

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA MAGISTRALE IN
SCIENZE STATISTICHE



**Previsioni giornaliere e clustering stagionale
nel processo di Sales and Operations Planning:
il caso aziendale fischer Italia**

Relatore Prof. Stefano Mazzuco
Dipartimento di Scienze Statistiche

Correlatrice Prof.ssa Mariangela Guidolin
Dipartimento di Scienze Statistiche

Laureanda Marta Polato
Matricola 2048581

Anno Accademico 2023/2024

Indice

Introduzione	1
1 Pianificazione delle Vendite e delle Operazioni	5
1.1 Pianificazione della Domanda	6
1.2 I Quattro Pilastri del processo di S&OP	7
1.3 Input del processo S&OP	8
1.4 Fasi del processo S&OP	10
2 Scomposizione serie storiche	13
2.1 Scomposizione STL	14
2.1.1 Loess	15
2.1.2 Ciclo interno ed esterno	17
2.1.3 Determinazione dei parametri	19
2.2 Scomposizione MSTL	21
2.3 Prophet	22
3 Clustering stagionale	27
3.1 Funzione TSclust	28
3.1.1 Distanza Euclidea	29
3.1.2 Distanza basata sulla forma	30
3.1.2.1 Algoritmo <i>k-Shape</i>	30
3.1.3 Distanza basata sul flusso temporale dinamico	33
4 Caso aziendale fischer Italia	35
4.1 Gruppo fischer	35
4.2 Azienda fischer Italia	36
4.2.1 Pianificazione delle Vendite e delle Operazioni	36
4.2.2 Tecnologia SO99+	37
4.3 Presentazione dataset	39
4.4 Analisi con scomposizioni STL e MSTL	41
4.5 Analisi con metodo Prophet	43
4.5.1 Confronto Prophet e MSTL	45
4.6 <i>Clustering</i> stagionale	47
4.7 Confronti Prophet e SO99+	51
4.7.1 Previsioni un passo in avanti	54

4.7.2	Caso giornaliero	56
4.7.3	<i>Topseller</i>	57
5	Conclusioni	63
	Appendice	67
	Bibliografia	71

Introduzione

Nell'attuale contesto di *business*, molte imprese si trovano ad affrontare sfide sempre più complesse per ottimizzare le proprie strategie di assortimento e soddisfare le esigenze dei clienti in maniera efficace. La natura dinamica del mercato richiede, quindi, un approccio avanzato nel processo di pianificazione delle attività operative. Adottare strategie flessibili e tempestive per garantire un equilibrio ottimale tra domanda e offerta è uno dei fattori chiave. In questo scenario, la gestione della domanda rappresenta un importante elemento di competitività. La selezione del metodo di sviluppo delle previsioni ed il controllo delle stagionalità dei prodotti sono parte integrante di questo processo.

Il presente lavoro si focalizza sull'analisi dei dati forniti dall'azienda *fischer Italia*. Nello specifico, si esamina la serie storica della quantità di merce ordinata dai clienti, con particolare attenzione alla famiglia di prodotto principale: il nylon. Attualmente, l'azienda elabora le proprie previsioni avvalendosi del programma *SO99+* che, tuttavia, presenta alcune limitazioni piuttosto rilevanti: la capacità di analisi del sistema si estende solamente ad un periodo storico di 5 anni ed il livello di dettaglio elaborato è mensile. Questo impone dei vincoli sulla granularità dell'analisi, permettendo la gestione di un solo tipo di stagionalità. Ciò limita il rilevamento di eventuali variazioni più specifiche e dettagliate nel comportamento dei dati nel corso del tempo.

L'azienda *fischer Italia* è costantemente impegnata nella ricerca di un approccio per ottimizzare il livello di dettaglio delle previsioni e per il raggruppamento dei propri prodotti in base alla loro stagionalità giornaliera. Una volta risolta tale questione, sarà d'interesse andare a sostituire le attuali previsioni mensili con un *forecast* giornaliero.

Il primo capitolo del presente lavoro è rivolto alla presentazione approfondita del processo di pianificazione delle vendite e delle operazioni, esplorando in dettaglio gli

elementi diversificati che contribuiscono alla sua struttura. Questa introduzione è utile per la comprensione del contesto generale in cui si inserisce l'intera analisi.

Il secondo capitolo si concentra sull'introduzione di approcci specifici finalizzati a migliorare le previsioni dei dati, ponendo particolare enfasi sulla gestione delle fluttuazioni stagionali. Dopo una panoramica dei diversi metodi esistenti, si approfondisce l'applicazione della scomposizione STL, una metodologia efficace nella gestione di dati giornalieri. Un'attenzione particolare sarà riservata al contesto di stagionalità multipla (MSTL), sottolineando come questo metodo possa essere implementato per affrontare le sfide legate a modelli di stagionalità più complessi. All'interno dello stesso capitolo, si presentano i punti chiave di un secondo metodo di scomposizione: Prophet. Questo modello offre un approccio innovativo e versatile alla gestione della stagionalità, specialmente nei casi con diverse fluttuazioni periodiche. Lo scopo è quello di fornire una panoramica completa delle opzioni disponibili per la gestione dei dati giornalieri, caratterizzati da fluttuazioni stagionali singole e multiple.

L'introduzione degli approcci STL e Prophet ha l'obiettivo di valutare le opzioni disponibili per ottimizzare la gestione delle fluttuazioni stagionali nei dati aziendali. Questo assume una rilevanza particolare nell'ottica di un passaggio ad un livello di previsione più dettagliato, trasformando l'analisi da mensile a giornaliera.

Nel terzo capitolo di questa ricerca, l'attenzione si sposta sull'illustrazione di diverse metodologie per il raggruppamento delle serie storiche, con particolare enfasi sulle dinamiche stagionali. Questo approfondimento è giustificato dal contesto aziendale delineato precedentemente, dove la gestione delle fluttuazioni stagionali assume un ruolo di fondamentale importanza. Si esaminano i due principali approcci di raggruppamento: gerarchico e di partizione. Questi metodi consentono l'organizzazione delle serie storiche in gruppi omogenei, facilitando l'identificazione di pattern e comportamenti simili nei dati. Per il processo di creazione dei gruppi, si presta attenzione a tre distanze: la distanza euclidea, la distanza basata sulla forma e la distanza basata sul flusso temporale dinamico. Queste misure di distanza sono oggetto di un'analisi dettagliata per comprendere il loro impiego nella valutazione della somiglianza tra le serie storiche.

Dopo aver introdotto il contesto e le diverse tecniche per affrontare le sfide aziendali, l'attenzione si sposta sull'applicazione pratica. Nel quarto capitolo si presenta il

quadro generale del Gruppo fischer e di fischer Italia, per poi approfondire il processo di Pianificazione della Domanda e delle Operazioni nel contesto aziendale specifico. Si confrontano le prestazioni dei metodi STL e Prophet, sia con una che con due stagionalità, utilizzando la metrica RMSE. Successivamente, si valutano i diversi metodi di *clustering* utilizzando le tre distanze presentate, analizzando i diversi casi sulla base dell'indice *silhouette*.

I risultati del metodo proposto che si dimostra più competitivo vengono confrontati con il sistema aziendale SO99+. Quest'ultimo algoritmo viene esaminato sia nel caso standard, sia arricchito degli indici stagionali ottenuti tramite il processo di *clustering*. I confronti vengono condotti a livello mensile, settimanale e giornaliero. È rilevante notare che il caso giornaliero è presentato esclusivamente per il metodo proposto migliore, poiché il sistema aziendale attuale, come già premesso, non raggiunge un livello di dettaglio così specifico. Come metrica di confronto si utilizza l'accuratezza previsiva, metrica di riferimento nell'ambito aziendale.

Un dettaglio rilevante da considerare durante la lettura del presente elaborato riguarda l'uso della parola "fischer": quando ci si riferisce all'azienda, sarà indicata con la lettera minuscola, perchè con la lettera maiuscola ci si riferirà al professor Klaus Fischer, figlio del fondatore, e attuale titolare e presidente del gruppo.

Capitolo 1

Pianificazione delle Vendite e delle Operazioni

Il processo di Pianificazione delle Vendite e delle Operazioni, comunemente definito *Sales and Operations Planning* (S&OP), rappresenta una metodologia aziendale fondamentale volta a sincronizzare le attività di produzione e vendita per ottimizzare l’allocazione delle risorse e garantire una gestione efficiente della catena di approvvigionamento. Nei contesti aziendali sempre più complessi e dinamici, tale processo di pianificazione emerge come un approccio strategico che mira a bilanciare la domanda del mercato con la capacità operativa, assicurando un allineamento efficace tra le diverse funzioni aziendali.

Il metodo introdotto da Richard Ling, presso la Oliver Wight Company verso la fine degli anni ottanta, ha acquisito rapidamente notorietà come approccio integrato, essenziale per armonizzare le attività all’interno delle organizzazioni. Il procedimento di S&OP è finalizzato alla formulazione di piani strategici che conferiscano alla gestione aziendale la capacità di guidare le proprie operazioni per conseguire un vantaggio competitivo in maniera continua. L’approccio consiste, quindi, nell’unire tutti i piani funzionali dell’azienda, compresi quelli di vendita, marketing, sviluppo, produzione, approvvigionamento e finanza, in un set integrato e coeso.

Tutto ciò presuppone che l’azienda disponga di un *business plan*, della possibilità di stimare i quantitativi di vendita, di un chiaro obiettivo in termini di redditività, produttività e tempi di consegna ai clienti, e di un processo di pianificazione che considera

le future richieste dei clienti come input primario.

In generale, l'implementazione di tale processo comporta i seguenti vantaggi.

- Favorire un miglior servizio ai clienti, alleggerire il magazzino e contribuire al raggiungimento degli obiettivi finanziari comportando una maggiore redditività
- Ripianificare e riaggiustare i piani operativi laddove necessario grazie agli strumenti di simulazione che offrono un supporto decisionale
- Misurare le prestazioni aziendali
- Migliorare il lavoro di squadra promuovendo la collaborazione interdipartimentale per consentire l'elaborazione di piani allineati in ogni area
- Avere una visione del medio e lungo periodo
- Integrare operativamente le diverse funzioni aziendali
- Incrementare l'efficienza operativa e strategica.

1.1 Pianificazione della Domanda

I protagonisti principali di tale contesto sono gli *stakeholder* interni che, attraverso l'utilizzo di specifici supporti tecnologici, collaborano al raggiungimento degli obiettivi di soddisfazione circa la disponibilità del mix di prodotti, massimizzando le vendite ed i profitti. Tra i responsabili di funzione ha un ruolo strategico il Pianificatore della Domanda (*Demand Planner*).

Questa figura professionale è incaricata di gestire, analizzare e prevedere lo stock di magazzino di un'azienda. Dato il suo incarico determinante, solitamente opera in collaborazione con diversi altri professionisti dell'impianto produttivo.

Gli obiettivi di tale incarico consistono nel miglioramento del livello di servizio al cliente, nella riduzione delle scorte di magazzino e dei costi d'acquisto, nell'ottimizzazione delle scorte stagionali e delle attività di distribuzione delle risorse produttive e logistiche. A tal proposito è necessario effettuare alcune azioni preliminari.

- Elaborare delle curve previsioni di vendita
- Interpretare correttamente le variazioni dei volumi di vendita
- Garantire un'efficiente gestione del ciclo di vita dei prodotti e adeguare i livelli di stock e la capacità produttiva

- Coordinare gli uffici coinvolti.

1.2 I Quattro Pilastri del processo di S&OP

Il successo del processo della Pianificazione delle Operazioni e delle Vendite si basa sulla comprensione dei Quattro Pilastri (*Four Fundamentals*): Domanda, Fornitura, Volume e Mix. Questi quattro, infatti, sono le fondamenta su cui si basa tutta l'attività di pianificazione.

Con il termine *Demand* si fa riferimento al fattore trainante del processo S&OP. La domanda determina le risorse, la capacità e il livello di inventario necessarie all'azienda per riuscire a soddisfare le esigenze dei clienti. Risulta necessario riuscire a cogliere tempestivamente i cambiamenti nel tempo, quali possono essere le tendenze stagionali, l'impatto delle attività promozionali e il ciclo di vita di un prodotto. Tale comportamento è dettato dalla strategia *Demand-driven* su cui si basa il mondo attuale: è la domanda ad influenzare la produzione.

Il secondo elemento fondamentale è definito *Supply*. È importante pianificare dettagliatamente la quantità di materiale e forza lavoro necessaria al fine di riuscire a soddisfare la domanda dei clienti interni ed esterni. In questo ambito vengono considerati, ad esempio, i costi, i tempi di consegna e tutti i vincoli di capacità, tecnologici e produttivi che costituiscono la catena di approvvigionamento.

Una situazione ideale prevede l'equilibrio tra domanda e fornitura, evitando così l'insoddisfazione di clienti e fornitori. Il prevalere di una componente sull'altra comporta diversi imprevisti riscontrabili nell'aumento dei costi, nella perdita di qualità e nel peggioramento del livello di servizio.

L'obiettivo del metodo S&OP consiste nell'identificare e nel mantenere un bilanciamento ottimale tra le due grandezze, principalmente attraverso due strategie. Un primo accorgimento essenziale consiste nell'essere in grado di anticipare la domanda futura nel modo più accurato possibile sulla base di un adeguato processo di pianificazione della domanda. A tale approccio si affianca una strategia di inclusione della variabilità della domanda nella fase *Supply*, al fine di gestire in modo efficace eventuali variazioni.

Per riuscire a raggiungere e mantenere l'equilibrio tra le prime due quantità dei Quattro Pilastri si introducono due nuovi elementi: Volume e Mix.

Il Volume rappresenta il livello più aggregato, in cui vengono considerate le famiglie di prodotti e non i singoli articoli. In questo contesto, l'azienda considera il Volume come indicatore di previsioni aggregate, che fornisce un quadro generale della domanda e della produzione nel medio-lungo periodo (da 18 a 36 mesi) con scansione mensile. In questo modo si riescono ad ottenere risultati più promettenti in termini di prestazioni dal momento che risulta più agevole focalizzarsi in un quadro più aggregato per quanto riguarda gli indicatori economici, i competitor, le modifiche delle quote di mercato e il lancio di nuovi prodotti.

Un livello più dettagliato è riscontrabile nel Mix, in cui viene considerato il singolo prodotto o ordine. Il periodo temporale considerato per l'applicazione di tale indicatore rientra nel breve termine (da 1 a 3 mesi), con frequenza di analisi settimanale o giornaliera. L'utilizzo del Mix offre un quadro altamente dettagliato e granulare della situazione, anziché una visione aggregata d'insieme.

1.3 Input del processo S&OP

Gli input del processo di Pianificazione delle Vendite e delle Operazioni sono suddivisibili in base all'elemento dei Quattro Pilastri di riferimento.

La componente principale dal lato della domanda è identificabile nelle previsioni di vendita (*Sales Forecasting*). I valori predetti possono essere formulati per qualsiasi tipo di attività aziendale e sono di competenza dei reparti Vendite e Marketing che, oltre a gestire il prezzo, hanno una conoscenza approfondita degli incentivi della forza vendita e dell'andamento del mercato. Queste due funzioni sono responsabili di sviluppare il piano di produzione e il piano di esecuzione sulla base della loro conoscenza della domanda migliorando, quindi, l'eventuale piano previsivo statistico proposto come punto di partenza. Per ottenere un livello di previsioni soddisfacente è consigliabile che il settore *Demand* (Vendite e Marketing) fornisca alle *Operations* (Acquisti, Logistica, Pianificazione e Produzione) una stima sufficientemente approssimata. Questa stima deve consentire la definizione delle capacità e degli ordini di materiali. A questo

compito, si affianca l'obiettivo di riduzione dell'errore di previsione che deve essere in continuo miglioramento. Per ottenere un risultato previsivo valido sulla base delle "4 R" è necessario ottenere un risultato Ragionevole, Realistico, Rivisto di frequente e Rappresentativo della domanda totale (Wallace, 2004).

Un'ulteriore caratteristica da esaminare in ambito previsivo è il livello di aggregazione. Un *forecast* troppo aggregato o troppo dettagliato non fornisce un aiuto concreto al calcolo in esame. In generale, il livello di dettaglio più appropriato è quello di "famiglia" o "sotto-famiglia" (suddivisioni più dettagliate dell'aggregazione totale dell'azienda e per *business unit*).

Infine, la frequenza più indicata per rivedere il calcolo previsivo è di una volta al mese.

Per quanto riguarda il lato della fornitura, il principale elemento di input è la capacità produttiva. L'esecuzione delle attività pianificate può essere messa in discussione dalla mancanza di tale risorsa. È indispensabile, quindi, dotarsi di uno strumento per il calcolo del fabbisogno di capacità. Tale controllo è inutile se contemporaneamente, soprattutto in realtà soggette a stagionalità, non si disponga di strumenti per garantire la flessibilità produttiva interna o esterna.

Il ruolo delle *Operations* è quello di accertare se è possibile soddisfare la domanda con le risorse a disposizione, verificando l'affidabilità del piano di approvvigionamento e produzione (detto Piano delle Operazioni). Il passaggio per sviluppare quest'ultimo programma per ogni suddivisione di prodotto di riferimento, partendo dalla previsione della domanda futura, si concretizza nella pianificazione approssimativa della capacità produttiva (*Rough-Cut Capacity Planning*). Questo strumento consente di relazionare la capacità richiesta e la capacità effettiva e, di conseguenza, di effettuare le modifiche richieste a livello di attrezzatura e di personale. Nello specifico, per progettare il piano con criterio è necessario redarre la *Bill of Resources*, una matrice in cui sono correlate le segmentazioni di prodotto con le rispettive risorse necessarie alla produzione.

In generale, è importante definire gli obiettivi e i *target* della strategia *Demand/-Supply* per avere una direzione da seguire per favorire un miglioramento continuo. È necessario, in questo contesto, definire l'approccio operativo: produrre sulla base della previsione futura, procedere con la produzione in corrispondenza dell'arrivo dell'ordine (con attività di progettazione già eseguita), oppure assemblare il prodotto una volta

raccolto l'ordine (con componenti dell'oggetto già fabbricate). Va, inoltre, stabilito qual è il livello di servizio medio che si vuole garantire al cliente, il livello di giacenze desiderato e il livello di ordini arretrati.

Ogni strategia comporta una diversa gestione delle scorte di sicurezza e di reazione agli ordini arretrati.

1.4 Fasi del processo S&OP

La pianificazione delle Vendite e delle Operazioni è ciclica ed affronta una successione di fasi realizzate su base mensile in accordo con le decisioni strategiche adottate dall'organizzazione. La creazione di un processo S&OP efficace richiede pianificazione e coordinamento tra tutti i dipartimenti, per predisporre nel breve-medio termine le attività di produzione, distribuzione e integrazione operativa con i fornitori di materiali. Tale pianificazione stabilisce un collegamento tra le funzioni orientate alla domanda, orientate alla fornitura e, in maniera marginale, orientate alle funzioni di supporto (Finanza e Risorse Umane).

Il processo prevede il susseguirsi di cinque fasi principali.

Lo stadio iniziale si concentra sulla raccolta dei dati: si acquisiscono le informazioni sulle prestazioni del piano in atto, vengono analizzate le tendenze, si consolidano i dati storici sulle vendite e gli approfondimenti dei team di vendita. Questo step include una revisione del mix di prodotti dell'organizzazione e dei fattori esterni che possono influenzare l'operato aziendale. Queste operazioni sono necessarie per procedere al calcolo delle previsioni di vendita: l'obiettivo principale consiste nel generare una previsione il più precisa possibile. Sviluppare un quadro completo e accurato della situazione attuale dell'organizzazione è il punto di partenza per ottenere il più precisamente possibile la potenziale domanda futura. Di conseguenza, una proprietà essenziale consiste nel disporre di un programma per rispondere ai cambiamenti delle esigenze del mercato o dei clienti, per riuscire a creare un piano operativo e di vendita realistico.

Questa prima fase di raccolta e revisione dei dati ha inizio al termine del mese e rappresenta il punto di partenza per il mese successivo.

La fase seguente al calcolo delle previsioni statistiche delle vendite consiste nell'integrazione delle informazioni interne ed esterne all'organizzazione. Questa fase di pianificazione della domanda interessa diverse funzioni aziendali. Il coinvolgimento umano rappresenta l'attributo determinante di questa operazione: vengono integrate le competenze logistiche e produttive, la conoscenza del mercato e dei comportamenti di acquisto dei consumatori nei segmenti in cui l'azienda opera.

La fase di pianificazione comporta l'analisi delle proiezioni iniziali, seguita dalla creazione di un piano della domanda che stabilisce chiaramente la quantità, il momento e il luogo in cui la domanda deve essere soddisfatta. Le organizzazioni devono esaminare regolarmente le previsioni, poiché possono subire cambiamenti rapidi. È importante integrare approfondimenti strutturali sulla domanda futura e identificare rischi strategici che potrebbero non essere completamente riflessi nelle previsioni di base, come, ad esempio, variazioni attese dovute a iniziative di marketing.

Il livello di aggregazione con cui effettuare la previsione è a discrezione dell'azienda, per i passaggi successivi invece è opportuno focalizzarsi a livello di singolo articolo. È conveniente, comunque, includere nel calcolo della previsione sia i prodotti esistenti sia i nuovi prodotti.

Soluzioni di successo per la previsione e la pianificazione della domanda permettono e facilitano il calcolo automatico della previsione della domanda di base, la pianificazione della promozione, la previsione causale della domanda, la previsione della domanda su più livelli e la previsione consensuale della domanda.

La terza fase del processo di *Sales and Operations Planning* si concretizza nella pianificazione della fornitura: *supply planning*. Una scelta logica consiste nell'applicazione di questo passaggio per ciascun stabilimento o linea produttiva. Questa strategia risulta appropriata dal momento che molti prodotti finiti sono realizzati con componenti con linee di produzione in comune e non. La pianificazione in questo ambito prevede la stesura del piano di produzione, del piano di acquisti e del piano di distribuzione. Viene riportato nel dettaglio l'utilizzo delle risorse e vengono evidenziati eventuali conflitti, vincoli di risorse, impossibilità di soddisfare i fabbisogni e i livelli di servizio che ne risultano. Il piano si traduce in una proiezione delle scorte sull'intero orizzonte temporale e nello sviluppo di un piano di fornitura che soddisfi le previsioni della domanda.

Una volta organizzati i piani della domanda e dell'offerta, si svolge una fase di riesame e riconciliazione degli stessi. Il pre-S&OP *meeting* è l'occasione d'incontro tra le principali parti interessate provenienti da diversi dipartimenti: responsabili dello Sviluppo Prodotto, pianificatori della catena di approvvigionamento, addetti del reparto Finanza e diversi rappresentanti dell'ambito della Domanda.

La discussione si pone come obiettivo l'allineamento tra tutti i dipartimenti sui piani e sugli obiettivi generali: si cerca di risolvere eventuali lacune e discrepanze nelle strategie, proponendo soluzioni alle criticità riscontrate nei passaggi precedenti. Il confronto permette di raggiungere il consenso sugli eventuali correttivi da apportare per utilizzare meglio le risorse, soddisfare la domanda attesa e realizzare gli scopi e gli obiettivi finanziari.

L'*executive meeting* rappresenta la conclusione del processo. Si tratta dell'incontro tra i rappresentanti di ciascuna funzione, il responsabile del processo di S&OP e il presidente dell'azienda. Durante la riunione vengono presentati i piani ottimizzati e finalizzati affinché possano essere approvati e rilasciati per l'implementazione. La riunione esecutiva offre ai dirigenti l'opportunità di finalizzare i piani, monitorare i progressi e apportare le modifiche necessarie. Per garantire il successo, la valutazione dell'S&OP dovrebbe essere regolare.

Capitolo 2

Scomposizione serie storiche

Nel contesto precedentemente analizzato, la scomposizione di una serie storica si configura come un'operazione cruciale nell'ambito previsivo. Le serie possono manifestare molteplici schemi di pattern, risulta quindi utile e vantaggioso suddividerle in diverse componenti. Questo procedimento viene effettuato sia per avere una maggiore comprensione della serie, sia per migliorare l'accuratezza previsiva.

La scomposizione classica di una serie storica è definita come:

$$Y_t = T_t + S_t + R_t, \quad t = 1, 2, \dots, N, \quad (2.1)$$

dove Y_t rappresenta l'osservazione al tempo t , T_t descrive il trend della serie, S_t esprime la stagionalità e R_t costituisce la componente residua.

In alternativa, si può definire una scomposizione moltiplicativa:

$$Y_t = T_t \times S_t \times R_t, \quad t = 1, 2, \dots, N \quad (2.2)$$

Nel primo caso, viene calcolato l'andamento del trend attraverso l'applicazione di medie mobili. Il risultato viene sottratto dalla serie storica di partenza per ottenere la serie detrendizzata. La stima della stagionalità, poi, si ottiene definendo la media dei valori detrendizzati raggruppati in base alla stagionalità, verificando che la somma di queste sia nulla. Infine, la componente residua viene calcolata sottraendo dalla serie storica di partenza le due quantità di interesse appena definite.

Nel caso di una scomposizione moltiplicativa, invece, la procedura differisce in due passaggi. L'eliminazione del trend e il calcolo della componente residua avvengono attraverso l'operazione di divisione: rispettivamente y_t/\hat{T}_t e $\hat{R}_t = y_t/(\hat{T}_t \times \hat{S}_t)$.

Tuttavia, tali procedimenti presentano alcune limitazioni.

- La stima del trend non è disponibile per le prime e le ultime osservazioni, rendendo impossibile il calcolo della componente residua in corrispondenza di tali istanti temporali
- La stima del trend tende ad attenuare i rapidi aumenti e le diminuzioni della serie
- Si assume che la stagionalità si ripeta ugualmente di anno in anno
- Può accadere che i valori delle serie storiche composte da pochi periodi risultino insoliti.

A tal proposito, il metodo *X-11* rappresenta un'alternativa per la scomposizione di una serie storica. Si tratta di una tecnica iterativa di stima delle diverse componenti che prevede l'utilizzo, ad ogni passo, di appropriate medie mobili. A differenza del caso precedente, questa scomposizione consente di ottenere delle stime relative al trend per tutte le osservazioni, e prevede maggior flessibilità per la componente stagionale. Questa, infatti, può leggermente variare tra i diversi momenti.

La scomposizione *X-11* risulta altamente robusta in caso di *outliers* e di cambiamenti nella serie. Tuttavia, la procedura presenta alcune restrizioni, quali la complessità in termini di implementazione, la mancata possibilità di trattare i dati mancanti e la condizione di poter analizzare solo serie storiche con stagionalità mensile o trimestrale.

Un comportamento simile è riscontrabile anche nella procedura *SEATS* (Seasonal Extraction in ARIMA Time Series).

2.1 Scomposizione STL

Una possibile soluzione alle problematiche presentate precedentemente è la scomposizione STL (*Seasonal and Trend decomposition using Loess*) (Cleveland et al., 1990).

L'obiettivo principale di tale scomposizione è suddividere la serie storica nelle relative tre componenti: trend, stagionalità e parte residua.

L'utilizzo di tale metodologia fa riferimento a particolari vantaggi, meno riscontrabili in altre tecniche. La scomposizione STL, infatti, presenta alcune caratteristiche favorevoli.

- Semplicità nell'implementazione e nell'utilizzo
- Velocità di calcolo, anche per serie storiche particolarmente estese
- Capacità di scomporre serie storiche che presentano valori mancanti
- Flessibilità nella specificazione del numero di variazioni nelle componenti di trend e stagionalità
- Robustezza delle stime delle componenti di trend e stagionalità rispetto a valori anomali e/o comportamenti atipici della serie
- Controllo da parte dell'utente riguardo il grado di lisciamiento del trend
- Possibilità di specificare qualsiasi numero di osservazioni per ciclo di stagionalità, consentendo quindi di poter gestire qualsiasi specificazione di quest'ultima
- Trattazione di serie storiche la cui componente stagionale varia nel corso del tempo, con relativa possibilità per l'utente di controllare tale alterazione.

2.1.1 Loess

La scomposizione STL consiste in una sequenza di operazioni di lisciamiento, ciascuna delle quali utilizza lo stesso metodo: la regressione locale pesata o loess (Azzalini & Scarpa, 2009).

La regressione loess (*Locally weighted least squares regression*) è una tecnica che integra la semplicità del metodo dei minimi quadrati con la flessibilità delle metodologie non lineari. Si tratta di una soluzione non parametrica che non richiede, quindi, nessuna specifica struttura della relazione tra la variabile dipendente e quelle indipendenti.

Una variazione alla tradizionale regressione locale consiste nel considerare un'ampiezza di banda non costante lungo l'asse delle x . Questa ampiezza variabile è determinata dalla sparsità dei dati. Nelle aree in cui vi è un minor numero di osservazioni, verrà data meno importanza alle oscillazioni, viceversa, nel caso di una maggior concentrazione verrà maggiormente considerata la variabilità dei punti per stimare una curva che segua al meglio i dati. Dal momento che la variabilità della stima dipende inversamente dalla

densità delle osservazioni, risulta vantaggioso permettere che l'ampiezza di banda sia variabile in base alla concentrazione locale delle x .

La stima non lineare ottenuta attraverso il loess viene calcolata prefissando una percentuale costante di punti rilevanti da includere nel kernel. Tale quantità definisce il parametro di lisciamiento.

Un'ulteriore importante caratteristica riguarda il fatto di poter utilizzare un metodo di stima locale robusto, in modo tale da evitare che le osservazioni anomale (*outliers*) influenzino la stima. Viene scelto, quindi, un nucleo a supporto limitato tricubico.

La curva di regressione $\hat{g}(x)$ può essere calcolata per qualsiasi valore di x , permettendo quindi di avere una stima del loess ovunque, e non solo in corrispondenza delle osservazioni. Si tratta di un'importante conclusione per consentire alla procedura STL di trattare anche valori mancanti.

Per la stima di $\hat{g}(x)$ viene scelto un intero positivo q , con $q \leq n$. Vengono, quindi, considerati i q valori delle x_i più vicini a x , assumendo che in valori vicini di x_i la variabile dipendente si comporti in maniera abbastanza simile. Risulta utile, a questo punto, pesare i punti in base alla loro distanza da x : viene pesato maggiormente il contributo delle x_i vicine attraverso una funzione peso.

Definita $\lambda_q(x)$ la distanza della q -esima x_i più lontana da x , sia W la funzione peso tricubica:

$$W(u) = \begin{cases} (1 - u^3)^3 & 0 \leq u < 1 \\ 0 & u \geq 1, \end{cases} \quad (2.3)$$

il peso per ciascuna x_i è

$$v_i(x) = W\left(\frac{|x_i - x|}{\lambda_q(x)}\right) \quad (2.4)$$

Tali pesi, quindi, decrescono man mano che ci si allontana dalla x : al punto più lontano viene assegnato peso nullo. Il passo finale consiste nell'adattare un polinomio di grado d ai dati con peso $v_i(x)$ in corrispondenza delle coordinate (x_i, y_i) .

Nel caso in cui venga scelto $q > n$, si definisce $\lambda_n(x)$ come la distanza tra x e il punto più lontano x_i e

$$\lambda_q(x) = \lambda_n(x) \frac{q}{n} \quad (2.5)$$

La regressione non parametrica loess prevede la definizione delle quantità d e q . All'aumentare della percentuale di dati considerati (q), la curva $\hat{g}(x)$ risulta più liscia. Per q che tende all'infinito, $v_i(x)$ tende a uno e $\hat{g}(x)$ tende ad un adattamento polinomiale di ordine d tramite i minimi quadrati ordinari.

2.1.2 Ciclo interno ed esterno

La scomposizione STL si compone di due procedure ricorsive: un ciclo interno ed uno esterno. Il primo si concentra sulla componente stagionale e di trend, mentre quello esterno minimizza l'effetto dei termini di errore.

Durante ogni iterazione del ciclo interno viene eseguito un processo di lisciamiento stagionale per stimare e aggiornare tale componente della serie storica. Ciò aiuta ad identificare ed isolare i pattern periodici. Successivamente, viene effettuato lo stesso procedimento considerando la componente di trend, per individuare e catturare le variazioni a lungo termine della serie. Ogni iterazione completa del ciclo interno si compone di $n_{(i)}$ passi.

Ad ogni ciclo esterno si procede con l'iterazione interna seguita dal calcolo di pesi di robustezza; tali pesi vengono poi impiegati nella successiva esecuzione del ciclo interno per mitigare l'influenza e l'impatto di comportamenti transitori e anomali sulle componenti di trend e stagionalità. I pesi vengono inizializzati al valore 1, e successivamente vengono eseguiti gli $n_{(o)}$ passaggi del ciclo esterno.

Vengono definite $S_v^{(k)}$ e $T_v^{(k)}$, per $v = 1, \dots, N$, rispettivamente la componente stagionale e di trend al termine del k -esimo passo. Tali quantità sono definite ad ogni istante, anche quelli in cui la serie presenta dati mancanti. L'aggiornamento al passo $(k + 1)$ delle due entità in considerazione viene presentato di seguito.

Il ciclo interno si compone di una prima fase in cui la serie storica viene detrendizzata: $Y_v - T_v^{(k)}$. Come valore iniziale di T_v si nota come $T_v^{(0)} \equiv 0$ permetta di ottenere risultati adeguati (Cleveland et al., 1990). La serie detrendizzata non assume valori negli istanti temporali in cui la serie originale non presenta osservazioni.

Successivamente, viene considerata ciascuna sotto-serie dei dati detrendizzati, alla quale viene adattata la procedura di lisciamiento utilizzando il loess con $q = n_{(s)}$ e $d = 1$. La specificazione di $n_{(s)}$ verrà esplicitata nel seguito. La procedura di lisciamiento

viene applicata per tutti gli istanti temporali delle sotto-serie, compresi quelli con valori mancanti, e per le posizioni appena prima della prima sottoserie e subito dopo l'ultima. L'insieme dei valori lisciati delle sotto-serie è definito $C_v^{(k+1)}$.

Il terzo passaggio prevede l'applicazione di un *low-pass filter* alla quantità $C_v^{(k+1)}$ per ridurre ulteriormente le variazioni ad alta frequenza e ottenere una stima più robusta della componente ciclica. Il filtro elimina le variazioni veloci o turbolente, concentrandosi sulle variazioni cicliche a frequenze più basse. Viene utilizzata, quindi, una media mobile di lunghezza $n_{(p)}$, seguita da un'ulteriore media mobile della medesima lunghezza, seguita da una terza di lunghezza 3, per concludere con una procedura di lisciamiento utilizzando il loess con parametri $d = 1$ e $q = n_{(l)}$. La scelta di $n_{(l)}$ verrà discussa in seguito. L'output $L_v^{(k+1)}$ è definito per ogni istante per $v = 1$ fino a N .

La fase successiva consiste nel detrendizzare le sotto-serie lisciate, ottenendo $S_v^{(k+1)} = C_v^{(k+1)} - L_v^{(k+1)}$.

La serie Y_v viene, quindi, destagionalizzata sottraendo la quantità $S_v^{(k+1)}$. Anche in questo caso, la serie ottenuta non assume valori negli istanti temporali in cui la serie originale non presenta osservazioni.

Infine, l'ultimo passaggio consiste nel lisciamiento della serie ottenuta allo *step* precedente utilizzando i parametri $q = n_{(t)}$ e $d = 1$. La procedura viene applicata anche negli istanti temporali in cui vi sono valori mancanti.

Una volta ottenute le stime T_v e S_v tramite l'esecuzione iniziale del ciclo interno, si definisce la componente residua come:

$$R_v = Y_v - T_v - S_v \quad (2.6)$$

Tale quantità non è calcolabile nei punti in un cui la serie ha valori mancanti.

Vengono, successivamente, costruiti dei pesi robusti per ogni osservazione di Y_v che riflettono quanto R_v sia estremo. Si definisce h come mediana del valore assoluto di R_v .

Il peso al punto v risulta

$$\rho_v = B(|R_v|/h) \quad (2.7)$$

dove B è la funzione peso quadratica

$$B(u) = \begin{cases} (1 - u^2)^2 & 0 \leq u < 1 \\ 0 & u \geq 1 \end{cases} \quad (2.8)$$

A questo punto viene ripetuto il ciclo interno, moltiplicando il peso ρ_v . Le iterazioni del ciclo esterno vengono ripetute per un totale di $n_{(o)}$ volte. La componente di trend, inizialmente posta pari a zero, assumerà il valore ottenuto all'ultimo passaggio del ciclo interno.

2.1.3 Determinazione dei parametri

La scomposizione STL prevede la determinazione di 6 parametri. La nomenclatura riportata segue quella indicata in Cleveland et al. (1990).

- $n_{(p)}$: il numero di osservazioni in ogni ciclo della componente stagionale. Tale parametro dipende dalle caratteristiche dell'applicazione
- $n_{(i)}$: il numero di passaggi del ciclo interno. Considerando il caso in cui $n_{(o)} = 0$, si vuole scegliere un valore per $n_{(i)}$ tale da permettere agli aggiornamenti delle componenti di trend e di stagionalità di convergere. Per ottenere tali risultati è sufficiente porre $n_{(i)}$ pari a 1 o 2 (Cleveland et al., 1990).

Nel caso in cui sia necessaria una stima robusta, e che quindi $n_{(o)} \neq 0$, risulta adeguato considerare un solo passaggio per il ciclo interno. Questa scelta risulta valida per motivi legati alla velocità di convergenza, e perchè non risulta vantaggioso o redditizio concentrarsi eccessivamente su una fase specifica del processo iterativo, poiché questo potrebbe non portare a benefici significativi nella convergenza generale del processo di ottimizzazione.

Come regola per valutare la convergenza si può utilizzare il seguente criterio: definite $U_v^{(k)}$ e $U_v^{(k+1)}$ come iterazioni successive della componente di trend o stagionale, $U_v^{(k)}$ raggiunge convergenza se:

$$\frac{\max_v |U_v^{(k)} - U_v^{(k+1)}|}{\max_v U_v^{(k)} - \min_v U_v^{(k)}} < 0.01 \quad (2.9)$$

- $n_{(o)}$: il numero di iterazioni del ciclo esterno. Nel caso in cui non sia richiesta una stima robusta, tale parametro viene considerato nullo. In questo caso, quindi, la scomposizione si concentra solo sul ciclo interno.

Nel caso in cui venga richiesta la robustezza dei valori di stima, si ricerca un valore tale da garantire la convergenza delle stime di trend e stagionalità. Considerando, come descritto in precedenza, $n_{(i)} = 1$, si può dimostrare che per ottenere la convergenza basta porre $n_{(o)}$ pari a 5 o 10 (Cleveland et al., 1990)

- $n_{(l)}$: il parametro di lisciamiento per il *low-pass filter*. Questo risulta pari all'intero più piccolo maggiore o uguale a $n_{(p)}$. Questa scelta semplifica l'implementazione computazionale, contribuisce a una scomposizione più fluida e consente al filtro loess di essere adattato in modo simile sia alle variazioni di trend che a quelle stagionali
- $n_{(t)}$: il parametro di lisciamiento per la componente di trend. Un aumento di tale valore comporta una minor capacità di estrazione del trend e, perciò, si otterrà una curva più liscia. Tale quantità assume sempre valori dispari.

È importante ricordare che il trend gioca un ruolo fondamentale nella stima della componente stagionale.

La prima funzione riguarda l'eliminazione delle variazioni di lungo periodo, raggiungibile considerando un valore di $n_{(t)}$ non troppo elevato. Il secondo vantaggio riguarda l'iterazione robusta. Il calcolo dei pesi robusti, infatti, si basa sulla grandezza della componente residua. In generale, non si vuole permettere che effetti a bassa frequenza, ovvero variazioni o tendenze che si verificano su una scala temporale più lunga, finiscano nei residui. Questo perchè si vuole dare minor peso solo a comportamenti estremi e transitori. A tal proposito, si sceglie un valore basso di $n_{(t)}$, tale da garantire che trend e stagionalità non competano per spiegare la variazione dei dati.

Come regola generale viene scelto il più piccolo valore intero dispari che soddisfa tale criterio:

$$n_{(t)} \geq \frac{1.5n_{(p)}}{1 - 1.5n_{(s)}^{-1}} \quad (2.10)$$

- $n_{(s)}$: il parametro di lisciamiento per la componente stagionale. Il suo aumento

comporta una maggiore lisciatura delle sotto-serie. Tale misura viene considerata almeno pari a 7. La scelta di $n_{(s)}$ determina la variazione dei dati che compongono la componente stagionale. Un valore più grande può includere variazioni stagionali su scale temporali più lunghe, mentre un valore più piccolo può evidenziare variazioni più immediate. La selezione ottimale dipende dalle caratteristiche specifiche della serie storica e dagli obiettivi dell'analisi.

2.2 Scomposizione MSTL

Al giorno d'oggi, è comune riscontrare insiemi di dati reali rappresentati da serie storiche caratterizzate da diverse modalità stagionali. A differenza delle serie tradizionali, i dati a frequenza più elevata possono presentare caratteristiche più intricate, come cicli stagionali multipli o stagionalità non intere. Di conseguenza, una scomposizione accurata risulta essenziale per fornire una base solida per le decisioni in diverse circostanze.

Il metodo di scomposizione MSTL (*Multiple Seasonal-Trend decomposition using Loess*) (Bandara et al., 2021) si presenta come un efficace strumento di lisciamiento automatico progettato per affrontare serie storiche con complesse strutture di stagionalità. In particolare, si estende la procedura STL applicando iterativamente la scomposizione per stimare le componenti di stagionalità multipla della serie. Questo processo è cruciale per regolare il livello di lisciamiento e separare in modo continuo le variazioni stagionali.

Spesso, i pattern stagionali sono strutturalmente disposti in maniera complessa, risultando intrecciati tra loro. Durante il processo di scomposizione, quindi, le componenti del ciclo stagionale più breve possono essere eccessivamente assorbite dal ciclo stagionale più lungo. Per ridurre la possibilità di tale eventualità, MSTL organizza i cicli stagionali in maniera crescente.

Il metodo restituisce una scomposizione additiva della serie storica includendo i pattern stagionali. Data l'osservazione al tempo t , definita \hat{X}_t , si ottiene il seguente risultato:

$$X_t = \hat{S}_t^1 + \hat{S}_t^2 + \dots + \hat{S}_t^n + \hat{T}_t + \hat{R}_t, \quad (2.11)$$

dove $\hat{S}_t, \hat{T}_t, \hat{R}_t$ rappresentano rispettivamente la stagionalità, il trend e la componente

residua. L'indice n identifica i diversi cicli stagionali, che possono essere anche non interi.

Dopo aver determinato le frequenze dei pattern stagionali della serie storica, ed averli riordinati in maniera crescente, la scomposizione MSTL applica in maniera iterativa il metodo STL ad ogni cadenza stagionale identificata. Il ciclo interno ripete quest'ultima procedura per estrarre le componenti stagionali dalla serie, che verranno poi raffinate e migliorate tramite diverse iterazioni di STL nel ciclo esterno.

Il trend della serie viene determinato durante l'ultima ripetizione, e sottratto dalla serie destagionalizzata per ricavare la componente residua.

Dal momento che MSTL rappresenta un'estensione di STL, i parametri utilizzati in quest'ultima procedura sono estendibili all'analisi della stagionalità multipla. In questo modo, MSTL si presenta come uno strumento flessibile e potente per la gestione di serie storiche complesse.

2.3 Prophet

Il metodo Prophet (Taylor & Letham, 2018), sviluppato da *Facebook*, è stato progettato per gestire le caratteristiche comuni delle serie storiche aziendali, configurandosi come una potente risorsa nell'ambito delle previsioni. Tale proposta offre un approccio innovativo e flessibile per modellare serie storiche complesse. La sua efficacia è particolarmente evidente nella gestione di dati giornalieri, una frequenza comune nell'ambito di *Facebook*, beneficiando di una metodologia robusta e automatizzata. Prophet, infatti, si distingue per la sua robustezza di fronte a dati mancanti e variazioni nel trend, mostrando tipicamente un'elevata capacità di gestione degli *outliers*.

Il modello si basa sulla scomposizione della serie storica, in maniera simile al caso classico in equazione 2.1, nelle seguenti componenti

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t, \quad (2.12)$$

dove $g(t)$ rappresenta la funzione del trend che modella le variazioni non periodiche, $s(t)$ identifica le stagionalità e $h(t)$ raccoglie gli effetti delle festività. Infine, il termine di

errore ε_t , con distribuzione normale, comprende tutti i cambiamenti presenti nella serie che non sono stati catturati dal modello, rappresentando pertanto la componente residua dell'analisi. Questa specificazione ricorda i modelli additivi generalizzati (GAM), che permettono di scomporre facilmente le serie e di introdurre nuove componenti secondo le necessità.

L'approccio Prophet offre diversi vantaggi in termini di flessibilità nella determinazione delle componenti, velocità di adattamento e facilità di interpretazione dei parametri.

Per quanto riguarda la componente $g(t)$, Prophet consente la scelta tra due diverse specificazioni: modello di crescita non lineare e modello lineare a tratti. Il primo caso fa riferimento alla crescita della popolazione in ecosistemi naturali, in cui la crescita non lineare satura ad una certa capacità. Il modello di crescita non lineare consiste, solitamente, nell'adattamento di un modello logistico di crescita, che nella forma più semplice risulta

$$g(t) = \frac{C}{(1 + \exp(-k(t - m)))}, \quad (2.13)$$

dove C esprime la capacità, k è il tasso di crescita e m il parametro di *offset*.

Tale rappresentazione può essere estesa considerando una capacità ed un tasso variabili nel tempo. La prima quantità risulta quindi espressa come $C(t)$, mentre la seconda viene definita stabilendo dei *changepoints* in cui il tasso di crescita cambia. Supponendo di avere S *changepoints* negli istanti s_j , $j = 1, \dots, S$, si definisce il vettore di aggiustamenti $\delta \in \mathbb{R}^S$, dove δ_j è il tasso di cambiamento che interessa il momento s_j . Il tasso, in qualsiasi istante t , risulta pari al tasso base k , a cui vengono sommati tutti gli aggiustamenti fino a tale momento: $t : k + \sum_{j:t>s_j} \delta_j$.

Per il modello logistico, il tasso di crescita al tempo t corrisponde a $k + a(t)^T \delta$, dove $a(t) \in \{0, 1\}^S$ è un vettore definito nella seguente maniera

$$a_j(t) = \begin{cases} 1, & t \geq s_j \\ 0, & \text{altrimenti} \end{cases} \quad (2.14)$$

A sua volta, anche il parametro m subisce un aggiustamento per connettere i punti

finali dei segmenti. Considerando il punto di cambiamento j , il risultato corretto è:

$$\gamma_j = \left(s_j - m - \sum_{l < j} \gamma_l \right) \left(1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l} \right), \quad (2.15)$$

dove γ è un parametro scelto per connettere i punti estremi. Il modello finale viene espresso come:

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T \delta)(t - (m + a(t)^T \gamma)))} \quad (2.16)$$

Alternativamente al modello logistico di crescita, il modello lineare a tratti prevede che il tasso di crescita sia costante a tratti. La componente di trend è definita nella forma

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma), \quad (2.17)$$

dove k è il tasso di crescita, δ raccoglie gli aggiustamenti del tasso, m rappresenta il parametro *offset* e γ_j viene posto pari a $-s_j \delta_j$ per rendere la funzione continua.

In generale, i punti di cambiamento possono essere specificati dall'analista sulla base di informazioni in suo possesso, oppure determinati automaticamente a partire da un insieme di valori. Quest'ultima opzione prevede la specificazione di una distribuzione a priori sparsa per il parametro δ . Ad esempio, si considera la distribuzione di Laplace: $\delta_j \sim \text{Laplace}(0, \tau)$.

Spesso le serie storiche aziendali presentano una stagionalità dovuta ai comportamenti che rappresentano. Per adattare e prevedere tali effetti è necessario specificare modelli di stagionalità che siano funzioni periodiche di t . Una serie di Fourier è uno strumento che trova naturale applicazione in tale ambito di modellazione. Una serie di Fourier può essere rappresentata come somma di seni e coseni:

$$s(t) = \sum_{n=1}^N \left(a_n \cos \left(\frac{2\pi n t}{P} \right) + b_n \sin \left(\frac{2\pi n t}{P} \right) \right), \quad (2.18)$$

dove P indica il periodo di stagionalità della serie (ad esempio, nel caso di dati settimanali $P = 7$). L'adattamento della componente stagionale richiede la stima di $2N$

parametri

$$\beta = [a_1, b_1, \dots, a_N, b_N]^T \quad (2.19)$$

È necessario, quindi, determinare la matrice dei vettori di stagionalità per ogni valore di t

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{P}\right), \dots, \sin\left(\frac{2\pi(N)t}{P}\right) \right] \quad (2.20)$$

La stagionalità si ottiene come:

$$s(t) = X(t)\beta \quad (2.21)$$

Il parametro *holiday*, $h(t)$, introduce una dimensione personalizzata alla modellazione della serie storica incorporando gli impatti specifici delle festività. Prophet permette di assorbire informazioni riguardo le date in cui si verificano festività e l'impatto che tali eventi possono avere sulla serie storica. Il metodo considera automaticamente le festività come eventi che possono influenzare la previsione, ed assegna loro un effetto additivo, consentendo di catturare variazioni anomale o picchi di attività che possono verificarsi durante le festività.

Per ciascuna ricorrenza i si considera la lista delle date passate e future D_i e una funzione indicatrice che specifica se l'istante temporale t coincide con una delle date in D_i . Ad ogni festività si assegna un parametro k_i che rappresenta la variazione nei valori osservati della serie dovuta al giorno di festivo. Si impone $k \sim \mathcal{N}(0, v^2)$ come distribuzione a priori e si genera la matrice di regressori

$$Z(t) = [\mathbf{1}(t \in D_1), \dots, \mathbf{1}(t \in D_L)], \quad (2.22)$$

dove L identifica l'ultima data futura relativa ad una festività, e si considera

$$h(t) = Z(t)k \quad (2.23)$$

Capitolo 3

Clustering stagionale

Nel contesto della continua ricerca di approcci per perfezionare le previsioni aziendali, fischer Italia ha posto particolare attenzione all'importanza del *clustering* stagionale dei propri prodotti. Questo capitolo si propone di esplorare in dettaglio le metodologie per raggruppare in modo strategico gli articoli in base alle loro fluttuazioni stagionali giornaliere. L'obiettivo primario di questa analisi è ottenere una comprensione più chiara della dinamica stagionale dei prodotti per ottimizzare l'accuratezza delle previsioni.

L'analisi dei *cluster* costituisce un approccio fondamentale nell'ambito dell'elaborazione dei dati, mirato a individuare strutture e raggruppamenti omogenei all'interno di un insieme di osservazioni. Un algoritmo di *clustering*, infatti, ha l'obiettivo di dividere i dati in gruppi affini (cioè contenenti oggetti simili tra loro) e/o identificare possibili *outliers* (oggetti considerati diversi da tutti gli altri).

Questa metodologia trova applicazione in diversi contesti, rivestendo particolare importanza anche nell'ambito delle serie storiche. L'obiettivo principale è quello di individuare pattern nascosti e comportamenti simili tra le osservazioni. In questo ambito, il raggruppamento diventa uno strumento cruciale per affrontare la complessità delle dinamiche temporali. Ciascuna serie può presentare caratteristiche uniche, ma spesso vi sono similitudini in termini di trend, ciclicità e stagionalità.

Il *clustering* stagionale delle serie storiche si concentra su una suddivisione in base a proprietà comuni legate alle variazioni ricorrenti nel corso del tempo. Questa segmentazione può rivelarsi preziosa per diverse finalità, tra cui l'ottimizzazione delle strategie

di previsione e l'adattamento di politiche gestionali specifiche. L'identificazione dei *cluster*, infatti, contribuisce a creare una mappa dettagliata delle dinamiche intrinseche alle diverse categorie di serie storiche. Questo approccio facilita una gestione più focalizzata e personalizzata, consentendo alle aziende di adattare le proprie strategie in base alle caratteristiche peculiari delle osservazioni.

La tecnica di *clustering* stagionale ricopre un importante aspetto nell'analisi dei dati, offrendo ampi vantaggi nella comprensione e nella gestione delle variazioni stagionali nei comportamenti o nei modelli di dati. Questa metodologia, in grado di individuare e raggruppare pattern di stagionalità, è rilevante in settori in cui le dinamiche variano significativamente a seconda delle stagioni. Comprendere l'esistenza di *cluster* fornisce alle organizzazioni una visione più approfondita delle tendenze e dei cicli ricorrenti presenti nei dati. Tale consapevolezza consente alle aziende di adattare le proprie strategie operative e di marketing in modo più mirato, rispondendo alle esigenze mutevoli del mercato. L'analisi si integra in diverse sfere aziendali: dall'ottimizzazione dei piani di approvvigionamento alla personalizzazione delle offerte e dei servizi, puntando al miglioramento dell'efficienza operativa e della soddisfazione del cliente.

In conclusione, il *clustering* stagionale non è solo una pratica analitica, ma si traduce in un potente strumento decisionale. L'identificazione di *cluster* offre un quadro più completo del comportamento dei dati nel corso del periodo in considerazione, consentendo alle aziende di adattarsi dinamicamente alle condizioni di mercato e di prosperare in un ambiente sempre più complesso.

3.1 Funzione TSclust

La formazione di *cluster* di serie storiche rappresenta un campo di ricerca in costante evoluzione con applicazioni estese in vari settori. La sfida principale in questo contesto è la definizione di una misura di dissimilarità tra le serie, un aspetto critico che influenza significativamente i risultati del *clustering*. Per valutare la similarità tra oggetti, è essenziale stabilire una misura di distanza adeguata. Nel contesto del *clustering* di serie storiche, questo compito è particolarmente complesso a causa della natura dinamica dei

dati nel tempo. Fortunatamente, la ricerca recente ha prodotto un crescente numero di misure di similarità progettate appositamente per affrontare questa complessità.

La funzione `TSclust`, del pacchetto `DTWclust` (Giorgino, 2009) disponibile per il programma R, rappresenta un importante contributo in questo campo, offrendo un assortimento ampio e diversificato di misure di similarità, fornendo agli analisti una gamma di opzioni per valutare la differenza tra due serie storiche. Questa diversità di criteri consente di adattare la scelta della misura di dissimilarità in base alle caratteristiche specifiche dei dati e agli obiettivi dell'analisi. Il pacchetto, infatti, contiene misure di dissimilarità comunemente usate, tra cui misure basate sulla complessità, su modelli, sulle caratteristiche e sulle previsioni. `TSclust` è un tentativo di integrare un ampio insieme di misure di diversità tra le serie storiche per permettere all'utente di confrontare le prestazioni e identificare i criteri di diversità utili in un determinato contesto.

3.1.1 Distanza Euclidea

Una delle misure più utilizzate nella definizione dei *cluster* di serie storiche è la distanza Euclidea. La determinazione di tale grandezza è una specificazione della più generale distanza di Minkowski, definita come:

$$d_{L_q}(X_T, Y_T) = \left(\sum_{t=1}^T (X_t - Y_t)^q \right)^{1/q}, \quad (3.1)$$

dove X_t ed Y_t sono le due serie storiche considerate, e q un intero positivo. La distanza Euclidea corrisponde al valore $q = 2$.

Tale metrica risulta competitiva in diverse applicazioni, è invariante alle permutazioni nel tempo e considera le osservazioni nei diversi istanti come valori indipendenti. Tale indice di prossimità, infatti, si basa sulla vicinanza dei valori osservati in ciascun momento temporale.

È fondamentale una fase iniziale di rimozione delle distorsioni e del rumore presente nei dati grezzi per gestire la sensibilità della misura, e confrontare serie storiche di uguale lunghezza e campionamento per comparare ciascuna misurazione effettuata nello stesso momento.

3.1.2 Distanza basata sulla forma

La distanza basata sulla forma delle serie storiche, definita *shape-based distance (SBD)*, si configura come un elemento centrale all'interno dell'algoritmo *k-Shape* (Paparrizos & Gravano, 2015), e rappresenta un'alternativa più efficiente rispetto al caso della distanza calcolata considerando il flusso temporale dinamico, spiegata successivamente.

Questa misura di similarità si basa sul calcolo della correlazione incrociata tra due serie storiche, con una fase preliminare di normalizzazione dei coefficienti per garantire una valutazione robusta. Si tratta, infatti, di una tecnica sensibile alla scala, in quanto tiene conto delle differenze di ampiezza tra le serie. Al fine di mitigare questa sensibilità e garantire un confronto più equo, è consigliata una fase di normalizzazione.

La distanza è definita come segue

$$SBD(x, y) = 1 - \max \left(\frac{NCCc(x, y)}{\|x\|_2 \|y\|_2} \right), \quad (3.2)$$

dove $\|\cdot\|_2$ rappresenta la norma l_2 delle serie e $NCCc$ è la sequenza di correlazioni incrociate delle due serie con i coefficienti normalizzati. La distanza comprende valori tra 0 e 2: un valore nullo indica una perfetta similarità.

Il calcolo di questa distanza può essere ottimizzato mediante l'impiego della *Fast Fourier Transform (FFT)* per ottenere la sequenza $NCCc$, come prevede l'implementazione adottata nella funzione `sbd` di `DTWclust`. Tale metodo si presenta come un'opzione che può essere molto veloce, che è simmetrica e che risulta molto competitiva riportando buoni risultati negli esperimenti eseguiti in Paparrizos & Gravano (2015).

3.1.2.1 Algoritmo *k-Shape*

L'algoritmo *k-Shape* è una tecnica di *clustering* di partizione basata sul centroide, in grado di preservare le forme delle sequenze di serie storiche. L'acquisizione di somiglianze basate sulla forma richiede misure di distanza in grado di gestire le distorsioni di ampiezza. Questo processo, solitamente, richiede un costo computazionale elevato. Per aggirare questa limitazione di efficienza, si adotta una versione normalizzata della misura di correlazione incrociata.

La correlazione incrociata è una misura statistica che può determinare la somiglianza di due sequenze $\vec{x}=(x_1, \dots, x_m)$ e $\vec{y}=(y_1, \dots, y_m)$, anche nel caso in cui non siano perfettamente allineate. Tale grandezza risulta invariante in presenza di cambiamenti perchè mantiene statico \vec{y} e fa scorrere \vec{x} su \vec{y} per calcolare il loro prodotto scalare per ogni traslazione s di \vec{x} . Un cambiamento s di una sequenza è definito come:

$$\vec{x}_{(s)} = \begin{cases} \overbrace{(0, \dots, 0, x_1, x_2, \dots, x_{m-s})}^{|s|}, & s \geq 0 \\ (x_{1-s}, \dots, x_{m-1}, x_m, \underbrace{0, \dots, 0}_{|s|}), & s < 0 \end{cases} \quad (3.3)$$

Con $\vec{x}_{(s)}$ si considerano tutti i possibili cambiamenti, con $s \in [-m, m]$. Si definisce la sequenza di correlazioni incrociate, di lunghezza $2m - 1$, come:

$$CC_w(\vec{x}, \vec{y}) = (c_1, \dots, c_w) = R_{w-m}(\vec{x}, \vec{y}) \quad w \in 1, 2, \dots, 2m - 1 \quad (3.4)$$

dove $R_{w-m}(\vec{x}, \vec{y})$ è calcolato come

$$R_{w-m}(\vec{x}, \vec{y}) = \begin{cases} \sum_{l=1}^{m-k} x_{(l+k)} \cdot y_l, & k \geq 0 \\ R_{-k}(\vec{y}, \vec{x}), & k < 0 \end{cases} \quad (3.5)$$

L'obiettivo è di individuare la posizione w per la quale $CC_w(\vec{x}, \vec{y})$ è massimizzata. Lo *shift* ottimale di \vec{x} rispetto a \vec{y} è quindi $\vec{x}_{(s)}$, dove $s = w - m$. Si possono distinguere diverse definizioni di normalizzazione di $CC_w(\vec{x}, \vec{y})$ in base all'applicazione in esame.

Le più comuni sono:

$$NCC_q(\vec{x}, \vec{y}) = \begin{cases} \frac{CC_w(\vec{x}, \vec{y})}{m}, & q = b \quad (NCC_b) \\ \frac{CC_w(\vec{x}, \vec{y})}{m-|w|}, & q = u \quad (NCC_u) \\ \frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{x}) \cdot R_0(\vec{y}, \vec{y})}}, & q = c \quad (NCC_c) \end{cases} \quad (3.6)$$

dove NCC_b rappresenta lo stimatore distorto, NCC_u quello non distorto e NCC_c il coefficiente di normalizzazione.

Le sequenze di correlazione incrociata prodotte dai confronti a coppie delle serie temporali differiscono in ampiezza in base alle normalizzazioni. Perciò, risulta vantaggioso utilizzare una normalizzazione che produce valori entro un determinato intervallo, in modo tale da garantire un confronto significativo. La soluzione è identificabile nella *shape based distance* presentata precedentemente.

L'algoritmo *k-Shape* si compone di un ciclo di iterazioni, per ciascuna delle quali si esegue un doppio passaggio.

- Fase di affinamento: si ricalcolano (o calcolano, se l'algoritmo è alla prima iterazione) i centroidi dei *cluster* per considerare i cambiamenti dei membri dei gruppi avvenuti nella fase di assegnazione (o nell'assegnazione casuale se si tratta della prima iterazione)
- Fase di assegnazione: si aggiorna l'appartenenza al *cluster* confrontando ciascuna serie con tutti i centroidi.

L'algoritmo si interrompe se nella fase di assegnazione non si rilevano cambiamenti nei membri dei *cluster*, oppure nel momento in cui si raggiunge il numero massimo di iterazioni.

L'operazione principale nella prima fase dell'algoritmo consiste nel calcolo dei centroidi. Nell'algoritmo *k-Shape*, questi riferimenti vengono definiti tramite l'utilizzo della distanza basata sulla forma: il centroide corrisponde alla ricerca della minima somma del quadrato delle distanze tra tutte le serie storiche all'interno di un *cluster*. Tale ricerca è conosciuta come "problema della sequenza di Steiner" (Petitjean & Gançarski, 2012): data una partizione $p_j \in P = p_1, \dots, p_k$, il corrispondente centroide \vec{c}_j deve soddisfare

$$\vec{c}_j = \arg \min_{\vec{w}} \sum_{\vec{x}_i \in p_j} \text{dist}(\vec{w}, \vec{x}_i)^2, \quad \vec{w} \in \mathbb{R}^m \quad (3.7)$$

Dato che la correlazione incrociata cattura la similarità delle serie storiche, piuttosto che la dissimilarità, il problema può essere riscritto come la massimizzazione del quadrato della similarità tra tutte le sequenze

$$\begin{aligned}
 \vec{\mu}_k^* &= \arg \max_{\vec{\mu}_k} \sum_{\vec{x}_i \in P_k} NCC_c(\vec{x}_i, \vec{\mu}_k)^2 \\
 &= \arg \max_{\vec{\mu}_k} \sum_{\vec{x}_i \in P_k} \left(\frac{\max_w [CC_w(\vec{x}_i, \vec{\mu}_k)]}{\sqrt{R_0(\vec{x}_i, \vec{x}_i) \cdot R_0(\vec{\mu}_k, \vec{\mu}_k)}} \right)^2
 \end{aligned} \tag{3.8}$$

3.1.3 Distanza basata sul flusso temporale dinamico

La distanza basata sul flusso temporale dinamico, definita *Dynamic Time Warping distance (DTW)*, è un algoritmo che compara due serie storiche con l'obiettivo di individuare il percorso ottimale per il confronto.

Questa metrica è stata introdotta per superare i limiti associati alla distanza Euclidea. La distanza DTW, infatti, risulta applicabile nel caso di serie storiche non sincronizzate che possono, quindi, presentare osservazioni simili in posizioni differenti nel tempo. La Figura 3.1 evidenzia in maniera intuitiva la differenza dei due metodi.

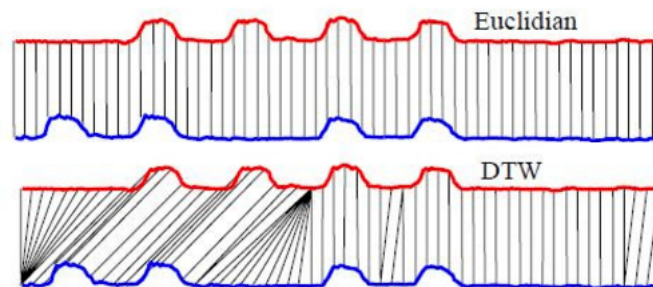


FIGURA 3.1: Confronto tra distanza Euclidea e DTW

Per determinare la distanza basata sulla sequenza dinamica delle osservazioni tra due serie storiche X_t e Y_t si calcola una matrice dei costi denominata *Local Cost Matrix (LCM)*. La matrice, di dimensione $n \times m$, raccoglie le distanze tra x_i e y_i , solitamente calcolate come differenze quadratiche: $d(x_i, y_i) = (x_i - y_i)^2$. In seguito, si definisce il percorso $W = w_1, w_2, \dots, w_K$, con $\max(n, m) \leq K \leq m + n - 1$, che deve attraversare la matrice rispettando le seguenti condizioni.

- Vincolo sul confine: il percorso deve iniziare e terminare agli angoli della matrice, ossia $w_1 = (1, 1)$ e $w_K = (n, m)$

- Vincolo di continuità: : solo gli elementi adiacenti possono essere considerati come passaggi successivi del percorso
- Vincolo di monotonia: gli *step* successivi del percorso devono essere spazialmente monotoni nel tempo.

La distanza totale risulta dalla somma degli elementi della matrice LCM interessati dal percorso W (*warping path*), mentre la distanza finale DTW è riscontrabile nel percorso con la minima distanza totale. Una possibile definizione della misura consiste nel considerare la radice quadrata della somma

$$d_{DTW}(X_T, Y_T) = \min \left(\sqrt{\sum_{k=1, \dots, K} w_k} \right) \quad (3.9)$$

La distanza basata sul flusso dinamico temporale è computazionalmente costosa. Se x ha lunghezza n e y ha lunghezza m , la distanza DTW risulta avere un costo di calcolo pari a $\mathcal{O}(nm)$, che è quadratica se m e n sono simili. Inoltre, è soggetta a distorsione e a lentezza di calcolo. Una soluzione più veloce è riscontrabile nella forma semplificata denominata `DTW_basic`.

Capitolo 4

Caso aziendale fischer Italia

4.1 Gruppo fischer

Il *fischer group of companies* è un gruppo industriale tedesco specializzato nel campo dei fissaggi strutturali, dell'*interior design* di autoveicoli e dei giochi di costruzioni. L'azienda si occupa della progettazione, produzione e distribuzione di sistemi di ancoraggio, viti, tasselli e altri elementi di fissaggio in vari settori, tra cui l'edilizia e l'industria.

Con oltre 14000 prodotti, costruiti ed assemblati nel rispetto delle più attuali normative europee e internazionali, fischer rappresenta da sempre il punto di riferimento per gli installatori e i progettisti nell'edilizia civile e industriale. Il Gruppo propone una gamma completa di fissaggi in nylon, ancoranti meccanici e chimici, fissaggi per idrotermosanitaria, linee vita, isolamento, sistemi per involucro edilizio, fotovoltaico e solare termico. Tali articoli possono vantare alta specializzazione, affidabilità, certificati di validità, sicurezza e velocità nell'installazione.

L'azienda viene fondata nel 1948 presso il comune di Waldachtal, in Germania, da Artur Fischer. Tra le invenzioni di maggior rilievo, attribuibili al fondatore, è essenziale menzionare le principali.

- Il sistema di flash sincrono per macchine fotografiche (1949)
- I sistemi di fissaggio in poliammide (1958)
- Il '*fischertechnik*' in scatola di montaggio, ossia un sistema di elementi plastici da montare per l'insegnamento della tecnica delle costruzioni (1964).

Ad oggi, la multinazionale tedesca, di cui è titolare e presidente il professor Klaus Fischer, fa affidamento a 50 consociate in 38 Paesi con siti produttivi in Germania, Italia, Argentina, Brasile, Cina, Repubblica Ceca e Stati Uniti d'America. L'azienda conta più di 5500 dipendenti, 2000 invenzioni e oltre 3000 brevetti registrati.

4.2 Azienda fischer Italia

L'azienda fischer Italia inizia la propria attività a Padova nel 1963 ed, attualmente, vanta il primo posto tra le consociate del Gruppo per fatturato. Tra le date da richiamare, il 1998 segna il lancio in Italia del tassello "fischer SX", che diventerà un *topseller*. È di rilievo menzionare il fatto che l'azienda applichi i principi "fPS" (*fischer Prozess System*) che tendono alla massima efficienza produttiva, logistica, distributiva, e alla riduzione degli sprechi per una creazione di valore per i propri clienti, attraverso i prodotti.

La sede italiana rappresenta il Centro di Competenza per tutto il gruppo nell'ambito sanitario e fotovoltaico, siede alla vicepresidenza del Consiglio Direttivo di ISI (Ingegneria Sismica Italiana), collabora con i più qualificati Istituti di Ricerca italiani, è partner di UniSMART e contribuisce allo sviluppo di progetti di ricerca con l'Università di Padova.

4.2.1 Pianificazione delle Vendite e delle Operazioni

L'azienda di tasselli si colloca come una realtà dinamica e consolidata nel panorama commerciale, con un fatturato totale del 2023 di circa 197.5 milioni di euro. Fondata sulla qualità e sull'innovazione, fischer Italia si distingue per la sua presenza radicata e il suo ruolo da *leader* nel settore edilizio ed industriale.

La maggior parte delle vendite, precisamente il 94%, deriva dal mercato locale, confermando una solida presenza a livello nazionale. Il rimanente 6% delle transazioni è costituito da vendite intra-gruppo. L'azienda dimostra una gestione efficace della propria catena di approvvigionamento, conseguendo un equilibrio strategico tra la produzione interna e le collaborazioni esterne. La ripartizione percentuale in tale ambito si configura nel modo seguente: 20% del fatturato proviene dalla produzione interna, 22% dal conto-lavoro, 28% da commercializzazione di prodotti di fornitura esterna e

marchio fischer e 30% da produzioni intra-gruppo. L'azienda coordina la propria rete di fornitura e distribuzione attraverso una gestione accurata e di *partnership* strategiche. La maggior parte delle spedizioni, pari all'80%, è affidata alla logistica interna, mentre la restante coinvolge *partner* logistici esterni.

Il ciclo di Pianificazione delle Vendite e delle Operazioni che interessa fischer Italia si articola principalmente in quattro stadi.

La prima fase del processo ha inizio alla fine del mese, occupando approssimativamente l'ultima settimana. Durante questo periodo vengono aggiornate le metriche per valutare le prestazioni dei mesi precedenti, analizzate le serie storiche di vendita, esaminate le previsioni statistiche, raccolti gli input di mercato per una valutazione qualitativa integrativa alle previsioni, rivisti i piani di sviluppo dei nuovi prodotti e lanciati i programmi promozionali.

Segue una fase di allineamento in cui vengono riesaminate le metriche e gli indicatori chiave di prestazione (*Key Performance Indicators*), si aggiornano le previsioni nel programma SO99+, i prezzi sono oggetto di analisi e validazione, e si confrontano le proiezioni elaborate dal processo S&OP con il *budget*.

Il terzo periodo riguarda la creazione dei piani di approvvigionamento e capacità, delle proiezioni di magazzino, del monitoraggio degli articoli a movimentazione lenta e della valutazione generale dei fornitori.

L'ultima fase coinvolge i reparti delle *Operations* con gli incontri di riconciliazione e le analisi dei livelli di magazzino.

4.2.2 Tecnologia SO99+

All'interno di questo ampio processo di pianificazione, svolge un ruolo cruciale la metodologia adottata per generare le previsioni. Attualmente, fischer Italia utilizza il programma SO99+ (*Service Optimizer 99+*). Il programma SO99+, sviluppato da ToolsGroup, si propone come supporto per la pianificazione della domanda, l'ottimizzazione delle scorte e del livello di servizio. L'obiettivo è di migliorare le previsioni di breve/medio periodo allocando correttamente gli stock, i piani di produzione e di distribuzione, garantendo un alto livello di servizio al cliente.

L'algoritmo proprietario presenta le seguenti caratteristiche.

- Autoadattività
- Capacità di gestire la domanda sia in termini di quantità che di frequenza
- Integrazione di una funzione di distribuzione di probabilità
- Smorzamento esponenziale con peso crescente sui mese recenti.

Gli elementi raccolti dal programma riguardano l'anagrafica del prodotto, la struttura del mercato e la domanda storica. I primi due *input* sono necessari per il calcolo della previsione statistica, che verrà poi adattata considerando gli effetti della pianificazione della domanda (forzature commerciali, stagionalità, promozioni, ecc.), mentre l'ultima variabile costituisce il filtro per gli eventi eccezionali.

La tecnologia SO99+ incorpora la metodologia definita *Frequency Based Forecasting* (May et al., 2023). Il sistema si basa sull'utilizzo delle quantità e delle frequenze, intese come righe ordine, per determinare una previsione più accurata ed un'allocazione ottimale delle risorse.

Nella fase iniziale viene eseguita una trasformazione di *Fourier* che interessa diversi intervalli di dati. Il passo successivo riguarda la determinazione dell'andamento di tali periodi: per ciascun intervallo si determina l'insieme popolato dalle osservazioni passate e si identifica la sezione di frequenze che restituisce il minor errore medio assoluto percentuale (MAPE). Il periodo selezionato viene considerato come comportamento di riferimento e utilizzato per il calcolo delle previsioni future. L'idea base dell'algoritmo è che ciascuna frequenza è assegnata ad un tipo di cambiamento. La classificazione viene eseguita per ogni fase di previsione poiché le frequenze variano tra i diversi periodi.

Il *software* permette di determinare una componente di stagionalità, che consente di applicare alla previsione una serie di indici correttivi per tenere conto di andamenti ricorsivi della domanda a livello mensile. Il calcolo automatico è eseguito confrontando la domanda media del mese in considerazione con la media mensile sull'orizzonte storico, con tecniche di protezione per eventi ad elevata variabilità e forte divergenza rispetto all'andamento medio.

4.3 Presentazione dataset

Il dataset da analizzare è reso disponibile dall'azienda fischer Italia e raccoglie gli ordini giornalieri ricevuti dai diversi clienti dal 2012 al 2023. Per ciascuna commessa vengono riportate informazioni identificative del cliente, del prodotto e della data di spedizione richiesta.

L'unità di misura considerata per l'analisi è la quantità di prodotto ordinata in pezzi e come riferimento temporale viene considerata la data di uscita merce dal magazzino fischer richiesta dal destinatario.

Il presente lavoro può essere suddiviso in base agli obiettivi. La prima parte di analisi si focalizza sul confronto tra i metodi STL e Prophet per determinare l'alternativa più competitiva all'algoritmo utilizzato dall'azienda. Le due proposte si distinguono per la capacità di poter prevedere su base giornaliera il numero di pezzi richiesti dai clienti, superando il limite del *bucket* mensile del software SO99+. Questa possibilità permette, inoltre, di considerare diverse e più tipologie di stagionalità, per riuscire a catturare al meglio l'andamento delle serie storiche di vendita.

Data l'elevata quantità di dati a disposizione, l'analisi si focalizza sulla famiglia di prodotto principale denominata "SBU 501": sistemi di fissaggio in nylon. Il dataset registra molteplici ordini per ciascun riferimento temporale, perciò si aggregano le quantità di pezzi richiesti per ciascuna data. L'obiettivo è di ottenere un piano previsivo a livello di Volume di "famiglia" e non di Mix per articolo. Come chiarito in precedenza, il successo della Pianificazione delle Operazioni e delle Vendite si basa sulla comprensione di Quattro Pilastri chiave, tra cui il Volume. Tale concetto rappresenta un livello di aggregazione più ampio, dove vengono prese in considerazione intere famiglie di prodotti anziché singoli articoli.

La fase iniziale di organizzazione del dataset comprende il popolamento delle lacune temporali assegnando loro quantità nulle. Questa scelta coinvolge in particolare i sabati e le domeniche, in cui l'azienda non registra richieste di merce, ma è dettata da aspetti tecnici ed intuitivi.

- Consentire il naturale adattamento dei parametri di stagionalità dei metodi considerati
- Allenare i modelli al riconoscimento dei *weekend* per ottenere previsioni giornaliere

più accurate e precise.

I dati riguardano il numero di pezzi ordinati, con la possibilità di presentare valori estremamente elevati o bassi, manifestando così una notevole variabilità. Il massimo registrato è di 13352158 pezzi, con una media di 640997 unità. Per mitigare la distanza rilevante tra la modalità minima e quella massima, si applica una trasformazione logaritmica

$$y = \log(x + 1) \quad (4.1)$$

Questa scelta mira a rendere più simmetrica la distribuzione della variabile, avvicinando i valori estremi a quelli centrali e, conseguentemente, a ridurre la variabilità complessiva.

In Figura 4.1 si riporta il risultato dell'autocorrelazione per la serie storica degli ordini relativi alla famiglia di prodotto "501": emerge un evidente segnale stagionale data la significatività e la cadenza regolare delle correlazioni. Questo aspetto conferma le intuizioni avanzate dall'azienda e verrà considerato nelle analisi successive.

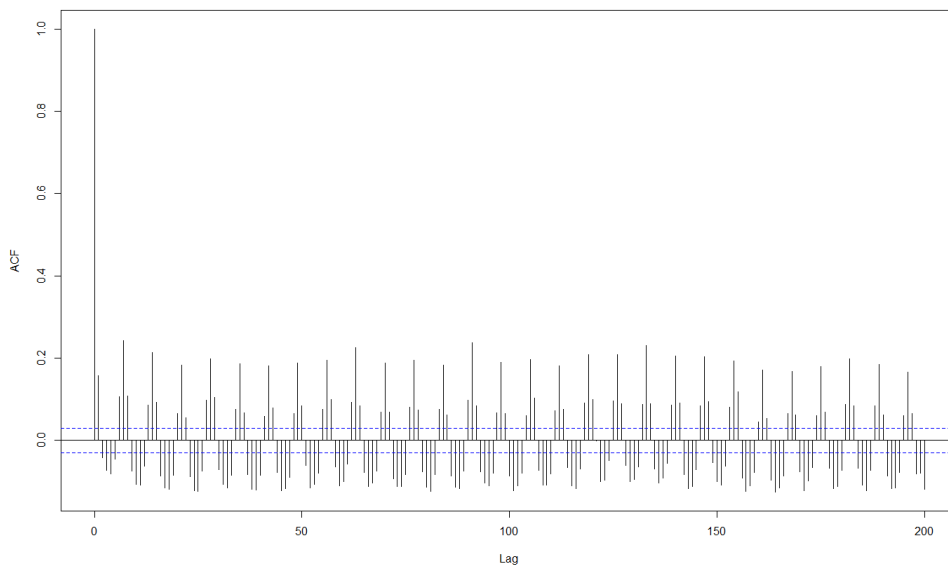


FIGURA 4.1: Funzione di autocorrelazione per la famiglia di prodotto "501"

Come metrica di confronto delle prestazioni dei diversi modelli si considera la radice

dell'errore quadratico medio

$$\sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (4.2)$$

dove Y_t è il valore della serie storica al tempo t e \hat{Y}_t è il suo valore stimato al medesimo istante temporale. Più il valore è basso e più il modello risulta preciso.

4.4 Analisi con scomposizioni STL e MSTL

La prima analisi avanzata si occupa dell'implementazione della scomposizione STL, con successivo confronto con il metodo MSTL per la gestione della stagionalità multipla. Si presenta sia il caso della serie storica con frequenza settimanale (`frequency = 7`), sia annuale (`frequency = 365`). L'insieme degli ordini registrati tra il 2012 e il 2022 viene considerato come insieme di stima da confrontare con l'insieme di verifica composto dalle commesse del 2023.

Nel caso della serie storica settimanale, i parametri sono scelti dopo considerazioni ragionevoli, pari a quelli in elenco.

- `s.window = 53`
- `t.window = 81`
- `l.window = 53`
- `inner = 1`
- `outer = 15`
- `robust = TRUE`.

Il valore relativo alla componente stagionale e al filtro loess si determina considerando le 52 settimane dell'anno e imponendo l'intero dispari più vicino. La quantità riferita al trend considera il criterio 2.10. I parametri riguardanti il numero di iterazioni per il ciclo interno ed esterno vengono selezionati in conformità alle indicazioni teoriche, al fine di favorire la convergenza delle componenti di trend e stagionalità.

Considerando, invece, la serie storica con frequenza annuale, i valori scelti sono i seguenti.

- `s.window = 365`
- `t.window = 550`

- `l.window = 365`
- `inner = 1`
- `outer = 15`
- `robust = TRUE`.

La quantità assegnata per la gestione della stagionalità riflette il numero di giorni in un anno.

Tra le due modalità di lettura dei dati, il caso settimanale restituisce un RMSE più basso, pari al 2.373, rispetto al risultato di 6.821 del caso annuale. Tuttavia, grazie anche all'analisi grafica riportata in Figura 4.2, emerge la limitazione nel considerare una sola stagionalità: il modello non coglie l'andamento generale della serie storica, ignorando le fluttuazioni che possono caratterizzare i diversi mesi. Nel mese di agosto, ad esempio, ci si aspetta un abbassamento dei risultati dovuto alla pausa estiva.

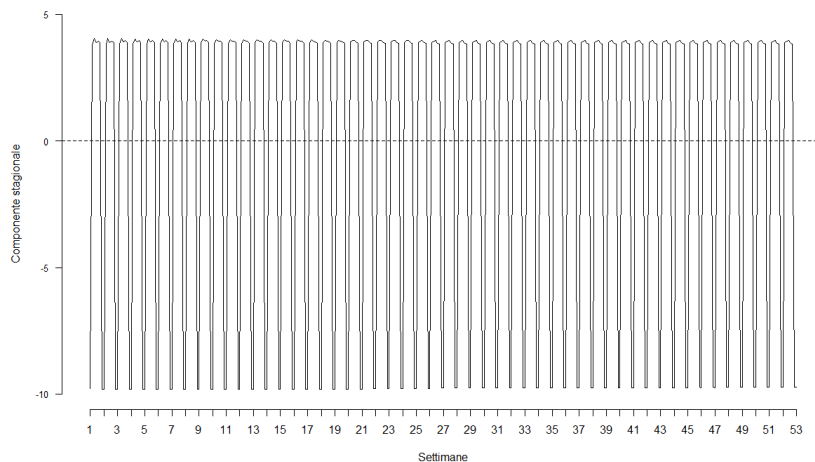


FIGURA 4.2: Stagionalità estratta con STL della serie storica settimanale del 2022

Si considera, quindi, la scomposizione MSTL per consentire la gestione di più di una stagionalità. Per il parametro `seasonal.periods` si specificano i valori 7 e 365 per supportare l'adattamento sia a livello di settimana che a livello annuale. Tale metodo restituisce un RMSE pari a 4.722, evidenziando un miglioramento rispetto alla considerazione annuale precedente ma un peggioramento se confrontato con il caso solo settimanale.

L'analisi grafica della componente stagionale analizzata in Figura 4.3, tuttavia, mostra un comportamento più verosimile e pertinente alla realtà aziendale.

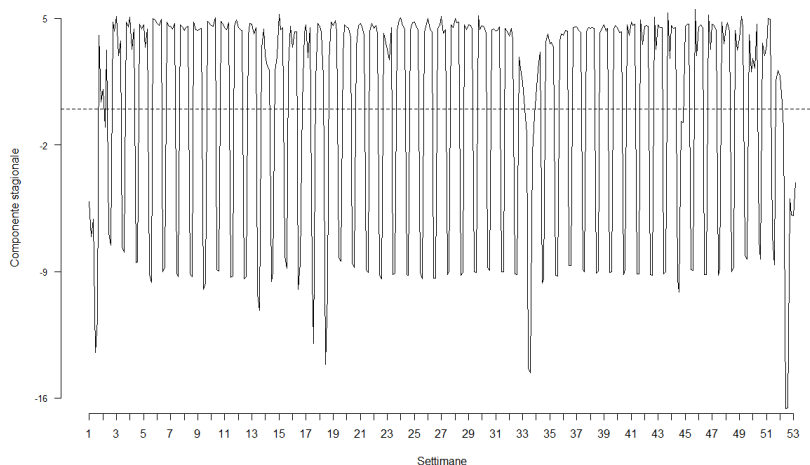


FIGURA 4.3: Stagionalità estratta con MSTL della serie storica del 2022

4.5 Analisi con metodo Prophet

La metodologia adottata da Prophet emerge come un' alternativa alla scomposizione STL per il calcolo delle previsioni giornaliere e l'analisi della componente stagionale. Per le successive indagini si utilizza il comando `prophet` del pacchetto `fable.prophet`.

Prophet consente di incorporare gli effetti delle festività tramite il parametro `holiday`. Il metodo propone una lista già popolata di eventi ricorrenti, suddivisa per nazione e anno. Nel caso specifico dell'applicazione, si selezionano le festività italiane a partire dal 2012, a cui si aggiungono le pause estive e natalizie che hanno caratterizzato la storia di fischer Italia. L'elenco ottenuto viene ampliato considerando anche i sabati e le domeniche.

Si adatta il modello specificando la tipologia di stagionalità e considerando gli effetti dei giorni festivi. Si confronta, come nel caso della scomposizione STL, l'adattamento con la componente stagionale settimanale (`season('week')`) con il modello con stagionalità annuale (`season('year')`).

L'output del metodo Prophet ottenuto con il software R risulta

$$y = trend \cdot (1 + multiplicative_terms) + additive_terms + .resid \quad (4.3)$$

dove y è la variabile risposta pari al logaritmo del numero di pezzi richiesti, come `trend` si considera un andamento lineare a tratti, il parametro `multiplicative_terms` risulta nullo perchè non specificato e la quantità `additive_terms` corrisponde alla somma della componente stagionale e degli effetti della quantità *holiday*. Nello specifico, è possibile estrarre l'indice di stagionalità corrispondente a ciascun giorno, elemento utile ai fini interpretativi e previsivi in ambito aziendale.

Il modello che considera una stagionalità settimanale restituisce un valore di RMSE nettamente più basso del caso annuale: 1.824 contro 6.137.

Analizzando le indicazioni stagionali estratte per il modello meglio performante in termini di radice dell'errore quadratico medio, non emergono sostanziali differenze tra le diverse settimane dell'anno, come dimostra la Figura 4.4.

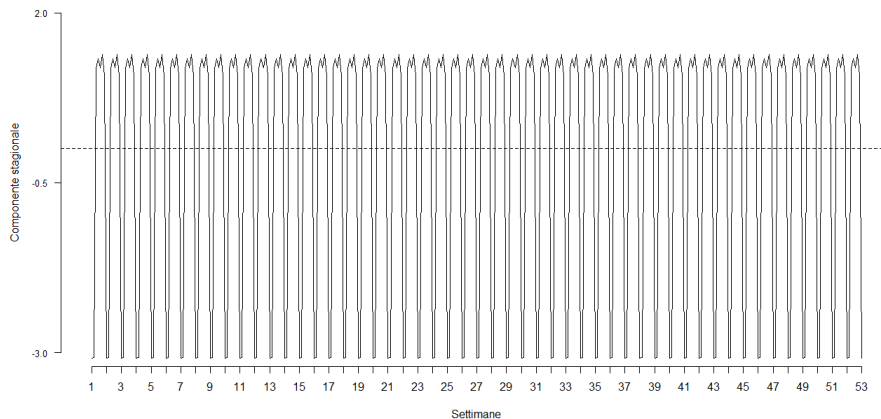


FIGURA 4.4: Stagionalità settimanale estratta con Prophet della serie storica del 2022

Di conseguenza, si ipotizza il caso in cui i dati siano caratterizzati da più di una stagionalità, valutando l'effetto totale degli indici di stagionalità relativi sia alle settimane e che all'anno in generale. Questo adattamento risulta più competitivo dei casi precedenti, restituendo un RMSE pari a 1.707. La componente di stagionalità dell'anno 2022 risultante dalla somma degli indici `'week'` e `'year'` calcolati con il metodo Prophet è

evidenziata in Figura 4.5. La curva ottenuta appare meno statica e più flessibile nel rincorrere l'andamento della serie storica.

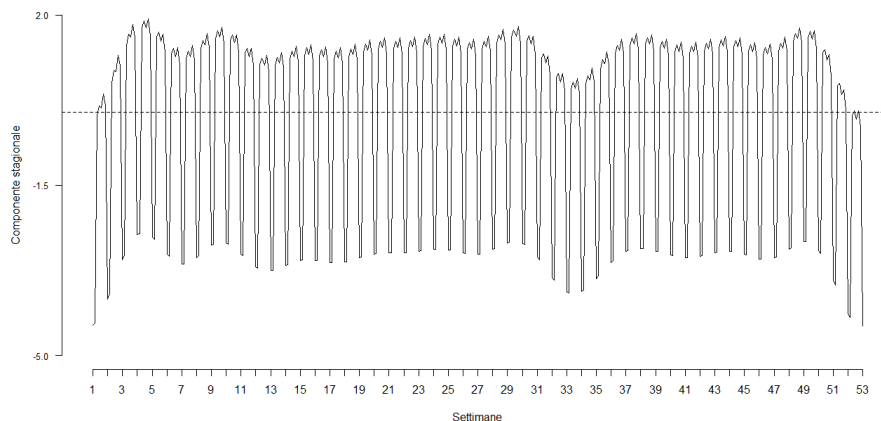


FIGURA 4.5: Stagionalità settimanale ed annuale estratta con Prophet della serie storica del 2022

4.5.1 Confronto Prophet e MSTL

Nel corso dell'analisi sono stati esaminati due diversi metodi: la scomposizione STL e Prophet. Per entrambe le tecniche, sono stati considerati sia il caso con una sola componente stagionale (settimanale o annuale) che l'opzione con più di una stagionalità (sia settimanale che annuale). L'adattamento di Prophet con stagionalità multipla si distingue per essere il più efficace, evidenziando un valore di RMSE inferiore rispetto alla metodologia basata sul loess. Questo risultato evidenzia la capacità del modello di cogliere le fluttuazioni ricorrenti nei dati anche in presenza di più di una componente stagionale.

In Figura 4.6, si nota chiaramente come le previsioni (riportate in blu) di Prophet siano coerenti con i dati storici, suggerendo un'elevata precisione nel catturare le tendenze passate. Al contrario, l'analisi di scomposizione MSTL, rappresentata in Figura 4.18, sembra manifestare una propensione alla sovrastima delle quantità future. Quest'ultimo risultato può essere causato da una struttura complessa dei dati che il modello non è in grado di cogliere adeguatamente, o dal numero elevato di parametri da specificare, che causano un buon adattamento ai dati di allenamento, comprese le fluttuazioni casuali o i rumori, con conseguente sovrastima in fase previsiva.

Gli intervalli di confidenza associati a MSTL risultano più ampi rispetto a quelli derivati da Prophet, indicando una maggiore incertezza nelle previsioni.

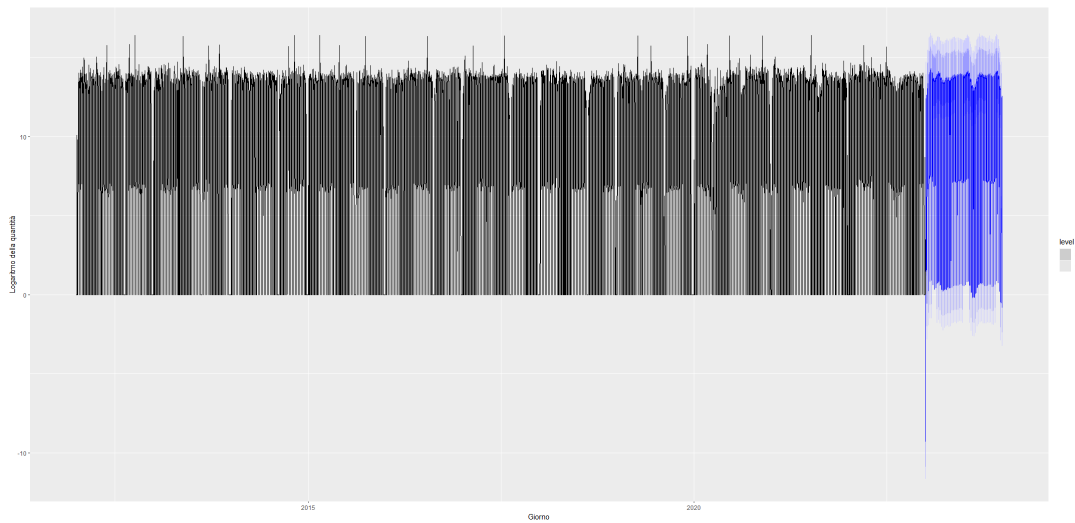


FIGURA 4.6: Previsioni con Prophet con doppia stagionalità

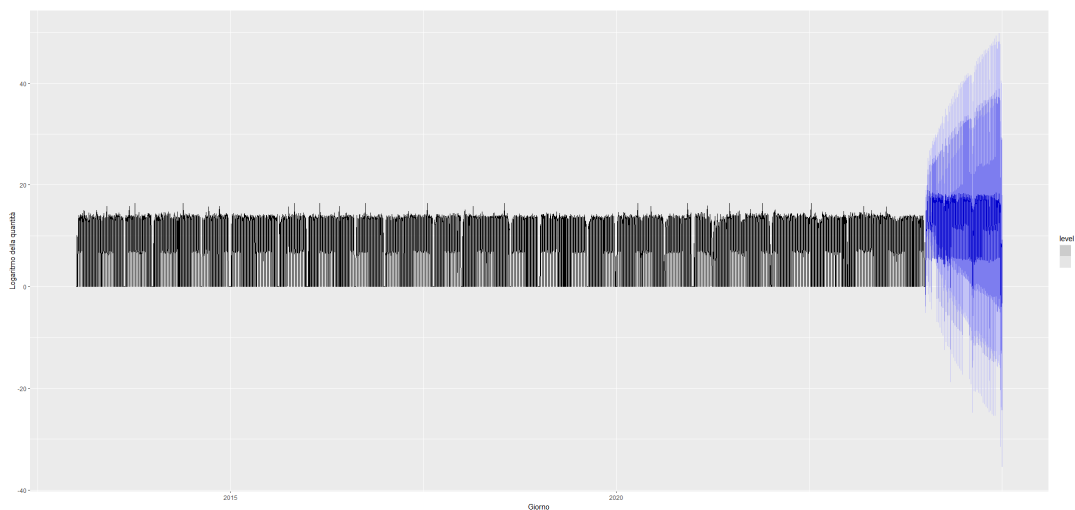


FIGURA 4.7: Previsioni con MSTL

In generale, il risultato favorevole di Prophet è riconducibile a diversi aspetti. Il metodo permette una gestione automatica delle festività, consentendo una migliore gestione di eventi ricorrenti che possono influenzare la serie storica. Il trattamento degli outliers consente un controllo robusto dei dati anomali o eventi eccezionali. Infine, Prophet si propone come tecnica di facile utilizzo, risultando meno soggetto a errori di specifica del modello stesso.

4.6 *Clustering* stagionale

Nella fase precedente, Prophet si è rivelato il metodo più competente per la fase previsiva. La seconda parte dello studio consiste nell'adattamento di tale modello per estrarre la componente stagionale delle serie storiche dei prodotti offerti da fischer Italia. L'obiettivo consiste nella segmentazione del portafoglio prodotti finiti in classi di stagionalità per migliorare l'accuratezza del processo previsivo. Il dato verrà utilizzato come *input* del sistema di pianificazione già presente in azienda.

L'analisi di *clustering* stagionale si focalizza sulla famiglia di prodotto "SBU 501". Le quantità di pezzi vengono raggruppate in base alla data richiesta di spedizione e al codice specifico del prodotto. Il totale degli articoli inclusi in questa analisi ammonta a 12371 identificativi. Si procede con il filtraggio dei dati per ottenere un insieme più appropriato e ridotto da analizzare.

Si selezionano solo i prodotti che registrano almeno 300 frequenze negli ultimi due anni a disposizione (2022-2023), rappresentando una media di circa 3 richieste di spedizione a settimana. Questa selezione permette di concentrare l'analisi sulle serie storiche più popolate, focalizzandosi sui codici che registrano una domanda più sostenuta rispetto ad altri. Si ottiene un risultato di 383 articoli.

La storia dei prodotti di fischer Italia si distingue per la presenza di articoli consolidati, venduti con successo per diversi anni, e altri frutto di recenti progettazioni. Questa diversità di prodotti nel catalogo dell'azienda può essere attribuita a diversi fattori: innovazione tecnologica, cambiamenti nelle esigenze dei clienti, competizione di mercato o, più semplicemente, ciclo di vita del prodotto.

Per garantire dei risultati attendibili, delle stime stabili e una comprensione approfondita delle dinamiche strutturali, si escludono le serie storiche con data iniziale successiva all'1 gennaio 2018. Questa scelta conduce ad un gruppo di 187 codici prodotto.

Dopo aver colmato le lacune temporali con quantità nulle, si adatta il modello Prophet con stagionalità multipla, sia settimanale che annuale, per ciascun prodotto. La variabile risposta considerata è la quantità di pezzi aggregata in precedenza. Diversamente dall'analisi previsiva precedente, non si adotta una trasformazione logaritmica perchè verrà eseguita una normalizzazione dei dati durante il processo di raggruppamento.

Da ciascun adattamento, si estrae la componente stagionale come somma dell'effetto

settimanale e dell'impatto annuale. La serie storica di stagionalità di ciascun articolo viene raccolta in un unico *dataframe* di cui si considerano solo i valori relativi all'anno 2023.

La definizione dei *cluster* stagionali coinvolge l'applicazione di due approcci principali: il metodo gerarchico e quello di partizione. Durante questo processo, vengono esaminate tre diverse metriche di distanza: la distanza euclidea, la distanza basata sulla forma (*sbd*) e la variante semplificata della distanza basata sul flusso dinamico del tempo (*dtw_basic*). Tale analisi consente di valutare le relazioni e le similarità tra le serie storiche stagionali in modi diversi, considerando aspetti come la forma e la dinamica temporale.

Si utilizza la funzione `tsclust` del pacchetto `dtwclust`, specificando la richiesta di normalizzazione dei dati attraverso il comando `preproc = zscore`.

Dopo un confronto avuto con l'azienda, si stabilisce che il numero massimo di *cluster* considerato appropriato per i 187 articoli in analisi non deve superare i 5 gruppi. Si confrontano i risultati ottenuti dai vari metodi di *clustering*, prendendo in considerazione questa specifica indicazione. Questa valutazione mira a fornire una suddivisione ragionevole dei prodotti in un numero gestibile di gruppi, agevolando così un'interpretazione chiara e pratica dei risultati nel contesto operativo dell'azienda.

Per valutare le prestazioni dei diversi metodi di *clustering*, ci si avvale dell'utilizzo dell'indice *silhouette*. Questa misura assume valori all'interno dell'intervallo $[-1, +1]$, dove un coefficiente uguale a $+1$ suggerisce una separazione netta e distintiva tra i vari *cluster*, un valore pari a 0 indica una neutralità tra i gruppi, implicando che essi possono essere sovrapponibili o che la distanza tra di essi non è significativa, e un valore negativo riflette un'assegnazione erronea dei dati ai *cluster*. L'analisi dell'indice *silhouette* fornisce, quindi, una chiara valutazione della qualità della suddivisione in *cluster*, permettendo di identificare la robustezza della struttura di raggruppamento proposta dai diversi metodi di analisi. L'indice è calcolato come segue

$$silhouette = \frac{(b - a)}{\max(a, b)} \quad (4.4)$$

dove “*a*” rappresenta la distanza media tra ciascun punto all'interno di un *cluster* e “*b*” indica la distanza media tra tutti i gruppi.

I valori dell'indice *silhouette* relativi ai vari scenari di analisi per la determinazione di 5 *cluster* sono presentati in Tabella 4.1. Per il caso gerarchico, si valutano i 4 diversi legami: medio, completo, singolo e il metodo di Ward.

	h-average	h-complete	h-single	h-ward.D2	partitional
euclidean	0.2621	0.0805	0.1632	0.1310	0.2348
sbd	0.4027	0.2451	0.2129	0.2698	0.0472
dtw_basic	0.2076	0.1288	0.1340	0.0971	-0.1333

TABELLA 4.1: Valori indice *silhouette*

La tecnica di *clustering* di serie storiche che restituisce un risultato migliore consiste nell'utilizzo della tecnica gerarchica con legame medio e distanza basata sulla forma. Di seguito vengono riportate le caratteristiche dei *cluster* in termini di grandezza dei gruppi e di distanza media *intra-cluster*. Tale distanza è tipicamente calcolata come la distanza media tra tutte le coppie di serie all'interno di un *cluster*.

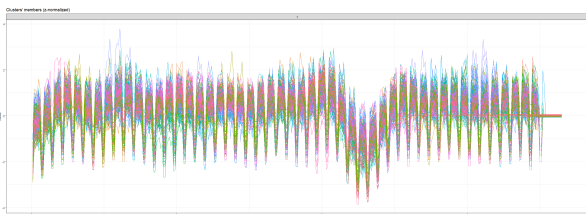
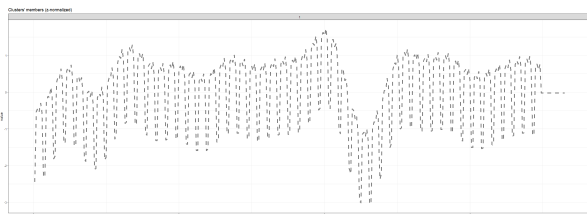
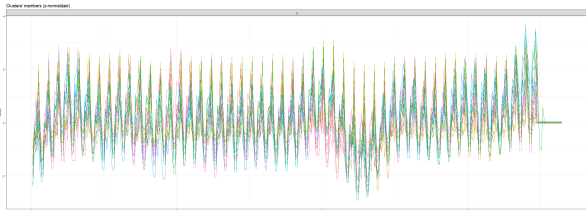
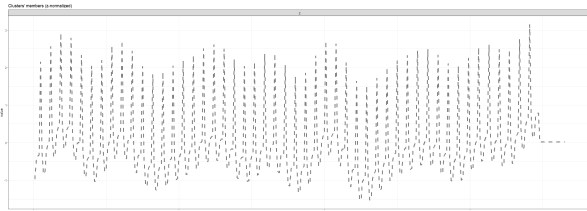
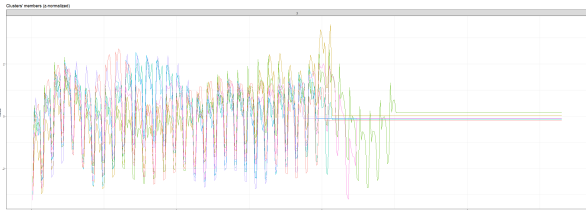
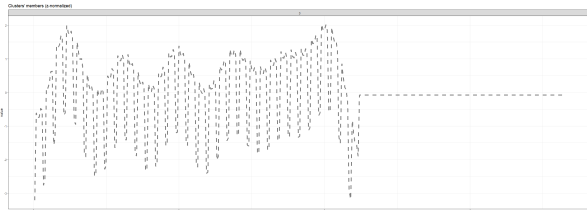
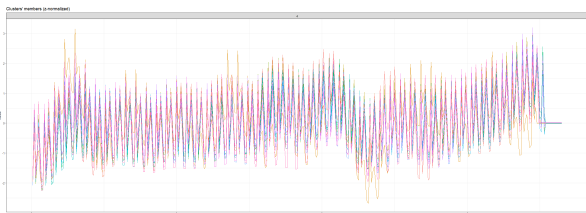
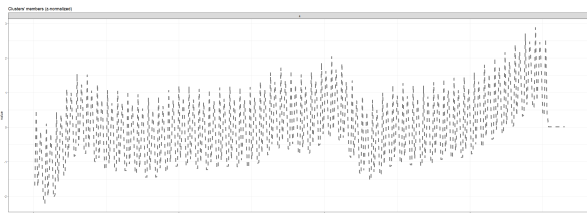
	size	av_dist
1	149	0.1312
2	17	0.2062
3	7	0.1929
4	12	0.0718
5	2	0.0510

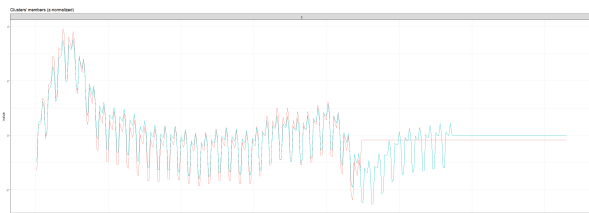
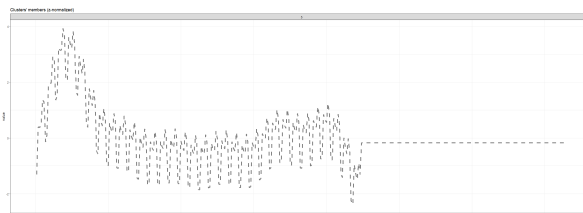
TABELLA 4.2: Informazioni dei *cluster* estratti

Le diverse suddivisioni di serie storiche stagionali ottenute e i relativi centroidi vengono riportati nelle Figure da 4.8 a 4.17.

Dall'esame dettagliato emergono evidenze di un gruppo considerevole di prodotti che manifesta una simile tendenza nella componente stagionale. Le serie storiche sono caratterizzate da fluttuazioni comuni tra i diversi mesi, ma non mostrano una cadenza giornaliera specifica condivisa da diversi articoli, tale da permettere la formazione di un insieme stagionale supplementare. Il secondo *cluster* si contraddistingue per un evidente apice di stagionalità che si manifesta in particolar modo una volta a settimana. I codici selezionati sono identificativi di prodotti esclusivi per la grande distribuzione, fanno quindi riferimento al calendario d'ordine dei clienti appartenenti a questo canale di vendita. Allo stesso modo, il terzo raggruppamento è popolato da serie storiche con un andamento stagionale caratterizzato da due picchi settimanali per la prima metà

dell'anno, e con una stagionalità nulla per i mesi restanti. Il metodo individua la merce le cui vendite vengono interrotte data la decrescita della domanda. Nel caso del quarto gruppo, emerge un rilevante picco stagionale settimanale, accompagnato da un trend generale in crescita che interessa l'ultimo trimestre dell'anno. L'algoritmo seleziona i codici prodotto che, all'interno del gruppo più ampio della grande distribuzione, vengono ordinati unicamente dal cliente Bricoman. Nel quinto *cluster* si osserva un andamento pressochè identico tra i due articoli selezionati. Si tratta di due tasselli in nylon contraddistinti da una fase promozionale per i primi due mesi dell'anno, seguita da un declino delle richieste che porta alla chiusura dei codici.

FIGURA 4.8: Serie stagionalità *cluster 1*FIGURA 4.9: Centroide *cluster 1*FIGURA 4.10: Serie stagionalità *cluster 2*FIGURA 4.11: Centroide *cluster 2*FIGURA 4.12: Serie stagionalità *cluster 3*FIGURA 4.13: Centroide *cluster 3*FIGURA 4.14: Serie stagionalità *cluster 4*FIGURA 4.15: Centroide *cluster 4*

FIGURA 4.16: Serie stagionalità *cluster 5*FIGURA 4.17: Centroide *cluster 5*

4.7 Confronti Prophet e SO99+

Nel corso delle fasi analitiche precedenti, il metodo Prophet e la tecnica di *clustering* basata sulla forma hanno manifestato una chiara competitività nell'analisi dei dati giornalieri forniti da fischer Italia. La scelta di adottare tali approcci è motivata dalla loro efficacia nell'interpretare e prevedere modelli per dati temporali. L'obiettivo di questa sezione è di utilizzare i risultati derivanti da queste metodologie come base per un confronto con il sistema di pianificazione attualmente in uso in azienda.

Nel contesto dell'analisi delle prestazioni dei diversi metodi impiegati, la metrica di confronto prioritaria è rappresentata dall'accuratezza previsiva (*forecast accuracy*). Questa scelta è motivata dalla sua adozione come misura principale all'interno dell'ambito aziendale di fischer Italia. L'accuratezza viene calcolata come $1 - \text{MAPE}$ (*Mean Absolute Percentage Error*)

$$FA = 1 - \frac{|Forecast_m^{lag} - Actual_m|}{Actual_m}, \quad (4.5)$$

dove la quantità *Forecast* raccoglie le previsioni, il valore *Actual* rappresenta i dati osservati, m il mese di riferimento e *lag* specifica a quanti periodi temporali precedenti alla previsione ci si riferisce nell'ambito dell'insieme di stima.

Un valore più elevato di *forecast accuracy* suggerisce una minore discrepanza tra le previsioni e i dati reali, indicando quindi una maggiore affidabilità del metodo adottato. Nel caso in cui si ottenga un risultato negativo, si considera un'accuratezza pari a zero.

In questo confronto si analizzano quattro scenari distinti.

- Metodo Prophet con insieme di stima ampio dal 2012: questa configurazione analizza l'intero insieme di dati disponibili, dal 2012 al 2022, per prevedere l'anno successivo. L'obiettivo è di valutare le prestazioni dell'algoritmo quando ha a disposizione un ampio arco temporale per l'apprendimento e la previsione

- Metodo Prophet con insieme di stima ristretto dal 2019: questa variante esamina l'insieme di stima a partire dal 2019. Questo approccio viene adottato per stabilire un confronto equo con il sistema aziendale SO99+, che per limitazioni di natura tecnologica, considera fino a 5 anni di storia. Questa configurazione mira a valutare la capacità del metodo Prophet di adattarsi e prevedere in condizioni di insieme di stima più limitate
- SO99+ senza indicazioni sugli indici stagionali: il sistema aziendale viene utilizzato senza considerare gli indici derivanti dal *clustering* stagionale
- SO99+ considerando gli indici stagionali proposti dall'analisi dei *cluster*: l'obiettivo è di valutare come l'introduzione di questi indici stagionali personalizzati possa influenzare la precisione delle previsioni rispetto al sistema SO99+ senza tali indicazioni.

Questi quattro scenari forniscono una panoramica completa delle prestazioni di Prophet e SO99+ in varie condizioni e configurazioni.

Si considerano i dati a disposizione a livello aggregato di famiglia di prodotto "SBU 501". Questa prospettiva a livello di Volume, inteso come uno dei Quattro Pilastri fondamentali del processo di S&OP, fornisce un quadro generale della domanda e della produzione nel medio-lungo periodo con orizzonte mensile.

In un primo confronto, l'elemento *forecast* (4.5) di ciascun mese si calcola considerando un insieme di stima che termina il 31 dicembre 2022. I risultati ottenuti non sono utili per la pianificazione operativa ma per l'organizzazione del *budget*.

Nella Tabella 4.3 si riportano i risultati percentuali delle *forecast accuracy* per ciascun mese. Per il metodo Prophet, in grado di gestire dati giornalieri, si aggregano i valori reali e previsti a seconda del mese. Si mettono in evidenza le combinazioni tra mese e metodo che restituiscono valori superiori all'85%, segnali di un'accuratezza previsiva molto alta. Il mese di dicembre è escluso dall'analisi per la limitata disponibilità di osservazioni al momento dell'estrazione della base dati (6 dicembre 2023), che avrebbe potuto influire sulla validità rappresentativa del mese all'interno dei risultati. Le righe finali della tabella restituiscono delle informazioni riassuntive delle diverse situazioni prese in considerazione.

Osservando la tabella di sintesi, si possono raccogliere diversi punti di riflessione.

Mese	Prophet 2012	Prophet 2019	SO99+ no <i>cluster</i>	SO99+ con <i>cluster</i>
01-2023	78,6%	80,0%	89,7%	88,6%
02-2023	86,7%	70,8%	82,6%	80,5%
03-2023	87,9%	66,0%	77,3%	66,2%
04-2023	93,9%	61,2%	62,0%	60,3%
05-2023	91,3%	88,0%	54,1%	58,6%
06-2023	81,9%	98,2%	72,8%	71,3%
07-2023	80,0%	62,4%	83,0%	83,1%
08-2023	83,6%	80,8%	62,3%	57,7%
09-2023	99,2%	74,7%	88,3%	87,2%
10-2023	93,1%	69,3%	89,3%	89,1%
11-2023	94,3%	77,1%	82,7%	83,5%

Minimo	78,6%	61,2%	54,1%	57,7%
Media	88,2%	75,3%	76,7%	75,1%
Massimo	99,2%	98,2%	89,7%	89,1%

TABELLA 4.3: Valori dell'accuratezza previsiva nei quattro scenari

L'implementazione del metodo Prophet, valutato in un contesto di stima più ampio, si traduce in una notevole riduzione delle discrepanze tra i dati effettivi e le previsioni rispetto alla tecnica aziendale standard basata su SO99+: il modello permette un miglioramento dell'accuratezza previsiva minima, media e massima rispettivamente di 24.5, 11.5 e 9.5 punti percentuali. È interessante notare come il metodo si mostri particolarmente reattivo quando si considerano dati storici più estesi, attenuando gli impatti sulla domanda derivanti dalla pandemia globale e mitigando gli effetti anomali di rallentamento. Le percentuali di precisione più basse, invece, si osservano nei mesi iniziali dell'anno, in passato affette da attività promozionali e cambi di listino che hanno generato variazioni nel comportamento di acquisto dei clienti. Anche i mesi estivi sono soggetti a chiusure variabili che impattano sulla prevedibilità.

Dopo aver determinato i *cluster* in base alla stagionalità giornaliera, l'informazione viene trasmessa al sistema aziendale che determina i relativi indici di stagionalità, successivamente utilizzati per il calcolo della previsione. L'accuratezza di SO99+ mostra una leggera diminuzione sul valore medio complessivo rispetto alla situazione in cui non si considerano i gruppi predefiniti. Questo peggioramento nella prestazione può essere attribuito al grado di specificazione della stagionalità: i *cluster* sono estratti sulla base di serie storiche di stagionalità con un intervallo giornaliero, mentre SO99+ opera su un

arco temporale mensile. L'incapacità di SO99+ nel catturare e rispondere a informazioni con tale livello di dettaglio rappresenta una limitazione chiave.

4.7.1 Previsioni un passo in avanti

Dal punto di vista della pianificazione operativa, risulta più utile effettuare una previsione seguendo la logica "un passo in avanti" (*one-step-ahead forecasts*). Questo calcolo consiste nel fornire un piano previsivo per il primo periodo successivo all'ultima osservazione presente nell'insieme di stima. Al passaggio successivo, si integra l'insieme con un nuovo dato e si calcola nuovamente la previsione. Si itera la procedura aggiungendo di volta in volta un'osservazione. Al termine, l'insieme di stima conterrà un numero di osservazioni pari al numero di previsioni prodotte. Con riferimento alla formula dell'accuratezza 4.5, si considera il parametro $lag = 1$. Nello specifico, ciascuna previsione mensile viene calcolata sulla base di un insieme di stima che termina il mese precedente al periodo in considerazione.

I risultati dell'accuratezza previsiva con $lag = 1$ per tutti i casi di interesse sono riportati in Tabella 4.4, evidenziando i casi di *accuracy* superiori all'85%.

Mese	Prophet 2012	Prophet 2019	SO99+ no <i>cluster</i>	SO99+ con <i>cluster</i>
01-2023	78,6%	79,6%	89,7%	88,6%
02-2023	87,5%	70,6%	73,5%	67,6%
03-2023	88,9%	66,4%	63,0%	51,7%
04-2023	94,4%	64,6%	48,3%	55,0%
05-2023	91,4%	96,9%	52,6%	56,1%
06-2023	87,2%	94,7%	74,4%	76,1%
07-2023	74,5%	65,7%	78,8%	76,6%
08-2023	95,8%	86,5%	13,3%	66,9%
09-2023	91,2%	82,3%	96,3%	98,6%
10-2023	84,3%	77,6%	99,3%	99,4%
11-2023	91,5%	87,3%	88,3%	95,9%
Minimo	74,5%	64,6%	13,3%	51,7%
Media	87,7%	79,3%	70,7%	75,7%
Massimo	95,8%	96,9%	99,3%	99,4%

TABELLA 4.4: Valori dell'accuratezza previsiva con $lag = 1$

Il metodo Prophet con insieme di stima più esteso si distingue, anche in questo caso, per l'ottima capacità predittiva. Rispetto al caso illustrato nel Capitolo 4.7, si

notano miglioramenti previsivi che interessano i mesi estivi. Focalizzando l'attenzione sul confronto equo tra il modello Prophet adattato utilizzando dati raccolti dal 2019 e il sistema previsivo aziendale senza indicazioni stagionali, il primo metodo supera il secondo in termini di prestazioni previsionali minime e medie rispettivamente di 61.2 e 17 punti percentuali. Questo conferma la competitività di Prophet nell'attuale contesto aziendale, impostato sul *one-step-ahead forecasts*: la previsione ottenuta sarà il punto di partenza per la pianificazione operativa di tutta la *supply chain*.

L'informazione estratta dal *clustering* stagionale e fornita al sistema SO99+ consente un miglioramento medio dell'accuratezza previsiva pari a 5 punti percentuali rispetto al caso standard. Tuttavia, questo dato è fortemente influenzato dai risultati relativi al mese di agosto: SO99+ nella configurazione standard restituisce un'accuratezza molto bassa che riduce il valore medio di precisione di quasi 6 punti percentuali.

Si confrontano le prestazioni a livello settimanale dei quattro diversi scenari, considerando la specificazione mensile $lag = 1$. Il sistema aziendale si limita a suddividere equamente la previsione mensile per il numero di settimane all'interno del mese. Il metodo Prophet, invece, restituisce un piano previsivo più accurato, aggregando le previsioni giornaliere in base alla settimana di riferimento. Si calcola l'accuratezza delle previsioni escludendo dall'analisi le settimane di dicembre per l'insufficienza di dati a disposizione.

Nella Tabella 4.5 si raggruppano le informazioni principali dei confronti.

	Prophet 2012	Prophet 2019	SO99+ no cluster	SO99+ con cluster
Minimo	23,3%	27,0%	9,6%	26,7%
Media	74,1%	70,8%	59,4%	61,3%
Massimo	99,5%	98,87%	99,9%	99,6%

TABELLA 4.5: Informazioni dell'accuratezza previsiva settimanale con $lag = 1$

In ciascuno dei quattro scenari considerati, si riscontrano valori di accuratezza settimanale pari a zero, sinonimo di una discrepanza sostanziale tra i valori reali e quelli previsti. I risultati nulli non vengono riportati nella Tabella 4.5 al fine di fornire un'indicazione più rilevante. Tutti gli adattamenti elaborano un livello di precisione nullo per la prima settimana dell'anno. In particolare, l'algoritmo SO99+ manifesta questo comportamento carente anche nelle ultime settimane di marzo, aprile e maggio, così come nel mese di agosto. Verso la fine del mese e durante la pausa estiva, si osserva

generalmente una diminuzione delle richieste di prodotti, apparentemente, di difficile previsione per il sistema aziendale.

La metodologia Prophet che interpreta i dati dal 2012 si conferma, mediamente, il miglior adattamento ai dati. Il metodo adottato da fischer Italia restituisce risultati migliori rispetto al caso standard se si introducono nel processo le informazioni derivanti dai *cluster* di stagionalità. In generale, Prophet gestisce i dati in modo ottimale grazie alla sua capacità di affrontarli a livello giornaliero, garantendo allo stesso tempo una gestione efficace anche a livello settimanale. D'altra parte, SO99+, distribuendo equamente le previsioni tra le settimane, mostra dei limiti nel conferire il giusto peso a ciascuna settimana.

Il superamento del vincolo di *bucket* mensile, e pseudo-settimanale, di SO99+ riscontrabile nella gestione di dati giornalieri tramite Prophet rivela un potenziale molto interessante in ambito aziendale. Una previsione dettagliata a livello giornaliero aiuta nell'ottimizzazione delle risorse, nella pianificazione delle attività e consente un presidio più accurato e potenzialmente rilevante per la gestione del *business*.

4.7.2 Caso giornaliero

Il sistema SO99+ non fornisce un piano previsivo giornaliero, perciò si analizzano solo i risultati delle *forecast accuracy* derivanti dall'implementazione di Prophet. Come insieme di stima si considerano i dati a disposizione dal 2012, superando il limite di capacità di cinque anni dell'algorithm aziendale.

Si analizza l'attinenza tra le previsioni e i valori osservati da gennaio a novembre 2023, con un *lag* mensile pari a 1. Dei 334 giorni a disposizione, si tralasciano i sabati e le domeniche, il Capodanno, l'Epifania, il Lunedì dell'Angelo, la Festa dei Lavoratori e la festa di Ognissanti. Con riferimento ai 235 giorni rimanenti, i valori di accuratezza previsiva si distribuiscono come nella Tabella 4.6.

	Numerosità
$FA \geq 85\%$	75
$60 \leq FA < 85\%$	78
$FA < 60\%$	82

TABELLA 4.6: Distribuzione dell'accuratezza previsiva giornaliera con *lag* = 1

Prophet restituisce delle previsioni con un'accuratezza superiore al 60% per la maggior parte dei casi (65%), di cui circa la metà mostra una precisione della previsione superiore all'85%. Questo risultato evidenzia la robustezza del metodo, nonostante il livello di dettaglio giornaliero sia molto sfidante. I livelli di accuratezza non presentano una cadenza regolare, ma le principali difficoltà di previsione si riscontrano nei primi e negli ultimi giorni del mese. In questi intervalli, infatti, le quantità di pezzi richieste risulta molto variabile. Anche in questo caso, il comportamento è influenzato da variabili esterne quali i termini di pagamento e le politiche di gestione delle scorte dei clienti.

4.7.3 *Topseller*

La gamma di prodotti offerta dall'azienda fischer Italia è estremamente diversificata e propone un'ampia varietà di articoli. Tuttavia, il fatturato complessivo della società è principalmente influenzato da un gruppo ristretto di prodotti, i quali contribuiscono a generare circa il 40% del totale dei ricavi. Questo insieme selezionato di 31 articoli comprende i veri e propri *topseller* dell'azienda, che guidano il successo finanziario complessivo.

I prodotti offerti possono essere categorizzati in base alla loro destinazione commerciale, che può essere orientata verso altre aziende (*business to business* o B2B) o verso il consumatore finale (*business to consumer* o B2C). Un'ulteriore segmentazione è legata alla presenza di alluminio nel prodotto: questo materiale è soggetto ad acquisti speculativi dovuti alle oscillazioni di prezzo. Di seguito sono elencati gli insiemi identificati.

- Categoria 1: B2B, settore cantieristico caratterizzato da grosse commesse episodiche
- Categoria 2: B2B, condizionato dalle logiche di produzione del cliente (OEM), con lotti e chiusure aziendali abbastanza stabili
- Categoria 3: B2B, condizionato dalla logica di produzione del cliente (OEM), con lotti e chiusure aziendali variabili
- Categoria 4: B2C, con consumo standard
- Categoria 5: B2C, con effetto superbonus 110% che ha aumentato drasticamente i volumi del 2022

- Categoria 6: B2C, con effetto superbonus 110%, prodotti che contengono alluminio
- Categoria 7: B2C, con effetto superbonus 110%, prodotti che contengono alluminio, caratterizzati da perdite di vendite intra-Gruppo nel corso del 2022/2023
- Categoria 8: B2C, con effetto superbonus 110% moderato.

Si analizzano le serie storiche dei *topseller* per comprendere le caratteristiche della base dati oggetto dello studio. Nello specifico, si verifica se le condizioni utilizzate nel capitolo precedente per selezionare gli articoli da analizzare sono soddisfatte: almeno 300 frequenze dal 2022 e data di inizio della serie storica antecedente al 2018. L'analisi mira a determinare la possibilità di ottenere risultati stabili o la necessità di considerare questi aspetti critici in fase di interpretazione dei risultati. Gli articoli classificati come "business to business" registrano meno di 300 frequenze negli ultimi due anni a disposizione. Al contrario, la quasi totalità dei 6 articoli con una serie storica che inizia dopo il 2018 è inclusa nel gruppo destinato al "business to consumer".

Nella Tabella 4.7 si riportano i valori medi dell'accuratezza delle previsioni mensile considerando un *lag* pari a 1. Per analizzare i dati si considera il metodo Prophet e il sistema SO99+ con e senza indici di stagionalità. Per ciascun articolo, si mette in evidenza l'adattamento che determina la miglior accuratezza.

L'algorithmo SO99+, potenziato con informazioni sulle caratteristiche stagionali, produce i risultati migliori in termini di accuratezza previsiva media mensile per metà dei gruppi considerati. Nella Categoria 1, si registra un'accuratezza media del 37,6%, mentre nella Categoria 2 l'algorithmo mostra una prestazione più promettente, con un risultato medio pari a 63,9%. L'insieme 6 evidenzia una precisione media del 60,9% mentre nella Categoria 8 si raggiunge il 79,4% di accuratezza media, indicando un'elevata affidabilità in questo specifico ambito. Questi risultati, benché variabili, forniscono spunti interessanti per l'analisi e l'ottimizzazione dell'algorithmo SO99+.

Nell'analizzare le prestazioni del metodo Prophet in questi insiemi, invece, è importante considerare le sfide legate alle specifiche caratteristiche delle serie storiche coinvolte. La bassa accuratezza riscontrata può essere ricondotta a diversi fattori chiave presenti in tali contesti. Alcune categorie mostrano poche osservazioni negli ultimi anni a disposizione. La limitata disponibilità di informazioni recenti incide notevolmente sulla capacità di adattamento del modello, rendendo difficile la fase di previsione.

Articolo	Categoria	Prophet	SO99+ con <i>cluster</i>	SO99+ no <i>cluster</i>
1	1	22,0%	37,6%	31,1%
2	2	50,4%	59,9%	57,1%
3	2	40,1%	60,3%	57,3%
4	2	65,9%	71,4%	67,4%
5	3	54,5%	54,8%	45,7%
6	3	63,3%	55,0%	53,3%
7	4	78,5%	84,6%	87,3%
8	4	87,4%	81,1%	80,0%
9	4	91,1%	87,8%	89,5%
10	4	84,7%	87,4%	89,6%
11	4	87,2%	79,7%	80,8%
12	4	87,0%	86,4%	88,6%
13	4	81,7%	82,0%	83,3%
14	4	84,9%	81,3%	85,1%
15	4	65,8%	67,0%	69,8%
16	4	86,2%	84,0%	84,9%
17	4	86,5%	79,9%	80,2%
18	4	83,7%	78,6%	80,0%
19	4	86,7%	76,2%	77,8%
20	5	53,2%	54,3%	56,8%
21	6	42,7%	50,4%	56,0%
22	6	41,7%	70,1%	66,5%
23	6	57,7%	74,6%	72,7%
24	6	47,8%	58,8%	56,7%
25	6	33,8%	48,1%	43,4%
26	6	43,5%	63,5%	62,8%
27	7	49,0%	59,8%	61,7%
28	7	41,3%	73,9%	69,5%
29	7	65,6%	71,3%	72,7%
30	7	56,6%	70,1%	73,1%
31	8	66,7%	79,4%	79,3%

Minimo	22,0%	37,6%	31,1%
Media	64,1%	70,0%	69,7%
Massimo	91,1%	87,8%	89,6%

TABELLA 4.7: Valore medio dell'accuratezza previsiva mensile con $lag = 1$ dei *top-seller*

Inoltre, la presenza di serie storiche che iniziano negli anni più recenti può introdurre complicazioni aggiuntive. Tali codici possono risentire dell'effetto della pandemia, che ha generato cambiamenti non trascurabili nei modelli di domanda e offerta, oppure limitare la capacità di apprendimento degli andamenti da parte del metodo Prophet.

Un altro aspetto rilevante è riscontrabile nei casi in cui si registra un aumento delle quantità di pezzi richieste fino al 2022, ultimo anno dell'insieme di stima, seguito da un crollo consistente. Questa dinamica rappresenta una variazione di trend rilevante, a cui il metodo Prophet non sembra reagire adeguatamente. La sua incapacità di catturare e adattarsi a queste variazioni repentine influenza negativamente le sue prestazioni.

Il codice "1" costituisce un esempio di andamento che risulta di difficile interpretazione per il metodo Prophet. Viene fornita di seguito la scomposizione nelle componenti di trend, stagionalità e componente residua per l'articolo preso in considerazione. Si procede al confronto tra il caso in cui l'insieme di stima si conclude nel 2022 e quello in cui termina nel 2023.

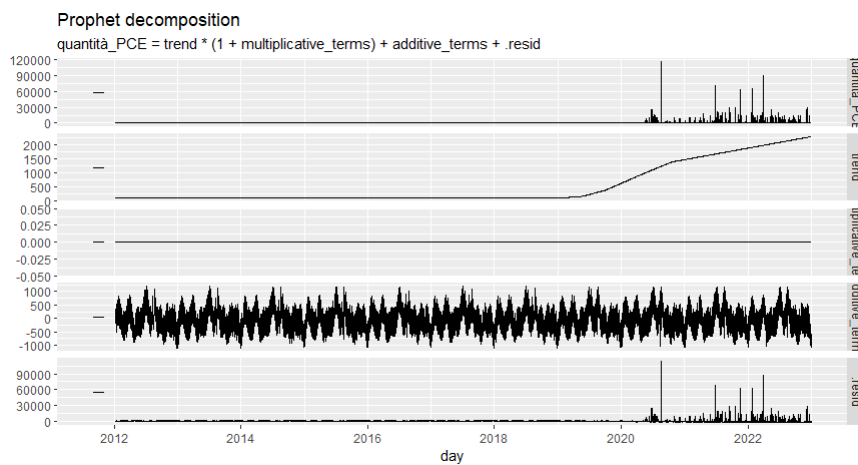


FIGURA 4.18: Scomposizione con Prophet, insieme di stima fino al 2022

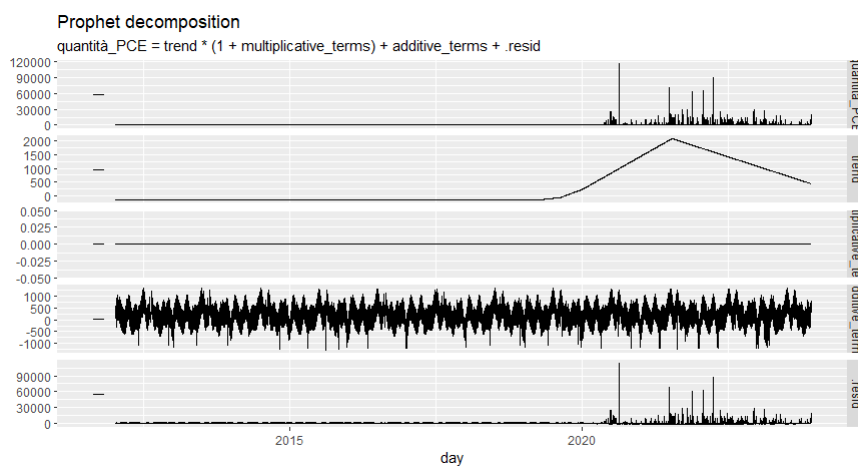


FIGURA 4.19: Scomposizione con Prophet, insieme di stima fino al 2023

Come si nota dalle Figure 4.18 e 4.19, l'andamento dell'articolo evidenzia una netta fase di crescita, che si traduce in una brusca diminuzione nel corso del 2023.

Negli scenari in cui il consumo di materiale segue un andamento standard, come nella Categoria 4, il modello Prophet emerge come la selezione più appropriata per stabilire un piano previsivo. La capacità di catturare le tendenze ricorrenti, insieme alla gestione efficace delle componenti stagionali, lo rendono particolarmente adatto a fornire stime accurate e affidabili.

Capitolo 5

Conclusioni

La presente tesi si è posta l'obiettivo di affrontare parte delle sfide connesse alla gestione del bilanciamento tra domanda e offerta in un contesto aziendale, proponendo approcci innovativi. In risposta alle specifiche richieste di Fischer Italia, sono state sviluppate strategie di previsione giornaliera e di categorizzazione dei prodotti in base alle loro fluttuazioni stagionali quotidiane. Nell'affrontare tali sfide, si è proposto di superare il limite di dettaglio mensile imposto dal sistema SO99+ in uso presso l'azienda, aprendo la strada a un'approfondita analisi della stagionalità dei dati per una gestione più fine e reattiva dei prodotti.

Sono stati comparati i metodi STL e Prophet, considerando sia una che due stagionalità. L'adozione di una sola stagionalità si è rivelata limitante, specialmente trattando dati giornalieri che, come era naturale aspettarsi, presentano diverse sfaccettature stagionali. In termini di metrica RMSE, Prophet con doppia stagionalità ha prodotto i risultati migliori. Il metodo si caratterizza per la sua notevole flessibilità nella configurazione dei parametri, garantendo la possibilità di affrontare aspetti specifici quali l'impatto delle festività e delle pause aziendali sulla domanda di prodotti. L'analisi delle previsioni di Prophet ha evidenziato, inoltre, intervalli di confidenza più limitati rispetto al caso MSTL, indicando una minore incertezza previsiva.

La scelta della tecnica di raggruppamento è ricaduta nel metodo gerarchico con distanza basata sulla forma. A parità di numero di *cluster*, questo metodo ha restituito il valore di *silhouette* più alto. Questo risultato è riconducibile alla sensibilità della

metrica. La rilevanza di questa scelta risiede nella sua capacità di considerare attentamente la forma delle serie storiche, risultando particolarmente sensibile a variazioni strutturali nei dati. Questo aspetto diventa cruciale quando si analizzano andamenti complessi, come quelli tipici delle dinamiche giornaliere della domanda. La distanza euclidea, sebbene ampiamente utilizzata, potrebbe non cogliere efficacemente le sottili variazioni nella forma delle curve temporali. D'altra parte, la distanza basata sul flusso dinamico temporale, pur essendo flessibile, potrebbe sovrastimare la somiglianza tra curve, considerando variazioni temporali che potrebbero non avere una corrispondenza diretta.

Sono stati esaminati i risultati del sistema aziendale SO99+ utilizzando, anche, l'informazione appena ricavata a livello di *cluster*. La nuova elaborazione è stata confrontata con i risultati ottenuti da Prophet. Per garantire un confronto equo, sono stati considerati due diversi insiemi di stima per Prophet: uno costituito da una serie storica di 5 anni dal 2019, per allinearsi alla configurazione standard di SO99+, e una serie più ampia con dati a partire dal 2012, al fine di valutare il metodo con la totalità di dati a disposizione. La metrica utilizzata per il confronto è stata la *forecast accuracy* (accuratezza previsiva), in linea con gli indicatori di prestazione utilizzati in azienda.

In una prima analisi, che replica l'attività aziendale di definizione del *budget*, si è mantenuto fisso l'insieme di stima, sono state calcolate le previsioni per ciascun mese del 2023. Il metodo Prophet, con un insieme di stima più ampio, ha dimostrato di ottenere i migliori risultati medi di accuratezza, registrando anche la più bassa variabilità tra i valori minimi e massimi. La *forecast accuracy* media è stata migliore di circa l'11.5% rispetto al valore ottenuto con SO99+. Tuttavia, nell'analisi di confronto con dati storici a partire dal 2019, l'algoritmo aziendale si è rivelato leggermente superiore, ma senza differenze sostanziali tra i metodi.

Da una prospettiva di pianificazione delle attività operative di *fischer*, è più utile valutare le previsioni con un *lag* temporale di un mese. Ciò implica che ciascuna previsione mensile venga calcolata sulla base di un insieme di stima che termina il mese precedente al periodo considerato. Prophet ha restituito risultati medi più accurati sia nel caso di un insieme di stima ampio che ristretto, dimostrando un rendimento migliore per un contesto applicativo in azienda.

Successivamente, si è proceduto a una valutazione più dettagliata delle previsioni, analizzando i risultati settimanali. In questo contesto, Prophet ha confermato la sua capacità di generare previsioni più accurate rispetto al sistema SO99+ standard che non supporta questo livello di dettaglio. Nel merito, si è evidenziato un miglioramento dell'11.4% nel considerare i dati a partire dal 2019 e del 14.7% considerando l'insieme di stima più ampio.

In tutti i confronti effettuati, l'impiego delle informazioni derivanti dai *cluster* di stagionalità non ha dimostrato di contribuire in maniera rilevante al miglioramento delle prestazioni del sistema aziendale. Questo risultato potrebbe essere riconducibile al fatto che i *cluster* sono stati calcolati considerando dati giornalieri, mentre SO99+ opera su scala mensile. Da ciò si deduce che il sistema potrebbe non essere in grado di beneficiare nella sua elaborazione di questa informazione troppo dettagliata.

Uno degli obiettivi di questa tesi consisteva nell'ottenere un piano previsivo giornaliero, il quale è stato realizzato attraverso l'applicazione del metodo Prophet. L'analisi dell'accuratezza delle previsioni giornaliere evidenzia un buon adattamento del metodo, poiché nel 65% dei casi si registrano valori di *forecast accuracy* superiori al 60%, con circa la metà di essi con precisione superiore all'85%. Si tratta di un risultato molto soddisfacente per l'azienda.

Nella prima fase dell'analisi, si è operato a un livello aggregato, considerando la famiglia di prodotto denominata "501": sistemi di fissaggio in nylon. Nella fase conclusiva, l'attenzione si è focalizzata sui prodotti *topseller* dell'azienda, ossia quelli che contribuiscono al 40% del fatturato totale. L'analisi è stata condotta con l'orientamento della pianificazione operativa, con un *lag* pari a 1. Prophet ha dimostrato una buona reattività nei casi in cui la serie storica di un articolo ha inizio almeno 6 anni prima e presenta un numero sostanziale di frequenze negli ultimi due anni di analisi. Al contrario, nelle situazioni che non soddisfano tali caratteristiche o evidenziano un repentino cambiamento di tendenza nel corso del 2023, Prophet sembra non reagire in modo ottimale.

In conclusione, le varie analisi delineano Prophet come uno strumento di grande potenza per supportare la creazione di previsioni in azienda. Attraverso confronti approfonditi con l'attuale sistema, si è rilevato un incremento nell'accuratezza delle previsioni. Questo permette di avere una visione più nitida delle dinamiche di domanda e consente

di orientare al meglio le attività di gestione dell'offerta. L'introduzione di questo modello si traduce in vantaggi tangibili per la pianificazione operativa e consente un utilizzo più efficace delle risorse, posizionandosi come un elemento fondamentale nella strategia e nel processo di elaborazione delle previsioni.

Appendice

Settimana	Prophet 2012	Prophet 2019	SO99+ no cluster	SO99+ con cluster
1 (gen)	42,8%	42,0%	0,0%	0,0%
2 (gen)	27,0%	27,0%	55,1%	55,6%
3 (gen)	71,6%	75,1%	71,4%	72,1%
4 (gen)	0,0%	0,0%	54,1%	52,7%
5 (gen-feb)	73,2%	78,4%	84,5%	83,4%
6 (feb)	56,2%	51,8%	91,4%	95,7%
7 (feb)	86,2%	66,6%	51,8%	44,9%
8 (feb)	93,2%	86,5%	53,2%	46,3%
9 (feb-mar)	57,2%	86,8%	36,9%	29,3%
10 (mar)	94,5%	80,0%	47,8%	35,2%
11 (mar)	67,6%	54,7%	77,2%	67,1%
12 (mar)	64,9%	42,8%	56,4%	44,5%
13 (mar-apr)	80,4%	63,6%	0,0%	0,0%
14 (apr)	66,5%	37,6%	99,9%	95,5%
15 (apr)	84,5%	52,0%	55,3%	61,7%
16 (apr)	90,5%	80,2%	43,7%	50,6%
17 (apr)	23,2%	50,8%	0,0%	0,0%
18 (mag)	72,4%	63,1%	81,7%	84,5%
19 (mag)	84,6%	73,9%	96,9%	99,4%
20 (mag)	46,4%	67,2%	22,5%	26,7%
21 (mag)	77,0%	89,4%	57,7%	61,1%
22 (mag-giu)	40,2%	46,5%	0,0%	0,0%
23 (giu)	86,3%	92,7%	68,2%	70,0%
24 (giu)	65,5%	72,6%	49,1%	51,1%
25 (giu)	97,7%	94,9%	80,1%	81,7%
26 (giu-lug)	99,5%	92,4%	74,5%	76,2%

Settimana	Prophet 2012	Prophet 2019	SO99+ no cluster	SO99+ con cluster
27 (lug)	70,4%	78,7%	69,4%	73,0%
28 (lug)	46,3%	41,2%	40,2%	39,1%
29 (lug)	89,5%	77,1%	69,6%	67,7%
30 (lug)	86,5%	74,1%	77,0%	74,9%
31 (lug-ago)	75,6%	65,4%	99,9%	97,2%
32 (ago)	91,7%	90,2%	0,0%	30,3%
33 (ago)	88,4%	72,1%	0,0%	0,0%
34 (ago)	94,2%	84,6%	0,0%	29,5%
35 (ago-set)	81,3%	97,3%	9,6%	64,2%
36 (set)	57,6%	51,9%	75,3%	73,6%
37 (set)	85,2%	91,8%	70,6%	73,5%
38 (set)	92,0%	83,2%	98,2%	99,6%
39 (set-ott)	67,6%	85,0%	44,7%	48,2%
40 (ott)	80,2%	68,3%	95,0%	94,8%
41 (ott)	70,6%	66,1%	96,0%	96,2%
42 (ott)	96,2%	98,9%	64,3%	64,1%
43 (ott)	84,8%	79,2%	95,4%	95,3%
44 (ott-nov)	95,9%	87,2%	82,0%	90,0%
45 (nov)	79,8%	71,0%	86,3%	80,5%
46 (nov)	95,3%	88,2%	90,6%	98,0%
47 (nov)	84,3%	83,8%	88,0%	82,1%
48 (nov-dic)	94,8%	92,9%	90,8%	84,6%

TABELLA 1: Valori dell'accuratezza previsiva settimanale con $lag = 1$

Giorno	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno
1	-	97,5%	11,8%	-	-	14,1%
2	0,0%	98,5%	93,3%	-	96,3%	7,0%
3	61,8%	67,7%	96,6%	73,4%	69,3%	-
4	26,2%	-	-	58,4%	68,6%	-
5	27,5%	-	-	44,8%	57,5%	88,4%
6	-	36,5%	86,0%	92,6%	-	82,9%
7	-	55,4%	82,6%	57,8%	-	95,7%
8	-	57,4%	81,3%	-	75,9%	60,1%
9	18,8%	79,8%	47,7%	-	77,6%	98,3%
10	28,7%	75,0%	74,1%	-	41,2%	-
11	26,3%	-	-	76,5%	90,8%	-
12	27,7%	-	-	71,5%	55,7%	91,3%
13	35,8%	56,0%	94,3%	86,2%	-	18,9%
14	-	88,6%	29,8%	83,4%	-	63,7%
15	-	93,8%	79,0%	-	71,3%	58,0%
16	78,3%	82,1%	61,6%	-	21,8%	34,4%
17	93,7%	95,5%	85,7%	91,6%	64,6%	-
18	82,4%	-	-	78,1%	16,1%	-
19	63,1%	-	-	74,4%	53,6%	95,6%
20	64,3%	96,7%	86,6%	96,5%	-	56,2%
21	-	89,5%	37,5%	93,5%	-	89,0%
22	-	99,2%	82,0%	-	81,4%	94,8%
23	68,0%	84,4%	69,5%	-	92,3%	99,0%
24	0,0%	69,8%	91,2%	39,6%	36,6%	-
25	0,0%	-	-	1,2%	92,8%	-
26	0,0%	-	-	13,1%	8,8%	68,5%
27	0,0%	3,0%	81,0%	0,0%	-	92,3%
28	-	4,3%	98,2%	36,4%	-	70,7%
29	-	-	80,2%	-	63,1%	74,5%
30	0,0%	-	68,7%	-	0,0%	87,6%
31	0,0%	-	0,0%	-	41,0%	-

Giorno	Luglio	Agosto	Settembre	Ottobre	Novembre
1	-	90,6%	81,4%	-	-
2	-	88,0%	-	9,6%	93,4%
3	34,9%	89,9%	-	80,8%	93,6%
4	0,0%	77,0%	48,0%	50,8%	-
5	96,6%	-	52,8%	90,2%	-
6	89,1%	-	42,6%	63,6%	72,9%
7	98,0%	64,8%	71,9%	-	65,2%
8	-	95,8%	88,9%	-	74,4%
9	-	17,4%	-	94,2%	89,2%
10	37,3%	95,8%	-	86,6%	64,7%
11	14,7%	0,0%	80,0%	68,3%	-
12	88,7%	-	95,9%	68,7%	-
13	91,4%	-	28,7%	52,9%	97,0%
14	84,3%	0,0%	79,6%	-	96,4%
15	-	0,7%	78,1%	-	86,4%
16	-	87,0%	-	81,6%	77,4%
17	84,8%	56,3%	-	91,0%	62,8%
18	88,7%	64,2%	84,5%	98,6%	-
19	41,0%	-	82,2%	79,1%	-
20	42,6%	-	91,0%	66,1%	96,7%
21	81,1%	87,5%	90,6%	-	59,9%
22	-	96,3%	59,6%	-	68,7%
23	-	45,8%	-	9,6%	79,4%
24	79,6%	79,2%	-	35,3%	86,9%
25	81,3%	0,0%	42,9%	31,0%	-
26	94,6%	-	62,4%	81,9%	-
27	92,6%	-	58,0%	99,8%	72,4%
28	73,2%	93,2%	0,7%	-	48,5%
29	-	71,3%	0,0%	-	85,2%
30	-	46,0%	-	95,1%	99,9%
31	46,0%	45,1%	-	95,4%	-

TABELLA 2: Valori dell'accuratezza previsiva giornaliera con $lag = 1$

Bibliografia

- AZZALINI, A. & SCARPA, B. (2009). *Analisi dei dati e data mining*. Springer Science & Business Media.
- BANDARA, K., HYNDMAN, R. J. & BERGMEIR, C. (2021). Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. *arXiv preprint arXiv:2107.13462* .
- CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E. & TERPENNING, I. (1990). Stl: A seasonal-trend decomposition. *J. Off. Stat* **6**, 3–73.
- GIORGINO, T. (2009). Computing and visualizing dynamic time warping alignments in r: the dtw package. *Journal of statistical Software* **31**, 1–24.
- MAY, M. C., KIEFER, L., FREY, A., DUFFIE, N. A. & LANZA, G. (2023). Applying frequency based forecasting for resource allocation. *Procedia CIRP* **120**, 147–152.
- PAPARRIZOS, J. & GRAVANO, L. (2015). k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*.
- PETITJEAN, F. & GANÇARSKI, P. (2012). Summarizing a set of time series by averaging: From steiner sequence to compact multiple alignment. *Theoretical Computer Science* **414**, 76–91.
- TAYLOR, S. J. & LETHAM, B. (2018). Forecasting at scale. *The American Statistician* **72**, 37–45.
- WALLACE, T. F. (2004). *Sales and operations planning: the how-to handbook*. TF Wallace & Co.

