DECISION SCIENCES INSTITUTE

Detecting and Predicting Phantom Inventory Using Random Forest and LSTM Models

Hung-Chen Hsu, Kratika Jain, Da Fang Lin, Yiran Liu, Kaushalya B. Naidu, Namitha Ramakrishnan, Rohan Ajay, Matthew A. Lanham
Purdue University Daniels School of Business
hsu367@purdue.edu, jain784@purdue.edu, lin1945@purdue.edu, liu387@purdue.edu, knaidu@purdue.edu, ramakr40@purdue.edu, rajay@purdue.edu, lanhamm@purdue.edu

ABSTRACT

Phantom inventory refers to discrepancies where inventory exists in records but not in reality, which can arise due to a variety of factors, such as misplacement, system glitches, theft, shrinkage, and inaccurate data entry. This paper will focus on exploring real business inventory data and modeling on detection and prediction. We aim to develop powerful machine learning models, random forest, time series, and clustering, by analyzing transaction data, stock level movement patterns, and sale anomalies to uncover hidden discrepancies and eventually manage to predict stockouts. Detecting and predicting phantom inventory case is crucial to optimize business performance and enhance profitability.

KEYWORDS: Phantom inventory, Detection, Prediction Random Forest, LSTM

INTRODUCTION

In the dynamic landscape of retail business, phantom inventory, referred to as ghost inventory or stockout, has been a problem for decades. Nowadays technological advances achieve customer convenience by offering online shopping and hence change shopping behavior. These changes result in new challenges for inventory management and supply chain operation. It is crucial to maintain accurate inventory records for both operational efficiency and customer satisfaction and is the driving motivation for our research.

In this paper, phantom inventory is defined as the merchandise that appears to be available in a company's records, but is not physically present on the shelves, indicating a discrepancy between the recorded inventory and actual stock. This issue can arise from various factors, including theft, misplacement, data entry errors, and system inaccuracies. When phantom inventory or stockout occurs, retailers will find that inventory records inaccurately reflect stock level, which lead to several challenges.

Loss of sales: Riaz (2021), points out that retailers worldwide lost more than $1.8 trillion annually due to inventory distortion issue. As one may imagine, offline retail accounts were the majority of this number. In the United States, a 2018 survey by the National Retail Federation (NRF) which is cited in by McCue (2019) reported a loss of $46.8 billion caused by inventory shrinkage.

Customer dissatisfaction: If a customer finds a product listed as available on a retailer's website, but encounters an empty shelf within store, this can lead to customer frustration and loyalty loss. According to Crowe (2023), after software provider, Retail Insight, surveyed more than 1000 U.S. shoppers, it was found that phantom inventory is negatively impacting sales and long-term loyalty.

Increased operational costs: Phantom inventory forces employees to take additional time searching for nonexistent merchandise and rectifying inventory records, largely and deeply diminishing business productivity and efficiency. As Paul Boyle, CEO of Retail Insight, pointed out, “Facing growing pressure on already razor-thin margins, phantom inventory risks becoming a profit-draining issue across every part of the retailer’s store operations, from wasted labor to out of stocks and increased shrink.” This comment was cited in the article by Mayer (2023) as well.

As we see from the past decade, phantom inventory continues to remain a persistent challenge in the retail industry. This issue leads to revenue loss, dissatisfied customers, and inefficiencies in operation and inventory management. Traditional inventory audits and manual checks fail to provide timely and accurate insights and forecasting into stock discrepancies, necessitating more advanced analytical approaches. With the rise of data-driven solutions, machine learning techniques such as time series forecasting, random forest, clustering, and other methods of anomaly detection offer potentials for addressing phantom inventory. By leveraging these techniques, it is now possible to accurately detect out-of-stock, effectively reduce revenue loss, and hence improve overall inventory accuracy.

In this study, we define and detect phantom inventory using key inventory features such as the daily balance on hand (BOH), which is recorded at the end of each day, and the ground truth label captured hourly by cameras. The primary objectives are to detect discrepancies between the BOH and the label data, and to predict the likelihood of out-of-stock in the future timeframe. By analyzing the discrepancies between inventory records and actual stock levels, we attempt to uncover patterns and trends that may signal phantom inventory. Time series models are explored to capture fluctuations in stock levels over time, allowing us to forecast potential stockouts before they occur. Additionally, we utilize random forest and clustering techniques for classification of phantom inventory merchandise, helping retailers understand which products or even locations are more prone to inaccuracies.

The integration of multiple modeling approaches enables a comprehensive analysis of phantom inventory. Ultimately, we aim to not only contribute to reducing loss of sales, improving customer experience and achieving operational excellence, but also provides a scalable solution that can also be adapted to various and broader industrial environments.

LITERATURE REVIEW

There are various definitions of academic phantom inventory in the academic literature. In retail operations, phantom inventory refers to inventory that appears in the records but is not actually available on the shelf. In other words, the system shows that an item is in stock when physically it is missing. Bassamboo et al. (2020) defines phantom inventory as “units of [a] good not available for sale” even though they are recorded in the inventory system. Farias et al. (2024) similarly describe phantom inventory events as instances where “inventory is recorded to be present on-shelf while in reality it is missing,” often due to causes like shrinkage or theft.

This concept is sometimes also called "ghost inventory" in practice, highlighting that the stock exists only in the computer records and not in the store's actual stock. All these definitions emphasize a discrepancy in which recorded on-hand quantities exceed the true physical inventory, leading the system to falsely assume items are available. Such discrepancies are a well-known form of inventory record inaccuracy in retail (DeHoratius & Raman, 2008) and have been reported to be widespread. Studies have found that 50–65% of inventory records can contain errors, with phantom inventory (overstated stock) being a common outcome (Farias et al., 2024; Bassamboo et al., 2020). This phenomenon can significantly undermine operations, as it causes stockouts that are invisible to the inventory system – the shelf is empty but the system doesn't trigger replenishment because it "thinks" products are still there (Rekik et al., 2019). In fact, undetected phantom inventory has been estimated to account for substantial lost sales (on the order of ~4% of annual revenue in the retail industry) because customers encounter empty shelves despite the system showing stock (Farias et al., 2024).

Since phantom inventory cannot be observed directly in systems until a correction is made, researchers have developed data-driven methods to detect or infer its occurrence from sales and inventory data. A key insight used in academic studies is that a phantom inventory often manifests as prolonged zero-sales periods for an item that is supposedly in stock. Tripathy (2019) explains that a "phantom inventory remains unnoticed unless an intervention (e.g., a shelf audit) occurs," and proposes a detection approach based on consecutive days of zero sales while the inventory-on-hand (IOH) remains positive. In this Bayesian model, if a product shows no sales for an extended stretch of time despite the system indicating stock on hand, it likely signals a phantom inventory situation (Tripathy, 2019). The logic is that if the item were truly on the shelf, some sales would normally occur given underlying demand. This approach models daily demand probabilistically (e.g., using a negative binomial distribution) to estimate the likelihood that zero-sales over X days could happen if the inventory were actually there. A very low likelihood (i.e., unexpectedly long sales drought while stock is recorded) points to phantom inventory with high probability.

More recent data-driven methods cast the phantom inventory detection as an anomaly detection problem. Farias et al. (2024) treat the pattern of sales across stores as a matrix and look for anomalies indicating that a particular store-SKU combination has abnormally low sales compared to expectation, given that it should have stock. In essence, they look for when the "rate of transactions has slowed or stopped" for an item, which, coupled with a known positive inventory record, suggests a phantom inventory event (Farias et al., 2024). Their approach uses cross-sectional data (i.e., multiple stores) to detect these anomalies more robustly, framing it as identifying outliers in a low-rank Poisson sales matrix (Farias et al., 2024). This and similar AI or algorithmic methods allow retailers to flag likely phantom inventory instances without waiting for manual audits. The common thread in these data-driven approaches is leveraging point-of-sale (POS) data and inventory records over time: when POS data show zero sales for a SKU that the inventory system lists as in-stock, an alert for phantom inventory can be raised (Tripathy, 2019). Such techniques have been shown to significantly improve detection, enabling timely corrective actions (like triggering a physical count or restocking) to reconcile the records (Tripathy, 2019; Farias et al., 2024).

DATA

In collaboration with a national retailer, we were granted to access to four different dataset categories. Namely, “consumables”, “dairy”, “dry_grocery”, and “frozen” from one of the store locations for the month of November 2024, structured into three CSV files for each category. Each category contains features-daily_2024-11.csv (“Features”), labels_2024-11.csv (“Labels”), and trxn_2024-11.csv (“Transaction”). The Features dataset (39 attributes) mainly captures product-level information, including various balance-on-hand (BOH) features and product sales data. We develop machine learning models to identify important features like product names, sales information, in-store location, and inventory level data (BOH). The Labels dataset (34 attributes), of which the stock level data is collected by the in-store shelf camera system on hourly basis, provides features such as out-of-shelf occurrences and their corresponding timestamps. Lastly, the Transaction dataset (17 attributes) provides transaction data that can help to identify potential demand level. These datasets offer a structured view of inventory flow, aiding in the detection and prediction of phantom inventory events.

METHODOLOGY

A structured approach is employed to detect and predict phantom inventory. We create two phases for the modeling. In Phase 1 we focus on detection of phantom inventory, while in Phase 2 we focus on prediction of product stockout day.

Phase 1: Detection of Phantom Inventory – Pre-processing and Exploratory Data Analysis

The dataset was given to us by one of the major players in U.S. grocery retailer. It consisted of three important tables that included Features, Labels and Transactions. To start EDA, we preprocessed the data by handling missing values. For feature engineering, K-means clustering was used to group the products on the basis of stock fluctuation patterns. Key variables such as SKU, BOH, transaction quantities and historical sales were selected.

Phase 1: Detection of Phantom Inventory – Model Building

After EDA, a random forest model was found to detect the phantom inventory by classifying the SKUs which have stock discrepancies. Many classification models were explored such random forest, CatBoost, XGBoost, and LightGBM. Random forest model was selected due to its robustness and interpretability. The cross-validation design used involved an 80-20% train-test split.

Phase 2: Prediction of Product Stockout Day – Pre=processing for LSTM-based Model

In Phase 2, we would adopt an LSTM-based forecasting approach for our predictive model. Since the dataset comprises daily inventory levels (DailyBOH) for each SKU, indexed by date, the preprocessing steps include the following two main steps: outlier capping and log transformation. We used outlier capping where the DailyBOH values were capped at the 99th percentile to limit the influence of extreme outliers, ensuring model stability. Log transformation was employed to cap the DailyBOH using a log1p transformation (np.log1p) to address skewness and zero-inflation. This created a more normalized distribution suitable for LSTM modeling.

2.2 Sliding Window Preparation

A sliding window of 7 days was used to create input sequences (X) and corresponding targets (y). For univariate forecasting, X contains only log-transformed DailyBOH; for multivariate forecasting, it includes both DailyBOH and transaction quantities. The data is scaled using either

MinMaxScaler (range 0-1) or RobustScaler, with the latter as an option for robustness against outliers.

MODELS

To build models for detecting phantom inventory, we required an implementation-based definition that will be carefully expounded in the remainder of this paper. We formulated a structured definition of phantom inventory using the following logic. Phantom inventory is identified by the coexistence of two conditions:

1. **Recorded On-Hand Stock > 0:** The inventory management system indicates that a given product is in stock (a positive quantity is recorded for that SKU).
2. **No Sales Over a Prolonged Period:** Despite the recorded stock, the point-of-sale data show zero sales for that product over a significant time frame (e.g., several consecutive days or weeks).

When both conditions hold, it implies the item is likely not actually present or available to customers, hence qualifying as phantom inventory. Essentially, the system “thinks” the item is on the shelf, but customers are not buying any – strongly suggesting the shelf is empty (the inventory exists only in phantom form). This definition reflects the operational logic used to flag phantom inventory in data-driven systems (Tripathy, 2019). It is consistent with academic characterizations: for instance, Bassamboo et al. notes that phantom inventory arises when products are “not available for sale” even though they appear in the records (Bassamboo et al., 2020), and Tripathy’s detection rule captures exactly that scenario of zero sales with positive stock (Tripathy, 2019).

In summary, **phantom inventory = recorded stock – actual stock**, when the recorded stock is erroneously higher and items that are supposed to be “there” are not actually on the shelf (Tripathy, 2019; Farias et al., 2024). This explicit definition guided by the implementation logic is supported by the literature and helps ensure that our understanding of phantom inventory is both conceptually sound and practically measurable.

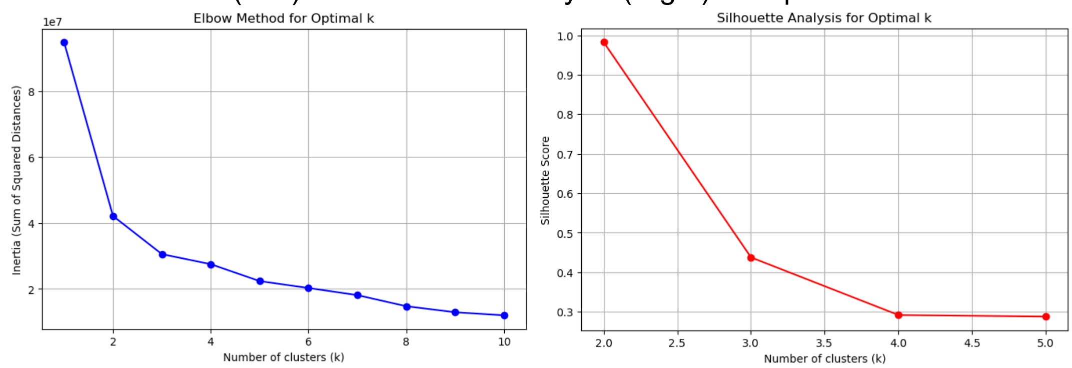
The key feature in the dataset is called “DailyBOH” (daily balance on hand), defined as “final recorded stock at the end of the day”. Another relevant feature that will be considered for the models is "CurrentBOH" (current balance on hand) defined as a mid-day snapshot or system-generated number. As the daily final records, "DailyBOH" is more reflective of actual stock levels, and it will be prioritized in the modeling. On the other hand, “CurrentBOH” still contributes some value data to the model, so we are not removing it completely but only use it when it is needed.

We create a missingness indicator for the purpose of handling missing values that exist in the dataset. The idea is that we label the missing entry as “missing” (by missingness indicator) and the machine learning model will learn from this indicator. Our national retail collaborator, has conducted powerful demand forecasting analysis and the data was provided in the dataset. Hence, we create sales gap feature to measure the difference between the predicted demand and the actual sales quantity.

K-means Clustering Model

Our first model is a K-means clustering. The goal is to further group the products with similar behaviors in supply and demand dynamics. Using a naïve elbow plot approach, we find k=2 and k=3 are strong candidates. Moreover, with the help of Silhouette Analysis, we discover that k=2 has the highest silhouette score (near 1.0), which may indicate one super-small cluster (e.g., just a handful of points) and one giant cluster. This partition might not be very informative. Meanwhile, k=3 whose silhouette score is around 0.45 has a reasonable drop in inertia on the elbow plot, indicating that it is a natural choice of the number of clusters and that it is a possibly more balanced partition as shown in Figure 1. In many practical cases, a silhouette score of 0.4–0.6 is considered okay if it produces more interpretable clusters. And this makes sense since the target variable is binary, 0 or 1 (phantom inventory flag). Considering that we also want a bit more granularity (e.g., “low,” “medium,” “high”) for the model, we will use k=3 for clustering the products.

Figure 1: Elbow Method (Left) and Silhouette Analysis (Right) for Optimal k

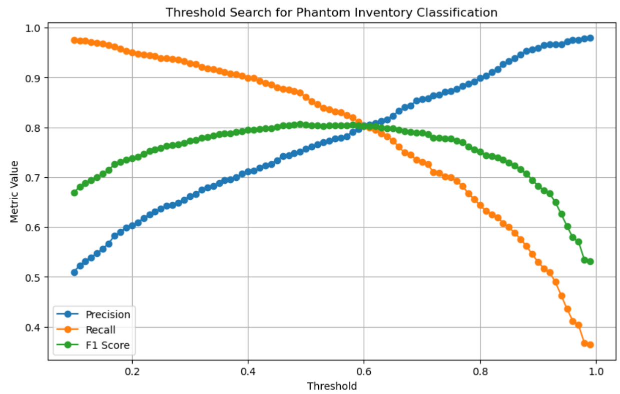


Random Forest Model

Based on the clusters, our second model was a Random Forest that will be used to classify phantom inventory and non-phantom inventory. Prior to building the random forest model, we tested a number of boosting algorithms which turned out to perform much worse than random forest. All of these algorithms resulted in precision, recall, and F1 scores strictly below 0.8, while random forest model outperformed all of them by giving all of the three evaluation metrics above 0.8.

We explored time sequential splitting techniques for our dataset which is inherently in time sequence. Counterintuitively, the random splitting turned to be a better choice. Thus, we randomly split the data into 80/20 for training-test set evaluation. We used the cross-validation process as an internal validation step. This approach saved us from having to explicitly carve out a separate validation set, especially when data provided was limited. The best threshold search for phantom inventory classification was found to have a threshold of 0.60 which provided a reasonable trade-off for balancing precision and recall. This is depicted in Figure 2.

Figure 2: Threshold Search for Phantom Inventory Classification



The results of random forest will be presented in the next section.

Long Short-Term Memory (LSTM) Networks

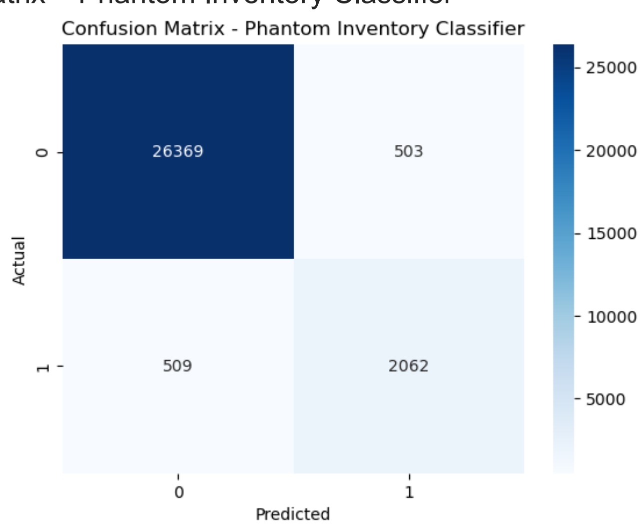
The LSTM time series model was developed and the configuration implemented was as follows. A single-layer LSTM comprised of 50 units with ReLU activation, followed by a dense layer outputting a single value. This was tested against a stacked LSTM configuration by adding a second LSTM layer (30 units) with return_sequences=True, enhancing the model's capacity to capture complex patterns. From our exploration we retained the single-layer LSTM as our final selection because it was more stable than stacked LSTM. Since this dataset is not that complicated for time series, the single layer performs well.

RESULTS

Phantom Inventory Detection

For phantom inventory detection, the result of random forest model is reasonable, as shown in the Figure 3 confusion matrix. The phantom inventory event is marked as Class 1. Precision is the proportion of detected phantom inventory events that are truly phantom inventory events, i.e., it tells how many of the phantom inventory alerts are correct, whereas recall is the proportion of actual phantom inventory events that are correctly identified.

Figure 3: Confusion Matrix – Phantom Inventory Classifier



Precision for phantom-inventory class (Class 1): 0.8039 while recall for Class 1: 0.8020. In phantom inventory detection, both precision and recall are important metrics, but their relative importance depends on the consequences of false positives and false negatives. If the retailer prefers to minimize the impact of false positives (e.g., to avoid unnecessary actions), then a high precision is needed. However, if failing to identify a phantom inventory is costly, then a high recall is required. We note that there is a trade-off between the two metrics, and hence we choose to put a result for both in a balanced way.

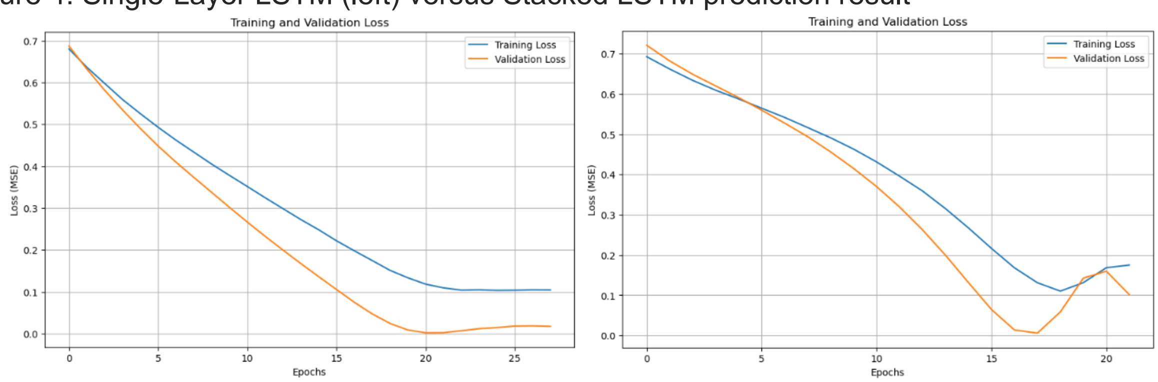
Phantom Inventory Prediction

The LSTM model predicts the next day's inventory iteratively over a 7-day horizon. Starting with the last observed window, each prediction updates the input sequence by shifting the window and appending the forecast. For multivariate inputs, additional features are held constant at their last observed values. Predictions are inverse transformed to return to the original DailyBOH scale.

We were able to predict OOS in the following way. The first day when forecasted DailyBOH falls to zero or below is identified as the out-of-stock date. If no such event occurs within the horizon, no run-out date is recorded. The predictive modeling performance is assessed using MSE on the original DailyBOH scale, comparing LSTM forecasts to a naive baseline (repeating the last observed value). The model's training loss (scaled MSE) is also monitored.

The performance of LSTM model is presented by the graph of training vs. validation loss in Figure 4. We include both Single-Layer LSTM and Stacked LSTM results for reference.

Figure 4: Single-Layer LSTM (left) versus Stacked LSTM prediction result



CONCLUSIONS

In this study, we defined and detected phantom inventory using key inventory features such as the daily balance on hand recorded at the end of each day, and the label data captured hourly by camera system. Furthermore, we were able to predict the likelihood of out-of-stock occurrences for the month of November 2024. In the end, we managed to detect and predict phantom inventory with the help of random forest and clustering techniques for categorizing merchandise, and time series models for forecasting the potential stockouts.

Beyond the used approaches in our study, other machine learning models may be investigated to improve the accuracy of phantom inventory detection. Supervised learning techniques may help to classify stock discrepancies based on historical data, while unsupervised learning techniques can help detect hidden patterns in inventory inaccuracies. We wish that our study not only contributes to reducing loss of sales, improving customer experience and achieving operational excellence, but also provides a scalable solution that may be adapted to various and broader industrial environments.

REFERENCES

Bassamboo, A., Moreno, A., & Stamatopoulos, I. (2020). *Inventory auditing and replenishment using point-of-sales data*. Production and Operations Management, 29(5), 1219–1231. DOI: 10.1111/poms.13153

Crowe, E. (2023, Oct 26). *Why Retailers Must Address Phantom Inventory Issues*. Progressive Grocer. https://progressivegrocer.com/why-retailers-must-address-phantom-inventory-issues

DeHoratius, N., & Raman, A. (2008). *Inventory record inaccuracy: An empirical analysis*. Management Science, 54(4), 627–641. DOI: 10.1287/mnsc.1070.0814

Farias, V. F., Li, A. A., & Peng, T. (2024). *Fixing inventory inaccuracies at scale*. Manufacturing & Service Operations Management, 26(3), 1102–1118. DOI: 10.1287/msom.2023.0146

Fleisch, E., & Tellkamp, C. (2005). *Inventory inaccuracy and supply chain performance: a simulation study of a retail supply chain*. International Journal of Production Economics, 95(3), 373–385. DOI: 10.1016/j.ijpe.2004.02.003

Mayer, M. (2023, Oct 26). *Phantom Inventory Haunts Retailers: Study*. Supply & Demand Chain Executive. https://www.sdcexec.com/warehousing/retail/news/22877459/retail-insight-phantom-inventory-haunts-retailers-study

McCue, T. (2019, Jan 31). *Inventory Shrink Cost The US Retail Industry $46.8 Billion*. Forbes. https://www.forbes.com/sites/tjmccue/2019/01/31/inventory-shrink-cost-the-us-retail-industry-46-8-billion/

Rekik, Y., Syntetos, A., & Glock C. (2019). *Inventory Inaccuracy in Retailing: Does it Matter?* ECR Community Shrinkage & On-shelf Availability Group. https://www.cardiff.ac.uk/__data/assets/pdf_file/0003/1729236/84486-ECR-Inventory-Inaccuracy-Report-v12.pdf

Riaz, O. (2021, Nov 15). *Amazon Inventory Management: Solutions To Common Problems*. Forbes. https://www.forbes.com/councils/forbesagencycouncil/2021/11/15/amazon-inventory-management-solutions-to-common-problems/

Tripathy, S. (2019). *Detection of phantom inventories at retail stores using a Bayesian approach* (Master's thesis, Massachusetts Institute of Technology). https://dspace.mit.edu/bitstream/handle/1721.1/122257/1119555264-MIT.pdf