6	3,03	5,42			
amar aranc	46	0			0	0	0			
amar marmell	46	0			0	0	0			
ambrat birr	63	0			0	0	0			
antipast tosc	746	34			1	2,94	4,56			
aranc marmell	46	0			0	0	0			
arrotoI pancett	1595	42			0	0	2,63			
artigianal birr	2584	110			0	0	4,26			
artigianal birrific	3	0			0	0	0			
artigianal cantucc	695	28			0	0	4,03			
artigianal colomb	6416	260			0	0	4,05			
artigianal frollin	257	14			0	0	5,45			
artigianal ital	375	17			0	0	4,53			
artigianal maiones	795	24			0	0	3,02			
artigianal onlin	295	13			0	0	4,41			
artigianal pasqu	2213	72			0	0	3,25			
artigianal pasqual	132	0			0	0	0			
artigianal past	1092	499			1	0,2	45,7			
artigianal torron	2641	77			0	0	2,92			
artigianal tosc	92	7			0	0	7,61			
artigianal uov	2213	72			0	0	3,25			
azi cest	1	0			0	0	0			
azi nataliz	1	0			0	0	0			
aziendal cest	22	2			0	0	9,09			
aziendal regal	22	2			0	0	9,09			
balsam glass	6	0			0	0	0			

bazon prosciutt	492	40		2	5	8,13			
befan biscott	48	3		0	0	6,25			
biadin liquor	6	0		0	0	0			
bianc chianin	6	0		0	0	0			
bianc cond	1	0		0	0	0			
bianc crem	432	6		0	0	1,39			
bianc ragù	6	0		0	0	0			
bianc tartuf	433	6		0	0	1,39			
bianc tosc	93	3		0	0	3,23			
bianc vin	93	3		0	0	3,23			
bio cappell	8	0		0	0	0			
bio marmell	2231	109		1	0,92	4,89			
bio past	8	0		0	0	0			
bio senator	8	0		0	0	0			
biodinam vin	28	0		0	0	0			
biolog cappell	26	1		0	0	3,85			
biolog collin	1	0		0	0	0			
biolog farin	381	119		0	0	31,23			
biolog ital	193	11		0	0	5,7			
biolog legum	193	11		0	0	5,7			
biolog lucches	1	0		0	0	0			
biolog marmell	2075	109		0	0	5,25			
biolog miel	5305	282		0	0	5,32			
biolog oli	597	21		0	0	3,52			
biolog past	26	1		0	0	3,85			
biolog senator	26	1		0	0	3,85			
biolog tosc	17	0		0	0	0			
biolog uva	27	2		0	0	7,41			
biolog vin	91	3		0	0	3,3			
birr chiar	95	3		0	0	3,16			
birr compr	445	26		0	0	5,84			
birr ital	506	23		0	0	4,55			
birr onlin	1795	82		0	0	4,57			
birr tosc	166	12		0	0	7,23			
birrif tosc	21	2		0	0	9,52			
birrific tosc	3	0		0	0	0			
biscott cantucc	270	8		2	25	2,96			
biscott cioccol	169	4		0	0	2,37			
biscott fondent	169	4		0	0	2,37			
biscott nocciol	129	0		0	0	0			
biscott tosc	678	25		0	0	3,69			
bosc confettur	100	2		0	0	2			
bosc frutt	100	2		0	0	2			
box food	225	19		0	0	8,44			
box prodott	97	14		0	0	14,43			
box regal	225	19		0	0	8,44			
box tosc	97	14		0	0	14,43			
bresaol filett	11	0		0	0	0			
buccell tosc	398	15		0	0	3,77			
cac pep	3316	85		1	1,18	2,56			
cac pic	1572	50		0	0	3,18			
cac sug	1744	35		1	2,86	2,01			

caff gran	988	29		0	0	2,94			
caff macin	597	9		0	0	1,51			
cannellin fagiol	899	24		0	0	2,67			
cantucc cioccol	30	2		0	0	6,67			
cantucc mandorl	21	0		0	0	0			
cantucc morbid	136	4		0	0	2,94			
cantucc tosc	351	11		0	0	3,13			
cappell dur	1	0		0	0	0			
cappell farin	1	0		0	0	0			
cappell gran	471	10		0	0	2,12			
cappell macin	1	0		0	0	0			
cappell past	1450	43		0	0	2,97			
cappell piotr	1	0		0	0	0			
cappell senator	2051	56		0	0	2,73			
cappell spaghet	130	3		0	0	2,31			
carn sals	3928	106		0	0	2,7			
cas cen	1626	97		0	0	5,97			
cas menu	49	2		0	0	4,08			
cas san	1675	99		0	0	5,91			
cas valentin	1675	99		0	0	5,91			
caseif tosc	195	107		0	0	54,87			
castagn farin	129	32		0	0	24,81			
castagn miel	1048	15		0	0	1,43			
cec farin	674	14		0	0	2,08			
cec macin	2	0		0	0	0			
cec piotr	2	0		0	0	0			
cen san	1626	97		0	0	5,97			
cen valentin	1626	97		0	0	5,97			
cest gastronom	13820	1224		45	3,68	8,86			
cest natal	7222	708		13	1,84	9,8			
cest nataliz	13784	965		13	1,35	7			
cest onlin	8794	696		11	1,58	7,91			
cest particular	602	39		1	2,56	6,48			
cest pasqu	708	28		0	0	3,95			
cest pasqual	4324	256		3	1,17	5,92			
cest regal	1510	84		2	2,38	5,56			
cest san	27	4		0	0	14,81			
cest valentin	27	4		0	0	14,81			
chianin ragù	6	0		0	0	0			
cib regal	58	1		0	0	1,72			
cib san	175	11		0	0	6,29			
cib valentin	175	11		0	0	6,29			
cinghial pat	99	4		1	25	4,04			
cinghial ragù	1026	27		0	0	2,63			
cinghial salam	1970	48		0	0	2,44			
cioccol fondent	169	4		0	0	2,37			
collin luches	1	0		0	0	0			
collin oli	1	0		0	0	0			
colomb pasqual	132	0		0	0	0			
compr onlin	445	26		0	0	5,84			
cond ner	2	0		0	0	0			
cond piccant	55	0		0	0	0			

cond tartuf	3	0		0	0	0			
confettur fich	409	6		0	0	1,47			
confettur frutt	100	2		0	0	2			
confettur mor	350	9		0	0	2,57			
confezion gastronom	434	24		0	0	5,53			
confezion prodott	752	48		4	8,33	6,38			
confezion regal	1681	110		5	4,55	6,54			
cost crud	285	10		0	0	3,51			
cost prosciutt	285	10		0	0	3,51			
crem lard	311	17		0	0	5,47			
crem sal	253	14		0	0	5,53			
crem tartuf	788	40		0	0	5,08			
crostin tosc	1874	88		0	0	4,7			
crud inter	73	0		0	0	0			
crud onlin	308	8		0	0	2,6			
crud prezz	559	8		0	0	1,43			
crud prosciutt	12701	426		3	0,7	3,35			
crud vend	308	8		0	0	2,6			
cucin tosc	6658	248		1	0,4	3,72			
culinar natal	22	1		0	0	4,55			
culinar regal	648	41		0	0	6,33			
cuor prosciutt	15	0		0	0	0			
decortic legum	182	130		0	0	71,43			
dolc tosc	828	38		0	0	4,59			
dur farin	2484	42		0	0	1,69			
dur gran	2484	42		0	0	1,69			
dur macin	1	0		0	0	0			
dur pietr	1	0		0	0	0			
dur senator	1	0		0	0	0			
ecommerc prodott	339	37		0	0	10,91			
ecommerc tipic	339	37		0	0	10,91			
eucalipt miel	293	4		0	0	1,37			
extra oli	215	11		0	0	5,12			
extra oliv	215	11		0	0	5,12			
extra vergin	215	11		0	0	5,12			
extravergin ital	204	7		0	0	3,43			
extravergin oli	491	15		0	0	3,05			
extravergin oliv	491	15		0	0	3,05			
fagiol lucches	5	0		0	0	0			
fagiol ross	548	19		1	5,26	3,47			
farin farr	753	19		1	5,26	2,52			
farin giall	336	7		0	0	2,08			
farin gran	5671	85		0	0	1,5			
farin macin	114	28		0	0	24,56			
farin pietr	114	28		0	0	24,56			
farin senator	1	0		0	0	0			
farin tener	3187	43		0	0	1,35			
farr fettuccin	242	5		0	0	2,07			
farr legum	45	1		0	0	2,22			
farr lucches	4	1		0	0	25			
farr zupp	45	1		0	0	2,22			
fest ide	3	0		0	0	0			

fest mamm	206	2		0	0	0,97			
fest regal	206	2		0	0	0,97			
food regal	225	19		0	0	8,44			
food salum	241	22		0	0	9,13			
food slow	241	22		0	0	9,13			
formagg onlin	387	27		0	0	6,98			
formagg pecor	4552	181		2	1,1	3,98			
formagg pecorin	10218	352		0	0	3,44			
formagg tosc	3673	263		6	2,28	7,16			
formagg vend	387	27		0	0	6,98			
fresc pecorin	1453	59		1	1,69	4,06			
fung linguin	25	1		0	0	4			
fung porcin	1173	10		0	0	0,85			
fung sug	1148	9		0	0	0,78			
garfagnan mondiol	24	1		0	0	4,17			
garfagnan prosciutt	16	1		0	0	6,25			
gastronom ide	820	47		1	2,13	5,73			
gastronom natal	176	10		0	0	5,68			
gastronom onlin	181	16		2	12,5	8,84			
gastronom regal	3946	218		3	1,38	5,52			
gastronom souven	10	1		0	0	10			
gastronom tosc	1785	129		0	0	7,23			
genoves pest	4029	137		0	0	3,4			
gioves san	444	18		0	0	4,05			
girasol miel	369	5		0	0	1,36			
gourmet san	267	14		0	0	5,24			
gourmet valentin	267	14		0	0	5,24			
gran macin	1	0		0	0	0			
gran pietr	1	0		0	0	0			
gran senator	471	10		0	0	2,12			
gran tener	3187	43		0	0	1,35			
ide mamm	3	0		0	0	0			
ide regal	890	48		1	2,08	5,39			
ide san	67	1		0	0	1,49			
ide valentin	67	1		0	0	1,49			
inter prezz	73	0		0	0	0			
inter prosciutt	627	102		0	0	16,27			
ital legum	193	11		0	0	5,7			
ital miel	6447	206		0	0	3,2			
ital oli	204	7		0	0	3,43			
ital oliv	204	7		0	0	3,43			
ital onlin	233	11		0	0	4,72			
ital prodott	233	11		0	0	4,72			
ital tipic	233	11		0	0	4,72			
lard tartuf	1	0		0	0	0			
legum onlin	120	12		0	0	10			
legum secc	1678	68		0	0	4,05			
legum zupp	45	1		0	0	2,22			
lepr ragù	234	9		0	0	3,85			
limon linguin	3	0		0	0	0			
limon marmell	1	0		0	0	0			
limon pep	3	0		0	0	0			

limon zenzer	1	0		0	0	0			
linguin pep	3	0		0	0	0			
linguin porcin	25	1		0	0	4			
linguin tartuf	749	6		0	0	0,8			
liquor lucc	4	0		0	0	0			
liquor tosc	142	12		0	0	8,45			
lucches oli	1	0		0	0	0			
lup urlo	1	0		0	0	0			
lup vin	1	0		0	0	0			
macin pietr	114	28		0	0	24,56			
macin senator	1	0		0	0	0			
maiones tosc	3	1		0	0	33,33			
mamm regal	206	2		0	0	0,97			
manz pozz	72	4		0	0	5,56			
marmell zenzer	1	0		0	0	0			
menu san	49	2		0	0	4,08			
menu valentin	49	2		0	0	4,08			
miel millefior	449	11		0	0	2,45			
miel onlin	10537	149		0	0	1,41			
miel produzion	2	0		0	0	0			
miel tartuf	16	1		0	0	6,25			
miel tosc	2	0		0	0	0			
miel vend	5442	163		0	0	3			
natal regal	198	11		0	0	5,56			
nataliz onlin	8613	680		9	1,32	7,9			
nataliz particular	602	39		1	2,56	6,48			
ner oliv	143	7		0	0	4,9			
ner pat	143	7		0	0	4,9			
ner sepp	90	3		0	0	3,33			
ner tagliatell	90	3		0	0	3,33			
ner tartuf	2	0		0	0	0			
noc uva	9	0		0	0	0			
oli oliv	2345	96		0	0	4,09			
oli peperoncín	279	4		0	0	1,43			
oli piccant	64	4		0	0	6,25			
oli tartuf	124	3		0	0	2,42			
oli tosc	1379	66		0	0	4,79			
oli vergin	215	11		0	0	5,12			
oliv pat	444	13		0	0	2,93			
oliv vergin	215	11		0	0	5,12			
onlin prodott	8239	972		13	1,34	11,8			
onlin prosciutt	6101	248		1	0,4	4,06			
onlin salum	14800	776		4	0,52	5,24			
onlin shop	952	118		1	0,85	12,39			
onlin tipic	2998	332		8	2,41	11,07			
onlin tosc	8393	988		13	1,32	11,77			
onlin vend	4890	282		6	2,13	5,77			
onlin vin	1004	22		0	0	2,19			
ortic tagliatell	134	1		0	0	0,75			
pacc regal	821	38		0	0	4,63			
pan puccin	26	1		0	0	3,85			
pancett stes	1839	62		1	1,61	3,37			

pancett tosc	361	9			0	0	2,49			
papp pomodor	884	17			0	0	1,92			
pasqu regal	1721	73			1	1,37	4,24			
pasqu uov	2213	72			0	0	3,25			
past senator	1450	43			0	0	2,97			
pat tartuf	442	4			0	0	0,9			
pat tosc	120	5			0	0	4,17			
pecorin pienz	4463	392			2	0,51	8,78			
pecorin stagion	291	14			0	0	4,81			
pecorin tartuf	316	11			0	0	3,48			
pecorin tosc	291	14			0	0	4,81			
penitent prosciutt	19	1			0	0	5,26			
pep pic	1572	50			0	0	3,18			
pep sug	1744	35			1	2,86	2,01			
peperoncin tagliatell	1	0			0	0	0			
piccant salam	1273	27			0	0	2,12			
pietr senator	1	0			0	0	0			
pistacc salam	147	6			0	0	4,08			
porcin sug	1148	9			0	0	0,78			
preg tosc	941	57			0	0	6,06			
preg vin	941	57			0	0	6,06			
prezz prosciutt	559	8			0	0	1,43			
prodott regal	1816	176			7	3,98	9,69			
prodott shop	952	118			1	0,85	12,39			
prodott tipic	18149	1877			28	1,49	10,34			
prodott tosc	35180	3444			41	1,19	9,79			
prodott tradizion	10	0			0	0	0			
prodott vend	817	95			4	4,21	11,63			
produzion tosc	2	0			0	0	0			
prosciutt salam	5276	132			3	2,27	2,5			
prosciutt trunc	1022	27			0	0	2,64			
prosciutt vend	308	8			0	0	2,6			
ragù tosc	680	17			0	0	2,5			
ragù verdur	151	1			1	100	0,66			
regal san	152	6			0	0	3,95			
regal tipic	1064	128			3	2,34	12,03			
regal tosc	1064	128			3	2,34	12,03			
regal valentin	152	6			0	0	3,95			
ros tosc	23	1			0	0	4,35			
ros vin	23	1			0	0	4,35			
ross tosc	371	22			0	0	5,93			
ross vin	371	22			0	0	5,93			
rustic tagliatell	9	0			0	0	0			
salam tartuf	445	33			2	6,06	7,42			
sals tosc	268	11			0	0	4,1			
salum slow	241	22			0	0	9,13			
salum tosc	4329	271			5	1,85	6,26			
salum vend	2374	130			2	1,54	5,48			
san valentin	2211	129			0	0	5,83			
sant vin	79	4			0	0	5,06			
sapor tipic	140	17			0	0	12,14			
sapor tosc	140	17			0	0	12,14			

selvaggin sug	171	6		0	0	3,51			
senator spaghet	130	3		0	0	2,31			
sepp tagliatell	90	3		0	0	3,33			
shop tosc	952	118		1	0,85	12,39			
special tosc	3688	270		5	1,85	7,32			
stagion tosc	291	14		0	0	4,81			
sug tosc	1071	63		1	1,59	5,88			
tartin tosc	12	1		0	0	8,33			
tartin veloc	1003	55		0	0	5,48			
tipic tosc	17717	1846		28	1,52	10,42			
tipic vend	817	95		4	4,21	11,63			
tosc tradizion	10	0		0	0	0			
tosc uva	44	1		0	0	2,27			
tosc vend	1204	122		4	3,28	10,13			
tosc vin	2932	194		0	0	6,62			
tosc zupp	1575	73		0	0	4,63			
urlo vin	1	0		0	0	0			
vend vin	1004	22		0	0	2,19			
acet balsam glass	6	0		0	0	0			
alimentar cest regal	1154	50		1	2	4,33			
alimentar confezion prodott	752	48		4	8,33	6,38			
alimentar confezion regal	1681	110		5	4,55	6,54			
alimentar pacc regal	821	38		0	0	4,63			
alimentar prodott regal	752	48		4	8,33	6,38			
amar aranc marmell	46	0		0	0	0			
artigianal birr ital	375	17		0	0	4,53			
artigianal birr onlin	295	13		0	0	4,41			
artigianal birr tosc	89	7		0	0	7,87			
artigianal birrific tosc	3	0		0	0	0			
artigianal colomb pasqual	132	0		0	0	0			
artigianal pasqu uov	2213	72		0	0	3,25			
azi cest nataliz	1	0		0	0	0			
aziendal cest regal	22	2		0	0	9,09			
bianc chianin ragù	6	0		0	0	0			
bianc cond tartuf	1	0		0	0	0			
bianc crem tartuf	432	6		0	0	1,39			
bianc tosc vin	93	3		0	0	3,23			
bio cappell past	8	0		0	0	0			
bio cappell senator	8	0		0	0	0			
bio past senator	8	0		0	0	0			
biolog cappell past	26	1		0	0	3,85			
biolog cappell senator	26	1		0	0	3,85			
biolog collin lucches	1	0		0	0	0			
biolog collin oli	1	0		0	0	0			
biolog ital legum	193	11		0	0	5,7			
biolog lucches oli	1	0		0	0	0			
biolog past senator	26	1		0	0	3,85			
biolog tosc vin	17	0		0	0	0			
birr compr onlin	445	26		0	0	5,84			
biscott cioccol fondent	169	4		0	0	2,37			
bosc confettur frutt	100	2		0	0	2			
box food regal	225	19		0	0	8,44			

box prodott tosc	97	14		0	0	14,43			
cac pep pic	1572	50		0	0	3,18			
cac pep sug	1744	35		1	2,86	2,01			
cappell dur farin	1	0		0	0	0			
cappell dur gran	1	0		0	0	0			
cappell dur macin	1	0		0	0	0			
cappell dur pietr	1	0		0	0	0			
cappell dur senator	1	0		0	0	0			
cappell farin gran	1	0		0	0	0			
cappell farin macin	1	0		0	0	0			
cappell farin pietr	1	0		0	0	0			
cappell farin senator	1	0		0	0	0			
cappell gran macin	1	0		0	0	0			
cappell gran pietr	1	0		0	0	0			
cappell gran senator	471	10		0	0	2,12			
cappell macin pietr	1	0		0	0	0			
cappell macin senator	1	0		0	0	0			
cappell past senator	1450	43		0	0	2,97			
cappell pietr senator	1	0		0	0	0			
cappell senator spaghet	130	3		0	0	2,31			
cas cen san	1626	97		0	0	5,97			
cas cen valentin	1626	97		0	0	5,97			
cas menu san	49	2		0	0	4,08			
cas menu valentin	49	2		0	0	4,08			
cas san valentin	1675	99		0	0	5,91			
cec farin macin	2	0		0	0	0			
cec farin pietr	2	0		0	0	0			
cec macin pietr	2	0		0	0	0			
cen san valentin	1626	97		0	0	5,97			
cest gastronom onlin	181	16		2	12,5	8,84			
cest gastronom regal	307	28		1	3,57	9,12			
cest nataliz onlin	8613	680		9	1,32	7,9			
cest nataliz particular	602	39		1	2,56	6,48			
cest regal san	27	4		0	0	14,81			
cest regal valentin	27	4		0	0	14,81			
cest san valentin	27	4		0	0	14,81			
cib regal san	58	1		0	0	1,72			
cib regal valentin	58	1		0	0	1,72			
cib san valentin	175	11		0	0	6,29			
collin lucches oli	1	0		0	0	0			
cond ner tartuf	2	0		0	0	0			
confezion prodott regal	752	48		4	8,33	6,38			
cost crud prosciutt	285	10		0	0	3,51			
crem lard tartuf	1	0		0	0	0			
crud inter prezz	73	0		0	0	0			
crud inter prosciutt	73	0		0	0	0			
crud onlin prosciutt	308	8		0	0	2,6			
crud onlin vend	308	8		0	0	2,6			
crud prezz prosciutt	559	8		0	0	1,43			
crud prosciutt vend	308	8		0	0	2,6			
culinar natal regal	22	1		0	0	4,55			
dur farin gran	2484	42		0	0	1,69			

dur farin macin	1	0		0	0	0			
dur farin pietr	1	0		0	0	0			
dur farin senator	1	0		0	0	0			
dur gran macin	1	0		0	0	0			
dur gran pietr	1	0		0	0	0			
dur gran senator	1	0		0	0	0			
dur macin pietr	1	0		0	0	0			
dur macin senator	1	0		0	0	0			
dur pietr senator	1	0		0	0	0			
ecommerc prodott tipic	339	37		0	0	10,91			
extra oli oliv	215	11		0	0	5,12			
extra oli vergin	215	11		0	0	5,12			
extra oliv vergin	215	11		0	0	5,12			
extravergin ital oli	204	7		0	0	3,43			
extravergin ital oliv	204	7		0	0	3,43			
extravergin oli oliv	491	15		0	0	3,05			
farin gran macin	1	0		0	0	0			
farin gran pietr	1	0		0	0	0			
farin gran senator	1	0		0	0	0			
farin gran tener	3187	43		0	0	1,35			
farin macin pietr	114	28		0	0	24,56			
farin macin senator	1	0		0	0	0			
farin pietr senator	1	0		0	0	0			
farr legum zupp	45	1		0	0	2,22			
fest ide mamm	3	0		0	0	0			
fest ide regal	3	0		0	0	0			
fest mamm regal	206	2		0	0	0,97			
food salum slow	241	22		0	0	9,13			
formagg onlin tosc	387	27		0	0	6,98			
formagg onlin vend	387	27		0	0	6,98			
formagg tosc vend	387	27		0	0	6,98			
fung linguin porcin	25	1		0	0	4			
fung porcin sug	1148	9		0	0	0,78			
gastronom ide regal	820	47		1	2,13	5,73			
gastronom natal regal	176	10		0	0	5,68			
gourmet san valentin	267	14		0	0	5,24			
gran macin pietr	1	0		0	0	0			
gran macin senator	1	0		0	0	0			
gran pietr senator	1	0		0	0	0			
ide mamm regal	3	0		0	0	0			
ide regal san	67	1		0	0	1,49			
ide regal valentin	67	1		0	0	1,49			
ide san valentin	67	1		0	0	1,49			
inter prezz prosciutt	73	0		0	0	0			
ital oli oliv	204	7		0	0	3,43			
ital onlin prodott	233	11		0	0	4,72			
ital onlin tipic	233	11		0	0	4,72			
ital prodott tipic	233	11		0	0	4,72			
limon linguin pep	3	0		0	0	0			
limon marmell zenzer	1	0		0	0	0			
lup urlo vin	1	0		0	0	0			
macin pietr senator	1	0		0	0	0			

menu san valentin	49	2		0	0	4,08			
miel produzion tosc	2	0		0	0	0			
ner oliv pat	143	7		0	0	4,9			
ner sepp tagliatell	90	3		0	0	3,33			
oli oliv vergin	215	11		0	0	5,12			
onlin prodott shop	952	118		1	0,85	12,39			
onlin prodott tipic	2998	332		8	2,41	11,07			
onlin prodott tosc	8006	961		13	1,35	12			
onlin prodott vend	817	95		4	4,21	11,63			
onlin prosciutt vend	308	8		0	0	2,6			
onlin salum vend	2374	130		2	1,54	5,48			
onlin shop tosc	952	118		1	0,85	12,39			
onlin tipic tosc	2765	321		8	2,49	11,61			
onlin tipic vend	817	95		4	4,21	11,63			
onlin tosc vend	1204	122		4	3,28	10,13			
onlin vend vin	1004	22		0	0	2,19			
pecorin stagion tosc	291	14		0	0	4,81			
preg tosc vin	941	57		0	0	6,06			
prodott regal tipic	1064	128		3	2,34	12,03			
prodott regal tosc	1064	128		3	2,34	12,03			
prodott shop tosc	952	118		1	0,85	12,39			
prodott tipic tosc	17577	1829		28	1,53	10,41			
prodott tipic vend	817	95		4	4,21	11,63			
prodott tosc tradizion	10	0		0	0	0			
prodott tosc vend	817	95		4	4,21	11,63			
regal san valentin	152	6		0	0	3,95			
regal tipic tosc	1064	128		3	2,34	12,03			
ros tosc vin	23	1		0	0	4,35			
ross tosc vin	371	22		0	0	5,93			
sapor tipic tosc	140	17		0	0	12,14			
tipic tosc vend	817	95		4	4,21	11,63			

Risultati completi del calcolo del valore di Shapley in base alla metrica

CTR

	Player	Shapley_value
1	decortic	62,11
2	caseif	47,05
3	maiones	11,02
4	castagn	9,81
5	bazzon	4,39
6	pienz	3,83
7	befan	3,16
8	box	3,03
9	past	2,96
10	shop	2,95
11	aziendal	2,64
12	sapor	2,61
13	cest	2,33
14	souven	2,04
15	farin	1,75
16	tipic	1,64
17	penitent	1,52
18	pistacc	1,38
19	slow	1,31
20	lepr	1,27
21	garfagnan	1,26
22	piccant	1,17
23	prodott	1,13
24	acac	1,06
25	tosc	1,04
26	compr	0,95
27	lucches	0,95
28	artigianal	0,92
29	selvaggin	0,78
30	ecommerc	0,78
31	food	0,72
32	farr	0,72
33	pietr	0,64
34	linguin	0,61
35	cioccol	0,61
36	salam	0,51
37	cinghial	0,49
38	extra	0,43
39	vergin	0,43
40	birrif	0,43
41	legum	0,37
42	ner	0,31
43	valentin	0,31
44	gastronom	0,30
45	mor	0,30
46	onlin	0,29

47	balsam	0,29
48	crem	0,28
49	san	0,27
50	stes	0,26
51	inter	0,26
52	tartin	0,26
53	sepp	0,23
54	fung	0,23
55	porcin	0,23
56	oliv	0,23
57	liquor	0,21
58	vend	0,18
59	nataliz	0,18
60	cen	0,15
61	pic	0,07
62	cas	0,04
63	cost	0,03
64	sal	0,02
65	bosc	0,01
66	frutt	0,01
67	bresaol	0,00
68	confettur	0,00
69	genoves	0,00
70	manz	0,00
71	pan	0,00
72	papp	0,00
73	pest	0,00
74	pomodor	0,00
75	pozz	0,00
76	puccin	0,00
77	affumic	-0,01
78	veloc	-0,04
79	pecor	-0,04
80	carn	-0,09
81	salum	-0,12
82	marmell	-0,13
83	extravergin	-0,16
84	birr	-0,17
85	cac	-0,20
86	macin	-0,20
87	pep	-0,21
88	biolog	-0,23
89	ros	-0,25
90	cannellin	-0,29
91	prosciutt	-0,29
92	sant	-0,30
93	natal	-0,32
94	ross	-0,38
95	tagliatell	-0,39
96	regal	-0,40
97	millefior	-0,41
98	oli	-0,42

99	spaghet	-0,42
100	fettuccin	-0,42
101	fondent	-0,46
102	sug	-0,48
103	arrotol	-0,48
104	ortic	-0,48
105	gourmet	-0,49
106	special	-0,50
107	fich	-0,51
108	preg	-0,54
109	pecorin	-0,54
110	morbid	-0,59
111	cantucc	-0,61
112	vin	-0,61
113	formagg	-0,63
114	tartuf	-0,67
115	biscott	-0,76
116	particular	-0,81
117	mondiol	-0,83
118	pat	-0,83
119	lucc	-0,88
120	confezion	-0,88
121	fresc	-0,89
122	pacc	-0,95
123	pasqu	-0,98
124	tener	-0,99
125	ragù	-0,99
126	frollin	-1,04
127	stagion	-1,05
128	acet	-1,07
129	tranc	-1,10
130	miel	-1,11
131	cib	-1,16
132	alimantar	-1,17
133	chiar	-1,18
134	copp	-1,18
135	bio	-1,19
136	peperoncin	-1,22
137	uov	-1,24
138	caff	-1,25
139	pancett	-1,32
140	limon	-1,41
141	chianin	-1,42
142	uva	-1,48
143	eucalipt	-1,49
144	gioves	-1,49
145	girasol	-1,50
146	amar	-1,67
147	aranc	-1,67
148	zenzer	-1,67
149	giall	-1,69
150	lard	-1,69

151	rustic	-1,71
152	menu	-1,74
153	antic	-1,76
154	lup	-1,79
155	urlo	-1,79
156	ital	-1,83
157	sals	-1,86
158	noc	-1,88
159	culinar	-1,91
160	verdur	-1,92
161	ide	-1,93
162	bianc	-2,02
163	cond	-2,16
164	glass	-2,34
165	fest	-2,62
166	mamm	-2,62
167	zupp	-2,64
168	colomb	-2,73
169	nocciol	-3,09
170	crostin	-3,12
171	dole	-3,23
172	antipast	-3,26
173	crud	-3,26
174	filett	-3,39
175	mandorl	-3,53
176	produzion	-3,56
177	torron	-3,57
178	azi	-3,65
179	collin	-3,66
180	fagiol	-3,66
181	cuor	-3,74
182	buccell	-4,05
183	cucin	-4,10
184	pasqual	-4,14
185	ambrat	-4,34
186	gran	-4,46
187	senator	-4,49
188	cappell	-4,49
189	secc	-5,27
190	dur	-5,36
191	biodinam	-5,36
192	prezz	-6,45
193	birrific	-7,31
194	biadin	-7,89
195	tradizion	-9,10
196	cec	-17,52

Risultati completi del calcolo del valore di Shapley in base alla metrica

CVR

	Player	Shapley_value
1	verdur	98,15
2	cinghial	5,76
3	bazzon	4,07
4	pat	3,43
5	cantucc	3,14
6	biscott	2,88
7	confezion	2,02
8	antipast	1,84
9	gastronom	1,50
10	fresc	1,33
11	alimantar	1,27
12	prodott	1,04
13	stes	1,03
14	fagiol	0,94
15	particolar	0,89
16	vend	0,86
17	farr	0,82
18	salam	0,82
19	special	0,75
20	sug	0,69
21	tipic	0,47
22	ross	0,46
23	onlin	0,40
24	pienz	0,15
25	regal	0,13
26	cac	0,11
27	pep	0,06
28	cest	0,06
29	farin	0,05
30	pecor	0,01
31	bio	0,01
32	ide	0,01
33	acac	0,00
34	acet	0,00
35	ambrat	0,00
36	antic	0,00
37	balsam	0,00
38	biadin	0,00
39	biodinam	0,00
40	bosc	0,00
41	bresaol	0,00
42	caff	0,00
43	carn	0,00
44	cas	0,00
45	cen	0,00
46	chiar	0,00

47	collin	0,00
48	confettur	0,00
49	copp	0,00
50	decortic	0,00
51	eucalipt	0,00
52	extra	0,00
53	extravergin	0,00
54	fich	0,00
55	filett	0,00
56	frutt	0,00
57	genoves	0,00
58	gioves	0,00
59	girasol	0,00
60	glass	0,00
61	gourmet	0,00
62	lucc	0,00
63	lup	0,00
64	manz	0,00
65	marmell	0,00
66	menu	0,00
67	millefior	0,00
68	mondiol	0,00
69	mor	0,00
70	noc	0,00
71	ortic	0,00
72	pan	0,00
73	papp	0,00
74	peperoncin	0,00
75	pest	0,00
76	pomodor	0,00
77	pozz	0,00
78	puccin	0,00
79	rustic	0,00
80	sal	0,00
81	sant	0,00
82	secc	0,00
83	sepp	0,00
84	spaghatt	0,00
85	tagliatell	0,00
86	urlo	0,00
87	veloc	0,00
88	vergin	0,00
89	macin	-0,01
90	gran	-0,01
91	pietr	-0,01
92	salum	-0,01
93	dur	-0,02
94	ros	-0,04
95	cappell	-0,05
96	senator	-0,05
97	cec	-0,06
98	past	-0,07

99	frollin	-0,09
100	torron	-0,09
101	biolog	-0,09
102	oli	-0,10
103	tener	-0,11
104	nataliz	-0,13
105	pic	-0,14
106	tartuf	-0,14
107	birr	-0,14
108	amar	-0,15
109	aranc	-0,15
110	zenzer	-0,15
111	castagn	-0,17
112	prosciutt	-0,20
113	uov	-0,22
114	tosc	-0,24
115	limon	-0,27
116	birrif	-0,28
117	oliv	-0,28
118	pancett	-0,28
119	slow	-0,28
120	pasqual	-0,31
121	cib	-0,32
122	compr	-0,32
123	pasqu	-0,33
124	giall	-0,33
125	ragù	-0,34
126	cond	-0,34
127	shop	-0,35
128	fung	-0,35
129	porcin	-0,35
130	liquor	-0,37
131	preg	-0,37
132	produzion	-0,37
133	uva	-0,37
134	artigianal	-0,38
135	inter	-0,39
136	prezz	-0,39
137	crud	-0,39
138	birrific	-0,40
139	vin	-0,40
140	miel	-0,42
141	colomb	-0,42
142	san	-0,45
143	bianc	-0,45
144	food	-0,46
145	linguin	-0,46
146	garfagnan	-0,47
147	ital	-0,48
148	stagion	-0,49
149	valentin	-0,49
150	ner	-0,49

151	lard	-0,51
152	natal	-0,52
153	legum	-0,54
154	sals	-0,55
155	tartin	-0,55
156	arrotol	-0,58
157	pecorin	-0,59
158	maiones	-0,60
159	chianin	-0,62
160	crem	-0,66
161	piccant	-0,68
162	cucin	-0,70
163	lucches	-0,77
164	cost	-0,78
165	affumic	-0,83
166	azi	-0,84
167	box	-0,91
168	cuor	-0,93
169	penitent	-0,93
170	tranc	-0,93
171	fest	-1,01
172	mamm	-1,01
173	formagg	-1,03
174	buccell	-1,10
175	caseif	-1,10
176	crostin	-1,10
177	dolc	-1,10
178	tradizion	-1,19
179	culinar	-1,23
180	zupp	-1,24
181	sapor	-1,37
182	ecommerc	-1,41
183	fondent	-1,67
184	selvaggin	-1,77
185	lepr	-1,85
186	pistacc	-2,03
187	cioccol	-2,19
188	aziendal	-2,20
189	cannellin	-2,33
190	pacc	-2,65
191	souven	-3,00
192	mandorl	-3,77
193	morbid	-3,77
194	fettuccin	-3,85
195	befan	-5,00
196	nocciol	-5,00

8. Bibliografia

Eric Enge, Stephan Spencer, Jessie C. Stricchiola. *The art of SEO, Mastering Search Engine Optimization*. Sebastopol, O'Really Media, 2015

Giorgio Taverniti. *SEO POWER, Strategie e strumenti per essere visibili sui motori di ricerca*. Milano, Ulrico Hoepli Editore, 2013

Marco Maltraversi. *SEO e SEM, Guida avanzata al web marketing*. Milano, Edizioni FAG, 2011

9. Sitografia

Google Search Central, voce *Guida introduttiva all'ottimizzazione per i motori di ricerca (SEO)*, <https://developers.google.com/search/docs/beginner/seo-starter-guide?hl=it> (visitato il 20 settembre 2021)

Google Search Central, voce *Creare titoli e snippet utili per i risultati della Ricerca*, <https://developers.google.com/search/docs/advanced/appearance/good-titles-snippets?hl=it> (visitato il 20 settembre 2021)

W3Schools, voce *HTML <title> tag*, [https://www.w3schools.com/tags/tag_title.asp#:~:text=The%20tag%20defines%20the,search%20engine%20optimization%20\(SEO\)](https://www.w3schools.com/tags/tag_title.asp#:~:text=The%20tag%20defines%20the,search%20engine%20optimization%20(SEO)) (visitato il 20 settembre 2021)

Wikipedia, voce *Conversion Rate* https://it.wikipedia.org/wiki/Conversion_rate (visitato il 21 settembre 2021)

Wikipedia, voce *Click Trough rate*, https://it.wikipedia.org/wiki/Click-through_rate (visitato il 21 settembre 2021)

Wikipedia, voce *Pay per click*, <https://en.wikipedia.org/wiki/Pay-per-click> (visitato il 21 settembre 2021)

Wikipedia, voce *Cost per mille*, https://it.wikipedia.org/wiki/Cost_per_mille (visitato il 21 settembre 2021)

Wikipedia, voce *Cost per action*, https://it.wikipedia.org/wiki/Cost_per_action (visitato il 21 settembre 2021)

Google Ads, voce *Come creare un annuncio*,
https://ads.google.com/intl/it_it/home/resources/come-creare-un-annuncio-online-con-google/ (visitato il 22 settembre 2021)

Guida di Google My Business,
<https://support.google.com/business/answer/6300665?hl=it> (visitato il 22 settembre 2021)

Wikipedia, voce *Search Engine Marketing*,
https://it.wikipedia.org/wiki/Search_engine_marketing (visitato il 22 settembre 2021)

Guida di Google Ads, voce *Informazioni sulle opzioni di corrispondenza delle parole chiave*, <https://support.google.com/google-ads/answer/7478529> (visitato il 30 settembre 2021)

Guida Google Ads, voce *Informazioni sullo stato delle parole chiave*,
<https://support.google.com/google-ads/answer/2453978?hl=it> (visitato il 30 settembre 2021)

Guida di Analytics, voce *Panoramica dei modelli di attribuzione*,
<https://support.google.com/analytics/answer/1662518?hl=it> (visitato il 4 ottobre 2021)

Guida di Analytics, voce *Metodologia dell'attribuzione basata sui dati*,
<https://support.google.com/analytics/answer/3191594> (visitato il 4 ottobre 2021)

Shapley Value Attribution Modeling,

<https://www.kaggle.com/jasonbrewster/shapley-value-attribution-modeling/notebook>, (visitato il 7 ottobre 2021)

Wikipedia, voce *Shapley Value*, https://en.wikipedia.org/wiki/Shapley_value (visitato il 7 ottobre 2021)

StatCounter Global Stats, aggiornato a dicembre 2021,

<https://gs.statcounter.com/search-engine-market-share> (visitato il 28 dicembre 2021)

9.1 Articoli

Solntseva, Daryana. 23 settembre 2020. *Che cos'è la SEO off-page? Una guida completa*. In “Semrush Blog”, <https://it.semrush.com/blog/seo-off-page-guida-completa/> (visitato il 23 settembre 2021)

Anindya Ghose, Vilma Todri. *Towards a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior*, settembre 2015, <https://bit.ly/3FTTJ1A> (visitato il 5 ottobre 2021)

Saeed R. Bagheri, Seyed Hanif Mahboobi, Kaifeng Zhao. *Shapley Value Methods for Attribution Modeling in Online Advertising*, <https://arxiv.org/ftp/arxiv/papers/1804/1804.05327.pdf> (visitato il 5 ottobre 2021)

Mancini, Gennaro. 9 marzo 2019. *Che cos'è una keyword e quali sono le parole chiave per la SEO*. In “Il Blog di SEOZoom”, <https://www.seozoom.it/keyword-parole-chiave-seo/> (visitato il 13 ottobre 2021)

Mancini, Gennaro. 24 febbraio 2020. *Keyword research, l'attività fondamentale per una strategia efficace*. In “Il Blog di SEOZoom”, <https://www.seozoom.it/keyword-research-seozoom-seo-copywriting/> (visitato il 13 ottobre 2021)

Haleua, Chris. 1° dicembre 2017. *Algorithmic Attribution: Choosing the Attribution Model That's Right for Your Company*. In “Adobe blog”, <https://blog.adobe.com/en/publish/2017/01/12/algorithmic-attribution-choosing-attribution-model-thats-right-company#gs.leh3i2> (visitato il 15 ottobre 2021)

Choudhary, Ankit. 25 novembre 2019. *A Unique Method for Machine Learning Interpretability: Game Theory & Shapley Values!* In “Blog di Analytics Vidhya”, <https://www.analyticsvidhya.com/blog/2019/11/shapley-value-machine-learning-interpretability-game-theory/> (visitato il 20 ottobre 2021)

Yuan, Jacky. 6 luglio 2020. *Multi-Touch Attribution Marketing Model - The Shapley value approach*. In “Medium.com”, <https://medium.com/analytics-vidhya/the-shapley-value-approach-to-multi-touch-attribution-marketing-model-e345b35f3359> (visitato il 20 ottobre 2021)

Affane, Reda. 11 agosto 2020. *Step Up Your Marketing Attribution With Game Theory*. In “Blog di dataiku”, <https://blog.dataiku.com/step-up-your-marketing-attribution-with-game-theory> (visitato il 20 ottobre 2021)

Sweeney, Michael. 25 novembre 2020. *Game Theory Attribution: The Model You've Probably Never Heard Of*. In “The Clearcode Blog”, <https://clearcode.cc/blog/game-theory-attribution/> (visitato il 23 ottobre 2021)

Wild, Cody Marie. 13 gennaio 2018. *One Feature Attribution Method to (Supposedly) Rule Them All: Shapley Values*. In “Towards Data Science”, <https://towardsdatascience.com/one-feature-attribution-method-to-supposedly-rule-them-all-shapley-values-f3e04534983d> (visitato il 23 ottobre 2021)

Kinley, James. 12 novembre 2019. *Data-Driven Marketing Attribution*. In “Towards Data Science”, <https://towardsdatascience.com/data-driven-marketing-attribution-1a28d2e613a0> (visitato il 23 ottobre 2021)