



UNIVERSITÀ DELLA VALLE D'AOSTA
UNIVERSITÉ DE LA VALLÉE D'AOSTE



DECISIONS_LAB

METODI E STRUMENTI DI INTELLIGENZA ARTIFICIALE PER IL FORECASTING DEI FLUSSI DELLA MOBILITÀ TURISTICA SPAZIALE E TEMPORALE

TIZIANA CIANO
MARCO ALDERIGHI
MASSIMILIANO FERRARA

Metodi e Strumenti di Intelligenza Artificiale per il forecasting dei flussi della mobilità turistica spaziale e temporale

Tiziana Ciano ¹– Marco Alderighi ¹– Massimiliano Ferrara²

Centro Stampa d'Ateneo
ISBN 978-88-99352-98-1

¹ Dipartimento di Scienze Economiche e Politiche, Università della Valle D'Aosta, email: t.ciano@univda.it; m.alderighi@univda.it;

² Dipartimento di Giurisprudenza, Economia e Scienze Umane, Università Mediterranea di Reggio Calabria, email: massimiliano.ferrara@unirc.it

Premessa

Il presente scritto passa in rassegna, con l'aggiunta di alcuni tratti di originalità su alcuni aspetti di strumentazione matematica, i modelli di forecasting di derivazione IA (Intelligenza artificiale) che a nostro parere troveranno ampia applicazione – nel prossimo futuro - nella modellistica previsionale finalizzata all'analisi di fenomeniche socio-economiche, con uno specifico focus in ambito turistico. A questo proposito, è doveroso evidenziare, come il contenuto del manoscritto rappresenti il “core” scientifico della Relazione a consuntivo relativa all'attività di ricerca riguardante il PROGETTO MONTUR -MONITORAGGIO REAL-TIME E PREVISIONE DEI FLUSSI TURISTICI IN VALLE D'AOSTA ATTRAVERSO SENSORI DISTRIBUITI E STRUMENTI DI MACHINE LEARNING E BIG DATA”.

Si ringrazia, altresì, il Piano Nazionale di Ripresa e Resilienza (PNRR), Progetto NODES – Nord Ovest Digitale e Sostenibile, Missione 4, Componente 2, Investimento 1.5, “Creazione e Rafforzamento di “Ecosistemi dell'innovazione”, costruzione di “leader territoriali di R&S”, finanziato dall'Unione europea NextGenerationEU.

Aosta e Reggio Calabria, marzo 2023

Tiziana Ciano, Marco Alderighi e Massimiliano Ferrara

Indice

1. Introduzione	4
2. Intelligenza Artificiale, Machine Learning e Big Data nel Turismo.....	5
3. Stato dell'arte	8
4. Individuazione dei principali tools, metodi e modelli utilizzabili per la realizzazione del sistema previsionale integrato.....	13
4.1 Machine Learning: training supervisionato e non supervisionato	13
4.2 Deep Learnig e Reti Neurali	16
4.3 Support Vector Machine.....	17
4.3.1 Iperpiani.....	18
4.3.2 Ottimizzazione del margine	20
4.3.3 Valutazione del classificatore.....	23
4.3.4 Matrice di confusione o errore	23
4.3.5 ROC e AUC	25
4.3.6 La convalida incrociata	26
4.4 L'Algoritmo Fuzzy C-Means (FCM) e le classificazioni dei "Turisti in entrata"	27
4.5 La logistica della Funzione Gompertz: dalla biologia uno strumento di analisi delle dinamiche di una popolazione.....	30
Bibliografia	32

1. Introduzione

La scienza dei dati (nota come *Data Science - DS*) ha offerto significative opportunità a molti settori e il turismo non fa eccezione. Sebbene il turismo sia noto come un campo interdisciplinare, che abbraccia la sociologia, l'economia, la geografia, la psicologia e le scienze della comunicazione, i ricercatori del turismo sono stati a lungo vincolati dalle metodologie di ricerca classiche (Egger & Luger, 2015). Oltre agli approcci quantitativi e qualitativi ampiamente applicati, con il passare del tempo, è stato possibile osservare progressi, specialmente nei metodi quantitativi.

Al giorno d'oggi, in questa era di digitalizzazione, i dati si presentano in nuove forme non strutturate e questo, insieme ai set di dati strutturati tradizionalmente, ha portato all'ascesa dei Big Data. Nel frattempo, i progressi nell'informatica e il rapido sviluppo degli algoritmi hanno portato all'emergere di analisi avanzate, che vanno oltre la *Business Intelligence* convenzionale per ottenere informazioni più approfondite e fare previsioni future. Tutto sommato, la scienza dei dati è molto più di una pletora di nuovi metodi e strumenti che possono elevare i modi tipici di fare ricerca empirica e consentire ai ricercatori di trovare risposte a domande precedentemente sconosciute; piuttosto, è un'area ricca e interdisciplinare, con l'obiettivo di ottenere approfondimenti relativi alle attività commerciali sui dati e contenente un insieme unico di metodi che possono, in futuro, essere sicuramente utilizzati in modo sufficiente nel turismo. Tuttavia, la *Data Science* deve ancora essere abbracciata dagli studiosi del turismo al massimo delle sue potenzialità, in parte a causa della vastità, del disordine e della natura non strutturata dei dati che alimentano confusione e incertezza. Come già detto in precedenza, il turismo è un campo di ricerca interdisciplinare in cui confluiscono a vario titolo sociologi, geografi, economisti, scienziati della comunicazione, psicologi, informatici, ecc. Ognuna di queste discipline porta sul tavolo i propri metodi, fornendo alla ricerca turistica un aspetto ricco di opzioni metodologiche per osservare e analizzare un fenomeno da un'ampia varietà di prospettive (Egger & Yu, 2022b). Proprio come i metodi classici della ricerca sociale empirica hanno gradualmente professionalizzato la ricerca turistica, è probabile che i metodi della Data Science vengano sempre più applicati per comprendere ulteriormente e risolvere i problemi del turismo. Ogni metodo ha i suoi punti di forza e di debolezza individuali ed è predestinato ad aggiungere valore a problemi. Aprendo questo scrigno di metodi, il turismo può scoprire schemi e correlazioni prima sconosciuti, analizzare in tempo reale e approfondire anche le conoscenze rivelate per comprendere e spiegare meglio fenomeni, sistemi, processi, strutture e comportamenti, per citarne solo alcuni. Soprattutto, il potere predittivo di alcuni approcci particolari rende possibile rispondere a domande orientate al futuro (Weihs & Ickstadt, 2018). Esempi di applicazioni della Data Science nel turismo includono ottimizzazione del percorso, analisi predittiva e previsione, personalizzazione e raccomandazione, *opinion mining* e analisi del *sentiment*, sistemi di allerta e monitoraggio, e molto altro (Egger, 2022a).

Ciò porta a una base migliore per il processo decisionale e una migliore pianificazione.

Oltre ai metodi, ai modelli e agli algoritmi esistenti, la qualità e la quantità dei dati sono di particolare importanza (Taleb *et al.*, 2018). Il turismo è un'industria ad alta intensità di informazioni (Buhalis & Amaranggana, 2015; Egger & Herdin Thomas, 2007) e, di conseguenza, è disponibile un'ampia gamma di diverse fonti di dati e formati di dati. Poiché noi umani siamo tracciati sia nel mondo reale che in quello fisico, ogni individuo produce enormi quantità di dati al giorno senza saperlo esplicitamente. I diversi sensori misurano e forniscono continuamente dati e, attraverso la nostra partecipazione attiva ai *social network*, contribuiamo alla crescita dei contenuti generati dagli utenti (Wise & Heidari, 2019).

Pertanto, i dati provenienti da sistemi e processi vengono registrati ininterrottamente e rilasciati per l'analisi in un'ampia varietà di formati. Sono disponibili come valori numerici, come testo, immagini o video, hanno riferimenti geografici o temporali o descrivono, come metadati, altri dati.

Inoltre, il turismo rappresenta un settore importante dell'economia in Italia e nel mondo, con un forte potenziale in termini di crescita e di occupazione nonché di integrazione sociale e culturale. Gli effetti della recente pandemia Covid-19 hanno reso particolarmente evidente la rilevanza di questo settore: in Italia l'improvvisa e drastica contrazione dei flussi turistici ha avuto significativi impatti sul PIL nazionale e conseguenze serie sulle imprese del settore e del suo indotto.

La rappresentazione statistica del turismo è particolarmente impegnativa, in quanto concerne la misurazione di un fenomeno che non riguarda un unico comparto ma un insieme di servizi anche fortemente eterogenei. Il turismo può essere analizzato sia dalla prospettiva della domanda sia da quella dell'offerta ed è, quindi, misurato attraverso diverse fonti informative. In generale, le statistiche che guardano al turismo dal lato dell'offerta si riferiscono alla capacità e all'occupazione delle strutture ricettive mentre quelle che misurano il fenomeno turistico dal lato della domanda si basano su informazioni raccolte direttamente dai viaggiatori, alle frontiere o mediante tradizionali indagini presso le famiglie. In Italia, nell'ambito del sistema delle statistiche ufficiali, sono condotte diverse rilevazioni sui fenomeni turistici, sia da parte dell'Istat sia della Banca d'Italia, ciascuna all'interno del proprio quadro normativo di riferimento: dal lato dell'offerta, l'Istat conduce l'indagine "Movimento dei clienti negli esercizi ricettivi"; dal lato della domanda, sempre l'Istat cura, all'interno dell'Indagine sulle spese delle famiglie, un focus su "Viaggi e vacanze", mentre la Banca d'Italia conduce un'"Indagine alle frontiere sul turismo internazionale dell'Italia".

Le indagini coprono dunque il fenomeno turistico secondo diverse prospettive, con alcune sovrapposizioni: il turismo interno all'Italia, effettuato sia da residenti (domestico) sia da stranieri (inbound); il turismo nazionale, sia domestico sia di viaggiatori italiani all'estero (outbound); il turismo internazionale, sia inbound sia outbound.

2. Intelligenza Artificiale, Machine Learning e Big Data nel Turismo

Il turismo è un settore complesso e multidimensionale, e tutte le sue caratteristiche contribuiscono a renderlo uno spazio ideale per l'applicazione di sistemi di Intelligenza Artificiale che sfruttano grandi *dataset*. In primo luogo, a causa dell'elevato numero di persone, processi e attività coinvolti, i dati sul turismo sono spesso big data. Inoltre, non sembra esserci alcun limite quando si tratta di tipi di fonti di dati e tecnologie di Intelligenza Artificiale che potrebbero essere rese utili per indagare e migliorare sia l'offerta che la domanda di turismo, comprese le strategie di marketing e la gestione dell'*overtourism*. Queste tecnologie potrebbero persino facilitare il recupero dai periodi di crisi e molto altro. Di conseguenza, nonostante l'impossibilità di una rassegna di tutte le esperienze e progetti significativi in questo campo, alcune linee guida possono essere dati:

- I progetti di Intelligenza Artificiale e big data possono trarre vantaggio da fonti di dati "esterne", cioè non solo quelle tradizionalmente considerate nelle applicazioni turistiche (Li *et al.*, 2018). Ad esempio, oltre ai contenuti generati dagli utenti, le previsioni meteorologiche e

i dati sui cambiamenti climatici potrebbero essere utilizzati per affrontare il problema del turismo stagionale.

- Strategie turistiche innovative possono essere progettate solo con nuove visioni e idee. Le tecniche di creatività possono successivamente essere applicate considerando le intuizioni ottenute dall'analisi di grandi set di dati. In alternativa, potrebbero essere utilizzati nuovi modi per visualizzarli; tra i principi più recenti dell'economia circolare, lo sviluppo sostenibile e l'economia ambientale possono ispirare governi e gestori di destinazioni, enti turistici locali e stakeholder del turismo in generale.
- Per quanto riguarda le numerose tecnologie di raccolta dati, quelle che identificano la posizione delle persone, o ugualmente, qualsiasi tipo di fenomeno sono molto rilevanti per la geografia del turismo. Nuove tecnologie e tecniche di (*action*) *tracking*, tra cui GPS, dispositivi mobili, sensori e dati dei droni, possono essere utilizzate per gestire i problemi di *overtourism*, ad esempio per l'ottimizzazione del percorso, la gestione del traffico, visite guidate personalizzate, ecc. (Mich, 2020b).
- L'Intelligenza Artificiale e i big data potrebbero anche contribuire a ridurre la frammentazione delle app turistiche (mobili) poiché, spesso, molte app sono troppo specializzate e non sufficientemente personalizzate. Uno dei motivi di ciò è che ognuno si basa su un singolo set di dati e non è progettato per soddisfare le reali esigenze delle parti interessate. I progressi dell'Intelligenza Artificiale devono essere costantemente monitorati per migliorare ulteriormente qualsiasi applicazione quasi classica dei big data. Ad esempio, i sistemi che analizzano i contenuti generati dagli utenti per le raccomandazioni dei clienti o il monitoraggio della reputazione web possono trarre vantaggio da alcuni dei progressi della *Natural Language Processing* sfruttando anche le tecniche per la classificazione delle immagini, il riconoscimento facciale, l'analisi video e l'identificazione vocale, in aggiunta al riconoscimento delle emozioni più recentemente sviluppato.
- Dato il ruolo centrale della comunicazione nel turismo, oltre all'analisi dei contenuti, è possibile applicare i cosiddetti sistemi generativi basati sull'Intelligenza Artificiale per supportare la generazione di contenuti. Testi, audio e immagini creati da questi sistemi non sono facilmente riconoscibili come prodotti automaticamente (e richiedono nuove soluzioni per diritto d'autore). Per quanto riguarda la qualità dei contenuti, l'Intelligenza Artificiale può anche utilizzare i set di dati per eseguire il controllo dei fatti e il rilevamento dei *deepfake*.
- Infine, le nuove e innovative applicazioni dell'Intelligenza Artificiale e big data possono sfruttare anche i tradizionali documenti (analogici) di qualsiasi tipo. Esistono grandi quantità di documenti che, una volta digitalizzati, potrebbero essere integrati con i set di dati esistenti, aggiungendo preziose conoscenze alle destinazioni turistiche, ad esempio per quanto riguarda il loro patrimonio culturale e storico. Anche le iniziative di *crowdsourcing*, in questo senso, potrebbero aiutare a costruire comunità, che sono fondamentali per il turismo.

Dunque, l'avvento dell'Intelligenza Artificiale ha notevolmente cambiato la prospettiva verso le applicazioni della potenza computazionale nella vita quotidiana. Grazie alla solida collaborazione di Statistica, Informatica e Ingegneria, le macchine possono ora adattarsi all'ambiente in evoluzione e ottimizzare facilmente i problemi che erano precedentemente non computazionali. I frequenti sviluppi nel campo della *data mining* e dei Big Data hanno ulteriormente dato impulso alle applicazioni della potenza di calcolo in campi più diversi. Pertanto, la tecnologia in continua evoluzione ha portato a

innovazioni prodigiose nella società (Bhilaré *et al.*, 2019). Infatti, la tecnica del *pre-processing* (pre-elaborazione) dei dati nel Machine Learning è un passaggio di fondamentale importanza che aiuta a migliorare la qualità dei dati. In particolare, si riferisce alla tecnica di preparazione, organizzazione e pulizia dei dati grezzi per renderli adatti alla costruzione e all'adattamento dei modelli. Pertanto, questo procedimento, viene definito come una tecnica di “*data mining*” che trasforma i dati originali in un formato leggibile e comprensibile dal modello. La pre-elaborazione dei dati, infatti, è il primo passaggio che segna l'inizio del processo di costruzione di un modello di Machine Learning. Solitamente, i dati del mondo reale sono imprecisi, incompleti, incoerenti (contengono valori anomali o errori) e spesso presentano valori nulli in riferimento ad attributi specifici. È qui che il *preprocessing* dei dati entra in gioco: pulisce, formatta e organizza i dati grezzi, rendendoli così pronti all'uso per i modelli di Machine Learning.

L'apprendimento automatico (Machine Learning) è senza dubbio uno degli sviluppi tecnologici più significativi e di vasta portata che plasmano attualmente i nostri tempi (Jamal *et al.*, 2018). Queste tecnologie possono essere trovate in tutte le aree della nostra vita, fornendoci informazioni e conoscenze derivate dai dati, anche se in modo poco appariscente. Fungono da spina dorsale degli assistenti vocali, come Siri, Cortana, Bixby o Alexa, sono i capisaldi dei chatbot, supportano il marketing personalizzato e prevedono il comportamento dei clienti. Ottimizzano i processi di ogni tipo, filtrano e classificano lo spam dalle nostre e-mail, costituiscono la base della prevenzione delle frodi e forniscono anche la base per i controlli antiplagio. Inoltre, le tecnologie Machine Learning vengono costantemente impiegate nell'uso delle piattaforme di social media senza che gli utenti se ne accorgano o ne siano consapevoli. Il Machine Learning fornisce l'infrastruttura centrale per l'Intelligenza Artificiale e spesso pone le basi per progetti di data science, collegando informatica e statistica al fine di sviluppare algoritmi e teorie basate su modelli statistici (Althbiti & Ma, 2020) con l'obiettivo di un'elevata capacità predittiva (Jordan & Mitchell, 2015). Sebbene gli inizi del Machine Learning risalgono agli anni '50, il suo potenziale è diventato evidente solo di recente poiché la quantità di dati oggi disponibili ha raggiunto proporzioni enormi. Il Machine Learning viene spesso citato nel contesto dei "Big Data" (Kelm *et al.*, 2020), dove è vero che la sua influenza aumenta parallelamente alla quantità di dati disponibili. Allo stesso tempo, però, va detto - e questo è particolarmente rilevante per i progetti di ricerca scientifica - che non sempre è necessario essere in possesso di diversi gigabyte di dati per poter applicare con successo il Machine Learning; anche set di dati più piccoli contenenti diverse centinaia o migliaia di istanze potrebbero essere sufficienti per utilizzare approcci di Machine Learning per identificare automaticamente modelli in dati complessi. La conoscenza di questi modelli può quindi aiutare a prevedere eventi futuri e prendere decisioni complesse con un certo grado di sicurezza. Numerosi sono stati i tentativi di definire il termine Machine Learning, come quello di Arthur Samuel, che per primo coniò il termine “Machine Learning” nel 1959, descrivendolo come “un campo di studio che dà ai computer la capacità di apprendere senza essere esplicitamente programmati”. Allo stesso modo, Akerkar (2019b) definisce il Machine Learning come "metodi computazionali che utilizzano l'esperienza per migliorare le prestazioni o per fare previsioni accurate. Il termine "esperienza" è usato qui per riferirsi a database esistenti e le loro proprietà (i dati di addestramento), che vengono utilizzate per apprendere e addestrare un modello (Mohri *et al.*, 2018). L'obiettivo è identificare modelli nei dati che consentano di descrivere meglio i dati, aumentare le prestazioni o eseguire la previsione più accurata possibile (Jamal *et al.*, 2018). Come ben sappiamo, i set di dati sono un insieme di esempi che contengono funzionalità per risolvere un problema. Se pensiamo ai dati in forma tabellare, ogni riga è un'istanza e ogni colonna è una caratteristica. Le *features*

sono pezzi di dati misurabili che vengono inseriti in un algoritmo di apprendimento automatico e aiutano a comprendere il problema. Il risultato è un modello, da intendersi come la rappresentazione addestrata di ciò che l'algoritmo ha appreso. Ad esempio, un algoritmo di *Random Forest* può essere addestrato con dati di addestramento e l'output è un modello di *Random Forest*. Ora è possibile inserire nuovi dati sconosciuti nel modello per ottenere previsioni, classificare i dati e molto altro, a seconda dell'algoritmo utilizzato. Pertanto, un algoritmo predittivo crea un modello predittivo che, se alimentato con nuovi dati, produce una previsione basata sui dati su cui è stato addestrato (Kelm *et al.*, 2020). Dunque, gli algoritmi di Machine Learning costruiscono modelli predittivi in grado di mappare le caratteristiche dell'utente a una classe (o decisione) grazie a una fase di apprendimento (Monreale, 2020). I modelli decisionali sfruttano spesso algoritmi molto sofisticati che permettono di ottenere delle ottime prestazioni in termini di accuratezza dei risultati; purtroppo le ottime prestazioni spesso sono ottenute con modelli molto complessi, come *Neural Network* e *Deep Neural Network*, che sono di difficile interpretazione anche per gli esperti del settore (Monreale, 2020). L'Intelligenza Artificiale, i big data e il Machine Learning sono spesso citati in relazione ai termini attualmente popolari di "turismo intelligente" e "destinazioni intelligenti". Secondo Gretzel, Sigala, Xiang & Koo (2015), ciò che presumibilmente rende le destinazioni, le città e il turismo "intelligenti", in generale, include le varie tecnologie dell'informazione e della comunicazione (ICT) integrate nell'infrastruttura fisica, esperienze intelligenti che tentano di ottimizzare le esperienze di viaggio attraverso la personalizzazione, la contestualizzazione e l'analisi in tempo reale (Buhalis & Amaranggana, 2015) nonché l'ecosistema aziendale orientato all'intelligenza. La *smartness* richiede, quindi, l'elaborazione di big data, disponibili come dati transazionali, contenuti generati dagli utenti, dati forniti da dispositivi integrati e misurati da sensori, ecc. (Gretzel *et al.*, 2015; Koo *et al.*, 2016). Tutti i tipi di dati, ad esempio immagini, audio, video, testo e metadati come valori di data e ora, dati geospaziali, tag e altro, sono di grande rilevanza. Per informazioni dettagliate sull'analisi nella progettazione del turismo intelligente, si consiglia di consultare Xiang & Fesenmaier (2017).

3. Stato dell'arte

Negli ultimi decenni, l'industria del turismo è diventata un settore importante dell'economia di un Paese. La domanda turistica, ha mostrato una crescita sostenuta. Questo mercato, però, ha subito diverse oscillazioni dovute alla volatilità dei fattori determinanti e degli interventi esterni. Ricercatori, professionisti e responsabili politici hanno prestato grande attenzione ai cicli di crescita del turismo e agli andamenti della domanda, cercando di prevedere i flussi futuri di turisti. Nella maggior parte degli studi esistenti sulla previsione della domanda turistica, l'attenzione dei ricercatori si è concentrata principalmente sui flussi turistici internazionali a causa del fatto che il turismo internazionale è statisticamente meglio richiamato rispetto al turismo interno, e solo pochi studi si sono concentrati sul turismo interno (Athanasopoulos & Hyndman, 2008; Blunk, Clark & McGibany, 2006), domanda di camere d'albergo (Pan, Wu, & Song, 2012; Yang, Pan, & Song, 2014), o domanda di siti ricreativi (Chen, Bloomfield, & Fu, 2003; Ellis & Doren, 1966). La domanda turistica internazionale è misurata in termini di arrivi turistici, spesa turistica o durata del soggiorno. Questi fattori sono generalmente analizzati utilizzando dati aggregati piuttosto che disaggregati. Questi dati sono correlati a diversi tipi di volatilità, come la stagionalità delle regioni di origine e di destinazione, i cicli economici associati ai tassi di cambio e ai livelli di reddito, o varie esternalità legate al cambiamento climatico o a eventi

speciali. Con la rapida espansione del turismo internazionale dovuta ai cambiamenti sociali, economici, politici e tecnologici, il turismo si sta diffondendo dai paesi sviluppati ai paesi di recente industrializzazione. Questo sviluppo può dar luogo a una combinazione di costi e benefici, poiché più paesi/regioni di destinazione turistica competono per le scarse risorse (Lim, 2006). Pertanto, previsioni accurate sono fondamentali per le destinazioni in cui i decisori cercano di capitalizzare gli sviluppi nel mercato del turismo e/o di bilanciare le loro capacità di carico ecologico e sociale. In questi contesti, i previsori della domanda turistica internazionale hanno provato a riflettere le condizioni generali dei mercati di origine, delle destinazioni e persino dei paesi/regioni vicini o concorrenti che possono influenzare i loro flussi turistici (Fotheringham, 1983; Song, Li & Cao, 2017). Data l'importanza di previsioni accurate per il dinamico e complesso mercato del turismo, negli ultimi decenni sono stati pubblicati oltre 600 studi sulla modellazione e la previsione della domanda turistica. Questi studi si sono concentrati principalmente sulla costruzione del modello e sulla valutazione delle prestazioni. Alcuni di loro, tuttavia, hanno proposto nuovi modelli ibridi o utilizzato varie combinazioni di metodi. La letteratura sulla previsione della domanda turistica, come si evince in prima battuta, è ampia (Li *et al.*, 2021). La maggior parte degli studi si è concentrata sulla previsione del flusso turistico internazionale utilizzando vari metodi quantitativi (Song *et al.*, 2019), comprese le serie temporali (Ma *et al.*, 2016; Park *et al.*, 2017), econometriche (Padhi & Pati, 2017) e Intelligenza Artificiale (IA) (Lv *et al.*, 2018; Sun *et al.*, 2019). In questa era dei big data, gli approcci basati sull'Intelligenza Artificiale hanno aumentato la popolarità (Song *et al.*, 2019) e sono stati ampiamente utilizzati per la previsione della domanda turistica grazie alla loro capacità di gestire dati non lineari (Law *et al.*, 2019; Sun *et al.*, 2019; Huang & Hao, 2020). La rete neurale artificiale (ANN), la Support Vector Regression (SVR) e la Random Forest (RF) sono tra i modelli basati sull'Intelligenza Artificiale più utilizzati (Sun *et al.*, 2019; Song *et al.*, 2019; Abellana *et al.*, 2020; Huang & Hao, 2020; Li *et al.*, 2020).

Pertanto, i metodi quantitativi esistenti per la previsione del turismo possono essere classificati in tre categorie: serie temporali, econometrica e Intelligenza Artificiale. A differenza dei modelli econometrici, i modelli basati sull'Intelligenza Artificiale possono descrivere dati non lineari senza una previa comprensione delle correlazioni tra variabili di input e output (Song *et al.*, 2019).

I modelli di serie temporali tradizionali includono modelli Naïve 1, modelli Naïve 2, modelli di livellamento esponenziale e modelli AR semplici (Song & Li, 2008; Wu, Song & Shen, 2017). Sono spesso usati come *benchmark* negli studi di previsione del turismo. I modelli a media mobile integrata autoregressiva (ARIMA) e i modelli SARIMA sono i modelli più comunemente utilizzati, a seconda della frequenza delle serie temporali. In letteratura sono state utilizzate anche varie estensioni del modello ARIMA.

Ad esempio, Chu (2009) ha introdotto un modello ARMA autoregressivo (ARARMA) e un modello ARMA frazionario integrato (ARFIMA) per prevedere gli arrivi di turisti in nove destinazioni nella regione Asia-Pacifico e ha scoperto che il modello ARFIMA ha superato i modelli SARIMA e ARARMA. Allo stesso modo, Assaf, Barros & Gil-Alana (2011) hanno utilizzato diversi modelli basati sull'integrazione frazionaria per prevedere gli arrivi di turisti in Australia, confermando che hanno sovraperformato i modelli ARIMA e SARIMA standard. Anche le serie temporali strutturali (Turner & Witt, 2001) e i modelli eteroschedastici condizionali autoregressivi generalizzati (Divino & McAleer, 2010) sono stati ampiamente utilizzati nella letteratura turistica. Negli ultimi anni, sono stati utilizzati modelli di serie temporali più avanzati per generare migliori prestazioni di previsione rispetto ai modelli di serie temporali tradizionali, come i modelli di stato dello spazio delle innovazioni per il livellamento esponenziale (ETS; Athanasopoulos *et al.* 2011), i modelli di analisi dello spettro

singolare (SSA) (Hassani *et al.*, 2017) e modelli di serie temporali strutturali dei parametri variabili nel tempo (Song *et al.* 2011).

I metodi di decomposizione, come SSA, *empirical mode decomposition* (Yahya, Samsudin & Shabri 2017) e *ensemble empirical mode decomposition* (Zhang *et al.*, 2017), hanno guadagnato molta popolarità negli ultimi anni e hanno dimostrato buone prestazioni di previsione. Queste tecniche sono state utilizzate in impostazioni di previsione di serie temporali univariate (Hassani *et al.*, 2017; Hassani *et al.*, 2015; Silva *et al.*, 2019) e impostazioni di previsione di serie temporali causali (Li & Law, 2020). A differenza dei modelli di serie temporali non causali, i modelli econometrici possono analizzare la relazione tra la domanda turistica e le sue determinanti chiave e le informazioni possono essere utilizzate per fornire raccomandazioni politiche. In letteratura sono stati identificati diversi fattori importanti che influenzano la domanda turistica, come il reddito del turista, i prezzi del turismo in una destinazione rispetto a quelli del paese di origine, i prezzi del turismo in destinazioni concorrenti e i tassi di cambio reali (Song & Li, 2008; Wu, Canzone & Shen, 2017). La regressione spuria è spesso presente nell'analisi di regressione tradizionale. Diversi modelli econometrici moderni sono stati introdotti nella modellazione e previsione del turismo, come il modello a ritardo distribuito autoregressivo (Song, Gartner & Tasci, 2012), il modello di correzione degli errori (Goh, 2012), il modello vettoriale autoregressivo (VAR) (Wong, Song & Chon, 2006), il modello a parametri variabili nel tempo (Page, Song & Wu, 2012), il modello del sistema di domanda quasi ideale (Li, Song & Witt 2006) e il modello VAR bayesiano (Gunter & Önder, 2015 ; Wong, Canzone & Chon, 2006). Numerosi studi hanno concluso che i modelli econometrici hanno prestazioni migliori (Song, Witt & Jensen, 2003), ma alcuni hanno confermato che i modelli di serie temporali superano i modelli econometrici nella previsione della domanda turistica (Athanasopoulos *et al.*, 2011). Oltre alle serie temporali e ai metodi econometrici, nella letteratura sulle previsioni del turismo sono stati introdotti una varietà di metodi basati sull'Intelligenza Artificiale. Il modello dominante è il modello di rete neurale artificiale (ANN). Consiste di diversi strati, ognuno dei quali può contenere più neuroni. Il modello ANN è un metodo non parametrico e basato sui dati che può essere utilizzato per modellare relazioni non lineari. È anche il metodo basato sull'Intelligenza Artificiale più frequentemente utilizzato negli studi di previsione della domanda turistica (Claveria, Monte & Torra, 2015; Law *et al.*, 2019; Sun *et al.*, 2019). Altri metodi basati sull'Intelligenza Artificiale utilizzati per prevedere il turismo la domanda include il modello Support Vector Machine (Chen *et al.*, 2015; Hong *et al.*, 2011), il modello del sistema fuzzy (Aladag *et al.*, 2014), il modello rough set (Goh, Law & Mok, 2008) e la teoria dei fuzzy set (Sole *et al.*, 2016).

Sebbene in letteratura siano stati introdotti e applicati vari metodi, vi è consenso sul fatto che nessun modello può sovraperformare costantemente altri modelli in tutte le condizioni (Song & Li, 2008). Utilizzando una meta-analisi, Peng, Song & Crouch (2014) hanno mostrato che le loro caratteristiche dei dati e le caratteristiche dello studio, come la misura della domanda, la frequenza dei dati e le coppie origine-destinazione, hanno influenzato l'accuratezza della previsione della domanda turistica.

Pianificatori turistici, responsabili politici, gestori di hotel, rivenditori e agenzie governative si affidano a previsioni accurate per sviluppare strategie, piani aziendali a lungo termine e gestire la capacità delle infrastrutture e il numero di visitatori a breve termine (Pan & Yang, 2017; Song, Qiu, & Park, 2019; Wan & Song, 2018). Nel 2019, le previsioni sulla domanda turistica globale hanno fruttato oltre 2750,7 miliardi di dollari, generando il 3,2% del PIL globale (WTTC, 2019). La previsione dei volumi della domanda turistica e la formulazione di ipotesi affidabili sulla crescita economica e sugli impatti sociali e ambientali richiedono modelli e metodologie di previsione della domanda accurati e robusti.

Nonostante i numerosi contributi alla letteratura, la ricerca sulla modellazione della domanda turistica rimane “una delle più interessanti e importanti aree di ricerca negli studi sul turismo” (Silva, Hassani, Heravi, & Huang, 2019). L'elevata rilevanza per l'ulteriore avanzamento della metodologia di previsione del turismo risiede nell'affrontare la sfida chiave per migliorare l'accuratezza e omettere l'*overfitting* del modello sviluppato (ad esempio Cai, Lu & Zhang, 2009; Li, Wu, Zhou & Liu, 2019; Wu, Law & Xu, 2012). Assaf, Li, Song & Tsionas (2019) hanno, inoltre, sottolineato che la nuova ricerca sulla previsione della domanda turistica dovrebbe concentrarsi sul miglioramento dell'accuratezza delle fonti di dati e sul miglioramento delle tecniche di elaborazione per generare una maggiore prevedibilità del modello.

Zhang *et al.* (2020) nel loro lavoro cercano di colmare queste lacune nella letteratura turistica esistente utilizzando innovative tecniche di Deep Learning basate sul *pooling* per ottenere una maggiore precisione e migliorare l'efficacia del modello di previsione. I metodi di Deep Learning sono modelli avanzati di reti neurali artificiali che utilizzano architetture di rete con un numero elevato di livelli di elaborazione interconnessi (si rimanda al paragrafo 4.2 per ulteriori approfondimenti). Ad esempio, Law, Li, Fong & Han (2019) hanno utilizzato il Deep Learning con un meccanismo di attenzione per costruire una rete neurale su larga scala. Tuttavia, il problema con le metodologie di Deep Learning esistenti rimane che questi modelli avanzati spesso affrontano il problema dell'*overfitting* (Shi, Xu & Li, 2017). Un problema chiave alla base che causa un modello sovradimensionato è in genere la disponibilità limitata dei dati (Hawkins, 2004; Tetko, Livingstone & Luik, 1995), che può essere risolta disponendo di un campione di dati più ampio o di una diversità di dati (Goodfellow, Bengio & Courville, 2016).

Essendo una delle aree di ricerca importanti nella gestione del turismo, la previsione del volume del turismo ha attirato molta attenzione sia degli accademici che dei professionisti (Song & Hyndman, 2011; Song & Li, 2008). La previsione accurata del volume del turismo è una parte importante delle operazioni commerciali efficienti e della gestione della destinazione (Palmer, Montano & Sesé, 2006). Finora, sono stati proposti modelli sulle previsioni del volume del turismo (Song, Qiu & Park, 2019). Dal punto di vista della granularità delle previsioni, la maggior parte di questi modelli si concentra sulla previsione a lungo termine (ad esempio mensile, trimestrale e annuale) di aree relativamente grandi (ad esempio province, paesi e regioni). I risultati delle previsioni di questi studi possono fornire un riferimento per lo sviluppo di politiche macro-economiche. Tuttavia, per un'area più piccola (ad esempio un'attrazione turistica), la previsione a breve termine (ad esempio giornaliera) e ad alta frequenza è più importante (Divino & McAleer, 2010; Pan & Yang, 2017). Sulla base dei risultati di previsione a breve termine (ad esempio giornalieri), da un lato, i gestori del turismo possono formulare pacchetti turistici o strategie di prezzo per aumentare gli arrivi di turisti in periodi di bassa domanda (Divino & McAleer, 2010); d'altra parte, i gestori del turismo possono prendere accordi per il personale e sviluppare piani di emergenza per prevenire il verificarsi di detenzione turistica in periodi di forte domanda (Li, Chen, Wang & Ming, 2018). Tuttavia, gli studi sulla previsione del volume a breve termine di un'area relativamente piccola sono molto rari.

Bi *et al.* (2020) nel loro lavoro si concentrano principalmente sulla previsione del volume turistico giornaliero per le attrazioni turistiche proponendo un nuovo approccio basato su reti di memoria a lungo termine (Long Short Term Memory - LSTM) in grado di incorporare dati di serie temporali multivariate, inclusi dati storici sul volume del turismo, dati dei motori di ricerca e dati meteorologici. Gli autori applicano tale approccio alla previsione del volume turistico giornaliero di Jiuzhaigou e HMA, due famose attrazioni turistiche in Cina.

Il settore turistico, con la sua natura deperibile dei prodotti, richiede una stima precisa della domanda. A tal fine, Kulshrestha, Krishnaswamy & Sharma, (2020) propongono una metodologia di deep learning, ovvero la rete Bayesian Bidirectional Long Short-Term Memory (BBiLSTM). BiLSTM è un modello di Deep Learning e l'ottimizzazione bayesiana viene utilizzata per ottimizzare gli iperparametri di questo modello. Gli autori hanno condotto cinque esperimenti utilizzando i dati sulla domanda turistica di Singapore per accertare la validità e confrontare il modello BBiLSTM proposto. I risultati sperimentali suggeriscono che il modello BBiLSTM supera altri modelli concorrenti come la rete Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Radial Basis Function Neural Network (RBFNN) e Autoregressive Distributed Lag Model (ADLM). Lo studio contribuisce alla letteratura turistica proponendo un metodo di Deep Learning superiore per la previsione della domanda.

Come accennato in precedenza, con il rapido sviluppo dell'industria del turismo, la previsione della domanda turistica è diventata sempre più importante per i governi e affari. Secondo l'Organizzazione mondiale del turismo delle Nazioni Unite (UNWTO) (2019), la crescita del turismo internazionale continua a superare l'economia globale, il che rende il settore del turismo una forza trainante globale per la crescita e lo sviluppo economico. La previsione della domanda turistica ha quindi attirato una crescente attenzione negli ultimi decenni, soprattutto alla luce della natura deperibile dei servizi turistici (Li, Song & Witt, 2005). Previsioni più accurate consentirebbero alle parti interessate di pianificare in anticipo e allocare le risorse in modo più efficiente e le aziende potrebbero adeguare le proprie strategie per migliorare le proprie prestazioni. Per migliorare l'accuratezza delle previsioni, sono stati continuamente sviluppati nuovi metodi nel campo della previsione della domanda turistica. Infatti come già detto, i modelli di serie temporali e i modelli econometrici sono utilizzati più di frequente e i modelli di Intelligenza Artificiale hanno iniziato a guadagnare popolarità nell'ultimo decennio (Jiao & Chen, 2019). Le revisioni dei modelli di previsione della domanda turistica hanno identificato una serie di nuove tendenze, compreso l'uso di modelli spaziali (Wu, Song & Shen, 2017). L'econometria spaziale, che riguarda le interazioni spaziali delle unità economiche (Lee & Yu, 2010b), è recentemente diventata un'area di ricerca popolare ed è stata applicata in molti diversi campi di studio. Poiché le località e le destinazioni sono molto rilevanti per l'industria del turismo, l'analisi spaziale è stata applicata in alcuni studi sul turismo per tenere conto dell'effetto di spillover spaziale tra le destinazioni o tra le origini e le destinazioni (Yang & Wong, 2012). Ad esempio, dati gli accordi sui visti Schengen nell'Unione Europea, i turisti a lungo raggio in particolare tendono a visitare più destinazioni durante un unico viaggio in Europa. Precedenti studi sul turismo (ad esempio, Romão, Guerreiro & Rodrigues, 2017; Romão & Nijkamp, 2019) hanno confermato l'esistenza di una significativa dipendenza spaziale tra i paesi europei. Jiao, Li & Chen (2020) tentano di esaminare se l'incorporazione sia della dipendenza spaziale che dell'eterogeneità spaziale può migliorare le prestazioni di previsione in 37 paesi europei. Inoltre, hanno confrontato le prestazioni di previsione con quelle di modelli di riferimento tra cui la media mobile autoregressiva, il livellamento esponenziale e i modelli Naïve. Per tutti gli orizzonti di previsione, i due modelli spaziali hanno sovraperformato i modelli non spaziali.

Dunque, le prestazioni di previsione superiori del modello locale suggeriscono che il pieno riflesso dell'eterogeneità spaziale può migliorare l'accuratezza delle previsioni turistiche.

Sebbene gli effetti spaziali sulla crescita del turismo e sui flussi turistici siano stati analizzati in molti studi, solo due hanno incorporato l'effetto di spillover spaziale nella previsione della domanda turistica. Long, Liu & Song (2019) e Yang & Zhang (2019) hanno utilizzato modelli spazio-temporali per

prevedere la domanda turistica interna in Cina e hanno confrontato le prestazioni di previsione con altri modelli di riferimento: Naïve 1, Naïve 2, minimi quadrati ordinari (OLS), autoregressivo integrato media mobile (ARIMA) e altri modelli. Il modello spazio-temporale autoregressivo spaziale (SAR) utilizzato in questi studi ha il vantaggio di incorporare ritardi sia spaziali che temporali in un modello di dati panel. A seguito di questi primi tentativi di introdurre modelli spazio-temporali nelle previsioni turistiche, nel lavoro di Jiao, Li & Chen (2020) mirano a sviluppare ulteriormente il modello spazio-temporale ed esaminare il miglioramento dell'accuratezza nelle previsioni turistiche.

Zheng, Huang & Lin (2021) nel loro contributo sviluppano un modello basato sull'Intelligenza Artificiale, noto come memoria a lungo termine orientata alle serie temporali correlate con meccanismo di attenzione (*Attention Mechanism*), per risolvere questo problema. I risultati mostrano che il modello supera significativamente i modelli di riferimento. Lo studio fa avanzare la letteratura sulla previsione della domanda turistica e offre implicazioni pratiche per l'ottimizzazione delle risorse, migliorando al contempo la soddisfazione del personale e dei clienti.

Come si evince, dallo studio dello stato dell'arte, ci sono diversi modelli e algoritmi predittivi che possono essere utilizzati per effettuare delle previsioni e dunque, adattabili alle casistiche rientranti nel progetto MONTUR.

4. Individuazione dei principali tools, metodi e modelli utilizzabili per la realizzazione del sistema previsionale integrato

In questa sezione offriamo una panoramica ragionata dei principali strumenti e metodi pensati in queste prime fasi del progetto di ricerca e dai quali partire per giungere alla successiva elaborazione del modello previsionale finale, al quale ad oggi non si è giunti, a causa dell'interruzione anticipata dell'assegno e soprattutto per la mancata acquisizione dei Big Data sui quali effettuare lo screening iniziale.

4.1 Machine Learning: training supervisionato e non supervisionato

I metodi e gli strumenti di Machine Learning si dividono in Supervised (supervisionati) e unsupervised (non supervisionati). Nel primo caso, i dati attraverso gli algoritmi vengono organizzati ed istruiti essendo già contrassegnati con degli output validi, o in ogni caso è tangibile la possibilità di addivenire ad una valutazione dei risultati al fine di validare o modificare il modello. Attraverso gli algoritmi non supervisionati al contrario ai dati non viene assegnata un'etichetta arrecante una risposta corretta; infatti è la macchina a dover ricavare dai dati tutte le informazioni che si rendono necessarie, riuscendo così a selezionare *pattern* di cui si disconosceva l'esistenza *ex ante*. Le casistiche di Machine Learning supervisionato si possono anche dicotomizzare in problemi di classificazione (previsione di risposte non numeriche) e di regressione (previsione di risposte numeriche). Nel Machine Learning non supervisionato abbiamo problemi di clustering o raggruppamento (ricerca di gruppi di oggetti simili), associazione (ricerca di sequenze comuni di oggetti) e di riduzione della dimensionalità (o *dimensionality reduction*, cioè la proiezione, selezione ed estrazione delle caratteristiche chiave di un

modello). Obiettivo del Machine Learning è attraverso l'utilizzazione di una piattaforma di algoritmi trasformare un set di dati in un modello operativo.

Quale sia l'algoritmo che si adatti meglio alla risoluzione del problema che si sta affrontando dipende dalle risorse informatiche disponibili e dalla natura dei dati a nostra disposizione.

Indipendentemente dall'algoritmo o dagli algoritmi utilizzati, è prima necessario ripulire e preparare i dati come detto in precedenza attraverso la fase del pre-processing.

Algoritmi di classificazione

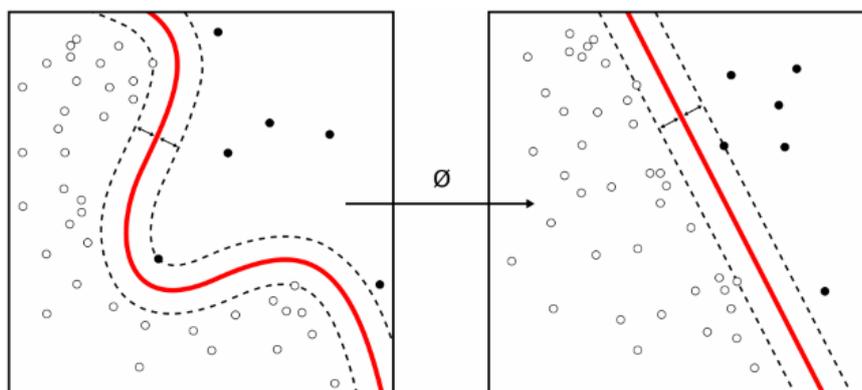


Figura 1: Un algoritmo di classificazione individua la funzione che meglio descrive le zone che separano i dati che presentano caratteristiche differenti tra loro. (Immagine di Alisneaky)

Un metodo di classificazione è una tipologia di problema di apprendimento supervisionato che richiede l'effettuazione di una scelta tra due o più classi da attribuire ai dati, in genere associando una probabilità ad ogni possibile classe. Mettendo *a latere* le reti neurali e il deep learning, per i quali si richiede un livello dispendioso di risorse informatiche, gli algoritmi più in uso sono il Naive Bayes Classifier³, l'Albero decisionale, la Regressione logistica⁴, K-Nearest Neighbours (K-

³ Naive Bayes è un algoritmo di classificazione probabilistico che utilizza il teorema della probabilità di Bayes o le regole di Bayes. L'algoritmo di Naive Bayes è chiamato "ingenuo" perché presuppone che l'occorrenza di una determinata caratteristica sia indipendente dall'occorrenza di altre caratteristiche ed è chiamato Bayes poiché utilizza la regola o la legge di Bayes per la classificazione. Questo algoritmo prevede le probabilità di appartenenza per ciascuna classe, ad esempio la probabilità che un determinato record o punto dati appartenga a una particolare classe. La classe con la probabilità più alta è considerata la classe più probabile. Questo è anche noto come Maximum A Posteriori (MAP).

⁴ L'algoritmo di Regressione Logistica può essere utilizzato sia per la classificazione che per la regressione. La regressione logistica è simile alla regressione lineare con il punteggio logistico. Si utilizzano funzionalità indipendenti dei dati per prevedere o classificare il valore target dipendente. Questo algoritmo è simile all'algoritmo neurale del perceptron in cui aggiorniamo i pesi per ottenere l'output previsto

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Qui x_1, x_2, \dots, x_n (variabile indipendente) sono le caratteristiche dei dati e $b_0, b_1, b_2, \dots, b_n$ sono i pesi che vengono aggiornati e y è l'output (variabile dipendente).

Questo strumento serve a valutare due aspetti fondamentali nella costruzione o ri-costruzione di un processo decisionale: Il legame esistente tra la variabile decisionale y e i dati (x_1, x_2, \dots, x_n) che aiutano i decision makers a decidere. Il grado di condizionamento tra la variabile indipendente x (l'informazione, diverse informazioni legate al caso in analisi, infatti indicate in un numero finito da 1 a n) e la variabile decisionale y , è misurato dal coefficiente b che può assumere valori tra 0 e 1. Se b è uguale a 0 l'informazione è inutile e annulla dal prodotto l'informazione che deve essere utilizzata dal decision maker se è uguale ad 1, l'informazione in oggetto conferma il peso dell'informazione utilissima ai fini del processo decisionale. Tutti i valori compresi tra 0 e 1 danno un peso variabile tra questi due estremi appena discussi.

NN)⁵ e Support Vector Machine (SVM). Si possono anche usare metodi abbinati (combinazioni di modelli), come Random Forest⁶ e altri metodi di potenziamento come AdaBoost e XGBoost.

Algoritmi di regressione

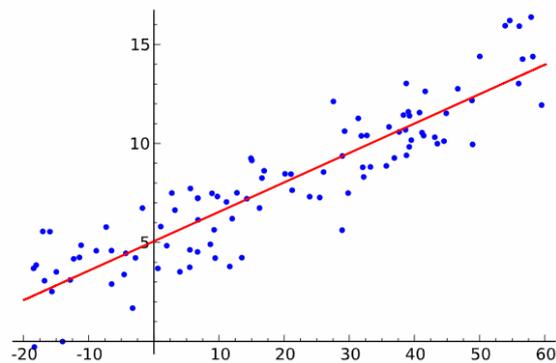


Figura 2: La regressione individua una funzione che traccia una linea che minimizzi la distanza media tra i punti.

Un metodo/modello di regressione è una tipologia di problema di apprendimento supervisionato che richiede una previsione di un numero. L'algoritmo più immediato e di immediata utilizzazione è la regressione lineare (metodo dei minimi quadrati).

Altri algoritmi di regressione di machine learning comuni (escluse le reti neurali) includono Naive Bayes, Albero decisionale, K-Nearest Neighbors, LVQ (Learning Vector Quantization), Least Angle Regression (LARS), Elastic Net, Random Forest, AdaBoost e XGBoost.

In questo contesto tecnicamente risulta altamente probabile notare una marcata sovrapposizione tra gli algoritmi di apprendimento automatico per la regressione e quelli per la classificazione. La scelta finale su quale metodo propendere rimane sempre in capo al *decision-maker*.

⁵ È un algoritmo di apprendimento supervisionato in cui il valore di k è predefinito e le variazioni del valore daranno origine a variazioni nelle metriche di prestazione. In questo algoritmo, i punti dati vengono tracciati in base alle loro caratteristiche indipendenti e quindi in base al valore di k , viene formato un raggio attorno al punto dati che deve essere classificato in modo da includere punti dati più vicini attorno ad esso. Il valore di k deve essere un numero dispari per la classificazione in due gruppi.

⁶ La foresta casuale è una classificazione supervisionata e un algoritmo di regressione. Questo algoritmo crea una serie di alberi decisionali utilizzando il sistema basato su regole. Come suggerisce il nome, forma una foresta; maggiore è il numero di alberi migliore è la sua misura delle prestazioni. Questo algoritmo crea prima i dati di *bootstrap* dai set di dati originali. Quindi, crea un albero decisionale sulla base dei dati di bootstrap. Possiamo creare un numero ' n ' di divisioni di caratteristiche (*features*) selezionando una caratteristica, due o n caratteristiche. Questi dati di funzionalità divisi vengono quindi convertiti in albero decisionale e questi alberi vengono analizzati per i dati da classificare e viene raggiunta una decisione sulla base dei voti massimi per la classe.

Algoritmi di clustering

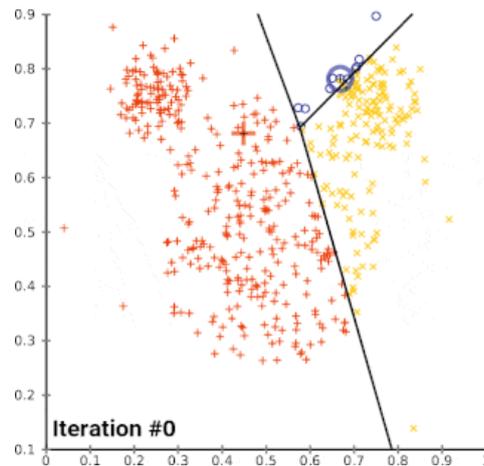


Figura 3: L'algoritmo K-Means affina il raggruppamento dei dati (Immagine di Chire.)

Un metodo di clustering è una tipologia di problema di apprendimento senza supervisione che richiede al modello di individuare gruppi di punti dai quali fare emergere delle similarità tra loro.

L'algoritmo più diffuso in questo ambito è il K-Means Clustering. Ricordiamo altri possibili esempi: clustering a spostamento medio, DBSCAN (cluster spaziale basato su densità di applicazioni con rumore), GMM (Gaussian Mixture Models) e HAC (Hierarchical Agglomerative Clustering o Clustering gerarchico).

4.2 Deep Learning e Reti Neurali

Il Deep Learning (apprendimento approfondito) è il ramo più avanzato del Machine Learning. Si tratta di un insieme di tecniche basate su reti neurali artificiali organizzate in diversi strati: ogni strato calcola i valori per quello successivo, in modo da elaborare l'informazione in maniera sempre più completa. Con una quantità sufficiente di dati, il sistema è in grado di apprendere la rappresentazione corretta e di risolvere problemi di apprendimento automatico senza la necessità di un pre-processamento dei dati, come invece avviene per le tecniche tradizionali di Machine Learning. Per quanto "tecnologicamente nuove" le tecniche di Deep Learning affondano le proprie radici nel passato, più esattamente nell'ambito delle reti neurali artificiali. Le reti neurali, note anche come reti neurali artificiali (o ANN, Artificial Neural Network) o reti neurali simulate (o SNN, Simulated Neural Network) sono un sottoinsieme del Machine Learning e sono l'elemento centrale degli algoritmi di Deep Learning. Il loro nome e la loro struttura sono ispirati al cervello umano, imitando il modo in cui i neuroni biologici si inviano segnali. Le reti neurali fanno affidamento sui dati di addestramento per imparare e migliorare la loro accuratezza nel tempo.

Dopo decenni di sviluppo, il Deep Learning ha registrato un ottimo successo in un'ampia gamma di complesse applicazioni di Intelligenza Artificiale che spaziano dalle attività di riconoscimento di modelli come la didascalia delle immagini (Le Cun, Bengio & Hinton, 2015) e l'elaborazione del

linguaggio naturale (Socher, Bengio & Manning, 2012), ai problemi di previsione per dati sequenziali come la previsione finanziaria e la previsione di cambiamenti di direzione nelle negoziazioni (Bao, Yue & Rao, 2017). Possiamo fare riferimento a due importanti architetture che hanno avuto un grande successo nella previsione di serie temporali; queste architetture sono RNN (Recurrent Neural Network) e LSTM (Long Short Term Memory). La RNN è un'architettura di rete profonda ampiamente adottata che utilizza le informazioni sequenziali (Cho *et al.*, 2014). La RNN funziona trasmettendo selettivamente le informazioni attraverso le fasi temporali durante l'elaborazione degli elementi di dati. Questa proprietà è essenziale per le applicazioni che includono la previsione della domanda turistica in cui la struttura incorporata nei dati di serie temporali sequenziali trasmette utili informazioni di contesto. Inoltre, la RNN conserva la sua memoria nel neurone dello strato nascosto di dimensioni fisse, che cattura tutte le informazioni precedentemente elaborate. L'output del neurone viene, quindi, generato in base all'input corrente e al precedente stato del neurone del livello nascosto attraverso il ciclo di feedback nella rete. La RNN può modellare la relazione di dipendenza tra sequenze di elementi tramite *loop* e ha avuto successo nelle applicazioni su dati non di serie temporali, inclusi dati genetici (Baldi & Pollastri, 2003). La LSTM è un'estensione della RNN che ha non solo l'unità di apprendimento ricorrente all'interno della rete ma anche diverse porte per catturare gli stati più lunghi dall'unità iniziale e gli stati più brevi dall'ultima unità. Avendo questa caratteristica, la LSTM è stato ampiamente utilizzato per risolvere problemi di previsione di serie temporali. Il meccanismo di attenzione è un metodo di ingegneria delle funzionalità che funziona insieme a varie architetture di rete profonde. Assegnando percentuali di peso diverse a diversi input, il modello potrebbe apprendere l'importanza dell'input senza doverlo fare prima di adattarsi al modello. Il meccanismo di attenzione su LSTM potrebbe adattarsi perfettamente alla previsione della domanda turistica e fornisce una soluzione *end-to-end* sia per la selezione delle caratteristiche che per la previsione.

4.3 Support Vector Machine

I data scientist possono applicare le Support Vector Machine (SVM) per risolvere sia i problemi di regressione che quelli di classificazione. Un SVM è un classificatore binario non probabilistico. L'aspetto non probabilistico è in contrasto con i classificatori probabilistici, come Naïve Bayes, che calcolano la probabilità di appartenenza alla classe sulla base degli esempi di addestramento (Aggarwal, 2015). Un SVM separa i dati attraverso un limite decisionale, che è un piano nello spazio delle caratteristiche multidimensionali. Solo un piccolo sottoinsieme dei dati tocca o supporta il confine decisionale, motivo per cui gli inventori li hanno chiamati vettori di support. Un SVM non può classificare i dati in più di due classi. La gestione di dati multiclasse con SVM è ancora un'area di ricerca attiva (Shao, Liu & He, 2021). Tuttavia, ci sono alcune soluzioni alternative. I metodi implicano la creazione di più SVM che confrontano i vettori di funzionalità tra loro utilizzando varie tecniche come *one-versus-Rest* (OVR) o *one-versus-one* (Bhavsar & Ganatra, 2012). Per k classi, il metodo OVR addestra k classificatori in modo che ogni classe discrimini le restanti $k - 1$ classi. OVO crea un problema di classificazione binaria per tutti i possibili accoppiamenti di classi, quindi richiede $k(k - 1)/2$ classificatori. Dopo aver costruito il numero di classificatori binari richiesti per i metodi OVR o OVO, l'algoritmo classifica un nuovo oggetto in base al voto di maggioranza tra l'insieme di classificatori. SVM è un metodo di Machine Learning supervisionato in cui ogni punto è un insieme

di caratteristiche $\{x_1, \dots, x_n\}$ e un'etichetta di classe y_i . Un SVM tratta ogni oggetto dati come un punto nello spazio delle caratteristiche che appartiene a una delle sole due classi. SVM definisce che le etichette di classe sono $y_i = 1$ o $y_i = -1$. Quindi, la rappresentazione matematica del set di dati è

$$\text{Data} = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in (-1, +1)\}_{i=1}^n \quad (1)$$

dove p è la dimensione del vettore della caratteristica e n è il numero di vettori.

Durante l'addestramento, il classificatore SVM trova un limite decisionale lineare nello spazio delle caratteristiche che meglio separa gli oggetti dati nelle due classi. Il problema di ottimizzazione equivalente trova due *iperpiani* paralleli che formano il divario più ampio privo di oggetti dati. Un iperpiano è un sottospazio con una dimensione inferiore allo spazio ambientale. La distanza perpendicolare tra gli iperpiani paralleli è il margine. L'iperpiano che biforca equidistante lo spazio tra gli iperpiani paralleli definisce un confine decisionale multidimensionale che separa i dati in due parti.

4.3.1 Iperpiani

Un SVM utilizza un iperpiano lineare piuttosto che uno non lineare perché in pratica quest'ultimo tenderà ad adattarsi troppo strettamente (*overfit*) a un confine che separerebbe perfettamente i dati di addestramento, ma non necessariamente i nuovi dati. In altre parole, l'*overfitting* del modello sui dati di addestramento può causare una scarsa generalizzazione del classificatore prevedendo in modo impreciso la classe dei nuovi dati (Aggarwal, 2015). L'equazione generale per un iperpiano lineare è

$$\mathbf{w} \cdot \mathbf{x} = 0 \quad (2)$$

Un iperpiano familiare nello spazio 2D è una linea

$$y = ax + b. \quad (3)$$

Riscritto nella forma standard per un iperpiano dà

$$y - ax - b = 0. \quad (4)$$

I vettori equivalenti sono

$$\mathbf{w} = \begin{pmatrix} -b \\ -a \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} 1 \\ x \\ y \end{pmatrix} \quad (5)$$

dove $w_0 = -b$, $w_1 = -a$ e $w_2 = 1$. Quindi, prendendo il prodotto scalare di \mathbf{w} e \mathbf{x} e ponendolo uguale a zero si ottiene l'equazione di forma standard per una linea. I valori specifici per le componenti del vettore di \mathbf{w} definiscono una linea specifica (iperpiano lineare) e qualsiasi punto definito da un (x, y) , che è il vettore della caratteristica, deve trovarsi da qualche parte su quella linea. Dato che i vettori sono vettori colonna, il prodotto scalare è equivalente all'operazione di matrice

$$\mathbf{w}^T \mathbf{x} = 0. \quad (6)$$

Le forme vettoriali e matriciali sono più facili da rappresentare nella memoria del computer per calcoli rapidi.

Il prodotto scalare di due vettori \mathbf{x} e \mathbf{y} è

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \times \|\mathbf{y}\| \times \cos(\theta) = \sum_i x_i y_i = |\mathbf{x}\mathbf{y}^T| = \alpha \quad (7)$$

che è un valore scalare. Quindi, perché il prodotto scalare di \mathbf{w} e \mathbf{x} sia zero, i vettori devono essere perpendicolari l'uno all'altro perché $\cos(90) = 0$. Cioè, se \mathbf{x} giace su un piano dello spazio, allora il vettore \mathbf{w} che definisce quel piano deve essere perpendicolare (normale) sia a \mathbf{x} che al piano. Quindi, il vettore unitario

$$\mathbf{u} = \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (8)$$

Dove $\|\mathbf{w}\|$ è la norma del vettore o la lunghezza deve anche essere perpendicolare al piano. Per l'esempio 2D,

$$\mathbf{u} = \frac{\mathbf{w}}{\|\mathbf{w}\|} = \left(\frac{w_1}{\|\mathbf{w}\|}, \frac{w_2}{\|\mathbf{w}\|} \right). \quad (9)$$

Questo vettore unitario di \mathbf{w} è importante per trovare la distanza di qualsiasi punto (caratteristica) dall'iperpiano proiettando quel punto su un vettore normale all'iperpiano. Ad esempio, il vettore \mathbf{p} nella Figura 1 è la proiezione del punto A sul piano del vettore \mathbf{w} che è normale all'iperpiano. Quindi la distanza dal punto A all'iperpiano è uguale alla lunghezza di \mathbf{p} , che è $\|\mathbf{p}\|$. In questo esempio, la proiezione del vettore \mathbf{a} sul piano di \mathbf{w} è

$$\mathbf{p} = (\mathbf{u} \cdot \mathbf{a})\mathbf{u} \quad (10)$$

Il prodotto scalare produce uno scalare, che è la grandezza (lunghezza) del vettore tale che

$$\mathbf{u} \cdot \mathbf{a} = \sum_i u_i a_i \quad (11)$$

e la direzione del vettore è \mathbf{u} .

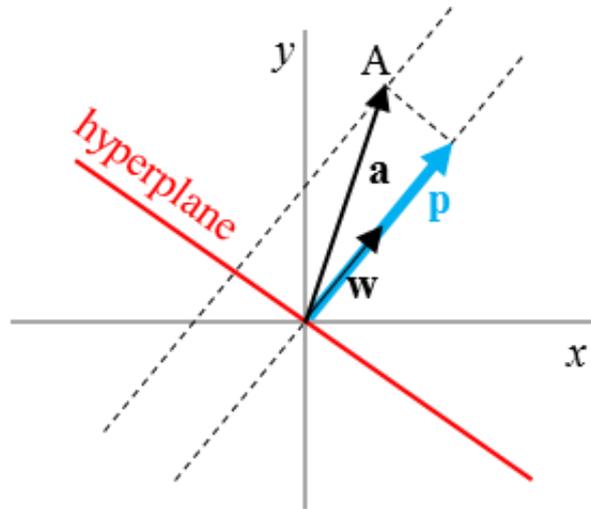


Figura 1: Proiezione di un vettore per calcolare la distanza da un iperpiano.

4.3.2 Ottimizzazione del margine

L'illustrazione della Figura 2 mostra che la regione delimitata dai due iperpiani H1 e H2 è priva di punti dati. I "vettori di supporto" sono gli oggetti dati ai confini dei due iperpiani. Figura 2 evidenzia i quattro vettori di supporto (punti) di questo esempio con i bordi neri. Il confine decisionale è un iperpiano lineare H0 equidistante tra i due iperpiani H1 e H2. Il margine dell'iperpiano è la distanza di separazione tra i due iperpiani H1 e H2. Equivalentemente, gli iperpiani H1 e H2 sono sfalsati di una certa quantità δ dall'iperpiano del confine decisionale H0. Associando \mathbf{w} all'iperpiano H0, l'iperpiano parallelo con offset $+\delta$ su un lato è

$$\mathbf{w} \cdot \mathbf{x} = +\delta \quad (12)$$

ed è un offset uguale e opposto all'altro lato dell'iperpiano

$$\mathbf{w} \cdot \mathbf{x} = -\delta \quad (13)$$

Dato un vettore di feature \mathbf{x}_i , la classe assegnata deve soddisfare

$$\mathbf{w} \cdot \mathbf{x}_i \geq +\delta \quad (14)$$

per un'etichetta di classe di $\mathbf{y}_i = 1$ e

$$\mathbf{w} \cdot \mathbf{x}_i \leq -\delta \quad (15)$$

per un'etichetta di classe di $\mathbf{y}_i = -1$. Risolvendo il problema di ottimizzazione si troverà l'optimum \mathbf{w} che massimizza l'offset o il margine. Per comodità matematica, l'impostazione dell'offset su $\delta = 1$ facilita la combinazione dei due vincoli di classificazione in un unico vincolo. Ovvero, moltiplicando entrambi i lati dell'equazione (14) per \mathbf{y}_i e assegnando il valore dell'etichetta di classe +1 ai risultati corretti

$$y_i(\mathbf{w} \cdot \mathbf{x}_i) \geq +1(+1) \Leftrightarrow y_i(\mathbf{w} \cdot \mathbf{x}_i) \geq 1 \quad (16)$$

Facendo lo stesso per l'equazione (15) otteniamo:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i) \leq -1(-1) \Leftrightarrow y_i(\mathbf{w} \cdot \mathbf{x}_i) \geq 1 \quad (17)$$

Di conseguenza, l'unico vincolo del problema di ottimizzazione diventa

$$y_i(\mathbf{w} \cdot \mathbf{x}_i) \geq 1 \quad \forall \quad 1 \leq i \leq n \quad (18)$$

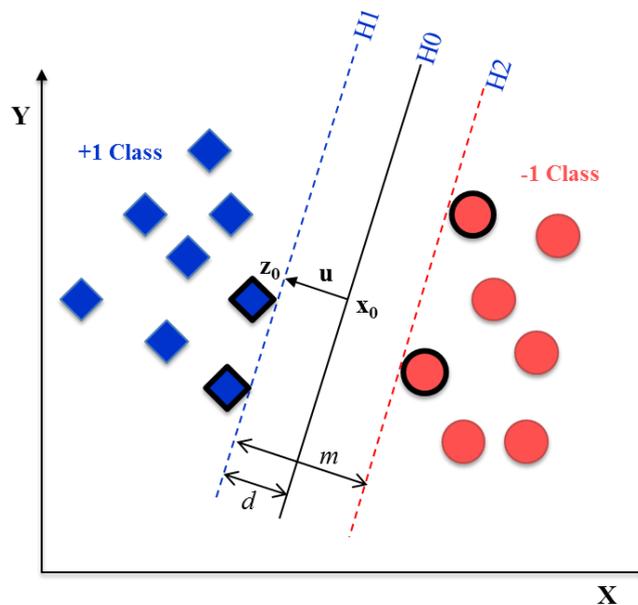


Figura 2: Massimizzazione del margine di un iperpiano.

Dato che il vettore \mathbf{w} è perpendicolare (normale) all'iperpiano H_0 , il suo vettore unitario $\mathbf{u} = \mathbf{w} / \|\mathbf{w}\|$ deve essere perpendicolare a H_0 con modulo 1. Il vettore $d\mathbf{u}$ è il vettore perpendicolare dall'iperpiano H_0 a un iperpiano parallelo H_1 a una certa distanza d . Sia \mathbf{x}_0 la coordinata di base del vettore $d\mathbf{u}$ sull'iperpiano e \mathbf{z}_0 la coordinata di punta che giace sull'iperpiano H_1 . Il vettore distanza è

$$\mathbf{z}_0 = \mathbf{x}_0 + d\mathbf{u} \quad (19)$$

Il fatto che \mathbf{z}_0 sia su H_1 significa che

$$\mathbf{w} \cdot \mathbf{z}_0 = +1 \quad (20)$$

Moltiplicando entrambi i lati dell'equazione (19) per il prodotto scalare di \mathbf{w} e sostituendo \mathbf{z}_0 si ottiene

$$\mathbf{w} \cdot (\mathbf{x}_0 + d\mathbf{u}) = +1 \quad (21)$$

Sostituendo \mathbf{u} dall'equazione (8) si ottiene

$$\mathbf{w} \cdot \left(\mathbf{x}_0 + d \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) = +1 \quad (22)$$

Equazione in espansione (22) produce

$$\mathbf{w} \cdot \mathbf{x}_0 + d \frac{\mathbf{w} \cdot \mathbf{w}}{\|\mathbf{w}\|} = +1 \quad (23)$$

Dato che

$$\mathbf{w} \cdot \mathbf{w} = \|\mathbf{w}\|^2 \quad (24)$$

L'equazione (23) diventa

$$\mathbf{w} \cdot \mathbf{x}_0 + d\|\mathbf{w}\| = +1 \quad (25)$$

Quindi,

$$\mathbf{w} \cdot \mathbf{x}_0 = 1 - d\|\mathbf{w}\| \quad (26)$$

Il fatto che \mathbf{x}_0 sia su H_0 significa che

$$\mathbf{w} \cdot \mathbf{x}_0 = 0 \quad (27)$$

Sostituendo l'equazione (27) nell'equazione (26) si ottiene

$$0 = 1 - d\|\mathbf{w}\| \quad (28)$$

Risolvendo per la distanza d si ottiene

$$d = \frac{1}{\|\mathbf{w}\|} \quad (29)$$

Pertanto il margine, che è la distanza tra gli iperpiani H_1 e H_2 , deve essere $2/\|\mathbf{w}\|$. Quindi il margine è

$$m = \frac{2}{\|\mathbf{w}\|} \quad (30)$$

Si noti ancora che l'utilizzo delle etichette di classe +1 e -1 come limiti di decisione per i vettori al di sopra e al di sotto di H_1 , rispettivamente, ha consentito una semplice derivazione del margine in termini della sola lunghezza del vettore \mathbf{w} . Dall'equazione (30) il problema di ottimizzazione della massimizzare del margine diventa equivalente a minimizzare la lunghezza del vettore \mathbf{w} . Quindi, il problema di ottimizzazione è di minimizzare la norma di \mathbf{w} soggetta al vincolo dell'equazione (18), che mantiene l'iperpiano vuoto di oggetti dati. Cioè, il problema di ottimizzazione diventa

$$\begin{aligned} \min f(w) &= \|w\| \\ \text{s. t. } g: y_i(w \cdot x_i + w_0) &\geq 1, \quad 1 \leq i \leq n. \end{aligned} \quad (31)$$

Si noti che questa notazione per il vincolo separa w_0 , la componente bias, dal vettore w , il che significa che gli indici di iterazione devono essere compresi tra 1 e n anziché 0 e n .

Al momento della convergenza alla soluzione ottimale, il classificatore SVM addestrato diventa

$$x \mapsto \text{signum}(w \cdot x) \quad (32)$$

In altre parole, l'etichetta di classe prevista per un nuovo vettore di caratteristiche x è il segno del prodotto scalare tra il vettore w e il nuovo vettore. La funzione signum produce -1 se il suo argomento è negativo (minore di zero), 0 se il suo argomento è uguale a 0 e +1 se il suo argomento è positivo (maggiore di zero).

4.3.3 Valutazione del classificatore

I classificatori spesso riportano una serie di parametri prestazionali come accuratezza, precisione e richiamo (Géron, 2019). Questa sezione definirà le varie metriche di rendimento utilizzando le seguenti abbreviazioni:

- TP = vero positivo
- TN = vero negativo
- FP = Falso positivo (errore di tipo I o falso allarme)
- FN = Falso negativo (errore di tipo II o mancato rilevamento)

4.3.4 Matrice di confusione o errore

Una matrice di confusione è una tabella che riassume le prestazioni del classificatore.

Tabella 1 illustra una matrice di confusione con le relazioni definite tra ciascuna metrica delle prestazioni.

N	Predetto Positivo	Predetto Negativo	
Effettivo Positivo	TP	FN	TP+FN
Effettivo Negativo	FP	TN	FP + TN
	PA + PA	FN+TN	

Tabella 1: Esempio di matrice di confusione

La precisione di un classificatore è

$$a = \frac{TP + TN}{n} = \frac{TP + TN}{(TP + FN) + (FP + TN)} \quad (33)$$

Il tasso di errore del tasso di errata classificazione è $1 - a$. La precisione p di classificazione per ogni classe è una misura della sua accuratezza di previsione per quella classe. La precisione è la proporzione di previsioni positive che sono corrette dove

$$p = \frac{TP}{TP + FP} \quad (34)$$

La precisione, quindi, fornisce un confronto relativo delle prestazioni per ciascuna classe per valutare gli errori. I professionisti a volte si riferiscono alla precisione come specificità. Intuitivamente, si può pensare alla specificità come alla capacità di un classificatore di identificare una classe specifica senza errori. La specificità è anche $1 - FP$, che è il vero tasso negativo.

Il richiamo r è la proporzione delle classi positive che il classificatore prevede correttamente dove

$$r = \frac{TP}{TP + FN} \quad (35)$$

Intuitivamente, si può pensare al richiamo come alla proporzione dei veri positivi che un classificatore può richiamare dal campione mediante le sue previsioni. Pertanto, i professionisti a volte si riferiscono al richiamo come sensibilità.

La Misura F_1 è la media armonica di precisione e richiamo tale che

$$F_1 = 2 \times \frac{p \cdot r}{p + r} \quad (36)$$

Quindi, la media armonica è una media ponderata della precisione e del richiamo dove 1 è il migliore e 0 è il peggiore. Si noti che F_1 è 1 se $p = r = 1$ e zero se p o r è zero. Riscriviamo la misura F_1 come

$$F_1 = \frac{TP}{TP + \left(\frac{FP + FN}{2}\right)} \quad (37)$$

chiarisce che il punteggio equivale a una precisione o a un richiamo basato sulla media degli errori totali, che è $(FP + FN)/2$. In pratica, c'è sempre un compromesso tra sensibilità e specificità. Quindi, la misura F_1 riflette l'ammontare del compromesso coinvolto.

4.3.5 ROC e AUC

ROC è spesso confuso con il termine regione di convergenza. Tuttavia, sta per Receiver Operating Characteristic nel contesto della valutazione delle prestazioni del classificatore. La ROC è un grafico del tasso TP rispetto al tasso FP (errore di tipo I), in funzione della soglia di decisione di un classificatore (Fawcett, 2006). L'asse y e l'asse x sono intervalli per i tassi TP e FP, rispettivamente.

Figura 3 fornisce un'illustrazione per derivare una ROC basata su ipotetiche distribuzioni di probabilità di due classi. In questo scenario, un ricevitore prevede la trasmissione di una cifra binaria in base a una caratteristica ricevuta. Il grafico copre il valore della caratteristica sull'asse x e la probabilità di ricezione per ogni classe sull'asse y . La soglia caratteristica T è impostata in modo tale che il classificatore, che in questo scenario è un ricevitore binario, preveda la trasmissione come una cifra zero e una per le tensioni ricevute rispettivamente al di sotto e al di sopra di T .

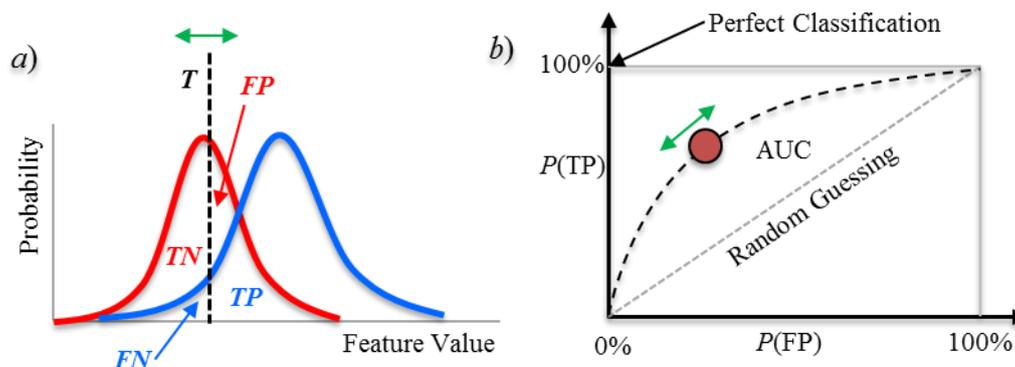


Figura 3: Determinazione della curva ROC per un classificatore.

La quantità di sovrapposizione tra le due distribuzioni è proporzionale alla quantità di rumore nel canale di trasmissione, che crea confusione sulle cifre binarie ricevute. Cioè, alcune delle tensioni che rappresentano zeri che superano la soglia faranno sì che il classificatore le preveda invece come uno, quindi quelle sarebbero previsioni FP. Allo stesso modo, alcune delle tensioni che rappresentano quelle che scendono al di sotto della soglia farebbero sì che il classificatore le preveda invece come zeri, quindi quelle sarebbero previsioni FN. Un canale a basso rumore produrrebbe distribuzioni per zeri e uno che non si sovrappongono, nel qual caso il classificatore può trovare una soglia chiara per separare le classi con errori FP o FN. La quantità di sovrapposizione nelle distribuzioni influisce sulla forma della curva ROC. Nessuna sovrapposizione sposta la curva verso la coordinata della classificazione perfetta (0, 1). La sovrapposizione completa sposta la curva ROC verso la diagonale, che rappresenta i risultati per l'ipotesi casuale. Quando la curva ROC scende al di sotto della diagonale, il classificatore

ha prestazioni peggiori rispetto all'ipotesi casuale. Data una certa quantità di sovrapposizione nella distribuzione dei valori delle caratteristiche per ciascuna classe (Figura 3a) spostare la soglia equivale a spostare il punto lungo la curva ROC (Figura 3b). L'obiettivo della progettazione del classificatore è massimizzare l'area sotto la curva ROC (AUC) e spostare la soglia di rilevamento verso il punto di classificazione perfetta. Il compromesso tra sensibilità e specificità diventa evidente immaginando di spostare la linea di soglia T . Aumentando il livello di soglia spostandolo verso destra si ridurrà il tasso FP perché la sensibilità verrà abbassata, riducendo così la probabilità di prevedere il rumore come positivo. Tuttavia, la soglia più alta (sensibilità inferiore) aumenterà anche il numero di rilevamenti mancati non rilevando i veri positivi che scendono al di sotto della soglia, aumentando così il tasso di FN. Riassumendo, diminuendo la sensibilità del ricevitore diminuirà la velocità FP ma aumenterà la velocità FN. Il contrario si verifica quando si aumenta la sensibilità abbassando la soglia. Cioè, il tasso FP aumenterà (diminuzione della specificità) perché il ricevitore può rilevare i picchi di rumore come positivi mentre il tasso FN diminuirà (aumento della sensibilità) perché il ricevitore avrà meno probabilità di perdere un segnale debole che è un vero positivo.

4.3.6 La convalida incrociata

L'obiettivo della convalida di un classificatore è determinare come si comporta su dati invisibili. Questo è un meccanismo per impedire l'impostazione di iperparametri che causano un *overfit del modello* con i dati di training, che può portare a una scarsa generalizzazione delle prestazioni sui nuovi dati. Il metodo più semplice di convalida delle prestazioni del classificatore suddivide il set di dati in un sottoinsieme di addestramento e di convalida. In genere, il classificatore si allena sul 70% dei dati e utilizza il restante 30% per convalidare le sue prestazioni utilizzando una o più misure delle prestazioni come la misura F1 e AUC. In letteratura ci si riferisce anche alla suddivisione dei dati in due parti come validazione *leave-p-out*. In altre parole, l'approccio lascia p osservazioni in un set di dati di convalida e utilizza il resto per l'addestramento. Alcune tecniche utilizzano la convalida *leave-one-out* ($p = 1$) per set di dati relativamente piccoli. Un'altra tecnica chiamata convalida incrociata valuta le prestazioni di un classificatore segmentando il set di dati in più parti quasi uguali in modo che il ciclo di valutazione utilizzi ciascuna parte per la convalida almeno una volta e l'unione delle parti rimanenti per addestrare un nuovo modello. Successivamente, il punteggio medio di ciascun modello diventa la prestazione riportata del classificatore. La convalida incrociata K-fold è una generalizzazione che suddivide i dati in un set di convalida e $k - 1$ set di addestramento. L'algoritmo crea le partizioni (sottoinsiemi o pieghe) selezionando casualmente punti dati per ogni sottoinsieme. L'utente può anche modificare la randomizzazione per produrre una stratificazione che equilibri la proporzione di classi in ogni sottoinsieme. Il vantaggio principale della convalida incrociata k-fold è che utilizza tutti i dati sia per l'addestramento che per la convalida per evitare distorsioni verso qualsiasi parte dei dati. Lo svantaggio è l'aumento del tempo di elaborazione per set di dati più grandi.

4.4 L'Algoritmo Fuzzy C-Means (FCM) e le classificazioni dei "Turisti in entrata"

L'analisi dei *cluster Fuzzy C-means* è un algoritmo di *clustering soft* (cioè appartiene alla famiglia degli algoritmi di partizione non esclusivi) proposto da Bezdek (1981). Questo algoritmo è caratterizzato da una "logica fuzzy" (il termine "fuzzy" significa letteralmente "sfocatura"), cioè permette di attribuire ai dati valori probabilistici di appartenenza ai cluster.

L'algoritmo di partizione *Fuzzy C-Means* permette, automaticamente, di classificare i dati di mercato in tempo reale, fornendo all'operatore non solo l'indicazione della presenza di potenziali anomalie (e quindi potenziali opportunità di trading interessanti) ma anche di associare a tali, informazione.

L'approccio metodologico *Fuzzy C-Means* (FCM) è una tecnica di *clustering* dei dati in cui, per ogni punto del training set, viene determinato un certo grado di appartenenza per ciascuno dei cluster identificati. Questo approccio è stato applicato per eseguire l'analisi finanziaria sociale. Ad esempio, Lindholm & Liu (2003) hanno utilizzato per la valutazione di allerta precoce degli indicatori di crisi finanziaria e Drobnic *et al.* (2000) lo hanno utilizzato per sviluppare regole fuzzy da un raggruppamento ottenuto da algoritmi di mappe auto-organizzanti.

Gli algoritmi di *clustering fuzzy* assegnano valori di appartenenza reali compresi tra 0 (non indica alcuna somiglianza con il cluster e quindi nessuna probabilità di appartenenza) e 1 (somiglianza completa). Per questa ragione:

- Eventuali dati possono appartenere in parte a cluster diversi;
- La somma dei punti relativi ad ogni possibile cluster deve essere pari a 1;
- La somma delle adesioni all'interno della matrice di appartenenza U , non deve superare il valore complessivo di N .

Fondamentalmente, dati N punti, l'obiettivo è suddividerli in C partizioni separate; l'aspetto particolare e innovativo rispetto al *clustering standard* è che questo approccio permette di determinare quali sono i gradi di appartenenza (livelli di sfumatura) dell' i -esimo punto relativo a ciascuna delle partizioni considerate.

L'algoritmo funziona assegnando casualmente un certo grado di appartenenza (μ_{ij}) a ciascun punto i rispetto a ciascuno dei j cluster e ottenendo così una certa distribuzione spaziale iniziale dei centri di massa delle partizioni C che si vogliono derivare.

Attraverso una procedura iterativa, la funzione sposta dinamicamente il baricentro verso la localizzazione ottimale che andrà a minimizzare un funzionale di costo che rappresenta la somma delle distanze euclidee di ciascun punto da ciascun centro di massa, opportunamente pesate con il corretto grado di appartenenza.

L'algoritmo si basa quindi sulla minimizzazione del seguente funzionale di costo $F(n, C, m)$:

$$F(n, C, m) = \sum_{i=1}^N \sum_{j=1}^C (\mu_{ij})^m \|x_i - c_j\|^2$$

Dove:

N = numero di punti del set di dati informativi;

C = il numero di cluster da identificare in termini di “ambiti decisionali”;

m = un parametro che controlla l'estensione della condivisione dell'appartenenza al cluster ed è qualsiasi numero reale maggiore di 1;

μ_{ij} = grado di appartenenza dell' i -esimo punto relativo al j -esimo gruppo;

c_j = centro del j -esimo cluster;

x_i = i -esimo punto del dataset.

Nel *clustering fuzzy*, gli elementi del set di dati possono appartenere a più di un cluster e ogni cluster è associato a un grado di appartenenza (livello di appartenenza).

A livello procedurale, possiamo evidenziare 4 passaggi distinti che devono essere effettuati dall'euristica per ottenere l'ottimale.

1. I gradi di appartenenza μ_{ij} devono essere inizializzati casualmente in modo da mantenere due condizioni:

$$\mu_{ij} > 0$$

$$\sum_{j=1}^c \mu_{ij} = 1$$

2. La distanza euclidea c_j si calcola con la seguente formula:

$$c_j = \sqrt{\sum_{k=1}^p (O_{ik} - X_{jk})^2}$$

Dove:

p = è il numero di variabili di input;

O_i = rappresenta la i^{th} osservazione;

X_j = è il centro del gruppo j .

Aggiornamento di tutti i nuovi μ_{ij} tramite:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^p \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^2}$$

2. Calcolo del valore del funzionale di costo J
3. Ripetizione dei passaggi 2-3-4 fino a quando non si ottiene praticamente alcun miglioramento nella funzione di costo o si raggiunge il numero massimo di iterazioni imposto dallo sperimentatore.

Sulla base di quanto sopra, si può considerare una procedura per identificare e riassegnare osservazioni incerte introducendo alcuni pesi nell'algoritmo FCM.

Tali pesi sono calcolati sulla base di alcune variabili socio economiche e per ciascuna di esse viene definita una variabile linguistica utilizzando cinque termini linguistici:

- a) Very low (VL),
- a) Low (L),
- b) Medium (A),
- c) High (H),
- d) Very high (VH).

Ogni termine linguistico è rappresentato da un numero sfocato trapezoidale. Per ogni variabile e per ogni termine linguistico si definisce il numero trapezoidale seguendo la procedura in Costea (2005).

Fondamentalmente, per ogni variabile, viene applicato l'algoritmo FCM e i valori di quella variabile sono raggruppati in cinque classi fuzzy. I termini linguistici sono definiti come segue (vedi Figura 7):

- Il termine linguistico VL è definito da tre punti: un punto minimo (A), un punto massimo (B) e un punto minimo per il termine linguistico L (C);
- I termini linguistici L, A, H sono definiti da quattro punti: termine linguistico precedente (ad esempio punto B per il termine linguistico L), un punto minimo (ad esempio punto C per il termine linguistico L), un punto massimo (ad esempio, punto D per il termine linguistico L) e il punto minimo per il termine linguistico successivo (ad esempio, E per il termine linguistico L);
- Il termine linguistico VH è definito da tre punti: il punto massimo per il termine linguistico (H), un punto minimo (I) e un punto massimo (J).

In altre parole, per definire tutti e cinque i numeri sfocati, dobbiamo definire un punto minimo e massimo per ogni termine linguistico. Il punto minimo/massimo per il termine linguistico LT è definito come il valore minimo/massimo di quell'intero data set ratio per il quale il titolo di appartenenza alla classe di termini linguistici LT è maggiore o uguale a 0,99.

Per ogni report e per ogni termine linguistico, il punto minimo viene inizializzato su *+inf* e il punto massimo viene inizializzato su *-inf*. È possibile che non ci siano agenzie con un numero di iscritti pari o superiore a 0,99, anche se questo caso è improbabile.

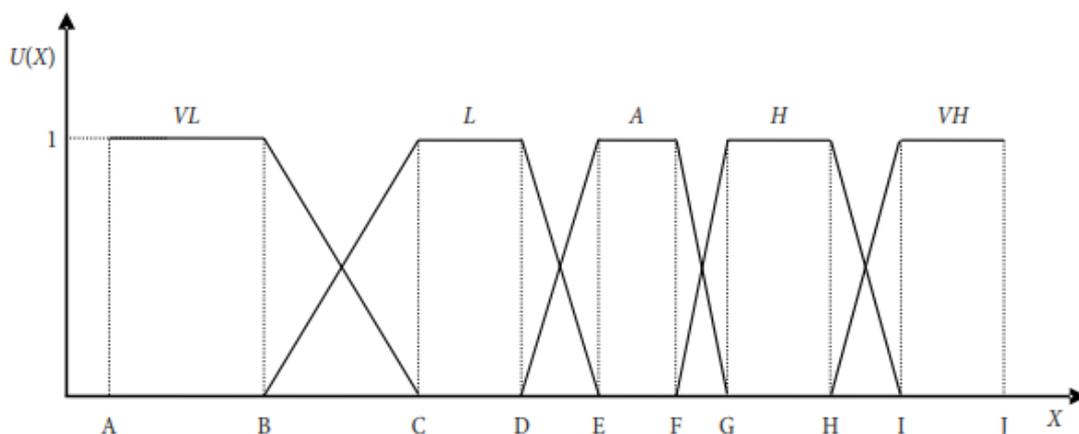


Figura 7: La rappresentazione trapezoidale dei cinque termini linguistici (VL, L, A, H, VH) per una variabile generica X.

4.5 La logistica della Funzione Gompertz: dalla biologia uno strumento di analisi delle dinamiche di una popolazione

Simile all'equazione logistica, il modello di Gompertz è formalizzato dall'equazione differenziale:

$$\frac{dN}{dt} = -rN \ln\left(\frac{N}{K}\right)$$

dove N rappresenta la popolazione, mentre le costanti r e K rispettivamente il tasso di crescita ed il *carrying capacity*, ovvero il termine asintotico della popolazione (definito dalle risorse disponibili nell'ambiente).

La legge di Gompertz è un modello dinamico unidimensionale di popolazione studiato da Benjamin Gompertz.

La soluzione generale di tale equazione differenziale rappresenta l'equazione della legge di Gompertz. Risolvendo si ottiene:

$$N(t) = Ke^{ce^{-rt}}$$

Imponendo la condizione iniziale

$$N(0) = N_0$$

si ottiene il valore di c : $c = \ln\left(\frac{N_0}{K}\right)$

Dalla formula si vede che il limite asintotico della funzione è K , ovvero:

$$\lim_{t \rightarrow \infty} N(t) = K$$

In un sistema che segue il modello di Gompertz si ha per popolazioni poco numerose rispetto alle risorse presenti nell'ambiente una crescita inizialmente esponenziale che in seguito si stabilizza, rallentando fino a diventare quasi lineare una volta che la popolazione si avvicina asintoticamente ad un valore di equilibrio.

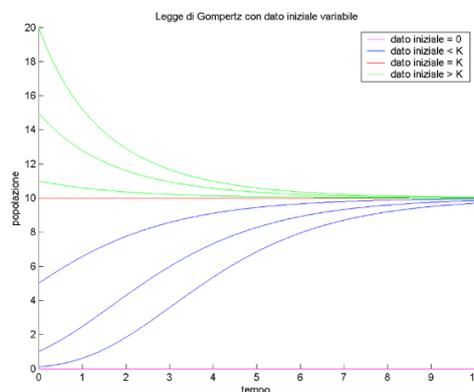


Figura 8: Legge di Gompertz

Se invece la popolazione è superiore a quanta ne può sostenere l'ambiente, questa diminuisce fino a raggiungere uno stato di equilibrio. La crescita dei tumori in un tessuto segue questa curva. Molti modelli di crescita economica fanno spesso riferimento a questo strumento. La nostra idea è di applicare questo tool per l'analisi di modelli di evoluzione di flussi turistici nel tempo e nello spazio territoriale oggetto di studio.

Bibliografia

- Abellana, D. P. M., Rivero, D. M. C., Aparente, M. E., & Rivero, A. (2021). Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario. *Journal of Tourism Futures*, 7(1), 78-97.
- Aggarwal, Charu C. (2015). *Data Mining*. New York, New York: Springer International Publishing.
- Akerkar, R. (2019b). Machine learning. In R. Akerkar (Ed.), *SpringerBriefs in business. Artificial intelligence for business* (pp. 19–32). Springer International. https://doi.org/10.1007/978-3-319-97436-1_2MO
- Aladag, C. H., E. Egrioglu, U. Yolcu, and V. R. Uslu. (2014). “A High Order Seasonal Fuzzy Time Series Model and Application to International Tourism Demand of Turkey.” *Journal of Intelligent and Fuzzy Systems* 26 (1): 295–302.
- Althbiti, A., & Ma, X. (2020). Machine learning. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of big data* (pp. 1–5). Springer International. https://doi.org/10.1007/978-3-319-32001-4_539-1
- Assaf, A. G., C. P. Barros, and L. A. Gil-Alana. (2011). “Persistence in the Short- and Long-Term Tourist Arrivals to Australia.” *Journal of Travel Research* 50 (2): 213–29.
- Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2019). Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research*, 58(3), 383–397.
- Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management*, 29(1), 19–31.
- Athanasopoulos, G., R. J. Hyndman, H. Song, and D. C. Wu. (2011). “The Tourism Forecasting Competition.” *International Journal of Forecasting* 27 (3): 822–44.
- Baldi, P., & Pollastri, G. (2003). The Principled design of large-scale recursive neural network architectures-DAG-RNNs and the protein structure prediction problem. *Journal of Machine Learning Research*, 4, 575–602.
- Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One*, 12(7), e0180944.
- Bezdek, J. C., Coray, C., Gunderson, R., & Watson, J. (1981). Detection and characterization of cluster substructure i. linear structure: Fuzzy c-lines. *SIAM Journal on Applied Mathematics*, 40(2), 339-357.
- Bhilare, S., Shah, J., Gaikwad, V., Gupta, G., Sanghvi, Y. S., Bhanage, B. M., & Kapdi, A. R. (2019). Pd/PTABS: An Efficient Catalytic System for the Aminocarbonylation of a Sugar-Protected Nucleoside. *Synthesis*, 51(22), 4239-4248.
- Bi, J. W., Liu, Y., & Li, H. (2020). Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research*, 83, 102923.
- Blunk, S. S., Clark, D. E., & McGibany, J. M. (2006). Evaluating the long-run impacts of the 9/11 terrorist attacks on US domestic airline travel. *Applied Economics*, 38(4), 363–370.
- Buhalis, D., & Amaranggana, A. (2015). Smart tourism destinations enhancing tourism experience through personalisation of services. In *Information and communication technologies in tourism 2015* (pp. 377–389). Cham: Springer. <https://doi.org/10.1007/978-3-319-14343-9\textunderscore>
- Cai, Z. J., Lu, S., & Zhang, X. B. (2009). *Tourism demand forecasting by support vector regression and genetic algorithm* (pp. 144–146). 8–11 August 2009. (Beijing, China).
- Chen, R. J., Bloomfield, P., & Fu, J. S. (2003). An evaluation of alternative forecasting methods to recreation visitation. *Journal of Leisure Research*, 35(4), 441–454.
- Chen, R., C. Y. Liang, W. C. Hong, and D. X. Gu. (2015). “Forecasting Holiday Daily Tourist Flow Based on Seasonal Support Vector Regression with Adaptive Genetic Algorithm.” *Applied Soft Computing* 26:435–43.

- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder
- Chu, F. L. (2009). "Forecasting Tourism Demand with ARMA-Based Methods." *Tourism Management* 30 (5): 740–51.
- Claveria, O., E. Monte, and S. Torra. (2015). "Tourism Demand Forecasting with Neural Network Models: Different Ways of Treating Information." *International Journal of Tourism Research* 17 (5): 492–500.
- Divino, J. A., & McAleer, M. (2010). Modelling and forecasting daily international mass tourism to Peru. *Tourism Management*, 31(6), 846–854.
- Drobics, M., Winiwater, W., & Bodenhofer, U. (2000, November). Interpretation of self-organizing maps with fuzzy rules. In *Proceedings 12th IEEE International Conference on Tools with Artificial Intelligence. ICTAI 2000* (pp. 304-311). IEEE.
- Egger, R., & Herdin Thomas (Eds.) (2007). *Tourismus Herausforderung Zukunft*. Münster: LIT. Retrieved from <https://scholar.google.com/citations?user!49jwau7qaaaaj&hl!4en&oi!4sra>
- Egger, R., & Luger, K. (Eds.) (2015). *Tourismus und mobile Freizeit. Lebensformen, Trends, Herausforderungen*: LIT.
- Egger, R., & Yu, C.-E. (2022a). Epistemological Challenges. In R. Egger (Ed.) *Tourism on the verge. Applied data science in Tourism* (pp. 17–34). Cham: Springer.
- Egger, R., & Yu, C.-E. (2022b). Data Science and Interdisciplinarity. In R. Egger (Ed.), *Tourism on the verge. Applied data science in Tourism* (pp. 35–49). Cham: Springer.
- Ellis, J. B., & Doren, C. S. (1966). A comparative evaluation of gravity and system theory models for statewide recreational traffic flows. *Journal of Regional Science*, 6(2), 57–70.
- Fawcett, Tom. (2006). "An introduction to ROC analysis." *Pattern Recognition Letters* (Elsevier) 27 (8): 861-874. <http://people.inf.elte.hu/kiss/11dwhdm/roc.pdf>. or statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734).
- Fotheringham, A. S. (1983). A new set of spatial-interaction models: The theory of competing destinations. *Environment and Planning A: Economy and Space*, 15(1), 15–36.
- Géron, Aurélien. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd. Sebastopol, California: O'Reilly Media.
- Goh, C. (2012). "Exploring Impact of Climate on Tourism Demand." *Annals of Tourism Research* 39 (4): 1859–83.
- Goh, C., R. Law, and H. M. Mok. (2008). "Analyzing and Forecasting Tourism Demand: A Rough Sets Approach." *Journal of Travel Research* 46 (3): 327–38.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Machine learning basics. *Deep Learning*, 1, 98–164.
- Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets*, 25(3), 179–188. <https://doi.org/10.1007/s12525-015-0196-8>
- Gunter, U., and I. Önder. (2015). "Forecasting International City Tourism Demand for Paris: Accuracy of Uni- and Multivariate Models Employing Monthly Data." *Tourism Management* 46:123–35.
- Hassani, H., A. Webster, E. S. Silva, and S. Heravi. (2015). "Forecasting U.S. Tourist Arrivals Using Optimal Singular Spectrum Analysis." *Tourism Management* 46:322–35.
- Hassani, H., E. S. Silva, N. Antonakakis, G. Filis, and R. Gupta. (2017). "Forecasting Accuracy Evaluation of Tourist Arrivals." *Annals of Tourism Research* 63:112–27.
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1), 1–12.
- Hong, W.-C., Y. Dong, L.-Y. Chen, and S.-Y. Wei. (2011). "SVR with Hybrid Chaotic Genetic Algorithms for Tourism Demand Forecasting." *Applied Soft Computing* 11 (2): 1881–90.
- Jamal, M., Ahmad, W., Andleeb, S., Jalil, F., Imran, M., Nawaz, M. A., ... & Kamil, M. A. (2018). Bacterial biofilm and associated infections. *Journal of the chinese medical association*, 81(1), 7-11.

- Jiao, E. X., & Chen, J. L. (2019). Tourism forecasting: A review of methodological developments over the last decade. *Tourism Economics*, 25(3), 469–492.
- Jiao, X., Li, G., & Chen, J. L. (2020). Forecasting international tourism demand: a local spatiotemporal model. *Annals of Tourism Research*, 83, 102937.
- Jordan, M. I. (2015). Artificial intelligence—The revolution hasn't happened yet Harvard. *Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.f06c6e61>
- Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand forecasting. *Annals of tourism research*, 83, 102925.
- Law, R., G. Li, D. K. C. Fong, and X. Han. (2019). “Tourism Demand Forecasting: A Deep Learning Approach.” *Annals of Tourism Research* 75:410–23.
- Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of tourism research*, 75, 410-423.
- Le Cun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Lee, L., & Yu, J. (2010b). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, 40(5), 255–271
- Li, C., Yang, Y., & Ren, L. (2020). Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species. *Infection, Genetics and Evolution*, 82, 104285.
- Li, G., H. Song, and S. F. Witt. (2006). “Time Varying Parameter and Fixed Parameter Linear AIDS: An Application to Tourism Demand Forecasting.” *International Journal of Forecasting* 22 (1): 57–71.
- Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, 44(1), 82–99.
- Li, G., Wu, D. C., Zhou, M., & Liu, A. (2019). The combination of interval forecasts in tourism. *Annals of Tourism Research*, 75, 363–378.
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301–323. <https://doi.org/10.1016/j.tourman.2018.03.009>
- Li, S., Chen, T., Wang, L., & Ming, C. (2018). Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tourism Management*, 68, 116–126.
- Li, X., and R. Law. (2020). “Forecasting Tourism Demand with Decomposed Search Cycles.” *Journal of Travel Research* 59 (1): 52–68.
- Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management*, 83, 104245.
- Lim, C. (2006). A survey of tourism demand modelling practice: Issues and implications. In L. Dwyer, & P. Forsyth (Eds.). *International handbook on the economics of tourism* (pp. 45–72). Cheltenham: Edward Elgar Publishing.
- Long, W., Liu, C., & Song, H. (2019). Pooling in tourism demand forecasting. *Journal of Travel Research*, 58(7), 1161–1174.
- Lv, H., Yang, Z., Wang, P. L., Ji, G., Song, J., Zheng, L., ... & Xu, Z. J. (2018). A voltage-boosting strategy enabling a low-frequency, flexible electromagnetic wave absorption device. *Advanced Materials*, 30(15), 1706343.
- Ma, X., & Hovy, E. (2016). End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*.
- Mich, L. (2020b). Systems and technologies for fluxes management. In M. Franch & R. Peretta (Eds.), *Tourism, fragilities and emergencies* (pp. 107–134). McGraw Hill.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- Monreale, A. (2020). Rischi etico-legali dell'Intelligenza Artificiale. *DPCE Online*, 44(3).

- Nicholas Wise, & Hadi Heidari (2019). Developing smart tourism destinations with the internet of things. In *Big data and innovation in tourism, travel, and hospitality* (pp. 21–29). Singapore: Springer. https://doi.org/10.1007/978-981-13-6339-9_2
- Padhi, S. S., & Pati, R. K. (2017). Quantifying potential tourist behavior in choice of destination using Google Trends. *Tourism Management Perspectives*, 24, 34–47.
- Palmer, A., Montano, J. J., & Sese, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, 27(5), 781–790.
- Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, 56(7), 957–970
- Pan, B., Wu, D. C., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*, 3(3), 196–210.
- Park, C. W., Han, S. Y., Namgung, H. W., Seo, P. N., Lee, S. Y., & Lee, S. H. (2017). Preparation and characterization of cellulose nanofibrils with varying chemical compositions. *BioResources*, 12(3), 5031–5044.
- Peng, B., H. Song, and G. I. Crouch. (2014). “A Meta-analysis of International Tourism Demand Forecasting and Implications for Practice.” *Tourism Management* 45:181–93.
- Romao, J., & Nijkamp, P. (2019). Impacts of innovation, productivity and specialization on tourism competitiveness – A spatial econometric analysis on European regions. *Current Issues in Tourism*, 22(10), 1150–1169.
- Romao, J., Guerreiro, J., & Rodrigues, P. M. (2017). Territory and sustainable tourism development: A space-time analysis on European regions. *Region*, 4(3), 1–17.
- Shao, Jianli, Xin Liu, and Wenqing He. (2021). "Kernel Based Data-Adaptive Support Vector machines for Multi-Class Classification." *Mathematics* 9 (9): 936. doi:10.3390/MATH9090936.
- Shi, H., Xu, M., & Li, R. (2017). Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Transactions on Smart Grid*, 9(5), 5271–5280
- Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research*, 74, 134–154.
- Socher, R., Bengio, Y., & Manning, C. D. (2012). Deep learning for NLP (without magic). Tutorial abstracts of ACL 2012, ACL '12 (pp. 5). Stroudsburg, PA: Association for Computational Linguistics.
- Song, H., & Hyndman, R. J. (2011). Tourism forecasting: An introduction. *International Journal of Forecasting*, 27(3), 817–821
- Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, 29(2), 203–220.
- Song, H., G. Li, S. F. Witt, and G. Athanasopoulos. (2011). “Forecasting Tourist Arrivals Using Time-Varying Parameter Structural Time Series Models.” *International Journal of Forecasting* 27 (3): 855–69.
- Song, H., Li, G., & Cao, Z. (2017). Tourism and economic globalization: An emerging research agenda. *Journal of Travel Research*. <https://doi.org/10.1177/0047287517734943>.
- Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75, 338–362.
- Song, H., S. F. Witt, and T. C. Jensen. (2003). “Tourism Forecasting: Accuracy of Alternative Econometric Models.” *International Journal of Forecasting* 19 (1): 123–41.
- Song, H., W. C. Gartner, and A. D. A. Tasci. (2012). “Visa Restrictions and Their Adverse Economic and Marketing Implications—Evidence from China.” *Tourism Management* 33 (2): 397–412.
- Sun, S., Y. Wei, K.-L. Tsui, and S. Wang. (2019). “Forecasting Tourist Arrivals with Machine Learning and Internet Search Index.” *Tourism Management* 70:1–10.
- Sun, Z., Vashishth, S., Sanyal, S., Talukdar, P., & Yang, Y. (2019). A re-evaluation of knowledge graph completion methods. *arXiv preprint arXiv:1911.03903*.

- Taleb, I., Serhani, M. A., & Dssouli, R. (2018). Big data quality: A survey. In 2018 IEEE International Congress on Big Data - IEEE BigData Congress 2018: Part of the 2018 IEEE World Congress on Services: 2–7 July 2018, San Francisco, California, USA: Proceedings (pp. 166–173). Piscataway, NJ: IEEE. <https://doi.org/10.1109/bigdatacongress.2018.00029>
- Tetko, I. V., Livingstone, D. J., & Luik, A. I. (1995). Neural network studies. 1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Computer Sciences*, 35(5), 826–833.
- Turner, L. W., and S. F. Witt. (2001). “Forecasting Tourism Using Univariate and Multivariate Structural Time Series Models.” *Tourism Economics* 7 (2): 135–47.
- Wan, S. K., & Song, H. (2018). Forecasting turning points in tourism growth. *Annals of Tourism Research*, 72, 156–167.
- Weih, C., & Ickstadt, K. (2018). Data Science: The impact of statistics. *International Journal of Data Science and Analytics*, 6(3), 189–194. <https://doi.org/10.1007/s41060-018-0102-5>
- Wong, K. K. F., H. Song, and K. S. Chon. (2006). “Bayesian Models for Tourism Demand Forecasting.” *Tourism Management* 27 (5): 773–80.
- WTTC (2019). Travel and tourism: Economic impact 2019. London: World Travel & Tourism Council (WTTC).
- Wu, D. C., H. Song, and S. Shen. (2017). “New Developments in Tourism and Hotel Demand Modeling and Forecasting.” *International Journal of Contemporary Hospitality Management* 29 (1): 507–29.
- Wu, Q., Law, R., & Xu, X. (2012). A sparse Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications*, 39(5), 4769–4774.
- Xiang, Z., & Fesenmaier, D. R. (2017). Big data analytics, tourism design and smart tourism. *Analytics in smart tourism design: concepts and methods*, 299–307.
- Yahya, N. A., R. Samsudin, and A. Shabri. (2017). “Tourism Forecasting Using Hybrid Modified Empirical Mode Decomposition and Neural Network.” *International Journal of Advances in Soft Computing and Its Applications* 9 (1):14–31.
- Yang, Y., & Wong, K. K. (2012). A spatial econometric approach to model spillover effects in tourism flows. *Journal of Travel Research*, 51(6), 768–778.
- Yang, Y., & Zhang, H. (2019). Spatial-temporal forecasting of tourism demand. *Annals of Tourism Research*, 75, 106–119.
- Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization’s web traffic data. *Journal of Travel Research*, 53(4), 433–447.
- Zhang, G., J. Wu, B. Pan, J. Li, M. Ma, M. Zhang, and J. Wang. (2017). “Improving Daily Occupancy Forecasting Accuracy for Hotels Based on EEMD-ARIMA model.” *Tourism Economics* 23 (7): 1496–514.
- Zhang, Y., Li, G., Muskat, B., Law, R., & Yang, Y. (2020). Group pooling for deep tourism demand forecasting. *Annals of Tourism Research*, 82, 102899.
- Zheng, W., Huang, L., & Lin, Z. (2021). Multi-attraction, hourly tourism demand forecasting. *Annals of Tourism Research*, 90, 103271.

ISBN 978-88-99352-98-1



9 788899 352981 >