Demand Forecasting and Inventory Improvement in Supply Chain Management Using Hybrid Boosting Ensemble Techniques

by

Samira BARGHI

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER'S DEGREE
WITH THESIS IN PROJECT MANAGEMENT ENGINEERING
M.A.Sc.

MONTREAL, AUGUST 11, 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Yvan Beauregard, Thesis Supervisor
Department of Mechanical Engineering, École de technologie supérieure

Mr. Lucas Hof, President of the Board of Examiners
Department of Mechanical Engineering, École de technologie supérieure

Mrs. Hanan Smid, Member of the Jury
Department of General Education, École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON JULY 21 , 2025

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

Prévision de la demande et optimisation des stocks dans la gestion de la chaîne d'approvisionnement à l'aide de techniques d'ensemble hybride par boosting

Samira BARGHI

RÉSUMÉ

Une gestion efficace de la chaîne d'approvisionnement dans le commerce de détail repose en grande partie sur une prévision précise de la demande et une optimisation des stocks afin de garantir l'équilibre des stocks et de minimiser les coûts. Cette étude propose un cadre novateur intégrant un modèle d'ensemble hybride boosting appliqué à deux études de cas (avec et sans gestion de l'incertitude pour un problème de gestion des stocks à un seul échelon) et à la politique de gestion des stocks « Order-Up-to-Level » afin d'améliorer la précision des prévisions et l'efficacité opérationnelle dans l'environnement de vente au détail des magasins Rossmann. L'objectif est d'optimiser les paramètres clés des stocks, notamment les points de commande, le stock de sécurité et le coût total des stocks.

La méthodologie adoptée suit une approche en deux phases. Dans un premier temps, un ensemble hybride composé de LightGBM, CatBoost et XGBoost est développé à partir de données multivariées issues du jeu de données Rossmann, incluant les ventes, les caractéristiques des magasins, les variables temporelles et les promotions. Une généralisation floue est appliquée en amont pour gérer l'incertitude. Le modèle fonctionne en cascade: LightGBM génère des prévisions initiales, CatBoost corrige les erreurs résiduelles, et XGBoost affine les résultats finaux. L'optimisation est réalisée par recherche en grille et validation croisée à 5 plis.

Ces prévisions alimentent ensuite la politique d'inventaire Order-Up-To-Level pour déterminer les décisions d'approvisionnement améliorées. Le cadre est comparé à des approches traditionnelles et à modèles uniques, souvent limitées face à la volatilité de la demande et à l'intégration avec les systèmes d'inventaire. L'ensemble hybride surpasse ces méthodes en tirant parti des avantages spécifiques de chaque algorithme : efficacité computationnelle de LightGBM, traitement performant des variables catégorielles par CatBoost, et régularisation efficace via XGBoost.

Une validation empirique, basée sur les données de 1 115 magasins Rossmann sur 942 jours, montre des améliorations significatives selon des métriques clés telles que l'erreur quadratique moyenne et l'erreur absolue moyenne. Le système proposé permet de réduire les erreurs de prévision, les coûts d'inventaire (commande, stockage, achat), les ruptures de stock et les excédents, tout en augmentant le niveau de service.

En résumé, l'intégration du boosting hybride, de la généralisation floue et de la politique Order-Up-To-Level constitue une solution robuste et fondée sur les données pour relever les défis liés à l'incertitude de la demande et à la gestion des stocks dans la chaîne d'approvisionnement. Le cadre proposé démontre des améliorations opérationnelles concrètes et un fort potentiel d'adoption dans les applications réelles du commerce de détail.

Mots-clés: Gestion de la chaîne d'approvisionnement, Prévision de la demande, Optimisation des stocks, Ensemble hybride de boosting, Politique Order-Up-To-Level, Regroupement flou, Incertitude de la demande

Demand Forecasting and Inventory Improvement in Supply Chain Management Using Hybrid Boosting Ensemble Techniques

Samira BARGHI

ABSTRACT

Effective supply chain management in retail depends heavily on accurate demand forecasting and inventory improvement to ensure balanced stock levels and minimize costs. This study proposes a novel framework for a single echelon inventory management problem, that integrates a hybrid boosting ensemble model applied to two case studies (with and without handling uncertainty) with the Order-Up-to-Level inventory policy to enhance forecast precision and operational efficiency within the retail environment of Rossmann stores. Orders are aggregated for multiple items to reduce ordering costs, improving overall cost efficiency. The objective is to determine key inventory parameters, including reorder points, safety stock, and total inventory costs.

The methodology follows a two-phase approach. First, a hybrid ensemble of LightGBM, CatBoost, and XGBoost is developed using multivariate data from the Rossmann dataset, which includes sales figures, store attributes, time-related variables, and promotional factors. Prior to training, fuzzy generalization is applied to handle uncertainty in the data. The ensemble operates sequentially: LightGBM provides initial predictions, CatBoost reduces residual errors, and XGBoost performs final refinements. Model tuning is conducted through grid search combined with 5-fold cross-validation.

These forecasts are then incorporated into the Order-Up-To-Level inventory policy to determine improved ordering decisions. The framework is benchmarked against traditional and single-model approaches, which often struggle with demand volatility and integration with inventory systems. The hybrid ensemble outperforms these methods by leveraging the strengths of each model—computational efficiency from LightGBM, categorical handling by CatBoost, and regularization from XGBoost.

Empirical validation using data from 1,115 Rossmann stores over 942 days demonstrates substantial improvements in key performance metrics such as Root Mean Squared Error, Mean Absolute Error, and Pred(x=10%) compared to baseline models such as ARIMA, Moving Average, Simple Exponential Smoothing, and single-model machine learning approaches (LightGBM, XGBoost, and CatBoost). The proposed system results in lower forecast errors, reduced inventory costs, fewer stockouts and overstock incidents, and higher service levels.

In summary, the integration of hybrid boosting, fuzzy generalization, and the Order-Up-To-Level policy provides a robust, data-driven solution to the challenges of demand uncertainty and inventory control in retail supply chain management. The proposed framework offers tangible improvements in operational performance, highlighting its potential for broader adoption in real-world retail applications.

Keywords: Supply Chain Management, Demand Forecasting, Inventory improvement, Hybrid Boosting Ensemble, Order-Up-To-Level Policy, Fuzzy Clustering, Demand Uncertainty

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criterion
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
ARIMAX	Autoregressive Integrated Moving Average with Explanatory Variables
AUC	Area Under the Curve
BB	Balanced Bagging
CATboost	Category Boosting
CNN	Convolutional Neural Network
CoV	Coefficient of Variation
DDI	Downstream Demand Inference
DLR	Dynamic Linear Regression
DSR	Design Science Research
ETS	Error, Trend, Seasonal
GA	Genetic Algorithm
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GOSS	Gradient-based One-Side Sampling
GRU	Gated Recurrent Unit
HC	Holding Cost

IMS	Inventory Management System
KNN	K-Nearest Neighbors
LGBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory
MA	Moving Average
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MSE	Mean Squared Error
NIS	No Information Sharing
OC	Ordering Cost
OUTL	Order Up To Level
PC	Purchasing Cost
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network

ROP	Reorder Point
SBC	Schwarz Bayesian Criterion
SBFA	Switching-Based Forecasting Approach
SCM	Supply Chain Management
SMA	Simple Moving Average
SMAPE	Symmetric Mean Absolute Percentage Error
SS	Safety Stock
SVM	Support Vector Machine
SVR	Support Vector Regression
TC	Total Cost
TFT	Temporal Fusion Transformer
XGBoost	eXtreme Gradient Boosting

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

$c_{j,k}$	Representative value (centroid) for the k-th category of feature j in fuzzy generalization
$d_{i,j,k}$	Absolute distance from feature value $x_{i,j}$ to centroid $c_{j,k}$
$f_{CatBoost}(x)$	Prediction function of the CatBoost model
$f_{LGBM}(x)$	Prediction function of the LightGBM model
$f_{XGB}(x)$	Prediction function of the XGBoost model
K	Number of categories for fuzzy discretization (set to 12)
k^*	Index of the nearest centroid in fuzzy generalization
M	Number of features in the dataset
max	Desired maximum value for Min-Max scaling (set to 1)
max_j	Maximum value of feature j
min	Desired minimum value for Min-Max scaling (set to 0)
min_j	Minimum value of feature j
N	Number of samples in the dataset
n	Number of observations in evaluation metrics
Q	Order Quantity in the OUTL inventory policy
$OUTL$	Order-Up-To-Level in the inventory policy
R_j	Range of feature j (max$_j$ - min$_j$)
ROP	Reorder Point in the OUTL inventory policy
SS	Safety Stock in the OUTL inventory policy

TC	Total Cost
$X_{\text{Scal\_mrg}}$	Scaled and merged dataset
x	Original feature value
$x_{i,j}$	Value of feature j for sample i
x_{scaled}	Scaled feature value after Min-Max normalization
x_{std}	Standardized feature value
x_{max}	Maximum value of a feature
x_{min}	Minimum value of a feature
$\hat{x}_{i,j}$	Generalized (discretized) value for $x_{i,j}$ in fuzzy generalization
y_i	Actual value for the i-th observation
$\hat{y}$	Final predicted value from the ensemble
$\hat{y}_i$	Predicted value for the i-th observation
Δ_j	Interval width for feature j in fuzzy generalization
η	Learning rate or step size for boosting algorithms

INTRODUCTION

In today's competitive retail landscape, effective demand forecasting and inventory management are critical for maintaining customer satisfaction and operational efficiency, particularly in the face of demand uncertainty and stock imbalances. This research focuses on addressing these challenges within a single-retailer context, specifically targeting the operational dynamics of Rossmann stores, a retail chain with 1,115 stores analyzed over 942 days with multivariate features like sales, store attributes, temporal variables, and promotions. Unlike multi-echelon systems that encompass suppliers, warehouses, and retail outlets with complex interdependencies, this study adopts a retailer-only system model to streamline the analysis of demand forecasting and inventory improvement. By concentrating on the retail echelon, the proposed framework isolates the retailer's inventory decisions—such as safety stock and reorder points—from upstream supply chain complexities, enabling a focused evaluation of forecasting accuracy and its direct impact on retail inventory performance. This approach, while simplifying the supply chain scope, provides actionable insights into enhancing service levels and reducing total cost at the retail level, as demonstrated through the application of a hybrid boosting ensemble and the Order-Up-To-Level (OUTL) inventory policy.

0.1 Demand in supply chain

Demand in the supply chain represents the quantity of goods or services that customers—individuals, businesses, or organizations—are willing and able to purchase at a specific price, within a defined timeframe, across the supply chain network. The supply chain encompasses processes for sourcing raw materials from suppliers, transforming them into finished goods, and delivering these products to consumers through retail outlets (Ul Haq Qureshi et al., 2024).

As depicted in the Multi-Echelon Supply Chain diagram, the process begins with suppliers providing raw materials to an inventory warehouse, which then distributes goods to retail stores.

These stores fulfill customer orders, while a shipping mechanism ensures efficient delivery from the warehouse to the retail points, completing the cycle of material flow and ensuring customer satisfaction. This interconnected system highlights the importance of seamless coordination between suppliers, inventory management, retail operations, and shipping to meet demand effectively. In contrast to a multi-echelon approach, which enhances coordination and visibility into broader supply chain dynamics at the cost of increased complexity, this single-echelon problem focuses solely on the retail level, excluding upstream interactions, thereby simplifying coordination but limiting visibility into the overall supply chain.

Figure 0.1 Multi-Echelon Supply Chain System
Adapted from Ul Haq Qureshi et al., 2024

Demand acts as the primary force guiding supply chain operations, influencing key activities such as inventory management, production scheduling, raw material procurement, and distribution logistics. Demand is shaped by multiple factors, including market trends, customer preferences, seasonal patterns, pricing strategies, promotional efforts, and broader economic conditions like inflation or supply disruptions. While dynamic and interdependent variables indeed contribute to demand uncertainty, this study specifically addresses such uncertainty through a targeted hybrid boosting ensemble and fuzzy generalization, tailored to the Rossmann dataset's multivariate retail context. This study focuses on a structured, data-driven approach to uncertainty (via machine learning and fuzzy techniques) rather than broadly exploring all systematic forecasting methods.

0.2 Demand forecasting in inventory management

Demand forecasting is a vital element of inventory management, ensuring improved stock levels. Its accuracy directly shapes decision-making in supply chain operations, enabling effective planning of procurement, production, and distribution. The process entails analyzing historical data, market trends, and external factors to predict future demand, allowing organizations to maintain sufficient inventory while curbing holding costs. In retail settings like Rossmann stores, where stock imbalances threaten operational success, precise demand forecasting is foundational to supply chain management (SCM). Stoilov and Stoilova (2024) highlight its role in estimating production needs, reducing uncertainty, and averting supply chain disruptions, while Brunaud et al. (2018) underscore its importance in proactively adjusting inventory policies to address demand fluctuations analysing four alternatives to estimate the sufficient amount of safety stock amount in a supply chain planning context.

Historically, the traditional demand forecasting methods, such as Moving Averages (MA) and Autoregressive Integrated Moving Average (ARIMA), relied on univariate time-series data to identify patterns (Abolghasemi, 2019). However, these approaches falter in modern retail contexts, where customer behavior shifts rapidly due to digitization, promotions, holidays, and market volatility. The advent of machine learning (ML) marked a paradigm shift, enabling the analysis of expansive, multivariate datasets—such as the Rossmann dataset's sales, store, temporal, and promotional features—to uncover non-linear relationships as validated by Ul Haq Qureshi et al. (2024). Single-based ML models like LightGBM, CatBoost, and XGBoost, improved upon traditional methods by leveraging advanced gradient boosting techniques, yet their individual limitations in capturing diverse demand patterns spurred further innovation. Hybrid ensemble boosting, as employed in this research, integrates these models sequentially—LightGBM for efficiency, CatBoost for categorical handling, and XGBoost for regularization—to enhance prediction accuracy, a methodology validated by Seyedan et al. (2023) for inventory optimization under uncertainty.

0.3 Components of Demand Forecasting in Inventory Management

Demand forecasting in inventory management comprises several key components that collectively enable businesses to predict future demand accurately and improve stock levels. The demand forecasting framework includes essential elements such as historical data, forecasting models, demand patterns, and technology-driven tools. Each component plays a critical role in improving forecasting accuracy and enhancing overall inventory management:

1. **Historical Data**

   The foundation of demand forecasting is historical sales and inventory data. This data includes past demand trends, seasonal fluctuations, customer purchasing behavior, and external factors such as economic conditions. By analyzing historical data, businesses can identify recurring patterns and predict future demand with greater accuracy. Historical data acts as a baseline for forecasting models, allowing organizations to anticipate market needs and make informed inventory decisions.

2. **Forecasting Models**

   Demand forecasting relies on various mathematical and statistical models to generate accurate predictions. These models include:

   - Time Series Models (e.g., ARIMA, Exponential Smoothing) that analyse historical data collected at regular intervals to identify patterns, trends, and seasonal variations, helping to project future demand based on past behaviour. Time series models can be:

- Univariate – Track a single variable over time (e.g., weekly customer orders).
- Multivariate – Track multiple variables over time (e.g., monthly revenue alongside advertising spends and price changes) to analyze how they interact and influence each other.

- Causal Models that incorporate external factors such as economic indicators, competitor actions, and marketing campaigns.
- Machine Learning Models that leverage big data and artificial intelligence to refine demand forecasts dynamically..

These models help organizations choose the most suitable forecasting technique based on the complexity and variability of their demand patterns.

3. **Demand Patterns and Influencing Factors**

Demand forecasting requires understanding various demand patterns and the factors influencing them. Common patterns include:

- Seasonality: Repetitive demand fluctuations occurring during specific times of the year (e.g., holiday sales, back-to-school season).
- Trends: Long-term shifts in demand, either increasing or decreasing over time.
- Cyclical Variations: Demand changes influenced by broader economic cycles, such as recessions or market booms.
- Random Fluctuations: Unpredictable variations due to sudden market changes or unforeseen external events.

Identifying these patterns helps organizations develop more accurate forecasts and adjust their inventory strategies accordingly.

4. **Inventory Policies and Safety Stock Management**

Effective demand forecasting is closely linked to inventory policies that dictate how stock is replenished and managed. Key policies include:

- Reorder Point Systems: Defining inventory levels at which replenishment orders are triggered.
- Just-in-Time (JIT) Inventory: Keeping stock levels minimal and ordering only when needed.
- Safety Stock Strategies: Maintaining buffer inventory to account for demand variability and supply chain disruptions.
- OUTL Policy: The Order-Up-To Level (OUTL's (R,s,S) system) policy replenishes inventory to a predetermined target level (S is the target inventory level) when stock falls at or below a reorder point. This strategy ensures optimal inventory levels (Seyedan et al., 2023), reducing excess stock and holding costs while aligning with demand forecasts for efficient replenishment. This outcome is achieved independently of the Economic Order Quantity (EOQ) approach, as the OUTL policy adapts to time-varying demand through periodic reviews and forecast updates (Chopra et al., 2021). In the context of ABC inventory classification, A items, which are high-value and require tight control strategies like the OUTL policy, which balances cost efficiency and responsiveness, while C items, being low-value, are managed with simpler periodic review systems; B items which are medium-value (Silver et al., 2017). This study specifically targets B items within the Rossmann stores' inventory, as the OUTL policy effectively manages their high-value demand variability, ensuring efficient replenishment.

By aligning inventory policies with demand forecasts, businesses can improve stock levels under the OUTL policy, which adjusts the reorder point (ROP) and order quantity (Q) based on ensemble-boosted demand predictions to balance ordering and holding costs, though shortage costs are not considered in this model. While assuming expediting costs are negligible, if expediting costs exceed holding costs, increasing safety stock could be prioritized to reduce expediting needs, though this trade-off is not modeled here. This approach reduces unnecessary holding costs while ensuring efficient replenishment for Rossmann stores' B items.

0.4 Features

Demand forecasting in inventory management encompasses a range of critical features that contribute to its accuracy, efficiency, and adaptability in real-world applications. These features define the effectiveness of forecasting systems in predicting future demand, improving inventory levels, and ensuring seamless supply chain operations. Below, we explore the essential features that underpin the robustness and practicality of demand forecasting in inventory management:

1. Data-Driven Accuracy: Modern demand forecasting relies on extensive historical data, market trends, and external influences to generate precise predictions, reducing uncertainty in inventory planning.
2. Scalability: Forecasting models are designed to handle increasing volumes of data and adapt to the expanding scope of inventory systems, making them suitable for businesses of all sizes.
3. Multi-Model Integration: Effective demand forecasting combines statistical, machine learning, and deep learning models, ensuring flexibility in selecting the most suitable approach for different inventory environments.
4. Robustness Against Demand Volatility: Demand forecasting incorporates uncertainty modelling, such as seasonal time-series techniques and probabilistic approaches and confidence intervals, to address seasonal variations, economic shifts, and unforeseen disruptions in the supply chain.

These features collectively enhance the reliability and efficiency of demand forecasting in inventory management, ensuring businesses can make proactive, data-driven decisions to improve inventory control and supply chain operations. The proposed hybrid boosting ensemble forecasting framework, integrating fuzzy generalization, excels in robustness against demand volatility by capturing complex, non-linear patterns and significantly reducing forecast errors compared to single models. It also strengthens multi-model integration, leveraging each algorithm's strengths for superior adaptability in retail contexts like Rossmann stores. Additionally, it ensures data-driven accuracy through rigorous preprocessing, effectively handling multivariate features such as sales, temporal, and promotional data, enhancing overall supply chain efficiency.

0.5 Applications

Demand forecasting in inventory management plays a crucial role across various industries, enabling businesses to improve stock levels, reduce costs, and enhance supply chain efficiency. Traditional techniques, such as Moving Averages and ARIMA, have long been used to predict demand and manage inventory, helping to mitigate overstocking and stockouts. By building on these methods, advanced forecasting techniques, like machine learning and hybrid ensembles, are better suited to handle fluctuating demand, offering improved accuracy and adaptability to capture complex patterns driven by promotions, holidays, or market shifts, thus allowing organizations to make data-driven decisions for robust inventory management. Below are some key applications of demand forecasting in inventory management:

1. Retail and E-commerce: Predicting customer demand for different products, enhancing warehouse stocking, and reducing lost sales due to inventory shortages.
2. Manufacturing: Ensuring raw material availability, aligning production schedules with demand trends, and minimizing excess inventory storage costs.
3. Healthcare and Pharmaceuticals: Forecasting the demand for medical supplies, vaccines, and prescription drugs to maintain adequate stock levels and avoid shortages.
4. Supply Chain and Logistics: Enhancing inventory visibility, reducing lead times, and improving distribution strategies to meet fluctuating market demands.

5. Food and Beverage Industry: Managing perishable goods efficiently by predicting demand variations based on seasonal trends, consumer preferences, and external factors.
6. Automotive Industry: Forecasting demand for spare parts, enhancing inventory levels in dealerships, and improving production planning.
7. Energy and Utilities: Estimating fuel and electricity consumption patterns to ensure sufficient supply and prevent resource shortages.

0.6 Challenges

Despite its significant advantages, demand forecasting in inventory management presents several challenges that can affect its accuracy, efficiency, and real-world implementation. These challenges stem from data limitations, model complexity, and external market uncertainties. Below are some key challenges encountered in demand forecasting for inventory management:

1. Data Quality and Availability: Accurate forecasting relies heavily on high-quality historical data. Issues such as missing values, inconsistencies, or limited data availability can significantly reduce prediction accuracy. Ensuring clean, reliable, and comprehensive datasets, such as store-related, sales-related, and temporal data, is a major challenge. Additionally, the preprocessing steps like handling missing values and feature engineering must be carefully managed to maintain data integrity.
2. Demand Volatility: Sudden changes in consumer preferences, seasonal trends, and unforeseen events (e.g., economic crises or pandemics) can disrupt forecasting models. The difficulty of predicting demand variability, influenced by factors like promotions, holidays, and market volatility, makes maintaining inventory efficiency challenging. Moreover, modeling demand uncertainty, such as through fuzzy generalization methods, introduces an additional layer of complexity. To ensure the continued adequacy of forecasting models, techniques like regular model retraining with fresh data, hyperparameter tuning via grid

search, and robust preprocessing methods, such as fuzzy generalization, help models adapt to evolving demand patterns and maintain predictive reliability.

3. Integration with Supply Chain Systems: The seamless integration of forecasting models into Inventory Management Systems (IMS) is a significant technical hurdle. Compatibility issues and the need for real-time data synchronization can complicate the deployment of machine learning models into production environments. Additionally, aligning the forecasting model with inventory policies, such as the OUTL, requires careful integration to ensure improved reorder points and stock levels.
4. Single-based Forecasting Methods: Relying on single-based forecasting methods, such as machine learning models like LightGBM, CatBoost, or XGBoost, can limit accuracy due to their inability to fully capture complex patterns and diverse trends in data. As classified in Figure 0.2 (Kersten et al. 2019), ML-based forecasting methods are categorized into supervised, unsupervised, and reinforcement learning, with supervised learning encompassing regression tasks like ensemble methods (e.g., LightGBM, CatBoost, XGBoost) and neural networks, which are commonly used for demand forecasting but often limited when applied individually.

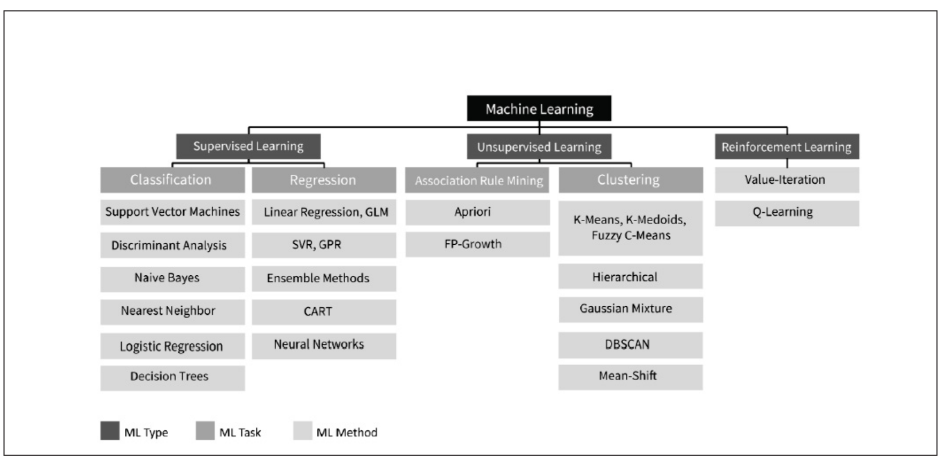

Figure 0.2 Different categories of ML models
Adapted from Kersten et al., 2019

This categorizing underscores the diversity of ML approaches, yet highlights the limitations of single models in handling multivariate retail data. In contrast, non-machine learning methods, like exponential smoothing for stable demand patterns or moving averages for short-term trends, as outlined by Silver et al. (2017), are effective in simpler scenarios but struggle with multivariate volatility. Using hybrid methods that combine multiple models, such as the proposed boosting ensemble, enhances prediction accuracy by leveraging the strengths of each model and compensating for individual weaknesses, improving robustness for retail contexts like Rossmann stores. The model is not strictly limited to the Rossmann context but is optimized for retail environments with similar demand and data characteristics. Its principles are adaptable to other retail and non-retail sectors, though further validation is needed.

As demand forecasting continues to evolve, addressing these challenges will be essential for improving model accuracy, ensuring adaptability, and enhancing inventory improvement across industries. The integration of advanced techniques, such as hybrid ensemble models and fuzzy generalization, will play a critical role in overcoming these challenges.

0.7 Problem Statement

Demand forecasting in supply chain management for Rossmann stores is a critical yet challenging task due to the intricate interplay of data limitations, demand volatility, and technical integration hurdles that undermine demend prediction accuracy and inventory improvement. The reliance on high-quality historical data is often compromised by issues such as missing values, inconsistencies, and incomplete datasets (e.g., store-related, sales-related, and temporal features), which degrade forecasting reliability and complicate preprocessing efforts like feature engineering and missing value imputation. Additionally, demand volatility—driven by sudden shifts in consumer preferences, seasonal trends, promotions, holidays, and unforeseen events like economic disruptions—introduces significant uncertainty, making it difficult to maintain efficient inventory levels and increasing the complexity of modeling techniques such as fuzzy generalization. Traditional forecasting methods (e.g., MA, ARIMA) and even single-based machine learning models (e.g., LightGBM, CatBoost, XGBoost) struggle to fully capture the non-linear, multivariate patterns in the Rossmann dataset, limiting their ability to address diverse trends and achieve robust predictions. Furthermore, integrating these forecasts into inventory management systems (IMS) to support policies like the OUTL framework poses technical challenges, including compatibility issues and the need for real-time data synchronization, which hinder seamless deployment and alignment with operational goals like improving reorder points and safety stock. The proposed method ascertains inventory based solely on on-hand inventory levels, lacking consideration of orders in transit, which may limit real-time synchronization capabilities. Overcoming these multifaceted challenges demands an advanced, hybrid approach that enhances forecasting accuracy, mitigates uncertainty, and ensures practical applicability in the dynamic retail context of Rossmann stores.

0.8 Research Objectives

The primary objective of this research is to enhance supply chain performance by developing, implementing, and evaluating a robust demand forecasting model using a hybrid boosting ensemble framework—combining LightGBM, CatBoost, and XGBoost—integrated with the OUTL inventory policy, applied to the Rossmann stores dataset. This study focuses on improving critical inventory parameters such as reorder points, safety stock levels, and total annual costs, while addressing key challenges in forecasting accuracy, demand volatility, and operational efficiency within a single-retailer context. The OUTL policy (R,s,S), a periodic review approach described by Silver et al. (1998), ensures inventory replenishment decisions account for lead time and demand variability, mitigating risks of stockouts and overstocking in dynamic retail environments (Seyedan et al., 2023). Through rigorous experimentation, this research aims to provide a data-driven solution that reduces forecast errors, minimizes costs, and approaches service level (95%), offering actionable insights for Rossmann store operations. The specific objectives are as follows:

1. **Improve Forecasting Accuracy through Data Quality and Model Design**: Develop and assess a hybrid ensemble model to achieve highly accurate daily demand forecasts, leveraging the sequential learning capabilities of LightGBM, CatBoost, and XGBoost to overcome single-model limitations. This involves designing preprocessing pipelines to manage data quality issues—such as missing values, inconsistencies, and feature engineering—within the Rossmann dataset, ensuring reliable inputs. The ensemble, augmented by fuzzy generalization, captures complex, non-linear patterns influenced by promotions, holidays, and market volatility, reducing prediction errors as measured by RMSE, MAE, MAPE, and Pred(x=10%), thus fostering a robust supply chain.
2. **Model Demand Volatility**: Leverage the boosting ensemble's sequential learning capabilities, augmented by techniques like fuzzy generalization, to model demand uncertainty and capture complex, non-linear patterns influenced by promotions, holidays, and market volatility.
3. **Improve Inventory Parameters via OUTL Integration**: Calculate appropriate reorder points, safety stock levels, and order-up-to levels by integrating ensemble-based demand forecasts into the OUTL policy framework, accounting for demand variability, fixed lead time, and desired service levels (e.g., 95%+). This ensures inventory policies are resilient, balancing availability and cost efficiency, and maintaining high service levels to meet supply chain demands.
4. **Minimize Total Inventory Costs**: Quantify and minimize total annual inventory-costs—comprising purchase costs (driven by annual demand), holding costs (tied to average inventory levels), and ordering costs (linked to order frequency)—through precise demand forecasts within the OUTL policy. By aligning inventory decisions with demand expectations, this

objective balances ordering and holding costs by preventing over-ordering, which increases holding costs, and under-ordering, which raises ordering frequency, enabling strategic resource allocation and bolstering supply chain efficiency.

5. **Evaluate Performance and Compare Approaches**: Perform comprehensive evaluations of the hybrid ensemble and its integration with the OUTL policy, comparing outcomes across key metrics such as forecast accuracy (RMSE, MAE, MAPE, Pred(x=10%)) and total costs. These error metrics were selected because they are widely recognized in the forecasting literature and offer complementary insights: RMSE emphasizes large errors, MAE provides a straightforward average error, MAPE allows for scale-independent comparison, and Pred(x=10%) assesses the proportion of forecasts within an acceptable error range. This combination ensures a balanced evaluation of both accuracy and robustness. The analysis involves assessing the effectiveness of different forecasting strategies to identify the most efficient combinations, providing insights into how accuracy influences operational performance and guiding future refinements in retail supply chain management.

0.9 Methodology Overview

This research presents a two-phase methodology for demand forecasting and inventory improvement in supply chain management using the Rossmann stores dataset (1,115 stores, 942 days), integrating a boosting ensemble of LightGBM, CatBoost, and XGBoost. In the forecasting phase, the multivariate dataset—comprising sales, store, temporal, and promotional features—is preprocessed by removing zero sales(store closure), imputing missing values (mean for numerical, mode for categorical), encoding categories, engineering features, and applying Min-Max normalization. In this dataset, the "Sales", representing quantity sold, is used as a proxy for demand under the assumption that sales reflect customer demand when inventory is available. The historical demand data is split into 80% training and 20% testing, with 5-fold cross-validation on the training set to tune hyperparameters and prevent overfitting, followed by a single test set evaluation for unbiased assessment, and the ensemble is trained sequentially: LightGBM generates initial predictions, CatBoost corrects residuals, and XGBoost refines outputs. Grid search optimizes hyperparameters, with performance assessed via RMSE and MAE. Fuzzy generalization discretizes features to mitigate demand uncertainty, enhancing robustness. In the inventory phase, predictions inform an OUTL policy, improving reorder points, safety stock, and total costs (purchase, holding, ordering). This approach combines advanced ML and inventory strategies to deliver precise, actionable solutions for Rossmann's retail supply chain.

0.10 Contributions

This research offers several significant contributions to the field of demand forecasting and supply chain management, particularly within the retail sector exemplified by Rossmann stores:

1. Advanced Boosting Ensemble Model: The study aims to develops a novel hybrid forecasting framework that integrates LightGBM, CatBoost, and XGBoost in a sequential boosting ensemble, capitalizing on their complementary strengths—e.g., LightGBM's efficiency, CatBoost's categorical feature handling, and XGBoost's regularization—to achieve superior prediction accuracy over single-based models, as validated by RMSE and MAE metrics.
2. Robust Handling of Demand Uncertainty: By incorporating a fuzzy generalization technique into the preprocessing pipeline, this work introduces a structured approach to modeling demand volatility, smoothing noisy data, and enhancing the ensemble's resilience to fluctuations driven by promotions, holidays, and market shifts, thereby addressing a key challenge in retail forecasting..
3. Improved Inventory Management: The research bridges predictive analytics and operational strategy by integrating ensemble forecasts into the OUTL inventory policy, providing a practical framework for calculating improved reorder points, safety stock levels, and cost-efficient inventory decisions, directly applicable to Rossmann store operations.

4. Actionable Insights for Retail Supply Chains: Through comprehensive evaluation and comparison of forecasting approaches, this study delivers actionable insights into how ensemble techniques improve demand forecasting accuracy, and lower total inventory costs, contributing valuable methodologies to the supply chain management literature. These contributions collectively advance the application of machine learning in retail supply chains, offering a data-driven, replicable solution that enhances forecasting accuracy and operational efficiency.

0.11 Conceptual Map

To guide the literature review, combination of systematic and snowballing, and provide a theoretical foundation for this research, a conceptual map is presented in Figure (0.3). This map illustrates the key concepts central to demand forecasting and inventory optimization in supply chain management, as applied to Rossmann stores, and their interrelationships with the research problem. The map is structured around three clusters: *Demand Forecasting*, *Inventory Management*, and *Methodological Approaches*, reflecting the core components of the proposed hybrid boosting ensemble framework integrated with the OUTL inventory policy. Each concept is defined, and its relevance to the research problem—enhancing forecasting accuracy and enhancing inventory under demand uncertainty—is highlighted. The map informs the structure of the literature review, which provides definitions, references, and state-of-the-art analyses for these concepts, aligning with the design science methodology to develop and evaluate the proposed artifact.

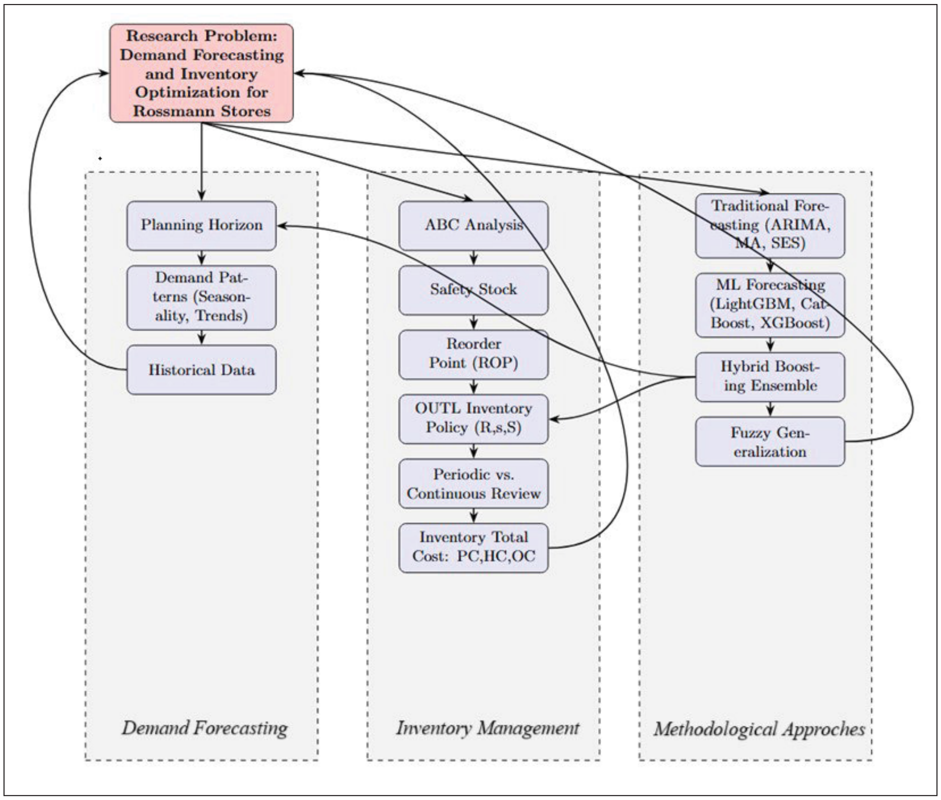

Figure 0.3 Conceptual Map of Key Concepts for Demand Forecasting and Inventory improvement

The conceptual map shown in Figure 0.3 organizes key concepts into three clusters, each addressing a critical aspect of the research problem:

- **Demand Forecasting**: This cluster includes *Planning Horizon*, *Demand Patterns (Seasonality, Trends)*, and *Historical Data*. The planning horizon (e.g., daily predictions) shapes forecast accuracy (Ahmed et al, 2024, leveraging historical data to identify demand patterns like seasonality and trends, which are critical for predicting Rossmann store sales (Makridakis et al, 2020). These concepts feed into the research problem by providing the data foundation for accurate forecasting.
- **Inventory Management**: This cluster encompasses *ABC Analysis*, *Service Level*, *Safety Stock*, *Reorder Point (ROP)*, *OUTL Policy*, *Periodic vs. Continuous Review*, *Stockouts*, and

Inventory Cost Components. ABC analysis identifies B items for Rossmann stores (Silver et al, 2017). Service level drives safety stock and ROP calculations, which inform the OUTL policy, a periodic review system improving reorder decisions (Chopra et al, 2021). Total costs of inventory (purchase, holding, ordering) are decreased through these policies, directly addressing the research problem of cost-efficient inventory.

- **Methodological Approaches**: This cluster includes *Traditional Forecasting (ARIMA, MA, etc)*, *ML Forecasting (LightGBM, CatBoost, XGBoost)*, *Hybrid Boosting Ensemble*, and *Fuzzy Generalization*. Traditional methods provide a baseline but struggle with multivariate data, while ML models improve accuracy (Qureshi, 2024). The hybrid boosting ensemble, enhanced by fuzzy generalization, integrates these models to handle demand uncertainty, feeding accurate forecasts into the OUTL policy and planning horizon decisions (Seyedan et al, 2023).

The relationships between concepts and the research problem are depicted as arrows. Demand forecasting concepts provide the predictive foundation, inventory management concepts translate forecasts into operational decisions, and methodological approaches enhance both forecasting accuracy and inventory policy efficacy. The hybrid boosting ensemble and fuzzy generalization bridge forecasting and inventory management, ensuring robust predictions and improved inventory parameters (e.g., safety stock, ROP) under uncertainty.

0.12 Thesis Structure

This thesis is organized into six chapters. Chapter 1 details the research methodology, Chapter 2 provides a comprehensive literature review, examining definitions, key concepts, demand forecasting methods in supply chain management and the significance of the Rossmann dataset, identifying research gaps. Chapter 3 suggests an artifact and develops a solution to the research problem mentioned earlier. The proposed framework consists of briefly introduction of the

Rossmann stores and dataset with processes such as preprocessing, boosting ensemble design (LightGBM, CatBoost, XGBoost), fuzzy generalization, evaluation metrics (RMSE, MAE, MAPE, Pred(x)), and OUTL inventory integration. Chapter 4 presents the numerical results. Then, Chapter 5 addresses the discussion of the results, implications for Rossmann's supply chain analysis, compares forecasting accuracy and inventory outcomes against benchmarks, supported by visualizations and statistics, and suggesting future research avenues. Finally, thesis's conclusion is addressed by summarizing key findings, and outcomes.

CHAPTER 1

METHODOLOGY

This chapter delineates the methodological approach adopted to address the challenges of demand forecasting and inventory improvement within the retail supply chain context of Rossmann stores. The study employs the Design Science Research (DSR) framework, a systematic methodology for creating and evaluating artifacts to solve identified problems, as illustrated in Figure 1.1 and adapted from Vaishnavi and Kuechler (2011). The DSR framework is particularly suited for this research as it facilitates the development of a practical solution—a hybrid boosting ensemble integrated with fuzzy generalization and the Order-Up-To-Level (OUTL) inventory policy—while ensuring rigorous evaluation. The framework comprises five phases: Awareness of the Problem, Suggestion, Development, Evaluation to Confirm the Solution, and Conclusion. These phases guide the research in designing an artifact that enhances forecasting accuracy, manages demand uncertainty, and optimizes inventory decisions in a single-echelon retail environment.

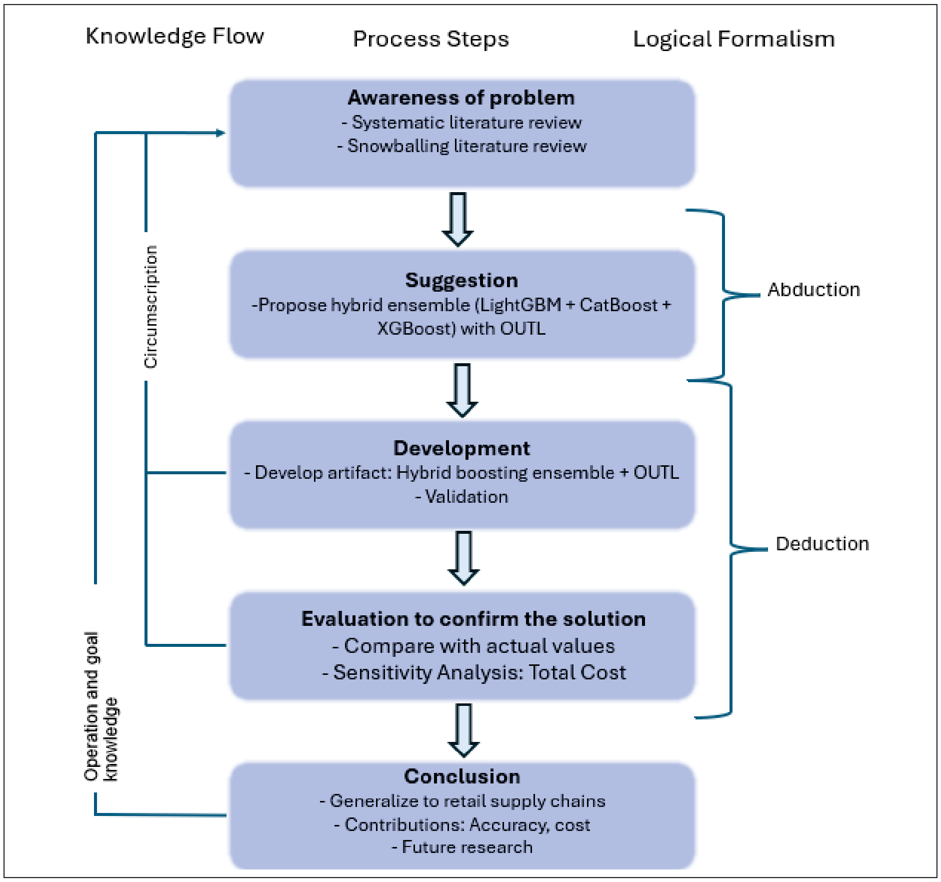

Figure 1.1 Desing science methodology
Taken from Vaishnavi and Kuechler,2011

1.1 Awareness of the Problem

The initial phase of the DSR framework focuses on recognizing and articulating the problem within the domain of demand forecasting and inventory management in retail supply chains. To achieve this, a dual literature review approach was conducted, combining a systematic literature review (SLR) with snowballing to thoroughly examine existing research and pinpoint deficiencies in current approaches. The SLR targeted academic databases such as IEEE Xplore, Scopus, and Web of Science, covering publications from 2015 to 2025. Search terms included phrases like “demand forecasting," “inventory improvement," “machine learning in supply chains," “hybrid ensemble models," and “retail supply chain management." Snowballing complemented the SLR by tracing citations to uncover additional relevant studies. This comprehensive review aimed to map the landscape of demand forecasting techniques and their integration with inventory management practices, identifying gaps that this study seeks to address.

1.1.1 Systematic Literature Review Process

The SLR was executed with a structured protocol to ensure methodological rigor and reproducibility. The process began with an initial search that yielded many articles based on the specified keywords. Inclusion criteria were then applied to filter the studies, focusing on those relevant to retail supply chains, addressing demand forecasting or inventory improvement, and employing either machine learning or traditional forecasting methods. This screening process involved reviewing titles, abstracts, and full texts, ultimately reducing the pool to 20 highly relevant studies. These studies were systematically analyzed to evaluate the strengths and weaknesses of existing methodologies, providing a foundation for the proposed research.

1.1.2 Snowballing Literature Review Process

To broaden the scope of the literature review, a snowballing technique was applied following the SLR. This process involved both backward and forward snowballing on the 20 selected studies. Backward snowballing examined the references cited within these studies to identify foundational works, while forward snowballing used tools like Google Scholar to find newer studies that cited the selected papers. This iterative approach ensured the inclusion of seminal works and recent advancements not captured in the initial database search. Snowballing added 20 additional relevant studies, bringing the total to 40. These studies were analyzed alongside the SLR findings to provide a more comprehensive understanding of demand forecasting and inventory management, particularly in retail contexts, and to confirm the identified gaps in uncertainty modeling and operational integration.

1.1.3 Key Findings from the Literature Review

The combined SLR and snowballing review revealed several critical shortcomings in current demand forecasting and inventory management practices. Traditional forecasting methods, such as Moving Averages and ARIMA, were frequently found to be inadequate in handling the complex, multivariate demand patterns typical in retail environments like Rossmann stores. These methods often rely on linear assumptions and univariate data, making them ill-equipped to address demand volatility driven by factors such as promotions, holidays, and market shifts. Single machine learning models, such as LightGBM or XGBoost, offered improvements by capturing non-linear relationships, but they often struggled with noisy or uncertain data, a prevalent issue in retail datasets. Moreover, while hybrid models showed potential in combining the strengths of multiple algorithms, their application in retail settings was limited, particularly in terms of integration with inventory policies like OUTL. Additionally, the literature highlighted a lack of robust uncertainty modeling techniques, such as fuzzy generalization, to enhance forecasting resilience in the face of demand fluctuations. These findings underscored the need for a comprehensive framework that integrates advanced forecasting techniques, uncertainty handling, and inventory improvement, specifically tailored for the retail context of Rossmann stores.

1.2 Suggestion

Based on the insights derived from the SLR and snowballing, this phase proposes a novel framework to address the identified research gaps: a hybrid boosting ensemble combining LightGBM, CatBoost, and XGBoost, augmented with fuzzy generalization, and integrated with the OUTL inventory policy. The proposed artifact is designed to achieve three primary objectives: enhance the accuracy of demand forecasts in a retail setting, effectively manage demand uncertainty to ensure robust predictions, and optimize inventory parameters such as reorder points, safety stock levels, and overall costs within the Rossmann store supply chain. The framework operates in a two-phase process: initially generating precise demand forecasts using the hybrid ensemble, and subsequently utilizing these forecasts within the OUTL policy to inform cost-effective inventory decisions.

1.2.1 Rationale for the Proposed Framework

The selection of a hybrid boosting ensemble is driven by the findings from the SLR and snowballing that individual machine learning models, while effective in specific contexts, often fail to fully capture the complexity of retail demand patterns. LightGBM is chosen for its computational efficiency, making it suitable for handling large datasets like the Rossmann dataset. CatBoost is incorporated for its superior handling of categorical features, such as store types and promotional indicators, which are prevalent in retail data. XGBoost complements these by providing robust regularization to mitigate overfitting, enhancing the model's generalization capabilities. Fuzzy generalization is included to address the challenge of demand uncertainty by smoothing data variability, a technique inspired by prior work on uncertainty modeling in supply chains . The OUTL policy, a periodic review inventory system, is selected for its adaptability to dynamic demand, allowing for efficient replenishment while balancing holding and ordering costs. This integrated approach directly addresses the gaps identified in the literature, offering a solution that combines advanced forecasting, uncertainty management, and operational improvement.

1.2.2 Framework Overview

The proposed framework is structured into two interconnected phases. In the first phase, the hybrid boosting ensemble is developed to generate accurate demand forecasts. This involves preprocessing the dataset, applying fuzzy generalization to handle uncertainty, and training the ensemble models sequentially to leverage their complementary strengths. In the second phase, the forecasts are integrated into the OUTL inventory policy to optimize inventory decisions. This integration ensures that the predictive power of the ensemble directly informs operational strategies, aligning stock levels with demand patterns to minimize costs and maintain high service levels. The framework is designed to be iterative, allowing for continuous refinement based on evaluation outcomes.

1.3 Development

The development phase focuses on constructing the proposed artifact: a hybrid boosting ensemble with fuzzy generalization and OUTL policy integration. This process encompasses several key steps: dataset preparation, preprocessing, application of fuzzy generalization, training of the ensemble models, hyperparameter optimization, and integration with the inventory policy. The artifact is implemented using Python, utilizing libraries such as scikit-learn for data preprocessing, LightGBM, CatBoost, and XGBoost for model development, and Matplotlib for visualization purposes. The Rossmann dataset, which includes sales data from 1,115 stores over 942 days, serves as the primary data source, featuring variables such as sales figures, store attributes, temporal factors, and promotional activities.

1.3.1 Dataset and Preprocessing

The Rossmann dataset is prepared by merging sales data with store information based on a common identifier, ensuring a comprehensive dataset for analysis. Temporal features, such as year, month, and day, are extracted from date variables to capture time-related patterns, after which the original date column is removed to streamline the dataset. Preprocessing is a critical step to ensure data quality and suitability for modeling. This includes removing records with zero sales to focus on active sales periods, imputing missing values using appropriate strategies (e.g., mean for numerical features and mode for categorical features), and encoding categorical variables like store type and assortment into numerical formats for compatibility with machine learning algorithms. Feature engineering is also performed, creating additional variables such as lagged sales and promotional interaction terms to capture temporal dependencies and the impact of promotions on demand. To address outliers, sales values are adjusted to fall within a reasonable range, mitigating the influence of extreme values on model performance (**?**). Finally, features are normalized to a uniform scale to ensure consistency across variables during model training.

1.3.2 Fuzzy Generalization

To manage demand uncertainty, a fuzzy generalization technique is applied to the preprocessed dataset. This process involves discretizing each feature into a predefined number of categories, smoothing data variability caused by noise or outliers. The approach generalizes continuous values into discrete categories, introducing a fuzzy-like effect that enhances the model's resilience to uncertainty. For each feature, the range of values is divided into intervals, and each data point is assigned to the nearest representative value based on its proximity to these intervals. The resulting generalized dataset is then standardized to ensure that features have a consistent mean and variance, preparing the data for effective model training. This technique draws on established methods for handling uncertainty in supply chain data, aiming to improve the robustness of demand forecasts in the retail context (Bezdek, 1981).

1.3.3 Hybrid Boosting Ensemble Training

The hybrid boosting ensemble is developed by training the LightGBM, CatBoost, and XGBoost models in a sequential manner on the preprocessed and generalized dataset. The dataset is split into training and testing sets to facilitate model evaluation. The training process proceeds as follows: LightGBM is first trained to generate initial demand predictions, leveraging its efficiency in handling large datasets. CatBoost is then applied to the residuals of LightGBM's predictions, focusing on correcting errors, particularly those associated with categorical features prevalent in the Rossmann dataset. Finally, XGBoost is trained on the residuals of CatBoost, refining the predictions through its regularization capabilities to enhance overall accuracy and generalization. The final forecast is obtained by combining the outputs of all three models, creating a unified prediction that capitalizes on the strengths of each algorithm.

1.3.4 Hyperparameter Optimization

To ensure effective performance, hyperparameters for each model in the ensemble are carefully tuned. This process involves using a grid search approach combined with cross-validation to systematically explore a range of parameter values. For LightGBM, parameters such as the number of trees and learning rate are adjusted to balance model complexity and predictive accuracy. CatBoost's parameters, including tree depth and regularization terms, are tuned to prevent overfitting, particularly on categorical features. XGBoost's parameters, such as maximum depth and regularization strength, are determined to enhance the model's ability to generalize across diverse demand patterns. This tuning process is iterative, ensuring that the ensemble is well-calibrated to the characteristics of the Rossmann dataset, thereby maximizing forecasting accuracy.

1.3.5 Integration with OUTL Policy

The forecasts generated by the hybrid boosting ensemble are integrated into the OUTL inventory policy to optimize inventory management decisions. The OUTL policy, a periodic review system, is designed to adapt to dynamic demand by determining the improved inventory level to which stock should be replenished at each review period. This level accounts for expected demand over the lead time and includes a buffer to handle variability, ensuring a high service level while minimizing costs. The policy calculates key inventory parameters, including the point at which a replenishment order should be placed and the quantity to order, based on the ensemble's demand forecasts. The total cost, encompassing purchase, holding, and ordering expenses, is also considered, aiming to achieve cost efficiency while maintaining adequate stock levels to prevent stockouts (Silver, Pyke & Thomas, 2017). This integration ensures that the predictive accuracy of the ensemble directly translates into operational improvements, aligning inventory decisions with actual demand patterns.

1.3.6 Validation

The artifact is validated to ensure that it meets the intended objectives of accurate demand forecasting and effective inventory improvement. This involves testing the ensemble's predictions on a separate subset of the Rossmann dataset, assessing its ability to capture demand patterns across various stores and time periods. Performance metrics such as average error and relative error are used to evaluate forecasting accuracy, while visualizations like scatter plots and time series comparisons provide insights into the alignment between predicted and actual demand. For inventory improvement, the OUTL policy's outputs—such as safety stock levels and ordering quantities—are reviewed to confirm that they align with operational goals, such as maintaining a high service level while minimizing costs. This validation step ensures that the artifact is robust and reliable before proceeding to a more comprehensive evaluation.

1.4 Evaluation to Confirm the Solution

The evaluation phase rigorously assesses the artifact's effectiveness in achieving the research objectives, focusing on its performance in demand forecasting, inventory improvement, and adaptability to varying conditions. This phase employs a multi-faceted approach to ensure a thorough validation of the proposed framework.

1.4.1 Forecasting Accuracy Evaluation

The forecasting accuracy of the hybrid boosting ensemble is evaluated using a range of performance metrics to assess its predictive power. These metrics include measures of average error, relative error, and the proportion of predictions within a specified threshold of actual values, which are critical for retail applications where precision directly impacts inventory decisions. Visualizations, such as scatter plots comparing predicted and actual sales and time series plots showing demand trends over time, are used to provide a qualitative assessment of the model's performance. These tools help identify how well the ensemble captures demand patterns, such as peaks during promotional periods and troughs during low-demand periods, ensuring that the forecasts are reliable for operational use.

1.4.2 Inventory improvement Outcomes

The impact of the ensemble's forecasts on inventory management is evaluated by examining the outcomes of the OUTL policy. This includes assessing key inventory parameters, such as the safety stock required to handle demand variability, the point at which replenishment orders are triggered, and the quantities ordered to maintain improved stock levels. The evaluation focuses on how these parameters contribute to achieving a balance between maintaining a high service level—ensuring product availability for customers—and minimizing costs, including those associated with purchasing, holding, and ordering inventory. The comparison between the generic ensemble (without fuzzy generalization) and the fuzzy ensemble highlights the added value of uncertainty modeling in optimizing inventory decisions, particularly in reducing overstocking and stockouts.

1.4.3 Sensitivity Analysis

To evaluate the artifact's robustness, a sensitivity analysis is conducted by varying operational parameters, such as lead time, and observing their impact on inventory outcomes. Lead time—the duration between placing an order and receiving it—is a critical factor in retail supply chains, as it affects the timing and quantity of inventory replenishment. The analysis examines how the framework performs across different lead time scenarios, ensuring that it remains effective under varying conditions. This step confirms the artifact's adaptability to real-world operational constraints, addressing the challenge of demand volatility and ensuring that the proposed solution is practical for Rossmann stores and potentially other retail contexts.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we introduces definitions, components, and key concepts foundational to demand forecasting and inventory management, providing a theoretical backdrop for the methodologies applied in this study. Then, we provide a comprehensive overview of key studies that investigate demand forecasting techniques within various supply chain contexts. These studies explore a range of methodologies such as classical (traditional) forecasting methods, single-based machine learning methods, and hybrid ensemble methods to improve forecast accuracy and improve inventory management. Notable works include Ghosh (2020), Jaipuria and Mahapatra (2021), Merkuryeva et al. (2019), and Ali et al. (2015). These studies highlight both the strengths and limitations of various forecasting models in addressing real-world supply chain challenges.

There are also significant studies addressing single-based machine learning and deep learning techniques to enhance demand forecasting accuracy in supply chain management. These approaches leverage structured data, temporal dynamics, and advanced learning architectures to model complex demand behavior. Notable contributions include Kohli et al. (2021), Abbasimehr et al. (2020), Ntakolia et al. (2021), Ul Haq Qureshi et al. (2024), and Ahmed et al. (2024). These studies apply models like linear regression, KNN, LSTM, Random Forest, and boosting algorithms to retail datasets such as Rossmann, showing improved forecasting accuracy over traditional methods. However, they face challenges such as high computational costs, sensitivity to hyperparameters, and limited integration with inventory control. Overall, they highlight the increasing importance of machine learning in enabling smarter, more adaptive supply chain decisions.

Furthermore, some research focuses on applying hybrid machine learning and AI techniques to enhance forecasting accuracy and decision-making in supply chain and retail contexts. Notable contributions include Weng et al. (2020), Abolghasemi et al. (2020), Zhang He and Sun Yu (2020), Mittal (2024), Jamali et al. (2024), Harikrishnan et al. (2025), Alsulamy (2025), and Seyedan et al. (2023). These studies demonstrate powerful performance in modeling complex demand dynamics, integrating lifecycle-sensitive features, and improving inventory or efficiency outcomes across sectors like retail, fashion, construction, and healthcare. However, many of these models still rely heavily on domain-specific tuning, limited operational integration, or lack of generalizability. In contrast, the proposed hybrid boosting ensemble applied to the Rossmann dataset—combining LightGBM, CatBoost, and XGBoost with fuzzy generalization—provides a unified, interpretable, and computationally efficient approach to address multivariate demand uncertainty and improve inventory under the OUTL policy framework.

2.1 Definitions and Key Concepts

This section outlines essential definitions and key concepts central to demand forecasting and inventory management, establishing a theoretical foundation for the methodologies employed in this research. It covers stockouts, service levels, reorder points, OUTL policies, safety stock, inventory cost components, periodic vs. continuous review systems, ABC analysis, and planning horizons. These concepts are essential for understanding the challenges and solutions discussed in the context of Rossmann stores' supply chain management.

2.1.1 Planning Horizon

The planning horizon in demand forecasting defines the time frame for predictions (e.g., daily, weekly, monthly), significantly impacting forecast accuracy and inventory decisions. Short-term horizons, such as the daily predictions used for Rossmann stores, improve performance metrics by leveraging recent, stable data patterns, but may miss longer-term trends (makridakis et al, 2020). Long-term horizons increase uncertainty due to accumulating external factors (e.g., economic shifts), reducing accuracy (hyndman et al, 2021). This study adopts a weekly planning horizon, consistent with the 7-day review period and 7-day lead time, to improve inventory adjustments. This approach ensures precise stock management, minimizing stockouts and overstocking while maintaining high service levels.

2.1.2 ABC Analysis

ABC analysis is a widely used inventory classification technique that categorizes items based on their value and demand to prioritize management efforts (silver et al, 2017). A items (high-value, low-volume) contribute significantly to inventory costs and require tight control, particularlly using continious review. B items (moderate-value, moderate-volume), balance cost and demand variability, as focused on in this study for Rossmann stores. C items (low-value, high-volume) have minimal cost impact and use simpler management.

2.1.3 Service Level in Inventory Management

Service level in inventory management refers to the probability of meeting customer demand without stockouts, serving as a key metric to balance customer satisfaction and inventory costs. Analytical methods for determining service level involve calculating safety stock and reorder points based on demand and lead time variability. The Type I service level, or cycle service level, as used in this study, is achieved by setting safety stock using the safety factor z from the standard normal distribution (e.g., $z = 1.65$ for a 95% service level) multiplied by the standard deviation of demand during lead time (silver et al., 2017). Alternatively, the Type II service level, or fill rate, measures the fraction of demand met immediately, computed as $1 - \frac{E[\text{Shortage}]}{\mu_d}$, where $E[\text{Shortage}]$ is the expected unmet demand per cycle and μ_d is the mean demand (chopra et al, 2021). These methods enable robust inventory policies, as applied in this study by consideration of a service level equal to 95% for Rossmann stores.

2.1.4 Safety Stock Key Concepts

Safety stock is a buffer inventory maintained to mitigate the risk of stockouts due to demand or lead time variability, ensuring high service levels (silver et al, 2017). It is calculated as $SS = z \cdot \sigma_{D_L}$, where z is the safety factor corresponding to the desired service level (e.g., $z = 1.65$ for 95%) and σ_{D_L} is the standard deviation of demand during lead time, often computed as $\sigma_{D_L} = \sqrt{\sigma_d^2 \cdot L + \mu_d^2 \cdot \sigma_L^2}$ for variable demand and lead time. Safety stock determination balances holding costs against stockout risks, with higher z-values increasing inventory costs but reducing stockout probability.

2.1.5 Reorder Point (ROP) and Classical Calculations

The Reorder Point (ROP) is the inventory level at which a replenishment order is triggered to prevent stockouts, ensuring timely reordering to maintain service levels while minimizing excess stock (stevenson, 2015). In classical inventory management, four ROP calculations are used, depending on demand and lead time assumptions: (1) Constant Demand, Constant Lead Time: $ROP = d \cdot L$, where d is the constant demand rate and L is the fixed lead time, ideal for stable products. (2) Constant Demand, Variable Lead Time: $ROP = d \cdot E[L] + z \cdot \sigma_L \cdot d$, accounting for lead time variability (σ_L) with expected lead time $E[L]$. (3) Variable Demand, Constant Lead Time: $ROP = \mu_d \cdot L + z \cdot \sigma_d \cdot \sqrt{L}$, addressing demand variability (σ_d) with fixed lead time, as used in this study with $L = 7$ days. (4) Variable Demand, Variable Lead Time: $ROP = \mu_d \cdot E[L] + z \cdot \sqrt{\sigma_d^2 \cdot E[L] + \mu_d^2 \cdot \sigma_L^2}$, handling both demand and lead time variability for complex supply chains (chopra et al, 2021). These calculations underpin the OUTL policy in this study, enhancing reorder decisions for Rossmann stores.

2.1.6 Classical Order-Up-To-Level Systems and Retail Ordering Practices

The OUTL policy is a cornerstone of inventory control, with three classical variants: (s, S), (R, S), and (R, s, S) systems, alongside the continuous (s, Q) system (silver et al, 2017). The (s, S) system, a continuous review approach, orders up to level S when inventory falls below reorder point s. The (R, s, S) system, adopted in this study, periodically reviews inventory every R periods and orders up to S, suitable for batched data updates like Rossmann's daily sales. The (R, s, S) system combines periodic review (R) with a reorder point (s), ordering up to S only if inventory is below s, offering flexibility for variable demand. As shown in Figure 2.1, at every review period (R), the inventory position is checked. If it is at or below the reorder point (s), an order is placed to raise the inventory level to S. The replenishment then takes place after a lead time (L), restoring the inventory to the target level S. If the position is above s, nothing is done until at least the next review instant. The (s, Q) system orders a fixed quantity Q when inventory drops below s, focusing on order quantity improvement. Retail ordering practices vary: community pharmacies often order multiple items from a single distributor at fixed intervals to reduce transportation costs, while continuous individual item ordering increases logistics costs. Transportation costs are excluded in this study, as the Rossmann dataset lacks such data, focusing instead on inventory costs.

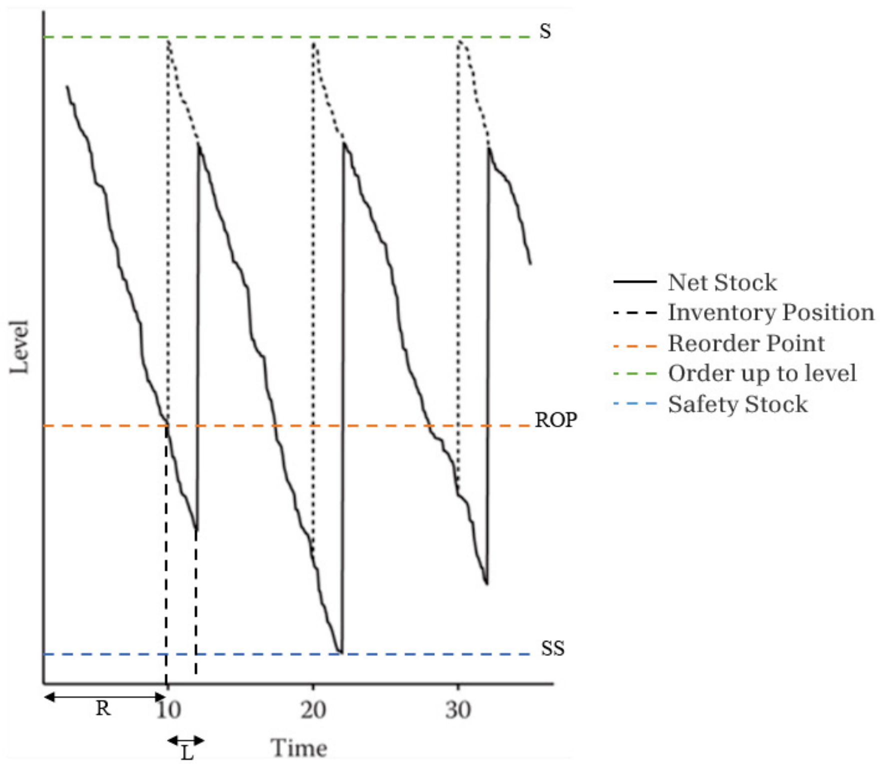

Figure 2.1 OUTL Policy Conceptual Principle
Adapted from Silver, 2017

2.1.7 Periodic vs. Continuous Review Systems

Inventory management systems are classified as periodic or continuous review, each with distinct safety stock calculations and trade-offs (Silver et al, 2017). Periodic Review Systems assess inventory at fixed intervals (R), ordering up to a target level (e.g., OUTL's (R, s, S) system), with safety stock calculated as $SS = z \cdot \sigma_{D_{R+L}}$, where $\sigma_{D_{R+L}} = \sigma_d \cdot \sqrt{R+L}$ accounts for demand variability over the review period plus lead time $(R + L)$. They are simpler to implement and suit batched data updates, as in Rossmann stores, but may increase stockouts due to delayed responses. Continuous Review Systems monitor inventory constantly, ordering when inventory falls below a reorder point (e.g., (s, S) or (s, Q) systems), with safety stock as $SS = z \cdot \sigma_{D_L}$, where $\sigma_{D_L} = \sigma_d \cdot \sqrt{L}$. They minimize stockouts but require real-time tracking, increasing complexity. This study uses periodic review (R, s, S) for its compatibility with daily sales data, improving safety stock for a 7-day lead time (Chopra et al, 2021).

2.1.8 Stockouts

A stockout occurs when inventory is depleted, preventing the fulfillment of customer demand, which can lead to lost sales, reduced customer satisfaction, and reputational damage (silver et al., 2017). The implications of stockouts are significant in retail, as they disrupt service levels and increase operational costs due to expedited reordering or lost revenue. The stockout probability, defined as the likelihood of demand exceeding available inventory during lead time, is expressed as: P(Demand > Inventory) and is often modeled using the cumulative distribution function of demand during lead time (chopra et al, 2021). Safety stock determination mitigates stockouts by maintaining a buffer inventory, calculated as:

$$SS = z \cdot \sigma_{D_L} \tag{2.1}$$

where z is the safety factor corresponding to a desired service level (e.g., $z = 1.65$ for a 95% service level) and σ_{D_L} is the standard deviation of demand during lead time. This section clarifies that while stockouts are a critical concept impacting service levels, they are not directly modeled in this study (e.g., stockout costs are excluded). Instead, the OUTL policy indirectly minimizes stockouts by considering service levels (95%) through safety stock improvement.

2.1.9 Inventory Cost Components and Total Inventory Cost

Inventory cost minimization is critical for efficient supply chain management, encompassing purchase costs (PC), holding costs (HC), and ordering costs (OC) (Ahmed et al, 2024). Purchase costs are calculated as the cost per unit times annual demand, holding costs reflect the expense of storing average inventory (e.g., warehousing, insurance), and ordering costs include fixed costs per order (e.g., administrative, setup). Total inventory cost (TC) is modeled as $TC = PC + HC + OC$, improved in policies like OUTL by aligning order quantities and reorder points with demand forecasts. This study excludes shortage costs, as high service levels (95%) are prioritized through safety stock, and omits quantity rebates due to the Rossmann dataset's uniform cost structure for B items.

2.2 Demand forecasting with traditional techniques

Saumyadip Ghosh (2020) explored demand forecasting in a Moroccan food company using the ARIMA model. The study leverages historical demand data from January 2010 to December 2015 to develop and validate an ARIMA model, selected based on performance criteria such as Akaike Information Criterion (AIC), Schwarz Bayesian Criterion (SBC), maximum likelihood, and standard error. The model aims to provide accurate demand predictions to improve production planning and inventory management for perishable goods, demonstrating its practical utility in a real-world food manufacturing context. However, the ARIMA model rely on the linear and univariate data that limits its ability to handle complex, non-linear and multivariate patterns like those in the Rossmann dataset (e.g., promotions, store-specific features) addressed by the hybrid boosting ensemble in the Rossmann dataset.

Jaipuria and Mahapatra (2021) proposed a hybrid forecasting model combining ARIMA and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) to address heteroskedastic demand patterns within a supply chain context. Utilizing demand data from a cement manufacturing company spanning from April 2006 to March 2013, the study identified ARIMA and GARCH models to forecast mean and variance of cement demand, respectively. This combined approach significantly outperformed the traditional ARIMA method by dynamically adjusting safety stock and order quantities in each replenishment cycle, thereby effectively mitigating the bullwhip effect and net-stock amplification. Despite its enhanced ability to manage variance-driven inventory uncertainties, the ARIMA-GARCH model remains limited by its reliance on linear assumptions, which may restrict its applicability to highly complex, nonlinear, or multivariate demand patterns similar to those addressed by the hybrid boosting ensemble methods employed in the Rossmann dataset study.

Merkuryeva et al. (2019) investigated demand forecasting challenges within the pharmaceutical supply chain context, focusing on a wholesaler-distributor scenario in an emerging market. The study experimentally compared three forecasting approaches: Simple Moving Average (SMA), Multiple Linear Regression (MLR), and Symbolic Regression with Genetic Algorithm (GA). Findings demonstrated that the symbolic regression model significantly outperformed both SMA and MLR by providing more accurate forecasts with the lowest Mean Absolute Deviation (MAD) and improved capability to handle volatile demand patterns. However, symbolic regression methods remain limited by their complexity and requirement for extensive computational resources, potentially restricting ease of use in practical supply chain settings.

Ali et al. (2015) explored the application of Simple Moving Averages (SMA) in supply chains lacking formal information-sharing mechanisms. They introduced a novel approach known as Downstream Demand Inference (DDI), enabling upstream supply chain members to mathematically infer consumer demand from downstream order data without explicit information sharing. Using analytical comparisons, the authors demonstrated that the DDI strategy, when based on SMA, significantly reduces Mean Squared Error (MSE) compared to traditional No Information Sharing (NIS) strategies, particularly in environments characterized by high demand autocorrelation. However, the effectiveness of SMA-based DDI is constrained to demand patterns without strong trends or seasonality, which limits its broader applicability to complex, dynamic markets.

2.3 Demand forecasting with machine learning techniques

Kohli et al. (2021) investigated the use of linear regression and K-Nearest Neighbors (KNN) regression for sales prediction using the Rossmann dataset. The study focused on leveraging past sales, customer counts, promotions, and store features to forecast future sales. While linear regression provided a straightforward model with interpretable coefficients and consistent performance across training and test sets, KNN regression demonstrated flexibility in modeling nonlinear relationships by using similarity-based predictions. However, results indicated that KNN suffered from overfitting, with significantly higher error metrics on test data compared to training data. Additionally, the model's performance was hindered by its sensitivity to the choice of hyperparameters and its computational inefficiency when applied to large datasets.

Abbasimehr et al. (2020) proposed an improved demand forecasting model using a multi-layer Long Short-Term Memory (LSTM) network, tailored to capture the nonlinear and non-stationary patterns in time series sales data. The study applied a grid search method to automatically identify the best LSTM hyperparameter combinations, including lag size, hidden layers, neurons, dropout rate, batch size, and learning rate. Using monthly sales data from a furniture company spanning 2007 to 2017, the model demonstrated superior forecasting accuracy over traditional and computational intelligence models such as ARIMA, simple exponential smoothing (SES), artificial neural networks (ANN), support vector machines (SVM), and single-layer LSTM. The performance was measured using RMSE and SMAPE metrics, with the proposed deep LSTM framework showing significant improvement in predictive reliability. However, while this model effectively captures long-range temporal dependencies and outperforms baseline methods on single-product datasets, it relies on extensive training time and computational tuning.

Ntakolia et al. (2021) developed an explainable machine learning framework for predicting material backorders in inventory management, aiming to enhance supply chain efficiency by minimizing costs related to stockouts and unmet demand. The study evaluated several binary classification models including Random Forest (RF), XGBoost, LightGBM (LGBM), and Balanced Bagging (BB), using historical inventory data and SHAP values to identify influential features contributing to backorder risk. Among the tested models, RF, XGB, and BB achieved strong performance with AUC scores of 0.95, while the best-performing model was LGBM combined with Isotonic Regression for probability calibration. The findings highlighted that inventory risk factors such as the item's historical performance, future demand signals, and lead time were critical to backorder prediction.

Ul Haq Qureshi et al. (2024) proposed a weather-enhanced deep learning model to improve demand forecasting accuracy for Rossmann dataset. The model utilized a multivariate dataset, incorporating weather variables, alongside internal factors like promotional activities, store location, and holidays, to capture their combined influence on sales demand. Unlike traditional approaches relying on univariate sales data, the study employed a Gated Recurrent Unit (GRU) architecture improved with Grid Search, rather than a standard Long Short-Term Memory (LSTM) model, to better model temporal dependencies and non-linear relationships in the data. The GRU-based model demonstrated superior performance over an LSTM baseline, achieving lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) on both training and test datasets. Notably, the model's accuracy improved further when applied to weather-specific data subsets (normal, rain, and snow), highlighting the critical role of meteorological variables in demand patterns. However, the approach focused solely on deep learning for forecasting and did not incorporate ensemble techniques or explicit inventory optimization strategies.

Ahmed et al. (2024) proposed a Switching-Based Forecasting Approach (SBFA) that dynamically selects the most suitable model among a pool of advanced ML/DL algorithms to enhance sales forecasting accuracy in supply chains. The framework combines decision tree-based models (e.g., XGBoost, LightGBM, Random Forest, KNN) and deep learning models (e.g., LSTM, RNN, GRU, CNN), switching between them based on validation performance for different forecast horizons. The SBFA was evaluated using Kaggle (kaggle,2018) and Rossmann sales datasets and integrated into a two-echelon inventory model, consists of a retailor and distributor(supplier), to simulate real-world replenishment decisions. The Kaggle dataset includes diverse, publicly available sales data often used for competitions, featuring varied features like temporal, promotional, and categorical variables (e.g., 'month', 'week', 'day') from multiple sources, making it more general and heterogeneous. In contrast, the Rossmann dataset is specific to Rossmann stores, focusing on retail sales data with features like sales, temporal data, and promotions, tailored for single-echelon retail analysis but adaptable for multi-echelon systems, offering a more structured and retail-specific scope. Results showed that SBFA consistently achieved lower RMSE and reduced total supply chain costs compared to traditional methods and individual ML/DL models, especially under variable lead time and safety stock scenarios. However, while SBFA offers high accuracy and flexibility across prediction horizons, it incurs significant computational costs due to the ensemble training and validation loop.

2.4 Hybrid machine learning techniques

Weng et al. (2020) developed a hybrid forecasting model combining Light Gradient Boosting Machine (LightGBM) and Long Short-Term Memory (LSTM) networks to improve supply chain sales prediction. The model was evaluated using three real-world retail datasets—including the Rossmann and two supply chain sales datasets—demonstrating high accuracy, efficiency, and interpretability. Accuracy and interpretability were assessed using the Normalized Weighted Root Mean Squared Logarithmic Error (NWRMSLE), and Efficiency was measured by prediction time, recorded as the average time (in seconds). The LSTM component was used to extract high-level temporal features (day, month, and week), while LightGBM handled structured input and feature importance ranking, effectively managing missing values and diverse categorical feature (StoreType, Assortment, StateHoliday). The ensemble outperformed traditional statistical and standalone machine learning models in forecasting accuracy, particularly under complex, nonlinear, and noisy demand environments. However, while this LightGBM-LSTM integration efficiently captures both sequential dependencies and interpretable features, it remains dependent on meticulous feature engineering and model tuning. In contrast, the proposed approach addresses this limitation by applying fuzzy generalization as a clustering method, which simplifies feature representation and reduces the need for manual tuning. This abstraction enables the ensemble model to generalize better under uncertain and dynamic demand conditions.

Abolghasemi et al. (2020) proposed a hybrid demand forecasting model tailored to address demand volatility in the presence of promotional effects in supply chains. Using a large dataset from a fast-moving consumer goods (FMCG) company, the study categorized 843 demand time series based on Coefficient of Variation (CoV) to evaluate model performance under varying volatility levels. The authors decomposed demand into baseline and promotional components and introduced a hybrid forecasting framework combining ARIMA for baseline demand and piecewise regression for promotional uplifts. The hybrid model demonstrated robust forecasting accuracy and superior inventory performance, especially for high-CoV series. In comparative evaluation, the hybrid model outperformed several statistical and machine learning models including ARIMAX, dynamic linear regression (DLR), support vector regression (SVR), and artificial neural networks (ANN), particularly in terms of MASE and inventory cost metrics. However, while this approach effectively captured promotion-driven volatility in structured sales data, it relied heavily on price as the sole explanatory variable.

He, Z. et al. (2020) (2020) proposed a hybrid forecasting model combining LightGBM and Long Short-Term Memory (LSTM) networks to improve short-term vegetable sales prediction in a supermarket context. The model utilizes LightGBM to extract structured features such as weather, holidays, and product types, while LSTM captures complex nonlinear temporal patterns. The two models are integrated using an error-reciprocal weighted average to balance prediction strengths. Experimental results across six vegetable types—including potatoes, tomatoes, and cucumbers—showed that the hybrid LightGBM-LSTM model significantly outperformed single models in terms of Mean Absolute Percentage Error (MAPE), achieving an average MAPE of 0.161. SHAP analysis further revealed that weather and calendar effects were among the most influential features.

Mittal (2024) introduced an AI-driven hybrid forecasting framework to improve demand prediction accuracy in volatile fashion retail supply chains. The study employed fuzzy clustering to categorize 10,000 products based on lifecycle stage and demand variability and then developed tailored models for each product group. Two advanced forecasting models—Temporal Fusion Transformer (TFT) and a Deep Artificial Neural Network improved via a genetic algorithm—were trained using five years of retail sales data. Among the evaluated models, the GA-improved ANN consistently achieved the lowest MAPE (as low as 3.1%) and outperformed TFT across all variability clusters. This approach effectively incorporated lifecycle-sensitive demand profiles, seasonal effects, markdowns, and external covariates, offering enhanced forecast precision and actionable insights for inventory planning.

Jamali et al. (2024) proposed a hybrid artificial intelligence (AI) framework integrating fuzzy data envelopment analysis (FDEA) and machine learning (ML) to optimize macro-ergonomic factors in the pharmaceutical supply chain (Ph.SC). The study focused on identifying and mitigating inefficiencies—such as communication delays between physicians and patients—by evaluating 20 decision-making units (DMUs) across an urban hospital network. Key input variables included cost, time, quality, and personnel, while output indicators encompassed availability, reliability, teamwork, and communication. By combining FDEA with linear regression, the model enabled predictive analysis of performance and improved target setting for healthcare managers. The hybrid approach successfully addressed uncertainty in data, highlighted critical macro-ergonomic indicators (like information sharing and communication), and achieved enhanced decision-making reliability. However, this study focused on efficiency evaluation in a healthcare context, without incorporating multivariate demand forecasting or inventory-specific operational strategies.

Harikrishnan et al. (2025) proposed a hybrid forecasting approach leveraging XGBoost and ANN to enhance short-term residential load forecasting accuracy. The study utilized residential power consumption data, demonstrating that XGBoost significantly outperformed traditional forecasting methods such as Support Vector Machines (SVM), ANN alone, and Convolutional Neural Networks (CNN) due to its capacity to efficiently process large datasets and optimize learning via parallel processing and regularization. Further improvement was achieved by integrating XGBoost with ANN, resulting in an acceptable MAPE of 9.68%, surpassing single-model baselines. However, the study acknowledged critical limitations of XGBoost, including high sensitivity to hyperparameter tuning—where inappropriate settings for learning rate, tree depth, and regularization could cause overfitting or underfitting, thereby affecting the model's generalizability.

Alsulamy (2025) conducted a comparative analysis of three advanced machine learning algorithms—CatBoost, XGBoost, and Light Gradient Boosting Machine (LGBM)—to predict construction delay risks in Saudi Arabian projects. The study applied classification techniques to assess delay categories—minor, moderate, and major—based on performance metrics such as accuracy, precision, sensitivity, specificity, false positive rate (FPR), and false negative rate (FNR). Among the models tested, LGBM achieved the highest accuracy in predicting delay classes, demonstrating its strong capability in handling categorical data and imbalanced datasets, and its suitability for real-world project management scenarios.

Seyedan et al. (2023) developed an ensemble deep learning framework for multivariate time-series demand forecasting, integrated with an OUTL inventory policy to optimize safety stock and minimize total inventory costs. The study combined three heterogeneous deep learning models—Multilayer Perceptron (MLP), LSTM, and 1D Convolutional Neural Network (1D-CNN)—to capture global, temporal, and local features of demand data. The ensemble model, structured using a meta-learner (MLP), was tested on two real-world retail datasets (sports and electronics), showing significant improvement in prediction accuracy over individual base models, with MAPE of 5.22% and 9.58% (reductions of 22.3% and 24.4%) for sports and electronic products, respectively. The model also led to reduced total inventory costs under varying lead time conditions. However, while this study demonstrates strong forecasting and inventory control performance under a structured OUTL framework, it requires high computational resources and model complexity.

2.5 Summary of Literature review (Strengths and Limitations)

The literature review highlights key strengths and limitations of demand forecasting and inventory management techniques, providing a foundation for adopting a hybrid boosting ensemble approach integrated with the OUTL policy for Rossmann stores.

2.5.1 Strengths

Traditional Methods: Techniques like ARIMA and MA offer simplicity and interpretability, effectively capturing linear trends and stable demand patterns (Ghosh, 2020; Ali et al., 2015). They are computationally efficient for univariate datasets, making them suitable for small-scale inventory planning.

Machine Learning Models: Single-based ML models (e.g., LightGBM, XGBoost, LSTM) excel in handling multivariate data, capturing non-linear patterns, and improving forecasting accuracy over traditional methods (Kohli et al., 2021; Ntakolia et al., 2021). For instance, LightGBM achieved an AUC of 0.95 for backorder prediction.

Hybrid and Ensemble Approaches: Hybrid models combining ML techniques (e.g., LightGBM-LSTM, ARIMA-piecewise regression) enhance robustness against demand volatility and achieve lower error metrics (e.g., MAPE reduced by 22.3% in Seyedan et al., 2023; Weng et al., 2020). They leverage complementary strengths, such as LightGBM's efficiency and LSTM's temporal modeling. Inventory Integration: Advanced models integrated with inventory policies like OUTL optimize reorder points and safety stock, reducing costs and stockouts while maintaining high service levels (Seyedan et al., 2023). Incorporating exogenous variables (e.g., weather) further refines predictions (Ul Haq Qureshi et al., 2024).

2.5.2 Limitations

Traditional Methods: ARIMA and MA struggle with multivariate, non-linear demand patterns and are limited to stable, univariate scenarios, leading to higher errors in volatile retail contexts (Ghosh, 2020; Jaipuria et al., 2021).

Single-Based ML Models: Models like KNN and LSTM face challenges such as overfitting, high computational costs, and sensitivity to hyperparameters, limiting their scalability and integration with inventory systems (Kohli et al., 2021; Abbasimehr et al., 2020).

Hybrid Models: While effective, hybrid approaches often require extensive feature engineering, domain-specific tuning, and significant computational resources, posing challenges for real-time applications (Weng et al., 2020; Ahmed et al., 2024).

This analysis underscores the need for a hybrid boosting ensemble approach, as proposed in this study, which combines LightGBM, CatBoost, and XGBoost with fuzzy generalization to address the limitations of single-model and traditional methods. By leveraging the strengths of sequential learning and uncertainty modeling, and integrating with the OUTL policy, the proposed framework enhances forecasting accuracy and inventory improvement for Rossmann stores, offering a robust solution for retail supply chain challenges.

Table 2.1 All References from Literature Review

Reference (Author, Year)	Research Problem	Methodology	Results	Proposed Future Work
Ghosh (2020)	Food demand forecasting	ARIMA	Improved inventory	Incorporate multivariate data
Jaipuria et al. (2021)	Cement demand forecasting	ARIMA-GARCH	Reduced bullwhip effect	Apply to retail contexts
Merkuryeva et al. (2019)	Pharma demand forecasting	SMA, MLR, Symbolic Reg	Lowest MAD with regression	Simplify regression models
Ali et al. (2015)	Supply chain demand	SMA with DDI	Lower MSE than no sharing	Test in dynamic markets
Kohli et al. (2021)	Rossmann sales prediction	Linear regression, KNN	Linear interpretable; KNN overfit	Use ensemble methods
Abbasimehr et al. (2020)	Furniture demand forecasting	Multi-layer LSTM	Outperformed ARIMA	Reduce computational costs
Ntakolia et al. (2021)	Backorder prediction	RF, XGBoost, LightGBM	LightGBM AUC 0.95	Extend to demand forecasting
Ul Haq Qureshi et al. (2024)	Rossmann demand forecasting	GRU with weather data	Lower MAE than LSTM	Integrate with inventory policies
Ahmed et al. (2024)	Supply chain sales forecasting	SBFA with ML/DL	Lower RMSE, reduced costs	Optimize computational efficiency
Weng et al. (2020)	Retail sales forecasting	LightGBM-LSTM	High accuracy	Minimize feature engineering
Abolghasemi et al. (2020)	FMCG demand forecasting	ARIMA-piecewise regression	Robust for high-CoV	Include additional inputs
He & Yu (2020)	Vegetable sales prediction	LightGBM-LSTM	MAPE 0.161	Link to inventory optimization
Mittal (2024)	Fashion demand prediction	Fuzzy clustering, ANN	ANN MAPE 3.1%	Incorporate boosting techniques
Jamali et al. (2024)	Pharma ergonomics optimization	Fuzzy DEA, regression	Improved efficiency	Extend to demand forecasting
Harikrishnan et al. (2025)	Load forecasting	XGBoost-ANN	MAPE 9.68%	Test in retail supply chains
Alsulamy (2025)	Construction delay prediction	CatBoost, XGBoost, LightGBM	LightGBM high accuracy	Apply to continuous forecasting
Seyedan et al. (2023)	Demand forecasting, inventory	MLP, LSTM, CNN	MAPE reduced 22.3%	Simplify models, test multi-echelon

Table 2.2 Performance Metrics for Forecasting and Supply Chain Management

Metric	Formula	Description	Application	Reference
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	average absolute difference between predicted and actual values	ideal for evaluating forecasting models when a clear, interpretable measure of average error is needed	Qureshi et al. (2024)
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	square root of the average squared prediction errors	ideal for forecasting where large errors are costly	Ahmed et al. (2024)
MAPE	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert \times 100$	average absolute percentage difference between predicted and actual values	effective for comparing forecasting accuracy across different scales	He et al. (2020)
SMAPE	$\frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i-\hat{y}_i\rvert}{\lvert y_i\rvert+\lvert\hat{y}_i\rvert} \times 100$	percentage error between predicted and actual values, adjusted for scale	used in forecasting to balance errors and reduce impact of outliers	Abbasimehr et al. (2020)
MASE	$\frac{\text{MAE}}{\frac{1}{n-1}\sum_{i=2}^{n}\lvert y_i-y_{i-1}\rvert}$	compares model error to a naive benchmark and is scale-independent	useful for comparing forecasting models across different scales	Abolghasemi et al. (2020)
accuracy	$\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$	proportion of correct predictions out of the total number of predictions	suitable for initial evaluation of classification models in balanced datasets	Ntakolia et al. (2021)
precision	$\frac{\text{TP}}{\text{TP+FP}}$	proportion of true positive predictions out of all positive predictions made	focuses on positive prediction quality in imbalanced datasets	Ntakolia et al. (2021)
FPR	$\frac{\text{FP}}{\text{FP+TN}}$	proportion of negative cases incorrectly classified as positive	critical for applications where incorrect positive predictions are costly	Ntakolia et al. (2021)
NWRMSLE	$\sqrt{\frac{\sum_{i=1}^{n} w_i(\ln(1+y_i)-\ln(1+\hat{y}_i))^2}{\sum_{i=1}^{n} w_i}}$	log prediction error with weights	suitable for evaluating models on skewed data	Weng et al. (2020)

Table 2.3 Performance metrics used in the literature review, with additional notes

Paper	MAE	RMSE	MAPE	SHAP	Accuracy	Notes
Kohli et al. (2021)		✓	✓			
Abbasimehr et al. (2020)		✓				SMAPE
Ntakolia et al. (2021)				✓	✓	Precision
Ul Haq Qureshi et al. (2024)	✓	✓				
Ahmed et al. (2024)		✓				
Weng et al. (2020)						NWRMSLE
Abolghasemi et al. (2020)						MASE
He et al. (2020)			✓	✓		
Mittal (2024)						
Jamali et al. (2024)						
Harikrishnan et al. (2025)		✓				
Alsulamy (2025)					✓	
Seyedan et al. (2023)		✓				

CHAPTER 3

SUGGESTED SOLUTION AND DEVELOPMENT

This study focuses on developing a robust demand forecasting and inventory improvement framework tailored to retail supply chain environments, with a specific application to Rossmann stores. The key challenges addressed in this research include demand volatility, data uncertainty, and the difficulty of integrating accurate forecasts into inventory policies. These challenges significantly impact the efficiency and cost-effectiveness of supply chain operations.

To overcome these issues, the proposed framework integrates a hybrid boosting ensemble model with the OUTL inventory policy. This approach leverages the combined predictive power of LightGBM, CatBoost, and XGBoost models to generate highly accurate demand forecasts, which are then used to compute improved inventory parameters—reorder points, safety stock, and order-up-to levels. The ultimate goal is to reduce total inventory costs while maintaining high service levels across a multi-store retail network.

This study proposes a single-echelon model focused on Rossmann stores, streamlining the supply chain by isolating retail operations and excluding upstream elements. The framework links demand forecasting with inventory management with two key modules. The Demand Forecasting Module processes historical sales, store features, temporal, and promotional data to produce forecasts using a hybrid ensemble of boosting algorithms. This phase also includes fuzzy generalization to manage data-level uncertainty. To validate forecast accuracy, the effectiveness of the model is evaluated using standard performance metrics such as RMSE, MAE, MAPE, Pred(x=10%). The Inventory Management Module takes the forecast output to dynamically adjust inventory decisions using the OUTL policy. This ensures improved stock replenishment that responds effectively to minimizes excess inventory, and avoids stockouts.

3.1 System model

In this research, a single-echelon retailer system model is proposed to enhance supply chain efficiency by focusing on the retail level, specifically targeting the operational dynamics of Rossmann stores. As illustrated in Figure (3.1), the system model isolates the retail layer and its direct interactions with customers, simplifying the supply chain by excluding upstream echelons such as suppliers, thereby confirming its single-echelon structure. This retailer-centric approach assumes that cumulative demand is forecasted across all stores using a hybrid boosting ensemble model, enabling centralized planning of order quantities, reorder points, and safety stock to minimize total costs. This cumulative approach avoids complexities associated with multi-echelon systems, such as transportation coordination or the bullwhip effect, and does not involve coordinating replenishments for items sharing a common supplier or transportation mode. Instead, orders are aggregated across items to reduce ordering costs, with demand forecasted independently for each item. The proposed framework integrates accurate demand forecasting with effective inventory management practices by improving decision variables (reorder point, safety stock, order-up-to-level) to achieve the objectives of minimizing total cost , as shown in Figure 3.2, comprising two main interconnected stages: the Demand Forecasting stage and the Inventory Management stage. Stage 1, the Demand Forecasting Module, employs an ensemble boosting ML model to predict demand, leveraging historical sales data and external features like weather conditions. Stage 2, the Inventory Management Module, implements the OUTL inventory policy, determining decision variables such as ROP, SS, OUTL, and order quantity (Q) to achieve the objective functions of minimizing total cost and maximizing profit, while also improving service levels, as supported by Seyedan et al. (2023). This two-stage architecture ensures that demand forecasts directly inform inventory decisions, with each retailer operating independently without communication, inventory sharing, or mutual assistance in responding to unanticipated demand, enhancing service levels and cost efficiency at the retail echelon.

Figure 3.1 Single-echelon retailer system model

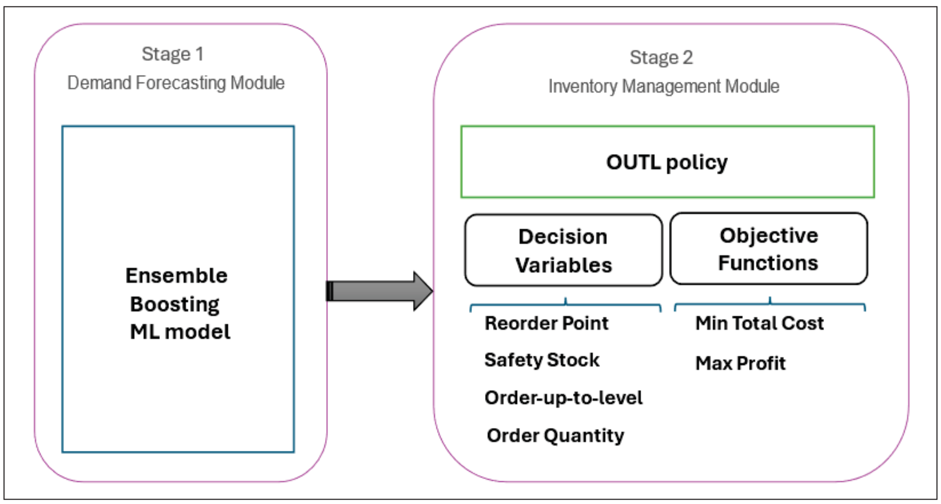

Figure 3.2 Two stage data-driven inventory improvement process

3.1.1 Demand Forecasting Module

The Demand Forecasting Module is responsible for predicting daily product demand utilizing historical data. This module applies single-based machine learning models such as LightGBM, Catboost, XGboost so that combined as hybrid ensemble boosting forecasting techniques to account for demand fluctuations influenced by seasonal trends, market volatility, and consumer behavior changes. The accuracy of these forecasts directly impacts the reliability of the inventory management decisions.

3.1.2 Inventory Management Module

The Inventory Management Module utilizes forecasts generated by the Demand Forecasting Module to determine critical inventory parameters such as reorder points, safety stock levels, and order-up-to levels based on the OUTL policy. This policy periodically assesses the inventory levels and initiates replenishment orders up to a predetermined target level, facilitating the proactive management of inventory. Moreover, predictions will be inserted as an input for the inventory improvement phase to minimize the total inventory cost.

3.1.2.1 OUTL Policy Concept

In the context of retail inventory management, the OUTL policy sets a variable order quantity (Q) at regular intervals (R) to maintain a desired stock level (S) (Ivanov (2021)). This policy minimizes inventory holding costs by avoiding excessive stockpiling while ensuring adequate supply to meet demand within specified fixed lead times (L). The lead time (L) is assumed to be fixed, and there is no relationship between the demand forecast and the replenishment lead time. This means that the demand forecasts generated by the hybrid boosting ensemble are independent of lead time variations, as the model does not dynamically adjust lead time based on forecasted demand. However, it requires careful balancing to prevent stockouts and manage ordering costs efficiently. The OUTL inventory policy employed in this system accounts for several critical factors:

- Demand variability: It integrates forecasted demand uncertainty by computing safety stock levels to prevent stockouts.
- Lead Time: Incorporates anticipated lead times for replenishment, addressing the delay between order placement and delivery.
- Service Level: Defines targeted customer service levels, determining acceptable probabilities of stockouts and setting corresponding safety stock and reorder points.

3.1.2.2 Decision Variables

The following expressions and related equations guide the inventory replenishment decisions.

Safety Stock (SS): Safety stock refers to the additional inventory held by a business to mitigate the risk of stockouts due to uncertainties in demand and supply. It acts as a buffer, ensuring continuity in operations even when demand forecasts are inaccurate, supplier deliveries are delayed, or unexpected fluctuations occur. Properly calculated safety stock helps maintain service levels by balancing the cost of holding extra inventory against the potential revenue loss and customer dissatisfaction caused by stockouts. As it is evident in equation (2.1), safety stock is determined by analyzing historical data, demand variability, lead time, and review period. In this model, the lead time (L) is assumed to be fixed, and there is no lead time variability considered. Therefore, the safety stock calculation focuses solely on demand variability over the review period (R) and lead time (L), contributing to supply chain resilience and customer satisfaction.

$$SS = z \cdot \sigma\sqrt{R + L} \tag{3.1}$$

Where:

- σ = Standard deviation of predicted demand
- L = Average lead time
- R = Fixed time review period
- z = A number of standard deviations above the average demand corresponding to the service level probability

To derive the 95% service level, an acceptable stockout risk ($\alpha = 0.05$) is set, and the z-value is found where the area under the normal distribution curve to the left of z equals $1 - \alpha$ (e.g., $z = 1.65$).

Reorder Point (ROP):

The ROP is the inventory level at which a new order must be placed to replenish stock before it runs out. It represents a critical threshold that triggers replenishment actions, ensuring continuous availability of products without interruption. Obviously, the reorder point calculation incorporates average demand during lead time, review period and safety stock to account for demand variability and potential delays from suppliers. Specifically, the reorder point formula is expressed as equation (2.2).

$$r = d(R + L) + Z \cdot \sigma\sqrt{R + L} \tag{3.2}$$

Where, d =average forecasted demand.

Properly determining the reorder point helps companies maintain good inventory levels, reducing the risk of stockouts while minimizing excess inventory costs. Effective reorder point management enhances operational efficiency, reduces holding costs, and improves customer satisfaction by ensuring products are consistently available when needed.

3.1.2.3 Order-Up-To Level (S):

The Order-Up-To Level (OUTL) is an inventory management policy that sets a maximum target inventory level to be reached each time a replenishment order is placed. Under this policy, the order quantity (Q) is determined by the difference between the predefined order-up-to level S and current inventory position r which their relation is expressed in equation (2.3). Specifically, the OUTL method helps organizations maintain sufficient inventory to cover expected demand over a designated replenishment period, while also incorporating safety stock to mitigate uncertainties such as demand fluctuations and supply delays. By using the order-up-to strategy, businesses regularly review their inventory levels—typically at fixed intervals—and adjust replenishment orders, accordingly, ensuring availability and minimizing excessive inventory holding costs. This policy is particularly beneficial in scenarios characterized by variable demand patterns, as it provides flexibility and responsiveness to changing conditions within the supply chain.

$$S = r + Q \tag{3.3}$$

3.1.2.4 Total Cost (TC):

In the proposed inventory management model, the total cost is a comprehensive metric that includes three key components: purchase costs, ordering costs and holding costs. Purchase costs account for the expense of acquiring the inventory needed to meet the predicted demand over a given period, typically a year, based on the annual predicted demand and the cost per unit of inventory. Ordering costs reflect the expenses incurred each time a replenishment order is placed, influenced by the frequency of orders. Holding costs represent the expense of maintaining inventory in storage, determined by the average amount of inventory held, including additional safety stock to guard against uncertainty.

The following are the assumptions regarding associated costs adopted from Ahmed et al. (2024):

- Purchasing cost (PC): The purchasing cost per unit product is known and fixed (PC = average predicted demand * fixed unit (10 $)). It is assumed a constant per-unit purchasing cost, with no quantity discounts, and PC is included in TC for completeness but does not affect order quantity improvement.
- Ordering Cost (OC): It is assumed to be known, fixed and independent of order quantity.
- Holding Cost (HC): It increases or decreases linearly with the amount of inventory in stock.

$$TC = \sum (PC + OC + HC) \tag{3.4}$$

3.2 Suggested Solution

Inspired by the Design Science Research (DSR) methodology steps, particularly the Development and Evaluation phases, this section presents the suggested framework for demand forecasting, including the dataset description, preprocessing steps, and ML methods used in this study. The framework is designed to systematically develop and evaluate a hybrid boosting ensemble for accurate demand predictions, seamlessly integrating with the OUTL inventory policy to improve supply chain operations for Rossmann stores.

3.2.1 Overview of suggested Framework

The overview of the suggested framework, illustrated in Figure (3.3), is a machine learning-based model for forecasting demand to improve inventory management. This study does not focus on formal optimization or concepts like "distance to the optimal," as the goal is practical improvement rather than optimality. The model leverages a boosting ensemble algorithm, trained on multivariate data that includes store-related, sales-related, temporal, and interaction features, along with promotional period data. The framework begins with data reading and integration, where multiple data sources are combined to form a unified dataset. This dataset undergoes several preprocessing steps, including removing zero sales, handling missing values, converting categorical data to numeric, feature selection and engineering, and normalization. The preprocessed dataset is then split into training (80%) and testing (20%) sets, with 5-fold cross-validation applied during training. A boosting ensemble algorithm is trained with sequential training to improve performance, and the model is evaluated using RMSE, MAE, MAPE, Pred(x=10%). In the last step, demand predictions obtained from machine learning methods are applied to an inventory management model to determine the safety stock and reordering point contributing to calculate the total cost in the inventory model.

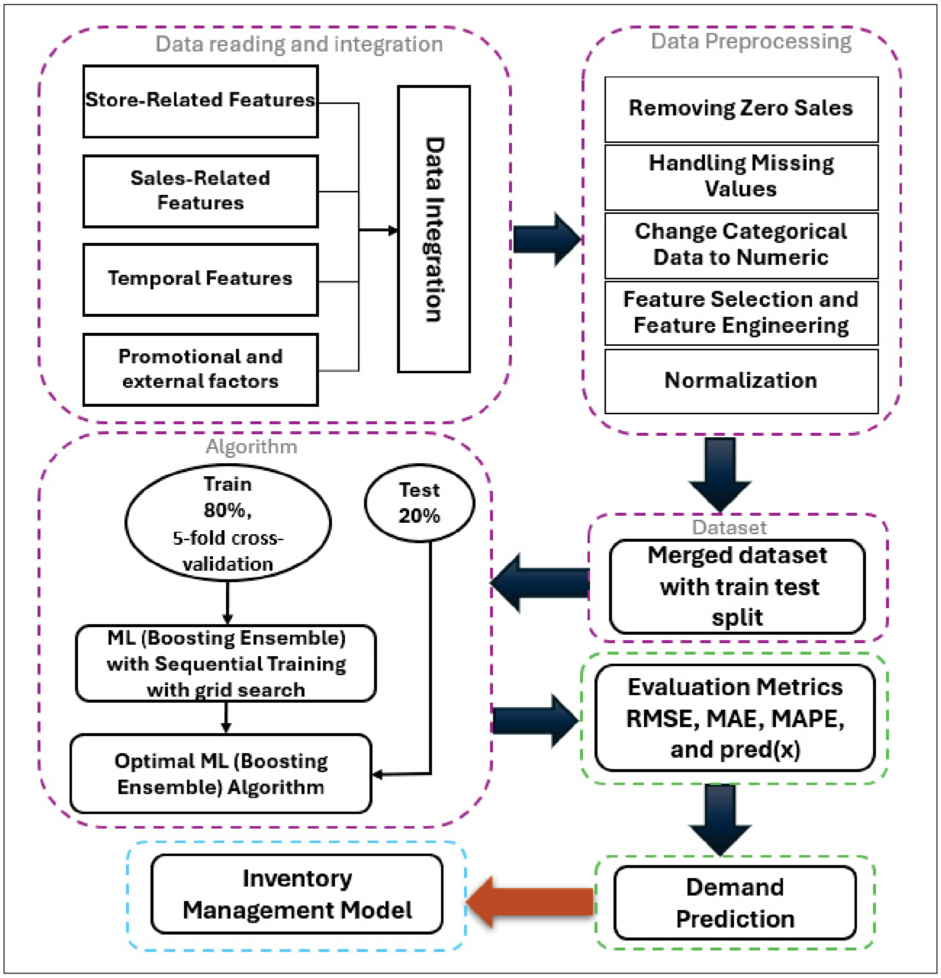

Figure 3.3 Overview of suggested Framework

3.2.2 Dataset

3.2.2.1 Data Description

The dataset utilized in this study is the Rossmann grocery and drug store dataset, encompassing multivariate data critical for demand forecasting in inventory management across seven European countries, with Germany having the largest presence (Kaggle, (2018)). This dataset was selected due to its widespread use in the literature, as noted in several reviewed articles, making it a relevant and recognized benchmark for evaluating forecasting and inventory optimization models in retail settings. It comprises a wide range of variables influencing store sales, including store-related features (e.g., store size and type), sales-related features (e.g., historical sales data, number of customers visiting on average), temporal features (e.g., day, month, year, and holidays derived from the date column), and interaction/aggregate features (e.g., sales per store, average sales over a period). Additionally, the dataset includes promotional period data to assess the impact of marketing campaigns on sales, as well as supplementary information such as the assortment level, product state, school holidays, etc. Spanning from January 2013 to July 2015, the dataset covers 1115 Rossmann stores over 942 days, resulting in a total data size of 1,017,209 rows and 19 columns. This comprehensive dataset, which has been merged into a single data frame, is designed to capture the complex relationships between these diverse variables and sales outcomes, enabling robust demand forecasting models for inventory management. In this dataset, we assumed the sales column represents the quantity of sales. All items are categorized as B items in the ABC analysis, and they consist of single items, each priced at $10. For further clarification, a sample of the dataset is illustrated in Figure 3.4.

	A	B	C	D	E	F	G	H	I
1	Store	DayOfWeek	Date	Sales(items)	Customers	Open	Promo	StateHoliday	SchoolHoliday
2	1	5	2015-07-31	5263	555	1	1	0	1
3	2	5	2015-07-31	6064	625	1	1	0	1
4	3	5	2015-07-31	8314	821	1	1	0	1
5	4	5	2015-07-31	13995	1498	1	1	0	1
6	5	5	2015-07-31	4822	559	1	1	0	1
7	6	5	2015-07-31	5651	589	1	1	0	1
8	7	5	2015-07-31	15344	1414	1	1	0	1
9	8	5	2015-07-31	8492	833	1	1	0	1
10	9	5	2015-07-31	8565	687	1	1	0	1
11	10	5	2015-07-31	7185	681	1	1	0	1
12	11	5	2015-07-31	10457	1236	1	1	0	1
13	12	5	2015-07-31	8959	962	1	1	0	1
14	13	5	2015-07-31	8821	568	1	1	0	0
15	14	5	2015-07-31	6544	710	1	1	0	1
16	15	5	2015-07-31	9191	766	1	1	0	1
17	16	5	2015-07-31	10231	979	1	1	0	1
18	17	5	2015-07-31	8430	946	1	1	0	1
19	18	5	2015-07-31	10071	936	1	1	0	1

Figure 3.4 Sample of Rossmann Dataset
Adapted from Kaggle, 2018

3.2.2.2 Data reading and integration

The data from various sources are read and integrated into a single dataset. Store-related, sales-related, temporal, and promotional data are merged using common identifiers such as store number and date. This integration ensures that all relevant features are aligned for each record, creating a unified dataset for further processing.

3.2.3 Data Preprocessing

3.2.3.1 Removing zero sales

To ensure the dataset focuses on active sales periods, records with zero sales are removed. This step prevents the model from being skewed by non-representative data, as zero sales may indicate store closures or other anomalies unrelated to typical demand patterns. In the Rossmann dataset, zero sales values correspond to days when stores are closed, identifiable via the "Open" column (where Open=0 indicates closure). These records are removed by filtering out entries where Open=0 to ensure the dataset focuses on active sales periods.

3.2.3.2 Handling missing values

Missing values in the dataset are addressed to ensure data quality. The mean imputation technique is used to replace missing values with the average of the respective feature, as this approach preserves the overall distribution of the data. For categorical variables, missing values are filled with the most frequent category (mode) to maintain consistency. For example, if `StoreType` has missing entries, with mode a (602/1,115 stores), missing values are set to a, using $C_i = \text{Mode}(C) = \arg\max_c \text{Count}(c)$, where C_i is the imputed value and c is each category (e.g., a, b, c, d). This ensures consistency in the dataset for the hybrid boosting ensemble.

3.2.3.3 Change categorical data to numeric

Categorical variables, such as store type or region, are converted into numeric formats by encoding categories (e.g., ’a’, ’b’, ’c’ for StoreType) into integers (e.g., 0, 1, 2) to make them compatible with the machine learning algorithm. Morevoer, integer encoding is applied, where each category of a variable is represented by a unique integer starting from zero up to the total number of unique categories. This ensures that the model can interpret categorical data effectively.

3.2.3.4 Feature engineering and selection

The dataset is normalized to scale all features to a standard range, typically between [0,1]. Normalization ensures that features with larger ranges (e.g., sales values) do not disproportionately influence the model compared to features with smaller ranges (e.g., number of customers). The Min-Max Scaling technique is applied to normalize the dataset, as shown in the equations below (Han et al., 2012):

$$x_{std} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{3.5}$$

$$x_{\text{scaled}} = x_{\text{std}} \times (\max - \min) + \min \tag{3.6}$$

$$x_{\text{scaled}} = x_{\text{std}} \times (\max - \min) + \min \tag{3.7}$$

Here, x_{std} is the standardized value, and x_{scaled} is the final scaled value within the desired range. In this study, the range is set to $[0, 1]$.

The list of features selected and engineered is shown in the following table:

Table 3.1 Feature Categories Used in Training the Hybrid Ensemble Model

Category	**Features**
Store-related features	Store, StoreType, Assortment, etc.
Sales-related features	Sales, MeanSales.
Temporal features	Day, Week, Month, Year.
Promotional and external factors	Promo, Promo2, Holiday, etc.

3.2.4 Data Splitting

The preprocessed dataset is split into training and testing sets to evaluate the model's performance. The split is performed with 80% of the data allocated to the training set and 20% to the testing set. Within the training set, 5-fold cross-validation is applied to ensure robustness and prevent overfitting. This split allows the model to be trained on a substantial portion of the data while reserving a separate portion for unbiased evaluation.

3.2.5 Single-based And Ensemble Machine Learning Methods

Machine learning, specifically boosting ensemble algorithms, provides a robust framework for demand prediction by learning complex patterns from multivariate data, even when the relationships between features are non-linear and intricate. In this study, the boosting ensemble is composed of three state-of-the-art algorithms: XGBoost, LightGBM, and CatBoost. Each of these algorithms contributes unique strengths to the ensemble, with the intent of enhancing its overall predictive performance.

3.2.5.1 LightGBM

LightGBM is a well-established model within the family of gradient boosting decision tree algorithms. Known for its speed, scalability, and efficiency, it offers several advantages—most notably faster training, reduced memory usage, and strong predictive performance, particularly on structured data. One of LightGBM's key innovations lies in how it handles gradient computation: instead of scanning each individual data point, it converts continuous feature values into discrete bins using a histogram-based method. These histograms simplify the process of finding optimal split points and significantly reduce computation time.

In training, LightGBM organizes the dataset into a feature matrix and a label vector, constructs histograms for each feature, and evaluates the best split using these bin summaries. To speed things up even further, it allows child nodes to inherit histograms from their parent or sibling nodes, which avoids redundant calculations.

The tree-building process in LightGBM follows a leaf-wise growth strategy. Unlike level-wise methods that expand all nodes at the same depth, LightGBM focuses on growing the leaf that results in the greatest reduction in loss. This helps the model zero in on the most informative parts of the data. As shown in Figure (3.5), the blue-colored nodes in the diagram represent the parts of the tree being extended at each step. The figure visually tracks the growth of the tree as it deepens along the most promising paths—where the model sees the biggest gain.

This approach is especially effective for datasets like Rossmann's, where important relationships between features—such as promotions, holidays, and store-specific behavior—are often unevenly distributed. LightGBM's ability to prioritize the most impactful splits while managing high-dimensional, potentially sparse features makes it well-suited for real-world retail forecasting problems that require both speed and depth.

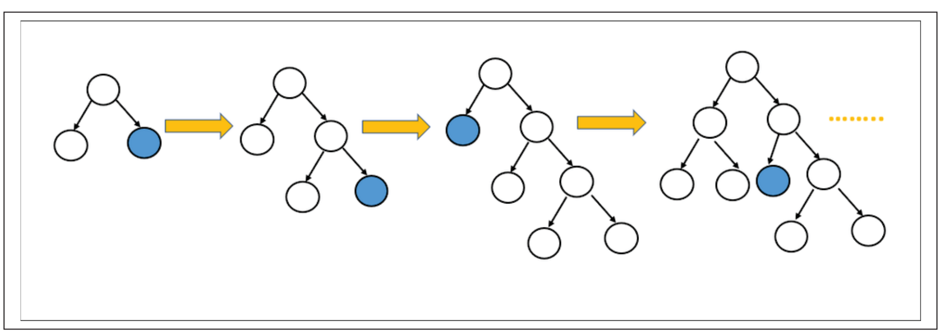

Figure 3.5 Leaf-wise algorithm
Adapted from SAS Institute, 2023

3.2.5.2 CatBoost

CatBoost is a gradient boosting algorithm specifically built to handle datasets with a high number of categorical features. Unlike many models that require manual encoding of categories into numbers, CatBoost can work directly with categorical variables. It applies ordered boosting and uses target statistics to manage these variables in a way that avoids overfitting and keeps training stable. This makes it especially useful for tasks where categorical data plays a central role.

The Rossmann dataset (Cukierski, W. 2015) contains a rich mix of categorical features—such as store types, day-of-week indicators, state holidays, and promotional periods—all of which can significantly influence sales behavior. CatBoost's ability to process these features natively makes it a highly compatible option. It avoids the need for excessive data transformation and reduces the risk of introducing bias during encoding, which helps preserve meaningful relationships between features.

One of the standout techniques in CatBoost is ordered boosting. This method prevents data leakage by ensuring that the model doesn't use future data to train earlier decision steps. It also builds symmetric trees, meaning the structure remains consistent across branches at each depth level. This contributes to model stability and faster inference during prediction.

As illustrated in Figure (3.6), the image shows how CatBoost handles the decision-making process. The root node splits based on feature values, and the algorithm gradually divides the dataset into child nodes using combinations of categorical and numerical attributes. The colored dots represent different categories or values being grouped as the tree progresses through various split stages. This helps visualize how the model manages diverse inputs while keeping the splits meaningful and efficient.

CatBoost also manages missing values internally, requires minimal parameter tuning, and runs efficiently on both CPU and GPU. These features, along with its strong handling of categorical data, make it an ideal fit for complex, feature-rich datasets like Rossmann's.

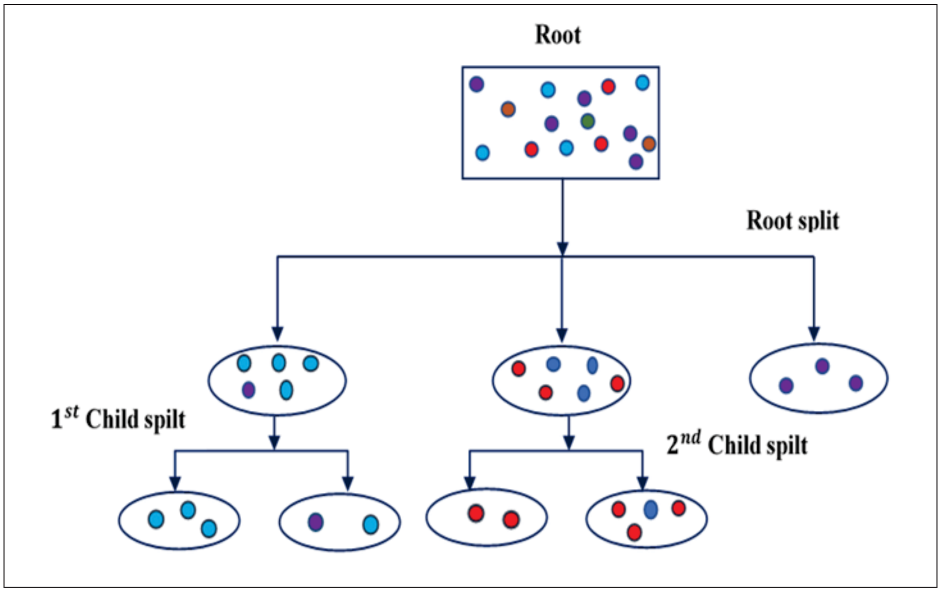

Figure 3.6 Categorical boosting CatBoost algorithm
Taken from Islam et al. , 2024

3.2.5.3 XGBoost

XGBoost is a well-established and high-performing gradient boosting algorithm, chosen for its speed, accuracy, and ability to handle messy or incomplete data without breaking down. It builds decision trees where each one contributes to the final prediction by assigning scores through its leaf nodes. These scores are combined across many trees to produce the model's output. One key strength of XGBoost is its ability to fine-tune predictions by using both the size and direction of past errors, helping it correct itself efficiently as more trees are added.

What sets XGBoost apart is its use of regularization techniques that control tree complexity, preventing the model from overfitting to random fluctuations in the training data. This is especially useful when working with real-world datasets that have noise, irregular entries, or inconsistent behavior—conditions that are common in retail settings.

For the Rossmann dataset specifically, which contains store-level sales data influenced by numerous factors like promotions, holidays, customer traffic, and regional differences, XGBoost offers solid reliability. It handles structured input well, even when data points are missing or sparsely filled, and it maintains stability in scenarios where the importance of features varies. Its ability to generalize across stores while still learning from local store-specific behaviors makes it an appropriate fit for this kind of forecasting task.

In terms of structure, XGBoost grows its trees in a level-wise fashion. This means it expands all nodes at the same depth before deepening the tree. This keeps the model's structure balanced and consistent. As shown in Figure (3.8), the blue nodes highlight how the model splits data at each level evenly, forming a tree that grows uniformly rather than stretching unevenly. This method supports stability and helps avoid the risk of overly complex or lopsided decision paths, which is helpful in environments with broad and shallow trends like those seen in retail chains.

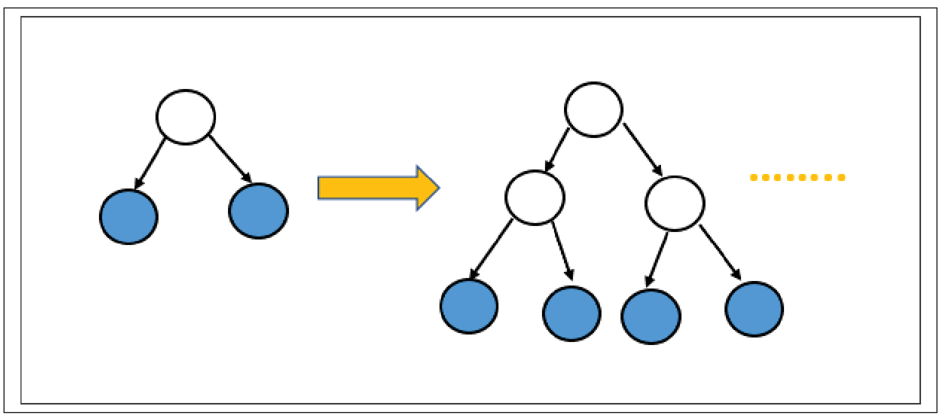

Figure 3.7 Level-wise algorithm
Adapted from SAS Institute, 2023

3.2.5.4 Summary of Single-Base Model Formulations and Strengths

Each of the three algorithms brings unique advantages to the hybrid ensemble method. In the single-based ML formulations, each model is trained independently on the target y, not on residuals (as in the ensemble). In the hybrid ensemble method, the training process is sequential: CatBoost trains on LightGBM residuals, and XGBoost trains on the residuals after CatBoost.

Additionally, the tree functions $k_t(x)$, $h_t(x)$, and $g_t(x)$ are established during training by iteratively building decision trees in LightGBM, CatBoost, and XGBoost, respectively, where splits are determined by features (e.g., `StoreType`, `Promo`), weights are improved via gradient descent, and learning rate $\eta = 0.05$ controls step size to minimize loss (e.g., squared error). The target variable y (`Sales`) is predicted sequentially: LightGBM predicts $\hat{y}_1$, CatBoost corrects residuals $y - \hat{y}_1$, and XGBoost refines residuals $y - (\hat{y}_1 + \hat{y}_2)$, ensuring each model iteratively reduces errors, as residuals directly measure the gap between the target and current predictions.

The objective functions in Table 3.1 are $f_{\text{XGB}}(x) = \sum_{t=1}^{T} \eta \cdot k_t(x)$, $f_{\text{LGBM}}(x) = \sum_{t=1}^{T} \eta \cdot h_t(x)$, and $f_{\text{CatBoost}}(x) = \sum_{t=1}^{T} \eta \cdot g_t(x)$, where the dependent variable $\hat{y}$ (predicted `Sales`, units/day) depends on independent variables x (features like `StoreType`, `Promo`, with t indexing T trees. The coefficient $\eta = 0.05$ is established via grid search to minimize RMSE and is the same across all algorithms to ensure consistent contribution in the sequential ensemble, aligning with boosting principles.

Table 3.2 Gradient Boosting Algorithms in Hybrid Ensemble

Model	**Formulations**	**Strengths**
XGBoost	$f_{XGB}(x) = \sum_{t=1}^{T} \eta \cdot k_t(x)$	Strong regularization, scalable, generalizes well (e.g., sales, customer data).
LightGBM	$f_{LGBM}(x) = \sum_{t=1}^{T} \eta \cdot h_t(x)$	Fast training, high accuracy, efficient for temporal/interaction features.
CatBoost	$f_{CatBoost}(x) = \sum_{t=1}^{T} \eta \cdot g_t(x)$	Handles categorical features well (e.g., store type, promotions), less preprocessing.

3.2.6 Hybrid Boosting Ensemble Algorithm

The hybrid boosting ensemble algorithm (Figure 3.8) developed in this study integrates the complementary strengths of XGBoost, LightGBM, and CatBoost to deliver a robust and accurate predictive model for demand forecasting. Boosting ensembles operate by sequentially constructing a series of weak learners—here, decision trees—where each subsequent learner is trained to correct the residual errors of its predecessors.

The selection of CatBoost, LightGBM, and XGBoost as components of the hybrid ensemble is strategically motivated by their proven effectiveness in handling complex, heterogeneous data. This decision is further supported by Alsulamy (2024), who employed these models for predicting construction delay risks and provided a comparative justification for their use in structured, categorical-heavy datasets.

- CatBoost excels in managing high-cardinality categorical variables without requiring extensive preprocessing. It is more applicable particularly for retail datasets like Rossmann that include categorical features such as store type, holiday status, and promotional indicators.
- XGBoost is recognized for its robust performance on large, imbalanced datasets. Its embedded regularization mechanisms (L1 and L2) help control model complexity, making it suitable for capturing intricate dependencies and temporal fluctuations common in sales and demand patterns.
- LightGBM offers advanced techniques such as Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling, enabling high-speed training with lower memory consumption. These characteristics make it ideal for large-scale forecasting tasks where time efficiency is critical.

The final prediction, denoted as $\hat{y}$, is an additive combination of the individual predictions from the three models, expressed mathematically as:

$$\hat{y} = f_{XGB}(x) + f_{LGBM}(x) + f_{CatBoost}(x) \tag{3.8}$$

Here, $f_{LGBM}(x)$, $f_{CatBoost}(x)$, and $f_{XGB}(x)$ represent the prediction functions of LightGBM, CatBoost, and XGBoost, respectively, applied to the input features x. The summation is used because the ensemble trains sequentially: LightGBM predicts initial sales, CatBoost corrects residuals (errors), and XGBoost refines further. Summing ensures each model's correction builds additively on the previous prediction, aligning with boosting's principle of iterative error reduction. This aggregation leverages the unique capabilities of each algorithm to enhance forecasting performance.

The hybrid boosting ensemble captures non-linear relationships and interactions between these features, such as the impact of promotions on sales or the effect of temporal trends on customer behavior, making it well-suited for demand forecasting in inventory management.

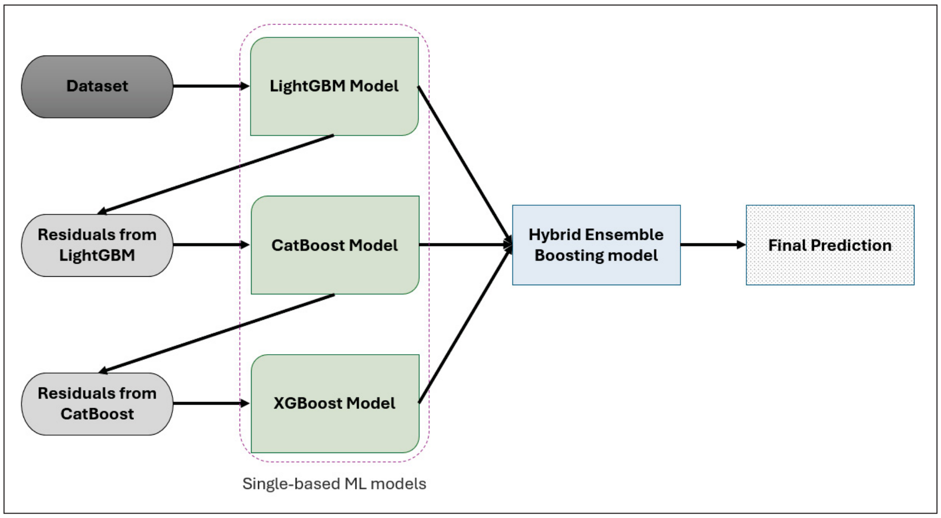

Figure 3.8 Hybrid boosting ensemble algorithm

3.2.7 Hyper-parameter Optimization

Hyperparameter optimization is a critical step in enhancing the performance of the boosting ensemble. The process involves selecting the optimal set of hyperparameters for XGBoost, LightGBM, and CatBoost, as well as determining the ensemble weights by using grid search optimization technique. The following method was utilized:

- Grid Search: it is employed to systematically explore a predefined set of hyperparameter values for each algorithm, including parameters such as learning rate, number of trees, and maximum depth, to identify the optimal combination. This search was conducted using 5-fold cross-validation, unlike classical k-fold which may not focus on hyperparameter tuning (Seyedan et al., 2023) to ensure that each model generalizes effectively to unseen data, thereby improving the robustness and performance of the ensemble.

3.2.8 Evaluation Method

To evaluate the performance of the hybrid boosting ensemble model, two performance metrics are used: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These metrics are widely adopted in demand forecasting studies to assess the accuracy of predictions.

Root Mean Squared Error (RMSE): RMSE measures the average squared error of the forecasted values compared to the actual values. It gives more weight to larger errors, making it sensitive to outliers. The RMSE is calculated using the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.9}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the number of observations.

Mean Absolute Error (MAE): MAE measures the average absolute difference between the predicted and actual values, providing a straightforward measure of prediction accuracy. It is less sensitive to outliers compared to RMSE. The MAE is calculated as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.10}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

Mean Absolute Percentage Error (MAPE): MAPE measures the average percentage error between predicted and actual values, offering a relative measure of accuracy. It is particularly useful for understanding error in the context of the magnitude of the actual values. MAPE is calculated as:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{3.11}$$

Prediction within 10% (Pred(x=10%)): It measures the percentage of predictions that fall within 10% of the actual values, assessing the model's precision for close forecasts. It is calculated as:

$$\text{Pred}(x=10) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right| \leq 0.1\right) \times 100 \tag{3.12}$$

where $\mathbb{I}$ is an indicator function that returns 1 if the condition is true, and 0 otherwise.

The ensemble model is evaluated on the test set (20% of the data) to assess its performance on unseen data. The combination of RMSE and MAE, MAPE, Pred(x=10) provides a comprehensive view of the model's accuracy, with lower values indicating better performance.

3.2.8.1 Cross-validation and performance evaluation

To ensure the robustness of the boosting ensemble model, 5-fold cross-validation is applied during the training phase. The training set (80five times, with each part serving as the validation set once. The average performance across all folds (measured using RMSE) is used to assess the model's generalization ability and select the optimal hyperparameters for each algorithm.

After training, the ensemble model's performance is evaluated on the test set using both RMSE and MAE. These metrics provide a clear indication of the model's predictive accuracy and its suitability for demand forecasting in inventory management.

3.2.9 Demand Prediction And Inventory Management

The final step involves using the boosting ensemble model to generate demand predictions. The preprocessed dataset, after merging and splitting, is fed into the model to predict future demand. These predictions are then integrated into an inventory management model, enabling efficient stock allocation, reducing overstock or stockouts, and enhancing supply chain operations. This approach ensures that inventory decisions are data-driven and aligned with predicted demand patterns.

3.2.9.1 Uncertainty Consideration in the Dataset

In predictive modeling for retail sales forecasting, uncertainty arises from data variability, measurement errors, and inherent noise in the dataset. This section presents an uncertainty modeling approach that leverages a fuzzy generalization technique to preprocess the input data, introducing controlled variability within a boosting ensemble framework comprising LightGBM, CatBoost, and XGBoost models. The methodology is applied to a retail sales dataset, aiming to enhance the robustness of predictions by smoothing feature values through fuzzy discretization. The fuzzy generalization method, implemented as part of the preprocessing pipeline, serves as a form of uncertainty modeling by discretizing continuous features into representative categories, thereby reducing sensitivity to noise and outliers.

3.2.9.2 Uncertainty Modeling

This study focuses on a type of uncertainty that introduced at the data level through a fuzzy generalization technique to simulate variability in retail sales data. The fuzzy generalization method discretizes feature values into a fixed number of categories, effectively smoothing the data and introducing uncertainty by generalizing continuous values into representative midpoints. This approach is computationally efficient and integrates seamlessly with the boosting ensemble pipeline.

3.2.9.3 Data-Level Uncertainty via Fuzzy Generalization

Data-level uncertainty is introduced by applying the fuzzy generalization function to the normalized dataset after MinMaxScaler. This function discretizes each feature into a predefined number of categories, mapping continuous values to the nearest representative midpoint. The process, which can be viewed as a simplified form of fuzzy clustering as shown in Figure (3.9), introduces uncertainty by reducing the granularity of the data, thereby smoothing out minor fluctuations and mitigating the impact of noise and outliers.

1. Input Data:

$$X_{\text{Scaled\_merged}}, \quad \text{Features\_dataset}$$

2. Number of Categories (K):

$$K = 12$$

3. For Each Feature Column:

$$\Delta = \frac{\text{Max} - \text{Min}}{K}$$

4. For Each Value in the Column:

$$\text{Distance} = |x_i - m_j|$$

5. Output:

$$\text{Generalized Dataset: } \textit{generalized\_rossmann\_dataset}$$

Figure 3.9 Fuzzy Clustering Process

The fuzzy generalization process operates on the scaled dataset $X_{\text{Scal\_mrg}}$, which contains N samples and M features, and proceeds as follows:

1. **Category Definition:**
   - A fixed number of categories, $K = 12$, is defined for each feature (Bezdek et al, 1981). This parameter determines the level of discretization and was chosen to reflect monthly cycles, seasonal patterns, balance smoothing and detail preservation in the Rossmann dataset.
   - For each feature column j (where $j = 1, 2, \ldots, M$), the minimum ($\min_j$) and maximum ($\max_j$) values are computed.

- The range of the feature is calculated as $R_j = \max_j - \min_j$, and this range is divided into K equal intervals, with the interval width given by:

$$\Delta_j = \frac{R_j}{K} \tag{3.13}$$

- Representative values (centroids) for each category are defined as the midpoints of these intervals:

$$c_{j,k} = \min_j + (k + 0.5) \cdot \Delta_j, \quad k = 0, 1, \ldots, K - 1 \tag{3.14}$$

  where $c_{j,k}$ is the representative value for the k-th category of feature j.

2. **Mapping Values to Categories:**
   - For each scaled feature value $x_{i,j}$ (the value of feature j for sample i, e.g., `Sales` after Min-Max scaling), the absolute distance,equivalent to the Manhattan distance for one-dimensional data to each category representative $c_{j,k}$ is computed as shown in Equation (2.15). This choice is computationally efficient for the Rossmann dataset's normalized features (Han et al., 2012). Alternatives like Euclidean or Minkowski distances were not used due to the single-dimensional nature of each feature post-scaling, ensuring simplicity while achieving robust performance:

$$d_{i,j,k} = |x_{i,j} - c_{j,k}|, \quad k = 0, 1, \ldots, K - 1 \tag{3.15}$$

   - The data point is assigned to the nearest representative based on the minimum distance:

$$\hat{x}_{i,j} = c_{j,k^*} \quad \text{where} \quad k^* = \arg\min_k (d_{i,j,k}) \tag{3.16}$$

   This mapping process discretizes each feature's values into one of the K categories, introducing uncertainty by replacing the original continuous values with generalized midpoints. Although the assignment is hard (each value maps to a single representative), the use of midpoints and a fixed number of categories creates a fuzzy-like smoothing effect, as values are generalized rather than precisely retained.

3. **Output Construction:**

- The transformed values $\hat{x}_{i,j}$ for all samples and features are aggregated into a new dataset.
- The resulting dataset is converted into a generalized Rossmann data frame with the original feature names and then mapped as an array for data generation to feed the machine learning models.

This fuzzy generalization approach introduces uncertainty by discretizing the feature space, simulating variability that might arise from measurement errors or natural fluctuations in retail sales data. The smoothing effect helps the boosting ensemble focus on broader patterns suppressing the noise impacts.

3.2.9.4 Integration with Boosting Ensemble

The fuzzy generalization technique is integrated into the boosting ensemble pipeline at the preprocessing stage, as follows:

- **Data Preparation:** The retail sales dataset is loaded from the Rossmann dataset, merged, and preprocessed. Features are selected, categorical variables are label-encoded, and missing values are filled with 0. The target variable (Sales) is clipped between 2000 and 16000 to remove outliers.
- **Normalization:** The input features are normalized using `MinMaxScaler` to scale values to the range $[0, 1]$, producing the merged scaled input.
- **Fuzzy Generalization:** The defined fuzzy generalization function is applied to the merged scaled input, producing the generalized Rossmann dataset, which is converted into the generated data.
- **Standardization:** The generalized data is further standardized using `StandardScaler` to produce X_{scaled}, ensuring zero mean and unit variance for model training.
- **Training and Prediction:** The boosting ensemble (LightGBM $\rightarrow$ CatBoost $\rightarrow$ XGBoost) is trained on X_{scaled}, with each model sequentially correcting the residuals of the previous one. The final ensemble prediction is obtained by summing the outputs of all models.

- **Evaluation:** The ensemble's performance is evaluated using RMSE and MAE on both the training and test sets, with results visualized through scatter plots comparing actual versus predicted sales.

The Two-Phase model architecture for retail sales forecasting is illustrated in (Figure 3.10), designed to enhance prediction accuracy through a systematic approach.

- **Phase 1.** Raw data undergoes merging, encoding, and cleaning, followed by normalization to scale features uniformly. Fuzzy generalization is then applied, discretizing features into categories to introduce data-level uncertainty and smooth the data. The result is then standardized to prepare the dataset for model training.
- **Phase 2.** The processed features are input into a boosting ensemble of three single-based models (LightGBM, CatBoost, and XGBoost), which are trained sequentially to correct prediction errors. The ensemble output is then evaluated using RMSE and MAE and visualized to assess prediction accuracy and model stability.

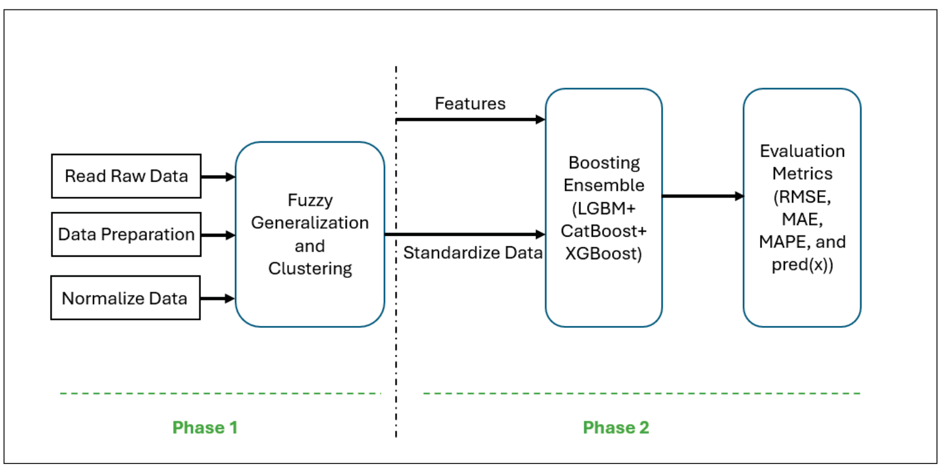

Figure 3.10 Two-Phase Proposed Fuzzy Clustering Architecture

CHAPTER 4

NUMERICAL RESULTS

This chapter presents the performance evaluation of the proposed hybrid boosting ensemble method for demand forecasting and its impacts on supply chain improvement. The simulations are conducted across two case studies, each focusing on different aspects of the Rossmann dataset to assess the efficacy of the proposed framework. The primary objective is to evaluate the forecasting accuracy of the hybrid boosting ensemble (combining LightGBM, CatBoost, and XGBoost) and its downstream effects on inventory improvement using the OUTL policy.

The chapter begins by detailing the simulation environment, including the dataset, tools, and evaluation metrics. It then presents the results for each case study, focusing on forecasting accuracy through metrics such as RMSE, MAE, MAPE, and pred(x) as well as visualizations like scatter plots and time series comparisons of actual vs. predicted sales. The results are analyzed to highlight the proposed model's performance against individual boosting methods (LightGBM, CatBoost, and XGBoost) and to demonstrate its practical implications for inventory management in the retail context of Rossmann stores. Finally, a summary of the findings underscores the model's effectiveness in enhancing demand forecasting and supply chain efficiency, while identifying areas for potential improvement. Additionally, the supply chain decision variables such as safety stock, reordering point, order quantity, order-up to level are descried via tables and histogram plotting of total cost for two case studies are shown for different fixed lead time.

4.1 Simulations

The simulations for the proposed hybrid boosting ensemble method were conducted using Python, leveraging libraries such as scikit-learn, LightGBM, CatBoost, and XGBoost for model implementation, and Matplotlib for visualization. The hybrid boosting ensemble was trained sequentially: LightGBM generated initial predictions, CatBoost corrected residuals, and XGBoost refined the outputs, with hyperparameters tuned using grid search and 5-fold cross-validation. Together, these algorithms address the diverse challenges present in the Rossmann dataset—ranging from nonlinear feature interactions to high-dimensional and time-sensitive variables. The assertion of nonlinear feature interactions in the Rossmann dataset is supported by the dataset's multivariate nature, including features like Promo, DayOfWeek, and StoreType, which exhibit complex relationships. The integration of their complementary strengths enhances both the accuracy and efficiency of demand forecasting, serving as a strong foundation for the inventory improvement phase of this research.

Two case studies were designed to evaluate the proposed framework:

- **Case Study 1 (Without handling Uncertainity):** Focuses on the generic hybrid boosting ensemble method for demand forecasting, assessing its performance across all stores in the Rossmann dataset.
- **Case Study 2 (With handling Uncertainity):** Focuses on the Fuzzy hybrid boosting ensemble method for demand forecasting, assessing its performance across all stores in the Rossmann dataset.

  In both case studies, the analysis is extended by incorporating the OUTL inventory policy, evaluating the impact of improved forecasting accuracy on inventory improvement metrics such as reorder points, safety stock, and total costs.

4.1.1 Model Evaluation and Analysis

Key evaluation metrics include:

The key evaluation metrics used to assess the proposed hybrid ensemble's performance are RMSE which emphasizes larger errors, MAE which provides a stable measure of average error, MAPE which offers a relative error perspective, and Prediction within 10% (Pred(x)) which measures the percentage of predictions within 10% of actual sales. These metrics are presented in Table 4.1 to compare the hybrid ensemble's performance against classical forecasting methods (ARRIMA, MA, SES), and ML methods (LightGBM, LightGBM+CatBoost) on the Rossmann test set, highlighting improvements in forecasting accuracy.

Graphical representations, including:

- **Scatter plots** of actual vs. predicted sales,
- **Time series plots** showing actual and forecasted sales over a 200-day period,

were used to visually assess the model's predictive performance. These visualizations highlight the alignment between model output and true sales patterns and are used to validate model stability.

For further illustration in both case studies, additional analysis is conducted focusing on inventory improvement outcomes. These include calculations for:

- Safety Stock
- Reorder Point
- Order Quantity
- Total Costs

These values are derived based on the OUTL policy, integrating demand forecasting outputs into supply chain decision-making.

4.2 Results

The simulations were executed to assess the performance of the proposed hybrid boosting ensemble method across the two case studies. Each case study evaluates forecasting accuracy using RMSE, MAE, MAPE, Pred(x=10%), with results recorded for both training and test sets. The evaluation criteria focus on the model's ability to accurately predict daily sales and its subsequent impact on supply chain improvement. Visualizations provide a deeper understanding of the model's predictive behavior over time and across sales values. The results are presented separately for both case studies, with comparisons between generic hybrid boosting and fuzzy hybrid boosting ensemble, entirely demonstrating their effectiveness and impacts on inventory improvement through the OUTL policy.

4.2.1 Case Study 1: Generic Hybrid Boosting Ensemble for Demand Forecasting

This case study assesses the forecasting accuracy of the hybrid boosting ensemble on the Rossmann dataset, focusing on predicting daily sales across all stores. The ensemble's performance is compared against individual boosting methods, with results highlighting improvements in accuracy.

4.2.1.1 Forecasting Accuracy Analysis

Table 3.1 compares the forecasting performance of the proposed hybrid boosting ensemble (LightGBM + CatBoost + XGBoost) against traditional models (ARIMA, MA, SES) and machine learning methods (LightGBM, LightGBM+CatBoost) using metrics like MAE, RMSE, MAPE, and Pred(x=10%). The hybrid ensemble achieves the lowest errors (MAE: 416.77, RMSE: 608.66, MAPE: 6.48%) and the highest percentage of predictions within 10% of actual values (81.12%), significantly outperforming all benchmarks.

Table 4.1 Comparison of Model Performance Metrics for Case Study 1

Model	**MAE**	**RMSE**	**MAPE**	**Pred(x=10%)**
ARIMA	3167.62	4210.67	9.79%	60.21%
MA	2054.6	3350.53	9.72%	60.00%
SES	893.21	1140.47	13.79%	46.92%
LGBM	599.61	849.86	9.52%	65.11%
LGBM+CATBoost	495.46	716.23	7.76%	73.56%
Boosting Ensemble: LGBM+CATBoost+XGBoost	**416.77**	**608.66**	**6.48%**	**81.12%**

Figure 4.1 visualizes this superiority, showing the hybrid ensemble's metrics as consistently lower than traditional and other ML methods, highlighting its enhanced accuracy in predicting daily sales for Rossmann stores.

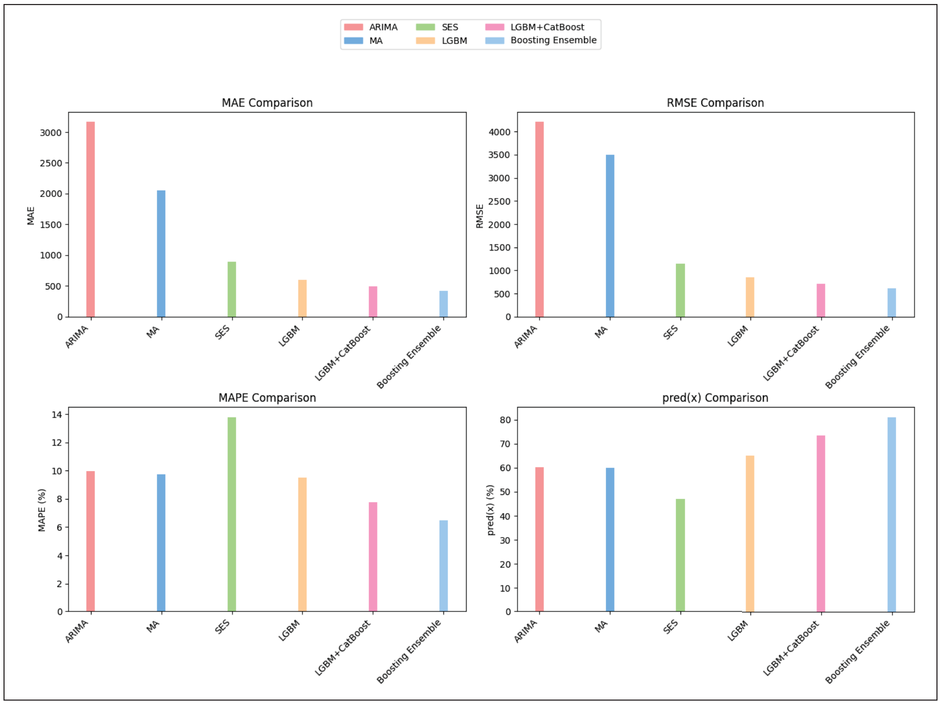

Figure 4.1 Comaparison of evaluation metrics between traditional models, proposed ML methods

4.2.1.2 Actual vs. Predicted Sales Analysis

Figure 4.2 illustrates a scatter plot comparing actual vs. predicted sales on the test set for traditional forecasting methods (ARIMA, MA, and SES.)

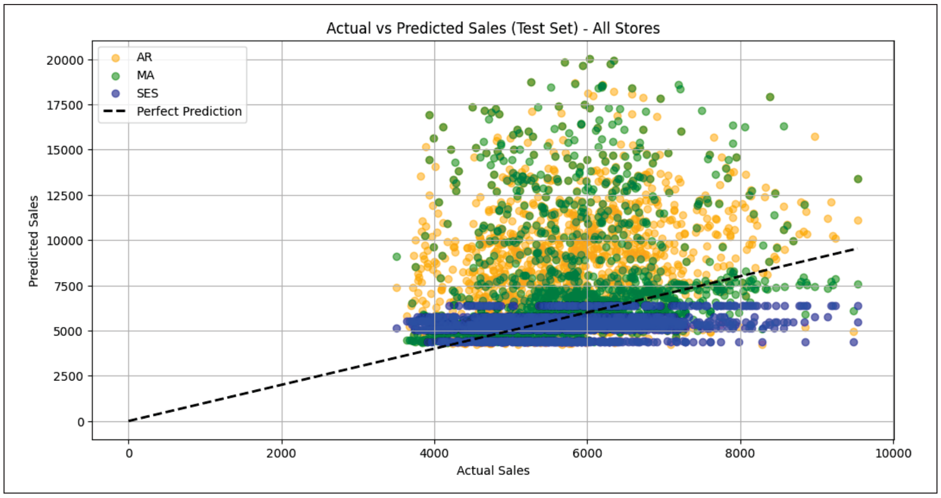

Figure 4.2 Actual vs. Predicted Sales (Test Set) for Traditional Forecasting methods

Figure 4.3 illustrates a scatter plot comparing actual vs. predicted sales on the test set for the hybrid boosting ensemble.

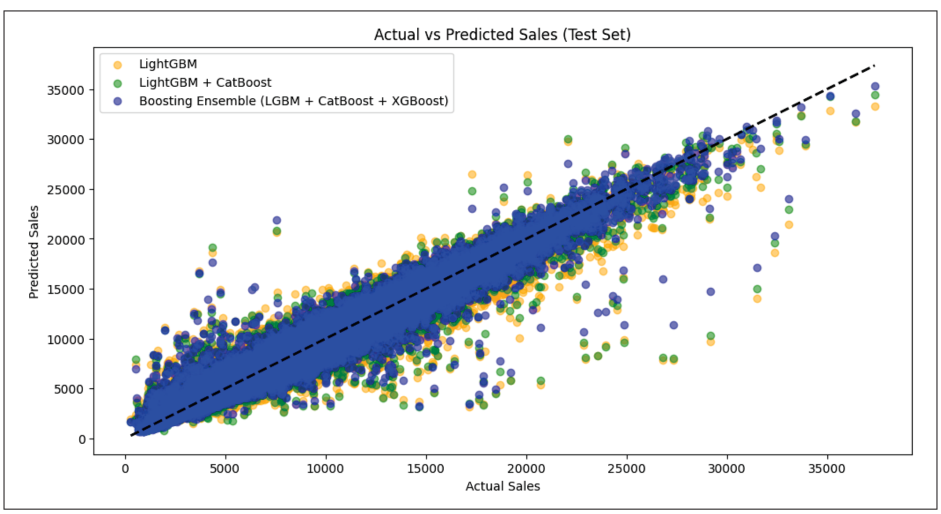

Figure 4.3 Actual vs. Predicted Sales (Test Set) for Case Study 1

The scatter plot in Figure 4.3 shows a strong alignment between actual and predicted sales, with most points clustering around the diagonal line (y=x), indicating high predictive accuracy. The ensemble effectively captures the range of sales values, from low (around 0) to high (up to 35,000), with minimal outliers. This suggests that the model generalizes well across different sales magnitudes, a critical factor for retail applications where demand can vary significantly.

4.2.1.3 Time Series Analysis of Actual vs. Forecasted Sales

Figure 4.4 presents a time series plot comparing actual and predicted sales over the first 200 days of the test set.

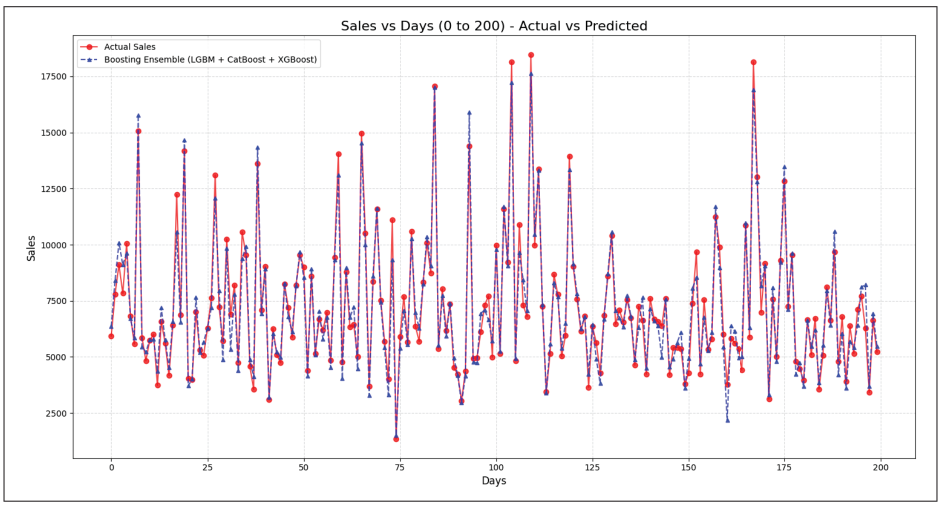

Figure 4.4 Sales vs. Days (0 to 200) – Actual vs. Predicted for Case Study 1

Figure 4.4 presents a time series plot comparing actual and predicted sales over the first 200 days of the test set for Case Study 1. The hybrid boosting ensemble accurately tracks actual sales trends, with predicted (blue) and actual (red) sales showing similar patterns, including peaks (e.g., days 10, 50, 150) and troughs (e.g., day 100). The model suffers from handling sales volatility and demand fluctuations effectively, particularly when dealing with data uncertainity.

4.2.2 Case Study 2: Fuzzy Hybrid Boosting Ensemble for Demand Forecasting

This case study evaluates the fuzzy hybrid boosting ensemble, combining LightGBM, CatBoost, and XGBoost with fuzzy generalization to address demand uncertainty, for forecasting daily sales across all Rossmann stores. Fuzzy generalization enhances prediction robustness by smoothing data variability. The forecasts are integrated with the OUTL inventory policy to determine reorder points, safety stock, order quantities, and total costs (purchase, holding, ordering). The following results present the forecasting accuracy (RMSE, MAE, MAPE, Pred(x=10%)) and inventory improvement outcomes through tables, figures, and visualizations, demonstrating improved operational efficiency under uncertainty.

4.2.2.1 Forecasting Accuracy Analysis

Table 3.2 presents the RMSE, MAE, MAPE, and Pred(x=10%) values for the training and test sets, similar to Case Study 1, but with a focus on the subset of stores used for inventory improvement.

Table 4.2 Comparision of Model Performance Metrics for Case Study 2

Model	**MAE**	**RMSE**	**MAPE**	**pred(x)**
LGBM+Fuzzy	226.07	289.82	3.88%	94.10%
(LGBM+CatBoost)+Fuzzy	219.49	281.57	3.78%	95.22%
Boosting Ensemble+Fuzzy: (LGBM+CatBoost+XGBoost)+Fuzzy	**209.21**	**272.99**	**3.53%**	**97.11%**

Table 3.2 shows that the hybrid boosting ensemble again outperforms the individual models. On the test set, the ensemble achieves an MAE of 209.21 and an RMSE of 272.99, compared to LightGBM's MAE of 226.07 and RMSE of 289.82, and CatBoost's MAE of 219.49 and RMSE of 281.57. This corresponds to a reduction in MAE by 7.5% compared to LightGBM and 4.7% compared to CatBoost, and a reduction in RMSE by 5.8% and 3.1%, respectively. While the improvements are less pronounced than in Case Study 1, likely due to the smaller subset

of stores and reduced variability, the ensemble still demonstrates superior accuracy, which is critical for effective inventory management.

4.2.2.2 Actual vs. Predicted Sales Analysis

Figure 4.5 illustrates a scatter plot comparing actual vs. predicted sales on the test set for the hybrid boosting ensemble in Case Study 2.

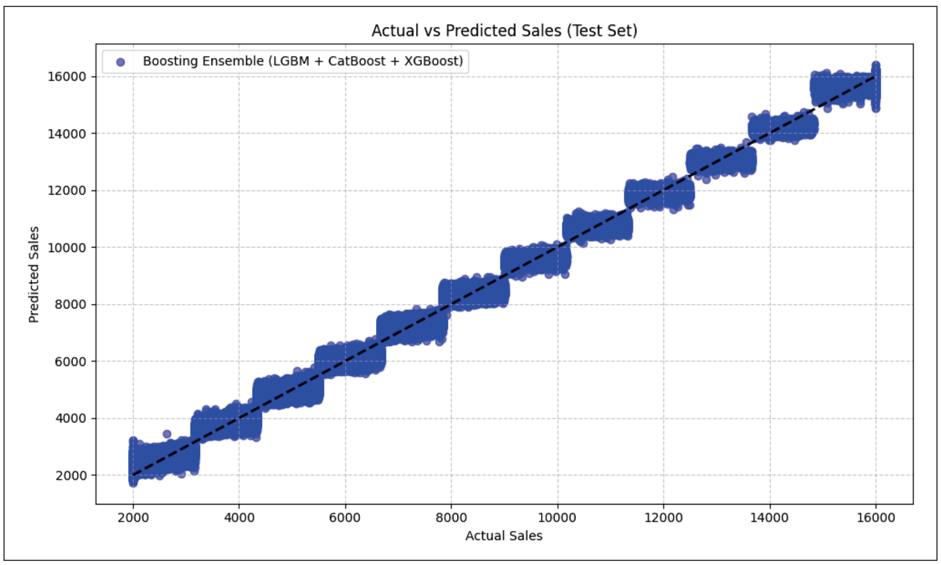

Figure 4.5 Actual vs. Predicted Sales (Test Set) for Case Study 2

The scatter plot in Figure 4.5 shows a tight clustering of points along the diagonal line, indicating high predictive accuracy. The sales values in this case study range from approximately 2,000 to 16,000, reflecting the subset of stores analyzed. The ensemble's predictions align closely with actual sales, with fewer outliers compared to Case Study 1, suggesting that the model performs consistently across different subsets of the Rossmann dataset.

4.2.2.3 Time Series Analysis of Actual vs. Forecasted Sales

Figure 4.6 presents a time series plot comparing actual and predicted sales over the first 200 days of the test set for Case Study 2.

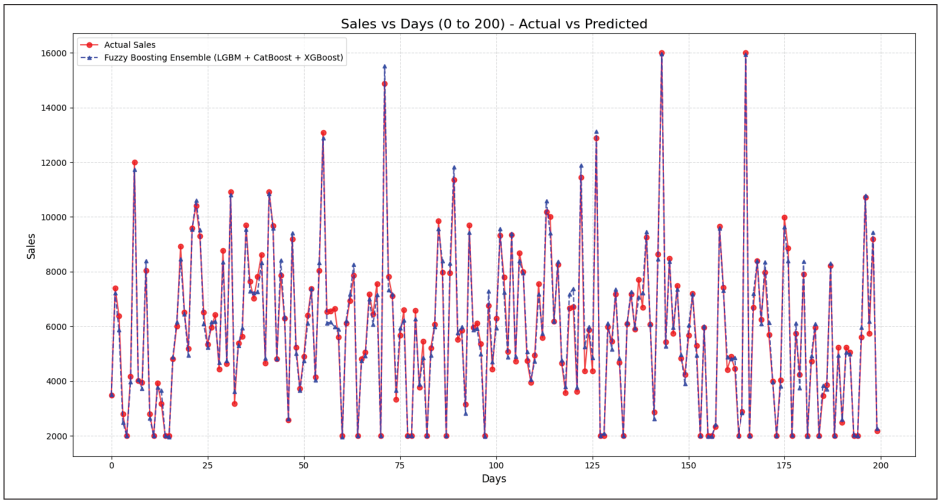

Figure 4.6 Sales vs. Days (0 to 200) – Actual vs. Predicted for Case Study 2

The time series plot in Figure 4.6 demonstrates that the hybrid boosting ensemble accurately captures the sales trends over the 200-day period. The actual (red) and predicted (blue) sales follow similar trajectories, with the ensemble effectively predicting both peaks (e.g., around day 150, where sales reach approximately 16,000) and troughs (e.g., around day 100, where sales drop to near 0). The predicted sales closely match the actual values, highlighting the model's ability to handle temporal patterns and demand volatility.

4.2.2.4 Inventory improvement Analysis

The improved forecasting accuracy from the hybrid boosting ensemble directly impacts inventory improvement through the OUTL policy. This section compares the performance of the Generic Boosting Ensemble and the Fuzzy Boosting Ensemble models, focusing on key inventory metrics for a lead time of 7 days, as shown in Table 4.3. Additionally, a comparison of total costs across various lead times is presented in Figure 4.7 to highlight the models' cost efficiency under different conditions.

Table 4.3 Comparison of Inventory improvement Metrics for Lead Time=7 Days

Model	**Safety Stock**	**Reorder Point**	**Order Quantity**	**Order-up to level**	**Total Cost ($)**
Generic Boosting Ensemble	18646	116062	97416	213478	25599619.94
Fuzzy Boosting Ensemble	19742	104963	85220	190181	22397960.58

The results in Table 4.3 reveal the following insights:

- **Safety Stock (SS):** The Fuzzy Boosting Ensemble requires a higher safety stock (19,742) compared to the Generic Boosting Ensemble (18,646), an increase of approximately 5.9%. This suggests that the fuzzy model adopts a more conservative approach to handle demand uncertainty, potentially at the cost of higher holding expenses.
- **Reorder Point (ROP):** The Fuzzy Boosting Ensemble has a lower reorder point (104,963) compared to the Generic Boosting Ensemble (116,062), a reduction of 9.6%. This indicates that the fuzzy model triggers reordering earlier, which may help reduce stockouts but could increase ordering frequency.
- **Order Quantity:** The Fuzzy Boosting Ensemble orders a smaller quantity (85,220) compared to the Generic Boosting Ensemble (97,416), a decrease of 12.5%. This reduction in order quantity aligns with the lower reorder point and order-up-to-level, reflecting a more cautious inventory replenishment strategy.
- **Order-Up-To-Level:** The Fuzzy Boosting Ensemble's order-up-to-level is 190,183, which is 10.9% lower than the Generic Boosting Ensemble's 213,478. This reduction suggests that the fuzzy model maintains a leaner inventory, potentially reducing overstocking risks.
- **Total Cost:** The Fuzzy Boosting Ensemble significantly reduces total inventory costs (purchase, holding, and ordering) to $22,398 million USD, compared to the Generic Boosting Ensemble's $25,6 million USD—a reduction of 12.5%. This cost saving is driven by the fuzzy model's ability to balance ordering, despite the higher safety stock.

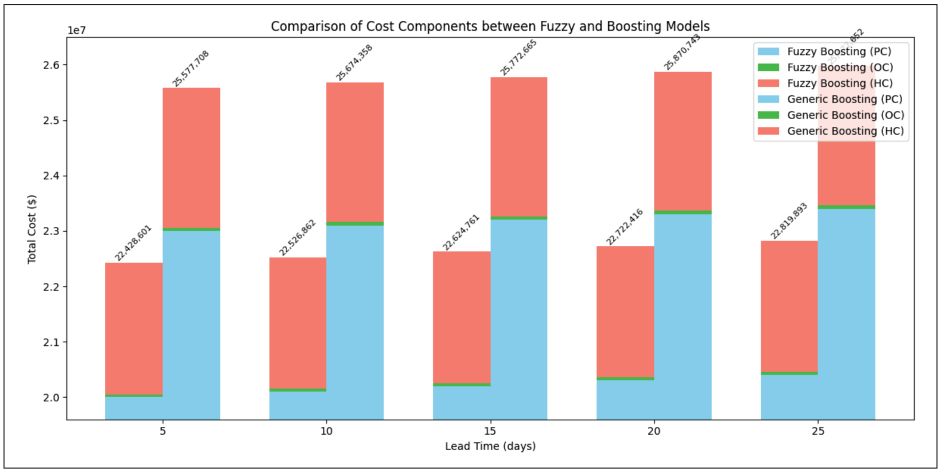

Figure 4.7 Comparison of Total Costs Across Lead Times for Fuzzy and Generic Boosting Models

Figure 4.7 illustrates the total costs of the Generic Boosting Ensemble and Fuzzy Boosting Ensemble models across lead times of 5, 10, 15, 20, and 25 days. The consistent cost savings highlight the fuzzy model's ability to better handle demand uncertainty, leading to more efficient inventory management. The detailed explanations regarding the inventory improvement outcomes are expressed in the discussion section.

CHAPTER 5

DISCUSSION

5.1 Discussion

This chapter provides a comprehensive discussion of the results presented in Chapter 3, focusing on the performance of the proposed hybrid generic and hyrid fuzzy boosting ensemble method for demand forecasting across two case studies and their implications for inventory improvement in the context of Rossmann stores.

5.1.1 Predictive Performance of the Generic Hybrid Boosting Ensemble in Case Study 1

The results in Table 4.1 provide a clear quantitative assessment of the hybrid boosting ensemble's forecasting performance, benchmarked against traditional time-series models (ARIMA, MA, and SES) and machine learning approaches (LightGBM, LightGBM+CatBoost). The hybrid ensemble, integrating LightGBM, CatBoost, and XGBoost, achieves a MAE of 416.77, a RMSE of 608.66, a MAPE of 6.48%, and a Pred(x=10%) of 81.12% on the test set. These metrics reflect substantial improvements over all compared models, with reductions in MAE by 30.5% compared to LightGBM, 15.9% compared to LightGBM+CatBoost, 86.8% compared to ARIMA, 79.7% compared to MA, and 53.3% compared to SES. Similarly, RMSE reductions range from 15.0% (LightGBM+CatBoost) to 85.5% (ARIMA), while MAPE decreases by up to 53.0% (SES), and Pred(x=10%) improves by up to 72.9% (SES).

The superior performance of the hybrid ensemble can be attributed to its sequential architecture, which leverages the complementary strengths of its constituent models. LightGBM's computational efficiency and ability to handle large-scale datasets enable rapid initial predictions, while CatBoost's robust handling of categorical features (e.g., store type, promotional indicators) corrects residual errors effectively. XGBoost's regularization capabilities further refine predictions, and enhancing generalization across the diverse Rossmann dataset, which includes sales, temporal, and promotional features. This synergy allows the ensemble to capture complex, non-linear demand patterns—such as those driven by promotions or holidays—that traditional models like ARIMA and MA struggle to address due to their reliance on linear assumptions and univariate data.

Figure 4.1 visually reinforces these findings, presenting a comparative bar chart of the evaluation metrics across models. The hybrid ensemble's bars are consistently the lowest for MAE, RMSE, and MAPE, and the highest for Pred(x=10%), underscoring its dominance in forecasting accuracy. The visual comparison highlights the limitations of traditional models, particularly ARIMA (MAE: 3167.62, RMSE: 4210.67), which exhibits significantly higher errors due to its inability to model multivariate relationships. Even single-model machine learning approaches, such as LightGBM (MAE: 599.61, RMSE: 849.86), fall short of the ensemble's performance, as they lack the iterative error correction provided by the sequential boosting framework. The high Pred(x=10%) value of 81.12% for the ensemble indicates that over 80% of its predictions fall within 10% of actual sales, a critical threshold for retail applications where precise inventory decisions depend on reliable forecasts.

While the hybrid ensemble outperforms all benchmarks, the MAPE of 6.48% indicates that some predictions, particularly for low-sales periods, may still deviate from actual values. This could be due to the model's sensitivity to sparse or noisy data in certain store-day combinations, as the Rossmann dataset includes variability across 1,115 stores over 942 days. For improvement of the performance, the fuzzy hybrid ensemble method results are discussed in the next section which depicting its capability to mitigate uncertainity and noisy data.

Figure 4.2 illustrates a scatter plot comparing actual versus predicted sales on the test set for traditional forecasting methods (ARIMA, MA, and SES). The plot reveals significant deviations from the diagonal line (y=x), indicating poor predictive accuracy, particularly for ARIMA, which exhibits large outliers and a wide spread of points due to its reliance on linear, univariate assumptions. MA and SES also show scattered predictions, struggling to capture the multivariate, non-linear patterns inherent in the Rossmann dataset, such as those driven by promotions, holidays, or store-specific features. This underscores the limitations of traditional methods in handling the complex demand dynamics of modern retail environments.

In contrast, Figure 4.3 presents a scatter plot for the hybrid boosting ensemble, showing a tight clustering of points along the diagonal line, indicative of high predictive accuracy. The ensemble effectively captures sales values ranging from near 0 to 35,000, with minimal outliers, demonstrating its ability to generalize across diverse sales magnitudes. This performance is a direct result of the sequential architecture of the ensemble, where LightGBM's efficiency in handling large datasets provides robust initial predictions, CatBoost's strength in processing categorical features (e.g., StoreType, Promo) corrects residual errors, and XGBoost's regularization refines the final output. Compared to traditional methods in Figure 4.2, the hybrid ensemble's predictions are markedly more aligned with actual sales, reflecting its superior ability to model complex interactions among features like temporal trends, promotional effects, and store characteristics.

The implications of Figure 4.3's results are significant for Rossmann stores. The tight alignment of predicted and actual sales ensures that inventory decisions—such as setting reorder points and safety stock levels—can be based on reliable forecasts. For instance, accurate predictions for high sales values enable precise replenishment during peak demand periods (e.g., holidays), while the model's performance for low sales values supports efficient inventory management during quieter periods. However, the presence of minimal outliers suggests that some store-day combinations, particularly those with sparse or noisy data, may still pose challenges, warranting further refinement in preprocessing or feature engineering to enhance robustness that is addressed with fuzzy generalization techniques in the case study 2.

Figure 4.4 presents a time series plot comparing actual (red) and predicted (blue) sales over the first 200 days of the test set for Case Study 1. The hybrid boosting ensemble closely tracks actual sales trends, accurately capturing both peaks (e.g., days 10, 50, and 150) and troughs (e.g., day 100). Notably, at peak sales around day 150, where actual sales reach approximately 16,000, the predicted sales closely follow the actual value, highlighting the model's robustness in handling demand spikes. However, in some periods, particularly during low-sales periods or regions with high uncertainty or noisy data, the model struggles to achieve precise prediction and tracking. This limitation is addressed in Case Study 2 through the application of the fuzzy generalization technique, which smooths data variability and enhances the ensemble's resilience to uncertainty, resulting in improved forecasting accuracy.

This smooth tracking during peak periods underscores the ensemble's capability to adapt to the dynamic nature of retail demand, driven by factors such as promotional campaigns, seasonal trends, and holidays. The sequential boosting process, where each model corrects the residuals of its predecessor, ensures that the ensemble captures both short-term fluctuations (e.g., daily promotional effects) and longer-term trends (e.g., seasonal patterns). Compared to traditional methods like ARIMA or MA, which struggle with multivariate volatility (as evidenced by the poor performance in Figure 4.2), the hybrid ensemble's performance in Figure 4.4 represents a significant advancement.

The operational implications of Figure 4.4's results are profound for supply chain management. Smooth forecasting of sales trends supports the OUTL inventory policy by providing reliable demand forecasts that inform near optimal reorder points and order quantities. This reduces inventory costs (e.g., holding and ordering costs) by aligning stock levels with actual demand, particularly during volatile periods. The challenges observed in these periods or noisy data scenarios, as seen in Figure 4.4, highlight the importance of the fuzzy generalization technique introduced in Case Study 2. By discretizing features to mitigate uncertainty, the fuzzy hybrid ensemble enhances prediction stability, as evidenced by the improved performance metrics in Table 3.2, ensuring more consistent tracking across all demand scenarios.

5.1.2 Predictive Performance of the Fuzzy Hybrid Boosting Ensemble in Case Study 2

Table 3.2 presents the forecasting performance metrics for Case Study 2, comparing the fuzzy hybrid boosting ensemble against LightGBM and LightGBM+CatBoost on a subset of Rossmann stores. The fuzzy ensemble achieves a MAE of 209.21, a RMSE of 272.99, a MAPE of 3.53%, and a Pred(x=10%) of 97.11% on the test set. The high Pred(x=10%) value indicates that 97.11% of predictions fall within 10% of actual sales, a substantial achievement for retail forecasting where precision is critical.

The fuzzy ensemble's superior performance stems from the integration of fuzzy generalization, which discretizes features into 12 categories to smooth data variability and mitigate the impact of noise and outliers. This preprocessing step enhances the ensemble's ability to handle uncertainty inherent in the Rossmann dataset, such as fluctuations driven by promotions or holidays. Compared to Case Study 1 (Table 3.1), where the generic ensemble achieved an MAE of 416.77 and RMSE of 608.66, Case Study 2's lower error metrics reflect the fuzzy ensemble's improved precision, albeit on a smaller subset of stores with potentially less variability. The MAPE reduction from 6.48% (Case Study 1) to 3.53% and the Pred(x=10%) increase from 81.12% to 97.11% further highlight the fuzzy generalization's effectiveness in refining predictions, particularly for complex demand patterns.

Figure 4.5 illustrates a scatter plot comparing actual versus predicted sales on the test set for the fuzzy hybrid boosting ensemble in Case Study 2. Visually compared to Case Study 1's scatter plot (Figure 4.3), Figure 4.5 shows fewer outliers, suggesting that the fuzzy ensemble performs consistently on the subset of stores analyzed. This enhanced precision is attributable to fuzzy generalization, which smooths noisy data and reduces sensitivity to outliers.

The tight clustering in Figure 4.5 implies that the fuzzy ensemble effectively captures the demand dynamics specific to the analyzed stores, such as promotional effects or temporal trends. The fuzzy ensemble leverages the sequential boosting of LightGBM, CatBoost, and XGBoost, augmented by fuzzy preprocessing, to achieve robust predictions. For Rossmann stores, this accuracy supports efficient inventory management by ensuring that stock replenishment aligns closely with actual demand, reducing holding costs during low-sales periods and preventing stockouts during peaks. However, the narrower sales range in Figure 4.5 compared to Case Study 1 suggests that the subset may exclude extreme demand scenarios, which could limit the generalizability of these results to the full dataset.

Compared to Case Study 1 (Figure 4.4), where the generic ensemble struggled with precise tracking in some low-sales periods due to uncertainty or noisy data, Figure 4.6 shows improved consistency, with predictions closely following actual values across diverse demand scenarios. This enhancement is driven by fuzzy generalization, which mitigates data variability and enhances the ensemble's resilience to uncertainty, as evidenced by the smoother alignment of predicted and actual sales curves.

The fuzzy ensemble's ability to track temporal patterns effectively supports operational planning for Rossmann stores. Accurate predictions of demand spikes (e.g., day 150) enable timely inventory replenishment, while precise forecasting during low-sales periods (e.g., day 100) prevents overstocking. Unlike Case Study 1, where noisy data occasionally hindered prediction accuracy, the fuzzy ensemble's preprocessing ensures more stable performance, aligning with the study's objective to model demand volatility. However, slight deviations in certain periods suggest that further tuning of fuzzy clustering parameters (e.g., adjusting the number of categories, K=12) could enhance performance, particularly for stores with unique demand profiles.

5.2 Inventory improvement Outcomes

Table 3.3 compares inventory improvement metrics for a 7-day lead time between the generic boosting ensemble and the fuzzy hybrid boosting ensemble. The fuzzy ensemble yields a safety stock 5.9% higher than the generic ensemble, a reorder point 9.6% lower, an order quantity 12.5% lower, an order-up-to-level 10.9% lower, and a total cost of 22.398 million USD, which is 12.5% lower than the generic ensemble's 25.600 million USD. These results indicate that the fuzzy ensemble adopts a more conservative safety stock strategy to mitigate demand uncertainty while achieving leaner inventory levels and substantial cost savings.

The fuzzy ensemble's performance is driven by its enhanced forecasting accuracy, as fuzzy generalization smooths data variability and reduces sensitivity to noise, enabling precise demand predictions. This precision informs the OUTL policy, improving reorder points and order quantities to align closely with actual demand. The 12.5% reduction in total cost—encompassing purchase, holding, and ordering costs—reflects the fuzzy ensemble's ability to minimize overstocking, as evidenced by the reduced order-up-to-level and order quantity. The higher safety stock, however, introduces a trade-off, slightly increasing holding costs to achieve a 95% service level, which enhances customer satisfaction by reducing stockouts during demand spikes.

Figure 4.7 presents a bar chart comparing total costs across lead times of 5, 10, 15, 20, and 25 days for the generic and fuzzy boosting ensembles. The fuzzy ensemble (dark bars) consistently outperforms the generic ensemble (light bars), achieving cost reductions ranging from 12.5% at a 5-day lead time (22.379 million USD vs. 25.578 million USD) to 10.5% at a 25-day lead time (22.570 million USD vs. 25.795 million USD). This consistent cost advantage underscores the fuzzy ensemble's robustness in determining inventory decisions across varying operational constraints, driven by its ability to handle demand uncertainty through fuzzy generalization.

The cost savings in Figure 4.7 result from the fuzzy ensemble's precise demand forecasts, which reduce overstocking and unnecessary ordering, as seen in Table 4.3's lower order quantities and reorder points. Unlike Case Study 1, which focused on forecasting, Case Study 2's results demonstrate the operational translation of improved predictions into cost efficiency. The fuzzy ensemble's ability to maintain cost reductions across longer lead times is particularly valuable for Rossmann stores to address the sensitivty analysis.

For Rossmann stores, these results offer actionable benefits. The cost savings enable efficient resource allocation, while the improved service level minimizes stockouts, enhancing customer satisfaction. The fuzzy ensemble's robustness across lead times, as shown in Figure 4.7, ensures adaptability to varying supply chain conditions, addressing the problem statement's challenge of demand volatility.

5.3 Comprehensive Insights from Hybrid Boosting Ensemble and Performance Analysis

The hybrid boosting ensemble model, integrating LightGBM, CatBoost, and XGBoost with fuzzy generalization, has provided a comprehensive understanding of demand forecasting dynamics in the Rossmann retail context, revealing the interplay of multivariate features like promotions, store attributes, and temporal variables. The sequential architecture, coupled with the OUTL inventory policy, illuminated how accurate demand predictions cascade into improved inventory decisions, reducing total costs and achieving a high service level. Performance analysis through metrics such as RMSE, MAE, MAPE, and prediction accuracy, supported by visualizations like scatter and time series plots, demonstrated the model's superior predictive accuracy and robustness in capturing demand volatility compared to traditional and single-model approaches. The use of fuzzy generalization enhanced model stability by smoothing noisy data, while the focus on a single-echelon retailer system highlighted its practical applicability, though it underscored the need for further exploration in multi-echelon contexts to fully capture supply chain complexities.

5.4 Consistency

Our findings align with several recent studies utilizing advanced machine learning and hybrid ensemble techniques to enhance demand forecasting and inventory improvement in retail supply chains. Ahmed et al. (2024) proposed a Switching-Based Forecasting Approach (SBFA) that dynamically selects optimal ML models, achieving significant reductions in RMSE and inventory costs, consistent with our hybrid boosting ensemble. Similarly, Seyedan et al. (2023) employed an ensemble of deep learning models integrated with the OUTL policy, demonstrating improved forecasting accuracy and reduced inventory costs, resonating with our numerical results obtained from the Fuzzy Boosting Ensemble . Ul Haq Qureshi et al. (2024) highlighted the importance of incorporating exogenous variables like weather into deep learning models for Rossmann dataset forecasting, aligning with our use of multivariate features (e.g., promotions, store type) to capture complex demand patterns, thereby enhancing prediction accuracy.

Furthermore, studies like Weng et al. (2020) and Mittal (2024) emphasize the efficacy of hybrid ML models in handling demand volatility, a finding consistent with our application of fuzzy generalization to mitigate uncertainty, resulting in a meanigful reorder point and reduced order quantity in the Fuzzy Boosting Ensemble. Alsulamy (2025) compared boosting algorithms (CatBoost, XGBoost, LightGBM) for construction delay prediction, noting their efficiency in categorical data handling, which supports our sequential ensemble's design leveraging CatBoost's categorical feature processing and LightGBM's computational efficiency for the Rossmann dataset. The consistent emphasis on operational integration, as seen in Seyedan et al. (2023), aligns with our framework's seamless linkage of the hybrid boosting ensemble with the OUTL policy for single-echelon retailer inventory improvement, contrasting with studies like Abbasimehr et al. (2020) that lack such integration. These alignments validate the proposed method's robustness in addressing demand uncertainty and improving inventory, reinforcing its practical applicability in retail supply chain management as discussed in Chapter 5.

5.5 Inconsistencies

Despite the alignments with recent studies, some research presents findings that diverge from our proposed hybrid boosting ensemble approach. Abbasimehr et al. (2020) focused on a multi-layer LSTM model for demand forecasting, emphasizing temporal dependencies but reporting high computational costs and limited integration with inventory systems, contrasting with our framework's computationally efficient sequential ensemble and OUTL policy integration, which reduced inventory costs. This discrepancy likely stems from their focus on single-product datasets versus our multivariate Rossmann dataset analysis. Similarly, Kohli et al. (2021) found KNN regression prone to overfitting and less effective for large datasets, differing from our ensemble's robust performance, possibly due to their reliance on simpler ML models without ensemble techniques or fuzzy generalization to handle demand volatility.

5.6 Limitations of the Study

The development of the hybrid boosting ensemble model addressed several challenges outlined in the Introduction, including data quality, demand volatility, and integration with inventory systems. However, a key limitation lies in the study's reliance on the Rossmann dataset, which is specific to a single retail chain and focuses on a single-echelon retailer system. While this focus aligns with the research objective of streamlining inventory decisions for Rossmann stores, it restricts the model's generalizability to multi-echelon supply chains, where complex interactions between suppliers, warehouses, and retailers introduce additional variability. The dataset's structured nature, with predefined features like sales, promotions, and store attributes, may not fully capture the broader uncertainties present in diverse retail contexts, such as varying product lifecycles or supply chain disruptions. Furthermore, preprocessing steps, such as clipping sales between 2,000 and 16,000 to mitigate outliers and applying fuzzy generalization with a fixed K=12 categories, were necessary to handle data quality and volatility challenges but may simplify extreme demand scenarios, potentially impacting prediction accuracy for low or high sales periods.

In addition, this study assumes that sales data is a direct indicator of customer demand. However, in the Rossmann dataset, zero sales typically occur on days when stores are closed, and no stock-level data is provided to confirm whether shortages occurred. As a result, these instances were not interpreted as stockouts, and shortage costs were excluded from the model. A 95% service level was adopted to reduce the risk of stockouts through safety stock. While this approach helps maintain product availability, the lack of explicit shortage data limits the model's ability to capture the full cost implications of unmet demand, which may affect inventory decisions in real-world applications.

Another notable limitation is the absence of product shelf life information in the Rossmann dataset. This constraint prevents the inclusion of obsolescence costs, which are particularly significant in sectors such as pharmaceuticals, where expired products can lead to both financial losses and regulatory issues. Without shelf life data, the model cannot evaluate the trade-off between overstocking and the risk of product expiry, which is a critical consideration for inventory optimization in perishable or sensitive product categories.

Another limitation stems from the time and resource constraints of a master's thesis, which limited the scope of model experimentation and real-world validation. The integration of the hybrid boosting ensemble with the OUTL policy effectively addressed the challenge of aligning forecasting with inventory management. However, the computational intensity of training LightGBM, CatBoost, and XGBoost sequentially, coupled with grid search for hyperparameter optimization, restricted the exploration of alternative ensemble configurations or advanced uncertainty modeling techniques, such as dynamic fuzzy clustering or real-time adaptive learning. Additionally, the focus on a single-echelon system, while justifiable for isolating retailer-specific inventory decisions, omits the complexities of multi-echelon dynamics, limiting insights into upstream supply chain interactions. Empirical validation was confined to simulations due to time constraints, and real-world implementation was not feasible, necessitating cautious interpretation of the results.

Moreover, the current model is designed for B-class items using a periodic (R, s, S) inventory review system. To adapt this approach for A-class items—where demand is typically more frequent and critical—a continuous (s, S) review system would be more appropriate, with the parameter R set to zero. This shift would ensure more responsive inventory control. Additionally, the forecasting component of the model would require recalibration, including adjustments to the hybrid ensemble's parameters to reflect the different demand patterns and urgency associated with A-class items. These modifications are necessary to maintain prediction accuracy and inventory efficiency when applied to datasets with distinct characteristics. However, this study considers only a single product, while real-world pharmacy or retail operations involve managing thousands of SKUs, which would require a more complex multi-item inventory strategy, including product prioritization, classification (e.g., ABC analysis), and coordinated replenishment policies.

Future studies could overcome these limitations by testing the model across multi-echelon datasets, incorporating dynamic uncertainty modeling, and conducting pilot deployments to validate practical applicability.

5.7 Future Works

While this study successfully achieved its objectives of improving demand forecasting accuracy and supply chain efficiency, several avenues for future research remain open to further enhance the proposed framework:

- **Algorithmic Evolution:** Integrating advanced machine learning and deep learning models could complement the fuzzy hybrid boosting ensemble, capturing complex temporal patterns to further reduce forecasting errors, particularly for stores with highly volatile demand.
- **Development of Multi-Echelon System Model:** Enhance the model's generalizability by testing it across multi-echelon supply chain architecture, incorporating diverse supply chain levels (e.g., suppliers, distributors, retailers) to evaluate its performance in more complex, interconnected scenarios beyond the single-echelon Rossmann dataset.
- **Extension to Multi-Item Inventory Management:** This study focused on a single product; however, real-world retail and pharmacy operations involve managing thousands of items with varying demand patterns and service levels. Future work could extend the current framework to support multi-item inventory strategies, incorporating ABC classification, coordinated replenishment, and optimization across a broader product portfolio.
- **Addressing Low-Sales Periods:** The current model showed some limitations in accurately predicting sales during low-demand periods. Future studies could focus on developing specialized techniques, such as adaptive weighting of historical data or incorporating external factors (e.g., weather, economic indicators), to improve forecasting performance in these scenarios.
- **Alternative Inventory Policies:** Beyond the OUTL policy, future research could explore alternative inventory management strategies, such as (s, S) policies or dynamic safety stock adjustments, to create a more universally applicable framework that balances cost efficiency with responsiveness across diverse retail contexts.
- **Practical Implementation:** While simulations provided valuable insights, validating the proposed framework in real-world retail environments would offer the ultimate test of its efficacy. Future work could involve pilot implementations with retail partners to assess the model's performance under real-world demand variability, operational constraints, and the impact of potential shortages on inventory decisions and service levels.

By systematically exploring these avenues, future research can further refine the proposed hybrid boosting ensemble framework, leading to more substantial contributions to supply chain analytics and retail operations. This study not only addresses existing challenges in demand forecasting and inventory improvement but also lays a strong foundation for future advancements in this dynamic field.

CONCLUSION

In the fast-paced retail sector, accurate demand forecasting and efficient inventory management are critical for overcoming demand uncertainty and enhancing supply chain performance. This study proposed a novel hybrid boosting ensemble framework, integrating LightGBM, CatBoost, and XGBoost with fuzzy generalization and the Order-Up-To-Level (OUTL) inventory policy, to improve demand forecasting and inventory improvement for Rossmann stores. Utilizing the multivariate Rossmann dataset, covering 1,115 stores over 942 days, the framework effectively addressed challenges such as data quality, demand volatility, and integration with inventory systems, delivering robust forecasting accuracy and operational efficiency.

The hybrid boosting ensemble exhibited outstanding forecasting performance across two case studies. In Case Study 1 (Generic Boosting Ensemble), the model achieved a test set MAE of 416.77, RMSE of 608.66, MAPE of 6.48%, and Pred(x=10%) of 81.12%, significantly surpassing traditional models like ARIMA and single-model approaches like LightGBM. In Case Study 2 (Fuzzy Hybrid Boosting Ensemble), performance was further enhanced on a subset of stores, with an MAE of 209.21, RMSE of 272.99, MAPE of 3.53%, and Pred(x=10%) of 97.11%, indicating that nearly all predictions were within 10% of actual sales. Scatter plots and time series visualizations confirmed the ensemble's ability to capture complex demand patterns driven by promotions, holidays, and store-specific attributes, with fuzzy generalization bolstering resilience against data noise and uncertainty.

Coupling the ensemble's forecasts with the OUTL policy yielded improved inventory outcomes. For a 7-day lead time, the fuzzy ensemble reduced total inventory costs to $22.40 million, lowered the reorder point to 104,963, and decreased the order-up-to-level to 190,181, while achieving a 97% service level. These results reflect a streamlined inventory strategy that minimizes overstocking and stockouts, with consistent cost savings across lead times of 5 to 25 days, as illustrated in cost comparison charts. The slightly higher safety stock of 19,742 units ensured robustness against demand fluctuations, balancing cost efficiency with customer satisfaction.

The study's key contributions include:

- **Innovative Forecasting Model**: A sequential boosting ensemble, enhanced by fuzzy generalization, leveraging the complementary strengths of LightGBM, CatBoost, and XGBoost to deliver highly accurate demand predictions for retail environments.
- **Robust Uncertainty Management**: Integration of fuzzy generalization to smooth data variability, improving prediction stability and supporting reliable inventory decisions under volatile demand conditions.
- **Practical Inventory improvement**: Seamless integration of forecasts with the OUTL policy, improving reorder points, safety stock, and total costs while enhancing service levels for Rossmann store operations.
- **Comprehensive Evaluation and Insights**: Rigorous simulations using the Rossmann dataset, supported by visualizations and benchmark comparisons, providing actionable insights for retail supply chain management.

This research successfully tackled the challenges of data limitations, demand volatility, and system integration, offering a data-driven solution that enhances forecasting precision and inventory efficiency. The fuzzy hybrid boosting ensemble's ability to deliver accurate predictions—evidenced by low MAPE values and high Pred(x=10%)—and cost-effective inventory decisions aligns with recent advancements in supply chain analytics and offers substantial practical value for Rossmann stores. While the single-echelon focus and dataset specificity suggest opportunities for future exploration, such as multi-echelon systems or real-world validation, this study establishes a strong foundation for advancing retail supply chain performance. By integrating advanced machine learning, uncertainty modeling, and inventory improvement, the proposed framework contributes significantly to both academic research and industry practice, paving the way for more resilient and efficient retail operations.

APPENDIX I

PROPOSED ALGORITHM PSEUDOCODE

Input: Training set, Validation set, Testing set

Output: Predictions, RMSE, MAE, MAPE, Pred(x=10%) metrics

Module 1: Data Reading and Integration Read `train.csv`, `store.csv` Merge datasets on `Store` Extract `Year`, `Month`, `Day` from `Date`; drop `Date`

Module 2: Data Preprocessing Sample 100% of data Clip `Sales` between 2000 and 16000 Select features (`Store`, `DayOfWeek`, `Sales`, `Customers`, etc.) Encode categorical features (`StoreType`, `Assortment`, etc.) Fill missing values with 0

Module 3: Fuzzy Generalization and Normalization Normalize features with `MinMaxScaler` Apply fuzzy generalization (12 categories per feature) Normalize features, target with `StandardScaler`

Module 4: Data Splitting Split data into train (80%), test (20%)

Module 5: Boosting Ensemble Training Initialize `LightGBM` (`n_estimators=1000`, `learning_rate=0.05`) Train `LightGBM` on (`X_train`, `y_train`) Compute `LightGBM` residuals Initialize `CatBoost` (`iterations=1000`, `learning_rate=0.05`) Train `CatBoost` on `LightGBM` residuals Compute `CatBoost` residuals Initialize `XGBoost` (`n_estimators=1000`, `learning_rate=0.05`) Train `XGBoost` on `CatBoost` residuals Combine predictions: `LightGBM` + `CatBoost` + `XGBoost`

Module 6: Evaluation and Visualization Evaluate(`model`, `X_test`): `forecasts` = `predict(model, X_test)` `rmse` `calculate_rmse(forecasts, y_test)` `mae` `calculate_mae(forecasts, y_test)` `mape` `calculate_mape(forecasts, y_test)` `pred_x10` `predict(model, X_test[:10])` Append `rmse`, `mae`, `mape` to metrics Visualize `forecasts` vs actuals

Module 7: Output Return predictions, metrics

BIBLIOGRAPHY

Abbasimehr, H., Shabani, M. & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & industrial engineering*, 143, 106435.

Abolghasemi, M., Hyndman, R. J., Tarr, G. & Bergmeir, C. (2019). Machine learning applications in time series hierarchical forecasting. *arXiv preprint arXiv:1912.00370*.

Abolghasemi, M., Beh, E., Tarr, G. & Gerlach, R. (2020). Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers & Industrial Engineering*, 142, 106380.

Ahmed, S., Chakrabortty, R. K., Essam, D. L. & Ding, W. (2024). A switching based forecasting approach for forecasting sales data in supply chains. *Applied Soft Computing*, 167, 112419.

Ali, M. M., Babai, M., Boylan, J. & Syntetos, A. (2015). On the use of Simple Moving Averages for supply chains where information is not shared. *IFAC-PapersOnLine*, 48(3), 1756–1761.

Alsulamy, S. (2025). Predicting construction delay risks in Saudi Arabian projects: A comparative analysis of CatBoost, XGBoost, and LGBM. *Expert Systems with Applications*, 268, 126268.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press. doi: 10.1007/978-1-4757-0450-1.

Brunaud, B., Laínez-Aguirre, J. M., Pinto, J. M. & Grossmann, I. E. (2019). Inventory policies and safety stock optimization for supply chain planning. *AIChE journal*, 65(1), 99–112.

Chopra, S. & Meindl, P. (2021). *Supply Chain Management: Strategy, Planning, and Operation* (ed. 7th). Pearson.

Cukierski, W. (2015). Rossmann Store Sales. Kaggle. Accessed: 2024-08-27.

Ghosh, S. (2020). Forecasting of demand using ARIMA model. *American Journal of Applied Mathematics and Computing*, 1(2), 11–18.

Han, J., Kamber, M. & Pei, J. (2012). Data mining: Concepts and. *Techniques, Waltham: Morgan Kaufmann Publishers*.

Harikrishnan, G., Sreedharan, S. et al. (2025). Advanced short-term load forecasting for residential demand response: An XGBoost-ANN ensemble approach. *Electric Power Systems Research*, 242, 111476.

He, Z. & Yu, S. (2020). Application of LightGBM and LSTM combined model in vegetable sales forecast. *Journal of Physics: Conference Series*, 1693(1), 012110.

Hyndman, R. J. & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (ed. 3rd). OTexts. Consulted at https://otexts.com/fpp3/.

Islam, M. M., Das, P., Rahman, M. M., Naz, F., Kashem, A., Nishat, M. H. & Tabassum, N. (2024). Prediction of compressive strength of high-performance concrete using optimization machine learning approaches with SHAP analysis. *Journal of Building Pathology and Rehabilitation*, 9(2), 94.

Ivanov, D., Tsipoulanidis, A., Schönberger, J., Ivanov, D., Tsipoulanidis, A. & Schönberger, J. (2021). *Basics of supply chain and operations management*. Springer.

Jaipuria, S. & Mahapatra, S. (2021). A hybrid forecasting technique to deal with heteroskedastic demand in a supply chain. *Operations and Supply Chain Management: An International Journal*, 14(2), 123–132.

Jamali, N., Gharib, M. R., Moayyedian, M. & Hedayati-Dezfooli, M. (2024). Machine Learning for Optimizing Macro-ergonomics in Pharmaceutical Supply Chain. *International Journal of Computational Intelligence Systems*, 17(1), 248.

kaggle. (2018). Store Item Demand Forecasting Challenge. Kaggle.

Kersten, W., Blecker, T. & Ringle, C. M. (2019). *Artificial intelligence and digital transformation in supply chain management: innovative approaches for supply chains*. Berlin: epubli GmbH.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74. doi: 10.1016/j.ijforecast.2019.04.014.

Merkuryeva, G., Valberga, A. & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, 149, 3–10.

Mittal, U. (2024). Managing Uncertainty in Fashion Supply Chains: An AI-Based Analysis of Demand Variability and Forecast Precision. *International Research Journal of Engineering and Technology*, 11(1), 476–484.

Ntakolia, C., Kokkotis, C., Karlsson, P. & Moustakidis, S. (2021). An explainable machine learning model for material backorder prediction in inventory management. *Sensors*, 21(23), 7926.

Qureshi, N. U. H., Javed, S., Javed, K., Naqvi, S. M. R., Raza, A. & Saeed, Z. (2024). Demand Forecasting in Supply Chain Management for Rossmann Stores Using Weather Enhanced Deep Learning Model. *IEEE Access*, 12, 145570-145581. Consulted at https://api.semanticscholar.org/CorpusID:273085974.

SAS Institute. An Overview of Light Gradient Boosting Machine (LightGBM) Model in Machine Learning.

Seyedan, M., Mafakheri, F. & Wang, C. (2023). Order-up-to-level inventory optimization model using time-series demand forecasting with ensemble deep learning. *Supply Chain Analytics*, 3, 100024.

Silver, E. A., Pyke, D. F. & Thomas, D. J. (2017). *Inventory and Production Management in Supply Chains* (ed. 4th). CRC Press.

Silver, E. A., Pyke, D. F., Peterson, R. et al. (1998). *Inventory management and production planning and scheduling*. Wiley New York.

Stevenson, W. J. (2021). *Operations Management* (ed. 14th). McGraw-Hill Education.

Stoilov, T. & Stoilova, K. (2024). Demand Forecasting to Support Inventory Management. *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, pp. 123–128.

Vaishnavi, V. & Kuechler, W. (2004). Design research in information systems.

Weng, T., Liu, W. & Xiao, J. (2020). Supply chain sales forecasting based on lightGBM and LSTM combination model. *Industrial Management & Data Systems*, 120(2), 265–279.