POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale
Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea Magistrale

Identificazione delle dimensioni di qualità latenti tramite l'analisi di "user generated contents" (UGC) con algoritmi di text mining: il caso delle strutture ospedaliere italiane

Relatore

Prof. Luca Mastrogiacomo

Candidato

Alessandro Visentin

Correlatore

Prof. Fiorenzo Franceschini

Marzo 2020

Sommario

Indice delle tabelle **2**
Indice delle figure **4**
Abstract **6**
Capitolo I **8**
Introduzione ai servizi 8
Caratteristiche dei servizi 9
Modelli di studio della qualità nei servizi 10
La trasformazione digitale nella sanità 13
Qualità nella sanità 14
Gli user-generated contents 17
Latent Dirichlet Allocation (LDA) 19
Capitolo II **28**
Applicazione di LDA per valutare le dimensioni di qualità 28
Estrazione del corpus di recensioni e preelaborazione del campione 28
Applicazione di LDA 30
Estrazione dei topic e assegnazione di questi ultimi alle recensioni 32
Validazione dell'algoritmo 34
Etichettatura dei topic 39
Capitolo III **40**
Presentazione dei risultati: Panoramica generale 40
Analisi regionale 44
Presentazione dei risultati: Nord Italia 45
Presentazione dei risultati – Centro Italia 48
Presentazione dei risultati – Mezzogiorno 51
Presentazione dei risultati – Roma 55
Presentazione dei risultati – Milano 58
Presentazione dei risultati – Napoli 61
Presentazione dei risultati – Torino 64
Capitolo IV **68**
Discussione dei risultati 68
Confronto dati tra Nord Italia, Centro Italia e Mezzogiorno 71

Confronto dati tra le città di Roma, Milano, Napoli e Torino 74
Limitazioni dello studio e sviluppi futuri 79
Capitolo V 81
Conclusioni 81
Appendice A 82
Fattori determinanti della qualità dei servizi (Parasuraman, Zeithaml and Berry, 1985) 82
Appendice B 83
Fattori determinanti della qualità dei servizi sanitari (Duggirala, Rajendran and Anantharaman, 2008) 83
Appendice C 84
Lista dei topic 84
Appendice D 88
Istogrammi della percentuale di prevalenza per ogni topic, quando il topic i-esimo è prevalente 88
Appendice E 94
Codice di Matlab – Algoritmo LDA 94
Appendice E 95
Lista delle parole eliminate 95
Codice Matlab – Preprocessamento del testo 96
Appendice F 97
Lista delle parole rimosse 97
Bibliografia 100

Indice delle tabelle

Tabella 1 Elenco dei parametri dell'algoritmo LDA e relativa spiegazione 21
Tabella 2 Esempio di una lista di parole appartenenti a due topic 26
Tabella 3 Modalità di assegnazione del topic a una recensione 34
Tabella 4 Lista delle etichette e il relativo numero di recensioni a esse collegate ... 36

Tabella 5 Lista degli indicatori e i relativi risultati sul campione utilizzato per la validazione 37
Tabella 6 Lista autori e relativi articoli scientifici e valori dei principali indicatori di bontà di validazione 39
Tabella 7 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere di tutta l'Italia 41
Tabella 8 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere di tutta l'Italia 42
Tabella 9 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere del Nord Italia 46
Tabella 10 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere del Nord Italia 47
Tabella 11 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere del Centro Italia 49
Tabella 12 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere del Centro Italia 50
Tabella 13 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere del Mezzogiorno 52
Tabella 14 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere del Mezzogiorno 54
Tabella 15 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Roma 56
Tabella 16 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Roma 57
Tabella 17 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Milano 59

Tabella 18 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Milano 60
Tabella 19 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Napoli 62
Tabella 20 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Napoli 63
Tabella 21 Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Torino 65
Tabella 22 Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Torino 66
Tabella 23 Numero di abitanti per ospedale per le tre zone geografiche: Nord Italia, Centro Italia e Mezzogiorno 72

Indice delle figure

Figura 1. Schema del modello PZB (Parasuraman, Zeithaml and Berry, 1985) 11
Figura 2 Rappresentazione grafica dell'algoritmo LDA (Blei, Ng and Jordan, 2003) 23
Figura 3 Relazione tra numero di topics e perplexity per i solver cvb0, savb, avb, cgs 31
Figura 4 Relazione tra Initial Topic Concentration e Perplexity con parametro NumTopics=15 32
Figura 5 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere di tutta Italia 43
Figura 6 Istogramma della percentuale di prevalenza quando il topic 1 è prevalente 44

Figura 7 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere del Nord Italia 48
Figura 8 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere del Centro Italia 51
Figura 9 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere del Mezzogiorno 54
Figura 10 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Roma 58
Figura 11 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Milano 61
Figura 12 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Napoli 64
Figura 13 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Torino 67
Figura 14 Confronto per ogni topic tra Nord Italia, Centro Italia e Mezzogiorno sulla percentuale media di prevalenza 73
Figura 15 Percentuale di recensioni confuse divise per zona geografica 74
Figura 16 Confronto per ogni topic tra Roma, Milano, Napoli e Torino sulla percentuale media di prevalenza 78
Figura 17 Percentuale di recensioni confuse divise per le quattro principali città italiane (Roma, Milano, Napoli, Torino) 79

Abstract

Many public and private organizations try to understand customers' perceptions and analyze their outcomes, they aim to expand the market and improve the quality of their services.

This is mainly achieved by intercepting consumers' needs. For what concerns services, the use of numerous quality measurement techniques, widely tested in the most mature manufacturing context, is suboptimal. To overcome this issue a considerable effort, for what concerns the research of innovative methods for measuring attitudes, was made by using computer science competence, specifically text mining techniques.

More precisely, the public health sector is rapidly changing due to the lower state expenditure allocated compared to the past, summed up with necessity to meet the increasingly demanding needs of citizens and constraints to optimize the resources.

My dissertation identifies latent quality scenarios through the analysis of Italian hospitals reviews published by users, with a text mining algorithm called Linear Dirichlet Allocation (LDA).

The outcome shows how data analyzed using LDA provide an output comparable to the classic survey methods and compares the different feels of perceived quality widespread through the Italian hospitals divided in three macro areas (North, Centre and South of Italy).

Data feed used for LDA algorithm analysis took the reviews on all Italian public hospitals extracted from Google. The answers provided show that the quality levels identified are mostly comparable with those available in the literature. Furthermore, it was possible to underline differences among the topics of interest covered by the reviews on a geographical basis. For example, in Northern Italy users discuss more about ER quality than the Southern, but in the latter the reviews are on average more negative in terms of hospital services perception.

It therefore emerges that data mining algorithms can be used to capture customer perceptions by analyzing the content published by the latter online. Future research may analyze other user generated contents such as video contents. This would require

more complex algorithm of image mining, which includes object recognition, image indexing and retrieval, image classification and clustering, association rules mining, and neural network (Tran, 2019).

Capitolo I

Introduzione ai servizi

Il settore dei servizi, (o settore terziario) sta acquisendo continuamente un'importanza maggiore nell'economia dei paesi industrializzati. Nel 2017 la percentuale di PIL globale apportato dai servizi è stata pari al 65% (*Services, value added (% of GDP) | Data*, 2017). In maniera analoga, secondo il fenomeno della servitization, i settori primario e secondario non limitano la propria offerta esclusivamente ai beni, bensì anche ai servizi (Vandermerwe and Rada, 1988).

Le ragioni alla base del suddetto fenomeno sono molteplici: in primo luogo, data la natura sempre più globale e competitiva del mercato ne consegue un assottigliarsi dei profitti derivanti dalla mera vendita dei prodotti; in aggiunta si realizza la necessità di differenziarsi dai concorrenti e non competere esclusivamente sulla base dei costi (Neely, 2008).

In base al grado di contatto del cliente con il fornitore di servizi si distinguono tre tipi di servizi (Gupta and Chen, 1995):

- Servizi puri: il cliente deve essere presente durante l'erogazione del servizio (e.g. i servizi alla persona, parrucchiere).
- Servizi misti: servizi in cui è presente l'interazione faccia a faccia con il consumatore (*front office*) e lavoro invisibile al cliente (back office) come ad esempio le compagnie aeree.
- Servizi semi-manifatturieri: questa tipologia di servizi, offerta ad esempio dalle compagnie telefoniche, non prevede l'interazione faccia a faccia con il cliente.

Un'altra possibile distinzione è basata sulla natura del fornitore di servizi e in questo caso si effettua una distinzione tra servizi privati e pubblici (Franceschini, 2001). I primi sono erogati da organizzazioni private, le quali hanno come fine ultimo il profitto. Diversamente, i servizi pubblici sono elargiti dallo Stato, pertanto nel rispetto dei principi di economicità, efficacia, imparzialità, pubblicità e trasparenza (*Legge 7 agosto 1990, n. 241*), in questo caso il fine ultimo non è il profitto bensì il benessere

collettivo. Tra i servizi pubblici, quello d'interesse ai fini dello studio è il servizio pubblico sanitario italiano.

Il Servizio Sanitario Nazionale (SSN) fornisce assistenza sanitaria a tutti i cittadini senza distinzioni di genere, residenza, età, reddito e lavoro. Esso si basa su principi fondamentali come la responsabilità pubblica della tutela della salute, universalità ed equità di accesso ai servizi, globalità di copertura e finanziamento pubblico attraverso la fiscalità generale (*Legge 23 dicembre 1978, n.833*). Il governo del sistema sanitario è esercitato da Stato e Regioni; al primo spetta la determinazione dei livelli essenziali delle prestazioni concernenti i diritti civili e sociali (i cosiddetti LEA, ovvero i Livelli Essenziali di Assistenza), mentre la tutela della salute rientra nella competenza concorrente affidata alle Regioni.

Caratteristiche dei servizi

Tradizionalmente la qualità dei prodotti è stata indagata più in profondità rispetto alla qualità nei servizi, questo poiché il sistema economico è fortemente incentrato sulla manifattura e l'industria (Franceschini, 2001).

La qualità dei prodotti è misurata come conformità alle specifiche e numero dei difetti in un campione. In particolare le principali metodologie utilizzate sono: il controllo di accettazione (in entrata, in progress, in uscita), il controllo statistico di processo e le carte di controllo (Franceschini and Galetto, 2011).

I servizi differiscono ampiamente dai prodotti per la loro natura intrinseca. Sono individuate cinque dimensioni (Mastrogiacomo, Barravecchia and Franceschini, 2016): innanzitutto, il servizio è di per sé intangibile e ciò rende difficile identificare le dimensioni di qualità proprie del mondo manifatturiero; in secondo luogo, il servizio è limitatamente standardizzabile date le sue caratteristiche, è unico e non è possibile l'esatta riproduzione dello stesso anche qualora richiesta dallo stesso cliente; il servizio, inoltre, non è immagazzinabile per un futuro riutilizzo a causa dell'intangibilità e del consumo immediato al momento della consegna; il cliente è tipicamente coinvolto durante la fruizione del servizio, diversamente dal mondo

manifatturiero in cui il coinvolgimento è minore; in ultima analisi, il fornitore del servizio e il cliente sono inseparabili durante l'erogazione.

Modelli di studio della qualità nei servizi

In un articolo di Parasuraman, Zeithaml e Berry (Parasuraman, Zeithaml and Berry, 1985) è esposto per la prima volta il modello elaborato per la descrizione della qualità nei servizi. Nel suddetto sono state analizzate quattro categorie di servizi: retail banking, carte di credito, intermediazione in titoli, servizi di manutenzione e riparazione di prodotti. Sono state condotte interviste dirette al top management di quattro imprese americane leader nei settori sopracitati e sono stati guidati dodici focus groups. Il risultato di queste analisi è che esistono differenze tra la qualità del servizio percepita dai manager delle società e la qualità percepita dai consumatori. Sulla base di questo assunto è stato elaborato il modello dei cinque gap, di seguito descritti.

GAP 1 (Aspettative del consumatore – qualità ipotizzata dal top management): i manager sovente non riescono a cogliere quali caratteristiche connotano alta qualità dal punto di vista del cliente finale.

GAP 2 (Qualità ipotizzata dal top management – qualità progettata): il più delle volte il top management conosce i requisiti richiesti dai clienti, ma esistono vincoli progettuali e di mercato che non consentono di trasporre ciò che i manager chiedono in requisiti tecnici.

GAP 3 (Qualità progettata – qualità realizzata): nella sfera dei servizi è difficile tradurre in realtà alcuni requisiti, poiché la variabilità è molto alta rispetto al mondo standardizzato dei prodotti.

GAP 4 (Qualità realizzata – Qualità comunicata ai consumatori): la pubblicità e le strategie di marketing possono modificare le aspettative dei clienti. Promettere più di quello che è consegnato, comporta un innalzamento delle aspettative iniziali ma diminuisce la percezione della qualità laddove le promesse non sono mantenute.

GAP 5 (Qualità attesa – qualità percepita): la qualità che un utente percepisce è funzione dell'intensità e della differenza tra qualità attesa del servizio e quella percepita, che è a sua volta funzione dei gap precedenti.

$$GAP\ 5 = f(GAP1, GAP2, GAP3, GAP4) \quad (1)$$

In Figura 1 è presente uno schema riassuntivo della teoria dei cinque gap.

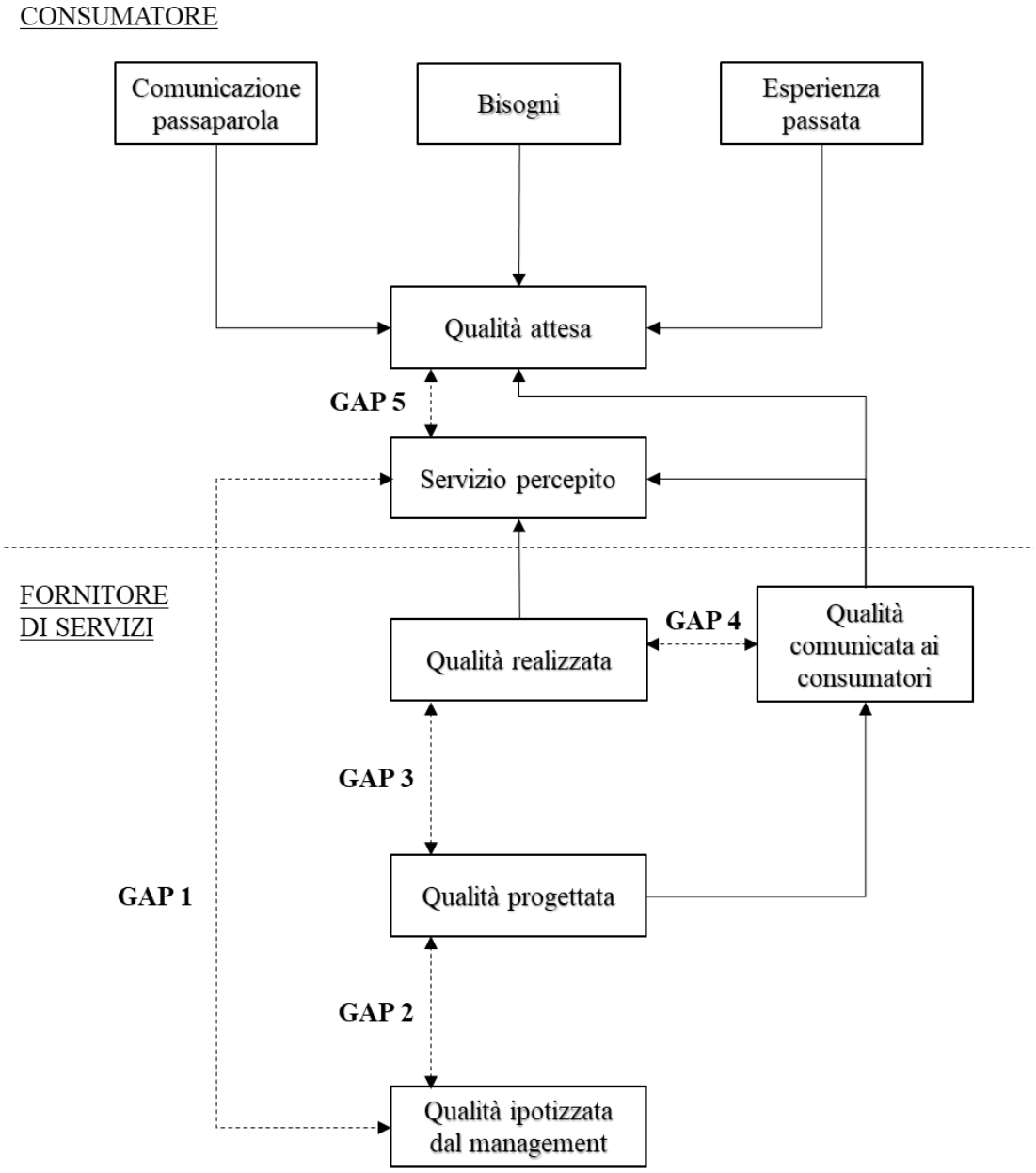


***Figura 1**. Schema del modello PZB (Parasuraman, Zeithaml and Berry, 1985)*

I focus groups dello studio di Parasuraman, Zeithaml e Berry hanno mostrato che, indipendentemente dal tipo di servizio offerto, i consumatori utilizzano criteri simili per la valutazione della qualità. Le dieci dimensioni che determinano la qualità nei servizi sono elencate nella in appendice A.

Il limite del modello descritto è che non fornisce gli strumenti operativi utili a misurare la qualità di un servizio. Per ovviare a ciò Parasuraman, Zeithaml e Berry, hanno sviluppato il SERVQUAL (Franceschini, 2001), ovvero un questionario per valutare le percezioni del cliente sulla qualità di uno specifico servizio (Parasuraman, Zeithaml and Berry, 1988).

In assenza di misure oggettive, un approccio appropriato per valutare la qualità del servizio consiste nel misurare la percezione della qualità da parte del consumatore. I passi che hanno portato alla creazione della scala sono di seguito descritti. Dopo l'enumerazione delle dieci dimensioni determinanti della qualità nei servizi, sono state generate novantasette voci rappresentative delle dieci dimensioni di qualità. Successivamente sono state raccolte le aspettative e le percezioni da un campione di duecento soggetti intervistati circa le tipologie di servizi precedentemente citati. Si è poi proceduto con la purificazione della scala attraverso l'eliminazione iterativa di alcune voci, con l'obiettivo di aumentare l'indicatore statistico α di Cronbach. A questo punto la scala consta di trentaquattro voci rappresentanti sette dimensioni di qualità. Per depurare ulteriormente la scala, sono state raccolte le aspettative e le percezioni da quattro campioni indipendenti di duecento individui intervistati. Infine, si è giunti alla definizione della scala SERVQUAL, costituita da ventidue voci rappresentative delle seguenti cinque nuove dimensioni (aspetti tangibili, affidabilità, tempo di risposta, fiducia, empatia). Le ultime due dimensioni contengono al loro interno sette dimensioni originali (comunicazione, credibilità, sicurezza, competenza, cortesia, capire il cliente, accesso).

Al fine di adattare il modello SERVQUAL al lavoro di tesi, saranno utilizzate come benchmark le dimensioni di qualità percepite dal paziente nel servizio dell'assistenza sanitaria presenti in appendice B, in quanto sono dimensioni specifiche del settore sanitario (Duggirala, Rajendran and Anantharaman, 2008). Per identificare le

dimensioni di qualità, è stato sviluppato un questionario formato da ottantasei item che indagano le sette dimensioni di qualità. I destinatari del questionario sono pazienti che hanno beneficiato di cure mediche nell'ultimo anno.

La trasformazione digitale nella sanità

Nei paesi industrializzati la spesa sanitaria è diventata nel corso degli anni ingente, pertanto pur garantendo la qualità del servizio offerto al paziente, la riduzione dei costi si configura come l'obiettivo principale da seguire. A tal fine, la strada intrapresa dai servizi sanitari è quella della digitalizzazione di molti processi. Le tendenze digitali ad oggi più adottate, in grado di apportare reali modifiche al settore sono la cartella clinica elettronica, la comunicazione digitale, la telemedicina, il data analytics, la cybersecurity ed il cloud (Watson, 2015).

La cartella clinica elettronica è un documento digitale creato dall'ospedale per il paziente con il fine di gestire in modo organizzato e continuativo il percorso di cura. Questo documento fornisce informazioni che possono aiutare i medici durante il percorso diagnostico-terapeutico e ridurre i tempi necessari per lo scambio di informazioni tra i professionisti sanitari (Hillestad *et al.*, 2005).

La comunicazione digitale per mezzo degli smartphone e delle app di messaggistica ha certamente velocizzato il flusso di informazioni all'interno degli ospedali, evitando al personale medico e infermieristico dispendio infruttuoso del proprio tempo, anche in situazioni di emergenza, per il reperimento di informazioni necessarie alla presa in carico del paziente.

Con il termine telemedicina si fa riferimento ad ogni attività medica effettuata a distanza. Questo concetto comprende sia la consulenza da remoto sia la trasmissione di dati a distanza per mezzo di dispositivi medicali avanzati e della rete 5G (Anwar and Prasad, 2018). Secondo l'Osservatorio Netics, i risparmi del SSN con la telemedicina ammonterebbero a 6,9 miliardi di euro all'anno.

Per organizzare e analizzare i dati di milioni di cartelle cliniche elettroniche non sono sufficienti le tecniche classiche di gestione dei dati, ma si rivelano necessarie moderne

metodologie di data analytics per gestire big data, al fine di migliorare l'efficienza del servizio sanitario e abbattere i costi (Raghupathi and Raghupathi, 2014). L'utilizzo di big data e analytics deve essere supportato da un'infrastruttura solida di cybersecurity per evitare la violazione e diffusione di informazioni sensibili (Kruse *et al.*, 2017).

Qualità nella sanità

In ambito sanitario, la qualità è definita come il grado con il quale i sistemi sanitari aumentano la probabilità *p* di raggiungere un risultato di salute atteso coerente con le conoscenze mediche correnti (Lohr and Schroeder, 1990). In questa definizione il termine "sistemi sanitari" comprende un'ampia gamma di servizi relativi alle malattie del corpo e della mente, il cui scopo è garantire la prevenzione e la promozione del benessere, le cure riabilitative e quelle palliative. Inoltre, la definizione si applica a tutti gli operatori e le strutture sanitarie. L'espressione "risultato di salute atteso", si riferisce ai risultati che i pazienti desiderano e pone in evidenza la correlazione esistente tra le cure fornite e i relativi effetti sul paziente. Emerge pertanto la necessità che i pazienti e le famiglie ricevano informazioni adeguate e sufficienti circa i possibili percorsi terapeutici da intraprendere, le eventuali cure alternative e i relativi esiti attesi, in modo tale da scegliere in autonomia e consapevolmente il trattamento sanitario a cui sottoporsi.

Inoltre, all'interno della definizione di "risultato di salute atteso" è presente il concetto di probabilità; questo implica che la qualità non coincide obbligatoriamente con un esito positivo. Risultati negativi, infatti, possono verificarsi anche se la qualità del servizio nella fattispecie sia molto alta. Al contrario, è possibile che i pazienti rispondano bene alle terapie nonostante la bassa qualità delle cure fornite. Pertanto, è importante che la qualità nella sanità tenga conto sia del processo di cura che del risultato.

L'espressione "coerente con le conoscenze mediche correnti" mette in evidenza che la qualità dei servizi sanitari è in funzione del sapere in un dato momento e che gli operatori professionali hanno il dovere di utilizzare la conoscenza al meglio. Sebbene la conoscenza e le best practices siano importanti per la qualità delle cure, nessun

professionista può garantire in autonomia un'ottima qualità delle cure fornite. Il buon funzionamento e l'efficienza di tutti gli attori della catena del valore della sanità permettono di prevenire e minimizzare gli errori, oltre che a consentire la disponibilità delle informazioni (Chassin, Galvin and Donaldson, 1998).

La letteratura propone tre pilastri su cui è fondata la teoria della qualità dei sistemi sanitari (Strome, 2013): la struttura, il processo e i risultati, di seguito saranno descritti.

La struttura della sanità comprende gli aspetti stabili: la fornitura dell'assistenza sanitaria, gli strumenti, le tecnologie e in generale le risorse sanitarie come l'organizzazione generale e le politiche sanitarie. Ad esempio, alcuni indicatori che descrivono la struttura di un ospedale sono:

- Numero di posti letto presenti in terapia intensiva.
- Copertura medica specialistica in terapia intensiva.
- Numero di scanner per tomografia computerizzata e la loro diponibilità.

Le suddette informazioni sono spesso utilizzate per paragonare le prestazioni erogate da due o più strutture sanitarie. Ad esempio, due reparti di emergenza possono assistere, durante una giornata lavorativa, un numero simile di pazienti con periodi di permanenza analoghi, ma è importante tenere in considerazione se un ospedale opera con meno letti o più medici.

I processi sono le varie attività svolte dai fornitori di assistenza sanitaria e l'interazione tra essi, i pazienti e le loro famiglie nel corso del trattamento sanitario. Il management delle strutture sanitarie spesso agisce sui processi per migliorare la qualità del servizio. Questo perché i processi sono relativamente facili da misurare: solitamente si calcolano gli intervalli di tempo delle varie attività e queste misurazioni costituiscono i dati prestazionali di base rispetto ai quali si cerca di applicare i miglioramenti individuati. La tecnologia rende il reperimento dei dati e l'analisi più facilmente eseguibili; ad esempio le cartelle cliniche elettroniche documentano in maniera impeccabile la storia clinica del paziente e costituiscono una fonte di dati molto preziosa per quanto riguarda l'analisi.

Il risultato è un punto di arrivo quantificabile e individuabile focalizzato sul paziente. Alcuni esempi di indicatori sono: la morbosità, la mortalità e la durata della degenza. Il risultato è spesso in funzione delle cure ricevute, è il punto di arrivo della fase del processo e spesso è ciò che più interessa al paziente. Secondo Porter i risultati devono (Porter, 2010):

- Includere le circostanze di salute più rilevanti per i pazienti.
- Coprire sia il breve che il lungo termine.
- Considerare i fattori di rischio e le condizioni iniziale per consentire operazioni di mitigazione del rischio.

Quando si sviluppano analisi per la misurazione della qualità, si deve assicurare che siano presi in considerazione tutti i risultati.

È utile proporre un esempio per comprendere la struttura, il processo e i risultati. Si consideri un'unità chirurgica ospedaliera. Gli elementi della struttura sono il numero di chirurghi, infermieri, sale operatorie e la tecnologia disponibile. Un indicatore che misuri un processo è la percentuale di operazioni iniziate puntualmente. Infine, esempi di misure di risultato sono il tasso di eventi critici e il tasso di mortalità per intervento chirurgico.

Quando si analizzano i dati relativi a uno stesso periodo di tempo, c'è il rischio di incorrere in una lettura inaccurata degli indicatori; alcuni mostrano delle risposte solo dopo un cambiamento sistemico. Ad esempio, si supponga che un miglioramento nel processo di dimissione dei pazienti dall'ospedale abbia come risultato quello di abbassare il tempo di degenza dei pazienti; gli indicatori di processo mostrano un tempo di degenza dei pazienti minore rispetto al periodo precedente all'implementazione, ma questa diminuzione potrebbe anche essere causata dal fatto che alcuni pazienti siano stati dimessi troppo presto in base alle cure che avrebbero dovuto ricevere, producendo un peggioramento della qualità percepita.

Molti istituti sanitari cercano di migliorare la qualità del servizio sia perché è un obiettivo da perseguire e sia per necessità competitive, in quanto la qualità è un obiettivo sempre in movimento. Ci sono molte sfide nel contesto in cui operano come ad esempio:

- Sfide esterne (cambi di regolazione e problemi di finanziamento pubblico).
- Sfide interne (gestione delle risorse umane).
- Cambiamento dei bisogni dei pazienti dovuto all'invecchiamento e all'aumento della popolazione.
- Adozione di nuove tecnologie che possono essere determinanti nel miglioramento della qualità ma che possono essere costose da un punto di vista finanziario.

Gli ospedali hanno la necessità di adattarsi a un ambito in continua mutazione, innovando e migliorando i propri processi. Si definisce l'espressione "Healthcare Quality Improvement" (Healtcare QI) la modifica del comportamento e dell'organizzazione del fornitore di servizi sanitari adottando strategie di cambiamento sistematico e migliorando l'esperienza del paziente e i risultati raggiunti (*Going lean in the NHS*, 2007). Healthcare QI può essere definito in modo alternativo come una serie di attività data-driven progettate per portare un miglioramento immediato della fornitura di assistenza sanitaria (Barton, 2009). Ciò che è più importante nell'healthcare QI è il focus sul paziente, che rappresenta il cliente primario del sistema sanitario. I pazienti non devono essere visti come destinatari passivi delle cure ma come individui che pretendono di ricevere cure e servizi di alta qualità.

Gli user-generated contents

Il punto iniziale dell'analisi effettuata è l'estrazione di valore da parte degli "User-Generated Contents" (UGC) o contenuti generati dagli utenti. Gli UGC sono definiti come qualsiasi tipo di contenuto (post, recensione, audio, video, immagine) creato dagli utenti e pubblicato su internet, spesso disponibile su piattaforme di social network (Shirky, 2009). Con l'avvento di Internet e del mobile Internet, la tecnologia ha permesso la facile creazione e generazione dei suddetti. Questa crescita di contenuti online ha avuto un forte impatto sulle organizzazioni pubbliche e private e la gestione della qualità dei prodotti e servizi. Al fine di chiarire cosa è UGC e cosa non è sono proposte tre caratteristiche centrali (Participative web and user-created content: Web 2.0, wikis and social networking, 2007).

- Contenuto pubblico: teoricamente un UGC potrebbe essere creato da un utente e mai pubblicato online. Lo studio si focalizza su tutto ciò che è pubblico o è accessibile pubblicamente su una pagina web o su un social network. Tutto ciò esclude e-mail e messaggi privati.
- Sforzo creativo: ciò implica che la creatività sia utilizzata per produrre un contenuto. Ad esempio, attività come caricare online fotografie, esprimere opinioni su un blog, postare video sono considerate UGC.
- Creazione dei contenuti al di fuori delle routine professionali: il creatore non pubblica online con l'aspettativa di profitto o remunerazione ma è spinto dalla necessità di comunicare la propria visione in merito ad un tema.

In particolare, tra gli UGC che interessano ai fini dello studio, ci sono le recensioni online presenti su piattaforme di e-commerce, piattaforme aggregatrici e blog di discussione. Le recensioni su aggregatori come Yelp o Trip Advisor hanno avuto un forte impatto sul settore dei viaggi e dell'ospitalità e sono uno degli elementi più importanti del processo decisionale che porta alla fruizione del servizio da parte dei consumatori (O'Connor, 2008).

Alcuni studi hanno dimostrato come sia molto importante la piattaforma sulla quale sono presenti le recensioni, poiché alcune sono più affidabili di altre, come ad esempio Trip Advisor, la quale adotta sistemi di filtraggio nel tentativo di evitare recensioni inadeguate (Chua and Banerjee, 2013).

È importante sottolineare che le recensioni non sono tutte uguali e alcune hanno più credibilità di altre. Nello specifico, le recensioni negative risultano più credibili di quelle positive, mentre quelle positive ispirano maggiore fiducia. Tutto ciò è vero laddove l'identità del recensore sia rivelata, mentre nel caso in cui questa rimanga nascosta non c'è differenza significativa tra recensioni positive o negative in termini di credibilità o fiducia percepita dal consumatore (Kusumasondjaja, Shanka and Marchegiani, 2012).

Le organizzazioni che sfruttano i dati generati online dai propri clienti possono non solo migliorare la misurazione della qualità del servizio, ma anche avere una migliore comprensione degli aspetti che influenzano la soddisfazione del cliente, che non è

basata esclusivamente sulla valutazione aggregata della recensione (Palese and Piccoli, 2016). La forma delle recensioni riveste un ruolo importante e la chiarezza di esposizione ha un effetto positivo sulla credibilità percepita dal consumatore, diversamente dalla lunghezza del testo (Korfiatis, García-Bariocanal and Sánchez-Alonso, 2012).

Numerosi studi sono focalizzati sul sentiment analysis a partire dagli UGC e in particolare dalle recensioni, per comprendere le percezioni degli utenti. Il sentiment analysis studia le opinioni espresse dagli utenti che implicano giudizi positivi o negativi (Liu, 2012).

Lo studio di Kasper e Vela (2011) mostra un'analisi sulle recensioni e i commenti presenti sul web riguardanti gli hotel, al fine di classificarli e facilitare l'accesso alle informazioni per la gestione del controllo qualità nel settore del turismo.

Un'analisi sui tweet relativi alla qualità degli ospedali (Greaves *et al.*, 2014) ha rivelato che l'11% dei tweet a proposito degli ospedali contengono informazioni sulla qualità delle cure ed il topic più frequente è l'esperienza del paziente (8%). Altri topic menzionati sono: messaggi di supporto ai pazienti, attività di raccolta fondi, promozione personale e comunicazione di cultura medica. Il 77% dei tweet hanno un'accezione positiva. Un altro studio effettuato su un'analisi dei commenti dei pazienti online (Greaves *et al.*, 2013), mostra che c'è una significativa associazione tra le valutazioni online sulle strutture ospedaliere presenti su siti web specializzati e i tradizionali questionari per la valutazione della soddisfazione dei pazienti. Questo indica che c'è una grande quantità di informazione presente online pronta per essere analizzata a scopo di monitoraggio e miglioramento dell'offerta dei servizi ospedalieri, purché sia affiancata dai classici metodi di indagine sulla qualità.

Latent Dirichlet Allocation (LDA)

Lo strumento utilizzato per analizzare gli UGC è chiamato Topic Modeling (Text Mining 101: Topic Modeling). Nell'era dei big data sempre più informazioni sono disponibili, pertanto è molto difficile estrarre valore da questi dati. Sono necessari

strumenti e tecniche per la ricerca e la comprensione di una vasta quantità di informazioni. Il Topic Modeling fornisce un metodo per organizzare, capire e sintetizzare una grossa mole di informazioni testuali. Questo metodo risulta utile per individuare pattern nascosti presenti all'interno della collezione di dati e classificare i documenti con dei topic. Il Topic Modeling si può descrivere come un metodo per trovare gruppi di parole (topic) da una collezione di documenti che meglio rappresentano le informazioni del corpus. Questa metodologia è definita anche "Text Mining". Esistono molteplici tecniche e algoritmi di Text Mining, quello utilizzato per questo studio è il Latent Dirichlet Allocation (LDA).

L'algoritmo LDA (Blei, Ng and Jordan, 2003) è uno dei metodi più utilizzati per l'estrazione di informazione dai documenti testuali. LDA è un modello probabilistico per collezionare dati discreti ed è un modello gerarchico bayesiano a tre livelli in cui ogni elemento di una raccolta di dati è modellato come una miscela di set di topic. Ogni topic è, a sua volta, modellato come una miscela di probabilità associate a ogni topic. Nel contesto dei dati testuali, le probabilità dei topic forniscono un'esplicita rappresentazione del documento.

Al fine di formalizzare la descrizione dell'algoritmo sono definiti i seguenti elementi:

- La *parola* è l'unità di base del modello, ed è definita come un elemento in un vocabolario indicizzato $\{1,\ldots,V\}$. I termini sono rappresentati da vettori che hanno una sola componente pari a 1 e tutte le altre componenti pari a 0. In particolare, la v-esima parola del vocabolario è rappresentata dal vettore V, nel quale $w^v=1$ e $w^u=0$ per $u\neq v$.
- Un *documento* è una sequenza di N parole indicate come $\mathbf{w}=(w_1,w_2,\ldots w_N)$ dove w_N è la n-esima parola della sequenza.
- Un *corpus* è una collezione di M documenti indicati come $D=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_M\}$, nel caso in esame $\mathbf{w}_M$ è la M-esima recensione.
- Un *topic* è una distribuzione di parole ed è indicato come z, in totale i topic sono k.

Di seguito è presente la Tabella 1 con la spiegazione delle variabili utilizzate dall'algoritmo LDA.

Variabile	**Spiegazione**
w_i	Parola i-esima nel vocabolario del corpus.
$\mathbf{w}=(w_1,w_2,\ldots,w_N)$	Documento **w**, il quale contiene N parole.
$D=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_M\}$	Corpus *D*, il quale contiene M documenti (recensioni).
z_i	Topic i-esimo

Tabella 1 *Elenco dei parametri dell'algoritmo LDA e relativa spiegazione*

I parametri dell'algoritmo sono (Intuitive Guide to Latent Dirichlet Allocation - Towards Data Science):

- α: parametro relativo alla distribuzione di Dirichlet che regola la miscela dei topic all'interno dei documenti.
- θ: matrice causale in cui θ(i,j) rappresenta la probabilità che il i-esimo documento (recensione) contenga il j-esimo topic.
- β: matrice casuale in cui β(i,j) rappresenta la probabilità che il topic i-esimo contenga la parola j-esima.

L'idea di base dell'algoritmo è che i documenti (recensioni) sono rappresentati come una miscela casuale di topic, dove ogni topic è caratterizzato da una distribuzione probabilistica di parole.

Di seguito, è illustrato l'algoritmo LDA per ogni documento ***w*** in un corpus *D*:

1. Scelta $N\sim$Poisson(ξ).
2. Scelta $\theta\sim$Dir(α).
3. Per ognuna delle *N* parole w_n:
    a. Scelta del topic $z_n\sim$Multinomiale(θ)

b. Scelta della parola w_n da $p(w_n|z_n,\beta)$, una probabilità multinomiale condizionata al topic z_n.

Una variabile casuale di Dirichlet θ ha la seguente densità di probabilità:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k}\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\theta_1^{\alpha_1-1}\dots\theta_k^{\alpha_k-1}, \quad (2)$$

dove il parametro α è un vettore di dimensione k con componenti $\alpha i>0$ e Γ(x) è la funzione Gamma. Dati i parametri α e β, la distribuzione congiunta della miscela di topic (Topic Mixtures) θ, un set di N topic $\boldsymbol{z}$ e n set di N parole $\boldsymbol{w}$ dati:

$$p(\theta, \boldsymbol{z}, \boldsymbol{w}|\alpha,\beta) = \mathrm{p}(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n,\beta), \quad (3)$$

dove $p(z_n|\theta)$ è θi dato che $z_n^i = 1$. Integrando su θ e sommando su z, si ottiene la distribuzione marginale di un documento (recensione):

$$p(\boldsymbol{w}|\alpha,\beta) = \int p(\theta|\alpha)(\prod_{n=1}^{N}\sum_{z_n} p(z_n|\theta)p(w_n|z_n,\beta))\,d\theta. \quad (4)$$

Infine, eseguendo il prodotto delle probabilità marginali di ogni singolo documento (recensione), si ottiene la probabilità del corpus:

$$p(D|\alpha,\beta) = \prod_{d=1}^{M}\int p(\theta_d|\alpha)(\prod_{n=1}^{N_d}\sum_{z_{dn}} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn},\beta))\,d\theta_d. \quad (5)$$

La Figura 2 illustra graficamente l'algoritmo LDA. Nello specifico, sono individuabili tre livelli: i parametri α e β sono campionati una sola volta poiché sono al livello del corpus totale, la variabile θ_d è al livello del singolo documento (recensione) ed è campionata una volta per documento, infine le variabili z_{dn} e w_{dn} sono al livello della singola parola e sono campionate una volta per ogni parola in ogni documento.

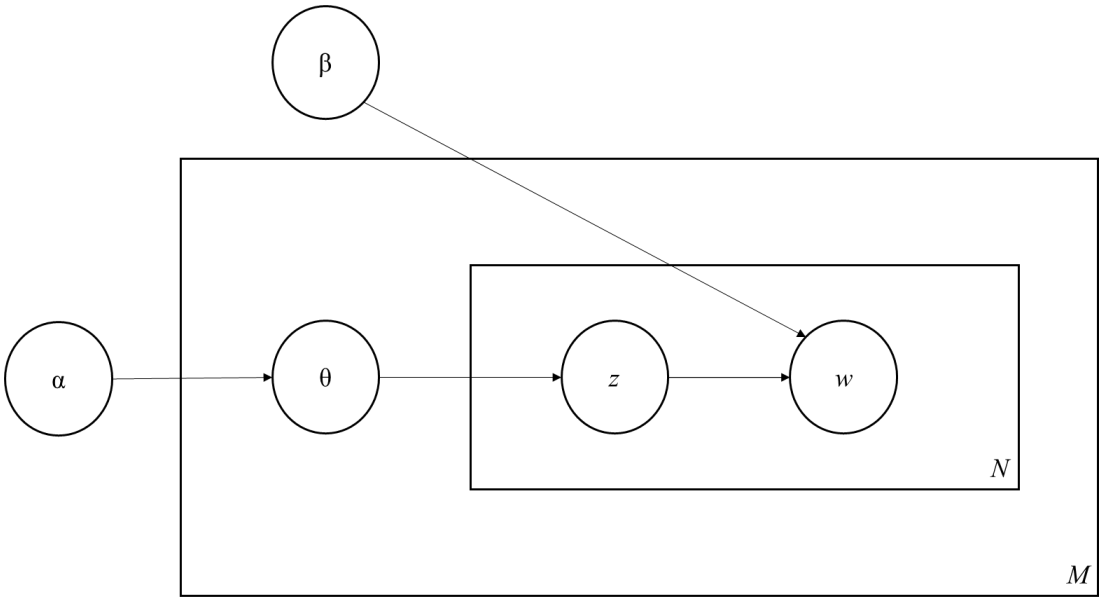


Figura 2 *Rappresentazione grafica dell'algoritmo LDA* (Blei, Ng and Jordan, 2003)

Resta ora da comprendere la modalità di valutazione delle performance di un algoritmo LDA. Nelle applicazioni dell'algoritmo, la funzione di verosimiglianza (indicata di seguito) ha un punto di massimo locale in cui per ogni documento la miscela dei topic è la stessa, in altre parole la prevalenza di ogni topic all'interno di ogni documento è costante.

$$l(\alpha|\beta) = \sum_{d=1}^{M} \log p(w_d|\alpha, \beta) \qquad (6)$$

L'obiettivo è evitare questa situazione, poiché è indice di scarsa capacità di classificazione dell'algoritmo e inerzia nell'assegnazione dei topic ai documenti. Prima di applicare l'algoritmo è importante inizializzare correttamente i parametri α, β e k. Per comprendere se i parametri sono inizializzati correttamente è utilizzato un indicatore chiamato *perplexity*. Questo è una funzione monotona decrescente nella funzione di verosimiglianza e coincide con l'inverso della media geometrica della

verosimiglianza per parola. Una minore perplexity indica una migliore performance dell'algoritmo. La funzione è definita di seguito:

$$perplexity(D_{test}) = exp\left\{-\frac{\sum_{d=1}^{M} log\ p(\boldsymbol{w}_d)}{\sum_{d=1}^{M} N_d}\right\}. \quad (7)$$

Per calcolare questo indicatore di bontà di adattamento, solitamente il corpus è diviso in due parti. Una parte è utilizzata per l'esecuzione dell'algoritmo, mentre un'altra con meno documenti è utilizzata per il calcolo della perplexity, quindi per allenare il modello. L'allenamento è una fase in cui discretamente si cerca di far variare i parametri per minimizzare la perplexity. Anche variando il parametro β, la perplexity non cambia molto, quindi è trascurabile. Importante è la scelta di α e del numero totale dei topic k.

Un altro metodo più rigoroso e corretto per risolvere questo problema è l'inferenza bayesiana (Blei, Kucukelbir and McAuliffe, 2016). La funzione bayesiana considerata è la seguente:

$$q(\theta, z|\gamma, \phi) = q(\theta|\gamma)\prod_{n=1}^{N} q(z_n|\phi_n), \quad (8)$$

dove il parametro di Dirichlet γ e i parametri multinomiali ($\phi_1,\ldots, \phi_n$) sono i parametri liberi dell'inferenza bayesiana. Dopo aver illustrato la famiglia della distribuzione di probabilità, il prossimo passo è costruire un problema di ottimizzazione che determini i valori dei parametri γ e ϕ, ciò avviene con il seguente problema di ottimizzazione:

$$(\gamma^*, \phi^*) = \arg\min D(q(\theta, \boldsymbol{z}|\gamma, \phi) \parallel p(\theta, \boldsymbol{z}|\boldsymbol{w}, \alpha, \beta)). \quad (9)$$

L'ottimizzazione dei valori di γ e ϕ avviene minimizzando la divergenza di Kullback-Leibler (KL) tra la distribuzione variazionale $q(\theta, \boldsymbol{z}|\gamma, \phi)$ e $p(\theta, \boldsymbol{z}|\boldsymbol{w}, \alpha, \beta)$, in particolare ponendo le derivate della divergenza di Kullback-Leibler uguali a zero. Per ogni i-esimo documento ($\boldsymbol{w_i}$) si ottiene così una stima di (γ^*,ϕ_i^*). Successivamente,

dato un corpus di documenti $D=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_M\}$, l'obiettivo è identificare i parametri α e β che massimizzano la funzione di massima verosimiglianza:

$$l(\alpha,\beta)=\sum_{d=1}^{M}\log p(\boldsymbol{w_d}|\alpha,\beta). \tag{10}$$

Come è stato precedentemente precisato, questa funzione (10) ha dei massimi locali, i quali non sono ideali per l'ottimizzazione dei parametri dell'algoritmo. L'inferenza variazionale fornisce un limite inferiore per la sopracitata funzione, il quale può essere massimizzato nel rispetto di α e β. È quindi possibile approssimare i parametri α e β per mezzo di una procedura chiamata *variational expectation maximization (EM)*, la quale alternativamente massimizza il limite inferiore della funzione in funzione di γ e ϕ, e durante l'iterazione successiva esegue la medesima azione tenendo fissi però i parametri γ e ϕ e massimizzando il limite inferiore in funzione di α e β. L'algoritmo EM è illustrato di seguito:

1. Expectation step. Per ogni documento d, si trovano i valori ottimali dei parametri variazionali $\{\gamma_d^*,\phi_d^*:d\in D\}$ come è eseguito nell'equazione (9).
2. Maximization step. Massimizzare il limite inferiore della massima verosimiglianza in funzione dei parametri α e β. Ciò è effettuato grazie all'output identificato nello step 1.

I due step dell'algoritmo EM sono ripetuti fino a che la funzione di massima verosimiglianza converge al limite inferiore. A quel punto i parametri α e β sono quelli ottimi per l'applicazione dell'algoritmo LDA.

Dopo questo excursus teorico è utile spiegare come funziona concretamente l'algoritmo LDA con un esempio. Si supponga di avere solamente due topic e la distribuzione delle parole principali, come descritto in Tabella 2.

TOPIC 1	**TOPIC 2**
mangiare	ufficio
pausa	lavoro
pasta	pausa
pranzo	scrivania
piatto	computer
…	…

Tabella 2 *Esempio di una lista di parole appartenenti a due topic*

Si ipotizzi di avere un corpus costituito dai quattro documenti di seguito riportati:

1. Io mangio la pasta.
2. Io lavoro in un ufficio.
3. Mangio in ufficio.
4. L'aereo decollerà domani.

L'algoritmo LDA analizza le frasi e associa ciascuna parola a un determinato topic. L'algoritmo consente di eliminare le cosiddette stop-words, ovvero articoli, proposizioni, congiunzioni e aggettivi. Dopo questa operazione, le frasi assumono la forma di seguito riportata:

1. Mangio pasta.
2. Lavoro ufficio.
3. Mangio ufficio.
4. Aereo decollerà domani.

Il primo documento contiene i termini "mangio" e "pasta", i quali appartengono al 100% al topic 1. Nel secondo documento le parole "lavoro" e "ufficio" appartengono al 100% al topic 2. La terza frase è particolare poiché le parole "mangio" e "ufficio" appartengono a due topic diversi, è quindi logico affermare che il documento appartiene per il 50% al topic 1 e per il 50% al topic 2. La quarta frase non ha attinenza

con i due topic, pertanto le parole “aereo”, “decollerà” e “domani” non appartengono a nessun topic ed è dunque ragionevole assegnare il documento per il 50% al topic 1 e il 50% al topic 2 seguendo la logica fuzzy, poiché non ci sono abbastanza informazioni per assegnare una percentuale di prevalenza a un topic o all’altro.

Capitolo II

Applicazione di LDA per valutare le dimensioni di qualità

Per effettuare l'analisi si è seguito un processo a cinque stadi (Zaki and McColl-Kennedy, 2019):

1. Fase di estrazione del corpus di recensioni e preelaborazione del campione.
2. Applicazione dell'algoritmo LDA sul campione, con il fine di individuare il numero ottimale di topic e il valore del parametro α che minimizzi la perplexity (l'indicatore di goodness of fit).
3. Estrazione dei topic e assegnazione di questi ultimi alle recensioni.
4. Validazione dell'algoritmo.
5. Attività di labelling, ovvero l'etichettatura di un titolo ad ogni topic.

È possibile osservare il codice del programma in Appendice E.

Estrazione del corpus di recensioni e preelaborazione del campione

Le recensioni sono state scaricate da Google con l'ausilio di un tool per l'estrazione di testo sul web, chiamato Data Toolbox. Inizialmente il dataset era composto da N=25224 documenti, ma successivamente, poiché lo scopo del lavoro è realizzare un'analisi delle recensioni in lingua italiana, sono state eliminate tutte quelle scritte in altra lingua o tradotte automaticamente da Google, in quanto al loro interno contenevano parole non riconducibili alla lingua italiana. Il dataset finale, pertanto, è composto da N=24060 recensioni. Le informazioni acquisite per ogni recensione sono state: testo, ospedale, città, provincia e regione.

La fase di preelaborazione del campione si è rivelata di esecuzione maggiormente complessa. Il fatto che l'analisi sia condotta su recensioni in lingua italiana, e pertanto non sia possibile garantire l'adattamento di numerose funzioni disponibili all'interno del Text Analytics Toolbox di Matlab alla lingua italiana, ha rappresentato la criticità

più rilevante di questa fase e ha reso necessario agire manualmente per adattare il testo all'algoritmo LDA.

In prima istanza, il documento contenente le recensioni è stato "tokenizzato", ovvero è stata applicata una trasformazione tale da permettere la scomposizione delle recensioni in una collezione di parole chiamate "tokens". Questa operazione è stata necessaria poiché i tokens si configurano come gli elementi base che rendono possibile l'analisi linguistica delle recensioni.

In seguito, alcune parole sono state rimosse manualmente dal corpo del testo, ovvero quelle che non apportavano alcun valore aggiunto all'analisi, come ad esempio pronomi, avverbi, nomi e numeri (Appendice F). In questa fase è stato necessario eseguire molte iterazioni poiché la lingua italiana ha un vasto vocabolario.

Purtroppo, non è stato possibile eseguire alcune funzioni per raffinare il testo su cui applicare successivamente l'agoritmo LDA, poiché sono disponibili solamente per la lingua inglese. Le funzioni non applicate sono *addPartOfSpeechDetails(.)* e *NormalizeWords(.).* La prima permette di separare le parole collegate da un apostrofo, ad esempio "you're" è trasformato in "you" e "are" e permette di aggiungere informazioni alle parole, "you" è identificato come pronome e "are" come verbo. La seconda funzione riduce le parole alla loro forma originale, ad esempio la parola "going" è ridotta a "go. Queste due funzioni aiutano molto l'algoritmo LDA, rendendo possibile l'analisi su un vocabolario ridotto migliorando conseguentemente gli indicatori di goodness of fit poiché comportano la presenza di meno parole raramente utilizzate nelle recensioni. Senza l'utilizzo di queste due funzioni, ad esempio le parole "ospedale" e "ospedali" sono conteggiate separamente e inserite all'interno del vocabolario come due parole distinte, ciò peggiora fortemente l'output dell'algoritmo.

Successivamente sono state eliminate la punteggiatura, i termini con meno di due caratteri e con più di quindici, con l'obiettivo di ridurre il vocabolario e rimuovere le parole non significative ai fini dell'analisi.

Infine, nono sono stati presi in considerazione i testi che contenevano due parole o meno, al fine di facilitare la futura assegnazione dei topic alle singole recensioni.

Al termine di questo processo, le recensioni disponibili per l'analisi sono N=20750. Il vocabolario del dataset iniziale (N=24060) era composto da 52388 parole diverse; al termine della preelaborazione (N=20750) il corpus contiene, invece, 15151 termini unici. La fase di preelaborazione ha pertanto portato a una diminuzione del 71,08% delle parole del dataset iniziale.

Applicazione di LDA

Prima di applicare l'algoritmo LDA si è resa necessaria la messa in atto di alcune operazioni preliminari. Inizialmente, il documento contenente tutte le recensioni era diviso in tre documenti separati:

- Il primo è documentsValidation, il quale contiene 2075 recensioni ed è il risultato dell'estrazione casuale del 10% delle recensioni del campione. Questo è stato utilizzato per il calcolo della perplexity.
- Il secondo è documentsValidation2, il quale contiene 50 documenti ed è stato utilizzato successivamente nella fase di validazione dell'algoritmo.
- Il terzo è documentsTrain, il quale contiene le rimanenti 18625 recensioni, è il documento sul quale è stato applicato l'algoritmo LDA.

A questo punto si è applicata la funzione bagOfWords(.) sul documento documentsTrain, la quale tiene traccia del numero di occorrenze di ogni parola all'interno del documento selezionato.

In seguito, sono state rimosse le parole che compaiono due volte o meno all'interno del documento per eliminare possibili termini contenenti errori di battitura o termini rari al fine di migliorare l'indicatore di bontà d'adattamento e avere risultati migliori come topic di output.

Dopo queste operazioni preliminari si è applicato l'algoritmo LDA sul campione. Inizialmente si è reso necessario scegliere il solver tra i quattro proposti da Matlab (*Text Analytics Toolbox - MATLAB*). Il primo solver è chiamato "savb" ed è definito stocastico, utilizza il metodo stochastic approximate variational Bayes. Questo si presta molto bene ai dataset corposi ed è veloce dal punto di vista del tempo

computazionale. I restanti tre solver sono definiti “batch solvers”. In particolare, “cgs” utilizza il collapsed Gibbs sampling, è molto accurato ma ha un costo computazionale maggiore. Il solver “avb”, il quale adopera approximate variational Bayes, solitamente risulta essere il più rapido, caratteristica che tutta via si riflette negativamente sull’accuratezza. Infine, il “cvb0” fa uso del collapsed variational Bayes di ordine zero e si dimostra il solver più accurato seppur il più lento.

L’unico solver preso in considerazione è stato il “cvb0”, poiché, quando utilizzato, è stato in grado di restituire un indicatore di perplexity minore per qualsiasi α e NumTopics scelti, pur comportando un tempo computazionale maggiore per l’esecuzione dello script. Come si può notare dalla Figura 3 i solver “savb” e “avb” hanno manifestato una perplexity crescente con l’aumentare del numero di topics. Il metodo “cgs” ha restituito una perplexity circa costante, mentre il “cvb0” si è dimostrato il solver migliore per tutti i valori di numero di topics presi in considerazione.

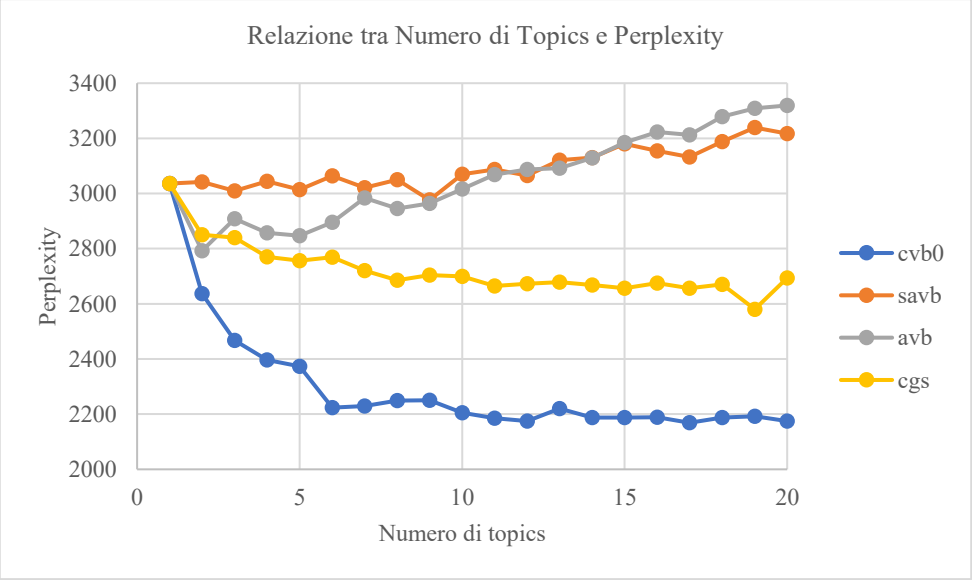


Figura 3 *Relazione tra numero di topics e perplexity per i solver cvb0, savb, avb, cgs*

Successivamente, scegliendo il numero di topics per l’analisi vera e propria, si è optato per una quantità pari a 15.

A questo punto si è resa necessaria la scelta del parametro α tale da minimizzare la perplexity. Quest'ultimo è chiamato anche Initial Topic Concentration (ITC), poiché indica la concentrazione media di topics all'interno di ogni documento. L'ITC ottimale al fine di minimizzare la perplexity è stato individuato nel valore 4,5 poiché è coincidente con il valore minore di perplexity ovvero 2068, come è possibile osservare in Figura 4; ciò significa che in media ogni recensione contiene al suo interno più di 4,5 topics.

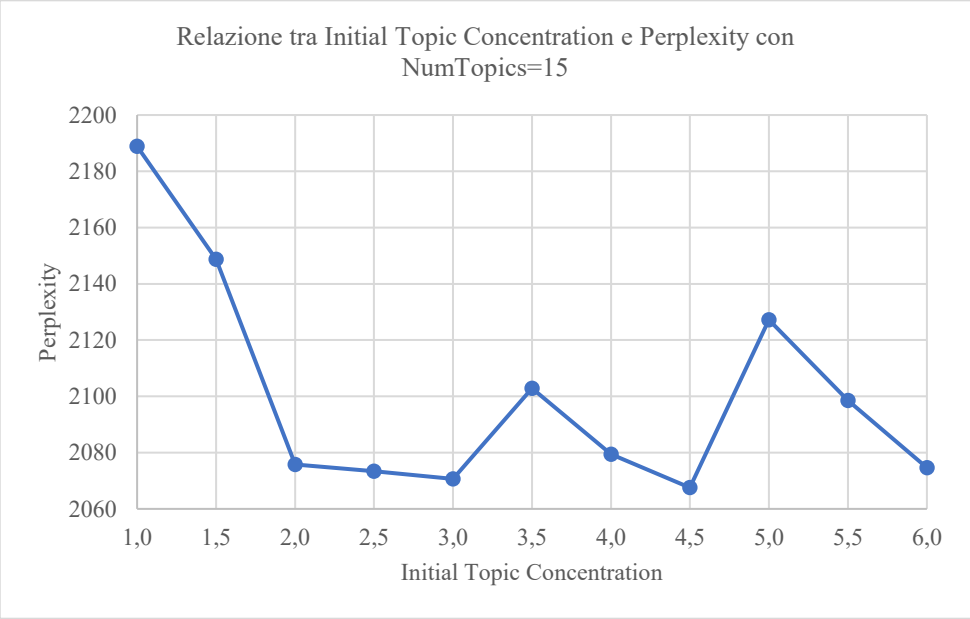


Figura 4 *Relazione tra Initial Topic Concentration e Perplexity con parametro NumTopics=15*

Estrazione dei topic e assegnazione di questi ultimi alle recensioni

Dopo aver scelto i parametri, si è proceduto all'esecuzione definitiva dell'algoritmo LDA sul documento DocumentsTrain.

L'output è stato composto da una lista numerata di quindici topic con altrettante parole ad essi associati, presenti in Appendice C. I termini all'interno dei topic sono stati disposti in ordine decrescente in base ad un valore di probabilità di una determinata

parola di appartenere effettivamente al topic selezionato. È stata utilizzata successivamente la funzione Transform per convertire l'intero corpus DocumentsTrain in una matrice rinominata TopicMixtures, di dimensioni pari a 18625x15. Ogni riga è corrispondente ad un vettore associato a una recensione, in grado di assegnare una percentuale di prevalenza di ogni topic alla recensione.

L'elemento generico della matrice è stato definito p_{ij} come percentuale di prevalenza del topic j-esimo associato alla recensione i-esima. Al fine di permettere una migliore comprensione della composizione di questa matrice, a titolo esemplificativo, si riporta di seguito una recensione tratta dal campione, precisamente la numero 9954, la quale recita:

*"Un ringraziamento a tutta l'equipe del reparto di nefrologia, al prof. *** e a tutti gli infermieri e medici che mi hanno assistito nel mio periodo di degenza soprattutto un grazie di cuore per la pazienza e professionalità che hanno nei miei confronti... Grazie di cuore a tutti".*

A questa recensione è stato associato questo vettore:

[0,0093 0,0093 0,0909 0,0121 0,0124 0,7724 0,0120 0,0093 0,0125 0,0103 0,0099 0,0092 0,0098 0,0113 0,0093].

Ogni componente indica la prevalenza del topic all'interno della recensione, in questo caso il topic 6 ha una prevalenza del 77,24%, quindi è il topic dominante, mentre altri topic sono irrilevanti come i topic 1, 2, 8, 11, 12, 13, i quali hanno prevalenza minore dell'1%. A questo punto è stata eseguita la fase di assegnazione dei topic prevalenti all'interno delle recensioni procedendo secondo le seguenti regole.

Innanzitutto, è stata definita la soglia $\theta = 6{,}667\%$ la quale è il risultato di $\frac{1}{15}$, dove il denominatore corrisponde al numero di topics. La logica che ha condotto all'individuazione di un tale valore soglia è così esplicabile: nel caso peggiore, la recensione più confusa in assoluto ha il seguente vettore di prevalenza:

[0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667 0,0667].

Ciò significa che nessun topic è contenuto all'interno della recensione, poiché l'algoritmo LDA utilizza una logica fuzzy. Da questa assunzione è stata stabilita l'assegnazione di un topic alla recensione qualora la differenza tra il topic prevalente (quello con percentuale di prevalenza maggiore) e il secondo topic prevalente sia maggiore di 6,667%. In caso contrario la recensione è stata definita come confusa, poiché presenti almeno due topic molto prevalenti. La Tabella 3 ivi riportata a fini esplicativi illustra la modalità di assegnazione di un topic.

A questo punto si è proceduto con l'assegnazione di un topic per ogni recensione non definibile "confusa" del documento TopicMixtures.

Caso	Assegnazione
p_{ij}-$p_{ik} > 6{,}667\%$	Topic j-esimo è assegnato alla recensione i-esima
p_{ij}-$p_{ik} \leq 6{,}667\%$	Recensione i-esima è definita confusa

Tabella 3 *Modalità di assegnazione del topic a una recensione*

Validazione dell'algoritmo

Successivamente alla fase di assegnazione dei topic alle recensioni, è seguita la validazione dell'algoritmo LDA.

Con il termine validazione si intende il processo di determinazione del grado di accuratezza dei risultati dell'algoritmo rispetto al mondo reale (Sornette *et al.*, 2007). Durante l'esecuzione di questa fase si è utilizzato il concetto di validità di convergenza, ovvero il grado di correlazione esistente tra le misure di uno stesso costrutto ottenute con metodi diversi ed indipendenti tra loro (Franceschini, 2001).

Il campione utilizzato per la validazione è documentsValidation2, originato da un'estrazione casuale di 50 recensioni dal corpus totale. In questo caso la validazione è manuale ed è stata effettuata dal tesista, il quale ha letto le recensioni e ha assegnato un topic ad ognuna di esse qualora ritenuto appropriato, oppure nel caso in cui la

recensione conteneva all'interno più topics o nessuno dei 15 elencati è stata definita confusa. In seguito, si è stabilita la definizione di alcuni costrutti tali da permettere il calcolo degli indici statistici ai fini della comprensione della bontà di validazione.

- Vero positivo (VP): il topic assegnato alla recensione dall'esperto coincide con il topic prevalente assegnato dall'algoritmo LDA.
- Vero negativo (VN): l'esperto non è in grado di assegnare un topic alla recensione e l'algoritmo non riesce a fare lo stesso (la recensione è classificata come confusa sia dall'esperto che dall'algoritmo).
- Falso positivo (FP): l'algoritmo assegna un topic prevalente alla recensione ma l'esperto giudica la recensione confusa o le assegna un topic differente.
- Falso negativo (FN): diversamente dall'esperto, l'algoritmo non è in grado di assegnare un topic prevalente alla recensione.

A titolo di esempio è riportata una recensione dal documento documentsValidation2:

"Sono 15 giorni che cerco di contattare il centro di prenotazioni, per fissare una prima visita con un gastroenterologo. ho fatto la prenotazione via mail, ma non sono mai stata contattata al telefono passo giornalmente qualche mezzora, ma non ho ancora avuto neppure il piacere di sentire una voce, tranne quella registrata, che mi dice di attendere in linea per non perdere la priorità acquisita. Qualcuno riesce a dirmi come devo fare? Devo necessariamente perdere una mezza giornata per venire di persona? Grazie per una risposta"

Si noti come il topic prevalente sia il topic 1, poiché sono presenti alcune delle parole più rappresentative di questo, come ad esempio "visita", "prenotazioni". L'opinione dell'esperto è supportata dall'algoritmo LDA, poiché come è possibile notare dal vettore di prevalenza sottostante, il topic più rappresentativo è il topic 1.

[0,6203 0,0118 0,0068 0,0119 0,0073 0,0193 0,0185 0,0097 0,0076 0,0316 0,0097 0,1140 0,0066 0,1180 0,0068]

L'output fornito dall'algoritmo e la scelta effettuata dal tesista in questo caso coincidono, pertanto la recensione è etichettata come un vero positivo.

Si è compiuta questa attività per tutte le 50 recensioni del corpus del documento utilizzato per la validazione (documentsValidation2).

I risultati sono illustrati in Tabella 4.

Etichetta	Numero di recensioni
Vero positivo	18
Vero negativo	12
Falso positivo	14
Falso negativo	6
Totale	50

Tabella 4 *Lista delle etichette e il relativo numero di recensioni a esse collegate*

Successivamente sono stati individuati degli indicatori per la comprensione della bontà di validazione (Zaki and McColl-Kennedy, 2019).

Il primo indicatore è chiamato recall ed è definito come il rapporto fra il numero di veri positivi (il numero di recensioni assegnate correttamente al topic da LDA) e la somma di veri positivi e falsi negativi.

Il secondo indicatore è chiamato precision ed è definito come il rapporto fra il numero di veri positivi e la somma di veri positivi e falsi positivi.

Il terzo indicatore è chiamato F-measure ed è la media armonica dei due indicatori precedenti (recall e precision).

Infine, l'ultimo indicatore è chiamato accuracy ed è definito come il rapporto tra la somma dei veri positivi e veri negativi e il totale degli elementi classificati. Di seguito sono mostrate le formule:

$$Recall\ (R) = \frac{VP}{VP + FN} \tag{11}$$

$$Precision\ (P) = \frac{VP}{VP + FP} \tag{12}$$

$$F - measure\ (F1) = \frac{2\ \times P \times R}{P + R} \tag{13}$$

$$Accuracy(A) = \frac{VP + VN}{VP + VN + FP + FN} \tag{14}$$

In Tabella 5 sono riassunti i risultati degli indicatori per il campione relativo alla validazione.

Indicatore	**Risultato**
Recall	75,00%
Precision	56,25%
F-measure	64,29%
Accuracy	60,00%

Tabella 5 *Lista degli indicatori e i relativi risultati sul campione utilizzato per la validazione*

A questo punto è necessario comprendere se i valori ottenuti di questi quattro indicatori sono in linea con gli studi in letteratura al fine di valutare la bontà dell'algoritmo. Lo studio di Pang, Lee e Vaithyanathan (2002) riporta una recall che varia, a seconda del campione, dal 50% al 69%, mentre l'accuracy si colloca in un range tra il 58% e il 65%. L'articolo scritto da Ye, Zhang e Law (2009) fornisce come output un valore di recall compreso tra il 67% e il 93% e accuracy tra il 51% e l'84% a seconda del campione selezionato. Il terzo articolo selezionato è scritto da Denecke (2008) e restituisce una precision del 67% e una recall del 63%. Il penultimo studio è stato condotto da Kasper e Vela (2011) e al suo interno vede l'utilizzo dell'indicatore di sintesi F-measure il quale è situato tra 0.66 e 0.81, mentre l'accuracy varia dal 54% al 67%. L'ultimo studio, elaborato da Greaves, Ramirez-Cano, Millett, Darzi e

Donaldson (2013), è molto accurato e presenta una F-measure tra lo 0.8 e lo 0.89 e un accuracy tra l'80% e l'89%. Nella Tabella 6 è illustrata una sintesi.

Autori	**Titolo**	**Recall**	**Precision**	**F-measure**	**Accuracy**
Pang, Lee, Vaithyanathan	Thumbs up? Sentiment Classification using Machine Learning Techniques	50%-69%			58%-65%
Ye, Zhang, Law	Sentiment classification of online reviews to travel destinations by supervised machine learning approaches	67%-93%			51%-84%
Denecke	Using SentiWordNet for multilingual sentiment analysis	63%	67%		
Kasper, Vela	Sentiment Analysis for Hotel Reviews			0.66-0.81	54%-67%
Greaves et al.	Use of sentiment analysis for capturing patient experience from free-text			0.8-0.89	80%-89%

	comments posted online.				

Tabella 6 *Lista autori e relativi articoli scientifici e valori dei principali indicatori di bontà di validazione*

In definitiva i valori degli indicatori individuati in questo studio si confermano in linea con la letteratura presa in analisi. Si segnala una recall maggiore rispetto agli studi comparati, mentre la precision è leggermente inferiore. L'F-measure risulta lievemente minore qualora comparata a quella riportata dagli altri studi e l'accuracy si colloca nella media. In generale un indicatore di accuracy maggiore del 55% indica un buon risultato di validazione (Khadjeh Nassirtoussi *et al.*, 2014).

Etichettatura dei topic

L'ultima fase di preparazione all'analisi è quella di etichettatura, la quale consiste nel dare un titolo a ogni topic. Questa fase è molto dipendente dal soggetto che la esegue, poiché non è riportata in letteratura una metodologia universalmente riconosciuta per l'attribuzione delle etichette. È possibile che persone diverse possano assegnare etichette fortemente differenti allo stesso topic.

In appendice C è riportata la lista delle etichette attribuite a ogni topic e delle prime quindici parole ad esso associate; inoltre per ogni termine è presente un peso che indica la probabilità che la parola appartenga effettivamente a quel topic.

Questa fase si è rivelata nel complesso non difficoltosa, seppur il topic 14 inizialmente fosse confuso, poiché l'individuazione di un termine in grado di rappresentarlo non si è dimostrata immediata.

Per semplificare tale azione, pertanto, si è proceduto alla lettura delle recensioni per le quali il topic 14 era il prevalente al fine di comprendere meglio la sua natura. In definitiva il topic sembra essere rappresentato bene dalla parola "Competenza", nonostante le recensioni trattassero di altri temi quali la cortesia, la credibilità e la comunicazione.

Capitolo III

Presentazione dei risultati: Panoramica generale

Il punto di partenza dell'analisi è l'applicazione di LDA al campione documentsTrain composto da 18625 recensioni relative a 648 strutture ospedaliere italiane e l'assegnazione dei topic alle recensioni. Successivamente sono state calcolate alcune statistiche. Per ogni topic si determina la percentuale media di prevalenza sul totale delle recensioni, la varianza e la deviazione standard, al fine di comprendere la dispersione della variabile "prevalenza". I risultati sono illustrati in Tabella 7. Come è possibile osservare, il topic in media più prevalente è il topic 6 con 8,41% di prevalenza media; questo risulta anche essere quello con la deviazione standard superiore. Al contrario, il meno prevalente è il topic 14 con 3,15% e la deviazione standard inferiore.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	7,31%	5,94%	7,90%	5,91%	5,27%
Varianza	1,56%	0,98%	1,16%	0,81%	0,63%
Deviazione standard	12,49%	9,91%	10,76	9,01%	7,93%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	8,41%	7,59%	4,32%	3,20%	3,33%
Varianza	1,77%	1,22%	1,11%	0,68%	0,74%
Deviazione standard	13,29%	11,06%	10,54%	8,26%	8,58%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Prevalenza media del topic su tutte le recensioni	4,18%	3,29%	3,59%	3,15%	5,18%

Varianza	1,30%	0,78%	0,78%	0,58%	1,32%
Deviazione standard	11,38%	8,81%	8,81%	7,62%	11,50%

Tabella 7 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere di tutta l'Italia*

Successivamente sono state conteggiate le recensioni per le quali il topic i-esimo è il prevalente e quelle per le quali non è presente un topic prevalente, ovvero le cosiddette recensioni confuse. Nella Tabella 8 sono illustrati i risultati. È evidente l'elevato numero di recensioni confuse, che corrispondono al 24,40% del totale. Il topic più discusso nelle recensioni è il topic 15 (affidabilità) poiché risulta essere prevalente nell'8,78% delle recensioni; a seguire il topic 6 (cortesia/empatia) con l'8,39% e in terza posizione il topic 3 (competenza 1) con il 7,55%. Tra i topic meno discussi si evidenziano il topic 14 (competenza 2) con il 2,40%, il quale è risultato anche il più arduo da etichettare, il topic 5 (operazione) con il 2,49% e infine il topic 9 (racconto dell'esperienza 2) con il 2,66%.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Numero di recensioni in cui il topic i-esimo è prevalente	1249	677	1407	730	464
Percentuale sul totale delle recensioni	6,71%	3,63%	7,55%	3,92%	2,49%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	1562	1216	1015	496	656
Percentuale sul totale delle recensioni	8,39%	6,53%	5,45%	2,66%	3,52%

	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	1263	542	722	447	1635
Percentuale sul totale delle recensioni	6,78%	2,91%	3,88%	2,40%	8,78%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	4544				
Percentuale sul totale delle recensioni	24,40%				

Tabella 8 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere di tutta l'Italia*

La Figura 5 mostra la percentuale di prevalenza media di un determinato topic prendendo in considerazione solo le recensioni in cui il suddetto è prevalente. I valori sono compresi nel seguente intervallo [34,6%, 56,1%]. Prendendo come esempio il topic 1, sono sommate le percentuali di prevalenza delle recensioni in cui questo è prevalente e tale risultato intermedio è stato diviso per il numero di recensioni in cui il topic 1 è prevalente. Il topic 12 (accesso/visita) ha la percentuale di prevalenza media maggiore che si attesta sul 56,1%, però il suddetto topic è prevalente soltanto in 542 recensioni. Il topic 9 (racconto dell'esperienza 2) ha una prevalenza media del 54,1% ed è prevalente in sole 496 recensioni. Il topic con una prevalenza media più bassa è il topic 3 con un valore di 34,6%.

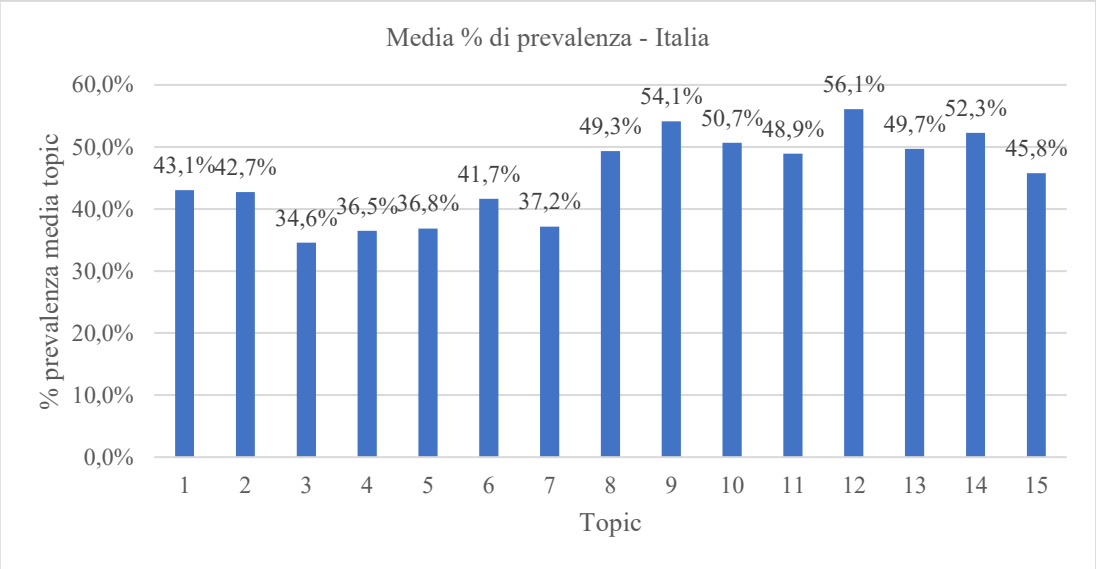


Figura 5 *Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere di tutta Italia*

Successivamente sono state analizzate le distribuzioni della percentuale di prevalenza per ogni topic. Una volta prese in considerazione tutte le recensioni per le quali un dato topic è risultato essere prevalente, è stato costruito un istogramma. Si è compiuta questa operazione per ognuno dei topic individuati, ottenendo pertanto quindici istogrammi. A titolo esemplificativo è riportato l'istogramma del topic 1 nella Figura 6, mentre per i restanti quattordici istogrammi si rimanda all'appendice D.

L'istogramma illustrato di seguito mostra il numero di recensioni per le quali quel topic è prevalente per ogni intervallo di percentuale di prevalenza. Per esempio, sono individuabili 2 recensioni in cui il topic 1 è prevalente con percentuale di prevalenza compresa nell'intervallo [7%, 12%]. La mediana è 41,04%.

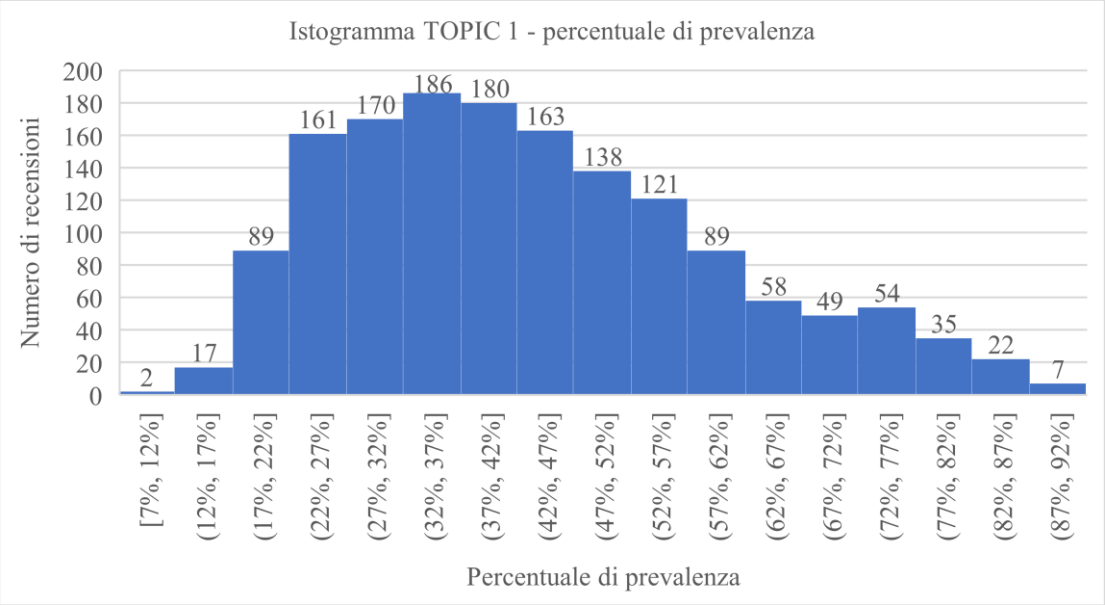


Figura 6 *Istogramma della percentuale di prevalenza quando il topic 1 è prevalente*

Analisi regionale

Le recensioni utilizzate per l'algoritmo LDA sono state divise per ospedale, città e regione. Ciò è stato compiuto per comprendere se sussistono differenze sostanziali tra l'Italia settentrionale, il Centro e il Meridione. Innanzitutto si è proceduto con la definizione delle zone, adoperando il criterio utilizzato dall'Istat (*Ripartizione geografica, Istat*). L'Italia è così ripartita:

- Nord: Piemonte, Valle d'Aosta, Liguria, Lombardia, Trentino-Alto Adige, Veneto, Friuli – Venezia Giulia, Emilia – Romagna.
- Centro: Toscana, Umbria, Marche, Lazio.
- Mezzogiorno: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna.

A questo punto è stata effettuata un'analisi sulle 20750 recensioni italiane. In particolare, quelle del Nord Italia sono 10789 (52,0% del totale), quelle del Centro Italia 5254 (25,3% del totale) e quelle del Sud Italia e Isole 4707 (22,7% del totale).

Nella prima fase del lavoro si sono analizzate separatamente le tre zone, mentre in una seconda fase è stata effettuata un'analisi specifica delle recensioni degli ospedali dei maggiori centri urbani italiani (Roma, Milano, Napoli e Torino).

Presentazione dei risultati: Nord Italia

Le recensioni analizzate del Nord Italia, relative a 294 strutture ospedaliere, sono 10789. Dalle statistiche calcolate, riportate nella Tabella 9 è possibile osservare la prevalenza media di ogni topic sul totale delle recensioni, la varianza di questa variabile e la deviazione standard. Il topic con una prevalenza media più alta è il topic 15 (affidabilità) con 8,89%, mentre il topic 1 (tempo di risposta) ha la varianza maggiore ovvero 1,55%. Il topic 5 (operazione) è quello con la prevalenza minore (5,29%) e il topic 14 (competenza 2) quello con la varianza minore (0,56%).

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	7,32%	6,03%	7,90%	5,82%	5,29%
Varianza	1,55%	1,02%	1,17%	0,77%	0,62%
Deviazione standard	12,45%	10,11%	10,84%	8,76%	7,85%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	8,43%	7,12%	7,47%	5,51%	5,74%
Varianza	1,75%	1,09%	1,26%	0,65%	0,68%
Deviazione standard	13,23%	10,46%	11,21%	8,06%	8,25%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15

Prevalenza media del topic su tutte le recensioni	7,18%	5,68%	6,19%	5,43%	8,89%
Varianza	1,26%	0,79%	0,79%	0,56%	1,38%
Deviazione standard	11,23%	8,87%	8,90%	7,50%	11,76%

Tabella 9 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere del Nord Italia*

Successivamente sono state conteggiate le recensioni confuse e quelle per le quali il topic i-esimo è il prevalente. Nella Tabella 10 sono illustrati i risultati. Il topic più prevalente nelle recensioni del Nord Italia è il topic 15 (affidabilità) con 1028 recensioni, ovvero il 9,53% del totale; la seconda posizione è occupata dal topic 6 (cortesia/empatia) con 920 recensioni (8,53% del totale) e in terza posizione si osserva il topic 3 (competenza) con 814 recensioni (7,54% del totale). Tra i topic meno prevalenti si annoverano il topic 14 (competenza 2) presente nel 2,36% delle recensioni, il topic 5 (operazione) con il 2,38% e il topic 9 (racconto dell'esperienza 2) con il 2,57%. È importante menzionare che 2661 recensioni (24,66% del totale) sono risultate confuse e non è stato possibile assegnare un topic prevalente.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Numero di recensioni in cui il topic i-esimo è prevalente	704	407	814	398	257
Percentuale sul totale delle recensioni	6,53%	3,77%	7,54%	3,69%	2,38%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	920	604	662	277	351

Percentuale sul totale delle recensioni	8,53%	5,60%	6,14%	2,57%	3,25%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	702	320	429	255	1028
Percentuale sul totale delle recensioni	6,51%	2,97%	3,98%	2,36%	9,53%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	2661				
Percentuale sul totale delle recensioni	24,66%				

Tabella 10 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere del Nord Italia*

In Figura 7 è possibile osservare per ogni topic la percentuale media di prevalenza solo quando quel topic è assegnato alle recensioni. Si noti come il topic 12 (accesso/visita) ha la percentuale di prevalenza maggiore (55,5%), a seguire il topic 9 (racconto dell'esperienza 2) con 54,2% e infine il topic 14 (competenza 2) con 51,1%. I topic con percentuali di prevalenza media più bassi sono il topic 3 (competenza) con 35,0%, il topic 4 (ostetricia e ginecologia) con 35,9% e il topic 5 (operazione) con 36,5%.

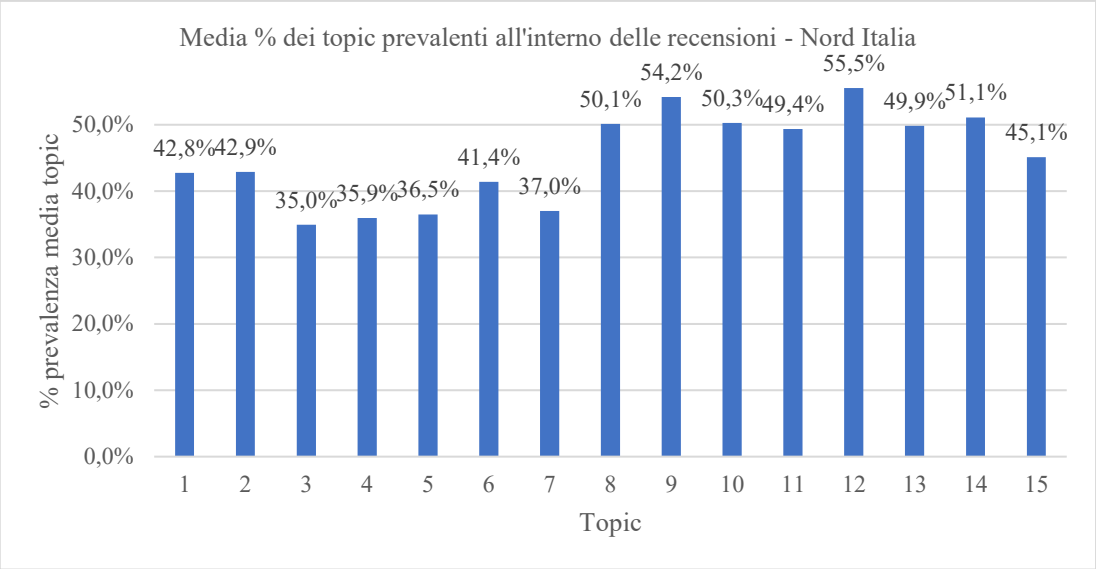


***Figura** 7 Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere del Nord Italia*

Presentazione dei risultati – Centro Italia

In questa fase sono state analizzate 5254 recensioni relative a 127 strutture ospedaliere del Centro Italia. Innanzitutto, sono state calcolate delle statistiche che ne descrivono la composizione, come illustrato in Tabella 11. Il topic più prevalente in media sembra essere il topic 15 (affidabilità) con una prevalenza media dell'8,45%, a seguire il topic 6 (cortesia/empatia) con una prevalenza media dell'8,36%. In terza posizione è apprezzabile il topic 7 (credibilità del personale) con una prevalenza media del 7,75%. Il topic mediamente meno presente è il 5 (operazione) con una prevalenza media pari a 5,15%. Il topic con la prevalenza più dispersa è il topic 6 (cortesia/empatia) con una varianza dell'1,79%.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	7,61%	5,90%	7,59%	6,14%	5,15%
Varianza	1,66%	0,97%	1,11%	0,90%	0,62%

Deviazione standard	12,90%	9,83%	10,51%	9,50%	7,85%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	8,36%	7,75%	6,43%	5,52%	6,07%
Varianza	1,79%	1,29%	0,97%	0,65%	0,75%
Deviazione standard	13,38%	11,36%	9,84%	8,03%	8,65%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Prevalenza media del topic su tutte le recensioni	7,64%	5,55%	6,04%	5,78%	8,45%
Varianza	1,36%	0,77%	0,76%	0,68%	1,33%
Deviazione standard	11,68%	8,80%	8,71%	8,24%	11,53%

Tabella 11 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere del Centro Italia*

Successivamente è stato effettuato per ogni topic il conteggio delle recensioni per le quali è prevalente e la percentuale di queste recensioni sul totale (5254 recensioni) come è possibile osservare in Tabella 12. I risultati mostrano che il topic 15 (affidabilità) è il più presente, essendo prevalente in 442 recensioni (8,41% del totale); a questo seguono il topic 6 (cortesia/empatia) prevalente in 429 recensioni (8,17%) del totale e il topic 11 (aspetti tangibili), prevalente in 388 recensioni (7,38% del totale). Il topic meno prevalente del Centro Italia è il topic 9 (racconto dell'esperienza 2), il quale è assegnato a sole 125 recensioni (2,38% del totale). È altresì importante menzionare le 1268 recensioni confuse, che costituiscono il 24,13% del totale.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5

Numero di recensioni in cui il topic i-esimo è prevalente	386	188	386	227	136
Percentuale sul totale delle recensioni	7,35%	3,58%	7,35%	4,32%	2,59%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	429	360	249	125	180
Percentuale sul totale delle recensioni	8,17%	6,85%	4,74%	2,38%	3,43%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	388	154	186	150	442
Percentuale sul totale delle recensioni	7,38%	2,93%	3,54%	2,85%	8,41%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	1268				
Percentuale sul totale delle recensioni	24,13%				

Tabella 12 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere del Centro Italia*

Nella Figura 8 è possibile osservare per ogni topic la percentuale media di prevalenza solo quando quel topic è assegnato alle recensioni. Si noti che il topic 9 (racconto

dell'esperienza 2) quando è assegnato a una recensione ha in media una prevalenza molto alta (55,78%), a seguire c'è il topic 14 (competenza 2) con una prevalenza media del 53,2% ed infine è presente il topic 10 (reputazione) con una prevalenza media del 52,3%. I topic con percentuali di prevalenza media più bassi sono il topic 3 (competenza) con 34,4%, il topic 5 (operazione) con 37,1% e il topic 7 con 37,6%.

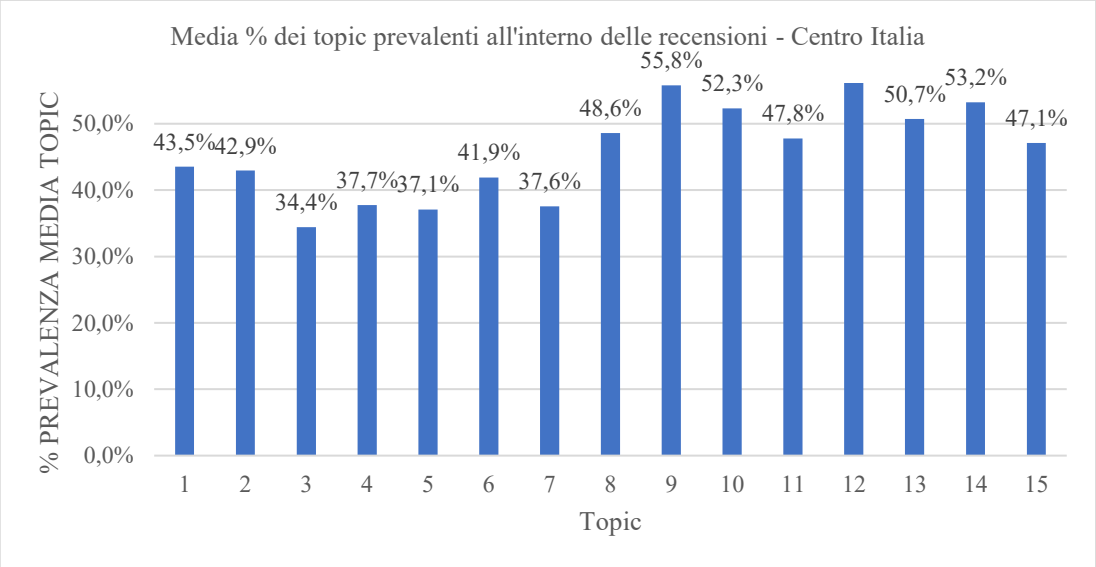


Figura 8 *Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere del Centro Italia*

Presentazione dei risultati – Mezzogiorno

Durante l'esecuzione di questa fase sono stati analizzati i dati forniti da 4707 recensioni relative a 227 strutture ospedaliere del Mezzogiorno. La Tabella 13 mostra per ogni topic la percentuale media di prevalenza calcolata su tutte le recensioni, la relativa varianza e la deviazione standard. Come è possibile notare, il topic con una prevalenza media più alta è il topic 7 (credibilità del personale) con un valore pari a 9,60%, in seconda posizione si trova il topic 6 (cortesia/empatia) con 8,34% e a seguire il topic 3 (competenza) con una prevalenza media di 8,02%. I topic con una prevalenza media minore sono il topic 14 (competenza 2) con 4,98%, il topic 5 (operazione) con 5,15% e infine il topic 12 (accesso/visita) con 5,49%. Il topic che ha la variabile

prevalenza più dispersa è il topic 6 (cortesia/empatia) con una varianza di 1,57%, mentre la varianza minore appartiene al topic 5 (operazione) con un valore di 0,51%.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	6,89%	5,82%	8,02%	5,63%	5,15%
Varianza della prevalenza	1,26%	0,72%	1,05%	0,61%	0,51%
Deviazione standard	11,23%	8,49%	10,27%	7,82%	7,14%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	8,34%	9,60%	6,24%	6,08%	6,72%
Varianza	1,57%	1,43%	0,74%	0,65%	0,72%
Deviazione standard	12,55%	11,95%	8,60%	8,05%	8,50%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Prevalenza media del topic su tutte le recensioni	7,23%	5,49%	6,45%	4,98%	7,36%
Varianza	1,09%	0,58%	0,63%	0,36%	0,96%
Deviazione standard	10,44%	7,60%	7,94%	6,03%	9,82%

Tabella 13 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere del Mezzogiorno*

A questo punto, come è possibile osservare nella Tabella 14, per ogni topic, è stato effettuato il conteggio delle recensioni per le quali quel topic è prevalente. La seconda riga è il risultato della divisione tra il numero in prima riga e le totali 4707 recensioni

del Mezzogiorno. Il topic 7 (credibilità del personale) risulta essere quello assegnato al maggior numero di recensioni (295) con un valore percentuale pari a 10,18%. A questo segue il topic 6 (cortesia/empatia) con 390 recensioni (8,29%), mentre in terza posizione c'è il topic 3 con 334 recensioni e in percentuale corrispondente al 7,10%. Il topic meno prevalente è il topic 14 (competenza 2) il quale è assegnato a sole 67 recensioni (1,42% del totale). In ultima istanza è utile notare che le recensioni a cui non è stato possibile assegnare un topic sono 1364, pari al 28,98% del totale.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Numero di recensioni in cui il topic i-esimo è prevalente	295	154	334	140	91
Percentuale sul totale delle recensioni	6,27%	3,27%	7,10%	2,97%	1,93%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	390	479	202	125	192
Percentuale sul totale delle recensioni	8,29%	10,18%	4,29%	2,66%	4,08%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	303	114	172	67	285
Percentuale sul totale delle recensioni	6,44%	2,42%	3,65%	1,42%	6,05%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	1364				

Percentuale sul totale delle recensioni	28,98%	

Tabella 14 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere del Mezzogiorno*

Nella Figura 9 è possibile osservare per ogni topic la percentuale media di prevalenza solo quando quel topic è assegnato alle recensioni. In particolare, il topic 14 (competenza 2) ha la percentuale di prevalenza maggiore (56,1%). La seconda percentuale di prevalenza maggiore, ovvero il 54,9%, si osserva per il topic 9 (racconto dell'esperienza 2), a cui segue, infine, il topic 12 (accesso/visita) con una percentuale di prevalenza del 53,3%. I topic con percentuali di prevalenza minori sono il topic 3 (competenza) con un valore percentuale di 32,7%, il topic 4 (ostetrica e ginecologia) con 33,1% e il topic 5 (operazione) con 34,1%.

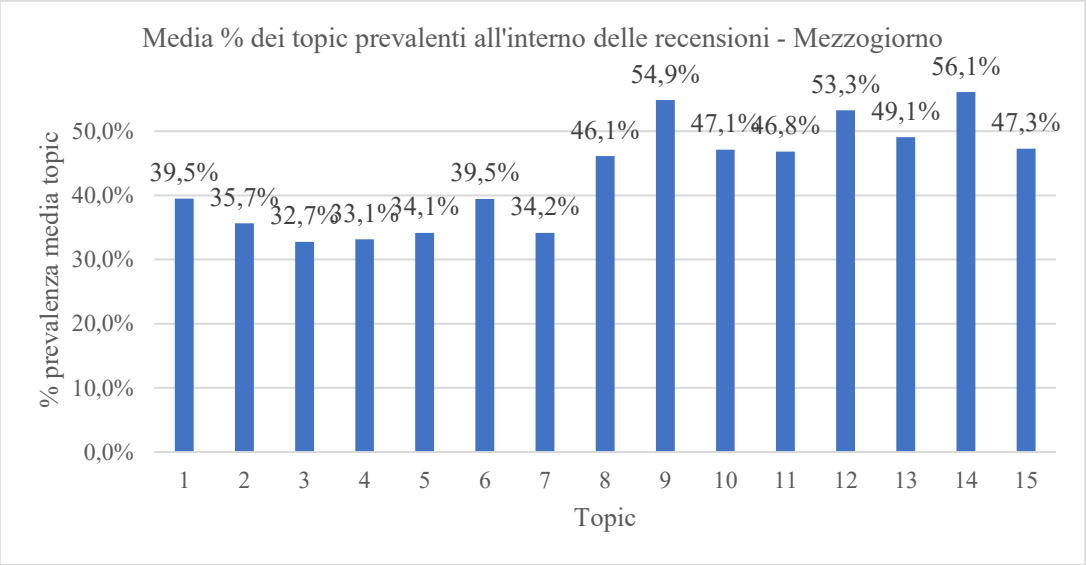


Figura 9 *Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere del Mezzogiorno*

Presentazione dei risultati – Roma

In questa fase è stato applicato l'algoritmo esclusivamente alle recensioni degli ospedali del comune di Roma. In particolare, sono state analizzate 2764 recensioni relative a 26 strutture ospedaliere. In Tabella 15 sono riportate alcune statistiche. Il topic 1 (tempo di risposta) è in media il più prevalente con un valore di 9,15%, in seconda posizione è presente il topic 6 (cortesia/empatia) con una prevalenza media dell'8,56%, in terza posizione si osserva il topic 15 (affidabilità) con una percentuale media del 7,79%. I tre topic meno prevalenti sono il topic 5 (operazione), il topic 10 (reputazione) e il topic 9 (racconto dell'esperienza 2) con una percentuale di prevalenza media rispettivamente del 5,33%, 5,49% e 5,63%.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	9,15%	5,78%	7,55%	6,30%	5,33%
Varianza della prevalenza	2,22%	0,94%	1,12%	1,05%	0,69%
Deviazione standard	14,91%	9,68%	10,58%	10,25%	8,32%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	8,56%	7,18%	6,54%	5,63%	5,49%
Varianza	1,84%	1,16%	1,05%	0,71%	0,61%
Deviazione standard	13,56%	10,76%	10,27%	8,42%	7,80%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Prevalenza media del topic su tutte le recensioni	6,92%	5,78%	5,84%	6,15%	7,79%
Varianza	1,23%	0,86%	0,75%	0,78%	1,19%

Deviazione standard	11,08%	9,29%	8,66%	8,81%	10,89%

Tabella 15 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Roma*

In Tabella 16 sono rappresentate per ogni topic il numero di recensioni per le quali quel topic è prevalente e la relativa percentuale sul totale delle recensioni. il topic 1 (tempo di risposta) è assegnato a 270 recensioni con un valore percentuale del 9,77%; a questo segue il topic 6 (cortesia/empatia) con 235 recensioni e 8,50%, ed infine si osserva il topic 15 (affidabilità) con 80 recensioni (7,42%). I topic meno assegnati alle recensioni sono il topic 9 (racconto dell'esperienza 2), il topic 5 (operazione) e il topic 10 (reputazione) rispettivamente con 68, 79 e 80 recensioni.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Numero di recensioni in cui il topic i-esimo è prevalente	270	97	202	135	79
Percentuale sul totale delle recensioni	9,77%	3,51%	7,31%	4,88%	2,86%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	235	162	136	68	80
Percentuale sul totale delle recensioni	8,50%	5,86%	4,92%	2,46%	2,89%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	170	86	89	83	205

Percentuale sul totale delle recensioni	6,15%	3,11%	3,22%	3,00%	7,42%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	667				
Percentuale sul totale delle recensioni	24,13%				

Tabella 16 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Roma*

Nella Figura 10, il topic 9 (racconto dell'esperienza 2) si conferma quello con una maggiore percentuale di prevalenza del 60,8%. Il topic 12 (aspetti tangibili) presenta invece una percentuale del 59,2% ed è seguito dal topic 14 (competenza 2). I topic con percentuali di prevalenza più basse sono il topic 3 (competenza), il topic 7 (credibilità del personale) e il topic 5 (operazione) con rispettivamente il 34,4%, 37,4% e 38,5%.

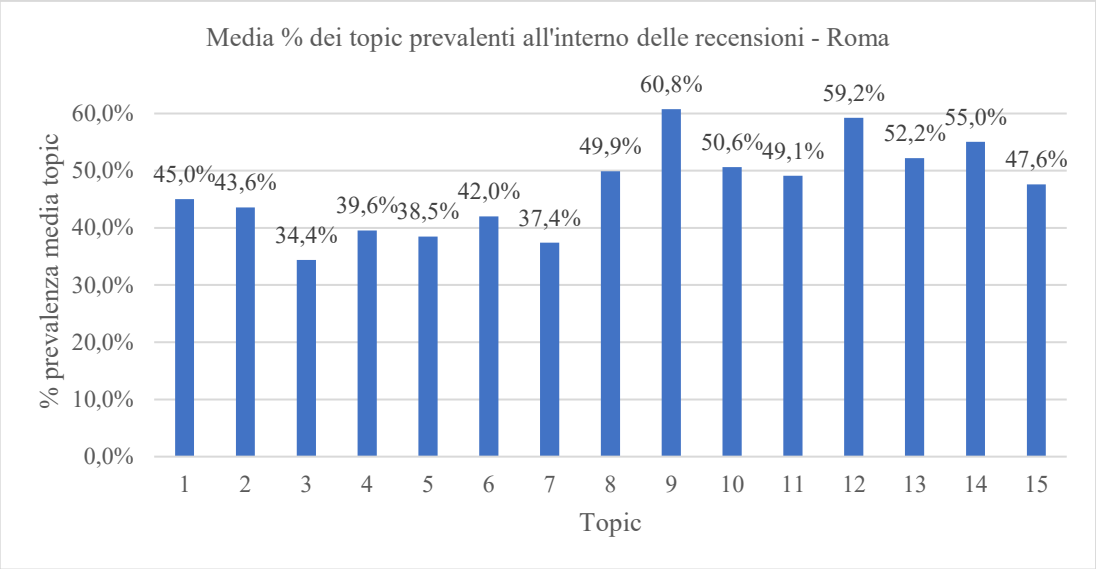


Figura 10 *Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Roma*

Presentazione dei risultati – Milano

Sono state analizzate 1532 recensioni relative a 14 strutture ospedaliere del comune di Milano. In Tabella 17 sono illustrate alcune delle statistiche eseguite. Il topic 1 (tempo di risposta) risulta il più prevalente in media con una percentuale del 9,64%, seguito dal topic 8 (pronto soccorso) con una percentuale del 9,39% e dal topic 7 (credibilità del personale) con una percentuale del 7,28%. I tre topic meno prevalenti sono il topic 5 (operazione), il topic 10 (reputazione) e il topic 14 (competenza 2) con una percentuale rispettivamente del 5,15%, 5,23% e 5,58%.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	9,64%	6,30%	7,06%	5,86%	5,15%
Varianza della prevalenza	2,54%	1,07%	1,01%	0,85%	0,64%
Deviazione standard	15,93%	10,33%	10,07%	9,21%	8,02%

	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	7,12%	7,28%	9,39%	5,71%	5,23%
Varianza	1,45%	1,06%	1,81%	0,74%	0,56%
Deviazione standard	12,04%	10,27%	13,46%	8,60%	7,47%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Prevalenza media del topic su tutte le recensioni	6,20%	6,39%	6,14%	5,58%	6,94%
Varianza	1,10%	0,92%	0,82%	0,64%	0,95%
Deviazione standard	10,47%	9,60%	9,07%	7,99%	9,74%

Tabella 17 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Milano*

In Tabella 18 per ogni topic è illustrato il conteggio delle recensioni per le quali quel topic è prevalente e la relativa percentuale sul totale delle recensioni. Il topic 1 (tempo di risposta) è assegnato a 153 recensioni (9,99%), mentre il topic 8 (pronto soccorso) è attribuito a 138 recensioni (9,01%). A questi seguono il topic 3 (competenza) e il topic 6 (cortesia/empatia) con 95 recensioni ciascuno (6.20%). I topic meno assegnati alle recensioni sono il topic 5 (operazione), 14 (competenza 2) e 10 (reputazione), rispettivamente con 41, 41 e 42 recensioni.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Numero di recensioni in cui il topic i-esimo è prevalente	153	60	95	65	41

Percentuale sul totale delle recensioni	9,99%	3,92%	6,20%	4,24%	2,68%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	95	84	138	44	42
Percentuale sul totale delle recensioni	6,20%	5,48%	9,01%	2,87%	2,74%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	69	57	65	41	89
Percentuale sul totale delle recensioni	4,50%	3,72%	4,24%	2,68%	5,81%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	394				
Percentuale sul totale delle recensioni	25,72%				

Tabella 18 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Milano*

Nella Figura 11 è possibile osservare per ogni topic la percentuale di prevalenza media quando il topic è prevalente. Il topic 9 (racconto dell'esperienza 2) ha una percentuale di prevalenza media del 60,9%, mentre il topic 11 (aspetti tangibili) ne ha una del

53,4%. Terzi per percentuale di prevalenza sono il topic 8 (pronto soccorso) e il topic 14 (competenza 2), con il 51,3%.

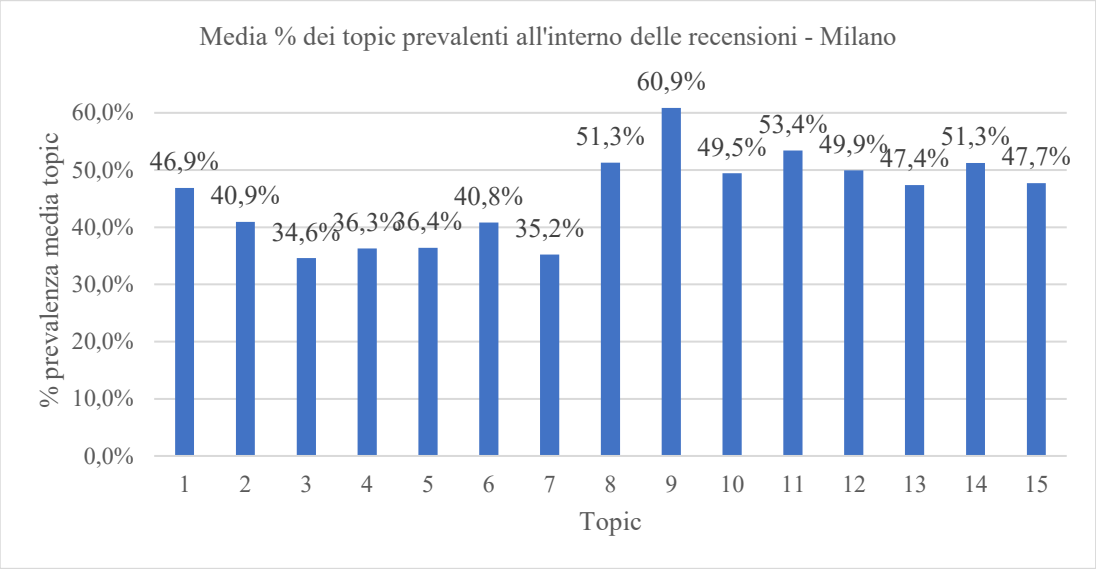


Figura 11 *Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Milano*

Presentazione dei risultati – Napoli

Sono state analizzate 721 recensioni relative a 16 strutture ospedaliere della città di Napoli. In Tabella 19 sono illustrate alcune statistiche come la media della prevalenza del topic su tutte le recensioni e la relativa varianza e deviazione standard. Il topic 3 (competenza) ha la percentuale media più alta con un valore di 9,29%. A questo segue il topic 11 (aspetti tangibili) con una percentuale pari a 8,98%. Infine, si osserva il topic 6 (cortesia/empatia) con una percentuale pari a 8,37%. I tre topic meno prevalenti sono il topic 12 (accesso/visita), il topic 8 (pronto soccorso) e il topic 2 (racconto dell'esperienza) con una percentuale rispettivamente pari a 4,80%, 4,99% e 5,02%.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	6,15%	5,02%	9,29%	6,29%	5,79%
Varianza della prevalenza	1,08%	0,52%	1,31%	1,13%	0,86%
Deviazione standard	10,38%	7,18%	11,43%	10,65%	9,30%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	8,37%	7,50%	4,99%	6,43%	7,15%
Varianza	1,74%	1,01%	0,51%	0,99%	1,03%
Deviazione standard	13,19%	10,07%	7,16%	9,94%	10,17%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Prevalenza media del topic su tutte le recensioni	8,98%	4,80%	6,22%	5,63%	7,40%
Varianza	1,78%	0,56%	0,79%	0,79%	0,97%
Deviazione standard	13,36%	7,51%	8,90%	8,88%	9,84%

Tabella 19 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Napoli*

In Tabella 20 è possibile notare che il topic 3 (competenza) è assegnato a 70 recensioni (9,71%), il topic 11 (aspetti tangibili) a 64 recensioni (8,88%) ed infine il topic 6 (cortesia/empatia) a 49 recensioni (6,80%). Al contrario, i topic meno assegnati sono il topic 2 (racconto dell'esperienza), il 12 (accesso/visita) e l'8 (pronto soccorso) con un numero di recensioni rispettivamente pari a 12, 16, 21.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Numero di recensioni in cui il topic i-esimo è prevalente	42	13	70	37	25
Percentuale sul totale delle recensioni	5,83%	1,80%	9,71%	5,13%	3,47%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	49	47	21	27	41
Percentuale sul totale delle recensioni	6,80%	5,96%	2,91%	3,74%	5,69%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	64	16	27	22	43
Percentuale sul totale delle recensioni	8,88%	2,22%	3,74%	3,05%	5,96%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	177				
Percentuale sul totale delle recensioni	24,55%				

Tabella 20 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Napoli*

Nella Figura 12 è illustrato come il topic 14 (competenza 2) presenti una percentuale di prevalenza del 57,9%; seguono il topic 13 (aspetti tangibili) con 53,6% e il topic 9 (racconto dell'esperienza 2) con 51,2%.

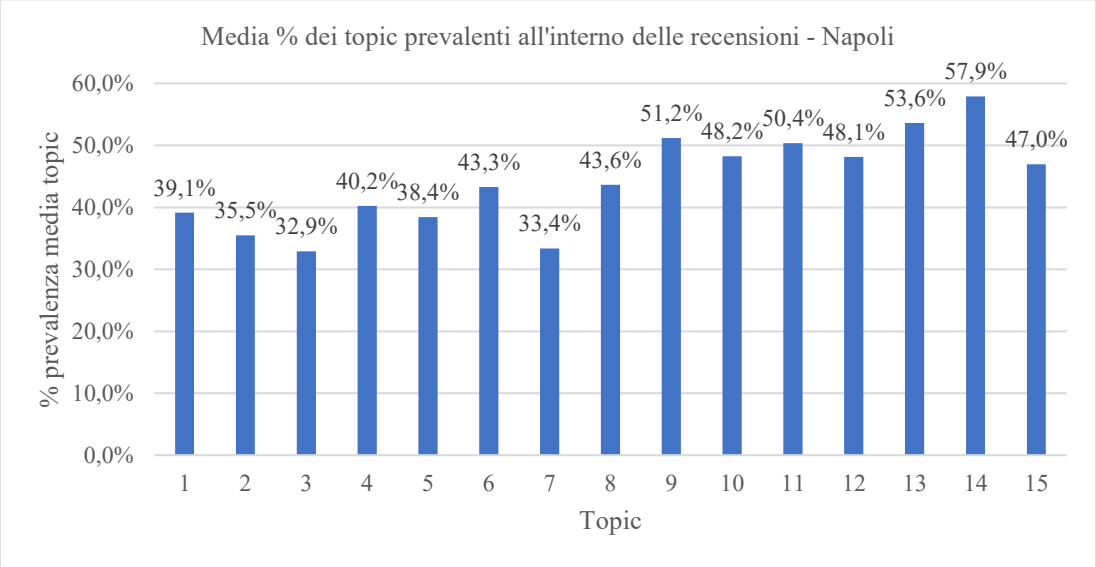


Figura 12 *Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Napoli*

Presentazione dei risultati – Torino

Sono state analizzate 502 recensioni relative a 7 strutture ospedaliere del comune di Torino. In Tabella 21 si osserva per ogni topic la prevalenza media calcolata su tutte le recensioni, la relativa varianza e deviazione standard. Il topic 8 (pronto soccorso) ha una percentuale di prevalenza media pari a 8,91% ed è seguito dal topic 7 (credibilità del personale) con un valore pari a 8,55%. Il terzo topic maggiormente prevalente è il 3 (competenza) con una percentuale pari a 8,44%. I topic in media meno prevalenti sono il topic 5 (operazione), il topic 14 (competenza 2) e il topic 10 (reputazione) con una percentuale rispettivamente pari a 4,73%, 4,99% e 5,17%.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Prevalenza media del topic su tutte le recensioni	6,92%	7,88%	8,44%	6,23%	4,73%
Varianza della prevalenza	1,56%	1,81%	1,35%	0,99%	0,52%
Deviazione standard	12,48%	13,45%	11,62%	9,93%	7,18%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Prevalenza media del topic su tutte le recensioni	7,69%	8,55%	8,91%	5,69%	5,17%
Varianza	1,46%	1,69%	1,73%	0,64%	0,54%
Deviazione standard	12,09%	13,01%	13,15%	8,00%	7,35%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Prevalenza media del topic su tutte le recensioni	5,96%	5,83%	5,57%	4,99%	7,46%
Varianza	1,05%	0,80%	0,65%	0,43%	1,05%
Deviazione standard	10,25%	8,94%	8,06%	6,57%	10,27%

Tabella 21 *Prevalenza media, varianza e deviazione standard di tutti i topic calcolati sul campione di strutture ospedaliere della città di Torino*

In Tabella 22 per ogni topic sono illustrate il numero di recensioni per le quali quel topic è prevalente e la percentuale sul totale delle recensioni. Il topic 3 (competenza) è prevalente in 45 recensioni (8,89%). In seconda posizione c'è il topic 7 (credibilità del personale) con 42 recensioni (8,30%). In terza posizione è presente il topic 8 (pronto soccorso) con 40 recensioni (7,91%). Al contrario, i tre topic meno assegnati alle recensioni sono il topic 5 (operazione), il topic 13 (aspetti tangibili) e il topic 14 (competenza 2) con rispettivamente 7, 12 e 12 recensioni.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Numero di recensioni in cui il topic i-esimo è prevalente	32	33	45	28	7
Percentuale sul totale delle recensioni	6,32%	6,52%	8,89%	5,53%	1,38%
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Numero di recensioni in cui il topic i-esimo è prevalente	37	42	40	15	13
Percentuale sul totale delle recensioni	7,31%	8,30%	7,91%	2,96%	2,57%
	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Numero di recensioni in cui il topic i-esimo è prevalente	24	14	12	12	35
Percentuale sul totale delle recensioni	4,74%	2,77%	2,37%	2,37%	6,92%
	Nessun topic				
Numero di recensioni in cui il topic i-esimo è prevalente	117				
Percentuale sul totale delle recensioni	23,12%				

Tabella 22 *Numero di recensioni in cui il topic i-esimo è prevalente e la relativa percentuale sul totale delle recensioni relative alle strutture ospedaliere della città di Torino*

Nella Figura 13 è possibile osservare come il topic 12 (accesso/visita) sia quello maggiormente prevalente, con una percentuale pari a 59,8%. A questo seguono il topic 13 (aspetti tangibili) con 59,0% e il topic 8 (pronto soccorso) con una percentuale pari a 52,6%. Al contrario, i tre topic con percentuali minori sono il 3 (competenza), il 4 (ostetricia e ginecologia) e il 5 (operazione) rispettivamente con una percentuale di prevalenza media pari a 34,9%, 37,4% e 40,0%.

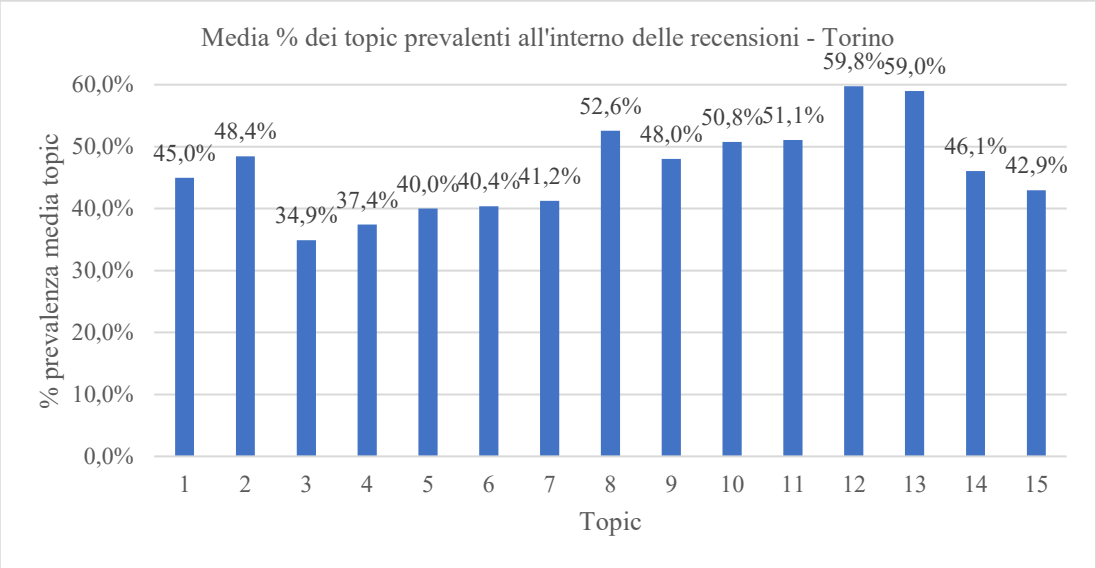


Figura 13 *Percentuale di prevalenza media per ogni topic relativa alle strutture ospedaliere della città di Torino*

Capitolo IV

Discussione dei risultati

Questo studio propone un approccio innovativo per l'analisi della qualità dei servizi ospedalieri, ossia l'applicazione di un algoritmo di machine learning a un campione di UGC, quali le recensioni delle strutture ospedaliere italiane, con lo scopo di individuare le dimensioni di qualità latenti e confrontarle con le dimensioni del modello SERVQUAL, in particolare il SERVQUAL adattato al settore dei servizi sanitari (Duggirala, Rajendran and Anantharaman, 2008).

In questa fase si confronteranno i topic individuati dall'algoritmo LDA (presenti in appendice C) e le dimensioni di qualità percepite dai pazienti individuate da Duggirala et al. (presenti in appendice B).

Il topic 1 (tempo di risposta) si riferisce alla rapidità e al tempo di attesa necessario per prenotare una visita, eseguire un'operazione oppure ritirare dei referti clinici. Il suddetto può essere confrontato con la dimensione di qualità "procedure amministrative", nonostante quest'ultima sia più ampia e non tenga conto solamente della variabile tempo.

Il topic 2 e il topic 9 (racconto dell'esperienza 1 e 2) sono relativi al racconto di un'esperienza clinica all'interno di una struttura ospedaliera, che può comprendere una visita specialistica, un'operazione o un trattamento. Questi ultimi topic menzionati sono paragonabili alla dimensione "processo di assistenza clinica", la quale comprende il processo di trattamento del paziente e l'output del processo terapeutico.

Il topic 3 e il 14 (competenza 1 e 2) si riferiscono alle recensioni in cui sono elencate e spiegate le qualità o i difetti del personale sanitario in relazione all'esperienza del recensore. Questi topic sono comparabili con la dimensione "qualità del personale" che è suddiviso in "assistenza medica", "assistenza infermieristica" e "qualità dello staff paramedico e di supporto" (Duggirala, Rajendran and Anantharaman, 2008).

Il topic 4 (ostetricia e ginecologia) è relativo a un reparto ospedaliero specifico, poiché sono numerose le recensioni che consigliano alle donne in gravidanza a quale struttura

rivolgersi al momento del parto. In particolare, questo topic non trova la collocazione in una sola dimensione, bensì risulta trasversale.

Il topic 5 (operazione) è relativo alla descrizione di un'operazione medico-chirurgica a cui il recensore è sottoposto. Quest'ultimo topic trova la sua collocazione naturale all'interno della dimensione "processo di assistenza clinica" in quanto risultano rilevanti sia il percorso terapeutico fornito sia il risultato al termine di esso.

Il topic 6 (cortesia/empatia) è relativo al modo nel quale i pazienti sono trattati e alla sensibilità dimostrata dal personale sanitario. Questo topic si identifica in maniera ottimale nella dimensione "qualità del personale".

Il topic 7 "qualità delle cure" si riferisce al risultato ottenuto dalle cure fornite dal personale ospedaliero. Il topic è comparabile con la dimensione di qualità "processo di assistenza clinica", in particolare con la sottodimensione "risultati dell'assistenza clinica".

Il topic 8 (pronto soccorso) è relativo alla descrizione di un'esperienza passata al pronto soccorso. Questo non è identificabile da una sola dimensione di qualità ma è trasversale a tutte le dimensioni.

Il topic 10 (reputazione) è relativo a quelle recensioni in cui sono descritti molti aspetti della struttura ospedaliera e per definizione risulta trasversale a tutte le dimensioni di qualità. Per esempio, una struttura ospedaliera può avere un'ottima reputazione per quanto riguarda la qualità del personale ma al contempo i tempi d'attesa per prenotare una visita potrebbero essere molto dilatati, portando i clienti a reputare pessime le procedure amministrative.

I topic 11 e 13 (aspetti tangibili 1 e 2) sono relativi alle strutture fisiche e la loro manutenzione come edifici, sale operatorie, spazi comuni, pulizia degli ambienti, infrastrutture e tecnologia. Questi topic si identificano perfettamente nella dimensione di qualità "infrastruttura".

Il topic 12 (accesso/visita) fa riferimento alle visite mediche e al relativo ritiro dei referti nelle strutture ospedaliere. Questo topic si colloca all'interno della dimensione di qualità "procedure amministrative" in quanto non è corrispondente alla qualità del

servizio ricevuto, ma alla fluidità e semplicità delle attività burocratiche da svolgere per accedere alle visite.

Il topic 15 (affidabilità) è una sintesi di tutta l'esperienza sanitaria, per questo motivo rientra nella dimensione di qualità più generica quale è "esperienza di assistenza medica nel complesso".

Come si evince da questa analisi i topic rispecchiano molto bene le dimensioni di qualità elaborate dagli studi in letteratura.

A seguire saranno discussi i risultati elaborati attraverso l'esecuzione dell'algoritmo sul documento documentsTrain (18625 recensioni). Osservando solamente la prevalenza media del topic su tutte le recensioni presente in Tabella 7 è possibile notare come i topic in media con un valore di prevalenza più alta (7%-8%) siano il topic 1 (tempo di risposta), 3 (competenza 1), 6 (cortesia/empatia) e 7 (qualità delle cure). In particolare, i recensori esprimono lamentele e insoddisfazione sul web per quanto riguarda le lunghe attese e la difficoltà incontrata per prenotare una visita specialistica nelle strutture ospedaliere pubbliche. Se da un lato i recensori si lamentano dei tempi eccessivamente dilatati della sanità, dall'altro elogiano la competenza e professionalità dei medici e dello staff sanitario. Collegato al topic "competenza 1" c'è il topic "cortesia/empatia", il più prevalente in assoluto (8,41%) e che fa riferimento a quelle recensioni in cui sono ringraziati i medici e lo staff sanitario per la loro umanità e sensibilità nel prestare servizio.

Il topic "qualità delle cure" contiene parole dispregiative nei confronti del personale sanitario e della qualità delle cure ricevute. In particolare, questo topic è relativo a tutte le recensioni in cui gli utenti si lamentano di una particolare esperienza vissuta durante una visita, un'operazione e soprattutto esprimono giudizi nei confronti del personale sanitario e del loro operato.

In ultima istanza se si osserva la Figura 5, i topic in media più prevalenti come il 9 (racconto dell'esperienza 2), il 12 (accesso/visita) e il 14 (competenza 2), i quali hanno rispettivamente una percentuale di prevalenza media pari a 54,1%, 56,1% e 52,3%, risultano i topic meno assegnati alle recensioni come si può evincere dalla Tabella 8, in quanto sono assegnati rispettivamente a 496, 542 e 447 recensioni. Questo è dovuto

al fatto che, sebbene siano prevalenti in un numero ridotto di recensioni, laddove lo sono hanno in media una percentuale di prevalenza molto alta, ossia la recensione è altamente influenzata da uno di questi topic, mentre gli altri risultano marginali.

Confronto dati tra Nord Italia, Centro Italia e Mezzogiorno

A questo punto si rivela utile confrontare i dati ottenuti dall'elaborazione statistica delle recensioni tra le tre zone geografiche precedentemente individuate, ovvero il Nord Italia, il Centro Italia e il Mezzogiorno. Difatti, sussistono significative differenze regionali in termini di qualità dei servizi sanitari erogati, nello specifico si riscontra un livello inferiore di quest'ultima e dei risultati ottenuti in seguito all'iter diagnostico-terapeutico intrapreso nelle strutture ospedaliere del Mezzogiorno. Inoltre, sono presenti significative differenze intra-regionali, sovente maggiori di quelle inter-regionali (*Studi dell'OCSE sulla Qualità dell'Assistenza Sanitaria: Italia 2014*, 2016).

Il primo dato meritevole di considerazione è relativo alla percentuale media della prevalenza dei topic, come è possibile osservare in Figura 14. Nel Nord Italia, diversamente dal resto del Paese, sono riscontrabili recensioni con una prevalenza media maggiore relativamente ai topic 2 (racconto dell'esperienza), 5 (operazione), 6 (cortesia/empatia), 8 (pronto soccorso), 12 (accesso/visita) e 15 (affidabilità). Di questi, il topic "pronto soccorso" risulta essere quello più prevalente nelle recensioni del Settentrione rispetto alle altre due zone. Il Sistema Sanitario Nazionale prevede 733 strutture ospedaliere pubbliche (*Elenco Aziende sanitarie locali e Strutture di ricovero*), suddivise in 318 al Nord, 147 al Centro e 268 al Mezzogiorno. Nella tabella di seguito riportata (Tabella 23) si illustra il rapporto che sussiste tra il numero di residenti in ognuna delle tre zone geografiche e il numero di strutture ospedaliere presenti in ciascuna di esse.

Zona	Abitanti/ospedali
Nord Italia	87262
Centro Italia	82092
Mezzogiorno	77005

Tabella 23 *Numero di abitanti per ospedale per le tre zone geografiche: Nord Italia, Centro Italia e Mezzogiorno*

I dati mostrano che il valore relativo al Nord Italia è nettamente superiore rispetto alle restanti due zone. Pertanto, una delle possibili spiegazioni al fenomeno che vede una prevalenza decisamente più alta del topic "pronto soccorso" al Nord potrebbe essere proprio il maggior numero di abitanti per ospedale, che comporterebbe pronti soccorsi più sovraffollati e con maggiori problematiche rispetto agli ospedali del Centro e del Mezzogiorno. A supporto di questa ipotesi, le città più grandi e più densamente popolate si trovano al Nord (Milano, Torino, Genova, Bologna) ed è proprio in questi centri che sono denunciate le problematiche appena descritte.

Il Centro Italia ha recensioni con una prevalenza media maggiore rispetto alle restanti due zone relativamente ai topic 1 (tempo di risposta), 4 (ostetricia e ginecologia), 11 (aspetti tangibili) e 14 (competenza 2), pur non evidenziandosi delle discrepanze eccessive rispetto all'andamento nazionale. Non sono presenti, infatti, topic per i quali questa zona ha una prevalenza nettamente minore o maggiore rispetto alle altre due.

Il Mezzogiorno presenta recensioni con una prevalenza media maggiore, in confronto alle restanti due zone relativamente ai topic 3 (competenza 1), 7 (qualità delle cure), 9 (racconto dell'esperienza 2), 10 (reputazione) e 13 (aspetti tangibili 2). In particolare, spicca il topic 7 (qualità delle cure), il quale ha la percentuale di prevalenza più alta in assoluto e pari a 9,60%. Analizzando le parole contenute nel suddetto topic in Appendice C, si nota che questo ha un'accezione negativa poiché contiene termini come "schifo", "vergogna", "male", "zero" e "stella", quest'ultimo utilizzato per spiegare che l'utente ha dato come valutazione della recensione il punteggio minore, ovvero uno. La percentuale alta di prevalenza di questo topic indica che in media le

recensioni del Mezzogiorno sono fortemente negative e testimoniano l'ampio divario di qualità presente tra le regioni meridionali del Paese e il resto dell'Italia.

Secondo l'Istituto Demoskopica il quale ha misurato il livello di efficienza e competitività dell'offerta sanitaria nelle Regioni italiane, meno di un terzo degli italiani (31,7%) si dichiara soddisfatto dei servizi sanitari legati correlati al ricovero ospedaliero, come ad esempio l'assistenza medica e infermieristica, il vitto e i servizi igienici. In particolare le ultime cinque posizioni (Basilicata, Puglia, Sicilia, Campania e Calabria) hanno un livello di soddisfazione molto basso, inferiore al 20% (*Indice di misurazione e valutazione dei sistemi regionali italiani*, 2019).

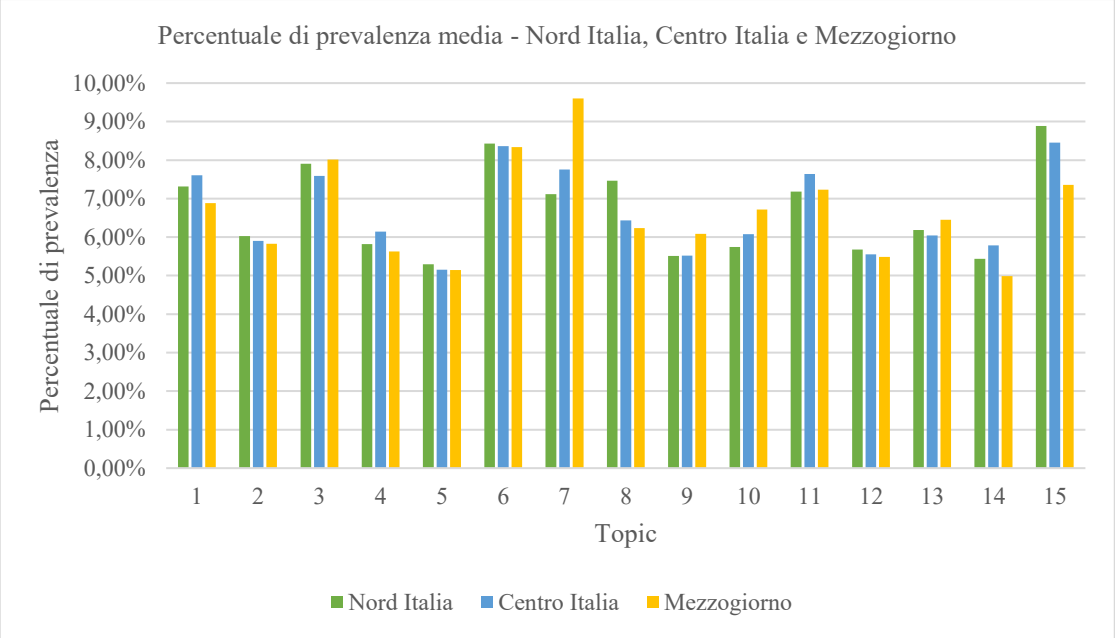


Figura 14 *Confronto per ogni topic tra Nord Italia, Centro Italia e Mezzogiorno sulla percentuale media di prevalenza*

In Figura 15 è possibile osservare per ogni zona geografica la percentuale di recensioni sul totale che non hanno un topic assegnato, ovvero sono risultate confuse. Il Mezzogiorno ha un valore nettamente più alto rispetto alle altre due zone, ciò significa che le recensioni provenienti dalla suddetta area geografica hanno all'interno in media una miscela di topic più omogenea e questo può trovare due spiegazioni. La prima è

che le recensioni contengono al loro interno più topic e quindi l'algoritmo non è in grado di assegnare un solo topic prevalente. La seconda motivazione potrebbe essere che l'algoritmo non riesca ad assegnare i topic alle recensioni e allora tutti i topic hanno una percentuale di prevalenza alta, questo può succedere quando le recensioni non sono attinenti ai topic presenti e l'algoritmo in questo caso utilizza una logica fuzzy, come mostrato nell'esempio al termine del Capitolo I. Questa seconda possibilità risulta essere tuttavia decisamente remota, poiché è praticamente impossibile che così tante recensioni non siano attinenti ai topic individuati, essendo state utilizzate queste stesse recensioni per allenare l'algoritmo a trovare i topic. Pertanto, la motivazione più plausibile sembra essere la prima descritta.

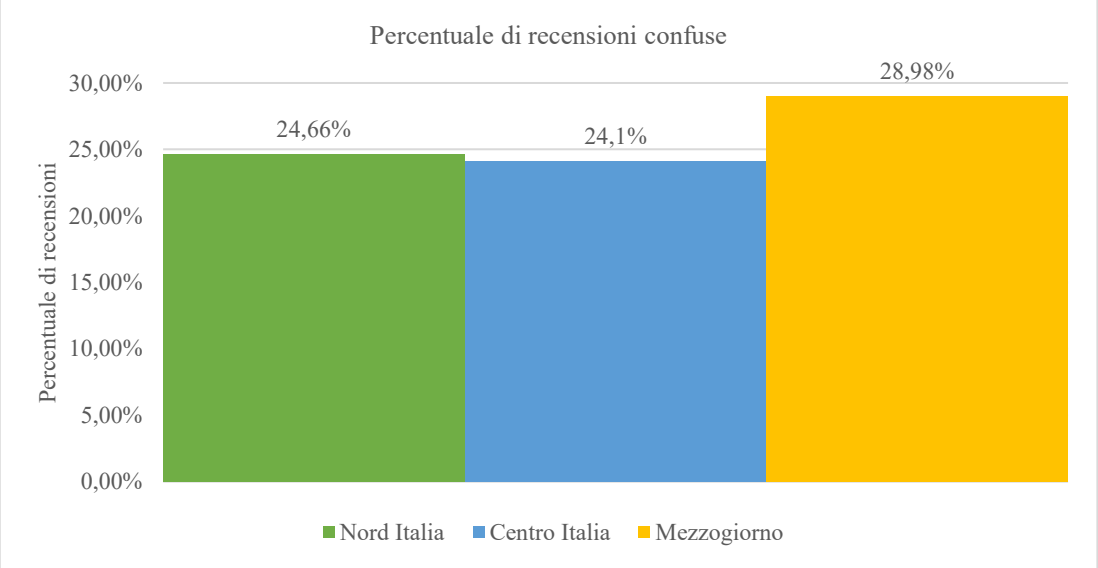


Figura 15 *Percentuale di recensioni confuse divise per zona geografica*

Confronto dati tra le città di Roma, Milano, Napoli e Torino

Sono effettuati ulteriori confronti tra le quattro città più popolose d'Italia, ovvero Roma, Milano, Napoli e Torino.

Il primo confronto è compiuto sulla percentuale di prevalenza media delle recensioni. Come è possibile apprezzare dalla Figura 16, Roma ha una prevalenza media più alta

rispetto alle restanti città relativamente ai topic 4 (ostetricia e ginecologia), 6 (cortesia/empatia), 15 (affidabilità). In particolare, i topic 6 e 15 hanno accezioni positive a giudicare dalle parole contenute in essi (Appendice C), come ad esempio "grazie", "umanità", "gentilezza", "competente", "preparato", "cortese" e "cordiale". Sembra pertanto che gli utenti abbiano una percezione positiva degli ospedali romani. Precisamente, il topic 6 (cortesia/empatia) e il 15 (affidabilità) hanno una prevalenza media rispettivamente pari a 8,56% e 7,79%.

La città di Milano ha una prevalenza media maggiore per quanto riguarda i topic 1 (tempo di risposta), 8 (pronto soccorso) e 12 (accesso/visita). In merito al topic 1 (tempo di risposta), la percentuale di prevalenza media è pari a 9,64%, ovvero la più alta tra le quattro maggiori città. Non è un caso che sia Milano il centro metropolitano con il primato di questo topic, poiché le strutture pubbliche hanno tempi d'attesa delle visite assai dilatati. Infatti, nonostante la Lombardia sia la Regione con il più alto budget nella sanità (18,5 miliardi), le liste d'attesa si rivelano molto lunghe. In particolare, all'ospedale Niguarda che è il più grande d'Europa, una prima visita dermatologica è fissata in media a 450 giorni dalla richiesta del cliente. All'ospedale Sacco, una prima visita gastroenterologica è pianificata in media a 332 giorni. L'ecografia alla mammella effettuata all'ospedale Bassini richiede in media 378 giorni (*Liste d'attesa, anche nella virtuosa Lombardia i conti non tornano. Strutture private: 'Tempi raddoppiati nel 2018' - Il Fatto Quotidiano*).

Il topic 8 (pronto soccorso) è molto più prevalente nelle recensioni degli ospedali milanesi, con una percentuale media pari a 9,39%. In particolare, l'attesa media al pronto soccorso in data 12 gennaio 2020 è stata pari a due ore per un codice minore (verde o bianco). Di prassi, i pazienti in codice rosso (pazienti in potenziale immediato pericolo di vita) sono sottoposti a visita medica specialistica senza attendere il proprio turno, mentre i codici gialli, che rappresentano comunque un'urgenza, devono essere visitati entro i primi quindici minuti dall'accesso al triage. Infine, i codici minori possono accedere alla valutazione medica specialistica solo successivamente alla gestione dei codici rossi e gialli.

Nel 2019 il DEA (dipartimento di emergenza e accettazione) dell'ospedale Niguarda è stato visitato da 95000 persone, il 20% di tutti gli individui che hanno effettuato un accesso ad un pronto soccorso dell'area milanese. Nel suddetto ospedale, i codici gialli attendono in media 26 minuti, mentre i codici verdi e bianchi 90 minuti. I tempi d'attesa sono più contenuti nelle ore notturne e durante la mattina, mentre si allungano nel pomeriggio, la sera e nel fine settimana. All'ospedale San Paolo di Milano nel 2019 l'attesa media per un codice verde è stata pari a 80 minuti la mattina e 130 minuti il pomeriggio. Nella giornata di sabato 12 gennaio 2020 in tutti gli ospedali milanesi ci sono stati 4500 accessi al pronto soccorso, di cui 270 all'ospedale Niguarda, 320 al Policlinico, 240 al Fatebenefratelli, 220 al San Carlo, 180 al San Paolo e 120 al Sacco (*Milano, due ore di attesa in Pronto soccorso: un pomeriggio in ospedale - Cronaca - ilgiorno.it*).

Il topic 12 (accesso/visita) è relativo a quelle recensioni che trattano il tema delle visite specialistiche e spesso sono consigli che gli utenti danno alla community in base alla propria esperienza vissuta.

La città di Napoli ha una prevalenza media maggiore relativamente ai topic 3 (competenza 1), 5 (operazione), 9 (racconto dell'esperienza 2), 10 (reputazione), 11 (aspetti tangibili) e 13 (aspetti tangibili 2). Gli ultimi tre topic citati sono collegati tra loro. In particolare, il topic reputazione è relativo ai commenti e alle recensioni all'interno dei quali sono effettuati dei paragoni con altre strutture sanitarie e si descrive la situazione generale della sanità italiana. I topic 11 e 13 (aspetti tangibili 1 e 2) sono relativi a quelle recensioni che hanno come argomento principale le strutture ospedaliere, gli edifici, la pulizia, i reparti, il parcheggio, l'organizzazione e tutti gli aspetti di contorno che non concernono specificatamente l'attività sanitaria, ma che rivestono un ruolo importante per gli utenti. Ciò è testimoniato dal fatto, in particolare per il topic 11, che la prevalenza è molto alta, pari a 8,98%. Il topic "aspetti tangibili 1" ha un'accezione più positiva poiché contiene parole come "eccellenza", "grande", "ottimo". Al contrario, il topic "aspetti tangibili 2" presenta un senso più negativo a causa di alcuni termini contenuti nel topic come "attese", "lunghe", "mancanza" e "pessima". Nello specifico, prendendo come periodo di riferimento dicembre 2019 e gennaio 2020, ci sono stati alcuni episodi spiacevoli accaduti in alcuni ospedali

napoletani. Ad esempio, il 7 gennaio 2020 sono state riscontrate irregolarità nei materiali utilizzati per il consumo e il confezionamento delle pietanze nell'ospedale Cardarelli, a seguito di alcune segnalazioni da parte dei pazienti (*Napoli: insetti nel cibo dei pazienti, l'Asl manda gli ispettori in ospedale - Il Mattino.it*). Un altro esempio significativo relativo al già nominato ospedale Cardarelli è che in concomitanza con il picco stagionale d'influenza (seconda settimana di gennaio), gli accessi giornalieri al pronto soccorso hanno raggiunto quota 250. In questa situazione, data la saturazione dei posti letto nei reparti, alcuni pazienti hanno trascorso parte della degenza su barelle stazionate nei corridoi. (*Cardarelli, riecco il caos: i corridoi dell'ospedale pieni di barelle - Il Mattino.it*).

La città di Torino ha una prevalenza maggiore rispetto alle altre per quanto riguarda i topic 2 (racconto dell'esperienza 1) e 7 (qualità delle cure). Il topic "racconto dell'esperienza" è relativo a un racconto di un'esperienza in ospedale, che può essere una visita, un'operazione o un accesso al pronto soccorso. Il topic "qualità delle cure" è simile al topic 2, tuttavia presenta un'accezione più negativa in quanto sono individuabili le parole "zero", "vergogna", "schifo" e "stella". È importante menzionare che anche il topic 8 (pronto soccorso) ha un indice di prevalenza molto alto pari a 8,91%.

Secondo il precedentemente citato report dell'istituto Demoskopica, il Piemonte è annoverata tra le migliori regioni secondo l'indice di performance sanitaria (*Indice di misurazione e valutazione dei sistemi regionali italiani*, 2019). Un esempio virtuoso in questo senso è rappresentato dall'ospedale Mauriziano di Torino, il quale ha adottato delle contromisure per decongestionare il pronto soccorso affollato. Nello specifico, i pazienti anziani sono seguiti a casa al fine di facilitare la gestione dei pazienti nel pronto soccorso. A causa dell'aumento costante della longevità, gli anziani costituiscono il 40% dei pazienti che giungono al DEA e di questi il 75% è ricoverato. Nel 2018 ci sono stati 23901 accessi di pazienti sopra i 65 anni (il 41% del totale) e 5975 di questi hanno portato al 25,5% dei ricoveri totali. L'obiettivo dell'iniziativa è ridurre il tempo di permanenza e i ritorni al pronto soccorso. Dal punto di vista pratico è attivata direttamente in pronto soccorso una struttura ambulatoriale alla quale i pazienti sono indirizzati su indicazione del medico di urgenza, in base alla valutazione

del rischio di "ritorno non programmato". In breve, il decorso del paziente è monitorato dopo la dimissione al fine di individuare le criticità che potrebbero condurlo nuovamente in pronto soccorso e proponendo correttivi sul fronte della terapia e dell'assistenza (*Mauriziano, anziani seguiti a casa: così si svuota il pronto soccorso - La Stampa*).

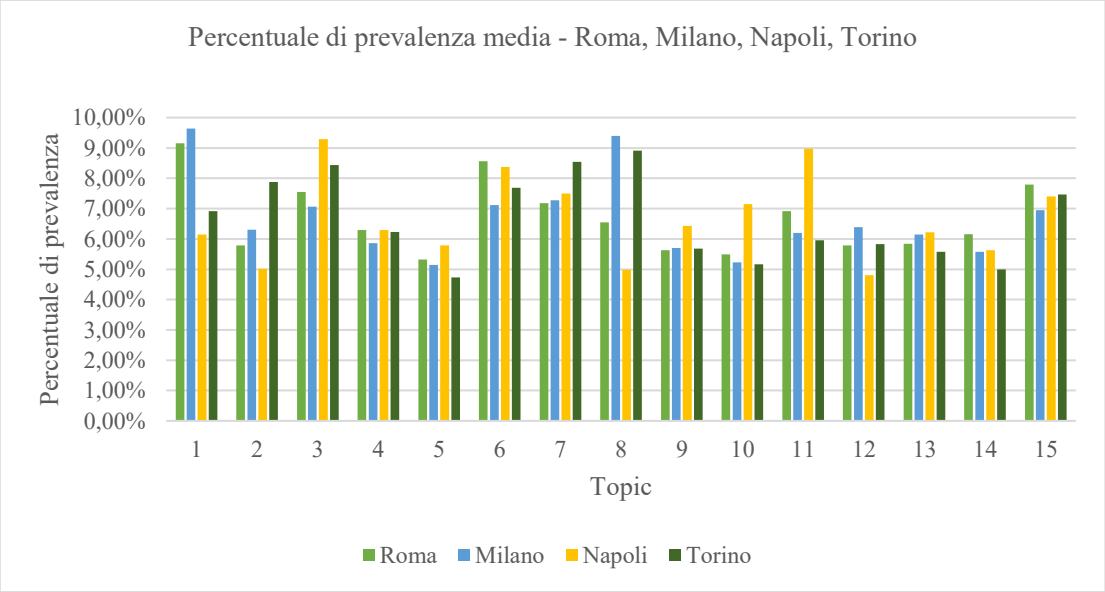


Figura 16 *Confronto per ogni topic tra Roma, Milano, Napoli e Torino sulla percentuale media di prevalenza*

In Figura 17 è illustrata per ogni città la percentuale di recensioni confuse sul totale. Si noti come Torino sia il centro abitato con meno recensioni confuse, per la quale l'algoritmo ha saputo meglio assegnare i topic. La città di Milano ha la percentuale di recensioni confuse maggiori.

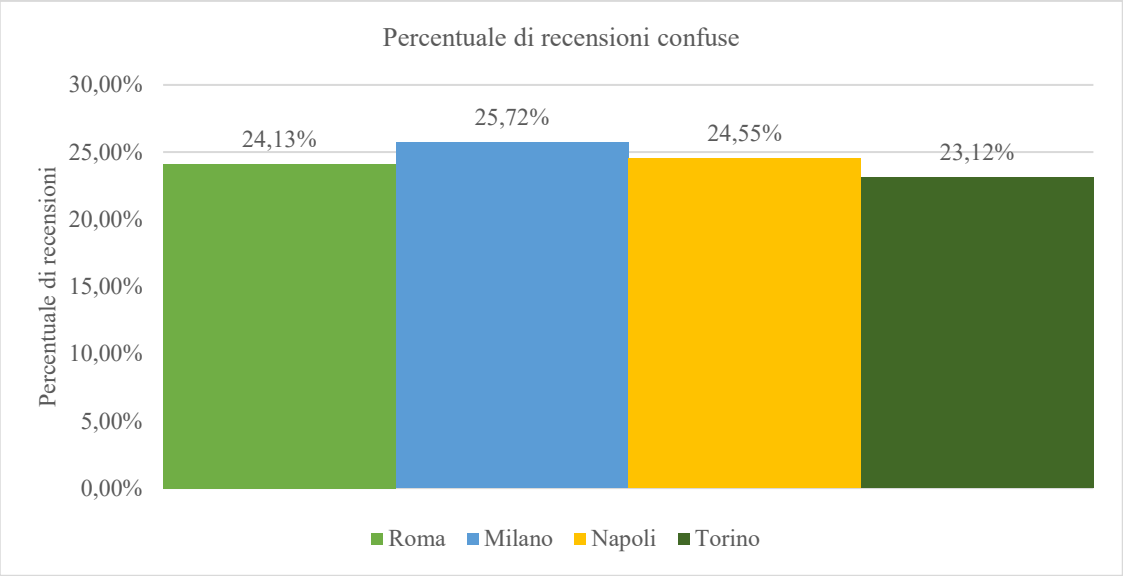


Figura 17 *Percentuale di recensioni confuse divise per le quattro principali città italiane (Roma, Milano, Napoli, Torino)*

Limitazioni dello studio e sviluppi futuri

L'algoritmo LDA è ampiamente riconosciuto in letteratura come strumento per estrapolare le informazioni dagli user generated contents e utilizzarle per migliorare il servizio offerto ai clienti (Guo, Barnes and Jia, 2017). Questo studio cerca di adoperare le tecniche di data mining per comprendere le percezioni dei pazienti e confrontarle con le dimensioni di qualità del settore sanitario individuate in letteratura.

La prima difficoltà incontrata nello sviluppo dell'elaborato è stata l'applicazione dell'algoritmo LDA. La stima dei parametri si è rivelata di lunga esecuzione e non molto accurata, in quanto è stata effettuata un'ottimizzazione discreta e non per mezzo del metodo variational expectation maximization (EM), il quale è molto oneroso in termini di potenza di calcolo utilizzata. L'esecuzione dell'ottimizzazione discreta comporta il forte rischio che i parametri scelti siano relativi a un punto di ottimo locale e non globale, con la conseguenza che ci siano altri valori di α, β e NumTopics che restituiscono un valore di perplexity minore.

Un problema molto importante è che il campione contiene recensioni di lingua italiana, mentre le funzioni utilizzate per raffinare il testo non supportano questa lingua e non

sono quindi state utilizzate nell'esecuzione dell'algoritmo; ciò ha comportato eliminare manualmente specifiche parole dalle recensioni. Nonostante questo accorgimento è possibile notare in Appendice C che alcuni termini sono presenti sia in forma singolare che plurale, come ad esempio i termini "struttura" e "strutture" contenuti nel topic 10 (reputazione). Per eliminare completamente questo problema si sarebbe potuta applicare la funzione *NormalizeWords(.)* come spiegato nel Capitolo II.

In secondo luogo, una limitazione dello studio si è incontrata nell'estrazione dei topic e nella successiva assegnazione di questi ultimi alle recensioni, in quanto in letteratura non è presente nessun metodo universalmente riconosciuto per eseguire questa operazione. Si è quindi proceduto nel definire la soglia θ=6,667. Analogamente, per quanto riguarda l'etichettatura dei topic in letteratura non è descritto nessun metodo per eseguire rigorosamente questa operazione. Pertanto, l'etichettatura risulta arbitraria e dipende fortemente dal soggetto che la esegue.

L'applicazione dell'algoritmo LDA ad intervalli di tempo regolari, potrebbe evidenziare variazioni nelle percezioni degli utenti e permettere al management sanitario di intercettare nuovi bisogni, individuando con più precisione le criticità del sistema e superandoli con l'adozione di nuove scelte gestionali.

Il naturale sviluppo futuro di questo studio è l'applicazione dell'algoritmo LDA a un campione più ampio di user generated content di carattere sanitario, nello specifico recensioni di ospedali provenienti da tutta Europa, per comprendere se esistono differenze intrastatali e interstatali e individuando i punti di forza e le debolezze al fine di migliorare la qualità del servizio offerto al cliente.

Capitolo V

Conclusioni

In questa tesi, si è sviluppata una tecnica di text mining per la classificazione semantica di documenti per mezzo dell'algoritmo LDA. In particolare, si è applicato l'algoritmo a un campione di recensioni estratte dalla piattaforma Google sulle strutture ospedaliere pubbliche italiane.

Successivamente si è proceduto con la validazione dell'algoritmo al fine di comprendere il grado di accuratezza del suddetto rispetto al mondo reale. I risultati di questa fase si sono dimostrati in linea con gli altri studi presenti in letteratura per quanto riguarda gli indicatori di Recall e Accuracy, mentre si sono rivelati leggermente inferiori per quanto riguarda gli indicatori di Precision e F-measure. Nello specifico l'Accuracy è pari a 60,00%, valore giudicato in base allo studio di Nassirtoussi et al. (2014) come un buon risultato di validazione in quanto è superiore al 55,00%.

Successivamente è stata effettuata un'analisi geografica delle recensioni del campione, in particolare si è cercato di comprendere quali differenze di prevalenza dei topic ci sono tra le recensioni delle strutture ospedaliere situate in diversi luoghi.

In primo luogo, sono state confrontate le recensioni provenienti dal Nord Italia, Centro Italia e Mezzogiorno e successivamente le recensioni dei quattro maggiori centri urbani italiani quali Roma, Milano, Napoli e Torino. Dai risultati è emerso che non esistono significative differenze di prevalenza per la maggior parte dei topic tra le differenti zone geografiche italiane, fatta eccezione per due: il 7 (qualità delle cure), il quale è molto più presente nelle recensioni del Mezzogiorno e denota un livello di qualità del servizio sanitario inferiore rispetto alle restanti due zone e il topic 2 (pronto soccorso), il quale è più prevalente nelle recensioni del Nord Italia. Al contrario, nel confronto effettuato tra le quattro città, esistono differenze nella prevalenza dei topic.

In conclusione, le tecniche di text mining risultano un valido strumento per estrarre valore da testi, in particolare dagli user generated content e sono sempre più utilizzate negli ambiti più disparati come il marketing e nello specifico nelle ricerche di mercato, nelle scienze politiche e in ambito accademico.

Appendice A

Fattori determinanti della qualità dei servizi (Parasuraman, Zeithaml and Berry, 1985)

Determinanti della qualità	Descrizione
Affidabilità	Coerenza tra le performance e la fiducia. Significa che la società eroga un buon servizio la prima volta e mantiene le promesse.
Tempo di risposta	La volontà e la prontezza a elargire il servizio.
Competenza	Possesso delle capacità e conoscenza richieste per fornire il servizio.
Accesso	Facilità di contatto e approccio al servizio.
Cortesia	Gentilezza, rispetto ed empatia del personale.
Comunicazione	Trasmissione delle informazioni attraverso l'utilizzo di un linguaggio facilmente comprensibile dal cliente. Adeguamento del registro linguistico in base alla tipologia del cliente.
Credibilità	Fiducia da parte del cliente nei confronti dell'erogatore del servizio. Il fornitore deve avere a cuore gli interessi del cliente.
Sicurezza	L'assenza del pericolo, rischio e dubbio e include la sicurezza fisica, sicurezza finanziaria e la riservatezza dei dati.
Capire/conoscere il cliente	Sforzo effettuato nel comprendere i bisogni e requisiti del cliente.
Aspetti tangibili	Gli oggetti fisici che fanno parte del servizio.

Appendice B

Fattori determinanti della qualità dei servizi sanitari (Duggirala, Rajendran and Anantharaman, 2008)

Determinanti della qualità	Descrizione
Infrastruttura	Questa dimensione esamina la disponibilità e la qualità delle strutture fisiche.
Qualità del personale	La prestazione fornita dai medici, infermieri e staff di supporto è una determinante importante della soddisfazione del paziente. Le loro abilità, esperienza, reattività sono caratteristiche proprie di questa dimensione.
Processo di assistenza clinica	L'esperienza del paziente con il processo clinico (trattamento e risultato) è coperta da questa dimensione.
Procedure amministrative	Questa dimensione esamina l'esperienza del paziente con la burocrazia dell'ospedale.
Indicatori di sicurezza	La percezione del cliente delle misure di sicurezza adottate durante i differenti stadi di permanenza del paziente in ospedale.
Esperienza di assistenza clinica nel suo complesso	Questa dimensione è relativa all'esperienza totale del paziente. Riguarda la probabilità che un paziente ha di consigliare l'ospedale a un membro della propria famiglia o a un conoscente, la differenza tra le cure attese e le cure ricevute, la qualità delle cure paragonate ad altri ospedali simili.
Responsabilità sociale	Il livello di responsabilità sociale e contribuzione dell'ospedale alla società, in termini di parità di trattamento dei pazienti, principi etici, mantenimento della privacy e riservatezza.

Appendice C

Lista dei topic

Topic	Parola	Peso	Topic	Parola	Peso
1 Tempo di risposta	visita	0.044628	2 Racconto dell' Esperienza 1	soccorso	0.02433
	cup	0.01702		casa	0.012782
	attesa	0.013809		notte	0.011578
	minuti	0.012694		ospedale	0.011468
	telefono	0.012356		male	0.011197
	servizio	0.011422		dolori	0.010969
	numero	0.011253		dolore	0.0099872
	appuntamento	0.010609		febbre	0.0068955
	visite	0.010356		medico	0.0064166
	prenotare	0.010139		giorno	0.0063166
	prenotazione	0.0098047		figlio	0.0056527
	sportello	0.0083954		esperienza	0.0055601
	pagare	0.0083796		figlia	0.005415
	nessuno	0.0083121		nessuno	0.0053737
	risponde	0.0080842		turno	0.0049606

3 Competenza 1	medici	0.082415	4 Ostetricia e ginecologia	reparto	0.046229
	infermieri	0.055914		ospedale	0.023319
	reparto	0.046429		ginecologia	0.022654
	gentili	0.035476		figlio	0.019304
	dottori	0.031833		bambino	0.016938
	disponibili	0.028928		figlia	0.014099
	professionali	0.028016		parto	0.013815
	bravi	0.026023		esperienza	0.013132
	competenti	0.021123		ostetricia	0.012813
	ottimo	0.019993		pediatria	0.01189
	ospedale	0.018659		partorito	0.01046
	personale	0.017042		infermiere	0.0086634
	preparati	0.015532		moglie	0.0086528
	soccorso	0.012452		ostetriche	0.0085856
	pazienti	0.010097		gravidanza	0.0076109

5 Operazione	intervento	0.034703	6 Cortesia / Empatia	grazie	0.056417
	reparto	0.021826		reparto	0.040953
	operato	0.018303		dott	0.022628
	ortopedia	0.013185		cuore	0.017963
	chirurgia	0.012686		umanità	0.015127
	day	0.010884		ringraziare	0.014705
	ospedale	0.010755		grande	0.014528
	lintervento	0.0091055		staff	0.01445
	operazione	0.0086602		ringrazio	0.014033
	operata	0.0086458		personale	0.013067
	ricovero	0.0085256		gentilezza	0.012957
	personale	0.0084219		dottor	0.012757
	operatoria	0.0081987		ringraziamento	0.011188
	sala	0.0079743		competenza	0.010795
	mesi	0.0078975		disponibilità	0.009981

7 Qualità delle cure	personale	0.024236	8 Pronto Soccorso	soccorso	0.07749
	pazienti	0.021478		attesa	0.068358
	soccorso	0.018899		tempo	0.03296
	persone	0.018536		codice	0.025654
	medici	0.014638		sala	0.014742
	pessimo	0.010878		persone	0.013521
	lavoro	0.010212		aspettare	0.013291
	gente	0.0091282		medico	0.011604
	male	0.0088681		verde	0.011584
	ospedale	0.0086986		visita	0.011246
	infermieri	0.0086257		triage	0.0097025
	zero	0.0078951		dattesa	0.0095433
	vergogna	0.0078136		visitato	0.0079396
	schifo	0.007804		bianco	0.0072847
	stella	0.0077583		accettazione	0.0071316

9 Racconto dell' esperienza 2	reparto	0.033033	10 Reputazione	personale	0.023033
	padre	0.016444		struttura	0.022951
	terapia	0.013082		sanità	0.021737
	ricoverato	0.011401		servizio	0.014119
	medicina	0.011001		medico	0.01028
	paziente	0.010505		paziente	0.0099451
	medici	0.010489		strutture	0.0085912
	esperienza	0.0093345		servizi	0.0082364
	vita	0.0084713		ospedale	0.0081665
	mamma	0.0076211		cortesia	0.0066183
	cura	0.0071657		esempio	0.006356
	ricoverata	0.0071041		sanitario	0.0060429
	ricovero	0.0065873		pubblica	0.0060167
	intensiva	0.0064243		italiana	0.0058306
	cure	0.0061739		lavoro	0.0054854

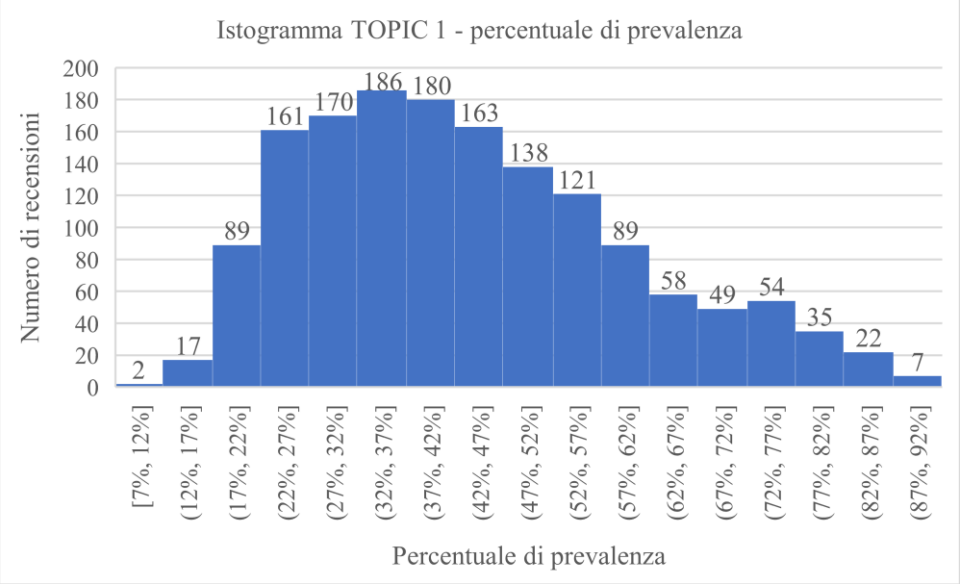
11 Aspetti tangibili 1	ospedale	0.037169	12 Accesso / Visita	medico	0.010545
	struttura	0.034414		visita	0.0096661
	parcheggio	0.017668		dottore	0.008786
	reparti	0.015748		giorno	0.0083642
	grande	0.013235		mattina	0.0078745
	lospedale	0.010431		paziente	0.0075064
	eccellenza	0.010267		frattura	0.0063764
	livello	0.009745		chiesto	0.0059988
	centro	0.0095826		persona	0.0056863
	servizi	0.0089837		mano	0.0054801
	punto	0.0082813		referto	0.0051029
	ospedali	0.0078018		piede	0.0050347
	soccorso	0.0077011		controllo	0.004505
	ottimo	0.0075242		casa	0.0043874
	zona	0.0072277		persone	0.0042884

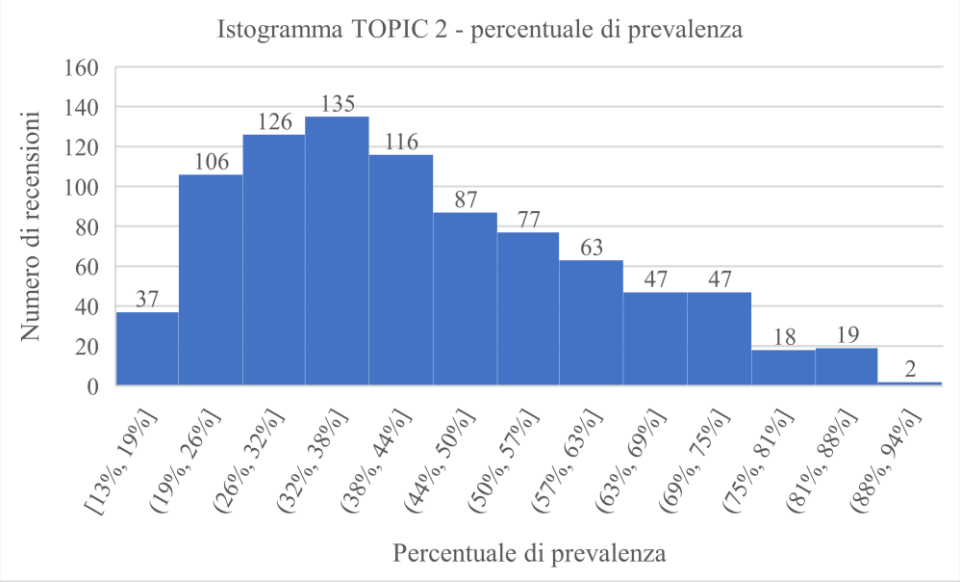
13 Aspetti tangibili 2	ospedale	0.037308	14 Competenza 2	dottoressa	0.013081
	personale	0.019982		problema	0.012596
	reparti	0.017937		esami	0.011596
	attese	0.012614		grazie	0.011494
	soccorso	0.010399		centro	0.0098111
	pazienti	0.0089785		grande	0.0094898
	parcheggio	0.0084613		soccorso	0.0093172
	organizzazione	0.0072988		esperienza	0.0089393
	struttura	0.0072733		visita	0.0084396
	persone	0.007081		medico	0.0082182
	lunghe	0.0068923		problemi	0.0072513
	visite	0.0066757		lavoro	0.0070982
	mancanza	0.0060195		dott	0.0070533
	manca	0.0059668		sangue	0.0069389
	pessima	0.0054585		gentilezza	0.0067621

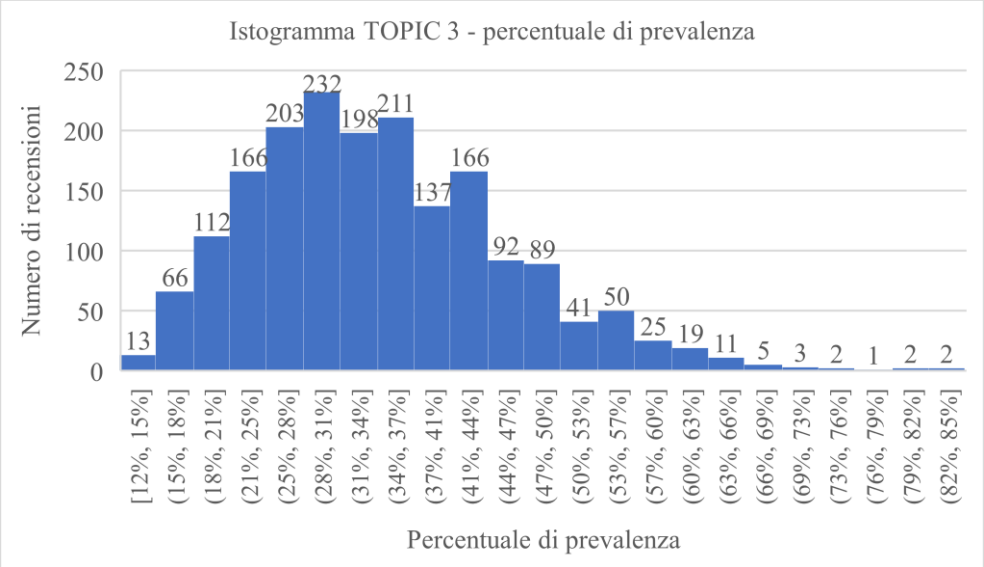
15 Affidabilità	personale	0.16324	
	gentile	0.048432	
	struttura	0.040419	
	disponibile	0.037872	
	ospedale	0.031195	
	professionale	0.029701	
	medico	0.026077	
	competente	0.023934	
	preparato	0.023325	
	ottima	0.022075	
	ottimo	0.020092	
	pulito	0.017872	
	efficiente	0.016745	
	cortese	0.013882	

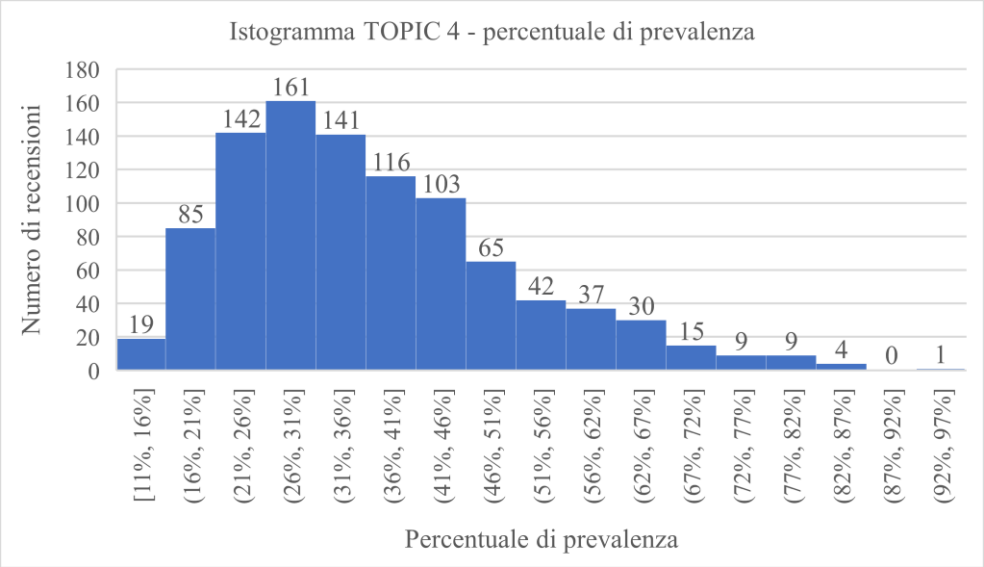
Appendice D

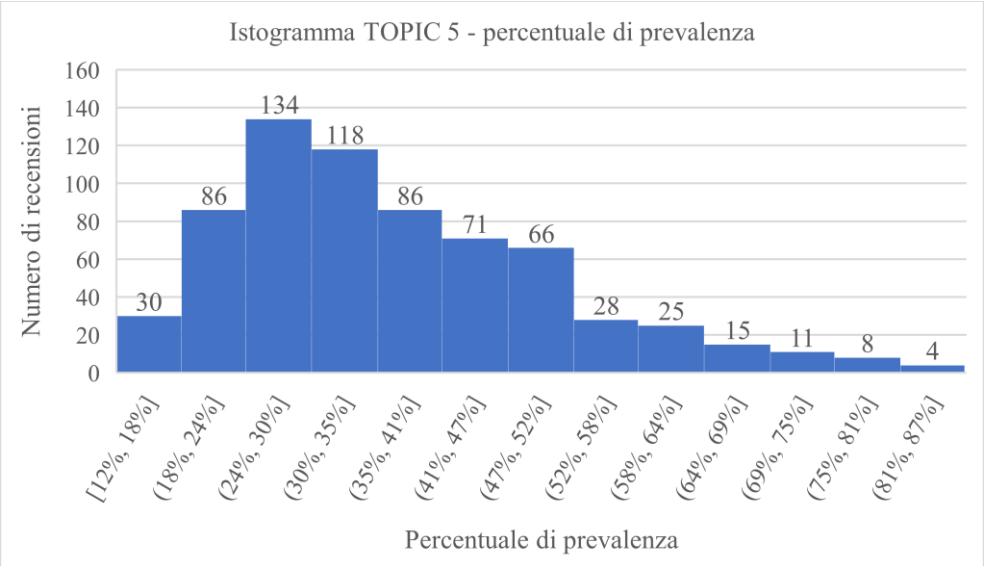
Istogrammi della percentuale di prevalenza per ogni topic, quando il topic i-esimo è prevalente

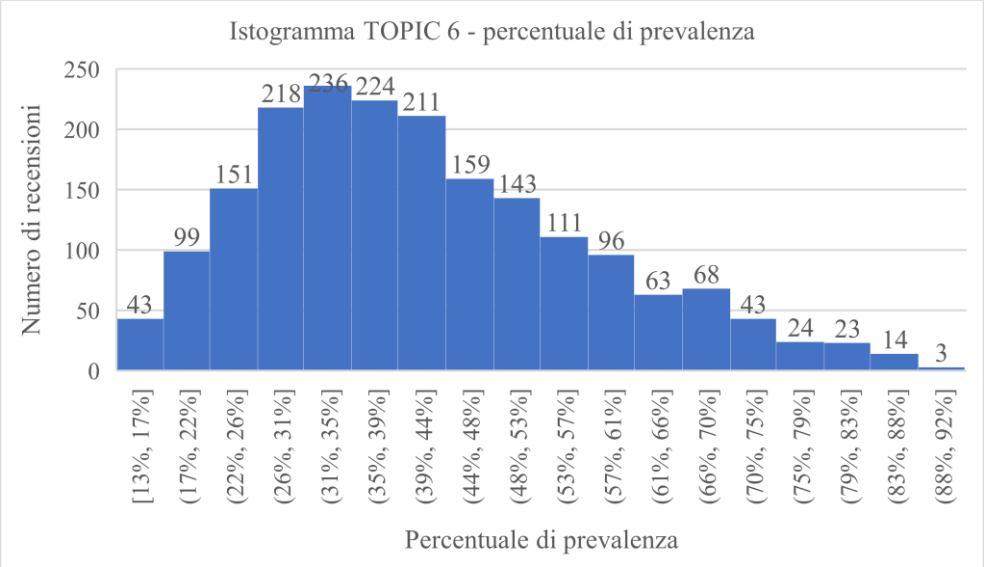


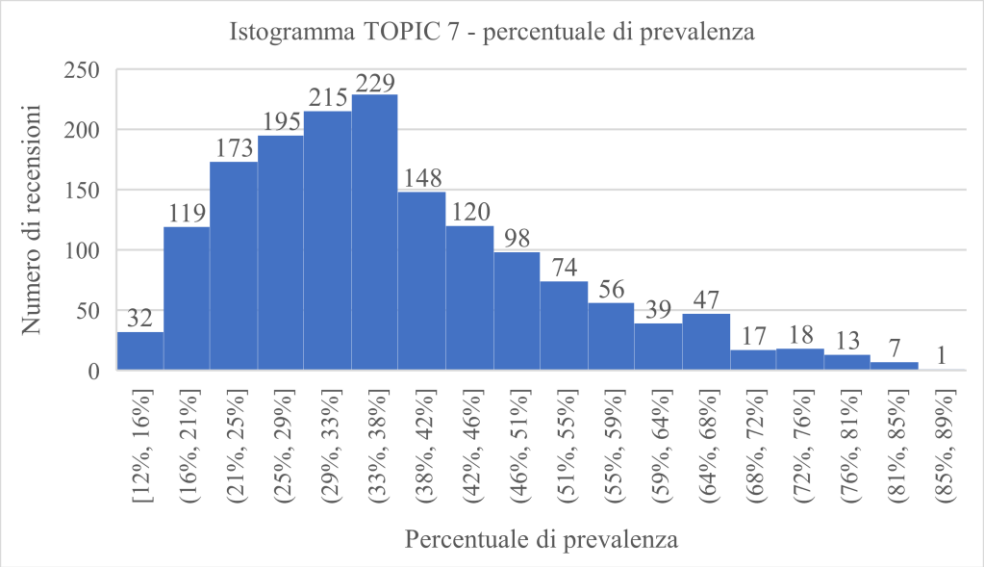


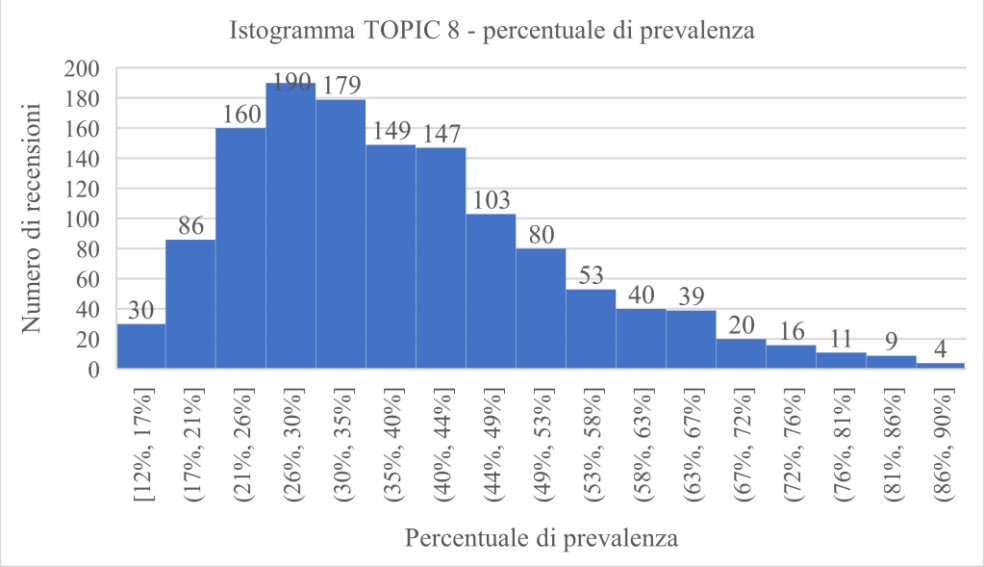

Istogramma TOPIC 3 - percentuale di prevalenza
Numero di recensioni
250
200
150
100
50
0
13
66
112
166
203
232
198
211
137
166
92
89
41
50
25
19
11
5
3
2
1
2
2
[12%, 15%]
(15%, 18%]
(18%, 21%]
(21%, 25%]
(25%, 28%]
(28%, 31%]
(31%, 34%]
(34%, 37%]
(37%, 41%]
(41%, 44%]
(44%, 47%]
(47%, 50%]
(50%, 53%]
(53%, 57%]
(57%, 60%]
(60%, 63%]
(63%, 66%]
(66%, 69%]
(69%, 73%]
(73%, 76%]
(76%, 79%]
(79%, 82%]
(82%, 85%]
Percentuale di prevalenza

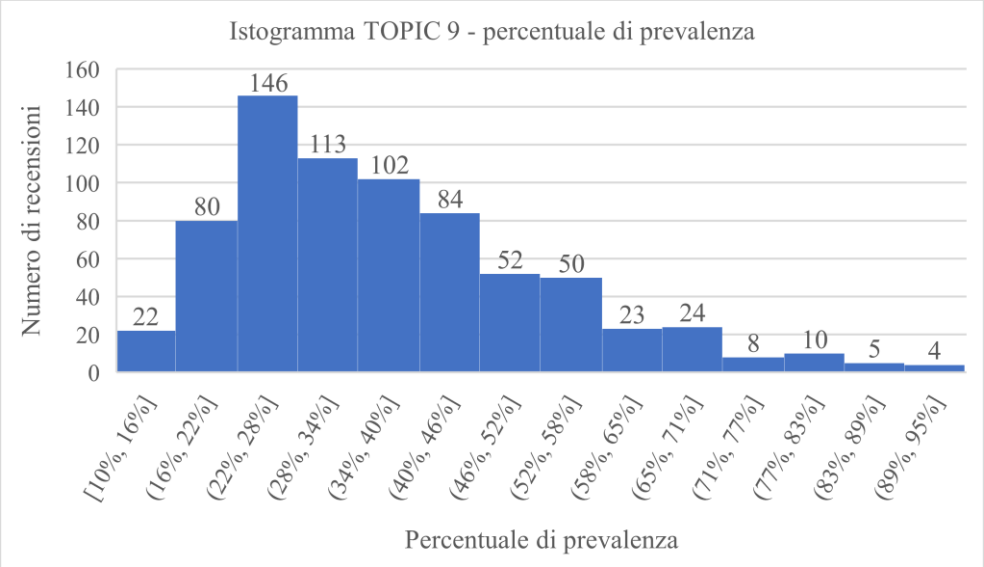

Istogramma TOPIC 4 - percentuale di prevalenza
Numero di recensioni
180
160
140
120
100
80
60
40
20
0
19
85
142
161
141
116
103
65
42
37
30
15
9
9
4
0
1
[11%, 16%]
(16%, 21%]
(21%, 26%]
(26%, 31%]
(31%, 36%]
(36%, 41%]
(41%, 46%]
(46%, 51%]
(51%, 56%]
(56%, 62%]
(62%, 67%]
(67%, 72%]
(72%, 77%]
(77%, 82%]
(82%, 87%]
(87%, 92%]
(92%, 97%]
Percentuale di prevalenza

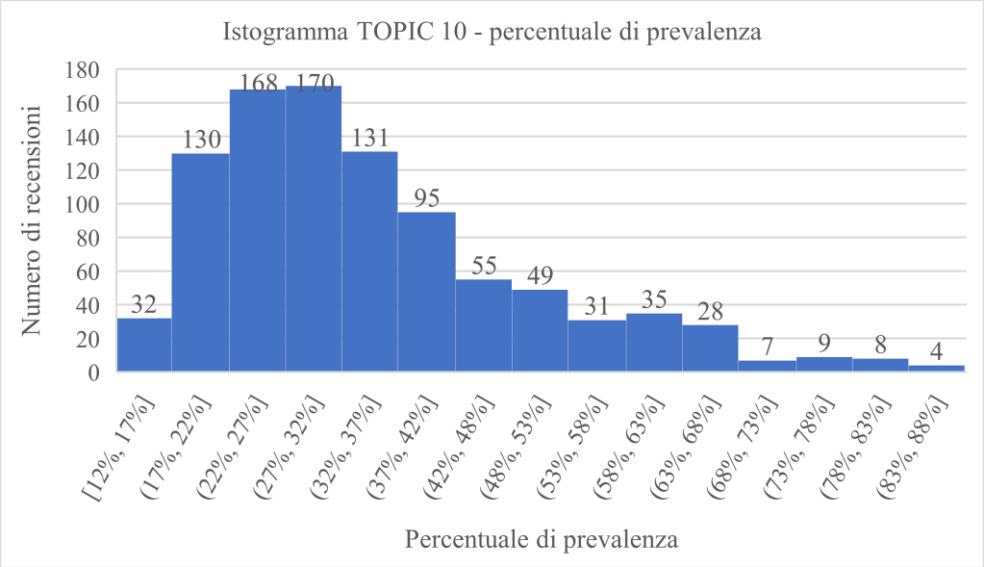

Istogramma TOPIC 5 - percentuale di prevalenza
Numero di recensioni
160
140
120
100
80
60
40
20
0
30
86
134
118
86
71
66
28
25
15
11
8
4
[12%, 18%]
(18%, 24%]
(24%, 30%]
(30%, 35%]
(35%, 41%]
(41%, 47%]
(47%, 52%]
(52%, 58%]
(58%, 64%]
(64%, 69%]
(69%, 75%]
(75%, 81%]
(81%, 87%]
Percentuale di prevalenza

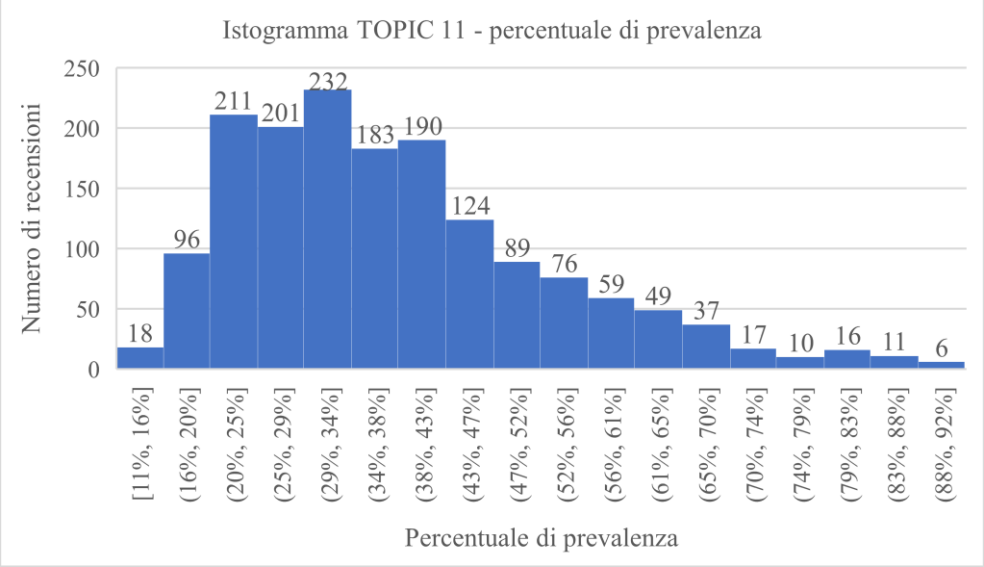

Istogramma TOPIC 6 - percentuale di prevalenza
Numero di recensioni
250
200
150
100
50
0
43
99
151
218
236
224
211
159
143
111
96
63
68
43
24
23
14
3
[13%, 17%]
(17%, 22%]
(22%, 26%]
(26%, 31%]
(31%, 35%]
(35%, 39%]
(39%, 44%]
(44%, 48%]
(48%, 53%]
(53%, 57%]
(57%, 61%]
(61%, 66%]
(66%, 70%]
(70%, 75%]
(75%, 79%]
(79%, 83%]
(83%, 88%]
(88%, 92%]
Percentuale di prevalenza

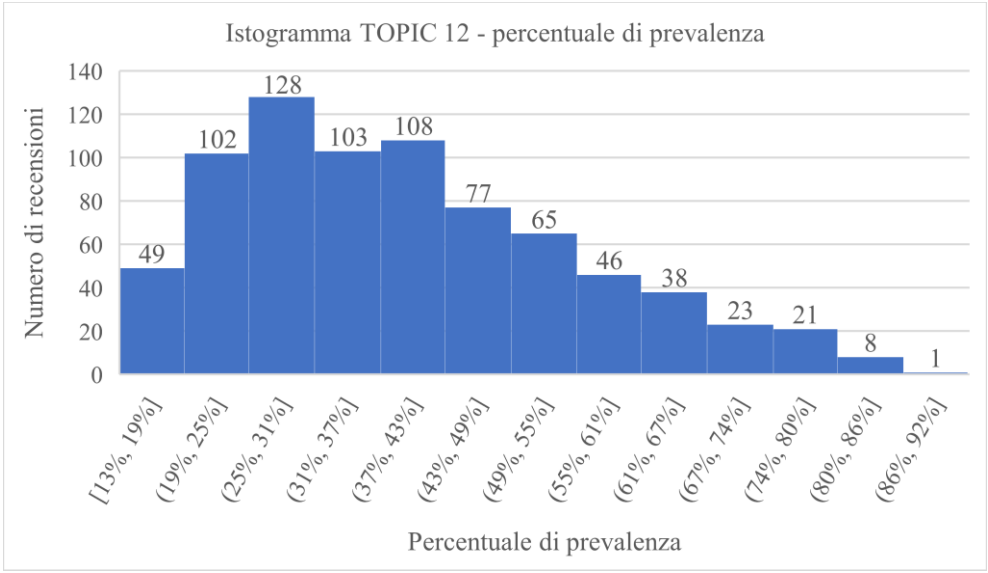

Istogramma TOPIC 7 - percentuale di prevalenza
Numero di recensioni
250
200
150
100
50
0
32
119
173
195
215
229
148
120
98
74
56
39
47
17
18
13
7
1
[12%, 16%]
(16%, 21%]
(21%, 25%]
(25%, 29%]
(29%, 33%]
(33%, 38%]
(38%, 42%]
(42%, 46%]
(46%, 51%]
(51%, 55%]
(55%, 59%]
(59%, 64%]
(64%, 68%]
(68%, 72%]
(72%, 76%]
(76%, 81%]
(81%, 85%]
(85%, 89%]
Percentuale di prevalenza

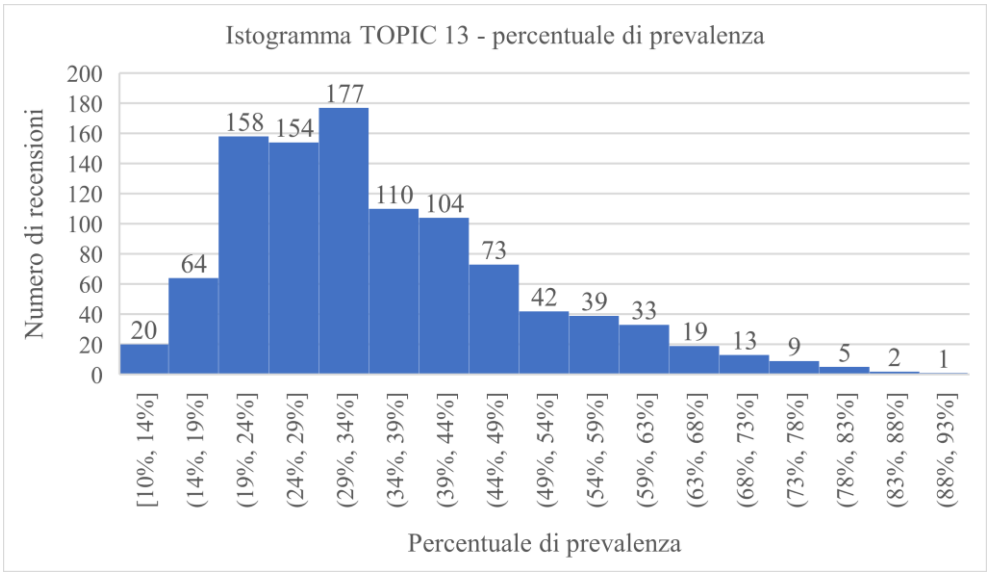

Istogramma TOPIC 8 - percentuale di prevalenza
Numero di recensioni
200
180
160
140
120
100
80
60
40
20
0
30
86
160
190
179
149
147
103
80
53
40
39
20
16
11
9
4
[12%, 17%]
(17%, 21%]
(21%, 26%]
(26%, 30%]
(30%, 35%]
(35%, 40%]
(40%, 44%]
(44%, 49%]
(49%, 53%]
(53%, 58%]
(58%, 63%]
(63%, 67%]
(67%, 72%]
(72%, 76%]
(76%, 81%]
(81%, 86%]
(86%, 90%]
Percentuale di prevalenza

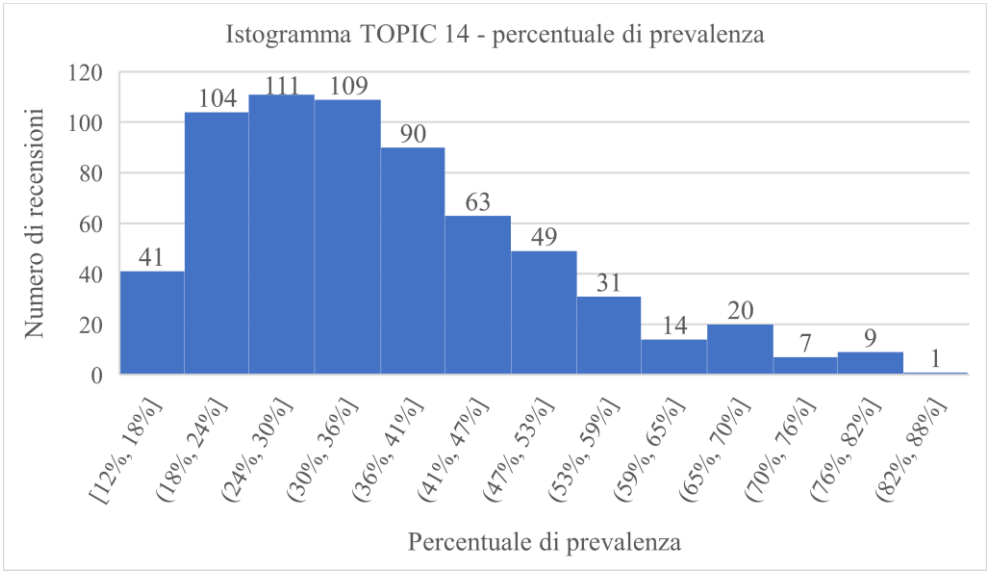

Istogramma TOPIC 9 - percentuale di prevalenza
Numero di recensioni
160
140
120
100
80
60
40
20
0
22
80
146
113
102
84
52
50
23
24
8
10
5
4
[10%, 16%]
(16%, 22%]
(22%, 28%]
(28%, 34%]
(34%, 40%]
(40%, 46%]
(46%, 52%]
(52%, 58%]
(58%, 65%]
(65%, 71%]
(71%, 77%]
(77%, 83%]
(83%, 89%]
(89%, 95%]
Percentuale di prevalenza

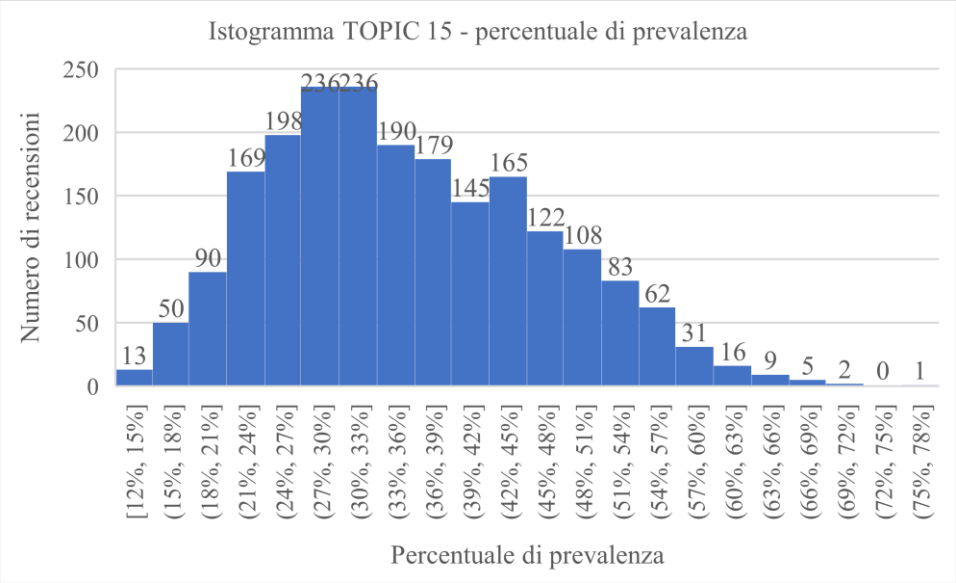

Istogramma TOPIC 10 - percentuale di prevalenza
Numero di recensioni
180
160
140
120
100
80
60
40
20
0
32
130
168
170
131
95
55
49
31
35
28
7
9
8
4
[12%, 17%]
(17%, 22%]
(22%, 27%]
(27%, 32%]
(32%, 37%]
(37%, 42%]
(42%, 48%]
(48%, 53%]
(53%, 58%]
(58%, 63%]
(63%, 68%]
(68%, 73%]
(73%, 78%]
(78%, 83%]
(83%, 88%]
Percentuale di prevalenza


Istogramma TOPIC 11 - percentuale di prevalenza
Numero di recensioni
250
200
150
100
50
0
18
96
211
201
232
183
190
124
89
76
59
49
37
17
10
16
11
6
[11%, 16%]
(16%, 20%]
(20%, 25%]
(25%, 29%]
(29%, 34%]
(34%, 38%]
(38%, 43%]
(43%, 47%]
(47%, 52%]
(52%, 56%]
(56%, 61%]
(61%, 65%]
(65%, 70%]
(70%, 74%]
(74%, 79%]
(79%, 83%]
(83%, 88%]
(88%, 92%]
Percentuale di prevalenza


Istogramma TOPIC 12 - percentuale di prevalenza
Numero di recensioni
140
120
100
80
60
40
20
0
49
102
128
103
108
77
65
46
38
23
21
8
1
[13%, 19%]
(19%, 25%]
(25%, 31%]
(31%, 37%]
(37%, 43%]
(43%, 49%]
(49%, 55%]
(55%, 61%]
(61%, 67%]
(67%, 74%]
(74%, 80%]
(80%, 86%]
(86%, 92%]
Percentuale di prevalenza


Istogramma TOPIC 13 - percentuale di prevalenza
Numero di recensioni
200
180
160
140
120
100
80
60
40
20
0
20
64
158
154
177
110
104
73
42
39
33
19
13
9
5
2
1
[10%, 14%]
(14%, 19%]
(19%, 24%]
(24%, 29%]
(29%, 34%]
(34%, 39%]
(39%, 44%]
(44%, 49%]
(49%, 54%]
(54%, 59%]
(59%, 63%]
(63%, 68%]
(68%, 73%]
(73%, 78%]
(78%, 83%]
(83%, 88%]
(88%, 93%]
Percentuale di prevalenza


Istogramma TOPIC 14 - percentuale di prevalenza
Numero di recensioni
120
100
80
60
40
20
0
41
104
111
109
90
63
49
31
14
20
7
9
1
[12%, 18%]
(18%, 24%]
(24%, 30%]
(30%, 36%]
(36%, 41%]
(41%, 47%]
(47%, 53%]
(53%, 59%]
(59%, 65%]
(65%, 70%]
(70%, 76%]
(76%, 82%]
(82%, 88%]
Percentuale di prevalenza


Istogramma TOPIC 15 - percentuale di prevalenza
Numero di recensioni
250
200
150
100
50
0
13
50
90
169
198
236
236
190
179
145
165
122
108
83
62
31
16
9
5
2
0
1
[12%, 15%]
(15%, 18%]
(18%, 21%]
(21%, 24%]
(24%, 27%]
(27%, 30%]
(30%, 33%]
(33%, 36%]
(36%, 39%]
(39%, 42%]
(42%, 45%]
(45%, 48%]
(48%, 51%]
(51%, 54%]
(54%, 57%]
(57%, 60%]
(60%, 63%]
(63%, 66%]
(66%, 69%]
(69%, 72%]
(72%, 75%]
(75%, 78%]
Percentuale di prevalenza

Appendice E

Codice di Matlab – Algoritmo LDA

```
clear all
close all
rng('default')

%% import  del corpus
filename = "provamatlab_NoEnglish.xlsx";
data = readtable(filename,'TextType','string');
data = data.Testo;

%% Eliminazione  delle righe vuote
idx = strlength(data) == 0;
data(idx,:) = [];

%% preparazione del testo
newDocuments = preprocessText(data);

% rimozione lista di parole (DA FILE parolerimosse.xlsx)
parole=readtable('parolerimosse.xlsx','TextType','string');
newDocuments = removeWords(newDocuments,table2array(parole));

%eliminazione delle recensioni con numero di parole minore o uguale a 2
idx=[];
for i= 1:size(newDocuments)
    if size(newDocuments(i).Vocabulary)<=2
        idx(i,1)=true;
    else
        idx(i,1)=false;
    end
end
idx=logical(idx);
newDocuments(idx,:) = [];

%% Bag of words
cleanBag = bagOfWords(newDocuments);
cleanBag = removeInfrequentWords(cleanBag,2);
[cleanBag,idx] = removeEmptyDocuments(cleanBag);
try
labels(idx) = [];
end

%% Confronto documenti prima (Raw) e dopo il pre-processing (Clean)
rawDocuments = tokenizedDocument(data);
rawBag = bagOfWords(rawDocuments);
numWordsClean = cleanBag.NumWords;
numWordsRaw = rawBag.NumWords;

%% Analisi numero di topic ottimale per LDA
% messa da parte del 10% random per la validazione e ancora 50 recensioni
numDocuments=numel(newDocuments);
cvp=cvpartition(numDocuments,'HoldOut',0.1);
documentsTrain=newDocuments(cvp.training);
documentsValidation=newDocuments(cvp.test);

numDocuments2=numel(documentsTrain);
cvp2=cvpartition(numDocuments2,'Holdout',50);
documentsTrain2=newDocuments(cvp2.training);
documentsValidation2=newDocuments(cvp2.test);
```

```
59
60  %crea un bag of words del documento per la validazione
61  bag = bagOfWords(documentsTrain2);
62
63  %rimuovi le parole che non sono usate più di 2 volte in totale e rimuovi
64  %i documenti non contenuti nel bag of words
65  bag = removeInfrequentWords(bag,2);
66  bag = removeEmptyDocuments(bag)
67
68  %%scelta del numero di topic, ITC e solver che minimizzi la perplexity
69  %e applicazione dell'algoritmo sul 90% dei documenti
70  numTopicsRange =[15] %1...n]
71  TopicConcentrationRange=[4.5] %0.1...1...n]
72  TAB= [];
73  solvers = ["cvb0"] %"avb" "cvb0" "savb"];
74
75  for i = 1:numel(numTopicsRange)
76      numTopics = numTopicsRange(i);
77          for k = 1:numel(TopicConcentrationRange)
78              ITC=TopicConcentrationRange(k);
79              for k=1:numel(solvers)
80                  solver=solvers(k);
81      mdl = fitlda(bag,numTopics, ...
82          'Solver','cvb0', ...
83          'FitTopicProbabilities', false,...
84          'FitTopicConcentration', false, ...
85          'InitialTopicConcentration',ITC ,...
86          'Verbose',0);
87
88      %fa la goodness of fit del 10% che avevamo precedentemente preso dal documento
89      [~,ValidationPerplexity] = logp(mdl,documentsValidation);
90      TAB = [TAB;numTopics ITC ValidationPerplexity]
91
92          end
93      end
94  end
95
96  %salva nel file FinalResult.txt la matrice review/topic
97  csvwrite('FinalResult.txt',mdl.DocumentTopicProbabilities)
98
99  %output grafico tridimensionale [numTopics, ITC, Perplexity]
100 surf(TAB);
101 writematrix(TAB,'OptimizationResults.txt');
102
103 %stampa le 15 parole che hanno più probabilità di appartenere al topic i-esimo
104  for n = 1:numTopics %VEDIAMO SE è GIUSTO
105      tbl = topkwords(mdl,15,n)
106  end
107
108 %Matrice di prevalenza del corpus con il 90% dei documenti e del corpus utilizzato per la validazione
109 topicMixtures = transform(mdl,documentsTrain2);
110 topicMixtures2=transform(mdl,documentsValidation2);
111
```

Codice Matlab – Preprocessamento del testo

```
function documents = preprocessText(textData)

%Conversione del corpus in minuscolo
cleanTextData = lower(textData);

%Cancella punteggiatura
documents = erasePunctuation(cleanTextData);

%Tokenizzazione del corpus
documents= tokenizedDocument(documents);

%Rimozione lista delle stop words (non applicabile in lingua ITA)
documents = removeStopWords(documents);

%Elimina parole con numero di lettere minore o uguale a 2 e
%maggiore e uguale a 15
documents = removeShortWords(documents,2);
documents = removeLongWords(documents,15);

%Normalizzazione del testo (non applicabile a lingua ITA)
documents = addPartOfSpeechDetails(documents);
documents = normalizeWords(documents,'Style','lemma');

end
```

Appendice F

Lista delle parole rimosse

2017	der	miei	resto
2018	detto	milano	riesce
2019	deve	mio	riguarda
abbastanza	devi	modo	san
abbia	devo	molta	sarà
abbiamo	dice	molti	sarebbe
agli	dicendo	molto	seconda
alcune	dico	neanche	sehr
alcuni	dicono	negli	sei
alla	die	nei	sembra
alle	dir	nel	sempre
allo	dire	nella	sente
alta	doctor	nelle	senza
altra	dopo	nemmeno	serum
altre	dove	nessun	sia
altri	dovevo	niente	siamo
altro	dovrebbe	noi	siano
anche	dovrebbero	nome	siete
ancora	dovuta	non	solo
andare	dovuto	nonostante	sono
andata	due	nostro	soprattutto
andate	durante	nulla	sotto
andato	ecc	oggi	spesso
andavate	entrare	ogni	ssa
andiamo	era	oltre	sta
anni	erano	ora	stata
anno	eravate	ore	stati
appena	ero	ormai	stato

appunto	essere	os	stessa
aprile	faccio	paolo	stesso
arrivata	fanno	papà	sto
assolutamente	far	parola	sua
aver	farci	parte	sul
avere	fare	particolare	sulla
avete	fate	per	suo
aveva	fatta	perche	tanta
avevano	fatto	perché	tante
avevo	filum	però	tanto
avuto	fin	piu	terzo
ben	fine	più	tipo
bene	fino	poca	tra
bisogna	forse	poche	tradotto
buon	forte	pochi	tre
buona	fosse	poco	troppa
buono	fuori	poi	troppo
caso	gennaio	portato	trova
cera	già	posso	trovare
cerano	giorni	possono	trovata
cero	giù	poter	trovato
certo	giugno	potete	trovo
che	giuseppe	potrebbe	tuo
chi	gli	prendere	tutta
circa	google	presso	tutte
come	gran	prima	tutti
comunque	hai	probabilmente	tutto
con	hanno	pronto	una
cosa	ich	proprio	und
cose	ieri	può	uno
cosi	inglese	pur	unora

così	inoltre	purtroppo	vado
credo	invece	qua	vari
cui	italia	qualche	varie
da	leu	quale	vengo
dai	loro	quando	vengono
dal	luglio	quanto	venivo
dalla	lui	quasi	veramente
dalle	lunedì	quattro	vero
dare	madre	quel	verso
dato	mai	quella	via
datum	mandato	quello	viene
davvero	meglio	questa	visto
degli	meno	queste	voi
dei	mentre	questi	volta
del	messo	questo	voto
della	metto	qui	wait
delle	mezzo	quindi	war
dello	mia	raggiungere	wurde
den	mie	recata	

Bibliografia

Anwar, S. and Prasad, R. (2018) 'Framework for Future Telemedicine Planning and Infrastructure using 5G Technology', *Wireless Personal Communications*. Springer New York LLC, 100(1), pp. 193–208. doi: 10.1007/s11277-018-5622-8.

Barton, A. (2009) 'Patient Safety and Quality: An Evidence-Based Handbook for Nurses', *AORN Journal*. Wiley-Blackwell, 90(4), pp. 601–602. doi: 10.1016/j.aorn.2009.09.014.

Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2016) 'Variational Inference: A Review for Statisticians'. doi: 10.1080/01621459.2017.1285773.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003) 'Latent Dirichlet allocation', *Journal of Machine Learning Research*, 3(4–5), pp. 993–1022. doi: 10.1016/b978-0-12-411519-4.00006-9.

Cardarelli, riecco il caos: i corridoi dell'ospedale pieni di barelle - Il Mattino.it (no date). Available at: https://www.ilmattino.it/napoli/cronaca/cardarelli_napoli_barelle_corridoi_ospedale-4988300.html (Accessed: 22 January 2020).

Chassin, M. R., Galvin, R. W. and Donaldson, M. S. (1998) 'The urgent need to improve health care quality: Institute of Medicine National Roundtable on Health Care Quality', *Journal of the American Medical Association*, pp. 1000–1005. doi: 10.1001/jama.280.11.1000.

Chua, A. Y. K. and Banerjee, S. (2013) 'Reliability of reviews on the Internet : The case of TripAdvisor'.

Denecke, K. (2008) 'Using SentiWordNet for multilingual sentiment analysis', in *Proceedings - International Conference on Data Engineering*, pp. 507–512. doi: 10.1109/ICDEW.2008.4498370.

Duggirala, M., Rajendran, C. and Anantharaman, R. N. (2008) 'Patient-perceived dimensions of total quality service in healthcare', *Benchmarking*, 15(5), pp. 560–583. doi: 10.1108/14635770810903150.

Elenco Aziende sanitarie locali e Strutture di ricovero (no date). Available at: http://www.salute.gov.it/portale/documentazione/p6_2_8_1_1.jsp?id=13 (Accessed: 19 January 2020).

Franceschini, F. (2001) *Dai prodotti ai servizi : le nuove frontiere per la misura della qualità*. UTET.

Franceschini, F. and Galetto, M. (2011) *Esercizi di ingegneria della qualità*. Celid.

Going lean in the NHS (2007).

Greaves, F. *et al.* (2013) 'Use of sentiment analysis for capturing patient experience from free-text comments posted online.', *Journal of medical Internet research*, 15(11), p. e239. doi: 10.2196/jmir.2721.

Greaves, F. *et al.* (2014) 'Tweets about hospital quality: A mixed methods study', *BMJ Quality and Safety*. BMJ Publishing Group, 23(10), pp. 838–846. doi: 10.1136/bmjqs-2014-002875.

Guo, Y., Barnes, S. J. and Jia, Q. (2017) 'Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation', *Tourism Management*. Elsevier Ltd, 59, pp. 467–483. doi: 10.1016/j.tourman.2016.09.009.

Gupta, A. and Chen, I. (1995) 'Service quality: Implications for management development', *International Journal of Quality & Reliability Management*, pp. 28–35. doi: 10.1108/02656719510093538.

Hillestad, R. *et al.* (2005) 'Can electronic medical record systems transform health care? Potential health benefits, savings, and costs', *Health Affairs*, 24(5), pp. 1103–1117. doi: 10.1377/hlthaff.24.5.1103.

Indice di misurazione e valutazione dei sistemi regionali italiani (2019).

Intuitive Guide to Latent Dirichlet Allocation - Towards Data Science (no date). Available at: https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-latent-dirichlet-allocation-437c81220158 (Accessed: 16 December 2019).

Kasper, W. and Vela, M. (2011) 'Sentiment Analysis for Hotel Reviews', *Proceedings of the Computational Linguistics-Applications Conference*, 231527(March), pp. 45–52.

Khadjeh Nassirtoussi, A. *et al.* (2014) 'Text mining for market prediction: A systematic review', *Expert Systems with Applications*. Elsevier Ltd, pp. 7653–7670. doi: 10.1016/j.eswa.2014.06.009.

Korfiatis, N., García-Bariocanal, E. and Sánchez-Alonso, S. (2012) 'Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content', *Electronic Commerce Research and Applications*, 11(3), pp. 205–217. doi: 10.1016/j.elerap.2011.10.003.

Kruse, C. S. *et al.* (2017) 'Cybersecurity in healthcare: A systematic review of modern threats and trends', *Technology and Health Care*. IOS Press, pp. 1–10. doi: 10.3233/THC-161263.

Kusumasondjaja, S., Shanka, T. and Marchegiani, C. (2012) 'Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence', *Journal of Vacation Marketing*, 18(3), pp. 185–195. doi: 10.1177/1356766712449365.

Legge 23 dicembre 1978, n.833, in materia di 'Istituzione del Servizio Sanitario Nazionale'. (1978).

Legge 7 agosto 1990, n. 241, in materia di 'Nuove norme sul procedimento amministrativo.' (1990).

Liste d'attesa, anche nella virtuosa Lombardia i conti non tornano. Strutture private: 'Tempi raddoppiati nel 2018' - Il Fatto Quotidiano (no date). Available at: https://www.ilfattoquotidiano.it/2018/12/28/liste-dattesa-anche-nella-virtuosa-lombardia-i-conti-non-tornano-strutture-private-tempi-raddoppiati-nel-2018/4852737/ (Accessed: 21 January 2020).

Liu, B. (2012) *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Lohr, K. N. and Schroeder, S. A. (1990) 'A Strategy for Quality Assurance in

Medicare', *New England Journal of Medicine*, 322(10), pp. 707–712. doi: 10.1056/NEJM199003083221031.

Mastrogiacomo, L., Barravecchia, F. and Franceschini, F. (2016) 'Service recycling and ecosystems: an intriguing similarity', *International Journal of Quality and Service Sciences*. Emerald Group Publishing Ltd., 8(4), pp. 555–562. doi: 10.1108/IJQSS-03-2016-0017.

Mauriziano, anziani seguiti a casa: così si svuota il pronto soccorso - La Stampa (no date). Available at: https://www.lastampa.it/torino/2019/12/30/news/mauriziano-anziani-seguiti-a-casa-cosi-si-svuota-il-pronto-soccorso-1.38270077 (Accessed: 23 January 2020).

Milano, due ore di attesa in Pronto soccorso: un pomeriggio in ospedale - Cronaca - ilgiorno.it (no date). Available at: https://www.ilgiorno.it/milano/cronaca/attesa-pronto-soccorso-1.4976155 (Accessed: 21 January 2020).

Napoli: insetti nel cibo dei pazienti, l'Asl manda gli ispettori in ospedale - Il Mattino.it (no date). Available at: https://www.ilmattino.it/pay/edicola/insetti_nel_cibo_in_ospedale_napoli_asl_manda_ispettori-4974826.html (Accessed: 22 January 2020).

Neely, A. (2008) 'Exploring the financial consequences of the servitization of manufacturing', *Operations Management Research*. Springer New York, 1(2), pp. 103–118. doi: 10.1007/s12063-009-0015-5.

O'Connor, P. (2008) 'User-Generated Content and Travel: A Case Study on Tripadvisor.Com', in *Information and Communication Technologies in Tourism 2008*. Springer Vienna, pp. 47–58. doi: 10.1007/978-3-211-77280-5_5.

Palese, B. and Piccoli, G. (2016) 'Online Reviews as a Measure of Service Quality', *Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics*. Available at: https://aisel.aisnet.org/sigdsa2016/9 (Accessed: 15 November 2019).

Pang, B., Lee, L. and Vaithyanathan, S. (2002) 'Thumbs up? Sentiment Classification using Machine Learning Techniques', in *Proceedings of the ACL-02 conference on*

Empirical methods in natural language processing - EMNLP '02. Morristown, NJ, USA: Association for Computational Linguistics, pp. 79–86. doi: 10.3115/1118693.1118704.

Parasuraman, A. P., Zeithaml, V. and Berry, L. (1988) 'SERVQUAL: A multiple- Item Scale for measuring consumer perceptions of service quality', *Journal of retailing*.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1985) 'A Conceptual Model of Service Quality and Its Implications for Future Research', *Journal of Marketing*. SAGE Publications, 49(4), p. 41. doi: 10.2307/1251430.

Participative web and user-created content: Web 2.0, wikis and social networking (2007) *Participative web and user-created content: Web 2.0*. Organisation for Economic Cooperation and Development (OECD). doi: 10.1787/9789264037472-en.

Porter, M. E. (2010) 'What is value in health care?', *New England Journal of Medicine*. Massachussetts Medical Society, pp. 2477–2481. doi: 10.1056/NEJMp1011024.

Raghupathi, W. and Raghupathi, V. (2014) 'Big data analytics in healthcare: promise and potential', *Health Information Science and Systems*. Springer Nature, 2(1). doi: 10.1186/2047-2501-2-3.

Ripartizione geografica secondo l'Istat (no date). Available at: http://dwcis.istat.it/cis/docs/4-8.htm (Accessed: 8 December 2019).

Services, value added (% of GDP) | Data (2017). Available at: https://data.worldbank.org/indicator/NV.SRV.TOTL.ZS (Accessed: 30 October 2019).

Shirky, C. (2009) *Uno per uno, tutti per tutti : il potere di organizzare senza organizzare*. Codice.

Sornette, D. *et al.* (2007) 'Algorithm for model validation: Theory and applications', *Proceedings of the National Academy of Sciences of the United States of America*, 104(16), pp. 6562–6567. doi: 10.1073/pnas.0611677104.

Strome, T. L. (2013) *Healthcare Analytics for Quality and Performance Improvement.*

Wiley.

Studi dell'OCSE sulla Qualità dell'Assistenza Sanitaria: Italia 2014 (2016) *Studi dell'OCSE sulla Qualità dell'Assistenza Sanitaria: Italia 2014*. OECD. doi: 10.1787/9789264268128-it.

Text Analytics Toolbox - MATLAB (no date). Available at: https://it.mathworks.com/products/text-analytics.html (Accessed: 18 November 2019).

Text Mining 101: Topic Modeling (no date). Available at: https://www.kdnuggets.com/2016/07/text-mining-101-topic-modeling.html (Accessed: 15 December 2019).

Tran, H. (2019) 'A Survey of Machine Learning and Data Mining Techniques used in Multimedia System', (113), pp. 13–21. doi: 10.13140/RG.2.2.20395.49446/1.

Vandermerwe, S. and Rada, J. (1988) 'Servitization of business: Adding value by adding services', *European Management Journal*, 6(4), pp. 314–324. doi: 10.1016/0263-2373(88)90033-3.

Watson, A. R. (2015) 'Impact of the digital age on transforming healthcare', in *Healthcare Information Management Systems: Cases, Strategies, and Solutions: Fourth Edition*. Springer International Publishing, pp. 219–233. doi: 10.1007/978-3-319-20765-0_13.

Ye, Q., Zhang, Z. and Law, R. (2009) 'Sentiment classification of online reviews to travel destinations by supervised machine learning approaches', *Expert Systems with Applications*. Elsevier Ltd, 36(3 PART 2), pp. 6527–6535. doi: 10.1016/j.eswa.2008.07.035.

Zaki, M. and McColl-Kennedy, J. (2019) *Text Mining Analysis Roadmap (TMAR) for Service Research*. doi: 10.1108/JSM-02-2019-0074.

Ringraziamenti

Desidero innanzitutto ringraziare il Prof. Mastrogiacocomo e il Prof. Franceschini, in ordine relatore e correlatore di questa tesi, per l'aiuto e i consigli forniti durante la stesura della stessa. Grazie soprattutto per aver stimolato dentro di me l'interesse nella qualità dei servizi durante il corso di studi e le attività del lavoro di gruppo.

Vorrei inoltre ringraziare tutte le persone che mi hanno supportato in questi anni di Poli, la lista è così lunga che dimenticherò sicuramente qualcuno: Prove, Luca S., Pietro, Fabio P., Schiuch, Simo Z., il signor Mattia, Gogo, Perpi, Sam, Giuli, Ludo, Enri, Dan, Kate, Francesca e tutte le altre persone che anche se incontrate brevemente hanno reso il viaggio magnifico. Un ringraziamento speciale va agli amici che mi accompagnano da tutta la vita e con cui ho condiviso tutto: il gruppo Borgata, Cola, Simone F., Marco M. Un ringraziamento unico va alla mia famiglia che mi ha permesso di realizzare questo sogno, in particolare ai miei genitori e mia sorella. Come ultima persona vorrei ringraziare la mia fidanzata Rossella, nei momenti più bui è stata faro, in quelli più felici gioia e oggi senza di lei sarei una persona diversa.

Grazie a tutti per aver reso questi anni i più belli della mia vita.