

POLITECNICO di TORINO

Collegio di Ingegneria Gestionale e della Produzione
Corso di Laurea Magistrale in Ingegneria Gestionale



Politecnico di Torino

Tesi di Laurea Magistrale

*Strumenti per il quality tracking attraverso l'analisi della Digital Voice of
Customer*

Relatore: Federico Barravecchia
Correlatore: Luca Mastrogiacomo

Candidato: Simone Bergesio
Matricola: S298535

Dicembre 2023

INDICE

INDICE DELLE FIGURE	5
INDICE DELLE TABELLE.....	5
INTRODUZIONE	7
CAPITOLO 1: DIGITAL VOC E METRICHE DI QUALITÀ	10
VOICE OF CUSTOMER	10
ANALISI DELLA DIGITAL VOC	12
TOPIC MODELING	13
STM.....	13
METODOLOGIA DI ANALISI APPLICATA NELLO STUDIO	19
ESTRAZIONE DEL SET DI DATI	19
PRE-PROCESSAMENTO	20
SELEZIONE DEL NUMERO OTTIMALE DI TOPIC.....	21
ETICETTATURA DEI DATI	21
VERIFICA DEI DATI.....	23
ANALISI DEI RISULTATI	30
MTP- Mean Topic Proportion	31
MRP-Mean Rating Proportion	31
IMTP - Interval Mean Topical Prevalence.....	33
CAPITOLO 2: APPLICAZIONE CASO STUDIO: SPOTIFY	36
INTRODUZIONE	36
METODOLOGIA	36
ESTRAZIONE DEL SET DI DATI	36
PRE-PROCESSAMENTO.....	38
SELEZIONE DEL NUMERO OTTIMALE DI TOPIC.....	39
ETICETTATURA DEI DATI	40
VERIFICA DEI DATI.....	42
ANALISI DEI RISULTATI	44
MTP-Mean Topic Proportion.....	44
MRP-Mean Rating Proportion	45
IMTP-Interval Mean Topical Prevalence	47
CAPITOLO 3: MONITORAGGIO NEL TEMPO DELLA RELAZIONE TRA DETERMINANTI DI QUALITÀ E VALUTAZIONI DEI CLIENTI	49
PROFILE MONITORING	49
CARTE DI CONTROLLO.....	50
CARTE X-S	52
CARTE P	53
TEST DI GRUBBS.....	54
TEST DI CORRELAZIONE DI SPEARMAN	56
CONFRONTO E ANALISI CRITICA DELLE METODOLOGIE DI ANALISI.....	57
ANALISI IMTP	59
CARTE X-S	59
CARTE P.....	67
ANALISI MRP	72
CARTE DI CONTROLLO.....	75
TEST DI GRUBBS.....	77

TEST DI CORRELAZIONE DI SPEARMAN	85
RELAZIONE TRA MRP MEDI E VALUTAZIONI CLIENTI	87
CONCLUSIONI	89
RIFERIMENTI	92
APPENDICE	97
CARTE X IMTP	97
CARTE S IMTP	98

INDICE DELLE FIGURE

Figura 1. Funzionamento di un algoritmo di topic modeling (Barravecchia et al., 2022).....	15
Figura 2. Sotto modelli successivi del STM (Roberts et al., 2016).....	16
Figura 3. Profili MRP rispettivamente negativo, neutro e positivo (Barravecchia et al., 2020).....	33
Figura 4. Profili IMTP.....	35
Figura 5. Relazione tra livello di rating e numero delle recensioni.....	38
Figura 6. Applicazione del pre-processing con riduzione del database iniziale.....	38
Figura 7. Relazione tra il valore dell'Held-Out Likelihood e il numero di topic.....	40
Figura 8. Livello di discussione medio dei topic (MTP).....	45
Figura 9. MRP per ogni topic sulla base del rating e del livello di discussione.....	46
Figura 10. IMTP nel corso dei mesi.....	48
Figura 11. Aggiornamenti di Spotify rilasciati nei mesi analizzati.....	48
Figura 12. Carta di controllo X per il topic 2.....	63
Figura 13. Carta di controllo X per il topic 5.....	64
Figura 14. Carta di controllo X per il topic 9.....	65
Figura 15. Carta di controllo S topic 2.....	66
Figura 16. Carta di controllo S topic 5.....	67
Figura 17. Carta di controllo p per il topic 1.....	68
Figura 18. Carte di controllo p per i topic dal 2 all'11.....	71
Figura 19. Rating delle recensioni nel tempo.....	74
Figura 20. Test di Grubbs nel mese di gennaio.....	78
Figura 21. Test di Grubbs nel mese di febbraio.....	79
Figura 22. Test di Grubbs nel mese di marzo.....	80
Figura 23. Test di Grubbs nel mese di aprile.....	81
Figura 24. Test di Grubbs nel mese di maggio.....	82
Figura 25. Test di Grubbs nel mese di giugno.....	83
Figura 26. Test di Grubbs nel mese di luglio.....	84
Figura 27. Ranghi di Spearman nel tempo.....	85
Figura 28. Valori MRP medi nel tempo.....	88
Figura 29. Carte X IMTP.....	98
Figura 30. Carte S IMTP.....	100

INDICE DELLE TABELLE

Tabella 1. Matrice di topical prevalence.....	17
Tabella 2. Matrice di topical content.....	18
Tabella 3. Assegnazione manuale dei topic alle recensioni.....	24
Tabella 4. Scenari individuati dal confronto tra assegnazione automatica dei topic e assegnazione manuale.....	26
Tabella 5. Valori di riferimento degli indicatori di validazione dell'algoritmo di topic modeling (Mastrogiacomo et al., 2021).....	30
Tabella 6. Cardinalità e percentuale di recensioni con uno specifico livello di rating.....	37
Tabella 7. Etichettatura finale dei topic con rispettive keywords associate.....	41
Tabella 8. Primi valori ottenuti dalla validazione dei risultati.....	43
Tabella 9. Valori ottenuti dalla validazione finale.....	43
Tabella 10. Valori MTP per ogni topic.....	44
Tabella 11. Valori MRP per ogni topic.....	46
Tabella 12. IMTP per ogni topic nel tempo.....	48

Tabella 13. Prospetto riassuntivo metodologie	57
Tabella 14. IMTP per mesi e per topic.	60
Tabella 15. Cardinalità delle recensioni per mesi.....	61
Tabella 16. Costruzione delle carte X.....	62
Tabella 17. Costruzioni delle carte di controllo S.	65
Tabella 18. Vittorie per ogni topic per ogni mese nella costruzione delle carte p.	67
Tabella 19. Costruzione carta p del topic 1.....	68
Tabella 20. Recensioni mese per mese con rispettivi livelli di rating,	73
Tabella 21. Costruzione carta X MRP.	76
Tabella 22. Valori test di Grubbs nel mese di gennaio.....	77
Tabella 23. Valori test di Grubbs nel mese di febbraio.	78
Tabella 24. Valori test di Grubbs nel mese di marzo.....	79
Tabella 25. Valori test di Grubbs nel mese di aprile.	81
Tabella 26. Valori test di Grubbs nel mese di maggio.....	81
Tabella 27. Valori test di Grubbs nel mese di giugno.	82
Tabella 28. Valori test di Grubbs nel mese di luglio.	83
Tabella 29. Ranghi di Spearman.....	85
Tabella 30. Valori MRP medio nel tempo.	87

INTRODUZIONE

In un contesto aziendale in continua trasformazione, in cui la customer experience è al centro dell'attenzione, l'ascolto attivo e attento della Voice-of-Customer, ovvero della “voce del cliente”, è diventato un approccio imprescindibile. Le imprese mirano a fornire prodotti e servizi di qualità per differenziarsi in un mercato sempre più competitivo, influenzato dall'evoluzione tecnologica. In particolare, con l'avvento del processo di servitizzazione (Wendermerwe e Rada, 1988), sovente il prodotto non viene più venduto singolarmente, ma in combinazione con un servizio (Ulaga, 2008), comportando un cambio significativo di prospettiva. Il controllo della qualità effettuato dalle aziende non si può più basare esclusivamente su fattori oggettivi, come quelli legati al contesto produttivo, che classificano gli elementi in base della difettosità (Fisher e Nair, 2006). Al contrario, inizia ad assumere un ruolo fondamentale il cliente e soprattutto la soddisfazione percepita in seguito all'utilizzo di un determinato servizio. Si introducono fattori soggettivi come i bisogni, le aspettative e i gusti personali, che variano rapidamente nel corso del tempo.

In questa circostanza, emergono complessità connesse alla quantificazione di tali variabili e, di conseguenza, alla valutazione della qualità del servizio offerto (Franceschini, 2001). Al fine di affrontare tali questioni, è stata adottata un'approfondita interazione con il consumatore finale attraverso la conduzione di interviste e la somministrazione di questionari.

Con l'avanzamento della tecnologica, il consumatore finale ha la possibilità di rilasciare recensioni sul web, trasmettendo quello che pensa e esprimendo la sua “voce”, la cosiddetta “Voice-of-Customer”.

Il cliente dispone di sempre più informazioni rispetto al passato, si documenta di più, diventa più critico e esigente. Monitorare nel tempo le sue aspettative e la sua soddisfazione risulta essere uno dei principali fattori critici di successo per le imprese.

L'obiettivo della tesi è di identificare gli aspetti chiave nella fruizione di un servizio, ovvero gli elementi determinanti per la soddisfazione del cliente finale e di come questi variano nel corso del tempo. Si intende esplorare la Voice-of-Customer (VoC) (Jach et al., 2021), che include le aspettative, le preferenze e i bisogni del consumatore che usufruisce del servizio (Gaskin et al., 2020). Per l'analisi, l'attenzione è focalizzata sui contenuti generati dagli utenti nel web, i così detti User-Generated Content (UGC) (Hopkins, 2013). Questo termine si riferisce ai contenuti creati e pubblicati spontaneamente dagli utenti in rete, come post nei blog, discussioni nei forum, commenti sui social network, immagini e video digitali per far emergere la propria opinione su un servizio di cui hanno usufruito. Nell'analisi proposta, si esaminano le recensioni rilasciate dagli utenti su un determinato caso di studio. Dopo l'esplorazione della VoC, e in particolare della digital VoC (con “digital” per

sottolineare l'aspetto digitale e online dei contenuti analizzati), si procede a trasformarla in metriche di qualità e in strumenti in grado di misurare nel tempo la relazione tra le determinanti di qualità e le valutazioni dei clienti.

Per effettuare l'analisi, si utilizza la tecnica del *topic modeling*, che sfrutta algoritmi di *machine learning* non supervisionati per individuare gli argomenti (topic) latenti presenti in una raccolta di documenti non strutturati (Muller et al., 2016). Da un grande set di documenti, questi algoritmi identificano un insieme di topic che descrivono accuratamente gli argomenti trattati all'interno del corpus testuale.

La scelta di utilizzare un algoritmo di *topic modeling* si basa sull'assunzione che i clienti valutino un servizio secondo criteri comuni per valutarne la qualità. Tali criteri emergono quando si rilasciano delle recensioni. Essendo elementi chiave per comprendere la qualità di un servizio, questi topic comuni più discussi, sono definiti come determinanti di qualità per uno specifico servizio.

Un obiettivo dell'analisi è individuare le determinanti di qualità tipiche del servizio e monitorarle nel tempo. Per osservare la loro evoluzione, si utilizzano degli indicatori, tra cui il Mean Rating Proportion (MRP) (Barravecchia et al., 2020) che spiega quanto è discusso un topic nei record di digital VoC associati con uno specifico livello di rating. I clienti del servizio, oltre a rilasciare recensioni testuali, attribuiscono generalmente un punteggio da 1 a 5 (dove 1 indica la minima soddisfazione percepita e 5 la massima). L'MRP consente di individuare se un topic è prevalentemente associato a recensioni con valutazioni positive, neutre o negative, e di conseguenza, come tale caratteristica viene percepita dai clienti.

La domanda di ricerca principale di questa tesi è la seguente:

- Esiste un metodo efficace per tracciare nel tempo l'andamento dell'indicatore MRP e identificare eventuali anomalie nei suoi valori?

In particolare, l'obiettivo è individuare uno strumento statistico che funzioni come “campanello d'allarme” e che sia in grado di individuare situazioni anomale. Questo strumento deve essere in grado di segnalare al management aziendale eventuali livelli di discussione anomali di un determinato topic per uno specifico livello di rating. Ad esempio, se un particolare argomento presenta valori che si discostano da quelli attesi per livelli di rating associati ad una bassa soddisfazione percepita, lo strumento deve essere in grado di rilevare tale anomalia nei dati. L'informazione ottenuta può essere

utilizzata come input per ottimizzare i processi aziendali esistenti o per riprogettarli al fine di migliorare l'efficienza complessiva.

Per condurre la ricerca, inizialmente si utilizzano strumenti di monitoraggio comuni come le carte di controllo (Shewart, 1924), successivamente si adottano tecniche meno convenzionali, tra cui l'analisi dei ranghi di Spearman (Spearman, 1904) e il test di Grubbs (Grubbs, 1950), attraverso le quali si esamina il comportamento dell'MRP. In particolare, il test di Grubbs è stato utilizzato per identificare i valori outliers che si discostano dalla distribuzione attesa dei dati. Quando si verifica questa situazione, viene attivato il "campanello d'allarme" che fa emergere dove si colloca l'eventuale problematica percepita dai clienti nell'usufruire del servizio.

L'organizzazione dell'elaborato prevede la suddivisione in tre capitoli:

1. Il primo capitolo descrive la digital VoC e presenta la metodologia adottata nel corso dello studio.
2. Nel secondo capitolo, la metodologia descritta nel primo si applica nel caso di studio Spotify. Si individuano le determinanti di qualità tipiche del servizio e si monitorano nel corso del tempo attraverso il calcolo di indicatori e metriche presentati nel primo capitolo.
3. Il terzo capitolo si focalizza sulla risposta alla domanda di ricerca iniziale. Si presentano le metodologie utilizzate dal punto di vista teorico e le si applicano al caso studio. In particolare, si presenta la soluzione proposta per l'analisi dell'indicatore MRP e per l'individuazione di valori anomali per il processo esaminato.

In conclusione, la tesi propone un'approfondita esplorazione dell'utilizzo della digital Voice-of-Customer come strumento di input per la definizione delle determinanti di qualità di un servizio e per il loro monitoraggio nel tempo. Attraverso una metodologia basata su algoritmi di *topic modeling* si possono analizzare dati e costruire indicatori personalizzati con l'obiettivo finale di migliorare la qualità dei prodotti o dei servizi offerti e per far percepire un livello di qualità stesso sempre più elevato per il cliente finale accrescendone la soddisfazione.

CAPITOLO 1: DIGITAL VOC E METRICHE DI QUALITÀ

VOICE OF CUSTOMER

Attualmente, risulta imperativo per qualsiasi organizzazione riconoscere e identificare le necessità, le preferenze, le richieste, gli obiettivi e le aspettative della propria clientela relativamente a un prodotto o servizio fornito. A tal fine, si fa ricorso alla Voice-of-Customer (VoC) e, in particolare, nell'era dell'evoluzione tecnologica, alla digital Voice-of-Customer (digital VoC). Quest'ultima si configura come il complesso di feedback e recensioni diffusi online dagli utenti che hanno fruito del prodotto o servizio in questione (Gaskin et al., 2020).

La digital VoC si presenta come una risorsa cruciale per le aziende, permettendo la gestione della qualità dei prodotti o servizi forniti e offrendo una visione completa delle percezioni dei clienti. Nel contesto produttivo, la gestione della qualità si basa su criteri connessi alla fisicità dei prodotti, quali la presenza di pezzi difettosi e fatti oggettivi. Al contrario, nel settore dei servizi, la qualità si fonda principalmente sulla percezione dei clienti, sottolineando l'importanza fondamentale della digital VoC. Il cliente assume un ruolo centrale nella fornitura dei servizi e nella definizione della qualità in un contesto prettamente orientato ai servizi.

La Voice-of-Customer (VoC) può essere considerata di primaria importanza nell'analisi della customer experience poiché abilita le aziende a comprendere la propria clientela, adattando di conseguenza i propri prodotti e servizi per soddisfarla. La soddisfazione del cliente, secondo la definizione dell'International Organization for Standardization (2015), costituisce un parametro soggettivo che riflette la percezione del grado di soddisfazione delle proprie aspettative.

Il ruolo cruciale svolto dalla soddisfazione del cliente per l'azienda fornitrice si manifesta nella capacità di instaurare relazioni durature e fedeli con la clientela esistente, oltre a generare recensioni positive e passaparola, divenendo un mezzo efficace per attirare nuovi clienti (Biesok e Wyrod-Wrobel, 2011).

La VoC riveste un ruolo di rilievo anche nell'ambito del marketing aziendale. Essa rappresenta uno strumento essenziale per la gestione della qualità, permettendo di identificare i fattori chiave, noti come determinanti di qualità, che influenzano l'esperienza del cliente. Attraverso il monitoraggio e il controllo delle percezioni di qualità dei clienti, è possibile sviluppare metodologie e soluzioni finalizzate al miglioramento continuo della percezione di prodotti o servizi nel corso del tempo.

La VOC può essere categorizzata in due tipologie:

- *Tradizionale*, questa modalità di VoC utilizza fonti quali questionari, interviste, sondaggi e focus group per raccogliere informazioni. Tuttavia, queste metodologie di indagine presentano limitazioni poiché richiedono un trade-off tra tempi e costi. L'arricchimento della ricerca e l'aumento del dettaglio comportano un notevole incremento in termini temporali e finanziari (Guo et al., 2016). L'aggiornamento delle informazioni con questa metodologia avviene periodicamente e in modo guidato, pertanto non è automatico. L'analisi dei dati raccolti in questo contesto impiega strumenti di analisi statistica.

Un aspetto critico è rappresentato dalle dimensioni del campione, vincolate e proporzionate alle risorse allocate per la raccolta dei dati.

- *Digitale*, la modalità di VoC impiegata durante lo sviluppo di questo lavoro, che supera le limitazioni associate al metodo tradizionale. Questa tipologia di VoC è più innovativa ed è resa possibile dall'evoluzione tecnologica e dal ruolo sempre più centrale giocato da internet nella quotidianità delle persone. In particolare, con l'avvento del Web 2.0 (O'Reilly, 2005) gli utenti del web hanno la possibilità di interagire e modificare i contenuti delle pagine internet, dando origine ai cosiddetti User-Generated Content (UGC). Questi rappresentano l'input principale per l'esplorazione della VoC in ambito digitale. Analizzando questi contenuti si possono individuare i bisogni dei clienti.

A differenza delle metodologie tradizionali, la VoC digitale è immediata, gratuita e costantemente aggiornata. Sfruttando il passaparola su Internet, garantisce un flusso continuo di informazioni senza necessità di campagne specifiche, basandosi sulla spontaneità degli utenti nel rilasciare recensioni. Il campione in questo contesto non è proporzionale al costo, poiché raccogliere pochi o molti contenuti comporta un costo pressoché identico, a causa dei bassi costi marginali associati (Timoshenko, 2019).

Il campione può raggiungere dimensioni assimilabili ai big data, rappresentando un vantaggio significativo rispetto alle analisi tradizionali con campioni di dimensioni limitate. Tuttavia, una parte considerevole di questi dati è caratterizzata da ripetitività e irrilevanza. Inoltre, poiché i dati sono non strutturati e generati liberamente dagli utenti, l'analisi manuale risulta impraticabile. Per ovviare a ciò, nell'utilizzo di dati estrapolati attraverso la digital VoC, si adottano tecniche di *data mining*. Queste permettono di estrarre informazioni significative da grandi volumi di dati, eliminando elementi ridondanti e non utili durante una fase di processamento.

Attraverso l'analisi della VoC, è possibile esaminare e monitorare le percezioni di qualità dei clienti, anticipando altresì i loro bisogni futuri e intervenendo per migliorare il design di nuovi servizi e prodotti.

L'analisi della VoC consente di identificare diversi tipi di requisiti (Wang & Tseng, 2011):

- Requisiti espliciti, requisiti che sono chiaramente espressi e facilmente individuabili.
- Requisiti impliciti, requisiti che non sono chiaramente espressi, ma possono considerarsi caratteristiche intrinseche relative al prodotto o servizio in questione.
- Requisiti latenti, requisiti che non sono né espressi chiaramente né intrinseci, ma sono comunque desiderati dal cliente. Identificare tali requisiti rappresenta uno dei principali fattori critici di successo per garantire un'elevata percezione di qualità del prodotto o del servizio offerto.

ANALISI DELLA DIGITAL VOC

Come precedentemente delineato, la digital VoC si presenta come una soluzione che supera le restrizioni della VoC tradizionale, consentendo la creazione di campioni estesi senza gravare eccessivamente sulle risorse. La creazione di grandi database, oltre a facilitare analisi più approfondite e significative, pone l'impossibilità di eseguire un'analisi manuale esaustiva dell'intero campione a disposizione, poiché comporterebbe tempi di esecuzione prolungati e delle intrinseche difficoltà nell'analisi.

Per affrontare questa problematica, si utilizzano tecniche di *data mining* che contribuiscono a ridurre i tempi, le risorse e i costi necessari per condurre l'analisi. Queste tecniche, basate su metodi di apprendimento automatico (*machine learning*), consentono l'estrazione di informazioni rilevanti da grandi quantità di dati strutturati (Delen e Crossland, 2008). I dati strutturati sono caratterizzati dalla loro organizzazione ordinata, agevolmente riconoscibile e adatta per il trattamento mediante modelli automatici, e possono riguardare dati di varia natura, come codici e numeri.

Un'implementazione specifica del *data mining* è rappresentata dalla tecnica di *text mining*, che si concentra sull'analisi di dati non strutturati, come sequenze di testo (recensioni) espresse in linguaggio naturale. Tale tecnica implica l'elaborazione del testo non strutturato all'interno di raccolte di documenti (database) strutturati. L'utilizzo di tecniche di *text mining* per estrarre informazioni significative dall'analisi della digital VoC è un approccio ampiamente adottato (Berry e Kogan, 2010).

TOPIC MODELING

Una delle metodologie impiegate per condurre il *text mining* è il *topic modeling*.

Il *topic modeling* è uno strumento di analisi non supervisionato finalizzato all'individuazione degli "argomenti" (topic) latenti presenti in una raccolta di documenti non strutturati (Müller et al., 2016).

Inizialmente, il modello genera i topic. Successivamente esamina la proporzione con cui ciascun topic individuato è discusso all'interno del corpus testuale.

Un elemento critico è la capacità di identificare un numero appropriato di argomenti in grado di delineare accuratamente il corpus testuale.

Nello studio proposto, il testo è estratto dalle recensioni rilasciate dai clienti su un servizio di cui hanno usufruito, utilizzando la tecnica del *topic modeling*.

Si sfruttano algoritmi di intelligenza artificiale che riescono a estrarre gli argomenti più discussi all'interno del database contenente le recensioni, permettendo l'individuazione delle determinanti di qualità tipiche del servizio. Queste determinanti sono le caratteristiche che influenzano il cliente nel percepire un servizio di qualità o meno. Avere un cliente soddisfatto si pone come base per il successo di un'azienda.

Tra gli algoritmi più diffusi per il *topic modeling* spiccano la "*Latent Dirichlet Allocation*" (LDA) (Blei et al., 2003) e lo "*Structural Topic Model*" (STM) (Roberts et al., 2013; Kherwa e Bansal, 2020).

Lo STM consente l'analisi, non solo del contenuto testuale delle recensioni, ma anche di altri metadati rilevanti, come la data della pubblicazione, il livello di rating associato e l'autore della recensione (Mastrogiacomo et al., 2021). In questo studio, per tali motivazioni, si usa l'algoritmo di STM per analizzare la digital-VoC.

STM

Lo STM può essere considerato un'estensione della LDA e dell'algoritmo di "*Correlated Topic Models*" (CTP) (Blei e Lafferty, 2007). Si caratterizza per il suo elemento di novità del caratterizzare in maniera flessibile i topic, includendo anche i metadati, ovvero informazioni aggiuntive rispetto al contenuto rilasciato dall'utente stesso. Si basa su un modello probabilistico che utilizza una variante dell'approccio LDA per estrarre gli elementi dai documenti (Roberts et al., 2016). Una caratteristica di questo algoritmo è la sua flessibilità nel poter analizzare documenti di lunghezze e formati diversi e trova applicazione in vari contesti di ricerca, analizzando contenuti postati sui social network, sui

forum online e nei sondaggi. È particolarmente utile per analizzare grandi quantità di dati tra loro correlati (Roberts et al., 2016).

L'algoritmo, ricevendo in input documenti espressi in forma testuale con relativi metadati estratti mediante la digital VoC, genererà 2 output:

1. Il topical content (φ) rappresenta il contenuto tematico di ciascun topic, caratterizzato da un elenco di parole chiave che lo descrivono.

Fondamentalmente, si tratta di una matrice in cui le righe rappresentano l'intero vocabolario degli argomenti, mentre le colonne identificano i diversi topic individuati. Ogni parola è associata a una probabilità basata sulla sua frequenza nei documenti relativi a un determinato topic. Le parole più rilevanti sono caratterizzate da valori più elevati, evidenziando la loro importanza nel descrivere il topic associato alla colonna corrispondente.

2. La topical prevalence (ϑ) rappresenta la distribuzione multinomiale di probabilità associata a ciascun documento. Indica la probabilità che le determinanti siano discusse all'interno del documento. Alcune probabilità sono significative, altre no e indicano che l'argomento non è discusso nel documento. In Figura 1 si può osservare come si presenta questo output.

Fondamentalmente, si tratta di una matrice in cui le righe corrispondono ai documenti analizzati, mentre le colonne rappresentano i vari topic individuati. Ogni cella contiene un valore che indica la frequenza con cui un determinato topic è trattato nel documento corrispondente. Un valore più elevato indica una probabilità maggiore che il topic sia ampiamente discusso in quel particolare documento.

Lo schema concettuale di questo algoritmo è illustrato nella Figura 1, in cui i documenti estratti tramite web scraping da svariate fonti costituiscono il dataset di digital-VoC.

Digital VoC collection

Algorithm

Output

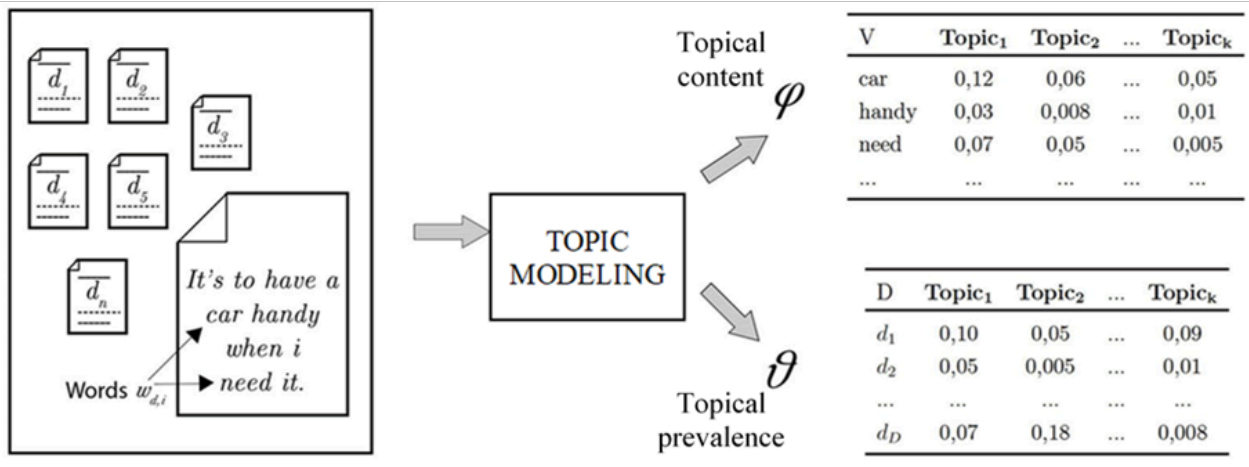


Figura 1. Funzionamento di un algoritmo di topic modeling (Barravecchia et al., 2022).

Per introdurre i metadati nel modello si usano covariate a livello documentale, in particolare i metadati che influenzano la prevalenza tematica sono introdotti come covariate di prevalenza tematica, mentre quelli che influenzano il contenuto tematico sono introdotti come covariate del contenuto tematico (Roberts et al., 2014; Roberts et al., 2019).

Per presentare il modello si utilizzano le seguenti notazioni (Roberts et al., 2019):

- d , per indicare i documenti che compongono il corpus testuale, con $d \in \{1 \dots D\}$.
- n , per indicare le parole che compongono i documenti, con $n \in \{1 \dots N_d\}$, dove N_d è il numero di parole del documento d -esimo.
- $w_{d,n}$, per indicare le parole dei documenti che costituiscono il vocabolario contenente i termini unici.
- k , per indicare i topic, con $k \in \{1 \dots K\}$, con K = numero di topic che si definisce a priori.

Come riportato in Figura 1 la topical prevalence e il topical content si possono rappresentare su due matrici, che verranno indicate rispettivamente con le lettere X e Y .

La matrice X presenta sulle righe i documenti del corpus D e sulle colonne le covariate che influenzano la topical prevalence P . Le sue dimensioni quindi sono $D \times P$.

La matrice Y presenta sulle righe analogamente alla matrice X i documenti del corpus D e sulle colonne le covariate che influenzano il topical content A . Le sue dimensioni quindi sono $D \times A$.

In ogni riga di entrambe le matrici, indicate rispettivamente con X_d e con Y_d è definito un vettore con tutte le covariate per ciascun documento.

In Figura 2 si riportano i tre sotto modelli successivi mediati i quali si può suddividere il modello STM (Roberts et al., 2016):

1. Il modello generatore della topical prevalence;
2. Il modello generatore del topical content;
3. Il modello generatore delle singole parole di un documento.

Nella Figura i rettangoli rappresentano le iterazioni effettuate dall' algoritmo per individuare come si distribuiscono le parole per ciascun topic e per individuare come si distribuiscono i topic stessi all'interno dei documenti. Il rettangolo esterno, invece indica che le variabili contenute al suo interno si differenziano per ogni documento. Le variabili contenute in quello più interno si differenziano per ogni parola.

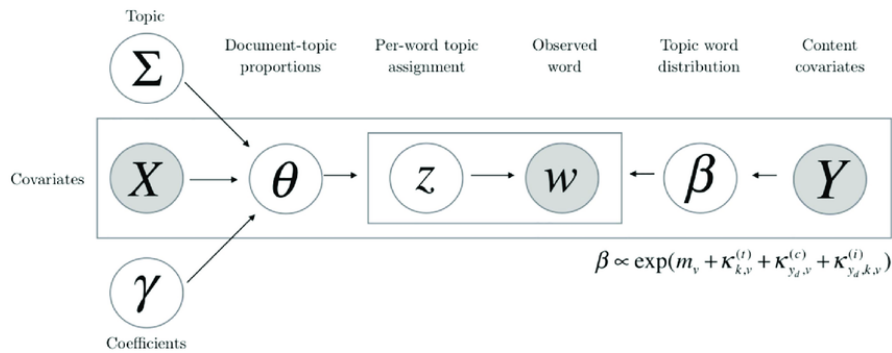


Figura 2. Sotto modelli successivi del STM (Roberts et al., 2016).

MODELLO GENERATORE DELLA TOPICAL PREVALENCE

Per determinare la proporzione di discussione di un topic per ciascun documento in analisi si assume che la topical prevalence di un documento ϑ_d segua una distribuzione Normale-Logistica.

La topic prevalence si esprime quindi in funzione delle covariate di documento X_d , dei topic k e dalla relazione covariata-topic Y , in Formula (1):

$$\vartheta_d | X_{d\gamma}, k \sim \text{LogisticNormal}(\mu = X_{d\gamma}, k) \quad (1)$$

Si definisce così una nuova matrice ϑ di dimensione $D \times K$ che per ciascun documento contiene i valori di proporzione dei topic presenti. Sulle righe si ha analogamente alla due matrice

precedentemente presentate D, ovvero i documenti che compongono il corpus, mentre sulle colonne si ha il numero di topic K. All'interno della matrice si avranno i valori ϑ_d che identificheranno i valori della topical prevalence per il d-esimo documento e per il k-esimo topic.

Ad un documento possono essere associati più valori di topical prevalence di topic diversi, ma la somma per riga, ovvero la somma per ciascun documento deve avere valore unitario.

Si riporta in Tabella 1 un esempio di come si può presentare una matrice di topic prevalence ϑ .

Tabella 1. Matrice di topical prevalence.

		Topic 1	Topic 2	...	Topic k
$\vartheta =$	Documento 1	0.10	0	...	0.5
	Documento 2	0.15	0.05	...	0

	Documento D	0	0	...	0.6

MODELLO GENERATORE DEL TOPICAL CONTENT

Soffermandosi sul topical content $\beta_{k,v}$, va notato che ciascun topic presenta una propria distribuzione di probabilità sulle parole, che può variare in base al valore dell'iterazione documento-covariata contenuta nella matrice Y. Queste covariate non sono integrate nel modello, come accade nel caso del modello generativo di topical prevalence, ma influiscono direttamente sulla parametrizzazione della distribuzione degli argomenti sulle parole. Il topical content è quindi determinato dalla distribuzione multinomiale delle occorrenze delle parole m_v , dalla quota di deviazione di un topic $k_v^{..k}$, dalla quota di deviazione di una covariata $k_v^{y..}$ e dall'interazione topic-covariata $k_v^{y,k}$, in Formula (2):

$$\beta_{k,v} \propto \exp (m_v + k_v^{..k} + k_v^{y..} + k_v^{y,k}) \quad (2)$$

La distribuzione Multinomiale delle occorrenze di ciascuna parola m_v si definisce come il logaritmo delle quote di deviazione rispetto alla distribuzione di tutte le parole del corpus. Queste si definiscono in funzione dei topic, delle covariate che si sono osservate e dalla interazione tra covariata e topic stessi.

Analogamente alla topical prevalence, si definisce una matrice, in questo caso di dimensione $V \times K$, con sulle righe i vocaboli che vanno a costituire termini unici, mentre sulle colonne i topic K . I valori contenuti all'interno della matrice, così come per il primo modello presentato per la topical prevalence, rappresenteranno i valori di topical content $\beta_{k,v}$ relativi al k -esimo topic considerato e al v -esimo vocabolo. Anche in questo caso la somma delle $\beta_{k,v}$ presenti per righe sarà unitaria. Si riporta in Tabella 2 un esempio di come si può presentare una matrice di topic content β .

Tabella 2. Matrice di topical content.

	Topic 1	Topic 2	...	Topic k
$\beta =$ "cane"	0.001	0	...	0.15
"gatto"	0.015	0.005	...	0
...
Vocabolo V	0	0	...	0.06

MODELLO GENERATORE DELLE SINGOLE PAROLE DI UN DOCUMENTO

Il terzo modello rappresenta una combinazione dei due modelli precedenti. Si combinano i valori di topical prevalence e di topical content. In questa fase ciascuna parola n viene associata ad uno specifico topic k , basandosi sulla distribuzione multinomiale di topical prevalence ϑ_d che si associa ad un determinato documento.

Si definisce $z_{d,n}$ come l'assegnazione di un topic nel documento d -esimo per l' n -esima parola, in Formula (3):

$$z_{d,n} | \vartheta_d \sim \text{Multinomial}(\vartheta_d) \quad (3)$$

Parallelamente per quanto concerne il topical content si determina la probabilità di estrazione delle parole $w_{d,n}$ per ogni topic, in Formula (4):

$$w_{d,n} | z_{d,n}, \beta_{d,k=z_{d,n}} \sim \text{Multinomial}(\beta_{d,k=z_{d,n}}) \quad (4)$$

Il modello descritto si implementa nello studio in R, un linguaggio di programmazione specifico per la statistica e la grafica computazionale. In particolare, si utilizza il pacchetto 'stm' (Roberts et al.

2019) e RStudio che permette di utilizzare lo Structural Topic Model per stimare modelli generativi tematici con valori delle covariate a livello del singolo documento.

METODOLOGIA DI ANALISI APPLICATA NELLO STUDIO

L'algoritmo di *topic modeling* che si è utilizzato per l'analisi si articola in più fasi (Barravecchia et al., 2021):

1. Estrazione del set di dati;
2. Pre-Processamento;
3. Selezione del numero ottimale di topic;
4. Etichettatura dei dati;
5. Verifica dei dati;
6. Analisi dei risultati.

ESTRAZIONE DEL SET DI DATI

Per estrarre il set di dati da analizzare, si procede con la raccolta degli UGC attraverso i social media o dalle piattaforme che aggregano recensioni. Per l'estrazione delle recensioni, la tecnica più efficiente è risultata essere il *web scraping* (Bressoud, 2020). Il meccanismo seguito con questa tecnica è pressoché analogo a quello compiuto dal crawler dei motori di ricerca come Google. Così come il crawler di Google visita, scarica e indicizza tutte le pagine web (tramite il loro linguaggio HTML estraendone tutto il contenuto al loro interno) e le memorizza in una memoria locale, così nella tecnologia si visita il database online e si estraggono tutti i contenuti e le informazioni presenti nelle recensioni. Un vantaggio significativo dell'utilizzare la tecnica di *web scraping* è la capacità di ottenere una grande quantità di informazioni (anche nell'ordine di big data) tramite poche semplici istruzioni (Mooney et al., 2015).

Per effettuare il *web scraping*, è possibile utilizzare svariate tecniche, a partire da estensioni di browser, a software sviluppati appositamente per eseguire tale metodologia, fino al più tradizionale utilizzo delle librerie di software comuni come Python e Matlab (Zaki e McColl-Kennedy, 2020).

Oltre al testo della recensione, è possibile estrarre anche ulteriori informazioni che sono correlate alla recensione stessa, quali il rating, la data di pubblicazione e ove presenti anche metadati come età, nazionalità, sesso dell'utente che ha rilasciato il contenuto. L'analisi risulterà tanto più dettagliata, tanto più si avranno metadati e informazioni associate alle recensioni.

PRE-PROCESSAMENTO

Durante questa fase, si attua il pre-processamento del testo estratto attraverso l'applicazione di algoritmi di pulizia e normalizzazione del testo. Tale procedura mira a ottimizzare l'efficienza dell'algoritmo di *topic modeling* (Meyer et al., 2008) mediante l'eliminazione delle porzioni delle recensioni ritenute non informative ai fini della determinazione degli argomenti trattati al loro interno. Le operazioni più frequentemente adottate in questa fase includono (Guo et al., 2017):

- La conversione del testo in carattere tutto minuscolo, per evitare che ci siano ambiguità e differenze con il maiuscolo.
- La rimozione dei numeri e della punteggiatura, in quanto non aggiungono molto significato al topic.
- La rimozione delle così dette “stop words”, ovvero quelle proposizioni e congiunzioni (come ad esempio “the”, “and”, “when”, “is”, “at”, “which”, “on” etc.).
- La rimozione delle parole che compaiono all'interno del database con una frequenza molto basse, queste infatti potrebbero avere un effetto confondente sui risultati o difficili da identificare in uno specifico topic. Le parole estremamente corte, che presentano meno di due caratteri non portano significato da sole e possono essere considerate rumore.
- La rimozione di parole che presentano o meno di due o più di 15 caratteri, per eliminare elementi che potrebbero avere un impatto limitato sulla comprensione del testo, o addirittura risultare indesiderati. Le parole con più di 15 caratteri potrebbero risultare essere errori di battitura o risultare poco informative. Inoltre riducendo le parole molto corte o lunghe, si riesce a ridurre la dimensione del vocabolario, risparmiando memoria e ottenendo efficienza computazionale.
- L'applicazione del processo di stemming (Jivani, 2011), in cui si elimina la desinenza finale delle parole per riportare tutte le parole stesse ad avere la stessa radice all'interno del vocabolario (ad esempio, le parole “likes”, “liked”, “likely” e “liking” vengono tutte ricondotte alla radice “like”), poiché parole che derivano da una stessa radice esprimono concetti affini o variazioni semantiche dello stesso concetto.
- La rimozione di quelle parole che non sono correlate ai contenuti del topic principale (ad esempio “another”, “mean”, “problem”, “review”, “made”, “did”, “done” etc.).
- La sostituzione dei n-grammi (insiemi di n parole che compaiono insieme in sequenza) con un unico termine (Durani et al., 2015) (ad esempio “customer service” viene replicato con “customerservice”).

SELEZIONE DEL NUMERO OTTIMALE DI TOPIC

Quando si utilizza l'algoritmo di *topic modeling* un parametro fondamentale di input è K , che rappresenta il numero di topic capaci di descrivere il corpus testuale. Per la sua definizione sono state proposte svariate alternative (Wallach et al., 2009).

Uno degli approcci impiega un metodo iterativo per determinare il parametro ottimale. In ogni iterazione, si valutano le prestazioni del modello. In questa fase, è cruciale identificare il numero ottimale di topic: un numero troppo basso non consentirebbe una distinzione sufficiente dei contenuti delle recensioni, generando topic troppo generalisti che descrivono concetti troppo ampi per effettuare un'analisi efficace; un numero troppo alto aumenterebbe eccessivamente la complessità del modello, producendo topic troppo specifici, di difficile interpretabilità e poco trattati nelle recensioni. Difficilmente un utente può stabilire a priori il numero di topics ideali per la sua analisi, spesso si può definire solo un range di valori ritenuti accettabili per conseguire gli obiettivi (nell'analisi proposta basandosi sulla dimensione del dataset individuato il range proposto varia tra 5 e 50). Si adotta un approccio iterativo tramite RStudio, eseguendo numerosi test (Roberts et al., 2014), applicando l'algoritmo *stm* per ogni valore k all'interno del vettore (contenente i valori di range individuati) e calcolando gli indicatori di performance.

In particolare nello studio, per individuare K si fa ricorso all'indicatore di performance Held-out likelihood, che misura l'efficacia del modello nel riuscire a spiegare la variabilità complessiva nel corpus testuale (Scott e Baldrige, 2013). Tale parametro ha un'applicazione diffusa negli algoritmi di *topic modeling* e valuta la qualità del modello generato.

Il calcolo di questo indicatore avviene utilizzando una parte del corpus dei documenti come training set (circa il 90%) e la restante parte dei documenti su cui non viene applicato l'algoritmo STM (Held-out documents) come set di test. Misurando la verosimiglianza tra i risultati ottenuti nel set di addestramento e nel set di test, si ottiene una metrica che valuta l'efficacia dello structural topic model per un numero specifico di argomenti K . Idealmente, si auspica l'utilizzo di K che massimizzi il valore di Held-out likelihood. Tale indicatore risulta cruciale per valutare in che misura il modello spieghi la variabilità all'interno del corpus di documenti e per prevenire l'overfitting, un fenomeno in cui un algoritmo di *topic modeling* tende a migliorare le sue performance eccessivamente aumentando il numero di argomenti, arrivando al punto di spiegare ogni documento con sé stesso.

ETICHETTATURA DEI DATI

In questa fase, si procede con l'etichettatura dei dati, ossia l'assegnazione di un'“etichetta” ai topic generati dall'algoritmo di *topic modeling*. Si attribuisce un titolo al topic, identificandolo sulla base dei termini più ricorrenti nei documenti che lo caratterizzano. È importante notare che topic diversi

possono presentare alcuni termini analoghi posti in contesti diversi, causando un'etichettatura diversa dei topic stessi.

Questa fase comporta numerosi vantaggi all'analisi tra cui:

- Ridurre il carico cognitivo necessario per interpretare uno specifico topic, etichettando infatti si assegna un significato a ciascun topic del modello, rendendo così più semplice comprendere di cosa tratti.
- Rimuovere la possibilità di interpretazione soggettiva dei singoli.
- Riassumere le informazioni riguardanti il topic utilizzando un titolo breve, in questo modo si ha una comunicazione chiara che comunica di cosa tratta il topic senza dover esaminare i documenti sottostanti.
- Comunicare i risultati ottenuti più agevolmente.
- Facilità nella ricerca e nell'organizzazione all'interno dei database.
- Miglioramento dell'usabilità, in quanto contribuisce a una migliore esperienza per l'utente semplificando la fruizione delle informazioni e la ricerca attraverso i dati.

La fase di etichettatura può essere di due tipologie:

- *Topic labelling automatico*,
utilizzando questo approccio, le etichette vengono assegnate automaticamente da algoritmi di intelligenza artificiale, quali ad esempio ChatGPT. Per l'assegnazione si utilizzano parole chiave rappresentative dei contenuti dei topic.
Un chiaro vantaggio di questa tipologia consiste nell'ottenere un'etichettatura rapida, che permette di risparmiare sui tempi e che si può adattare a dataset di grandi dimensioni.
Un contro potrebbe essere quello di ottenere un'etichettatura di non semplice comprensione o che non descriva a pieno il contenuto dei documenti, quindi alcuni topic potrebbero subire un'etichettatura errata o ambigua.
- *Topic labelling manuale*,
si intende l'operazione di assegnazione del titolo del topic svolta manualmente da una o più persone, che esaminano l'intero contenuto presente all'interno dei documenti.
Una delle caratteristiche distintive di questa tipologia è la sua elevata precisione e flessibilità. Gli annotatori umani responsabili dell'etichettatura manuale riescono a cogliere particolari che una macchina non può garantire. Si riescono in questo modo a individuare anche i topic più complessi. Questa metodologia è particolarmente indicata quando si mira a condurre un'analisi estremamente accurata.

Va notato che un'analisi così dettagliata richiede un considerevole impegno umano e un notevole periodo di tempo, comportando costi più elevati. Inoltre, la sua applicazione su dataset di dimensioni considerevoli può risultare problematica e potrebbe dipendere dall'esperienza dell'operatore responsabile dell'etichettatura.

Solitamente si predilige l'etichettatura automatica quando si ha a che fare con quantità di dati troppo elevati per poter essere gestiti manualmente. Quando invece si vuole un'analisi più accurata e, in particolar modo, quando non si può sacrificare la precisione l'etichettatura manuale risulta essere l'approccio prediletto. Talvolta si può considerare una combinazione delle due tipologie, con una prima etichettatura automatica, e una successiva revisione manuale.

Per generare un'etichetta semanticamente corretta e appropriata si devono fornire alcuni input manuali (Blei 2012).

Per l'analisi del caso studio, la precisione è ritenuta una caratteristica imprescindibile e pertanto si usa l'etichettatura dei topic manuale. Come input per l'analisi si prendono le parole chiave e i documenti più rappresentativi.

Per trovare l'elenco di parole chiave che descrivono al meglio il topic preso in analisi, si prendono in considerazione alcuni parametri (Roberts et al., 2019), tra cui principalmente:

- *Highest probability*, ovvero quelle parole che hanno la più alta probabilità di essere contenute all'interno del topic.
- *FREX*, ovvero quelle parole chiave che sono più presenti e sono esclusive del topic.
- *Recensioni più rappresentative*, ovvero si va a fare un'analisi qualitativa esaminando alcune recensioni che rappresentano al meglio il topic in analisi, andando anche a individuare il “tono” e il tema trattato nello specifico.

VERIFICA DEI DATI

La metodologia adottata in questa fase prevede il confronto tra i risultati ottenuti da una valutazione umana e quelli generati automaticamente basandosi sull'output dell'algoritmo (ovvero la topical prevalence). Come già menzionato in precedenza, l'assegnazione manuale è considerata il valore di riferimento corretto. Si effettua una comparazione tra il valore assegnato automaticamente e il valore di riferimento per valutare l'efficacia del modello.

Per l'assegnazione manuale dei topic, si effettua un'analisi qualitativa su ciascun documento. Per ciascuno di essi, per ogni topic individuato, si assegnano i seguenti valori in Formula (5):

$$\begin{cases} 0 & \text{se il topic non è trattato all'interno del documento} \\ 1 & \text{se il topic è trattato all'interno del documento} \end{cases}$$

(5)

Si riporta un esempio sul caso studio che si affronterà nel secondo capitolo di questo procedimento di assegnazione manuale dei topic a ciascuna recensione in Tabella 3:

Tabella 3. Assegnazione manuale dei topic alle recensioni.

Recensioni	Topic	1	2	3	4	5	6	7	8	9	10	11
I had the app but I got it back. My problem is that it says FREE but when I used the app after 14 days it says your free trial has ended. And it wants me to pay money. I don't have the money for that. ☹️ Please fix this.		0	0	0	1	0	0	0	0	0	0	0
It good i just wish you could customize the mood/genera filters on the liked songs. Also the shuffle and play thing i want to click shuffle multiple times not only being able to press it once.		0	1	0	0	0	0	0	0	0	0	0
This is amazing. Im listening to it right now on my xbox while on discord! (That made no sense) And i love the app so much! I listen to it like 24/7 because Its so goooood like SHEEESH		0	0	1	0	0	0	0	0	0	0	0
Overall the app is good but i have a feature request, please add play next for songs and even add to queue for songs because everytime the album needs to be accessed		0	1	0	0	0	0	0	1	0	0	0
Why is it asking for Bluetooth access each time I open the app now? Annoying		0	0	0	0	1	0	0	0	0	0	0
Very worst app, I got too much of ad when move to another song. I wish to say use offline music app for better Experience		0	0	0	0	0	0	0	1	0	0	0

Dati i seguenti topic relativi al caso studio Spotify:

- TOPIC 1-Connection problem: wifi;
- TOPIC 2-Subscription not premium: shuffle;
- TOPIC 3-App rating: section music;
- TOPIC 4- Fees and payments;
- TOPIC 5-Bug user experience/ app reliability;
- TOPIC 6-podcast and information;
- TOPIC 7-User friendly;
- TOPIC 8-Song accessibility;
- TOPIC 9-Huge selection;

- TOPIC 10-Account and log in problem;
- TOPIC 11-Update features;

Si esamina un campione rappresentativo del dataset, garantendo un numero sufficiente di documenti per una rappresentatività significativa. Ad esempio, l'analisi delle prime n recensioni (dove n è un numero sufficientemente elevato per descrivere approssimativamente l'intero dataset) potrebbe essere influenzata dal fattore tempo. Per superare questa limitazione, si estrae casualmente un numero n di documenti ritenuti sufficienti, in relazione al dataset di partenza, che saranno sottoposti ad un'analisi qualitativa approfondita di uno o più valutatori umani.

Per l'assegnazione automatica dei topic, a partire da una distribuzione multinomiale di probabilità assegnata dall'algorithmo di *topic modeling* è necessario identificare i topic più rappresentativi per ogni recensione. Per tale scopo, è possibile utilizzare diversi approcci, tra cui (Barravecchia et al., 2021):

- *Massima probabilità*: in questo approccio, a ciascun documento viene assegnato il topic più probabile, ovvero quello con la probabilità di appartenenza più elevata. In questo modo, a ciascun documento verrà assegnato uno ed un solo topic. Questo approccio viene scartato per questa analisi in quanto ignora le sfumature nelle probabilità di assegnazione, ovvero pur avendo un'elevata probabilità un topic può essere comunque escluso dal documento. Questo metodo può essere adottato in analisi più semplici, grazie alla sua immediata e facile applicazione.
- *Soglia statica*: in questo approccio, si stabilisce una soglia di probabilità minima. Un documento è assegnato a un topic solo se la probabilità di appartenenza supera la soglia predeterminata per quel topic.
- *Soglia dinamica*: a differenza della soglia statica, che rimane costante durante l'intero processo, la soglia dinamica si adatta alle fluttuazioni delle probabilità dei topic nei documenti.

In questo caso, è possibile temporaneamente fissare la soglia a un determinato valore e successivamente aggiornarla in base alle dinamiche osservate durante l'analisi.

Questo metodo consente di assegnare un topic a un documento solo se la probabilità di appartenenza a quel topic è significativamente superiore rispetto alle altre.

Nell'analisi proposta, è stato adottato l'approccio della soglia dinamica. Questo metodo, analogamente alla soglia statica consente di superare le limitazioni della massima probabilità,

fornendo criteri più flessibili per l'assegnazione del topic. La scelta della soglia dinamica è stata preferita rispetto alla soglia statica in considerazione di documenti ambigui o caratterizzati da probabilità simili per più topic.

La Formula (6) utilizzata per il calcolo della soglia dinamica è la seguente:

$$\text{Soglia dinamica}(SD) \rightarrow SD_i = Q3_i + (1,5 * IQR_i) \quad (6)$$

Dove IQR_i rappresenta il range interquartile i-esimo, calcolato come differenza tra il terzo quantile Q3 e il primo quartile Q1. Il range interquartile indica il grado di dispersione della funzione. In questa Formula (6), i indica la i-esima recensione.

ANALISI DEI RISULTATI

Dalla comparazione tra l'assegnazione automatica dei topic ai documenti e l'assegnazione manuale, utilizzata come riferimento, è possibile ottenere indicazioni sulla bontà del modello impiegato, ovvero si può comprendere l'efficacia dell'algoritmo di *topic modeling* nell'assegnazione automatica dei topic ai documenti.

Per ogni documento analizzato mediante entrambe le metodologie, si possono verificare 4 possibili casistiche (Barravecchia et al., 2022) riportate in Tabella 4.

Tabella 4. Scenari individuati dal confronto tra assegnazione automatica dei topic e assegnazione manuale.

		Assegnazione manuale dei topic (condizioni effettive)	
		Presenza del topic i	Non presenza del topic i
Assegnazione automatica dei topic	Presenza del topic i	Veri positivi (true positive= tp)	Falsi positivi (false positive= fp)
	Non presenza del topic i	Falsi negativi (false negative= fn)	Veri negativi (true negative= tn)

Si individuano 4 scenari possibili:

- Veri positivi (tp), in questo caso si ha inferenza corretta. Si ha infatti l'accordo tra l'assegnazione automatica e l'assegnazione manuale. Entrambe rilevano la presenza del topic all'interno della recensione.
- Falsi positivi (fp), in questo caso si ha l'errore di prima specie α . Si ha disaccordo tra l'assegnazione automatica, che rileva la presenza di un topic, e l'assegnazione manuale, che non la rileva.

- Falsi negativi (fn), in questo caso si ha l'errore di seconda specie β . Si ha disaccordo tra l'assegnazione manuale, che rileva la presenza di un topic, e l'assegnazione automatica, che non la rileva.
- Veri negativi (tn), in questo caso si ha inferenza corretta. Si ha infatti l'accordo tra l'assegnazione automatica e l'assegnazione manuale. Entrambe non rilevano la presenza di un topic nella recensione.

INDICATORI DI VALIDAZIONE

In questo paragrafo si indicheranno i 4 scenari individuati dal confronto tra l'assegnazione manuale e l'assegnazione automatica nel seguente modo:

- tp=true positive;
- tn=true negative;
- fp=false positive;
- fn=false negative.

Per valutare la bontà dei risultati ottenuti, si utilizzeranno i seguenti indicatori di validazione (Costa et al., 2007):

- *Accuracy*: è l'indicatore più significativo e si ottiene dal rapporto tra la sommatoria dei veri positivi e dei veri negativi con la sommatoria dei veri positivi, dei veri negativi, dei falsi positivi e dei falsi negativi, in Formula (7). Un valore più elevato di questo indicatore, idealmente tendente all'unità, riflette una maggiore capacità dell'algorithmo nel spiegare il database analizzato.

$$Accuracy = \frac{\sum_{i=1}^n tp_i + \sum_{i=1}^n tn_i}{\sum_{i=1}^n tp_i + \sum_{i=1}^n tn_i + \sum_{i=1}^n fp_i + \sum_{i=1}^n fn_i} \quad (7)$$

- *Precision*: stima la probabilità che una previsione positiva sia effettivamente positiva, quindi corretta (infatti viene denominato valore predittivo positivo). Misura il numero di totali veri positivi riscontrati sul totale dei positivi trovati (ovvero la somma dei veri positivi e dei falsi positivi), in Formula (8).

$$Precision = \frac{\sum_{i=1}^n tp_i}{\sum_{i=1}^n tp_i + \sum_{i=1}^n fp_i} \quad (8)$$

- *Recall*: misura il rapporto tra i veri positivi e la somma tra i veri positivi e i falsi negativi, in Formula (9). Viene anche denominata sensibilità del modello o tasso di positività reale. Mediante questo indicatore si ha una misura di quanto un algoritmo riesca a rilevare un determinato topic, ove presente in una recensione. Il *Recall* può essere calcolato anche come il complemento all'unità dell'errore di seconda specie β , in Formula (10).

$$Recall = \frac{\sum_{i=1}^n tp_i}{\sum_{i=1}^n tp_i + \sum_{i=1}^n fn_i} \quad (9)$$

$$Recall = 1 - \beta \quad (10)$$

- *F1 Score*: è una misura dell'accuratezza del test. Per la sua determinazione, necessita dei valori di *Precision* e di *Recall*, in Formula (11).

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

- *False discovery rate*: è la probabilità di identificare erroneamente un topic. Si ottiene dal rapporto tra i falsi positivi e la sommatoria tra veri positivi e falsi positivi, in Formula (12).

$$False\ discovery\ rate = \frac{\sum_{i=1}^n fp_i}{\sum_{i=1}^n tp_i + \sum_{i=1}^n fp_i} \quad (12)$$

- *Negative predictive Value*: si ottiene dal rapporto tra i falsi negativi e la sommatoria tra falsi negativi stessi e i veri negativi, in Formula (13), ovvero coglie la probabilità che l'algoritmo non rilevi un topic che non è attualmente presente.

$$Negative\ predictive\ value = \frac{\sum_{i=1}^n tn_i}{\sum_{i=1}^n tn_i + \sum_{i=1}^n fn_i} \quad (13)$$

- *False omission rate*: è il complemento all'unità del *negative predict value* e indica la proporzione di topic non rilevati quando un topic è già presente. Si ottiene dal rapporto tra i falsi negativi e la sommatoria tra falsi negativi e veri negativi, in Formula (14).

$$False\ omission\ rate = \frac{\sum_{i=1}^n fn_i}{\sum_{i=1}^n tn_i + \sum_{i=1}^n fn_i} \quad (14)$$

- *Fall-out*: è una misura che si ottiene dal rapporto tra i falsi positivi e la sommatoria tra falsi positivi stessi e veri negativi, in Formula (15). Viene anche definito tasso di falsi positivi, in quanto indica la probabilità di trovare un falso positivo, ove il valore sia falso per il modello. Il *fall-out* è equivalente all'errore di prima specie α , in Formula (16).

$$Fall - out = \frac{\sum_{i=1}^n fp_i}{\sum_{i=1}^n fp_i + \sum_{i=1}^n tn_i} \quad (15)$$

$$Fall - out = \alpha \quad (16)$$

- *Miss-Rate*: questo indicatore è l'equivalente del *fall-out*, ma ponendo i falsi negativi come protagonisti della misurazione, in Formula (17). È l'equivalente dell'errore di seconda specie β . Viene anche definito come tasso di falsi negativi e si misura dal rapporto tra quelli che sono i falsi negativi e la sommatoria dei falsi negativi stessi e i veri positivi.

$$Miss - Rate = \frac{\sum_{i=1}^n fn_i}{\sum_{i=1}^n tp_i + \sum_{i=1}^n fn_i} \quad (17)$$

- *Specificity*: questo indicatore misura la proporzione di veri negativi, in Formula (18). Nell'algoritmo di topic modelling si interpreta come la proporzione di topic che non sono identificati perché non sono presenti nel modello attuale.

$$Specificity = \frac{\sum_{i=1}^n tn_i}{\sum_{i=1}^n tn_i + \sum_{i=1}^n fp_i} \quad (18)$$

VALORI TARGET

In Tabella 5 si riportano i valori target che ci si aspetterebbe di ottenere dagli indicatori di validazione appena definiti (Mastrogiacomo et al., 2021). Tutti gli indicatori sono espressi in un range di valori compreso tra $[0;1]$ (o espressi in valore percentuale) e, per ognuno di essi, è espresso il valore target da soddisfare e la direzione verso la quale ci si avvicina ad un buon valore dell'indicatore.

Tabella 5. Valori di riferimento degli indicatori di validazione dell'algoritmo di topic modeling (Mastrogiacomo et al., 2021).

Indicatori	Range	Direzione	Valori target
ACCURACY	[0;1]	Alto è buono	>0.95
PRECISION	[0;1]	Alto è buono	>0.70
FALSE DISCOVERY RATE	[0;1]	Basso è buono	<0.05
FALSE OMISSION	[0;1]	Basso è buono	<0.05
NEGATIVE PREDICTIVE VALUE	[0;1]	Alto è buono	>0.90
RECALL	[0;1]	Alto è buono	>0.70
FALL-OUT	[0;1]	Basso è buono	<0.05
MISS RATE	[0;1]	Basso è buono	<0.20
SPECIFICITY	[0;1]	Alto è buono	>0.90
F1 SCORE	[0;1]	Alto è buono	>0.70

Nel caso gli indicatori calcolati si discostassero dai valori di riferimento si potrebbero seguire diverse strade (Mastrogiacomo et al., 2021):

- Verificare che i topic o le etichette assegnate rappresentino al meglio il topic in questione.
- Verificare che i parametri in input siano corretti, come ad esempio il numero di topic.
- Verificare la correttezza dell'interpretazione del modello sui vari metadati.
- Verificare l'input iniziale, ovvero verificare la pertinenza e l'omogeneità del database analizzato.

ANALISI DEI RISULTATI

L'algoritmo di *topic modeling* produce principalmente due risultati, ovvero trova le determinanti di qualità latenti del servizio analizzato posto sotto analisi e quanto in proporzione uno specifico topic sia associato ad ogni recensione.

L'applicazione dello STM consente una varietà di analisi (Mastrogiacomo et al. 2021):

- Consente di effettuare un'indagine sulle correlazioni tra metadati e la prevalenza dei topic;
- Consente di effettuare uno studio sulla prevalenza dei topic all'interno del campione di recensione analizzato;
- Consente di esaminare le relazioni tra il giudizio dei clienti e la prevalenza dei topic;
- Consente di analizzare le correlazioni tra i topic;
- Consente di analizzare la dinamica temporale delle proporzioni degli argomenti.

L'obiettivo finale nell'analizzare i risultati consiste nell'individuazione specifica delle determinanti di qualità che caratterizzano il servizio e riuscire a utilizzarle come strumenti per migliorare la qualità

dei processi aziendali o per eventuali riprogettazioni. A tal proposito, si costruiscono degli indicatori per esaminare questi aspetti che si propongono nelle prossime sezioni.

MTP- Mean Topic Proportion

Il Mean Topic Proportion (MTP) è un indicatore che fornisce quanto in media è discusso un topic nei documenti in analisi. Per calcolarlo, si utilizza la seguente Formula (19):

$$MTP = \frac{\sum_{i=1}^N TP_{i,t}}{N} \quad (19)$$

Dove:

- N il numero di recensioni/documenti analizzati;
- $TP_{i,t}$. la topical prevalence del t-esimo topic chiave nell'i-esima recensione.

La somma degli MTP relativi a tutti i topics identificati è pari a 1, come si evince dalla seguente Formula (20):

$$\sum_{t=1}^T MTP_t = 1 \quad (20)$$

È il primo indicatore che si prende in considerazione nell'analisi, nonché il più basilico. Indica una proporzione media della discussione del topic che si sta analizzando e si ottiene facendo il rapporto tra il valor medio della topic prevalence e il numero totale di recensioni analizzate.

È un indicatore che non va considerato isolato, in quanto potrebbe presentare delle distorsioni in base alla piattaforma utilizzata per creare il database da analizzare. Un esempio che può far comprendere questa distorsione, sono le recensioni prese tramite l'App Store che si concentreranno maggiormente sulla performance dell'applicazione, concentrandosi quindi solo su un aspetto e non sull'intero servizio e contenuto proposto.

MRP-Mean Rating Proportion

Il Mean Rating Proportion (MRP) (Barravecchia et al., 2020) è un indicatore che spiega quanto è discusso un topic nei record di digital VoC associati con uno specifico livello di rating. Per calcolarlo, si utilizza la seguente Formula (21):

$$MRP = \frac{\sum_{i \in R_k} TP_{i,t}}{|R_k|} \quad (21)$$

Dove:

- t = topic t-esimo;
- k = il livello della scala di valutazione (di rating) della recensione (si prendono i valori su cui sono espresse le valutazioni, solitamente da 1 a 5, con 1 minimo livello di soddisfazione percepita e 5 massimo livello);
- R_k =il sottoinsieme di recensioni della digital VoC associate al livello di rating k ;
- $TP_{i,t}$ =la topical prevalence del topic t-esimo nell'i-esima recensione;
- $|R_k|$ =la cardinalità delle recensioni associate a un livello di valutazione (rating) pari a k .

Si noti che la somma degli MRP relativi a tutti i topics identificati e a un livello di valutazione specifico è pari a 1, come si evince dalla Formula (22):

$$\sum_{t=1}^T MRP_{t,k} = 1 \quad \forall k \quad (22)$$

I clienti esprimono le loro valutazioni tramite la digital VoC, rilasciando recensioni sul web, sui vari store o su piattaforme di recensioni etc. mediante l'utilizzo di scale di valutazioni di tipo ordinale, rilasciando valori che possono variare da 1 a 5 (più comunemente noti come da 1 stella a 5 stelle). Con 1 stella si indica la peggior valutazione possibile e con 5 stelle una valutazione di eccellenza. Questo indicatore permette di avere una visione completa dell'andamento delle recensioni. Risponde alla domanda "come viene percepito un topic all'interno della digital VoC?"

Ciascun topic può presentare un profilo diverso. I profili, rappresentati in Figura 3, possono essere di tre tipologie (Barravecchia et al., 2020):

- *Profilo positivo*, il topic analizzato è maggiormente discusso nelle recensioni che presentano un rating positivo. Pertanto, è importante continuare a mantenere una discussione alta e positiva per questo topic che rappresenta un fattore critico di successo per l'azienda, in quanto genera soddisfazione da parte del cliente.
- *Profilo neutro*, il topic analizzato è maggiormente discusso nelle recensioni che presentano un rating neutro. Si tratta di topic che sono allo stesso tempo minacce e opportunità. Si concentrano principalmente nelle valutazioni intermedie (2, 3, 4) e meno nelle valutazioni di minima e massima soddisfazione (rispettivamente 1 e 5). Focalizzandosi su questi aspetti,

l'obiettivo deve essere quello di renderli profili positivi e far in modo che il cliente sia soddisfatto e percepisca valore anche da questi aspetti. Allo stesso modo, se trascurati e non monitorati continuamente nel corso del tempo rischiano di trasformarsi in profili negativi ed entrare in contrasto con gli obiettivi aziendali.

- *Profilo negativo*, il topic analizzato è maggiormente discusso nelle recensioni che presentano un rating negativo. Questi topic si caratterizzano principalmente di recensioni con il minimo livello di soddisfazione (1). Sono gli elementi che ostacolano la crescita aziendale e su cui si crea conflitto tra il consumatore finale e l'azienda stessa. Vanno monitorati costantemente nel tempo e possono essere la causa di allontanamento di un cliente dal servizio offerto.

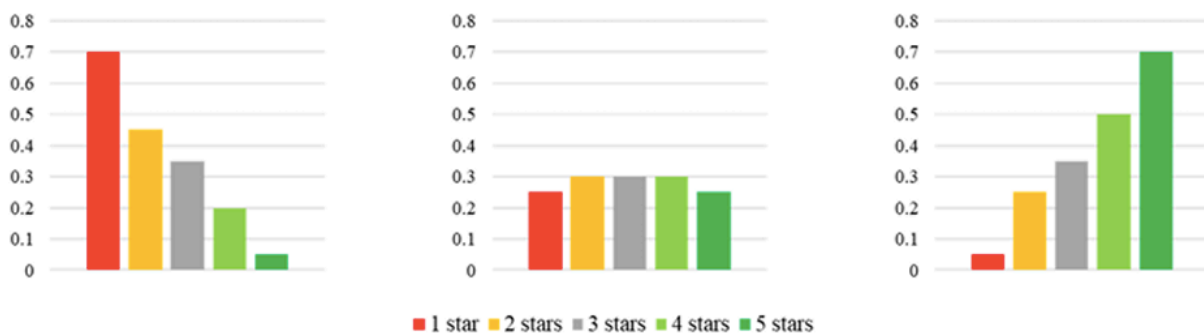


Figura 3. Profili MRP rispettivamente negativo, neutro e positivo (Barravecchia et al., 2020).

Questo indicatore risulta essere molto importante per l'analisi, in quanto l'informazione che produce mostra il legame che si ha tra il topic considerato con il livello di soddisfazione del cliente. Se monitorato nel tempo consente di far emergere l'evoluzione dei bisogni dei consumatori del servizio proposto, e in alcuni casi può permettere di anticipare i comportamenti e ottenere un vantaggio competitivo nei confronti delle imprese rivali.

Per questa serie di motivazioni e per l'importanza di questo indicatore, nel terzo capitolo si svolgerà un approfondimento e ci si concentrerà sull'individuazione dei valori anomali che si possono riscontrare. Si cercherà inoltre di individuare uno strumento che possa aiutare il management aziendale nel prendere le decisioni e che si comporti da "campanello d'allarme" qualora qualche valore si discostasse dalla distribuzione attesa dei dati.

IMTP - Interval Mean Topical Prevalence

Per osservare la variazione di un determinato topic nel corso del tempo si utilizza l'indice Interval Mean Topical Prevalence (IMTP) che calcola la topic prevalence di ciascun topic in un determinato

periodo di tempo (Barravecchia et al., 2022). Nell'analisi proposta ciascun periodo corrisponderà a un mese. Per calcolarlo si utilizza la seguente Formula (23):

$$IMTP_{t,n} = \frac{\sum_j^{R_t} TP_{i,t}}{|R_n|} \quad (23)$$

Dove:

- t =topic t-esimo;
- n =periodo n-esimo;
- R_n =il sottoinsieme delle recensioni della digital VoC associate al periodo n-esimo;
- $TP_{i,t}$ =si riferisce alla topic prevalence associata al topic t-esimo nella recensione i-esima;
- $|R_n|$ =indica la cardinalità delle recensioni della digital VoC associate al periodo n-esimo.

Come per MTP e MRP, vale la seguente Formula (24):

$$\sum_{d=1}^D IMTP_{t,n} = 1 \quad \forall n \in (1, \dots, N) \quad (24)$$

Questo indicatore è di notevole importanza non solo perché fornisce una fotografia dei bisogni dei clienti, ma anche poiché evidenzia quanto essi siano importanti e come cambino nel corso del tempo. Inoltre, assume un'importanza cruciale perché potrebbe portare rilevare eventuali anomalie all'interno dei risultati analizzate.

Quindi, l'importanza di questo indicatore si rispecchia nell'evoluzione temporale che da l'informazione sull'evoluzione dei bisogni dei clienti nel tempo. Le determinanti di qualità possono essere classificate in base al loro valore di IMTP in 3 categorie, come si riporta in Figura 4 (Barravecchia et al., 2022):

- Determinanti stazionarie: elementi di qualità che mantengono una stabilità nel tempo, mantenendo costante la loro rilevanza e impatto sulla qualità senza variazioni significative.
- Determinanti con IMTP crescente nel tempo: si riferiscono agli elementi di qualità la cui importanza, misurata attraverso l'IMTP, cresce progressivamente all'aumentare del tempo. Questo suggerisce che tali determinanti diventano progressivamente più cruciali o influenti nella valutazione generale della qualità.

- Determinanti con IMTP decrescente nel tempo: Indica gli elementi di qualità la cui importanza, misurata tramite l'IMTP, diminuisce con il passare del tempo. Ciò implica che tali determinanti potrebbero ridursi in rilevanza o impatto nella valutazione complessiva della qualità nel corso del tempo.

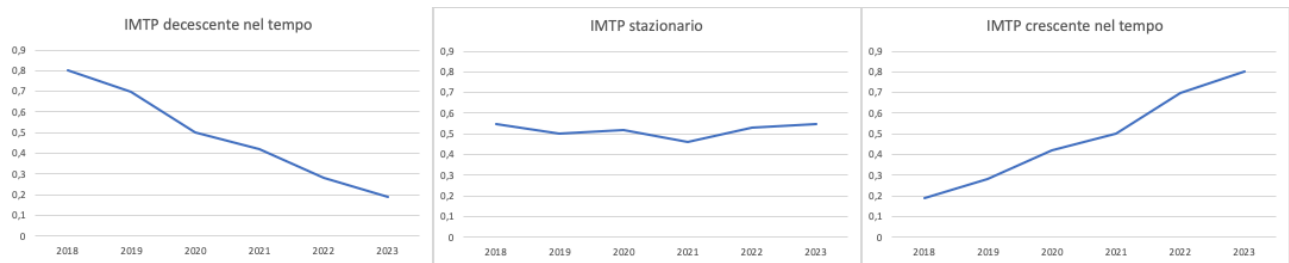


Figura 4. Profili IMTP.

CAPITOLO 2: APPLICAZIONE CASO STUDIO: SPOTIFY

INTRODUZIONE

Si intende applicare i concetti teorici esposti nel primo capitolo a un caso pratico, nello specifico al contesto di Spotify. Spotify rappresenta attualmente una delle principali piattaforme di streaming musicale, offrendo la riproduzione digitalmente di brani, podcast e video. Vanta un catalogo molto ampio, offrendo la possibilità ai propri utenti di accedere immediatamente a milioni di brani e milioni di contenuti provenienti da qualsiasi parte del mondo e da autori di ogni angolo del pianeta. Spotify si presenta come un'applicazione detta "freemium", ovvero che nasce come applicazione completamente gratuita (free) e offre la possibilità di accedere a contenuti esclusivi tramite la sottoscrizione di un abbonamento premium, per il quale bisogna pagare un corrispettivo mensile. Spotify offre inoltre la possibilità di effettuare un periodo di prova gratuito per il piano a pagamento che suddivide in abbonamento "Family", o in abbonamento "Student", per avere accesso ai piani a pagamento con tariffa ridotta. La differenza tra il piano gratuito e il piano a pagamento, corrisponde sostanzialmente nel fatto che l'utente con un piano gratuito possa ascoltare tutto il catalogo a disposizione, ma con l'obbligo di dover visualizzare inserzioni pubblicitarie e con il vincolo di non poter scegliere l'ordine in cui usufruire dei contenuti; questa modalità viene definita come "shuffle mode" e rappresenta una modalità casuale di ascolto non on-demand. L'utente con un piano gratuito inoltre, non può usufruire della funzione di download dei brani con possibilità di ascolto offline. Ad oggi, Spotify vanta più di 500 milioni di utenti attivi mensilmente. Focalizzandosi esclusivamente sugli utenti premium, registra oltre 210 milioni di abbonati paganti, che corrispondono ad una crescita media su base annua del 15%. Per condurre l'analisi di questo caso studio, si segue la metodologia presentata nel primo capitolo applicando passo dopo passo i concetti teorici proposti.

METODOLOGIA

ESTRAZIONE DEL SET DI DATI

In questa fase iniziale, è stato costituito un database di digital-VoC attraverso l'analisi di recensioni fornite dagli utenti. Il risultato dell'estrazione è rappresentato da un campione di UGC composto da 61257 recensioni e metadati di due categorie:

1. Data di pubblicazione della recensione.

2. Livello di rating con un valore che varia da 1 a 5 (dove 1 rappresenta il minimo livello di soddisfazione percepito e 5 il massimo). Durante l'analisi, per facilitare la visualizzazione e l'interpretazione, si useranno le stelle per indicare i valori discreti da 1 a 5.

Il periodo posto in analisi si estende da gennaio 2022 a luglio 2022 e tutte le recensioni sono redatte in lingua inglese. Un punto di forza del campione è la consistente mole di recensioni selezionate, aspetto che conferisce al campione un notevole valore, poiché consente di condurre un'analisi specifica e significativa.

Dall'analisi del database emerge che la maggior parte delle recensioni presenta valutazioni estreme, ossia 1 o 5 stelle. Nel 35,86% delle recensioni, la valutazione di rating ottenuta è la massima, con 5 stelle, seguita dalle valutazioni con 1 stella, presenti nel 28,69% dei casi.

La Tabella 6 fornisce una panoramica riassuntiva della distribuzione delle valutazioni delle recensioni all'interno del database.

I valori ottenuti e le percentuali di distribuzione delle valutazioni, in particolare quelli con 1 stella e 5 stelle, riflettono il comportamento tipico dell'utente medio, che tende a rilasciare una recensione principalmente quando è molto soddisfatto del servizio ricevuto o quando è fortemente insoddisfatto, usandola come forma di lamentela. Per una visualizzazione più agevole delle distribuzioni di valutazioni all'interno del database, la Figura 5 pone in relazione il numero delle recensioni con le valutazioni associate.

Tabella 6. Cardinalità e percentuale di recensioni con uno specifico livello di rating.

RATING	Cardinalità	%
★	17576	28,69%
★★	7094	11,58%
★★★	6840	11,17%
★★★★	7781	12,70%
★★★★★	21966	35,86%



Figura 5. Relazione tra livello di rating e numero delle recensioni.

PRE-PROCESSAMENTO

Dopo l'estrazione del set di dati, è stata condotta la fase di pre-processamento degli UGC. Il database è stato sottoposto a un processo di "ripulitura" e filtraggio, eliminando i termini ritenuti non significativi per l'analisi in questione.

Seguendo un approccio iterativo, mirato ad individuare un numero crescente di parole prive di significato nel documento analizzato (chiamate "stopwords") attraverso l'algoritmo utilizzato sono state rimosse un totale di 711 parole. Si è individuato un vocabolario finale di 2016 termini.

La Figura 6 presenta un output dell'algoritmo utilizzato, mostrando i risultati ottenuti in questa fase di pre-processamento. L'esito evidenzia il fatto che l'analisi sia stata condotta su un database con 61257 documenti validi, in cui si sono individuati 23188 termini e rimossi (o nella maggior parte dei casi normalizzati e ricondotti ad un'unica parola, grazie al processo di stemming) 21172. Inoltre 98 documenti sono risultati vuoti, ossia privi di qualsiasi termine.

```
Removing 21172 of 23188 terms (40985 of 620289 tokens) due to frequency
Removing 98 Documents with No Words
Your corpus now has 61257 documents, 2016 terms and 579304 tokens.
```

Figura 6. Applicazione del pre-processing con riduzione del database iniziale.

Un aspetto particolare, è stato la gestione delle emoticons, presenti in un gran numero di recensioni. Le emoticons sono state tradotte come parole virgolettate dal processore. È stata quindi prestata particolare attenzione per risalire alle parole corrispondenti alle emoticons più utilizzate per poi inserirle nella lista di stopwords. (Un esempio è la traduzione dell'emoticon del cuore con il termine "love").

SELEZIONE DEL NUMERO OTTIMALE DI TOPIC

In questa fase è stato determinato il numero di topic da estrarre mediante l'algoritmo. Questa selezione risulta particolarmente importante, poiché stabilisce la quantità di determinanti di qualità che spiegheranno il contenuto delle recensioni.

Per individuare il numero ottimale di topic K , è stato applicato l'algoritmo di STM. Come anticipato nel primo capitolo, si è preso in considerazione il parametro Held-Out Likelihood. Questo parametro riflette la capacità del modello tematico di descrivere efficacemente la variabilità complessiva nel corpus testuale, con un valore più elevato indicativo di una migliore performance del modello.

Il profilo ottenuto dall'Held-Out Likelihood in Figura 7, può essere chiaramente ricondotto ad un andamento quasi stazionario, infatti una volta raggiunto il primo massimo $K=11$, inizia ad assumere un comportamento molto simile all'aumentare di del numero di topic. Per scegliere la quantità K più corretta da estrarre, si è fatta una valutazione effettuando un trade-off tra la qualità del modello sulla base del numero di topic stessa e sulla complessità nel gestire il modello. Si presentavano infatti alternative che manifestavano un maggior valore di Held-Out Likelihood rispetto al numero di topic K scelto, ovvero 11, in particolare $K=28$ e $K=36$. Queste si sono scartate in quanto, per una variazione di circa 1% del parametro Held-Out Likelihood, presentavano un numero di topic maggiore rispettivamente al doppio e al triplo rispetto a quello inizialmente selezionato. Un numero così alto di topic, non è giustificato da un aumento così basso di Held-Out Likelihood. In definitiva, la scelta ritenuta più corretta è stata quella di selezionare un numero di topic ottimali per descrivere il corpus testuale di $K=11$.

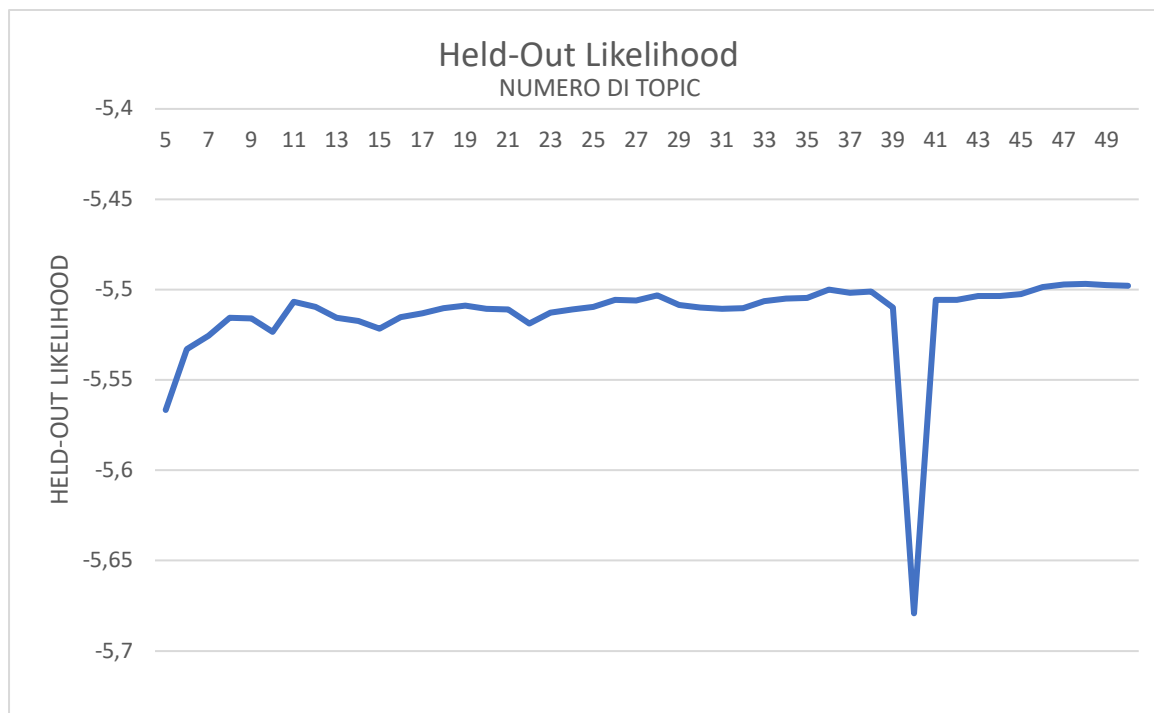


Figura 7. Relazione tra il valore dell'Held-Out Likelihood e il numero di topic.

ETICETTATURA DEI DATI

In questa fase, si procede all'etichettatura dei vari topic individuati, analizzandoli uno ad uno per identificare l'etichetta che li caratterizza al meglio.

In Tabella 7 si riportano i risultati di Highest Probability e FREX ottenuti con RStudio con valore K impostato sul valore ottenuto nella fase precedente, ovvero 11.

Prestando particolarmente attenzione sulle Highest Probability ottenute per ogni topic, ovvero tutte le parole chiave con più alta probabilità di essere discusse nel topic stesso, e sul FREX, indicante le parole chiave frequenti ed esclusive del singolo topic, si è effettuata una prima etichettatura manuale per ciascuno di essi. Questa fase mira principalmente a ridurre il carico cognitivo necessario per comprendere una determinante di qualità e per non lasciare spazio di interpretazione ai singoli, rendendo il processo più oggettivo. Nella Tabella 7 sono riportate le etichettature manuali finali assegnate per ogni topic.

Un aspetto di particolare interesse in questa fase del lavoro è stata la modifica all'etichettatura relativa ai topic 3 e 8, inizialmente etichettati rispettivamente come “positive feedback” e come “advertising” e solo successivamente ridefiniti “App rating: section music” e “Song Accessibility”. È evidente che il focus posto inizialmente sui due topic in esame, si è rivelato totalmente in contrasto con il risultato finale. Per giustificare questo cambiamento, è stata condotta un'analisi qualitativa su un piccolo

campione di recensioni contenenti i topic in questione, basandosi sui risultati ottenuti nella fase di validazione che verrà presentata nel paragrafo successivo.

Tabella 7. Etichettatura finale dei topic con rispettive keywords associate.

Topic	Criterio	Keywords	Etichettatura
1	Highest Probability	work, download, problem, connect, app, offlin, lyric, phone, internet, load	Connection problem: wi-fi
	FREX	internet, connect, wi-fi, offlin, load, storag, mode, network, data, lyric	
2	Highest Probability	song, play, playlist, shuffl, random, search, skip, list, differ, album	Subscription not premium: shuffle
	FREX	shuffl, list, playlist, click, order, repeat, search, pick, instead, specif	
3	Highest Probability	music, app, spotifi, experi, youtub, enjoy, kind, feel, daili, place	App rating: section music
	FREX	enjoy, fantast, appl, world, experi, relax, amazon, hand, know, conveni	
4	Highest Probability	premium, pay, month, version, subscript, servic, money, spotifi, cancel, buy	Payments and fees
	FREX	premium, plan, price, money, buy, subscript, pay, trial, membership, paid	
5	Highest Probability	app, play, updat, stop, fix, keep, paus, phone, crash, bug	Bug user experince/ app reliability
	FREX	paus, bar, crash, disappear, close, stop, control, bug, bottom, latest	
6	Highest Probability	podcast, spotifi, support, episod, littl, bit, watch, allow, video, custom	Podcast and information
	FREX	misinform, episod, video, littl, speech, bit, spread, drive, compani, neil	
7	Highest Probability	use, spotifi, app, year, user, sinc, friend, function, applic, improv	User friendly
	FREX	everyday, interfac, applic, smooth, navig, brillant, use, impress, intuit	
8	Highest Probability	listen, song, app, add, star, annoy, minut, rate, skip, advertis	Songs accessibility

Topic	Criterio	Keywords	Etichettatura
	FREX	min, listen, advertis, preview, irrit, uninterrupt, minut, add, interrupt, gave	
9	Highest Probability	spotifi, find, reccommend, artist, new, select, stream, hear, avail, old	Huge selection
	FREX	varieti, stream, choic, collect, find, genr, artist, wide, high, rang	
10	Highest Probability	tri, account, log, qualiti, sound, help, audio, someth, spotifi, wrong	Account and log-in problem
	FREX	email, password, gmail, username, email, valid, sent, mail, facebook, atmo	
11	Highest Probability	app, time, chang, feature, spotifi, make, thing, option, may, reason	Features updates
	FREX	thing, chang, sort, wors, small, abil, featur, actual, histori, besid	

VERIFICA DEI DATI

In questa fase del lavoro, sono stati validati i risultati, ovvero è stato valutato se il modello generato dall' algoritmo di *topic modeling* rappresentasse correttamente il contenuto dei documenti analizzati. Per effettuare questa validazione, sono stati confrontati i risultati ottenuti da una valutazione manuale su un campione di recensioni n (con n=100) estratte casualmente dal data set analizzato, con i risultati generati automaticamente sulla base dell' algoritmo di topical prevalence.

Dal confronto sono emersi i valori di:

- True positive (tp);
- False positive (fp);
- False negative (fn);
- True negative (tn).

Attraverso questi valori sono stati costruiti gli indicatori riportati in Tabella 8, con le Formule da (7) a (18) riportate nel primo capitolo. Per semplicità di visualizzazione, i valori che rispettano i valori target sono stati evidenziati con il colore verde, quelli fuori dai target invece di colore rosso. Per il calcolo degli indicatori sono state utilizzate le formule presentate nel primo capitolo.

Tabella 8. Primi valori ottenuti dalla validazione dei risultati.

		Target
ACCURACY	0,88	>0.95
PRECISION	0,89	>0.70
FALSE DISCOVERY RATE	0,11	<0.05
FALSE OMISSION	0,06	<0.05
NEGATIVE PEDICTIVE	0,94	>0.90
RECALL	0,59	>0.70
FALL-OUT	0,01	<0.05
MISS RATE	0,41	<0.20
SPECIFICITY	0,99	>0.90
F1 SCORE	0,71	>0.70

Da un primo confronto si sono ottenuti risultati non in linea con quelli attesi e con i valori target, come si può notare dalla presenza delle celle colorate in rosso in Tabella 8, che rappresentano appunto i valori fuori target.

Per risolvere questa problematica sono stati analizzati i valori più critici, risalendo ad un elevato valore di False negative. In questo caso l'assegnazione automatica dei topic ai documenti in analisi non riconosce un topic che invece è presente, ovvero è stato rilevato dalla assegnazione di riferimento manuale. Si va incontro all'errore di seconda specie. Questo è stato considerato la causa scatenante delle criticità emerse. In particolare, esaminando più in profondità le principali problematiche, emergevano nei topic 3 e 8, apparentemente soggetti ad un'etichettatura non corretta. Per ovviare a ciò è stata effettuata una nuova etichettatura per i topic indagati e si osservano i nuovi risultati.

Effettuando ex novo il confronto con le nuove etichettature, i risultati si presentano notevolmente più soddisfacenti come si evince in Tabella 9.

Tabella 9. Valori ottenuti dalla validazione finale.

		Target
ACCURACY	0,95	>0.95
PRECISION	0,89	>0.70
FALSE DISCOVERY RATE	0,11	<0.05
FALSE OMISSION	0,05	<0.05
NEGATIVE PEDICTIVE	0,95	>0.90
RECALL	0,65	>0.70
FALL-OUT	0,01	<0.05
MISS RATE	0,35	<0.20
SPECIFICITY	0,99	>0.90
F1 SCORE	0,75	>0.70

Gli unici valori che non rispettano i valori target, nonostante la nuova etichettatura dei topic errati, sono il *false discovery rate*, che dovrebbe presentare un valore inferiore a 0,05, e il *miss rate*, che

dovrebbe presentare un valore inferiore a 0,20. Il valore errato del primo indicatore fa emergere l'identificazione di più topic tramite l'assegnazione automatica rispetto all'assegnazione manuale. Un valore elevato di *miss rate*, pur essendo diminuito rispetto alla prima analisi, sottolinea la presenza dell'errore di seconda specie, con l'assegnazione automatica che non riesce a cogliere un topic effettivamente presente con l'assegnazione manuale.

ANALISI DEI RISULTATI

In questa sezione si analizza la digital VoC, ponendo particolare attenzione sulle determinanti di qualità individuate nelle fasi precedenti e sulle topical prevalence dei documenti analizzati.

Per analizzare i risultati emersi, si sono calcolati i valori di MTP, MRP e IMTP presentati nel primo capitolo.

MTP-Mean Topic Proportion

Il primo indicatore calcolato è l'MTP che fornisce quanto in media è discusso un topic all'interno della base di dati analizzata.

Come si evince dai valori in Tabella 10 e dalla Figura 8 in questo caso i topic più discussi risultano essere il topic 2: "Subscription not premium: shuffle", il topic 5: "Bug user experience/app reliability" e il topic 8: "songs accesibility". Bisogna tuttavia osservare che il database tratta i diversi topic individuati in maniera comunque abbastanza omogenea.

Tabella 10. Valori MTP per ogni topic.

TOPIC LEGENDA	
1	Connection problem: wifi
2	Abbonamento non premium: randomizzazione (Subscription not premium:shuffle)
3	App rating: section music
4	Tariffe e i pagamenti (Payments and fees)
5	Bug relativi all'user experience/ affidabilità dell'app (app reliability)
6	Uso di spotify per podcast e informazione (Podcast and information)
7	User friendly
8	Accessibilità ai brani (songs accessibility)
9	Stop selection/ huge selection / varietà catalogo
10	Account and log in problem
11	Aggiornamento features (update)

Topic	1	2	3	4	5	6	7	8	9	10	11
MTP	0,09	0,12	0,10	0,08	0,15	0,06	0,07	0,11	0,10	0,06	0,05

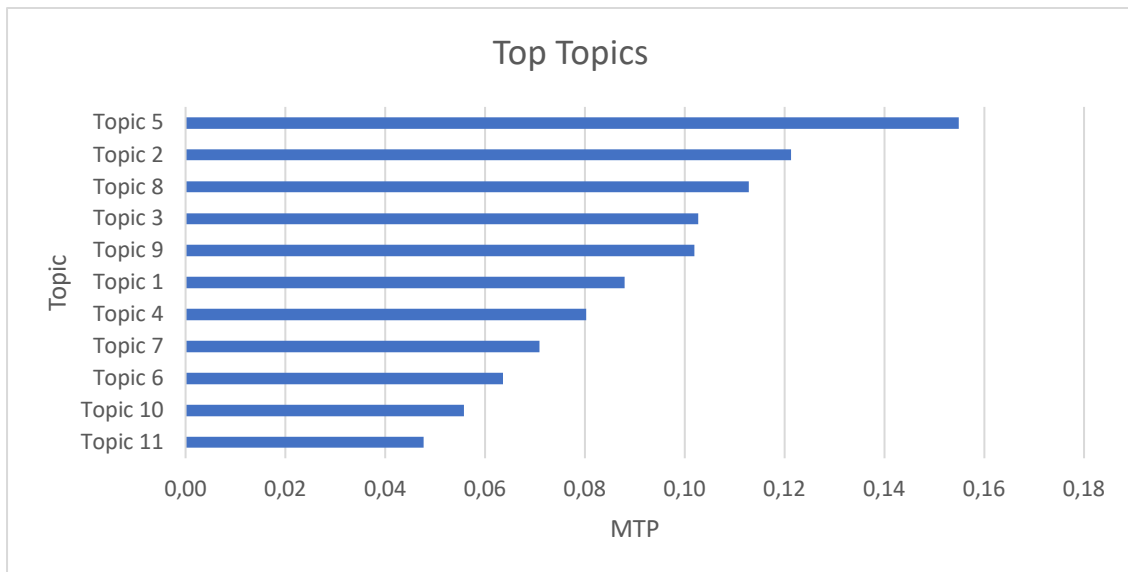


Figura 8. Livello di discussione medio dei topic (MTP).

MRP-Mean Rating Proportion

Il secondo indicatore calcolato è l'MRP che dice in media quanto è discusso un topic nei record di digital VoC associati con uno specifico livello rating.

Osservando i valori riportati in Tabella 11 e conseguentemente riportati in Figura 9 si può notare come analizzando il topic con maggior discussione, ovvero il 5, relativo ai problemi legati all'user experience e all'affidabilità dell'applicazione, questo sia associato tendenzialmente ad un rating negativo, ovvero sia trattato principalmente come elemento di poca soddisfazione per i clienti.

Analogo ragionamento si può effettuare per il topic relativo all'abbonamento non premium e alla caratteristica di ascolto shuffle, trattato principalmente in termini negativi dagli utenti.

Osservando i successivi topic più discussi, rispettivamente 8 (che tratta dell'accessibilità ai brani), 3 (che tratta dell'app rating: la sezione musica) e 9 (che tratta dell'ampia varietà del catalogo) sono associati per lo più a recensioni con rating positivo.

Emerge chiaramente come nel complesso gli utenti di Spotify siano soddisfatti dell'esperienza offerta sia dal punto di vista dei contenuti (topic 9) sia dal punto di vista dell'utilizzabilità (topic 3 e 7) e che le recensioni negative interessino per lo più aspetti legati all'affidabilità e dei bug riscontrati durante la fruizione dei contenuti come nel caso del topic 5.

Soffermendosi sul topic 2, il quale descrive l'aspetto relativo alla randomizzazione dei brani per gli utenti "non premium", si può osservare come questo presenti un profilo dell'MRP negativo. Il profilo negativo per questo topic non rispecchia un campanello d'allarme per Spotify, in quanto è l'azienda stessa ad inserire, volutamente, un aspetto che generi insoddisfazione per il cliente non abbonato alla versione premium in modo da stimolare il passaggio all'abbonamento premium.

Sintetizzando, il topic 2 risulta meritevole di un approfondimento in quanto il rating negativo al quale è associato è frutto di una strategia aziendale e non di una faglia nella qualità del servizio. La strategia aziendale è quella di incentivare gli utenti a sottoscrivere l'abbonamento premium inserendo degli ostacoli che aumentano il numero di detrattori nella versione gratuita.

Tabella 11. Valori MRP per ogni topic.

TOPIC LEGENDA											
1-Connection problem: wifi											
2-Abbonamento non premium: randomizzazione (Subscription not premium:shuffle)											
3-App rating: section music											
4-Tariffe e i pagamenti (Payments and fees)											
5-Bug relativi all'user experience/ affidabilità dell'app (app reliability)											
6-Usò di spotify per podcast e informazione (Podcast and information)											
7-User friendly											
8-Accessibilità ai brani (songs accessibility)											
9-Stop selection/ huge selection / varietà catalogo											
10-Account and log in problem											
11-Aggiornamento features (update)											

MRP											
Topic	1	2	3	4	5	6	7	8	9	10	11
★	0,10	0,13	0,07	0,09	0,20	0,07	0,06	0,10	0,06	0,07	0,05
★★	0,10	0,14	0,06	0,08	0,24	0,06	0,06	0,10	0,06	0,05	0,05
★★★	0,10	0,15	0,07	0,08	0,20	0,06	0,06	0,11	0,07	0,05	0,05
★★★★	0,09	0,13	0,10	0,08	0,13	0,06	0,07	0,13	0,11	0,05	0,05
★★★★★	0,07	0,10	0,15	0,08	0,09	0,06	0,09	0,12	0,16	0,05	0,05

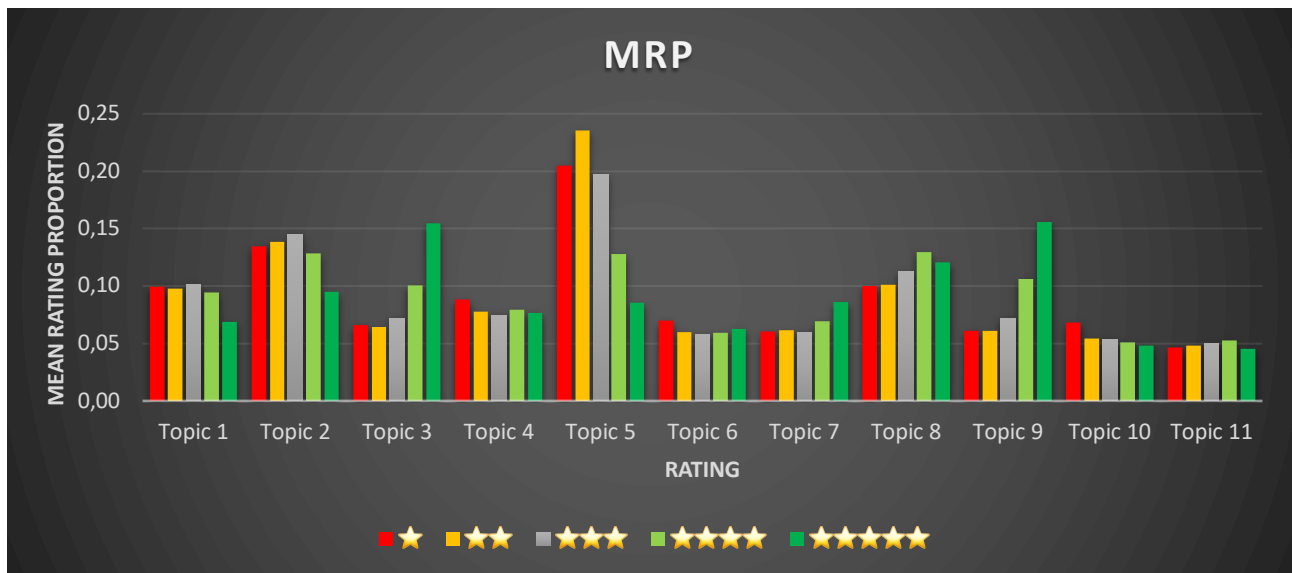


Figura 9. MRP per ogni topic sulla base del rating e del livello di discussione.

Analizzando questi indicatori, anche in maniera incrociata, si può dedurre come gli utenti abbiano una valutazione nel complesso positiva dell'app. Ciò è testimoniato dalla correlazione con rating positivo che i topic relativi ad app e contenuti presentano. Tuttavia si può notare, riprendendo la Figura 9, che il topic 5 che fa riferimento all'affidabilità dell'app, sia il più discusso. Si giustifica

questa evidenza considerando che l'utente percepisca una maggiore necessità di recensire nel momento in cui si imbatte in problematiche piuttosto che durante un'esperienza piacevole.

Questo viene ulteriormente enfatizzato nei servizi digitali, in quanto, in questo ambito, la recensione risulta essere anche uno strumento di segnalazione delle problematiche alla piattaforma oltre che uno strumento di pubblicità (positiva o negativa).

Affinché un utente effettui una recensione positiva deve innescarsi un meccanismo di engagement che stimoli il soggetto a dedicare tempo al positive feedback, mentre risulta spontaneo ed immediato effettuare recensioni negative nel momento in cui ci si imbatte in delle problematiche, proprio per il duplice valore che assume la recensione in questo ambito, ossia di feedback e di segnalazione.

IMTP-Interval Mean Topical Prevalence

Questo indicatore permette di comprendere come varia il livello di discussione dei topic nel tempo, in questo modo fornisce una fotografia dell'evoluzione dei bisogni dei clienti nel tempo e consente di individuare eventi anomali e/o malfunzionamenti.

Per condurre l'analisi, si possono utilizzare strumenti come le carte di controllo o altri metodi di individuazione di outliers.

In questo caso in Tabella 12 e nel grafico plottato in Figura 10 si può osservare un andamento regolare per numerosi topic, e solo per alcuni, si evidenziano anomalie.

Incrociando questa analisi con quella dell'MRP, si può notare come l'anomalia si presenta per i topic che hanno una distribuzione non neutrale, ovvero i topic che hanno una forte correlazione con rating o positivo o negativo. Si ha infatti, per i topic 5 (relativo ai bug relativi all'user experience e problemi di affidabilità dell'app), 8 (relativo all'accessibilità ai brani), 9 (relativo alla varietà del catalogo) e 10 (relativo ai problemi di account e log-in) il palesarsi di un'anomalia positiva che comporta l'aumento di discussione del topic 5 e 10 e un'anomalia negativa con riduzione di discussione per gli altri topic (8,9). Inoltre sempre incrociando l'analisi con l'indicatore MRP si evidenzia come i topic 5 e 10 sono legati a rating negativi mentre 8 e 9 a rating positivi. Dall'analisi del grafico in Figura 12 emerge chiaramente che nel mese di aprile si è verificato uno shock negativo.

Approfondendo la ricerca per risalire alle cause di queste anomalie, si è individuato in Figura 11 il lancio di 3 aggiornamenti nel mese di aprile. Si suppone che probabilmente il primo aggiornamento abbia creato problemi per gli utenti e le 2 versioni successive siano state lanciate proprio per risolvere i problemi; le evidenze dimostrano il successo di questa manovra attraverso un annullamento dello shock e un ritorno a condizioni normali nel mese di maggio.

Tabella 12. IMTP per ogni topic nel tempo.

<p style="text-align: center;">TOPIC LEGENDA</p> <p style="text-align: center;">1-Connection problem: wifi 2-Abbonamento non premium: randomizzazione (Subscription not premium: shuffle) 3-App rating: section music 4-Tariffe e i pagamenti (Payments and fees) 5-Bug relativi all'user experience/ affidabilità dell'app (app reliability) 6-Usò di spotify per podcast e informazione (Podcast and information) 7-User friendly 8-Accessibilità ai brani (songs accessibility) 9-Stop selection/ huge selection / varietà catalogo 10-Account and log in problem 11-Aggiornamento features (update)</p>											
IMTP											
Topic	1	2	3	4	5	6	7	8	9	10	11
Gennaio	0,08	0,12	0,10	0,09	0,12	0,09	0,07	0,12	0,11	0,05	0,05
Febbraio	0,09	0,12	0,10	0,08	0,12	0,08	0,07	0,12	0,11	0,05	0,05
Marzo	0,09	0,12	0,10	0,08	0,13	0,06	0,07	0,11	0,10	0,09	0,05
Aprile	0,08	0,11	0,09	0,07	0,25	0,06	0,07	0,10	0,09	0,05	0,05
Maggio	0,09	0,13	0,11	0,08	0,14	0,06	0,07	0,12	0,11	0,05	0,05
Giugno	0,09	0,13	0,11	0,08	0,13	0,06	0,07	0,12	0,11	0,05	0,05
Luglio	0,10	0,12	0,11	0,08	0,14	0,06	0,07	0,12	0,10	0,05	0,05

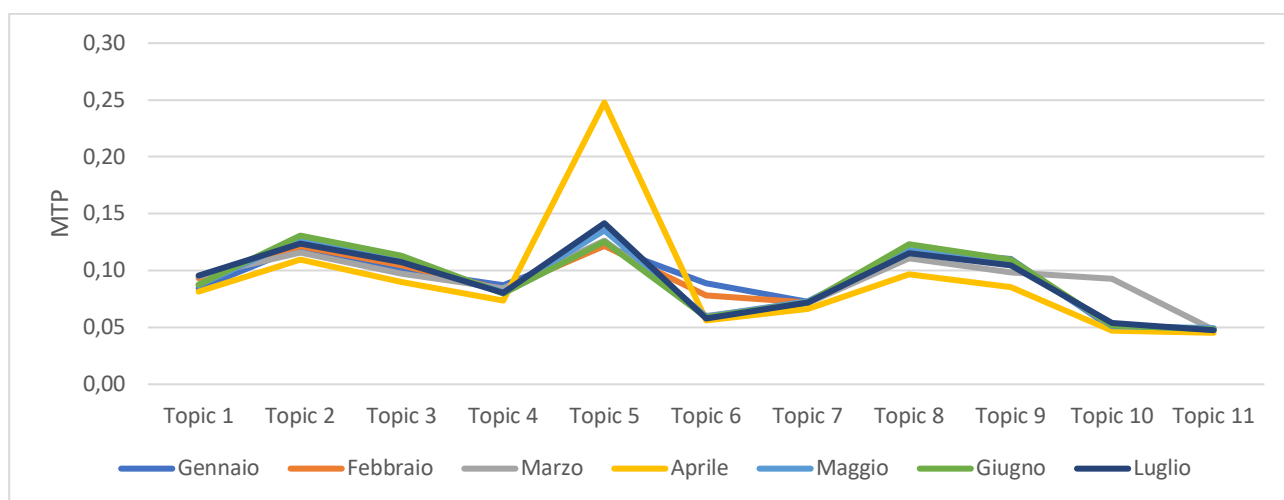


Figura 10. IMTP nel corso dei mesi.

VERSIONI PRECEDENTI

<p>8.7.26.901</p> <p>APK</p> <p>05/05/2022</p>	<p>70.35 MB</p>	<p>8.7.24.1111</p> <p>APK</p> <p>29 apr 2022</p>	<p>69.97 MB</p>	<p>8.7.22.1125</p> <p>APK</p> <p>18 apr 2022</p>	<p>69.88 MB</p>
<p>8.7.20.1261</p> <p>APK</p> <p>8 apr 2022</p>	<p>69.57 MB</p>	<p>8.7.18.1138</p> <p>APK</p> <p>31 mar 2022</p>	<p>69.56 MB</p>	<p>8.7.16.1354</p> <p>APK</p> <p>24 mar 2022</p>	<p>69.24 MB</p>

Figura 11. Aggiornamenti di Spotify rilasciati nei mesi analizzati.

CAPITOLO 3: MONITORAGGIO NEL TEMPO DELLA RELAZIONE TRA DETERMINANTI DI QUALITÀ E VALUTAZIONI DEI CLIENTI

In questo capitolo, si vogliono analizzare gli strumenti per il monitoraggio nel tempo della relazione tra le determinanti di qualità e le valutazioni dei clienti, applicandoli al caso studio.

Controllare e monitorare nel tempo l'evoluzione dei bisogni e delle aspettative dei clienti risulta è fondamentale per ogni azienda, in quanto consente di comprendere a fondo ciò che il cliente desidera e ciò che percepisce come di qualità, influenzando la sua decisione di acquistare il servizio offerto. Il monitoraggio continuo costituisce anche la base del meccanismo di retention, attraverso il quale un cliente riutilizza il servizio offerto e attraverso cui si instaura una relazione di fedeltà con l'azienda. Attraverso questa pratica, l'azienda è in grado di valutare la reazione del cliente a determinate innovazioni, fornendo una misura di riferimento per valutare la direzione del proprio sviluppo.

In questo capitolo, la tecnica che si adotta è il profile monitoring, che consente il monitoraggio nel tempo di una specifica caratteristica. In molte applicazioni, la qualità di un processo si caratterizza e si spiega attraverso una relazione funzionale tra una variabile dipendente e una o più variabili indipendenti esplicative. Il profile monitoring, o monitoraggio dei profili, si impiega per comprendere e verificare la stabilità di questa relazione nel corso del tempo (Woodall 2007).

In particolare, si utilizzerà dapprima l'indicatore IMTP e successivamente l'MRP (Barravecchia et al., 2020). L'obiettivo è di individuare uno strumento statistico per il monitoraggio dell'indicatore MRP, che funzioni come un "campanello d'allarme", segnalando in caso di valori che si discostano dalla distribuzione dei dati attesa. Per raggiungere questo obiettivo, si condurrà un'analisi approfondita dell'indicatore stesso, cercando i valori che si possono considerare outliers e quindi anomali.

Per lo sviluppo di questo capitolo, si utilizzeranno nuovamente i dati estratti nel secondo capitolo riguardanti il caso studio Spotify.

PROFILE MONITORING

Il profile monitoring, o il monitoraggio dei profili, rappresenta uno strumento essenziale per valutare la stabilità della relazione tra una variabile dipendente e una o più variabili indipendenti esplicative nel corso del tempo. Al fine di studiare la qualità all'interno di un processo o durante la realizzazione di un prodotto, si creano dei grafici o si utilizzano strumenti che evidenzino la relazione tra una variabile di risposta e le altre variabili che spiegano la variazione.

Per realizzare questo obiettivo, si generano “curve”, ovvero i profili che permettono di osservare questa relazione. In molte applicazioni, il modello di regressione lineare è lo strumento più utilizzato (Chang e Gan, 2006; Zhu e Lin, 2007; Norossana et al., 2007). Questo modello, si presenta come una tecnica statistica relativamente semplice ed è utilizzata in moltissime applicazioni. Questa tecnica permette di modellizzare la relazione tra una variabile dipendente (d’interesse/target) con una o più variabili indipendenti (che la spiegano). Essendo una regressione lineare, il suo compito principale è ricercare una relazione lineare tra le variabili coinvolte. Mediante essa, si possono effettuare previsioni o stime sulla variabile dipendente partendo dai valori della variabile indipendente. Nel corso dello studio si considera l’utilizzo della regressione lineare per studiare i parametri indagati.

Nelle prossime sezioni si utilizzano 3 diverse metodologie:

- Le Carte di controllo (in particolare le carte X-S e le carte p);
- Il Test di Grubbs;
- Test di correlazione di Spearman.

CARTE DI CONTROLLO

Le carte di controllo sono strumenti molto utili per comprendere al meglio lo stato di un processo e la sua evoluzione nel tempo (Shewart, 1924). Questi strumenti infatti, permettono di controllare e monitorare nel tempo determinate caratteristiche di un processo, oppure nel caso analizzato nel secondo capitolo, permettono di monitorare i livelli di discussione dei topic individuati mediante l’algoritmo di topic modelling (ovvero le determinanti di qualità del processo in analisi).

Le carte di controllo sono composte da due assi cartesiani: sull’asse delle ascisse si riportano i valori misurati del campione in analisi seguendo l’ordine temporale (nel caso analizzato saranno i mesi); sull’asse delle ordinate si riporteranno i valori della caratteristica misurata. Si hanno tre linee parallele all’asse delle ascisse rappresentati rispettivamente:

- LCS-ovvero il limite superiore di controllo.
- LC-ovvero il limite centrale di controllo.
- LCI-ovvero il limite inferiore di controllo.

Queste tre linee determinano quattro zone all’interno dello spazio rappresentante la carta di controllo. Se i valori misurati ricadono all’interno di una delle due zone comprese tra LCS e LC, o tra LC e LCI, allora il processo si dice “in controllo”, viceversa se i valori misurati ricadono all’esterno del

LCS o del LCI allora il processo si dice “fuori controllo”. Si ha una variabilità intrinseca del processo che porta alla fluttuazione casuale dei punti all’interno dei limiti individuati. Qualora i valori ricadessero al di fuori dei limiti di controllo oppure, seppur rimanendo entro i limiti di controllo evidenziassero andamenti anomali non casuali, bisognerebbe intervenire con azioni correttive volte a migliorare i processi.

Le carte di controllo consentono di osservare il funzionamento di sistemi manifatturieri o di sistemi di servizi. Si distinguono dai piani di campionamento per via del fattore “tempo”. I primi infatti osservano i risultati di un processo già avvenuto, mentre le carte di controllo consentono di osservare i risultati di un processo in corso.

Si hanno due fasi principali per la costruzione delle carte di controllo (Franceschini e Galetto, 2011):

- Fase di impianto, ovvero “come si costruisce la carta” in cui si identificano la centratura e la tolleranza naturale del processo, in cui quindi si identificano i limiti di controllo. Nell’analisi della digital VoC in questa fase si costruiranno i limiti specifici per ogni tipologia di topic individuato.
- Fase di monitoraggio (o di controllo), ovvero la valutazione dei valori misurati, in cui si osserva se un processo è “in controllo” o “fuori controllo”. Nell’analisi della digital VoC emergono i topic che si discostano dai comportamenti tipici e sui quali si devono effettuare azioni correttive.

Le carte di controllo sono un’applicazione nel tempo del test di ipotesi. Inoltre, bisogna notare come un processo “in controllo” non sia equivalente a dire che le specifiche tecniche del prodotto o del servizio in analisi siano rispettate, infatti un processo “in controllo” indica solamente che un processo è stabile nel tempo, ma potrebbe essere fuori specifica. Si possono anche avere delle anomalie di processo, ovvero un qualsiasi comportamento che si allontani dalle condizioni di variabilità naturale del processo. Per individuare le condizioni di “fuori controllo” si utilizzano le Regole Western Electric (Western Electric 1956; Montgomery 2009). Quando si è in questa situazione, bisogna andare ad interrogarsi su quelle che possono essere le cause che hanno portato a questa condizione e bisogna far in modo di rimuoverle effettuando azioni correttive e andando a rendere efficienti i processi. In questo studio, si usano le carte di controllo per variabili X-S e le carte di controllo per attributi p.

CARTE X-S

Le carte X-S sono carte di controllo per variabili e lavorano in coppia, poiché hanno funzioni complementari: la carta S analizza la variabilità all'interno del campione, mentre la carta X misura la variabilità tra campioni. In sostanza, la carta X si occupa del valore centrale di un processo, mentre la carta S si occupa della sua dispersione (Montgomery, 2009). Questa carta si ritiene idonea quando la dimensione campionaria è grande ($n > 10$) o quando la dimensione del campione è variabile (Montgomery, 2009). Si utilizzeranno nello studio per l'analisi della topical prevalence e di come questa vari nel corso del tempo.

La carta X media si presenta nel seguente modo in Formula (25):

$$\begin{cases} LCS_X = X + A_3 * S \\ LC_X = X \\ LCI_X = X - A_3 * S \end{cases} \quad (25)$$

La carta S si presenta nel seguente modo in Formula (26):

$$\begin{cases} LCS_S = B_4 * S \\ LC_S = S \\ LCI_S = B_3 * S \end{cases} \quad (26)$$

Essendo il numero di recensioni in ogni periodo variabile, varieranno anche i limiti di controllo, quindi ogni punto del processo avrà limiti propri.

Essendo i campioni di recensioni analizzati con numerosità diverse le Formule (27) e (28) per calcolare rispettivamente la media X e lo scarto tipo (o deviazione standard) S saranno le seguenti:

$$X = \frac{\sum_i^k n_i * X_i}{k} \quad (27)$$

$$S^2 = \frac{\sum_i^k (n_i - 1) * s_i^2}{\sum_i^k n_i - k} \quad (28)$$

Modificando il valore di n , cambiano i valori di $A3$, $B4$ e $B3$, che sono parametri che cambiano sulla base della dimensione del campione n e sono indicati sulla Tavola dei parametri per il calcolo dei limiti delle carte di controllo (Montgomery, 2009). Modificando questi valori, cambiano i valori dei limiti. Più sarà grande il valore di n più i limiti si avvicineranno al valore centrale. Per la carta S si effettuerà la sottrazione di un'unità in quanto rappresenta la perdita di un grado di libertà per aver calcolato la media del campione.

CARTE P

Per analizzare la topical prevalence si possono utilizzare anche le carte p . Si tratta di carte di controllo per attributi. Solitamente queste carte trattano informazioni più povere rispetto alle carte di controllo per variabili, infatti trattano caratteristiche che si possono esprimere attraverso un numero finito di stati (Montgomery, 2009).

Per utilizzare questa carta nello studio, si fa un'assunzione piuttosto importante: si stabilisce una determinante di qualità "vincente" tra quelle individuate, ovvero quella determinante di qualità che presenta la topical prevalence maggiore rispetto a tutte le altre (Barravecchia et al., 2022). A questa determinante "vincente" verrà assegnato un punteggio di 1, mentre a tutte le restanti un punteggio di 0. In questo modo la determinante "vincente" ritrarrà la caratteristica più rappresentativa della recensione contenuta all'interno della digital VoC. A partire dalle singole recensioni del dataset analizzato, si fa la forte assunzione di considerare la singola recensione associata esclusivamente a una delle 11 determinanti di qualità trovate, ovvero quella con la topical prevalence più alta. Chiaramente, questa assunzione, risulta essere piuttosto forte in quanto la maggioranza delle recensioni tratta più di una singola determinante.

Per ogni recensione j -esima presente nella digital VoC si effettuerà quindi la seguente trasformazione binaria in Formula (29) (con TP = topical prevalence del j -esimo topic e del d -esimo documento):

$$R_j \begin{cases} 1 & \text{se } TP_{j,d} = \text{MAX}_j(TP_{j,d}) \\ 0 & \text{se } TP_{j,d} \neq \text{MAX}_j(TP_{j,d}) \end{cases} \quad (29)$$

Quindi, la sommatoria relativa alla topical prevalence di una specifica recensione sarà pari all'unità e sarà dovuta interamente alla determinante "vincente" in Formula (30):

$$\sum_{d=1}^D R_{j,d} = 1 \quad \forall j \in (1, \dots, J) \quad (30)$$

Per trovare i limiti di controllo di questa carta si utilizza la seguente Formula (31):

$$\left\{ \begin{array}{l} LCS_p = p + 3 * \sqrt{\frac{p(1-p)}{n}} \\ LC_p = p \\ LCI_p = p - 3 * \sqrt{\frac{p(1-p)}{n}} \end{array} \right.$$

(31)

TEST DI GRUBBS

Come precedentemente menzionato, uno degli obiettivi di questo capitolo consiste nello studiare il comportamento dell'indicatore MRP nel tempo e identificare uno strumento statistico in grado di fungere da "campanello d'allarme" nel caso in cui si osservi un valore che si discosti dalla normale distribuzione dei dati.

Lo strumento selezionato per questo scopo è il "test di Grubbs" (Grubbs E.Frank, 1950). Questo test si rivela particolarmente idoneo all'analisi in corso, poiché consente di individuare gli outliers all'interno dei dati MRP. Un aspetto rilevante di questo test è la sua natura univariata che verifica se un valore è significativamente diverso dagli altri valori presenti in un set di dati, esaminandolo singolarmente.

Trova applicazione in svariati campi in cui è utilizzato per individuare valori outliers all'interno di un set di dati univariati. Questo test è applicabile in contesti come:

- Controllo qualità, consente di identificare eventuali problemi nei processi di produzione (Montgomery, 2009).
- Ricerca scientifica, in esperimenti o raccolte dati identificando valori errati che potrebbero distorcere i risultati.
- Ambito medico, per individuare alcuni valori anomali in laboratorio o per identificare esami fuori scala o incoerenti.
- Scienze ambientali, per effettuare alcune misurazioni a livello ambientale quali la qualità dell'aria, dell'acqua o nella misurazione del clima.
- Finanza, identifica anomalie nell'utilizzo di determinati strumenti finanziari, di indicatori economici o di performance (Chan et al., 2015).

Per applicare questo test, si procede con la formulazione delle ipotesi, denotando con H_0 l'ipotesi nulla e con H_1 l'ipotesi alternativa in Formula (32):

$$\begin{cases} H_0: \text{Il valore in questione non è un outlier.} \\ H_1: \text{Il valore in questione è un outlier.} \end{cases} \quad (32)$$

Si prosegue calcolando la statistica di Grubbs, utilizzando la seguente Formula (33):

$$G = \frac{|X - \text{mean}(X)|}{s} \quad (33)$$

Dove:

- X è il valore che si sta testando.
- $\text{mean}(X)$ è la media dei dati.
- s è la deviazione standard dei dati.

Successivamente, si determina il valore critico G_C dal test di Grubbs in base al livello di significatività (α) desiderato e al numero di dati (n). Questo processo coinvolge l'utilizzo della distribuzione t di Student (Student, 1908), con un determinato livello di significatività scelto arbitrariamente e un numero specifico di gradi di libertà. I gradi di libertà dipendono dalla dimensione del campione. Nel caso in esame, si ha un campione di dimensione 11, ovvero il numero di topic individuati, a cui si andranno a sottrarre 2 gradi di libertà, usati per determinare il valor medio del campione e la deviazione standard, per ottenere un totale di 9 gradi di libertà. La Formula (34) per calcolare il valore critico G_C è la seguente:

$$G_C = \frac{(n-1)}{n} * \sqrt{\frac{t^2}{(n-2+t^2)}} \quad (34)$$

Dove:

- n è la dimensione del campione in analisi, quindi $n=11$ ovvero il numero di topic;
- t è il valore della distribuzione t di Student associato al livello di significatività $\alpha/2$ e al numero di gradi di libertà (dof=degree of freedom), ovvero 9.

Una volta determinato il valore critico G_C , si confrontano tutti i valori di G ottenuti con esso. Se il valore di G è maggiore del valore critico, è possibile rifiutare l'ipotesi nulla e considerare il valore come un outlier. L'identificazione di un valore come outlier attiva il campanello d'allarme, segnalando un'anomalia nei dati, e di conseguenza, all'interno del processo.

TEST DI CORRELAZIONE DI SPEARMAN

Questa metodologia è stata impiegata per esaminare la possibile correlazione tra il numero di valutazioni dei clienti e i corrispondenti punteggi associati alle recensioni. È importante notare che le valutazioni dei clienti sono state raccolte su scale ordinali con punteggi che variano da 1 a 5. Al fine di analizzare la correlazione tra i punteggi e il numero di valutazioni dei clienti, considerando la natura ordinale dei dati, è stato adottato un approccio non parametrico, in particolare il test di correlazione di Spearman. Questo test è idoneo per valutare la relazione tra due variabili ordinali, prescindendo dalla scala di misurazione effettiva dei dati.

È importante sottolineare che tale test fornisce informazioni sulla presenza di una correlazione monotona tra il numero di valutazioni dei clienti e i punteggi, senza specificare la natura precisa di questa relazione (lineare, curvilinea, ecc.). Tuttavia, è uno strumento utile per individuare un possibile legame tra le variabili in questione.

È cruciale notare che questo test da solo non è in grado di stabilire relazioni di causa-effetto o di spiegare le motivazioni sottostanti ai risultati osservati. Pertanto, è consigliabile interpretare i risultati in sinergia con altri strumenti analitici, come l'analisi di regressione, l'analisi di segmentazione, l'analisi qualitativa o l'analisi di serie storiche. L'adozione di un approccio integrato tra questi metodi consente di ottenere risultati più significativi e approfonditi. La Formula (35) di Spearman è la seguente:

$$\rho_S = 1 - \frac{6 * \sum_{i=1}^n (R(X_i) - R(Y_i))^2}{n * (n^2 - 1)} \quad (35)$$

Dove:

- $R(X_i)$ indica il rango delle valutazioni;
- $R(Y_i)$ indica il rango di $MRP_{t,k}$, ovvero i ranghi della percentuale media di un topic con un determinato rating.
- n indica il numero di livelli di valutazione considerati.

Il valore che può assumere il coefficiente di Spearman ha un valore compreso tra -1 e 1. Il valore di -1 indica una correlazione perfettamente monotona negativa, ovvero un aumento del numero di valutazioni dei clienti tende ad essere associato a una diminuzione dei punteggi, mentre il valore di 1 indica una correlazione perfettamente monotona positiva, ovvero un aumento del numero di valutazioni dei clienti tende ad essere associato a un aumento dei punteggi. Se il valore è compreso tra -1 e 1, e se si avvicina allo 0, questo suggerisce una mancanza di correlazione tra le due variabili.

CONFRONTO E ANALISI CRITICA DELLE METODOLOGIE DI ANALISI

Si riporta in Tabella 13 uno schema riassuntivo di tutte le metodologie che si utilizzeranno in questo capitolo, con una breve descrizione in cui si indica il parametro tenuto in considerazione, e una serie di pro e contro per ognuna di esse.

Tabella 13. Prospetto riassuntivo metodologie.

Strumento	Descrizione	Pro e Contro
Carte di controllo X-S (IMTP)	Sono carte di controllo per variabili e lavorano in coppia, poiché hanno funzioni complementari, la carta S analizza la variabilità all'interno del campione, mentre la carta X misura la variabilità tra campioni. La carta X si occupa del valore centrale di un processo, mentre la carta S si occupa della sua dispersione.	Queste carte sono molto importanti perché permettono di capire se un processo sia in controllo o meno. Tuttavia, un limite da tenere in considerazione è quello per cui per la costruzione della carta si utilizzano i valori di discussioni medi nel tempo per ogni determinante di qualità. In questo modo si creano limiti ad-hoc per ogni determinante e i risultati possono perdere significatività statistica. Si verificano problematiche nel caso si avessero dei trend: se l'andamento dell'IMTP è crescente, la carta con linea centrale statica non è adatta perché non tiene in

Strumento	Descrizione	Pro e Contro
		considerazione il trend verificatosi.
Carta di controllo p (IMTP)	<p>Si tratta di una carta di controllo per attributi. Questa carta tratta informazioni più povere rispetto alle carte di controllo per variabili. Per utilizzarla si fa una forte assunzione: si stabilisce una determinante di qualità “vincente” tra quelle individuate, ovvero quella determinante che presenta la maggior topical prevalence rispetto a tutte le altre. A questa determinante “vincente” verrà assegnato un punteggio di 1, mentre a tutte le restanti un punteggio di 0. In questo modo, la determinante “vincente” ritrarrà la caratteristica più rappresentativa della recensione contenuta all’interno della digital VoC.</p> <p>Per ogni recensione j-esima presente nella digital VoC si effettuerà quindi la seguente trasformazione binaria:</p> $R_j \begin{cases} 1 & \text{se } TP_{j,a} = \text{MAX}_j(TP_{j,a}) \\ 0 & \text{se } TP_{j,a} \neq \text{MAX}_j(TP_{j,a}) \end{cases}$	<p>Le carte di controllo che si creano possono presentare numerosi fuori controllo. Queste carte, trattando solitamente informazioni più povere, possono presentare varie problematiche, dovute alle forti assunzioni che si fanno per essere create. Quindi, il rischio nell’utilizzare questo approccio è avere un’approssimazione eccessiva che comporta una perdita di significatività dell’analisi. Si verificano problematiche nel caso si avessero dei trend: se l’andamento dell’IMTP è crescente, la carta con linea centrale statica non è adatta perché non tiene in considerazione il trend verificatosi.</p>
Test di Grubbs (MRP)	<p>Permette di individuare gli outliers all’interno dei dati MRP utilizzando un test di ipotesi:</p> $\begin{cases} H_0: \text{Il valore in questione non è un outlier.} \\ H_1: \text{Il valore in questione è un outlier.} \end{cases}$ <p>Calcola la statistica di Grubbs G e i valori critici G_c utilizzando il test T di Student, con una determinata numerosità del campione, che influenzerà il numero di gradi di libertà, e un</p>	<p>Una caratteristica importante di questo test è il fatto di essere un test univariato che verifica se un valore è significativamente diverso dagli altri valori in un set di dati. Questo test acquista rilevanza qualora sia abbinato ad altre analisi.</p>

Strumento	Descrizione	Pro e Contro
	<p>livello di significatività prestabilito. Se i valori di Grubbs G calcolati superano il valore critico G_C il valore in questione è un outlier e quindi anomalo.</p>	
<p>Test di correlazione di Spearman (MRP)</p>	<p>Questo test permette di valutare la correlazione che si ha tra due variabili ordinali, ignorando di fatto la scala di misurazione effettiva dei dati.</p> <p>Il valore che può assumere il coefficiente di Spearman ha un valore compreso tra -1 e 1. Più i valori tenderanno verso -1, più si avrà correlazione monotona negativa, ovvero un aumento del numero di valutazioni dei clienti tende ad essere associato a una diminuzione dei punteggi e viceversa qualora i valori tendessero a +1.</p>	<p>Il test non fornisce informazioni specifiche sulla natura della relazione tra i dati, quindi non dice la natura della relazione, tuttavia è utile per identificare una eventuale relazione monotona tra le variabili analizzate. Inoltre, individua un rapporto di causa-effetto o spiega le ragioni che si celano dietro determinati risultati. Per questo solitamente, si consiglia di interpretare i risultati generati da questo test in combinazione con altri strumenti.</p>

ANALISI IMTP

CARTE X-S

La variabile considerata per l'analisi è la topical prevalence (TP) delle determinanti di qualità trovate, ovvero dei topic individuati. Questa analisi tiene conto del numero di recensioni rilasciate per ogni periodo e, di conseguenza, della dimensione del campione considerato per la costruzione della carta stessa.

La variabile di riferimento è l'IMTP, che fornisce informazioni sulla variazione della discussione di una specifica determinante di qualità nel corso di un sotto periodo di tempo. La Formula (36) dell'IMTP è la seguente:

$$IMTP_{t,n} = \frac{\sum_j^{R_t} TP_{i,t}}{|R_n|} \quad (36)$$

Dove:

- t indica il topic t-esimo a cui ci si sta riferendo;
- n indica il periodo n-esimo;
- $TP_{i,t}$ indica la topical prevalence associata al topic t-esimo nella recensione i-esima;
- $|R_n|$ indica la cardinalità delle recensioni della digital VoC associate al periodo n-esimo.

Questa variabile tiene conto dell'evoluzione dei bisogni dei clienti nel tempo per ogni determinante di qualità individuata, rendendola particolarmente adatta a un'analisi nel tempo della tipologia di profile monitoring.

L'IMTP fornisce informazioni sulla discussione media in un intervallo di tempo quindi saranno associate due variabili:

1. d- rappresentante il numero di determinanti di qualità individuate, ovvero d=11 nel caso di Spotify;
2. t-rappresentante il numero di periodi che verranno analizzati. Essendo analizzati i mesi da gennaio a luglio t=7.

Nella Tabella 14 sono riportati i valori dell'indicatore IMTP ottenuti nel caso studio Spotify del secondo capitolo. La Tabella 15 presenta la cardinalità delle recensioni suddivise per mese nel corso del periodo considerato.

Tabella 14. IMTP per mesi e per topic.

TOPIC LEGENDA											
1-Connection problem: wifi											
2-Abbonamento non premium: randomizzazione (Subscription not premium:shuffle)											
3-App rating: section music											
4-Tariffe e i pagamenti (Payments and fees)											
5-Bug relativi all'user experience/ affidabilità dell'app (app reliability)											
6-Usò di spotify per podcast e informazione (Podcast and information)											
7-User friendly											
8-Accessibilità ai brani (songs accessibility)											
9-Stop selection/ huge selection / varietà catalogo											
10-Account and log in problem											
11-Aggiornamento features (update)											
IMTP	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11
Gennaio	0,08	0,12	0,10	0,09	0,12	0,09	0,07	0,12	0,11	0,05	0,05

IMTP	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11
Febbraio	0,09	0,12	0,10	0,08	0,12	0,08	0,07	0,12	0,11	0,05	0,05
Marzo	0,09	0,12	0,10	0,08	0,13	0,06	0,07	0,11	0,10	0,09	0,05
Aprile	0,08	0,11	0,09	0,07	0,25	0,06	0,07	0,10	0,09	0,05	0,05
Maggio	0,09	0,13	0,11	0,08	0,14	0,06	0,07	0,12	0,11	0,05	0,05
Giugno	0,09	0,13	0,11	0,08	0,13	0,06	0,07	0,12	0,11	0,05	0,05
Luglio	0,10	0,12	0,11	0,08	0,14	0,06	0,07	0,12	0,10	0,05	0,05

Tabella 15. Cardinalità delle recensioni per mesi.

MESE	Nome	Cardinalità
1	Gennaio	5228
2	Febbraio	7449
3	Marzo	8020
4	Aprile	13746
5	Maggio	10828
6	Giugno	12132
7	Luglio	3854

Si calcolano i valori di $s^2_{d,t}$ con $d=1, \dots, 11$ che rappresentano le determinanti di qualità e con $t=1, \dots, 7$ che rappresentano i mesi da gennaio a luglio considerati per l'analisi.

Utilizzando la formulazione delle carte di controllo X e delle carte S precedentemente descritte nelle Formule (25) e (26), si sono costruite le carte di controllo. Nella Tabella 16 sono riportati i calcoli e i passaggi intermedi coinvolti nella loro costruzione.

Nel contesto dell'analisi delle recensioni della Digital Voice of Customer (VoC), la decisione di adottare un piano di campionamento mensile è stata motivata da ragioni pratiche e analitiche. L'obiettivo principale è catturare variazioni significative nelle opinioni dei clienti nel corso del tempo. Tali variazioni possono derivare da fattori stagionali, eventi promozionali o modifiche nei prodotti/servizi.

La scelta di un piano mensile si è dimostrata appropriata per garantire una gestibilità efficace delle risorse disponibili per la raccolta e l'analisi delle recensioni. Questa periodicità si adatta alle esigenze di monitoraggio regolare senza richiedere un'impraticabile frequenza di aggiornamento. Inoltre, un

approccio mensile agevola l'aggregazione dei dati, fornendo una visione complessiva delle tendenze nel tempo.

È importante notare che l'analisi mensile è stata preferita per la sua capacità di adattarsi a contesti in cui le informazioni cambiano lentamente. Questo piano di campionamento risulta quindi adeguato per identificare cambiamenti sostanziali nelle recensioni senza la necessità di frequenze più elevate. In sintesi, la scelta del piano di campionamento mensile è stata guidata dalla volontà di garantire un equilibrio tra la frequenza di aggiornamento, la gestibilità delle risorse e la capacità di catturare in modo significativo le dinamiche temporali delle recensioni nella Digital Voice of Customer.

A scopo di chiarezza, i valori all'interno dei limiti di controllo sono stati evidenziati con uno sfondo verde, indicando che il processo è in controllo. In contrasto, i valori al di fuori dei limiti di controllo sono stati evidenziati con uno sfondo rosso, richiedendo un'analisi più approfondita.

Tabella 16. Costruzione delle carte X.

TOPIC LEGENDA											
1-Connection problem: wifi											
2-Abbonamento non premium: randomizzazione (Subscription not premium:shuffle)											
3-App rating: section music											
4-Tariffe e i pagamenti (Payments and fees)											
5-Bug relativi all'user experience/ affidabilità dell'app (app reliability)											
6-Usò di spotify per podcast e informazione (Podcast and information)											
7-User friendly											
8-Accessibilità ai brani (songs accessibility)											
9-Stop selection/ huge selection / varietà catalogo											
10-Account and log in problem											
11-Aggiornamento features (update)											

Carta X	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11
Gennaio	0,084	0,121	0,100	0,087	0,122	0,089	0,073	0,117	0,110	0,049	0,048
Febbraio	0,093	0,122	0,104	0,083	0,122	0,078	0,072	0,117	0,109	0,051	0,049
Marzo	0,095	0,116	0,097	0,084	0,126	0,060	0,071	0,111	0,099	0,093	0,048
Aprile	0,081	0,110	0,090	0,074	0,248	0,056	0,066	0,097	0,086	0,047	0,045
Maggio	0,087	0,128	0,110	0,081	0,135	0,060	0,073	0,118	0,107	0,051	0,049
Giugno	0,087	0,131	0,113	0,080	0,125	0,059	0,072	0,123	0,110	0,052	0,048

Carta X	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11
Luglio	0,096	0,124	0,108	0,081	0,142	0,058	0,072	0,115	0,104	0,054	0,048
LC	0,089	0,122	0,103	0,081	0,146	0,066	0,071	0,114	0,103	0,057	0,048
LCS	0,100	0,138	0,110	0,088	0,170	0,072	0,075	0,122	0,112	0,064	0,049
LCI	0,078	0,105	0,096	0,074	0,121	0,060	0,067	0,106	0,095	0,050	0,047

In appendice, nella Figura 29, sono presentate le carte X costruite a partire dai valori riportati nella Tabella 16. Come evidenziato dai valori con sfondo rosso, i principali elementi fuori controllo si riscontrano nel mese di aprile. La carta X misura la variabilità tra campioni. Per l'analisi di tali valori, sono stati identificati i 3 tipi di comportamento manifestati. A tale scopo, sono forniti alcuni esempi:

1. Nel caso del primo topic, rappresentato dalla carta di controllo in Figura 12, corrispondente al topic 2 (relativo abbonamento non premium: randomizzazione), si osserva che queste determinanti, così come il topic 1 (relativo ai problemi di connessione wi-fi), mostrano un processo in controllo nel tempo. Infatti, tutti i valori ottenuti sono collocati all'interno dei limiti di controllo superiore e inferiore.

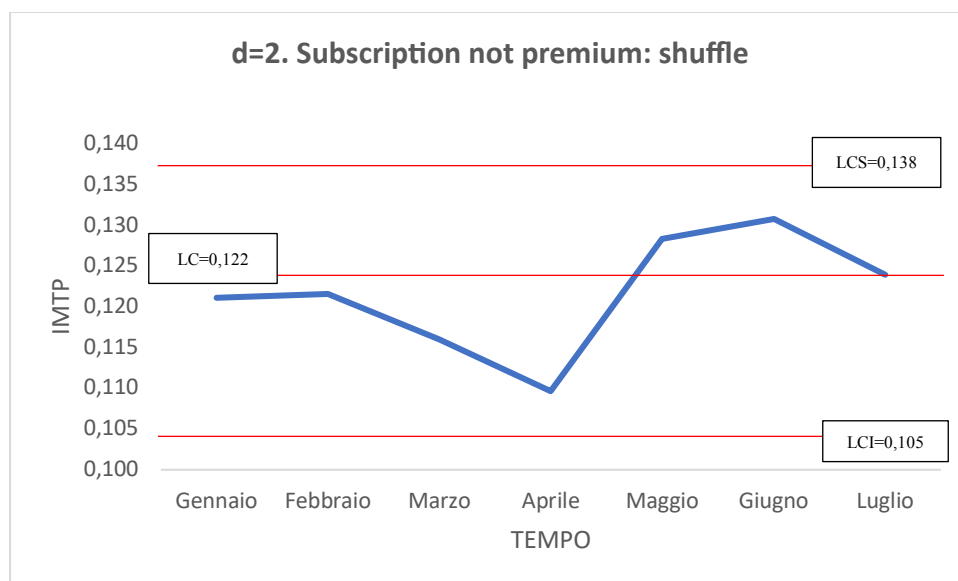


Figura 12. Carta di controllo X per il topic 2.

2. Il topic 5, relativo ai bug relativi nell' user experience e all'affidabilità dell'app, rappresentato in Figura 13, mostra un comportamento atipico rispetto agli altri topic individuati. Questa

topic risulta in controllo nel corso del tempo, fatta eccezione per il mese di aprile, durante il quale la sua discussione media assume valori significativamente superiori al limite di controllo superiore. Questo incremento nei valori durante il mese di aprile è notevole, quasi raddoppiando rispetto ai i valori dei mesi precedenti, per poi rientrare in controllo nei mesi successivi. Emerge chiaramente che, durante il mese di aprile, moltissimi utenti abbiano riscontrato problemi simili tra di loro relativi a questa determinante, amplificando notevolmente l'entità del problema. È interessante notare che, una volta individuato un fuori controllo del processo, particolarmente negativo in questo caso, esso sia stato prontamente affrontato e prontamente risolto già a partire dal mese di maggio.

Questa determinante ha registrato un picco così importante nel mese di marzo, in cui ha quasi monopolizzato le recensioni rilasciate in quel mese, a discapito di quasi tutti gli altri topic. Infatti, tutti i topic (eccezion fatta per il topic 1 e 2) nel mese di aprile sono risultati fuori controllo con valori al di sotto del limite di controllo inferiore. I risultati ottenuti confermano l'analisi del secondo capitolo e le considerazioni effettuate sull'importanza dell'user experience e dell'affidabilità da garantire all'utente durante l'utilizzo dell'applicazione. La presenza di punti fuori controllo nei processi deve fungere da campanello d'allarme per le aziende. In particolare, ove la caratteristica discussa rappresentasse una problematica per l'utente finale, come in questo caso, è opportuno trovare una soluzione e porre rimedio nel minor tempo possibile. Nel caso avvenisse una situazione opposta, ovvero un fuori controllo al di sopra dei limiti di controllo superiori, per una caratteristica prettamente positiva, bisognerebbe comunque porre attenzione, ma si potrebbe intervenire meno tempestivamente, ma anzi bisognerebbe studiare le cause dei valori ottenuti e fare in modo che persistano nel tempo, andando di fatto a rendere più efficienti i processi nel tempo.

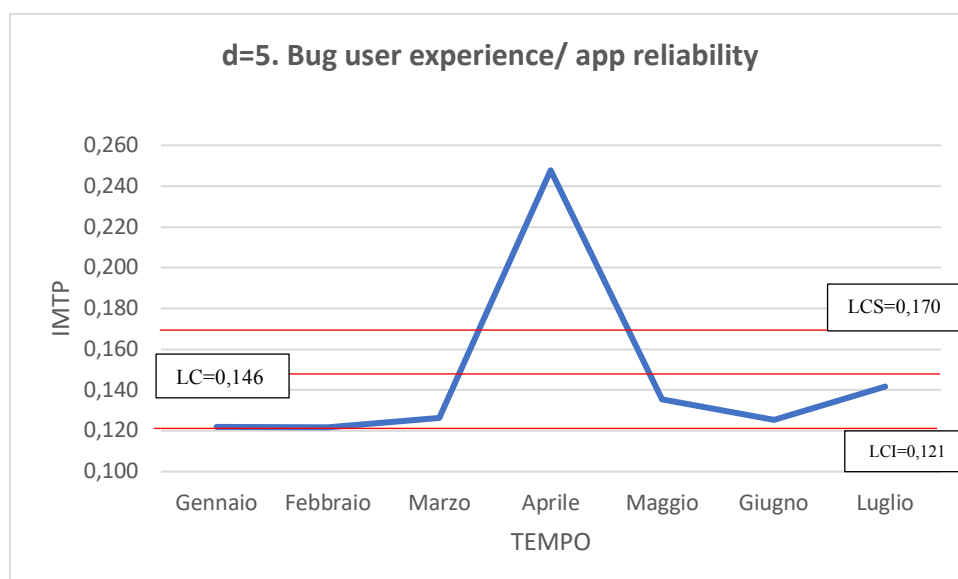


Figura 13. Carta di controllo X per il topic 5.

3. In Figura 14 è rappresentato il comportamento del topic 9 (varietà del catalogo) come esempio del comportamento dei topic 3,4 e dal 6 all'11 nel mese di aprile. Tutti questi rappresentavano determinanti in controllo nel tempo, con un picco opposto a quella del topic 5 nel mese di aprile. La maggior parte delle determinanti presentava processi in controllo e comportamenti quasi stazionari prima dello “shock” causato dal topic 5 nel mese di aprile. Una volta risolto il problema, si è tornati nella situazione pre-shock, con processi nuovamente in controllo e andamenti di tipo quasi stazionario.

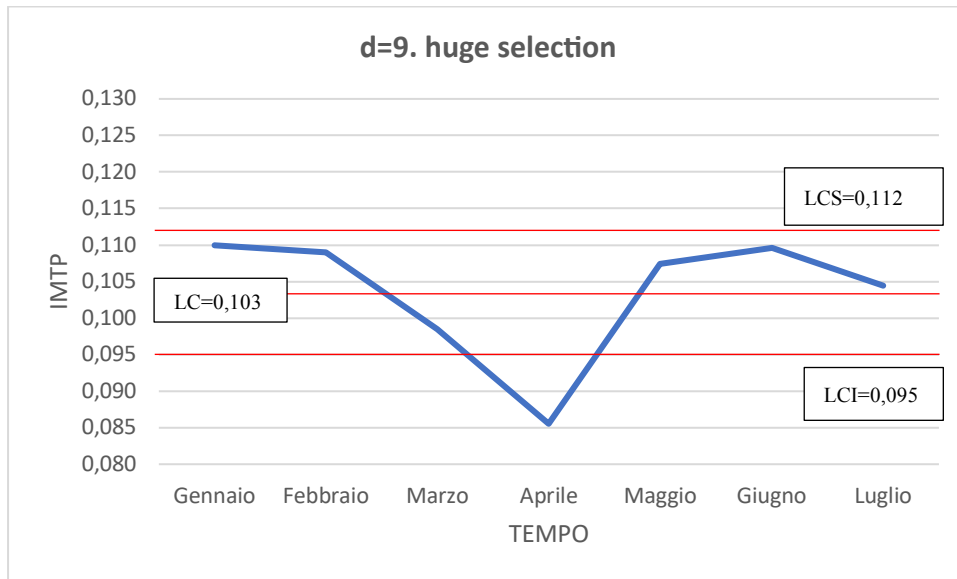


Figura 14. Carta di controllo X per il topic 9.

Successivamente si sono costruite le carte S, rappresentanti lo scarto tipo (o deviazione standard) per ogni campione. Nella Tabella 17 sono rappresentati i calcoli intermedi per la costruzione della carta. Per migliorare la chiarezza visiva, sono stati evidenziati su sfondo rosso i valori delle determinanti fuori controllo, mentre su sfondo verde i valori in controllo.

Tabella 17. Costruzioni delle carte di controllo S.

CARTA S											
s^2	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11
Gennaio	0,009	0,015	0,005	0,007	0,016	0,009	0,003	0,008	0,009	0,004	0,001
Febbraio	0,010	0,014	0,006	0,006	0,014	0,009	0,003	0,007	0,008	0,004	0,001
Marzo	0,010	0,013	0,005	0,006	0,015	0,004	0,003	0,007	0,007	0,004	0,001
Aprile	0,007	0,012	0,005	0,005	0,049	0,003	0,003	0,006	0,007	0,003	0,001
Maggio	0,009	0,014	0,007	0,006	0,017	0,003	0,003	0,006	0,007	0,004	0,001
Giugno	0,009	0,015	0,007	0,005	0,015	0,003	0,003	0,007	0,007	0,004	0,001
Luglio	0,011	0,014	0,006	0,006	0,021	0,003	0,003	0,006	0,007	0,004	0,001
LC	0,009	0,014	0,006	0,006	0,021	0,005	0,003	0,007	0,007	0,004	0,001
UCL	0,017	0,026	0,011	0,011	0,039	0,009	0,006	0,012	0,014	0,007	0,002
LCL	0,001	0,002	0,001	0,001	0,002	0,001	0,000	0,001	0,001	0,000	0,000

Nell'appendice, nella Figura 30, sono rappresentate le carte S costruite utilizzando i valori riportati in Tabella 17.

Nel contesto delle carte S, sono stati identificati due comportamenti distinti:

1. Nella Figura 15, è stato come esempio il topic 2, che manifesta il medesimo comportamento di tutte le altre determinanti di qualità individuate nell'analisi, ad eccezione del topic 5. Tutti questi topic mostrano processi in controllo, nonostante l'influenza del mese di aprile sui valori della carta X. Come previsto, in questi topic la dispersione è in controllo e non si osservano valori significativamente dai valori centrali. Tale andamento non si verifica per il topic 5, che emerge come protagonista dell'analisi.

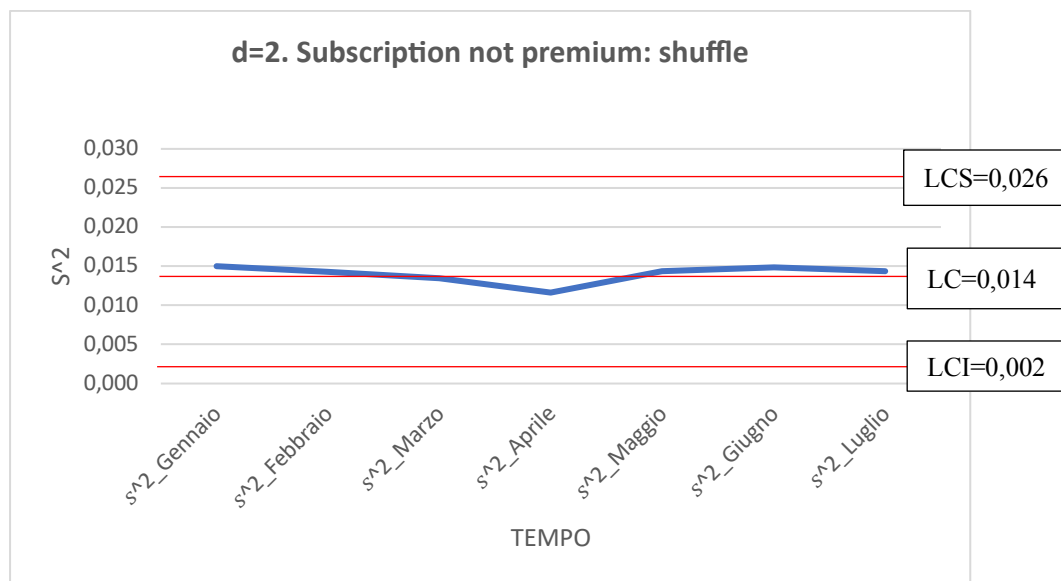


Figura 15. Carta di controllo S topic 2.

2. Il topic 5, relativo ai bug relativi dell'user experience e all'affidabilità dell'app, è emerso come il topic con comportamento anomalo nell'analisi. Tale comportamento, è osservabile anche nella carta S rappresentata in Figura 16. I valori mostrano un andamento in controllo, ad eccezione del mese di aprile, che come si è rilevato determinante e caratteristico nell'analisi. In questo mese, si sono verificati valori anomali, causati da bug importanti all'interno dell'applicazione che hanno portato malessere e lamentele da parte degli utenti. Tuttavia, tali problematiche sono state prontamente risolte già nel mese successivo, in cui i valori tornano a stabilizzarsi e in controllo. Ergo, si può affermare che i risultati ottenuti dalle carte X-S, confermano l'analisi della digital-VoC presentata nel secondo capitolo. Si rende ancora più noto e chiaro quanto siano importanti l'user experience e l'affidabilità e il monitoraggio continuo dei servizi forniti dall'app, e come, anche una piccola problematica possa influenzare rapidamente il giudizio dei clienti finali, che sono estremamente sensibili a tali questioni e pronti a esprimere la propria insoddisfazione. Quindi, questa determinante è di vitale importanza per l'azienda e richiede un monitoraggio costante nel tempo per prevenire

il ripresentarsi di una situazione simili a quelle registrate ad aprile, che seppur risolte prontamente, può allontanare svariati clienti, anche in maniera definitiva dal servizio offerto.

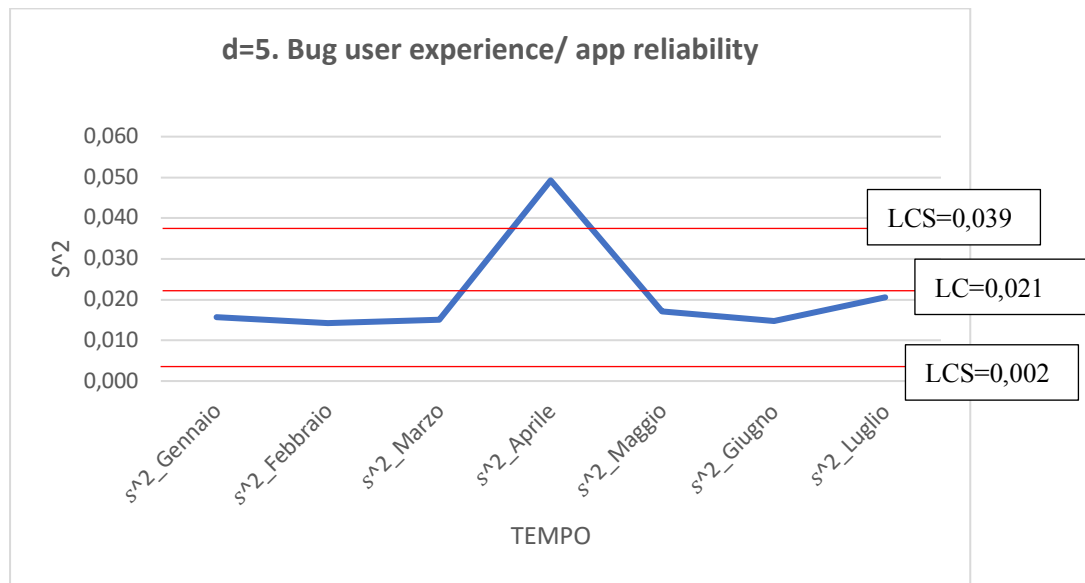


Figura 16. Carta di controllo S topic 5.

CARTE P

Queste carte, come precedentemente menzionato, trattano informazioni di basso livello informativo e richiedono campioni molto ampi. In riferimento a prodotti, queste carte sono spesso chiamate “per difettosi”. In particolare, queste carte valutano la percentuale di elementi difettosi all’interno di un campione, utilizzando variabili binarie (0=elemento non difettoso, 1=elemento difettoso). Nel contesto dei servizi, come nel caso studio analizzato, si applica un ragionamento simile, ma focalizzato sulla caratteristica “vincente”.

In un primo passaggio intermedio, per ogni periodo analizzato t (mesi da gennaio a luglio) e per ogni determinante di qualità individuata (gli 11 topic), si è sommato il numero di “vittorie” (ovvero, la presenza di 1). La Tabella 18 mostra i risultati di questo passaggio.

Tabella 18. Vittorie per ogni topic per ogni mese nella costruzione delle carte p.

Topic \ Mesi	1	2	3	4	5	6	7	8	9	10	11	Rt
Gennaio	452	782	465	464	772	565	224	619	690	151	55	5228
Febbraio	847	1121	744	594	1046	624	310	864	983	230	86	7449
Marzo	893	1120	717	633	1242	297	327	845	888	987	70	8020
Aprile	973	1503	1126	777	5621	384	492	1115	1247	383	125	13746
Maggio	997	1766	1396	789	1829	344	532	1277	1415	350	133	10828
Giugno	1134	2043	1661	825	1772	370	564	1598	1649	396	120	12132

Topic \ Mesi	1	2	3	4	5	6	7	8	9	10	11	Rt
Luglio	440	581	523	271	664	115	169	441	485	130	35	3854
SOMMA	5736	8916	6632	4353	12946	2699	2618	6759	7357	2627	624	61257

In seguito, utilizzando la Formula (31), si sono definiti i limiti di controllo superiore e inferiore per ogni topic.

Un esempio di costruzione di carta per il topic 1 è presentato in Tabella 19:

Tabella 19. Costruzione carta p del topic 1.

CARTA P TOPIC 1						
Campione	"vittorie"	Rt	Valore	LCS	LCI	LC
Gennaio	452	5228	0,086	0,110	0,085	0,097
Febbraio	847	7449	0,114	0,108	0,087	0,097
Marzo	893	8020	0,111	0,107	0,087	0,097
Aprile	973	13746	0,071	0,105	0,090	0,097
Maggio	997	10828	0,092	0,106	0,089	0,097
Giugno	1134	12132	0,093	0,106	0,089	0,097
Luglio	440	3854	0,114	0,112	0,083	0,097

p medio	0,097
---------	-------

E ponendo i risultati su una carta di controllo, si ottiene la seguente Figura 17:

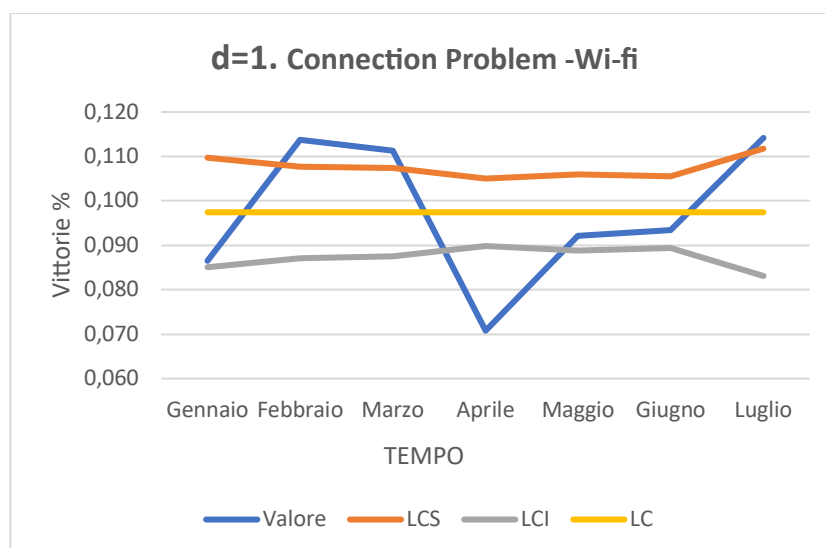
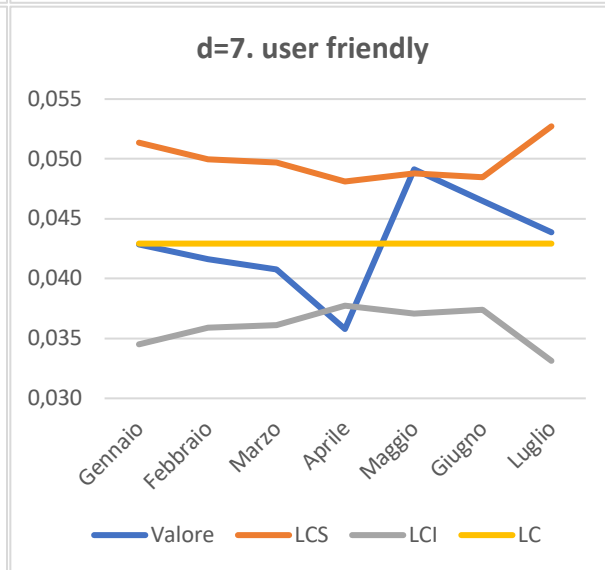
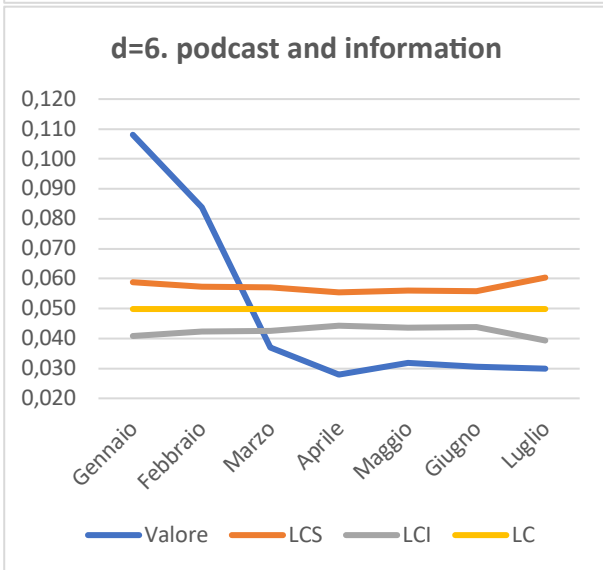
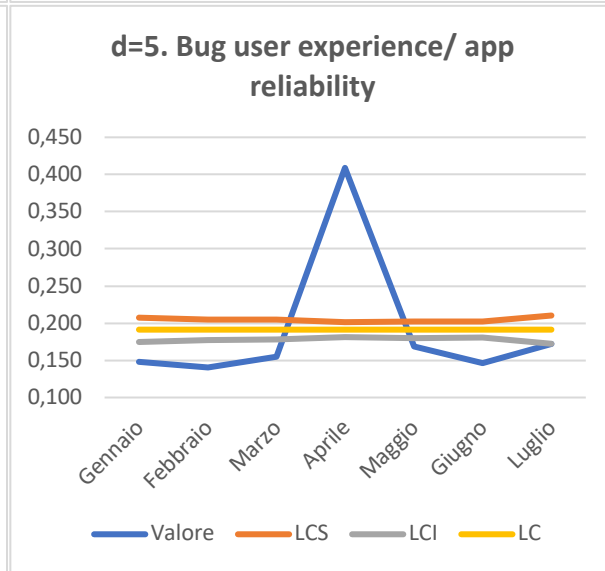
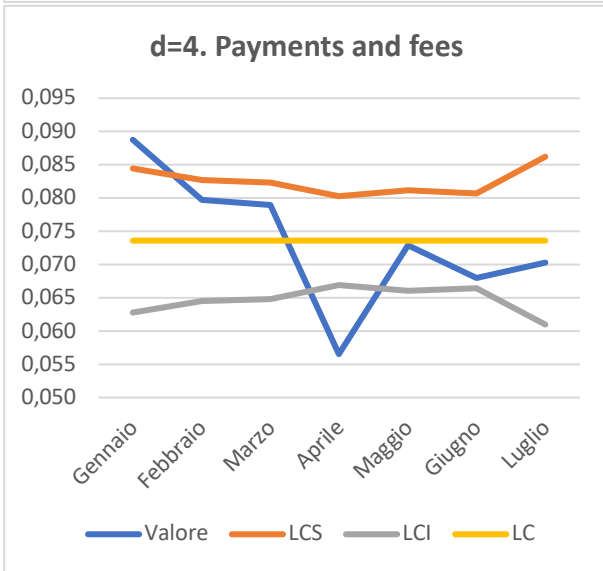
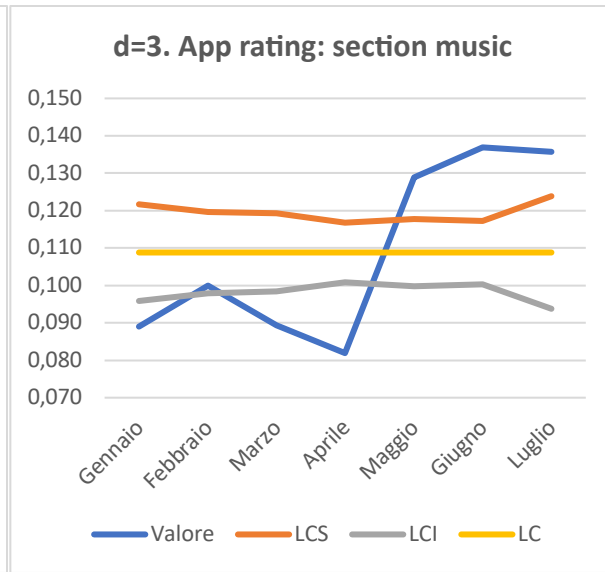
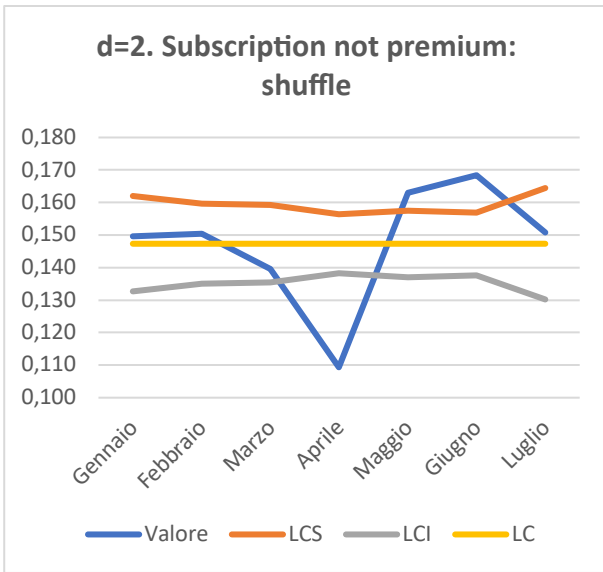


Figura 17. Carta di controllo p per il topic 1.

Si nota immediatamente che il processo non è in controllo. Ciò è evidenziato dal superamento del limite di controllo superiore in 3 punti: febbraio, marzo e luglio, oltre a scendere al di sotto del limite di controllo inferiore nel mese di aprile. Questo topic ha una discussione prettamente negativa, infatti

tratta problematiche alla connessione e in particolare al wi-fi. Quindi, il valore ottenuto dovrebbe rimanere all'interno dei limiti di controllo e preferibilmente mantenere una discussione che decresce nel tempo o che comunque tenda al limite di controllo inferiore, volta a sottolineare la diminuzione dei problemi nel tempo. Tuttavia, ciò non si verifica nei mesi di febbraio e marzo, durante i quali si riscontrano più problemi del previsto. A aprile, si osserva una drastica diminuzione della discussione su questo topic, che potrebbe suggerire la risoluzione dei problemi. Tuttavia, come è emerso dall'analisi precedente, si evince che questa riduzione di discussione della determinante in analisi sia causata dal manifestarsi di altre problematiche. Tale ipotesi viene confermata osservando la carta di controllo relativa al topic 5 (che tratta bug e affidabilità dell'app). Dopo la risoluzione delle problematiche che hanno influenzato le recensioni di aprile infatti, la discussione media di questo topic torna a crescere nel tempo, entro limiti di controllo fino a giugno, per poi evidenziare un nuovo fuori controllo a luglio. Pertanto, è essenziale prestare particolare attenzione a questo topic, rappresentando un aspetto negativo per il cliente, e cercare di risolvere le problematiche di connessione e di wi-fi nell'applicazione, mantenendo il processo in controllo.

Ripetendo i passaggi per ogni topic, si ottengono le seguenti carte di controllo p, rappresentate in Figura 18, con sull'asse delle ascisse il tempo (mesi da gennaio a luglio) e sull'asse delle ordinate i valori p ottenuti:



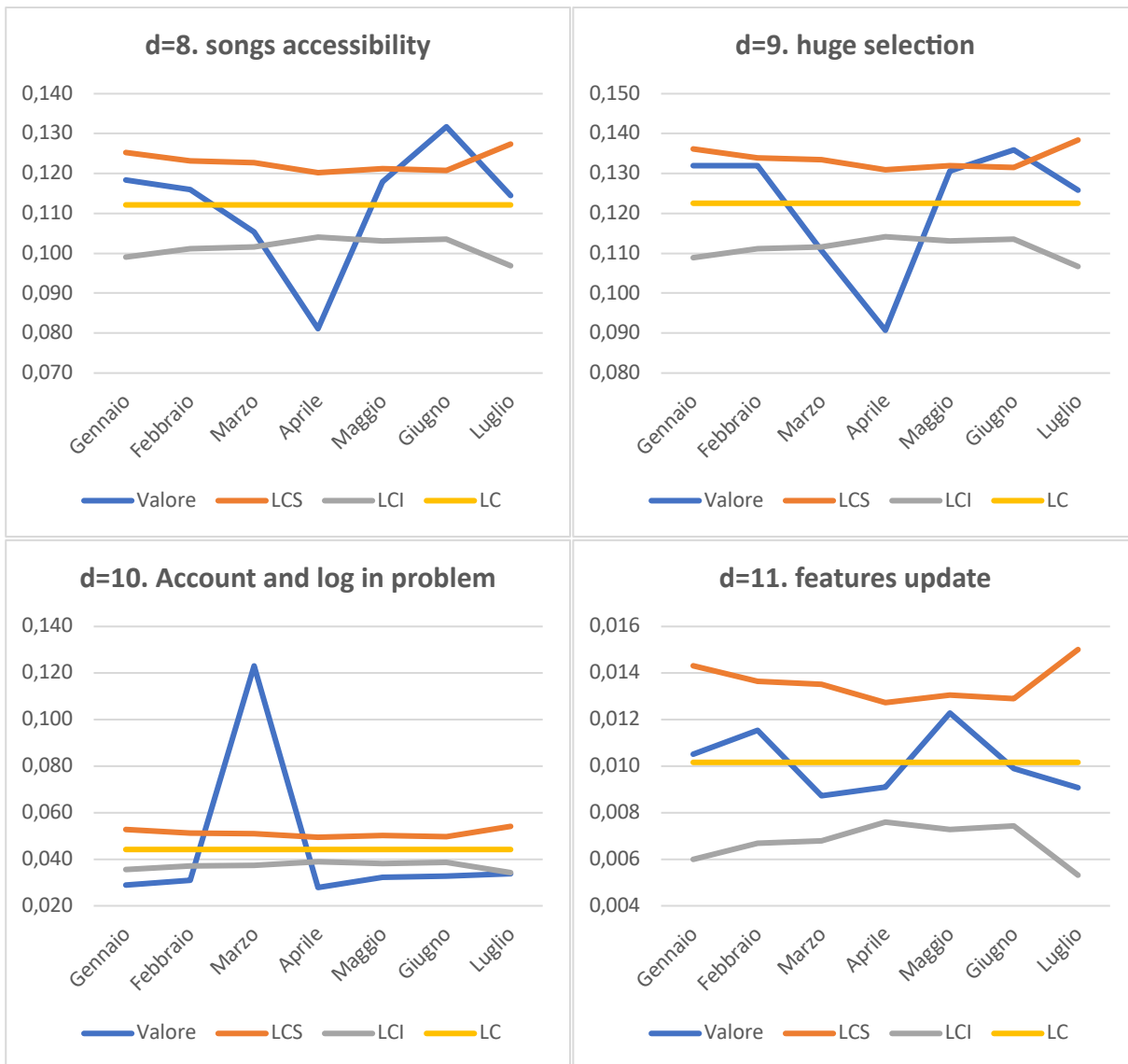


Figura 18. Carte di controllo p per i topic dal 2 all'11.

Il topic 11 risulta essere l'unica determinante di qualità che rimane in controllo per l'intero periodo analizzato, sebbene sia anche il topic meno discusso nel database. È evidente come i principali eventi fuori controllo si verificano nel mese di aprile, in cui praticamente tutti i topic individuati subiscono una diminuzione della discussione a vantaggio del topic 5, relativo ai problemi relativi di bug dell'applicazione e all'affidabilità dell'app. Queste osservazioni sono coerenti con l'analisi condotta nel secondo capitolo e sottolineano l'importanza del monitoraggio continuo e dell'efficienza dell'App, dell'user experience e della fornitura di un servizio di alta qualità.

Analizzando i risultati di queste carte nel tempo, si nota un aumento della discussione del topic 10, riguardante i problemi di log-in e dell'account, nel mese di marzo, prontamente risolto ad aprile. L'Osservazione delle carte di controllo nel tempo evidenzia l'importanza di alcune determinanti di qualità, alcune delle quali appaiono come positive, mentre altre come negative. Ad esempio il topic

10, associato a una problematica, richiede una soluzione tempestiva se si verifica un fuori controllo oltre il limite superiore. Viceversa, un valore di fuori controllo al di sotto del limite inferiore rappresenta un aspetto positivo che dovrebbe essere replicato per migliorare l'efficienza del processo nel tempo. Lo stesso ragionamento si applica alle determinanti di qualità associate a una soddisfazione positiva, come il topic 3 (app rating: section music), il topic 7 (user-friendly experience in-app), e il topic 9 (vastità del catalogo e ampia scelta). La presenza di curve crescenti in questi topic suggerisce una buona performance aziendale e un aumento della percezione di qualità da parte degli utenti.

Osservando le carte di controllo, emerge la presenza di numerosi fuori controllo in quasi tutti i topic individuati. Tuttavia, si sottolinea che le carte p non sembrano essere sufficientemente efficaci per l'analisi condotta, presentando notevoli problematiche. Una delle ragioni di questo "errore" è la forte assunzione iniziale che ciascuna recensione tratti al massimo una determinante di qualità. Le carte p presentano caratteristiche più stringenti rispetto alle carte per variabili, richiedendo la presenza mutuamente esclusiva di un topic per ogni recensione. La scelta di effettuare una trasformazione binaria, indicando se il topic rappresentasse la caratteristica "vincente" o meno, comporta criticità che si riflettono nella costruzione delle carte.

ANALISI MRP

Dopo aver utilizzato l'indicatore IMTP per monitorare nel tempo le determinanti di qualità, ora si studia l'MRP.

Lo scopo di questa sezione è di analizzare inizialmente un indicatore molto importante come l'MRP, che fornisce una comprensione approfondita di come le determinanti di qualità siano percepite basandosi sui valori di rating rilasciati dagli utenti, e quindi di capire come varia la soddisfazione percepita rispetto al servizio offerto nel corso del tempo. Tuttavia, quello che si cerca di individuare, è uno strumento statistico che funga da "campanello d'allarme" nel caso si riscontrassero valori che si discostano dalla distribuzione attesa dei dati. Dopo una prima analisi generale sull'indicatore posto osservazione, si cercherà uno strumento che permetta di individuare valori anomali e outliers, fornendo un avviso al fornitore del servizio.

In sintesi, i clienti esprimono le loro valutazioni tramite la digital-VoC, rilasciando recensioni sul web, sui vari store o su piattaforme di recensioni etc. e per farlo utilizzano scale di valutazioni di tipo ordinale, rilasciando valori che possono variare da 1 a 5 (più comunemente noti come valori di rating da 1 stella a 5 stelle). Con 1 stella si indica la minor valutazione di soddisfazione possibile e con 5 stelle una valutazione di eccellenza. L'indicatore, già discusso in precedenza, consente una visione

completa dell'andamento delle recensioni è l'MRP. Questo indicatore indica quanto è discusso un topic nei record di digital-VoC associati con uno specifico livello di rating. In questo modo, si riesce a capire dove è principalmente discusso il topic all'interno del rating.

Le determinanti di qualità individuate si dividono in tre categorie (Barravecchia et al., 2020):

- Determinanti neutrali rispetto ai livelli di soddisfazione. Queste determinanti non sono caratterizzate da valori estremi di rating, ma bensì sono riconducibili prettamente a valori intermedi (2, 3 e 4 stelle). Su queste determinanti bisogna riporre particolare interesse e attenzione in quanto rappresentano valori intermedi, e quindi allo stesso tempo potenzialità e minacce.
- Determinanti associate a livelli di soddisfazione negativi. Queste determinanti sono caratterizzate da una monotonia decrescente, ovvero sono presenti principalmente recensioni con un basso livello di rating rispetto a quelle con un alto livello di rating. Queste determinanti sono quelle che sono più riconducibili alle percezioni di insoddisfazione da parte del cliente nel confronto del servizio e rappresentano delle vere e proprie debolezze che bisognerà cercare di risolvere nel corso del tempo.
- Determinanti associate a livelli di soddisfazione positivi. Queste determinanti sono caratterizzate da una monotonia crescente, ovvero sono presenti principalmente recensioni con un alto livello di rating. Queste determinanti sono quelle più riconducibili alle percezioni di soddisfazione da parte del cliente nel confronto del servizio e rappresentano dei veri e propri punti di forza che andranno mantenuti tali nel corso del tempo.

In Tabella 20 si riporta uno schema riassuntivo delle recensioni analizzate all'interno del database. Si può notare come il mese di aprile risulti essere il mese con il numero più alto di recensioni, mentre quello di luglio quello con il numero più basso.

Tabella 20. Recensioni mese per mese con rispettivi livelli di rating,

rating	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio
★	2171	2489	1788	2597	2824	4084	1623
★★	583	1032	749	1016	1067	1869	778
★★★	582	878	754	1181	1115	1658	672
★★★★	551	952	1097	1835	1487	1489	370
★★★★★	1341	2098	3632	7117	4335	3032	411
TOTALI	5228	7449	8020	13746	10828	12132	3854

In Figura 19, invece, si riportano questi valori su un grafico per semplicità di visualizzazione. Si possono osservare così, il rating delle recensioni nel tempo e un confronto con le recensioni totali.

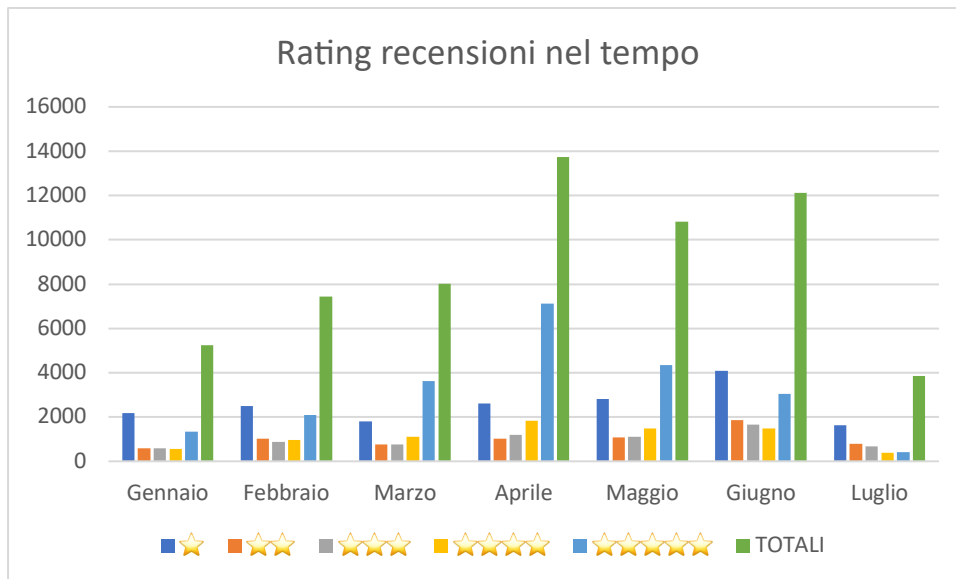


Figura 19. Rating delle recensioni nel tempo.

Mediante la Figura, è possibile ottenere una rappresentazione visiva delle valutazioni presenti all'interno del database analizzato. Si possono vedere le cardinalità delle recensioni su uno specifico livello di rating. Nei mesi di gennaio, febbraio, giugno e luglio prevalgono le recensioni di insoddisfazione, caratterizzate da una valutazione di 1 stella. Al contrario, nei mesi di marzo, aprile e maggio prevalgono le recensioni di soddisfazione, con una valutazione massima di 5 stelle. È evidente che ci sia stato un cambiamento positivo di tendenza tra febbraio e marzo, seguito da un'inversione negativa tra maggio e giugno. È importante controllare e monitorare questa metrica nel tempo, poiché fornisce indicazioni al management aziendale sul comportamento dell'azienda. Anche da questa visualizzazione, rappresentata nella Figura precedente, si può notare la prevalenza di recensioni con 1 stella e di recensioni con 5 stelle, confermando la tendenza dei clienti a rilasciare una recensione per manifestare una lamentela importante o un'importante soddisfazione.

L'indicatore che si vuole studiare per effettuare un profile monitoring è l'MRP. Questo indicatore, utilizza una scala ordinale discreta, con valori compresi tra 1 e 5 per esprimere le preferenze. Le scale con variabili ordinali, nonostante presentino categorie che risultano ordinate (1, 2, 3, 4 e 5), non implicano necessariamente una distanza equidistante tra le categorie stesse.

Uno degli strumenti più utilizzati per effettuare profile monitoring, come precedentemente riportato, è la regressione lineare. In questo caso però, questa soluzione non si presenterebbe come la scelta migliore, infatti, essa non è adatta per studiare variabili ordinali, ma bensì necessita di scale di misurazione continue per le variabili indipendenti, che spiegano la variabile dipendente.

Quindi, l'ipotesi di utilizzare una regressione lineare per analizzare variabili ordinali è esclusa principalmente per 3 ragioni principali:

1. L'assunzione su cui si basa la regressione lineare di scala continua, ovvero sul fatto che i valori delle variabili indipendenti possano variare in modo costante e assumere valori continui. L'utilizzare variabili ordinali implicherebbe invece l'utilizzo di categorie discrete e ordinate, con distanze tra le categorie non costanti.
2. Si potrebbero perdere informazioni. Usando la regressione lineare infatti, le variabili ordinali verrebbero assunte come continue e ciò potrebbe portare a conclusioni errate e fuorvianti, così come a previsioni che si discostano dalla realtà.
3. Violazione delle assunzioni su cui si fonda la regressione lineare stessa, tra cui la linearità della relazione e l'omoschedasticità, assunzione fondamentale che richiede l'uniformità della varianza degli errori. Violare assunzioni necessarie su cui si basa il modello porterebbe chiaramente a risultati errati e non validi.

Ergo, per l'analisi che si vuole condurre, è necessario evitare l'utilizzo della regressione lineare. Bisogna prestare particolare attenzione nella scelta della tecnica da utilizzare, basandosi sui dati a propria disposizione e sul modo in cui le variabili sono presentate in modo tale da ottenere i risultati voluti e affinché essi siano accurati e con una valenza significativa.

Trattando variabili ordinali, bisogna cercare di utilizzare tecniche statistiche più specifiche, che superino i limiti della regressione lineare (Ding et al., 2006; Williams et al., 2007; Morguerza et al., 2007; Chen e Nembhard, 2007), tra questi si potrebbe tenere in considerazione la regressione logistica ordinale o la regressione ordinale, che tengono conto della tipologia ordinale con cui sono espresse le variabili e forniscono quindi risultati più appropriati all'analisi che si sta effettuando. Altre opzioni che si possono effettuare sono analisi più classiche quali le carte di controllo, o l'esplorazione di strumenti alternativi.

CARTE DI CONTROLLO

Come precedentemente riportato nell'analisi dell'IMTP, le carte di controllo rappresentano un elemento ampiamente utilizzato per condurre un'analisi di profile monitoring.

Ripetendo i passaggi svolti per la costruzione delle carte di controllo per l'indicatore IMTP, si costruisce la carta di controllo X per l'indicatore MRP. Lo scopo è verificare se il servizio è in controllo statistico, ovvero se opera in modo coerente e privo di deviazioni significative o cause speciali che influenzano i dati.

Per la costruzione della carta si sono applicati gli stessi passaggi effettuati per l'indicatore IMTP. Analogamente si sono costruiti i limiti di controllo per ogni topic. Se tutti i valori rientrano all'interno

dei limiti di controllo il processo si definisce “in controllo statistico”. In Tabella 21 si riportano i valori ottenuti eseguendo i calcoli appena citati.

Tabella 21. Costruzione carta X MRP.

TOPIC LEGENDA											
1-Connection problem: wifi											
2-Abbonamento non premium: randomizzazione (Subscription not premium:shuffle)											
3-App rating: section music											
4-Tariffe e i pagamenti (Payments and fees)											
5-Bug relativi all'user experience/ affidabilità dell'app (app reliability)											
6-Usò di spotify per podcast e informazione (Podcast and information)											
7-User friendly											
8-Accessibilità ai brani (songs accessibility)											
9-Stop selection/ huge selection / varietà catalogo											
10-Account and log in problem											
11-Aggiornamento features (update)											

MRP	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11
★	0,10	0,13	0,07	0,09	0,20	0,07	0,06	0,10	0,06	0,07	0,05
★★	0,10	0,14	0,06	0,08	0,24	0,06	0,06	0,10	0,06	0,05	0,05
★★★	0,10	0,15	0,07	0,08	0,20	0,06	0,06	0,11	0,07	0,05	0,05
★★★★	0,09	0,13	0,10	0,08	0,13	0,06	0,07	0,13	0,11	0,05	0,05
★★★★★	0,07	0,10	0,15	0,08	0,09	0,06	0,09	0,12	0,16	0,05	0,05
MEDIA	0,09	0,13	0,09	0,08	0,17	0,06	0,07	0,11	0,09	0,06	0,05
dev std	0,01	0,02	0,04	0,01	0,06	0,00	0,01	0,01	0,04	0,01	0,00
LCS	0,13	0,19	0,21	0,09	0,35	0,08	0,10	0,15	0,21	0,08	0,06
LCI	0,05	0,07	0	0,06	0	0,05	0,03	0,08	0	0,03	0,04

Come si evince dai risultati, tutti i valori rientrano entro i limiti di controllo, quindi si potrebbe pensare di poter concludere che il processo è in controllo statistico analizzando l'indicatore MRP. Questo rappresenterebbe un segnale positivo e indicherebbe che il processo funzioni in modo stabile e prevedibile rispetto ai dati storici. Non emergono, infatti, segni evidenti di problemi o anomalie riguardanti il processo che abbiano portato a variazioni significative nei dati. Anche il topic 5, pur presentando un livello di discussione più elevato rispetto agli altri, risulta comunque entro i limiti di controllo.

In conclusione, essendo tutti i valori all'interno dei limiti di controllo, si potrebbe pensare che il processo sia stabile e funzioni come previsto sulla base dei dati storici. Tuttavia, è importante considerare un limite utilizzando queste carte. Infatti, per la costruzione della carta, si utilizzano i valori di discussioni medi nel tempo per ogni determinante di qualità. In questo modo, si creano limiti ad-hoc per ogni determinante e i risultati possono perdere significatività statistica, come si evince dal fatto che tutto il processo risulti in controllo statistico, nonostante si sappia, tramite analisi precedenti,

che il mese di aprile sia stato la causa di problematiche all'interno delle discussioni delle determinanti stesse.

Ergo, per queste problematiche non si sono costruite le carte S per i topic analizzati, in quanto queste tipologie di carte presentando limiti ad hoc propri per ogni topic e non si possono ritenere lo strumento adatto per l'analisi di questo indicatore. Analogo discorso vale per le carte p, che presentano le stesse problematiche riscontrate nella costruzione per l'indicatore IMTP.

Per superare tali limitazioni, nella prossima sezione si presenta un nuovo strumento statistico progettato per avvisare e segnalare eventuali anomalie all'interno del processo nel corso del tempo.

TEST DI GRUBBS

Si esegue il test di Grubbs con un livello di significatività $\alpha = 0,01$ e con un livello di significatività $\alpha = 0,05$. I valori outliers individuati relativi al livello di significatività al 1% saranno evidenziati con sfondo della cella rosso, mentre i valori outliers relativi al livello di significatività al 5% saranno evidenziati con il testo in rosso.

Nella Tabella 22 sono riportati i valori trovati per il mese di gennaio. Come si evince dall'assenza di celle evidenziate di colore rosso, non verificano outliers per un livello di significatività al 1%. Con un livello di significatività al 5%, invece, si attiva un campanello dall'allarme per il topic 5 con livello di rating pari a 2 e per il topic 9 con livello di rating pari a 5. Un livello di discussione più elevato del normale per il topic 9 non rappresenta una situazione problematica per l'analisi in questione, in quanto si ha una discussione più elevata del normale ma di natura positiva con un rating di 5 stelle. Quindi, bisognerà concentrarsi sul topic 5 con livello di rating pari a 2. In Figura 20 si riportano graficamente i valori trovati.

Tabella 22. Valori test di Grubbs nel mese di gennaio.

Gennaio	1	2	3	4	5	6	7	8	9	10	11	Media	Dev.st
★	0,03	0,04	0,02	0,03	0,04	0,04	0,02	0,03	0,02	0,02	0,01	0,0276	0,0094
★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0098	0,0043
★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,0104	0,0041
★★★★	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,01	0,01	0,0122	0,0038
★★★★★	0,02	0,03	0,05	0,03	0,03	0,03	0,03	0,04	0,06	0,02	0,02	0,0309	0,0124

GRUBBS	1	2	3	4	5	6	7	8	9	10	11
★	-0,08	1,31	-0,66	0,25	1,71	1,04	-0,85	0,50	-0,64	-1,15	-1,44
★★	0,21	1,32	-0,53	-0,12	2,31	-0,41	-0,72	0,61	-0,57	-1,08	-1,01
★★★	0,15	1,78	-0,45	-0,33	1,64	-0,63	-0,84	1,12	-0,27	-1,13	-1,04
★★★★	0,26	1,35	0,13	-0,29	0,69	-0,93	-0,67	1,58	0,74	-1,58	-1,28
★★★★★	-0,67	-0,03	1,53	-0,26	-0,36	-0,30	-0,09	0,62	2,08	-1,26	-1,27

$t_{9,0.05}$	2,262
Gc	1,9517

$t_{9,0.01}$	3,250
Gc	2,3966

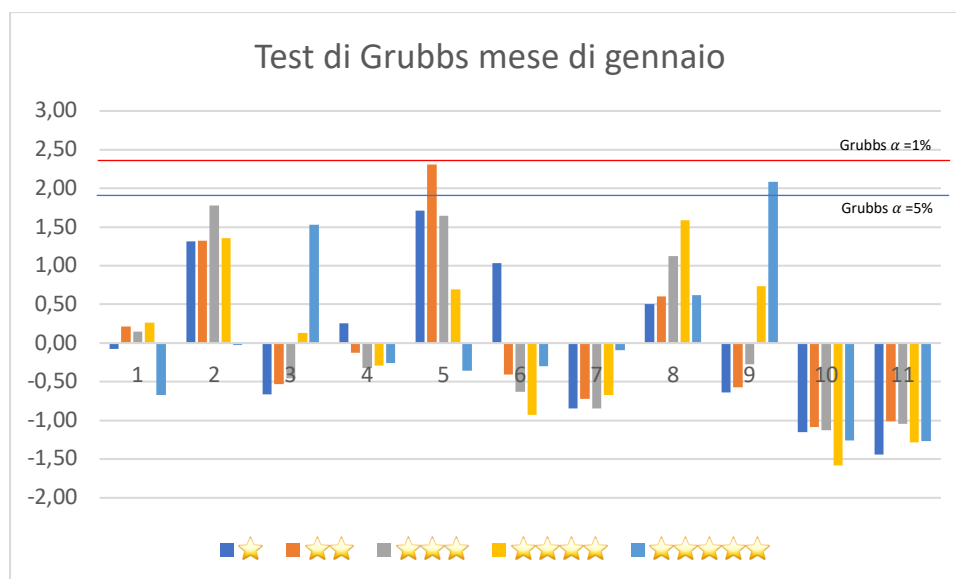


Figura 20. Test di Grubbs nel mese di gennaio.

In Tabella 23 sono riportati i valori trovati per il mese di febbraio. L'unico outlier presente, analogamente al mese di gennaio è il topic 5 con rating 2 al livello di significatività del 5%. Non si accende nessun campanello d'allarme in questo mese con un livello di significatività all'1%. In Figura 21 si riportano i valori trovati su un grafico.

Tabella 23. Valori test di Grubbs nel mese di febbraio.

Febbraio	1	2	3	4	5	6	7	8	9	10	11	Media	Dev.st
★	0,03	0,04	0,02	0,03	0,04	0,03	0,02	0,03	0,02	0,02	0,01	0,0251	0,0092
★★	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0094	0,0040
★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0100	0,0041
★★★★	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,01	0,01	0,0125	0,0041
★★★★★	0,03	0,03	0,06	0,03	0,03	0,03	0,03	0,04	0,06	0,02	0,02	0,0339	0,0137

GRUBBS	1	2	3	4	5	6	7	8	9	10	11
★	0,38	1,40	-0,61	0,10	1,84	0,18	-0,88	0,53	-0,66	-1,02	-1,25
★★	0,64	1,28	-0,53	-0,14	2,01	-0,63	-0,70	0,62	-0,50	-1,02	-1,03
★★★	0,63	1,61	-0,45	-0,32	1,68	-0,69	-0,74	0,73	-0,33	-1,10	-1,02
★★★★	0,35	1,40	0,37	-0,26	0,56	-0,92	-0,74	1,31	0,68	-1,43	-1,33
★★★★★	-0,56	0,01	1,59	-0,41	-0,25	-0,37	-0,14	0,73	1,82	-1,18	-1,25

$t_{9,0.05}$	2,262
Gc	1,9517

$t_{9,0.01}$	3,250
Gc	2,3966

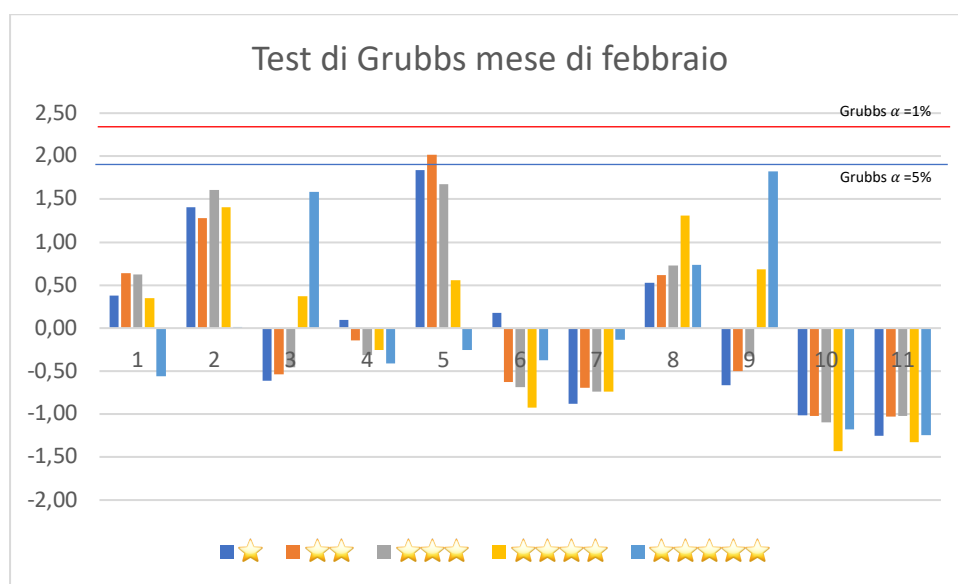


Figura 21. Test di Grubbs nel mese di febbraio.

In Tabella 24 sono riportati i valori trovati per il mese di marzo. Analogamente ai mesi precedenti, l'unico outlier è nuovamente il topic 5 con un livello di rating pari a 2, rilevato con un livello di significatività al 5%. Non emerge nessun campanello d'allarme in questo mese a un livello di significatività all'1%. In Figura 22 si riportano graficamente i valori trovati.

Tabella 24. Valori test di Grubbs nel mese di marzo.

Marzo	1	2	3	4	5	6	7	8	9	10	11	Media	Dev.st
★	0,03	0,04	0,02	0,03	0,05	0,02	0,02	0,03	0,02	0,05	0,01	0,0283	0,0109
★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0107	0,0045
★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0099	0,0038
★★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,0121	0,0039
★★★★★	0,02	0,03	0,05	0,03	0,03	0,02	0,03	0,04	0,05	0,02	0,02	0,0299	0,0119

GRUBBS	1	2	3	4	5	6	7	8	9	10	11
★	0,49	0,70	-0,77	0,13	1,60	-0,83	-0,84	0,11	-0,86	1,54	-1,26
★★	0,44	1,05	-0,71	-0,17	2,23	-0,78	-0,71	0,22	-0,84	0,37	-1,09
★★★	0,42	1,58	-0,45	-0,36	1,82	-0,99	-0,84	0,57	-0,50	-0,07	-1,18
★★★★	0,18	1,29	0,28	-0,45	1,04	-1,11	-0,78	1,49	0,39	-1,02	-1,29
★★★★★	-0,48	0,14	1,54	-0,33	-0,26	-0,85	-0,09	0,83	1,78	-1,05	-1,24

$t_{9,0.05}$	2,262
Gc	1,9517

$t_{9,0.01}$	3,250
Gc	2,3966

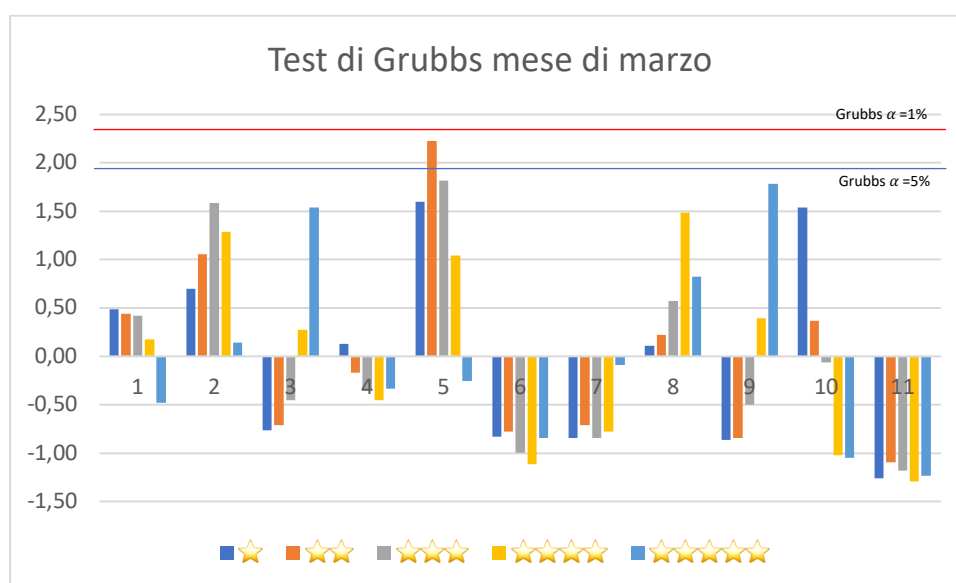


Figura 22. Test di Grubbs nel mese di marzo.

In Tabella 25 sono riportati i valori trovati per il mese di aprile. Come ci si aspettava dalle analisi precedenti, si osservano gli unici campanelli d'allarme per un livello di significatività al 1% per il topic 5 per i 3 livelli di rating più bassi. Ovviamente questi campanelli permangono con un livello di significatività del 5%, estendendosi addirittura per il quarto rating.

È evidente, analizzando i valori in questa tabella, notare come nel mese di aprile si sia verificato un evento che ha influenzato i valori di rating più bassi per il topic 5, ovvero qualcosa che ha portato a discutere maggiormente in maniera negativa questo topic in questo mese. Tornando all'analisi sull'IMTP, si possono ricondurre i risultati ottenuti alle problematiche riscontrate nell'affidabilità e nell'user experience in questo mese che hanno portato numerose lamentele e insoddisfazione nei clienti. In Figura 23 si riportano graficamente i valori trovati.

Tabella 25. Valori test di Grubbs nel mese di aprile.

Aprile	1	2	3	4	5	6	7	8	9	10	11	Media	Dev.st
★	0,03	0,04	0,02	0,03	0,11	0,02	0,02	0,03	0,02	0,02	0,01	0,0302	0,0266
★★	0,01	0,02	0,01	0,01	0,06	0,01	0,01	0,01	0,01	0,01	0,01	0,0143	0,0142
★★★	0,01	0,02	0,01	0,01	0,04	0,01	0,01	0,01	0,01	0,01	0,01	0,0119	0,0096
★★★★	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0094	0,0040
★★★★★	0,02	0,03	0,04	0,02	0,03	0,02	0,02	0,03	0,04	0,01	0,01	0,0252	0,0107

GRUBBS	1	2	3	4	5	6	7	8	9	10	11
★	-0,04	0,26	-0,38	-0,19	2,92	-0,43	-0,42	-0,14	-0,50	-0,48	-0,60
★★	-0,12	0,26	-0,37	-0,26	2,93	-0,42	-0,37	-0,12	-0,46	-0,56	-0,51
★★★	-0,06	0,41	-0,33	-0,30	2,87	-0,51	-0,46	0,03	-0,41	-0,64	-0,59
★★★★	-0,02	0,75	0,03	-0,41	2,30	-0,82	-0,61	0,77	0,15	-1,09	-1,06
★★★★★	-0,57	0,11	1,64	-0,37	0,06	-0,83	-0,12	0,75	1,68	-1,17	-1,18

t _{9,0.05}	2,262
G _c	1,9517

t _{9,0.01}	3,250
G _c	2,3966

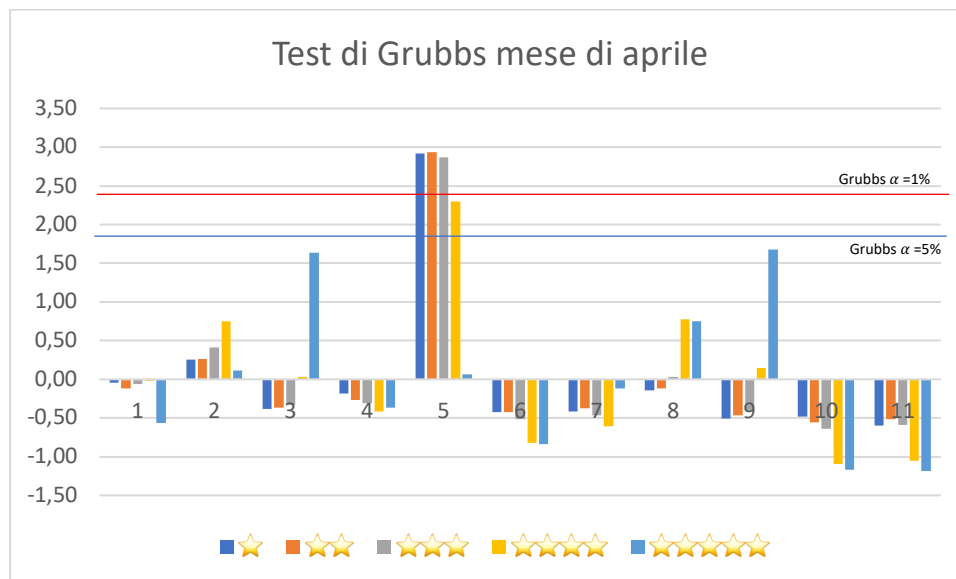


Figura 23. Test di Grubbs nel mese di aprile.

In Tabella 26 sono riportati i valori trovati per il mese di maggio. Gli unici outliers presenti in questo mese si hanno sempre per il topic 5 per i primi due rating con un livello di significatività del 5%. Non si accende nessun campanello d'allarme in questo mese con un livello di significatività all'1%. In Figura 24 si riportano graficamente i valori trovati.

Tabella 26. Valori test di Grubbs nel mese di maggio.

Maggio	1	2	3	4	5	6	7	8	9	10	11	Media	Dev.st
★	0,03	0,04	0,02	0,02	0,05	0,02	0,02	0,03	0,02	0,01	0,01	0,0237	0,0109
★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0097	0,0049
★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0095	0,0044
★★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,0120	0,0041
★★★★★	0,03	0,04	0,06	0,03	0,03	0,02	0,03	0,05	0,06	0,02	0,02	0,0361	0,0156

GRUBBS	1	2	3	4	5	6	7	8	9	10	11
★	0,29	1,51	-0,55	-0,06	2,07	-0,62	-0,70	0,45	-0,62	-0,82	-0,96
★★	0,22	1,33	-0,55	-0,21	2,27	-0,64	-0,57	0,41	-0,52	-0,87	-0,87
★★★	0,26	1,64	-0,41	-0,34	1,89	-0,81	-0,67	0,65	-0,33	-0,96	-0,91
★★★★	0,06	1,33	0,40	-0,34	1,07	-1,15	-0,64	1,21	0,60	-1,35	-1,20
★★★★★	-0,60	0,11	1,78	-0,36	-0,13	-0,84	-0,12	0,79	1,59	-1,08	-1,15

t _{9,0.05}	2,262
G _c	1,9517

t _{9,0.01}	3,250
G _c	2,3966

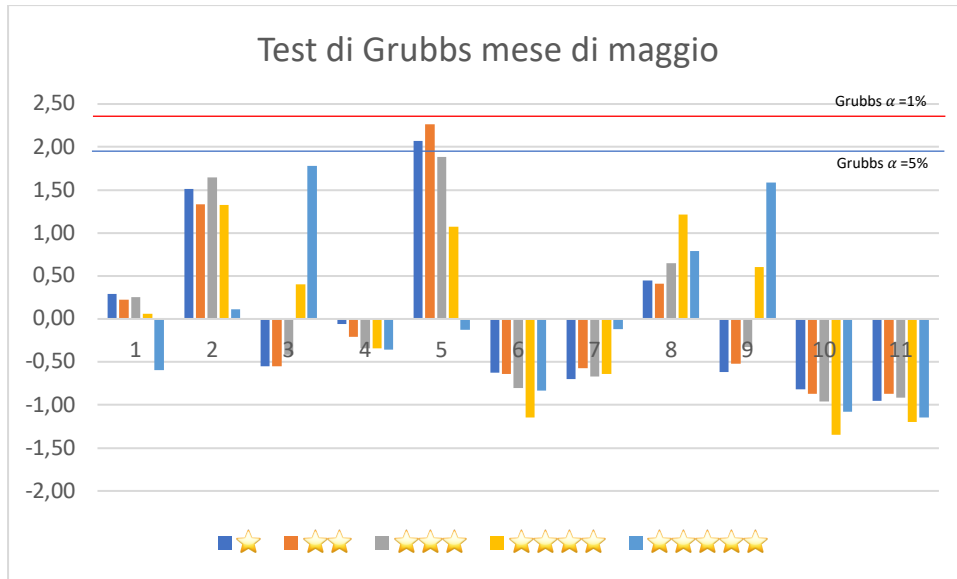


Figura 24. Test di Grubbs nel mese di maggio.

In Tabella 27 sono riportati i valori trovati per il mese di giugno. In questo mese, non si ha alcun valore di outlier per alcun livello di significatività. In questo mese quindi, tutti i valori rispettano una distribuzione normale dei dati. In Figura 25 si riportano graficamente i valori trovati.

Tabella 27. Valori test di Grubbs nel mese di giugno.

Giugno	1	2	3	4	5	6	7	8	9	10	11	Media	Dev.st
★	0,03	0,04	0,02	0,02	0,04	0,01	0,01	0,03	0,02	0,01	0,01	0,0225	0,0105
★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,00	0,00	0,0085	0,0042
★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,0092	0,0039
★★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,0121	0,0040
★★★★★	0,03	0,04	0,07	0,03	0,04	0,02	0,04	0,05	0,06	0,02	0,02	0,0387	0,0166

GRUBBS	1	2	3	4	5	6	7	8	9	10	11
★	0,30	1,74	-0,52	-0,01	1,76	-0,71	-0,74	0,61	-0,63	-0,78	-1,02
★★	0,21	1,73	-0,49	-0,22	1,87	-0,68	-0,65	0,59	-0,49	-0,93	-0,94
★★★	0,41	1,63	-0,33	-0,32	1,67	-0,79	-0,79	0,86	-0,32	-1,00	-1,03
★★★★	0,10	1,26	0,57	-0,45	0,83	-1,08	-0,71	1,39	0,61	-1,31	-1,22
★★★★★	-0,57	0,17	1,73	-0,43	-0,12	-0,85	-0,15	0,87	1,57	-1,09	-1,14

t _{9,0.05}	2,262
G _c	1,9517

t _{9,0.01}	3,250
G _c	2,3966

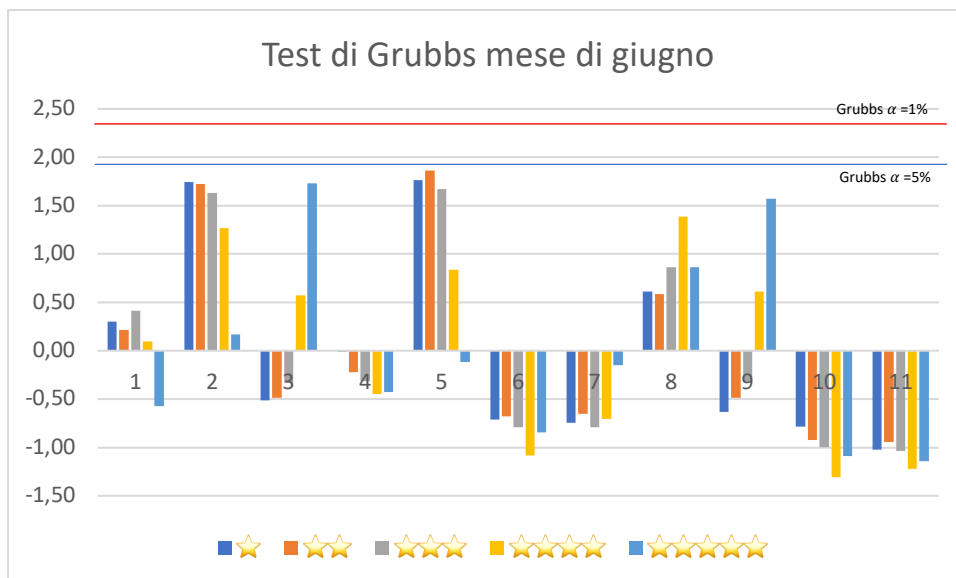


Figura 25. Test di Grubbs nel mese di giugno.

In Tabella 28 sono riportati i valori trovati per il mese di luglio. Anche in questo mese non si ha alcun outlier per un livello di significatività al 1%, mentre si accende un campanello d'allarme con rating basso per il topic 5 a un livello di significatività al 5%. In Figura 26 si riportano graficamente i valori trovati.

Tabella 28. Valori test di Grubbs nel mese di luglio.

Luglio	1	2	3	4	5	6	7	8	9	10	11	Media	Dev.st
★	0,03	0,04	0,02	0,02	0,05	0,02	0,02	0,03	0,02	0,02	0,01	0,0251	0,0127
★★	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,00	0,00	0,0087	0,0049
★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,00	0,0094	0,0046
★★★★	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,0122	0,0039
★★★★★	0,03	0,04	0,06	0,03	0,03	0,02	0,03	0,05	0,06	0,02	0,02	0,0356	0,0154

GRUBBS	1	2	3	4	5	6	7	8	9	10	11
★	0,68	1,07	-0,56	-0,02	2,28	-0,72	-0,68	0,26	-0,69	-0,67	-0,96
★★	0,46	1,14	-0,54	-0,35	2,36	-0,71	-0,49	0,32	-0,56	-0,78	-0,84
★★★	0,63	1,34	-0,48	-0,41	2,07	-0,86	-0,60	0,52	-0,48	-0,75	-0,97
★★★★	0,08	1,44	0,45	-0,03	0,75	-1,13	-0,68	1,27	0,51	-1,35	-1,31
★★★★★	-0,60	0,14	1,75	-0,42	-0,13	-0,80	-0,17	0,78	1,64	-1,05	-1,15

$t_{9,0.05}$	2,262
Gc	1,9517

$t_{9,0.01}$	3,250
Gc	2,3966

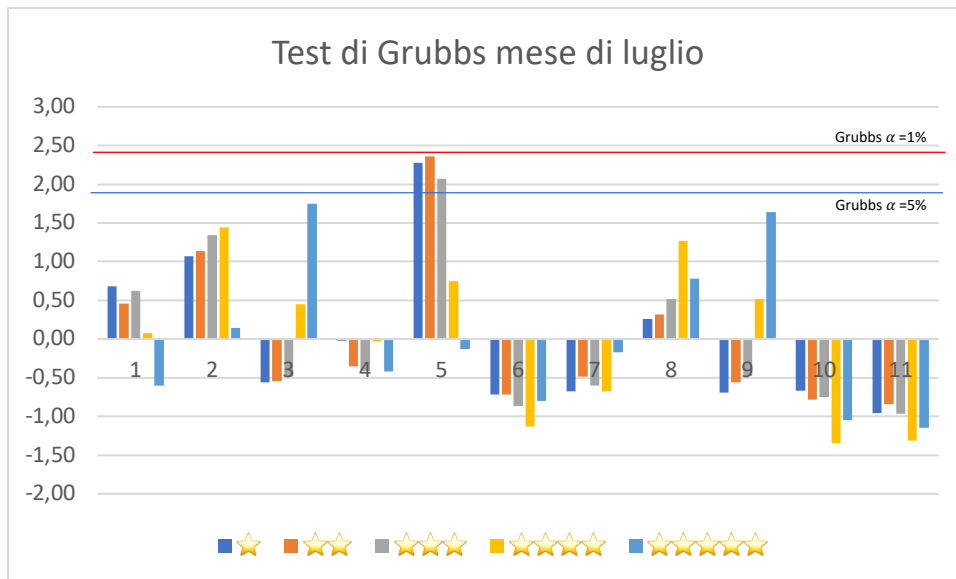


Figura 26. Test di Grubbs nel mese di luglio.

Dai dati emersi, è possibile dedurre che con un livello di significatività al 1%, gli outliers e quindi i campanelli dall'allarme si manifestano esattamente quando ce li si aspetta, ovvero nel mese di aprile per il topic 5, relativo all'affidabilità e all'user experience, quando è avvenuto lo shock discusso in tutte le analisi precedenti.

Utilizzando questo strumento, è possibile individuare valori anomali rispetto alla distribuzione normale dei dati: ad aprile, si sono riscontrati problemi di user experience e di affidabilità che hanno generato insoddisfazione al cliente. In questo mese, è aumentata la discussione del topic che tratta questo aspetto, in particolar modo è aumentata per i livelli di rating basso (da 1 a 3), ovvero si è posta maggior attenzione a aspetti negativi rispetto alle condizioni normali.

Rilassando il livello di significatività al 5%, emergono ulteriori valori sempre associabili al topic 5, che rappresenta il principale protagonista dell'analisi. L'unica eccezione si verifica nel mese di gennaio, con un outlier per il topic 9 con un rating di 5 stelle. Questa situazione, più che rappresentare un campanello d'allarme, deve essere considerata come esempio e da replicare nel tempo, in quanto ha generato una discussione superiore rispetto alla norma per il topic al massimo livello di rating. Questo topic, è riconducibile alla varietà e alla vastità offerta dal catalogo, quindi un eccesso di discussione con un livello di rating massimo, dovuto all'aver rilasciato una serie di contenuti particolarmente apprezzati dai clienti in questo mese.

Ergo, ha portato la massima soddisfazione comunicabile dal cliente e ciò rappresenta un aspetto molto positivo per il processo, che dovrebbe mirare a mantenere tale situazione nel tempo e cercare di replicarla.

TEST DI CORRELAZIONE DI SPEARMAN

I risultati dell'applicazione della formula del test di correlazione di Spearman sono riportati graficamente in Figura 27. Questa analisi, a differenza di quelle precedenti, non si focalizza sui singoli topic, ma sulle distribuzioni dei rating. I valori dei ranghi calcolati nel tempo, tenendo in considerazione il numero di recensioni e del rating associato a ciascuna di esse, è riportato invece in Tabella 29.

Tabella 29. Ranghi di Spearman.

rating	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio	TOTALE
1	2171	2489	1788	2597	2824	4084	1623	17576
2	583	1032	749	1016	1067	1869	778	7094
3	582	878	754	1181	1115	1658	672	6840
4	551	952	1097	1835	1487	1489	370	7781
5	1341	2098	3632	7117	4335	3032	411	21966
TOTALI	5228	7449	8020	13746	10828	12132	3854	61257
r(x)	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio	TOTALE
1	1	1	2	2	2	1	1	2
2	3	3	5	5	5	3	2	4
3	4	5	4	4	4	4	3	5
4	5	4	3	3	3	5	5	3
5	2	2	1	1	1	2	4	1
ρ_s	0,4	0,3	-0,3	-0,3	-0,3	0,4	0,9	-0,3

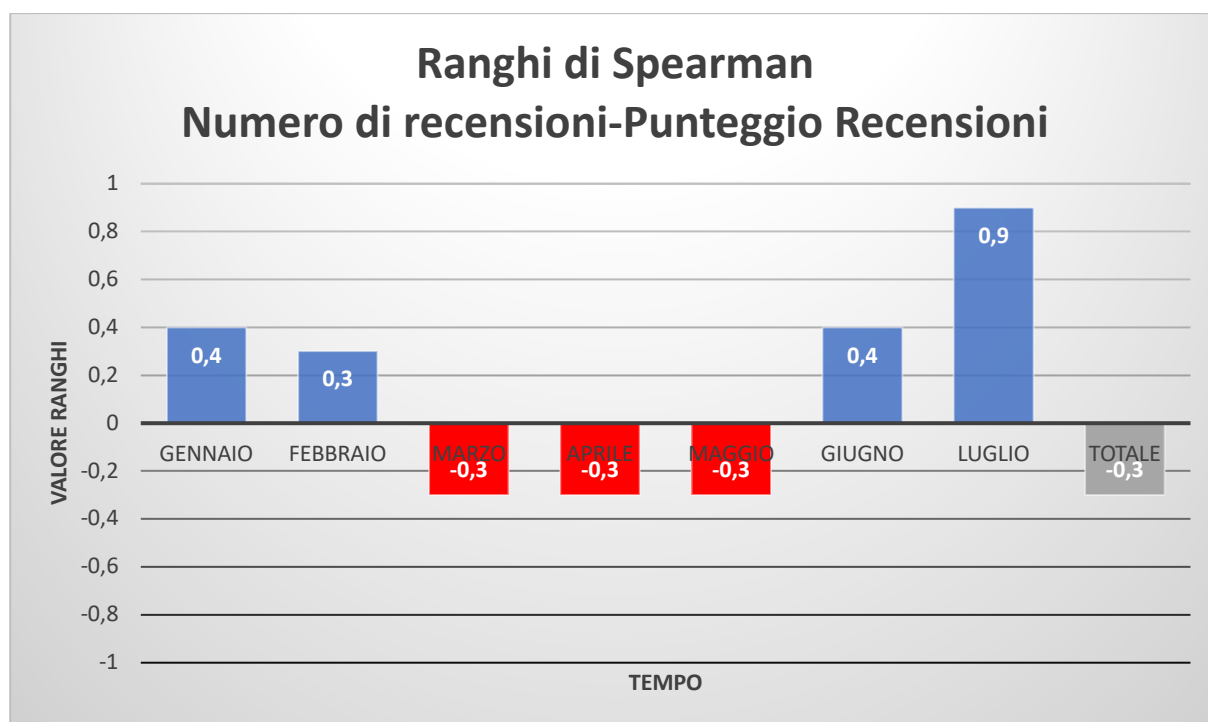


Figura 27. Ranghi di Spearman nel tempo.

Analizzando i risultati ottenuti, si osserva quanto segue:

- Gennaio, febbraio, giugno e luglio: presentano coefficienti positivi, ciò indica che si ha una correlazione tra le valutazioni dei clienti (rango dei rating) e i punteggi delle recensioni. Ovvero, a un aumento del numero di recensioni è associato un incremento delle valutazioni medie dei clienti.
- Marzo, aprile, maggio: presentano coefficienti negativi, ciò indica una correlazione negativa tra i ranghi dei rating e i punteggi delle recensioni. Ovvero, a un aumento del numero di recensioni è associata una diminuzione delle valutazioni medie dei clienti.
- Totale: presenta un valore negativo (-0,3) ciò significa che si ha una correlazione negativa complessiva tra i ranghi dei rating e i punteggi delle recensioni per i mesi considerati. Tuttavia, questa correlazione non appare particolarmente forte, ed è guidata principalmente dai mesi tra marzo e maggio.

In generale, i risultati suggeriscono che la relazione tra il numero di recensioni e le valutazioni delle recensioni può variare da mese a mese. È evidente come il risultato complessivo sia influenzato principalmente dai valori che si ottengono tra marzo e maggio, poiché le valutazioni in questi mesi hanno un impatto più significativo sui risultati finali. Questo, si presenta in linea con le valutazioni e le considerazioni effettuate in precedenza: l' user experience e l' affidabilità rappresentano un tema fondamentale per Spotify. Infatti, al manifestarsi di problemi correlati a tali aspetti, si è osservato un aumento del numero di recensioni e una diminuzione delle valutazioni medie dei clienti. Risolvendo le problematiche, la correlazione si presenta nuovamente come positiva e simile alla situazione precedente all' insorgere dei problemi.

Questa analisi quindi, ha consentito di individuare una correlazione leggermente positiva nel tempo tra il numero di recensioni e i punteggi associati, fino all' insorgere di problematiche che hanno completamente invertito questa tendenza e che hanno manipolato i valori ottenuti. Dopo la risoluzione delle problematiche e l' implementazione degli aggiornamenti, si è verificato un ritorno alla situazione pre "shock" e ad avere una correlazione nuovamente positiva. Tuttavia, nell' analisi complessiva, le valutazioni negative riscontrate nei mesi tra marzo e maggio guidano il risultato, poiché più rilevanti in termini di peso specifico, determinando un risultato complessivo negativo.

RELAZIONE TRA MRP MEDI E VALUTAZIONI CLIENTI

In questa ultima sezione, si procederà a confermare ulteriormente i risultati ottenuti dalle precedenti analisi. Si andrà ad osservare la relazione che vige tra l'indicatore MRP e le valutazioni dei clienti. Si partirà dai valori MRP precedentemente calcolati e riportati in Tabella 30.

L'MRP è associato a uno specifico topic. In questa sezione, invece, si è calcolata una media degli MRP per ogni valutazione per ogni livello di rating.

I risultati sono presentati graficamente nella Figura 28, con l'asse delle ascisse dedicato al tempo e sull'asse delle ordinate rappresentante i valori di MRP.

Tabella 30. Valori MRP medio nel tempo.

MRP	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio
★	0,03	0,03	0,03	0,03	0,02	0,02	0,03
★★	0,01	0,01	0,01	0,01	0,01	0,01	0,01
★★★	0,01	0,01	0,01	0,01	0,01	0,01	0,01
★★★★	0,01	0,01	0,01	0,01	0,01	0,01	0,01
★★★★★	0,03	0,03	0,03	0,03	0,04	0,04	0,04

Da questa analisi si può osservare il trend individuato nei paragrafi precedenti. È evidente sin da subito che nel mese di aprile aumenta il valore di MRP nelle recensioni caratterizzate da valutazioni prettamente negative (da 1 a 3 stelle), mentre diminuiscono quelle prettamente positive (da 4 a 5). Questo conferma ulteriormente il trend sviluppatosi nel mese di aprile, durante il quale si è verificato un cambiamento rispetto ai mesi passati, che porta maggior insoddisfazione da parte degli utenti. Inoltre, si osserva, a conferma dei risultati precedentemente ottenuti, come dal mese di maggio i valori tendano a ristabilirsi sui valori precedenti ad aprile.

In sintesi, il comportamento individuato per queste determinanti nelle recensioni è prettamente stazionario, ma uno shock può generare cambiamenti improvvisi che portano a distribuzioni anomale. Allo stesso tempo, la risoluzione delle problematiche consente un ritorno repentino alla situazione stazionaria di partenza.



Figura 28. Valori MRP medi nel tempo.

CONCLUSIONI

La ricerca ha fatto emergere l'importanza della digital Voice of Customer (VoC) come risorsa fondamentale per la comprensione delle aspettative, delle esigenze, dei bisogni e delle opinioni dei clienti. È una risorsa dal valore inestimabile per le aziende, poiché consente di migliorare la qualità dei propri servizi o prodotti attraverso un continuo feedback degli utilizzatori effettivi. Per valutare questi aspetti gioca un ruolo fondamentale la soddisfazione del cliente.

È inoltre emerso che la trasformazione di questi continui feedback in strumenti e metriche di qualità è fondamentale per misurarli e ad avere una rilevanza significativa. Avere metriche di qualità misurabili e rilevanti fornisce un quadro chiaro per il management aziendale dal punto di vista della qualità, che guida e consente di indirizzare le scelte strategiche e di modificare, o eventualmente, riprogettare, i processi aziendali.

L'utilizzo di strumenti di monitoraggio consente di comprendere l'evoluzione nel tempo della digital-VoC, offrendo una chiara prospettiva del cambiamento delle percezioni e delle richieste dei clienti nel tempo. Attraverso l'utilizzo di un algoritmo di *topic modeling* si sono individuate le determinanti di qualità di un caso di studio specifico.

Monitorare nel tempo la digital-VoC corrisponde a monitorare nel tempo queste determinanti di qualità individuate nelle recensioni. Il monitoraggio nel tempo di queste metriche di qualità è alla base del miglioramento continuo nelle aziende, poiché consente di far emergere gli elementi di debolezza su cui le aziende devono effettuare interventi correttivi. Nel caso studio di Spotify preso in esame, la determinante di qualità volta a individuare l'user experience e l'affidabilità del servizio ha generato valori anomali nel mese di aprile, condizionando inevitabilmente i valori di rating e le soddisfazioni percepite dai clienti finali. Spotify è riuscita prontamente ad individuare questa area di debolezza e ad effettuare un'azione correttiva nell'immediato come si dimostra con il ritorno a valori che seguono la distribuzione attesa dei dati già a partire dal mese successivo. Le aziende dovrebbero comportarsi allo stesso modo, ove si presentassero valori anomali negativi all'interno dei propri parametri monitorati, poiché la soddisfazione del cliente e la capacità di risolvere eventuali problematiche sono cruciali per la fidelizzazione del cliente, elemento essenziale per il successo aziendale.

L'importanza della digital-VoC va oltre al monitoraggio e controllo, estendendosi a capacità di condurre analisi predittive con campioni di dati sempre più ampi.

Questa è la base del vantaggio competitivo, ovvero riuscire ad anticipare i bisogni e le richieste dei propri clienti, per trarne vantaggio nei confronti dei competitors. La prospettiva per il futuro prevede

che dall'introduzione dell'intelligenza artificiale in campo aziendale possa esasperare ulteriormente l'intero processo.

Tuttavia, una problematica riscontrata utilizzando la digital-VoC è la qualità dei dati analizzati e la loro rappresentatività del reale pensiero e delle reali percezioni dei clienti finali. I clienti infatti, tendono a rilasciare una recensione, o un contenuto, principalmente per esprimere una lamentela o per esprimere una valutazione di eccellenza. Si ottengono in questo modo principalmente valutazioni estreme. Si rischiano così di perdersi le valutazioni intermedie percepite da alcuni clienti che non rilasciano spontaneamente un commento perché potrebbero non sentirne il bisogno.

Nello studio, è emersa l'importanza dell'indicatore MRP, che consente di comprendere quanto sia discusso un topic nei record di digital VoC associati con uno specifico livello di rating. Mediante l'utilizzo di questo indicatore, si riescono ad individuare i profili di discussione degli argomenti all'interno dei contenuti rilasciati dagli utenti e si riescono a classificare in (Barravecchia et al., 2020):

- Topic associati a livelli di soddisfazione positiva.
- Topic associati a livelli di soddisfazione neutra.
- Topic associati a livelli di soddisfazione negativa.

Questa classificazione fornisce un primo approccio per comprendere ciò che i clienti percepiscono come di valore nella fruizione del servizio o nell'utilizzo del prodotto offerto. Inoltre, fornisce una panoramica degli elementi critici che generano insoddisfazione e per gli elementi neutri che non influenzano la percezione del cliente.

L'obiettivo del lavoro consisteva nell'analizzare i valori dell'indicatore MRP in un determinato periodo, per individuare uno strumento che consentisse di individuare eventuali comportamenti anomali riscontrati. Tale analisi mira a risalire le cause di tali anomalie con il fine di apportare azioni correttive per migliorare i processi aziendali.

Per il monitoraggio temporale dell'indicatore MRP, è stato adottato il test statistico di Grubbs, un test che permette di individuare gli outliers e i valori anomali all'interno di un campione in analisi. Questo test funge da "campanello d'allarme" nel rilevare eventuali valori che si discostano dalla distribuzione dei valori attesi. Il test è stato sottoposto a un esempio pratico, il caso studio di Spotify, confermando pienamente i risultati emersi dalle analisi precedenti.

Lo studio dell'MRP nel corso del tempo inoltre, fornisce una fotografia dettagliata dell'evoluzione dei bisogni dei clienti, osservando come variano nel tempo le percezioni e le aspettative, offrendo una prospettiva dinamica e informativa per la gestione aziendale.

L'analisi del cliente e delle sue percezioni attraverso la digital-VoC rappresenta attualmente un elemento imprescindibile per ogni azienda. L'analisi dell'indicatore MRP attraverso il test di Grubbs è risultata essere uno strumento molto prezioso per l'individuazione di valori anomali all'interno della distribuzione attesa dei dati sulla valutazione dei clienti. Mediante questo strumento, si possono migliorare la qualità e l'affidabilità delle analisi che si basano sulle valutazioni dei clienti, si può avere inoltre un monitoraggio e un controllo continuo dei processi aziendali. Questo processo mira a migliorare l'efficienza e la produttività dell'azienda.

In generale, il presente studio offre un contributo rilevante per le organizzazioni che intendono introdurre approcci innovativi nell'analisi dei dati. In particolare, l'impiego di metriche avanzate come l'MRP e l'applicazione del test statistico di Grubbs per individuare anomalie rappresentano metodologie attuali e di spicco.

Le potenziali applicazioni pratiche di questo studio nell'ambito aziendale sono diversificate e promettenti. L'utilizzo di strumenti avanzati come l'MRP può promuovere una maggiore efficienza nei servizi offerti ai clienti, basandosi sulla comprensione dettagliata delle loro esigenze. Questo approccio risulta particolarmente utile nell'identificare gli elementi responsabili di recensioni positive e, di conseguenza, di elevata soddisfazione clienti.

Inoltre, le implicazioni pratiche abbracciano il miglioramento del design del prodotto, orientato alle percezioni, richieste e bisogni emersi dai clienti. Questa metodologia non solo contribuisce allo sviluppo di prodotti più aderenti alle aspettative del cliente, ma può anche conferire vantaggi strategici nel contesto del marketing e della pubblicità. Utilizzando queste informazioni per plasmare strategie di marketing mirate, le aziende possono massimizzare il coinvolgimento dei clienti, guadagnandone la fiducia e aumentando le prospettive di successo sul mercato.

In conclusione, l'approccio proposto si configura come uno strumento prezioso per informare decisioni aziendali, ottimizzando processi e risultati nel contesto operativo.

RIFERIMENTI

- Abbasi S., Yeganeh A., Shongwe S., 2022. Monitoring non-parametric profiles using adaptive EWMA control chart, *Scientific reports* 12, 14336.
- Adikaram K., Hussein M., Effenberger M., Becker T., 2015. Data Trasformation Technique to improve the outlier detection power of Grubbs' Test for Data Expected to follow linear relation. *Journal of Applied Mathematics*. 1–9.
- Barravecchia F., Mastrogiacomo L., 2022. Statistical Process Control techniques to monitor quality determinants in digital Voice-of-Customer, *Proceedings book of the 5th International Conference on Quality Engineering and Management*, 120-141.
- Barravecchia F., Mastrogiacomo L., Franceschini F., 2020, Identifying car sharing quality determinants: a data-driven approach to improve engineering design. *International Conference on Quality Engineering and Management*, 125–140.
- Barravecchia F., Mastrogiacomo L., Franceschini F., 2020. Categorizing Quality Determinants in Mining User-Generated Contents. *Sustainability*, 12(23), 1-12.
- Barravecchia F., Mastrogiacomo L., Franceschini F., 2022. KA-VoC Map: Classifying product Key-attributes from digital voice-of-customer, *Quality Engineering* 34 (3), 344-358.
- Barravecchia F., Mastrogiacomo L., Franceschini F., 2023. Product quality tracking based on digital Voice-of-Customers, *Total Quality Management & Business Excellence*, 1-24.
- Barravecchia F., Mastrogiacomo L., Franceschini F., 2022. Digital voice of-customer processing by topic modelling algorithms: insights to validate empirical results, *International Journal of Quality & Reliability Management* 36 (6), 1453-1470.
- Berry M., Kogan J., 2010. *Text mining: applications and theory*. John Wiley & Sons.
- Biesok G., Wyrod-Wrobel J., 2011. Customer Satisfaction: Meaning and Methods of Measuring. In G. Biesok, & J. Wyrod-Wrobel (Eds.), *Marketing and Logistic Problems in the Management of Organization*, 23-41.
- Blei D. M., Lafferty J. D., (2007) Correlated topic models. *The Annals of Applied Statistics* 1(1), 17-35.
- Blei D., 2012. Probabilistic Topic Models. Conference Paper: *Communications of the ACM*, 55, 77-84.
- Blei D., Ng A., Jordan M., 2003. Latent Dirichlet Allocation, *Journal of Machine Learning Research* 3, 993-1022.

- Bressoud T., White D., 2020. Web Scraping. Introduction to Data Systems. *Springer*. 681-714.
- Chang T., Gan F., 2006. Monitoring Linearity of Measurement Gauges. *Journal of Statistical Computation and Simulation* 76, 889-911.
- Chen S., Nembhard H., 2007. A High-Dimensional Control Chart for Profile Monitoring, *Technometrics* 27(4), 451-464.
- Costa E., Lorena A., Carvalho A., Freitas A., 2007. A review of performance evaluation measures for hierarchial classifiers. *Evaluation Methods for Machine Learning II: Papers from the AAAI-2007 Workshop*, 1-6.
- Delen D., Crossland M., 2008. Revenue and Prot Implications of industrial service strategies. *Journal of Service Research* 17(1):23-39
- Ding Y., Zeng L., Zhou S., 2006. Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes. *Journal of Quality Technology* 38, 199- 216.
- Durrani N., Schmid H., Fraser A., Koehn P., Schütze H., 2015. The operation sequence model—combining n-gram-based and phrase-based statistical machine. *Computational Linguistic* 41(2), 157-186.
- Feinerer I., Hornik K., Meyer. D., 2008. Text mining infrastructure in R. *Journal of statistical software*, 25, 1-54.
- Forina M., 2014. Fondamenta per la chimica analitica: Applicazioni della statistica, *ResearcGate*.
- Franceschini F., Galetto M., 2011. Esercizi di ingegneria della qualità, *Celid*, Torino.
- Franceschini F., 2001. Dai prodotti ai servizi. Le nuove frontiere per la misura della qualità. *UTET*, Torino, IT.
- Franceschini F., Galetto M., 2003. Tecniche di Quality Tracking per il supporto alla pianificazione di nuovi prodotti. *QUALITÀ*, 2003(6), 25-28.
- Franceschini F., Galetto M., Maisano D., Mastrogiacomo L., 2019. Ingegneria della Qualità. Applicazioni ed esercizi (Quarta edizione), *CLUT*, Torino.
- Glen S., 2022. Grubbs' Test for Outliers (Maximum Normed Residual Test). *StatiticsHowTo.com*, URL: .
- Gosset M., 1921. An experimental determination of the probable error of Dr. Spearman's correlation coefficients. *Biometrika* 13.
- Grubbs Frank E., 1950. Sample criteria for testing outlying observations. *Annals of Mathematical Statistics* 21 (1), 27–58.

- Guo Y., Barnes S. J., Jia Q., 2017. Mining meaning from online ratings and reviews: information systems research: challenges, promises and guidelines. *European Journal of Information Systems*, 25(4), 289-302.
- Jach P., Antony J., Thomson S., Cudney B., Furterer S., 2021. Voice of the customer as a tool for service quality analysis in public transport. *The TQM Journal* 34(3), 448-475.
- Jivani A., 2011. A comparative study of stemming algorithms. *Int. J. Comp. Tech* 2(6), 1930-1938.
- Kherwa P., Bansal P., 2019. Topic modeling: a comprehensive review. *EAI Endorsed transactions on scalable information systems*, 7(24).
- Mastrogiacomo L., Barravecchia F., Franceschini F., Marimon F., 2021. Mining quality determinants of product-service systems from user-generated contents, *Quality Engineering* 33 (3), 425-442.
- Meyer D., Hornik K., Feinerer I., 2008. The Role of Text Pre-processing in Sentiment Analysis, Elsevier, *Information Technology and Quantitative Management* 17, 26-32.
- Moguerza J., Munoz A., Psarakis S., 2007. Monitoring Nonlinear Profiles using Support Vector Machines. *Lecture Notes in Computer Science*, 4756, 574-583.
- Montgomery D. C., 2009. Introduction to Statistical Quality Control. 6th ed. *Wiley*, Hoboken, NJ.
- Müller. O., Junglas I., Brocke J., Debortoli S., 2016. Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Communications of the Association for Information Systems* 39(1).
- Myers J., Well A., Lorch R., 2013. Research design and statistical analysis. *Routledge*.
- Norossana R., Amiri A.; Soleimani P., 2007. On the Monitoring of Autocorrelated Linear Profiles. *Communications in Statistics-Theory and Methods* 37(3), 425-442.
- Norossana R., Saghaei A., Amiri A., 2011. Statistical Analysis of Profile Monitoring, *Wiley*.
- Norossana R., Soleimani P., 2007. Effect of Within Profile Autocorrelation on the Performance of Linear Profiles. *Proceedings of the 5th International Industrial Engineering Conference*, Tehran, Iran.
- O'Reilly T., 2007. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies* (1), 17.
- Roberts M., Stewart B., Airoidi E., 2016. A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association*, 111, 1-49.

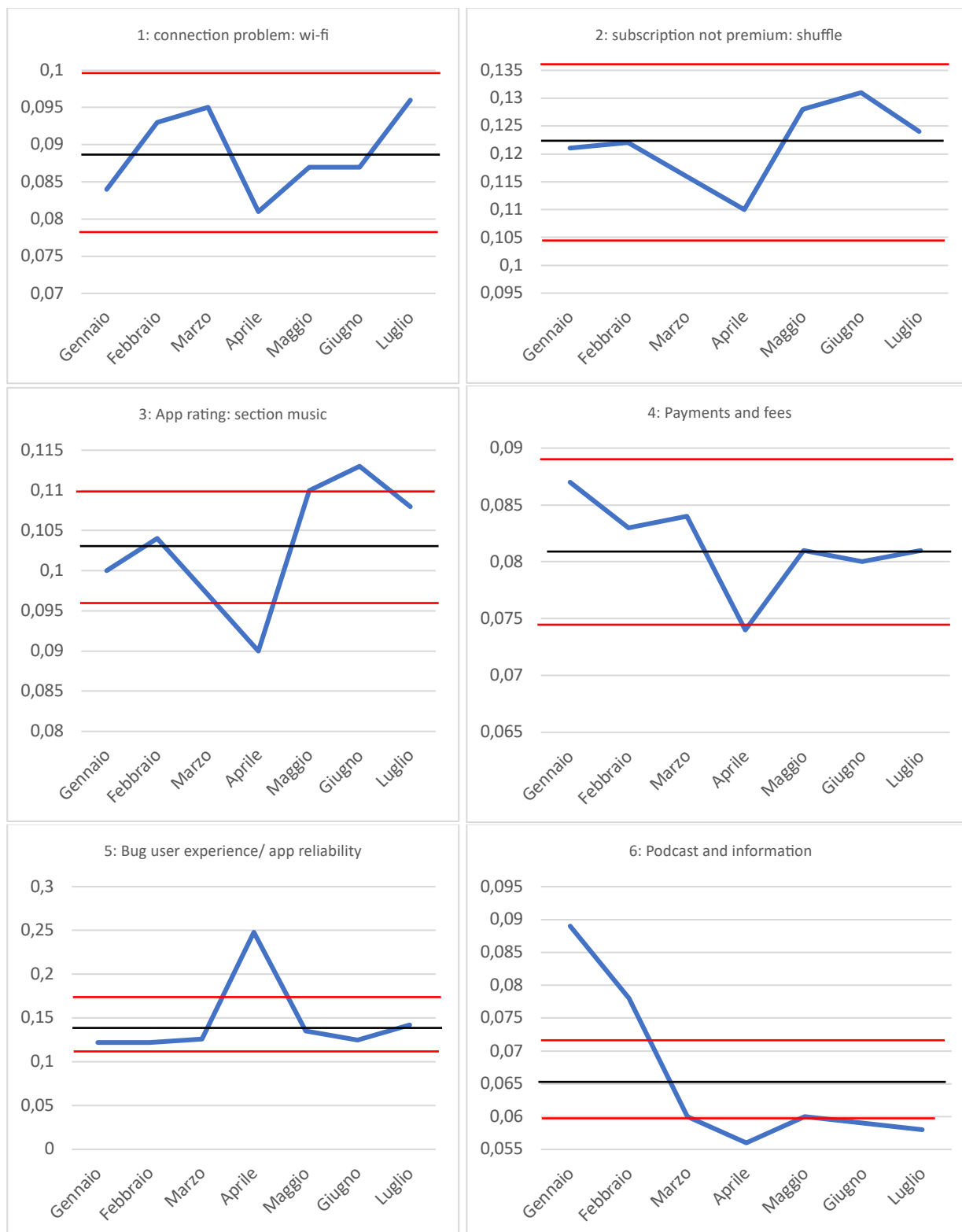
- Roberts M., Stewart B., Tingle D., 2019. stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91(2), 1-40.
- Roberts M., Stewart B., Tingle D., Airoldi E., 2013. The Structural Topic Model and Applied Social Science. *ICONIP*, 2013, 1-4.
- Roberts M., Stewart B., Tingle D., Lucas C., Leder-Luis J., Gadarian S. K., Albertson B., Rand D. G., 2014. Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064-082.
- Roberts M., Stewart B., Tingley D., 2019. STM: An R Package for Structural Topic Models, *Journal of Statistical Software* 91(2), 1-40.
- Scott J., Baldridge J., 2013. A recursive estimate for the predictive likelihood in a topic model. *Journal of Machine Learning Research* 31, 527–535.
- Spotify (2023), sito web aziendale, URL: <https://open.spotify.com/intl-it>.
- Spearman C., 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15, 72–101.
- Student (Gosset W.), 1908. The probable error of a mean. *Biometrika* (6), 1-25.
- Timoshenko A., Hauser J., 2019. Identifying customer needs from user-generated content. *Marketing Science* 38(1), 1-20.
- Ulaga W., Reinartz W., 2008. How to sell services more profitably. *Harvard Business Review* 86(5), 129.
- UNI, UNI EN ISO 9000:2015 Sistemi di Gestione della qualità- Principi e linee guida per la loro applicazione, *Ente Nazionale Italiano di Unificazione*, 2015.
- Universität Göttingen (2023), sito web universitario, URL: <http://www.sediment.uni-goettingen.de/staff/dunkl/software/pep-grubbs.pdf>.
- Vandermerwe S., Rada J., 1988. Servitization of Business: Adding Value by Adding Services, *European Management Journal* (6), 314-324.
- Wallach H., Mimno D., McCallum A., 2009. Rethinking LDA: Why priors matter. *Advances in neural information processing systems* 22, 1973-1981.
- Wang Y., Tseng M., 2011. Integrating comprehensive customer requirements into product design. *CIRP Annals – Manufacturing Technology* 60(1), 175-178.
- Western Electric, 1956. Statistical Quality Control Handbook. *Western Electric Corporation*, Indianapolis, IN.
- Williams J., Woodall W., Birch J., 2007. Statistical Monitoring of Nonlinear Product and Process Quality Profiles. *Quality & Reliability Engineering International* 23, 925-941.

- Woodall W., 2007. Profile Monitoring, *Encyclopedia of Statistics in Quality and Reliability*.
- Woodall W., Spitzner D.; Montgomery D., Gupta S., 2004. Using Control Charts to Monitor Process and Product Quality Profiles. *Journal of Quality Technology* 36, 309-320.
- Zaki M., J. McColl-Kennedy., 2020. Text Mining Analysis Roadmap (TMAR) for Service. *Journal of Services Marketing* 34(1), 30–47.
- Zhu J., Lin D., 2007. Monitoring the Slopes of Linear Profiles, *Quality Engineering* 22(1), 1-12.

APPENDICE

CARTE X IMTP

Le linee rosse delimitano i limiti di controllo superiore e inferiore, mentre la linea nera indica il limite di controllo centrale.



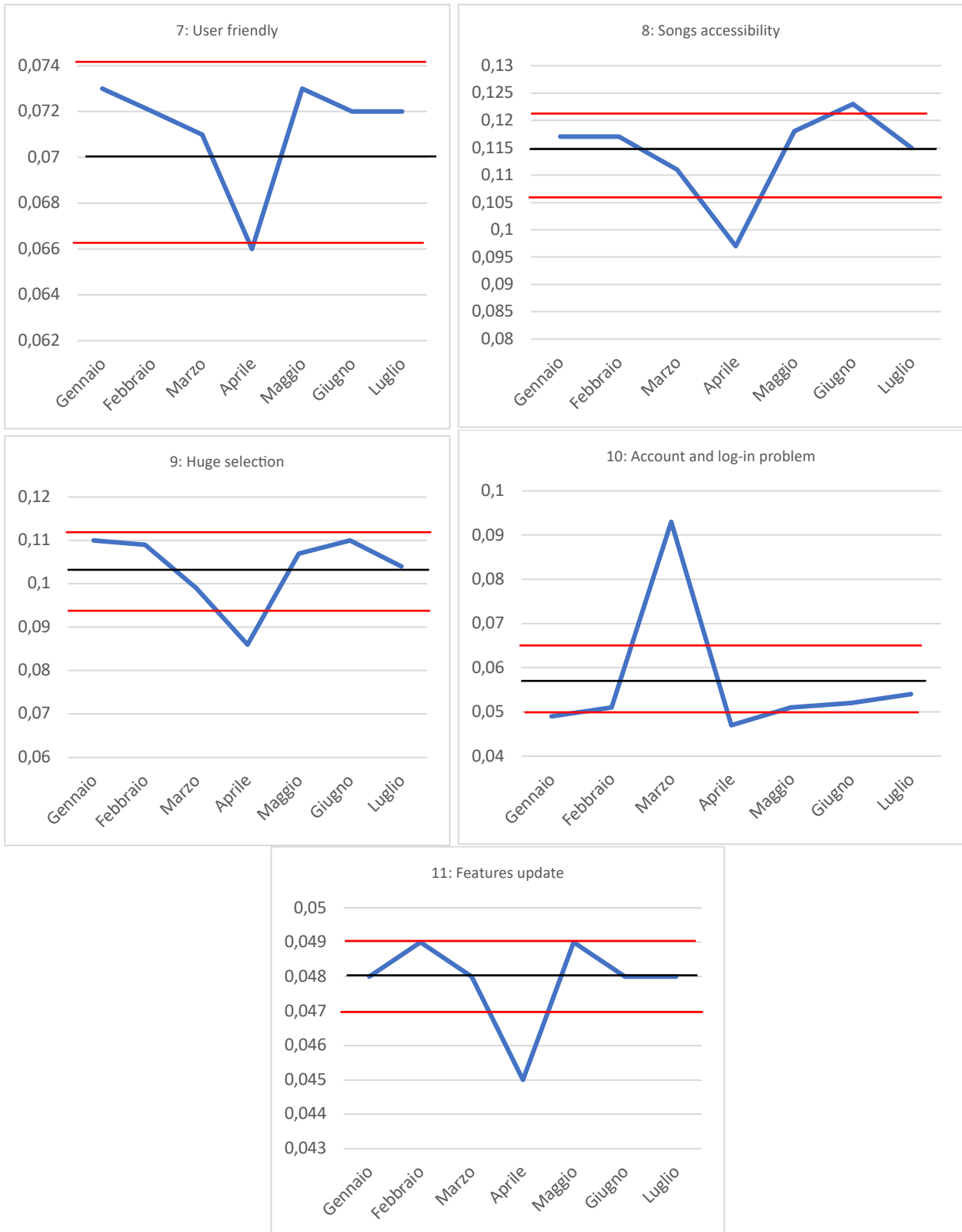


Figura 29. Carte X IMTP.

CARTE S IMTP

Le linee rosse delimitano i limiti di controllo superiore e inferiore, mentre la linea nera indica il limite di controllo centrale.

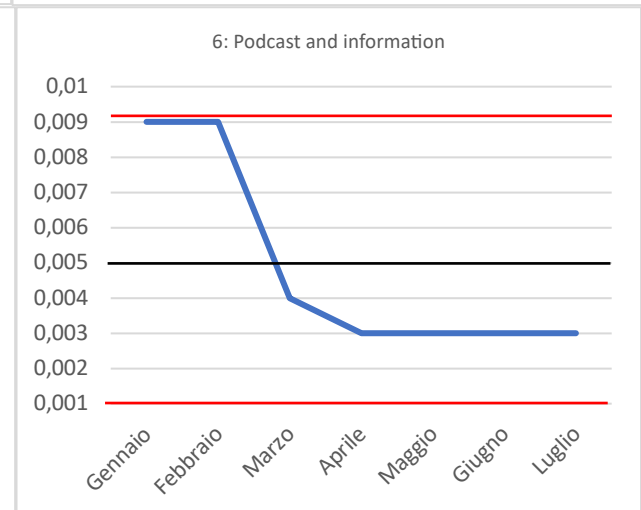
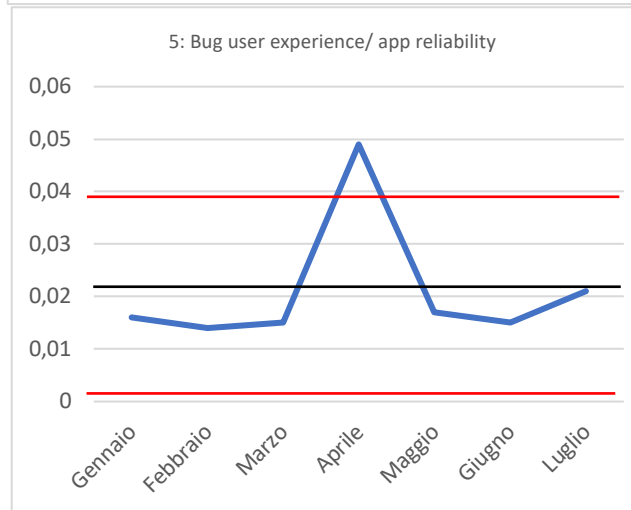
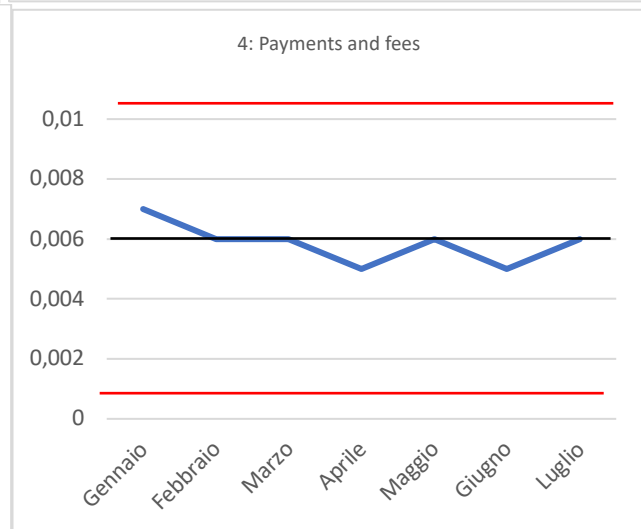
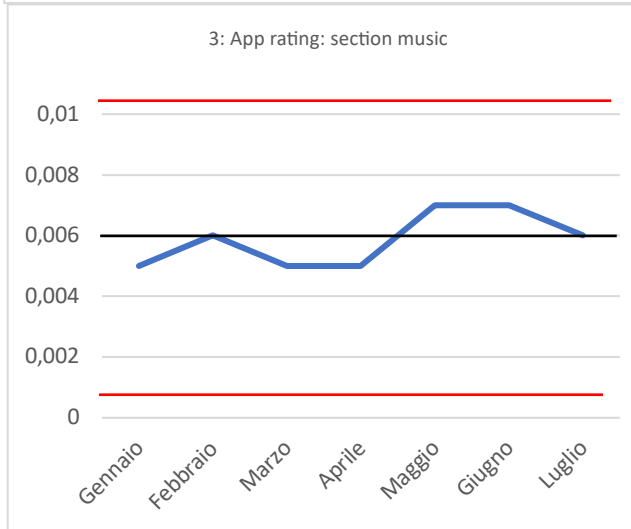
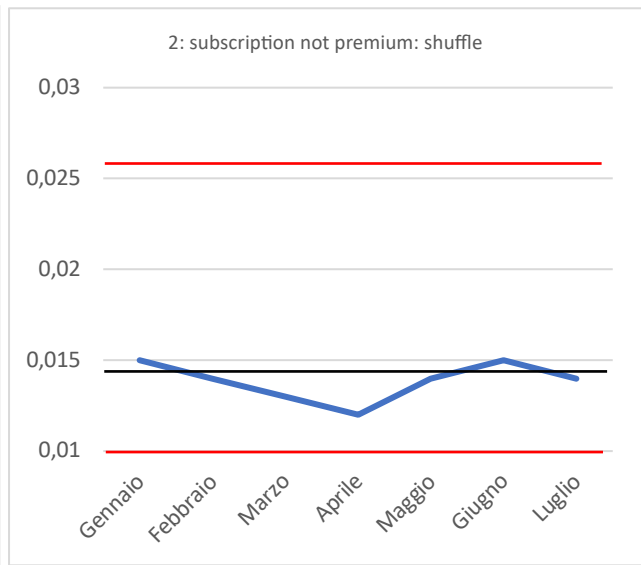
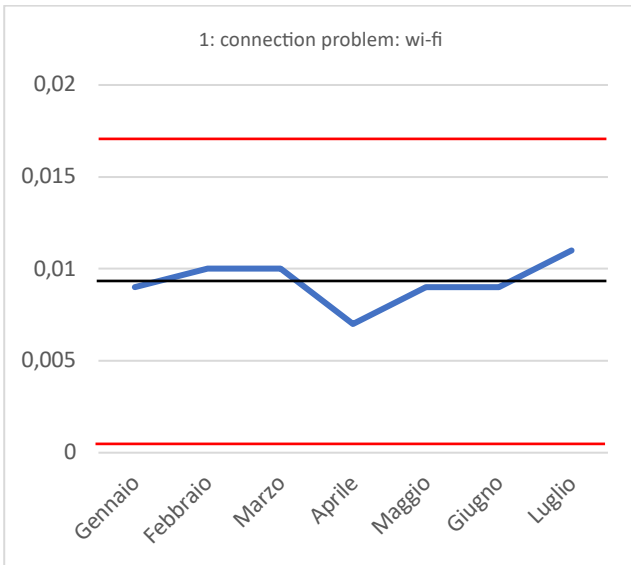




Figura 30. Carte S IMTP.