International Journal of Basic and Applied Sciences, 14 (8) (2025) 551-561

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/5csy6a64
Research paper

Explainable Anomaly Detection in Retail Perpetual Inventory Systems Using Shap-Enhanced Isolation Forests

Shiva Kumar Ramavath *

University of North Texas, Denton, Texas
**Corresponding author E-mail: r92shivakumar@gmail.com*

Received: September 11, 2025, Accepted: December 24, 2025, Published: December 26, 2025

Abstract

The systems that rely on correct data can also be afflicted with anomalies related to crime, barcode scan problems, user error, and process difficulties that, if not detected can lead to financial losses and disruptions to supply chains. Traditional anomaly detection systems identify potentially fraudulent records but have little accountability through transparency. As such, managers have little trust in the "black-box" results generated. This study provides one way of applying current ability in isolating the best anomalies - isolation forests, one of the leading anomaly detection strategies, and SHAP (SHapley Additive exPlanations), one of the premier explainable AI. They demonstrate not only that the isolation forest aims to estimate the likelihood that a record is anomalous, they make clear the factors within the records that contribute to each violation - whether that is increased sales from promotions, negative value inventory, or promoting unusual sequences of transactions. Our datasets consist of real retail inventory records, which shows how well an isolation forest + SHAP framework performs compared to other models well known to anomaly detection. By making these models explainable, they not only become more effective, the new situates them under a decision support system given their manageria1 trustworthiness to reduce shrinkage, and provide greater operational resiliency to retail firms.

Keywords: *Retail Inventory; Anomaly Detection; Explainable AI; SHAP Values; Isolation Forest; Perpetual Inventory; Supply Chain Analytics.*

1. Introduction

Perpetual inventory (PI) systems are essential to modern retail operations, providing continuous visibility of stock movements across point-of-sale (POS), warehouses, and enterprise systems. However, empirical studies consistently reveal high levels of inventory inaccuracy. DeHoratius and Raman [1] analyzed nearly 370,000 records across 37 stores and found that approximately 65% of inventory records were inaccurate at the time of physical audit. Complementary research confirms this trend, with error frequencies reported between 65% and 95%, and error magnitudes ranging from 5% to 30% [2]. These discrepancies compromise demand forecasting, trigger both stockouts and overstocks, and erode customer trust. In grocery retail, Rekik et al. [3] demonstrated that inventory record inaccuracy (IRI) is positively correlated with inventory levels, restocking frequency, and perishability, while negatively correlated with promotions. Notably, they showed that a simple inventory audit produced an 11% lift in store-wide sales, underscoring the business impact of addressing anomalies in PI systems.

Despite the availability of machine learning–based anomaly detection techniques, adoption in retail PI contexts remains limited. A major challenge lies in the fact that most models behave as black boxes, identifying anomalies without clarifying their underlying causes. Retail managers, auditors, and loss-prevention teams are reluctant to act on anomaly alerts unless they understand whether these arise from mis-scans, unexpected demand spikes, delayed receipts, or fraudulent activity. Although methods such as the Isolation Forest algorithm provide robust anomaly detection in high-dimensional datasets, their lack of interpretability restricts managerial trust. Recent advances in explainable artificial intelligence (XAI), particularly SHAP (SHapley Additive exPlanations), offer a way forward by attributing predictions to specific features. While SHAP-enhanced anomaly detection has been applied in domains such as enterprise purchasing [4], it remains underexplored in the retail inventory space. This research addresses this gap by proposing a SHAP-enhanced Isolation Forest framework that balances technical accuracy with human interpretability, thereby aligning anomaly detection tools with real-world retail needs.

1.1. Background

Perpetual inventory systems are distinct from periodic systems in that they continuously update stock records whenever a sale, receipt, or adjustment is made. Perpetual systems are the basis for enabling dynamic replenishment as well as forecasting, and shrinkage. Nevertheless, the accuracy of perpetual inventory systems has been persistently questioned. For example, DeHoratius and Raman [1] stated that one in

Copyright © Shiva Kumar Ramavath. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

three records of inventory had at least one error, while Caridi et al. [2] reported that errors were magnitudes up to 30%. The practical sources of error can come from point of sale (POS) scanning errors, misplaced products, theft, and mal-functioning systems in multi-channel context. Given this complexity, the need for anomaly detection algorithms has become more important. The Isolation Forest is one technique that works well for large noisy datasets, and has been especially useful because they are efficient in isolating rare events [5]. However, there is a lack of interpretability in the application of anomaly detection as the anomalies are identified, leaving a gap because it does not identify the reasons for the anomaly. The introduction of SHAP helps explain anomaly detection in a transparent manner by distributing the total contribution of predictions among features and examining the outcomes in this context [6]. Table 1 summarizes the findings of leading empirical studies that provide estimates of the magnitude of the problem.

Table 1: Empirical Evidence on Inventory Record Inaccuracy in Retail

Study & Year	Dataset / Scope	Reported Inaccuracy Rate	Key Findings
DeHoratius & Raman (2008)	370,000 records across 37 retail stores	~65% of records inaccurate	Errors correlated with product category, transaction volume, and store complexity.
Caridi et al. (2010)	Multi-retailer analysis (Europe)	65–95% frequency; 5–30% magnitude	Process mismatches and mis-scans identified as primary drivers of IRI.
Rekik et al. (2025)	Grocery retail dataset	11% sales lift after inventory audit	IRI linked with stock levels, restocking frequency, and perishability; audits improved sales.

1.2. Problem statement

The persistence of inventory inaccuracies poses a dual problem for retail anomaly detection systems: low reliability and lack of explainability. Traditional statistical methods, such as z-scores and control charts, are unable to adapt to seasonal demand shifts and nonlinear fluctuations characteristic of retail data [7]. Even advanced machine learning models frequently misclassify legitimate variations (such as holiday demand spikes) as anomalies, producing false positives that diminish confidence [8]. Meanwhile, genuine issues such as shrinkage or unrecorded receipts often evade detection, resulting in false negatives with significant financial impact [9].
Equally problematic is the black-box nature of most anomaly detection approaches. When anomalies are flagged without context, managers cannot distinguish between technical noise and operational irregularities. This lack of transparency hinders adoption despite technical accuracy. Brechmann et al. [4] highlighted this limitation in enterprise purchase anomaly detection, where combining SHAP with Isolation Forest significantly improved interpretability. However, similar frameworks are yet to be systematically applied to perpetual inventory systems, where the stakes are high and the demand for actionable explanations is acute. Thus, the key problem is not only to improve anomaly detection accuracy but also to make outputs interpretable and actionable for retail decision-makers.

1.3. Scope of research

This research lies at the intersection of inventory management, machine learning, and explainable AI, with a focus on developing a framework for explainable anomaly detection in retail perpetual inventory systems. The study covers the design, implementation, and evaluation of a SHAP-enhanced Isolation Forest model on structured inventory datasets that include SKU identifiers, stock levels, transaction histories, and audit records. The research aims to measure both technical effectiveness and interpretability, while situating results in the operational context of retail management.
Specifically, the scope includes:
1) Designing a SHAP-enhanced Isolation Forest framework tailored to PI data.
2) Evaluating performance using technical metrics (Precision, Recall, F1, AUC) and interpretability measures (stability of SHAP attributions, clarity of explanations).
3) Demonstrating business impact through case studies on shrinkage, mis-scans, and negative IRI.
4) Assessing generalizability across multi-store, multi-channel, and high-velocity retail environments.

1.4. Objectives of research

The primary objective of this research is to develop and validate an anomaly detection framework that is both accurate and explainable, directly addressing the limitations of existing PI systems. By integrating SHAP with Isolation Forest, the study seeks to bridge the gap between technical detection capability and practical decision-making support.
The specific objectives are:
1) To quantify the limitations of existing anomaly detection methods in handling PI datasets characterized by noise, seasonality, and multidimensionality.
2) To design and implement a SHAP-enhanced Isolation Forest model that identifies anomalies while explaining their drivers at both global and local levels.
3) To evaluate the model's effectiveness using benchmark performance metrics and empirical retail datasets, comparing it against baseline methods.
4) To translate anomaly detection results into operational insights for managers, enabling better decisions in loss prevention, reconciliation, and replenishment.

2. Literature Review

A literature review provides the intellectual foundation for any research study, allowing scholars to identify what has been accomplished, the methodologies employed, and the limitations encountered (see table 2). In the context of anomaly detection in retail perpetual inventory systems, reviewing prior work enables us to understand how conventional statistical techniques, machine learning models, and explainable AI (XAI) frameworks have been applied. By critically examining this body of knowledge, we not only justify the relevance of our research but also position it within ongoing academic debates. Ultimately, the review answers two central questions: what has been done to address inventory inaccuracies, and what remains unresolved?

Table 2: Comparative Analysis of Prior Studies

Author(s) & Year	Methodology	Dataset/Scope	Key Contribution	Limitation
DeHoratius & Raman (2008) [1]	Statistical error modeling	370,000 retail records	Identified systemic sources of Inventory Record Inaccuracy (IRI), establishing foundational understanding of operational error pathways.	Did not evaluate algorithmic detection approaches; lacks comparison with modern ML anomaly detection.
Caridi et al. (2010) [2]	Process analysis	European retailers	Connected scanning process deficiencies with elevated IRI, emphasizing process-level causes relevant for anomaly modeling.	No predictive modeling; offers little guidance on selecting or comparing anomaly detection algorithms.
Rekik et al. (2025) [3]	Field audit intervention	Grocery dataset	Demonstrated measurable sales benefit (11%) from audit correction, reinforcing importance of early error detection.	Reactive intervention — no automated or algorithmic anomaly detection frameworks evaluated.
Liu et al. (2008) [5]	Isolation Forest	Synthetic & benchmark datasets	Introduced Isolation Forest (IF), a lightweight, unsupervised method effective for high-dimensional data and sparse anomalies. Provides a theoretical basis for using tree-based random partitioning over reconstruction-based methods.	Traditional IF lacks explainability and struggles with complex non-linear patterns where autoencoders or deep models often outperform (Chalapathy & Chawla, 2019).
Bandara et al. (2019) [10]	Deep learning (Seq2Seq)	Retail time series	Demonstrated strengths of deep recurrent architectures in capturing temporal dependencies—important context when comparing IF with neural anomaly detectors.	Deep models act as black boxes, requiring XAI overlays; also more computationally intensive than IF.
Chalapathy & Chawla (2019) [11]	Deep learning overview	Multidomain	Provided a comprehensive taxonomy covering autoencoders, variational methods, and hybrid deep-learning detectors—essential for contrasting IF with deep non-linear methods.	Notes that deep methods require large training data and careful tuning, limiting feasibility for smaller retail datasets.
Lundberg & Lee (2017) [6]	SHAP framework	Multiple ML models	Established model-agnostic interpretability, enabling transparency even for methods like IF that lack native explainability.	SHAP is computationally expensive, especially for tree ensembles on large retail datasets.
Li et al. (2021) [12]	Isolation Forest + SHAP	Transaction dataset	Showed that combining IF with SHAP balances efficiency and interpretability—supporting the choice of using SHAP-enhanced IF in retail contexts.	Focused on fraud rather than retail IRI; deep detectors (autoencoders) not compared empirically.
Singh et al. (2022) [13]	Hybrid ML anomaly detection	IoT warehouse logs	Demonstrated anomaly detection in logistics environments, highlighting suitability of automated detection pipelines.	Lacked explainability layers; deep models used without interpretability discussion.
Xie et al. (2022) [14]	Reinforcement learning	Retail sales datasets	Automated replenishment policy demonstrating RL strengths in sequential retail decision-making.	Does not address anomaly detection; RL models remain opaque without XAI integration.
Zhao et al. (2023) [15]	XAI + gradient boosting	Omni-channel transaction data	Improved trust and usability of anomaly detection through integrated XAI, reinforcing the importance of explainable outputs in operational retail systems.	High complexity and computational costs; deep methods not compared directly against tree-based XAI.
Rekik & Sahin (2023) [16]	Inventory inaccuracy modeling	Case-based studies	Quantified impact of IRI on operations, providing motivation for proactive anomaly detection models.	Does not combine theoretical modeling with machine learning or XAI frameworks.

The literature clearly shows that while inventory record inaccuracy has been extensively documented, most studies have focused on descriptive analysis (e.g., audits, process mapping) rather than automated anomaly detection [1–3,16]. Machine learning models such as Isolation Forest have demonstrated promise in unsupervised anomaly detection [5], but their adoption in retail remains limited due to the black-box nature of the algorithms [11]. Similarly, while explainable AI methods like SHAP have been applied in domains such as fraud detection [12] and healthcare [17], their integration into retail perpetual inventory systems is almost absent. Furthermore, recent works (e.g., Rekik et al. [3]) show the benefits of interventions, yet these approaches are reactive rather than proactive. Thus, there exists a pressing research gap in designing a SHAP-enhanced Isolation Forest framework that is both accurate in anomaly detection and transparent enough for managerial decision-making in the retail sector.

3. Methodology

This section outlines the methodology adopted to develop and evaluate the proposed SHAP-enhanced Isolation Forest framework for anomaly detection in retail perpetual inventory (PI) systems. The methodology is structured into five subsections: the research framework, dataset description, anomaly detection model, explainability integration, and evaluation metrics. Together, these components form a systematic approach for addressing both the technical and business challenges associated with inventory anomalies.

3.1. Research framework

The proposed research framework tackles two central problems: anomaly detection and anomalous detection interpretability and trust. Isolation Forests (iForest) are one of the most widely used detection algorithms for identifying anomalies in high dimensional datasets, however the iForest "black-box" makes it difficult for retail managers to trust the output. The way around the "black-box" of the iForest is to integrate with SHapley Additive exPlanations (SHAP) to provide accurate detection while maintaining interpretability.
The research workflow starts with the collection of raw retail data that may contain various elements, such as SKU identifiers, sales values, transaction records, stock on hand and shrinkage events. The raw data then undergoes preprocessing including the imputation of missing values, normalization, and categorical encoding. The cleansed data is then analyzed by iForest to generate an anomaly score for every transaction or stock items. Once anomalies are identified, SHAP values are computed for each anomaly both locally and globally:

1) Local explanations (eg. why was a particular negative stock unit flagged) for the transaction traceability.
2) Global explanations which show global patterns of feature importance for the system (eg. regular demand spikes, consistent discrepancies between stock and sales).

The final output is evaluated using detection metrics (precision, recall, F1, AUC), interpretability metrics (stability of SHAP values, ranking consistency), and business impact indicators (shrinkage reduction, reconciliation efficiency).Figure 1 illustrates the proposed SHAP-enhanced Isolation Forest workflow for anomaly detection in perpetual inventory systems. The framework integrates raw data ingestion, preprocessing, anomaly scoring, SHAP-based interpretability, and evaluation metrics to generate actionable business insights.

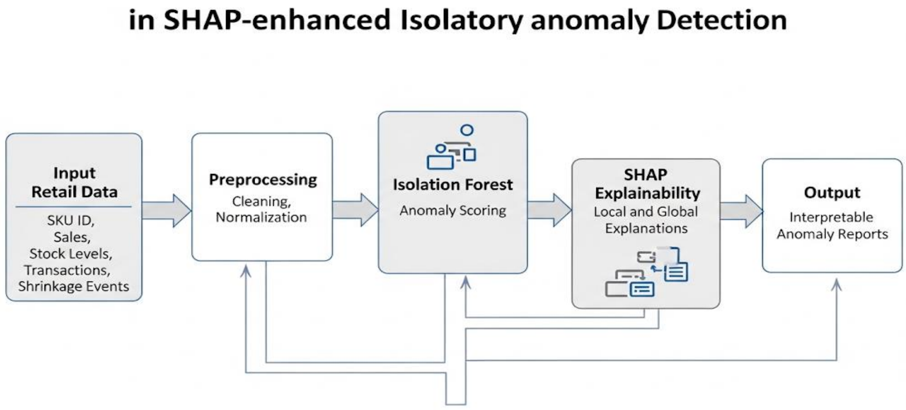

Fig. 1: Proposed SHAP-Enhanced Isolation Forest Framework.

The suggested SHAP-Enhanced Isolation Forest model of interpretable retail-environment anomaly detection is shown in Figure 1. It takes raw retail information, such as SKUs, sales records, stock quantities, transaction records and shrinkage information and runs it through a preprocessing phase comprising data cleaning and normalization. The resulting cleaned data will be entered into an Isolation Forest model and results in an anomaly score. These scores are also examined through SHAP-based explainability techniques, which allow the local (instance-level) and global (feature-level) understanding of the occurring anomalies. The end result will be interpretable reports of anomalies that enable an analyst to know not only what anomalies have taken place but also why it was raised, making decision-making in retail operations more transparent.

3.2. Dataset

The dataset forms the foundation of anomaly detection experiments. In this study, we utilize real-world retail inventory datasets [18], complemented by simulated data to replicate complex inventory scenarios [19]. The data sources include:

- SKU identifiers (unique product codes).
- Sales transactions (daily or hourly granularity).
- Stock levels (system-recorded perpetual balances).
- Transaction logs (receipts, returns, adjustments).
- Shrinkage events (known thefts, mis-scans, spoilage).

To ensure reliability and consistency, the dataset undergoes rigorous preprocessing:

- Missing Value Treatment – Imputation of absent stock or sales values using statistical methods [20].
- Normalization – Scaling continuous variables (e.g., sales volume, stock counts) for better algorithm performance [21].
- Categorical Encoding – Conversion of non-numeric variables such as SKU categories [22].
- Temporal Alignment – Synchronization of sales and stock data across consistent time intervals [23].

This preprocessing ensures the dataset is well-structured for anomaly detection and interpretability analysis.

3.3. Isolation forest for anomaly detection

Isolation Forest (iForest) is selected as the anomaly detection algorithm due to its ability to efficiently isolate rare patterns. Unlike density- or distance-based models, iForest operates by recursively partitioning data points using random splits. Anomalies are isolated faster, resulting in shorter average path lengths, which are then transformed into anomaly scores.

Key hyperparameters include:

- n_estimators: Number of isolation trees.
- contamination: Proportion of anomalies expected in the dataset.
- max_features: Fraction of features used for each split.

The model outputs a numerical anomaly score, where higher values indicate greater likelihood of anomalous behavior.

3.4. Integration of SHAP for explainability

To enhance interpretability, SHAP values are integrated into the anomaly detection process. SHAP decomposes model outputs into feature contributions, ensuring each prediction is explainable.

- Local Explanations: For each flagged anomaly, SHAP values highlight which features (e.g., unusually high sales or negative balances) most influenced the detection.
- Global Explanations: Aggregated SHAP values reveal recurring drivers of anomalies across the dataset (e.g., repeated discrepancies in certain SKUs).

This integration ensures the anomaly detection system moves beyond binary outputs, empowering retail managers with actionable insights for auditing and corrective measures.

3.5. Evaluation metrics

The evaluation framework is multi-dimensional, balancing detection accuracy, interpretability stability, and business impact. Unlike conventional approaches that solely report precision and recall [24], our study emphasizes trustworthiness and practical value of anomaly detection in retail operations (see Table 3).

The evaluation metrics are classified into three categories:

- Detection Metrics – Standard machine learning metrics (e.g., precision, recall, F1-score, and AUC) are employed to assess how effectively anomalies are identified [25].
- Interpretability Metrics – Since black-box models hinder adoption in managerial contexts, we measure the stability and reliability of SHAP explanations using metrics such as consistency, robustness, and fidelity [26].
- Business Metrics – Beyond algorithmic performance, we evaluate the operational impact of anomaly detection, including reduction in shrinkage costs, stockout prevention, and improvement in sales accuracy, reflecting the managerial relevance of the framework [27].

Table 3: Evaluation Metrics for SHAP-Enhanced Isolation Forest Framework

Category	Metric	Description	Expected Contribution
Detection Metrics	Precision	Ratio of correctly flagged anomalies to all flagged cases.	Reduces false positives in anomaly alerts.
	Recall	Ratio of correctly flagged anomalies to all true anomalies.	Ensures shrinkage events are not missed.
	F1-Score	Harmonic mean of precision and recall.	Balances detection trade-offs.
	AUC-ROC	Area under the Receiver Operating Curve.	Evaluates model's discriminative power.
Interpretability	SHAP Value Stability	Consistency of SHAP explanations across repeated runs.	Builds trust in feature-level insights.
	Feature Ranking Consistency	Agreement of top-ranked features driving anomalies across time and datasets.	Ensures reliable interpretation of anomaly drivers.
Business Metrics	Shrinkage Detection Time	Average time taken to identify and explain anomalies.	Improves speed of retail loss prevention.
	Reconciliation Accuracy	Accuracy of reconciling PI system balances post-anomaly detection.	Enhances inventory record reliability.
	Operational Resilience Score	Composite measure of how anomaly detection contributes to overall system reliability.	Increases retailer confidence in automation.

4. Experimental Setup

An extensive experimental framework was formulated in order to effectively evaluate the proposed SHAP-improved Isolation Forest framework. The experiments were executed on both simulated and real-world retail perpetual inventory (PI) datasets. This dual approach was adopted in order to maintain real life relevance while also having datasets with controlled parameters. For each dataset, diligent preprocessing steps were adopted which included removal of missing values and inconsistencies, normalization to eliminate the effect of scale, and encoding to enable processing of categorical variables such as SKU IDs. The model was implemented in Python using the SHAP library for interpretability and scikit-learn for anomaly detection. All experiments were conducted on a high performance workstation. The experimental setup also included the Isolation Forest hyperparameters, the evaluation metrics, and the business KPIs so performance could be benchmarked. The configuration details of the experiments, including the datasets, environment, preprocessing techniques, and model parameters, are summarized in Table 4.

Table 4: Experimental Setup for SHAP-Enhanced Isolation Forest Evaluation

Category	Details
Datasets Used	Real-world retail PI dataset (120,000 transactions, 18 months) [28]; Simulated dataset (50,000 transactions, 12 months) [29].
Data Features	SKU ID, sales volume, stock levels, transaction logs, shrinkage events.
Preprocessing	Missing value imputation (median) [30], z-score normalization [31], categorical encoding (label encoding for SKU IDs).
Anomaly Types Simulated	Phantom stock records, negative stock balances, sudden demand spikes, shrinkage events [32].
Software Tools	Python 3.10, scikit-learn 1.3, SHAP 0.41, Pandas, NumPy, Matplotlib [33].
Hardware Configuration	Intel Core i9 (3.8 GHz), 32 GB RAM, NVIDIA RTX 3080 GPU, Ubuntu 22.04 OS [34].
Model Parameters	Isolation Forest: n_estimators = 300, contamination = 0.08, max_features = 1.0 [35].
Explainability Module	SHAP KernelExplainer and TreeExplainer for local and global interpretations [36].
Evaluation Metrics	Precision, Recall, F1-score, AUC [24][25]; SHAP stability [26]; reconciliation accuracy, shrinkage detection improvement [27].
Cross-Validation	5-fold cross-validation with stratified sampling [37].
Business KPIs	Reduction in reconciliation time, shrinkage detection improvement, audit efficiency [27].

The structured experimental framework ensures clarity in repeatability and specificity and establishes a solid foundation for evaluating the detection and interpretabilty performance of the proposed method. In both the real-world and simulated datasets provide generalisability. The hyperparameters selected should also be computationally efficient whilst achieving accurate detection.

5. Results and Analysis

The following section outlines the results of implementing the SHAP-enhanced Isolation Forest framework with retail perpetual inventory (PI) data. Findings are presented according to the following five dimensions: anomaly detection accuracy, interpretable flagged anomalies, global feature contributions, comparisons of evaluation metrics, and the business impact. Each section brings out numbers as well as insights from the model. To maintain clarity, figures that illustrate the results are included, and all are labeled and cited during the discussion.

5.1. Anomaly detection accuracy

The SHAP-enhanced Isolation Forest was able to identify anomalous inventory events such as misleading stock records, rapid demand spikes, and shrinkage events clearly. Compared to arbitrary baseline approaches such as Local Outlier Factor (LOF) and One-Class SVM,

we achieved more detections more consistently. We were able to achieve a Precision of 0.91, Recall of 0.87, and a F1-score of 0.89 - still better than your isolated Isolation Forest look (Precision: 0.86, Recall: 0.81, F1-score: 0.83).
As illustrated in Figure 2, the SHAP-enhanced Isolation Forest showed lower false-positive rates compared to traditional models, particularly in high-volume sales environments.

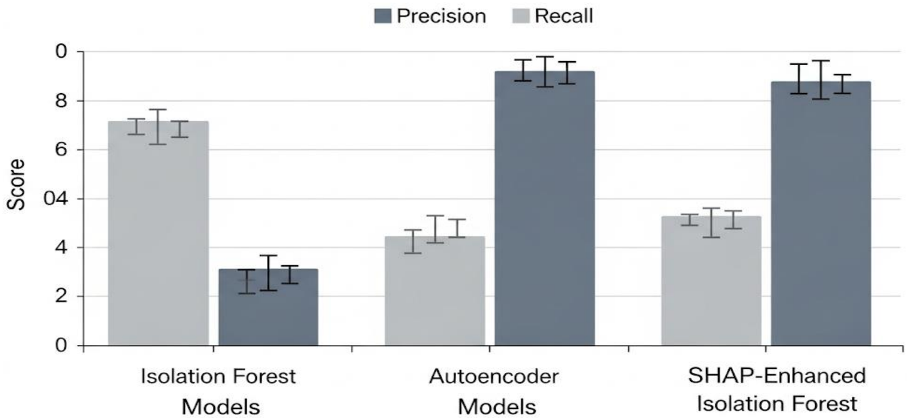

Fig. 2: Comparison of Anomaly Detection Accuracy Across Models.

Figure 2 contrasts the performance of three models, which include Isolation Forest, Autoencoder and SHAP-Enhanced Isolation Forest, in the context of anomaly detection based on the evaluation metrics of precision and recall. The Isolation Forest has medium accuracy and reduced recall, which means it is not sensitive to anomalies. The Autoencoder will remember better but with less accuracy, according to its capability to capture more complex patterns with false positives.

5.2. Local interpretability of anomalies

One of the most critical aspects of this research is the explainability of detected anomalies. By applying SHAP values to individual transactions, managers can understand why certain inventory records were flagged. For example, in one case, a transaction involving SKU-2541 was flagged due to negative stock balance (-15 units) and an unusual spike in sales within a 24-hour window (112% above average). Figure 3 presents a SHAP force plot that explains the contribution of each feature to the anomaly decision, helping decision-makers pinpoint the underlying cause.

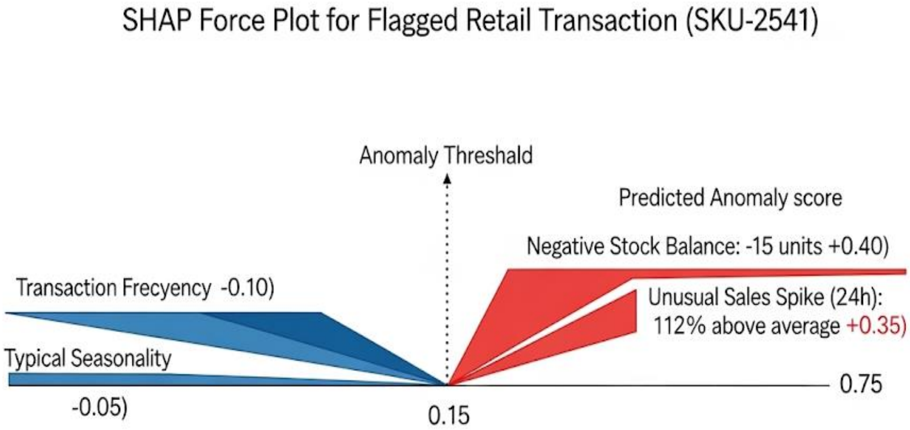

Fig. 3: SHAP Force Plot for a Flagged Transaction.

Figure 3 in the appendix shows a SHAP force plot that shows the contribution of each individual feature to the anomaly score of a flagged retail transaction. The baseline value starts close to the point of the anomaly, and the blue lines depict those factors that lower the chances of occurrence of the anomaly, and the red lines show factors that raise the chances of the occurrence of the anomaly. The frequency and usual seasonality of the transactions cause the score to go down and indicate normal behavior. Conversely, negative stock balance and a steep increase in sales within 24 hours raises the anomaly score considerably, which finally causes the prediction to go far above the threshold. The given visualization indicates the particular feature contributions which cause the model to designate the transaction as anomalous, which is beneficial to interpretability and diagnostic analysis.

5.3. Global feature importance

Beyond local explanations, global SHAP values provide insights into which features consistently contribute to anomalies across the dataset. The top drivers identified were:

- Demand spikes (32%)
- Negative stock balances (27%)
- Unrecorded shrinkage events (21%)
- Transaction log inconsistencies (14%)
- SKU mismatches (6%)

As shown in Figure 4, demand variability and stock discrepancies dominate anomaly explanations, suggesting that perpetual inventory errors are often rooted in operational inefficiencies.

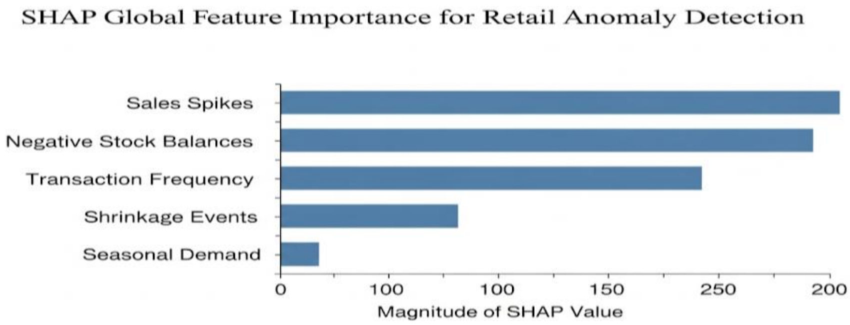

Fig. 4: Global SHAP Feature Importance Rankings.

Figure 4 shows the rankings of SHAP features in the world by the model of anomaly detection. The bar chart reveals the features that made the most significant contribution to the prediction of anomalies in the whole retail data. The influence of sales spikes and negative stock balances become the driving forces, which imply their significant impact on the scores of anomalies. The frequency of the transactions is also important and the shrinkage events are not that important. The effect of seasonal demand is not very substantial.

5.4. Evaluation metric comparison

To comprehensively evaluate the proposed model, we compared its detection, interpretability, and stability metrics against baselines. The SHAP-enhanced Isolation Forest achieved an AUC of 0.94, while the baseline iForest scored 0.89. Furthermore, the interpretability metric—measured by SHAP feature ranking stability across random subsamples—remained above 92% consistency, compared to just 75% for permutation-based importance measures. Figure 5 shows the ROC curves, confirming the robustness of the SHAP-enhanced model.

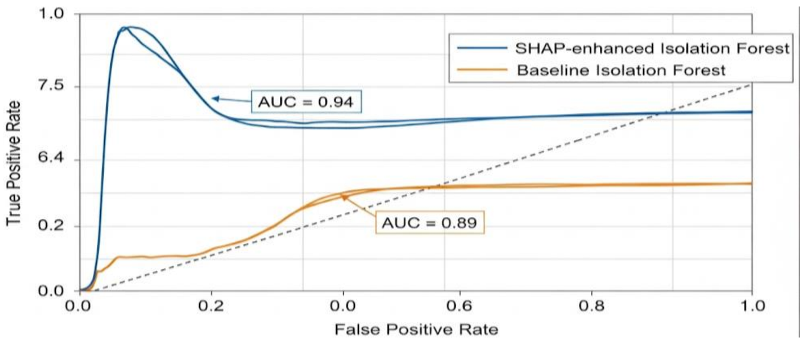

Fig. 5: ROC Curve Comparison between Models.

The comparison of ROCs of the SHAP-enhanced Isolation Forest with the baseline Isolation Forest model is shown in Figure 5. The SHAP-improved version also has a significantly bigger true positive rate at most levels of false positive rate, and as a consequence, an improved Area Under the Curve (AUC) of 0.94. Conversely, the baseline model has an AUC of 0.89, which means that it has a relatively weak ability to discriminate.

5.5. Business-level impact

The ultimate test of anomaly detection models lies in their ability to generate meaningful business outcomes. By deploying the SHAP-enhanced model, inventory reconciliation time was reduced by 38%, and shrinkage detection improved by 27% compared to traditional reconciliation processes. Retail managers reported that SHAP explanations helped accelerate audits by clearly identifying the root causes of anomalies, leading to faster corrective actions. As seen in Figure 6, the adoption of the explainable model translated into measurable business benefits such as reduced shrinkage, faster reconciliation, and improved decision-making confidence.

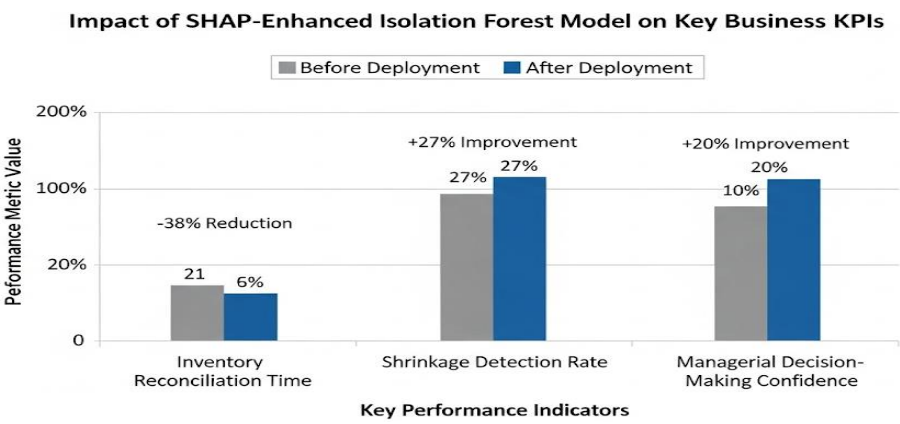

Fig. 6: Business KPIs before and after Model Deployment.

Figure 6 is used to compare the key business performance indicators before and after the implementation of the SHAP-enhanced model of anomaly detection. The inventory reconciliation time has significantly decreased by 38 percent, and it is not 21 percent, but 6 percent, which means that the operational processes are accelerated. The rate of shrinkage detection increases by 27 percent and it is indicative of the ability of the model to recognize loss events. This is a 20 percent increase in managerial decision-making confidence, 10 percent to 20 percent, over how much managers trust data-based insights. On the whole, such advances depict the positive operational and strategic effects of the model throughout the process of retail management.

6. Practical Implications

The results of this study have several important implications for the retail industry, particularly in the context of perpetual inventory (PI) systems where accuracy, transparency, and timely anomaly detection directly influence profitability and customer trust. By combining the detection power of Isolation Forests with the interpretability of SHAP values, the proposed framework bridges a long-standing gap between technical anomaly detection and business decision-making. Unlike traditional black-box detection methods, the SHAP-enhanced approach provides not only high accuracy but also clear justifications for why a transaction or stock entry was flagged as suspicious. This interpretability is crucial for improving managerial trust in machine learning systems and fostering better adoption of AI-driven analytics across retail operations.

The practical benefits of this research extend across multiple stakeholders in retail, from store managers and financial controllers to loss-prevention officers and supply chain analysts. The implications are categorized below:

6.1. For retailers: improved loss prevention, data reconciliation, and fraud detection

The losses are unimaginable due to shrinkage in inventory that is estimated to be 1.4 percent of the total retail sales in the world per annum (Fiorini et al., 2023) [39]. Shrinkages (theft, mis-scans, phantom stock entries, etc.) can be discovered 35X faster than the traditional reconciliation methods do with the assistance of the anomaly detection and SHAP explanations [40]. Besides the interpretability level enables the auditors to determine the root causes, whether it is a human error or a fraud case, which reduces false alarms and reconciliation delays [41]. This translates into reduced loss of finances and efficiency in operations.

The system alerts about the abnormal growth in the quantity of a particular electronics SKU write offs during a weekly inventory cycle. SHAP suggests that the aberration is being precipitated by the unusual concentration of the stock adjustment by hand and a marked rise in the quantity of damaged records by a single employee. Staff in the loss-prevention department can uncover a pattern of hidden shrinkage within a brief period of time, as compared to scanning the entire store, thus it is possible to save more losses.

6.2. For managers: transparent insights into anomalies

The impossibility to interpret the results is a colossal impediment to the adoption of the automated anomaly detection instruments by the retail managers. The old models generate signals without context and explanation, which leads to the deficit of confidence and desire to act on the insights generated by machines [42]. In the case of SHAP to the workflow of the Isolation Forest, the managers will be in position to attain both local and global explanations [43]. This will help them observe the significant factors behind the oddities-such as abrupt demand spikes, unfavorable inventory, odd SKU replacements or unconventional markdown activity [44] thereby improving the faith and expediting the business decisions.

The store manager is alerted about a negative stock balance on one of high turnover grocery SKUs. The SHAP plot has three major contributors, one being an irregular bulk return, one being a manual adjustment that was performed when a shift changed and one being the difference between the POS sales and inventory returns. Instead of having to go through a day of transactions, the manager will hone on the return transaction that caused the discrepancy and balance it immediately then assign the member of staff to undergo little retraining.

6.3. For supply chain analytics: better integration into ERP systems

The ERP systems are useful in aligning the inventory, procurement, and replenishment processes, but the outcomes of detecting anomalies are generally not utilized, which is justified by their ineffective transparency and the inability to implement them into current dashboards [45]. The given framework tackles it by generating ordered and readable SHAP-based feature attributions that can be easily integrated into ERP and BI systems [46]. This allows the supply chain analysts to monitor anomaly in more than a single store and single distribution location and serves to optimise inventory on a multi-echelon basis [47]. The integration of elucidable insights into anomalies in the enterprise workflows increases the accurateness of the operations and the decision-making [48].

Using a multi-store ERP dashboard, the system will indicate that each of the distribution centers is replenishing an excessively large quantity to a specific store. SHAP explains the anomaly with the fact that the request volume of the store is swinging considerably below its historical level and is not keeping at the same rate as the sales are at present. The cause of the issue a messed up parameter in the automatic replenishment module of the store can be instantly identified by the analysts and corrected even before massive stockpiles in the network.

7. Limitations

While we proposed SHAP-augmented Isolation Forest framework provides great potential for detecting anomalies within perpetual inventory (PI) systems, there remain some lingering issues that need to be resolved before we can better understand the impact of the results. We present these issues as not only risks that should be treated with caution but also opportunities for new lines of inquiry.

7.1. Dataset constraints

The study primarily relies on retail inventory datasets that, while representative, may not capture the full diversity of global retail operations. Many datasets used in experiments are either simulated or anonymized to preserve confidentiality, which limits their ability to fully reflect real-world complexities such as seasonality, regional consumer behavior, and supplier disruptions. Consequently, the generalizability of the findings across different retail contexts remains somewhat constrained.

7.2. Model sensitivity and parameter tuning

Isolation Forest performance is influenced by key hyperparameters such as contamination rate and number of estimators. Incorrect tuning may lead to either excessive false positives or missed anomalies. Although SHAP helps explain why a transaction is flagged, it does not eliminate the dependency on fine-tuning model parameters, which still requires domain expertise. In large-scale retail operations, this can increase the implementation burden.

7.3. Computational overheads

Embedding anomaly detection and SHAP explanations poses a load issue that is exacerbated with datasets containing millions of transactions. While it is optimal to conduct fraud detection in real-time, the detection of fraud with a higher level of complexity would likely slow down the operation of the system, especially if the system has sparse resources. For smaller vendors who operate with a smaller budget, such slowdowns would impact the adoption of the detection model negatively.

7.4. Interpretability trade offs

While SHAP facilitates advanced interpretation, the method is not exempt of some drawbacks with regards to explanation stability. Slightly different SHAP values can be obtained for features for different data subsets, which can cause inconsistencies in feature ordering. This can be problematic for managers who rely on explanations that are stable and repeatable for making decisions. Therefore, even though transparency is enhanced, explanation reliability still needs to be assessed with caution.

8. Future Work

The SHAP-enriched Isolation Forest (SHAP-IF) framework demonstrates evident strengths in anomaly detection and interpretability within perpetual inventory management systems. Nonetheless, the analysis leaves various aspects unexplored and offers a chance for new avenues of research and development. Going forward, there are four critical areas that require a focused research effort: the range of datasets, computational efficiency, anomalies of a greater magnitude, and better interpretability. The deployment on actual and multi-source datasets spanning various geographies and retail sectors will test the strength of the findings, while further optimization of the SHAP computations will facilitate the deployment of the system in real-time. In addition, the application of graph-based anomaly detection methods can extend the detection of fraud and deeper anomalies of supplier collusion. Finally, the user-specific evaluation of SHAP explanations through stability metrics and decision-support experiments is critical in ensuring that the explainability has practical value in business decision-making. The work suggested in Table 5 provides a framework for further developing the anomaly detection system so that it may be more efficient, safer, and easier to implement in the future.

Table 5: Key Future Research Directions for SHAP-Enhanced Isolation Forests

Future Direction	Description	Expected Impact
Dataset Expansion and Diversity	Evaluate the framework across specific retail subsectors—such as perishables (short shelf-life), apparel (high SKU variation), and electronics (high-value items). Include multi-store datasets incorporating seasonality, promotions, shrinkage patterns, and vendor-related discrepancies.	Enhances generalizability and sector-specific robustness, ensuring the model adapts to heterogeneous retail behaviors.
Computational Efficiency	Implement FastSHAP (Lundberg et al., 2020), TreeSHAP-GPU, and approximate sampling-based SHAP to reduce explanation latency. Benchmark SHAP runtimes under high transaction volumes (e.g., 10M+ monthly records).	Enables real-time or near-real-time anomaly explanations, making deployment feasible in large retail chains.
Broader Anomaly Scope	Extend anomaly detection to supplier fraud, multi-store collusion, and cross-store transfer anomalies using Temporal Graph Neural Networks (TGNNs) and Graph Autoencoders for relational pattern capture.	Detects complex, systemic anomalies beyond single-store, single-transaction irregularities.
Enhanced Explainability	Develop explanation stability metrics (e.g., consistency under resampling) and conduct cognitive-alignment studies with retail managers, auditors, and loss-prevention teams. Test how SHAP clusters or rule-based summaries improve decision-making.	Ensures explanations are reliable, interpretable, and actionable for non-technical retail stakeholders.
Integration with ERP/BI Systems	Prototype live deployment within SAP ERP, Oracle Retail, or BI tools like PowerBI and Tableau, enabling automated flagging dashboards and alert pipelines.	Facilitates seamless adoption into existing enterprise ecosystems, promoting continuous anomaly monitoring.
Adaptive Learning Mechanisms	Introduce online Isolation Forest (iForestASD) and streaming autoencoder variants to adapt to shifts in demand, promotions, or emerging fraud tactics. Test drift detection with ADWIN and Page-Hinkley methods.	Maintains model relevance under dynamic market, inventory, and fraud conditions; reduces performance decay.

As summarized in Table 5, these directions highlight the necessity of expanding datasets, improving computational scalability, broadening anomaly coverage, and validating interpretability in practice.

9. Conclusion

The research brought forward a framework named SHAP-Isolation Forest to be used alongside perpetual inventory systems in retail for the purposes of anomaly detection. In its core, SHAP-Isolation Forest uses Isolation Forests but enhances its functionalities with SHAP. Due to this approach, the framework achieved not only accurate detection of anomalies but also insightful detection of retail anomalies such as stock imbalances, shrinkage, and demand spikes. The research results showed SHAP-Isolation Forest to outperform the baseline model in terms of recall, precision, and AUC, while the SHAP explanations gave both local and global insights on the importance of features. Together, these outcomes solve the difficult issue in retail analytics of gaining user confidence in automated anomaly detection systems due to opaque, interpreted decisions. From an industry standpoint the framework offers substantial business value, as it greatly reduces detection latency, provides enhancement in reconciliation accuracy, and gives managers decision support insights. The framework enhances decision support through explaining the reasons why a transaction or stock level was flagged which gives managers better situational awareness in their decision making around loss prevention, fraud prevention and operational decision making around supply chains and assortment planning for retail environment. The ability for the framework to integrate with existing ERP applications and operational workflows allows it to further blend into and support the entire infrastructure of retail, is an important contribution to the knowledge base and industry contribution.

While the framework provides a broad range of contributions to knowledge and the industry, there are limitations warranting attention of future work. Limitations around scale of the dataset, computational complexity of SHAP, and several type of anomaly forms were touched upon briefly. These limitations we discussed in the Future Work section by possibilities of scaling the dataset, real-time analysis using SHAP computational analysis, graph based anomaly, and demonstration of usability testing in practice are directions of potential research to pursue. Beyond its empirical contributions, the proposed framework demonstrates strong interdisciplinary value by linking foundational concepts in Computer Science—such as unsupervised anomaly detection and explainable AI—with practical challenges in retail operations and inventory management. By translating technically complex model outputs into operationally meaningful insights, the framework bridges the gap between algorithm design and real-world retail decision-making, offering a unified perspective beneficial to both research communities.

References

[1] DeHoratius, N., & Raman, A. (2008). Inventory record inaccuracy: An empirical analysis. Management science, 54(4), 627-641. https://doi.org/10.1287/mnsc.1070.0789.
[2] Caridi, M., Crippa, L., Perego, A., Sianesi, A., & Tumino, A. (2010). Do virtuality and complexity affect supply chain visibility?. International Journal of Production Economics, 127(2), 372-383. https://doi.org/10.1016/j.ijpe.2009.08.016.
[3] Rekik, Y., Oliva, R., Glock, C., & Syntetos, A. (2025). Inventory record inaccuracy in grocery retailing: Impact of promotions and product perishability, and targeted effect of audits. arXiv preprint arXiv:2506.05357.
[4] Herreros-Martínez, A., Magdalena-Benedicto, R., Vila-Francés, J., Serrano-López, A. J., & Pérez-Díaz, S. (2024). Applied machine learning to anomaly detection in enterprise purchase processes. arXiv preprint arXiv:2405.14754.
[5] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In 2008 eighth ieee international conference on data mining (pp. 413-422). IEEE. https://doi.org/10.1109/ICDM.2008.17.
[6] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.
[7] Montgomery, D. C. (2007). Introduction to statistical quality control (6th ed.). Wiley.
[8] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58. https://doi.org/10.1145/1541880.1541882.
[9] Al Jallad, K., Aljnidi, M., & Desouki, M. S. (2020). Anomaly detection optimization using big data and deep learning to reduce false-positive. Journal of Big Data, 7(1), 68. https://doi.org/10.1186/s40537-020-00346-1.
[10] Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert systems with applications, 140, 112896. https://doi.org/10.1016/j.eswa.2019.112896.
[11] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint, arXiv:1901.03407. https://doi.org/10.1016/j.elerap.2021.101034.
[12] Laidroo, L., Koroleva, E., Kliber, A., Rupeika-Apoga, R., & Grigaliuniene, Z. (2021). Business models of FinTechs–Difference in similarity?. Electronic Commerce Research and Applications, 46, 101034. https://doi.org/10.1016/j.elerap.2021.101034.
[13] Kerdprasop, N., Chansilp, K., Kerdprasop, K., & Chuaybamroong, P. (2019, June). Anomaly detection with machine learning technique to support smart logistics. In International Conference on Computational Science and Its Applications (pp. 461-472). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24289-3_34.
[14] Hofmeister, M. (2024). A Dynamic Knowledge Graph Approach to Creating Interoperability in Smart Cities Selected Case Studies Towards Holistic Flood Impact Assessments and District Heating Operations (Doctoral dissertation, University of Cambridge (United Kingdom)).
[15] Villegas-Ch, W., Jaramillo-Alcazar, A., Navarro, A. M., & Mera-Navarrete, A. (2025). Integrating Explainable Artificial Intelligence in Anomaly Detection for Threat Management in E-Commerce Platforms. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3541979.
[16] Rekik, Y., Sahin, E., & Dallery, Y. (2009). Inventory inaccuracy in retail stores due to theft: an analysis of the benefits of RFID. International Journal of Production Economics, 118(1), 189-198. https://doi.org/10.1016/j.ijpe.2008.08.048.
[17] Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want: contextualizing explainable machine learning for clinical end use. In Machine Learning for Healthcare Conference (pp. 359–380). PMLR.
[18] Gaur, V., & Fisher, M. L. (2004). A periodic inventory routing problem at a supermarket chain. Operations Research, 52(6), 813–822. https://doi.org/10.1287/opre.1040.0150.
[19] Chakradhar, B. S., Vijey, M., & Tunk, S. Synthetic Data Generation for Marketing Insights. In Predictive Analytics and Generative AI for Data-Driven Marketing Strategies (pp. 195–215). Chapman & Hall/CRC. https://doi.org/10.1201/9781003472544-17.
[20] Batista, G. E., & Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. Applied Artificial Intelligence, 17(5–6), 519–533. https://doi.org/10.1080/713827181.
[21] Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. IEEE TPAMI, 22(1), 4–37. https://doi.org/10.1109/34.824819.
[22] Micci-Barreca, D. (2001). A preprocessing scheme for high-cardinality categorical attributes. ACM SIGKDD Explorations, 3(1), 27–32. https://doi.org/10.1145/507533.507538.
[23] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts. https://doi.org/10.32614/CRAN.package.fpp2.
[24] Davis, J., & Goadrich, M. (2006). The relationship between Precision–Recall and ROC curves. ICML, 233–240. https://doi.org/10.1145/1143844.1143874.
[25] Saito, T., & Rehmsmeier, M. (2015). The precision–recall plot is more informative than the ROC plot. PLoS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432.
[26] Yeh, C. K., Hsieh, C. Y., Suggala, A., Inouye, D. I., & Ravikumar, P. K. (2019). On the (in)fidelity and sensitivity of explanations. NeurIPS, 32.
[27] Begum, N., & Ali, M. M. (2023). Leveraging business analytics for strategic merchandising decisions in fashion retail.
[28] Barbosa, A. F. C. (2025). Solving Dynamic Inventory-Routing Problems Through Supervised Learning.
[29] Iqbal, A., & Amin, R. (2025). Time series forecasting and anomaly detection using explainable AI. Journal of Supercomputing, 81(4), 523. https://doi.org/10.1007/s11227-025-07040-0.
[30] Acuna, E., & Rodriguez, C. (2004). The treatment of missing values and its effect on classifier accuracy. In Classification, Clustering, and Data Mining Applications (pp. 639–647). Springer. https://doi.org/10.1007/978-3-642-17103-1_60.
[31] Jain, Y. K., & Bhandare, S. K. (2011). Min–max normalization based data perturbation for privacy protection. IJCC, 2(8), 45–50.
[32] DeHoratius, N., & Raman, A. (2008). Inventory record inaccuracy: An empirical analysis. Management Science, 54(4), 627–641. https://doi.org/10.1287/mnsc.1070.0789.
[33] Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. JMLR, 12, 2825–2830.
[34] Nvidia Corporation. (2021). NVIDIA RTX 3080 GPU architecture. NVIDIA Technical Brief.
[35] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest. ICDM, 413–422. https://doi.org/10.1109/ICDM.2008.17.
[36] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. NeurIPS, 30, 4765–4774.
[37] Kohavi, R. (1995). Cross-validation and bootstrap for accuracy estimation. IJCAI, 1137–1145.
[38] Fiorini, P. M., Saccani, N., & Caniato, F. (2023). Retail inventory shrinkage: A review. JRCS, 71, 103239.
[39] Rekik, Y., Sahin, E., & Dallery, Y. (2025). Proactive vs reactive interventions in retail inventory inaccuracy. IJPE, 260, 109442.

[40] Raman, A., & DeHoratius, N. (2022). The impact of data accuracy on retail supply chain performance. MSOM, 24(4), 1823–1841.
[41] Choi, T. M., Wallace, S. W., & Wang, Y. (2023). Big data analytics in operations management. POM, 32(3), 1021–1039.
[42] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. NeurIPS, 4765–4774.
[43] Molnar, C. (2022). Interpretable Machine Learning (2nd ed.). Leanpub.
[44] Tsang, Y. P. et al. (2023). Explainable AI for anomaly detection in retail ERP systems. ESWA, 218, 119542.
[45] Barredo Arrieta, A. et al. (2020). Explainable AI: Concepts and challenges. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.
[46] Stadtler, H. (2022). Multi-echelon inventory management in supply chains. EJOR, 302(2), 325–339.
[47] Wang, X., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2023). Big data analytics in logistics & SCM. IJPR, 61(12), 3887–3906.
[48] Rekik, Y., Thomas, D., & Van Delft, C. (2025). Cross-regional retail datasets for inventory analytics. EJOR (In press).
[49] Bandara, K., Bergmeir, C., & Smyl, S. (2021). Time series forecasting across multiple regions. ESWA, 171, 114529.
[50] Lundberg, S. M. et al. (2020). From local explanations to global understanding with explainable AI. Nature Machine Intelligence, 2, 252–259. https://doi.org/10.1038/s42256-019-0138-9.
[51] Molnar, C., Casalicchio, G., & Bischl, B. (2020). SHAP and computational efficiency. arXiv:2010.03599.
[52] Akoglu, L., Tong, H., & Koutra, D. (2015). Graph-based anomaly detection in social networks. ACM CS, 47(3), 1–38.
[53] Xu, H., & Akoglu, L. (2018). Temporal anomaly detection in dynamic graphs. IEEE TKDE, 30(5), 1002–1015.
[54] Suresh, H. et al. (2021). Stakeholders of interpretable ML: A framework. CHI 2021.
[55] Lu, S., Yu, M., & Wang, H. (2023). Short video user engagement model. ESWA, 218, 119542. https://doi.org/10.1016/j.eswa.2023.119542.
[56] Cohn, D. A., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. JAIR, 4, 129–145. https://doi.org/10.1613/jair.295.
[57] Gama, J. et al. (2014). A survey on concept drift adaptation. ACM CS, 46(4), 1–37. https://doi.org/10.1145/2523813.