



**Politecnico  
di Torino**

# POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Gestionale – LM/31

Percorso ICT and data analytics

Tesi di Laurea Magistrale

## **Applicazione di algoritmi di machine learning per la previsione degli acquisti e la gestione delle scorte**

**Relatori:**

Prof.re Maurizio Galetto

Prof.ressa Elisa Verna

**Candidato:**

Martina Martire

Anno Accademico 2025-2026

Sessione di Laurea – Marzo 2026



## Sommario

<b>Introduzione</b> .....	<b>4</b>
<b>Capitolo 1 – Il problema della domanda incerta nel settore agroalimentare</b> .....	<b>6</b>
1.1 Il settore agroalimentare secondario .....	6
1.2 La previsione della domanda.....	7
1.3 Obiettivi di sostenibilità: riduzione sprechi e ottimizzazione dei costi .....	9
1.4 La struttura del lavoro .....	10
<b>Capitolo 2 – Stato dell'arte: dai Metodi Statistici al Machine Learning</b> .....	<b>12</b>
2.1 Modelli Statistici e tecniche di Regressione Lineare .....	12
2.2 L'ascesa del <i>Machine Learning</i> e l'efficacia degli algoritmi <i>Ensemble</i> .....	15
2.3 Il ruolo del <i>Deep Learning</i> : Reti Neurali e modelli LSTM .....	16
2.4 Metriche di valutazione e criteri di confronto tra modelli .....	18
<b>Capitolo 3: Metodologia di ricerca</b> .....	<b>20</b>
3.1 Protocollo PRISMA.....	20
3.2 Strategia di Ricerca .....	20
3.2.1 Processo di selezione degli studi .....	22
3.3 Analisi preliminare e raccolta di informazioni .....	24
<b>Capitolo 4 – Interpretazione di primo livello: tassonomia dei risultati</b> .....	<b>26</b>
4.1 Applicabilità e finalità degli algoritmi predittivi nel settore agroalimentare.....	26
4.2 Mappatura lungo la <i>Food Value Chain</i> .....	26
4.3 Classificazione merceologica e cluster di prodotto .....	30
4.4 Analisi comparativa degli algoritmi: utilizzo vs performance .....	33
4.4.1 Valutazione degli algoritmi più utilizzati .....	35
4.4.2 Valutazione degli algoritmi più performanti.....	37
4.5 Caratterizzazione dei dati e architetture informative .....	41
4.6 Verso l'integrazione: introduzione alle correlazioni multidimensionali.....	45
<b>Capitolo 5 – Interpretazione di secondo livello: analisi delle correlazioni</b> .....	<b>47</b>

5.1 Obiettivi dell'analisi multidimensionale .....	47
5.2 Correlazione tra <i>value chain</i> e obiettivi della ricerca.....	47
5.3 Correlazione tra tipologia di prodotto esaminato e obiettivi della ricerca.....	49
5.4 Efficacia degli algoritmi in base al contesto operativo .....	52
5.4.1 Correlazione tra algoritmi analizzati e tipologia di prodotto esaminato nella fase della <i>value chain</i> ....	52
5.4.2 Correlazione tra algoritmi più performanti e tipologia di prodotto esaminato nella fase della <i>value chain</i> .....	56
5.5 Dipendenza tra <i>value chain</i> e caratteristiche dei dati.....	58
<b>Capitolo 6 – Conclusioni</b> .....	<b>60</b>
6.1 Risultati ottenuti.....	60
6.2 Prospettive future e possibili implementazioni .....	65
<b>Riconoscimenti</b> .....	<b>67</b>
<b>Ringraziamenti</b> .....	<b>68</b>
<b>Allegati</b> .....	<b>69</b>
Allegato A – Scheda informativa degli articoli analizzati .....	72
Allegato B – Caratterizzazione dei dati.....	81
Allegato C – Features analizzate negli articoli .....	83
<b>Bibliografia</b> .....	<b>87</b>
<b>Indice delle Figure</b> .....	<b>93</b>
<b>Indice delle Tabelle</b> .....	<b>94</b>



# Introduzione

Il settore agroalimentare rappresenta uno dei pilastri fondamentali dell'economia contemporanea, integrando attività produttive, industriali e distributive essenziali per il sostentamento della popolazione globale. Tuttavia, l'epoca dei mercati stabili può ormai considerarsi conclusa: l'attuale contesto competitivo è caratterizzato da un'elevata complessità e da un marcato grado di incertezza, determinati da fattori quali crisi economiche, cambiamenti climatici e frequenti interruzioni delle catene di approvvigionamento. In tale scenario, la pianificazione efficiente della supply chain assume un ruolo cruciale, in particolare nel settore alimentare secondario, dove la natura deperibile dei prodotti impone una gestione estremamente accurata dei flussi logistici.

Il problema centrale affrontato in questo lavoro riguarda la previsione della domanda in condizioni di incertezza. L'incapacità di stimare correttamente i volumi di vendita genera significative inefficienze: una sovrastima della domanda conduce a surplus produttivi e a un incremento degli sprechi alimentari, mentre una sottostima determina carenze di prodotto e una conseguente perdita di competitività. Sebbene i metodi statistici tradizionali abbiano rappresentato per lungo tempo lo standard di riferimento, essi mostrano oggi evidenti limiti nella gestione della non linearità e dell'elevata dimensionalità dei dati, spesso di natura multivariata, che caratterizzano i mercati moderni.

Il presente lavoro di tesi si propone di analizzare in modo sistematico e approfondito lo stato dell'arte dei modelli di previsione della domanda nel settore agroalimentare, con l'obiettivo di individuare gli approcci di Machine Learning e Deep Learning più efficaci per l'ottimizzazione della gestione delle scorte e la riduzione dell'impatto ambientale. Al fine di garantire rigore metodologico e replicabilità dei risultati, la ricerca è stata condotta seguendo il protocollo PRISMA 2020, che ha consentito la selezione di un corpus finale composto da 52 articoli scientifici pertinenti.

L'indagine non si limita a una semplice ricognizione tecnologica, ma sviluppa una tassonomia strutturata degli studi analizzati, classificandoli lungo le diverse fasi della Food Value Chain (dal processing al retail) e secondo sette specifici cluster di prodotto. Attraverso un'analisi di secondo livello, la tesi esplora le correlazioni multidimensionali tra le fasi della filiera, le tipologie di dati impiegate e le prestazioni degli algoritmi utilizzati, offrendo un quadro organico delle migliori pratiche per l'implementazione di sistemi predittivi resilienti e sostenibili.

In conclusione, il lavoro intende dimostrare come l'integrazione di tecniche avanzate di intelligenza artificiale non costituisca soltanto un progresso tecnologico, ma rappresenti una necessità strategica per garantire un equilibrio duraturo tra efficienza economica e responsabilità ecologica lungo l'intera filiera agroalimentare.

# Capitolo 1 – Il problema della domanda incerta nel settore agroalimentare

## 1.1 Il settore agroalimentare secondario

Il settore agroalimentare rappresenta uno dei pilastri fondamentali dell'economia moderna, in quanto integra attività produttive, industriali e distributive legate all'alimentazione, come affermano Nouran Nassibi e Heba Fasihuddin nella loro analisi *“Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches”* [1]. Esso si articola in più fasi che compongono una filiera complessa, dalla produzione primaria alla trasformazione industriale e, infine, alla distribuzione dei prodotti finiti ai consumatori.

All'interno di questa filiera, il settore agroalimentare secondario – spesso indicato anche come settore alimentare – comprende tutte le attività di trasformazione, conservazione, confezionamento e distribuzione dei prodotti agricoli. Si tratta, quindi, della fase industriale del sistema agroalimentare, nella quale le materie prime provenienti dall'agricoltura, dall'allevamento o dalla pesca vengono trasformate in alimenti destinati al consumo diretto o in semilavorati utilizzabili in ulteriori processi produttivi.

Questa fase assume un ruolo cruciale non solo per la creazione di valore aggiunto, ma anche per l'impatto che esercita in termini di occupazione, innovazione tecnologica e sostenibilità, come viene descritto da Himanshu Sharma in *“Data-Driven Forecasting and Inventory Optimization using Machine Learning Models and Methods”* [2]. Anche Mugdha e Dr. Durgansh Sharma [3] hanno concentrato le loro ricerche nel settore alimentare secondario, nello specifico la loro analisi si sofferma sull'economia circolare e sulla competitività dei sistemi territoriali, promuovendo l'efficienza nell'uso delle risorse e la riduzione degli sprechi.

Dal punto di vista operativo, le imprese appartenenti al comparto secondario agroalimentare comprendono un'ampia varietà di attività, tra cui:

- la macinazione dei cereali e la produzione di farine e derivati;
- la trasformazione del latte e la produzione di formaggi e latticini;
- la lavorazione delle carni e dei prodotti ittici;
- la produzione di conserve vegetali, bevande e prodotti da forno;
- la realizzazione di prodotti surgelati o pronti al consumo.

Queste aziende si collocano al centro della catena di approvvigionamento alimentare, fungendo da ponte tra il settore primario e il mercato finale. Da un lato, dipendono fortemente dalla disponibilità e dalla qualità delle materie prime agricole; dall'altro, sono condizionate dalla variabilità della domanda e dalle esigenze dei consumatori finali.

La complessità gestionale del settore è accentuata dalla natura peribile dei prodotti e dalla necessità di gestire la volatilità dei consumi per evitare perdite economiche, sfide discusse nella revisione sistematica *“Quantitative Models for Forecasting Demand for Perishable Products: A Systematic Review.”* di Jonathan Vinicius Kaizer, Rodrigo Clemente Thom de Souza e Linnyer Beatrys Ruiz Aylon [4]. In questo scenario, l'adozione di tecniche di *machine learning* è diventata necessaria per permettere alle aziende di adattarsi ai rapidi cambiamenti del mercato grazie alla capacità di elaborare grandi volumi di dati in modo efficiente.

In sintesi, il settore agroalimentare secondario svolge un ruolo strategico nel garantire l'equilibrio tra efficienza economica, sicurezza alimentare e sostenibilità ambientale, costituendo un elemento chiave per la resilienza e la competitività dell'intera filiera alimentare.

## 1.2 La previsione della domanda

L'epoca dei mercati stabili appartiene ormai al passato: oggi il contesto competitivo è estremamente complesso e dinamico, caratterizzato da incertezze di varia natura come crisi economiche e monetarie, pandemie, cambiamenti climatici e interruzioni nelle catene di approvvigionamento. Questa crescente complessità riduce la trasparenza dei flussi materiali e informativi, condizione essenziale per la buona riuscita della pianificazione della catena di approvvigionamento.

La gestione risulta particolarmente difficile nel settore alimentare a causa della breve durata di conservazione dei prodotti, richiedendo decisioni precise sui livelli di inventario per evitare rincari operativi, come riportato da E. Anitha nel lavoro *“Effective Food Demand Forecasting Using Machine Learning Algorithms”* [5]. Sebbene comunemente associato agli alimenti, il prodotto deperibile include ogni bene che non può essere immagazzinato a lungo o che diventa rapidamente obsoleto, un'osservazione approfondita da Grigorios Tsoumakas nello studio *“A survey of machine learning techniques for food sales prediction”* [6].

La variabilità e l'incertezza della domanda generano due principali categorie di inefficienze operative:

- Sovrastima della domanda, che comporta surplus produttivi, alti costi di gestione e sprechi;
- Sottostima della domanda, che causa carenze di prodotto, perdita di vendite e insoddisfazione dei clienti.

Per le aziende del settore alimentare, come i servizi di consegna pasti o i produttori di beni freschi, la sfida principale consiste nel regolare i livelli di produzione e di scorte in modo da ridurre le perdite dovute alla rapida deperibilità delle materie prime. L'attuale catena di approvvigionamento alimentare risulta spesso inefficiente poiché l'incapacità di reagire efficacemente alle interruzioni o alle fluttuazioni della domanda genera surplus produttivi o carenze di prodotto, un concetto espresso da Sandeep Kumar Panda nell'articolo "*Time Series Forecasting and Modeling of Food Demand Supply Chain Based on Regressors Analysis*" [7].

La progettazione della supply chain per i prodotti deperibili richiede un equilibrio tra molteplici fattori: costi di trasporto, costi di carenza, costi di inventario, spreco di prodotto, deterioramento della qualità e perdita di durata di conservazione. Queste problematiche sono aggravate da fattori esterni che influenzano la domanda, come le preferenze dei consumatori, le attività promozionali, le festività e la stagionalità.

Per quanto riguarda la previsione, i metodi statistici classici – come *ARIMA* e *Holt-Winters* – sono stati a lungo la base del *forecasting* e hanno fornito risultati soddisfacenti in contesti relativamente stabili. Tuttavia, con l'aumento della complessità e del volume dei dati, tali metodi mostrano oggi limiti evidenti: difficoltà a gestire grandi quantità di dati, a integrare variabili esterne o a trattare dati non stazionari. Questo ha spinto la ricerca verso l'utilizzo di tecniche più avanzate basate sull'intelligenza artificiale e sul *machine learning*, capaci di apprendere dai dati storici e migliorare progressivamente le previsioni.

Gli approcci più recenti si orientano verso modelli ibridi, che combinano metodi statistici e algoritmi di apprendimento automatico. Tali modelli sono in grado di catturare sia le relazioni lineari sia quelle non lineari, migliorando così la precisione della previsione. Inoltre, l'integrazione di variabili esterne come clima, località, promozioni o dati provenienti dai social media consente di ottenere modelli più realistici e performanti.

L'obiettivo finale è disporre di un sistema di previsione accurato ed efficiente, capace di riflettere la domanda reale e di supportare una gestione dell'inventario più snella ma adeguata, riducendo sprechi e carenze.

### 1.3 Obiettivi di sostenibilità: riduzione sprechi e ottimizzazione dei costi

L'importanza di una previsione della domanda accurata va ben oltre il mero obiettivo di ottimizzazione finanziaria, ponendosi come base solida per i processi di pianificazione e approvvigionamento che rendono la catena di approvvigionamento più reattiva ed efficiente.

Storicamente, il focus principale del *Supply Chain Management* (SCM) è stato limitato alla minimizzazione dei costi operativi e delle spese, spesso senza considerare le implicazioni ecologiche, un obiettivo analizzato da Samiul Islam, Saman Hassanzadeh Amin e Leslie J. Wardley nello studio "*Machine learning and optimization models for supplier selection and order allocation planning*" [8]. Tuttavia, le crescenti preoccupazioni ambientali e le normative hanno spinto le aziende ad adottare la *Green Supply Chain Management* (GSCM). La GSCM mira a stabilire un equilibrio tra le dimensioni finanziarie, sociali e ambientali, incorporando pratiche ecocompatibili come la selezione di fornitori "green".

Nel settore alimentare, la sostenibilità è direttamente collegata alla capacità di ridurre l'abbondante spreco di cibo e mantenere un livello nutrizionale più elevato. La catena di approvvigionamento alimentare attuale è spesso inefficiente a causa dell'alto spreco e della sua incapacità di gestire le interruzioni. La previsione inadeguata della domanda è identificata come una delle cause cruciali dello spreco alimentare, insieme alla mancanza di coordinamento, alla gestione inefficace dello stoccaggio e alla sovrapproduzione. Le aziende che migliorano l'accuratezza del *forecasting* possono ottenere risultati economici vantaggiosi, ottimizzare la gestione della catena e, di conseguenza, ridurre significativamente gli sprechi.

L'accuratezza della previsione è strettamente correlata al processo di Selezione dei Fornitori e Allocazione degli Ordini (SSOA), una fase in cui le assunzioni tradizionali sulla domanda spesso si rivelano imprecise nella pratica, rendendo necessaria l'integrazione di dati storici e stime quantitative, come analizzato nell'articolo precedentemente citato. La SSOA è un problema tipico di ottimizzazione multi-obiettivo, che richiede di bilanciare criteri economici (costo totale) con fattori operativi (capacità, difettosità, puntualità di consegna) e ambientali (distanza, emissioni di carbonio). L'integrazione tra le previsioni della domanda basate su ML e i modelli di SSOA è considerata un approccio innovativo che fornisce decisioni più solide e sostenibili, specialmente in un settore dove la domanda è intrinsecamente incerta.

A livello strategico, l'attuale instabilità del mercato impone alle catene di approvvigionamento lo sviluppo della *Supply Chain Resilience*, intesa come la capacità di ridurre la probabilità di interruzioni e accelerare il recupero delle prestazioni operative, un concetto approfondito da

S. Hosseini nell'articolo *“Review of quantitative methods for supply chain resilience analysis”* [9]. In questo contesto, l'adozione di strategie di ridondanza e preparazione alle fluttuazioni dipende in modo cruciale dalla condivisione delle informazioni e dall'accuratezza dei modelli predittivi. La formazione di cluster logistici (collaborazioni formali) facilita la condivisione di informazioni, fornendo economie di scala e migliorando la capacità di reazione. Il miglioramento dell'allineamento tra domanda e offerta è possibile solo attraverso la collaborazione e la condivisione delle informazioni tra tutti i membri della catena di approvvigionamento.

La collaborazione estesa fino al consumatore finale, considerato un vero e proprio partner strategico, rappresenta l'elemento chiave per minimizzare l'effetto frusta e ottimizzare la pianificazione complessiva della catena di approvvigionamento.

L'applicazione di algoritmi di machine learning nel processo di previsione conferisce al sistema la capacità di apprendere dalle fluttuazioni della domanda e di adattarsi. Questa emergenza della resilienza del sistema (*system resilience*) si traduce in un miglioramento dei livelli di inventario, nella riduzione degli sprechi alimentari e, nel lungo termine, offre opportunità di sviluppo sostenibile per i membri del cluster logistico. L'integrazione e la collaborazione dovrebbero estendersi fino a includere i consumatori finali, trattati come *partner* strategici, per minimizzare l'effetto frusta (*bullwhip effect*)<sup>1</sup> e semplificare la pianificazione complessiva.

## 1.4 La struttura del lavoro

La presente tesi si propone di analizzare in modo sistematico e approfondito lo stato dell'arte relativo ai modelli di previsione della domanda nel settore agroalimentare, con l'obiettivo di individuare gli approcci più efficaci per migliorare l'accuratezza previsionale e ottimizzare la gestione delle scorte. A tal fine, è stata condotta una revisione strutturata della letteratura secondo il protocollo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), al fine di garantire trasparenza, replicabilità e rigore metodologico nel processo di selezione degli studi.

Nella fase iniziale, l'attenzione si è concentrata sulla definizione di una strategia di ricerca adeguata, attraverso la sperimentazione di diverse combinazioni di query booleane volte a

---

<sup>1</sup> L'effetto frusta descrive un fenomeno di instabilità nella catena di approvvigionamento in cui l'incertezza della domanda finale genera oscillazioni via via più ampie negli ordini man mano che si risale la filiera verso i produttori. Questo avviene perché ogni attore del sistema tenta di proteggersi dalle fluttuazioni del mercato basandosi su stime spesso imprecise, portando a inefficienze operative sistemiche.

individuare i contributi accademici più pertinenti. Il confronto tra i risultati ottenuti, caratterizzati da volumi differenti di documenti, ha permesso di identificare una specifica combinazione di termini in grado di selezionare un corpus finale di 52 articoli scientifici, ritenuti pienamente coerenti con i temi della previsione della domanda e della gestione delle scorte nel comparto agroalimentare.

Una volta consolidata la base bibliografica, il lavoro è proseguito con una fase di analisi sistematica degli studi selezionati, durante la quale sono state raccolte informazioni dettagliate relative al contesto applicativo, alla fase della catena del valore coinvolta, alla tipologia di prodotto considerata, nonché agli aspetti tecnici dei modelli adottati, includendo algoritmi utilizzati, caratteristiche dei dati e metriche di valutazione delle performance. Tale approccio ha consentito di organizzare un insieme eterogeneo di contributi in un database strutturato, idoneo a un'analisi comparativa.

Successivamente, è stata condotta un'interpretazione di primo livello dei risultati, finalizzata a delineare lo stato dell'arte delle tecnologie di previsione nel settore agroalimentare. Gli studi sono stati classificati secondo la *food value chain*, parallelamente, sono stati definiti sette cluster di prodotto per uniformare l'analisi delle diverse categorie merceologiche. Dal punto di vista metodologico, è stato effettuato un confronto tra modelli statistici tradizionali e tecniche di *machine e deep learning*.

Infine, l'ultima parte del lavoro è stata dedicata all'analisi delle relazioni tra le diverse dimensioni considerate, con l'obiettivo di individuare eventuali correlazioni tra la fase della filiera e il tipo di analisi svolta. L'indagine ha, inoltre, approfondito la natura e la complessità dei dati necessari a supportare i diversi modelli predittivi, offrendo un quadro complessivo delle migliori pratiche per l'implementazione di sistemi di previsione della domanda nel settore agroalimentare.

# Capitolo 2 – Stato dell'arte: dai Metodi Statistici al *Machine Learning*

## 2.1 Modelli Statistici e tecniche di Regressione Lineare

La previsione della domanda e la gestione delle scorte nel settore agroalimentare si basano storicamente su un solido corpus di modelli statistici e tecniche di regressione lineare, strumenti fondamentali per trasformare dati grezzi in modelli interpretabili di fenomeni complessi. L'obiettivo primario di tali approcci consiste nella predizione di una variabile target numerica, come il volume delle vendite o il fabbisogno di scorte, attraverso la definizione di una relazione matematica tra variabili indipendenti e dipendenti. All'interno della letteratura scientifica analizzata, i metodi statistici rappresentano circa il 24% degli algoritmi complessivamente utilizzati, confermando la loro rilevanza come base metodologica, nonostante la crescente competizione da parte delle tecniche di intelligenza artificiale.

La regressione lineare semplice costituisce il punto di partenza di questa disciplina, proponendo un modello in cui la variabile target è spiegata da un unico predittore attraverso una linea retta, dove l'intercetta rappresenta il valore di base e la pendenza misura la variazione media della risposta per ogni unità di cambiamento della variabile indipendente.

$$y = \beta_0 + \beta_1 x$$

Figura 1: Formula regressione lineare semplice

Per stimare i parametri ottimali di tali modelli, si ricorre comunemente al metodo dei minimi quadrati, il quale mira a minimizzare la somma dei residui quadratici (RSS), ovvero la discrepanza tra i valori reali osservati e quelli previsti dal modello. Infatti, si vorrebbe sempre ottenere che la differenza fra la 'y' dei dati di training e la 'y' predetta, comunemente indicata come ' $\hat{y}$ ', sia il più possibile minima, in modo tale da limitare la discrepanza tra i valori reali e quelli ottenuti. La differenza viene definita residuo [10].

Quando la complessità del fenomeno aumenta, si passa alla regressione lineare multipla, che integra una moltitudine di variabili esplicative per affinare la precisione predittiva, sebbene la presenza di correlazioni tra i predittori possa incrementare la varianza dei coefficienti e rendere più difficile l'interpretazione dei risultati.

$$y = f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n + \xi$$

Figura 2: Formula regressione multipla

La regressione multipla, infatti, non modella la correlazione tra i dati perché ogni coefficiente è stimato e testato separatamente: per definizione  $\beta_i$  rappresenta l'effetto medio su  $y$  di un aumento di un'unità di  $x_i$ , mantenendo fissi tutti gli altri elementi. Tuttavia, ci sono dei casi in cui esistono delle correlazioni fra i dati e, di conseguenza, alcune variabili non rimangono fisse: se ci sono forti dipendenze si hanno problemi sia nel modello che nell'interpretazione dei  $\beta_i$ . Per risolvere questa casistica, ad esempio, si possono effettuare delle analisi di correlazione per eliminare le *features* che presentano criticità affinché sia possibile ottenere un problema più semplice, interpretabile e controllabile. Questa esigenza diventa fondamentale per dati di alta dimensionalità<sup>2</sup> in quanto potrebbero esserci informazioni ridondanti e questo si ripercuote nel modello finale causando:

- Un'interpretazione più complessa;
- Parametri più difficili da stimare.

Si dovrebbe ridurre la dimensionalità per semplificare la computazione, migliorare la stima dei coefficienti e l'interpretazione, al contempo, all'aumentare del numero di *features* si potrebbero ottenere *performance* migliori: bisogna, quindi, trovare il giusto *trade-off* attraverso una *features selection* guidata da esperti di business e dominio [10].

In questo contesto, tecniche come il modello lineare con selezione *stepwise*<sup>3</sup> permettono di ottimizzare l'inserimento delle variabili, migliorando le performance computazionali e la spiegabilità del modello finale.

Per gestire set di dati ad alta dimensionalità o prevenire il problema dell'*overfitting*<sup>4</sup>, la statistica moderna introduce varianti regolarizzate della regressione lineare, tra cui spiccano la *Ridge regression* e la *Lasso regression*. La regressione *Ridge* aggiunge una penale basata sulla somma dei quadrati dei coefficienti (penalità L2), riducendo la complessità del modello senza eliminare variabili. Al contrario, la regressione *Lasso* (*Least Absolute Shrinkage and Selection Operator*) utilizza una penalità basata sul valore assoluto dei coefficienti (penalità

---

<sup>2</sup> Con l'espressione dati di alta dimensionalità si indicano insiemi di dati caratterizzati da un elevato numero di variabili (o attributi) rispetto al numero di osservazioni. Questa condizione può aumentare la complessità computazionale e rendere più difficile l'analisi e la modellazione, rendendo spesso necessarie tecniche di selezione o riduzione delle variabili

<sup>3</sup> Il modello lineare con selezione *stepwise* è una procedura di regressione multipla che prevede l'inclusione o l'esclusione iterativa delle variabili esplicative sulla base di criteri statistici predefiniti (quali il valore dei *p-value*, l'AIC o il BIC).

<sup>4</sup> L'*overfitting* si verifica quando un modello statistico si adatta eccessivamente ai dati di addestramento, catturando anche il rumore casuale anziché la struttura sottostante del fenomeno analizzato. Ciò comporta un'elevata accuratezza sui dati osservati ma una ridotta capacità di generalizzazione su nuovi dati.

L1), che ha la proprietà di annullare i parametri meno significativi, effettuando di fatto una selezione automatica delle *features*. L'integrazione di questi due approcci dà vita all'*Elastic Net*, mentre varianti come la *Bayesian Ridge* permettono di includere la distribuzione di probabilità dei parametri nel processo di stima. Per le situazioni in cui la relazione tra le variabili non segue una linea retta, si ricorre alla Regressione Polinomiale (PR) o alla *spline regression*, che utilizzano funzioni di potenza o segmenti curvi per approssimare curvature complesse, pur richiedendo cautela per evitare di interpretare il rumore come un pattern significativo.

$$y = f(x) = \beta_0 + \beta_1x + \beta_2x^2 + \dots + \beta_nx^n + \xi$$

Figura 3: Formula regressione polinomiale

La regressione polinomiale trasforma le *feature* in uno spazio con dimensionalità diversa e, sulle *feature* trasformate, applica la regressione lineare. L'ideale è partire dalla regressione lineare, calcolarsi la qualità del dato e, man mano che si cresce con il livello di ordine del modello, valutare il guadagno ottenuto: questo approccio viene definito *forward*. Il metodo opposto è chiamato *backward*: si parte dall'ordine maggiore per poi diminuire in modo progressivo. Solitamente si predilige il primo approccio in quanto permette di raggiungere prima il giusto *trade-off* tra qualità del modello e complicazione nel gestire ordini minori.

L'obiettivo è ottenere il livello di ordine minore possibile per avere un modello interpretabile e per evitare il fenomeno dell'*overfitting* [10].

Un capitolo distinto ed essenziale è rappresentato dai modelli statistici di serie temporali, che costituiscono il 13% degli algoritmi censiti nella ricerca e sono specificamente progettati per dati osservati a intervalli regolari. Questi modelli analizzano le componenti intrinseche del segnale, come il trend (movimento a lungo termine), la stagionalità (oscillazioni periodiche) e le fluttuazioni irregolari o casuali, spesso identificate come rumore. Tra le tecniche più semplici, ma efficaci, si trovano i metodi *Naive* e *Seasonal Naive*, che utilizzano l'ultima osservazione disponibile come previsione per il futuro, e le tecniche basate sulla Media Mobile (MA) o sulla Mediana, utili per stabilizzare serie storiche volatili. Per fenomeni più strutturati, il modello di *Holt-Winters* e i metodi di smorzamento esponenziale (*exponential smoothing*) offrono strumenti avanzati per catturare contemporaneamente tendenze e pattern stagionali.

L'evoluzione verso modelli stocastici più rigorosi ha portato alla diffusione della famiglia dei modelli autoregressivi, come l'AR (*Autoregression*), in cui il valore futuro dipende linearmente

dai valori passati, e il VAR (*Vector Autoregression*) per casi multivariati. I modelli ARIMA (*Autoregressive Integrated Moving Average*) e la loro variante stagionale SARIMA rappresentano lo standard per il *forecasting* statistico, combinando componenti *autoregressive* e medie mobili per gestire serie non stazionarie. L'estensione ARIMAX permette, inoltre, di includere variabili esogene esterne, come indicatori economici o dati meteorologici, per arricchire la capacità predittiva del modello.

In conclusione, nonostante la solidità teorica e l'elevata interpretabilità che li caratterizzano, i modelli statistici tradizionali risultano complessivamente meno efficaci rispetto alle moderne tecniche di intelligenza artificiale nel rispondere alle sfide del settore agroalimentare.

Questa minore efficacia deriva principalmente dal fatto che i metodi classici faticano a modellare adeguatamente le forti non-linearità e le interazioni multidimensionali che caratterizzano la domanda di prodotti alimentari, la quale è intrinsecamente soggetta a elevata volatilità e incertezza. Mentre gli approcci statistici necessitano spesso di complessi processi preliminari di pulizia, normalizzazione e rimozione del rumore per poter operare correttamente, gli algoritmi di Machine Learning dimostrano una maggiore resilienza e capacità di apprendimento dai dati, riuscendo a identificare pattern complessi anche in presenza di variabili esterne eterogenee.

Di conseguenza, laddove la relazione tra i dati diventa curvilinea o influenzata da numerosi fattori concomitanti, i modelli statistici perdono precisione, rischiando di non riflettere la reale dinamica di un mercato caratterizzato da prodotti con vita utile limitata e domanda instabile.

## 2.2 L'ascesa del *Machine Learning* e l'efficacia degli algoritmi *Ensemble*

L'adozione delle tecniche di *Machine Learning* (ML) rappresenta una risposta necessaria alla crescente instabilità economica globale e all'incertezza che rende difficile la previsione delle vendite nel settore alimentare. Rispetto ai modelli lineari tradizionali, gli approcci basati sul ML possiedono un potere predittivo superiore grazie alla loro capacità intrinseca di elaborare e valutare grandi volumi di dati in modo efficiente. Questa transizione tecnologica permette ai membri della catena di approvvigionamento di beneficiare di capacità adattive, migliorando la sincronizzazione tra domanda e offerta e riducendo sensibilmente gli sprechi derivanti da prodotti scaduti o eccessi di magazzino. Mentre i sistemi basati sull'esperienza umana sono soggetti a errori e pregiudizi individuali, i modelli di Machine Learning sono dinamici, non risentono di bias soggettivi e possono adattarsi rapidamente ai cambiamenti nei modelli di consumo.

La superiorità del *Machine Learning* rispetto alle tecniche statistiche classiche, come ARIMA o ARMA, risiede nella sua maggiore flessibilità. Questi modelli permettono infatti l'inclusione di variabili esterne esogene, quali condizioni meteorologiche, festività, promozioni ed eventi sociali,

che influenzano profondamente la domanda alimentare ma che non possono essere integrate nei modelli statistici univariante. Il *machine learning* si dimostra particolarmente efficace nel gestire la non-linearità delle sequenze temporali e i complessi pattern di domanda tipici dei prodotti deperibili, dove la volatilità è elevata e la vita utile dei prodotti è limitata. Inoltre, queste tecniche mostrano una resilienza superiore nel trattamento di dati rumorosi o incompleti, identificando autonomamente relazioni e tendenze senza la necessità di conoscenze pregresse sui parametri di distribuzione.

All'interno del panorama del *machine learning*, le tecniche di *ensemble learning* si sono affermate come strumenti di eccellenza per massimizzare l'accuratezza delle previsioni. Questi metodi si basano sulla combinazione di molteplici modelli, spesso alberi di decisione, per costruire un sistema predittivo più robusto e preciso rispetto ai singoli componenti. Gli algoritmi di *ensemble Decision Tree* (EDT), nelle varianti *Boosted* e *Bagged*, sono ampiamente raccomandati per la loro capacità di migliorare la capacità di previsione e ridurre l'errore quadratico medio. In particolare, il modello EDT *Boosted* ha dimostrato in diversi contesti operativi, come la ristorazione collettiva e i refettori universitari, di poter superare significativamente le prestazioni di altri algoritmi, garantendo stabilità anche in assenza di sistemi di prenotazione preventiva.

Tra i modelli ensemble più diffusi, il *Random Forest* si distingue per la sua abilità nel gestire variabili multivariate e variazioni della domanda non cicliche o irregolari, spesso causate da eventi promozionali o festività. Parallelamente, algoritmi di *boosting* stocastico come *XGBoost* e *LightGBM* eccellono nel catturare trend complessi attraverso il miglioramento progressivo di modelli più semplici, permettendo alle imprese di anticipare le fluttuazioni stagionali e ottimizzare i livelli di inventario in tempo reale. L'integrazione di approcci ibridi, che combinano ad esempio *Random Forest* con reti neurali come LSTM, permette di catturare contemporaneamente correlazioni temporali e spaziali, superando costantemente le soluzioni discrete basate su singoli modelli. In definitiva, l'efficacia di queste strategie ensemble risiede nella loro capacità di minimizzare la varianza e l'errore di previsione, fornendo ai decisori strumenti analitici avanzati per una gestione della filiera alimentare più sostenibile e profittevole.

## 2.3 Il ruolo del *Deep Learning*: Reti Neurali e modelli LSTM

Il *Deep Learning* (DL) si è affermato come una delle evoluzioni più significative all'interno del *machine learning*, offrendo una capacità superiore nel processare ed elaborare grandi volumi di dati attraverso architetture complesse che superano i limiti dei modelli lineari. Le Reti Neurali Artificiali (ANN) costituiscono il fondamento di questa tecnologia e operano mappando le relazioni tra variabili di input e output quando queste non risultano

immediatamente evidenti attraverso i metodi convenzionali. Tra le varianti più comuni figurano le *Feed-Forward Neural Networks* (FFNN), capaci di regolare i pesi delle connessioni tra neuroni per minimizzare l'errore di previsione, e le *Cascade Forward Neural Networks* (CFNN), che accelerano il processo di apprendimento includendo connessioni dirette tra lo strato di input e i livelli successivi. Queste strutture permettono di catturare tendenze dinamiche non lineari e stagionalità nei dati di vendita, superando spesso la precisione degli esperti umani.

Un limite intrinseco delle reti neurali tradizionali riguarda la gestione di dati sequenziali, compito solitamente affidato alle Reti Neurali Ricorrenti (RNN), le quali però soffrono della problematica del gradiente evanescente che ne impedisce l'apprendimento su sequenze temporali estese. Per risolvere questo ostacolo sono stati introdotti i modelli *Long Short-Term Memory* (LSTM), una forma specializzata di RNN progettata per memorizzare informazioni critiche all'interno di una memoria a lungo termine. L'architettura delle LSTM si basa su una cella di memoria gestita da tre porte specifiche denominate *input gate*, *forget gate* e *output gate*, le quali decidono quali informazioni conservare, eliminare o trasmettere allo stato successivo. Questo meccanismo permette alle LSTM di descrivere le proprietà dei dati senza richiedere conoscenze pregresse sui parametri o sulle distribuzioni delle caratteristiche, rendendole ideali per modellare la domanda alimentare che presenta dipendenze temporali complesse.

Nella letteratura scientifica, l'efficacia dei modelli LSTM è ampiamente documentata dalla loro capacità di superare sistematicamente algoritmi come le Support Vector Machines (SVM) e i modelli statistici ARIMA, specialmente in presenza di pattern di domanda non lineari e sequenze temporali estese. Le LSTM eccellono nel catturare sia le correlazioni temporali che quelle covariate, mostrando una varianza e un bias inferiori rispetto ad altre tecniche di previsione. Tuttavia, l'implementazione di tali sistemi richiede un'attenta fase di preelaborazione<sup>5</sup> dei dati, inclusa la normalizzazione e la rimozione del rumore, nonché un'accurata ottimizzazione degli iperparametri per evitare il rischio di *overfitting*, che si verifica quando il modello si adatta eccessivamente ai dati di addestramento senza riuscire a generalizzare su nuove informazioni.

---

<sup>5</sup> Per fase di preelaborazione dei dati si intende l'insieme di operazioni volte a preparare i dati grezzi per l'analisi e l'applicazione di modelli di *Machine e Deep Learning*. Essa comprende attività come la pulizia dei dati, la gestione dei valori mancanti, la normalizzazione delle variabili e la trasformazione dei dati in un formato idoneo all'addestramento dei modelli, con l'obiettivo di migliorarne accuratezza e affidabilità.

L'evoluzione più recente nel campo del *Deep Learning* applicato al settore agroalimentare vede l'integrazione dei modelli LSTM con le Reti Neurali Convoluzionali (CNN) in architetture ibride definite CNN-LSTM. In questo approccio, la componente CNN viene utilizzata per estrarre caratteristiche locali e relazioni spaziali dai dati multivariati, mentre lo strato LSTM modella le dipendenze temporali tra tali caratteristiche. Tale configurazione si è dimostrata particolarmente promettente per la previsione della domanda su base geografica, riuscendo a identificare variazioni rapide e *pattern* spazio-temporali che sfuggono ai modelli statistici regressivi di base. L'impiego di queste tecniche avanzate consente alle aziende della filiera alimentare di ottimizzare la gestione dell'inventario in tempo reale, riducendo drasticamente gli sprechi e garantendo una maggiore stabilità operativa.

## 2.4 Metriche di valutazione e criteri di confronto tra modelli

La valutazione dell'efficacia dei modelli quantitativi per la previsione della domanda si affida prevalentemente a indicatori statistici che permettono di misurare i risultati ottenuti e determinare la necessità di eventuali aggiustamenti ai dati di input. All'interno della letteratura scientifica analizzata, il *Mean Absolute Percentage Error* (MAPE) risulta essere la metrica più ampiamente documentata e utilizzata come indicatore relativo, risultando particolarmente utile quando si devono confrontare set di dati caratterizzati da scale differenti. Infatti, esprime l'errore in termini percentuali rispetto ai valori osservati, diventando indipendente dall'unità di misura.

$$MAE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Figura 4: Formula MAPE

Parallelamente, il *Root Mean Square Error* (RMSE) viene impiegato per confrontare gli errori di previsione tra modelli che operano sulla medesima base dati e corrisponde alla media degli errori al quadrato sotto radice quadrata e penalizza maggiormente gli errori più grandi.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Figura 5: Formula RMSE

Qualora l'obiettivo primario sia la minimizzazione della deviazione assoluta, si ricorre solitamente al *Mean Absolute Error* (MAE): questa metrica misura direttamente l'errore medio in valore assoluto tra i valori osservati e quelli stimati, senza penalizzare eccessivamente gli estremi.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Figura 6: Formula MAE

Il *Mean Square Error* (MSE), invece, funge frequentemente da indicatore della qualità specifica per i modelli statistici di tipo ARIMA. Altri criteri di valutazione includono l'utilizzo del *Median Absolute Deviation* (MAD) e del *Max Absolute Error* (MaxAE) per catturare diverse sfumature della distribuzione degli errori e penalizzare i picchi di scostamento.

Il confronto tra le prestazioni avviene attraverso protocolli rigorosi che prevedono la suddivisione dei dati in set di addestramento e di test, oppure l'applicazione della validazione incrociata per garantire la robustezza delle stime ed evitare problemi di overfitting. Mentre i modelli classici richiedono una fase preliminare di preelaborazione per la normalizzazione e la rimozione degli outlier, gli algoritmi di machine learning dimostrano una superiore capacità di gestire tali discrepanze durante il processo di valutazione. In ultima analisi, la scelta della metrica e della metodologia di confronto dipende spesso dalle caratteristiche del prodotto, come la sua vita utile, e dalla necessità di bilanciare la precisione con l'efficienza operativa per ridurre sprechi e rotture di stock.

## Capitolo 3: Metodologia di ricerca

### 3.1 Protocollo PRISMA

La fase iniziale del lavoro di tesi è stata dedicata all'individuazione sistematica degli articoli riguardanti la previsione della domanda e la gestione delle scorte, adottando il rigore metodologico del protocollo PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) [11]. Matthew J Page, nell'articolo precedentemente citato, spiega come questa metodologia abbia lo scopo primario di guidare i ricercatori nel documentare in modo rigoroso il motivo per cui lo studio è stato effettuato, le procedure seguite e i risultati finali ottenuti. Il protocollo si fonda su due strumenti operativi essenziali: una checklist composta da 27 elementi e un diagramma di flusso a quattro fasi che illustra visivamente il percorso delle informazioni attraverso i processi di identificazione, screening, inclusione ed esclusione degli studi [12].

Un pilastro fondamentale di PRISMA 2020 è la distinzione netta tra la condotta metodologica della revisione e la sua rendicontazione (*reporting*); mentre la condotta riguarda il modo in cui la ricerca viene eseguita, PRISMA si concentra esclusivamente su come essa viene descritta, partendo dal presupposto che una descrizione trasparente sia l'unico modo per permettere ai lettori di valutare la qualità dello studio. L'adozione di questo standard mira a garantire la riproducibilità della ricerca, offrendo tutte le informazioni necessarie per giudicare la validità dei metodi e l'applicabilità dei risultati nel proprio contesto. Rispetto alla versione originale del 2009, il protocollo 2020 riflette i significativi progressi tecnologici dell'ultimo decennio, integrando indicazioni specifiche sull'uso di strumenti di automazione, elaborazione del linguaggio naturale e *machine learning* per la selezione delle prove.

### 3.2 Strategia di Ricerca

In linea con i requisiti del protocollo, che impone di identificare chiaramente ogni banca dati consultata e la relativa piattaforma di accesso, la ricerca è stata svolta prevalentemente su Scopus: è stato preferito a Web of Science per la sua maggiore ampiezza nel settore e per la capacità di includere la maggior parte dei contributi presenti in Web of Science, rendendolo lo strumento più idoneo per identificare gli articoli scientifici di rilievo nel campo della gestione delle scorte e della previsione della domanda.

Il primo passaggio operativo ha riguardato la definizione della query di ricerca, un processo critico che, secondo le estensioni del protocollo come il PRISMA-S, deve essere documentato in modo robusto per minimizzare i *bias* e stabilire con precisione i dati disponibili per l'analisi. La strategia di ricerca è stata strutturata con l'obiettivo di formulare una query che garantisca la massima completezza e precisione nell'estrazione dei dati, in conformità con la metodologia PRISMA. La tecnica utilizzata ha previsto l'integrazione dei tre pilastri fondamentali dello studio attraverso l'operatore logico AND; ciascuno di questi pilastri è stato poi esteso includendo sinonimi e categorie supplementari mediante l'operatore OR.

Il nucleo centrale della query si articola su tre aree tematiche distinte, ognuna rappresentata da una parola chiave specifica:

- *Machine Learning*: definisce il contesto metodologico e tecnologico;
- *Food*: delimita l'ambito applicativo della ricerca;
- *Demand forecasting*: identifica il tema centrale e l'obiettivo dell'indagine.

Questa stringa di ricerca iniziale è stata progressivamente affinata e arricchita con vari campi di applicazione, come rappresentato nella *tabella 1*, che illustra la logica della query finale impiegata.

Punti core della ricerca	Query completa
<b>Machine Learning</b>	“machine learning” OR “deep learning” OR “artificial intelligence” OR “data mining” <b>AND</b>
<b>Food</b>	“food” OR “agro-food” OR “food industry” OR “food supply chain” <b>AND</b>
<b>Demand forecasting</b>	“demand forecast” OR “sales forecast”

*Tabella 1: Query utilizzata*

Successivamente alla definizione della query, la stringa nella sua versione completa è stata inserita all'interno del database selezionato attraverso la funzione di ricerca avanzata. Tale operazione è stata mirata specificamente ai campi *Title*, *Abstract* e *Keywords*, garantendo così che i risultati estratti fossero strettamente correlati ai pilastri tematici della ricerca.

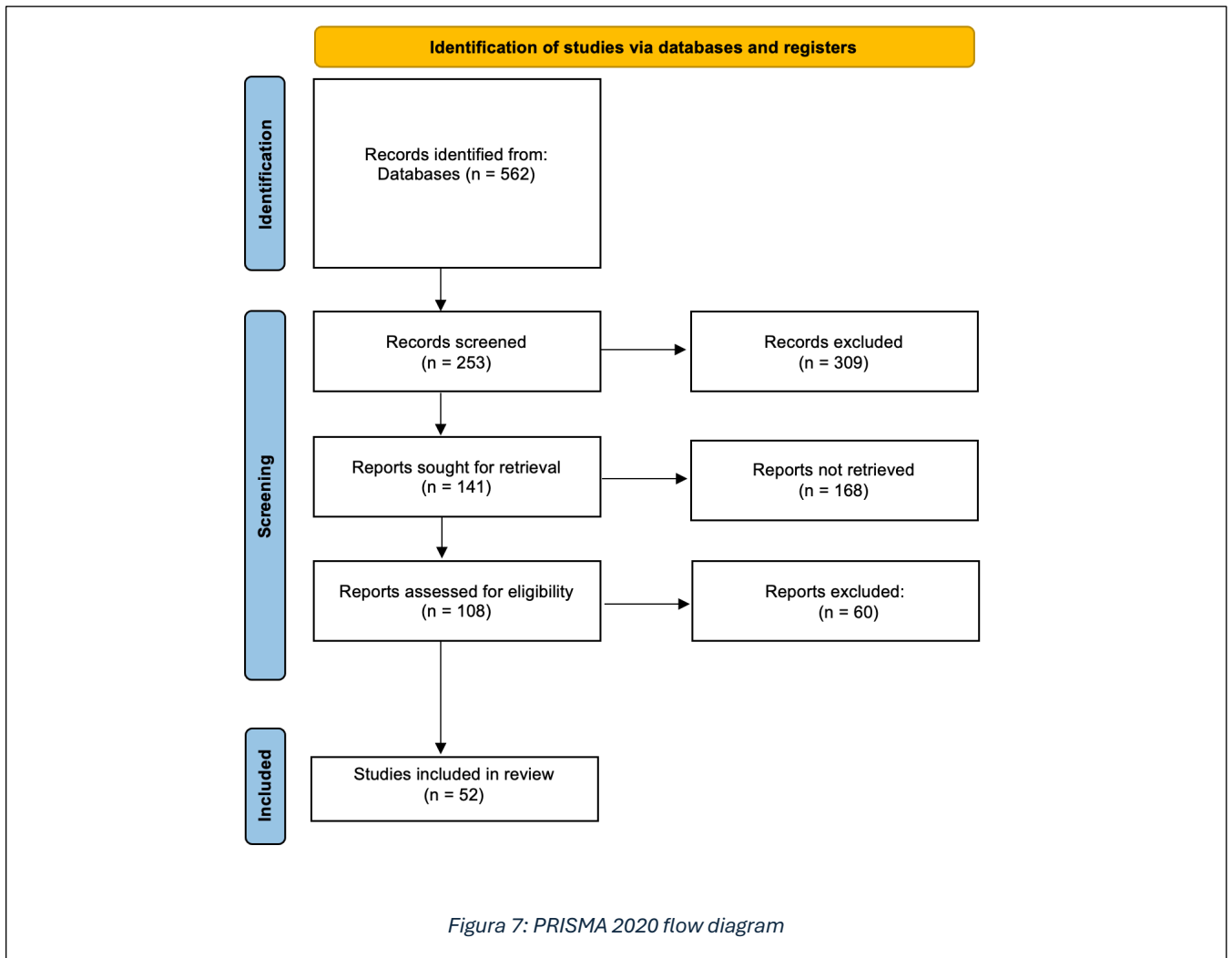
Dall'applicazione di questa strategia sono stati ottenuti inizialmente 562 record totali. Questi risultati hanno costituito il punto di partenza per la fase di identificazione prevista dal protocollo PRISMA.

### 3.2.1 Processo di selezione degli studi

Per garantire la trasparenza metodologica, ridurre il rischio di bias e assicurare la replicabilità dell'analisi, sono stati definiti criteri di inclusione ed esclusione chiari. La selezione è stata circoscritta a specifiche aree disciplinari coerenti con gli obiettivi della ricerca e limitata a tipologie documentali riconosciute come affidabili, quali articoli *peer-reviewed*, *conference paper* e *review* sistematiche.

Per delimitare il campo d'indagine, è stato selezionato un intervallo temporale compreso tra il 2015 e il 2025: la motivazione risiede nel rapido sviluppo tecnologico degli ultimi anni, poiché, come riportato in 'A survey of machine learning techniques for food sales prediction' [6], i metodi di *machine learning* e *deep learning* hanno visto un'evoluzione considerevole solo in questo periodo, superando i tradizionali approcci statistici prevalenti in precedenza. Inoltre, è stata applicata una restrizione alla sola lingua inglese per garantire una standardizzazione linguistica ed evitare potenziali errori di interpretazione tecnica derivanti dall'uso di strumenti di traduzione esterni.

Parallelamente, i criteri di esclusione hanno permesso di scartare gli studi i cui obiettivi non erano coerenti con il focus della ricerca. In particolare, la maggior parte degli studi scartati si concentrava su fasi della catena del valore non rilevanti per il presente lavoro, come la previsione delle quantità da coltivare per soddisfare la domanda agricola [13], oppure faceva riferimento a settori completamente differenti, quali l'energia [14] o l'industria cinematografica [15].



L'intero processo di selezione è stato strutturato in conformità con le linee guida PRISMA, garantendo la massima trasparenza, replicabilità e tracciabilità di ogni fase decisionale e delle motivazioni alla base delle esclusioni. Partendo da un corpus iniziale di 562 record identificati attraverso le interrogazioni dei database, è stata operata una prima scrematura sistematica per allineare i risultati agli obiettivi della ricerca. In questa fase, la selezione è stata limitata al settore agroalimentare secondario e a tipologie documentali certificate, sopra citate, riducendo il corpus a 253 record sottoposti a screening e comportando l'esclusione di 309 unità.

Il raffinamento è proseguito attraverso l'applicazione di parametri di esclusione tematici riguardanti le fasi della catena del valore non oggetto di studio, come l'ottimizzazione dei raccolti o le analisi focalizzate esclusivamente sul consumatore finale. Questo passaggio ha portato alla definizione di 141 report, sebbene 168 documenti non siano stati recuperati. Di conseguenza, sono stati valutati per l'idoneità 108 report complessivi.

Durante l'analisi di titoli e abstract, è stata fondamentale l'estrazione e la catalogazione di variabili tecniche specifiche — identificate dalle fonti come data items essenziali — tra cui gli algoritmi implementati, l'ambito applicativo e la tipologia di dati, elementi necessari per garantire la corretta interpretazione e applicabilità delle conclusioni della revisione.

Infine, la lettura integrale dei testi ha permesso di verificare con precisione la coerenza con gli obiettivi scientifici, portando all'esclusione di ulteriori 60 report. Al termine di questo processo di selezione multilivello, il corpus finale è risultato composto da 52 studi inclusi nella revisione, i quali approfondiscono i temi della previsione della domanda e della gestione delle scorte. L'intero flusso decisionale, che ha guidato la riduzione progressiva del materiale iniziale fino alla definizione del campione finale, è visualizzato sinteticamente nel diagramma di flusso PRISMA nella *figura 7*.

### 3.3 Analisi preliminare e raccolta di informazioni

In seguito alla selezione della combinazione ottimale di query, che ha permesso di circoscrivere il dataset a 52 articoli totali, si è resa necessaria una fase di analisi sistematica per estrarre il valore informativo da ciascun contributo. Questa fase, definita come analisi preliminare e raccolta di informazioni, ha avuto l'obiettivo di catalogare i contenuti in modo rigoroso per consentire un successivo confronto quantitativo e qualitativo tra i risultati.

Per ogni articolo selezionato, è stata condotta una lettura approfondita procedendo poi alla registrazione di commenti specifici circa l'inerenza dello studio rispetto agli obiettivi della ricerca. Nello specifico, per ogni pubblicazione è stata creata una scheda informativa strutturata che raccoglieva i seguenti pilastri analitici riportata nell'Allegato A:

- inquadramento del contesto: è stato definito l'ambito di applicazione specifico e la fase della *food value chain* analizzata, andando a identificare il soggetto su cui veniva svolta l'analisi;
- ambito funzionale: ogni studio è stato classificato in base alla sua finalità operativa, distinguendo tra previsione della domanda, gestione delle scorte o una combinazione di entrambe affinché fosse poi possibile identificare una direttrice sulla quale svolgere il lavoro di tesi;
- caratterizzazione del prodotto: per rendere l'analisi più ampia e comparabile, i prodotti oggetto degli studi sono stati raggruppati in 7 cluster;

- profilo tecnico e algoritmico: per ciascun caso sono stati censiti gli algoritmi impiegati, suddividendoli in macrocategorie quali *machine learning*, *deep learning*, metodi statistici o modelli ibridi, identificando inoltre quale tra questi risultasse il più performante secondo i test svolti dagli autori;
- analisi dei dati: sono state estratte informazioni critiche sulla natura dei dati utilizzati, come la dimensione (univariata o multivariata), la frequenza di rilevamento (es. giornaliera, settimanale), il range temporale (espresso in mesi) e le specifiche features considerate;
- metriche di valutazione: per garantire la comparabilità delle performance, sono state annotate le metriche utilizzate per validare i modelli.

Questa meticolosa attività di raccolta ha permesso di trasformare dati testuali non strutturati in un dataset pronto per l'analisi. L'obiettivo finale non è stato solo quello di riassumere i singoli lavori, ma anche di identificare delle possibili correlazioni significative tra le diverse dimensioni.

## Capitolo 4 – Interpretazione di primo livello: tassonomia dei risultati

### 4.1 Applicabilità e finalità degli algoritmi predittivi nel settore agroalimentare

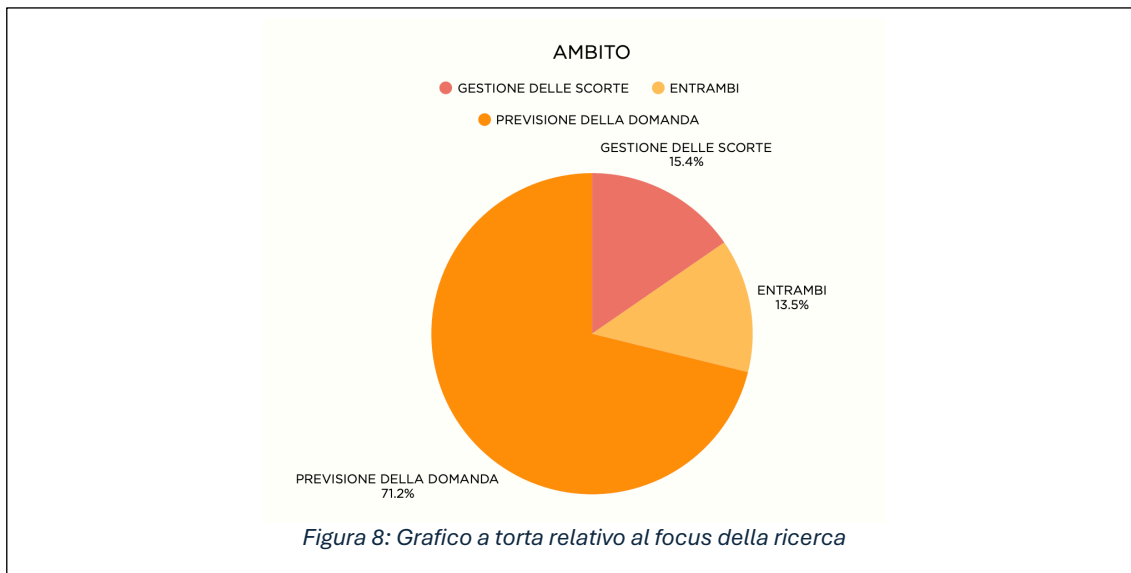
L'analisi sistematica dei contributi scientifici raccolti permette di confermare con certezza che gli algoritmi di previsione sono pienamente applicabili nel settore agroalimentare, fornendo strumenti efficaci per affrontare le incertezze intrinseche della filiera. Attraverso l'impiego di tecniche computazionali avanzate, è infatti possibile implementare modelli capaci di supportare concretamente il *forecasting* della domanda per prodotti caratterizzati da alta deperibilità e volatilità [4].

Sulla base delle *review* della letteratura analizzate, emerge un'evoluzione tecnologica netta: nel corso degli anni, gli approcci basati su *machine learning* e *deep learning*, o modelli ibridi che combinano entrambi, sono diventati significativamente più performanti rispetto ai metodi statistici classici. Questo incremento di efficacia è riconducibile a due fattori principali: la potenza e la flessibilità. Mentre le tecniche statistiche tradizionali, come i modelli ARMA e ARIMA, sono spesso limitate all'analisi di serie temporali *univariate*, l'approccio basato sul *machine learning* consente di integrare una vasta gamma di variabili esogene ed esterne (come dati meteorologici, festività o indicatori economici), permettendo al sistema di riflettere con maggiore accuratezza la realtà del mercato [6]. Inoltre, l'adozione di tecniche di apprendimento automatico offre il vantaggio di costruire sistemi dinamici e adattivi, capaci di evolvere insieme ai dati e di superare in precisione l'intuito umano, spesso influenzato da pregiudizi o limitazioni cognitive.

In sintesi, la transizione verso modelli di ML e DL non rappresenta solo un miglioramento tecnico, ma una necessità strategica per ridurre gli sprechi e ottimizzare i livelli di inventario in un settore dove la finestra temporale per la vendita è estremamente ridotta.

### 4.2 Mappatura lungo la *Food Value Chain*

Un passaggio fondamentale della tassonomia ha riguardato la classificazione degli studi in base alla loro finalità operativa. Ogni articolo del dataset è stato analizzato per determinare se il focus della ricerca fosse orientato alla previsione della domanda, alla gestione delle scorte o a un approccio integrato che comprendesse entrambi gli aspetti.



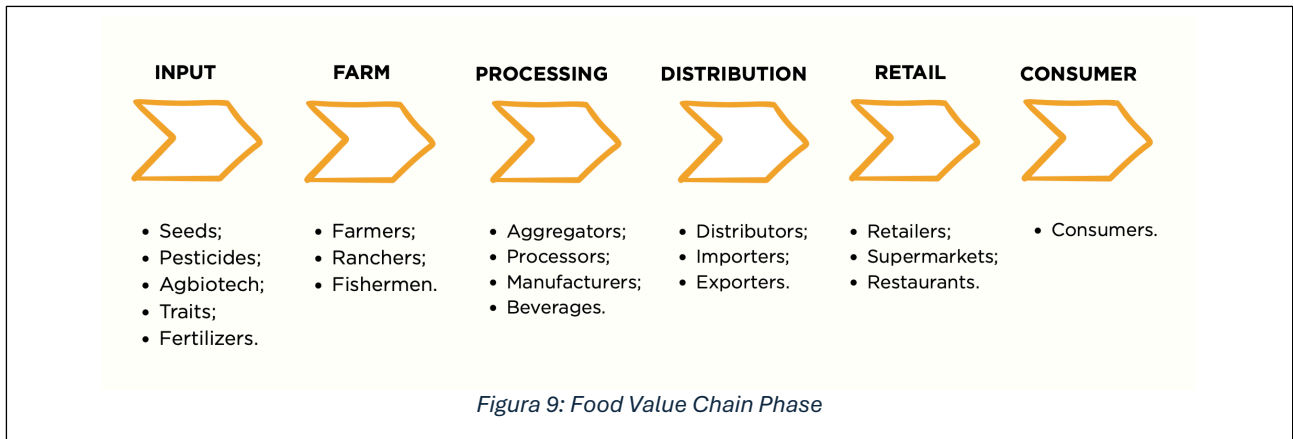
Dall'analisi dei dati, emerge una netta polarizzazione verso l'attività predittiva:

- previsione della domanda (71,2%): la maggioranza dei contributi scientifici si concentra esclusivamente sulla stima accurata dei volumi di vendita o dei fabbisogni futuri. Questo dato evidenzia come la letteratura scientifica identifichi nel *forecasting* la sfida tecnologica principale e il punto di partenza imprescindibile per l'efficientamento della filiera;
- gestione delle scorte (15,4%): una porzione significativamente minore di studi si focalizza in via esclusiva sulle politiche di riordino e sull'ottimizzazione fisica dei magazzini;
- approccio combinato (13,5%): solo una quota ridotta di articoli affronta il problema in modo olistico, integrando la fase di previsione con quella decisionale relativa allo stoccaggio.

Questa distribuzione conferma che, nel settore agroalimentare, la precisione nella previsione della domanda è considerata il fattore critico di successo. Data l'alta deperibilità dei prodotti, l'attenzione dei ricercatori non si sofferma tanto sulla gestione statica del magazzino, quanto sulla capacità di anticipare il mercato per ridurre alla radice il rischio di sovrapproduzione o di rotture di stock. In sintesi, la gestione delle scorte viene spesso trattata come una variabile dipendente, la cui efficacia è subordinata alla qualità dell'algorithmo di previsione adottato.

Prendendo come riferimento per la definizione della filiera alimentare l'articolo "*Teaching professional ethics in culinary studies*", pubblicato da Amir Shani, Yaniv Belhassen e Daniel Soskolne nell'*International Journal of Contemporary Hospitality Management* [17], gli autori adottano la *Food Value Chain* (catena del valore alimentare) come quadro teorico per

analizzare le questioni etiche che emergono in ogni fase della produzione e del consumo di cibo. Secondo le fonti, la catena del valore è definita come l'insieme di tutti i processi e le attività necessari per trasformare le materie prime in un prodotto o servizio, fino al suo smaltimento finale dopo il consumo.



All'interno dell'industria alimentare, questa analisi coinvolge sei attori o fasi generiche, ciascuna con componenti e problematiche specifiche:

- *Input Industry* (Industria degli input): rappresenta l'inizio della catena e comprende la fornitura di semi, pesticidi, fertilizzanti e tecnologie legate all'agrobiotecnologia. Le principali preoccupazioni etiche in questa fase riguardano l'inquinamento, i problemi legati al lavoro e la tracciabilità;
- *Farmer* (Fase agricola/Farm): coinvolge direttamente agricoltori, allevatori e pescatori. In questo passaggio, l'attenzione si focalizza sulla sostenibilità agricola, sul reddito dei produttori, sul benessere animale, sulla sicurezza idrica e sul commercio equo e globale;
- *Processing* (Trasformazione): questa fase riguarda la trasformazione delle materie prime in prodotti finiti o semilavorati. Include aggregatori, trasformatori (*processors*), produttori (*manufacturers*) e l'industria delle bevande (*beverages*). È un passaggio critico per la creazione di valore aggiunto e per la stabilizzazione dei prodotti deperibili;
- *Distribution* (Distribuzione): si occupa del movimento fisico e della logistica dei prodotti su scala globale e locale, coinvolgendo distributori, importatori ed esportatori. Questa fase è tra le più studiate perché rappresenta il punto di giunzione logistico dove l'accuratezza della previsione della domanda è vitale per evitare sprechi;

- *Retail* (Vendita al dettaglio): rappresenta l'interfaccia diretta con il mercato. Include rivenditori (*retailers*), supermercati e ristoranti. Insieme alla distribuzione, costituisce il fulcro della ricerca scientifica sul forecasting, poiché è qui che si manifesta la volatilità della domanda dei consumatori finali;
- *Consumer* (Consumatore): rappresenta l'anello finale della catena, ovvero l'utente che acquista e consuma il prodotto. In questa fase, l'attenzione si sposta sulle abitudini di acquisto, sulla salute e sullo smaltimento dei rifiuti alimentari.

Per tutti gli articoli selezionati è stato individuato l'ambito di applicazione specifico all'interno della *food value chain*, precedentemente definita. L'obiettivo di questa classificazione è comprendere in quali segmenti della filiera l'applicazione di algoritmi predittivi e di *machine learning* sia più matura e richiesta dalla letteratura scientifica. Nell'analisi sono stati inclusi anche 2 articoli che, pur non trattando un singolo caso studio, rappresentano delle *review* della letteratura, offrendo una prospettiva trasversale sul tema.

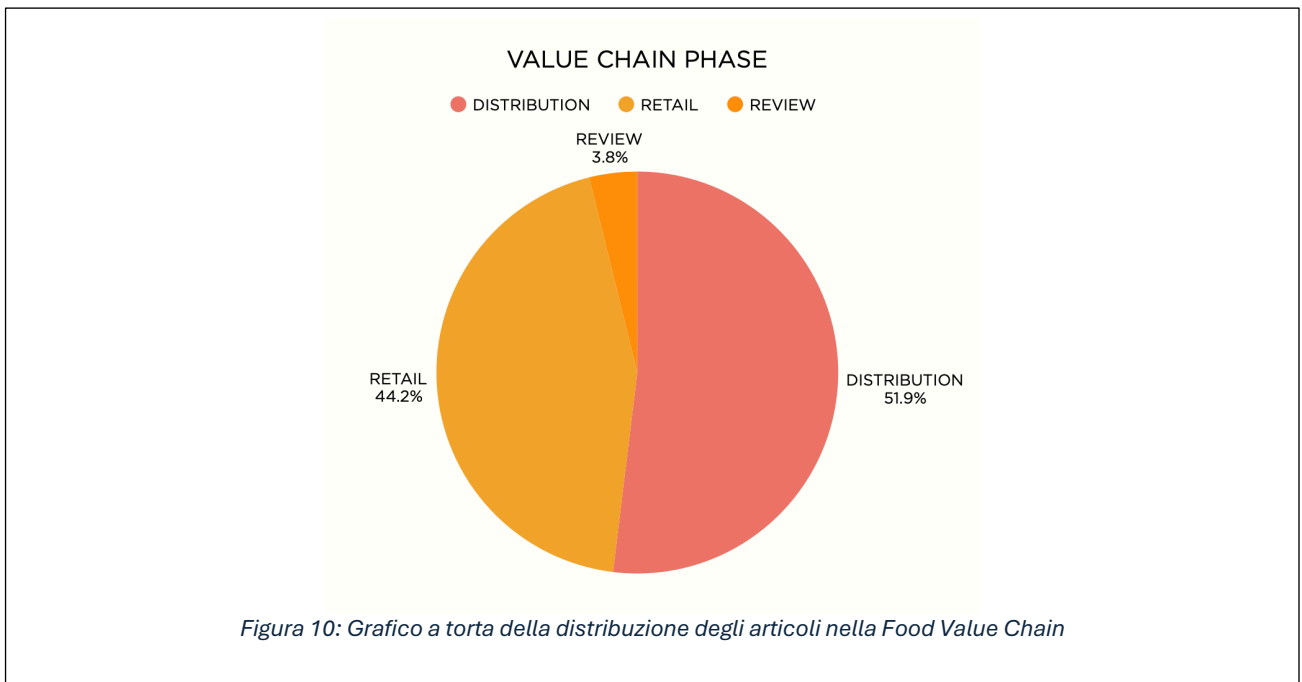
	<b>Review</b>	<b>Distribution</b>	<b>Retail</b>
<b>Utilizzo [unità]</b>	2	27	23

*Tabella 2: Ambito di applicazione degli articoli [unità]*

Dall'analisi dei dati emerge una netta polarizzazione della ricerca verso le fasi "*downstream*" della catena del valore, ovvero quelle più vicine al mercato e al consumatore finale. Nello specifico:

- *Distribution* (51,9%): questa fase rappresenta l'ambito di applicazione predominante, con un totale di 27 articoli dedicati. La centralità della distribuzione sottolinea l'importanza critica della logistica e della gestione dei flussi intermedi nel settore alimentare, dove il coordinamento tra produzione e punti vendita è vitale per preservare la freschezza del prodotto;
- *Retail* (44,2%): il commercio al dettaglio costituisce il secondo pilastro della ricerca, con 23 articoli analizzati. In questo contesto, l'attenzione è focalizzata sull'ottimizzazione degli scaffali e sulla riduzione degli sprechi direttamente nel punto vendita, dove la variabilità della domanda dei consumatori è massima;
- *Review della letteratura* (3,8%): una piccola ma significativa porzione della selezione è composta da 2 articoli di revisione. Questi contributi sono essenziali per inquadrare

l'evoluzione tecnologica globale e le tendenze metodologiche che accomunano i diversi ambiti applicativi.



Il fatto che oltre il 96% degli studi si concentri su *distribution* e *retail* evidenzia come la comunità scientifica identifichi in questi due stadi i principali colli di bottiglia dell'efficienza alimentare. È proprio in queste fasi, infatti, che la deperibilità dei prodotti e l'incertezza della domanda generano i maggiori rischi economici e ambientali.

### 4.3 Classificazione merceologica e cluster di prodotto

Un passaggio cruciale per l'organizzazione dei risultati della ricerca è stata l'analisi della tipologia di prodotto o del servizio analizzato in ciascun contributo scientifico. Inizialmente, per ogni articolo è stata individuata con precisione la categoria merceologica specifica su cui verteva l'analisi. Successivamente, al fine di garantire un'analisi più ampia e statisticamente significativa, questi singoli soggetti sono stati raggruppati in 7 cluster definiti, riportati in *figura 11*, capaci di includere l'intera eterogeneità dei casi studio rilevati.

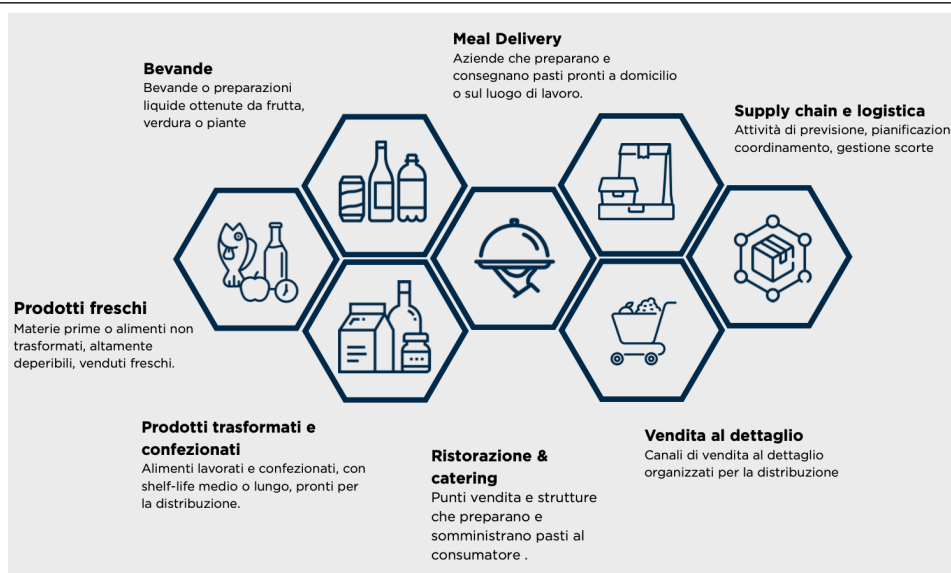


Figura 11: I 7 cluster di casi d'uso identificati

I sette raggruppamenti sono stati definiti come segue:

1. **Prodotti freschi:** comprende materie prime o alimenti non trasformati, caratterizzati da un'alta deperibilità, come pesce fresco, frutta e prodotti da forno;
2. **Prodotti trasformati e confezionati:** include alimenti lavorati con una *shelf-life* medio-lunga, pronti per la distribuzione, come prodotti dolciari e cioccolato;
3. **Bevande:** raggruppa bevande o preparazioni liquide pronte al consumo, come succhi di frutta, verdura e caffè;
4. **Ristorazione/Catering:** si riferisce a punti vendita e strutture (come mense universitarie o servizi per eventi) che preparano e somministrano pasti direttamente al consumatore finale;
5. **Meal Delivery:** include le aziende specializzate nella preparazione e consegna di pasti pronti a domicilio o sul luogo di lavoro;
6. **Vendita al dettaglio:** si focalizza sui canali di vendita organizzati per la distribuzione alimentare, tipicamente supermercati e panetterie;
7. **Supply chain e logistica:** comprende le attività trasversali di coordinamento, pianificazione della distribuzione e ottimizzazione delle scorte all'interno delle reti distributive.

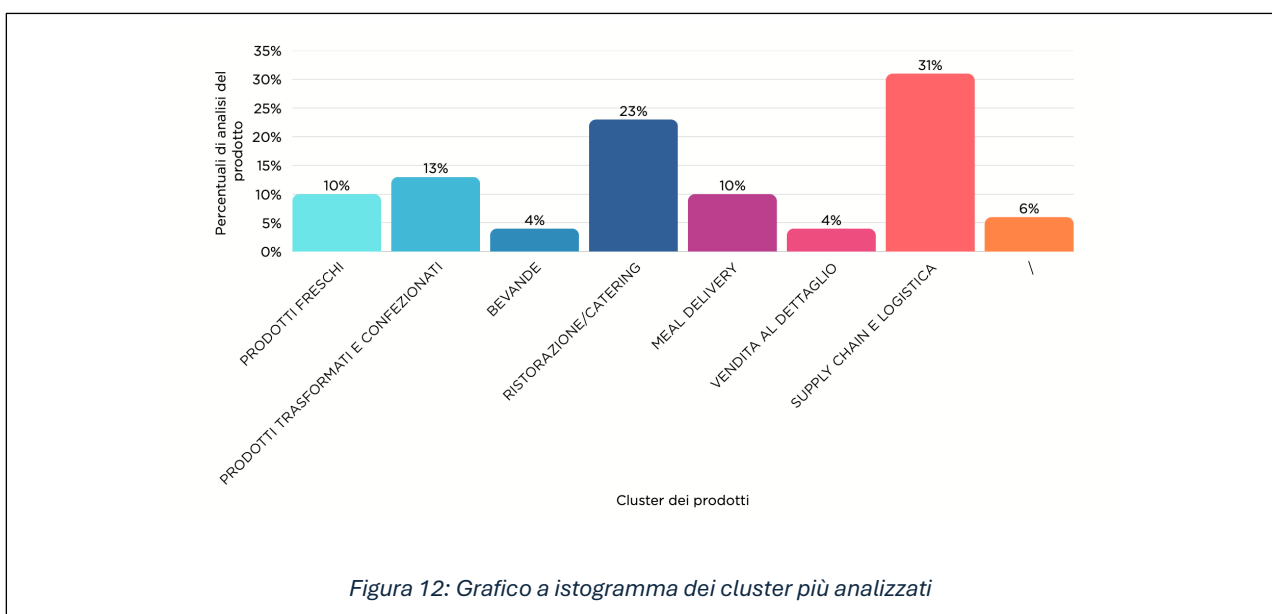
<b>Cluster</b>	<b>Esempio voci assegnate</b>	<b>Articoli di riferimento</b>
<b>Prodotti freschi</b>	Fresh fish, Fresh food market, Bananas, Bread, Sunflower seeds	[18], [19], [20], [21], [22]
<b>Prodotti trasformati e confezionati</b>	Packaged food products, Processed food, Long-shelf-life products (especially confectionery), Chocolate products	[8], [23], [24], [1], [25], [26], [27]
<b>Bevande</b>	Liquid products based on fruits and vegetables with different recipe, Coffee company	[16], [28]
<b>Ristorazione/Catering</b>	Restaurant, University refectory, Peruvian food manufacturing, Food catering services	[29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40]
<b>Meal Delivery</b>	Meal delivery service	[7], [41], [5], [42], [43]
<b>Vendita al dettaglio</b>	Supermarket, Bakery	[44], [45]
<b>Supply chain e logistica</b>	Logistical coordination in distribution networks, Supply chain forecasting and inventory optimization, Optimizing inventory management in a meal supply business, Profit-making fulfillment centers	[46], [47], [2], [3], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59]

Tabella 3: Assegnazione di ogni articolo al cluster di appartenenza

L'inserimento di ciascuna tipologia di prodotto nel relativo cluster ha permesso di osservare la distribuzione dell'interesse scientifico. Dall'analisi quantitativa rappresentata nella *figura 12* emerge quanto segue:

- il settore più analizzato è quello della Supply chain e logistica, che copre il 31% dei casi studiati. Questo dato conferma che la ricerca è fortemente orientata all'ottimizzazione dei flussi e del coordinamento logistico piuttosto che al singolo prodotto fisico;

- al secondo posto si colloca la Ristorazione e il Catering, con il 23% degli articoli. La complessità della gestione dei pasti in strutture collettive rappresenta dunque un terreno fertile per l'applicazione di modelli predittivi;
- seguono i prodotti trasformati e confezionati (13%), mentre i prodotti freschi e il *meal delivery* incidono entrambi per il 10%.
- le percentuali minori si registrano per il settore delle bevande (4%) e della vendita al dettaglio (4%), indicando ambiti in cui l'applicazione specifica descritta negli articoli è meno frequente rispetto ai temi logistici generali.



In conclusione, i risultati evidenziano come la ricerca sul *forecasting* alimentare tenda a privilegiare la gestione dei processi complessi (logistica e ristorazione) rispetto alla previsione legata a singole categorie di prodotto.

#### 4.4 Analisi comparativa degli algoritmi: utilizzo vs performance

L'analisi degli articoli selezionati ha seguito un processo metodologico rigoroso suddiviso in due fasi principali per quanto riguarda lo studio degli algoritmi: in primo luogo, per ogni studio è stato effettuato un censimento completo di tutti gli algoritmi impiegati per affrontare il caso specifico e, in un secondo momento, è stato isolato quello che ha garantito la migliore performance tra quelli messi a confronto.

Complessivamente sono stati individuati 69 algoritmi differenti, i quali sono stati successivamente raggruppati in 11 cluster di livello superiore per facilitarne la comparazione. Questi cluster sono stati, infine, ricondotti a quattro macrocategorie tecnologiche: metodi statistici, *machine learning*, *deep learning* e *hybrid*.

Nella *tabella 4* è possibile visualizzare la divisione effettuata:

- **Metodi statistici:** questo cluster rappresenta l'approccio più tradizionale alla previsione. Include la Regressione Lineare, estesa alle sue varianti regolarizzate per una maggiore precisione, e i Modelli Statistici di Serie Temporal. Nelle fonti, questi modelli fungono spesso da termine di paragone per le tecnologie più recenti;
- **Machine Learning:** si tratta di una macrocategoria estremamente variegata che spazia da modelli logici semplici a tecniche aggregate. Comprende gli Alberi di Decisione singoli e la più robusta famiglia degli Ensemble di Alberi (che raggruppa tecniche di *Bagging*, *Boosting* e il *Random Forest*). Altri approcci inclusi sono i metodi basati sulla distanza come il KNN, le *Support Vector Machines* (SVM) e i modelli probabilistici come i *Gaussian Process*;
- **Deep Learning:** questo gruppo si concentra sulle architetture ispirate al funzionamento del cervello umano, classificate nelle fonti come Reti Neurali Classiche e Avanzate. Tra le varianti specifiche citate compaiono le *Neural Networks* (NN) e le *Long Short-Term Memory* (LSTM), queste ultime particolarmente adatte a gestire sequenze di dati temporali complessi;
- **Hybrid:** la categoria rappresenta l'integrazione di diverse filosofie algoritmiche. Include Ensemble Avanzati e tecniche di *Stacking*, che combinano modelli diversi per compensarne le reciproche debolezze, oltre ad Algoritmi evolutivi e Modelli Temporal ibridi progettati per contesti specifici della catena del valore.

Questa suddivisione è stata fondamentale per analizzare come i diversi approcci vengano applicati a vari ambiti, come la previsione della domanda o la gestione delle scorte, e a diverse fasi della *Food Value Chain*.

<b>Metodi statistici</b>	Regressione Lineare e Varianti Regolarizzate
	Modelli Statistici di Serie Temporal
<b>Machine Learning</b>	Alberi di Decisione e Regressione su Alberi (non ensemble)
	Ensemble di Alberi (Bagging, Boosting, Random Forest)

	Support Vector Machines
	KNN e Metodi a Base di Distanza
	Gaussian Process e Modelli Probabilistici
<b>Deep Learning</b>	Reti Neurali Classiche e Avanzate
<b>Hybrid</b>	Ensemble Avanzati / Stacking
	Algoritmi evolutivi
	Modelli Temporali Ibridi / Specifici

Tabella 4: Suddivisione per cluster degli algoritmi

#### 4.4.1 Valutazione degli algoritmi più utilizzati

Il diagramma di Sankey rappresentato nella *figura 13* analizza la diffusione delle diverse metodologie computazionali impiegate negli studi selezionati, raggruppando i 69 algoritmi individuati in quattro macrocategorie e 11 cluster di livello superiore.

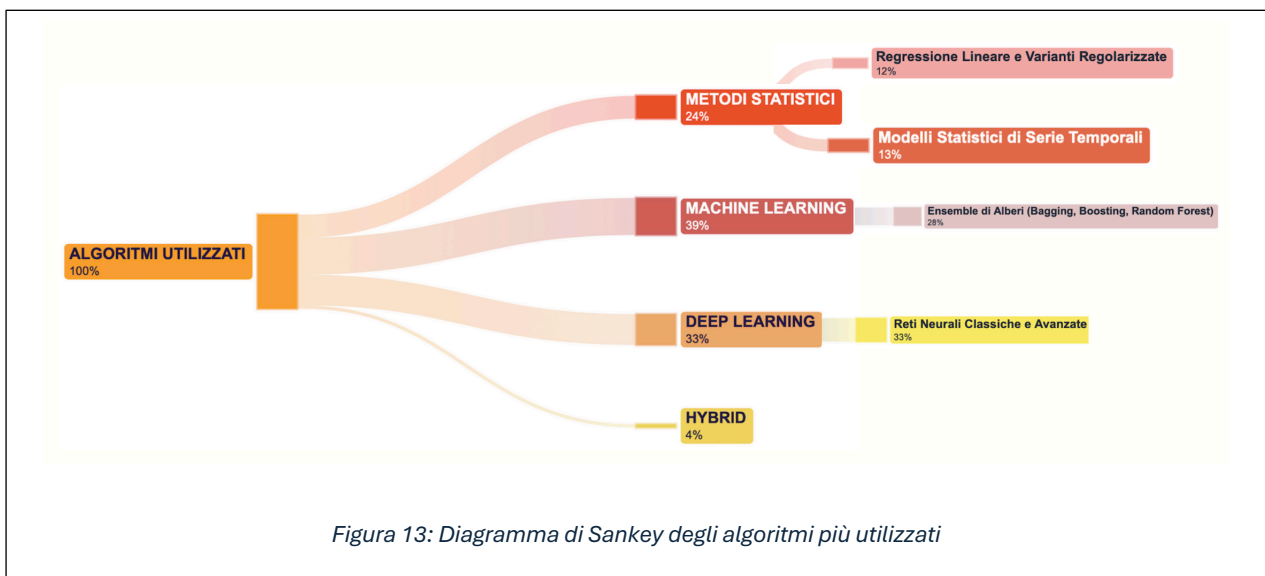
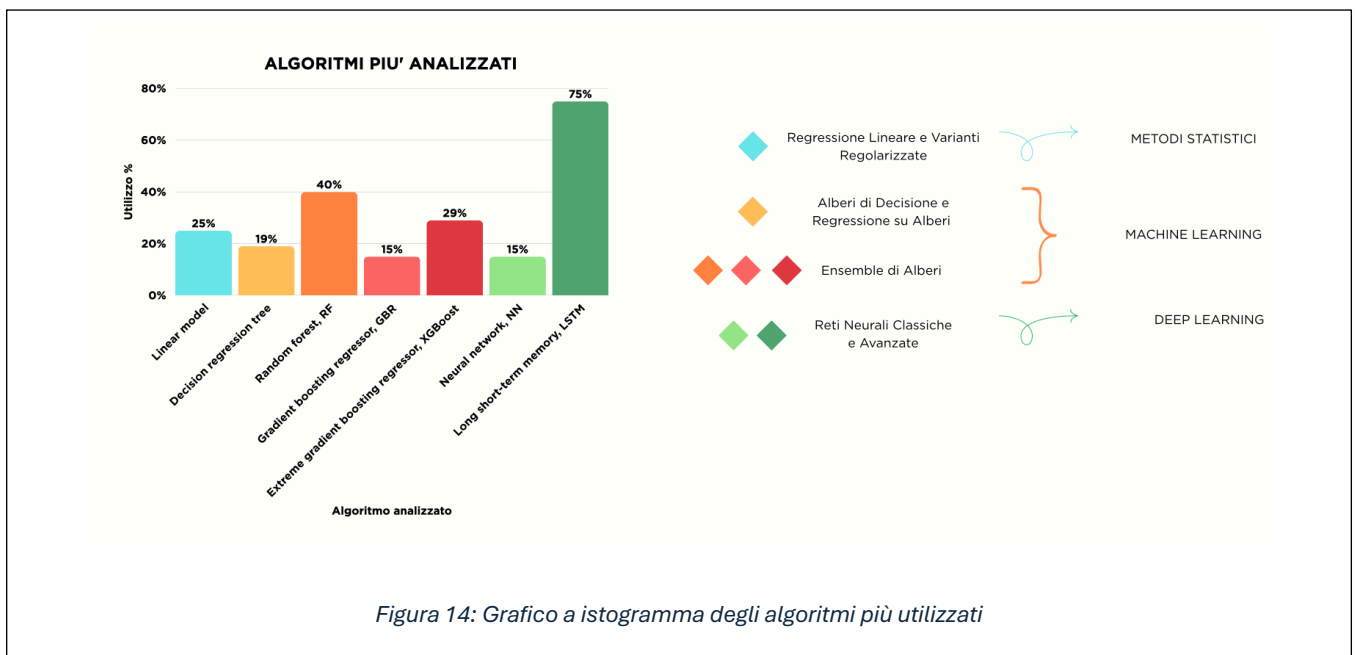


Figura 13: Diagramma di Sankey degli algoritmi più utilizzati

A livello macroscopico, il *machine learning* si conferma la scelta predominante tra i ricercatori rappresentando il 39% degli algoritmi totali utilizzati, con un ruolo centrale ricoperto dal cluster degli Ensemble di Alberi — comprendente *Random Forest*, *Bagging* e *Boosting* — che incide per il 28%.

Un aspetto di particolare rilievo emerge, tuttavia, dalla scomposizione dei singoli cluster, poiché, sebbene il *machine learning* sia la categoria più ampia nel complesso, il gruppo di algoritmi più frequente in assoluto risulta essere quello delle "Reti Neurali Classiche e Avanzate" appartenente al *deep learning*, il quale da solo copre il 33% del totale degli impieghi.

Parallelamente, i metodi statistici mantengono una presenza significativa del 24%, suddivisa in modo quasi paritario tra i modelli di serie temporali (13%) e la regressione lineare con le sue varianti regolarizzate (12%), mentre le soluzioni Ibride, pur rappresentando una frontiera avanzata, occupano attualmente una quota minoritaria pari al 4%. In sintesi, emerge un panorama tecnologico fortemente orientato verso l'intelligenza artificiale, in cui la flessibilità delle Reti Neurali e la robustezza degli Ensemble di Alberi dominano la scena applicativa nel settore agroalimentare secondario.



Dall'analisi della *figura 14*, emerge una netta predominanza di alcune tecniche specifiche che guidano la ricerca scientifica nel settore. L'algoritmo *Long Short-Term Memory (LSTM)*, appartenente al cluster delle Reti Neurali Classiche e Avanzate (*Deep Learning*), si posiziona come il leader indiscusso, con un utilizzo nel 75% degli studi analizzati. Al secondo posto per diffusione si trova il *Random Forest (RF)*, impiegato nel 40% dei casi, il quale rappresenta l'esponente di punta del cluster Ensemble di Alberi all'interno della macrocategoria *Machine Learning*.

Sempre nell'ambito degli ensemble, riscontra un notevole successo l'*Extreme Gradient Boosting (XGBoost)* con un utilizzo del 29%, mentre il *Gradient Boosting Regressor (GBR)* si attesta su una frequenza più contenuta del 15%. Per quanto concerne i metodi più tradizionali, il modello lineare (Metodi Statistici) mantiene una quota di adozione rilevante pari al 25%, dimostrandosi ancora più diffuso rispetto agli Alberi di Decisione singoli (19%) e alle Reti Neurali (NN) classiche (15%). Questa gerarchia di utilizzo riflette una chiara preferenza della

letteratura per modelli avanzati capaci di catturare dipendenze temporali complesse e di integrare molteplici variabili decisionali.

#### 4.4.2 Valutazione degli algoritmi più performanti

Dopo aver esaminato la diffusione degli strumenti computazionali, l'analisi si è focalizzata sull'individuazione dell'algoritmo che ha garantito la migliore performance per ciascuno studio, procedendo poi a inserirlo nel rispettivo cluster di appartenenza, come viene mostrato nella *tabella 5*. In ogni articolo esaminato, infatti, veniva indicato esplicitamente quale tra i modelli testati risultasse il più efficace, una valutazione resa possibile grazie all'impiego di diverse metriche di misurazione.

<b>Famiglia di algoritmi</b>	<b>Cluster di appartenenza</b>	<b>Algoritmo più performante</b>	<b>Articolo di riferimento</b>
<b>Metodi statistici</b>	Regressione Lineare e Varianti Regolarizzate	Linear model with stepwise selection	[22], [48]
	Modelli Statistici di Serie Temporali	SARIMA	[26]
		RRC	[8]
		Tracking Patterns	[51], [55]
		STL	[37]
		Moving Average, MA	[21]
Holt-Winters statistical model	[21], [37]		
<b>Machine Learning</b>	Alberi di Decisione e Regressione su Alberi (non ensemble)	Decision regression tree	[28]
	Ensemble di Alberi (Bagging, Boosting, Random Forest)	(Ensemble Decision Tree) EDT Models Bagged	[32]

		Random forest, RF	[29], [31], [47], [5], [20], [48], [50], [22], [59]
		Gradient boosting regressor, GBR	[16], [41]
		Light gradient boosting regressor, LightGBM	[33]
		Extreme gradient boosting regressor, XGBoost	[19], [3], [42], [39], [57]
		Extra Trees Regressor	[59]
		Ensemble (XGBoost + CatBoost)	[36]
		Support Vector Machines	SVM
<b>Deep Learning</b>	Reti Neurali Classiche e Avanzate	Neural network, NN	[53]
		RNN	[44]
		Feedforward ANN, FFNN	[46]
		Weight-Constrained Neural Network, WCNN	[38]
		Long short-term memory, LSTM	[7], [18], [29], [1], [2], [54]
		Blended-LSTM-GRU	[58]
		LSTM + RF	[23], [2]
		Nonlinear autoregressive exogenous neural network, NARXNN	[24], [35]

		DeepAR, DAR	[27]
		Multi-headed Transformer, MhT	[27]
		Temporal Fusion Transformer, TFT	[27]
<b>Hybrid</b>	Ensemble Avanzati / Stacking	Stacking ensemble con XGBoost come meta-modello	[34]
		Stacking of RF, XGBoost, ANN, SVM	[30]
	Algoritmi evolutivi	TOR AI	[45]
	Modelli Temporal Ibridi / Specifici	Prophet	[43]

Tabella 5: Relazione tra l'algoritmo più performante, il cluster di appartenenza e l'articolo esaminato

Dall'elaborazione dei dati presentati nelle fonti emerge una netta prevalenza delle tecniche di *Machine Learning*, come si può notare della *figura 15*, che si confermano le più efficaci nel 41% dei casi analizzati. In particolare, il cluster degli "Ensemble di Alberi" si distingue come il più performante in assoluto con una quota del 37%, superando le "Reti Neurali Classiche e Avanzate" che si attestano al 33%.

È rilevante notare come i Metodi Statistici tradizionali risultino i più efficaci solo nel 19% degli studi, mentre le architetture Ibride chiudono la classifica con il 7% di casi in cui sono risultate superiori alle altre. Questi risultati indicano una chiara tendenza della ricerca verso modelli capaci di gestire l'elevata complessità dei dati, premiando la robustezza degli algoritmi di ensemble rispetto alla popolarità delle reti neurali o alla semplicità dei modelli lineari.

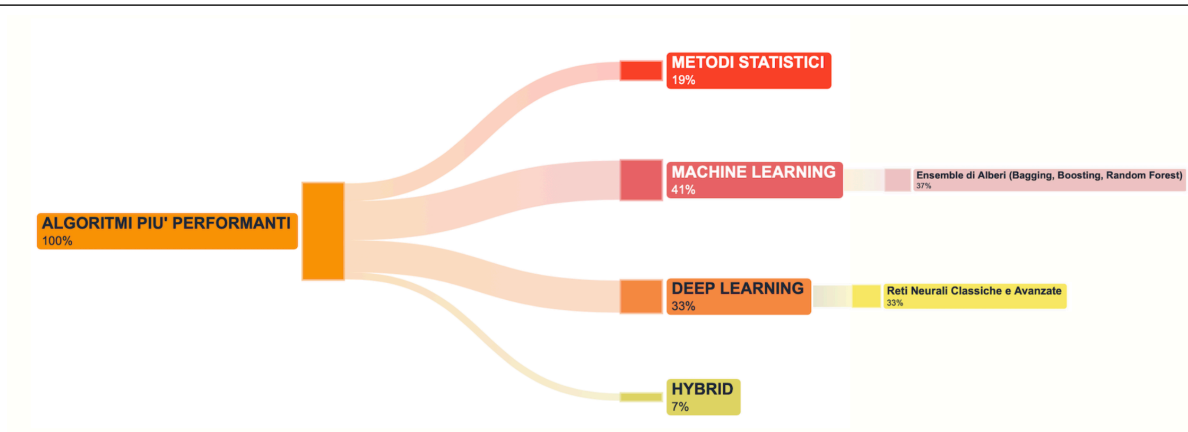


Figura 15: Diagramma di Sankey degli algoritmi più performanti

Dall'analisi dei singoli algoritmi si evince come il *Random Forest* si sia distinto con una frequenza del 17% come il più performante.

Seguono in questa classifica, l'*Extreme Gradient Boosting Regressor (XGBoost)*, con una frequenza di successo del 9%, e le *Neural Networks (NN)* classiche, che sono risultate le migliori nel 6% degli studi. Questi dati confermano, infatti, la superiorità operativa del cluster Ensemble di Alberi (Machine Learning), rispetto alle Reti Neurali Classiche e Avanzate (Deep Learning).

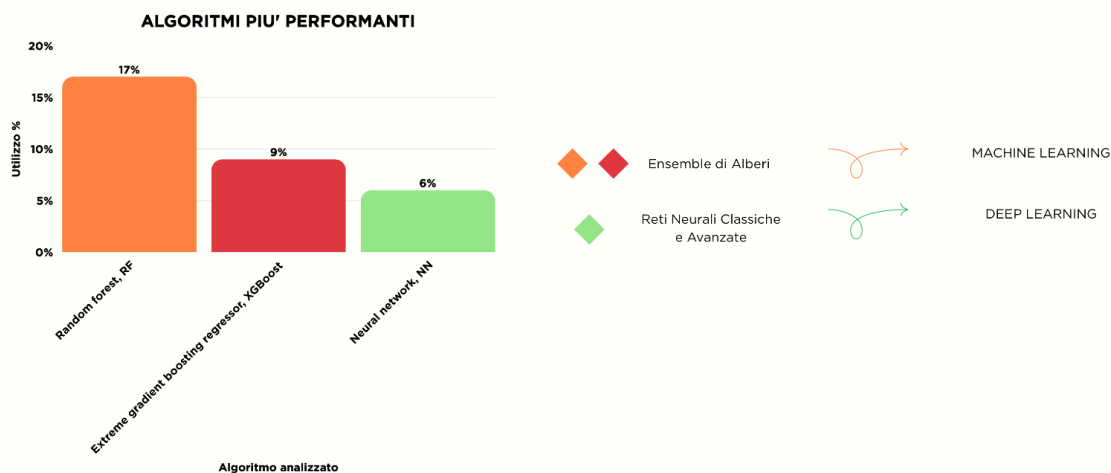


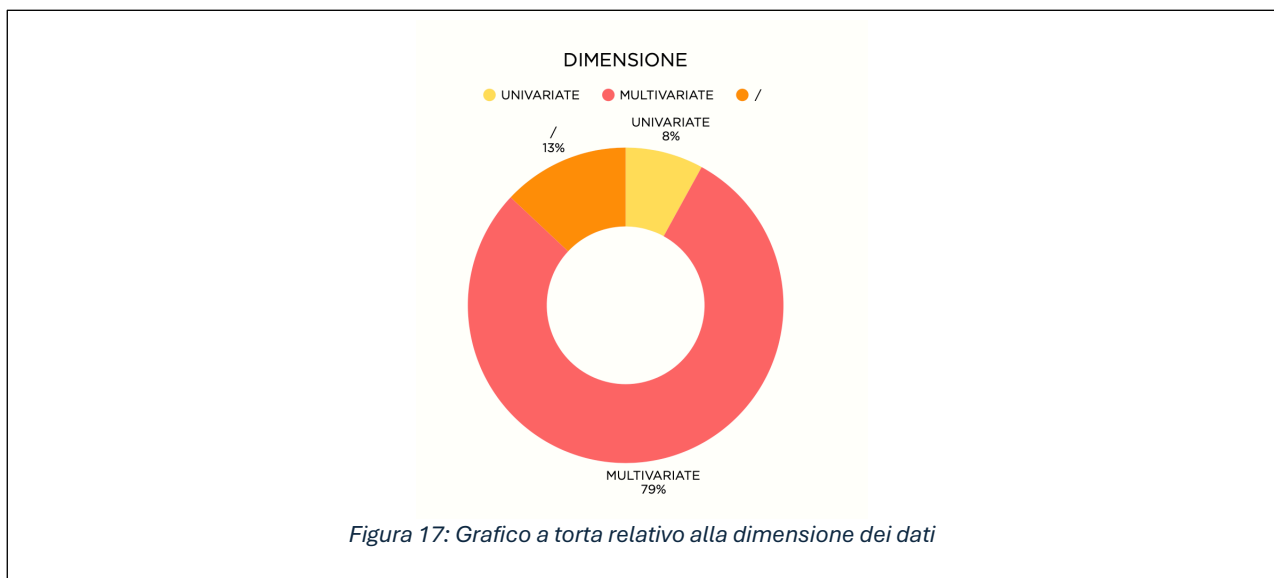
Figura 16: Grafico a istogramma degli algoritmi più performanti

## 4.5 Caratterizzazione dei dati e architetture informative

Il lavoro di ricerca ha, successivamente, approfondito le caratteristiche dei set di dati impiegati, riportate nell'Allegato B, focalizzandosi in particolare sulla loro fonte, sull'estensione temporale, sulla dimensione e sulla frequenza di rilevamento.

Nel contesto della dimensione dei dati analizzati, è fondamentale distinguere tra la natura univariata e quella multivariata delle informazioni raccolte per i modelli di previsione:

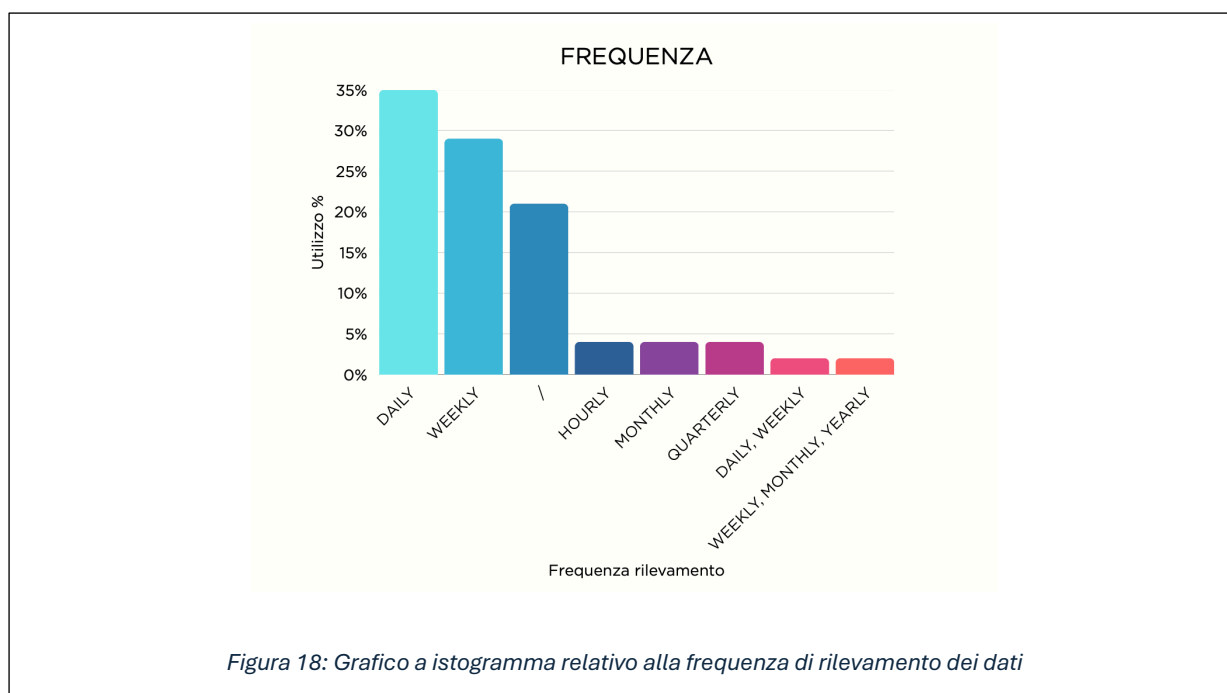
- dati Univariati: si riferiscono a studi che prendono in considerazione una singola variabile (come, ad esempio, lo storico delle vendite di un solo prodotto) per effettuare una previsione. Nelle fonti analizzate, questo approccio risulta essere minoritario, rappresentando solo l'8% dei casi, come viene riportato nella *figura 17*;
- dati Multivariati: si caratterizzano per l'utilizzo di più variabili contemporaneamente per formulare una stima. Questo approccio è largamente predominante, con una diffusione del 79%, poiché i ricercatori preferiscono integrare diversi fattori e variabili per migliorare l'accuratezza della previsione della domanda;
- informazione non indicata: è importante specificare che in alcuni articoli questa informazione non era presente: nello specifico, per il 13% del campione analizzato.



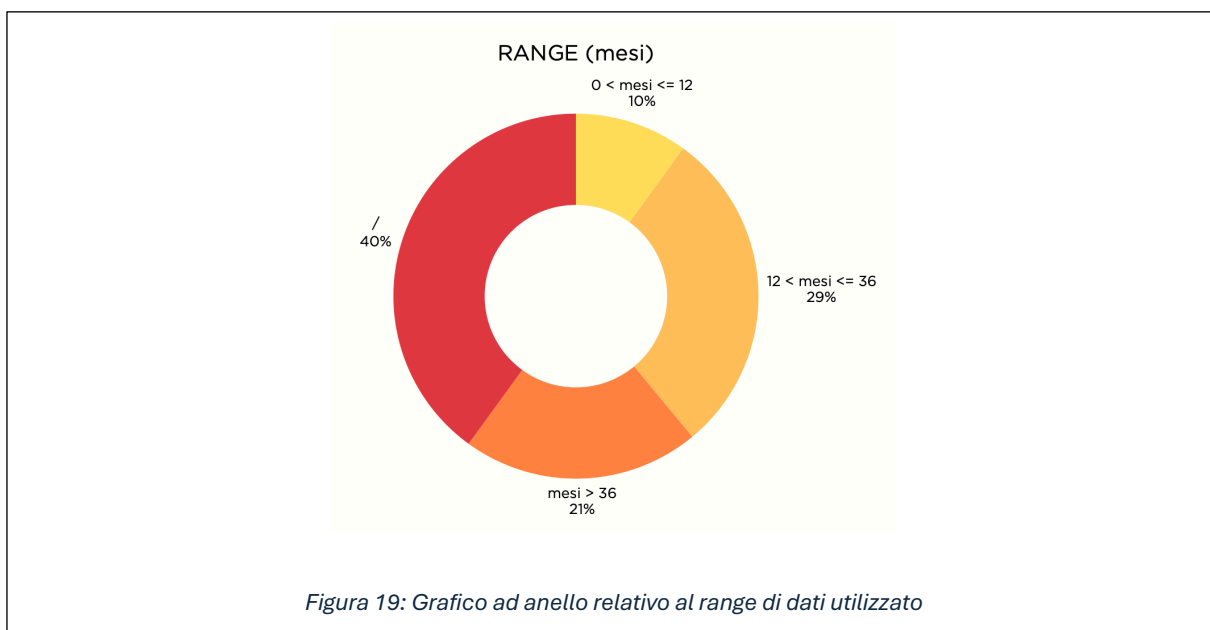
In sintesi, la preferenza per i dati multivariati riflette la necessità di catturare la complessità dei fenomeni reali, dove un singolo parametro non è quasi mai sufficiente a spiegare o prevedere correttamente l'andamento del mercato agroalimentare.

Per quanto riguarda la frequenza di rilevamento dei dati, l'analisi delle fonti evidenzia una distribuzione variegata che rispecchia la necessità di monitorare i fenomeni agroalimentari a diversi intervalli temporali per affinare la previsione della domanda. Le frequenze più comuni emerse negli studi, come riporta il grafico in *figura 18*, sono quella giornaliera (*Daily*), settimanale (*Weekly*) e mensile (*Monthly*), che rappresentano i principali standard per l'osservazione dei flussi operativi. Accanto a questi, i ricercatori utilizzano anche granularità più specifiche, come quella oraria (*Hourly*) per analisi di estremo dettaglio o quella trimestrale (*Quarterly*) per una prospettiva più ampia.

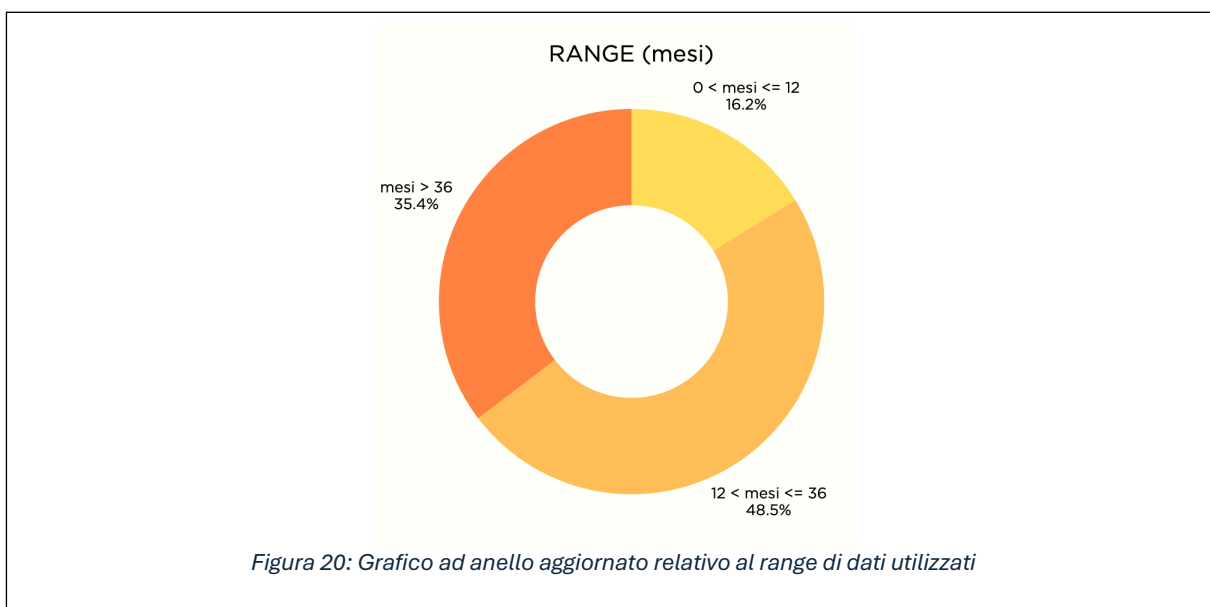
Un aspetto rilevante è l'impiego di approcci combinati, dove nello stesso studio vengono integrate frequenze diverse, come quella giornaliera e settimanale o un mix di dati settimanali, mensili e annuali. Questa scelta metodologica suggerisce la volontà di catturare contemporaneamente sia le variazioni rapide del mercato sia le tendenze stagionali di lungo periodo, fornendo una visione d'insieme più robusta e precisa.



L'analisi iniziale relativa all'arco temporale dei dati, illustrata nella *figura 19*, rivela un panorama in cui l'informazione risulta spesso assente, dato che per ben il 40% degli articoli analizzati il range non viene specificato. Tra gli studi che invece forniscono questo dettaglio, il 29% utilizza dati che coprono un periodo compreso tra i 12 e i 36 mesi, il 21% si spinge oltre i 36 mesi, mentre solo il 10% si concentra su un orizzonte temporale inferiore all'anno.



Per superare il limite imposto dall'elevata percentuale di dati mancanti e rendere i risultati più significativi, nelle fonti è stata compiuta la scelta di svolgere l'analisi esclusivamente sui 31 articoli restanti, escludendo quelli privi di specifica. Questa operazione di pulizia del campione permette di osservare la reale distribuzione delle preferenze metodologiche tra i ricercatori che dichiarano il range temporale utilizzato, ottenendo così il grafico aggiornato riportato in *figura 20*.



È evidente come la preferenza per il medio periodo diventi ancora più marcata, con il 48,5% degli studi che utilizza dati compresi tra i 12 e i 36 mesi. Anche l'adozione di serie storiche a lungo termine (oltre i 36 mesi) assume un peso rilevante arrivando al 35,4%, mentre la fascia

dei dati a breve termine, ovvero entro i 12 mesi, rimane la meno rappresentata con una quota del 16,2%.

In conclusione, quando il dato è dichiarato, emerge chiaramente come la ricerca nel settore agroalimentare tenda a basarsi su osservazioni che coprono almeno un anno intero, probabilmente per gestire meglio la stagionalità del settore.

Data la netta prevalenza dell'approccio multivariato (79% degli studi), l'analisi si è spinta a identificare quali variabili specifiche, definite *features*, vengano integrate nei modelli per affinare la precisione delle previsioni nella filiera agroalimentare. Queste variabili non riguardano solo il prodotto in sé, ma abbracciano una vasta gamma di fattori esterni e logistici che influenzano la domanda.

Dalla mappatura delle caratteristiche considerate negli studi esaminati, emergono quattro categorie principali per frequenza di utilizzo:

1. *Time series features* (40% di utilizzo): questa è la categoria più frequente e comprende informazioni temporali fondamentali come il giorno della settimana, il mese, l'anno e l'indicazione se si tratti di un giorno festivo. L'alta percentuale di utilizzo sottolinea come la stagionalità e il calendario siano i piloni principali della domanda nel settore agroalimentare;
2. *Transaction details* (29% di utilizzo): include i dati economici diretti della transazione, ovvero il prezzo (attuale e massimo), il volume di vendita e la classificazione in categorie e sottocategorie merceologiche;
3. Informazioni relative al negozio/stabilimento (27% di utilizzo): queste variabili contestualizzano la vendita analizzando la città, lo stato e la presenza di infrastrutture come il parcheggio disponibile;
4. Caratteristiche sul prodotto (25% di utilizzo): scendono nel dettaglio qualitativo e fisico del bene, prendendo in esame il tipo di cucina, gli ingredienti e il peso.

Oltre a queste categorie dominanti, i ricercatori attingono a un ventaglio di informazioni ancora più granulari, la cui distribuzione è riportata nell'Allegato C:

- Fattori esterni e ambientali: vengono considerati gli *Economy indicator* (prezzo della benzina, del diesel e indici economici generali) e il tempo meteorologico, monitorando temperatura massima, minima e velocità del vento;
- Dinamiche operative e di mercato: si analizzano la domanda storica del prodotto  $p$  nel tempo  $q$ , i costi di inventario e l'impatto di promozioni o scontistiche;

- Produzione e target: troviamo variabili legate alle caratteristiche della produzione, alla tipologia di packaging e a fattori sociodemografici indicati come *customer features* e *population features*.

Per determinare quale algoritmo risulti più performante tra quelli confrontati negli studi, nei diversi articoli sono state utilizzate le metriche di valutazione.

Dall'analisi delle fonti emerge una gerarchia precisa, rappresentata in *tabella 6*, basata sulla frequenza di utilizzo di questi parametri:

1. RMSE (52%) e MAE (52%): sono le metriche più diffuse e vengono utilizzate con la medesima frequenza. L'RMSE (*Root Mean Square Error*) è preferito quando si vuole penalizzare maggiormente gli errori più grandi, al contrario, il MAE (*Mean Absolute Error*) misura l'errore medio;
2. MAPE (31%): rende l'errore confrontabile universalmente;
3.  $R^2$  (19%) e MSE (19%): l' $R^2$  indica quanto il modello sia in grado di spiegare la varianza dei dati, mentre l'MSE (*Mean Squared Error*), come l'RMSE, serve a evidenziare i grandi errori più di altre metriche;
4. *Accuracy* (17%): rappresenta la metrica più semplice, indicando puramente la percentuale di previsioni corrette rispetto al totale delle osservazioni effettuate.

<b>Metrica di valutazione</b>	<b>Utilizzo [unità]</b>	<b>Utilizzo %</b>
<b>RMSE</b>	27	52%
<b>MAE</b>	27	52%
<b>MAPE</b>	16	31%
<b>R<sup>2</sup></b>	10	19%
<b>MSE</b>	10	19%
<b>ACCURACY</b>	9	17%

*Tabella 6: Metriche di valutazione più utilizzate*

## 4.6 Verso l'integrazione: introduzione alle correlazioni multidimensionali

Questo paragrafo conclusivo pone le basi per un'interpretazione di secondo livello, necessaria per superare l'analisi isolata delle singole dimensioni finora osservate.

L'analisi si sposta sull'interazione reciproca tra tre pilastri fondamentali:

- La fase della value chain analizzata;
- La tipologia di prodotto agroalimentare;
- L'algoritmo più performante, la cui efficacia non è assoluta ma validata attraverso le metriche di valutazione emerse dalle fonti.

# Capitolo 5 – Interpretazione di secondo livello: analisi delle correlazioni

## 5.1 Obiettivi dell'analisi multidimensionale

L'obiettivo fondamentale di questo capitolo è quello di indagare l'esistenza di correlazioni significative tra le diverse dimensioni emerse durante la fase di analisi iniziale. Mentre l'interpretazione di primo livello si è concentrata sulla descrizione isolata delle variabili, l'analisi multidimensionale mira a determinare come questi elementi interagiscano tra loro per delineare tendenze specifiche nel panorama della ricerca.

Nello specifico, gli obiettivi prefissati per questa sezione sono:

- individuare relazioni tra l'ambito di applicazione, la tipologia di prodotto e la fase della *value chain*: si vuole comprendere se determinati settori o prodotti tendano a concentrarsi in fasi specifiche della catena del valore;
- analizzare il nesso tra algoritmi e contesto operativo: l'analisi punta a verificare se esista una corrispondenza tra gli algoritmi più citati (e successivamente quelli più performanti) e la tipologia di prodotto esaminata nelle diverse fasi della *value chain*;
- approfondire la dipendenza tra dati e applicazioni: l'obiettivo è esaminare come l'ambito di applicazione degli articoli influenzi la tipologia e la natura dei dati necessari per lo svolgimento dello studio.

In sintesi, il capitolo cercherà di rispondere alla domanda centrale se vi siano effettive interdipendenze tra le dimensioni esaminate, fornendo una visione d'insieme più articolata e profonda dei risultati ottenuti.

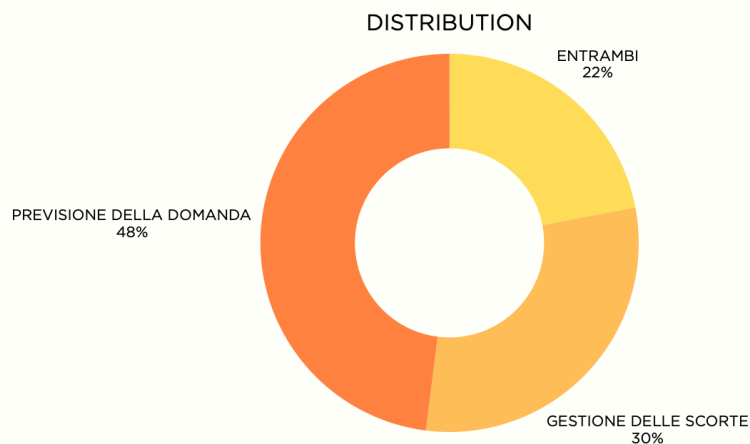
## 5.2 Correlazione tra *value chain* e obiettivi della ricerca

In questo paragrafo viene analizzata la correlazione tra le fasi della catena del valore e la tipologia di analisi prevalente negli studi esaminati, concentrandosi specificamente sulla distinzione tra la previsione della domanda e la gestione delle scorte.

Dall'analisi dei dati emerge una differenza significativa nel focus della ricerca a seconda che l'articolo tratti della fase di distribuzione o di quella di vendita al dettaglio:

- *Distribution*: in questo ambito, gli obiettivi della ricerca appaiono più distribuiti, sebbene la previsione della domanda rimanga l'attività principale interessando il 48%

degli articoli, come si evince dalla *figura 21*. Una parte consistente della letteratura, pari al 30%, si focalizza specificamente sulla gestione delle scorte, mentre il 22% degli studi adotta un approccio integrato che affronta entrambi gli aspetti. Ciò indica che nella fase distributiva esiste un interesse bilanciato tra la necessità di anticipare le richieste del mercato e l'esigenza di ottimizzare i livelli di stock.



*Figura 21: Grafico a torta sulla correlazione tra articoli sulla "Distribution" e obiettivi della ricerca*

- *Retail*: il panorama cambia radicalmente quando l'analisi si sposta sulla fase di vendita al dettaglio. Qui si riscontra una polarizzazione quasi totale: il 96% degli articoli si pone come obiettivo primario la previsione della domanda futura. Sorprendentemente, la ricerca dedicata esclusivamente alla gestione delle scorte in questa fase è inesistente (0%), mentre solo una minima parte degli studi (4%) analizza entrambe le dimensioni.

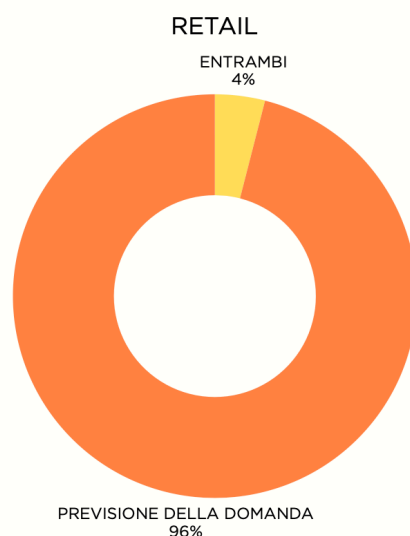


Figura 22: Grafico a torta sulla correlazione tra articoli sul "Retail" e obiettivi della ricerca

Questo divario evidenzia come, nel contesto del *Retail*, la priorità della ricerca scientifica sia quasi esclusivamente rivolta alla comprensione e all'anticipazione del comportamento d'acquisto del consumatore finale, lasciando la gestione operativa delle scorte come tematica secondaria o strettamente dipendente dalla qualità della previsione stessa.

### 5.3 Correlazione tra tipologia di prodotto esaminato e obiettivi della ricerca

Successivamente è stata esaminata la correlazione tra la natura merceologica del prodotto e gli obiettivi analitici perseguiti dalla letteratura scientifica. Tale analisi viene applicata in modo sistematico a tutti e sette i cluster di prodotti precedentemente affrontati nelle fasi descrittive della tesi, al fine di identificare pattern specifici per ogni raggruppamento:

- **Prodotti freschi:** in questo cluster, l'80% degli articoli si concentra esclusivamente sulla previsione della domanda futura. La gestione delle scorte come unico obiettivo non è trattata (0%), mentre il restante 20% degli studi adotta un approccio integrato che include entrambe le finalità;

<b>PRODOTTI</b>	Gestione delle	<b>Previsione della</b>	Entrambi
<b>FRESCHI</b>	scorte	<b>domanda</b>	
Utilizzo	0	<b>4</b>	1
Utilizzo %	0%	<b>80%</b>	20%

Tabella 7: Relazione tra articoli sui "Prodotti freschi" e analisi svolta

- Prodotti trasformati e confezionati: sebbene la previsione della domanda rimanga l'obiettivo prevalente (57%), emerge una maggiore varietà rispetto ai prodotti freschi. Il 14% degli articoli si focalizza specificamente sulla gestione delle scorte, mentre il 29% analizza entrambe le dimensioni;

<b>PRODOTTI TRASFORMATI E CONFEZIONATI</b>	Gestione delle scorte	Previsione della domanda	Entrambi
Utilizzo	1	4	2
Utilizzo %	14%	57%	29%

Tabella 8: Relazione tra articoli sui “Prodotti trasformati e confezionati” e analisi svolta

- Bevande: la ricerca in questo ambito appare perfettamente bilanciata, dividendosi equamente (50% e 50%) tra studi dedicati alla gestione delle scorte e studi sulla previsione della domanda;

<b>BEVANDE</b>	Gestione delle scorte	Previsione della domanda	Entrambi
Utilizzo	1	1	0
Utilizzo %	50%	50%	0%

Tabella 9: Relazione tra articoli sulle “Bevande” e analisi svolta

- Ristorazione / catering: la letteratura mostra una forte polarizzazione verso la previsione della domanda, che interessa il 92% degli articoli. Solo una minima parte (8%) estende l'analisi anche alla gestione delle scorte;

<b>RISTORAZIONE / CATERING</b>	Gestione delle scorte	Previsione della domanda	Entrambi
Utilizzo	0	11	1
Utilizzo %	0%	92%	8%

Tabella 10: Relazione tra articoli sulla “Ristorazione/Catering” e analisi svolta

- Vendita al dettaglio: in questo cluster, la totalità degli studi analizzati (100%) ha come unico obiettivo la previsione della domanda;

<b>VENDITA AL DETTAGLIO</b>	Gestione delle scorte	Previsione della domanda	Entrambi
Utilizzo	0	2	0
Utilizzo %	0%	100%	0%

Tabella 11: Relazione tra articoli sulla “Vendita al dettaglio” e analisi svolta

- Supply chain e logistica: questo ambito presenta una distribuzione più articolata in quanto il 44% degli articoli si occupa di previsione della domanda, ma resta molto rilevante la quota dedicata alla gestione delle scorte (38%). Il 19% degli studi tratta entrambi i temi;

<b>SUPPLY CHAIN E LOGISTICA</b>	Gestione delle scorte	Previsione della domanda	Entrambi
Utilizzo	6	7	3
Utilizzo %	38%	44%	19%

Tabella 12: Relazione tra articoli sulla “Supply chain e logistica” e analisi svolta

- Meal delivery: analogamente alla vendita al dettaglio, il 100% degli articoli esaminati per questo cluster è finalizzato alla previsione della domanda.

<b>MEAL DELIVERY</b>	Gestione delle scorte	Previsione della domanda	Entrambi
Utilizzo	0	5	0
Utilizzo %	0%	100%	0%

Tabella 13: Relazione tra articoli sul “Meal delivery” e analisi svolta

In sintesi, emerge chiaramente come i settori caratterizzati da un contatto diretto con il consumatore finale o da prodotti altamente deperibili (freschi, ristorazione, *delivery*) privilegino quasi esclusivamente la previsione della domanda. Al contrario, contesti più legati

ai flussi logistici o a prodotti a lunga conservazione (bevande, *supply chain*) dedicano un'attenzione significativa anche alla gestione ottimale dei livelli di stock.

## 5.4 Efficacia degli algoritmi in base al contesto operativo

Successivamente l'indagine si focalizza sulla valutazione dell'efficacia degli algoritmi in relazione alle specificità del contesto operativo esaminato nelle fonti. L'obiettivo primario è determinare se esistano soluzioni tecnologiche specializzate in grado di rispondere meglio alle sfide di previsione e gestione tipiche di ogni segmento della filiera alimentare.

Per fornire una visione completa e approfondita, l'analisi verrà condotta seguendo due prospettive complementari: in primo luogo, si analizzerà la correlazione tra gli algoritmi più utilizzati (ovvero quelli più frequentemente citati ed esaminati negli studi) e il contesto operativo. In secondo luogo, verrà svolta la stessa analisi ma tra gli algoritmi risultati più performanti e il medesimo contesto di applicazione.

Questa distinzione è cruciale per determinare se la scelta di un algoritmo dipenda dalla fase della *value chain* considerata, ovvero se un modello sia più indicato per la distribuzione o per la vendita al dettaglio. Inoltre, l'analisi permetterà di identificare se, all'interno di una specifica tipologia di analisi, l'applicazione di determinati algoritmi prevalga in un cluster di prodotti particolare.

### 5.4.1 Correlazione tra algoritmi analizzati e tipologia di prodotto esaminato nella fase della *value chain*

In questa sezione viene approfondito il legame tra gli strumenti computazionali impiegati nella letteratura e le caratteristiche operative degli studi selezionati.

Dall'analisi degli articoli, gli algoritmi sono stati classificati in tre macrocategorie principali: metodi statistici, *machine learning* e *deep learning*. Di seguito si riporta il dettaglio della loro applicazione in base alla fase della catena del valore (*distribution* o *retail*) e alla prevalenza nei cluster di prodotto:

- *Linear Model*: questo algoritmo appartiene al cluster dei metodi statistici e viene impiegato in 13 articoli, con una prevalenza nella fase di *distribution* con 8 casi rispetto ai 5 del *retail*. La sua applicazione è fortemente concentrata nel primo caso nel cluster *Supply Chain* e Logistica per il 75% mentre, per il *retail*, nel 40% dei casi l'algoritmo linear model è stato utilizzato nell'ambito della ristorazione e del catering;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Linear Model</b> 13	Distribution	Supply chain e logistica
	8	100%
	Retail	Ristorazione/catering
	5	40%

Tabella 14: Correlazione tra Linear Model e tipologia di prodotto esaminato nella fase della value chain

- **Decision Regression Tree:** questo algoritmo rientra nel cluster del *Machine Learning* ed è utilizzato in 10 articoli analizzati. La sua applicazione risulta più frequente nella fase di *distribution*, con 6 casi, rispetto alla fase di *retail*, che ne conta 4. Dal punto di vista degli ambiti applicativi, l'algoritmo è prevalentemente impiegato nel cluster *Supply Chain* e Logistica, che rappresenta il 67% dei casi totali. Al contrario, nei settori della ristorazione, delle bevande, della vendita al dettaglio e del *meal delivery*, l'utilizzo risulta più contenuto, attestandosi al 25% per ciascun ambito;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Decision Regression Tree</b> 10	Distribution	Supply chain e logistica
	6	67%
	Retail	Ristorazione/catering, 25%
	4	Bevande, 25%
		Vendita al dettaglio, 25%
		Meal delivery, 25%

Tabella 15: Correlazione tra Decision Regression Tree e tipologia di prodotto esaminato nella fase della value chain

- **Random Forest:** inserito nel cluster del *Machine Learning* (*Ensemble* di Alberi), conta 21 citazioni, suddivise tra 12 casi in *distribution* e 9 nel *retail*. Il 58% delle sue applicazioni nella distribuzione riguarda la *Supply Chain* e Logistica, mentre nell'altro ambito di applicazione la ristorazione/catering ricoprono il 33% e il meal delivery ha una quota del 22%;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Random Forest</b> 21	Distribution	Supply chain e logistica
	Retail	Ristorazione/catering, 33% Meal delivery, 22%

Tabella 16: Correlazione tra Random Forest e tipologia di prodotto esaminato nella fase della value chain

- *Gradient Boosting Regressor*: appartiene al cluster del *Machine Learning* ed è presente in 8 articoli, con una netta concentrazione nella fase di *distribution* (7 casi) con un 57% di utilizzo nella *Supply Chain* e Logistica, rispetto a una sola applicazione nel *retail* in ambito *meal delivery*;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Gradient Boosting Regressor</b> 8	Distribution	Supply chain e logistica
	Retail	Meal delivery

Tabella 17: Correlazione tra Gradient Boosting Regressor e tipologia di prodotto esaminato nella fase della value chain

- *Extreme Gradient Boosting Regressor*: appartenente al cluster del *Machine Learning*, e in particolare alla categoria degli *ensemble di alberi*, questo algoritmo è analizzato in 15 articoli della letteratura considerata. La sua applicazione risulta pressoché bilanciata tra la fase di *retail*, con 8 casi dei quali per il 63 % in ambito ristorazione/catering, e quella di *distribution*, con 7 casi. Per quanto riguarda gli ambiti applicativi, XGBoost trova un impiego prevalente nel cluster *Supply Chain e Logistica*, che concentra il 71% delle ricerche nella distribuzione;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Extreme Gradient Boosting Regressor</b> 15	Distribution	Supply chain e logistica
	Retail	Ristorazione/catering
	7	71%
	8	63%

Tabella 18: Correlazione tra Extreme Gradient Boosting Regressor e tipologia di prodotto esaminato nella fase della value chain

- *Neural Network*: appartenente al cluster del *Deep Learning*, questo algoritmo è oggetto di analisi in 8 studi della letteratura considerata. La sua applicazione risulta maggiormente concentrata nella fase di *retail*, con 5 casi, rispetto alla fase di *distribution*, che ne conta 3 tutti in ambito *Supply Chain e Logistica*. Per la fase *retail*, invece, il settore della ristorazione/catering viene indagato nel 40% dei casi;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Neural Network</b> 8	Distribution	Supply chain e logistica
	Retail	Ristorazione/catering
	3	100%
	5	60%

Tabella 19: Correlazione tra Neural Network e tipologia di prodotto esaminato nella fase della value chain

- *Long Short-Term Memory*: appartenente al cluster del *Deep Learning*, questo algoritmo risulta essere il più ampiamente analizzato nella letteratura considerata, con un totale di 39 articoli. Il suo impiego appare sostanzialmente bilanciato tra la fase di *distribution*, con 19 casi, e quella di *retail*, che ne conta 18. Per quanto riguarda gli ambiti applicativi, LSTM mostra una prevalenza nel cluster *Supply Chain e Logistica*, che concentra il 58% dei casi legati alla fase della distribuzione, mentre per il *retail* il 50% degli articoli analizzando l'ambito ristorazione/catering.

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Long Short-Term Memory</b> 39	Distribution	Supply chain e logistica
	Retail	Ristorazione/catering
	19	58%
	18	50%

Tabella 20: Correlazione tra Long Short-Term Memory e tipologia di prodotto esaminato nella fase della value chain

In conclusione, l'analisi sistematica dei dati evidenzia che non è stata riscontrata una correlazione rilevante tra gli algoritmi analizzati e la specifica tipologia di prodotto all'interno della catena del valore. La letteratura scientifica tende a utilizzare un'ampia varietà di modelli in modo trasversale, adattandoli a diversi contesti senza una specializzazione esclusiva per cluster merceologico.

Tuttavia, il *Gradient Boosting Regressor* rappresenta l'unica eccezione significativa, mostrando una specializzazione marcata verso la fase di *distribution* e nello specifico nel cluster relativo alle *Supply Chain* e Logistica.

Inoltre, indipendentemente dalla categoria di appartenenza dell'algoritmo (Metodi Statistici, *Machine Learning* o *Deep Learning*), la maggior parte di essi trova la sua massima applicazione nel cluster *Supply Chain* e Logistica per ciò che riguarda la fase della distribuzione, mentre in ambito retail vengono considerate maggiormente le fasi relative alla ristorazione/catering e al *meal delivery*.

#### 5.4.2 Correlazione tra algoritmi più performanti e tipologia di prodotto esaminato nella fase della value chain

Le informazioni raccolte dagli articoli analizzati permettono di definire la specializzazione tecnica degli algoritmi che hanno mostrato le migliori performance predittive. In questa fase, l'indagine non si limita a osservare quanto spesso un algoritmo venga citato, ma punta a identificare quali soluzioni garantiscano i risultati più efficaci in relazione alla fase della *value chain* e alla categoria merceologica trattata.

Di seguito viene analizzato il dettaglio delle performance per i principali algoritmi emersi dalle fonti:

- *Random Forest* (RF): questo algoritmo si attesta come il modello più performante in assoluto, con un totale di 9 occorrenze positive. La sua efficacia è marcatamente orientata verso la fase di *distribution* con 6 casi, di cui il 67% in ambito *Supply Chain* e Logistica. Per quanto concerne il *retail*, invece, trova applicazione in 3 casi con una presenza al 67% in ambito ristorazione/catering e la restante nel *meal delivery*;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Random Forest</b> 9	Distribution	Supply chain e logistica
		67%
	Retail	Ristorazione/catering, 67%
	3	Meal delivery, 33%

Tabella 21: Correlazione tra *Random Forest* e tipologia di prodotto esaminato nella fase della *value chain*

- *Long Short-Term Memory*: questo modello di *deep learning* si conferma l'algoritmo più performante in 6 casi. La sua applicazione risulta prevalente nella fase *distribution* con 4 casi, i restanti 2 riguardanti la fase *retail* avevano come soggetto della ricerca il cluster dei prodotti freschi e quello della ristorazione/catering;

Algoritmo	Food value phase	Cluster di prodotto principale
<b>Long Short-Term Memory</b> 6	Distribution	Supply chain e logistica
		50%
	Retail	Prodotti freschi, 50%
	2	Ristorazione/catering, 50%

Tabella 22: Correlazione tra *Long Short-Term Memory* e tipologia di prodotto esaminato nella fase della *value chain*

- *Extreme Gradient Boosting Regressor*: questo modello si distingue per una distribuzione operativa inversa rispetto ai precedenti. Mostra, infatti, una maggiore efficacia nel *retail* con 3 casi rispetto ai 2 nella *distribution*, la cui specializzazione è al 100% nel cluster *Supply Chain* e Logistica. Per quanto riguarda il *retail*, invece, trova

un'applicazione bilanciata nei segmenti ristorazione/catering, prodotti freschi e *meal delivery*.

Algoritmo	Food value phase	Cluster di prodotto principale
Extreme Gradient Boosting Regressor 5	Distribution 2	Supply chain e logistica 100%
	Retail 3	Ristorazione/catering, 33% Prodotti freschi, 33% Meal delivery, 33%

Tabella 23: Correlazione tra Extreme Gradient Boosting Regressor e tipologia di prodotto esaminato nella fase della value chain

In conclusione, emerge una chiara specializzazione funzionale: se la fase di distribuzione beneficia maggiormente della robustezza di Random Forest e della precisione temporale di LSTM, il settore del retail trova in XGBoost lo strumento più performante e mirato.

## 5.5 Dipendenza tra *value chain* e caratteristiche dei dati

L'efficacia dei modelli di analisi nella *supply chain* alimentare non dipende esclusivamente dalla scelta dell'algoritmo, ma è strettamente legata alla natura e alla struttura dei dati utilizzati, che variano significativamente in base alla fase della *value chain* considerata. Lo studio degli articoli esaminati permette di identificare dei pattern ricorrenti per quanto riguarda la frequenza, la dimensione (univariata o multivariata) e l'estensione temporale (range) delle rilevazioni. Per questa analisi sono stati esclusi i 2 articoli classificati come "review".

Le principali dinamiche emerse sono le seguenti:

- *Distribution* (27 articoli): in questo segmento, i dati mostrano una preferenza per la frequenza giornaliera al 37%, suggerendo la necessità di un monitoraggio costante e granulare per gestire i flussi logistici. La struttura dei dati è prevalentemente multivariata (70%), indicando che le analisi non si basano solo sulla serie storica di una variabile, ma integrano molteplici fattori influenzanti. Mentre, per quanto riguarda l'estensione temporale, la maggioranza degli studi (53%) utilizza dataset che coprono un arco di tempo compreso tra 12 e 36 mesi;

Frequenza	Dimensione	Range (mesi)
-----------	------------	--------------

<b>Distribution</b>	Daily	Multivariate	12 < mesi <= 36
(27 articoli)	37%	70%	53%

Tabella 24: Correlazione tra la fase "Distribution" e dati di supporto all'analisi effettuata nell'articolo

- **Retail** (23 articoli): nel commercio al dettaglio, si osserva uno spostamento verso una frequenza settimanale al 33%. La natura dei dati è ancor più marcatamente multivariata (78%) rispetto alla distribuzione, evidenziando come le vendite al dettaglio siano condizionate da un numero elevato di variabili esterne (promozioni, stagionalità, prezzi dei concorrenti). Anche in questo caso, il range temporale più diffuso è quello tra 12 e 36 mesi (43%), sebbene con una concentrazione minore rispetto alla fase distributiva.

	Frequenza	Dimensione	Range (mesi)
<b>Retail</b>	Weekly	Multivariate	12 < mesi <= 36
(23 articoli)	33%	78%	43%

Tabella 25: Correlazione tra la fase "Retail" e dati di supporto all'analisi effettuata nell'articolo

In conclusione, entrambe le fasi si affidano a dataset multivariati e a un orizzonte temporale di medio termine (1-3 anni), ma la distinzione principale risiede nella granularità temporale: la distribuzione richiede una precisione quotidiana per gestire il movimento fisico delle merci, mentre il retail si focalizza su una visione settimanale, più allineata alle dinamiche di acquisto del consumatore finale.

## Capitolo 6 – Conclusioni

### 6.1 Risultati ottenuti

L'analisi sistematica delle informazioni raccolte permette di delineare un quadro articolato e approfondito sull'utilizzo degli algoritmi nel settore alimentare secondario. I risultati emersi evidenziano come l'orientamento tecnologico della filiera sia guidato da necessità specifiche che variano drasticamente a seconda del posizionamento all'interno della *value chain* e della natura dei processi gestionali richiesti.

Il primo elemento di rilievo emerso dall'indagine riguarda le finalità applicative degli algoritmi, evidenziando una prevalenza assoluta della previsione della domanda come obiettivo principale degli studi analizzati. Tuttavia, la rilevanza di questo obiettivo non è uniforme lungo la catena del valore, mostrando una netta divergenza tra le fasi di distribuzione e vendita al dettaglio.

Nella fase di *retail*, l'attenzione della letteratura è quasi totalmente polarizzata sulla capacità di anticipare le richieste del mercato, coinvolgendo ben il 96% degli articoli esaminati, il che corrisponde a 22 articoli su un totale di 23. In questo specifico segmento, i dati mostrano l'assenza totale di studi dedicati esclusivamente alla gestione delle scorte, mentre solo il 4% degli articoli adotta un approccio che integra entrambi gli obiettivi.

<b>RETAIL</b>	Gestione delle scorte	Previsione della domanda	Entrambi
Utilizzo	0	22	1
Utilizzo %	0%	96%	4%

Tabella 26: Correlazione tra tipologia di analisi svolta per la value phase "retail"

Al contrario, nella fase di *distribution*, la previsione della domanda riguarda il 48% degli studi, ovvero 13 articoli su un campione di 27, indicando una ripartizione molto più equilibrata delle finalità operative. In tale ambito, la maggiore complessità gestionale emerge chiaramente dal 30% di contributi focalizzati unicamente sulla gestione delle scorte (8 articoli) e dal 22% che combina simultaneamente previsione e gestione (6 articoli).

DISTRIBUTION	Gestione delle	Previsione della	Entrambi
	scorte	domanda	
Utilizzo	8	13	6
Utilizzo %	30%	48%	22%

Tabella 27: Correlazione tra tipologia di analisi svolta per la value phase “distribution”

Questa discrepanza suggerisce che, mentre il commercio al dettaglio è quasi interamente assorbito dalla gestione della variabilità dei consumi finali, la distribuzione deve affrontare sfide logistiche più articolate che richiedono di bilanciare la previsione dei fabbisogni con l'ottimizzazione fisica dei magazzini.

I dati raccolti permettono di delineare un quadro piuttosto netto circa gli orientamenti della letteratura scientifica attuale per ogni tipologia di prodotto esaminata. Emerge, infatti, una preminenza della previsione della domanda per tutti i cluster descritti.

Tuttavia, nell'ambito del cluster relativo alla *supply chain* e alla logistica si riscontra l'eccezione più significativa e interessante a questa tendenza generale. In questo settore, pur mantenendosi la previsione della domanda come l'argomento principale con il 44% degli studi, si osserva una distribuzione decisamente più equilibrata rispetto alle altre categorie. La gestione delle scorte, infatti, non viene relegata a un ruolo marginale, ma mantiene una rilevanza sostanziale interessando il 38% degli articoli.

Ancor più indicativo è il dato relativo all'approccio integrato: il 19% della letteratura in ambito logistico sceglie di non scindere i due aspetti, trattando entrambi i temi contemporaneamente. Questa evidenza suggerisce che, nel contesto operativo della catena di approvvigionamento, la pura previsione non è considerata sufficiente se non accompagnata da una solida strategia di controllo fisico e gestione del magazzino.

L'analisi delle correlazioni tra gli algoritmi di *machine learning* e *deep learning* e la tipologia di prodotto esaminata nelle diverse fasi della *food value chain* evidenzia come non esista una relazione univoca e generalizzabile tra specifici modelli e determinati contesti applicativi. Dall'analisi emerge che, per tutti e sette gli algoritmi maggiormente utilizzati - *Linear Model*, *Decision Regression Tree*, *Random Forest*, *Extreme Gradient Boosting Regressor*, *Gradient Boosting Regressor*, *Neural Network*, *Long Short-Term Memory* - l'ambito di applicazione prevalente è rappresentato dalla fase *distribution* della *food value chain*.

Un risultato particolarmente significativo riguarda l'algoritmo *Gradient Boosting Regressor*, che appare fortemente associato alla fase della *distribution*. In particolare, su 8 casi analizzati, 7 risultano collocati in questo ambito, evidenziando un trend rilevante e suggerendo una specifica propensione di tale algoritmo all'analisi di problematiche legate alla *supply chain* e alla logistica. Questo dato rafforza l'ipotesi che il GBR sia particolarmente adatto a contesti caratterizzati da elevata complessità operativa e da una forte interdipendenza tra le variabili analizzate.

Fanno eccezione le reti neurali classiche *Neural Network* e l'*Extreme Gradient Boosting Regressor*, per le quali si osserva una lieve prevalenza della fase di *retail*, sebbene la differenza rispetto alla *distribution* risulti contenuta.

Con riferimento alla fase della *distribution*, in tutti i casi la maggior parte degli articoli associati ai diversi algoritmi analizzati si concentra sul cluster relativo alla *supply chain* e alla logistica, evidenziando come tali modelli vengano impiegati prevalentemente per supportare attività di pianificazione, coordinamento dei flussi e ottimizzazione operativa. Questa tendenza risulta particolarmente marcata per le reti neurali, impiegate nel 100% dei casi in ambito logistico, e per l'extreme gradient boosting, che raggiunge il 71% di focalizzazione in questo settore.

Food value phase	Algoritmo	Cluster di prodotto principale
<b>DISTRIBUTION</b>	Linear model	Supply chain e logistica 75%
	Decision regression tree	Supply chain e logistica 67%
	Random forest	Supply chain e logistica 58%
	Gradient boosting regressor	Supply chain e logistica 57%
	Extreme gradient boosting regressor	Supply chain e logistica 71%
	Neural network	Supply chain e logistica 100%
	Long short-term memory	Supply chain e logistica 58%

Tabella 28: Relazione tra la fase 'distribution' e il cluster di prodotto principale analizzato dai singoli algoritmi

Al contrario, per la fase di *retail*, i cluster di prodotto esaminati risultano più eterogenei, pur mostrando una predominanza degli studi focalizzati sul consumatore finale. In questo contesto, il cluster "Ristorazione/catering" emerge come il fulcro principale per la maggior parte dei modelli. Tuttavia, il settore retail mostra anche nicchie di specializzazione tecnica molto precise, come nel caso del "Meal delivery", che costituisce l'unico ambito di applicazione (100%) per il Gradient Boosting Regressor. In definitiva, i dati mostrano che la scelta di un algoritmo non dipende solo dalle sue prestazioni matematiche, ma dalla sua capacità di risolvere le sfide operative specifiche di ogni segmento della filiera.

Food value phase	Algoritmo	Cluster di prodotto principale
<b>RETAIL</b>	Linear model	Ristorazione/catering 40%
	Decision regression tree	Ristorazione/catering, 25% Bevande, 25% Vendita al dettaglio, 25% Meal delivery, 25%
	Random forest	Ristorazione/catering 33%
	Gradient boosting regressor	Meal delivery 100%
	Extreme gradient boosting regressor	Ristorazione/catering 63%
	Neural network	Ristorazione/catering 60%
	Long short-term memory	Ristorazione/catering 50%

Tabella 29: Relazione tra la fase 'retail' e il cluster di prodotto principale analizzato dai singoli algoritmi

Analizzando gli algoritmi ritenuti più performanti e le caratteristiche dei dati impiegati nelle diverse fasi della *food value chain* al fine di individuare eventuali pattern tra le variabili considerate, emerge che due dei tre algoritmi con le migliori performance, *Random Forest* e *Long Short-Term Memory*, risultano essere prevalentemente applicati alla fase di *distribution*. Al contrario, l'*Extreme Gradient Boosting Regressor* viene utilizzato principalmente nell'ambito del *retail*, con 3 applicazioni su 5.

Algoritmo	Distribution	Retail
Random Forest	67%	33%
LSTM	67%	33%
XGBR	40%	60%

Tabella 30: Correlazione tra l'efficacia degli algoritmi predittivi e il posizionamento operativo nelle fasi di Distribution e Retail

Nella distribuzione, l'esigenza di gestire flussi logistici complessi e monitoraggi quotidiani premia la robustezza di *Random Forest* nel trattare dati multivariati e la capacità delle reti LSTM di interpretare lunghe sequenze temporali senza necessità di conoscenze pregresse. Nel retail, settore focalizzato quasi esclusivamente sull'anticipazione dei consumi finali, XGBoost emerge come lo strumento più mirato grazie alla sua superiore efficacia nel gestire variabili esterne eterogenee e cicli di riassortimento settimanali. Questa distinzione tecnica permette di ottimizzare le performance predittive allineando la potenza di calcolo agli obiettivi gestionali specifici di ogni fase della catena del valore.

Una tendenza analoga a quella riscontrata con gli algoritmi più utilizzati, si osserva analizzando quelli più performanti: nella fase di *distribution*, le percentuali più elevate di utilizzo si concentrano nel cluster *supply chain* e logistica. Nella fase di *retail*, invece, assumono particolare rilevanza i cluster relativi ai prodotti freschi, alla ristorazione/catering e al *meal delivery*.

Infine, l'analisi condotta sulle fonti, evidenzia una chiara correlazione tra la specifica fase della *value chain* e le caratteristiche dei dati utilizzati a supporto delle analisi, mostrando in primo luogo come la complessità multivariata sia l'approccio predominante in entrambi i settori considerati. Nello specifico, la dimensione multivariata viene adottata nel 70% degli studi sulla distribuzione e nel 78% di quelli relativi al retail, suggerendo che un'analisi efficace della supply chain richieda necessariamente l'integrazione di molteplici variabili simultanee anziché l'osservazione di un singolo fattore.

Per quanto riguarda la granularità temporale, emerge una distinzione significativa: la fase di distribuzione predilige una frequenza giornaliera (37%), probabilmente per rispondere a necessità logistiche di breve termine, mentre nel retail prevale una frequenza settimanale (33%), che meglio si adatta ai cicli di riassortimento e ai comportamenti d'acquisto. Nonostante questa differenza nella frequenza, si riscontra una convergenza riguardo

all'orizzonte temporale dei dati, poiché in entrambe le fasi il range più comune si attesta tra i 12 e i 36 mesi (53% per la distribuzione e 43% per il *retail*).

	Frequenza	Dimensione	Range (mesi)
<b>Distribution</b>	Daily	Multivariate	12 < mesi <= 36
<b>Retail</b>	Weekly	Multivariate	12 < mesi <= 36

Tabella 31: Sintesi dei requisiti ottimali di dimensione, range e frequenza dei dati

In conclusione, per una tesi focalizzata su questi ambiti, i dati indicano che un modello analitico robusto deve essere strutturato su una base multivariata, attingendo a uno storico consolidato di almeno uno o tre anni e calibrando la frequenza di aggiornamento dei dati in base al focus operativo, prediligendo il monitoraggio quotidiano per la logistica distributiva e quello settimanale per le dinamiche di vendita al dettaglio.

## 6.2 Prospettive future e possibili implementazioni

Le prospettive future di questa ricerca si orientano verso una fase di validazione pratica e applicativa dei modelli teorici analizzati, con l'obiettivo di trasformare le evidenze statistiche in strumenti di supporto decisionale per il contesto industriale. Un primo ambito di sviluppo riguarda l'ottimizzazione delle scorte per le linee di produzione, dove l'integrazione di modelli predittivi avanzati può ridurre drasticamente le inefficienze logistiche e i costi di stoccaggio. Parallelamente, una sfida cruciale è rappresentata dalla previsione della domanda per i prodotti stagionali, che richiede un affinamento degli algoritmi per gestire picchi di consumo concentrati in periodi limitati dell'anno, sfruttando la capacità dei modelli di apprendere da serie storiche multivariate.

Il percorso di implementazione proposto si articola in fasi sequenziali che garantiscono la robustezza dei risultati. La fase iniziale prevede la raccolta e l'analisi di dataset reali, mirata all'identificazione di un caso di studio specifico e dei fattori critici che influenzano la previsione, come i KPI aziendali definiti in collaborazione con il management. Successivamente, si passerà allo sviluppo tecnico vero e proprio, caratterizzato da un'intensa attività di feature engineering per adattare i dati alle specificità del settore. In questa fase, l'implementazione ciclica di algoritmi come il *Gradient Boosting Regressor*, l'*XGBoost* e le reti

neurali permetterà di confrontare le performance in un ambiente operativo reale, testando la loro capacità di gestire la complessità dei flussi distributivi.

Un elemento distintivo delle implementazioni future è la validazione industriale, che non si limiterà alla mera precisione statistica. Sarà fondamentale procedere con una calibrazione dei modelli insieme agli esperti di settore, i quali apporteranno la loro conoscenza del mercato per affinare le previsioni algoritmiche. L'analisi finale dell'impatto operativo permetterà di misurare i benefici tangibili dell'adozione di queste tecnologie, valutando come una previsione più accurata possa tradursi in una gestione più sostenibile e reattiva dell'intera supply chain. Questo approccio integrato promette di colmare il divario tra la ricerca accademica e le necessità quotidiane della distribuzione e del *retail*.

## Riconoscimenti

Questa tesi è realizzata nell'ambito del progetto NODES, finanziato dal MUR sui fondi M4C2 - Investimento 1.5 Avviso "Ecosistemi dell'Innovazione", nell'ambito del PNRR finanziato dall'Unione europea - NextGenerationEU (Grant agreement Cod. n. ECS00000036).

Desidero ringraziare sinceramente il mio relatore, il Professor Galetto, e la correlatrice, la Professoressa Verna, per la guida costante e i preziosi consigli offerti durante questo percorso. Un ringraziamento speciale va all'Ingegnere Piovano per l'incessante supporto tecnico, la disponibilità e la pazienza dimostrate in ogni fase della ricerca. Grazie per la competenza e la fiducia che mi avete dedicato.

# Ringraziamenti

## Allegati

Per migliorare l'interpretabilità delle tabelle in allegato ad ogni articolo è stato associato un identificativo.

<b>Id</b>	<b>Titolo</b>
1	<i>Application of Machine Learning to support production planning of a food industry in the context of waste generation under uncertainty</i>
2	<i>Machine learning and optimization models for supplier selection and order allocation planning</i>
3	<i>Predictive analytics for demand forecasting: A deep learning-based decision support system</i>
4	<i>Time Series Forecasting and Modeling of Food Demand Supply Chain Based on Regressors Analysis</i>
5	<i>Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning</i>
6	<i>Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network</i>
7	<i>Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste</i>
8	<i>Machine Learning based Food Demand Estimation for Restaurants</i>
9	<i>Enhanced Demand Forecasting System for Food and Raw Materials Using Ensemble Learning</i>
10	<i>Restaurants store management based on demand forecasting</i>
11	<i>Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches</i>
12	<i>Supporting of manufacturer's demand plans as an element of logistics coordination in the distribution network</i>
13	<i>Machine-learning-based demand forecasting against food waste: Life cycle environmental impacts and benefits of a bakery case study</i>
14	<i>Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory</i>
15	<i>Prediction of Restaurant Sales during High Demand States Using Population Statistical Data</i>
16	<i>A Proposed Demand Forecasting Model by Using Machine Learning for Food Industry</i>
17	<i>A Stack-based Ensemble Model with Explainability for Food Demand Forecasting</i>

18	<i>Quantitative Models for Forecasting Demand for Perishable Products: A Systematic Review</i>
19	<i>A Production Model That Combines Lean And Industry 4.0 Principles To Enhance The Productivity Of Small And Medium-Sized Enterprises (SMEs) In Peru's Food Manufacturing Sector</i>
20	<i>A Strategic Analysis of Food Demand Using Machine Learning and Explainable AI</i>
21	<i>Demand Forecasting of Perishable Products Using Traditional Statistical and Machine Learning Methods</i>
22	<i>Effective Food Demand Forecasting Using Machine Learning Algorithms</i>
23	<i>Demand Forecasting for Daily Retail Orders in Fresh Food Market</i>
24	<i>Machine Learning-Driven Daily Demand Forecasting for Fresh Produce: A Case Study with Bananas</i>
25	<i>Restaurant Demand Forecasting Using Dynamic Neural Network for Business Networking Application</i>
26	<i>An Efficient Approach for Food Demand Forecasting Using an Ensemble Technique and Statistical Analysis</i>
27	<i>Data-Driven Forecasting and Inventory Optimization using Machine Learning Models and Methods</i>
28	<i>Transforming Inventory Management Through Machine Learning based Demand Forecasting for a Circular Economy</i>
29	<i>Inventory Optimization Using Data Science Technologies for Supply Chain 4.0</i>
30	<i>A survey of machine learning techniques for food sales prediction</i>
31	<i>Machine learning in predicting demand for fast-moving consumer goods: An exploratory research</i>
32	<i>A hybrid artificial intelligence sales-forecasting system in the convenience store industry</i>
33	<i>Prediction of Quality Food Sale in Mart Using the AI-Based TOR Method</i>
34	<i>Machine Learning based Food Sales Prediction using Random Forest Regression</i>
35	<i>A Comparison of Data Mining Approaches for Forecasting Sales of FMCG Food Products</i>
36	<i>Time series analysis for supply chain planning in restaurants</i>
37	<i>Supporting of manufacturer's demand plans as an element of logistics coordination in the distribution network</i>
38	<i>An Efficient Novel Approach on Machine Learning Paradigms for Food Delivery Company through Demand Forecasting in societal community</i>

39	<i>Employing constrained neural networks for forecasting new product's sales increase</i>
40	<i>Sales forecasting for fresh foods: A study in Indonesian FMCG</i>
41	<i>Demand forecasting procedure for short life-cycle products with an actual food processing enterprise</i>
42	<i>Introduction of AI-based sales forecasting: how to drive digital transformation in food and beverage outlets</i>
43	<i>Explainable artificial intelligence to improve the resilience of perishable product supply chains by leveraging customer characteristics</i>
44	<i>A STUDY ON DATA MINING TECHNIQUES FOR FORECASTING FMCG PRODUCT SALES</i>
45	<i>Food Sales Analysis and Prediction Using Machine Learning</i>
46	<i>Data-Driven Mechanisms for a Newsvendor Problem: A Case Study</i>
47	<i>Sales Forecasting During the COVID-19 Pandemic for Stock Management</i>
48	<i>An Efficient Approach for Food Demand Forecasting Using an Ensemble Technique and Statistical Analysis</i>
49	<i>Large-scale Probabilistic Forecasting of Consumer Engagement of CPG Products using Heterogeneous Web Data</i>
50	<i>Bayesian-Optimized Ensemble Deep Learning Models for Demand Forecasting in the Volatile Situations: A Case Study of grocery Demand during Covid-19 Outbreaks</i>
51	<i>Predictive Analytics: A Machine Learning Approach for Insights in Food Production and Sales</i>
52	<i>Forecasting Necessary Levels Of Bread Stock For "Direct-To-Shelf" Delivery Using CNN And LSTM Artificial Neural Networks</i>

## Allegato A – Scheda informativa degli articoli analizzati

Id	Anno	Ambito	Value chain phase	Cluster tipologia prodotto	Algoritmi	Metriche
1	2020	Gestione delle scorte	Distribution	Bevande	Nine different regression models have been applied in this study: linear model with stepwise selection, regression tree, bagged tree, random forest, gradient boosting, ridge regression, lasso regression, elastic net, and spline regression.	RMSE, R <sup>2</sup> , MAE
2	2021	Entrambi	Distribution	Prodotti trasformati e confezionati	HTL, ARIMA, RRC // RRC, SVR, PR	RMSE, MAPE, average total error
3	2022	Previsione della domanda	Retail	Prodotti trasformati e confezionati	ARIMA, ARIMAX, RF, NN, LSTM, ARIMA+NN, ARIMA+RF, LSTM+RF	ME, MAE, RMSE
4	2023	Previsione della domanda	Distribution	Meal delivery	5 of ML: random forest regressor, GBR, XGB, LightGBM and CatBoost and 2 of DP: LSTM and Bi-LSTM	RMSLE, RMSE, MAPE, MAE
5	2022	Previsione della domanda	Retail	Prodotti freschi	Holt-Winters, RF, SVR, Feedforward ANN, LSTM	RMSE, MAE, MPE, MNE
6	2021	Previsione della domanda	Distribution	Prodotti trasformati e confezionati	nonlinear autoregressive exogenous neural network (NARXNN)	MAE, MAPE, RMSE, R <sup>2</sup>

7	2024	Previsione della domanda	Retail	Ristorazione/catering	The present paper proposes four food demand forecasting models that can be divided into two categories. The first category comprises two casual models, one based on a random forest (RF) algorithm and the other based on the light gradient-boosting machine (LightGBM) algorithm. The second category comprises two time series models, one based on a long short-term memory (LSTM) neural network and the other based on a transformer neural network.	RMSE, MAPE, R <sup>2</sup>
8	2023	Previsione della domanda	Retail	Meal delivery	Random Forest Regressor, Gradient Boosting Tree Regressor are used for the prediction and optimization	/
9	2021	Previsione della domanda	Retail	Ristorazione/catering	RF, XGBoost, ANN, SVM, Stacking (RF, XGBoost, ANN, SVM)	MAE
10	2020	Entrambi	Retail	Ristorazione/catering	RF regression	/
11	2023	Previsione della domanda	Distribution	Prodotti trasformati e confezionati	SVM, LSTM	MAPE, RMSE
12	2023	Gestione delle scorte	Distribution	Supply chain e logistica	The author proposes to examine two packages of forecasting methods: a traditional one, which includes the use of exponential time series smoothing methods and ARIMA methods, and a contemporary	MAPE

					one, which includes the use of methods based on artificial neural networks and machine learning	
13	2024	Previsione della domanda	Retail	Vendita al dettaglio	neural network (RNN)	/
14	2023	Previsione della domanda	Retail	Ristorazione/catering	"The goal of this study is to develop prediction models for food demand based on MLAs, such as Artificial Neural Network (ANN), Gaussian Process Regression (GPR), Support Vector Regression (SVR), Regression Tree, and	MAE, MSE, R
15	2021	Previsione della domanda	Retail	Ristorazione/catering	Ensemble Decision Tree (EDT). "	MAE
16	2022	Entrambi	Distribution	Prodotti trasformati e confezionati	we used LightGBM	MSE, MAE
17	2022	Previsione della domanda	Retail	Ristorazione/catering	The model will use the Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) algorithms because they give the most accurate results	RMSLE, MAE
18	2022	Previsione della domanda	Review	/	Linear Regression, Decision Tree Regressor, XGBoost, Stacking Ensemble, LIME (per interpretabilità)	RMSE, MAE, MSE and MaxAE
19	2023	Gestione delle scorte	Distribution	Prodotti trasformati e confezionati	/	upper case R squared, MAE, MAPE, MSE, RMSE

20	2024	Previsione della domanda	Distribution	Supply chain e logistica	The ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal ARIMA)	RMSE, MAE and R-squared
21	2025	Previsione della domanda	Retail	Bevande	Linear Regression, Ridge Regression, Lasso Regression, Bayesian Ridge, Random forest regressor, XGBRegressor, Gradient Boosting Regressor	MAE, sMAPE, RelMAE
22	2023	Previsione della domanda	Retail	Meal delivery	Decision Tree Regression (DT) and Random Forests (RF), Simple Exponential Smoothing (SES), Autoregressive Integrated Moving Average (ARIMA)	RMSE, R2, MAE, MAPE
23	2023	Previsione della domanda	Retail	Prodotti freschi	This project tries to implement a random forest algorithm	NMAE, MSE and MAE
24	2024	Previsione della domanda	Distribution	Prodotti freschi	In this paper, we introduce multiple prominent methodologies in the time-series analysis: Long Short-Term Memory (LSTM), Transformers, Prophet, Feedforward Neural Networks (FNN), and eXtreme Gradient Boosting (XGBoost).	accuracy, recall, precision, and F1-score
25	2024	Previsione della domanda	Retail	Ristorazione/catering	For each experiment, 6 ML models including logistic regression (LR), RF, SVM, K-Nearest Neighbor (KNN), decision tree (DT), and gradient boosting (GB), are trained and tested on the three dataset variations, incorporating the most impacting variables only as indicators	MSE

26	2025	Previsione della domanda	Retail	Ristorazione/catering	this paper integrates forecasting restaurant demand with NARX ANN.	R <sup>2</sup> , MAE, RMSE
27	2024	Gestione delle scorte	Distribution	Supply chain e logistica	XGBoost, CatBoost	MAE RMSE, ME, variance, accuracy, and bias
28	2024	Entrambi	Distribution	Supply chain e logistica	LSTM (Long Short-Term Memory), Random Forest, ARIMA, Reti Neurali, Regressioni Lineari, LSTM + RF	RMSE, R-squared and MSE
29	2024	Gestione delle scorte	Distribution	Supply chain e logistica	XGBoost, LightGBM, Random Forest, and Long short term memory(LSTM)	MSE, MAE, RMSE
30	2019	Previsione della domanda	Review	/	After using regression machine learning techniques (linear regression, decision tree, and random forest) for OUT Inventory and IN inventory,	MSE, RMSE, RRMSE, MAE, MAPE, MASE
31	2019	Previsione della domanda	Distribution	Supply chain e logistica	/	/
32	2012	Previsione della domanda	Distribution	/	/	RETRACTED
33	2022	Previsione della domanda	Retail	Vendita al dettaglio	RETRACTED	mean squared error and variance
34	2022	Previsione della domanda	Distribution	Supply chain e logistica	The proposed TOR AI approach is compared with several machine-learning approaches, such as linear regression	accuracy score, mean absolute error and max error.

35	2023	Previsione della domanda	Distribution	Supply chain e logistica	Gradient Boosting Regression, and Random Forest Regression	The study compares the accuracy and efficiency of these algorithms using various performance metrics.
36	2020	Previsione della domanda	Retail	Ristorazione/catering	We explored various techniques such as Linear Regression, K- means clustering, and Tracking patterns and evaluated their effectiveness in predicting future sales of FMCG food products	RMSE
37	2023	Gestione delle scorte	Distribution	Supply chain e logistica	To make sales forecasts, the time series data is used to build machine learning models using the Holt Winter's and STL algorithms.	MAPE
38	2023	Previsione della domanda	Retail	Meal delivery	The author proposes to examine two packages of forecasting methods: a traditional one, which includes the use of exponential time series smoothing methods and ARIMA methods, and a contemporary one, which includes the use of methods based on artificial neural networks and machine learning	accuracy, RMSE
39	2019	Previsione della domanda	Retail	Ristorazione/catering	Xgboost Regression, Decision Tree, Linear Regression	F1-score and accuracy
40	2020	Previsione della domanda	Retail	Prodotti freschi	In the sequel, we compare the classification accuracy of WCNN algorithm against state-of-the-art ANN training algorithms. Next, we evaluate the performance of the weight-	MAE, MSE, MAPE, MAD

					constrained neural networks trained with WCNN against state-of-the-art machine learning algorithms. The Naive Bayes algorithm [3] was the representative of the probabilistic classifiers while kNN algorithm [1] was selected as instance-based learner with Euclidean distance as distance metric.	
41	2014	Previsione della domanda	Distribution	Supply chain e logistica	The algorithms used in this study are specific algorithms for time series data such as Moving Average, Multiple Regression, Holt-Winter Method (Multiplicative / Additive), ANN, and SVR	Minimun error, maximun error, mean error, average mean error, std. Deviation, linear correlation, frequency
42	2024	Previsione della domanda	Retail	Ristorazione/catering	This study adopts two methods to forecast demand trends of new short life-cycle products, called Neural Network and Decision Tree	RMSE, MAE, MAPE
43	2024	Previsione della domanda	Distribution	Supply chain e logistica	Therefore, we defined our prediction task as a regression problem using XGBoost as the learning algorithm.	RMSE
44	2024	Previsione della domanda	Distribution	Supply chain e logistica	Autoregression (AR) model, Vector autoregression (VAR), Support vector regression (SVR), Univariate LSTM	accuracy
45	2024	Entrambi	Distribution	Supply chain e logistica	linear regression, decision trees, Pattern Tracking	Accuracy
46	2024	Entrambi	Distribution	Prodotti freschi	Based on accuracy measures, the XGBoost and SVM with Positive and Negative	RMSE, MAPE, MAE

					Binomial distributions were compared in terms of algorithm performance.	
47	2023	Gestione delle scorte	Distribution	Supply chain e logistica	We first present five traditional approaches followed by the firm, which are Naive, Seasonal Naive (S-Naive), Median, Seasonal Median (S-Median), and Moving Average (MA). we choose six procedures that have been heavily performed in the relevant literature: Linear Regression, Random Forest, XGBoost (eXtreme Gradient Boosting), LightGBM, LSTM (Long Short-Term Memory)	MAPE and RMSE
48	2025	Previsione della domanda	Retail	Ristorazione/catering	Here, we decided to use the XGBoost and LGBMRegressor model with the tree-based gradient boosting feature, which is the most popular of the machine learning algorithms, and the linear-based Ridge model.	R- squared, RMSE, MAE
49	2024	Previsione della domanda	Distribution	Prodotti trasformati e confezionati	The creation of an effective model for predicting food demand is at the center of the process. XGBoost and CatBoost,	NRMSE, RMSE, MSE, MASE,MAPE,sMAPE, MSIS
50	2025	Gestione delle scorte	Distribution	Supply chain e logistica	Encoder-Decoder LSTM (EDL), Neural Basis Expansion Analysis for Time Series (NBEATS), DeepAR (DAR), Multi-headed Transformer (MhT), Temporal Fusion Transformer (TFT)	RMSE, MAE
51	2025	Entrambi	Distribution	Supply chain e logistica	The proposed model considers a combination of Bayesian-optimized LSTM (BO-LSTM) and GRU (BO-GRU) into an ensemble with the purpose of reducing the	MSE and RMSE, Average Absolute Error MAE, R-squared, or R <sup>2</sup>

					forecasting error. The suggested model was statistically compared to 6 benchmark models namely RF, GBRT, FFNN, RNN, GRU, LSTM, in addition to its contributors. This work proposed the Blended-LSTM-GRU model which consists of three main parts including the BO-LSTM, BO-GRU, and the blender.	
52	2023	Previsione della domanda	Retail	Meal delivery	Decision Tree Regressor, Extra Trees Regressor Random Forest Regressor, XGBoost Regressor, Bagging Regressor, Gradient Boosting Regressor, CatBoost, LightGBM, Polynomial Regression, MLP Regressor, AdaBoost Regressor, SVR, Linear Regression, Ridge Regression, Bayesian Ridge, Lasso Regression	RMSE, MAE

## Allegato B – Caratterizzazione dei dati

<b>Id</b>	<b>Dimensione</b>	<b>Range</b>	<b>Mesi</b>	<b>Frequenza</b>
1	Multivariate	10	10	Daily
2	Multivariate	40 anni	480	Quarterly
3	Multivariate	3 anni	36	Weekly
4	Multivariate	145 settimane	36	Weekly
5	Multivariate	34 mesi	34	Daily
6	Univariate	24 mesi	24	Daily
7	Multivariate	dai 2 anni a 8 anni	da 24 a 96	Daily
8	Multivariate	145 settimane	36	Weekly
9	Multivariate	24 mesi	24	Daily
10	Multivariate	circa 48 mesi	48	Daily
11	Multivariate	36 mesi	36	Daily
12	Univariate	tra 18 e 90 mesi	da 18 a 90	/
13	Multivariate	1 mese	1	Daily
14	Multivariate	12 mesi	12	Daily
15	Multivariate	/	/	Hourly
16	Multivariate	/	/	Quarterly
17	Multivariate	/	/	Weekly
18	Review	/	/	/
19	/	/	/	/
20	Multivariate	145 settimane	36	Weekly
21	/	/	/	Weekly
22	Multivariate	/	/	/
23	Multivariate	7 anni	84	Weekly, Monthly, Yearly
24	Multivariate	/	/	Daily
25	Multivariate	2 mesi	2	Monthly
26	Multivariate	/	/	Weekly
27	Multivariate	/	/	Weekly

28	Multivariate	145 settimane	36	Weekly
29	Multivariate	5 anni	60	/
30	Review	/	/	/
31	/	/	/	/
32	/	/	/	/
33	/	/	/	Daily
34	Univariate	/	/	Daily
35	Multivariate	10 anni	120	/
36	Multivariate	145 settimane	36	Weekly
37	/	da 1,5 a 5 anni	da 18 a 90	Daily
38	Multivariate	/	/	Weekly
39	Multivariate	1 mese	1	Weekly
40	Multivariate	18 mesi	18	Daily, Weekly
41	Multivariate	2 anni e 3 mesi	27	/
42	Multivariate	5.75 anni	69	Hourly
43	Multivariate	3 anni	36	Daily
44	Multivariate	/	/	Weekly
45	Multivariate	/	/	/
46	Multivariate	167 settimane	42	Weekly
47	Multivariate	2 anni	24	Daily
48	Multivariate	/	/	Weekly
49	Multivariate	110 mesi	110	Monthly
50	Univariate	533 giorni	17	Daily
51	Multivariate	/	/	Daily
52	Multivariate	/	/	Daily

## Allegato C – Features analizzate negli articoli

Id	Dimensione	Characteristics of the production	Type of packaging	Characteristics of the product	Transaction details	Economy indicator	Demand related to product p in period q	Cost of inventory	Promotions	Weather	Store related	Customer features	Population features	Time series features
1	Multivariate	1	1	1										
2	Multivariate	1			1	1	1	1						
3	Multivariate				1	1	1		1	1	1			1
4	Multivariate				1		1		1		1			1
5	Multivariate						1		1	1				1
6	Univariate	/	/	/	/	/	/	/	/	/	/	/	/	/
7	Multivariate			1						1				1
8	Multivariate			1	1						1			
9	Multivariate			1						1				1
10	Multivariate									1				1
11	Multivariate				1						1			
12	Univariate	/	/	/	/	/	/	/	/	/	/	/	/	/
13	Multivariate			1										

14	Multivariate			1											1	
15	Multivariate														1	1
16	Multivariate													1		
17	Multivariate													1		1
18	Review															
19	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
20	Multivariate														1	1
21	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
22	Multivariate														1	
23	Multivariate														1	1
24	Multivariate															1
25	Multivariate														1	
26	Multivariate														1	1
27	Multivariate														1	1
28	Multivariate														1	1
29	Multivariate														1	1

30	Review														
31	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
32	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
33	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
34	Univariate	/	/	/	/	/	/	/	/	/	/	/	/	/	/
35	Multivariate					1									
36	Multivariate														1
37	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
38	Multivariate			1	1							1			1
39	Multivariate								1			1			
40	Multivariate														1
41	Multivariate				1										
42	Multivariate								1	1	1				
43	Multivariate													1	
44	Multivariate														
45	Multivariate														1



## Bibliografia

- [1] Nouran Nassibi, Heba Fasihuddin: “Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches” *International Journal of Advanced Computer Science and Applications*, Vol. 14, No. 3, 2023
- [2] Himanshu Sharma: “Data-Driven Forecasting and Inventory Optimization using Machine Learning Models and Methods” 2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET) | 979-8-3503-6772-0/24/\$31.00 ©2024 IEEE | DOI: 10.1109/ACET61898.2024.10730235
- [3] Mugdha, Dr. Durgansh Sharma: “Transforming Inventory Management Through Machine Learning based Demand Forecasting for a Circular Economy” 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS) | 979-8-3315-2820-1/24/\$31.00 ©2024 IEEE | DOI: 10.1109/CCIS63231.2024.10932031
- [4] Jonathan Vinicius Kaizer, Rodrigo Clemente Thom de Souza e Linnyer Beatrys Ruiz Aylon: “Quantitative Models for Forecasting Demand for Perishable Products: A Systematic Review” The Author(s), under exclusive license to Springer Nature Switzerland AG 2022 A. C. Bicharra Garcia et al. (Eds.): IBERAMIA 2022, LNAI 13788, pp. 393–404, 2022. DOI: 10.1007/978-3-031-22419-5\_33
- [5] E. Anitha: “Effective Food Demand Forecasting Using Machine Learning Algorithms” 2023 IEEE Engineering Informatics | 979-8-3503-3852-2/23/\$31.00 ©2023 IEEE | DOI: 10.1109/IEEECONF58110.2023.10520567
- [6] Grigorios Tsoumakas: “A survey of machine learning techniques for food sales prediction” *Artif Intell Rev* (2019) 52:441–447 DOI: 10.1109/IACIS65746.2025.11211065
- [7] Sandeep Kumar Panda: “Time Series Forecasting and Modeling of Food Demand Supply Chain Based on Regressors Analysis” Received 1 March 2023, accepted 30 March 2023, date of publication 11 April 2023, date of current version 4 May 2023 DOI: 10.1109/ACCESS.2023.3266275
- [8] Samiul Islam, Saman Hassanzadeh Amin e Leslie J. Wardley: “Machine learning and optimization models for supplier selection and order allocation planning” *Int. J. Production Economics* 242 (2021) 108315 DOI: 10.1016/j.ijpe.2021.108315
- [9] S. Hosseini: “Review of quantitative methods for supply chain resilience analysis” March 2019 DOI: 10.1016/j.tre.2019.03.001

- [10] Materiale del corso di Laurea Magistrale di Ingegneria Gestione “Business Intelligence per big data” della Professoressa Eliana Pastor
- [11] Matthew J Page: “PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews” Published 29 March 2021 DOI: 10.1136/bmj.n160
- [12] <https://www.prisma-statement.org/prisma-2020-flow-diagram>
- [13] Panpatte S.; Ganeshkumar C: “Artificial Intelligence in Agriculture Sector: Case Study of Blue River Technology” DOI: 10.1007/978-981-15-9689-6\_17
- [14] Gond S.; Krishnan N.; Kumar R.K: “FORECASTING THERMAL ENERGY DEMANDS FOR VARIOUS PROCESS INDUSTRIES USING MACHINE LEARNING TECHNIQUES” DOI: 10.1615/InterJEnerCleanEnv.2023047190
- [15] Venkataramani S.; Ramesh A.; Sundar S.; Jain A.K.; Gudur G.K.; Vijayaraghavan V: “A dynamically adaptive movie occupancy forecasting system with feature optimization” DOI: 10.1109/ICDMW.2019.00118
- [16] Garre A.; Ruiz M.C.; Hontoria E.: "Application of Machine Learning to Support Production Planning of a Food Industry in the Context of Waste Generation under Uncertainty" DOI: 10.1016/j.orp.2020.100147
- [17] Amir Shani, Yaniv Belhassen: "Teaching professional ethics in culinary studies" International Journal of Contemporary Hospitality Management Vol. 25 No. 3, 2013 DOI: 10.1108/09596111311311062
- [18] Miguéis V.L.; Pereira A.; Pereira J.; Figueira G: “Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning” DOI: 10.1016/j.jclepro.2022.131852
- [19] Bozdogan U.; Alptekin G.I: “Demand Forecasting for Daily Retail Orders in Fresh Food Market” 2023 4th International Informatics and Software Engineering Conference (IISEC) | 979-8-3503-1803-6/23/\$31.00 ©2023 IEEE | DOI: 10.1109/IISEC59749.2023.10391047
- [20] Seyam A.; Barachi M.E.I.; Mathew S.S.; Du B.; Shen J: “Machine Learning-Driven Daily Demand Forecasting for Fresh Produce: A Case Study with Bananas” 2024 Twelfth International Conference on Advanced Cloud and Big Data (CBD) | 979-8-3315-1107-4/24/\$31.00 ©2024 IEEE | DOI: 10.1109/CBD65573.2024.00071
- [21] Zhao M.A.; Setyawan B: “Sales forecasting for fresh foods: A study in Indonesian FMCG” 2020 International Conference on Information Science and Communications Technologies

(ICISCT) | 978-1-7281-9969-6/20/\$31.00 ©2020 IEEE | DOI:  
10.1109/ICISCT50599.2020.9351484

[22] Sancaktaroglu A.; Gokgur B.; Kocabiyikoglu A: “Data-Driven Mechanisms for a Newsvendor Problem: A Case Study” DOI: 10.35378/gujs.1334184

[23] Punia S.; Shankar S: “Predictive analytics for demand forecasting: A deep learning-based decision support system” DOI: 10.1016/j.knosys.2022.109956

[24] Lutoslawski K.; Hernes M.; Radomska J.; Hajdas M.; Walaszczyk E.; Kozina A: “Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network” DOI: 10.1109/ACCESS.2021.3123255

[25] Nassibi N.; Fasihuddin H.; Hsairi L: “A Proposed Demand Forecasting Model by Using Machine Learning for Food Industry” DOI: 10.1145/3584202.3584307

[26] Komori-Zevallos A.R.; Montedoro-Garay F.M.; Garcia-Lopez Y.J.; Flores J.C.Q: “A Production Model That Combines Lean And Industry 4.0 Principles To Enhance The Productivity Of Small And Medium-Sized Enterprises (SMEs) In Peru’s Food Manufacturing Sector” DOI: <https://doi.org/10.15488/1525>

[27] Abdullah Al I.; Perera K.S: “Large-scale Probabilistic Forecasting of Consumer Engagement of CPG Products using Heterogeneous Web Data” DOI: 10.1016/j.procs.2024.05.124

[28] Vi T.D.; Nhan H.T: “Demand Forecasting of Perishable Products Using Traditional Statistical and Machine Learning Methods” DOI: 10.1007/978-981-97-7839-3\_33

[29] Rodrigues M.; Miguéis V.; Freitas S.; Machado T: “Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste” DOI: 10.1016/j.jclepro.2023.140265

[30] Harshini K.; Madhira P.K.; Chaitra S.; Reddy G.P: “Enhanced Demand Forecasting System for Food and Raw Materials Using Ensemble Learning” DOI: 10.1109/AIMV53313.2021.9671005

[31] Tanizaki T.; Hoshino T.; Shimmura T.; Takenaka T: “Restaurants store management based on demand forecasting” DOI: 10.1016/j.procir.2020.05.101

[32] Aci M.; Yergök D: “Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory” DOI: 10.17559/TV-20230117000232

- [33] Shinoda K.; Yamada M.; Takanashi M.; Tsuboi T.; Hasegawa D.; Fukazawa Y.; Kimoto M: “Prediction of Restaurant Sales during High Demand States Using Population Statistical Data” DOI: 10.23919/ICMU50196.2021.9638955
- [34] Chatterjee S: “A Stack-based Ensemble Model with Explainability for Food Demand Forecasting” DOI: 10.1109/IATMSI56455.2022.10119320
- [35] Khalifeh A.; Abubaker F: “Restaurant Demand Forecasting Using Dynamic Neural Network for Business Networking Application” DOI: 10.1109/SSD61670.2024.10549333
- [36] kumar D.A.; Rao B.T.; Rangaswamy B.; Meghana K: “An Efficient Approach for Food Demand Forecasting Using an Ensemble Technique and Statistical Analysis” DOI: 10.1007/978-3-031-77075-3\_6
- [37] Dilkhush Mihirsan D.; Joseph J.T.; Renisha B: “Time series analysis for supply chain planning in restaurants” DOI: 10.1109/ICCCS49678.2020.9276872
- [38] Livieris I.E.; Kiriakidou N.; Kanavos A.; Vonitsanos G.; Tampakas V: “Employing constrained neural networks for forecasting new product's sales increase” DOI: 10.1007/978-3-030-19909-8\_14
- [39] Groene N.; Zakharov S: “Introduction of AI-based sales forecasting: how to drive digital transformation in food and beverage outlets” DOI: 10.1007/s44163-023-00097-x
- [40] kumar D.A.; Rao B.T.; Rangaswamy B.; Meghana K: “An Efficient Approach for Food Demand Forecasting Using an Ensemble Technique and Statistical Analysis” DOI: 10.1007/978-3-031-77075-3\_6
- [41] Pandey N.K.; Mishra A.K.; Kumar V.; Kumar A.; Diwakar M.; Tripathi N: “Machine Learning based Food Demand Estimation for Restaurants” DOI: 10.1109/ISCON57294.2023.10112059
- [42] Yerragudipadu S.; Gurram V.R.; Rayapudi N.S.; Bingi B.; Gollapalli L.; Peddapatlolla U: “An Efficient Novel Approach on Machine Learning Paradigms for Food Delivery Company through Demand Forecasting in societal community” DOI: 10.1051/e3sconf/202339101089
- [43] Grogan L.; Reus V.; O'Carroll B.; Fallon S: “Forecasting Necessary Levels Of Bread Stock For "Direct-To-Shelf" Delivery Using CNN And LSTM Artificial Neural Networks” DOI: 10.1109/IEIT59852.2023.10335581
- [44] Hübner N.; Caspers J.; Coroamă V.C.; Finkbeiner M: “ Machine-learning-based demand forecasting against food waste: Life cycle environmental impacts and benefits of a bakery case study” DOI: 10.1111/jieec.13528

- [45] Irfan D.; Tang X.; Narayan V.; Mall P.K.; Srivastava S.; Saravanan V: “Prediction of Quality Food Sale in Mart Using the AI-Based TOR Method” DOI: 10.1155/2022/6877520
- [46] Kmiecik M: “Supporting of manufacturer's demand plans as an element of logistics coordination in the distribution network” DOI: 10.30657/pea.2023.29.9
- [47] Adith M.; Katti P.; Gupta D: “A Strategic Analysis of Food Demand Using Machine Learning and Explainable AI” DOI: 10.1109/I2CT61223.2024.10543689
- [48] Bouazizi E.; Khedr A.; Elfaïoumy S.; Belal M: “Inventory Optimization Using Data Science Technologies for Supply Chain 4.0” DOI: 10.34028/iajit/21/6/4
- [49] Tarallo E.; Akabane G.K.; Shimabukuro C.I.; Mello J.; Amancio D: “Machine learning in predicting demand for fast-moving consumer goods: An exploratory research” DOI: 10.1016/j.ifacol.2019.11.203
- [50] Naik H.; Yashwanth K.; Suraj P.; Jayapandian N: “Machine Learning based Food Sales Prediction using Random Forest Regression” DOI: 10.1109/ICECA55336.2022.10009277
- [51] Doshi M: “A Comparison of Data Mining Approaches for Forecasting Sales of FMCG Food Products” DOI: 10.1109/ICCCNT56998.2023.10307432
- [52] Kmiecik M: “Supporting of manufacturer's demand plans as an element of logistics coordination in the distribution network” DOI: 10.30657/pea.2023.29.9
- [53] Gaku R: “Demand forecasting procedure for short life-cycle products with an actual food processing enterprise” DOI: 10.1080/18756891.2014.947121
- [54] Jauhar S.K.; Harinath S.; Krishnaswamy V.; Paul S.K: “Explainable artificial intelligence to improve the resilience of perishable product supply chains by leveraging customer characteristics” DOI: 10.1007/s10479-024-06348-z
- [55] Doshi M.; Hirapara J: “A STUDY ON DATA MINING TECHNIQUES FOR FORECASTING FMCG PRODUCT SALES” DOI: 10.1109/ICCCNT61001.2024.10726077
- [56] Sadasivam V.R.; Narendharakumar G.; Ragavan M.; Prabhuram A: “Food Sales Analysis and Prediction Using Machine Learning” DOI: 10.1109/AIMLA59606.2024.10531342
- [57] Yildirim E.; Cam V.; Balki F.; Sarp S: “Sales Forecasting During the COVID-19 Pandemic for Stock Management” DOI: 10.1007/978-3-031-37649-8\_12
- [58] Al-Theeb N.; Smadi H.; Al-Qaydeh N: “Bayesian-Optimized Ensemble Deep Learning Models for Demand Forecasting in the Volatile Situations: A Case Study of grocery Demand during Covid-19 Outbreaks” DOI: 10.3926/jiem.6571

[59] Sivakumar P.; Elango A.; Sai P.C.; Jyotsna C: “Predictive Analytics: A Machine Learning Approach for Insights in Food Production and Sales” DOI: 10.1109/COMP-SIF65618.2025.10969882

## Indice delle Figure

Figura 1: Formula regressione lineare semplice.....	12
Figura 2: Formula regressione multipla .....	13
Figura 3: Formula regressione polinomiale.....	14
Figura 4: Formula MAPE .....	18
Figura 5: Formula RMSE .....	18
Figura 6: Formula MAE .....	19
Figura 7: PRISMA 2020 flow diagram .....	23
Figura 8: Grafico a torta relativo al focus della ricerca.....	27
Figura 9: Food Value Chain Phase.....	28
Figura 10: Grafico a torta della distribuzione degli articoli nella Food Value Chain .....	30
Figura 11: I 7 cluster di casi d’uso identificati .....	31
Figura 12: Grafico a istogramma dei cluster più analizzati.....	33
Figura 13: Diagramma di Sankey degli algoritmi più utilizzati .....	35
Figura 14: Grafico a istogramma degli algoritmi più utilizzati .....	36
Figura 15: Diagramma di Sankey degli algoritmi più performanti.....	40
Figura 16: Grafico a istogramma degli algoritmi più performanti.....	40
Figura 17: Grafico a torta relativo alla dimensione dei dati .....	41
Figura 18: Grafico a istogramma relativo alla frequenza di rilevamento dei dati.....	42
Figura 19: Grafico ad anello relativo al range di dati utilizzato.....	43
Figura 20: Grafico ad anello aggiornato relativo al range di dati utilizzati .....	43
Figura 21: Grafico a torta sulla correlazione tra articoli sulla “Distribution” e obiettivi della ricerca .....	48
Figura 22: Grafico a torta sulla correlazione tra articoli sul “Retail” e obiettivi della ricerca ..	49

## Indice delle Tabelle

Tabella 1: Query utilizzata .....	21
Tabella 2: Ambito di applicazione degli articoli [unità].....	29
Tabella 3: Assegnazione di ogni articolo al cluster di appartenenza .....	32
Tabella 4: Suddivisione per cluster degli algoritmi .....	35
Tabella 5: Relazione tra l'algoritmo più performante, il cluster di appartenenza e l'articolo esaminato .....	39
Tabella 6: Metriche di valutazione più utilizzate .....	45
Tabella 7: Relazione tra articoli sui “Prodotti freschi” e analisi svolta .....	49
Tabella 8: Relazione tra articoli sui “Prodotti trasformati e confezionati” e analisi svolta .....	50
Tabella 9: Relazione tra articoli sulle “Bevande” e analisi svolta.....	50
Tabella 10: Relazione tra articoli sulla “Ristorazione/Catering” e analisi svolta .....	50
Tabella 11: Relazione tra articoli sulla “Vendita al dettaglio” e analisi svolta .....	51
Tabella 12: Relazione tra articoli sulla “Supply chain e logistica” e analisi svolta .....	51
Tabella 13: Relazione tra articoli sul “Meal delivery” e analisi svolta .....	51
Tabella 14: Correlazione tra Linear Model e tipologia di prodotto esaminato nella fase della value chain.....	53
Tabella 15: Correlazione tra Decision Regression Tree e tipologia di prodotto esaminato nella fase della value chain .....	53
Tabella 16: Correlazione tra Random Forest e tipologia di prodotto esaminato nella fase della value chain.....	54
Tabella 17: Correlazione tra Gradient Boosting Regressor e tipologia di prodotto esaminato nella fase della value chain .....	54
Tabella 18: Correlazione tra Extreme Gradient Boosting Regressor e tipologia di prodotto esaminato nella fase della value chain .....	55
Tabella 19: Correlazione tra Neural Network e tipologia di prodotto esaminato nella fase della value chain.....	55
Tabella 20: Correlazione tra Long Short-Term Memory e tipologia di prodotto esaminato nella fase della value chain .....	56
Tabella 21: Correlazione tra Random Forest e tipologia di prodotto esaminato nella fase della value chain.....	57

Tabella 22: Correlazione tra Long Short-Term Memory e tipologia di prodotto esaminato nella fase della value chain .....	57
Tabella 23: Correlazione tra Extreme Gradient Boosting Regressor e tipologia di prodotto esaminato nella fase della value chain .....	58
Tabella 24: Correlazione tra la fase "Distribution" e dati di supporto all'analisi effettuata nell'articolo.....	59
Tabella 25: Correlazione tra la fase "Retail" e dati di supporto all'analisi effettuata nell'articolo .....	59
Tabella 26: Correlazione tra tipologia di analisi svolta per la value phase "retail" .....	60
Tabella 27: Correlazione tra tipologia di analisi svolta per la value phase "distribution" .....	61
Tabella 28: Relazione tra la fase 'distribution' e il cluster di prodotto principale analizzato dai singoli algoritmi .....	62
Tabella 29: Relazione tra la fase 'retail' e il cluster di prodotto principale analizzato dai singoli algoritmi.....	63
Tabella 30: Correlazione tra l'efficacia degli algoritmi predittivi e il posizionamento operativo nelle fasi di Distribution e Retail .....	64
Tabella 31: Sintesi dei requisiti ottimali di dimensione, range e frequenza dei dati .....	65