

Università degli Studi di Napoli “Federico II”



Dipartimento di Scienze Biomediche Avanzate

Tesi di Dottorato di Ricerca in

Scienze Biomorfologiche e Chirurgiche XXXIV Ciclo

Progettazione, implementazione e realizzazione di strumenti software destinati all'e-Public Health, per processi di business intelligence di dati sanitari ed il supporto al controllo di gestione nelle aziende sanitarie.

Design, implementation and software tools development for e-Public-Health, for business intelligence processes of health data and supporting the management control in healthcare companies.

Coordinatore del corso

Ch.mo Prof. Alberto Cuocolo

Candidata

Dott.ssa Antonella Fiorillo

Tutor

Prof. Giovanni Improta

Dott. Mario Russo

Anno accademico 2020-2021

A chi ha creduto in me!

Indice

Acronimi	4
Scopo del lavoro.....	6
1 Introduzione: Le sfide e le opportunità di sviluppo del sistema sanitario .8	
1.1 Introduzione alla Digital Health	10
1.2 Business intelligence.....	15
1.2.1 Architettura di un sistema di Business Intelligence	17
1.3 Uno sguardo ravvicinato al Data Analytics.....	25
1.3.1 Data mining	27
1.3.1.1 Machine learning.....	31
1.3.2 Logica Fuzzy.....	39
1.4 Il Controllo di Gestione nelle aziende sanitarie.....	46
1.4.1 Gli obiettivi e finalità del controllo di gestione	46
1.4.2 Architettura del sistema di Controllo di Gestione.....	49
1.4.3 Gli strumenti di Controllo di Gestione	52
1.5 Metodologie per la gestione dei processi sanitari.....	54
1.5.1 I principi e le logiche del Lean Thinking	55
1.5.2 Lean Healthcare	57
1.5.3 I fondamenti del Six Sigma	61
1.5.4 Il Lean Six Sigma.....	66
2 Progettazione di una Piattaforma di Business Intelligence in Sanità a	
supporto del controllo di gestione	69
2.1 Scopo del lavoro	69
2.2 Gli strumenti.....	70
2.3 I Metodi.....	75
2.3.1 Progettazione del Data Warehouse	75

2.3.2	Processo di BI	78
2.4	Risultati: Piattaforma di BI.....	79
3	Produzione scientifica	85
3.1	Lean Six Sigma: Applicazioni in ambito sanitario	85
3.2	Data analytics: Applicazioni in ambito sanitario	94
4	Conclusioni e Sviluppi futuri	98
	Indice delle Tabelle.....	101
	Indice delle Figure	101
	Bibliografia	103

Acronimi

ACP: Analisi delle Componenti Principali

ADL: Analisi Discriminante Lineare

ADQ: Analisi Discriminante Quadratica

ASA: American Society of Anaesthesiologists

BI: Business Intelligence

BN: Bayesian Networks

CCE: Cartella Clinica Elettronica

CTQ: Critical To Quality

DBMS: Date Base Management System

DMAIC: Define, Measure, Analyse, Improve and Control

DT: Decision Trees

DW: Data Warehouse

ECG: Elettrocardiogramma

ERD: Entity Relationship Diagram

ETL: Extract, Transform, Load

FCM: Fuzzy c-means

FSE: Fascicolo Sanitario Elettronico

HTA: Health Technology Assessment

ICT: Information and Communication Technology

ISTAT: Istituto Nazionale di Statistica

IT: Information Technology

kNN: K-Nearest Neighbours

LEA: Livelli Essenziali di Assistenza

LSS: Lean Six Sigma

LT: Lean Thinking

ML: Machine Learning

MLP: Multi-Layer Perceptron

NHS: National Health Service

NN: Neural Networks

OLAP: On-line Analytical Processing

PDTA: Percorso Diagnostico Terapeutico Assistenziale

RF: Random Forest

SIO: Sistema Informativo Ospedaliero

SS: Six Sigma

SSN: Servizio Sanitario Nazionale

SVM: Support Vector Machine

TAC: Tomografia Assiale Computerizzata

TPS: Toyota Production System

VSM: Value Stream Map

Scopo del lavoro

L' enorme quantità di dati sanitari gestiti ed immagazzinati dalle organizzazioni sanitarie sta crescendo ad un tasso esponenziale. Quasi tutti i dati generati sono memorizzati in banche dati difficilmente esportabili verso altri sistemi sia per ragioni burocratiche che per ragioni tecnologiche. Infatti, i dati memorizzati all'interno di un ospedale non sono quasi mai integrati con altri sistemi Information Technology (IT) come quelli dedicati alla Business Intelligence. Ad esempio, se considereremo tutti i dati disponibili in un ospedale dal punto di vista del singolo paziente, le informazioni sul paziente esisteranno nei differenti sistemi che gestiscono le cartelle cliniche elettroniche, laboratorio, sistema di imaging e nei database di prescrizione. Tuttavia, nella stragrande maggioranza dei casi, tutte le fonti di dati menzionate sono memorizzate in siti separati, comportando l'impossibilità di ricavare valore dall'aggregazione di questi set di dati in maniera intelligente. L'unico modo per raggiungere questo obiettivo è riuscire ad aggregare le diverse fonti di dati non solo all'interno dell'infrastruttura IT di una singola clinica ospedaliera o di una applicazione circoscritta, ma anche su più fornitori di assistenza sanitaria. Questi insiemi di dati unificati non andrebbero a vantaggio solo di tutti i soggetti del settore sanitario, ma porterebbero anche un beneficio importante al paziente fornendo un trattamento di prima scelta. Il presente lavoro di ricerca ricade nella circostanza appena descritta e si articola nei seguenti capitoli.

Nel capitolo 1 viene effettuata un'introduzione sulla Digital Health, inquadrando il contesto in cui opera il progetto di ricerca. Sono state trattate tematiche che gettano le basi per raggiungere l'obiettivo finale, o meglio, la progettazione e implementazione di strumenti software di Business Intelligence per la gestione dei dati sanitari a supporto del controllo di gestione.

Tuttavia, è stata effettuata una panoramica su concetti di: Business Intelligence e sull'architettura di tale sistema; le tecniche di Big Data Analytics per l'analisi dei dati presenti nei sistemi di Business Intelligence, come il Data Mining ed il Machine Learning. Per completare, è stata descritta la Logica Fuzzy, in quanto sperimentata nelle attività di ricerca all'estero, presso l'azienda iMed engineering s.r.o. situata a Praga.

In seguito, è stato introdotto il concetto di sistema di controllo di gestione nelle aziende sanitarie, definendo l'obiettivo dell'attività, l'architettura del sistema e gli strumenti per implementarlo. Questo inquadramento è stato utile per chiarire le caratteristiche tipiche

del controllo di gestione da considerare nella progettazione della piattaforma di BI, con la consapevolezza delle criticità presenti nelle attività di controllo di gestione e quindi, contribuire al miglioramento della performance del servizio offerto.

In aggiunta, sono illustrate le metodologie utilizzate nelle ricerche condotte presso l'Università degli Studi di Napoli Federico II. Pertanto, è stata effettuata una panoramica sulle tecniche del Lean Thinking, Six Sigma e la combinazione di entrambi i metodi, il Lean Six Sigma. In particolar modo, sono state evidenziate le applicazioni di tali metodologie per affrontare problemi di gestione dei processi sanitari nelle aziende ospedaliere. In effetti, le criticità del settore sanitario non sono solo la raccolta e la gestione dei dati sanitari, ma è anche importate intervenire sui processi per migliorare la qualità del servizio offerto.

Nel capitolo 2 è racchiusa l'attività di ricerca e sviluppo effettuata presso l'azienda Gesan S.r.l. L'obiettivo finale è stato quello di giungere ad un'innovazione per la gestione e analisi dei dati sanitari che, ad oggi, non è presente nelle strutture sanitarie campane. Pertanto, sono state descritte le fasi di progettazione e implementazione della piattaforma di Business Intelligence per la gestione dei dati sanitari, al supporto del controllo di gestione. Più specificamente, sono illustrati gli strumenti di sviluppo e le attività di progettazione del Data Warehouse fino alla creazione della piattaforma di BI. In particolare, l'analisi e la ricognizione delle fonti dati è stata necessaria per individuare i sistemi informativi da cui attingere i dati, dopo che è stato compreso il contesto e gli utenti destinatari della piattaforma. Invece, la progettazione del Data Warehouse ha consentito di definire il modello relazionale usando una collezione di tabelle per modellare sia le entità che le relazioni tra di esse. Dunque, utilizzando il Data Warehouse ben progettato è stata creata una dashboard di Business Intelligence in grado di fornire informazioni attraverso report dinamici e indicatori di riferimento, uno strumento utile per verificare e monitorare la qualità dell'assistenza nei confronti della generalità dei cittadini, come prevede l'art. 10, comma 1 del d. lgs. 502/1992.

In conclusione, nel capitolo 3 sono stati discussi i principali articoli scientifici pubblicati in importanti riviste e conferenze nel corso delle attività di dottorato. Infatti, sono riportati i risultati raggiunti nel settore dei servizi sanitari impiegando le metodologie descritte nel primo capitolo, dal Lean Six Sigma alle tecniche di Data Analytics come, Machine Learning, Regressione Lineare e Logica Fuzzy.

1 Introduzione: Le sfide e le opportunità di sviluppo del sistema sanitario

Il settore sanitario sta affrontando molti cambiamenti che pongono nuove sfide. In particolare, le normative governative in rapida evoluzione, le innovazioni tecnologiche e le aspettative dei pazienti creano un nuovo ambiente in cui la gestione di uno studio medico non riguarda più solo la cura dei pazienti. Guardando al presente e oltre, le principali sfide che il settore sanitario sta cercando di affrontare sono:

- *Aumento dei costi*: i principali fattori sono, in primo luogo, l'aumento della domanda, a causa dell'invecchiamento della popolazione (gran parte della spesa sanitaria è sostenuta da pazienti over 65) e, di conseguenza, dei casi di malattie croniche associate all'invecchiamento. A questo va aggiunto l'aumento delle malattie legate allo stile di vita (come l'abuso di alcol, fumo e droghe e l'obesità). Un altro fattore determinante per l'aumento della domanda e della spesa è l'introduzione di nuove tecnologie e terapie; in effetti, è probabile che ciò aumenti la domanda di nuovi trattamenti, ma anche i costi, almeno nel breve periodo, legati alla loro attuazione.
- *Aumento della trasparenza*: la facilità di accesso alle informazioni mediche attraverso internet sta aumentando la consapevolezza dei clienti sulle loro esigenze e richieste mediche, riducendo l'asimmetria dell'informazione. Di conseguenza, le organizzazioni sanitarie dovrebbero concentrarsi profondamente su ciò che i pazienti desiderano. Infatti, i pazienti nel loro percorso di cura richiedono di avere la massima qualità possibile del servizio al minor costo possibile e con la massima rapidità.
- *Sicurezza informatica*: a causa del ruolo crescente dei sistemi informativi per raccogliere dati sensibili sui pazienti, il settore è diventato un facile bersaglio della criminalità informatica. Si registra, infatti, una crescente tendenza alla violazione dei dati, con 32 milioni di cartelle cliniche dei pazienti violate solo nella prima metà del 2019. Le organizzazioni sanitarie dovrebbero aumentare la loro reattività nei confronti dei criminali informatici, non solo per garantire la riservatezza delle cartelle cliniche, ma anche per prevenire ingiunzioni per violazione degli standard di sicurezza.
- *Pagamento*: per migliorare l'esperienza del paziente è necessario offrire una varietà di opzioni di pagamento (ad es. eCheck, carta di credito, ecc.) tramite portali online e le più recenti tecnologie di pagamento, come il mobile e il text-to-pay. Tuttavia, è

spesso difficile per gli studi medici impostare internamente tali sistemi di fatturazione e di elaborazione dei pagamenti. Non solo devono negoziare i termini con ciascun processore di pagamento e costruire l'infrastruttura (ad esempio, portale per i pazienti, elaborazione sicura dei pagamenti), ma devono anche assorbire i costi amministrativi correnti per la manutenzione di tali tecnologie. È necessario assicurarsi che il portale di pagamento e il sistema di elaborazione dei pagamenti siano pienamente conformi, altrimenti si rischia di incorrere in pesanti sanzioni.

- *Big data*: con l'introduzione delle tecnologie vengono generati sempre più dati sanitari ed essi sono sparsi tra diversi soggetti e sistemi. Inoltre, i dati sanitari provengono da molte fonti in una varietà di formati. Attualmente non esiste un unico sistema o un'unica infrastruttura tecnologica per recuperare, archiviare e analizzare i dati provenienti da varie fonti. Affinché le organizzazioni sanitarie possano sfruttare con successo il potere dei Big Data, la leadership deve abbracciare tale processo decisionale e supportarlo a livello esecutivo.

In aggiunta, l'Italia dall'inizio del 2020, sta affrontando l'emergenza sanitaria COVID-19 che ha spinto la diffusione di strumenti digitali nel settore sanitario, accelerandone anche la conoscenza e l'uso da parte di cittadini, medici e strutture sanitarie nelle diverse fasi del percorso di cura. L'obiettivo a medio lungo termine deve essere quello di superare il concetto di ospedale come luogo fisico, per sostituirlo come un insieme di servizi e attività che, grazie alla tecnologia, raggiungono il paziente nella propria vita quotidiana e nella propria casa, limitando la necessità di spostamenti talvolta logisticamente difficili, aumentando l'efficienza, riducendo i costi e liberando maggiori risorse da poter utilizzare su casi più complessi.

La Digital Health è l'argomento centrale in cui lo scopo del lavoro si articola. Infatti, l'obiettivo finale del lavoro mira alla progettazione ed implementazione di strumenti software che consentono di acquisire ed elaborare in tempo reale dati sanitari provenienti dalle diverse unità dell'azienda ospedaliera.

1.1 Introduzione alla Digital Health

Il termine Digital Health ha iniziato a diffondersi intorno al 2000 per caratterizzare ampiamente l'impatto della Information and Communication Technology (ICT) nell'ambito dell'assistenza sanitaria [1], [2]. La Digital Health ha molti sinonimi: health IT, health innovation, e-Health, m-Health (mobile-health), online health, smart health, telehealth or telemedicine, e così via [3].

La Commissione Europea ha definito la Sanità Digitale, in inglese e-Health, come «l'uso delle ICT nei prodotti, servizi e processi sanitari accompagnato da cambiamenti di ordine organizzativo, il tutto finalizzato a un miglioramento della salute dei cittadini, dell'efficienza e della produttività in ambito sanitario, nonché a un maggiore valore economico e sociale della salute. L'e-Health riguarda l'interazione tra i pazienti e chi offre i servizi sanitari, la trasmissione di dati tra le varie istituzioni o la comunicazione peer-to-peer tra pazienti e/o professionisti in ambito sanitario».

Secondo l'Osservatorio ICT Salute della School of Management del Politecnico di Milano, *“L'innovazione digitale nella Sanità italiana oggi è una soluzione obbligata, l'unica in grado di modernizzare il sistema e di consentirgli di resistere all'impatto della crescita della domanda fermando il continuo declino della qualità e dell'efficienza”*. Questa necessità di innovazione tecnologica è evidenziata anche dalle linee guida del Ministero della Salute che stabilisce che *“l'adozione di soluzioni basate sulle ICT diventa un'operazione strumentale, volta a migliorarne l'appropriatezza, l'efficienza e l'efficacia, attraverso l'efficienza complessiva del Servizio Sanitario Nazionale (SSN)”*.

Il ruolo dell'e-Health è quello di affiancare a strumenti più tradizionali altri più tecnologici, in modo da fornirli a medici e pazienti per curare e tutelare la salute e il benessere. Il fenomeno dell'e-Health interessa principalmente tre aree:

- ambito tecnologico: sono idee innovative sia in termini software che hardware;
- ambito sanitario: nuovi processi e strumenti che grazie alla loro capacità innovativa generano miglioramenti nella qualità delle prestazioni sanitarie offerte ai cittadini;
- ambito umano: serve a garantire che le tecnologie dell'e-Health non perdano il filo conduttore che li guida, ossia l'usabilità di fruitori come medici, pazienti e infermieri.

Nel dettaglio si possono riassumere quelle che sono le potenziali caratteristiche positive dell'e-Health:

- efficienza: riduce l'uso di risorse, aumenta il rendimento delle prestazioni grazie alla diminuzione degli errori commessi dai medici e delle cure non necessarie, riducendo anche le liste d'attesa e facilitando il cittadino nell'accesso ai propri dati sanitari;
- trasparenza: maggiore tracciabilità circa l'utilizzo e l'accesso ai dati clinici dei pazienti;
- adattabilità: possibilità di sfruttare le nuove tecnologie da diversi dispositivi per migliorare la produttività e diminuire gli errori.

L'e-Health comprende un ampio spettro di tecnologie tra cui soluzioni innovative, dispositivi personali indossabili e dispositivi interni, ma anche sensori di vario tipo. Può aiutare a identificare i rischi per la salute e l'assistenza nella diagnosi, nel trattamento e nel monitoraggio delle condizioni di salute e di malattia, offrendo nuovi modi per acquisire dati continui su individui e popolazioni.

Nello specifico, l'obiettivo principale è quello di migliorare la customer experience dell'utente (medico e paziente), durante l'utilizzo degli strumenti e dei servizi di assistenza. La digitalizzazione del settore sanitario coinvolge diverse figure professionali (medici, ricercatori, personale sanitario) e richiede lo studio e l'approfondimento di diverse discipline come la Medicina, l'ingegneria informatica e la gestione d'impresa.

In Italia, l'Osservatorio di Innovazione Digitale in Sanità ha presentato un rapporto in cui dimostra che il SSN ha una propensione alla digitalizzazione ed alle sperimentazioni tecnologiche, grazie all'utilizzo di tecnologie nei sistemi di cura ed assistenza, come ad esempio la Cartella Clinica Elettronica, il Fascicolo Sanitario Elettronico, il monitoraggio biometrico ed i servizi di Telemedicina, che permettono la trasmissione dei dati ai pazienti via mail, sms e WhatsApp. Di seguito, esamineremo nel dettaglio ognuna delle tecnologie precedentemente elencate.

La Cartella Clinica Elettronica

La Cartella Clinica Elettronica (CCE), introdotta in Italia nel 2012¹, raccoglie le informazioni sanitarie (es. dati anagrafici, visite specialistiche, referti e risultati degli esami) relative ad un paziente, le quali sono conservate attraverso supporti informatici per essere a disposizione delle aziende ospedaliere. Le informazioni vengono costantemente condivise e aggiornate e i dati clinici e sanitari di un paziente possono essere visualizzati in modo facile e veloce.

Il suo utilizzo è in grado di ridurre il tempo delle operazioni a livello burocratico, garantendo un risparmio circa le spese sostenute e rendendo quindi il sistema sanitario più efficiente. Possiamo riassumere i vantaggi della CCE nell'elenco che segue:

- Consultazione e aggiornamento dei dati di un paziente da parte di coloro che lo curano;
- Condivisione immediata dei dati tra gli operatori sanitari;
- Possibilità di ricercare dati ed eseguire analisi grazie alle informazioni disponibili;
- Consente di uniformare le modalità operative delle strutture sanitarie ospedaliere.²

Le informazioni contenute nelle CCE sono gestite nel totale rispetto della privacy del paziente e solo gli operatori sanitari che hanno in cura il paziente possono accedervi.

Il Fascicolo Sanitario Elettronico

Il Fascicolo sanitario elettronico (FSE), è definito dall'art. 12 del Decreto-legge n. 179/2012³ come “*l'insieme dei dati e documenti digitali di tipo sanitario e socio sanitario generati da eventi clinici presenti e trascorsi, riguardanti l'assistito*”, a differenza della CCE, traccia l'intera storia clinica del paziente. Attraverso una card elettronica, il cittadino possiede le credenziali d'accesso per condividere informazioni con il medico curante ed aggiornare i propri dati clinici in totale sicurezza, mentre gli operatori sanitari possono consultarlo attraverso un sistema di autenticazione. I dati del paziente possono

¹ DL 16/2012 – Decreto-legge del 2 marzo 2012. “Disposizioni urgenti in materia di semplificazioni tributarie e di potenziamento delle procedure di accertamento”. L'emendamento al DL Semplificazioni prevede che nei piani di sanità nazionali e regionali venga messa in pratica la gestione elettronica delle pratiche cliniche. Fonte: Parlamento Italiano - Camera dei Deputati.

² Mauro Moruzzi, 2018, “Il sistema sanitario verso la cartella clinica virtuale: cos'è e come farla”.

³ Decreto-legge 18 ottobre 2012, n. 179 recante Ulteriori misure urgenti per la crescita del Paese, c.d. Decreto Crescita bis, coordinato con la legge di conversione 17 dicembre 2012, n. 221 e pubblicato in G.U. n. 294 del 18/12/2012. La sezione IV del decreto è dedicata alla Sanità digitale.

essere utilizzati anche per ricerche scientifiche e statistiche, a patto di garantirne l'anonimato.

La Telemedicina

La Telemedicina dà la possibilità di erogare assistenza sanitaria, attraverso l'utilizzo di nuove tecnologie, in particolare dell'ICT, anche se il medico o qualsiasi altro professionista sanitario, ed il paziente, sono fisicamente distanti.

Ciò comporta la trasmissione di informazioni e dati di carattere medico nella forma di testi, suoni, immagini o altre forme necessarie per la prevenzione, la diagnosi, il trattamento e il successivo controllo dei pazienti⁴. Questa forma di erogazione non mira alla sostituzione della classica prestazione sanitaria, bensì vuole essere uno strumento che consenta di rendere più continuativo il rapporto tra il paziente ed il professionista sanitario. Può in sostanza essere adoperata per scopi sanitari quali:

- Prevenzione secondaria: possibilità di dedicare servizi specifici a cittadini considerati a rischio o già malati;
- Diagnosi: strumento per valutare i sintomi di un paziente per capirne la malattia o la patologia;
- Cura: per stabilire la terapia ed analizzare lo sviluppo delle cure;
- Riabilitazione: pazienti che necessitano di cure riabilitative, possono riceverle a domicilio;
- Monitoraggio: monitoraggio dei parametri vitali.

Per quanto riguarda, invece, le modalità con le quali la telemedicina può essere erogata, vi sono tre macro-categorie:

- Telemedicina specialistica
- Telesalute
- Teleassistenza

La prima ingloba le molteplici possibilità con cui vengono erogati i servizi medici specifici di una particolare scienza medica, tra medico e paziente oppure tra medici e altri professionisti sanitari. Le prestazioni possono essere realizzate in differenti modalità:

⁴ Ministero della Salute: "TELEMEDICINA Linee di indirizzo nazionali".

- Televisita: è un atto sanitario grazie al quale un dottore può relazionarsi con il paziente da remoto;
- Teleconsulto: È una parere a distanza fra dottori, i quali, collaborando, stabiliscono la prognosi o dispongono una cura;
- Telecooperazione sanitaria: corrisponde ad un aiuto fornito da un professionista sanitario nei confronti di un collega che sta prestando un atto sanitario.

Per quanto concerne la seconda categoria, la telesalute, questa riguarda soprattutto l'assistenza primaria, poiché prevede l'assistenza nella diagnosi, nel monitoraggio, nella gestione e nella responsabilizzazione dei pazienti affetti da una patologia cronica. Questa categoria prevede il ruolo attivo sia del paziente che del medico. Infine, l'ultima macro-categoria, ossia la teleassistenza, si identifica in una struttura dalle caratteristiche socio-assistenziali che si fa carico, a domicilio, delle persone anziane o particolarmente vulnerabili, utilizzando allarmi, servizi di emergenza e chiamate di “supporto”.⁵

Tra le tecnologie utilizzate nell'ambito dell'e-Health rientrano anche i dispositivi elettronici indossabili (wearables), utilizzati principalmente nel settore del fitness (es. Smartwatch o bracciali Fitbit), che hanno comportato lo sviluppo di tecniche avanzate di analisi e monitoraggio dei parametri vitali durante l'attività fisica o nella quotidianità, utili anche in ambito medico per tenere sotto controllo i pazienti con patologie croniche.

L'attuale periodo di emergenza Covid-19 ha accresciuto la consapevolezza tra cittadini, professionisti sanitari e manager delle strutture sanitarie sul contributo del digitale nel processo di prevenzione, cura e assistenza. Ma è ancora lontana l'evoluzione del sistema sanitario italiano verso il modello della “Connected Care”, ovvero un sistema salute connesso e personalizzato, grazie a un utilizzo maturo delle tecnologie digitali, alla valorizzazione dei dati e all'empowerment dei cittadini e dei professionisti⁶.

In Italia, nel 2019 la spesa per la Sanità Digitale era cresciuta del 3%, raggiungendo un valore di 1,43 miliardi di euro e confermando il trend di crescita già osservato negli ultimi anni. Per il 2020, il 45% delle aziende sanitarie stima un aumento delle spese correnti e il 47% una crescita degli investimenti per la Sanità Digitale. Il boom di interesse per la Telemedicina durante il lockdown ha portato a un aumento delle sperimentazioni: il 37%

⁵ Michele Gentili, 2019, “Telemedicina 2.0, ecco dove l'Italia deve investire o la Sanità pubblica è senza futuro”.

⁶ Osservatori.net digital innovation. “Telemedicina: eppur si muove! Ma la vera sfida è un nuovo modello di Connected Care”.

delle strutture sanitarie sta sperimentando il Tele-monitoraggio (27% nel 2019) e il 35% la Tele-visita (15% nel 2019).⁷

Ulteriori tematiche che rientrano nella Digital Health, sono la Business Intelligence, i Big Data Analytics, le metodologie di gestione dei processi sanitari e il controllo di gestione nelle aziende sanitarie. Tali elementi sono fondamentali per la progettazione di strumenti software ICT integrati di Business Intelligence per la valutazione e monitoraggio di dati sanitari, che si colloca come soluzione innovativa e flessibile in grado di innalzare il livello di efficienza e efficacia dei servizi sanitari.

1.2 Business intelligence

Il termine Business Intelligence (BI), introdotto da *Howard Dresner* nel 1989 [4], è un termine di gestione aziendale utilizzato per descrivere applicazioni e tecnologie utilizzate per raccogliere, fornire accesso e analizzare dati e informazioni di un'organizzazione, al fine di aiutare a migliorare il processo decisionale [5], [6].

Un sistema BI è un sistema di supporto alle decisioni che integra enormi quantità di dati provenienti da diverse fonti di dati eterogenee e fornisce un insieme di metodologie e strumenti in grado di raccogliere, integrare, analizzare e presentare dati sulle attività e sui processi che avvengono all'interno di un'organizzazione[7]–[9].

In Sanità si affrontano molte sfide: come analizzare enormi quantità di informazioni a diversi livelli negli ospedali nel tentativo di affrontare problemi complessi di assistenza sanitaria [10], aumento della popolazione, scarso ruolo del paziente nella decisione sul trattamento, mancanza di cooperazione nell'attuale sistema di assistenza sanitaria, lunghi tempi di attesa per ricevere il trattamento e molti pazienti insoddisfatti [11].

Date le sfide, la complessità e la natura sensibile associate alla Sanità, le organizzazioni sanitarie potrebbero trarre vantaggio da sistemi e tecnologie, inclusa l'adozione di un sistema di BI [12], [13]. Tuttavia, attraverso i sistemi informativi ospedalieri già è migliorata l'assistenza sanitaria fornendo una documentazione leggibile, una migliore condivisione delle informazioni e migliore gestione degli avvisi [14].

⁷ Osservatori.net digital innovativo. “Covid e Sanità: cittadini, medici e ospedali più digitali dopo l'emergenza”.

La BI è un concetto emergente nel settore sanitario, ma è una pratica manageriale consolidata in altri settori industriali, dove è vista come una componente chiave del processo decisionale strategico e operativo [15], [16]. I vantaggi previsti includono: accesso più semplice ai dati [8]; processo decisionale migliore [17]; risparmio di tempo [18] e miglioramento della performance [19], [20].

Man mano che le normative cambiano e la quantità di dati aumenta, le organizzazioni sanitarie si rivolgono a soluzioni di BI per accedere ai dati raccolti da sistemi informativi in real time, permettendo di prendere decisioni precise, aiutare a migliorare i risultati e la cura dei pazienti, ridurre i costi e garantire delle prospettive di cura migliori nel settore di riferimento.

Dallo studio della letteratura è emerso che le organizzazioni sanitarie utilizzano i sistemi di BI come soluzione robusta per quattro obiettivi principali: miglioramento delle performance dei processi ospedalieri, miglioramento delle cure cliniche e delle malattie, analisi delle performance dei processi ospedalieri, gestione dei dati e migliore comunicazione e collaborazione tra il personale sanitario [21].

Shailam et al. (2018) [22], hanno creato una dashboard nella sala di lettura della radiologia pediatrica che fornisce in tempo reale una visualizzazione che mostra informazioni aggiornate sugli esami radiologici programmati e in corso, che potesse aiutare i radiologi a migliorare il flusso di lavoro clinico e l'efficienza. Infatti, l'implementazione della dashboard ha portato alla diminuzione del tempo impiegato per protocollare, monitorare e scannerizzare gli esami dal radiologo, nonché la riduzione dei tempi di attesa e di esame del paziente e un migliore utilizzo delle risorse.

Wickramasinghe et al. (2011) [23], hanno sviluppato un nuovo modulo di BI per analizzare, visualizzare ed estrarre conoscenza dai dati dalla rete di gestione delle malattie croniche. Lo scopo del modulo BI è facilitare il processo decisionale a breve e lungo termine e migliorare la comprensione dei modelli di assistenza collaborativa, delle politiche e dei modelli economici alla base della rete di gestione delle malattie croniche (cdmNet). Il modulo BI è stato costituito da tre sotto moduli: pre-elaborazione, dashboard, e data mining. Lo scopo del sotto modulo di pre-elaborazione è convertire i dati cdmNet in maniera da poter effettuare l'accesso rapido. Il sotto modulo dashboard fornisce interfacce con componenti grafiche, roll up e drill down, consentendo alle parti interessate di navigare tra le informazioni e comprendere come gestire nel migliore dei

modi le malattie croniche. Il sotto modulo di data mining mira a estrarre modelli dai dati che potrebbero potenzialmente fornire soluzioni a molte domande aperte alla base della gestione delle malattie croniche.

Silva et al. (2017) [24], hanno applicato le tecnologie BI per analizzare un processo specifico nella loro organizzazione per identificare le aree problematiche e per decidere se è necessario apportare miglioramenti. L'obiettivo principale di questo progetto è stato lo sviluppo di indicatori per fornire informazioni su pazienti ricoverati nell'unità di Malattie Cerebrovascolari. Gli indicatori sono stati creati con l'utilizzo di soluzioni di BI per consentire l'analisi di dati multidimensionali in modo interattivo da più prospettive e il confronto tra i dati ricevuti e le statistiche ottenute da diversi studi.

Carte et al. (2008) [25], hanno utilizzato una soluzione di BI nell'organizzazione "Cardinal Health" per gestire diversi tipi di dati e per migliorare le comunicazioni interne e le collaborazioni tra il personale, pertanto ha portato a una riduzione del costo dell'infrastruttura IT e una migliore gestione dei dati.

Pertanto, in tutti i settori, essere ricchi di dati non è direttamente correlato ad ottenere una buona informazione, ma per permettere ai dati di esprimere il massimo potenziale informativo, è necessario saperne estrarre tutto il valore. Senza competenze tecniche e strumenti adeguati di analisi e gestione, i Big Data rischiano di restare solamente dei numeri interessanti, ma non in grado di divenire supporto significativo al processo decisionale [26].

1.2.1 Architettura di un sistema di Business Intelligence

Una tipica architettura di un ambiente di Business Intelligence è formata dai seguenti livelli [27], [28], mostrati in Figura 1:

- *Livello delle sorgenti*: rappresenta le fonti dati archiviate generalmente in database di produzione, spesso utilizzando anche diverse tecnologie o risorse software, siano essi interni o esterni all'organizzazione;
- *Livello dell'alimentazione*: gli strumenti ETL (Extract, Transform and Loading) sono gli attori principali di questo livello, utili per estrarre i dati dalle diverse sorgenti, pulirli e caricarli nel Data Warehouse (DW);

- *Livello del DW*: i flussi dei dati vengono convogliati in un unico contenitore che può essere direttamente consultato, ma che lascia spazio anche alla possibilità di costruire dei blocchi dipendenti dal DW primario, chiamati data mart. I data mart sono più facilmente consultabili in quanto contengono aggregazioni di dati ben strutturati e contenenti le informazioni rilevanti per una particolare categoria o area del business;
- *Livello di analisi*: rappresentato dagli strumenti di data mining, On-line Analytical Processing (OLAP) e di reportistica utili per l'analisi dei dati contenuti nel data warehouse.

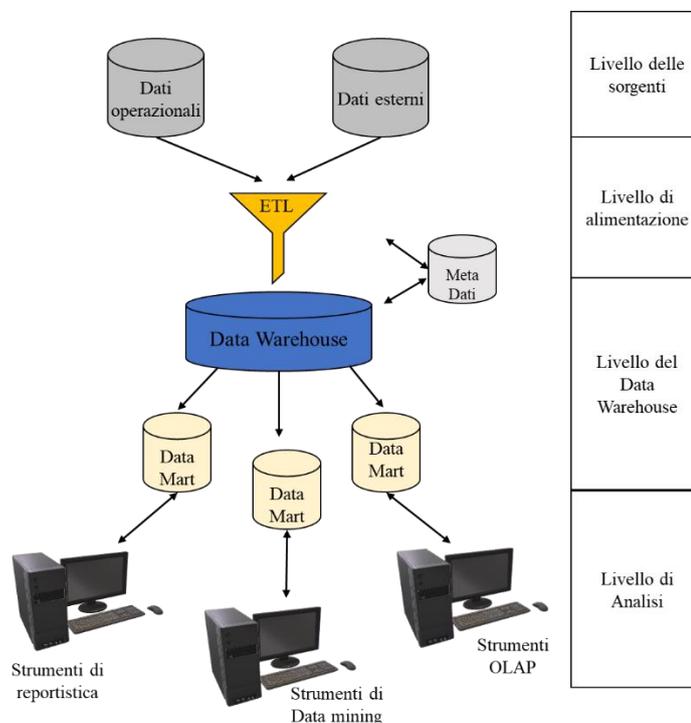


Figura 1. Architettura di un sistema di BI

Livello sorgenti dati

Nel primo livello delle sorgenti dati occorre raccogliere e integrare i dati presenti presso le diverse sorgenti primarie e secondarie, eterogenee per provenienza e tipologia. Esse si basano prevalentemente, sui dati dei sistemi operazionali (ERP, CRM, SCM, etc.), ma possono includere anche documenti non strutturati, quali le e-mail e dati acquisiti

esternamente da agenzie e istituti economici. In generale ne fanno parte tutte le informazioni provenienti sia dall'interno che dall'esterno dell'organizzazione.

Livello di alimentazione

I dati acquisiti nel primo livello, devono poi sottoporsi ad un processo di estrazione, pulizia e trasformazione, ETL, prima di essere caricati nel DW (Livello dell'alimentazione). Le procedure di popolamento del DW possono raggiungere elevati livelli di complessità, in relazione alle discrepanze esistenti tra le sorgenti, al loro livello di correttezza e al livello di precisione rappresentativa nel tempo che si desidera mantenere nel sistema informazionale [29]. Le attività di estrazione, trasformazione e caricamento vengono svolte in modo automatico da strumenti software specifici.

Nella fase di estrazione i dati vengono estratti dalle diverse fonti interne ed esterne disponibili. È possibile distinguere sotto il profilo logico tra una prima estrazione iniziale, nel corso della quale un DW vuoto viene alimentato con i dati disponibili riferiti a tutti i periodi passati, e le successive estrazioni di natura incrementale rivolte all'aggiornamento del DW mediante i dati che si rendono disponibili con il trascorrere del tempo [30].

La fase intermedia di pulitura dei dati ha un ruolo importante, in quanto, ha l'obiettivo di aumentare la qualità mediante la correzione di eventuali inconsistenze, inesattezze e carenze. I principali problemi che vengono riscontrati sono: dati duplicati, dati mancanti, inconsistenze tra campi correlati. Nella fase di pulitura vengono applicate regole automatiche predefinite per la correzione di eventuali errori [31].

La fase di trasformazione, invece, prevede ulteriori conversioni dei dati che ne garantiscono l'omogeneità, rispetto all'integrazione delle diverse fonti. Le principali operazioni realizzate durante questa fase sono [32]:

- **Conversione:** le sorgenti relazionali potrebbero utilizzare formati differenti per la memorizzazione delle informazioni, e quindi è necessaria un'operazione di conversione al fine di ottenere un formato uniforme;
- **Integrazione:** si effettua l'integrazione dei dati provenienti da sorgenti differenti;
- **Aggregazione:** i dati vengono aggregati ad un livello di dettaglio differente (ad esempio solitamente su base mensile);
- **Misure derivate:** vengono calcolate nuove misure a partire da quelle già esistenti;
- **Selezione:** si seleziona un sottoinsieme dei campi contenuti nelle sorgenti, in modo tale da ridurre la quantità di dati da elaborare.

Infine, il caricamento dei dati sul DW è l'ultima fase. Questa procedura può avvenire in due modi [33]:

- **Refresh:** i dati vengono completamente riscritti all'interno del DW. Questa tecnica di caricamento è utilizzata nella fase di inizializzazione del DW, quando si procede all'estrazione di tutti i dati storici;
- **Update:** vengono aggiunti al DW solo i nuovi dati e gli aggiornamenti dei dati storici. Questa tecnica è solitamente utilizzata in relazione all'estrazione incrementale, al fine di ottenere un aggiornamento periodico del DW.

A seguito delle procedure ETL, i dati entrano nel DW a tutti gli effetti. Il DW corrisponde al livello intermedio dell'architettura di un sistema di BI, in cui i dati vengono integrati e archiviati.

Livello Data Warehouse

Il livello del DW è il componente principale di un sistema di BI. Questo componente è un repository di dati provenienti da diverse fonti che viene utilizzato per memorizzare informazioni [34].

William H. Inmon (2005) [35], è stato il primo che ha parlato esplicitamente di DW, lo definisce come "una raccolta di dati "integrata, orientata al soggetto, variabile nel tempo e non volatile" di supporto ai processi decisionali". Pertanto, secondo Inmon la raccolta di dati deve essere:

- *Integrata:* in un DW confluiscono dati provenienti da più sistemi e fonti esterne. È ,pertanto, fondamentale l'integrazione dei dati raccolti attraverso due modi: utilizzando metodi di codifica uniformi, mediante l'utilizzo delle stesse unità di misura o attraverso l'omogeneità semantica di tutte le variabili;
- *Orientata al soggetto:* il DW è orientato a temi aziendali specifici piuttosto che alle applicazioni o alle funzioni. In un DW i dati vengono archiviati in modo da essere facilmente letti o elaborati dagli utenti. L'obiettivo è quello di fornire dati organizzati in modo tale da favorire la produzione di informazioni. Si passa dalla progettazione per funzioni ad una modellazione dei dati che consenta una visione multidimensionale degli stessi;
- *Variabile nel tempo:* i dati archiviati all'interno di un DW coprono un orizzonte temporale molto più esteso rispetto a quelli archiviati in un sistema operativo. Nel DW sono contenute una serie di informazioni relative alle aree di interesse che

colgono la situazione relativa ad un determinato fenomeno in un determinato intervallo temporale piuttosto esteso. Ciò comporta che i dati contenuti siano aggiornati fino ad una certa data che, nella maggior parte dei casi, è antecedente a quella in cui l'utente interroga il sistema;

- *Non volatile*: tale caratteristica indica la non modificabilità dei dati contenuti nel DW che consente accessi in sola lettura. Ciò comporta una semplicità di progettazione del database rispetto a quella di un'applicazione transazionale. In tale contesto non si considerano le possibili anomalie dovute agli aggiornamenti, né tanto meno si ricorre a strumenti complessi per gestire l'integrità referenziale o per bloccare record a cui possono accedere altri utenti in fase di aggiornamento.

È importante notare che queste proprietà distinguono i DW dai database operativi. A differenza dei database operativi, un DW non è volatile e, quindi, i suoi dati non vengono modificati o cancellati, ma vengono aggiunti al DW quando entrano nel sistema. In questo modo, i DW variano nel tempo e consentono l'archiviazione temporale e l'analisi dei dati. Di conseguenza, l'analisi temporale dei dati consente la presentazione di informazioni sull'evoluzione delle attività e dei processi che si verificano all'interno dell'organizzazione in un determinato periodo di tempo. I dati possono anche essere archiviati in repository più piccoli orientati al soggetto noti come data mart. Questi data mart sono strutture che contengono quantità di dati inferiori rispetto ai DW, permettono un'analisi più orientata ai loro obiettivi perché hanno solo dati su un determinato argomento [35].

La progettazione di un DW può avvenire secondo due modalità diverse [36]:

- *top-down*: introdotto da *Inmon* (2005)[35], consiste nel progettare l'intero DW e, solo successivamente, estrarre da esso i singoli data mart. Questa soluzione si basa su una visione globale dell'obiettivo e garantisce la realizzazione di un DW consistente e ben integrato. Gli svantaggi sono legati all'impossibilità di prevedere nel dettaglio tutte le esigenze che potrebbero sorgere nelle diverse aree aziendali e ai lunghi tempi di realizzazione. Tale soluzione è anche conosciuta nella letteratura come *data-driven*, cioè guidata dai dati;
- *bottom-up*: introdotto da *Kimball e Ross* (2011) [37], è un approccio di tipo incrementale: il DW viene costruito assemblando iterativamente i diversi data mart. Questa soluzione permette, in fase di progettazione, di spostare il focus sulla singola area di business e quindi di realizzare in tempi brevi i singoli data mart.

Tale approccio è anche conosciuto nella letteratura come approccio analysis-driven, cioè guidato dalle analisi.

Il processo per la realizzazione del DW può essere descritto come segue [38], [39]:

1. *Analisi e riconciliazione delle sorgenti*: questa fase prevede la ricognizione e mappatura delle sorgenti disponibili (ricognizione), l'integrazione e omogeneizzazione delle stesse per eliminare le inconsistenze;
2. *Analisi dei requisiti*: il progettista deve coinvolgere tutte le figure professionali dell'azienda per raccogliere e stilare gli obiettivi da raggiungere;
3. *Progettazione concettuale*: in questa fase si caratterizza nel dettaglio il DW ed è il passo fondamentale per rappresentare concettualmente il modello multidimensionale. Per questo motivo è necessario costruire il Dimensional Fact Model (DFM), ovvero si costruisce uno schema di fatto in cui sono modellati: i fatti⁸, le dimensioni⁹, gli attributi¹⁰ e le misure¹¹. Uno schema di fatto è del tipo di Figura 2.
4. *Progettazione logica*: la fase che permette di definire, a partire dal modello concettuale, uno schema della base di dati nel modello di dati implementabile dal Date Base Management System (DBMS). Il prodotto che si ottiene al termine di questa fase è denominato schema logico, in cui si mette in evidenza la multidimensionalità del DW. La struttura multidimensionale dei dati può essere rappresentata utilizzando schemi a stella, fiocco di neve e costellazione. Lo schema a stella consiste in una tabella del fatto che riferenzia due o più tabelle dimensionali; lo schema a fiocco di neve è un'estensione dello schema a stella dove le tabelle dimensionali riferenziano a loro volta altre tabelle; infine, lo schema a costellazione consiste in due o più tabelle del fatto che condividono alcune dimensioni, Figura 3.
5. *Progettazione dell'alimentazione*: in questa fase vengono definite le modalità di caricamento dei dati provenienti dalle sorgenti dati nel DW.

⁸ Un fatto modella un insieme di eventi che si verificano quando un obiettivo viene raggiunto.

⁹Una dimensione è una proprietà di un fatto che contiene una serie di attributi.

¹⁰ Sono dei campi la cui valorizzazione si accompagna alla registrazione di un fatto ad opera di un obiettivo, quindi sono propriamente le caratteristiche del fatto.

¹¹ Una misura è una proprietà numerica di un fatto che ne descrive un aspetto quantitativo.

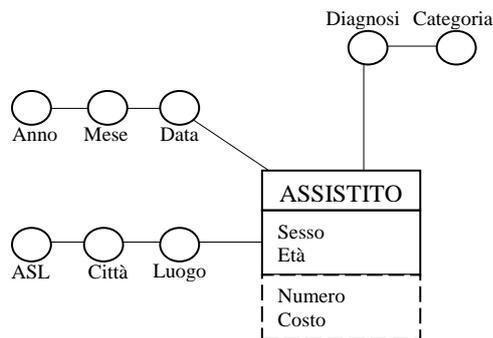


Figura 2. Progettazione concettuale. Schema di fatto (DFM).

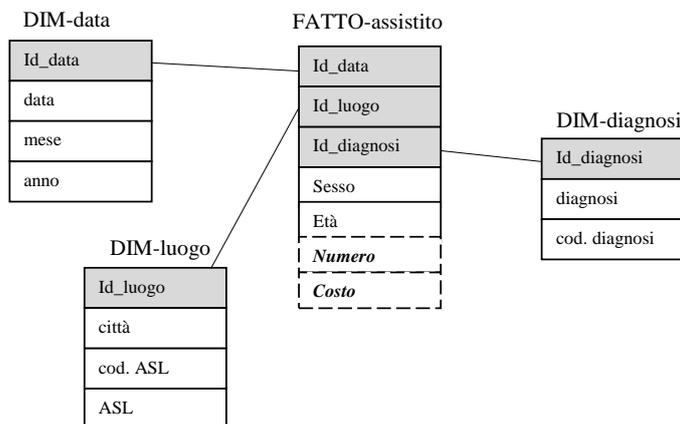


Figura 3. Progettazione logica. Schema a stella.

Livello d'analisi

Al livello di analisi, con l'ausilio di software di Business Intelligence, vengono utilizzati diversi elementi, in particolare strumenti di visualizzazione dei dati (grafici a barre e a torta, istogrammi, ecc.), strumenti di data mining e OLAP, report e query, creazione di dashboard.

Tra le tecniche più utilizzate per accedere e analizzare i dati archiviati in un DW o data mart, emerge *On-line Analytical Processing* (OLAP) [40]. La tecnologia OLAP consente agli analisti di acquisire informazioni sui dati attraverso un accesso rapido, coerente e interattivo a un'ampia varietà di possibili visualizzazioni delle informazioni. I dati da

analizzare con OLAP sono organizzati in cubi che consentono la visualizzazione delle informazioni secondo ogni dimensione del modello dati. Il modello multidimensionale è costituito dai fatti, la descrizione quantitativa di ciascun evento, ovvero il verificarsi di una ricorrenza per un determinato fatto, fa uso delle misure. La numerosità degli eventi, che possono occorrere comporta la necessità di raggrupparli in uno spazio n-dimensionale i cui assi sono detti dimensioni. La Figura 4 mostra un esempio dell'entità "Paziente" con le n-dimensioni (anno, tipo accesso, diagnosi). Gli strumenti OLAP supportano l'analisi in tempo reale, consentendo all'utente di effettuare ricerche più strutturate e veloci per generare grafici e tabelle. In base alle esigenze di analisi dell'utente, questi strumenti consentono di eseguire operazioni sui dati come *drill-down*, *roll-up*, *slice and dice* [41]:

- l'operatore di *roll-up* serve ad aggregare i dati, accumulandoli su una o più dimensioni;
- l'operatore di *drill-down*, al contrario, espone i dati aumentandone il livello di dettaglio;
- il terzo operatore seleziona un sottoinsieme dei dati (*slice*) e lo analizza sotto diversi aspetti (*dice*).

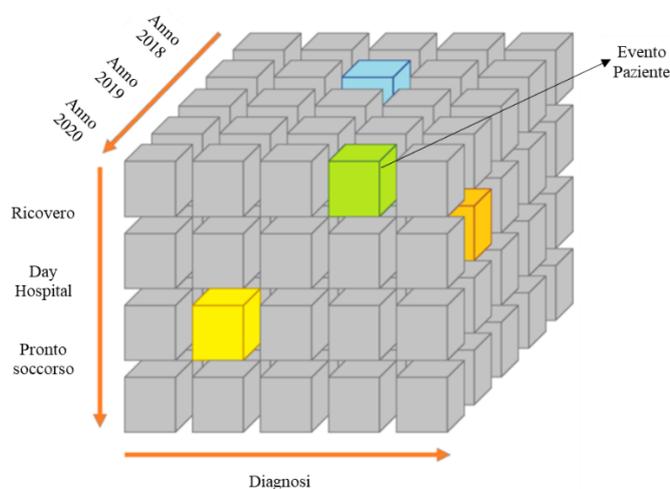


Figura 4. Cubo OLAP. L'evento paziente ha dimensioni: Anno, Tipo di accesso, Diagnosi.

Tuttavia, il Data Mining ha come obiettivo quello di aiutare l'utente nella ricerca dei patterns¹² e per farlo utilizza un insieme di tecniche e metodologie statistiche e di intelligenza artificiale come il Machine Learning. Gli strumenti di Data Mining

¹² Per pattern si intende una struttura, un modello, o più in generale una rappresentazione sintetica dei dati.

permettono di estrarre conoscenza da banche dati di dimensioni molto elevate mediante l'applicazione di algoritmi che individuano le associazioni nascoste tra le informazioni e le rendono visibili. Questo fa sì che non occorra più effettuare manualmente l'analisi dei dati, ma basti indicare con una certa approssimazione cosa e dove si vuole ricercare e lasciare che uno strumento automatico si accoli il peso della computazione [42] (*Vedere il paragrafo 1.3.1. per ulteriori approfondimenti*).

Dopo le fasi di elaborazione e gestione dei dati, si passa alla parte relativa ai report e dashboard. La sua finalità è stata la rappresentazione per via grafica dei dati e delle loro relazioni, al fine di generare una conoscenza accessibile e fruibile. Intento principale, del resto, è la realizzazione di uno strumento di supporto allo stadio decisionale che può conferire un valore aggiunto alle scelte e all'interpretazione dei diversi contesti e ambiti, costituenti e coinvolti. Pertanto, le dashboard sono le vesti di documenti di sintesi comprendenti un insieme di report precedentemente realizzati, immagazzinando in un solo colpo d'occhio informazioni immediate. Infatti, sono progettate per conseguire il massimo impatto visivo in una piattaforma, ideata per la comprensione rapida che sfrutta una combinazione di tabelle, grafici e altri indicatori, giungendo a una visione immediata della performance resasi in tempo reale rispetto ai trend vigenti [43].

1.3 Uno sguardo ravvicinato al Data Analytics

La raccolta dei dati è una parte importante per ogni organizzazione, più informazioni si hanno, più è possibile organizzarsi in modo ottimale per fornire i migliori risultati. Oggi si è di fronte a una situazione in cui si è inondati di tonnellate di dati relativi ad ogni aspetto della nostra vita come attività sociali, scienza, lavoro, salute, ecc.. Ciò ha portato alla creazione del termine "Big Data" per descrivere dati di grandi dimensioni e ingestibili. Come ogni altro settore, le organizzazioni sanitarie stanno producendo dati a un ritmo straordinario che presenta molti vantaggi e sfide allo stesso tempo [44].

Inoltre, ciò che caratterizza i Big Data non è solo la quantità ma anche la complessità riconducibile alla loro varietà. Secondo la definizione ufficiale della Comunità Europea, *“i Big Data in Sanità si riferiscono a grandi set di dati raccolti periodicamente o automaticamente, che vengono archiviati elettronicamente, riutilizzabili allo scopo di migliorare le prestazioni del sistema sanitario”*. Tuttavia, i Big Data della Sanità digitale (cartelle cliniche elettroniche, dossier sanitari, fascicolo sanitario elettronico) possono

diventare importanti se, una volta archiviati, vengono rielaborati e trasformati in informazioni.

Con il termine “Big Data Analytics” si fa riferimento al processo che include la raccolta e l’analisi dei Big Data per ottenerne informazioni. Le informazioni raccolte e rielaborate tramite le tecniche di Data Analytics permettono non solo di curare il paziente in maniera più mirata, ma anche di utilizzare questi dati ai fini della ricerca scientifica, per la prevenzione e la cura di gravi patologie, per la ricerca di nuovi farmaci, ecc.[45].

La Data Analytics prevede, in generale, quattro tipi principali di analisi dei dati [46]:

- L'analisi descrittiva aiuta a rispondere alle domande su cosa è successo precedentemente. Queste tecniche usano grandi set di dati per descrivere i risultati ottenuti. Sviluppando indicatori di performance, Key Performance Indicator (KPI), che possono aiutare a tenere traccia di successi o fallimenti. Questo processo richiede la raccolta di dati rilevanti, l'elaborazione dei dati, l'analisi dei dati e la visualizzazione dei dati, fornendo informazioni essenziali sulle performance passate.
- L'analisi diagnostica aiuta a rispondere alle domande sul perché le cose sono accadute. Queste tecniche integrano l'analisi descrittiva di base, prendono i risultati dall'analisi descrittiva e scavando più a fondo per trovare la causa. Ciò avviene generalmente in tre fasi: identificare le anomalie nei dati, che possono essere cambiamenti imprevisti in una metrica; raccolta dei dati relativi a queste anomalie; utilizzo di tecniche statistiche per trovare relazioni e tendenze che spieghino queste anomalie.
- L'analisi predittiva aiuta a rispondere alle domande su cosa accadrà in futuro. Queste tecniche utilizzano dati storici per identificare le tendenze e determinare se è probabile che si ripetano. Gli strumenti analitici predittivi forniscono preziose informazioni su ciò che potrebbe accadere in futuro e le sue tecniche includono una varietà di tecniche statistiche e di Machine Learning, come: reti neurali, alberi decisionali e regressione.
- L'analisi prescrittiva aiuta a rispondere alle domande su cosa dovrebbe essere fatto. Utilizzando le informazioni provenienti dall'analisi predittiva, è possibile prendere decisioni basate sui dati e consente agli utenti di prendere decisioni informate di fronte all'incertezza. Le tecniche di analisi prescrittiva si basano su

strategie di Machine Learning in grado di trovare modelli in set di dati di grandi dimensioni. Analizzando le decisioni e gli eventi passati, è possibile stimare la probabilità di esiti diversi.

1.3.1 Data mining

Lo sviluppo del Data Mining è stato favorito nel corso del tempo da tre fattori principali, ovvero il grande accumulo di dati in formato elettronico, il costo sempre minore di dispositivi di storage e lo sviluppo di nuove tecniche dell'intelligenza artificiale come il Machine Learning.

Il Data Mining usa i metodi di programmazione e l'analisi matematica-statistica per trovare patterns e relazioni nascoste all'interno dei dati. Pertanto, ha due funzioni fondamentali; quella di estrazione e quella di analisi. Nella prima, le tecniche utilizzate estraggono informazioni implicite e nascoste per poterle rendere utilizzabili; nella seconda, si esplorano e si analizzano questi dati in maniera automatica o semi-automatica per scoprire degli schemi e delle regolarità nei comportamenti analizzati.

Per definire il Data Mining, sono state estrapolate alcune definizioni dalla letteratura che insieme permettono di chiarire il concetto:

- Il Data Mining è uno step nel processo di Knowledge Discovery in Database (KDD), che consiste nell'applicare analisi di dati e algoritmi di individuazione che, sotto accettabili limiti di efficienza computazionale, producono una particolare enumerazione di patterns (modelli) sui dati [47].
- Il Data Mining è la ricerca di relazioni e modelli globali che sono presenti in grandi database, ma che sono nascosti nell'immenso ammontare di dati, come le relazioni tra i dati dei pazienti e le loro diagnosi mediche. Queste relazioni rappresentano una preziosa conoscenza del database e, se il database è uno specchio fedele del mondo, tali relazioni sono una valida rappresentazione della realtà [48].
- Il termine Data Mining si riferisce quindi all'intero processo che si articola nella raccolta e nell'analisi dei dati, nello sviluppo di modelli matematici e

nell'adozione di decisioni e azioni concrete basate sulle conoscenze acquisite [49].

Dalle definizioni sopra citate si possono individuare i punti chiave del concetto di Data Mining:

- È uno step dell'ampio processo chiamato Knowledge Discovery in Database (KDD). Lo scopo di tale processo è di estrarre la conoscenza dai dati. La definizione di KDD fornita da *Usama Fayyad et al.*, (1996) [47] è: “*La KDD è il processo non banale di identificazione di patterns validi, nuovi, potenzialmente utili e comprensibili nei dati*”, i quali oltre a fornire una definizione gli hanno identificato cinque fasi che caratterizzano il processo di KDD: selezione dei dati, preelaborazione, trasformazione, data mining, interpretazione dei risultati del modello di data mining;
- Analizza grosse mole di dati, poiché oggi non è difficile disporre di macchine con una buona potenza di calcolo che consente di scavare attraverso un sacco di dati. Questo a sua volta, permette agli utenti di fare ciò che sanno fare meglio, ovvero impostare il problema e comprendere i risultati [50];
- Si appoggia a dei modelli matematici, quindi si fa riferimento all'insieme di modelli e metodi matematici che sono il fulcro di ogni analisi di data mining e che vengono utilizzati per generare nuove conoscenze [49];
- Si basa su metodi di apprendimento induttivo, in quanto cerca di ricavare regole generali a partire da una sequenza di esempi rappresentati dalle osservazioni riferite al passato e accumulate nei database. In altre parole, le analisi si propongono di trarre alcune conclusioni a partire da un campione di osservazioni passate, che vengono generalizzate all'intera popolazione, in modo che possano essere il più accurate possibili;
- I risultati ottenuti devono essere patterns (modelli) e regole significative, pertanto le attività operative di un'organizzazione generano dati e, di conseguenza, generano dei patterns. Tuttavia, l'obiettivo del data mining non è quello di trovare ogni schema presente nei dati, ma di estrarre quei patterns che rivelano utili informazioni per l'attività. Questo significa trovare schemi per aiutare le operazioni di routine.

In genere, ci sono due tipologie di modelli di Data Mining: predittivo e descrittivo. I modelli predittivi usano alcune variabili dei dati analizzati per predire valori sconosciuti

e futuri di altre variabili di interesse, mentre i modelli descrittivi si focalizzano sul trovare schemi interpretabili che forniscono informazioni sulla struttura dei dati. Sebbene i confini tra predizione e descrizione non siano netti, alcuni dei modelli predittivi possono essere descrittivi nella misura in cui possono essere comprensibili e viceversa.

Le tecniche del data mining sono: classificazione (predittiva), clustering (descrittiva), ricerca di regole associative (descrittiva), ricerca di pattern sequenziali (descrittiva), regressione (predittiva), individuazione di deviazioni (predittiva).

I principali modelli utilizzati nel settore sanitario sono utilizzati per prevedere varie malattie e per assistere i medici nella diagnosi e nella predizione della loro decisione clinica [51].

Dalla letteratura emergono le seguenti tecniche prevalentemente utilizzate in sanità:

Regressione

La Regressione è una tecnica di Data Mining utilizzata per adattare un'equazione a un set di dati. La forma più semplice di regressione, la regressione lineare [52], utilizza la formula di una linea retta ($y = mx + b$) e determina i valori appropriati per m e b per prevedere il valore di y in base a un dato valore di x [53], [54]. La tecnica di regressione multipla è un'estensione della regressione lineare che coinvolge più di una variabile predittiva. Consente di modellare la variabile di risposta Y come una funzione lineare del vettore di feature ultra dimensionale che abbiamo, ovvero $Y = \alpha + \beta_1 X_1 + \beta_2 X_2$, dove α , β_1 e β_2 sono coefficienti di regressione. Fondamentalmente vengono utilizzati modelli di regressione lineare per mostrare o prevedere la relazione tra due variabili o fattori. Il fattore che viene previsto (il fattore per cui l'equazione si risolve) è chiamato variabile dipendente. I fattori utilizzati per prevedere il valore della variabile dipendente sono chiamati variabili indipendenti. Xie et al. (2015) [55], hanno proposto un algoritmo di albero decisionale di regressione per prevedere il numero di giorni di ospedalizzazione in una popolazione. I risultati sperimentali hanno mostrato che l'algoritmo proposto è stato eseguito in modo significativo nella popolazione generale e nelle sottopopolazioni al fine di prevedere il futuro ricovero.

Classificazione

Il processo di classificazione consiste nell'assegnazione di etichette di classe ai dati non ancora etichettati. Dato un insieme di classi, individuate per mezzo di etichette di classe,

e una collezione di dati precedentemente classificati, l'obiettivo è quello di generare un modello di classificazione che consenta di assegnare le corrette etichette di classe ai dati non ancora etichettati. Un modello di classificazione viene generato utilizzando uno specifico algoritmo di classificazione a cui in input vanno dei dati di training (o training set) che consistono in una collezione di dati, chiamati istanze, costituiti da un insieme di valori per un insieme finito attributi e da un valore per l'attributo di classe. Dall'analisi dei dati di training viene generato il modello di classificazione e questo conclude la prima fase del processo di classificazione. Successivamente, utilizzando questo modello di classificazione è possibile assegnare ai dati non ancora etichettati, denominati dati di test (o test set), le rispettive etichette di classe [56]. La classificazione è divisa in due tipologie: supervised e unsupervised (o clustering). Nella classificazione di tipo supervised l'insieme delle possibili etichette di classi è conosciuta a priori, quindi, nel processo di classificazione, alle istanze di test viene assegnata un valore di etichetta di classe tra quelle indicate a priori. Invece, nella classificazione di tipo unsupervised non ci sono etichette di classe note a priori, ma solo dopo aver effettuato la classificazione, dividendo i dati in gruppi secondo attributi simili, è possibile dare un nome ai diversi gruppi generati e creare così le etichette di classe. Quest'ultima è meglio nota sotto il nome di clustering. In letteratura esistono diverse tecniche di classificazione di tipo supervised, tra le quali sono presenti: alberi di decisione, regole di classificazione, regole di associazione, reti neurali, naïve bayes e reti bayesiane, K-Nearest Neighbours (kNN), Support Vector Machine (SVM). (*Vedere il paragrafo 1.3.1.1 per ulteriori approfondimenti*).

Clustering

Il clustering è definito come l'apprendimento non supervisionato che si verifica osservando solo le variabili indipendenti mentre l'apprendimento supervisionato analizza sia le variabili indipendenti che quelle dipendenti. È diverso dalla classificazione che è un metodo di apprendimento supervisionato. Infatti, non ha classi predefinite e per questo motivo, il clustering può essere utilizzato al meglio per studi di natura esplorativa, principalmente se tali studi comprendono una grande quantità di dati. L'obiettivo del raggruppamento è di tipo descrittivo mentre l'obiettivo della classificazione è di tipo predittivo. Il metodo di clustering è quello di formare i cluster da un database di grandi dimensioni sulla base della misura di similarità [57]. L'obiettivo del clustering è scoprire un nuovo insieme di categorie. Il clustering raggruppa le istanze di dati in sottoinsiemi in

modo tale che istanze simili siano raggruppate insieme, mentre istanze diverse appartengano a gruppi diversi. L'approccio di clustering viene utilizzato per identificare le somiglianze tra i punti dati. Ogni punto dati all'interno dello stesso cluster ha una maggiore somiglianza rispetto ai punti dati appartenenti ad altri cluster. *Belciug et al.* (2009) [58], hanno utilizzato il clustering gerarchico agglomerativo approccio per raggruppare i pazienti in base alla loro lunghezza di rimanere in ospedale al fine di fornire un migliore utilizzo delle risorse ospedaliere e fornire servizi migliori ai pazienti.

Dunque, il Data mining grazie all'analisi dei dati, la scoperta di patterns ricorrenti e l'utilizzo di algoritmi per costruire i modelli predittivi sono sempre più importanti nel settore dell'assistenza sanitaria, in quanto migliorano le condizioni di salute dei pazienti. Infatti, consentono di prevedere patologie, prevenire abusi nell'utilizzo di medicinali ma anche errori da parte di medici aumentando l'aderenza alla terapia, riducendo il rischio di ospedalizzazione e abbassando i costi sostenuti dal servizio sanitario nazionale [54], [59]. Pertanto, utilizzando il Data Mining nell'assistenza sanitaria, si possono individuare i migliori percorsi di cura che i pazienti dovrebbero seguire e creare dei veri e propri piani di trattamento che permettono di scoprire la metodologia migliore per curare persone con una determinata patologia e che presentano tra loro caratteristiche simili, garantendo così una Digital Health sempre più personalizzata e su misura delle esigenze dei singoli.

1.3.1.1 Machine learning

Il Machine Learning (ML) si occupa di apprendimento automatico della macchina, fornisce gli algoritmi per utilizzare le tecniche di data mining per migliorare il processo di apprendimento. Il data mining e ML possono, riguardare entrambi l'apprendimento dai dati e il prendere decisioni migliori. Ma il modo in cui lo affrontano è diverso. Il data mining è un processo più manuale che si basa sull'intervento umano e sul processo decisionale. Ma, con il ML, una volta stabilite le regole iniziali, il processo di estrazione di informazioni, apprendimento e perfezionamento è automatico e avviene senza l'intervento umano. Il data mining viene utilizzato su un set di dati esistente (come un data warehouse) per trovare modelli. Il ML, d'altra parte, viene addestrato su un set di dati di training, che insegna al computer come dare un senso ai dati e quindi fare previsioni su nuovi set di dati (data test).

Nel 1959, Arthur Samuel [60], un pioniere nel campo del ML lo definì come il "campo di studio che dà ai computer la capacità di apprendere senza essere esplicitamente programmati".

Il ML si trova al crocevia tra informatica, statistica e una varietà di altre discipline interessate al miglioramento automatico nel tempo e all'inferenza e al processo decisionale in condizioni di incertezza [61]. L'incertezza di una diagnosi dedotta da un segno potrebbe essere vista come un problema matematico che ha a che fare con le probabilità. L'applicazione della teoria della probabilità all'apprendimento dai dati è chiamata apprendimento bayesiano [62]. Un algoritmo è addestrato per eseguire un compito (ad esempio, per distinguere il cancro dai tessuti non cancerosi) e per migliorarne le prestazioni e l'accuratezza con l'esperienza. A seconda della quantità di dati e risorse computazionali, questi processi possono guadagnare tempo e precisione rispetto al giudizio soggettivo umano. Le tecniche di ML spesso trattano una grande quantità di dati e quindi hanno un approccio più realistico ai problemi biologici e medici rispetto alla semplice scienza matematica e modellizzazione. Si distinguono due paradigmi del ML: apprendimento supervisionato e apprendimento non supervisionato.

Apprendimento supervisionato

Nell'apprendimento supervisionato, l'algoritmo si concentra su un output noto; in tal caso l'obiettivo è produrre una previsione probabilistica y' in risposta a una query x' . I modelli di apprendimento automatico supervisionato vengono addestrati su dati pre-etichettati denominati set di addestramento di coppie (x, y) . L'errore di addestramento è approssimativamente definito dalla differenza tra il risultato previsto y' e il risultato effettivo y . L'apprendimento supervisionato include diversi algoritmi, come decision trees, decision forests, logistic regression, SVM, neural networks, kNN e Bayesian classifiers [61]. L'esempio classico in medicina è l'interpretazione automatizzata di un elettrocardiogramma (ECG) o di una Tomografia Assiale Computerizzata (TAC). Questi sono tutti compiti che un medico esperto può svolgere e quindi il computer spesso cerca di approssimare le prestazioni umane e forse superarlo [63].

Le applicazioni mediche dell'apprendimento supervisionato sono l'assistenza alla diagnosi per patologie rare e difficili, e anche il follow-up, che richiede una valutazione accurata dell'esito del trattamento: l'applicazione del metodo SVM a un set di dati di bambini (68 sani e 88 con forma di paralisi cerebrale) ha dimostrato che l'algoritmo SVM

è stato in grado di classificare i gruppi di bambini con un'accuratezza complessiva del 96,80% in diagnosi utilizzando solo due parametri [64]. Allo stesso modo, la diagnosi di cardiopatia valvolare attraverso insiemi di reti neurali è stata effettuata con un'accuratezza del 97,4% [65].

Apprendimento non supervisionato

Al contrario, nell'apprendimento non supervisionato, non ci sono risultati da prevedere. L'obiettivo è trovare modelli o raggruppamenti all'interno dei dati. Non è necessaria alcuna formazione perché non c'è l'output desiderato. Le classi degli algoritmi del ML sono analisi di associazione, clustering, riduzione delle dimensioni, reti neurali artificiali. Tra tutti quello più utilizzato è il concetto di clustering, vale a dire che nel dataset la macchina possa automaticamente raggruppare caratteristiche (sintomi, immagini, geni).

L'aspetto "fuzzy"¹³ conferisce all'algoritmo la flessibilità di classificare un punto dati per ciascun cluster in una certa misura relativa alla probabilità di appartenenza a quel cluster. Questo metodo può rilevare modelli nascosti in set di dati grandi e complessi.

L'apprendimento non supervisionato può essere molto utile per identificare meccanismi o fattori di rischio per malattie complesse [63]. Fuzzy c-means (FCM), un metodo di clustering efficace sviluppato da Bezdek [66], è uno degli strumenti più comuni per la diagnosi medica [67], è stato proposto per la classificazione delle malattie della tiroide [68] o tumori per caratterizzazione genica [69]. Anche questo metodo ha i suoi punti deboli: la necessità di determinare il cutoff di appartenenza, cioè la distanza dall'elemento al centro dei cluster, rappresentata dalla media dei dati considerati e il fatto che i cluster si formano qualunque siano gli input, cioè, se i dati non appartengono affatto a nessun cluster, l'algoritmo lo farà comunque, portando potenzialmente a errori. La FCM ha dimostrato la sua efficacia nella cefalea primaria, dove la diagnosi è spesso difficile. L'algoritmo ha mostrato un'accuratezza (proporzione dei risultati veri, positivi o negativi) di 0,97 per l'emivrania [67].

Di seguito, si approfondiscono i più comuni algoritmi del ML supervisionato.

- *k- Nearest Neighbor*: È un algoritmo di classificazione non parametrico. Assegna ad un punto campione non etichettato, la classe del più vicino di un

¹³ Per Fuzzy, si intende un aspetto "sfocato" dell'elemento a cui non è possibile attribuire una categoria specifica.

insieme di punti precedentemente etichettati [70]. *k-Nearest Neighbour* non prevede la creazione di un modello di classificazione per etichettare le istanze di test, ma conserva in un grafico le istanze di training e successivamente, utilizzando proprio questo grafico, riesce ad effettuare la classificazione delle istanze di test. Il grafico utilizzato per il salvataggio delle istanze di training è un grafico n-dimensionale, in cui n corrisponde al numero di attributi che caratterizzano le istanze appartenenti alla collezione dei dati di training. Nel grafico le istanze vengono rappresentate come dei punti. La classificazione, secondo questa tecnica, avviene rappresentando l'istanza di test nel grafico e scegliendo i k punti più vicini a questa. All'istanza di test viene assegnata il valore della classe più presente tra quelle delle k istanze di training più vicine. È adatto per classi multimodali e applicazioni in cui l'oggetto può avere molte etichette. Inoltre, le prestazioni dipendono dalla selezione di un buon valore di 'k'. Non esiste un modo di principio per scegliere "k", se non attraverso tecniche computazionali costose come la convalida incrociata. È influenzato negativamente dai dati irrilevanti. Le prestazioni variano anche in base alle dimensioni poiché tutti i dati devono essere rivisitati [71], [72]. Gli autori *Armañanzas et al.* (2013) [73], *Jen et al.* (2012) [74], *Bagui et al.* (2003) [75], hanno utilizzato il kNN nei rispettivi modelli predittivi della diagnosi medica. Tuttavia, con questo algoritmo se il set di dati di addestramento è grande richiede molto tempo quando viene elaborato durante la classificazione di nuovi dati e questo processo richiede un tempo di classificazione più lungo. Dal lavoro degli autori citati, l'accuratezza della classificazione è ciò che vorrebbero ottenere invece del tempo di classificazione poiché l'accuratezza della classificazione è più importante nella diagnosi medica.

- *Support Vector Machine*: È un algoritmo complesso, ma può fornire un'elevata precisione. *Support Vector Machine* (SVM) consente di classificare collezioni di dati sia lineari che non lineari. Una SVM rappresenta tutte le istanze della collezione dei dati di training su un piano formato da un numero di assi (dimensioni) pari al numero degli attributi che costituiscono le istanze. Le tre principali caratteristiche di un classificatore SVM sono: linee o iperpiani, a seconda se il classificatore presenta, rispettivamente, un grafico bidimensionale o n-dimensionale, margini e vettori di supporto. Una linea o

un iperpiano costituisce un “confine” che permette di classificare le istanze appartenenti a classi diverse dividendole tra loro. Un margine è una distanza tra le due istanze di classi diverse più vicine. Invece, i vettori di supporto corrispondono alle istanze più difficili da classificare per una SVM, poiché sono quelle che si trovano all'interno dei margini di un iperpiano. Una SVM si applica in maniera differente a seconda della tipologia delle istanze di una collezione di dati da classificare, linearmente separabili e non. In presenza di istanze separabili in maniera lineare, occorre trovare tra tutte le rette o tra tutti gli iperpiani che le separano tra le diverse classi quelle che massimizzano il valore del margine. Infatti, viene selezionata la retta o l'iperpiano con valore di margine massimo, poiché consente di minimizzare l'errore di classificazione. Per quanto riguarda collezioni di dati separabili in maniera non lineare il procedimento di classificazione è più complesso, in quanto occorre operare in due fasi separate ed estendere il comportamento del precedente approccio. Nella prima fase le istanze della collezione di dati vengono mappate su uno spazio dimensionale più grande in modo da renderle separabili in maniera lineare. Infine, nella seconda fase è possibile utilizzare l'approccio precedente di ricerca una linea o un iperpiano che massimizza la grandezza del margine, dato che adesso le istanze sono linearmente separabili. Si basano sul concetto di massimizzare la distanza minima dall'iperpiano al punto campione più vicino [76].

A differenza di k-NN, l'accuratezza e le prestazioni sono indipendenti dalla dimensione dei dati, ma dal numero di cicli di training. Particolarmente popolare nei problemi di classificazione del testo in cui gli spazi dimensionali molto elevati sono la norma. La complessità rimane inalterata dal numero di caratteristiche. È robusto per dati ad alta dimensione e ha una buona capacità di generalizzazione. Tuttavia, la velocità di training è inferiore e le sue prestazioni dipendono dalla scelta dei parametri [77]. Poiché la selezione dei parametri influisce sulle prestazioni, viene utilizzata una tecnica nota come Particle Swarm Optimizer (PSO) per la selezione dei parametri ottimali. Questo modello ibrido è noto come PSO-SVM [30]. Le applicazioni includono la selezione delle funzionalità e la classificazione delle immagini. *Gudadhe et al.* (2010) [78], svilupparono un sistema di diagnosi usando il

SVM per la classificazione delle malattie cardiache e ottenne un'accuratezza del 80.41%.

- *Decision Trees e Random Forest*: Gli Decision Trees (DT) sono facili da interpretare e spiegare, possono gestire facilmente le interazioni tra le funzionalità. Poiché non è parametrico, i valori anomali non influiscono molto sul modello e può gestire dati linearmente inseparabili. Alcuni algoritmi famosi sono: ID3, C4.5, C5.0 e CART secondo diversi criteri di suddivisione come Gini Coefficient, Gain Ratio e Info Gain [79]. Gli DT possono gestire una varietà di dati (nominali, numerici, testuali), valori mancanti e attributi ridondanti; avere una buona capacità di generalizzazione; sono robusti ai dati irrilevanti, forniscono prestazioni elevate per uno sforzo computazionale relativamente piccolo. Un DT prevede la costruzione di un albero che presenta tre principali elementi: i nodi, gli archi e le foglie [52]. Un nodo è costituito da un attributo di split, attributo scelto tra quelli dei dati di training, ovvero il miglior attributo locale in grado di produrre una distribuzione omogenea dei dati di training, in accordo con le etichette di classe. Ogni arco consente la connessione tra un nodo padre e un nodo figlio e presenta uno specifico valore dell'attributo di split presente nel nodo padre. Ogni foglia, invece, rappresenta un'etichetta di classe. Gli DT possono essere facilmente convertiti in regole di classificazione. Gli autori *Hashi et al.* (2017) [80], hanno implementato un sistema che utilizza gli algoritmi DT e kNN come modello di classificazione supervisionato. Il risultato sperimentale dimostra che DT fornisce una migliore accuratezza per la diagnosi del diabete.

Tuttavia, è difficile gestire dati ad alta dimensione con DT, sebbene il tempo di calcolo sia inferiore, viene impiegato un tempo considerevole per costruire l'albero. Questi usano un approccio "divide et impera" che funziona bene se esistono pochi attributi altamente rilevanti ma non molto bene se sono presenti molte interazioni complesse. Gli errori si propagano attraverso gli alberi, il che diventa un problema serio all'aumentare del numero di classi [81], [82]. Inoltre, man mano che l'albero cresce, il numero di record nei nodi foglia potrebbe essere troppo piccolo per prendere decisioni statisticamente significative sulla rappresentazione della classe. Per cui è stato sviluppato il modello Random Forest (RF). Questo metodo opera addestrando un numero

di alberi decisionali e restituendo la classe con la maggioranza su tutti gli alberi nell'insieme [83]. Le RF, di solito leggermente più avanti delle SVM, sono le vincitrici di molti problemi di classificazione. Sono veloci, scalabili, robusti all'influenza dei dati irrilevanti, non si adattano eccessivamente, facili da interpretare e visualizzare senza parametri da gestire. Tuttavia, all'aumentare del numero di alberi, l'algoritmo diventa lento per la previsione in tempo reale.

- *Neural Networks*: sono dispositivi computazionali che si basano vagamente sulla struttura neuronale, sul metodo di elaborazione e sulla capacità di apprendimento del cervello umano, ma su scale molto più piccole. Questa tecnica è applicabile a problemi in cui le relazioni possono essere non lineari o piuttosto dinamiche. Le Neural Networks (NN) forniscono una potente alternativa alle tecniche convenzionali che sono spesso limitate da forti assunzioni di normalità, linearità, indipendenza dalle variabili, ecc. Le NN sono una tecnica di classificazione la cui struttura è composta da un insieme di nodi interconnessi tra loro che ne costituiscono l'unità computazionale. All'interno di una NN sono sempre presenti due tipologie di nodi: nodi di input, prendono in ingresso le istanze sia per la fase di apprendimento della rete che per quella di classificazione, e nodi di output che assegnano la classe all'istanza sotto analisi. Oltre a questi due tipi di nodi possono essere presenti anche degli altri, chiamati nodi nascosti, organizzati in vari livelli. Ad ogni connessione tra i vari nodi che formano una NN è associato un peso che inizialmente assume un valore casuale, ma durante la fase di apprendimento della rete viene continuamente aggiornato a seguito di un errore di classificazione di un'istanza rispetto al valore di classe previsto. Inoltre, ad ogni nodo è associato anche un valore di offset. In una NN gli input inviati ai nodi di input sono costituiti dal valore degli attributi dell'istanza che deve essere analizzata. L'output di questo primo livello della rete rimane invariato, poiché in uscita dai nodi di input sono presenti gli stessi valori che vengono forniti in input. In ogni nodo appartenente ai livelli successivi al primo, nodi nascosti e nodi di output, avviene l'effettiva computazione. Infatti, gli input di questi livelli corrispondono agli output dei livelli precedenti in cui però bisogna considerare il peso associato al collegamento tra i due nodi ed un

valore caratteristico del nodo, l'offset. Humar et al. (2008) [58], progettarono un sistema di classificazione delle malattie cardiache utilizzando una neural network con l'integrazione della Fuzzy Logic. La classificazione ottenne una accuratezza del 87.4%. Perceptron è la forma più semplice di NN, utilizzata per la classificazione di pattern detti linearmente separabili. Consiste di un singolo neurone con pesi regolati. La presenza di caratteristiche irrilevanti rende la formazione molto inefficiente e poco pratica. Il Multi-Layer Perceptron (MLP), è il classificatore NN più utilizzato, in grado di modellare funzioni complesse ed è robusto per input e dati irrilevanti, [84], [85] [87].

- *Bayesian Networks*: è un modello grafico per rappresentare le relazioni di probabilità tra insiemi di variabili. Inizialmente, la rete (Directed Acyclic Graph) deve essere rappresentata, quindi vengono determinati i parametri che rendono difficile l'implementazione senza un'opinione di esperti. L'informazione a priori sul problema può essere rappresentata come relazione strutturale tra le sue caratteristiche [71]. Tuttavia, le Bayesian Networks (BN) non hanno molto successo con insiemi di dati ad alta dimensione perché reti di grandi dimensioni non sono realizzabili in termini di tempo e spazio.
- *Naïve Bayes*: è un algoritmo probabilistico, calcola la probabilità di ogni etichetta per un determinato oggetto osservando le sue caratteristiche. Poi, sceglie l'etichetta con la probabilità maggiore. Per calcolare la probabilità delle etichette usa il teorema di Bayes. A differenza delle Neural Networks o SVM, non ci sono parametri liberi da impostare, il che semplifica notevolmente NB [86]. Non è applicabile quando si devono prendere in considerazione le interazioni tra le caratteristiche [72], [81], [82].
- *Regressione logistica*: modello statistico utilizzato quando al dataset si adatta una curva logistica [83]. Questa tecnica viene applicata quando la variabile dipendente o la variabile target è dicotomica. A differenza di Decision Trees o SVM, esiste una buona interpretazione probabilistica e il modello può essere aggiornato per acquisire facilmente nuovi dati. Poiché restituisce probabilità e le soglie di classificazione possono essere facilmente modificate. Il modello logistico può essere un'alternativa per Discriminant Analysis [77]. Ha meno assunzioni: nessuna ipotesi sulla distribuzione delle variabili indipendenti, nessuna relazione lineare tra i predittori e la variabile target deve essere data.

Può gestire effetti di interazione, effetti non lineari e termini di potenza. Tuttavia, sono necessarie grandi dimensioni del campione per ottenere risultati stabili.

- *Discriminant Analysis*: Combina le variabili in modo tale da massimizzare le differenze tra i gruppi predefiniti. Le variabili possono essere combinate in modo lineare o quadratico fornendo rispettivamente l'analisi discriminante lineare (ADL) e l'analisi discriminante quadratica (ADQ). ADL viene applicato quando si assume che i predittori siano distribuiti normalmente e che la covarianza di ciascuna classe sia la stessa. ADQ non ha tali ipotesi. ADQ separa le classi utilizzando una superficie quadratica (cioè una sezione conica). Entrambi sono usati come classificatori. Tuttavia, l'ADL è più usato come tecnica di riduzione dimensionale.

1.3.2 Logica Fuzzy

La logica fuzzy (dall'inglese letteralmente sfocato o confuso) fu introdotta da Lofti Zadeh nel 1965, ed è un'estensione della logica booleana classica. La logica fuzzy, infatti, si basa sulla teoria degli insiemi sfocati. Gli insiemi sfocati indicano una classe di oggetti per i quali la proprietà di appartenenza alla classe stessa è rappresentata da una funzione continua piuttosto che da una funzione del tipo vero-falso/sì-no. Per questi insiemi non è più valido il principio di non contraddizione secondo cui un elemento appartiene oppure non appartiene ad una determinata classe assegnandoli un grado di appartenenza uno o zero, ma piuttosto vi appartiene con un grado che varia con continuità fra zero e uno [87].

Per comprendere meglio la logica fuzzy, è necessario comprendere i semplici insiemi che la fanno funzionare. Sappiamo bene che un "set" è un gruppo di elementi con una definizione o identità condivisa. Gli elementi fuzzy non appartengono a un singolo set, piuttosto, hanno gradi di appartenenza (degrees of membership) per ciascuno dei suoi insiemi. Un grado di appartenenza indica semplicemente quanto un elemento appartiene a un insieme. Ad esempio, un elemento fuzzy può avere un'appartenenza di 35% nell'insieme A e 65% nell'insieme B. Esiste contemporaneamente in entrambi gli insiemi, creando così una sovrapposizione. La capacità di esistere in più insiemi contemporaneamente rende la logica fuzzy perfettamente adatta per gestire input linguistici naturali. Tali input linguistici naturali che gli esseri umani potrebbero fornire

possono essere, ad esempio, in riferimento alla temperatura "molto caldo", "caldo" o "leggermente caldo". La logica fuzzy può definire arbitrariamente i livelli di parole non avendo confini discreti, ma avendo gradi di appartenenza e sovrapposizioni creando insiemi fuzzy. Si parla allora di insiemi fuzzy come estensione degli insiemi tradizionali (crisp sets), in quanto gli insiemi fuzzy hanno una frontiera che non è più una linea netta di demarcazione tra gli elementi che vi appartengono e quelli che non vi appartengono, ma un'area in cui si trovano elementi classificabili come appartenenti all'insieme con un certo grado[88].

Sulla base delle specifiche proprietà e delle operazioni sugli elementi, gli insiemi fuzzy (o fuzzy sets) sono usati per trattare l'incertezza e per rappresentare la conoscenza mediante regole [89], infatti la logica fuzzy permette l'interpretazione dei dati, sulla base di variabili linguistiche¹⁴ predefinite, secondo opportune regole. Le regole (fuzzy rules) sono un modo per mettere in relazione la descrizione, in termini linguistici, di una situazione con un'azione da svolgere, cioè una regola è scritta come "*If situation Then conclusion*". La situazione è chiamata antecedente o premessa, mentre l'azione da svolgere è chiamata conclusione. Quanto appena detto si traduce in diverse composizioni delle funzioni di appartenenza. La logica fuzzy definisce i meccanismi inferenziali con cui arrivare all'azione da svolgere (per esempio, segnalazione o meno dell'allarme di avvenuto incidente), costituendo quindi un sistema di inferenza: Fuzzy Inference System (FIS). Un FIS è composto da blocchi, come mostrato nella successiva Figura 5, facendo riferimento al metodo Mamdani, il metodo di inferenza più usuale. Le fasi fondamentali nell'approccio fuzzy sono: la definizione delle funzioni di appartenenza, la fuzzificazione, l'inferenza e l'output fuzzy, vedi Figura 5.

¹⁴ Una variabile linguistica è una variabile fuzzy, i cui valori sono termini linguistici (per esempio, "piccolo", "medio", "grande").

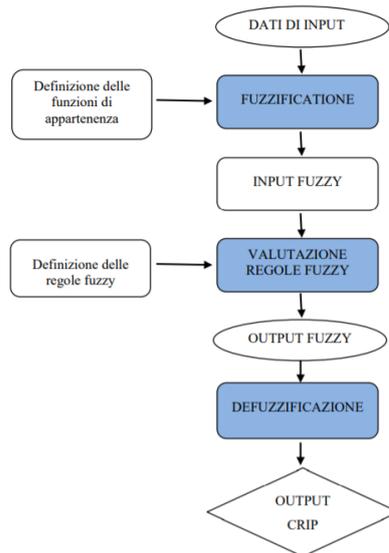


Figura 5. Le principali fasi in un sistema fuzzy

Il primo passo dell'analisi fuzzy riguarda l'individuazione delle variabili in ingresso al sistema. La definizione delle funzioni di appartenenza è il passo successivo su cui poi si basano tutte le altre operazioni. Tali funzioni, rappresentative degli insiemi fuzzy, possono assumere forme differenti (trapezoidali, triangolari, gaussiane, ecc.) a seconda delle situazioni e per convenzione possono assumere valori compresi tra 0 e 1. La fase successiva, ossia quella di fuzzificazione, consiste nell'attribuire ad un dato parametro in ingresso il suo grado di appartenenza ai diversi insiemi fuzzy in cui è stato suddiviso il dominio di esistenza del parametro; con tale operazione si normalizzano tutti i dati nell'intervallo $[0,1]$, perciò sono possibili anche confronti tra quantità tra loro diverse e misurate in scale differenti. L'inferenza è la fase in cui sono applicate le regole di combinazione tra gli insiemi fuzzy e da cui è possibile dedurre un risultato. Le regole sono delle semplici espressioni linguistiche che sono convertite in formalismo matematico con il linguaggio "if...then" della logica stessa. L'output fuzzy è anch'esso un valore di appartenenza che può essere usato sia "puro", come proprietà qualitativa, sia "defuzzificato", come numero reale compatibile con approcci non-fuzzy [90].

Insiemi fuzzy

Un fuzzy set è definito come una classe di oggetti, con un continuum di gradi di appartenenza, caratterizzato da una funzione di appartenenza (caratteristica) che assegna ad ogni oggetto un grado di appartenenza compreso fra zero ed uno. L'insieme, sia per la teoria classica sia per quella degli insiemi sfocati, è visto come un aggregato di elementi caratterizzati da una certa proprietà. La differenza nella definizione tra le due teorie

riguarda il tipo di relazione che lega gli oggetti all'insieme. Infatti, la teoria classica parte dal presupposto che si possano distinguere nettamente gli oggetti che appartengono ai diversi insiemi, la teoria degli insiemi sfuocati, invece, considera delle classi dai confini incerti, e ammette che gli oggetti appartengono ad esse solo in una certa misura. Tale sfumatura si ripercuote sui valori assunti dalla funzione di appartenenza degli insiemi. Nonostante, sia abbastanza semplice concepire un insieme come insieme sfuocato, non è invece immediato determinare e definire la funzione di appartenenza. Infatti, proprio nell'ambito della classificazione automatica, ad esempio, sono stati messi a punto algoritmi in grado di individuare sotto-insiemi fuzzy e di determinare le funzioni di appartenenza. Il grado di appartenenza ad una determinata funzione individua quanto un elemento appartiene ad un generico insieme A, inteso anche come il grado di verità dell'affermazione. A questo punto si può affermare che i numeri fuzzy sono strumenti utili, specialmente quando si lavora con quantità numeriche imprecise. L'utilizzo della teoria degli insiemi fuzzy, quindi, permette di rappresentare tali numeri correttamente, n quanto nasce dall'esigenza di trovare una strumentazione logico-matematica, capace di unire le capacità proprie del linguaggio naturale, di rappresentazione della polivalenza e dell'indeterminazione, con i vantaggi della formalizzazione algebrica e della rappresentazione numerica [91].

Funzioni di appartenenza

Una funzione di appartenenza è una funzione che associa un valore (generalmente numerico) al grado di appartenenza all'insieme. Per convenzione, il numero reale che rappresenta il grado di appartenenza $[\mu(x)]$ assume valore 0 quando l'elemento non appartiene all'insieme, 1 quando vi appartiene del tutto.

Le funzioni di appartenenza riferite ad una variabile, inoltre, possono essere in numero maggiore di una. L'unico fondamento delle funzioni di appartenenza è che esse devono essere in grado di descrivere nel modo più realistico possibile il fenomeno da modellare. Le funzioni di appartenenza possono assumere diversi andamenti: lineari (come le funzioni triangolari o trapezoidali), gaussiane e a campana. Nella definizioni delle funzioni di appartenenza è necessario rappresentare in modo soddisfacente l'andamento del fenomeno analizzato e individuare il numero di parametri ossia i gradi di libertà di ogni curva. Tra le funzioni di appartenenza quelle maggiormente utilizzate sono le triangolari (Figura 6) e le trapezoidali (Figura 7): le prime sono caratterizzate da un

andamento triangolare mentre le seconde da uno trapezoidale. Il vantaggio di queste funzioni sta nella loro semplicità [92].

La funzione di appartenenza triangolare dipende da tre parametri scalari a , b e c ed è data dalla seguente espressione:

$$f(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$$

mentre quella trapezoidale dipende da quattro parametri scalari (a , b , c e d), come mostrato nella seguente espressione:

$$f(x; a, b, c, d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right)$$

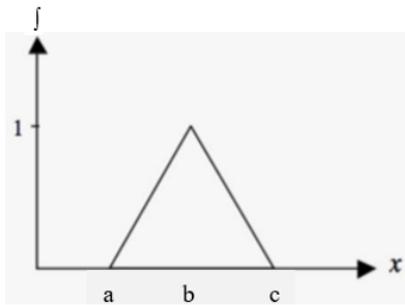


Figura 6. Funzione di appartenenza triangolare

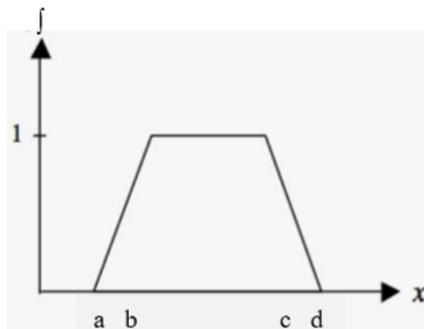


Figura 7. Funzione di appartenenza trapezoidale

Le Regole fuzzy

Le regole fuzzy rappresentano un modo per mettere in relazione una descrizione di un fenomeno in termini linguistici con un'azione da svolgere, espressa anch'essa in termini linguistici. In pratica, tali regole sono rappresentazioni di inferenze logiche fatte su

composizioni di proposizioni fuzzy. Ad esempio, una regola può essere: se la temperatura (T) è alta e la pressione (P) è alta, allora la quantità di gasolio (Qg) da fornire al bruciatore dell'impianto deve essere ridotta di molto, in termini logici si scrive: if (T is Alta) and (P is Alta) then (Qg is Ridotta). Con l'ausilio di regole del tipo "if...then", sarà possibile ricavare l'output per qualsiasi problematica perché con tale istruzione vengono associati gli ingressi alle uscite [93].

Fuzzificazione

La fuzzificazione è una delle fasi dell'approccio fuzzy necessaria a determinare l'output finale. In particolare, è il procedimento attraverso il quale le variabili di ingresso vengono convertite in misure fuzzy in funzione della loro appartenenza a determinate classi come ad esempio Molto bassa, Basso, Medio, Alto, Molto alta. Tale conversione da grandezze deterministiche a fuzzy viene effettuata attraverso le funzioni di appartenenza predefinite per quelle classi. Quindi data una funzione di appartenenza μ_A e un valore x° la conversione di x° in un insieme fuzzy consiste nel complemento dell'insieme A in corrispondenza dell'ordinata $\mu_A(x^\circ)$ ovvero nell'insieme. Se consideriamo una funzione di appartenenza μ_A e un valore x° , la conversione di x° in un insieme fuzzy consiste nel complemento dell'insieme A in corrispondenza dell'ordinata $\mu_A(x^\circ)$ ovvero nell'insieme [94]:

$$A' = \{x | \mu_{A'}(x) = \min(\mu_A(x^\circ), \mu_A(x))\}$$

Se il valore x° intercetta più di una funzione di appartenenza si avrà un grado di appartenenza per ciascuna funzione pari al valore dell'intercetta con l'asse delle ordinate, vedi Figura 8.

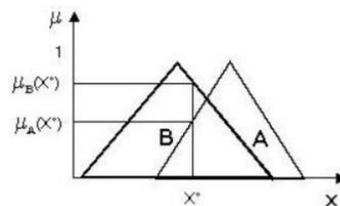


Figura 8. Funzione di appartenenza della variabile X.

Infatti, se consideriamo tre variabili di ingresso P, T, U e un'uscita Z, ognuna delle quali con funzioni di appartenenza Basso, Medio ed Alto, vedi Figura 9.

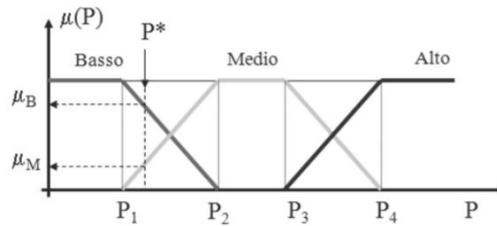


Figura 9. Processo di fuzzificazione per la variabile P.

Una prima regola tipo di questo sistema fuzzy sarà del tipo seguente: if (P è Basso) and (T è Medio) and (U è Alto) then (Z è Medio), dove l'operatore tra gli antecedenti è l'AND, cioè questa regola ha un peso se e solo se ognuno dei tre antecedenti (P è Basso), (T è Medio) e (U è Alto) sono verificati. Se invece dell'operatore AND ci fosse stato l'OR una seconda regola sarebbe stata: if (P è Basso) or (T è Medio) or (U è Alto) then (Z è Medio), quindi avrebbe restituito un risultato solo se uno dei tre antecedenti si fosse verificato.

L'operazione che deve compiere il sistema fuzzy nella fase di fuzzificazione consiste nell'individuare i fuzzy set delle variabili di input in funzione del valore attuale delle variabili stesse (valore delle variabili nell'istante desiderato) e calcolarne il grado di appartenenza a ciascun insieme. Guardando all'esempio in Figura 9 relativo alla variabile P, si ha che per il valore di input pari a P* l'operazione di fuzzificazione individua l'appartenenza alla classe Basso con grado pari a μ_B e alla classe Medio pari a μ_M , per cui secondo la prima regola il grado di appartenenza che la soddisfa è μ_B [94].

Inferenza fuzzy

L'inferenza fuzzy è il processo di mappatura partendo da uno spazio di input e giungendo ad un opportuno spazio di output utilizzando le regole fuzzy. Tali regole non sono rappresentate da complicati modelli matematici ma da semplici espressioni linguistiche, che vengono convertite in formalismo matematico con il linguaggio "if-then-else" della logica fuzzy. Solitamente le regole si compongono di due parti (if-then): l'antecedente che definisce la condizione e il conseguente che definisce l'azione. In realtà è possibile definire anche una regola "else" che entra in causa nel momento in cui nessuna delle regole è stata soddisfatta ovvero quando le "then" sono debolmente soddisfatte. Quindi l'inferenza è quel procedimento attraverso cui viene determinato l'output fuzzy a partire dagli antecedenti [87].

Defuzzificazione

L'output ottenuto dall'inferenza fuzzy spesso risulta inutilizzabile, quindi occorre pertanto riconvertirlo in un valore deterministico attraverso un'operazione detta di defuzzificazione. La defuzzificazione è il processo inverso della fuzzificazione, ovvero dato un insieme fuzzy restituisce un valore numerico rappresentativo di quell'insieme. Esistono diversi metodi di defuzzificazione e la scelta del metodo più opportuno dipende, nei sistemi di supporto alle decisioni, dal contesto della decisione stessa. Nel caso di decisioni di carattere quantitativo è preferibile ricorrere a metodi di defuzzificazione che scelgano un valore intermedio che tenga conto di tutto il risultato sfocato dell'output ("miglior compromesso"), mentre nel caso di decisioni di tipo qualitativo risultano più adatti a metodi che scelgano il valore tipico del termine linguistico dell'output con maggiore grado di attivazione ("maggior risultato possibile"). Si tratta di due metodi che, per loro natura forniscono risultati con caratteristiche molto differenti: l'uno propone una mediazione tra i vari risultati, l'altro premia "il più forte" [95].

Tuttavia, l'implementazione di tale logica ha buone potenzialità nel contesto sanitario, infatti può aiutare nella classificazione di descrizioni vaghe fornite dal personale sanitario, come ad esempio per valutare delle diagnosi e dire se i pazienti sono malati o per determinare il rapporto di rischio di una malattia [96].

1.4 Il Controllo di Gestione nelle aziende sanitarie

Prima di passare a discutere della progettazione di una piattaforma di BI a supporto del controllo di gestione, è bene precisare il contesto in cui è allocato tale sistema di BI in grado di soddisfare a pieno le esigenze di valutazione e monitoraggio delle attività.

1.4.1 Gli obiettivi e finalità del controllo di gestione

Nell'ambito del processo di "aziendalizzazione" (d. lgs. 502/1992), che da circa un decennio interessa la Pubblica Amministrazione, assumono rilevanza i concetti di efficienza ed efficacia, in relazione ai quali e per la realizzazione degli stessi si afferma un concetto di controllo.

In un'azienda il controllo di gestione è il sistema teso a guidare la gestione verso il conseguimento degli obiettivi stabiliti in sede di pianificazione operativa, rilevando,

attraverso la misurazione di appositi indicatori, lo scostamento tra obiettivi pianificati e risultati conseguiti e informando di tali scostamenti gli organi responsabili, affinché possano decidere e attuare le opportune azioni correttive. Il controllo di gestione è quindi:

- Uno strumento in grado di fornire alla direzione dell'impresa informazioni utili per comprendere meglio la situazione aziendale al fine di decidere in maniera più efficace;
- Un processo di raccolta, analisi e diffusione di informazioni utili per dirigere un'impresa;
- Il feedback sull'andamento dell'azienda poiché individua le funzioni e i reparti che hanno contribuito al raggiungimento degli obiettivi, previene situazioni difficili e consente di intervenire con correzioni gestionali per migliorare l'utilizzo delle risorse;
- L'attività di guida e orientamento della gestione, in grado di assicurare che le risorse economiche ed i fattori produttivi a disposizione dell'azienda siano impiegati in modo efficace ed efficiente coerentemente agli obiettivi prestabiliti.

Il controllo di gestione risulta, dunque, uno strumento di government, di monitoraggio e di valutazione. Esso risponde ad esigenze informative interne ed è organizzato e utilizzato da ogni impresa nel modo più appropriato rispetto alla tipologia di attività svolta, allo stile direzionale del management. Lo scopo del controllo di gestione non è quello di sanzionare i comportamenti difformi dalle regole quanto, piuttosto, quello di aiutare il personale ad indirizzare il proprio comportamento verso il conseguimento degli obiettivi aziendali. Il sistema di controllo di gestione è strettamente connesso al sistema di pianificazione, tanto che di solito, sia in ambito teorico che nella pratica aziendale, si parla di "sistema di pianificazione e controllo".

Il sistema di controllo di gestione è solitamente legato:

- con il sistema di contabilità analitica che permette di ripartire i costi aziendali tra i singoli centri di responsabilità e i singoli obiettivi aziendali;
- con il sistema informativo che permette di raccogliere, archiviare e trattare i dati relativi a costi e indicatori e distribuire le relative informazioni ai soggetti utilizzatori. Nell'ambito dei sistemi informativi aziendali particolarmente utili ai fini del controllo di gestione sono i sistemi di reportistica.

Dal punto di vista organizzativo, il sistema di controllo di gestione è normalmente progettato e gestito da un organo posto in staff al vertice aziendale (o al vertice di una divisione aziendale). La persona preposta a questo organo viene a volte qualificata controller, con un termine importato dalla pratica aziendale anglosassone.

Le fasi del controllo di gestione si possono riassumere in:

- Pianificazione: elaborazione delle strategie aziendali di medio-lungo periodo e dei piani a più breve termine (spesso sotto forma di budget);
- Attuazione: mobilitazione delle risorse aziendali per il raggiungimento degli obiettivi fissati nei piani;
- Controllo: verificare che i piani siano stati realizzati e che gli obiettivi siano stati raggiunti;
- Correzione: revisione dei piani per individuare nuove strategie più efficaci e nuovi obiettivi di miglioramento.

Nell'ambito delle organizzazioni sanitarie il controllo di gestione, fornisce informazioni utili non solo alla Direzione Generale, ma anche a tutti coloro che hanno responsabilità di programmazione, organizzazione e gestione dei servizi sanitari e, quindi, anche ai medici ed agli altri operatori sanitari dirigenti e non. Inoltre, esso agevola il processo di aziendalizzazione, favorendo l'affermazione del management in sanità, tenendo presente che l'obiettivo per chi opera in questo settore è l'erogazione delle prestazioni, rientranti nei Livelli Essenziali di Assistenza (LEA) che deve essere garantita nel rispetto di vincoli di efficacia, efficienza ed economicità della gestione. Pertanto, il Controllo di Gestione, in campo sanitario acquista preminenza sia in termini di contenimento dei costi, sia al fine di porre le opportune premesse per l'adeguato soddisfacimento dei bisogni. In tal senso, il controllo di gestione consiste "nella procedura diretta a verificare lo stato di attuazione degli obiettivi programmati e, attraverso l'analisi delle risorse acquisite e della comparazione tra i costi e la quantità/qualità dei servizi offerti, la funzionalità dell'organizzazione dell'ente, l'efficacia, l'efficienza, il livello di economicità e di qualità dell'attività di realizzazione dei predetti obiettivi". Tutto ciò per consentire ai dirigenti di ottimizzare, anche mediante tempestivi interventi di correzione, il rapporto tra costi e risultati. Si evince che, da un lato risulta indispensabile fornire servizi e prestazioni sanitarie appropriate, cioè in grado di soddisfare i cittadini, dall'altro lato è parimenti indispensabile il razionale impiego delle risorse e, quindi, realizzare gli aspetti

dell'economicità e dell'efficienza, nonché la necessità di gestire con appropriatezza gli aspetti organizzativi, amministrativi e, in linea più generale, di supporto all'attività assistenziale (qualità gestionale). In particolare, nelle aziende sanitarie il cui fine è l'erogazione dei servizi essenziali per gli utenti, il giudizio di economicità non può esaurirsi nella determinazione del risultato d'esercizio, ma richiede il ricorso a ulteriori elementi di valutazione, tra cui assumono particolare rilievo le verifiche relative al grado di soddisfacimento dei bisogni e alla qualità dei servizi erogati. Alle aziende di oggi non è più sufficiente conoscere il reddito d'esercizio, o nel caso del no profit, di pareggiare i conti (entrate uguali uscite), ma si richiede un governo del valore, quel valore che si riversa nel servizio offerto all'utente e che l'utente percepisce, quel valore incorporato nelle risorse materiali costituenti l'aspetto tecnico dell'azienda e nelle risorse immateriali di conoscenze e competenze, ma anche di comportamenti.

1.4.2 Architettura del sistema di Controllo di Gestione

Le aziende ospedaliere operano in un ambiente altamente imprevedibile e dinamico, caratterizzato da un'elevata complessità strutturale. In questo contesto il processo decisionale, che presenta un alto livello di decentramento, deve essere supportato da un sistema di controllo in una logica di feed-forward, attraverso la quale il processo di controllo viene utilizzato per indirizzare la gestione non tanto verso gli obiettivi precedentemente stabiliti, bensì verso le condizioni che l'ambiente sembra prospettare per il futuro. Oppure, altrimenti, da un sistema di controllo a funzione risposta, grazie al quale i dirigenti, ogni volta che sorge un problema, utilizzano i dati forniti dal sistema informativo per sviluppare calcoli economici e promuovere una soluzione più adatta in relazione agli obiettivi definiti.

Il sistema di controllo deve utilizzare strumenti in grado di fornire idonee informazioni ai responsabili per rispondere agli stimoli che l'ambiente esterno esercita continuamente sull'organizzazione; informazioni che permettano di valutare adeguatamente le alternative future per preordinarne l'esecuzione. Lo stile di controllo deve assumere una logica di responsabilizzazione diffusa, affinché assicuri flessibilità nei comportamenti.

Il sistema di controllo si compone di due dimensioni che sono: quella organizzativa e quella informativo-contabile. È necessario mantenere un corretto equilibrio tra loro, evitando di attribuire un'eccessiva rilevanza agli strumenti tecnico-contabili, in modo tale da evitare il rischio che la componente non amministrativa, consideri il controllo di

gestione come uno strumento lontano dalla professionalità medica e privo di una valenza tecnica capace di essere di supporto all'analisi operativa dell'attività svolta.

La progettazione di un sistema di controllo in sanità deve basarsi sul riconoscimento delle variabili chiave e sull'individuazione delle responsabilità ad esse collegate, così da formare quel comportamento organizzativo utile a valorizzare gli aspetti relativi all'efficacia, alla qualità e alla soddisfazione dell'utente. È importante che consideri la congiunzione nell'erogazione del servizio, che avviene nel momento in cui ad un servizio principale se ne associano altri, così come la congiunzione tecnica dei processi, anche se presente in misura ridotta nelle aziende ospedaliere.

I meccanismi di controllo devono porsi in maniera diversa a seconda che l'attività controllata sia di back-office o di front-office. L'erogazione di certi servizi avviene, infatti, in tempi brevi, il meccanismo di feed-back non consente di evitare gli effetti negativi derivanti dalla mancata soddisfazione del cliente, quindi è opportuno mettere in atto meccanismi di tipo anticipato.

L'attenzione rivolta dal processo di controllo solamente sul servizio non è sufficiente per esaurire le esigenze informative della gestione poiché talvolta risulta difficile creare relazioni specifiche tra risultati ottenuti e risorse impiegate. È quindi necessario analizzare anche altre dimensioni, come quella territoriale o quella della tipologia dei clienti per conoscere al meglio come certi servizi o comportamenti possono influire sui risultati.

La quasi impossibilità di standardizzare i servizi offerti dalle aziende ospedaliere, incide in maniera notevole sulla possibilità di creare correlazioni specifiche tra risorse e servizi offerti su cui basare giudizi di convenienza.

Il sistema di controllo deve essere, inoltre, uno strumento di diffusione degli orientamenti e della mentalità strategica a tutti i livelli, in modo da assicurare sia un coordinamento delle azioni e sia la conoscenza da parte degli operatori di quale comportamento risulti più consono per il raggiungimento dell'equilibrio complessivo, dal momento che esiste nell'erogazione del servizio un'elevata discrezionalità che porta alla soluzione del problema in base ad una soggettiva interpretazione legata alla competenza professionale. La partecipazione coordinata degli operatori è necessaria, poiché la soddisfazione dei bisogni richiede una serie di azioni legate tra loro da un rapporto di complementarità spazio-temporale, non si tratta, dunque, di ricercare solo un equilibrio economico, ma di

associargli il soddisfacimento del bisogno espresso, in modo da attribuire all'equilibrio anche un valore qualitativo.

Tramite il controllo di gestione s'intende realizzare una focalizzazione, un'integrazione e una convergenza dei singoli attori organizzativi nei riguardi degli obiettivi e delle finalità aziendali, ciò comporta delle conseguenze anche a livello operativo, poiché tale sistema non può trovare il proprio sviluppo nella gerarchia, bensì nell'utilizzo di altri meccanismi, come quello di appartenenza ad un gruppo. Questo fenomeno diventa rilevante in un'azienda ospedaliera, dal momento che sono presenti diversi gruppi professionali all'interno di essa. Il sistema affinché risulti efficace deve concentrarsi principalmente sullo sviluppo dell'autocontrollo, basato sulla condivisione dei valori e delle aspettative, rispetto al controllo amministrativo, che prevede soluzioni formalizzate e istituzionali di funzionamento, in modo da creare una maggiore motivazione dei singoli.

Nei sistemi di controllo delle aziende ospedaliere è importante considerare il ruolo e il peso che l'aspetto economico assume, ma focalizzarsi solamente su parametri economico-finanziari vuol dire avere una visione parziale dei vari aspetti che caratterizzano l'attività. Nelle combinazioni economico-sanitarie il risultato economico assume una rilevanza piuttosto sfumata come indicatore dell'andamento gestionale a causa delle condizioni operative in cui le aziende si trovano ad agire. L'azienda sanitaria deve perseguire i propri scopi di tutela della salute garantendosi la sopravvivenza e lo sviluppo delle attività con un efficiente impiego delle risorse ed un equilibrio economico. Si tratta, quindi, di raggiungere un equilibrio in funzione delle finalità. È necessaria, dunque, anche un'attenzione verso variabili diverse da quelle economico-finanziarie, che monitorano l'andamento di alcuni aspetti, quali ad esempio la soddisfazione del cliente, la qualità e l'ambiente. Il sistema deve controllare non solo l'efficienza, ma anche l'efficacia di breve e medio- lungo periodo e la qualità.

In base all'art.4 del D.lgs.286/99 ai fini della progettazione di un sistema di controllo di gestione, ciascuna amministrazione dovrà definire:

- le unità responsabili della progettazione e della gestione del controllo di gestione;
- le unità organizzative dove si intende misurare l'efficacia, l'efficienza e l'economicità dell'azione amministrativa (centri di responsabilità e di costo);
- la definizione degli obiettivi gestionali e dei soggetti responsabili;
- l'insieme dei servizi e delle finalità dell'azione amministrativa;

- le modalità di rilevazione e ripartizione dei costi tra le diverse unità organizzative, nonché gli indicatori specifici per misurare l'efficacia, l'efficienza e l'economicità.

Il controllo di gestione dovrà in ogni caso avere nel nuovo contesto manageriale valore di guida, di responsabilizzazione e di coordinamento in quanto il controllo di gestione è costituito da più funzioni distinte e, al tempo stesso, strettamente correlate. Il controllo deve quindi fungere da guida delle attività dei responsabili, da supporto per le decisioni, deve servire a monitorare l'andamento dell'attività e a coordinare l'attività dei diversi responsabili, infine deve servire anche per valutare le prestazioni dei manager.

1.4.3 Gli strumenti di Controllo di Gestione

I principali strumenti utilizzati dal controllo di gestione sono:

- gli indicatori di performance, ovvero tutti quelle analisi usate per confrontare tra loro diverse alternative di comportamento;
- la contabilità analitica dei costi;
- la programmazione e il controllo del budget.

Per l'obiettivo del progetto di ricerca e per evitare di sfociare dal contesto, ci concentriamo solo sugli indicatori di performance.

Gli indicatori non sono altro che delle variabili misurabili che servono a confrontare un fenomeno nel tempo (tra momenti diversi) e nello spazio (tra realtà diverse) o rispetto un obiettivo da raggiungere o da mantenere. Per essere considerati di valore, devono rispettare le seguenti caratteristiche:

- **riproducibilità:** a fronte dello stesso fenomeno, osservatori diversi o lo stesso osservatore in tempi diversi attribuiscono alla misura lo stesso valore o valori simili;
- **accuratezza o validità:** garantisce la corrispondenza tra valore osservato e valore "vero" del fenomeno;
- **sensibilità:** individua anche modeste variazioni del fenomeno, nel tempo e nello spazio;

- la specificità: coglie solo il fenomeno di interesse, senza influenze da fattori estranei;
- tempestività: prontezza nel rilevare i cambiamenti del fenomeno.

In ambito sanitario la vera sfida non è tanto avere a disposizione un ampio spettro di variabili misurabili, quanto dare loro un ordine e una corretta interpretazione. In tutto questo gli strumenti digitali come il DW e la BI giocano un ruolo fondamentale anche, e soprattutto, in ambito sanitario. Tra i principali indicatori in sanità riferiti al controllo di gestione troviamo:

- Processi aziendali interni: movimenti di magazzino, prodotti impiegati (medicinali, strumenti di lavoro, dispositivi, ecc.), le giacenze e il lead time dei fornitori. Tutti dati importanti che permettono di calcolare indici fondamentali come la rotazione di magazzino, i giorni di scorta degli articoli e il loro consumo medio.
- Controllo dei costi: una tra le difficoltà maggiori è l'indispensabile controllo costante del budget, del bilancio e delle marginalità. I fogli Excel risultano pesanti e dispersivi per tutti questi conti, mentre dei pratici cruscotti, report e grafici che evidenziano i dati rilevanti rendono la consultazione più smart e immediata, evidenziando immediatamente criticità e dati positivi.
- Risorse umane: verificare le presenze/assenze del personale può diventare un lavoro veloce se gestito in un unico cruscotto dove la direzione potrà così confrontare dati, estrarli e analizzarli con semplicità.
- Performance del pronto soccorso: è diventato importante per l'azienda ospedaliera determinare tutti i dati del pronto soccorso tra cui il numero di codici rossi/gialli/verdi/bianchi e quali sono i giorni in cui è più affollato per ricavare poi il tasso di ammissione e i tempi di attesa medi.
- Tasso occupazione posti letto: è l'indicatore non-economico più importante per valutare l'efficienza dell'attività ospedaliera. Calcolare la % di occupazione dei posti letto per reparto e nel tempo non è semplice in quanto bisogna incrociare i dati dei ricoveri e dei transiti dei pazienti tra reparti. Il risultato mi consente di capire quali reparti sono prossimi alla saturazione (tasso del 100%) e quali sono sottoimpiegati (tasso inferiore al 75%). A parità di risorse impiegate posso dunque rimodulare le risorse dei vari reparti riducendo le liste d'attesa dei pazienti in attesa di ricovero.

1.5 Metodologie per la gestione dei processi sanitari

L'analisi e lo studio di big data sanitari per ottenere la conoscenza implicita nei dati è fondamentale per migliorare la qualità e l'efficienza dei servizi sanitari. Ma, per gli esponenti del management sanitario, è sicuramente di ausilio anche l'utilizzo di metodologie di carattere manageriale, in grado di valutare la performance e gestire i processi sanitari.

Le inefficienze, che contribuiscono alle scarse prestazioni degli ospedali, risiedono quindi sostanzialmente nei fallimenti operativi, momenti in cui al dipendente non vengono fornite le attrezzature, le informazioni o le persone necessarie a soddisfare il compito assegnato. La difficoltà che si nasconde dietro l'eliminazione di tali inefficienze è che si manifestano come problemi su piccola scala, pertanto il loro impatto è considerato limitato; tuttavia questi, accumulati, generano una notevole influenza sul processo. Per le loro piccole dimensioni si rivelano infatti difficili da affrontare, poiché nella pratica, l'attenzione è diretta solo a problemi su larga scala, ma la loro elevata ricorrenza genera problemi soprattutto quando si raggiungono le ultime fasi del processo. Altre inefficienze sono collegate all'ambiente di lavoro, ideale per la proliferazione degli errori; infatti le attività svolte all'interno dell'ospedale sono scollegate e i dipendenti hanno una conoscenza specifica solo di ciò che fanno nella loro routine giornaliera. Questo implica un difficile passaggio di informazioni tra i professionisti e una difficile collaborazione tra di loro [97]. Tutto questo può essere riassunto in un unico grande problema: assenza di definizione dei processi; i dipendenti non hanno linee guida chiare su cui operare e questo porta al grave problema della variabilità. A conferma di questo, un'indagine dell'Istituto Nazionale di Statistica (ISTAT) mostra che nel 2017 l'Italia si è posizionata tra le prime nazioni nel panorama europeo in termini di qualità del sistema sanitario offerto; tuttavia, per il 45,1% della popolazione, la qualità del servizio sanitario regionale è diminuita. Quindi esiste un gran numero di specialisti e strutture in generale che offrono un servizio di alta qualità ma in parallelo c'è una gestione dispendiosa di queste eccellenze, che deteriora la qualità complessiva percepita dai pazienti. In Italia, inoltre, la maggior parte dei processi sanitari sono organizzati con una gestione verticale della singola specializzazione; questo permette lo svilupparsi di eccellenze all'interno dello stesso reparto quindi confinate solo a singole attività. La causa di ciò è l'assenza di un responsabile dell'intero processo di cura del paziente in quanto le attività, come detto,

sono frammentate e focalizzate sulla realizzazione di precisi successi, mentre le connessioni tra di loro sono contraddistinte da inefficienze e difetti. Essendo presente un'organizzazione dipartimentale e verticalizzata, il paziente sarà assegnato al percorso di cura dando maggior rilevanza alla condizione clinica e alla natura del danno. Per aumentare la qualità e diminuire la durata e i costi potrebbe essere utile utilizzare ruoli di coordinamento e collegamento manageriale, e team di progetto in grado di sviluppare la collaborazione lungo una dimensione orizzontale dell'organizzazione, ovvero tra quelle attività che formano processi finalizzati all'erogazione di specifiche prestazioni o servizi [98]. Detto questo, è evidente come il settore sanitario sia invaso da procedure dispendiose e inefficienti; per questo è reputata necessaria un'azione di miglioramento. Prendendo spunto da altri settori e studiandone i casi di successo si è cercato di applicare anche nel settore sanitario tecniche manageriali, al fine di migliorare il modello di business e garantire un servizio soddisfacente per i pazienti.

1.5.1 I principi e le logiche del Lean Thinking

Per rispondere alle problematiche precedentemente citate già agli inizi degli anni 2000 in alcuni ospedali americani venne testata l'applicazione della filosofia Lean. Da allora la Lean Healthcare si è diffusa rapidamente nel campo, diventando oggi una delle strategie di gestione chiave negli ambienti sanitari volte a migliorare l'efficienza e la qualità delle procedure [99].

Il Lean Thinking (LT) è emerso negli anni '40 nell'industria automobilistica giapponese, Toyota. Il Toyota Production System (TPS) si basa sul concetto di eliminazione degli sprechi o degli interventi non necessari sull'intero processo industriale [100].

Il termine "Lean" è stato utilizzato per la prima volta da *Womack et al.* (2007) [101] per descrivere il TPS, che prevede una serie di metodi e strumenti per migliorare continuamente l'efficienza e l'efficacia di un sistema eliminando gli sprechi. LT è implementato secondo cinque principi [102]:

1. Definizione del valore secondo la prospettiva del cliente.
2. Mappare il flusso del valore, ovvero delineare le attività in cui si articola il processo, distinguendo quelle a valore aggiunto e quelle non a valore aggiunto.
3. Creare il flusso: il processo di creazione del valore deve fluire continuamente (nessuna attesa e ritardo).

4. Pull production: il flusso di valore dovrebbe essere basato sulla pull production, ovvero produrre solo quando e quanto richiesto dal cliente.
5. La ricerca della perfezione è il punto a cui dobbiamo tendere attraverso il miglioramento continuo e l'eliminazione degli sprechi.

Pertanto, LT si concentra sulla riduzione degli sprechi (chiamati Muda, in giapponese) e delle attività che assorbono tempo e risorse senza aggiungere valore al processo produttivo, sincronizzando i flussi di lavoro [103]. Toyota mette in evidenza due importanti tecniche per la riduzione degli sprechi e il miglioramento continuo del processo: “just-in-time” (produzione di merce quando richiesta dal cliente e nella quantità necessaria) e “jidoka” (automazione a misura d'uomo, attraverso apparecchiature che si interrompe automaticamente quando vengono rilevati problemi di qualità) [104].

Tra i numerosi strumenti Lean, i più citati negli studi, ci sono la Value Stream Map (VSM), Visual Management, diagramma di Ishikawa, Kanban, Standardization, Kaizen Process map [105]. Pertanto, LT consente di effettuare una valutazione qualitativa e la sua forza risiede nella sua serie di soluzioni standard a problemi comuni [106], [107]. Il Lean si è diffuso in altri settori e organizzazioni passando gradualmente ai servizi, infatti, dall'inizio del 21° secolo, gli strumenti Lean più utilizzati nel settore sanitario sono VSM, 5S e Kaizen [108].

I rifiuti indicati da LT nei servizi di produzione sono ben descritti in letteratura e possono essere interpretati dai ricercatori per affrontare problemi in ambito sanitario settore [109].

Secondo *Usman* (2020) [110], le sette categorie di rifiuti del TPS possono essere adattate all'assistenza sanitaria, come riportati successivamente, nonché include una ottava categoria “Talento”, ovvero la mancata formazione dei tecnici dell'emergenza e dei medici nelle nuove tecniche diagnostiche [111]. Ridurre questi sprechi nelle strutture sanitarie porta a un servizio più veloce, meno errori e meno costi [112]. Dalla letteratura è possibile inquadrare le tipologie di sprechi in sanità:

- Trasporti: Tiene conto di quei movimenti di materiali e lavoratori che non aggiungono valore alla merce: spostamenti inutili di pazienti e personale.
- Inventario: Porta a costi di finanziamento e di stoccaggio più elevati, tassi di difettosità più elevati e pazienti in attesa (cioè farmaci con scorte eccessive o scarse).

- **Movimento:** Consiste in movimenti fisici o camminate da parte dei lavoratori che li distraggono dal lavoro effettivo di elaborazione: personale alla ricerca di attrezzature, farmaci e scartoffie.
- **Attesa:** È tempo di inattività che si verifica quando il tempo viene utilizzato in modo inefficace: eccessiva attesa dell'esito degli esami, del personale o dei farmaci; attesa causata da un errore di pianificazione.
- **Sovraproduzione:** Non è necessario produrre più di quanto richiesto dal cliente o troppo presto prima che ci sia la necessità. Aumenta il tempo di piombo e stoccaggio, vale a dire diagnostiche non necessarie, agendo per programma.
- **Over-processing:** Si verifica quando i lavoratori eseguono involontariamente più lavorazioni di quelle richieste dal cliente in termini di qualità o caratteristiche del prodotto: ripetere inutilmente i test, chiedere più volte i dettagli del paziente.
- **Difetti:** Può includere errori nelle pratiche burocratiche, ritardi di trasporto, produzione secondo specifiche errate, spreco di materie prime o generazione di scarti, ripetizione dei test a causa di errori, riammissione per mancato scarico.

1.5.2 Lean Healthcare

Una volta analizzata il LT dal punto di vista teorico, è interessante capire come essa viene calata in ambito operativo e i risvolti che essa ha sui vari processi presenti all'interno dell'ospedale. I principali step di tale filosofia applicati nell'assistenza sanitaria sono stati introdotti da *J.P. Womack e D.T. Jones* [103] e sono i seguenti:

- assegnare ai pazienti le cure utili a soddisfare le proprie esigenze nel minor tempo possibile in modo da creare valore. In un'azienda sanitaria, la quantificazione di questo valore dipende dalla discrepanza che c'è tra il trattamento che il paziente riceve e il trattamento che lo stesso paziente si aspettava di ricevere; più è ampio questo gap, più il paziente è insoddisfatto del servizio ricevuto. L'obiettivo di ogni azienda sarà quindi quello di generare un servizio che possa essere all'altezza delle richieste dei propri clienti.
- identificare i processi attraverso i quali i pazienti ricevono le cure. Per processo si intende una sequenza di attività, logicamente correlate, svolte secondo una determinata sequenzialità e/o simultaneità, che ha un'origine (un punto di partenza) e che permette di raggiungere un determinato risultato finale (punto di arrivo) attraverso l'impiego di risorse (persone, macchine, materiale). Tale

sequenza è caratterizzata da: input misurabile (risorse tangibili o intangibili che innescano il processo); vincoli (regole, condizioni, tempi che influenzano le attività); risorse (persone e mezzi utilizzati nel processo); output misurabile (risultati del processo). Si possono distinguere due tipologie di processi, quelli che contribuiscono direttamente alla creazione dell'output di un'organizzazione denominati processi core (cura e assistenza del paziente), e quelli che non intervengono direttamente alla produzione di beni o erogazione di servizi ma sono altrettanto indispensabili affinché questo avvenga (processi di supporto che sono ad esempio il processo logistico, il processo di approvvigionamento dei materiali o il processo di gestione delle risorse umane).

- eliminazione dei “Muda” presenti all'interno dei processi e conseguente riprogettazione degli stessi, in modo che tutti i passi siano sincronizzati e procedano allo stesso ritmo. Esistono due tipi di Muda: l'impiego di risorse per una operazione che in sé non crea valore ma risulta necessaria per attuarne altre produttrici di valore. In questo caso possono essere individuate delle azioni di miglioramento organizzativo ma, non è possibile prevedere una eliminazione completa; un dispendio di risorse del tutto inutile, qui è necessario attuare un piano per la completa rimozione. Alcuni problemi presenti all'interno di un ospedale sono stati classificati da *Ohno*, 1988 [100] in: sovrapproduzione (ricovero dei pazienti più lungo del necessario, in quanto sottoposti a operazioni ridondanti o inutili), scorte (mancato utilizzo o non utilizzo di risorse), trasporti di merci e informazioni (es. cartelle cliniche o altre documentazioni, report medici), attese (mancata sincronizzazione tra gli operatori che causa attese nell'iter ospedaliero a cui è sottoposto il paziente), movimenti di personale e difetti (es. diagnosi o trattamenti non corretti, violazione privacy, errata raccolta dati).
- fornire servizi in base alla domanda esterna. Utilizzo quindi di una strategia “pull” che si oppone a quella “push” in base alla quale la disponibilità è definita ex-ante senza tenere conto della richiesta proveniente dall'esterno. Il sistema pull, comunemente detto a trazione, è un metodo in cui le attività a valle segnalano i loro bisogni alle attività a monte. Lavorando così, è il paziente che “tira” il processo; ciò permette di individuare ed eliminare una grande quantità di sprechi, ma richiede un'attenta e continua analisi dei flussi interni ed esterni all'azienda. Per ottenere una strategia pull è necessario seguire delle linee guida e degli

accorgimenti tipici della filosofia Lean: lavorare al Takt Time (il Takt Time è definito come il tempo nel quale deve essere eseguito il servizio per il paziente, affinché le prestazioni vengano erogate al ritmo della domanda) cioè capire come sincronizzare le attività per soddisfare gli input in ingresso; sviluppare, per quanto possibile, un flusso continuo e bilanciato; livellare la produzione in corrispondenza del processo pace maker, quello che dà il passo all'intero flusso; solitamente è situato verso il fondo del processo.

- raggiungere la perfezione dei processi; questo obiettivo richiede un miglioramento continuo riducendo il più possibile gli errori. Facendo scorrere velocemente il processo emergono sempre più sprechi nascosti nel flusso, e quanto più esso è tirato dal cliente tanto più vengono messi in evidenza gli ostacoli da rimuovere. Il traguardo della perfezione non va inteso come se fosse possibile individuare da subito, una volta per tutte; si può semmai intendere come un asintoto che ha lo scopo di mantenere attivo un processo di miglioramento continuo. Tale miglioramento può manifestarsi a volte attraverso grandi innovazioni e consistenti balzi tecnologici e organizzativi, ma molto più frequentemente è il frutto di tanti piccoli ma sistematici affinamenti. Tra gli indicatori più significativi per questa valutazione ritroviamo: l'efficienza, rappresentata dal numero di prestazioni realizzate da un'unità di fattore produttivo impiegato; l'efficacia, misura il contributo dei servizi sanitari al miglioramento dello stato di salute; la capability, identificata nella capacità di riprodurre nel lungo periodo, lo stesso prodotto o servizio senza alterarne il risultato; la flessibilità, che indica la capacità del processo di adattarsi ai cambiamenti interni all'organizzazione o esterni come il mutamento delle condizioni del mercato o la variazione dei requisiti della clientela.

I processi riguardo ai quali vengono applicati i principi Lean, sono classificabili, in base a quanto stabilito da J. Black nel suo libro "The Toyota way to Healthcare Excellence"[99], in sei principali flussi che riguardano essenzialmente:

- pazienti: tale flusso va dall'arrivo degli interessati all'ospedale alla loro dimissione. Esso dovrebbe essere facile e veloce; invece i pazienti e le loro famiglie si trovano molto spesso a lottare per essere presi in considerazione. Un accorto design, eliminando le attività che non aggiungono valore, riduce il tempo che essi passeranno all'interno dell'ospedale e la loro soddisfazione. Alcune

indicazioni per implementare un flusso più snello sono: progettare in anticipo il miglior percorso dei pazienti, portando il servizio verso di loro, piuttosto che farli rimbalzare da una parte e dall'altra; creare cellule aperte e flessibili, con strumentazioni e staff intercambiabili; aggiustare sul posto il processo che si interrompe, senza rinviare il problema; disegnare il processo in modo tale che non si debbano ripetere più volte le stesse cose.

- dipendenti: i movimenti di specialisti, infermieri e personale per realizzare i loro compiti. Troppo spesso accade che il giro per raggiungere i diversi luoghi di visita e consulenza sia talmente lungo, complesso e richieda tanto tempo, da rendere il contatto coi malati una frazione minima di questo. Le principali contromisure individuate per risolvere questi problemi sono: creare un ambiente ergonomico; eliminare sedie e poltrone che inducono allo stazionare; evitare sovraccarichi di impegni, ma distribuire gli impegni fra il personale di assistenza, da dove è presente meno pressione verso dove c'è più domanda; evitare il movimento inutile, così presente nelle aree di cura.
- farmaci: spostamenti di dispositivi medici dai magazzini centrali ai vari reparti per poi arrivare ai carrelli dove sono dispensati. Essi devono essere orchestrati seguendo il flusso dei pazienti (accompagnandoli da vicino) in modo da avere l'ammontare necessario di farmaci al momento e nel posto giusto. Accortezze da adottare sono: localizzare medicazioni e farmaci vicini ai pazienti come in Terapia Intensiva; creare piccole farmacie periferiche vicine ai pazienti, tutte governate dal sistema Kanban (prudente riempimento al consumo); pianificare la dimissione il giorno prima, preparando la farmacia alla consegna dei farmaci di chi viene dimesso.
- informazioni: sono registrate, man mano che il paziente avanza nel suo iter di cura, nel sistema informativo dell'ospedale, il quale deve permettere di visualizzare i dati corretti al momento giusto. L'elettronica può servire per mettere tutte le informazioni di un paziente in un unico documento che viene ripreso in ogni occasione, per comunicare ed aggiornarne la storia.
- apparecchiature mediche: dimensionare il numero di apparecchiature necessarie in base al numero di esami e interventi da effettuare e metterle in luoghi dove non necessitano tempi eccessivi per la mobilitazione. Inoltre, cercare di ottimizzare le

dimensioni: leggeri, trasportabili da una sola persona senza fatica, piccoli e fatti nella maniera più semplice.

Se ben gestiti, i processi possono avere un impatto strategico nell'azienda su quattro versanti: costi, ricavi, investimenti, competenze, come validato con studio scientifici di cui si discuterà nel paragrafo successivo.

1.5.3 I fondamenti del Six Sigma

Il Six Sigma (SS) è una metodologia che si focalizza su obiettivi di miglioramento, ottimizzazione e re-design dei processi produttivi utilizzando allo scopo strumenti matematico-statistici. Nello specifico, il Six Sigma fornisce una misura statistica del numero di difetti che caratterizzano un determinato processo.

Sviluppata originariamente negli Anni '80 grazie all'opera dell'azienda Motorola, essa ha successivamente riscontrato un notevole successo, venendo proficuamente impiegata in realtà produttive quali Allied Signal e General Electric [113], [114] [115]. Sono proprio le applicazioni di successo di queste grandi aziende che hanno attratto l'interesse di altre organizzazioni e studiosi, permettendone una progressiva diffusione ed applicazione, tanto che nel corso degli anni il SS è stata utilizzata in numerosi contesti, quali quello manifatturiero, il settore dei servizi e, più di recente, in congiunzione con altre metodologie, come nel caso del Lean Six Sigma.

L'approccio prende la sua denominazione dalla lettera dell'alfabeto greco Sigma. Tale termine è utilizzato per descrivere la variabilità, maggiormente nota in termini matematici come deviazione standard.

La metodologia, in particolare, viene utilizzata come misura statistica della probabilità che un errore si verifichi in un determinato processo. In tal senso, prendendo in considerazione la misura di un Critical To Quality (CTQ) e l'intervallo di tolleranza definito, essa fornisce in modo chiaro il seguente messaggio: maggiore è il numero di sigma contenuti in una tolleranza intorno al target, minore sarà la probabilità che siano fabbricati beni od erogati servizi che risultino non conformi.

In tal senso, la seguente Figura 10 evidenzia graficamente la portata statistica di quanto appena detto.

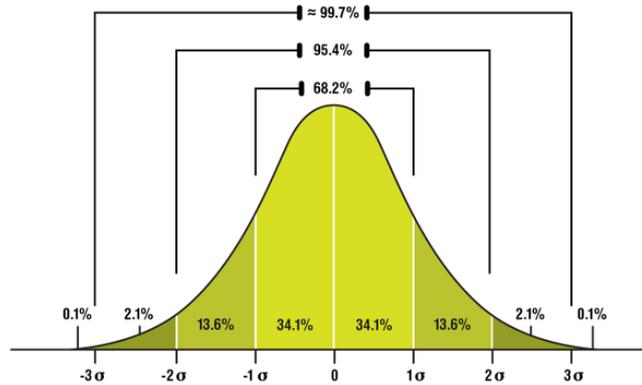


Figura 10. Distribuzione normale ed intervalli +/- sigma

La misura, in sostanza, serve a valutare le dimensioni di efficienza ed efficacia raggiunte dall'azienda in oggetto. Nello specifico si tratta di:

- di efficienza, perché la misura prende in esplicita considerazione la capacità di saper soddisfare le esigenze del cliente con le risorse a disposizione, ovvero con le potenzialità del processo come attualmente gestito;
- di efficacia, poiché la valutazione ottenuta esprime se le esigenze dei clienti sono state soddisfatte ed a quale grado.

Statisticamente, la misura deve essere interpretata nel seguente modo: un determinato processo raggiunge un livello di SS quando esso produce 3,4 prodotti o servizi difettosi (difetti) ogni milione di prodotti o servizi. In altri termini, SS identifica chiaramente ed in modo tecnico il numero di clienti insoddisfatti per milione di prodotti/servizi venduti: espresso diversamente, SS è equivalente a 3,4 esperienze negative per i clienti, ogni milione di prodotti/servizi loro venduti.

È interessante approfondire l'analisi per comprendere l'esatta portata dell'informazione che stiamo trattando: ad esempio, se i difetti (le “esperienze negative”) salissero a 233, l'azienda in questione rispetterebbe solo una soglia di Five Sigma; ove i difetti salissero a 6.210 l'azienda sarebbe Four Sigma e così via, come mostrato in Tabella 1.

Tabella 1. Conversione in Sigma

Processo Six Sigma	Rendimento	Difetti per milione
6	99,99966	3,4
5	99,98	233
4	99,4	6.210
3	93,4	66.087
2	69,1	308.538
1	30,9	691.538
0	6,7	933.193

Il SS ha uno strumento potentissimo per andare a trovare le cause che hanno generato un problema: il modello DMAIC. Il DMAIC è un modello dall'approccio per processi che si compone di 5 fasi:

- Define (definizione)
- Measure (misurazione)
- Analyze (analisi)
- Improve (miglioramento)
- Control (controllo)

Durante la fase chiamata “Define” i responsabili del progetto dovranno individuare lo scopo del lavoro che si vuole fare, per chiarire a tutti quali miglioramenti si vogliono apportare al processo sotto esame. In questa fase occorrerà fissare obiettivi che siano realistici sia per quanto riguarda le tempistiche sia per quanto riguarda i costi.

Nella fase di pianificazione occorrerà definire non solo quali dati raccogliere ma anche assicurarsi che il modo di raccogliarli sia valido. Un'altra cosa da fare assolutamente in questa fase è quella di chiarire i ruoli di ogni partecipante al progetto. Ogni persona deve sapere cosa ci si aspetta e quali sono i suoi compiti. Per ultimo, andranno fissate le risorse necessarie per lo svolgimento del lavoro. Gli strumenti in questa fase sono:

- Il project charter: Lo strumento principale da utilizzare nella fase “Define” del ciclo DMAIC è il project charter. Questa rappresentazione sintetica del progetto deve contenere le informazioni standard per la gestione del progetto (scopo, ruoli ricoperti dalle persone coinvolte nel lavoro, budget, obiettivi, ecc.)
- SIPOC: Per cambiare il processo, occorre comprenderlo a fondo per chiarire quali sono le cause di eventuali insuccessi e poterle correggere. Bisognerà chiarire gli obiettivi per far sì che le persone sappiano quali sono i loro traguardi. Allo stesso modo andranno chiariti il punto di partenza e quello di arrivo del progetto, l'individuazione di clienti e fornitori (interni ed esterni) e una descrizione anche sommaria dei vari step di processo. A questo scopo si usa spesso uno strumento denominato SIPOC. Il nome deriva dall'acronimo delle parole inglesi: Suppliers (fornitori), Inputs (elementi in ingresso), Process (processo), Outputs (elementi in uscita) e Customers (clienti) e individua la rappresentazione grafica di ognuno di questi elementi.
- Diagramma di Pareto e caratteristiche “critical”: Le origini del SS mostrano un'attenzione spasmodica al fatto che è il cliente che definisce la Qualità. Prima di affrontare qualsiasi iniziativa di miglioramento è cruciale, dunque, individuare le aspettative dei clienti e il livello di soddisfazione attuale. Troppo spesso, infatti, si crede di sapere ciò che vuole il cliente mentre, in realtà, non si sa affatto. Il ciclo DMAIC prevede proprio che siano i clienti a dire se sono soddisfatti o meno dell'operato. Una volta ricavati i dati che ci servono sulle sensazioni dei clienti, possiamo valutarli in molti modi, con l'obiettivo di stabilire quali siano le caratteristiche davvero importanti per la definizione di Qualità che ne dà il cliente. Le caratteristiche “CTQ” andranno definite per fissare delle vere e proprie specifiche, misurate e monitorate costantemente. Un semplice diagramma di Pareto aiuterà a catalogare le segnalazioni dei clienti e a dare loro il giusto peso così come un diagramma delle affinità che aiuterà a rifinire il lavoro per capire fino in fondo cosa la clientela si aspetti da un certo processo.

La fase “Measure”: I componenti principali di questa fase comprendono la creazione di una mappa del processo sotto esame e, la raccolta di tutti i dati necessari per poter impostare un lavoro efficace di analisi. Gli strumenti più utilizzati sono i flow chart e control chart. L'obiettivo principale della fase di misurazione del ciclo DMAIC è quello di ottenere più informazioni possibili dal processo corrente cioè dal processo così com'è

(prima dell'implementazione di qualsiasi azione di miglioramento). In questa fase, dunque, si cerca di capire sia come lavora il processo sia quanto in realtà funzioni bene. Una volta che la flow chart sarà completata, il team di lavoro la utilizzerà per ricavare informazioni utili all'analisi successiva del processo e per individuare come nascono i problemi e come possono essere eliminati o, almeno, ridotti. Occorrerà, dunque, cercare di individuare step mancanti, azioni ridondanti, colli di bottiglia, piccole variazioni e ogni altra cosa che può portare al verificarsi di difetti, inefficienze e problemi grandi o piccoli.

A seconda dei dati da rappresentare, si utilizzeranno strumenti grafici diversi. La control chart è uno strumento potentissimo, che riesce ad essere nello stesso tempo rappresentazione visuale e strumento statistico permettendo, così, di capire quali componenti della varianza di processo fanno parte del processo stesso (cause comuni) e quali, invece, sono esterni (cause specifiche).

La fase “Analyze”: Gli obiettivi di questa terza fase saranno quelli di verificare se le cause potenziali, identificate in precedenza come quelle che hanno scatenato il problema in esame, siano effettivamente quelle giuste e avere il supporto della conferma che deriva dall'analisi dei dati. Questa fase, come si può intuire, è importantissima perché sarà impossibile fare qualsiasi tipo di miglioramento fino a quando non avremo identificato le cause contro cui dobbiamo combattere. Può essere che l'intuizione giusta sulla causa scatenante del problema sia già arrivata nelle prime due fasi del ciclo ma, sospetti e ipotesi non possono bastare per impostare un lavoro serio di miglioramento e dovranno essere confermati dai dati. Sono molti gli strumenti che possono essere di aiuto in questa fase e che appartengono alla sfera della metodologia SS. Uno di questi è il Brainstorming, accompagnato da tutti quegli strumenti utili per visualizzare una mappa del processo sotto esame. Altri strumenti utilissimi allo scopo sono la Tecnica dei 5 Perché e il Diagramma di causa ed effetto. Una volta creata una lista di cause potenziali, il passo successivo sarà quello di organizzarla in modo che sia semplice dare ad ogni causa la giusta priorità. Una volta realizzato il diagramma, sarà facile verificare se una causa potenziale ha conseguenze in più di un'area ed affidare, di conseguenza, ad essa la giusta priorità. I metodi utilizzati per analizzare i dati dipenderanno dalla tipologia di dati raccolti e potranno essere di tipo grafico o statistico. La fase di Analisi terminerà quando il project team riuscirà ad individuare con certezza almeno una delle cause che hanno scatenato il problema in studio.

La fase “Improve”: Solo dopo aver raccolto ed esaminato tutte le evidenze oggettive si passa alla fase di miglioramento vera e propria. La fase di analisi, nel frattempo, avrà fornito una comoda base di dati da cui partire a progettare iniziative di miglioramento. Lo scopo di questa fase è quella di progettare la soluzione più adatta a risolvere il problema che si prende in esame.

La fase “Control”: Ogni iniziativa di miglioramento ha bisogno di un meccanismo di feedback e di controllo per assicurare che non si torni, lentamente, nella situazione che vigeva prima del cambiamento introdotto. In questa fase andrà avviato un meccanismo di monitoraggio periodico che misuri l'impatto delle modifiche fatte.

1.5.4 Il Lean Six Sigma

Per Lean Six Sigma (LSS) si intende una metodologia che ha come obiettivo quello di ottimizzare i processi attraverso l'eliminazione degli sprechi accrescendone le performance in termini di efficienza ed efficacia operativa. Mentre il SS risulta sostanzialmente finalizzato ad incrementare il livello di qualità dei processi gestiti, da parte sua il Lean si focalizza, invece, anche su target di efficienza ed economicità, rendendo, di fatto, le scelte manageriali sostenibili e convenienti. Inoltre, il SS si caratterizza maggiormente come approccio di problem solving focalizzato sulla variabilità all'interno dei processi, il Lean pone grande enfasi sui concetti di flusso e valore che sono nevralgici per la definizione e la gestione di quei medesimi processi, perseguendo conseguentemente obiettivi di eliminazione degli sprechi ed incremento delle attività a valore. L'apparente differenza tra i due approcci, quindi, rappresenta un vantaggio, proprio grazie alla loro integrazione ed alla conseguente esaltazione dei rispettivi punti di forza.

Dalle originali applicazioni in ambito manifatturiero, con specifico riferimento all'industria automobilistica, il Lean si è progressivamente diffuso in altri contesti operativi, compreso il settore dei servizi (es. aziende operanti in sanità, banche, enti locali e governativi, ecc.). In tali contesti operativi, le logiche, i principi e gli strumenti del Lean si sono ampiamente diffusi nel corso degli ultimi e recenti anni, a causa di una serie di motivazioni. In primo luogo, in virtù dell'elevata incidenza sui costi operativi di una varietà di sprechi (i cosiddetti muda); come osserva a questo proposito *George* (2003) [106], esiste una grande opportunità per l'applicazione dei principi Lean nei servizi; in secondo, poiché nel settore dei servizi l'incidenza di attività che nella percezione del

cliente non creano valore è molto elevata; basti pensare ai costi del personale legati alle numerosissime attività di back-office (lontano, quindi, dagli occhi del cliente) che sono necessarie in numerose realtà aziendali operanti nei servizi. In terzo luogo, i processi sottostanti alla produzione ed erogazione dei servizi sono spesso lenti, assai strutturati e complessi. Ciò si traduce in elevati costi e, talvolta, in una bassa qualità del servizio e, conseguentemente, in scarsi livelli di soddisfazione del cliente e più esigui risultati economico-finanziari. In tutte queste organizzazioni, di fatto, si aprono notevoli spazi ed opportunità per l'applicazione dei principi Lean , in questo specifico caso, del LSS.

In tal senso, l'approccio offre tutte le potenzialità metodologiche ed operative che possano consentire di:

- internamente alle aziende in cui viene implementato, migliorare i processi produttivi attraverso l'eliminazione dei difetti, la riduzione dei tempi di erogazione dei servizi, l'abbattimento dei costi di produzione di tali servizi;
- esternamente a tali organizzazioni, misurare e comunicare i livelli di qualità raggiunti; in questo specifico senso, il LSS permette di misurare esplicitamente la qualità come percepita dal cliente, allo stesso tempo identificando con chiarezza le caratteristiche fondamentali che quei dati servizi devono possedere per assicurare elevati livelli di customer satisfaction.

A margine di quanto appena proposto, del resto, dobbiamo far rilevare la varietà delle caratteristiche operative, organizzative e gestionali che possono caratterizzare aziende di servizi tra loro diverse, od operanti in settori di business differenti. In tal senso, ad esempio, abbiamo già ricordato che nell'etichetta «azienda di servizi» possono ricomprendersi allo stesso modo una struttura ospedaliera, una banca o, ancora, un'azienda operante nel settore della logistica.

Nel tentativo, comunque, di offrire una visione di sintesi sui vantaggi e le opportunità del LSS per le aziende di servizi, evidenziamo che essa, nel suo complesso, si caratterizza sempre e comunque come una metodologia tesa al miglioramento dei processi aziendali ed all'ottenimento di positivi risultati in termini di riduzione dei costi e dei tempi ciclo, nonché di incremento della qualità dei servizi erogati, della soddisfazione del cliente, dei ritorni economico finanziari sugli investimenti realizzati.

Congiuntamente, le due metodologie operano quindi in modo integrato e possono consentire di importare nel settore dei servizi una serie di insegnamenti e strumenti tali da:

1. imparare a riconoscere ed eliminare gli sprechi (muda); in tale direzione i principi Lean offrono un prezioso framework in grado di supportare il management nell'identificazione e riduzione dei 7 muda;
2. coniugare in unico approccio i vantaggi ed i guadagni in termini di rapidità e di azioni di immediata efficacia ed implementazione tipici del Lean con gli strumenti metodologici di miglioramento caratteristici del SS;
3. sfruttare le due metodologie nell'abbattere una serie di costi, legati alle varie tipologie di sprechi, compresi quelli legati all'eccessiva ed inutile elevata complessità di numerosi processi legati all'erogazione dei servizi.

Sostanzialmente, il LSS persegue l'obiettivo di intervenire sul sistema di produzione ed erogazione del servizio, migliorandolo tramite la riduzione del numero di difetti e delle attività non a valore aggiunto per il cliente (per di più, in modo rapido), in tal modo incrementando la customer satisfaction e, a seguire, i risultati economico-finanziari per l'azienda.

Detto questo, sottolineiamo come rimangano alcune perplessità relative alle modalità operative migliori per rendere fattiva e funzionale una tale integrazione: come osservano *Proudlove et al. (2008) [116]* “*The distinctions between Lean and Six Sigma in practice are not as clear cut as the academic literature might suggest (...) there is no wide/y-accepted common or integrated methodology*”.

In questo senso, tra i vari strumenti utilizzabili allo scopo di implementare un sistema di LSS, quello basato sulla metodologia DMAIC ha senza dubbio ricevuto ampi consensi in letteratura ed assunto una notevole diffusione nella prassi operativa.

2 Progettazione di una Piattaforma di Business Intelligence in Sanità a supporto del controllo di gestione

2.1 Scopo del lavoro

In questo capitolo è racchiuso il risultato dell'attività di ricerca e sviluppo effettuata presso l'azienda Gesan S.r.l. L'obiettivo finale è stato quello di giungere ad un'innovazione nella gestione dei dati sanitari che, ad oggi, non è presente nelle strutture sanitarie campane. Pertanto, si è prefissato di costruire una dashboard di BI per l'analisi e il monitoraggio della produzione sanitaria, ovvero uno strumento in grado di estrarre informazioni dai dati sanitari eterogenei per supportare il controllo di gestione.

Dalla letteratura sono emersi studi che descrivono implementazioni di processi di BI nelle organizzazioni sanitarie italiane. *Pecoraro* (2013) [117], ha proposto un sistema di BI progettando un DW basato sui dati del FSE, con particolare riferimento al percorso assistenziale dei pazienti affetti da diabete mellito. Infatti, sono stati identificati una serie di indicatori sia clinici sia di equità nell'erogazione del servizio sul territorio, considerando le linee guida delle cure primarie. Invece, *Costa et al.* (2020), [118] hanno utilizzato strumenti di BI per effettuare l'analisi e valutazione del fabbisogno di personale nelle aziende sanitarie, col fine di migliorare la performance dei processi sanitari.

Sebbene, sono presenti soluzioni software di BI nell'ambito della Sanità, esse sono limitate a dare un supporto nel contesto di riferimento. Invece, la piattaforma di BI, che si intende realizzare, mira a supportare il controllo di gestione nelle aziende ospedaliere, e quindi, si colloca come strumento innovativo e flessibile in grado di innalzare il livello di efficienza degli ospedali. Infatti, può consentire di:

- Acquisire ed analizzare dati relativi a diverse aree di interesse, cliniche e non;
- Creare modelli di dati;
- Analizzare e predire l'incidenza di malattie;
- Ottimizzare la gestione delle risorse, riducendo gli sprechi e individuando le possibili aree di azione.

Tuttavia, per garantire le suddette funzioni la piattaforma di BI ha previsto l'integrazione dei dati sanitari eterogenei raccolti dai sistemi informativi dell'azienda Gesan. Pertanto,

può rappresentare un'opportunità anche per Gesan, poiché i loro software potranno essere dotati di una soluzione di BI progettata per supportare il controllo di gestione.

L'architettura che si intende realizzare è rappresentata in Figura 11, come mostrato, verrà progettato un DW ottenuto dall'integrazione delle base dati provenienti dai sistemi informativi ospedalieri (SIO) di Gesan (Bus dell'integrazioni), il quale verrà connesso ad uno strumento software di BI in grado di realizzare report e dashboard interattive. Naturalmente, in questo progetto di ricerca, la piattaforma comprenderà solo alcune delle base dati dei sistemi informativi Gesan, per questioni legati alla complessità dell'integrazione. Pertanto, nei paragrafi successivi sono descritti gli strumenti utilizzati, le attività di Data warehousing e l'implementazione delle dashboard di BI.

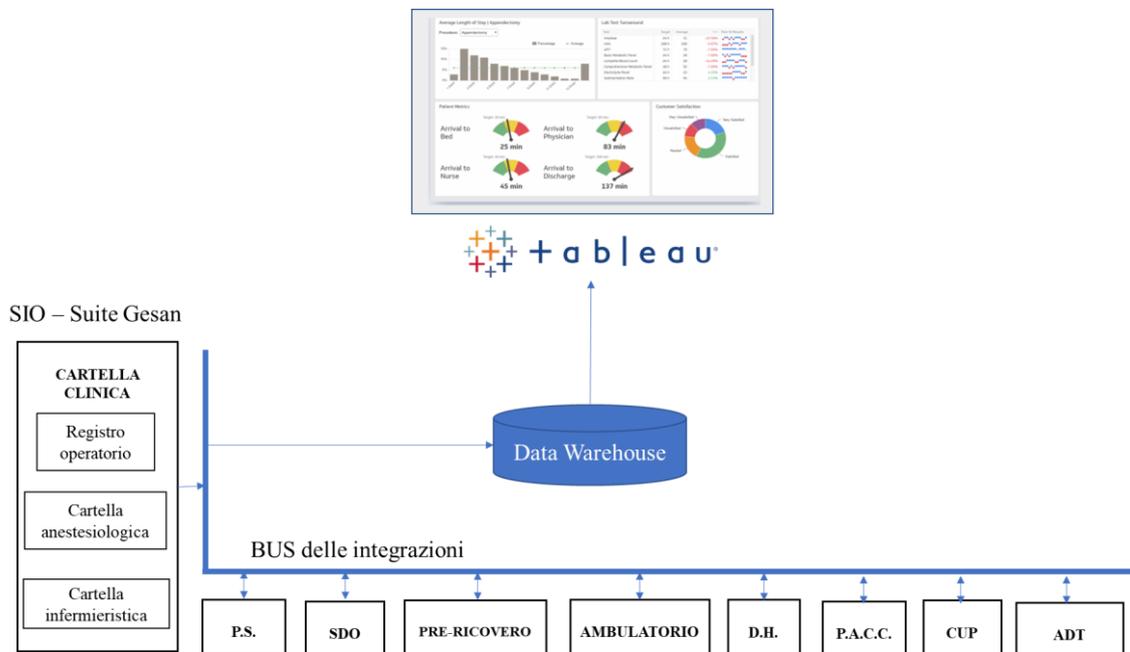


Figura 11. Architettura della piattaforma di BI

2.2 Gli strumenti

Per lo sviluppo del progetto di ricerca sono stati utilizzati i seguenti strumenti:

- Visual Paradigm: è uno strumento di progettazione e gestione potente, multipiattaforma e tuttavia facile da usare per i sistemi IT. Inoltre, fornisce agli sviluppatori di software una piattaforma di sviluppo all'avanguardia per creare

applicazioni di qualità più velocemente, migliori e più economiche. Facilita un'eccellente interoperabilità con altri strumenti CASE¹⁵ e la maggior parte dei principali IDE¹⁶ che eccelle l'intero processo di sviluppo Model-Code-Deploy in questa soluzione one-stop-shopping, vedi Figura 12. Le principali operazioni che esegue sono:

- Modellazione UML: consente di disegnare tutti i tipi di diagrammi UML, che includono: Diagramma di classe; Diagramma dei casi d'uso; Diagramma di sequenza; Schema di comunicazione; Diagramma delle attività; Schema dei componenti; Diagramma del tempo, ecc.
- Progettazione database: è possibile disegnare i seguenti tipi di diagrammi per facilitare la modellazione del database: Diagramma delle relazioni tra entità (Entity Relationship Diagram - ERD); Diagramma Object Relation Model (visualizza la mappatura tra modello a oggetti e modello dati). È possibile modellare non solo tabelle del database, ma anche procedure di stored, trigger, sequenza e visualizzazione del database in un ERD. Oltre a disegnare un ERD da zero, si può decodificare un diagramma da un database esistente.
- Modellazione dei processi aziendali: È possibile disegnare i seguenti tipi di diagrammi per facilitare la modellazione dei processi aziendali: Diagramma dei processi aziendali; Diagramma di flusso dei dati; Diagramma della catena del processo event-drive; Diagramma della mappa del processo; Diagramma organizzativo.

¹⁵ Gli strumenti C.A.S.E. (acronimo di Computer-Aided Software Engineering, ovvero sviluppo del software assistito dal computer) sono quelli che supportano lo sviluppo del software attraverso interfacce grafiche (shell) e visuali e librerie di funzionalità.

¹⁶ Un IDE, o ambiente di sviluppo integrato, è un software progettato per la realizzazione di applicazioni che aggrega strumenti di sviluppo comuni in un'unica interfaccia utente grafica.

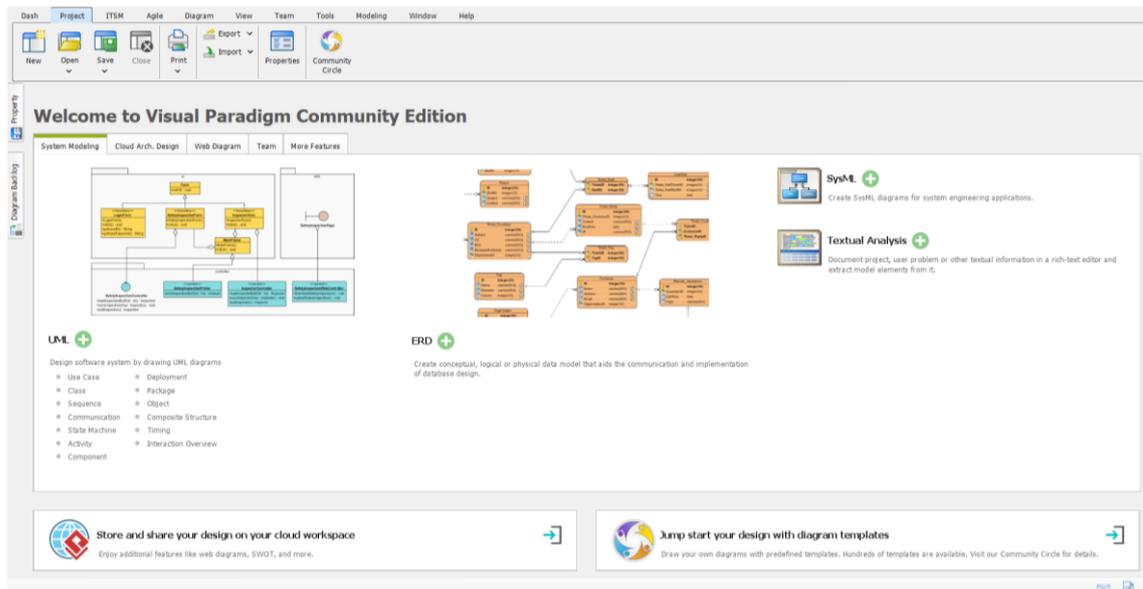


Figura 12. Workspace Visual Paradigm

- DBeaver: è uno strumento di gestione di database universale open source adatto a coloro che hanno bisogno di lavorare con i dati in modo professionale. È in grado di manipolare i dati come in un normale foglio di calcolo, creare report analitici basati su record da diversi archivi di dati ed esportare informazioni in un formato appropriato. Per gli utenti di database avanzati, DBeaver suggerisce un potente editor SQL, numerose funzionalità di amministrazione, capacità di migrazione di dati e schemi, monitoraggio delle sessioni di connessione al database e molto altro. DBeaver supporta più di 80 connessioni a database, vedi Figura 13. Avendo l'usabilità come obiettivo principale, DBeaver offre:
 - Interfaccia utente accuratamente progettata e implementata,
 - Supporto di origini dati Cloud,
 - Supporto per standard di sicurezza aziendale,
 - Capacità di lavorare con varie estensioni per l'integrazione con Excel, Git e altri,
 - Grande numero di funzioni,
 - Supporto multiplatforma.

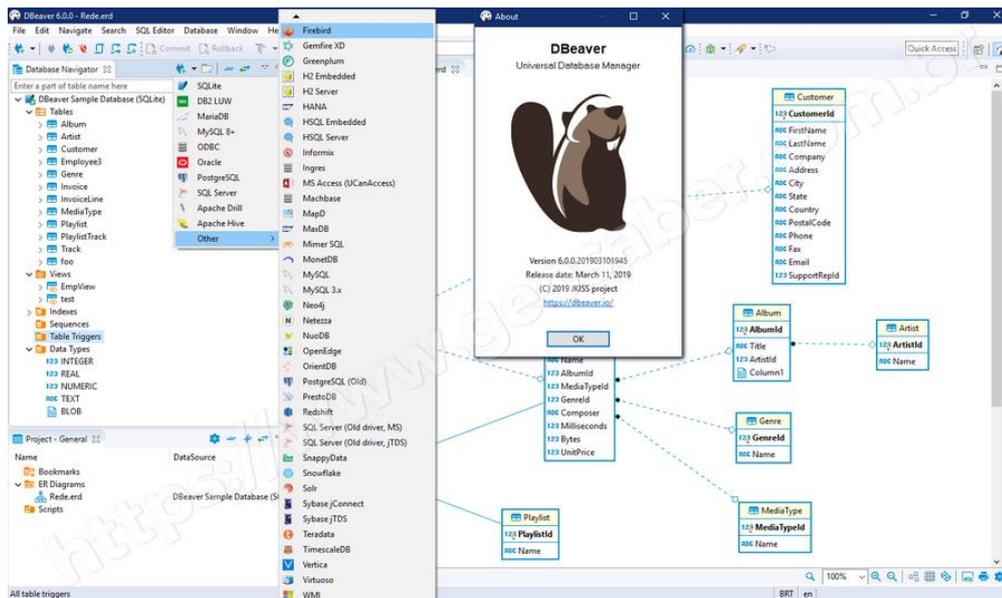


Figura 13. Interfaccia workspace DBeaver

- Tableau Prep: è uno strumento della suite di prodotti Tableau progettato per rendere la preparazione dei dati facile e intuitiva. Il processo ETL si costruisce come un flusso di attività per combinare, modellare e pulire i dati per l'analisi, vedi Figura 14. Consente la connessione a dati provenienti da una varietà di file, server o estrazioni di Tableau. Pertanto, una volta connessi all'origine dei dati, si trascinano le tabelle nel riquadro di flusso in cui si effettuano operazioni aggiungendo passaggi di flusso come: filtro, divisione, rinomina, pivot, join, unione e altro per pulire e modellare i dati. Ogni fase del processo è rappresentata visivamente in un diagramma di flusso che si crea e controlla. Tableau Prep tiene traccia di ogni operazione in modo che si possa controllare il lavoro e apportare modifiche in qualsiasi momento del flusso. Al termine del flusso, si esegue per applicare le operazioni all'intero set di dati. In qualsiasi momento del flusso, si può creare un'estrazione dei dati, pubblicare l'origine dati su Tableau Server o Tableau Online, pubblicare il flusso su Tableau Server o Tableau Online per continuare a modificare sul Web o aggiornare i dati utilizzando una pianificazione. Si può anche aprire Tableau Desktop direttamente da Tableau Prep per visualizzare in anteprima i dati.

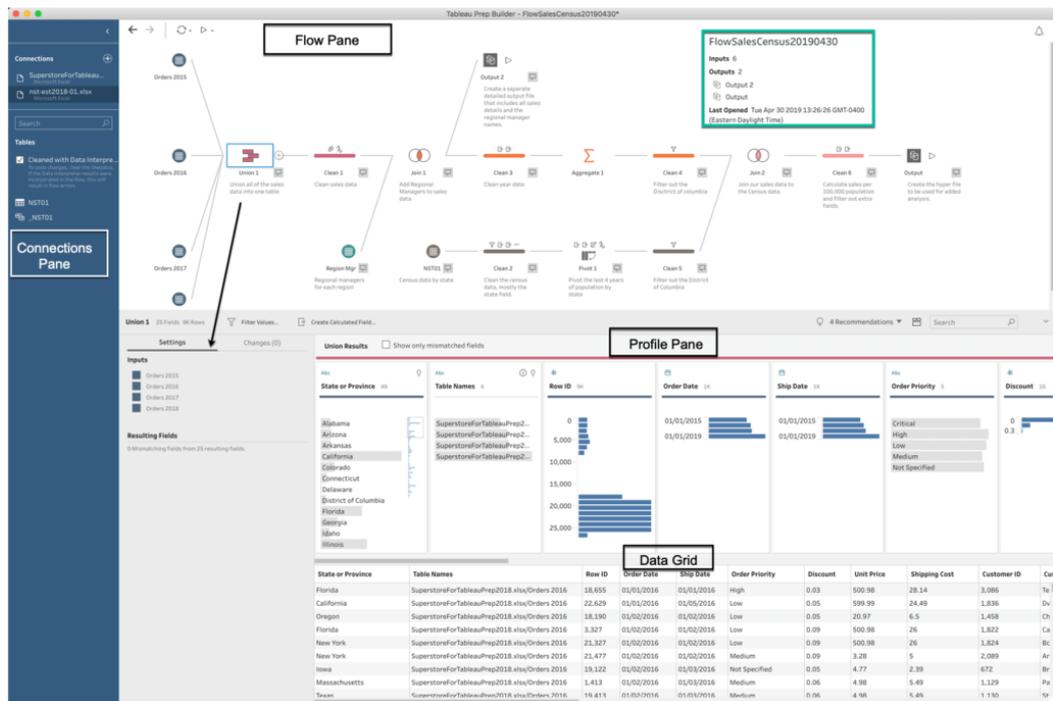


Figura 14. Area di lavoro di Tableau Prep

- Tableau Desktop: è una piattaforma software di visualizzazione interattiva dei dati focalizzata sulla BI, non richiede competenze di programmazione, permettendo a chiunque di esplorare e analizzare grandi volumi di dati, identificare trend nascosti, rappresentare le grandezze più importanti, costruire dashboard interattive. Quindi, una volta effettuata la connessione alla sorgente dati, si ha una vista dati e struttura a cartella di lavoro (workbook) composta da uno o più fogli (struttura simile ad Excel). Un foglio può essere rappresentato da:
 - Un Worksheet: singola vista con analisi, schede, legende, filtri e pannello dei dati
 - una Dashboard: raccolta di viste provenienti da uno o più worksheets che possono interagire tra di loro.
 - una Story: sequenza di worksheets o dashboards raccolte in una sequenza prestabilita al fine di «narrare una storia».

Inoltre, per creare le visualizzazioni nel foglio di lavoro sono disponibili le seguenti sezioni, vedi Figura 15: Toolbar; Data & Analytics Panel; Columns e Rows; Filters; Marks; Vista centrale delle visualizzazioni, la barra di navigazione tra i WorkBooks.

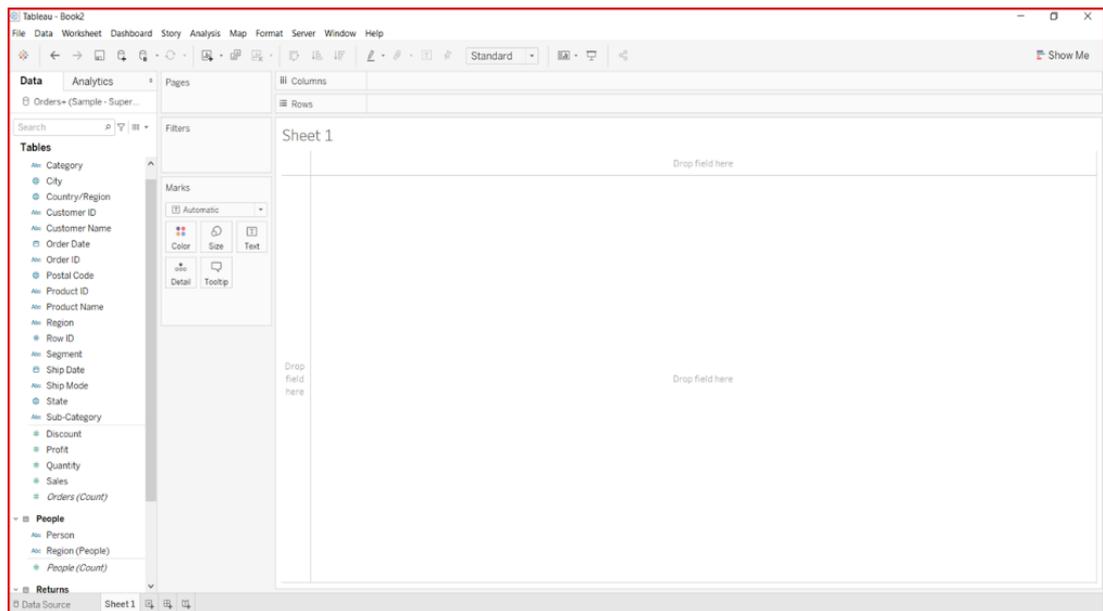


Figura 15. Interfaccia Tableau Desktop

2.3 I Metodi

2.3.1 Progettazione del Data Warehouse

Analisi del contesto e ricognizione delle fonti

Come primo passo, si sono individuate le fonti dati prendendo in considerazione le intenzioni degli utenti finali. Tra i vari interessi, si è preso in considerazione l'attività del Pronto Soccorso (PS) e la gestione dei posti letti nei reparti ospedalieri. Quindi, si sono individuati le base dati dei sistemi informativi Gesan che contengono i dati di riferimento al PS e all'Ammissione Dimissione Trasferimento (ADT)¹⁷ del paziente. Il sistema informativo Gesan preso in considerazione è Wirgilio - Web-Hospital, che raccoglie i seguenti dati:

- Centro Unico di Prenotazione
- Cassa/ Ticket
- ADT
- Pronto Soccorso
- Prericovero

¹⁷ L'Accettazione Dimissione e Trasferimento è uno strumento atto ad organizzare gli accessi in una struttura ospedaliera, dalla prenotazione di un ricovero mediante l'utilizzo di liste d'attesa, al successivo percorso, dall'accettazione alla dimissione.

- P.A.C.C¹⁸.

Successivamente, dopo un'attenta riflessione concettuale dei fatti su cui modellare il DW si è compreso di realizzare due schemi, uno che abbia il PS come fatto centrale con le rispettive dimensioni e misure e l'altro che abbia l'ADT come fatto centrale.

Processo ETL

Il processo di ETL per le base dati scelti per le analisi, ha previsto non molte operazioni. Lo strumento utilizzato è stato Tableau Prep, attraverso il quale una volta connesso il database di Wirgilio- Web Hospital, sono state estratte le tabelle di interesse. In questa fase, è stata fatta una pulizia dei dati, intesa come eliminazione dei dati "null" e modifica del formato di determinati campi, come ad esempio attribuire il formato data, testo o generale. Inoltre, con l'ausilio di tale strumento, partendo dalla selezione di determinate misure (giornate di degenza, numero pazienti, numero posti letto, ecc..) è stato possibile costruire metriche per eseguire calcoli più complessi. Le metriche calcolate si possono classificare nei seguenti modi:

- Metriche semplici: contengono una formula per determinare i calcoli da eseguire sui dati. Un esempio: $\text{sum}(\text{pazienti})$, $\text{sum}(\text{posti letto})$, $\text{sum}(\text{giornate di degenza})$;
- Metriche composte: sono state derivate e calcolate a partire dalle metriche semplici, ad esempio, tasso di utilizzo dei posti letto = $(\text{Giornate Degenza complessive} / (\text{Posti Letto} \times 365)) \times 100$.

Queste tipo di operazioni appena descritte sono state effettuate per entrambi le base dati presi in considerazione (ADT e PS).

Progettazione concettuale e logica del Data Warehouse

I DW saranno alimentati da uno strumento di caricamento ETL (Tableau Prep), una volta progettata la struttura dello schema multidimensionale considerate le esigenze degli utenti. A tal fine, è stata seguita la prospettiva bottom-up. Tale approccio è anche conosciuto nella letteratura come approccio analysis-driven, cioè guidato dalle analisi. Tuttavia, sono stati progettati due schemi multidimensionali di DW a seconda del fatto

¹⁸ Percorsi Ambulatoriali Complessi e Coordinati, sono lo strumento operativo del Day Service, consistono in gruppi selezionati di prestazioni, coerenti al raggiungimento degli obiettivi diagnostico-terapeutici, incluse nel catalogo regionale della specialistica ambulatoriale.

centrale a cui è destinata l'analisi. A questo scopo sono stati utilizzati due strumenti, Visual Paradigm (Figura 16) e DBeaver (Figura 17).

Nella Figura 16, è stato realizzato uno schema a fiocco di neve, ovvero un'estensione dello schema a stella, in cui la tabella del fatto centrale contiene i posti letto liberi/occupati distinti per genere e denominata "Bed Management_status", la quale riferenzia più tabelle dimensionali (la tabella "org_hospital") che a loro volta riferenziano altre tabelle (la tabella "org_department" e a sua volta "org_ward" e "org_branch"). Per la realizzazione è stato utilizzato Visual Paradigm.

Mentre, nella Figura 17, è stato realizzato uno schema a fiocco di neve in cui la tabella centrale contiene i dati di accesso al PS denominata "gesan_ps_data", la quale riferenzia più tabelle dimensionali ("gesan_struttura") che a loro volta riferenziano altre tabelle ("gesan_tipo_struttura" e "gesan_parametri_edwin"). Per la realizzazione è stato utilizzato DBeaver.

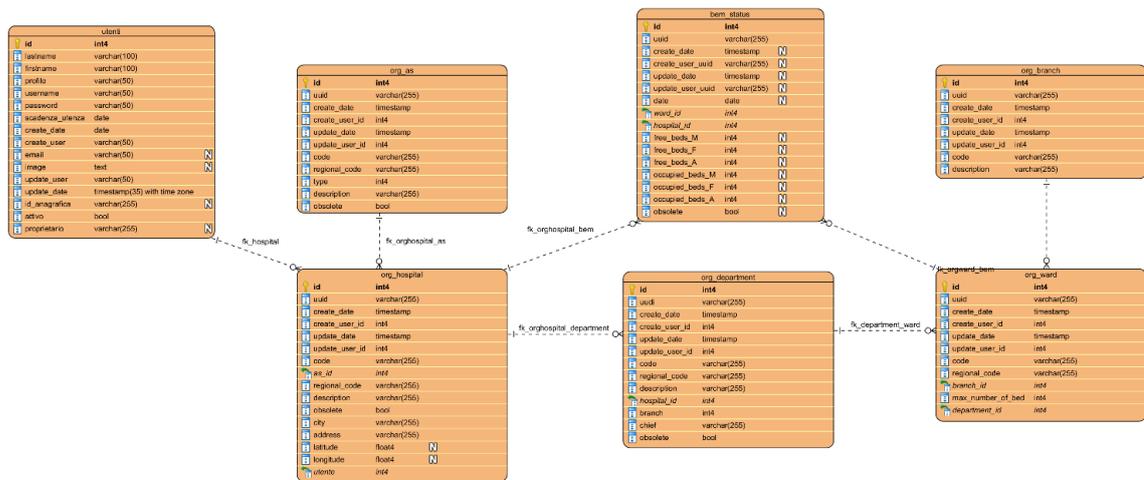


Figura 16. Schema a fiocco di neve - ADT

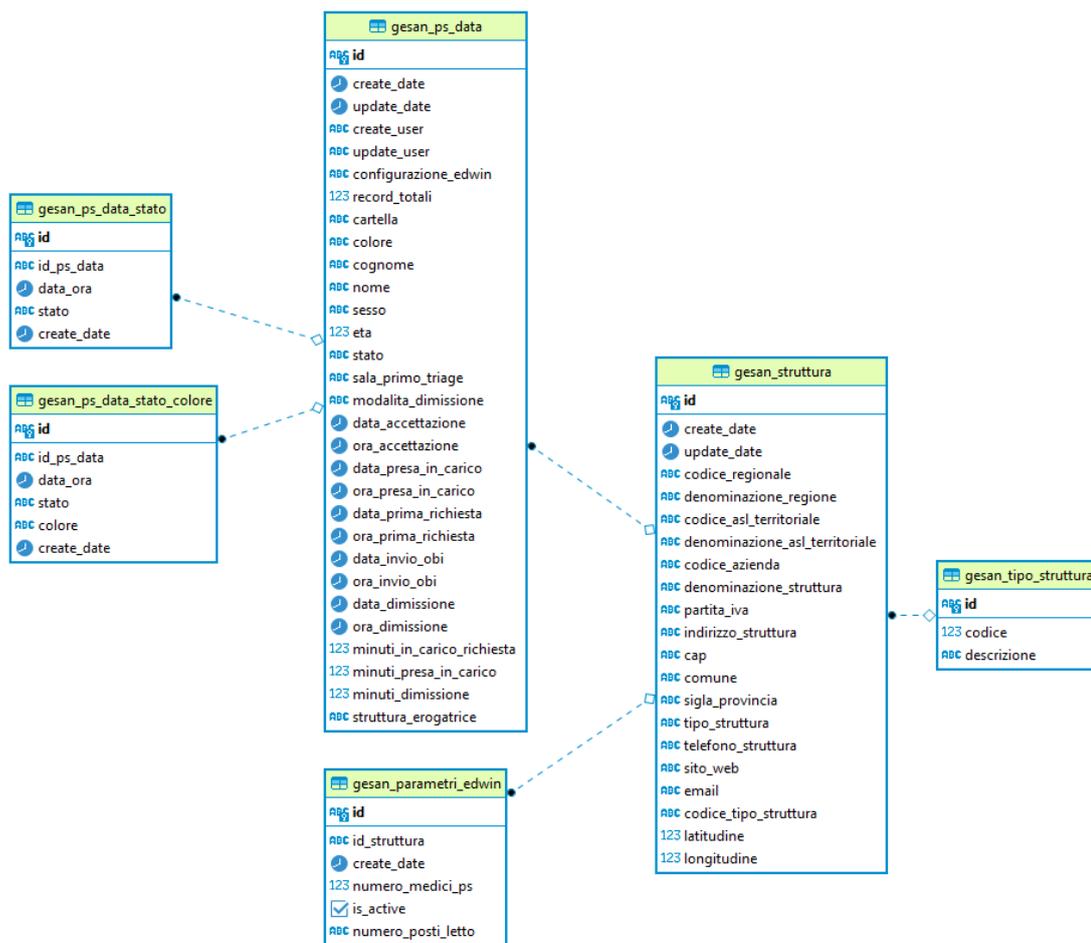


Figura 17. Schema a fiocco di neve - Pronto Soccorso

Dopo la progettazione dello schema multidimensionale del DW, esso viene alimentato con lo strumento di caricamento ETL (Tableau Prep).

2.3.2 Processo di BI

Il DW così realizzato si connette allo strumento software di BI, come mostrato nell'architettura in Figura 11. In questa fase si ottengono i risultati finali, ovvero è il momento in cui si realizzano delle ottime visualizzazioni per poter esporre le informazioni celate dietro ai SIO. Le rappresentazioni visuali offrono la possibilità di presentare le grandi quantità di informazioni disponibili. Infatti, la presentazione dei dati in modo grafico e con altri oggetti visivi permette all'utente di poter individuare nuove e utili proprietà, le loro connessioni e i possibili margini dai valori attesi. L'utilizzo del colore può far emergere aggregazioni presenti fra i dati, mentre l'uso di animazione può dare la capacità di districarsi rapidamente tra livelli di dettagli concatenati.

Le dashboard assumono le vesti di documenti di sintesi comprendenti un insieme di report precedentemente realizzati, immagazzinando in un solo colpo d'occhio informazioni immediate sulle prestazioni di tutto il contesto. Le informazioni risultanti sono sfruttate dagli utenti finali che hanno bisogno di una visione generale delle performance delle attività, traendo così vantaggi enormi dalla visualizzazione tempestiva e immediatamente comprensiva di dati strategici. Infatti, le dashboard sono progettate per conseguire il massimo impatto visivo della piattaforma, ideata per la comprensione rapida che sfrutta una combinazione di tabelle, grafici e altri indicatori. Quindi, si giunge a una visione immediata della performance resasi attuale rispetto ai trend vigenti.

2.4 Risultati: Piattaforma di BI

La piattaforma di BI è costituita di dashboard, le quali sono state implementate con il software di BI "Tableau Desktop". Le dashboard presentano una serie di visualizzazioni grafiche fornendo indicazioni agli addetti al controllo di gestione. Sono state realizzate due tipologie: una relativa al monitoraggio della distruzione dei posti letto nei reparti ospedalieri e l'altra in riferimento al monitoraggio della produttività del PS e statistiche sociodemografico sui pazienti che hanno fatto accesso.

Nella Figura 18, Figura 19, Figura 20, sono mostrati i casi d'uso della dashboard per l'analisi della produzione sanitaria del PS. Il PS è la struttura ospedaliera dedicata al trattamento delle urgenze e delle emergenze sanitarie, e che quindi si trova spesso a gestire grandi flussi di pazienti. Infatti, con i filtri a disposizione nella barra menu (data, sesso, cittadinanza, colore triage, sala triage, specialistica del pronto soccorso) è possibile simulare differenti scenari per comprendere l'andamento delle attività al PS oppure visualizzare le informazioni in real time.

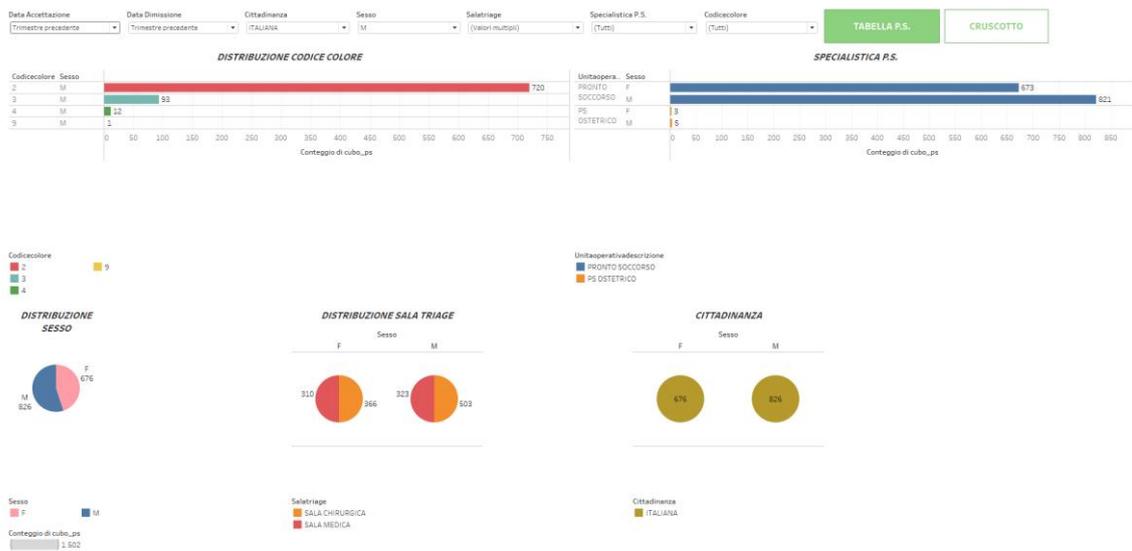


Figura 18. Cruscotto per l'analisi della produzione sanitaria. Quadro del Pronto Soccorso, caso d'uso n.1

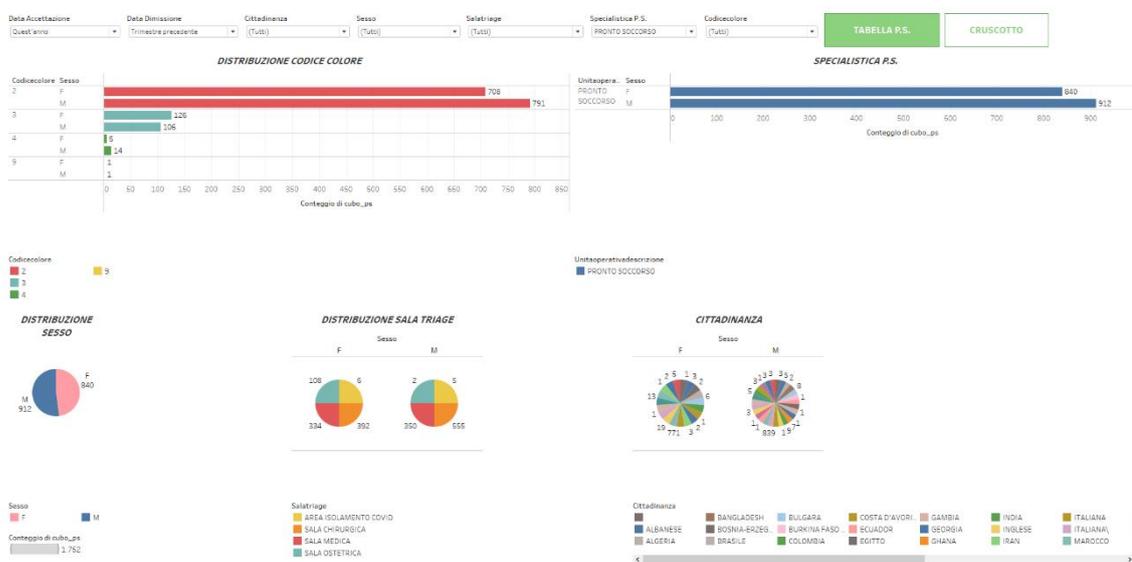


Figura 19. Cruscotto per l'analisi della produzione sanitaria. Quadro del Pronto Soccorso, caso d'uso n.2

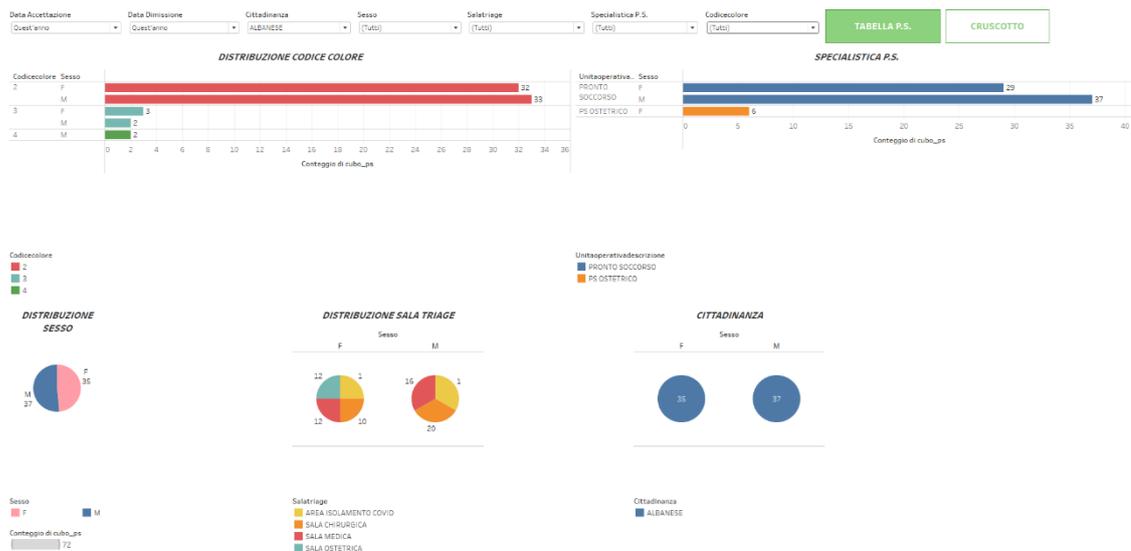
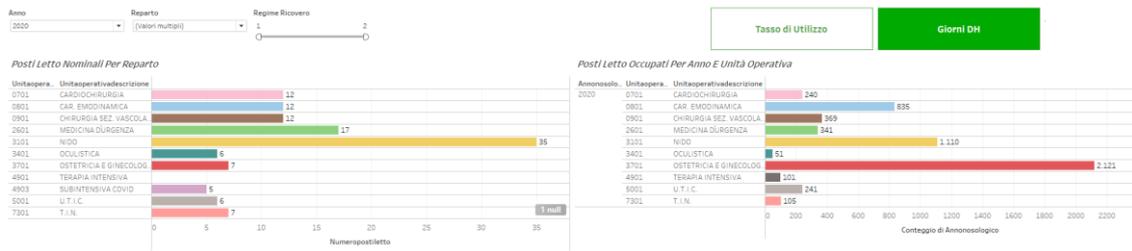


Figura 20. Cruscotto per l'analisi della produzione sanitaria. Quadro del Pronto Soccorso, caso d'uso n.3

Nella Figura 21, Figura 22, Figura 23, Figura 24, Figura 25, Figura 26, sono rappresentati i casi d'uso della seconda tipologia di dashboard indirizzata per la gestione dei posti letto. In questi casi, invece, sono mostrati visualizzazioni per monitorare e analizzare informazioni sulla disponibilità dei posti letto e il tasso di utilizzo per reparto ospedaliero (Figura 21), inoltre, è presente in dettaglio una sezione in cui sono indicati i giorni di day hospital per ogni reparto (Figura 22).



Legenda

CAR. EMO DINAMICA | CHIRURGIA SEZ. VASCOLARE | NIDO | OSTETRICIA E GINECOLOGIA | T.I.N. | U.T.I.C.

CARDIOCHIRURGIA | MEDICINA D'URGENZA | OCULISTICA | SUBINTENSIVA COVID | TERAPIA INTENSIVA

Tasso di Utilizzo

$Tu = ((O) / (P) \times 365) \times 100$

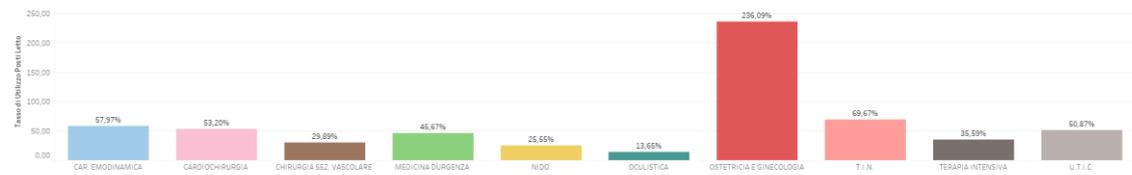


Figura 21. Dashboard controllo di gestione, caso d'uso n.1 - Posti letto per reparto anno 2020 regime ricovero 1

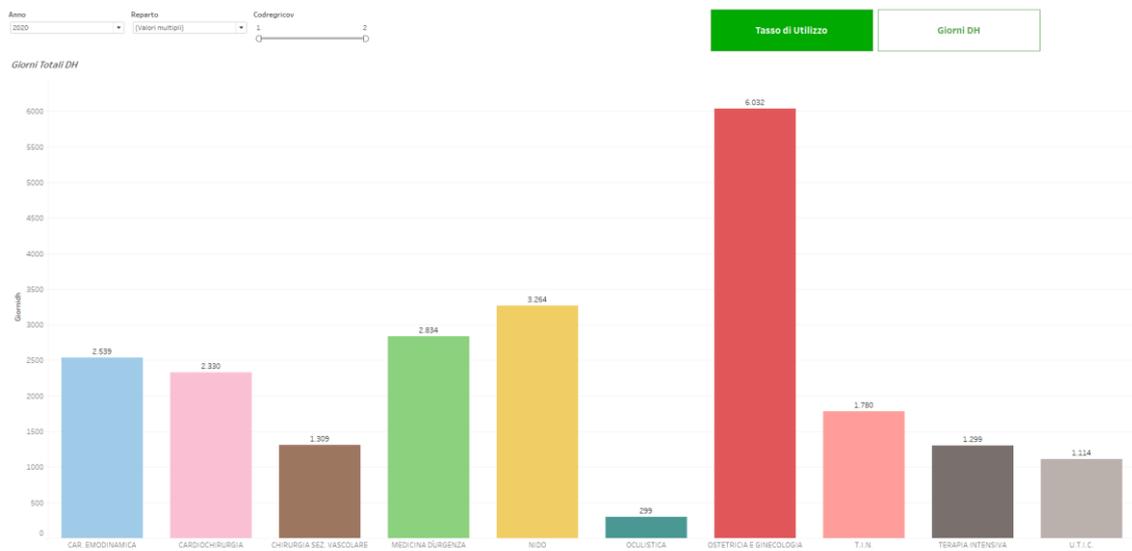


Figura 22. Dashboard controllo di gestione, caso d'uso n.1.1 - Giorni di Day Hospital per reparto

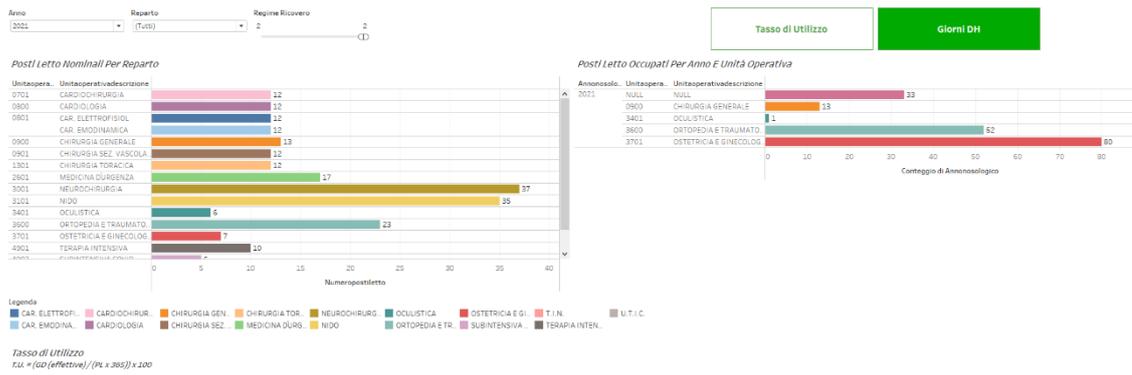


Figura 23. Dashboard controllo di gestione, caso d'uso n.2 - Posti letto per reparto anno 2021 regime ricovero 2

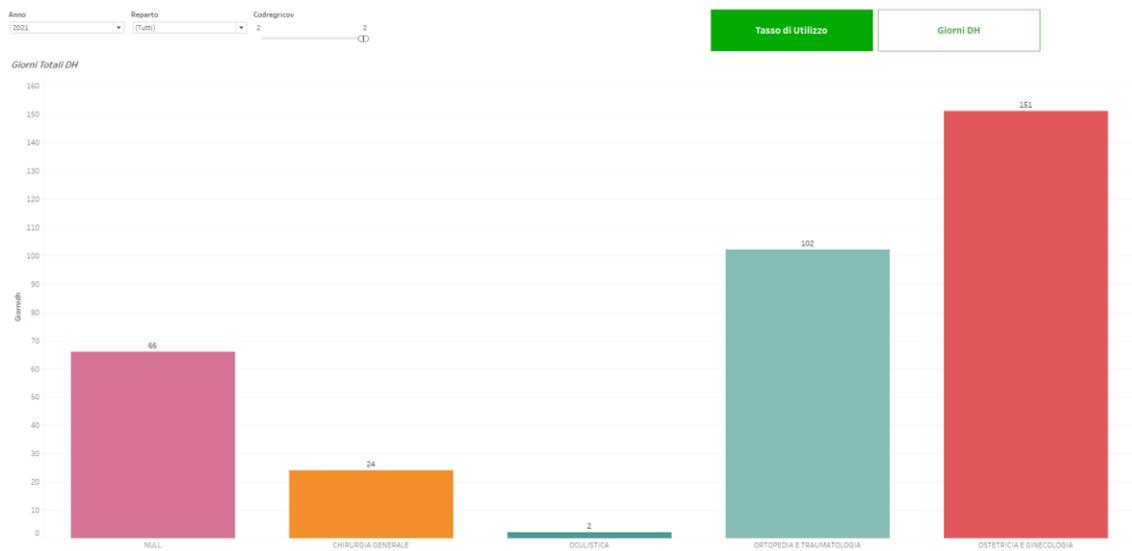


Figura 24. Dashboard controllo di gestione, caso d'uso n.2.1 - Giorni di Day Hospital per reparto

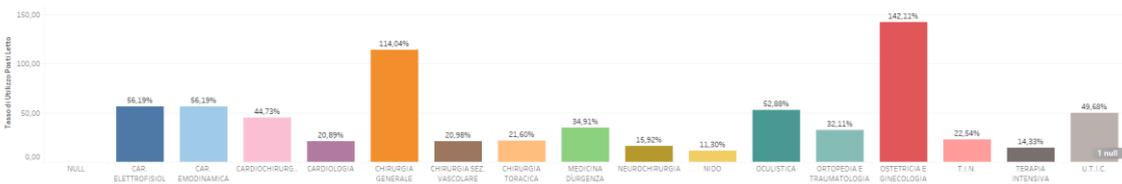
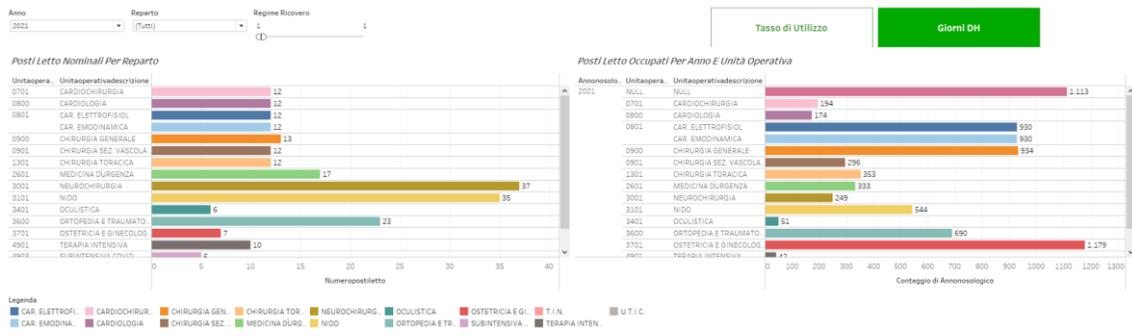


Figura 25. Dashboard controllo di gestione, caso d'uso n.3 - Posti letto per reparto regime di ricovero 1 anno 2021

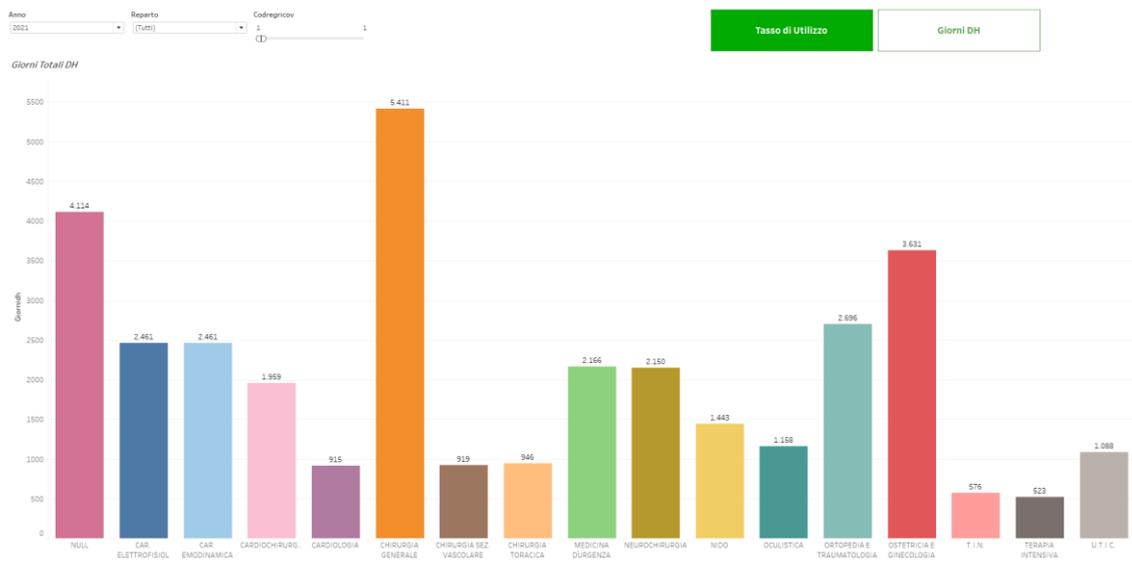


Figura 26. Dashboard controllo di gestione, caso d'uso n.3.1 - Giorni di Day Hospital per reparto

3 Produzione scientifica

Questa sezione è dedicata alla discussione dei principali articoli scientifici pubblicati in importanti riviste e conferenze nel corso delle attività di dottorato, dimostrando la validità dell'introduzione della metodologia del Lean e del SS nel settore dei servizi sanitari e delle tecniche di data analytics come, ML, Regressione Lineare e FL.

3.1 Lean Six Sigma: Applicazioni in ambito sanitario

- L'obiettivo di *Fiorillo et al.* (2021) [119], è stato di migliorare la qualità del processo di ricovero e la gestione dei pazienti attraverso la riduzione dei costi e la minimizzazione della durata della degenza preoperatoria.

Materiali e Metodi: I dati utilizzati nello studio sono stati raccolti attraverso osservazioni personali, interviste ai pazienti, brainstorming e da cartelle cliniche di pazienti sottoposti a chirurgia del cancro. Dalle cartelle cliniche sono state considerate le seguenti variabili cliniche:

- genere (maschio, femmina);
- età (<55; 55≤70; >70);
- punteggio American Society of Anaesthesiologists (ASA) (livello basso, livello alto);
- igiene orale (livello basso, livello alto);
- diabete (si, no);
- malattie cardiovascolari (si, no).

Il metodo utilizzato per lo studio è stato il LT, in particolare, nella prima fase, le analisi del processo di ospedalizzazione hanno previsto:

- Analisi qualitativa attraverso gli strumenti Lean; ovvero con la VSM sono stati identificati sprechi e inefficienze durante le attività preoperatorie che influenzano di conseguenza la degenza preoperatoria; e il diagramma di Ishikawa con il quale sono state inquadrate le cause principali degli sprechi.
- Analisi quantitativa è stata eseguita per supportare la definizione delle soluzioni Lean. Infatti, sono stati eseguiti test statistici per confrontare la media della degenza delle categorie di ogni singola variabile clinica, per investigare quali variabili clinica potrebbero influenzare la degenza ospedaliera.

Per proseguire, grazie ai risultati ottenuti dalla prima fase, il team di ricerca ha progettato e implementato le soluzioni Lean per ridurre gli sprechi.

Nella seconda fase, sono state eseguite analisi quantitative per misurare il miglioramento del processo di ospedalizzazione analizzando la degenza ospedaliera. Quindi, sono state effettuate analisi statistiche descrittive della degenza e test statistici per verificare l'ipotesi di riduzione della degenza dopo l'implementazione delle soluzioni Lean. Lo strumento utilizzato per le analisi statistiche è stato IBM SPSS Statistics.

Risultati: Gli autori attraverso la VSM hanno individuate sprechi e inefficienze durante le attività preoperatorie, di conseguenza influenzando la degenza preoperatoria. Le principali cause identificate con il diagramma di Ishikawa hanno permesso di riflettere sulle possibili soluzioni. La principale azione di miglioramento è stata l'introduzione del servizio di pre-ricovero. Tuttavia, per valutare le azioni Lean implementate è stata eseguita un'analisi statistica comparativa della degenza preoperatoria (pre e post azioni di miglioramento), in cui è stato dimostrato che la degenza media preoperatoria è diminuita da 4,90 a 3,80 giorni (22,40%). In particolare, tutte le categorie presentano una differenza statisticamente significativa prima e dopo le soluzioni Lean, ad eccezione della categoria dei pazienti di età compresa tra 55 e 70 anni e degli over 70, pazienti con un punteggio ASA di livello alto, con una scarsa igiene orale e in presenza di diabete. Tra le variabili statisticamente significative, i pazienti con punteggio ASA di livello basso mostrano la diminuzione della degenza più significativa, circa il 67,70% in termini di valori medi.

Conclusioni: il LT ha permesso di evidenziare il miglioramento del processo di ospedalizzazione con l'introduzione del servizio di pre-ricovero. Quindi, con l'adozione della cultura di miglioramento continuo il processo di ricovero è stato progettato nuovamente. I benefici delle soluzioni Lean implementate sono rivolti ai pazienti in termini di minore degenza e maggiore soddisfazione del servizio, ma anche all'ospedale prevedendo minori costi di gestione dei pazienti e la qualità del processo migliorato.

- *Latessa et al.* (2021) [120], hanno condotto uno studio con l'obiettivo di migliorare il processo chirurgico dei pazienti sottoposti a sostituzione delle protesi

al ginocchio e all'anca, riducendo la durata media della degenza e, di conseguenza, i costi ospedalieri.

Materiali e Metodi: il LSS è stato applicato per valutare il processo di chirurgia artroplastica. In conformità con un tipico processo di miglioramento LSS, il ciclo DMAIC è stato adottato per analizzare a fondo il nuovo protocollo e migliorare le prestazioni del processo. I dati dei pazienti nel presente studio sono stati raccolti da cartelle cliniche, in particolare, le variabili osservate sono state:

- genere (maschio, femmina)
- età (<60; 60≤75; >75)
- punteggio ASA (I–II; III–IV)
- diabete (si, no)
- allergie (si; no)
- malattie cardiovascolari (si, no)

Le analisi statistiche sono state eseguite con IBM SPSS Statistics.

Il ciclo DMAIC ha previsto le seguenti fasi:

Nella fase di Define è stato definito lo scopo del progetto inquadrando gli attori, la CTQ (degenza ospedaliera) e il miglioramento che si intende raggiungere. Per fare ciò è stata utilizzata una Project Chart e poi lo schema SIPOC per descrivere sinteticamente i processi.

Nella fase di Measure, l'obiettivo è misurare la performance del processo per implementare misure correttive per ridurre la CTQ. Infatti, è stata effettuata una analisi descrittiva della degenza, misurata in giorni e creato un istogramma per osservare la distribuzione media della degenza.

Nella fase di Analyse, secondo il LSS, sono stati identificati attraverso una Basic VSM i fattori che influenzano il processo; inoltre, è stata condotta un'analisi statistica per comprendere le variabili che influenzano quantitativamente la degenza. In aggiunta, per indagare sulle cause dell'eccessiva durata della degenza durante il processo, grazie ad una sessione di brainstorming, è stato possibile inquadrare le cause con il diagramma di Ishikawa.

Nella fase dell'Improve, dopo che sono state individuate, misurate e analizzate le cause del problema (eccessiva degenza) è stato descritto e implementato il protocollo Fast Track Surgery (FTS), il quale rappresenta l'azione di miglioramento.

La fase di Control, ha l'obiettivo di monitorare il processo, testare la validità dell'introduzione del FTS e pianificare azioni di revisione per garantire un risultato a lungo termine. Infatti, sono state effettuate analisi statiche comparative pre e post FTS e una run chart per monitorare graficamente l'andamento della degenza pre e post FTS.

Risultati: le analisi statistiche comparative (pre e post FTS) hanno mostrato una diminuzione statisticamente significativa del 12,70% della degenza media complessiva di 8,72 a 7,61 giorni. In particolare, le pazienti donne senza allergie, con un ASA score basso, non affette da diabete e malattie cardiovascolari hanno mostrato una significativa riduzione delle giornate di degenza con l'implementazione del protocollo FTS. Solo la variabile età non è stata statisticamente significativa. Infine, è stato condotto uno studio demografico per trovare differenze significative in termini di frequenza tra i due gruppi analizzati. È emerso che ci sono state differenze statisticamente significative tra i gruppi analizzati, ad eccezione delle variabili di genere e diabete.

Conclusioni: con il LSS è stato possibile valutare l'introduzione dell'FTS in ambito ortopedico, evidenziando un impatto positivo sul processo chirurgico ed in particolar modo della riduzione della degenza media e, di conseguenza, dei costi ospedalieri. Pertanto, i benefici ottenuti da questa ricerca si rivolgono sia alle esigenze gestionali degli ospedali che dei pazienti. Inoltre, il pre-ricovero ha rappresentato una fase cruciale del protocollo FTS e per questo motivo dovrebbe essere implementato con l'obiettivo di ridurre l'impatto dei fattori di rischio del paziente sul recupero post-operatorio. L'unico suggerimento che viene dato è quello di implementare questo percorso clinico dopo aver già introdotto il pre-ricovero in ospedale; la combinazione di FTS e pre-ricovero massimizzerà il suo impatto sui costi e sulla degenza.

- Nell'articolo di *Ricciardi et al.* (2019) [112], l'obiettivo è stato di verificare l'efficacia dell'introduzione del percorso diagnostico terapeutico assistenziale (PDTA) dei pazienti anziani con frattura al femore, ricoverati all'A.O.R.N. Cardarelli di Napoli.

Materiali e Metodi: Il LSS è stato utilizzato come metodo di studio. Sono stati osservati due gruppi di pazienti operati a causa di una frattura al femore prima e

dopo l'introduzione del PDTA. I dati sono stati raccolti dalle cartelle cliniche. Le caratteristiche osservate sono state:

- genere (maschio, femmina);
- età (<75; 75≤90; >90);
- punteggio ASA (I–II; III–IV) ;
- diabete (si, no);
- allergie (si; no);
- malattie cardiovascolari (si, no).

Lo strumento utilizzato per le analisi è stato IBM SPSS Statistics.

Il ciclo DMAIC ha previsto le seguenti fasi:

Nella fase di Define è stato definito lo scopo del progetto inquadrando gli attori, la variabile da osservare (degenza ospedaliera) e il miglioramento che si intende raggiungere. Per fare ciò è stata utilizzata una Project Chart e poi lo schema SIPOC per descrivere sinteticamente i processi.

Nella fase di Measure, l'obiettivo è misurare la performance del processo prima del miglioramento. Infatti, è stata effettuata una analisi descrittiva della degenza, misurata in giorni e creato un istogramma per osservare la distribuzione della degenza.

Nella fase di Analyse, è stata realizzata la VSM del processo che ha permesso di individuare gli sprechi e le inefficienze che influenzano la degenza. Inoltre, è stata eseguito un test statistico per verificare quale variabile clinica può significativamente influenzare la degenza. Infine, è stato utilizzato il diagramma di Ishikawa per individuare le principali cause del problema, ovvero l'eccessiva degenza ospedaliera.

Nella fase Improve, è stata proposta l'azione di miglioramento al problema, descrivendo il nuovo percorso, il PDTA.

Nella fase di Control, si è monitorata la performance del PDTA introdotto. A tale fine sono state realizzate due control run che esprimono graficamente l'andamento della degenza tra i due periodo (pre e post PDTA).

Risultati: Dall'analisi comparativa (pre e post PDTA) è mostrato che la degenza media dei pazienti è diminuita da 13,14 a 9,21 giorni (-29,90 %). I pazienti con un livello di punteggio ASA I-II hanno mostrato la diminuzione più significativa, circa il 38,20% in termini di media. Una buona riduzione in termini percentuali si

è verificata anche nei pazienti con diabete (-33,20 %), senza malattie cardiovascolari (-33,30 %) e, come prevedibile, di età inferiore a 75 anni (-32,50 %). Nonostante i notevoli risultati, i pazienti con allergie e senza diabete hanno ottenuto i risultati più bassi (-28,80 %) così come le donne (-28,90 %).

Conclusioni: L'implementazione del PDTA ha permesso al team di centrare l'obiettivo, ovvero migliorare il risultato dei pazienti, riducendo la degenza, e di conseguenza, i relativi costi ospedalieri. Il metodo del LSS è stato un ottimo supporto per l'analisi e la valutazione del problema ed ha consentito di introdurre una soluzione di miglioramento.

- Nel lavoro di *Ponsiglione et. al* (2021) [121], lo scopo è stato quello di realizzare uno studio Health Technology Assessment (HTA)¹⁹ che confronti due terapie farmacologiche in termini di degenza post-operatoria, misurata in giorni, e fornisca ai medici due modelli per predire la degenza ospedaliera dei pazienti sottoposti a chirurgia del cancro sul tessuto osseo del cavo orale.

Materiali e metodi: Il metodo SS è stato utilizzato come strumento di HTA, ovvero è una tecnica di gestione della qualità e miglioramento dei processi che combina l'uso delle statistiche con il ciclo DMAIC. Successivamente, è stata utilizzata la regressione lineare multipla per creare due modelli di predizione della degenza ospedaliera. Il ciclo DMAIC ha previsto:

La fase Define ha permesso di identificare gli attori del processo e gli obiettivi da raggiungere identificando il problema attraverso la Project Chart;

La fase di Measure ha definito le principali caratteristiche del processo di ricovero. Infatti, le variabili oggetto dello studio sono state:

- genere (maschio, femmina)
- età (<51; 50<61; >60)
- punteggio ASA (livello basso; livello alto)
- diabete (si, no)
- igiene orale (livello basso; livello alto)
- malattie cardiovascolari (si, no)

¹⁹ La valutazione delle tecnologie sanitarie (Health Technology Assessment - HTA) è un processo multidisciplinare che sintetizza le informazioni sulle questioni cliniche, economiche, sociali ed etiche connesse all'uso di una tecnologia sanitaria, in modo sistematico, trasparente, imparziale e solido. Il suo obiettivo è contribuire all'individuazione di politiche sanitarie sicure, efficaci, incentrate sui pazienti e mirate a conseguire il miglior valore.

In aggiunta, è stato utilizzato un istogramma per mostrare la distribuzione della degenza ospedaliera post operatoria (CTQ) per entrambi gli antibiotici (Cefazolin /Clindamycin e Ceftriaxone), con la relativa analisi statistica descrittiva.

Nella fase di analisi, è stato analizzato il percorso di ricovero con una flow chart ed eseguiti test statistici per verificare l'influenza delle variabili cliniche sulla degenza per entrambi i gruppi di pazienti trattati rispettivamente con, Cefazolin /Clindamycin e Ceftriaxone.

Nella fase di Improve, sono state indicate le due terapie antibiotiche, mentre nella fase di Control è stato utilizzato un Box Plot per evidenziare il decremento della degenza per i pazienti trattati con Ceftriaxone.

Successivamente, è stata utilizzata la regressione lineare multipla per creare due modelli di predizione della degenza ospedaliera.

Risultati: Dall'analisi statistica comparative tra i due antibiotici(Cefazolin /Clindamycin e Ceftriaxone), nel gruppo trattato con Ceftriaxone è stata osservata una riduzione della degenza media complessiva dei pazienti per tutte le variabili considerate. In particolare, i migliori risultati si sono ottenuti nei pazienti più giovani (-54,1%) e nei pazienti con scarsa igiene orale (-52,4%). Mentre, i risultati della regressione hanno mostrato che i migliori predittori della degenza per cefazolin/clindamycin sono il punteggio ASA e il lembo, mentre per il Ceftriaxone oltre a questi ultimi, anche l'igiene orale e la linfadenectomia sono i migliori predittori.

Conclusioni: La metodologia SS, utilizzata come strumento di HTA, ci ha permesso di comprendere le performance degli antibiotici e ha fornito variabili che influenzano maggiormente la degenza ospedaliera. Inoltre, i modelli ottenuti possono migliorare il risultato dei pazienti, riducendo la degenza e i relativi costi, aumentando di conseguenza la sicurezza del paziente e migliorando la qualità delle cure fornite.

- L'obiettivo degli autori *Latessa et.al* (2020) [122], è stato quello di valutare l'efficacia della somministrazione di due antibiotici (Ceftriaxone e Cefazolin/Clindamycin) osservando la degenza dei pazienti sottoposti a chirurgia del cancro orale presente sulla mucosa e l'osso.

Materiali e Metodi: Il metodo del SS, attraverso il ciclo DMAIC ha permesso di effettuare test statistici per indagare le variabili che potrebbero effettivamente

influenzare la degenza dei due gruppi di pazienti trattati, rispettivamente, con Ceftriaxone e Cefazolin/Clindamycin. Per le analisi statistiche è stato utilizzato il software di calcolo IBM SPSS Statistics. Le variabili osservate nello studio sono state:

- genere (maschio, femmina)
- età (≤ 59 ; ≥ 60)
- punteggio ASA (1–2; 3–4)
- diabete (si, no)
- igiene orale (si; no)
- malattie cardiovascolari (si, no)
- procedure chirurgiche (rimozione del cancro; rimozione del cancro e ricostruzione dei tessuti)
- lembo (si; no)
- linfadenectomia (si; no)
- tracheotomia (si; no)
- infezioni (si; no)
- deiscenze (si; no)
- fistole (si; no)

Risultati: I risultati dell'analisi statistica comparativa della degenza dei due gruppi di pazienti hanno mostrato: una diminuzione statisticamente significativa della degenza, in particolare, per i pazienti non sottoposti a linfadenectomia e in pazienti non esposti a infezioni del sito chirurgico trattati con Ceftriaxone. Inoltre, dallo studio demografico è emerso una differenza statisticamente significativa nella frequenza (N) per entrambi i gruppi per la caratteristica punteggio ASA. Infatti, Cefazolin e Clindamycin è spesso utilizzata per i pazienti con punteggio ASA 3-4, mentre Ceftriaxone è preferito per i pazienti con punteggio ASA 1-2.

Conclusioni: In questo lavoro è stato dimostrato la validità del metodo SS per confrontare l'utilizzo di due antibiotici, piuttosto che per il miglioramento della qualità dei processi.

- Nell' articolo di *Ricciardi et al.* (2021) è stata eseguita una valutazione delle tecnologie sanitarie (HTA) con l'obiettivo di confrontare le performance di due protocolli antibiotici considerando la degenza postoperatoria.

Materiali e Metodi: Il ciclo DMAIC è stato utilizzato come strumento di HTA. A tal fine, sono stati analizzati due gruppi di pazienti sottoposti a chirurgia del cancro alla mucosa orale trattati, rispettivamente, con Ceftriaxone e Cefazolin più Clindamycin. Per le analisi statistiche è stato utilizzato il software di calcolo IBM SPSS Statistics. Le variabili osservate nello studio, che potenzialmente possono influenzare la degenza, sono le seguenti:

- genere (maschio, femmina)
- età (<50; 50 ≤70; ≥70)
- punteggio ASA (livello basso; livello alto)
- diabete (si, no)
- igiene orale (livello basso; livello alto)
- malattie cardiovascolari (si, no)
- procedure chirurgiche (rimozione del cancro; rimozione del cancro e ricostruzione dei tessuti)
- lembo (si; no)
- linfadenectomia (si; no)
- tracheotomia (si; no)
- infezioni (si; no)
- deiscenze (si; no)
- fistole (si; no)

Risultati: Dall'analisi statistica comparativa risulta che è la differenza nella performance antibiotica, in termini di degenza postoperatoria su tutti i pazienti non è statisticamente significativa. Mentre, osservando le categorie delle singole variabili, i pazienti con linfadenectomia e tracheotomia trattati con Cefazolin più Clindamycin hanno, in media, una degenza postoperatoria inferiore statisticamente significativa. In generale, le differenze percentuali spiegano che, in media, la degenza postoperatoria non segue una linea di tendenza a favore di un singolo antibiotico. In particolare, le differenze percentuali più rilevanti indicano che i pazienti di età inferiore ai 50 anni trattati con Ceftriaxone hanno avuto un aumento della degenza del 60,1% rispetto a quelli trattati con Cefazolin più Clindamycin; al contrario, i pazienti con diabete trattati con Cefazolin più Clindamycin hanno avuto una diminuzione della degenza di 30,3%.

Invece, dallo studio demografico è emerso che solo alcune variabili hanno mostrato una differenza statisticamente significativa in frequenza tra i due gruppi di pazienti. Tuttavia, la maggiore frequenza (N) di pazienti trattati con Cefazolin più Clindamycin è statisticamente significativa nei seguenti casi: pazienti con rimozione del cancro più ricostruzione; l'uso del lembo, l'esecuzione della linfadenectomia o una tracheotomia; mentre è stata rilevata la maggiore frequenza di pazienti trattati con Ceftriaxone in caso di punteggio ASA basso.

Conclusioni: Alla luce delle intuizioni fornite da questo studio per quanto riguarda il confronto di due protocolli antibiotici, l'utilizzo del ciclo DMAIC e degli strumenti SS per eseguire studi HTA sono risultati strumenti idonei a raggiungere l'obiettivo.

3.2 Data analytics: Applicazioni in ambito sanitario

Tra i lavori di ricerca e produzione scientifica, è stata effettuata la validazione anche di alcune delle tecniche di data analytics come di seguito discusso.

- Nello studio di *Fiorillo et al. (2021)* [123], l'obiettivo è stato di creare modelli in grado di predire la degenza ospedaliera complessiva, misurata in giorni, dei pazienti a rischio di infezioni nosocomiali ricoverati presso l'Ospedale Universitario Federico II di Napoli, per supportare i medici nel processo decisionale.

Materiali e Metodi: I dati sono stati raccolti dalle cartelle cliniche dei pazienti nei reparti di chirurgia generale e medicina clinica. Le variabili cliniche osservate nello studio sono state: età, degenza pre-operatoria, peso DRG, numero di procedure, presenza di colonie. Inizialmente, è stata eseguita un'analisi di regressione lineare multipla, con lo strumento IBM SPSS Statistics, per prevedere la degenza. Sebbene, le ipotesi della regressione lineare non possano essere perfettamente verificate in uno scenario medico, è il modello più semplice per trattare i dati in via preliminare. Quindi, la regressione è stata eseguita con algoritmi di ML con Knime Analytics Platform, poiché i modelli ML sono più potenti e non hanno ipotesi rispetto alla regressione lineare multipla.

Risultati: I risultati ottenuti con la regressione lineare multipla mostra un coefficiente di determinazione $R^2 = 0.288$, invece con gli algoritmi DT e KNN hanno mostrato, rispettivamente, $R^2=0.264$ e $R^2 =0.162$. I risultati ottenuti

mostrano che tutti gli R^2 sono stati piuttosto bassi sia per la regressione lineare multipla che per le analisi di regressione ML.

Conclusioni: Il dataset ha rappresentato un'ovvia limitazione a causa delle poche variabili raccolte, poiché non hanno consentito una previsione accurata della degenza, nonostante l'ampio numero del campione (20mila pazienti). Nonostante ciò, lo studio ha fornito modelli in grado di dimostrare l'esistenza della relazione tra LOS e variabili cliniche considerate nello studio.

- Nell'articolo di *Picone et al.* (2021) [124], è stato condotto uno studio pilota per creare modelli che prevedono la degenza, misurata in giorni, per supportare i medici nella valutazione del ricovero dei pazienti a rischio di infezioni nosocomiali in terapia intensiva, considerando le informazioni cliniche preoperatorie.

Materiali e Metodi: I dati utilizzati nello studio sono stati raccolti dalle cartelle cliniche dei pazienti ricoverati in terapia intensiva per adulti e neonatali presso l'Ospedale Universitario 'Federico II' di Napoli. Le variabili considerate sono state: età, genere, peso DRG, numero di procedure, presenza di colonie. Come primo passo, sono state eseguite analisi di regressione lineare multipla con lo strumento IBM SPSS Statistics e poi, analisi di regressione con algoritmi di ML con Knime Analytic Platform. Successivamente, la degenza è stata raggruppata per settimane e classificata con gli algoritmi di classificazione ML.

Risultati: La regressione lineare multipla ha ottenuto un coefficiente di determinazione (R^2) pari a 0,343. Il miglior R^2 è stato ottenuto dagli algoritmi di regressione ML Random Forest ($R^2=0.414$) e Gradient Boosted Tree ($R^2=0.382$). Per quanto riguarda l'analisi di classificazione, gli algoritmi RF e Multi-Layer Perceptron hanno mostrato una precisione rispettivamente del 49,398% e del 46,988%, un errore del 50,602% e del 53,012%. *Conclusioni:* Pertanto, i modelli forniti dallo studio possono essere di ausilio nella predizione della degenza dei pazienti e supportare i medici nel processo decisionale.

- Nell'articolo di *Ricciardi et al.* (2020) [125], lo scopo è di classificare un ampio set di dati di pazienti come sano o con malattia coronarica utilizzando fattori di rischio convenzionali.

Materiali e Metodi: Un algoritmo di ADL è stato inizialmente applicato da solo per classificare i pazienti in due gruppi: sani e patologici. Successivamente

l'analisi delle componenti principali²⁰ (ACP) e ADL sono stati utilizzati insieme per ridurre il numero di caratteristiche e riclassificare i pazienti. I fattori di rischio convenzionali considerati sono stati: età, sesso, storia della malattia, diabete, ecc.. Successivamente, l'algoritmo ADL è stato calcolato sia sulle caratteristiche cliniche, fornite e suggerite dai medici, sia sui componenti principali (una combinazione lineare di caratteristiche cliniche) ottenute per ridurre il numero precedente di caratteristiche.

Risultati: L'accuratezza della classificazione ottenuta è stata dell'84,5% utilizzando solo ADL e dell'86,0% utilizzando la combinazione di ADL e ACP. Il richiamo e la sensibilità non hanno mostrato percentuali elevate (inferiori al 70%), mentre la specificità e la precisione hanno avuto risultati migliori: 94,2 e 97,7%, rispettivamente, solo da ADL e 96,2 e 98,4%, rispettivamente, da ADL e ACP combinati.

Conclusioni: Le tecniche di ML implementate consentirebbe ai clinici di ottenere risultati comparabili a quelli ottenuti attraverso la pratica clinica quotidiana, pur utilizzando solo variabili anamnestiche e nessuna di quelle ottenute attraverso esami strumentali. È probabile che questo metodo si applichi alle diagnosi dei pazienti, ma supporta anche i medici nella formulazione delle prognosi attraverso algoritmi predittivi numerici. Pertanto, la possibilità di prevedere la diagnosi a partire da poche variabili (magari limitate a quelle anamnestiche) consentirebbe alle strutture sanitarie di spendere meno per esami costosi, ottenendo comunque diagnosi affidabili.

Infine, grazie a un team di ricerca, è attualmente in corso uno studio scientifico per l'applicazione della Logica Fuzzy. In questo studio si propone un sistema di monitoraggio della qualità dell'aria in sala operatoria basato sulla FL, lo strumento utilizzato per implementare lo studio è stato Matlab. L'obiettivo è ottimizzare l'ambiente chirurgico per prevenire le infezioni associate all'assistenza ospedaliera (infezioni nosocomiali) e quindi ridurre il relativo tasso di mortalità controllando la contaminazione aerea. Questo scopo può essere raggiunto introducendo opportuni controlli e allarmi che permettano di avere un chiaro report del livello di stato dell'aria interna. Pertanto, viene proposto un FIS che

²⁰ La ACP è una tecnica di riduzione della dimensionalità lineare che può essere utilizzata per ridurre un ampio insieme di variabili a un insieme più piccolo che contiene ancora la maggior parte delle informazioni originali.

aiuta il personale ospedaliero responsabile della sicurezza e della prevenzione ad analizzare e ripristinare rapidamente l'efficienza l'impianto di condizionamento e ventilazione dell'aria. La soluzione proposta si basa sulla raccolta di parametri ambientali che vengono elaborati dal FIS fornendo informazioni complessive sulla qualità dell'aria in uscita. I parametri valutati dal FIS sono: la carica delle particelle in sala operatoria (ISO 5, ISO 7), pressione differenziale, ricambi d'aria per ora, temperatura e umidità relativa, percorso del personale sanitario. Un FIS, ottenuto combinando le potenzialità di FL e Expert Systems, dà la possibilità di monitorare le condizioni delle sale operatorie modellando opportunamente i dati di input di natura linguistica e riproducendo il processo cognitivo degli esperti attraverso tecniche inferenziali.

4 Conclusioni e Sviluppi futuri

In questo lavoro di ricerca, sono state effettuate considerazioni introduttive sulla Digital Health che grazie alle tecnologie correlate consente di raccogliere grosse quantità di dati sanitari e quindi fa nascere il bisogno di analizzare i dati e comprendere il valore dell'informazione. Quindi, si è trattato di Data Analytics, ovvero delle tecniche in grado di analizzare le grosse mole di dati con l'ausilio di strumenti tecnologici. Inoltre, tra le sfide del sistema sanitario sono emerse le criticità di gestione dei processi sanitari pertanto si è discusso delle metodologie del Lean e del SS, validate con una serie di articoli scientifici discussi nell'ultimo capitolo. Per proseguire, a valle dell'obiettivo finale del lavoro di ricerca è stato analizzato il controllo di gestione nelle aziende ospedaliere e il suo funzionamento per poter cogliere l'importanza delle sue attività.

Dunque, lo scopo finale del lavoro di ricerca è stato la progettazione e la realizzazione di una piattaforma di BI per l'analisi dei dati sanitari, al supporto del controllo di gestione nelle aziende sanitarie. Come è emerso dalla letteratura, in Italia, sono stati proposti diversi sistemi di BI che consentono l'analisi dei dati sanitari e il monitoraggio di indicatori di performance relativi a determinati processi [117], [118]; ma non si sono mostrati in grado di supporto il controllo di gestione nel suo complesso. A tal fine, la piattaforma di BI realizzata si colloca come strumento innovativo e flessibile in grado di innalzare il livello di efficienza degli ospedali garantendo il supporto al controllo di gestione, poiché consente:

- Acquisire grosse moli di dati provenienti da sorgenti eterogenee;
- Integrare i dati per sviluppare dashboard interattive per il monitoraggio delle performance che, consentono di evidenziare le criticità ed accelerano il processo di decision making;
- Incrementare la comunicazione e collaborazione tra diversi presidi medici, fornendo supporto al medico in fase di diagnosi e favorendo la produzione di referti con caratteristiche omogenee;
- Fornire un indispensabile strumento di management, attraverso il riconoscimento e l'estrazione di pattern;
- Eseguire operazioni di forecasting, utili a determinare, con buona approssimazione, i valori futuri di opportune variabili estratte dai dati.

Questa piattaforma è stata realizzata progettando prima il DW dei dati sanitari provenienti da sorgenti eterogenee, seguendo le esigenze di analisi degli utenti. Poi, si sono definite le metriche e gli indicatori utili per sviluppare le dashboard interattive. Tramite le dashboard, dotate di visualizzazioni dinamiche, è stato possibile fornire informazioni su un arco temporale pregresso o in real time, in riferimento ad esempio: alla distribuzione dei pazienti che hanno fatto accesso al PS divisi per codice triage, alla distribuzione dei pazienti secondo la specialistica (PS generale o ostetrico), alla distribuzione dei pazienti nelle sale di triage, ed andando più nel dettaglio, ottenere informazioni di tipo sociodemografico dei pazienti. Inoltre, in altri esempi le dashboard, invece, sono in grado di fornire indicazioni sulla distribuzione della disponibilità dei posti letto per reparto, e in dettaglio, il tasso di utilizzo dei posti letto per reparto.

L'architettura progettata permette di creare un grande numero di report differenti ed il DW ha la possibilità di fornire tante dashboard utili al decision-making, senza dimenticare l'introduzione di modelli di data mining e ML non usati in questo caso. Infatti, può essere un'espansione ad esempio, analizzare la contabilità analitica e quindi, individuare i costi diretti e indiretti per reparto, in maniera tale da poter valutare i consumi e gli sprechi o prevedere i costi e l'andamento della produttività dei servizi.

Inoltre, sono stati inquadrati gli sviluppi futuri della piattaforma. Più specificamente, per quanto riguarda la modellazione dei dati può essere sicuramente innovativa l'introduzione della tecnologia FHIR, acronimo di Fast Healthcare Interoperability Resources. È l'ultimo standard proposto da HL7²¹, nato con l'intento di ereditare le migliori caratteristiche dai suoi predecessori, descrive i formati e gli elementi dei dati (noti come "risorse") e un'interfaccia di programmazione dell'applicazione per lo scambio dei dati sanitari e quindi, delle cartelle cliniche elettroniche.

FHIR può essere usato come strumento per scambiare informazioni sia "stand-alone" sia insieme alle precedenti versioni. Caratteristica principale sono le risorse, le quali sono state create per facilitare il trasferimento dei dati sanitari da un sistema all'altro. Ogni entità del mondo reale è una "risorsa" in modo univoco. Con la tecnologia FHIR si organizzano i dati in base a «risorse» quali paziente, condizioni, farmaci e fornisce una

²¹ Lo standard HL7 (Health Level Seven) è uno standard approvato dall'ANSI (American National Standards Institute) per la generazione di messaggi standardizzati nel settore dell'ICT sanità, il suo nome deriva dal particolare livello 7 dello standard ISO/OSI (International Standards Organization / Open System Interconnection) per le telecomunicazioni.

struttura standardizzata per l'organizzazione e l'interpretazione di tali dati da parte dei diversi sistemi informatici o dalle applicazioni.

Nella HL7 v3 lo stesso identico concetto può essere rappresentato con differenti modelli, ad esempio, esistono dieci modi diversi per rappresentare un "paziente". FHIR, invece, identifica il paziente in modo univoco. Ma, in ogni caso, la tecnologia FHIR consente di creare più profili, ma tutti con lo stesso schema e la stessa serializzazione. Inoltre, FHIR permette di utilizzare due differenti formati per la trasmissione del flusso dei dati, XML²² e JSON²³.

Molti dei principali sistemi sanitari negli Stati Uniti hanno già adottato le FHIR nelle loro pratiche per l'IT del settore sanitario, pertanto può essere considerato una traccia di un futuro sviluppo anche in Italia.

In conclusione, considerata la natura complessa dell'erogazione dei servizi sanitari e l'organizzazione delle strutture sanitarie, le potenzialità offerte dalla BI sono numerose e garantiscono un valore aggiunto alla pratica del manager nella gestione dei processi e nelle fasi di decision-making.

²² XML è un linguaggio di markup creato dal World Wide Web Consortium (W3C) per definire una sintassi per la codifica dei documenti che sia gli umani che le macchine potrebbero leggere.

²³ JSON (acronimo di JavaScript Object Notation) è un formato utilizzato per lo scambio dati tra applicazioni all'interno di un rapporto di tipo client-server. Attraverso JSON è possibile rappresentare flussi dati, anche molto complessi, organizzati all'interno di una struttura multidimensionale (esattamente come accade in un array).

Indice delle Tabelle

Tabella 1. Conversione in Sigma	63
---------------------------------------	----

Indice delle Figure

Figura 1. Architettura di un sistema di BI	18
Figura 2. Progettazione concettuale. Schema di fatto (DFM).	23
Figura 3. Progettazione logica. Schema a stella.....	23
Figura 4. Cubo OLAP. L'evento paziente ha dimensioni: Anno, Tipo di accesso, Diagnosi.	24
Figura 5. Le principali fasi in un sistema fuzzy	41
Figura 6. Funzione di appartenenza triangolare.....	43
Figura 7. Funzione di appartenenza trapezoidale.....	43
Figura 8. Funzione di appartenenza della variabile X.....	44
Figura 9. Processo di fuzzificazione per la variabile P.	45
Figura 10. Distribuzione normale ed intervalli +/- sigma	62
Figura 11. Architettura della piattaforma di BI	70
Figura 12. Workspace Visual Paradigm	72
Figura 13. Interfaccia workspace DBEaver	73
Figura 14. Area di lavoro di Tableau Prep	74
Figura 15. Interfaccia Tableau Desktop	75
Figura 16. Schema a fiocco di neve - ADT	77
Figura 17. Schema a fiocco di neve - Pronto Soccorso.....	78
Figura 18. Cruscotto per l'analisi della produzione sanitaria. Quadro del Pronto Soccorso, caso d'uso n.1	80
Figura 19. Cruscotto per l'analisi della produzione sanitaria. Quadro del Pronto Soccorso, caso d'uso n.2	80

Figura 20. Cruscotto per l'analisi della produzione sanitaria. Quadro del Pronto Soccorso, caso d'uso n.3	81
Figura 21. Dashboard controllo di gestione, caso d'uso n.1 - Posti letto per reparto anno 2020 regime ricovero 1.....	82
Figura 22. Dashboard controllo di gestione, caso d'uso n.1.1 - Giorni di Day Hospital per reparto.....	82
Figura 23. Dashboard controllo di gestione, caso d'uso n.2 - Posti letto per reparto anno 2021 regime ricovero 2.....	83
Figura 24. Dashboard controllo di gestione, caso d'uso n.2.1 - Giorni di Day Hospital per reparto.....	83
Figura 25. Dashboard controllo di gestione, caso d'uso n.3 - Posti letto per reparto regime di ricovero 1 anno 2021	84
Figura 26. Dashboard controllo di gestione, caso d'uso n.3.1 - Giorni di Day Hospital per reparto.....	84

Bibliografia

- [1] S. R. Frank, J. R. Williams, e E. L. Veiel, «Digital health care: where health care, information technology, and the Internet converge», *Manag Care Q*, vol. 8, n. 3, pagg. 37–47, gen. 2000.
- [2] G. E. Iyawa, M. Herselman, e A. Botha, «Digital Health Innovation Ecosystems: From Systematic Literature Review to Conceptual Framework», *Procedia Computer Science*, vol. 100, pagg. 244–252, gen. 2016, doi: 10.1016/j.procs.2016.09.149.
- [3] K. Gray e C. Gilbert, «Digital Health Research Methods and Tools: Suggestions and Selected Resources for Researchers», in *Advances in Biomedical Informatics*, D. E. Holmes e L. C. Jain, A c. di Cham: Springer International Publishing, 2018, pagg. 5–34. doi: 10.1007/978-3-319-67513-8_2.
- [4] H. Dresner, «Howard Dresner predicts the future of business intelligence», *Retrieved January*, 1989.
- [5] M. Ghazanfari, M. Jafari, e S. Rouhani, «A tool to evaluate the business intelligence of enterprise systems», *Scientia Iranica*, vol. 18, n. 6, pagg. 1579–1590, dic. 2011, doi: 10.1016/j.scient.2011.11.011.
- [6] D. J. Power, «Understanding Data-Driven Decision Support Systems», *Information Systems Management*, vol. 25, n. 2, pagg. 149–154, mar. 2008, doi: 10.1080/10580530801941124.
- [7] A. Popovič, R. Hackney, P. S. Coelho, e J. Jaklič, «Towards business intelligence systems success: Effects of maturity and culture on analytical decision making», *Decision Support Systems*, vol. 54, n. 1, pagg. 729–739, dic. 2012, doi: 10.1016/j.dss.2012.08.017.
- [8] W. Bonney, «Applicability of business intelligence in electronic health record», *Procedia-Social and Behavioral Sciences*, vol. 73, pagg. 257–262, 2013.
- [9] B. Kaur e V. Singh, «Business Intelligence: Need and Usage in Indian Corporate Sector», *Journal of Critical Reviews*, vol. 7, pag. 2020, ago. 2020, doi: 10.31838/jcr.07.19.369.
- [10] D. Coelho, J. Miranda, F. Portela, J. Machado, M. F. Santos, e A. Abelha, «Towards of a business intelligence platform to Portuguese Misericórdias», *Procedia Computer Science*, vol. 100, pagg. 762–767, 2016.

- [11] M. E. Porter e E. O. Teisberg, *Redefining Health Care: Creating Value-based Competition on Results*. Harvard Business Press, 2006.
- [12] M.-L. Ivan, M. Raducu, M. Velicanu, e C. Ciurea, «Using Business Intelligence Tools for Predictive Analytics in Healthcare System», *ijacsa*, vol. 7, n. 5, 2016, doi: 10.14569/IJACSA.2016.070527.
- [13] B. G. Geetha, I. M. Umesh, e R. R. Varma, «A STUDY ON APPLICATIONS OF DATA ANALYTICS IN NON-BUSINESS DOMAINS», *vol*, vol. 6, pagg. 1249–1251, 2017.
- [14] J. E. Han *et al.*, «Effect of Electronic Health Record Implementation in Critical Care on Survival and Medication Errors», *The American Journal of the Medical Sciences*, vol. 351, n. 6, pagg. 576–581, giu. 2016, doi: 10.1016/j.amjms.2016.01.026.
- [15] N. Foshay, A. Taylor, e A. Mukherjee, «Winning the Hearts and Minds of Business Intelligence Users: The Role of Metadata», *Information Systems Management*, vol. 31, n. 2, pagg. 167–180, apr. 2014, doi: 10.1080/10580530.2014.890444.
- [16] B. H. Wixom e H. J. Watson, «An Empirical Investigation of the Factors Affecting Data Warehousing Success», *MIS Quarterly*, vol. 25, n. 1, pagg. 17–41, 2001, doi: 10.2307/3250957.
- [17] Z. Jourdan, R. K. Rainer, e T. E. Marshall, «Business Intelligence: An Analysis of the Literature», *Information Systems Management*, vol. 25, n. 2, pagg. 121–131, mar. 2008, doi: 10.1080/10580530801941512.
- [18] J. M. Ferranti, M. K. Langman, D. Tanaka, J. McCall, e A. Ahmad, «Bridging the gap: leveraging business intelligence tools in support of patient safety and financial effectiveness», *Journal of the American Medical Informatics Association*, vol. 17, n. 2, pagg. 136–143, mar. 2010, doi: 10.1136/jamia.2009.002220.
- [19] M. Karami, M. Fatehi, M. Torabi, M. Langarizadeh, A. Rahimi, e R. Safdari, «Enhance hospital performance from intellectual capital to business intelligence», *Radiology management*, vol. 35, pagg. 30–5; quiz 36, nov. 2013.
- [20] J. Glaser e J. Stone, «Effective use of business intelligence: leveraging your organization’s business data could improve financial and operational performance--and quality of care», *Healthcare Financial Management*, vol. 62, n. 2, pagg. 68–73, feb. 2008.

- [21] M. Isazad Mashinchi, A. Ojo, e F. J. Sullivan, «Analysis of Business Intelligence Applications in Healthcare Organizations», gen. 2019. doi: 10.24251/HICSS.2019.503.
- [22] R. Shailam, A. Botwin, M. Stout, e M. S. Gee, «Real-Time Electronic Dashboard Technology and Its Use to Improve Pediatric Radiology Workflow», *Current Problems in Diagnostic Radiology*, vol. 47, n. 1, pagg. 3–5, gen. 2018, doi: 10.1067/j.cpradiol.2017.03.002.
- [23] L. K. Wickramasinghe, P. Schattner, K. Jones, e L. P. Am, «Chronic disease management: a Business Intelligence perspective», *Proceedings of the Fifth Australasian Workshop on Health Informatics and Knowledge Managemen (HIKM 2011), Perth, Australia Chronic, 2011, no. Hikm, pp. 55–62.*, pagg. 55–62, 2011.
- [24] C. Silva, J. Pereira, L. Costa, H. Peixoto, J. Machado, e A. Abelha, «Business Intelligence for Cardiovascular Disease Assessment», in *2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, ago. 2017, pagg. 186–193. doi: 10.1109/FiCloudW.2017.90.
- [25] T. A. Carte, A. B. Schwarzkopf, T. M. Shaft, e R. W. Zmud, «Advanced business intelligence at Cardinal Health», *MIS Quarterly Executive*, vol. 4, n. 4, pag. 5, 2008.
- [26] T. H. Davenport, J. G. Harris, D. W. De Long, e A. L. Jacobson, «Data to knowledge to results: building an analytic capability», *California management review*, vol. 43, n. 2, pagg. 117–138, 2001.
- [27] M. N. Dehkordi, «A novel association rule hiding approach in OLAP data cubes», *Indian Journal of Science and Technology*, vol. 6, n. 2, pagg. 4063–4075, 2013.
- [28] R. Akbar, M. Silvana, M. H. Hersyah, e M. Jannah, «Implementation of Business Intelligence for Sales Data Management Using Interactive Dashboard Visualization in XYZ Stores», in *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*, ott. 2020, pagg. 242–249. doi: 10.1109/ICITSI50517.2020.9264984.
- [29] L. Cardoso, F. Marins, F. Portela, M. Santos, A. Abelha, e J. Machado, «The next generation of interoperability agents in healthcare», *International journal of environmental research and public health*, vol. 11, n. 5, pagg. 5349–5371, 2014.
- [30] K. Kakish e T. Kraft, *ETL Evolution for Real-Time Data Warehousing*. 2012.

- [31] J. Sreemathy, S. Priyadharshini, K. Radha, K. Sangeerna, e G. Nivetha, «Data Validation in ETL Using TALEND», in *2019 5th International Conference on Advanced Computing Communication Systems (ICACCS)*, mar. 2019, pagg. 1183–1186. doi: 10.1109/ICACCS.2019.8728420.
- [32] J. Sreemathy, I. Joseph V., S. Nisha, C. Prabha I., e G. Priya R.M., «Data Integration in ETL Using TALEND», in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, mar. 2020, pagg. 1444–1448. doi: 10.1109/ICACCS48705.2020.9074186.
- [33] S. Vyas e P. Vaishnav, «A comparative study of various ETL process and their testing techniques in data warehouse», *Journal of Statistics and Management Systems*, vol. 20, n. 4, pagg. 753–763, lug. 2017, doi: 10.1080/09720510.2017.1395194.
- [34] M. Vaish e P. Gupta, *Business Intelligence: Escalation of Data Warehousing and Data Mining for effective Decision Making*. 2020.
- [35] W. H. Inmon, *Building the Data Warehouse, 4th Edition | Wiley*. 2005. Consultato: ott. 21, 2021. [Online]. Disponibile su: <https://www.wiley.com/en-us/Building+the+Data+Warehouse%2C+4th+Edition-p-9780764599446>
- [36] A. Hamoud, A. S. Hashim, e W. A. Awadh, «Clinical Data Warehouse: A Review», Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3308892, 2018. Consultato: ott. 21, 2021. [Online]. Disponibile su: <https://papers.ssrn.com/abstract=3308892>
- [37] R. Kimball e M. Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. John Wiley & Sons, 2011.
- [38] I. N. A. Prabawa, D. Agung, K. Arimbawa, e I. G. N. Janardana, «Analysis and Design Data Warehouse For E-Travel Business Optimization», *Int. J. Eng. Emerg. Technol*, vol. 4, n. 1, 2019.
- [39] B. Shiyal, «Modern Data Warehouses and Data Lakehouses», in *Beginning Azure Synapse Analytics: Transition from Data Warehouse to Data Lakehouse*, B. Shiyal, A c. di Berkeley, CA: Apress, 2021, pagg. 21–48. doi: 10.1007/978-1-4842-7061-5_2.
- [40] S. Mathur, S. Lal Gupta, e P. Pahwa, «Optimizing OLAP Cube for Supporting Business Intelligence and Forecasting in Banking Sector», *Journal of Information*

Technology Management, vol. 13, n. 1, pagg. 81–99, gen. 2021, doi: 10.22059/jitm.2021.80026.

[41] A. Konikov, E. Kulikova, e O. Stifeeva, «Research of the possibilities of application of the Data Warehouse in the construction area», *MATEC Web Conf.*, vol. 251, pag. 03062, 2018, doi: 10.1051/mateconf/201825103062.

[42] M. S. Amin, Y. K. Chiam, e K. D. Varathan, «Identification of significant features and data mining techniques in predicting heart disease», *Telematics and Informatics*, vol. 36, pagg. 82–93, mar. 2019, doi: 10.1016/j.tele.2018.11.007.

[43] J. Fanzo *et al.*, «A research vision for food systems in the 2020s: Defying the status quo», *Glob Food Sec*, vol. 26, pag. 100397, set. 2020, doi: 10.1016/j.gfs.2020.100397.

[44] S. Dash, S. K. Shakyawar, e M. Sharma, *Big data in healthcare: management, analysis and future prospects. J Big Data* 6, 54 (2019).

[45] P.-T. Chen, C.-L. Lin, e W.-N. Wu, «Big data management in healthcare: Adoption challenges and implications», *International Journal of Information Management*, vol. 53, pag. 102078, 2020.

[46] N. A. Ghani, S. Hamid, I. A. Targio Hashem, e E. Ahmed, «Social media big data analytics: A survey», *Computers in Human Behavior*, vol. 101, pagg. 417–428, dic. 2019, doi: 10.1016/j.chb.2018.08.039.

[47] U. Fayyad, G. Piatetsky-Shapiro, e P. Smyth, «From Data Mining to Knowledge Discovery in Databases», *AI Magazine*, vol. 17, n. 3, Art. n. 3, mar. 1996, doi: 10.1609/aimag.v17i3.1230.

[48] M. Holsheimer e S. Arno, «Data mining: The search for knowledge in databases.», *Centrum voor Wiskunde en Informatica*, pag. 1, 1994.

[49] C. Vercellis, *Business intelligence. Modelli matematici e sistemi per le decisioni*. McGraw-Hill., 2006.

[50] G. S. Linoff e M. J. A. Berry, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons, 2011.

[51] P. Ahmad, S. Qamar, S. Qasim, e A. Rizvi, «Techniques of Data Mining In Healthcare: A Review».

- [52] J. Han e M. Kamber, *Data Mining Concepts and Techniques*, Morgan Kauffman, 2nd ed. 2006.
- [53] M. Rathi, «Regression Modeling Technique on Data Mining for Prediction of CRM», in *Information and Communication Technologies*, Berlin, Heidelberg, 2010, pagg. 195–200. doi: 10.1007/978-3-642-15766-0_28.
- [54] N. Jothi, N. A. Rashid, e W. Husain, «Data Mining in Healthcare – A Review», *Procedia Computer Science*, vol. 72, pagg. 306–313, gen. 2015, doi: 10.1016/j.procs.2015.12.145.
- [55] Y. Xie *et al.*, «Predicting Days in Hospital Using Health Insurance Claims», *IEEE Journal of Biomedical and Health Informatics*, vol. 19, n. 4, pagg. 1224–1233, lug. 2015, doi: 10.1109/JBHI.2015.2402692.
- [56] H. Leopord, D. W. K. Cheruiyot, e D. S. Kimani, «A Survey and Analysis on Classification and Regression Data Mining Techniques for Diseases Outbreak Prediction in Datasets», pag. 11.
- [57] C. McGregor, C. Catley, e A. James, «A process mining driven framework for clinical guideline improvement in critical care», 2011.
- [58] S. Belciug, «Patients length of stay grouping using the hierarchical clustering algorithm.», *Annals of the University of Craiova-Mathematics and Computer Science Series*, vol. 36, n. 2, pagg. 79–84.
- [59] H. Almarabeh e E. Amer, «A Study of Data Mining Techniques Accuracy for Healthcare», *International Journal of Computer Applications*, vol. 168, pagg. 12–17, giu. 2017, doi: 10.5120/ijca2017914338.
- [60] A. L. Samuel, «Some Studies in Machine Learning Using the Game of Checkers», *IBM Journal of Research and Development*, vol. 3, n. 3, pagg. 210–229, lug. 1959, doi: 10.1147/rd.33.0210.
- [61] M. I. Jordan e T. M. Mitchell, «Machine learning: Trends, perspectives, and prospects», *Science*, vol. 349, n. 6245, pagg. 255–260, lug. 2015, doi: 10.1126/science.aaa8415.
- [62] Z. Ghahramani, «Probabilistic machine learning and artificial intelligence», *Nature*, vol. 521, n. 7553, pagg. 452–459, mag. 2015, doi: 10.1038/nature14541.

- [63] R. C. Deo, «Machine learning in medicine», *Circulation*, vol. 132, n. 20, pagg. 1920–1930, 2015.
- [64] J. Kamruzzaman e R. K. Begg, «Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait», *IEEE Transactions on Biomedical Engineering*, vol. 53, n. 12, pagg. 2479–2490, dic. 2006, doi: 10.1109/TBME.2006.883697.
- [65] R. Das, I. Turkoglu, e A. Sengur, «Diagnosis of valvular heart disease through neural networks ensembles», *Computer methods and programs in biomedicine*, vol. 93, n. 2, pagg. 185–191, 2009.
- [66] Y. Kanzawa, «Generalization Property of Fuzzy Classification Function for Tsallis Entropy-Regularization of Bezdek-Type Fuzzy C-Means Clustering», in *Modeling Decisions for Artificial Intelligence*, Cham, 2020, pagg. 119–131. doi: 10.1007/978-3-030-57524-3_10.
- [67] Y. Wu, H. Duan, e S. Du, «Multiple fuzzy c-means clustering algorithm in medical diagnosis», *Technology and Health Care*, vol. 23, n. s2, pagg. S519–S527, gen. 2015, doi: 10.3233/THC-150989.
- [68] A. T. Azar, S. A. El-Said, e A. E. Hassanien, «Fuzzy and hard clustering analysis for thyroid disease», *Computer methods and programs in biomedicine*, vol. 111, n. 1, pagg. 1–16, 2013.
- [69] L. Sun e J. Xu, «Feature selection using mutual information based uncertainty measures for tumor classification», *Bio-Medical Materials and Engineering*, vol. 24, n. 1, pagg. 763–770, gen. 2014, doi: 10.3233/BME-130865.
- [70] S.-W. Lin, K.-C. Ying, S.-C. Chen, e Z.-J. Lee, «Particle swarm optimization for parameter determination and feature selection of support vector machines», *Expert Systems with Applications*, vol. 35, n. 4, pagg. 1817–1824, nov. 2008, doi: 10.1016/j.eswa.2007.08.088.
- [71] F. Pernkopf, «Bayesian network classifiers versus selective k-NN classifier», *Pattern recognition*, vol. 38, n. 1, pagg. 1–10, 2005.
- [72] M. J. Islam, Q. M. J. Wu, M. Ahmadi, e M. A. Sid-Ahmed, «Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers», in *2007*

International Conference on Convergence Information Technology (ICCIT 2007), nov. 2007, pagg. 1541–1546. doi: 10.1109/ICCIT.2007.148.

[73] R. Armañanzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin, e P. Larrañaga, «Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach», *Artificial Intelligence in Medicine*, vol. 58, n. 3, pagg. 195–202, lug. 2013, doi: 10.1016/j.artmed.2013.04.002.

[74] C.-H. Jen, C.-C. Wang, B. C. Jiang, Y.-H. Chu, e M.-S. Chen, «Application of classification techniques on development an early-warning system for chronic illnesses», *Expert Systems with Applications*, vol. 39, n. 10, pagg. 8852–8858, 2012.

[75] S. C. Bagui, S. Bagui, K. Pal, e N. R. Pal, «Breast cancer detection using rank nearest neighbor classification rules», *Pattern Recognition*, vol. 36, n. 1, pagg. 25–34, gen. 2003, doi: 10.1016/S0031-3203(02)00044-4.

[76] M. Aly, «Survey on Multiclass Classification Methods». 2005.

[77] R. Caruana e A. Niculescu-Mizil, «An empirical comparison of supervised learning algorithms», in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, giu. 2006, pagg. 161–168. doi: 10.1145/1143844.1143865.

[78] M. Gudadhe, K. Wankhade, e S. Dongre, «Decision support system for heart disease based on support vector machine and Artificial Neural Network», in *2010 International Conference on Computer and Communication Technology (ICCCCT)*, set. 2010, pagg. 741–745. doi: 10.1109/ICCCCT.2010.5640377.

[79] L. Rokach e O. Maimon, «Top-down induction of decision trees classifiers - a survey», *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, n. 4, pagg. 476–487, nov. 2005, doi: 10.1109/TSMCC.2004.843247.

[80] E. K. Hashi, Md. S. U. Zaman, e Md. R. Hasan, «An expert clinical decision support system to predict disease using classification techniques», in *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, feb. 2017, pagg. 396–400. doi: 10.1109/ECACE.2017.7912937.

[81] D. Xhemali, C. J HINDE, e R. G STONE, «Naïve bayes vs. decision trees vs. neural networks in the classification of training web pages», *D. XHEMALI, CJ HINDE and Roger G. STONE," Naive Bayes vs. Decision Trees vs. Neural Networks in the*

Classification of Training Web Pages", *International Journal of Computer Science Issues, IJCSI, Volume 4, Issue 1, pp16-23, September 2009*, vol. 4, n. 1, 2009.

[82] N. B. Amor, S. Benferhat, e Z. Elouedi, «Naive Bayes vs decision trees in intrusion detection systems», in *Proceedings of the 2004 ACM symposium on Applied computing*, New York, NY, USA, mar. 2004, pagg. 420–424. doi: 10.1145/967900.967989.

[83] A. C. Lorena *et al.*, «Comparing machine learning classifiers in potential distribution modelling», *Expert Systems with Applications*, vol. 38, n. 5, pagg. 5268–5275, mag. 2011, doi: 10.1016/j.eswa.2010.10.031.

[84] Z.-H. Zhou, «Rule extraction: Using neural networks or for neural networks?», *J. Comput. Sci. & Technol.*, vol. 19, n. 2, pagg. 249–253, mar. 2004, doi: 10.1007/BF02944803.

[85] S. B. Kotsiantis, «Supervised Machine Learning: A Review of Classification Techniques», pag. 20.

[86] L. I. Kuncheva, «On the optimality of Naïve Bayes with dependent binary features», *Pattern Recognition Letters*, vol. 27, n. 7, pagg. 830–837, mag. 2006, doi: 10.1016/j.patrec.2005.12.001.

[87] V. Ojha, A. Abraham, e V. Snášel, «Heuristic design of fuzzy inference systems: A review of three decades of research», *Engineering Applications of Artificial Intelligence*, vol. 85, pagg. 845–864, ott. 2019, doi: 10.1016/j.engappai.2019.08.010.

[88] A. Bonarni, «Sistemi Fuzzy», *Mondo digitale*, 2003.

[89] E. H. Mamdani e S. Assilian, «An experiment in linguistic synthesis with a fuzzy logic controller», *International Journal of Man-Machine Studies*, vol. 7, n. 1, pagg. 1–13, gen. 1975, doi: 10.1016/S0020-7373(75)80002-2.

[90] W. Silvert, «Fuzzy indices of environmental conditions», *Ecological Modelling*, vol. 130, n. 1, pagg. 111–119, giu. 2000, doi: 10.1016/S0304-3800(00)00204-0.

[91] L. A. Zadeh, «Fuzzy sets as a basis for a theory of possibility», *Fuzzy Sets and Systems*, vol. 1, n. 1, pagg. 3–28, gen. 1978, doi: 10.1016/0165-0114(78)90029-5.

- [92] P. harliana e R. Rahim, «Comparative Analysis of Membership Function on Mamdani Fuzzy Inference System for Decision Making», *J. Phys.: Conf. Ser.*, vol. 930, pag. 012029, dic. 2017, doi: 10.1088/1742-6596/930/1/012029.
- [93] J. A. Mendez *et al.*, «Improving the anesthetic process by a fuzzy rule based medical decision system», *Artificial Intelligence in Medicine*, vol. 84, pagg. 159–170, gen. 2018, doi: 10.1016/j.artmed.2017.12.005.
- [94] S. Thaker e V. Nagori, «Analysis of Fuzzification Process in Fuzzy Expert System», *Procedia Computer Science*, vol. 132, pagg. 1308–1316, gen. 2018, doi: 10.1016/j.procs.2018.05.047.
- [95] T. J. Ross, «Membership Functions, Fuzzification and Defuzzification», in *Fuzzy Systems in Medicine*, P. S. Szczepaniak, P. J. G. Lisboa, e J. Kacprzyk, A c. di Heidelberg: Physica-Verlag HD, 2000, pagg. 48–77. doi: 10.1007/978-3-7908-1859-8_3.
- [96] G. Arji *et al.*, «Fuzzy logic approach for infectious disease diagnosis: A methodical evaluation, literature and classification», *Biocybernetics and Biomedical Engineering*, vol. 39, n. 4, pagg. 937–955, ott. 2019, doi: 10.1016/j.bbe.2019.09.004.
- [97] D. Jones e A. Mitchell, *Lean thinking for the NHS.*, Lean Enterprise Academy. 2006.
- [98] E. Vignati e P. Bruno, *Organizzazione per processi in sanità: un approccio trasversale all'organizzazione per vincere le resistenze al cambiamento.* Franco Angeli, 2003.
- [99] J. R. Black, D. Miller, e J. Sensel, *The Toyota way to healthcare excellence: increase efficiency and improve quality with Lean*, vol. 1. Health Administration Press Chicago, 2008.
- [100] T. Ohno e N. Bodek, *Toyota Production System: Beyond Large-Scale Production.* Boca Raton: Productivity Press, 1988. doi: 10.4324/9780429273018.
- [101] J. P. Womack, D. T. Jones, e D. Roos, *The Machine That Changed the World: The Story of Lean Production-- Toyota's Secret Weapon in the Global Car Wars That Is Now Revolutionizing World Industry.* Simon and Schuster, 2007.
- [102] P. Amaro, A. C. Alves, e R. M. Sousa, «Lean Thinking: From the Shop Floor to an Organizational Culture», in *Advances in Production Management Systems. Towards*

Smart and Digital Manufacturing, Cham, 2020, pagg. 406–414. doi: 10.1007/978-3-030-57997-5_47.

[103] J. P. Womack e D. T. Jones, «Lean Thinking—Banish Waste and Create Wealth in your Corporation», *Journal of the Operational Research Society*, vol. 48, n. 11, pagg. 1148–1148, nov. 1997, doi: 10.1057/palgrave.jors.2600967.

[104] P. Östlund, «Improving Materials Supply Processes to Assembly Lines Through Toyota Production System and Lean Manufacturing», 2020, Consultato: ott. 21, 2021. [Online]. Disponibile su: <http://lup.lub.lu.se/student-papers/record/9024451>

[105] M. babu Purushothaman, J. Seadon, e D. Moore, «Waste reduction using lean tools in a multicultural environment», *Journal of Cleaner Production*, vol. 265, pag. 121681, 2020.

[106] M. L. George, *Lean Six Sigma for Service*. McGraw-Hill, New York., 2003.

[107] H. De Koning, J. P. Verver, J. van den Heuvel, S. Bisgaard, e R. J. Does, «Lean six sigma in healthcare», *Journal for Healthcare Quality*, vol. 28, n. 2, pagg. 4–11, 2006.

[108] Hisham Kelendar, «Lean Thinking from Toyota Manufacturing to the Healthcare Sector», *Res Med Eng Sci*, vol. 8, n. 5, 2020.

[109] H. Kelendar e M. A. Mohammed, «LEAN AND THE ECRS PRINCIPLE: DEVELOPING A FRAMEWORK TO MINIMISE WASTE IN HEALTHCARE SECTORS», *International Journal of Public Health and Clinical Sciences*, vol. 7, n. 3, Art. n. 3, lug. 2020.

[110] I. Usman, «Lean hospital management implementation in health care service: A multicase study», *Systematic Reviews in Pharmacy*, vol. 11, n. 3, Art. n. 3, 2020.

[111] N. Catalyst, «What is lean healthcare?», *NEJM Catalyst*, 2018.

[112] C. Ricciardi *et al.*, «Lean Six Sigma approach to reduce LOS through a diagnostic-therapeutic-assistance path at A.O.R.N. A. Cardarelli», *The TQM Journal*, vol. 31, n. 5, pagg. 657–672, gen. 2019, doi: 10.1108/TQM-02-2019-0065.

[113] R. Slater, «Get better or get beaten: 31 leadership secrets from GE’s Jack Welch», *Irwin Professional Pub.*, 1994.

[114] C. Busco, *Performance measurement, (un)learning and change – an interdisciplinary perspective*. CEDAM, 2005.

- [115] K. Narasimhan, «The Six Sigma Way: How GE, Motorola, and Other Top Companies Are Honing Their Performance», *The TQM Magazine*, vol. 14, n. 4, pagg. 263–264, gen. 2002, doi: 10.1108/tqmm.2002.14.4.263.1.
- [116] N. Proudlove, C. Moxham, e R. Boaden, «Lessons for Lean in Healthcare from Using Six Sigma in the NHS», *Public Money & Management*, vol. 28, n. 1, pagg. 27–34, feb. 2008, doi: 10.1111/j.1467-9302.2008.00615.x.
- [117] F. Pecoraro, «Il Fascicolo Sanitario Elettronico come strumento di Business Intelligence in sanità», 2013.
- [118] C. Costa, G. Franco, M. Maria, D. B. Domenico, e B. Simone, «Analisi e valutazione del fabbisogno di personale nelle aziende sanitarie attraverso l'utilizzo di strumenti di Business Intelligence: l'esperienza della Regione del Veneto», *Analisi e valutazione del fabbisogno di personale nelle aziende sanitarie attraverso l'utilizzo di strumenti di Business Intelligence: l'esperienza della Regione del Veneto*, pagg. 107–133, 2020.
- [119] A. Fiorillo, A. Sorrentino, A. Scala, V. Abbate, e G. Dell'aversana Orabona, «Improving performance of the hospitalization process by applying the principles of Lean Thinking», *The TQM Journal*, vol. 33, n. 7, pagg. 253–271, gen. 2021, doi: 10.1108/TQM-09-2020-0207.
- [120] I. Latessa *et al.*, «Implementing fast track surgery in hip and knee arthroplasty using the lean Six Sigma methodology», *The TQM Journal*, vol. 33, n. 7, pagg. 131–147, gen. 2021, doi: 10.1108/TQM-12-2020-0308.
- [121] A. M. Ponsiglione *et al.*, «Application of DMAIC Cycle and Modeling as Tools for Health Technology Assessment in a University Hospital», *Journal of Healthcare Engineering*, vol. 2021, pag. e8826048, ago. 2021, doi: 10.1155/2021/8826048.
- [122] I. Latessa, I. Picone, A. Fiorillo, A. Sorrentino, G. D. Orabona, e A. S. Valente, «DMAIC Approach to Reduce LOS in Patients Undergoing Oral Cancer Surgery», in *8th European Medical and Biological Engineering Conference*, Cham, 2020, pagg. 424–433. doi: 10.1007/978-3-030-64610-3_49.
- [123] A. Fiorillo, I. Picone, I. Latessa, e A. Cuocolo, «Modelling the length of hospital stay in medicine and surgical departments», in *2021 International Symposium on Biomedical Engineering and Computational Biology.*, 2021.

- [124] I. Picone, I. Latessa, A. Fiorillo, A. Scala, T. Angela Trunfio, e M. Triassi, «Predicting length of stay using regression and Machine Learning models in Intensive Care Unit: a pilot study», in *2021 11th International Conference on Biomedical Engineering and Technology*, New York, NY, USA, mar. 2021, pagg. 52–58. doi: 10.1145/3460238.3460247.
- [125] C. Ricciardi *et al.*, «Linear discriminant analysis and principal component analysis to predict coronary artery disease», *Health Informatics J*, vol. 26, n. 3, pagg. 2181–2192, set. 2020, doi: 10.1177/1460458219899210.