

ALMADONNA MATER STUDIORUM · UNIVERSITÀ DI  
BOLOGNA

---

SCUOLA DI SCIENZE  
Corso di Laurea in Informatica

Title thesis:  
Subtitle thesis

Relatore:  
Chiar.mo Prof.  
Maurizio Gabbrielli

Presentata da:  
Luca Genova

Correlatori:  
Dott.  
Stefano Pio Zingaro  
Saverio Giallorenzo

Sessione II  
Anno Accademico 2020/2021

# Sommario

Il Machine Learning è un sottoinsieme dell'intelligenza artificiale (AI) che si occupa di creare sistemi che apprendono o migliorano le performance in base ai dati che utilizzano. Intelligenza artificiale è un termine generico e si riferisce a sistemi o macchine che imitano l'intelligenza umana.

Definire in maniera semplice le caratteristiche e le applicazioni del machine learning non è sempre possibile, visto che questo ramo è molto vasto e prevede differenti modalità, tecniche e strumenti per essere realizzato. Inoltre, le differenti tecniche di apprendimento e sviluppo degli algoritmi danno vita ad altrettante possibilità di utilizzo che allargano il campo di applicazione dell'apprendimento automatico rendendone difficile una definizione specifica. Si può tuttavia dire che quando si parla di machine learning si parla di differenti meccanismi che permettono a una macchina intelligente di migliorare le proprie capacità e prestazioni nel tempo. La macchina, quindi, sarà in grado di imparare a svolgere determinati compiti migliorando, tramite l'esperienza, le proprie capacità, le proprie risposte e funzioni.

Alla base dell'apprendimento automatico ci sono una serie di differenti algoritmi che, partendo da nozioni primitive, sapranno prendere una specifica decisione piuttosto che un'altra o effettuare azioni apprese nel tempo.

L'obiettivo di questa tesi è rispondere alla domanda: *Si possono autogenerare dei modelli?* Per arrivare alla risposta ho analizzato in letteratura due casi studio, al fine di tirare fuori i relativi workflow per arrivare alla generalizzazione. Infatti è stato applicato un approccio bottom-up, partendo dai casi studio specifici fino alla generalizzazione.

I casi studio in questione sono:

- AI in radiology
- Demand forecasting

Alla fine scopriremo che non è sempre possibile, dipende fortemente dall'applicazione del modello.

# Indice

# Capitolo 1

## Introduzione

Le prime sperimentazioni per la realizzazione di macchine intelligenti risalgono agli inizi degli anni Cinquanta del Novecento, quando alcuni matematici e statistici iniziarono a pensare di utilizzare i metodi probabilistici per realizzare macchine che potessero prendere decisioni proprio tenendo conto delle probabilità di accadimento di un evento. Il primo grande nome legato al machine learning sicuramente quello di Alan Turing, che ipotizzò la necessità di realizzare algoritmi specifici per realizzare macchine in grado di apprendere. In quegli stessi anni, anche gli studi sull'intelligenza artificiale, sui sistemi esperti e sulle reti neurali vedevano momenti di grossa crescita alternati da periodi di abbandono, causati soprattutto dalle molte difficoltà riscontrate nelle possibilità di realizzazione dei diversi sistemi intelligenti, nella mancanza di sussidi economici e dallo scetticismo che circondava spesso chi provava a lavorarci. A partire dagli anni Ottanta, una serie di interessanti risultati ha portato alla rinascita di questo settore della ricerca: una rinascita che è stata resa possibile da nuovi investimenti nel settore. Alla fine degli anni Novanta l'apprendimento automatico trova nuova linfa vitale in una serie di innovative tecniche legate ad elementi statistici e probabilistici: si trattava di un importante passo che permise quello sviluppo che ha portato oggi l'apprendimento automatico ad essere un ramo della ricerca riconosciuto e altamente richiesto.

A seconda del tipo di algoritmo utilizzato per permettere l'apprendimento alla macchina, ossia a seconda delle modalità con cui la macchina impara ed accumula dati e informazioni, si possono suddividere tre differenti sistemi di apprendimento automatico: supervisionato, non supervisionato e per rinforzo. I tre modelli di apprendimento sono utilizzati in maniera differente a seconda della macchina su cui si deve operare, garantendo così sempre la massima performance e il migliore risultato possibile per la risposta agli stimoli esterni.

L'apprendimento supervisionato consiste nel fornire al sistema informatico della macchina una serie di nozioni specifiche e codificate, ossia di modelli ed esempi che permettono di costruire un vero e proprio database di informazioni e di esperienze. In questo modo, quando la macchina si trova di fronte ad un problema, non dovrà fare altro che attingere

alle esperienze inserite nel proprio sistema, analizzarle, e decidere quale risposta dare sulla base di esperienze già codificate. Questo tipo di apprendimento, in qualche modo, fornito già confezionato e la macchina deve essere solo in grado di scegliere quale sia la migliore risposta allo stimolo che le viene dato.

Gli algoritmi che fanno uso di apprendimento supervisionato vengono utilizzati in molti settori, da quello medico a quello di identificazione vocale: essi, infatti, hanno la capacità di effettuare ipotesi induttive, ossia ipotesi che possono essere ottenute scansionando una serie di problemi specifici per ottenere una soluzione idonea ad un problema di tipo generale.

**Apprendimento automatico** L'apprendimento non supervisionato o senza supervisione prevede invece che le informazioni inserite all'interno della macchina non siano codificate, ossia la macchina ha la possibilità di attingere a determinate informazioni senza avere alcun esempio del loro utilizzo e, quindi, senza avere conoscenza dei risultati attesi a seconda della scelta effettuata. Dovrà essere la macchina stessa, quindi, a catalogare tutte le informazioni in proprio possesso, organizzarle ed imparare il loro significato, il loro utilizzo e, soprattutto, il risultato a cui esse portano. L'apprendimento senza supervisione offre maggiore libertà di scelta alla macchina che dovrà organizzare le informazioni in maniera intelligente e imparare quali sono i risultati migliori per le differenti situazioni che si presentano.

L'apprendimento per rinforzo rappresenta probabilmente il sistema di apprendimento più complesso, che prevede che la macchina sia dotata di sistemi e strumenti in grado di migliorare il proprio apprendimento e, soprattutto, di comprendere le caratteristiche dell'ambiente circostante. In questo caso, quindi, alla macchina vengono forniti una serie di elementi di supporto, quali sensori, telecamere, GPS eccetera, che permettono di rilevare quanto avviene nell'ambiente circostante ed effettuare scelte per un migliore adattamento all'ambiente intorno a loro. Questo tipo di apprendimento è tipico delle auto senza pilota, che grazie a un complesso sistema di sensori di supporto è in grado di percorrere strade cittadine e non, riconoscendo eventuali ostacoli, seguendo le indicazioni stradali e molto altro.

# Capitolo 2

## AI in Radiology

Interpretare immagini mediche e riassumerle sotto forma di referti radiologici è un compito impegnativo, noioso e complesso. Un radiologo fornisce una descrizione completa di un'immagine medica sotto forma di referto radiologico descrivendo reperti normali o anormali e fornendo un riepilogo per il processo decisionale. La ricerca mostra che la pratica della radiologia è soggetta a errori a causa del numero limitato di esperti, dell'aumento del volume dei pazienti e della natura soggettiva della percezione umana. Per ridurre il numero di errori diagnostici e per alleviare il compito dei radiologi, è necessario un sistema di generazione di referti computerizzato in grado di generare automaticamente un referto radiologico per una determinata immagine medica.

### 2.1 Overview

L'imaging medico è la fonte di dati più grande e in più rapida crescita nel settore sanitario che rappresenta circa 90% di tutti i dati medici [1]. I radiologi interpretano abitualmente le immagini mediche e raccontano i loro risultati sotto forma di referti radiologici. A causa della natura soggettiva della percezione umana, i radiologi possono non rilevare sottili scoperte che portano a errori diagnostici. La ricerca mostra che questi errori diagnostici possono raggiungere il 30%, il che rappresenta un gran numero di casi gravi [3], [4]. Per alleviare il ruolo dei radiologi e ridurre l'onere della stesura dei referti, è necessario un sistema di generazione di referti di radiologia assistita da computer in grado di redigere un referto di radiologia data un'immagine medica in modo che un radiologo possa rivederlo, modificarlo se necessario eventuali modifiche. L'apprendimento automatico fornisce un modo efficace per automatizzare l'analisi e la diagnosi per le immagini mediche. Può potenzialmente ridurre l'onere per i radiologi nella pratica della radiologia.

Le applicazioni dell'apprendimento automatico in radiologia includono la segmentazione di immagini mediche (ad es. cervello, colonna vertebrale, polmone, fegato, rene, colon); registrazione di immagini mediche (ad es. registrazione di immagini di organi da diverse

modalit  o serie temporali); sistemi di rilevamento e diagnosi computerizzati per immagini TC o RM (ad es. mammografia, colongrafia TC e CAD nodulo polmonare TC); analisi della funzione o dell'attivit  cerebrale e diagnosi di malattie neurologiche da immagini fMR; sistemi di recupero di immagini basati sul contenuto per immagini TC o RM; e analisi del testo dei referti di radiologia utilizzando l'elaborazione del linguaggio naturale (NLP) e la comprensione del linguaggio naturale (NLU).

### 2.1.1 Deep Learning

Come si pu  vedere le applicazioni del machine learning in radiologia sono molte. Ma c'  una cosa in comune con tutte le applicazioni: il fulcro   imaging diagnostico per arrivare ad una diagnosi.

Per creare un modello di classificazione che mi permette di partire da dati in input (immagini DICOM) fino ad avere un referto medico attendibile bisogna lavorare molto sui dati e nel contesto dell'assistenza sanitaria, il deep learning mostra grandi promesse per l'analisi di dati strutturati (ad es. database, tabelle) e non strutturati (ad es. immagini, testo).

## 2.2 Workflow

Mi sono concentrato principalmente sul processo di assistenza alla creazione di referti medici, analizzando un framework in particolare, e di seguito mostro il workflow che sono riuscito a tirarne fuori

### 2.2.1 Raccolta dei dati

Questo passaggio, insieme al successivo (cura dei dati), vengono eseguiti per standardizzare e migliorare qualit  del set di dati per le successive deep neural network training.

La *raccolta* dati si riferisce al processo di raccolta di informazioni da una o pi  fonti per variabili predefinite per testare ipotesi di ricerca e valutare i risultati.   un prerequisito per la formazione di modelli di deep learning. I dati possono essere dati clinici (biobanca), immagini e relativi metadati (DICOM - standard internazionale che   usato per salvare delle immagini mediche come TC e scansioni a risonanza magnetica) e/o annotazioni (referti di radiologia). Questi ultimi rappresentano annotazioni umane e caratteristiche generate dalla macchina.

Analizziamo il caso specifico in cui viene utilizzata una collezione di referti di radiografie del torace, dove ad ogni referto radiologico   associata una coppia di immagini aventi visualizzazione frontale e laterale. Ogni referto di radiologia   composto da quattro sezioni, vale a dire, confronto, indicazione, risultati e impressione:

- La sezione confronto indica se l'attuale studio di imaging viene confrontato con uno qualsiasi dei precedenti studi di imaging del paziente.
- La sezione indicazione elenca le informazioni sul paziente come et, sesso e informazioni cliniche rilevanti, comprese eventuali malattie e sintomi esistenti.
- La sezione risultati indica se ciascuna area dell'immagine normale, anormale o potenzialmente anormale.
- La sezione impressione riassume i risultati, l'anamnesi clinica del paziente e le indicazioni per lo studio ed considerata la parte più indicativa e importante di un referto radiologico per il processo decisionale.

Il processo di raccolta e quello di cura dei dati sono i passaggi più dispendiosi in termini di tempo in un progetto di intelligenza artificiale, ma fondamentale per qualsiasi addestramento del modello.

Uno dei grandi problemi in questo campo proprio la reperibilità dei dati per l'addestramento. Il problema della disponibilità dei dati per motivi legali e quello della disponibilità dei dati con annotazioni (etichette) di esperti possono creare delle barriere ai fini dell'addestramento.

## **Approvazione istituzionale**

Se un progetto si basa sull'utilizzo di dati di imaging, l'approvazione da parte dei comitati di revisione istituzionali deve essere conforme alle normative regionali. I comitati di revisione istituzionali devono imporre il rispetto dell'autonomia del paziente (consenso libero, informato e continuo) o rinunciare alla necessità del consenso del paziente e trovare un equilibrio tra rischi (ad esempio, prevenzione di violazioni dei dati su larga scala e divulgazione involontaria) e benefici (p. es., migliorare la diagnosi e migliorare la selezione del trattamento).

Una soluzione all'approvazione costituzionale può essere il Transfer Learning, spiegato successivamente nella fase di cura dei dati.

### **2.2.2 Cura dei dati**

La *cura* dei dati si riferisce al processo di esplorazione e pulizia dei dati ai fini di addestrare, convalidare e testare gli algoritmi.

Sebbene siano stati compiuti sforzi per sviluppare strumenti di cura automatica, questo passaggio richiede ancora la conoscenza e la supervisione umana per ottenere set di dati di alta qualità. Ad esempio, dopo la selezione dei casi ammissibili da una coorte basata su una biobanca, l'acquisizione dei dati richiederebbe la raccolta di tutte le immagini corrispondenti pertinenti dal sistema di comunicazione di immagini e archiviazione locale



(PACS). Successivamente, la cura pu richiedere la selezione delle sequenze appropriate, delle fasi vascolari e dei piani di imaging. Questo passaggio pu anche richiedere l'esclusione di casi anomali a causa di artefatti di imaging.

Di seguito descriver alcuni passaggi/regole applicati ai dati.

### **Anonimizzazione dei dati**

Pratiche che garantiscono la riservatezza delle informazioni relative al paziente sono di fondamentale importanza per i progetti di deep learning perch le informazioni mediche sensibili possono essere reidentificate. Quindi, tre concetti devono essere tenuti a mente durante la pianificazione e l'esecuzione del progetto: identificazione, anonimizzazione e pseudonimizzazione. In questo caso il set di dati completamente anonimizzato e non possibile estrarre alcuna informazione specifica del paziente.

### **Esplorazione dei dati e controllo qualit**

La fase di esplorazione dei dati consiste nel valutare le propriet qualitative generali (ad es. tramite visualizzazione) o quantitative (ad es. tramite statistiche) del set di dati grezzi iniziale, al fine di mostrare caratteristiche specifiche, tendenze globali o valori anomali.

### **Etichettatura dei dati per l'addestramento**

I radiologi in genere eseguono misurazioni, disegnano regioni di interesse e commentano le immagini con annotazioni. Il markup si riferisce a "simboli grafici posizionati su un'immagine per rappresentare un'annotazione", mentre l'annotazione si riferisce a informazioni esplicative o descrittive relative al significato di un'immagine generata da un osservatore umano. Dopo la selezione delle immagini appropriate, l'etichettatura dei dati pu richiedere la delineazione delle lesioni, tramite riquadri di delimitazione o maschere di segmentazione accompagnate da annotazioni sul tipo di lesioni e sulla loro posizione.

### **Strategie di campionamento dei set di dati**

Il campionamento dei dati si riferisce alla selezione di sottoinsiemi di dati a scopo di formazione. La capacit di un algoritmo di eseguire un compito specifico su dati non visti chiamata generalizzazione. Per ottimizzare e misurare queste prestazioni, l'intero set di dati disponibile deve essere suddiviso in diversi set. I campioni in tutti i set dovrebbero condividere lo stesso processo di generazione dei dati, pur essendo indipendenti l'uno dall'altro e distribuiti in modo identico. La strategia di campionamento pi frequente nel deep learning dividere il set di dati in set di addestramento, convalida e test. Il rapporto ottimale di campioni distribuiti in ciascuno set varia per ogni problema. Ma come regola empirica, viene comunemente utilizzata una suddivisione dell'80% di formazione, del 10% di convalida e del 10% di divisione del test. Questa divisione consente a pi corsi

di formazione che utilizzano lo stesso set di formazione per cercare gli iperparametri ottimali per massimizzare le prestazioni sul set di convalida. Quando si ottengono le migliori prestazioni sul set di convalida, l'algoritmo viene infine utilizzato una volta sul set di test per misurare e confermare la prestazione finale. Per set di dati pi piccoli, la strategia di campionamento pi comunemente utilizzata la convalida incrociata k-fold. Il set di dati diviso equamente in k pieghe. Per ogni addestramento, l'algoritmo viene addestrato su quasi tutte le pieghe ma testato su una singola piega di controllo dei dati. L'allenamento viene ripetuto k volte utilizzando pieghe di controllo variabili. La prestazione finale la media delle k prestazioni misurate. Gli algoritmi di deep learning generalmente introducono due limitazioni significative per utilizzare sistematicamente la convalida incrociata k-fold. In primo luogo, l'addestramento di algoritmi di deep learning su grandi set di dati di solito implica un carico computazionale intenso che impedisce in pratica un numero elevato di iterazioni di addestramento con risorse limitate. In secondo luogo, l'addestramento delle reti neurali profonde dipende da molti pi iperparametri rispetto agli algoritmi di apprendimento automatico meno profondi.

## Transfer learning

Il Transfer Learning (TL) un problema di ricerca nell'apprendimento automatico (ML) che si concentra sulla memorizzazione delle conoscenze acquisite durante la risoluzione di un problema e l'applicazione a un problema diverso ma correlato. Poich, come detto precedentemente, difficile raccogliere dati medici annotati su larga scala a causa dell'interoperabilit, della privacy e di questioni legali, il TL svolge un ruolo fondamentale nel campo dell'imaging medico, dove difficile accedere a set di dati di immagini etichettati. Nel TL una rete neurale viene inizializzata dai pesi del modello che gi stato addestrato su un set di dati su larga scala.

### DA TENERE?

Per consentire il trasferimento di conoscenza per l'estrazione di incorporamenti di immagini, si inizializza la CNN con pesi pre-addestrati del modello che stato addestrato su un set di dati di classificazione delle immagini su larga scala come ImageNet. Inoltre, dal lato del testo, utilizziamo gli incorporamenti di parole Glove con un vettore di incorporamento di 300 dimensioni per ogni parola e sono stati addestrati su un corpus di testo generico denominato Common Crawl con corpus di testo denominato Common Crawl con 42 miliardi di token, 1,9 milioni di vocaboli. Poich la natura del testo medico diversa dal testo generale, abbiamo anche applicato la conoscenza del dominio iniziando LSTM con RadGlove (Radiology Glove), vettori con incorporamenti di parole a 100 dimensioni e addestrati su 4,5 milioni di rapporti di radiologia presso la Stanford University

## Analisi del linguaggio

Per comprendere la natura del compito, confrontiamo un set di dati generico di sottotitoli di immagini con il nostro. Sebbene il compito della generazione del referto radiologico sembri simile al compito di didascalia delle immagini, ad es. data un'immagine di input, generare una descrizione testuale del suo contenuto. Tuttavia, il compito di generare report molto più complesso rispetto alla generazione di didascalie per immagini generiche. I modelli di sottotitoli delle immagini all'avanguardia generano una breve frase con una lunghezza media di 11.78 parole, mentre per la generazione di report il modello deve generare tipicamente da 5 a 6 frasi. Inoltre, i referti radiologici hanno una media di 36.66 parole mentre una didascalia generica di un'immagine ha una media di 11.78 parole per descrizione. Sulla base di ulteriori analisi in cui vengono riportati i referti di radiologia normale e anormale, abbiamo scoperto che la generazione di referti per immagini anormali è più impegnativa rispetto alla loro controparte normale, dato che i referti anormali sono più lunghi del referto normale poiché i radiologi forniscono una descrizione dettagliata di tutte le anomalie e delle loro regioni salienti.

## Esclusione dei dati

Nei referti di radiologia, due sono le sezioni più importanti:

- Risultati → evidenziano tutte le principali anomalie presenti nell'immagine
- Impressione → riassume i risultati evidenziando se lo studio è positivo o negativo

Nel nostro caso, quindi, ci concentriamo solo sui risultati e sulle sezioni delle impressioni. Alcuni set di dati hanno sezioni di confronto e indicazione mancanti a causa delle quali non sono usate nello studio.

Nel set di dati, ci sono anche referti radiologici in cui mancano reperti e sezioni di impressione, quindi escludiamo tutti i referti di radiologia e le relative immagini mediche in cui mancano sia i risultati che le sezioni di impressione. Questo riduce il set di dati.

Poiché la sezione di impressione è la più indicativa se una scansione è positiva o negativa e non è presente un set di dati etichettato, viene applicata l'elaborazione del testo basata su regole per classificare un referto radiologico in Normale (Negativo) o Anormale (Positivo). In questo studio, viene combinato il testo dei risultati e le sezioni delle impressioni per formulare la didascalia finale per un'immagine medica.

### 2.2.3 Creazione del modello

Il processo di assistenza alla creazione di referti medici può essere suddiviso in due parti:

- Codifica dell'immagine
- Decodifica in referto radiologico

Il codificatore ha la forma di una rete neurale convoluzionale (CNN) che codifica le immagini DICOM, mentre il decodificatore è una rete neurale ricorrente a più stadi (RNN) che traduce le caratteristiche dell'immagine medica in un referto radiologico.

In questo caso andiamo quindi a creare un framework codificatore-decodificatore, le caratteristiche dell'immagine globale vengono prima estratte utilizzando la CNN e quindi inserite in un LSTM (Long short-term memory che è un'architettura di rete neurale ricorrente artificiale utilizzata nel campo dell'apprendimento profondo) per generare una sequenza di parole. Quando si applica la struttura del codificatore-decodificatore alla generazione di referti di radiologia, l'attività può essere considerata come la traduzione di immagini mediche in referti di radiologia.

La figura seguente mostra il diagramma del nostro approccio alla generazione di report da immagini mediche. Per generare referti radiologici da immagini mediche, ci viene fornita un'immagine di input  $I$  e il modello dovrebbe generare una sequenza di parole  $y = (y_1, y_2, \dots, y_N)$ , dove  $y$  denota la descrizione sotto forma di referto radiologico e  $y_1 \dots y_N$  denotano le parole nella relazione.

Dato un set di allenamento  $D = (I, y)$  che consiste in  $(I, y)$  coppie, dove  $I$  rappresenta una data immagine medica e  $y$  rappresenta un referto radiologico accompagnato per l'immagine sotto di  $y = (y_1, y_2, \dots, y_N)$ , addestriamo il modello rispetto ai suoi parametri  $\theta$  al fine di massimizzare un modello probabilistico  $p_\theta(y_1, y_2, \dots, y_N | I)$

## Encoder

Le reti neurali convoluzionali (CNN) sono comunemente utilizzate per le rappresentazioni di immagini in quanto hanno la capacità di catturare caratteristiche a vari livelli di astrazione. Le CNN hanno mostrato prestazioni all'avanguardia nella classificazione delle immagini e nel riconoscimento degli oggetti. I livelli in una CNN includono livelli convoluzionali, livelli di pooling e livelli completamente connessi. In questo studio, viene utilizzato il modello Inception-v3 di Google come codificatore perché ha una struttura significativamente più profonda e, grazie ai blocchi di Inception, apprende una rappresentazione semanticamente ricca delle immagini.

## Decoder

Le reti neurali ricorrenti (RNN) sono reti neurali specializzate che eccellono nell'apprendimento e nell'elaborazione di dati sequenziali come testo e parlato. Gli RNN sono generalmente addestrati dalla retropropagazione nel tempo, il che può portare a problemi di gradienti evanescenti ed esplosivi. Due degli RNN più popolari e importanti che superano questi problemi sono Memoria a lungo termine (LSTM) e Unità ricorrente chiusa (GRU).

## 2.2.4 Formazione del modello

Il framework, quindi, costituito da un codificatore che trasforma l'immagine grezza in una rappresentazione vettoriale e un decodificatore che trasforma il vettore dell'immagine codificata in una sequenza di parole.

In questo framework della rete neurale, una CNN viene solitamente impiegata come codificatore per estrarre una rappresentazione globale di immagini e il decodificatore solitamente un RNN che modella il  $p(S_t|I, S_0, S_1, \dots, S_{t-1})$ , e genera le frasi corrispondenti. L'obiettivo generale del framework del codificatore-decodificatore trovare la sequenza di parole (frasi) pi probabile che descriva una data immagine. Durante l'addestramento, vogliamo trovare il modello  $\theta^*$  che soddisfa lequazione 1:

$$\theta^* = \sum_{(I,S)} \log p(S|I; \theta) \quad (2.1)$$

dove  $\theta$  sono i parametri del modello,  $I$  l'immagine medica di input, e  $S$  il rapporto di verit fondamentale corrispondente all'immagine. La probabilit della descrizione corretta  $S$  per una data immagine  $I$  pu essere modellato come la probabilit congiunta sulle sue parole  $S_0, S_1, \dots, S_{t-1}$  come indicato nell'equazione 2:

$$\log p(S|I) = \sum_{t=0}^N \log p(S_t|I, S_0, S_1, \dots, S_{t-1}) \quad (2.2)$$

dove  $S_0, S_1, S_{t-1}$  denota le parole nell'immagine descrizione.  $S_0$  denota uno speciale  $< start >$  token e  $S_N$  denota uno speciale  $< end >$  token.

La figura seguente mostra il diagramma schematico di una cella LSTM:

## 2.2.5 Analisi dei risultati e miglioramento

Text

## 2.2.6 Distribuzione

Text

## 2.3 Analisi

In questa fase di analisi ho cercato di confrontare alcuni casi che ho trovato in letteratura, sono andato alla ricerca di quali potessero essere le invarianti e le versatilit tra loro in modo da centrare il nostro obiettivo.

### 2.3.1 Invarianti e versatilit  nei dati

#### Invarianti

La cosa principale che   comune a tutti i tipi di modelli che vogliono portare un'immagine DICOM in un referto medico sono i set di dati.

Set di dati formati sempre da: immagine DICOM e/o annotazione.

Ovviamente alcuni input devono essere scartati perch  non presentano le caratteristiche di cui abbiamo bisogno.

#### Versatilit 

La differenza principale   che in alcuni modelli, per i dati di addestramento viene eseguito un aumento. L'aumento dell'immagine sarebbe preferibile per aumentare il livello di robustezza del modello. Ovviamente, l'aumento delle immagini si traduce in un aumento del numero di set di dati per la formazione e potrebbe esserci il rischio del problema di sovradattamento.

Anche se viene esaminato lo stesso paziente, le immagini non sarebbero identiche tra gli esami a causa di lievi differenze di posizionamento. Pertanto, l'uso di immagini ruotate o spostate parallelamente renderebbe il modello robusto per lievi differenze nelle posizioni del paziente. A volte vengono utilizzate anche immagini speculari. L'uso di immagini con rumore aggiunto sarebbe utile per rendere il modello robusto per diversi gradi di rumore dell'immagine. Come gi  detto in precedenza, un modello potrebbe differire da un altro anche per le strategie di campionamento del set di dati:

- Addestramento, convalida e test
- Convalida incrociata k-fold

Anche se il primo   quello pi  utilizzato.

### 2.3.2 Invarianti e versatilit  nel modello

#### Invarianti

Una delle cose comuni a diversi modelli   che per estrarre le caratteristiche dall'immagine viene sempre utilizzata un CNN per le sue elevate prestazioni nel riconoscimento delle immagini.

#### Versatilit 

Per estrarre le parole?? cosa viene usato -; RNN (LSTM)?? sempre lo stesso??

  possibile trovare alcuni modelli (ex: r.AID.ologist) concepiti come modello di ragionamento basato sui casi (Case-Based Reasoning - CBR), alimentato da diversi modelli di

deep learning che sono stati inclusi per migliorarne le prestazioni, oltre a fornire funzionalità aggiuntive. Sebbene CBR sia integrato come un ciclo continuo, le sue sottoparti sono progettate per essere completamente modulari, quindi possono essere facilmente sostituite per adattarsi a nuovi dati.

Il ragionamento basato su casi (CBR) il processo di risoluzione di nuovi problemi basandosi sulle soluzioni di problemi anteriori ed è stato formalizzato, per il ragionamento automatico, come un processo suddiviso in quattro fasi:

1. Recupero: Dato un problema, recuperare in memoria dei casi rilevanti (ricordi) per risolverlo. Un caso composto da un problema, la soluzione e tipicamente alcune annotazioni su come si è arrivati alla soluzione.
2. Riutilizzo: Mappare la soluzione del caso precedente al problema attuale. Ci può comportare alcune modifiche al caso precedente per adattarlo al caso attuale.
3. Revisione: Avendo mappato la soluzione precedente al caso attuale, bisogna provare la nuova soluzione (nel mondo reale o con una simulazione) e, se necessario, rivedere la nuova soluzione.
4. Conservazione: Dopo che la soluzione è stata adattata al problema attuale, memorizzare l'esperienza come nuovo caso.

# Capitolo 3

## Demand forecasting

La previsione della domanda è uno dei problemi principali delle catene di approvvigionamento. Mira ad ottimizzare le scorte, ridurre i costi e aumentare le vendite, i profitti e la fedeltà dei clienti. A tal fine, i dati storici possono essere analizzati per migliorare la previsione della domanda utilizzando vari metodi come tecniche di apprendimento automatico, analisi di serie temporali e modelli di deep learning.

In questo studio viene analizzato il workflow della previsione della domanda con i relativi modelli.

### 3.1 Overview

Poiché la concorrenza aumenta di giorno in giorno tra i rivenditori sul mercato, le aziende stanno concentrando più tecniche di analisi predittiva per ridurre i costi e aumentare la produttività e il profitto.

Esistono diversi tipi di previsione della domanda:

- A breve termine (viene effettuato per un periodo più breve da 3 mesi a 12 mesi).
- Da medio a lungo termine (in genere viene effettuato da più di 12 mesi a 24 mesi in anticipo (36-48 mesi in alcune attività)).
- Livello aziendale interno (questo tipo di previsione si occupa delle operazioni interne dell'azienda come categoria di prodotto, divisione vendite, divisione finanziaria e gruppo di produzione. Ci include previsioni di vendita annuali, stima di COGS, margine di profitto netto, flusso di cassa, ecc.)
- Passivo (Si effettua per imprese stabili con piani di crescita molto conservativi. Semplici estrapolazioni di dati storici vengono effettuate con ipotesi minime.)



- Attivo (viene eseguito per ridimensionare e diversificare le aziende con piani di crescita aggressivi in termini di attività di marketing, espansione del portafoglio di prodotti e considerazione delle attività della concorrenza e dell'ambiente economico esterno).

I rivenditori in diversi settori sono alla ricerca di soluzioni automatizzate di previsione della domanda e rifornimento che utilizzino big data e tecnologie di analisi predittiva [3]. I metodi di previsione tradizionali si basano su approcci di previsione basati su serie temporali. Questi approcci di previsione prevedono la domanda futura in base a dati di serie temporali storiche, ovvero una sequenza di punti dati misurati a intervalli successivi nel tempo. Questi metodi utilizzano un numero limitato di dati storici di serie temporali relativi alla domanda. Negli ultimi due decenni, I modelli di data mining e machine learning hanno attirato maggiore attenzione e sono stati applicati con successo alla previsione delle serie temporali.

I metodi di previsione dell'apprendimento automatico possono utilizzare una grande quantità di dati e funzionalità correlate alla domanda e prevedere la domanda e i modelli futuri utilizzando diversi algoritmi di apprendimento. Tra molti metodi di apprendimento automatico, i metodi di deep learning (DL) sono diventati molto popolari e sono stati recentemente applicati a molti campi come il riconoscimento di immagini e parole, l'elaborazione del linguaggio naturale e la traduzione automatica. I metodi DL hanno prodotto previsioni e risultati migliori, rispetto ai tradizionali algoritmi di apprendimento automatico in molte ricerche. L'Ensemble Learning (EL) è anche un'altra metodologia per aumentare le prestazioni del sistema.

Un sistema di insieme composto da due parti:

- generazione di insieme
- integrazione di insieme [4].

Nella parte di generazione dell'ensemble, viene generato un insieme diversificato di modelli di previsione di base utilizzando metodi o campioni diversi. Nella parte di integrazione, le previsioni di tutti i modelli vengono combinate utilizzando una strategia di integrazione.

Oltre ai metodi delle serie temporali, gli approcci di intelligenza artificiale stanno diventando popolari con la crescita delle tecnologie dei big data. Nei problemi di classificazione, le prestazioni degli algoritmi di apprendimento dipendono principalmente dalla natura della rappresentazione dei dati [23].

## 3.2 Workflow

### 3.2.1 Raccolta dati

Il primo compito quando si avvia il progetto di previsione della domanda fornire al cliente informazioni significative. Il processo include i seguenti passaggi:

1. Raccogli i dati disponibili
2. Esamina brevemente la struttura dei dati, l'accuratezza e la coerenza
3. Esegui alcuni test e progetti pilota sui dati
4. Guarda attraverso un riepilogo statistico

In questo caso studio il tipo di dato pu essere molto differente in base all'obiettivo aziendale. Distinguiamo i dati in:

- Strutturati
- Non strutturati

I seguenti dati possono essere utilizzati per costruire modelli di previsione.

Nei dati strutturati possiamo trovare dati interni come i dati di vendita e-commerce, le transazioni di vendita, gli ordini d'acquisto, le proprie informazioni POS e dati esterni come tempo metereologico, spedizioni/ricevute del negozio del cliente, le informazioni POS cliente e dati sindacati di terze parti.

Nei dati non strutturati possiamo trovare dati interni come siti web, recensioni, campagne di marketing, app, dispositivi in negozio e dati esterni come social media, stream di click, IoT, dispositivi di geolocalizzazione, video, linguaggio naturale.

### 3.2.2 Comprensione e preelaborazione dei dati

Indipendentemente da ci che vorremmo prevedere, la qualit dei dati una componente fondamentale di una previsione accurata della domanda.

#### Parametri di qualit dei dati

Quando si costruisce un modello di previsione, i dati vengono valutati in base ai seguenti parametri:

- Consistenza
- Precisione

- Validit 
- Rilevanza
- Accessibilit 
- Completezza
- Detalizzazione

### **Preelaborazione dei dati**

In realt , i dati raccolti dalle aziende spesso non sono ideali. Questi dati di solito devono essere puliti, analizzati per lacune e anomalie, verificati per pertinenza e ripristinati. La preelaborazione pu  essere applicata in tre forme: adeguamenti stagionali, log o trasformazioni di potenza e rimozione della tendenza.

- Dati originali: non viene applicata alcuna preelaborazione.
- Trasformazione dei dati: Ai dati originali viene applicato il log o la trasformazione di potenza Box-Cox [66] per ottenere la stazionariet  nella varianza.
- Destagionalizzazione dei dati: i dati sono considerati stagionali se esiste un coefficiente di autocorrelazione significativo al lag 12. In tal caso i dati vengono destagionalizzati utilizzando il classico approccio di decomposizione moltiplicativa [39]. L'addestramento dei pesi ML, o l'ottimizzazione dei metodi statistici, viene successivamente effettuato sui dati destagionalizzati. I pronostici ottenuti vengono poi ristagionali per determinare i pronostici finali. Ci non avviene nel caso dei metodi ETS e ARIMA poich  includono modelli stagionali, selezionati utilizzando test relativi e criteri informativi che si occupano direttamente della stagionalit  e della complessit  del modello.
- Detrending dei dati: viene eseguito un test di Cox-Stuart [67] per stabilire se sia necessario utilizzare un trend lineare deterministico, o in alternativa un primo differenziamento, per eliminare il trend dai dati e ottenere la stazionariet  nella media.
- Combinazione dei tre precedenti: i vantaggi delle singole tecniche di pre-elaborazione vengono applicati simultaneamente per regolare i dati originali. Al fine di accelerare i calcoli, prima determiniamo la migliore alternativa di pre-elaborazione per migliorare le prestazioni di previsione post-campione in un passo avanti del metodo MLP (il pi  popolare tra quelli ML) e poi applichiamo quella ai restanti modelli ML.

## Clustering dei dati

Identificare e separare diversi modelli nei dati e utilizzare ciascuno di essi come set di dati indipendenti. Questo noto come clustering e pu essere ottenuto utilizzando algoritmi di apprendimento automatico. Nell'apprendimento non supervisionato, un problema comune identificare e classificare correttamente un certo numero di classi identiche (cluster) presenti nei dati (Bishop 2006). All'inizio, il numero di cluster,  $k$ , solitamente sconosciuto. Trovarlo pu essere ottenuto eseguendo l'algoritmo di clustering per diversi valori di  $k$  e valutando i risultati per ciascun valore.

Il numero ottimale di cluster quello che ottimizza un determinato criterio, come la varianza tra osservazioni nello stesso cluster. Complessivamente, il numero di cluster e il modo in cui vengono trovati, la formulazione della distanza e l'implementazione dell'algoritmo di clustering stesso sono reciprocamente influenzati e diverse combinazioni di essi produrranno risultati diversi.

Qui, l'analisi di clustering pu essere utilizzata per verificare la stagionalità e altre relazioni tra i dati di input.

Una volta che i dati sono stati puliti, generati e verificati per pertinenza, vengono strutturati in un modulo completo.

La comprensione dei dati il compito successivo una volta completate la preparazione e la strutturazione. Non ancora un modello, ma un modo eccellente per comprendere i dati tramite la visualizzazione. Prima di iniziare a creare modelli, bisogna dividere il set di dati per l'addestramento, la convalida e il test.

### 3.2.3 Costruzione del modello

Non esistono algoritmi di previsione "taglia unica". Spesso, le funzionalità di previsione della domanda consistono in diversi approcci di apprendimento automatico. La scelta dei modelli di apprendimento automatico dipende da diversi fattori, come l'obiettivo aziendale, il tipo di dati, la quantità e la qualità dei dati, il periodo di previsione, ecc.

Mostreremo innanzitutto i metodi statistici tradizionali e poi ci concentreremo su quelli utilizzando il ML, in particolare sull'approccio per serie temporali, che quello più utilizzato.

Nei modelli di previsione delle serie temporali, l'approccio classico consiste nel raccogliere dati storici, analizzare questi dati sottostanti e utilizzare il modello per prevedere il futuro [28]. Questi algoritmi sono algoritmi di previsione comunemente usati nel dominio di previsione della domanda di serie temporali.

#### Perceptron multistrato (MLP)

Il perceptron multistrato (spesso chiamato semplicemente rete neurale) forse l'architettura di rete più popolare oggi in uso sia per la classificazione che per la regressione

(Bishop, 1995). Il MLP è dato come segue:

$$\hat{y} = v_0 + \sum_{j=1}^{NH} v_j g(w_j^T x'), \quad (3.1)$$

dove  $x'$  il vettore di input  $x$ , aumentato di 1, cioè  $x' = (1, x^T)^T$ ,  $w_j$  il vettore dei pesi per  $j$ -esimo nodo nascosto,  $v_0, v_1, \dots, v_{NH}$  sono i pesi per l'output nodo e  $\hat{y}$  l'output di rete. La funzione  $g$  rappresenta l'output del nodo nascosto, ed è data in termini di una funzione di schiacciamento, ad esempio la funzione logistica:  $g(u) = 1/(1 + \exp(-u))$ . Un modello correlato nella letteratura econometrica è il modello di autoregressione a transizione graduale che si basa anche sulla costruzione di funzioni lineari e transizioni di funzioni logistiche (Medeiros e Veiga, 2000; van Dijk et al., 2002). L'MLP è un modello fortemente parametrizzato e selezionando il numero di nodi nascosti  $NH$  possiamo controllare la complessità del modello. La svolta che ha dato credito alla capacità delle reti neurali è la proprietà di approssimazione universale (Cybenko, 1989; Funahashi, 1989; Hornik et al., 1989; Leshno et al., 1993). In determinate condizioni, se la funzione del nodo nascosto  $g$ , ogni data funzione continua su un insieme compatto può essere approssimata il più vicino possibile a quella data arbitrariamente utilizzando una rete con un numero finito di nodi nascosti. Sebbene questo sia un risultato rassicurante, è fondamentale evitare un'eccessiva parametrizzazione, specialmente nelle applicazioni di previsione che in genere hanno una quantità limitata di dati altamente rumorosi. La selezione del modello (attraverso la selezione del numero di nodi nascosti) ha quindi suscitato molto interesse nella letteratura sulle reti neurali (vedi Anders e Korn, 1999; Medeiros et al., 2006 per esempio). Viene usata una procedura di convalida  $K$ -fold per selezionare il numero di nodi nascosti. Per ottenere i pesi, viene definito l'errore quadratico medio e i pesi vengono ottimizzati utilizzando tecniche di gradiente. Il metodo più noto, basato sul concetto di discesa più ripida, è l'algoritmo di backpropagation. Un metodo di ottimizzazione del secondo ordine chiamato Levenberg Marquardt è generalmente noto per essere più efficiente dell'algoritmo di retropropagazione di base.

### **Rete neurale bayesiana (BNN)**

Una rete neurale bayesiana (BNN) è una rete neurale progettata sulla base di una formulazione probabilistica bayesiana (MacKay, 1992a,b). In quanto tali, i BNN sono correlati al concetto di statistica classica della stima dei parametri bayesiani e sono anche correlati al concetto di regolarizzazione come nella regressione della cresta. Le BNN hanno goduto di un'ampia applicabilità in molte aree come l'economia/finanza (Gencay e Qi, 2001) e l'ingegneria (Bishop, 1995). L'idea di BNN è di trattare i parametri o i pesi di rete come variabili casuali, obbedendo a una distribuzione a priori. Questa distribuzione è progettata in modo da favorire modelli a bassa complessità, cioè modelli che producono accoppiamenti lisci. Una volta osservati i dati, viene valutata la distribuzione a posteriori

dei pesi e pu essere calcolata la previsione della rete. Le previsioni rifletteranno quindi sia l'aspetto di levigatezza imposto dal precedente sia l'aspetto di accuratezza del fitness imposto dai dati osservati.

Un concetto strettamente correlato l'aspetto della regolarizzazione, per cui la seguente funzione obiettivo costruita e minimizzata

$$J = \alpha E_D + (1 - \alpha) E_W \quad (3.2)$$

dove  $E_D$  la somma dei quadrati degli errori nelle uscite di rete,  $E_W$  la somma dei quadrati dei parametri di rete (cio i pesi) ed il parametro di regolarizzazione. Per l'approccio bayesiano, la scelta tipica del priore (credenze iniziali su un evento in termini di distribuzione di probabilit) la seguente densit normale che d pi peso ai valori dei parametri di rete pi piccoli

$$p(w) = \left( \frac{1 - \alpha}{\pi} \right)^{\frac{L}{2}} e^{-(1-\alpha)E_W} \quad (3.3)$$

dove  $L$  indica il numero di parametri (pesi). Il posteriore quindi dato da

$$p(w|D, \alpha) = \frac{p(D|w, \alpha)p(w, \alpha)}{p(D, \alpha)} \quad (3.4)$$

dove  $D$  rappresenta i dati osservati. Assumendo errori normalmente distribuiti, la densit di probabilit dei dati dati i parametri pu essere valutata come

$$p(D|w, \alpha) = \left( \frac{\alpha}{\pi} \right)^{\frac{M}{2}} e^{-(1-\alpha)E_D} \quad (3.5)$$

dove  $M$  il numero di punti dati di addestramento. Sostituendo le espressioni per le densit in (3) e (5) in (4), otteniamo

$$p(w|D, \alpha) = c \exp(-J) \quad (3.6)$$

dove  $c$  una costante di normalizzazione. La costante di regolarizzazione  $\alpha$  anche determinata utilizzando concetti bayesiani, da

$$p(\alpha|D) = \frac{p(D|\alpha)p(\alpha)}{p(D)} \quad (3.7)$$

Entrambe le espressioni (6) e (7) dovrebbero essere massimizzate per ottenere rispettivamente i pesi ottimali e il parametro  $\alpha$ . Il termine  $p(D|\alpha)$  in (7) ottenuto da un'approssimazione quadratica di  $J$  in termini di pesi e quindi integrando i pesi.

## Reti neurali di regressione generalizzata (GRNN)

Nadaraya e Watson hanno sviluppato questo modello (Nadaraya, 1964; Watson, 1964). comunemente chiamato stimatore Nadaraya-Watson o stimatore di regressione del kernel. Nella comunit del machine learning, viene in genere utilizzato il termine rete neurale di regressione generalizzata (o GRNN). Useremo quest'ultimo termine. Il modello GRNN un modello non parametrico in cui la previsione per un dato punto dati  $x$  data dalla media degli output target dei punti dati di addestramento nelle vicinanze del dato punto  $x$  (Hardle, 1990). La media locale costruita pesando i punti in base alla loro distanza da  $x$ , usando alcune funzioni del kernel. La stima solo la somma ponderata delle risposte osservate (o output target) data da

$$\text{equation} - \text{one} \tag{3.8}$$

dove i pesi  $w_m$  sono dati da

$$\text{equation} - \text{two} \tag{3.9}$$

dove  $y_m$  l'output di destinazione per il punto dati di training  $x_m$  e  $K$  la funzione del kernel. Abbiamo usato il tipico kernel gaussiano *formula*. Il parametro  $h$ , chiamato larghezza di banda, un parametro importante in quanto determina la levigatezza dell'adattamento, poich aumentandolo o diminuendolo controller la dimensione della regione di levigatura.

## Regressione K-Nearest Neighbor (KNN)

KNN un metodo non parametrico che basa la sua previsione sugli output di destinazione dei  $K$  vicini pi vicini del dato punto di interrogazione (vedi Hastie et al., 2001). Nello specifico, dato un punto dati, calcoliamo la distanza euclidea tra quel punto e tutti i punti del training set. Quindi selezioniamo i punti dati di allenamento  $K$  pi vicini e impostiamo la previsione come media dei valori di output target per questi  $K$  punti. Quantitativamente parlando, sia  $I(x)$  l'insieme di  $K$  vicini pi prossimi del punto  $x$ . Allora la previsione data da

$$\hat{y} = \frac{1}{K} \sum_{m \in I(x)} y_m \tag{3.10}$$

dove di nuovo  $y_m$  l'output di destinazione per il punto dati di training  $x_m$ . Naturalmente  $K$  un parametro chiave in questo metodo e deve essere selezionato con cura. Un  $K$  grande porter a un adattamento pi fluido, e quindi a una varianza inferiore, ovviamente a scapito di un bias maggiore, e viceversa per un  $K$  piccolo.

## Regressione del vettore di supporto (SVR)

La regressione del vettore di supporto (Scholkopf e Smola, 2001; Smola e Scholkopf, 2003) è un metodo di successo basato sull'utilizzo di uno spazio di caratteristiche ad alta dimensione (formato trasformando le variabili originali) e penalizzando la conseguente complessità utilizzando un termine di penalità aggiunto alla funzione di errore. Consideriamo prima per l'illustrazione un modello lineare. Quindi, la previsione data da

$$\text{equation} - \text{one} \tag{3.11}$$

dove  $w$  è il vettore del peso,  $b$  il bias e  $x$  il vettore di input. Sia  $x_m$  e  $y_m$  denotano, rispettivamente, il  $m$ -esimo vettore di input di addestramento e output di destinazione,  $m = 1, \dots, M$ . La funzione di errore data da

$$\text{equation} - \text{two} \tag{3.12}$$

Il primo termine nella funzione di errore è un termine che penalizza la complessità del modello. Il secondo termine è la funzione di perdita  $\epsilon$  - *insensitive*, definita come

$$\text{equation} - \text{three} \tag{3.13}$$

Non penalizza gli errori sottostanti, consentendo un po' di spazio di manovra per lo spostamento dei parametri per ridurre la complessità del modello. Si può dimostrare che la soluzione che minimizza la funzione di errore data da

$$\text{equation} - \text{four} \tag{3.14}$$

dove  $\alpha_m$  e  $\alpha_m^*$  sono moltiplicatori di Lagrange. I vettori di addestramento che forniscono moltiplicatori di Lagrange diversi da zero sono chiamati vettori di supporto e questo è un concetto chiave nella teoria SVR. I vettori non di supporto non contribuiscono direttamente alla soluzione e il numero di vettori di supporto è una misura della complessità del modello (vedi Chalimourda et al., 2004; Cherkassky e Ma, 2004). Questo modello viene esteso al caso non lineare attraverso il concetto di kernel  $K$ , dando una soluzione

$$\text{equation} - \text{five} \tag{3.15}$$

Un kernel comune è il kernel gaussiano. Assumiamo che la sua larghezza sia  $\sigma K$  (la deviazione standard della funzione gaussiana).

## Rete Neurale Ricorrente (RNN)

L'RNN semplice, noto anche come rete di Elman, ha una struttura simile all'MLP, ma contiene connessioni di feedback per tenere conto degli stati precedenti insieme all'input



corrente prima di produrre l'output o gli output finali. Questo viene fatto salvando una copia dei valori precedenti del livello contenente i nodi ricorrenti e utilizzandoli come input aggiuntivo per il passaggio successivo. A questo proposito, la rete pu esibire un comportamento temporale dinamico per una sequenza temporale. In questo studio il modello utilizzato per implementare la RNN quello sequenziale. composto da due livelli, uno nascosto contenente nodi ricorrenti e uno di output contenente uno o pi nodi lineari. A causa degli elevati requisiti computazionali, non stata utilizzata la convalida k-fold per scegliere l'architettura di rete ottimale per serie, ma piuttosto tre nodi di input e sei unit ricorrenti, che formano lo strato nascosto, per tutte le serie temporali del set di dati.

### **Rete neurale di memoria a lungo termine (LSTM)**

La rete LSTM simile alla RNN discussa sopra ed stata proposta per evitare il problema della dipendenza a lungo termine. Il vantaggio delle unit LSTM rispetto alle normali unit RNN la loro capacit di conservare le informazioni per periodi di tempo pi lunghi grazie alla loro architettura complessa che consiste in diverse porte con il potere di rimuovere o aggiungere informazioni allo stato dell'unit. Simile a RNN, il modello utilizzato per implementare la rete LSTM quello sequenziale composto da uno strato nascosto e uno di output. Allo stesso modo, a causa dell'elevato tempo di calcolo, l'architettura del modello costituita da tre nodi di input, sei unit LSTM che formano lo strato nascosto e un singolo nodo lineare nello strato di output. La funzione di attivazione lineare viene utilizzata prima dell'uscita di tutte le unit e quella del sigmoide rigido per il passo ricorrente.

### **Stacking (Insieme)**

Stacking o Stacked Generalization un algoritmo di apprendimento automatico di insieme. Utilizza un algoritmo di meta-apprendimento per apprendere come combinare al meglio le previsioni di due o pi algoritmi di apprendimento automatico di base. Potremmo usare Stacking per combinare vari modelli e fare nuove previsioni. L'architettura di un modello di impilamento (stacking model) coinvolge due o pi modelli di base, spesso indicati come modelli di livello 0 e un meta-modello che combina le previsioni dei modelli di base indicati come modello di livello 1:

- Modelli di livello 0 (modelli base): i modelli si adattano ai dati di addestramento e le cui previsioni vengono compilate.
- Modello di Livello 1 (Meta-Modello): modello che apprende come combinare al meglio le previsioni dei modelli base.

Il meta-modello addestrato sulle previsioni fatte dai modelli di base su dati fuori campione. In altre parole, i dati non utilizzati per addestrare i modelli di base vengono inviati

ai modelli di base, vengono effettuate previsioni e queste previsioni, insieme agli output previsti, forniscono le coppie di input e output del set di dati di addestramento utilizzato per adattarsi al meta-modello. Gli output dei modelli di base utilizzati come input per il meta-modello possono essere valori reali in caso di regressione e valori di probabilità, valori simili alla probabilità o etichette di classe in caso di classificazione. La regressione a stack è una tecnica di apprendimento di insieme per combinare più modelli di regressione tramite un meta-regressore. I singoli modelli di regressione vengono addestrati in base al training set completo; quindi, il meta-regressore viene adattato in base agli output - meta-caratteristiche - dei singoli modelli di regressione nell'insieme.

### **3.2.4 Formazione del modello**

Una volta sviluppati i modelli di previsione, il momento di iniziare il processo di formazione. Quando si addestrano i modelli di previsione, i data scientist di solito utilizzano dati storici. Elaborando questi dati, gli algoritmi forniscono modelli addestrati pronti per l'uso.

### **3.2.5 Validazione del modello**

Questo passaggio richiede l'ottimizzazione dei parametri del modello di previsione per ottenere prestazioni elevate. Utilizzando un metodo di ottimizzazione della convalida incrociata in cui il set di dati di training è suddiviso in dieci parti uguali. I data scientist addestrano modelli di previsione con diversi set di iperparametri. L'obiettivo di questo metodo è capire quale modello ha la previsione più accurata.

### **3.2.6 Miglioramento**

Quando si ricercano le migliori soluzioni aziendali, gli scienziati dei dati di solito sviluppano diversi modelli di apprendimento automatico. Poiché i modelli mostrano diversi livelli di accuratezza, gli scienziati scelgono quelli che soddisfano meglio le loro esigenze aziendali. La fase di miglioramento prevede l'ottimizzazione dei risultati analitici. Ad esempio, utilizzando tecniche di insieme di modelli, è possibile ottenere una previsione più accurata. In tal caso, l'accuratezza viene calcolata combinando i risultati di più modelli di previsione.

### **3.2.7 Distribuzione**

Questa fase presuppone l'integrazione dei modelli di previsione nell'uso della produzione. Ti consigliamo inoltre di impostare una pipeline per aggregare nuovi dati da utilizzare per le tue prossime funzionalità di intelligenza artificiale. Ci puoi fare risparmiare molto

lavoro di preparazione dei dati in progetti futuri. In questo modo aumenta anche la precisione e la varietà di ciò che potresti essere in grado di prevedere.

### **3.3 Analisi**

In questa fase di analisi ho cercato di confrontare tutti i casi che ho trovato in letteratura, sono andato a ricercare quali potessero essere le invarianti e le versatilità tra loro.

#### **3.3.1 Invarianti e versatilità nei dati**

- Ci sono molti modi di pretrattare i dati e dipendente anche dal tipo di modello utilizzato (stagionalità? Serie temporali?)
- il successo di un modello/algoritmo dipende fortemente da come i dati vengono pretrattati
- Dipende dai casi ma ci sono molti dati non strutturati che provengono anche da variabili esterne (la temperatura o il meteo se si tratta di previsione della domanda dell'acqua)

#### **3.3.2 Invarianti e versatilità nel modello**

- Moltissimi modelli (ovviamente scelta in base al caso di previsione), spiegare quali metodi sono migliori di altri in determinati contesti.
- Negli ultimi anni i più utilizzati sono RNN, K- nearest e LSTM.
- I metodi statistici tradizionali sono migliori di quelli con ML.
- Modello migliore = combinazione tra i diversi modelli (inserendo anche quelli tradizionali)

## Capitolo 4

### Analisi tra i diversi casi studio

# Capitolo 5

## Frameworks