CREATING A SAFE AND SUSTAINABLE OMNI-CHANNEL RETAIL ECOSYSTEM USING PREDICTIVE AND PRESCRIPTIVE MACHINE LEARNING

by

Somedip Karmakar, M.STAT.

DISSERTATION
Presented to the Swiss School of Business and Management Geneva
In Partial Fulfillment
Of the Requirements
For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

OCTOBER, 2024

CREATING A SAFE AND SUSTAINABLE OMNI-CHANNEL RETAIL ECOSYSTEM USING PREDICTIVE AND PRESCRIPTIVE MACHINE LEARNING

by

Somedip Karmakar

Supervised by

Dr. Anuja Shukla

APPROVED BY

____

Dissertation chair

RECEIVED/APPROVED BY:

____

Admissions Director

Dedication

This research thesis is dedicated with heartfelt gratitude to my remarkable parents, whose unwavering love and support have been my anchor during this journey. Your sacrifices, motivation, and belief in my capabilities have been the driving force behind my research. I am immensely grateful to my teachers and mentors who have been instrumental in shaping my academic career. Your guidance, constructive criticism, and encouragement have not only equipped me with the necessary knowledge and skills but also instilled in me the confidence to overcome challenges in my pursuit of knowledge. Your collective contribution to my research journey has been invaluable. This work is a testament to your unwavering faith in me, and for that, I am forever thankful. Your teachings will continually echo in my future endeavors. Thank you for helping me turn an ambitious dream into a tangible reality.

Acknowledgements

I would like to express my deepest gratitude to my research guide, whose profound knowledge, guidance, and patience have been critical throughout this research journey. Your constant intellectual challenges and unwavering support have made this thesis possible. My sincere thanks also go to my mentors and teachers, who have nurtured my curiosity and fueled my passion for learning. Your insightful advice, invaluable feedback, and constant encouragement have been fundamental in shaping this research project. A special thank you to all the participants from the retail industry who willingly shared their time, experiences, and insights. Your contributions have been pivotal in enhancing the depth and breadth of this study. Your willingness to participate and share your knowledge has not only enriched this research but also contributed significantly to my understanding of the retail industry. I am eternally grateful for your invaluable assistance, without which this thesis would not have been possible. Your influence on this work extends far beyond what is seen on the pages of this thesis, and for that, I sincerely thank you.

ABSTRACT

CREATING A SAFE AND SUSTAINABLE OMNI-CHANNEL RETAIL ECOSYSTEM USING PREDICTIVE AND PRESCRIPTIVE MACHINE LEARNING

Somedip Karmakar
2024

Dissertation Chair: Gualdino Miguel Cardoso
Co-Chair: Apostolos Dasilas

The retail ecosystem has significantly evolved from simple neighborhood stores to large omni-channel retail systems. With razor-thin margins, the retailers need to constantly innovate to reduce losses from fraud, shrink, waste, lost sales opportunities, out-of-stock and many other scenarios. Recent advances in omni-channel retailing have also led to the rise of a new group of fraudsters who abuse the gap between the online and offline channels and take advantage of the relaxed retail company policies. Fraud can be related to payment, account take over, refund and cancellation abuse, associate collusion, marketplace, and even call-center abuse. Waste can be due to over-production, sub-optimal pricing or discounts, damaged products, throwaways, availability issues, packaging waste.

The paper tries to identify the need for better research on improvement in fraud detection systems for sparse data, improvement of the customer experience and reduction of returns. The paper discusses how the advanced predictive and prescriptive ML models can help in building tools and recommendations which reduce the losses and make the retail ecosystem safe from fraud and sustainable against wastage.

The data analysis is performed on omni-channel retail transactions over a period of two years across multiple countries, and feedback is collected from total loss analysts and store associates through interviews, surveys, and interactive tools. Machine Learning methods of supervised and unsupervised learning including an ensemble of anomaly detection, graph analytics, classification models with gradient boosting, reinforcement learning, can be used to highlight the high priority cases to focus upon, and then identify the root causes of such losses. Then, Causal Discovery models can be used to identify root causes and provide prescriptive recommendations. The ecosystem of ML assisted solutions can support in multiple aspects of running retail store operations starting from inventory management, predictive replenishment, shrink reduction, effective disposition, waste management, and finally combating different malicious fraud and abuse activities, all of which lead to different types of loss to the retailers. The paper provides a comprehensive evaluation of the challenges in omni-channel retail ecosystem and the proposed machine learning tools which can help resolve the issue at large. It also proposes an effective impact measurement solution based on Causal Inference techniques which can easily measure the impact of the changes made on important KPIs like savings, wastage, returns, etc. and tailored to both online and offline retail channels.

The implementation of ML techniques in retail total loss management can lead to multi-million-dollar savings through better fraud detection, inventory management and waste management. Performance of supervised models improves to close to 98% over the course of months by using feedback loop.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

1.1 Introduction to Retail Safety and Sustainability

Retail is a business which sells goods and services to customers, and it is one of the most essential services with high importance for all. Irrespective of economic situation, purchase of goods and products has been a core tenant of civilization. With time, retail has grown into a booming business with widespread impact in everyday life. Its reach has grown from major supercenters to neighborhood markets and now its reach has reached all homes through computers and mobiles, powered by the internet. Retail is a difficult business, with very little to negative margins on large volume of grocery to moderate to large margins on seldom purchased electronics and general merchandise. Such razor-thin margins are further compromised by stiff competition, growing cost and infrastructure expenses. Also, customers expect ease and convenience of shopping, with super-fast delivery, expectation of high quality and support, easy return, and cancellation policies. It is no wonder, that many retailers are not able to survive especially with disruptive changes like pandemic or inflation. Retail loss can be attributed to many different reasons like fraud, shrink, returns, waste, item availability, over-production. National Retail Security survey has identified retail loss to cause $94.5 billion loss, growing more than 4% year over year (NRF, 2023). The explosive growth of e-Commerce and Omni-channel retail has seen a rise in fraudulent activities, especially those exploiting the gap between the different systems, and it is growing at 30% annually according to LexisNexis® Risk Solutions' paper "True Cost of Fraud" in 2018. In US, out of every $100 in turnover, fraudsters currently snatch 5.65 cents. Recent economic conditions have led to an increase in violence, shoplifting, collusion, and fraud across the globe. The omni-channel retail revolution has led to even more growth of retail loss. Poor inventory management and planning has led to huge crisis with over-flowing backrooms in stores, item unavailability due to incorrect inventory recording and over-production and over-procurement of items leading to huge waste. It is very important for retailers to provide a safe, trustworthy, frictionless, and sustainable omni-channel shopping experience to the customers. Since the margins are already very low, it is also important for retailers to prevent loss as much as feasible.

1.2 Research Problem

Retail fraud like shoplifting and collusion is an age-old problem, but this has now been amplified as we move to omni-channel retail. Since retail companies were not developed in an omni-first mindset, they are either traditional brick and mortar stores who adopt the online ecosystem or ecommerce companies setting up new physical capabilities to interact with customers better. There exist a lot of gaps in these systems and a small group of fraudsters are taking advantages of these gaps to exploit the system, which erodes the trust of the retailers and creating unnecessary friction for the majority of genuine customers to shop on the platform. Similar scenarios lie in the waste management and sustainability. Items which are bought online and returned at stores may not find the appropriate avenues to be appropriately resold at stores or proper disposition path, which can increase overall waste for the stores. Moreover, most retailers don’t provide prescriptive guidelines to the store associates for waste management, and any action being taken is subjective and not uniform. Here is a great opportunity of unifying and identifying best practices to be followed to become a sustainable retailer.

1.3 Purpose of Research

The role of Machine Learning and Artificial Intelligence in solving retail loss and making it safe and sustainable is critical, although existing research has focused only on solving parts of the problem, mainly focusing on payment fraud prevention. According to Nanduri et. al (2020), dynamic nature of fraud patterns makes it challenging to tackle ecommerce fraud and models can miss out on prediction of previously unseen patterns. These ML systems also face the problem of ignoring the follow-up decisions made by other associated parties during implementation phases. Weber and Schutte (2019), mentions that there is a multitude of possible applications of ML in all areas of retail and wholesale. Due to the circumstances of the retail trade, the number of obstacles to data analysis is particularly high. The huge assortment and fast sales pattern across multiple stores create huge quantity of data, which is not possible for manual analysis. One path forward can be to develop testing of hypothesis on different fraud modus operandi to identify the features related to fraud.

Although ML based solutions can have a huge opportunity to reduce the cost and optimize the business process, yet such processes are not adopted to full capacity. Retailers need to review their ways of working with different partners like suppliers, negotiate better and help to improve forecasts of suppliers. Through these steps, they can develop sustainable and long-lasting improvements using the power of Machine Learning. This begs the question of how ML techniques can be effectively used to reduce waste and fraud and make the retail ecosystem safe and self-sustaining. The research will showcase the importance of predictive and prescriptive Machine Learning techniques in retail loss prevention.

1.4 Significance of the Study

In the rapidly evolving retail industry, omni-channel retail has emerged as a growing means of engaging with customers and gaining market share. At the same time, gaps in the different parts of retail ecosystem led to inefficiencies of processes to prevent loss in such multi-channel shopping journey. Although advanced Machine Learning and Artificial Intelligence techniques are being used in different parts of retail industry, there is a clear gap in terms of prescriptive ML solutions to prevent omni-channel loss, especially in terms of pro-active fraud prevention and waste reduction. There is a huge opportunity in terms of hundreds of millions of dollar savings through implementation of tools and techniques which can act as safeguards and provide clear recommendations to store associates to reduce such shrink and wastages.

While existing ML solutions focus only on eCommerce card payments risk, there are no available techniques to solve for omni-channel fraud with multiple points of collusion of customers, associates, drivers, agents. Any external fraud engines without proper retail context can lead to false positives, and poor customer experience. Retailers find difficulty in accurate forecasting of demand especially for short-lived fresh items across stores and e-Commerce, leading to huge wastage. Major problem for retailers in 2022 came out in the form of inventory management. Due to inflation and changing customer preferences, overstocking and over-production became a huge problem across the world and led to huge wastage. On the other hand, just having the item in the backroom and not properly stocked on the shelves can cause item availability and Out-of-Stock problems, leading to missed sales opportunities.

With rapid advances in retail landscape, the gap between physical and digital world has come down. As such, the customers expect frictionless and seamless purchase experience in all modes of purchase. In this evolving scenario, the retailers can stay competetive by taking advantage of the data and deriving actionable insights from it using advanced ML models. Such models can help the retailer to improve on its thin margins and this in-turn helps to provide better service and convenience to the customers.

With the end of the pandemic, more customers have started visiting stores, and also enjoy convenience of online shopping. The divergent and non-uniform systems that power these two separate part of the retail business can lead to gaps that can be exploited by fraudsters, or can lead to process gaps that lead to wastage and unknown shrink in the physical stores. Many a times, the inventory remains in backroom, and is not accessible to customers, which can cause both lost sales opportunity as well as poor customer experience.

The goal of this paper is to uncover the instances of opportunities that can be discovered and resolved using advanced ML techniques. This can include different types of unintentional and malicious loss opportunities and providing actionable and prescriptive recommendations to prevent future losses.

1.5 Research Purpose and Questions

Supervised ML models depend heavily on good quality labelled data, but such is rarely available in omni-retail scenario, due to lack of time and bandwidth of subject matter experts. This problem gets aggravated during rare-event scenarios like fraud or waste. While AI and ML solutions have progressed a lot, there are challenges in adoption by business due to lack of interpretability of the solutions. Also, most of the predictive ML models fail to provide prescriptive recommendations that store associates can follow. There have been recent advances in Causal Inference, ML and Discovery models, but they have not been studied or applied in context of retail loss prevention.

There is a need for a better understanding of how predictive and prescriptive Machine Learning solutions can help to reduce retail loss and make omni-channel retail safe and sustainable. More specifically, the following research questions need to be addressed:

1. How Machine Learning can help in eliminating gaps between offline and online channels which lead to fraudulent activities?
2. What are the major causes of Retail Waste and how prescriptive Machine Learning recommendations can help to reduce waste?
3. How can early Detection of Out-of-Stock using Computer Vision help in real-time Inventory management to reduce waste due to overstocking?
4. How can better demand planning for Fresh production help in reducing wastage due to over-production ?

Therefore, our research aims to find the optimal approach to reduce retail loss by using advanced ML models.

CHAPTER II:

REVIEW OF LITERATURE

2.1. Introduction

Retail Loss is a major concern especially in uncertain times of inflation. Loss can be due to several reasons like fraud, shrink, returns, waste, item availability, over-production. According to National Retail Security Survey (2019), shrink rate is around 1.38%, which has an estimated $50.6 billion impact on the retail industry. The explosive growth of e-Commerce and Omni-channel retail has seen a rise in fraudulent activities, especially those exploiting the gap between the different systems, and it is growing at an alarming rate. In US, out of every $100 in turnover, fraudsters currently snatch 5.65 cents. In a competitive retail industry operating on razor thin margins, this represents significant loss and essentially increases prices for other non-fraud customers.

Retail loss prevention executives engage in a constant battle to detect and prevent inventory shrinkage. Inventory shrinkage is the financial loss attributable to a combination of employee theft, shoplifting, administrative error, and vendor fraud. (Holliger & Davis, 2002 National Retail Security Survey).

In the book "The Challenges to Preventing Losses in Retailing" (2018), Martin Gill informs us that there has been relatively little research on the ways in which crime prevention decisions in the commercial environment take account of the need to protect people and property and other assets on the one hand, and yet don't impede the commercial imperative on the other.

According to Adrian Beck (2020, Total Retail Loss 2.0), challenges that retailers face for effective total loss prevention are : data being spread across the business (44%) and not having the necessary data (23%). This is then followed by issues relating to the current organizational structure (14%). Less than one in ten felt that a lack of senior management support was a barrier (9%). Beck has classified Retail Total Loss as coming from different sources like Store, eCommerce, Supply Chain and Corporate. These are further divided into known and unknown stock loss. Even within the known loss there are malicious loss (fraud) or non-malicious loss which are process opportunities.

According to the Food and Agriculture Organization of the United Nations (FAO, 2018), over 1.3 billion metric tons of food goes to waste worldwide every year. This amounts to a full third of the world's food production, and represents an immense amount of wasted resources, including water, energy, land, and labour. According to the Natural Resource Defence Counsel, 40 percent of the food in the U.S. goes uneaten. The NRDC says the food is tossed for a variety of reasons including overstocking, low presentation standards, and misleading date labels. One of the major problems for any retail operation is managing the flow of waste out of their brick and mortar operations. The amount of waste generated can fluctuate based on shifts in their inventory or other internal events. The result is an inefficient system, where there is either too much or too little waste whenever pick-ups occur.

As long as there is inventory in the display shelves and in stockrooms, retail stores face the challenges of retail waste management and disposal (Actenviro, 2022) . Retail waste begins when the ordered items are prepared for shipping from the manufacturer to the retailer. Upon arrival at the retailers, the boxes are often wrapped again for inventory and stock before displaying on the shelf. The disposal of these plastics, stuffing, and containers—whether from the retailers or consumers—all contribute to the retail waste problems. For example, Amazon shipped 5 billion items around the world in 2017.

The role of Machine Learning and Artificial Intelligence to solving retail loss and making it safe and sustainable is critical, although existing research has focused only on solving parts of the problem, mainly focusing on payment fraud prevention. Hence it is important to review the past applications of Machine Learning models in Fraud prevention, Return reduction, Inventory Management and Waste Reduction, to make omni-channel ecosystem safer and more sustainable.

2.2 Machine Learning Methods for Omni-Channel Fraud Prevention

2.2.1 Anomaly Detection to detect emerging unknown risks

Anomaly detection is a very important problem which has been researched within diverse research areas and application domains. Many anomaly detection techniques have been specifically developed for certain application domains, while others are more generic. (Chandola et al,2009).

According to Li et. al (2016), in the big data era, the digital revolution has driven the entire financial industry to collect, store and analyze massive volumes of data nowadays than it ever has in history. With the overwhelming scale of data, new technologies are needed to derive competitive advantage and unlock the power of the data, including the approaches people use for financial risk management. Recent years have witnessed a paradigm shift in the financial industry as awareness of the importance of data is becoming widespread. With the overwhelming scale of data, new technologies are needed to derive competitive advantage and unlock the power of the data, including the approaches people use for financial risk management. Risk management contains various domains and application scenarios, e.g., credit risk- the risk of default on a debt;

market risk- the possibility of losses due to the movements in market prices; frauds deliberately committed by human beings- trading fraud, financial statement fraud, and insurance fraud; and project management risk- the potential failure factors in project lifecycles, spread from the design, development, production, and sustainment of a project. In this dissertation, the main focus is on financial risk management, which means the violation of good behaviors in regards to financial losses. Data mining techniques are deployed to scour large databases in order to find novel and useful patterns that might otherwise remain unknown. They also provide capabilities to predict the outcomes of future observations. The focus is to develop effective and efficient data analysis techniques to detect financial anomalies and mitigate potential risks. The key challenge is how to address the unique characteristics of different data repositories and develop suitable techniques to meet the specific needs of a particular business application. Specifically, we aim to provide case studies to apply advanced data mining techniques in the following three applications, trading fraud detection, contract risk management, and f inancial statement fraud detection. With the development of information technology, it is easy to collect and store massive amount of data in financial service industry. The detection of trading ring pattern is of great value to regulators, since it can help to reduce their workload by preliminarily filtering out a huge number of normal patterns automatically, thus let them be more focused on the suspicious patterns that may commit frauds. In addition, the detected trading ring patterns can provide a view of the interactions among some accounts in the trading network, which can help to detect the collaborative fraud activities for suspicious accounts with trading volumes under the threshold of the system. Financial fraud detection is one of the most important aspects in financial risk management. In recent years, there is an increasing trend in financial losses due to financial fraud and bankruptcy. However, traditional auditing process and risk analysis could not match the increased demand of automatic and efficient detection of financial statement fraud. Hence, how to develop an efficient and effective financial fraud detection framework draws the best interests among investors, public, researchers, auditors and regulators. In general, the published financial statements are one of the most pervasive and consistently available predictor of a companys future performance, since they provide the basis for understanding and evaluating the financial status of a company. However, fraudulent financial reporting can destroy the true picture of a companys financial situation. For example, by manipulating elements such as liabilities, expenses, or losses in the financial statement, the unhealthy financial status can be covered. Nowadays, financial statement fraud is one of the most notable management frauds in the U.S. due to the large financial losses it yields. Therefore, financial statement fraud detection becomes a crucial and pervasive issue in financial risk management. Given a directed graph, a blackhole pattern is a group which contains a set of nodes in a way such that the average in-weight of this group is significantly larger than the average out-weight of the same group. In contrast, a volcano pattern contains a set of nodes where the average out-weight is significantly larger than the average in-weight. The problem of finding volcano patterns is a dual problem of mining blackhole patterns. Therefore, we focus on discovering the blackhole patterns and develop a blackhole mining framework. Specifically, we design two pruning schemes for reducing the computational cost by reducing both the number of candidate patterns and the average computation cost for each candidate pattern. The f irst pruning scheme is to exploit the concept of combination dominance to reduce the exponential growth search space. Based on this pruning approach, we develop the gBlackhole algorithm. On the other hand, the second pruning scheme follows an approximate strategy. We improve the computational efficiency by first screening out nodes with small diff-weights to reduce the size of the graph, and then mining the top-K blackhole patterns in the subgraph induced by the rest of the nodes. This approach strikes a balance between the efficiency and the completeness of blackhole mining. Existing classification methods kNN can be used to partially solve our problem. However, a critical challenge of applying this method is to define the right metric that can be used to gauge similarity (or distance) between contracts. Along this line, we extend the Mahalanobis distance metric learning framework, and formulate a constrained optimization problem such that the learnt distance for each pair of similar points is close to their actual distance, while each pair of dissimilar points can be well separated. A key advantage of the proposed method is the ability to train model with not only continuous distance measures between contract pairs, but also the binary side information of dissimilar pairs. Finally, experimental results on real-world service contract data show that our proposed approach greatly outperforms existing benchmarks, and can provide more accurate contract risk assessment.

Traditional auditing process and risk analysis could not match the increased demand of automatic and efficient detection of financial statement fraud. Data mining techniques are used to help address this challenge. In recent years, there is an increasing trend in financial losses due to financial fraud and bankruptcy. A number of high-profile companies, such as Enron, World Corm, Xerox and Lucent, were charged with fraud by the US Security and Exchange Commission (SEC) in the past decade. Hence, how to develop an efficient and effective financial fraud detection framework draws the best interests among investors, public, researchers, auditors and regulators. Data mining techniques, with advanced clustering and predictive capabilities, could facilitate auditors and managers in accomplishing the task of management fraud detection and early-warning decision making. The aim of the study is to utilize clustering techniques to extract useful knowledge from bankrupt companies using published financial statement data. Li et. al (2016) proposed two pruning approaches to reduce the computational cost by decreasing both the number of combinations and the average computational cost for each combination.

According to Vanini et. al (2023), the goal of anomaly detection is to detect fraudulent activities in e-banking systems and to maintain the number of false alarms at an acceptable level. The implementation of the model consists of three steps: pre-filter, feature extraction, and machine learning. Non-learning pre-filters ensure that both obvious fraud and normal transactions are sorted early to reduce the false positive rate. Only transactions that pass the pre-filter step are passed on to the machine-learning model. Banks utilize non-customer-based static if-then rules, such as blacklists or whitelists. Pre-filters free the algorithms from obvious cases. The adaptability and flexibility of the machine-learning model is necessary to counter the ever-improving attacks of fraudsters with effective fraud detection. Large transaction volume with the need for real-time fraud detection, a highly imbalanced dataset, dynamic fraud behavior, limited forensic information, and varying customer behavior. Given the extremely imbalanced data, fully supervised algorithms typically struggle. For each customer participating in an e-banking session, we assess whether the agent's behavior is consistent with the account holder's normal behavior. The key information for behavioral analysis lies in the sequence of the customer's clicks during the session. As there are very few known fraud cases, all base learners are trained on fraud-free data only in step one. Fraud cases are only utilized in step two of ensemble aggregation when the base learners are combined to form the final predictive function. The first step is to define base learners who are rich enough to detect a wide range of suspicious transactions or online user sessions. T he fraud detection model calculates scores and, in comparison with a threshold value, decides whether a transaction is flagged as an anomaly. This process results in the probability of detection for a given investigation effort, as indicated by the ROC curve. By making the threshold dependent on the transaction size, we can ensure that larger transaction amounts are more likely detected than smaller ones. This gives up part of the true positive rate (TPR) to reduce overall economic losses (i.e., the TPR decreases for a given FPR). This economic optimization that leads to adjusted ROC curves defines the triage model. To minimize expected cumulative losses, the constant fraud anomaly detection threshold becomes a function of the transaction amount. Here, the transaction amounts are random variables whose distributions are estimated. In the optimization problem, the transaction function is chosen to maximize the average cumulative sum of the detected fraudulent transactions, where the expected FPR must not exceed a certain threshold. Utilizing this optimal threshold function, the fitted ROC curves were obtained. The optimization problem has a unique solution if the ROC curve is a concave function of the false positive rate function of the threshold and if the acceptance set of the expected false positive function constraint is convex. With the chosen piecewise linear false positive constraint function, the assumptions regarding the existence of an optimum are satisfied. The ROC curves that result when fraud anomalies are detected serve as inputs for the optimization. However, because only a vanishingly small number of fraud cases exist, the TPR values for certain FPR levels are subject to considerable uncertainty. Hence, cubic spline functions were utilized for the ROC curve of the optimization. Losses from transaction fraud are included in operational risk incurred by banks. As for other operational risks, one of the key questions from a risk-management perspective is whether the allocated resources and countermeasures are adequate. To answer this, one needs some way of quantifying the risk incurred, ideally a Value-at-Risk (VaR)-model that fits the general risk framework of the bank. Unlike market or credit risk, fraud risk is borne by comparatively few individuals or groups who utilize very specific strategies and technologies to overcome vulnerability in the payment process. Simultaneously, defenders analyze attack plans and update their countermeasures. In this constantly changing environment, neither the frequency of attacks nor the transaction amounts can be assumed to be statistically regular with great certainty. Therefore, we propose a simple, flexible stochastic model composed of basic building blocks. With such a model, risk managers can quickly adjust the model as needed, perform what-if analyses, or simulate changes in payment infrastructure. Defense against sophisticated online banking fraud involve several resources and methods. These include risk models, algorithms, human action, knowledge, computer tools, web technology, and online business systems in the context of risk management. As fraud is part of regulated operational risk, our model allows us to analytically capture these operational risks without crude benchmarking. This also provides a microeconomic foundation for capital adequacy. In the area of operational risk, these results put internal models that are not risk sensitive or difficult to verify on a solid footing. A complicated problem, such as online payment fraud detection, requires a comprehensive understanding. A prerequisite for this is access to a large dataset. To evaluate our method, we utilized a real dataset from a private bank. Regardless of the chosen algorithm, feature extraction is an essential part of developing an effective fraud detection method. We utilized historically observed and confirmed fraudulent transaction identifiers as the ground truth. Each feature in the feature vectors for each e-banking session aims to encode deviations from normal customer behavior. Thus, behavioral, transactional, and customer-specific features are important.

According to Nanduri et. al (2020), one major challenge in tackling e-commerce fraud results from dynamic fraud patterns, which can degrade the detection power of risk models and can lead to them failing to detect fraud that has emerging unrecognized patterns. The problem is further exacerbated by the conventional decision frameworks that ignore the follow-up decisions made by other associated parties (e.g., payment-instrument-issuing banks and manual review agents). These kinds of problems can be handled well using anomaly detection. Many existing fraud detection systems typically operate by adding fraudulent transactions/accounts to “blacklists”, to match for likely frauds in the new instances (Phua et al, 2010). Unsupervised approaches such as outlier detection, spike detection, and other forms of scoring have been applied. Yamanishi et al (2005) demonstrated the unsupervised SmartSifter algorithm which can handle both categorical and continuous variables, to detect statistical outliers using Hellinger distance. Bolton and Hand (2002) also suggest Break Point Analysis to monitor intra-account behavior over time. In the long term, less complex and faster algorithms produce equal, if not better results on population-drifting, concept-drifting, adversarial-ridden data. Hence it is important to identify and estimate the proper underlying distribution before we can use anomaly detection techniques. Multivariate Anomaly detection techniques like Isolation Forest outperforms one-class SVM, LOF and Random Forests in terms of AUC, processing time, and it is robust against masking and swamping effects. (Liu et al, 2012). According to Shih (2021), reverse logistics services (returns and exchanges) have become the target of abuse or fraudulent activities which have caused a lot of economic losses for many online retail companies. While traditional anomaly detection systems rely on the underlying distribution to be Normal or Uniform, but such is rarely the case for rare event scenarios like fraud.

2.2.2 Payments Fraud using Supervised Models

Credit card has become popular mode of payment for both online and offline purchase, which leads to increasing daily fraud transactions. Most fraud-detection systems monitor credit card transactions by means of classifiers which return alerts for the risky payments (Pozzolo et al, 2015). These differ from conventional classification because only a small set of supervised samples is provided by human investigators who have limited bandwidth. Labels of many transactions are made available only several days later, when customers have possibly reported unauthorized transactions. The delay in obtaining accurate labels and the interaction between alerts and supervised information must be carefully taken into consideration when learning in a concept-drifting environment. Various supervised models like Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB) have been used for credit card fraud detection (Wang et al, 2019). The advantage of supervised learning is being capable to achieve very promising results given sufficient training data, while the disadvantage is being dramatically affected by the data imbalance issue and the data labeling processing. Kulatilleke (2022) notes that another challenge in supervised models for payments fraud is that human annotation errors compound with machine classification errors. The cost of fraud detection should not be higher than the fraud volume itself. In addition to a severe lack of suitable machine learning data sets for fraud detection research, of those available, each credit card data set gives conflicting characteristics as well as fraud ratios which could be the cause of the inability to obtain an ideal general classification algorithm. While majority of the existing methods of fraud prevention focus heavily on credit card fraud, there is very minimal research on other payment modes like Paypal, wallet, Pay at store or pay on delivery and seller payment fraud. Also, there is not much research which can segregate refund abuse and fraud from the genuine returns through supervised learning.

2.2.3 Linkages and Graph Analytics

Sadowksi (2014) informs us that sophisticated fraudsters have developed a variety of ways to elude discovery, both by working together, and by leveraging various other means of constructing false identities. Graph databases offer new methods of uncovering fraud rings and other sophisticated scams with a high-level of accuracy. Consider an online transaction with the following identifiers: user ID, IP address, geo location, a tracking cookie, and a credit card number. One would typically expect the relationships between these identifiers to be close to one-to-one. Some variations are naturally to be tolerated to account for shared machines, families sharing a single credit card number, individuals using multiple computers, and the like. However as soon as the relationships begin to exceed a reasonable number, fraud is often at play. A layered approach has emerged as a best practice for detecting fraud. While existing analysis techniques are sufficient for catching certain fraud scenarios, sophisticated criminals often elude these methods by collaborating. Criminal rings are very skilled at concealing collusion and inventing and staging complex “paper collisions” that do not arouse suspicion. In the recent years money laundering schemes have grown in complexity and speed of realization, affecting financial institutions and millions of customers globally. Strengthened privacy policies, along with in-country regulations, make it hard for banks to inner- and cross-share, and report suspicious activities for the AML (Anti-Money Laundering) measures. (Stavarache et al, 2019). Graph networks can be used to identify rings of fraudsters through exact and approximate linkages, however there is a possibility of creating false linkages if the attributes like names, email, etc. are too common, and it becomes even more difficult for approximate similarity. Another major challenge is managing the huge data volume for pairwise linkages if the number of customers is too high. Online marketplaces are now a popular way for users to

buy and sell goods over the Internet. On these sites, user reputations—based on feedback from other users concerning prior transactions—are used to assess the likely trustworthiness of users. However, because accounts are often free to obtain, user reputations are subject to manipulation through white-washing, Sybil attacks, and user collusion. This manipulation leads to wasted time and significant monetary losses for defrauded users, and ultimately undermines the usefulness of the online marketplace (Post et al, 2011). Graph analytics can also be potentially used to find similar sellers who have been suspended or terminated in the past, but attempts to come back using a false identity, but there is not enough research on this field.

2.2.4 Account Takeover Detection

With billions of usernames and passwords readily accessible via the black market, account takeover poses a significant threat to services that rely solely on passwords for authentication (Milka, G., 2018). Account takeover is a type of malicious attack where a fraudster steals accounts and passwords from normal users, causing the loss of money and the exposure of personal information. Existing solutions either rely on extensive manual labelling, and require behaviour sequences and context graphs of accounts (Gao, M., 2022). Identity providers can use risk analysis, in conjunction with ‘login challenges’, to bridge the security gap between two-factor authentication and password-only users with minimal additional friction. There are ongoing challenges such as how public opinion may be at odds with actions identity providers should take to improve overall account security. Coppolino (2015) informs us about using an approach based on the Dempster–Shafer (DS) theory, that results in high performance of the detection process, i.e. high detection rates and low false positive rates. The approach is based on combining multiple (and heterogeneous) data feeds to get to a degree of belief that takes into account all the available evidence. The proposed approach has been validated with respect to a challenging demonstration case, specifically the detection of frauds performed against a mobile money transfer (MMT) service. Doerfler (2019) has showcased that knowledge-based challenges prevent as few as 10% of hijacking attempts rooted in phishing and 73% of automated hijacking attempts. Device-based challenges provide the best protection, blocking over 94% of hijacking attempts rooted in phishing and 100% of automated hijacking attempts. However these methods hamper the usability limitations of legitimate users. The login challenges act as an important barrier to hijacking, but that friction in the process leads to 52% of legitimate users failing to sign-in-though 97% of users eventually access their account in a short period. This is an vital challenge, since in the process of preventing fraud, the process creates friction for genuine customers, leading to lost sales opportunity.

2.2.5 Challenges of Getting Labelled Data

One of the primary challenges of fraud prevention system is having good quality labelled data. Labelled data is a primary requirement for any type of supervised learning problem. Only unsupervised algorithms can achieve to detect very prominent cases of anomalies, however there is no priority order nor any indication as to if those cases are truly fraudulent. Getting even small sample of fraudulent cases is not sufficient, since it is a highly imbalanced problem, and the observed sample may be a biased representation. The lack of availability of training datasets for fraud detection and the difficulty in identifying the behaviour of sophisticated fraudsters hinder research on such fraud detection (Elshaar et al, 2020). At time of crisis, when fraud permanently frightens the basis of modern societies, the existence of effective tools to prevent it, or just to identify it in time, is critical. However, the detection of fraud is naturally impaired (among other issues) by the difficulty

on labelling data, due to the cost of identifying and attest fraud. Moreover, the inability to incorporate domain knowledge in the mining process makes classifiers to use inadequate attributes to distinguish entities, ignoring most of existing relevant information, like entities' social relations. (Botelho et al, 2011). Classifiers trained with traditional techniques for fraud detection usually present accuracy levels significantly below the average. The difficulty on correctly labelling data and the fact that agents in these domains are humans, also aggravate the challenge. In fraud detection any non-caught fraud just can be labelled as non-fraud. This means that an unknown amount of instances are incorrectly classified, which difficult the training process. The quantity of labelled instances may become the most difficult issue to overcome, since it depends on the availability and ability of domain experts to label existing data. In this context, semi-supervised clustering may be an important tool, due to its ability to deal with a reduce amount of labelled data. Manual labelling of data is both costly and time consuming. Therefore, in a real streaming environment, where huge volumes of data appear at a high speed, labelled data may be very scarce. Thus, only a limited amount of training data may be available for building the classification models, leading to poorly trained classifiers.

(Masud et al, 2008). Semi-supervised learning and active learning techniques have been used to deal with situations like this, however these techniques also require some amount of reliable labelled data and incorporation of domain knowledge to define the non-fraudulent aspects clearly. One approach could be to provide appropriate recommendations to the agents tagging the sample observations, however this approach requires further research.

2.2.6 Challenges of False Positives

Fraud detection problems are well-defined supervised learning problems, and data scientists have long been applying machine learning to help solve them. However, false positives still plague the industry with rates as high as 10-15%. Only 1 in 5 transactions declared as fraud be truly fraud. Analysts have pointed out that these

high false positives may be costing merchants more than fraud itself. To mitigate this, most enterprises have adopted a multi-step process that combines work by human analysts and machine learning models. This process can potentially reduce the false positive rate by 5% – but this improvement comes only with high (and very costly) levels of human involvement. Even with such systems in place, a large number of false positives remain (Wedge et al, 2018). According to Wallny(2022), in domain of credit card fraud, analysts have reported that fraud prediction in e-commerce still has to deal with false positive rates of 30-70%, and many cardholders reduce card usage after being wrongly declined. State-of-the-art fraud detection systems combine a confidential set of transaction blocking rules with a data-driven ML model. This model typically employs a classifier, trained by ML algorithms on examples of genuine and fraudulent transactions, which can assign a suspiciousness score to an incoming transaction. Transactions exceeding a pre-set threshold are declined. Due to the economic potential, ML for fraud detection has received a lot of attention, however, it remains challenging to achieve satisfying performance in practice. In particular, the susceptibility of the models to generate false declines (or false positives) is a major problem. They cause embarrassing situations for customers, lost revenue for merchants, and administrative overhead for issuers. Especially during the ongoing COVID-19 pandemic and the connected increase in online transactions, false positives became an even bigger problem. Quantification of the cost associated with false positives can be an underestimation, because a customer might switch to a new card issuer permanently. Further, it is likely that not only false positives have a deteriorating effect on customer experience and therefore cause cost. Also false negatives are likely to cause cost additional to the lost transaction amount. A customer who repeatedly found his card charged by fraudsters, without the interference of his bank, might also consider abandoning that card and switching to a new issuer even if she gets reimbursed. This requires further research on the financial impacts of false positives.

2.3 Machine Learning Methods for Waste Management

2.3.1 Time Series Forecasting for Over Production

Food waste is a global challenge that has significant environmental, social and economic implications. Food retailers are in a powerful position to influence food waste reduction by producers, manufacturers and consumers. (Huang et al., 2021). The food processing and manufacturing industry produces food losses and food waste throughout the entire production phase due to reasons such as: damage during transport or nonappropriate transport systems, problems during storage, losses during processing or contamination, inappropriate packaging. (Parfitt et al., 2010). Food retailers are more concerned with sales forecasting due to their special characteristics, such as the short shelf-life of their products, need to maintain high product quality and the uncertainty and fluctuations in consumer demands. As products can only be sold for a limited period of time, both shortage and surplus of goods can lead to loss of income for the company. The variations in consumer demand are caused by factors like price, promotions, changing consumer preferences or weather changes (Van der Vorst, Beulens, De Wit, & Van Beek, 1998). An initiative of several large companies in the food industry, which aimed to improve forecasting practice, identified that 48% of food companies are poor at forecasting (Adebanjo & Mann, 2000). Fresh products like milk, bread have a shelf-life of less than a week and it is delivered and stocked on shelves daily. Therefore, stockpiling by customers is unlikely to happen and it could only be limited to a small extent. Kondo and Kitagawa (2000) presented a methodology for time series analysis on milk sales that allows close examination of some factors that influence milk sales, such as trend, regular variation during weekdays and promotions. Successful sales forecasting reduces considerably the lost sales and products returns, which is very important not only for the improvement of net income of the company, but also for environmental reasons since the returned food products are usually discarded. Consumers' demand for low quantities, high variety and a direct delivery to their homes with minimal costs has made the whole supply chain more sensitive to costs. This problem is an even bigger concern in last-mile delivery (e-commerce) which constitutes 28% of current total logistics costs due to uncertainty and disturbances (Lau, 2014). Currently, the food supply chain is inefficient because of abundant food waste and low nutrition level due to inability to cope with disruptions (Borrello et al., 2017; Managa et al., 2018). In order to properly cope with such possible disruptions as demand fluctuation, the collaboration of the whole supply chain is essential. The integration of the whole supply chain would also improve information sharing in other aspects not limited to inventory management. For example, improved forecasting accuracy along with higher collaboration could help in the improvement of both production processes and warehousing or product delivery (Gružauskas et al., 2019).

2.3.2 Out-of-Stock Detection using Computer Vision

In any retail business, it is very important that the customers are able to find the required items on shelf. At retail stores, out-of-stock shelves inevitably reduce sales and customers and cannot be considered a temporary loss. Out-of-stock (OOS) problem is a significant reason of the decline of goods sales in offline supermarkets since the frequent lack of goods on the shelves can reduce the enthusiasm of shoppers. For this purpose, it is necessary to effectively detect the OOS situation, which can ensure that the products are replenished in time (Chen et al., 2019). Traditional methods require labour-intensive human work dedicated to checking for products to refill raising the requirement of automatic solutions to detect OOS. Recent studies in this context have considered image sequences acquired with cameras mounted on shopping carts carried by retail's customers (Santarcangelo et al., 2018). In a retail scenario, the continuous monitoring and detection Out-Of-Stock (OOS) of products is a key factor to better manage the spaces and to improve stores profits. In the last decade, several strategies for automatic OOS detection in retail environments have been proposed. The most recent works in the field proved that deep learning strategies outperform previous approaches on solving this task Some approaches were quite simple and employed physical sensors or analytical observations Convolutional Neural Network (CNN) can be used to predict attention maps useful to find OOS in retail areas and hence suggest the retail employers where to intervene. There is opportunity to improve upon existing approaches by employing different architectures as well as using depth information in order to achieve better performances (Allegra et al, 2021).

2.3.3 Real-time Overstated Inventory Correction

The dynamics of inventory record inaccuracy (IRI) remain under-explored in a multi-channel distribution center (MCDC) setting. Additionally, how employees interact with the rest of an MCDC's inventory system to impact IRI is yet to be explored. There are substantial levels of IRI on a daily basis, with varying levels of consistency and a bias toward negative errors, raising concerns as to the suitability and stability of MCDC inventory systems. Without the real-time perspective, such evidence would otherwise have remained unobserved (Barratt, 2018). A product can be OOS because of either (i) it is not present in the store warehouse or (ii) it is present in the store warehouse but the shelf is not been refilled yet. In the former case the shortage of the product for the replenishment is due to external logistic reasons (e.g., mistakes during the orders, delay of the suppliers, out-of-date of perishable goods, etc). In the latter case the OOS depends on internal logistic issues that involve employees, store manager and refill planning policy. Internal logistic issues mainly face short-term decisions as warehouse management, replenishment organizations and costs, and staff-related issues . Refilling costs depend largely from frequency and management of the shelf replenishment operations, store layout and personnel costs. The employees are usually a scarce resource and the refill task is not scheduled with high priority in the task list. Under these conditions the optimal store managing is a very dynamic context in which routing and scheduling aspects play an important role. (Frontoni et al, 2017). Inventory accuracy is one of the keys to an efficient and effective supply chain, yet is often referred to as the 'missing link' in retail execution. Forecasting, ordering, and replenishment use inventory records as input, and the quality of these functions is impacted by inventory accuracy. To study the impact of RFID on inventory accuracy, a study was conducted to examine the store-level influence of RFID on perpetual inventory, which indicates that RFID does reduce inventory inaccuracy in the presence of normal business processes for on-hand adjustments (Hardgrave, 2009). There are challenges related to availability of ground truth information to corroborate if any OOS event really indicates an overstated PI or if the items are present elsewhere in the stores. Manual review by store associates would be provide limited opportunity, hence it is important to identify the most crucial out events to validate, to build any Machine Learning solution to solve the overstated PI problem.

2.3.4 Causal Discovery models for Waste & Prescriptive Models

In a consumer society, the retail sector contributes significantly to waste production. Supermarkets play a central role in the challenges of resources efficiency and waste prevention. Comparison of different waste treatments highlights the importance of recycling, particularly in the context of the circular economy. However, despite the importance of the topic, the academic studies are still scarce (Marrucci et al, 2020). Worldwide, demand for food is growing. The demand for dairy and meat products is especially expected to increase. This will lead to a rise in demand for food production, from 60% to 110% by 2050 (estimations from 2011 to 2012) (Garnett, 2013), and a need for an expansion of global food production. In the meantime, food is lost and wasted all over the world (Richter and Bokelmann, 2016) while undernutrition persists in developing countries (Pingali et al., 2017). The identified barriers preventing waste reduction include a lack of data, an unclear definition of waste, a lack of public awareness, undervaluation of the hidden costs of unsustainability, and global negative ecological impacts non-representative of the economic impact of each stakeholder (Lemaire, 2019). One major problem associated with food waste is that its costs are often undervalued and underreported (Binyon, 2007) and hence they remain “hidden”. Raising awareness of these “hidden” costs could be a catalyst for resolving the problem as businesses will start to realize the scale of the predicament and its impact on the bottom line. In an industry which is traditionally known for low margins, effective waste management critical to increase profitability levels of the chain members, especially by reductions in energy and raw materials usage and improvements in recycling and re-usage activities (Hyde et al., 2001). It is crucial to investigate the root causes of waste in the supplier/retailer interface. Waste at this stage has a significant impact because products have already gone through most of their value adding activities, accumulating costs and embedded energy. Analysis into the causes of waste revealed two important issues, that many causes are common across products and that some of the causes are not the result of management practices or human decision, such as short shelf-life and weather fluctuations. The analysis also showed that most causes have interdependencies, but that they are part of a complex web of interdependent causes and effects (Mena et al, 2011). Hence it becomes very important to find the causal relationship for waste and use them to build prescriptive recommendations for waste reduction. There is a lack of research on use of causal discovery models to identify the true root causes of waste which will be a core focus of our subsequent research.

2.3.5 Return Reduction based on Text Analysis and Supervised Models

Owing to the generous refund policy ubiquitous among today's retailers, returns are prevalent and expensive to manage. Implementing buyer assistance programs and adjusting return time windows are two common methods that retailers use to actively influence return rates. A buyer assistance program can improve consumer understanding and reduce returns, but to be effective it must be focused on the products that are most likely to be returned. Managing return rates by adjusting return time windows, on the other hand, requires an understanding of the time dimension of returns, a topic for which there is scant guidance in the literature (Shang, 2019). Customer product returns are key cost drivers that eat into online retailers' profits. Management research has neglected to examine ways of reducing return rates without causing a concomitant decrease in sales. Studies have shown important contingencies of the causal relationship by considering three variables: purchase frequency; retailer type; and customer gender. Overall findings indicate that an online retailer's reputation is a powerful means of reducing product return rates (Walsh, 2016). Text-mining based analysis of evaluation comments indicate that store layout and facilities, as well as product availability and waiting time had a great impact on consumer satisfaction. There is a direct relationship between consumer's perceptions and attitude regarding a retail store's service quality and their satisfaction level. The usage of big data in the form of User generated Content and online reviews offers huge potential for analysing consumer satisfaction and reducing returns. (Brandtner, 2021). There has been very little past research on using Machine Learning and Text analysis to reduce the overall returns for retail business.

2.3.6 Recycling as a Service using Computer Vision

Sustainability issues are particularly sensitive to the fashion supply chain, given current fierce competition, intensive resource use, and the exposure of penurious labour conditions in some regions. Since Brudtland's report in 1987, sustainable development (SD) and sustainability have progressively been incorporated in governmental policy and corporate strategy. Defined then as aiming to meet "the needs of the present generation without compromising … future generations," it became the basic framework of United Nation's (UN) Agenda 21. (De Brito, 2008). In principle, the recycling of packaging waste should reduce the consumption of raw materials. Moreover, the resulting decrease in waste disposal is likely to increase the lifespan of sanitary landfills (Fullerton and Kinnaman, 1995). Recycling also generates additional financial costs either for the private (e.g. the industry) or public (e.g. waste management operators) sector stakeholders (Massarutto et al., 2011). These extra-costs are often translated into higher prices for goods or additional waste management tariffs or taxes. Moreover, recyclables and, more specifically, sorted packaging waste can have a rather low market value, sometimes even negative. This is particularly true when raw materials are inexpensive. Among other aspects, the net economic sustainability of the recycling of packaging waste is therefore connected with the type of packaging material recycled. The logistics chain of recycling is usually quite complex. To set up an effective system requires high up-front costs (investments in new infrastructure for selective collection and sorting of packaging waste) and additional transport costs. Refuse collection can have direct links between drop-off containers and landfills, but the separated waste must first be transported to the sorting facility then to the storing or recycling facilities (da Cruz et al., 2012). Hence it is necessary to build a system that would serve as a tool to help humans recycle properly in order to reduce pollution and waste. The system should be able to classify waste objects into 4 separate categories: glass, metal, plastic and paper with high accuracy in a short period of time. The number of photos and variation of photos being fed into the network is a very important aspect of convolutional neural networks. The more photos and variation in them, the higher the success rates will be (Tomaselli, 2019) .Such a recycling system can be provided as a service to customers by the online retailers to help in reduction of waste and improve sustainability.

2.4 Summary of the Literature Review

Through the literature review we can conclude that retail total loss is a varied field of study ranging from fraud prevention, waste management, return reduction, inventory management and many others. In all of these applications, machine learning already plays a very significant role, but in omni-channel retail ecosystem, there is a significant gap in terms of application of prescriptive ML models and causal discovery models. Another significant area of research is to solve the problems of getting good quality labelled data for the supervised learning problems in fraud detection or overstated inventory correction. Reduction of false positives in fraud prevention remains a persistent unsolved problem. More research is required in terms of efficient use of graph analytics for collusion discovery and linkage of different entities. Computer vision solutions can be used for detection of out-of-stock in retail stores and to correctly classify the recyclable materials which can be offered as a service to customers by the retailers.

Through the literature review we identified that retail total loss is a varied field of study ranging from fraud prevention, waste management, return reduction, inventory management and many others. In all of these applications, machine learning already plays a very significant role, but in omni-channel retail ecosystem, there is a significant gap in terms of application of prescriptive ML models, models in absence of good quality labelled data, reduction of false positives, etc.

The following research gaps were identified –

- Major issue lies in having access to good quality labelled data for ML models, and no effective ways of handling false positives. No effective holistic ML techniques were available in omni-channel fraud.
- Prescriptive ML models to provide efficient and actionable recommendations to store employees are not available. No clear technique exist to measure the impact and effectiveness of waste reduction strategies.
- Challenges exist for early detection and prediction before out-of-stock(OOS) scenarios, availability of ground truth information to corroborate if any OOS event really indicates an overstated inventory issue.
- Existing research doesn't address challenges of short shelf life, manual adjustments, non-adherence to forecasts in production, rolling demand across shifts.

CHAPTER III:
METHODOLOGY

3.1 Overview of the Research Problem

The long-term goal of the research is to develop a standardized intelligent system for any omni-channel retailer which can minimize the loss due to wastage and reduce fraudulent loss. Data-driven AI products start with gathering all required data from different sources, transforming, and aggregating them using Big Data ecosystem like Spark, and creating intelligent and actionable insights and recommendations that the business users can easily consume to improve their efficiency of business processes. The objective of the current study is to create a comprehensive AI powered system of intelligent recommendations in relation to omni-channel retail loss prevention and fraud prevention. Particularly, the research has the following sub-objectives:

1. To provide a comprehensive review of advanced ML solutions in online and offline fraud prevention and identifying the gaps in omni-channel retail.
2. To develop a semi-supervised ML solution for fraud prevention along with collusion detection and linkage analysis.
3. To identify root-cause analysis of store waste and Out-of-stock based inventory adjustments.

4. To outline a conceptual framework for recommendations to store associates and impact measurement for waste management.

The result of this study will be valuable to the retail industry practitioners as well as Total loss prevention software developers in developing better tools for fraud prevention and waste management in omni-channel retail ecosystem. This system once implemented can lead to multi-million-dollar savings in retail loss.

3.3 Research Purpose and Questions

The research attempts to answer the following questions:

1. How Machine Learning can help in eliminating gaps between offline and online channels which lead to fraudulent activities?
2. What are the major causes of Retail Waste and how prescriptive Machine Learning recommendations can help to reduce waste?
3. How can early Detection of Out-of-Stock using Computer Vision help in real-time Inventory management to reduce waste due to overstocking?
4. How can better demand planning for Fresh production help in reducing wastage due to over-production?

3.4 Research Design

The primary research method for this study is literature review to understand the state-of-the-art systems in this field, gathering feedback from experts through interviews and tools, followed by conceptual modeling and evaluation of the models on real world. The research methodology for this study was primarily Quantitative Research, which was further aided by literature review of systems in the field. This was supplemented with expert feedback through interviews and tools, conceptual modelling and evaluation of these models in real-world scenarios. The research methodology for a study on Retail Fraud Prevention and Waste Reduction included a literature review, expert interviews, conceptual modelling, statistical hypothesis testing, and the use of machine learning models. The researchers formulated hypotheses about potential fraud indicators and tested these using past known fraud cases. They used both supervised and unsupervised machine learning models to identify and predict fraudulent activities. To overcome the challenge of insufficient labelled data, they proposed a reinforcement and active learning-based recommendation system. For waste reduction, the researchers used a diagnostic approach to identify major sources of waste and applied a regression-based framework and causal discovery model to pinpoint the root causes. They also used forecasting to recommend exact production quantities to minimize waste and computer vision to assess product quality and freshness. To improve search ranking algorithms, prescriptive analytics were used. Optimization techniques were also employed to identify the optimal path once items were returned.

A combination of Experimental and Case Study Research Design was adopted to study the impact of specific ML model techniques and interventions that retailers can apply to reduce overall loss across multiple dimensions of fraud, waste and shrink. This involved testing different ML models or algorithms, comparing their performance, and evaluating their effectiveness in identifying and reducing fraud or waste .

The first step towards building an omni-channel fraud prevention system is to understand the modus operandi of the fraud. These can be formulated as hypothesis, for example, we want to study if the number of cards used, number of address changes, time of the day, refund value etc. are related to fraud. We have performed statistical testing of hypothesis to validate based on past cases of known fraud. Next, we started with unsupervised modelling techniques like anomaly detection. Usual anomaly detection techniques rely on the underlying distribution being Gaussian Normal, however this is hardly the case with rare event scenarios like fraud or waste. So we have proposed novel techniques of identifying the underlying distributions and then finding the outliers as potentially risky cases.

Another unsupervised approach is to hypothesize a theoretical best candidate which is completely non-fraudulent and create features which are monotonically increasing with the risk, then find appropriate distance measures like Mahalanobis distance to create an unsupervised risk score. Another approach is to try Supervised ML models like classification, which is only possible if we have good quality labelled data. This is an adaptive ML model where the parameters and weights are learnt over the period of time, and gradually the focus shifts from unsupervised to supervised models. We have used tree-based models like Gradient Boosting and Neural Network approaches to learn from past cases of fraud. However, supervised models alone cannot detect new types of fraud, hence it needs to be applied together with the unsupervised models to create a combined risk model. Next we looked at potential collusion of customers with different agents like drivers, store associates, sellers, etc. using a Graph based approach like GNN and entropy based measures to create a collusion detection algorithm.

Customers may also create multiple identities to avoid detection, hence we formulated a linkage solution based on Graph based approaches like community detection and priority based connected components. In 3P ecosystem, one interesting type of fraud is sales breakout where sellers have a spike is sales followed by huge refunds and can leave the system before recovery, such systems can be detected based on a time dependent classification model.

False positives are usually very dangerous in these scenarios, and one approach can be to add additional data points from omni-channel customers' successful store purchases to improve the false positives in an ecommerce payment fraud system. One major challenge for any supervised models in such rare event scenarios is the lack of proper labelled data, so we propose a reinforcement learning and active learning-based recommendation system, which continuously learns based on the previous feedback, and provides better recommendation through explore and exploit mechanism.

In return reduction, we identified root causes of return and used prescriptive analytics to improve the search ranking algorithm to avoid low quality items and sellers, improve the arrival time prediction. We also applied optimization techniques to identify optimal disposition path after items are returned. We also improved the availability of items by identifying Out of Stock scenarios using computer vision, and in case the item is misclassified as out-of-stock, since it may be in other store locations, we can use a real-time classification system for inventory adjustments.

In Waste management, the primary approach is to perform a diagnostic approach to understand which stores or items are majorly driving huge waste and what are the types like throw away, donation, dispose, loss due to markdown, loss due to improper placement, out-of-stock, etc. Then we used a regression-based framework and a Causal discovery model to identify the true root causes, then provide appropriate recommendations to reduce the waste. In Fresh items, one of the major causes of waste is overproduction, where large quantities of items like bakery or meat is being produced in sales, and if they are not sold off it leads to throwaways. One approach is to use a forecasting approach to recommend the exact quantities that need to be produced and minimize the waste. Another approach is to use Computer Vision(CV) to detect the freshness of produce and vegetables and appropriately provide markdowns. Such CV techniques could also be used to detect the quality of returned items to determine if they can be resold. Another interesting approach is to provide recycling services to the customers, where they can provide the pictures of the items to donate or dispose and we use image processing to determine the quality and quantity of such recycling activity and also provide some incentive to customers. Additional techniques to make the retail ecosystem more sustainable is to detect early the electricity consumption of various machines in stores through forecasting and regression models and identify the defective ones to reduce carbon footprint.

3.5 Population and Sample

The data used in this research is from top omni-channel retailers like Walmart, Target, Amazon, Costco, and others from multiple countries like United States, Canada, Mexico, United Kingdom, Central America, Chile and China, and it was collected from April 2022 to March 2024. The data points on fraudulent transactions were collected either from the retailers directly or 3rd party sources, 3rd party fraud gateways tied to retailers like ThreatMetrix , LexisNexis, Accertify, Signifyd, and others.

For data gathering on past fraudulent cases or the waste accrued in stores, we relied heavily on Asset Protection and Total Loss analysts from individual stores across different countries, as well as central teams who view the complete picture across the country and ecommerce orders. The primary Subject Matter Experts (SMEs) were around 100 loss prevention analysts from across UK, US, Canada, Chile, Mexico and Central American countries. In addition to gathering the historical data on past cases, they also helped in validation of the model results using several different survey tools and dashboards to provide their feedback. The store images for identifying Out-of-Stock scenarios were collected from on-shelf cameras provided by Focal, and automated shelf-scanning robots provided by Brain Corp. The recommendations around Waste reduction and the Out-of-stock validation were performed by 1000+ associates across different stores, and recommendations from our models were shared through Mobile applications on Associates devices, which allowed them to apply the recommendation and provide feedback.

3.6 Participant Selection

The following approach was taken to choose the associates who provide the input for research validation:

1. Stores within each country (like US, Canada, Mexico, Central America, Chile etc) are further clustered into different groups based on different features like store size, type, footfall, sales, returns, waste, neighboring population demographics (census), geolocation. There have been agreement and approval from the Asset protection team for gathering necessary input data.
2. Based on statistical techniques like silhouette width, optimal cluster size was chosen, say, 10 clusters for a country.
3. Within the country, a Stratified Random Sampling without replacement was performed to choose 40 pilot stores, (4 stores are chosen from each cluster), which covered around 10% of all stores.
4. Within each store, for each Department (around 10 Fresh departments and 10 non-Fresh departments) , 2 store associates were chosen at random, which accounted for around 40 associates per store, bringing the total to around 1600 associates.
5. Each such associate were provided daily top 3 personalized recommendations on mobile application, and based on their feedback and completion of actions, the model re-learnt and provided better recommendations. The associates marked an action as complete or provide feedback for no-action in the device.

Such activities of gathering data and feedback through tools and interviews have been performed from April 2022 to March 2024.

3.7. ML models for online and offline Fraud Prevention

This research aims to examine the influence of applications of Machine Learning (ML) based fraud prevention systems in the retail sector, especially on online, offline and hybrid models. It is based on feedback and validation from a sample of risk analysts who are subject matter experts on such decision making process. The primary objective of this study is to understand the benefits of Machine Learning in detecting and preventing fraudulent activities in retail, as well as the potential challenges. The study has adopted a mixed-method approach, integrating both quantitative and qualitative methods to ensure a very comprehensive analysis. The quantitative part of the study has involved conducting multiple surveys and feedback collection through intelligent tools, among risk analysts to statistically assess the impact of ML-based fraud prevention systems. The qualitative part involved conducting semi-structured interviews to gain in-depth insights into the experiences and perspectives of risk analysts regarding the use of ML in preventing fraud. The feedback from the qualitative interviews were further taken as enhancements to the approach to improve the ML algorithms as well as improve the quantitative approach.

The research design involved collection of data across multiple risk analysts, across a long period of time at a weekly cadence. The population for this study consisted of risk analysts who have experience or exposure to ML-based fraud prevention systems in the retail sector.

A purposive sampling method has been used to select 100 participants who meet the criteria. For the quantitative part of this study, a structured questionnaire embedded in an web application has been administered to the risk analysts. The questionnaire was based on the top risky recommendations from the model, along with justifications and had been designed to capture the perceived effectiveness, feedback on the riskiness and reasons, actions required, benefits, and challenges of the different fraud prevention algorithms.

The qualitative part involved conducting semi-structured interviews with a subset of the respondents. The interviews have been designed to capture in-depth insights on specific instances where ML has aided in fraud detection and prevention, any issues with the data or accessibility of information, as well as any challenges encountered. The study has been guided by the Technology Acceptance Model (TAM) to understand the adoption of ML in fraud prevention. This model suggests that perceived usefulness and perceived ease of use influence the acceptance of a new technology. In this case, the TAM can help to understand if risk analysts find ML-based fraud prevention systems useful and easy to use. Quantitative data has been analysed using statistical software. Descriptive statistics has been used to summarize the data, and inferential statistics have been used to test the hypotheses developed for the study. Qualitative data from the interviews have been transcribed and analysed using thematic analysis. To ensure reliability and validity, the study has adhered to strict data collection procedures and used validated instruments. The research instrument of the Web UI powered by ML driven recommendations has been pilot-tested to ensure they capture the intended data accurately. Moreover, the findings from the quantitative data have been triangulated with the qualitative data to validate the results. All participants have been informed about the purpose of the study, and their consent has been obtained before data collection. Anonymity and confidentiality has been maintained throughout the study. This study has helped to provide valuable insights into the impact of Machine Learning-based fraud prevention systems in the retail sector. The findings could guide future implementation and development of ML technology in fraud prevention. With the rapidly growing retail environment, the effective use of ML in fraud prevention is crucial. Thus, through this research, we aim to shed light on the effectiveness and challenges of ML-based fraud prevention systems, which can provide valuable insights for retailers, risk analysts, and policymakers.

The following sections highlight the different ML techniques used in the study.

3.7.1. Iterative robust parametric anomaly detection

Anomaly Detection is the identification of rare items, events or observations which raise suspicions by differing significantly from most of the data. Majority of anomaly detection techniques rely on the underlying data to follow a Normal Distribution. For our use case, we want to identify unusually high cases of (i.e. anomalies in) Refund Amount, Refund Frequency, and Refund Rate.

Now, based on the data distribution, we have arrived at the following hypothesis:

Refund Amount **~ Exponential(λ),** since only a small proportion of all customers will refund a high amount.

***Refund Frequency* ~ Poisson(λ):** Refund frequency is a discrete event and with similar analogy to refund rate, this can be modelled as a rare event problem.

***Refund Rate* ~ Beta(α,β):** Refund Rate, calculated as refund amount as proportion to purchased amount, is a proportion taking values between 0 and 1, and the shape of the distribution varies rapidly and is very well captured by a Beta distribution with varying parameters.

Once the parameters have been identified, next steps involve robust parameter estimation and then determination of thresholds for outlier detection.

***Positive Refund Amount*:** For an Exponential Distribution, the Maximum Likelihood Estimator of λ is given by $\hat{\boldsymbol{\lambda}} = \frac{\mathbf{1}}{\bar{\boldsymbol{x}}}$, where x_i represents the refund amount for the i^{th} customer, we take the trimmed mean of the refund amounts. Once we have estimated the parameters of the model, we choose the threshold for anomaly detection based on Tukey's criteria for Exponential Family of Distributions as

$$\boldsymbol{Q3 + 1.5\ IQR = \frac{\ln(4)}{\hat{\lambda}} + 1.5\frac{\ln(3)}{\hat{\lambda}}}$$

***Refund Frequency*:** For a Poisson Distribution, the Maximum Likelihood Estimator of λ is given by $\hat{\boldsymbol{\lambda}} = \bar{\boldsymbol{x}}$, where x_i represents the refund frequency for the i^{th} customer. The threshold for anomaly detection is given by the chosen percentile values of the fitted distribution, e.g., 95^{th} percentile.

***Refund Rate*:** For Beta Distribution parameter estimation, we use the method of moments estimators, given by:

$$\hat{\boldsymbol{\alpha}} = \bar{\boldsymbol{x}}\left(\frac{\bar{\boldsymbol{x}}(\mathbf{1} - \bar{\boldsymbol{x}})}{\bar{\boldsymbol{v}}} - \mathbf{1}\right) \quad , \quad \boldsymbol{if} \quad \bar{\boldsymbol{v}} < \bar{\boldsymbol{x}}(\mathbf{1} - \bar{\boldsymbol{x}})$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{1} - \bar{\boldsymbol{x}})\left(\frac{\bar{\boldsymbol{x}}(\mathbf{1} - \bar{\boldsymbol{x}})}{\bar{\boldsymbol{v}}} - \mathbf{1}\right) \quad , \boldsymbol{if} \quad \bar{\boldsymbol{v}} < \bar{\boldsymbol{x}}(\mathbf{1} - \bar{\boldsymbol{x}})$$

For our case, we replace the estimates by Winsorized mean and Winsorized variances. After the parameter estimation, the threshold for anomaly detection is given by the chosen percentile values of the fitted distribution, e.g., 95^{th} percentile.

Firstly, we consider the refund rates, and find the thresholds for high risk in terms of refund rate. Next, customers are grouped by high-risk refund rates in groups of 5% intervals. Within each of the refund rate buckets, the high-risk customers are decided based on the variable threshold for refund amounts and refund frequencies. A stricter threshold is given towards the low refund rate buckets, and a relaxed threshold for the high buckets. Another implementation is finding the risk by grouping across geographical locations like cities or postcodes.

The benefits of this technique are to identify which customers, drivers, employees, stores, cities, areas are of greater risk, so that the Fraud Investigation team can take appropriate action against refund fraud committed by those customers and other entities.

3.7.2. Semi-supervised Customer Risk Scoring

The single metric anomaly detection described in previous section helps in identification of risky cases based on some chosen metrics, and data-driven thresholds. However, there may be some fraudulent customers who are missed out by these thresholds. Also, sense of prioritization is missed out in simple univariate outlier detection. Hence there is need for creating a holistic customer risk scores which considers a multitude of features and generates a Machine-Learning driven score. The first and most important part of said risk score model is creating customized features which can holistically capture all aspects of fraudulent behavior of the customers. We divide the features into four broad sets:

Usual Risk KPIs*:* These include refund amount/frequency, cancellations amount, refund rate, goods not returned refunds, refunds channel (e.g. through web or call center), recency of refunds.

Risk by Association*:* Share of high-refunding stores, cities, postcodes; refunds made in high value items and in high-risk stores.

Multi-party Collusion: Customer may collude with certain drivers and store employees, and hence associations of customers, drivers, and employees

Suspicious Behavior: Repeated refunds of same item; recent spike in refunds; doesn't return damaged Items or unwanted substitutes; refunds at a higher price

Risk score is adaptable multi-variate anomaly detection framework which can also handle other business relevant metrics of interest as well.

The following steps are involved in the risk scoring process:

Metric Aggregation: In the feature list some metrics are at different levels of granularity like customer-employee, store, postcode, item, etc. The first challenge is identifying the proper features which can represent the feature at a customer level.

Feature Selection & Cleaning: An essential task of any Machine Learning modeling exercise is to identify which features are important. In our case, we had over 60 features generated corresponding to the mentioned sets. As a first step, we remove the redundant variables by choosing to keep only one out of highly correlated feature sets to reduce multi-collinearity in the model. To bring parity to the features, they are normalized to lie in the range of (0,1).

Choice of Weights: The initial version of the customer risk score is weighted aggregate of metrics. Weights are based on the following:

Precision**:** The variation of the features will differ significantly. Some of the metrics will have a smooth transition from lowest to highest value across all the customers. Others will have values for only few customers. Hence the variation of the features across the customers will be a crucial in determining the importance of the feature. The inverse of the variance or the precision will be the first criteria for deciding the feature weight. This metric will be important to identify the feature as important even if it has not been seen as a causal measure for any of the previously identified fraudulent cases.

Suspension propensity Importance: In this section we discuss a supervised Machine-Learning technique, which helps us learn about the importance of features based on previously suspended customer accounts. Now only a miniscule proportion of all customers will have been previously suspended. This leads to a highly unbalanced dataset. The first step in such a case, is to generate synthetic data and perform techniques like over and under sampling, which balances out the bias in the data. Next the augmented data is split into training, validation, and test sets. The next step is to train an Extreme Gradient Boosting (XGBoost) model on this data. This model has been chosen out of several competing models because it has a more regularized model formalization to control over-fitting, which gives it better performance. Also, the model is easy to scale and run on a huge dataset. The hyper-parameters of the XGBoost model which needed to be tuned were the maximum number of iterations, learning rate for stochastic gradient descent, and the boosting method. These hyper-parameters are tuned based on the validation set. Now, in our case, because of the unbalanced data issue, one way of handling it is using the "Area Under the Precision-Recall Curve" as the metric of prediction accuracy. So, the model learns while trying to maximize the said area. Now this measure gives a better chance of capturing the true fraudulent customers as compared to misclassification error or AUC. Now the feature importance coming out of this model is defined by the information gain attributed to that feature. Since this is a binary classification problem, the decrease of Gini impurity is taken as the measure of information gain. For our purpose of risk scoring, we only take the variable importance part of the best fitted model, as the second part of weightage. This part will try to mimic the cases similar to earlier suspended accounts and those customers who behave in a similar fashion will be given a higher risk score. For our case, the best model yielded a score of 96% in terms of Area under Precision-Recall curve. The final risk score will be the combined average of the weighted average of the scaled features, with the inner weights coming from the precision and the ML model.

$$Initial\ Risk\ Score_i = \sum_{j=1}^{m} \big(\alpha w_{1j} + (1-\alpha) w_{2j}\big) f_{ij} ,$$

where f_{ij} is the jth scaled feature for the ith customer, w_{1j} is the weight of the jth feature based on precision and w_{2j} is the weight of the jth feature based on suspension propensity,

α is the importance weight, chosen according to the available proportion of previously suspended cases and the model accuracy.

Risk Buckets and Reasons: Once the risk scores have been identified for all the customers, the next step in grouping the scores into appropriate buckets of Very High, High, Medium and Low risk buckets. The thresholds for such buckets can be derived through change-point detection. One technique which has been used to transform the risk scores into more meaningful and interpretable is by looking into the percentile rank of the individual score and projecting the actual score into the same space as that of its percentile rank using estimated parameters from a fitted linear regression model. We also identified the reasons for the customer obtaining high risk score as this helps investigation to focus on relevant behavior instances. The technique used for this is finding the percentile rank of the customer for each feature across all the customers, then sorting the features based on said percentile ranks, thresholding features at a 3rd quartile, and choosing the top 5 features. These features are mapped to meaningful reasons, which can be easily interpreted by the Investigation personnel to take the appropriate action against the prioritized list of customers based on the risk score. The Risk score model will also learn from feedback provided by the user marking a customer as risky or non-risky. After such a model, the feature importance of this model can be fed back into the original model to improve the overall risk score. The model can be further enriched by the new features arising from Fraud signature discovery techniques discussed in the subsequent sections.

3.7.3. Linkage of Duplicate Customer Accounts

The fraudulent customers can create multiple duplicate accounts, and refund through such accounts to avoid being tagged as risky. To detect such cases, we have arrived at the solution of linking the duplicate accounts using the following approaches:

***Hierarchical Exact Linkage*:** Here we consider the following customer dimensions and in the mentioned order of hierarchy: Same "Doorstep & Zip code", Same Phone number, Same Mobile Number, Same Card/Payment Reference number, and finally Same "Zip code & name". Now for each of these features we create a hierarchical web of linkage. Starting with the first feature we link all customers having the same zip code and doorstep, in at least one of their multiple orders, into one group. Then among all the different phone numbers used by these set of customers we match from the remaining subset of customers, thus growing the group. In this manner with each subsequent feature of linkage, the group of linked accounts will keep growing. The final web of linked customers will be represented by one of the accounts as a primary linked account and the refunds and risk would be aggregated to this level. This will help in capturing other related accounts of the suspicious customers. This method becomes particularly useful in case one of the accounts in the web has been previously suspended, thereby increasing the risk for all the linked accounts in the set. The technique used here follows the same analogy of a hierarchical agglomerative clustering technique but based on nominal features.

Fuzzy String based Matching: Although the hierarchical exact linkage method is very useful and reliable in case of linking duplicate customer accounts, it may fail to identify cases where none of the mentioned customer features match exactly. Customers can change their email addresses very slightly by a few numbers of letters and create a new duplicate account. The first level of fuzzy string matching is applied on the customer email addresses based on Levenshtein distance, Jaro-Winkler and Hamming distance. The technique which fared the best was Levenshtein distance, which is the minimum number of single-character edits like insertions, deletions, or substitutions, that are required to change one word into the other. This is computed using the Wagner–Fischer algorithm, which is a dynamic programming algorithm. The next part is considering the Haversine distance of the different addresses given by these accounts which are matched in the first part. Given the latitudes and longitudes of these addresses, the Haversine distance is the distance measured by the arc joining these two points on a sphere. This will help us understand the cases where similar customer emails have been used to order to nearby locations, and identify such cases, even if there are no common attributes.

3.7.4. Graph based Collusion Detection

Collusion can exist among multiple parties like customers, delivery drivers, store employees, neighbors, etc. These relations can escalate to cases of Organized Retail Crime. Let us start looking into the case of collusion the customer and the delivery driver. One possible method is that the customer claims that product is damaged on transit or missing, or the product delivered is different from the ordered item or a substitute which is not wanted by the customer. In such cases, the driver and the customer may be colluding internally giving rise to the opportunity of fraudulent refunds. The scenario can be best represented using the following Bi-partite Graph as shown below in Figure 1:

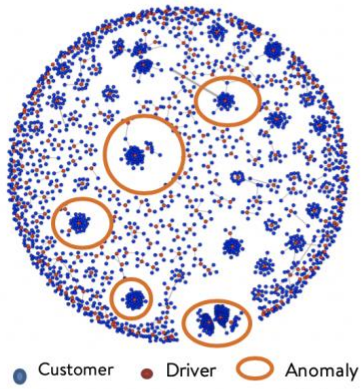


Figure 1: The Bi-partite Graph captures the relationship between Customers and Drivers

The Graph represents the two types of nodes as the Customers and Drivers, and the edge weights are indicative of the refund amount of their interaction. Now from this graph we are interested in the following metrics:

Nodes with High Degrees of connectivity: This metric captures either customers or drivers who interact with an unusual number of the opposite entity.

Edges with High Edge-Weight: This metric helps in identifying interactions between drivers and customers with unusually high refund amounts.

High Entropy Ratio: This metric is used to capture unusually disproportional behavior. The diversity of a vertex is defined as the scaled Shannon entropy of the weights of its incident edges:

$D_i = \frac{H_i}{\log(k_i)}$, where

$H_i = -\sum_{j=1}^{k_i} p_{i,j} \log(p_{i,j}), p_{i,j} = \frac{w_{i,j}}{\sum_{l=1}^{k_i} w_{i,l}}$

and k_i is the total degree of vertex i, $w_{i,j}$ is the weight of the edge(s) between vertices i and j.

This helps in capturing Drivers having high refunds with small number of customers; and Customers who are refunded by large number of drivers

Anomalous Structure in the Graph through community extraction: Community detection and extraction algorithms can be run on this bi-partite graph to determine regions or sub-graphs with anomalous behavior.

This graph can be extended to the cases of multi-party collusion to include collusion with store employees who are part of collection and refund as well as neighboring customers.

3.7.5. Fraud Signature Discovery

The purpose of fraud signature discovery is to identify how the suspicious customers were committing the fraud and finding their signature moves or *modus operandi*. This will help in creating a library of rules against which all future suspicious transactions will be flagged, and the security personnel will be alerted in real-time. We can explain the steps involved in this discovery process using an illustration of one fraud signature, namely geo-spatial outliers.

In this case, our initial hypothesis is the following: Customers for whom the distance between billing and shipping distance is unusually high or non-conforming to usual pattern are at a greater risk of high fraudulent refunds. To discover the rule for fraud signature we perform the following steps:

Identify the baseline thresholds for distance measures using Isolation forests: At first, from the raw distance metrics we create several derived metrics like Median Distance, IQR of Distances, Maximum Distance and Number of address changes. Next, we run an Isolation Forest algorithm to find the cut-offs based on multi-variate outlier detection. Here the anomaly score for outlier detection is defined as

$$S_{x,n} = 2^{-\frac{E(h(x))}{c(n)}}$$

where $h(x)$ is the path length of observation x, $c(n)$ is the average path length of unsuccessful search in a Binary Search Tree and n is the number of external nodes. A score close to 1 indicates anomalies and score much smaller than 0.5 indicates normal observations. Suppose for our case at 10 km distance we find that the score becomes close to 1, so we take the baseline threshold for median distance as 10km.

Hypothesis Testing: The customers are divided into two populations, one with median distance greater than 10 km (Population A), and the others (Population B). Now we test the following hypothesis:

$$H_0 : Refunds\ of\ both\ populations\ are\ same$$

$$H_1 : Refunds\ of\ Poplulation\ A > Refunds\ of\ Population\ B$$

Shapiro-Wilk Test and Q-Q plot easily reveal that the distribution of refunds is non-Normal. Hence, we move to non-parametric test of Equality of refunds of the two populations. In this case we have used One-sided Mann-Whitney U Test [6]. Since we are dealing with a large sample size, the test statistic U is given by:

$U = \min(U1, U2)$ = min(R1 - $\frac{n1(n1+1)}{2}$, R2 - $\frac{n2(n2+1)}{2}$),

where n1 is the sample size for sample 1, and R1 is the sum of the ranks in sample 1.

For large samples, U is approximately Normal with mean $\frac{n1n2}{2}$, and variance $\frac{n1n2(n1+n2+1)}{12}$.

We can claim that we reject the null Hypothesis at 95% level of significance if the p-value of the test is <0.05.

Iterative Hypothesis Testing and Uniformly Most Powerful Test: We perform the test iteratively over a range of values around the baseline threshold. Power is the probability of rejecting the null hypothesis when the alternative is true. For our case, we can create an empirical version of power through bootstrapping, by taking repeated samples with replacement from both the population.

$$\text{Empirical Power} = \frac{\#\ (\text{Sample Refunds of Pop B} > \text{Sample Refunds of Pop A})}{\#\ \text{Samples}}$$

We choose that value of threshold as final rule for which the H_0 is rejected, and the test is Uniformly most powerful.

The rule generated from this process is treated as a fraud signature and used for future real-time fraud prevention. Also, the features may be added into the Customer Risk scoring model.

3.7.6. Account Takeover Fraud detection and Prevention

As cyber threats continue to evolve in complexity, the retail industry has experienced a rise in Account Takeover (ATO) fraud. ATO fraud occurs when a fraudster gets unauthorized access to a genuine customer's account and makes fraudulent purchases. This criminal activity not only affects the financial stability of the business, but also significantly damages its reputation and customer trust. To combat ATO fraud, machine learning has emerged as a potent tool due to its ability to learn from data, identify patterns, and make decisions with minimal human intervention.

The first stage in using machine learning to detect and prevent ATO fraud involves the collection of relevant data. This data is used to train, validate, and test the machine learning models. The data can be obtained from various sources such as transaction details, payment method and masked payment ids, bank authentication, user behaviour logs, and historical fraud data.

Transaction records include details such as the amount of transaction, location, time, frequency, and the type of goods or services purchased. User behaviour logs capture the actions of users on the platform, such as login times, IP addresses, device fingerprints, and browsing patterns. Historical fraud data consists of previously identified instances of fraud and chargebacks which are vital for supervised learning models. The collected data is then pre-processed to remove any inconsistencies, missing values, and outliers.

The next step involves the selection of features, which are the variables that the machine learning model uses to make predictions. The chosen features should be relevant and contribute significantly to the detection of ATO fraud.

Important features for ATO fraud detection include:

1. Transactional features: The amount, frequency, and timing of transactions can be indicative of fraudulent activity. For example, a sudden increase in transaction frequency or amount might signal ATO fraud.

2. Behavioural features: These include login frequency, time spent on the platform, and browsing patterns. Unusual behaviour, such as logins at odd hours, frequent transactions, multiple failed authentication attempts can highlight potential fraud.

3. Contextual features: The device used, unique device fingerprints, IP address, geo-location, and other contextual information can also be used to detect ATO fraud. For instance, a sudden change in the device or location might indicate that an account has been compromised.

The performance of machine learning models in detecting ATO fraud is evaluated using various validation techniques. These techniques measure the accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC) of the models.

Machine Learning Methodologies for ATO Fraud Detection:

1. Supervised Learning: In supervised learning, the model is trained on a labelled dataset based on chargeback and historic normal and fraudulent transactions. The model learns to distinguish between the two based on the features. Various supervised learning algorithms such as logistic regression, decision trees, random forest, and support vector machines have

been used for ATO fraud detection. Each of these models has its unique strengths and weaknesses, and their effectiveness in ATO fraud detection varies.

Logistic regression is a statistical method that works on the principle of probability. It is often used for binary classification problems, like predicting whether a transaction is fraudulent or not. The algorithm calculates the probability of an event occurring and classifies the data based on this probability. The simplicity of logistic regression makes it a popular choice for many applications. However, it assumes a linear relationship between the input variables and the log-odds of the output variable, which may not always hold true.

Decision Trees, on the other hand, are non-parametric supervised learning methods used for classification and regression. They create a model of decisions based on actual values of attributes in the data. Decision trees are easy to understand and interpret but are prone to overfitting, especially when dealing with data that has many features.

Random Forest is a type of ensemble learning method, which combines multiple decision tree models to improve prediction accuracy. It operates by constructing numerous decision trees at training time and outputs the mode of the classes for classification or mean prediction for regression. This makes the Random Forest algorithm less prone to overfitting compared to single decision trees. However, the computation cost of Random Forest is significantly higher.

Support Vector Machines (SVM) are a set of supervised learning methods used for classification and regression. SVMs seek to find the hyperplane that maximizes the margin between classes in the feature space, which can be particularly effective in high-dimensional spaces. However, SVMs are not suitable for large datasets due to their high training time, and they don't directly provide probability estimates.

When comparing these approaches, logistic regression and SVM are more suitable for datasets with fewer features and instances due to their computational efficiency. However, they can underperform when dealing with complex datasets with numerous features and non-linear relationships. On the other hand, decision trees and random forests handle high-dimensional data and can model non-linear relationships better, making them more suited for complex ATO fraud detection tasks.

The choice of the algorithm should depend on the specific requirements of the ATO fraud detection task. If interpretability is crucial, decision trees would be a good fit as they provide clear rules for their decisions. If the task requires handling high-dimensional data and modelling complex relationships, random forests might be more suitable.

However, considering the complexity and evolving nature of ATO frauds, an ensemble approach combining multiple models' strengths might be the most effective solution. For instance, a combination of random forests (for their ability to handle complex data and provide robust predictions) and logistic regression (for its computational efficiency and interpretability) might work well. The ensemble model would balance the trade-off between prediction accuracy and computational efficiency, providing a robust and scalable solution for ATO fraud detection. In conclusion, while each of these supervised learning algorithms has its merits, a hybrid or ensemble approach that combines the strengths of multiple models could potentially offer the most effective solution for ATO fraud detection.

Supervised Machine Learning models can be trained to detect ATO fraud using a variety of features:

i. Login Frequency: Increases in login frequency can indicate suspicious activity.
ii. Password Reset Frequency: A high number of password resets can be a red flag.
iii. Change in User Behavior: Any significant deviation from a user's normal browsing or purchasing behavior can be indicative of ATO.
iv. Time Spent on Site: A sudden decrease in the average time spent on the site might be a sign of ATO.
v. IP Address: If the account is accessed from a new or suspicious IP address, it can indicate ATO.
vi. Device ID: Similar to IP address, a new or unusual device ID can indicate fraud.
vii. Number of Failed Logins: Multiple failed login attempts can indicate a brute force attack.
viii. Time of Login: Logins at unusual hours can be a sign of ATO.
ix. Account Activity: Unusual account activities like deleting or adding new addresses, changing personal details, etc., can indicate ATO.
x. Location: A login from a new or unusual location can be suspicious.
xi. Purchase Frequency: Sudden increases in purchase frequency can be indicative of ATO.

xii. Purchase Value: A sudden increase in the average purchase value can be a red flag.

xiii. Type of Purchased Products: Purchasing items that the user has never shown interest in before can be suspicious.

xiv. Number of Items per Purchase: A sudden increase in the number of items per purchase can indicate ATO.

xv. Payment Method: The use of a new or unusual payment method can be a red flag.

xvi. Shipping Address: A new or unusual shipping address can indicate fraud.

xvii. Email Activity: Changes in email activity, such as opening rates, can indicate that a fraudster has taken over the account.

xviii. Account Age: New accounts are often more susceptible to ATO.

xix. Login Sequence: Unusual or complex login sequences might be a sign of ATO.

xx. Browser Type: A sudden switch in browser type can be suspicious.

xxi. Operating System: Similar to browser type, a new or unusual operating system can indicate ATO.

xxii. VPN Usage: Use of VPNs can also be a red flag for ATO.

xxiii. Social Media Activity: Changes in associated social media account activity can indicate ATO.

xxiv. Customer Service Interaction: Increase in customer service interactions can be a sign of ATO.

xxv. Cart Abandonment Rate: A sudden increase in cart abandonment rate can indicate ATO.

These features can help a supervised ML model identify patterns and anomalies that are indicative of ATO fraud. The model can be trained to recognize these patterns and predict the likelihood of an account being compromised, allowing for timely detection and prevention of such fraud.

2. Unsupervised Learning: In unsupervised learning, the model is trained on an unlabelled dataset. It learns to identify patterns and anomalies in the data, which can indicate potential fraud. Clustering algorithms like K-means and hierarchical clustering, and outlier detection algorithms like Local Outlier Factor (LOF) are commonly used in unsupervised learning for ATO fraud detection. Unsupervised learning techniques are extremely useful for detecting Account Takeover (ATO) fraud, especially when there are no pre-defined labels or historic fraud patterns available. These methods involve training models on unlabelled data to identify patterns, anomalies, or clusters, which can be indicative of potential fraud. In the context of ATO fraud, unsupervised learning algorithms can analyse vast amounts of transactional data, user behaviour data, and network data. They can identify abnormal behaviour or transactions that deviate significantly from the norm, thus flagging potential ATO fraud incidents.

The types of data and features used in unsupervised learning for ATO fraud detection can vary widely, but some of the most common include:

i. Transactional Data: This includes details about purchases, such as the value, frequency, items purchased, and methods of payment.

ii. User Behaviour Data: This can include details about the user's browsing behaviour, such as the time spent on site, pages visited, items viewed, and interactions with the site.

iii. Network Data: This includes information about the user's network, such as IP addresses, device IDs, and locations.

iv. Temporal Data: This includes time-related data, such as the time of transactions, logins, and other account activities.

v. Account Data: This includes information about the account itself, such as the age of the account, the frequency of password resets, and the number of failed logins.

Clustering algorithms like K-means and hierarchical clustering are often used in unsupervised learning for ATO fraud detection. K-means clustering algorithm partitions the data into K distinct clusters based on the distance from the centroid of the cluster. It can help group similar user behaviours together and identify outliers that may indicate fraudulent behaviour.

Hierarchical clustering, on the other hand, creates a tree of clusters, which allows for a more nuanced understanding of the data. It can be useful for identifying sub-groups within the data that may represent different types of fraudulent behaviour.

In addition to clustering, outlier detection algorithms like Local Outlier Factor (LOF) can be used. LOF measures the local deviation of a given data point with respect to its neighbours. It can be used to identify data points that are significantly different from the rest of the data, which may indicate potential ATO fraud.

Validating the results of unsupervised learning models for ATO fraud detection can be challenging due to the lack of pre-defined labels or ground truth. However, there are several methods that can be used, including:

- Visual Inspection: This involves inspecting the clusters or outliers identified by the model to see if they make sense and align with known fraud patterns.
- Subject Matter Expert Review: This involves having a fraud expert review the results to provide feedback and validation.
- Performance on Labelled Data: If some labelled data is available, it can be used to evaluate the performance of the model. For instance, the model's clusters or outliers can be compared to the actual fraud labels to see how well they align.
- Anomaly Score Thresholds: This involves setting a threshold for the anomaly score, above which a data point is considered an outlier and potential fraud.

Unsupervised learning techniques provide a valuable tool for detecting ATO fraud, especially in situations where labelled data is scarce or non-existent. By identifying patterns, anomalies, and clusters in the data, these methods help detect potential fraud incidents and protect e-commerce customer accounts.

3. Deep Learning: In an era where digital transactions are becoming the norm, Account Takeover (ATO) fraud has emerged as a significant problem. Deep learning and graph-based techniques have shown promise in their ability to detect these fraudulent activities. These techniques leverage the power of Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Graph Neural Networks (GNNs) to analyse vast amounts of data and identify complex patterns that can indicate potential ATO fraud.

RNNs are a type of artificial neural network designed to recognize patterns in sequences of data, such as time series data. They are particularly effective in analysing temporal data and can capture patterns across different time steps. For instance, sudden changes in the frequency or amount of transactions from a particular account can be detected using RNNs.

CNNs are predominantly used in image and video processing but can also be employed to analyse structured data like transaction records. CNNs can process this data to identify spatial patterns and correlations that may not be identifiable using traditional analytical techniques. For instance, they can analyse patterns in customer behaviour data, such as their spending habits, transaction patterns, and more.

GNNs, on the other hand, are designed to work with data structured as graphs. They are particularly useful in detecting ATO fraud because they can model relationships between different entities, such as users, accounts, and transactions. By analysing these relationships, GNNs can identify suspicious patterns like unexpected links between accounts or abnormal transaction chains.

The types of data and features used in these deep learning models for ATO fraud detection typically include:

i. Transactional Data: This includes details about transactions, such as the time, amount, frequency, and type of purchases made.
ii. User Behaviour Data: This includes browsing behaviour, interaction with the site, and past transaction history.
iii. Temporal Data: This includes time-related data, such as the time of transactions, login, and other account activities.
iv. Account Data: This includes details about the account itself, such as the age of the account, number of password resets, number of failed logins, etc.
v. Network Data: This includes information about the user's network, such as IP addresses, device information, and geographic location.

Validation of the model results is a crucial step to ensure the accuracy and reliability of the model. This can be done through various methods, such as:

- Cross-Validation: This involves partitioning the data into subsets, training the model on one subset, and then testing it on another. This process is repeated multiple times with different partitions to ensure the model's robustness.
- Precision-Recall Analysis: This involves analysing the precision (the proportion of true positive results among all positive results) and recall (the proportion of true positive results among all actual positives). A model with high precision and recall is considered reliable.

- ROC Curve: The Receiver Operating Characteristic (ROC) curve plots the true positive rate against the false positive rate, providing a measure of the model's discrimination capability.
- AUC Score: The Area Under the ROC Curve (AUC) score provides a single measure of the model's performance. A score close to 1 indicates a high-performing model.

In summary, deep learning and graph-based techniques offer powerful tools for ATO fraud detection. By analysing diverse data types and features, these models can detect complex patterns and potential fraud indicators. Nonetheless, validation of these models is crucial to ensure their effectiveness and reliability in a real-world setting.

Machine learning can also aid in the prevention of ATO fraud through real-time monitoring and anomaly detection. By continuously monitoring user behaviour and transactions, machine learning models can detect any suspicious activity and flag it for further investigation. Additionally, these models can also predict the likelihood of an account being compromised based on the user's behaviour and transaction history, thereby allowing preventive measures to be taken.

A systematic approach to data collection, pre-processing, feature selection, and model validation is crucial for effective detection of ATOs. Furthermore, as fraudsters continually evolve their strategies, ongoing research and development in machine learning techniques are necessary to stay ahead.

3.7.7. Seller Fraud and collusion

The exponential growth of e-commerce retailing has given rise to the pressing issue is seller fraud and collusion. Sellers on e-commerce platforms engage in a broad spectrum of fraudulent activities, which can range from sales break-out fraud, money laundering, returns and cancellation abuse, quality issues, counterfeit products, and seller linked accounts through account details and catalogue. These frauds are not only detrimental to the customers but also the e-commerce platforms, hence necessitating the need for efficient detection and prevention systems. This paper explores the application of Machine Learning (ML) techniques in identifying and combating different types of seller fraud and collusion in e-commerce retail.

Sellers often engage in deceptive practices to manipulate the payment and sales systems of the e-commerce platforms. They may create fake orders or inflate sales volumes to win higher search rankings or to defraud the platform of advertising revenues.

e-commerce platforms are becoming an increasingly common channel for money laundering. Sellers may collude with buyers to engage in circular trading, where goods are bought and sold repeatedly to create the illusion of legitimate sales.

Some sellers abuse the return and cancellation policies of e-commerce platforms. They may refuse to honour legitimate return requests or cancel orders arbitrarily to avoid negative reviews or to manipulate their sales records.

Quality fraud involves selling substandard products, misrepresenting product features, or selling counterfeit goods. This type of fraud not only harms the customers but also tarnishes the reputation of the e-commerce platform and the legitimate manufacturers.

Some fraudulent sellers operate multiple linked accounts to engage in deceptive practices such as self-reviewing or to evade detection after being banned for fraudulent activities.

The first step to applying ML techniques is to collect a large amount of data. Data can be collected from various sources such as transaction records, customer reviews, seller profiles, and product catalogues. The collected data should then be pre-processed and cleaned to remove any outliers or irrelevant information.

Several ML algorithms can be applied to detect different types of seller fraud and collusion. For instance, supervised learning algorithms such as decision trees, random forests, and support vector machines can be trained on labelled data to classify transactions as fraudulent or legitimate. Unsupervised learning algorithms such as clustering can be used to detect unusual patterns or anomalies in the data. Deep learning techniques such as recurrent neural networks can be used to analyse sequential data such as customer reviews or transaction histories.

The performance of the ML models should be evaluated using appropriate metrics such as accuracy, precision, recall, or Area Under the Receiver Operating Characteristic (AUROC). The models should also be continually monitored and updated to adapt to changing fraud patterns.

The outputs of the ML models should be interpreted and actioned by fraud analysts and marketplace business users. For instance, if a transaction is classified as fraudulent, the system could automatically cancel the transaction or flag it for further investigation.

Sales breakout in the context of eCommerce refers to a sudden, unexpected surge in sales, often due to a particular season, holiday, or a marketing campaign. However, unscrupulous sellers may artificially create a sales breakout by manipulating transactions to appear more successful than they actually are. For example, a seller may create multiple fake accounts to buy their own products, creating a false impression of high demand. Once the sales break-out occurs, the sellers can get the payment from the retailer, and leave the system. This causes a huge burden to the retailers to pay back the refund amounts to the customers, especially in cases of spike in returns. Payments fraud by sellers typically involves manipulation of the payment processing system to divert funds to their own accounts. This could involve tactics such as overcharging for products, creating fake transactions, or using stolen credit card information. Sellers may also engage in 'triangulation fraud', where they use another person's credit card information to purchase goods on behalf of a legitimate customer, then pocket the difference in price.

Features to Use in ML Models to Capture Such Types of Fraud:

There are several features that can be used in machine learning (ML) models to capture these types of fraud:

i. Transaction Frequency: A sudden increase in the frequency of transactions could indicate a sales breakout or other fraudulent activity.

ii. Payment Method: The use of the same payment method for multiple transactions could suggest fraudulent activity.

iii. Price Fluctuation: Significant fluctuations in the price of a product could suggest that a seller is engaging in fraudulent activity.

iv. Customer Reviews: Negative customer reviews, or a lack of reviews, could suggest a fraudulent seller.

v. Product Return Rate: A high rate of product returns could suggest that a seller is not delivering the promised goods.

vi. Shipping Details: If the shipping address is frequently changing or if it is different from the billing address, it could be a sign of fraud.

vii. Account Activity: Unusual account activity, such as a new account making a large number of purchases, could indicate fraud.

viii. IP Address: If multiple transactions are coming from the same IP address, it could be a sign of a seller using multiple fake accounts to boost sales.

Money Laundering: This happens when a seller uses their online store to 'clean' illegally obtained money. They do this by setting up fake transactions that appear legitimate to the outside observer. Essentially, the seller purchases their own products with the dirty money and then withdraws the 'clean' money that has been deposited into their bank account.

Collusion: This is when two or more sellers work together to defraud customers or the ecommerce platform. They may coordinate to artificially inflate prices, manipulate reviews, or create fake transactions.

Returns and Cancellation Abuse: This occurs when a seller manipulates the return and cancellation policies of the ecommerce platform to their advantage. They might falsely claim that the customer returned an item, issue a refund, and then keep the item to sell again. Alternatively, they might cancel orders arbitrarily, causing inconvenience and financial loss to the customers.

Machine learning models can play a vital role in identifying and preventing such fraudulent activities. Here are some features that can be used in ML models:

a. Transaction patterns: Unusual buying patterns such as frequent high-value transactions, purchasing of the same product multiple times, or transactions at odd hours could indicate money laundering or collusion.

b. Seller-Buyer relationship: If a particular buyer is repeatedly purchasing from the same seller, it might be a sign of money laundering or collusion.

c. Return and Cancellation History: A high rate of returns or cancellations from a single seller could be a red flag for abuse of return and cancellation policies.

d. Product Pricing: Overpriced or underpriced products could indicate manipulation by sellers, potentially in a case of collusion.

e. Payment Methods: The use of the same payment method for multiple transactions might suggest fraudulent activity.

f. Customer Reviews: Negative reviews or a sudden influx of positive reviews could indicate fraudulent activity.

g. Shipping Details: Frequent changes in shipping address or multiple orders shipped to the same address could also be indicators of fraud.

h. Account Activity: Sudden changes in account activity, such as an increase in transactions, could be a sign of fraud.

By training machine learning models on these features, we could predict and identify fraudulent activities with a high degree of accuracy. These models learn from historical data and adapt to new patterns of fraud, helping to protect ecommerce platforms and their customers from potential financial loss.

Quality fraud occurs when the product delivered to the customer is of inferior quality than what was promised. Sellers might use misleading product descriptions or images to deceive customers. Some sellers may sell fake or counterfeit products that imitate well-known brands. These products are typically of lower quality and may be illegal, depending on the jurisdiction. Some fraudulent sellers may create multiple accounts to sell their products. They do this to avoid detection, especially if one of their accounts gets banned for fraudulent activities. This happens when a seller manipulates the catalogue listings to deceive customers. They might list an item under the wrong category, use misleading tags, or falsely claim that the product has certain features. Machine Learning Models can use several features to detect these types of fraud:

a. Product Reviews: Negative reviews or reviews mentioning inferior quality, fake products, or mismatch with product descriptions can help identify quality issues and counterfeit products.

b. Image Analysis: Machine learning models can analyse product images and compare them with the product description. Differences may indicate quality issues or counterfeit products.

c. Seller Information: Details like the seller's address, IP address, device ID, and other account information can be used to link multiple accounts to the same seller.

d. Price Analysis: If a branded product is listed at a price significantly lower than its usual market price, it might be a counterfeit product.

e. Category and Tag Analysis: Machine learning models can analyse the product category and tags to detect inconsistencies that might indicate catalogue fraud.

f. Seller's History: A seller with a history of negative reviews, frequent account creation, or previous fraudulent activities is more likely to engage in these types of fraud.

Sellers deliberately create false shipments to cheat customers in the following.

- Invalid Tracking information: Customers finds that the tracking information shared by the sellers is invalid
- Tracking Reuse: Some sellers reuse the past order tracking information multiple times

- Address Mismatch: Orders with a shipping address that are different from the delivery address

The seller may have received a lot of customer complaints. After reviewing the orders, it has been found that the tracking number is not valid, or the orders are being shipped to different addresses across the country. This leads to poor customer experience affecting the company reputation apart from generating a high refunds and cancellations.

We have utilized shipment data, which includes order-related information such as tracking numbers, shipment carriers, and customer complaints received per order.

To calculate the risk associated with orders, we created several key feature attributes to identify the risky orders and assess the overall risk of the sellers

- is non-compliant length: tracking nbr's length is different from the corresponding carriers standards
- is invalid tracking: tracking nbr contains gibberish
- is generic sequence: consecutive or repetitive sequence is present, like 012345, aaaa, abcde etc
- is non-compliant carrier format: the format matches with what carrier has specified, eg. ups starts with 1Z followed by 16 alphanumeric characters
- is same as purchase order: sales order nbr is used as tracking nbr
- is tracking nbr reused: multiple reuse of past order's tracking nbr
- lob complaints count: complaints recieved for an order from customers

- item risk score: item abuse factors
- seller risk score: sales breakout and money laundering aspect of sellers

It's worth noting that we have incorporated complaint information and the seller-item risk score in our analysis. Higher values for these factors indicate an elevated likelihood of a seller engaging in fraudulent activity during the fulfilment process.

Also, for now we have only considered customer complaints related to the fulfilment process in our analysis like 'Missing Item', 'Lost After Delivery (S2H)', 'Lost in Transit', 'Undeliverable', 'Damaged Order', 'Incorrect item received', 'Late to Arrive', 'Customer Refused Delivery', 'Delivered to Incorrect Address', 'Damaged Packaging', '3P Damaged'.

Seller fraud and collusion pose significant challenges to the growth and sustainability of e-commerce retailing. However, with the advent of ML techniques, it has become increasingly feasible to detect and prevent such fraudulent activities. Additionally, e-commerce platforms need to establish effective legal and policy frameworks to discourage seller fraud and collusion.

3.7.8. Online Payments, Accounts and Returns Transactions Fraud

The proliferation of e-commerce retailing has brought with it an unprecedented rise in online payment fraud and collusion, posing a significant challenge to businesses and consumers alike. Fraudsters are exploiting system vulnerabilities across a spectrum of payment methods, including credit and debit cards, PayPal, 'pay later' options, 'pay at store', and 'pay on delivery'. This paper explores different types of fraud and abuse in these channels, and presents machine learning techniques as a potential solution for identifying and mitigating such fraudulent activities.

i. Identity Theft: Fraudsters often use stolen credentials to impersonate genuine customers and make unauthorized transactions.
ii. Account Takeover: This involves gaining unauthorized access to a victim's e-commerce account to make fraudulent transactions.
iii. Friendly Fraud: Also known as chargeback fraud, this occurs when a customer makes an online shopping purchase with their own credit card, and then requests a chargeback from the issuing bank after receiving the purchased goods or services.
iv. Merchant Collusion: Dishonest merchants collude with customers or create fake customer accounts to make transactions and then claim a chargeback.
v. Triangulation Fraud: This involves using stolen credit card information to purchase goods for a customer who has purchased an item from a fake website.

Machine Learning (ML) offers a powerful toolset to detect and prevent these types of fraud. The application of ML algorithms can help in identifying patterns and anomalies that may indicate fraudulent activities.

Data collection is the first step in developing a machine learning model for fraud detection. Various data sources can be utilized, including transaction data, user behaviour data, and historical fraud data. Transaction data includes details such as transaction amount, payment method, and time of transaction. User behaviour data captures the user's navigation patterns on the e-commerce platform. Historical fraud data, on the other hand, includes past instances of identified fraud which can be used to train the model.

Several ML algorithms can be applied to detect fraud, each with its strengths and limitations.

1. Supervised Learning Algorithms: These algorithms, such as Logistic Regression and Decision Trees, are trained on a labelled dataset where fraudulent and non-fraudulent transactions are already identified.

2. Unsupervised Learning Algorithms: In scenarios where labelled data is scarce, unsupervised algorithms like Clustering can be used to identify unusual behaviour or outliers which may indicate fraud.

3. Deep Learning: Neural networks, a type of deep learning, are particularly effective in detecting complex fraud patterns by learning from large datasets.

Several features have been used in these ML models to detect transaction fraud –

a. Transaction Amount: The amount involved in the transaction can be used to detect fraud. Unusually high or low transaction amounts could be a sign of fraudulent activity.
b. Location of Transaction: The geographical location where the transaction is made can be an indicator of fraud. If the transaction location is different from the user's usual location, it might be fraudulent.
c. Time of Transaction: The time when the transaction is made can also be an indicator. Fraudulent transactions might occur at unusual times.

d. Frequency of Transactions: The frequency of transactions made by a user can be used to detect fraud. A sudden increase in transaction frequency might indicate fraudulent activity.
e. Type of Product Purchased: The type of product purchased can also be an indicator. Certain products might be more likely to be purchased fraudulently.
f. Payment Method: The payment method used in a transaction can be an indicator of fraud. Certain payment methods might be more likely to be used in fraudulent transactions.
g. User Behavior: The user's browsing and purchasing behavior can be used to detect fraud. For example, a user who makes a lot of purchases in a short amount of time might be a fraudster.
h. Device Information: Information about the device used for the transaction, such as IP address, device type, and operating system, can be indicators of fraud.
i. Customer Information: Information about the customer, such as age, gender, and previous transaction history, can be used to detect fraud.
j. Shipping Address: If the shipping address is different from the billing address, it might indicate a fraudulent transaction.
k. Account Age: Newly created accounts may have a higher likelihood of being fraudulent.
l. Number of Items Purchased: Buying a large number of the same item can be a sign of fraud.

m. Return and Chargebacks: Frequent returns or chargebacks could indicate a fraudulent user.

The performance of these models is evaluated using metrics such as accuracy, precision, recall, and F1-score. The models are also continuously monitored and updated as new data comes in to accommodate changing fraud patterns.

The output of the ML model is typically a fraud score that indicates the probability of a transaction being fraudulent. Fraud analysts then review these flagged transactions and decide whether to approve, decline, or flag them for further investigation.

While fraud and collusion pose significant threats to e-commerce retail, machine learning offers a promising avenue for combating these challenges. By adopting a data-driven approach, businesses can effectively identify and mitigate fraudulent activities, thereby enhancing security and trust in online transactions. However, the complexity of fraud detection necessitates continuous model training, evaluation, and adjustment to keep up with evolving fraud tactics.

3.7.9. Conclusions

The fraud detection engine is built using a combination of the different methods discussed above, including iterative robust parametric anomaly detection, semi-supervised customer risk scores, linking duplicate customer accounts, collusion detection and Fraud signature discovery. The research methodology also identifies the inter-relationship among the components and how they can enrich one another. The resultant ensemble solution can detect most types of fraudulent activities pertaining to retail refund fraud and can lead to significant savings in terms of loss prevention, and we continue to update our system to identify new types of fraud.

3.7.10. Active Learning Techniques for Sparse Labelled Data

Many of the classification-based problems we deal with are rare event scenarios. To start with we do not have enough labels (especially positive labels) to train our machine learning models. We are dependent on subject matter experts for labelling who have limited bandwidth for this activity. Usually an unsupervised anomaly detection model followed by semi-supervised learning can help to solve this issue, but it can be time-taking with limited labels. To deal with this, we have developed a generic framework called 'Concurrent Arm Beta Sampling based Reinforcement Learning Recommendation System for Unlabeled Data' to actively recommend the data points (to the subject matter experts) to be labelled and improve the hit rate (number of positive labels detected / number of recommended samples). This can be applied to any business problem having unlabeled data points like recommendation system, loss prevention or other similar problems. The framework built is unique in following ways: It can identify the explore and exploit buckets(samples) dynamically using feature space reduction and segmentation. The algorithm used generates a beta distribution for every explore and exploit bucket based on past rewards and dynamically assigns the explore and exploit sampling rate (recommended data points) for next iteration from each bucket based on concurrent arm beta sampling. Based on sampling constraints, a suitable size of recommendations are provided proportional to the sampling rate of the bucket. It also dynamically gives more weightage to rewards from recent iterations compared to old ones. The combination of these ML techniques can ensure faster learning rate by providing labels for appropriate data points.

3.7.10.1. Need for Active Learning Techniques for Data Generation

Recommendation systems are used commonly in various industries to provide meaningful, actionable, and business value driven recommendations towards desirable goal of personalization. In retail e-commerce business, recommendation systems are used for personalizing the shopping cart, personalizing the landing page of website, prioritizing customer service requests, etc. Typically, recommendation systems leverage some form of Machine Learning algorithm to learn preferences and patterns from past instances of known preferences (aka "labelled data") towards things to be recommended (aka "items") and generalize those to new instances of unknown preferences.

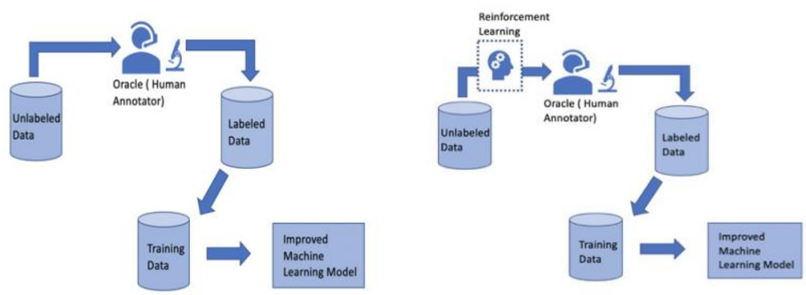


Figure 2: The Improvement over Usual system for Training a Supervised Machine Learning Model using recommendations.

In diagram above (Figure 2), role of human annotator may be played by different entities depending on the context. For instance, a customer who makes purchases can be considered annotator who is providing implicit preference for products in retail store. “Training data” refers to data which is fed into Machine Learning model based on characteristics of items to make Machine learn about distinguishing characteristics of preferred items and other items. A distinction is important that other items may not always be less desirable or non-preferred but are only items for which preferences are unknown. In Machine Learning, such problems are referred as those belonging to class called Positive-Unlabeled Learning [1], where some positive preferences are known, and remainder instances are unlabeled or of unknown preferences. One typical way such Machine Learning models are trained are to consider unlabeled preferences are negative preferences through set of algorithms belonging to Machine Learning category called “Classification Models”.

Once recommendations are algorithmically generated, human annotator is presented with them, and can “annotate” them by classifying them in preferred and non-preferred class. Machine-Learning model can then be re-trained and improved using new set of labelled data with Positive preferences, Negative preferences, and Unknown preferences. This cycle continuously improves quality and predictive capabilities of recommendation system.

Multiple Machine-Learning algorithms are available to train Classification Models, discussion of which is outside the scope of this innovation. However, all of them rely on sufficient availability of data in both classes. In cases where data is imbalanced and preferred class is very small minority (say, about 0.1% of observations) these models do not perform well, and accuracy of recommendations is low. This is because model learns from very limited observations and cannot truly understand human annotator's preferences resulting in high false positive rate (false positive instances are where system predicts annotator will prefer certain items, but they do not) which reduces quality of recommendation and human annotator's satisfaction from recommendations.

There is another challenge as well: Model learns from limited known preferences and recommends more items of same type continuing to exploit similar preferences but fails to explore other items which human annotator would have preferred had s/he gotten chance to use them. This stagnates the predictive performance of recommendation system and requires too many rounds of learning loop (recommendation –> annotation –> re-learn –> recommendation) to truly learn human annotator's preference.

This innovation proposes concurrent arm beta sampling approach[2] to active learning[3] to get out of this stagnation and improve learning significantly thereby improving performance of recommendations as seen in reduced false positive rate.

Novelty of this method lies in specific application and specific arrangement of sequence of steps required to use concurrent-arm beta sampling based active learning to improve performance of recommendation for cases with sparse labelled data.

3.7.10.2. Application of the algorithm in real world classification models

This algorithm is applied in context of identification of items with high repeated refund risk. Instances of known risky returns are extremely rare, with a positive rate of around 0.05% to 0.3% only.

Classification Model in use in this application is XGBoost which predicts likelihood that given item is potentially risky based on training on this imbalanced data, using 60 input predictors pertaining to refund rates, refund amounts, and refund reasons to generate the risk probabilities.

This application presents all the challenges described in the context of this innovation –

i. Known instances of risky returns are very sparse giving recommendation model little to learn from
ii. Recommending new potential risky items which are similar to known risky items leaves new patterns of risk completely undetected

Moreover, human annotator, ie. the investigators of this application have limited capacity to investigate new recommendations – due to obvious time and resource reasons – and hence should be presented recommendation which really benefit from efforts in improving the model significantly and rapidly.

Goal of this innovation is to recommend new set of items which could be potentially risky, which have not only low false positive rate but also explore other risk patterns not seen in known instances while limiting the number of items to be investigated.

3.7.10.3. Approach in Details

For each iteration of learning loop, three groups of items are available:

i. Positive labelled – known preferences
ii. Negative labelled – known non-preferences

iii. Unlabeled – unknown preferences, assumed to be non-preferences for Classification Models

Firstly, all these items are clustered into multiple similar segments based on similarity across item characteristics (e.g. refund rate, refund amount, transaction history etc, in context of our application, but can be a generic predictor set of any application) using off the shelf k-Means clustering algorithm [5] (or any other clustering technique). Rest of the process is to be followed independently for each of the segment. For each segment, two independent sub-groupings are undertaken:

First, sub-group items into multiple buckets based on defining characteristics likely to be associated with goal of recommendation. Say, we have n buckets.

Second, sub-group items based on pair wise Euclidean distance where dimensions are defined by characteristics of those items. Items closer to known Positive labelled items are sub-grouped into exploit set and rest in explore set.

As illustrated in Figure 3, for each segment of items, we have 2n categories where each bucket has both exploit and explore sets[6]

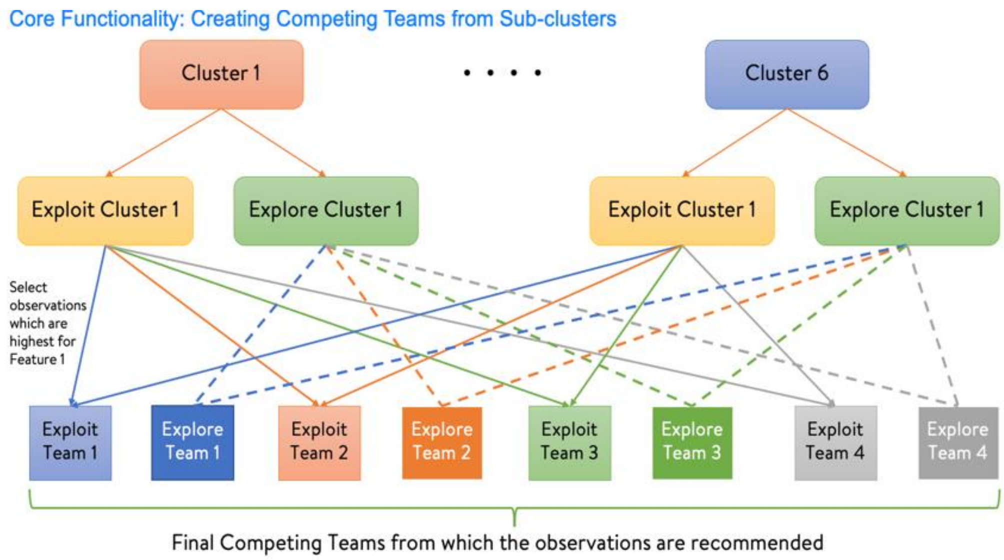


Figure 3: The process of creating Teams from the clusters

Rest of the approach describes how limited number of items are selected from each of the category to achieve conflicting goals described below.

i. Reward, a number between 0 and 1, is metric associated with each category which defines how useful this category is for providing valuable recommendations which satisfy human annotator and has fewer false positives.

ii. Sampling Rate, a number between 0 and 1, is metric associated with each category which defines what fractions of items in that category should be recommended.

iii. Recommendation Cap is total number of items to be recommended to not to overwhelm human annotator's time and resources.

In general, having high sampling rate for categories with high rewards will achieve higher recommendation performance and reduced false positive rate at the cost of exploration of other categories which may have higher reward should they be recommended. This is tradeoff between exploiting what is known, but perhaps settle for sub-optimal outcome, and exploring what is unknown to potentially seek greater optimal outcome.

Initial Sampling Rate for each category is defined equally, i.e., 1/2n, which is 1/8 th in our application with 8 categories.

Initial Reward Rate is defined as S/R, where S is number of known positive preferences in the category, and R is number of total items in the category, i.e., proportion of positive preferences in the category.

Now, number of items are sampled from each category such that total number equals Recommendation Cap (say, N). We use **Beta Distribution** to select the Sampling Rate. Beta Distribution[7] is a parametric distribution which needs 2 parameters (inputs). These 2 parameters are called alpha (α) and beta (β). When $\alpha = \beta = 5$, then there is equal chance of choosing a lower and higher sampling rate from the range 0 to 1. We want this for categories whose rewards are neither lower nor higher in the recent past.

When $\alpha \ll \beta$, i.e. beta is much higher than alpha, there is higher chance of choosing a lower sampling rate. We need this for categories having lower average reward in the recent past. When $\alpha \gg \beta$, i.e. alpha is much higher than beta, there is higher chance of choosing a higher sampling rate. We need this for categories having higher average reward in the recent past.

Starting with fixed sampling rate in first iteration, alpha and beta are updated at each iteration, and then an observation is sampled from corresponding Beta Distribution which becomes our sampling rate for next iteration of learning loop.

Parameter estimation for the Beta Distribution:

- $\hat{\alpha}$ = 1 + (mean percentage rewards for last few iterations * 10)
- $\hat{\beta}$ = 1 + (1 - mean percentage rewards for last few iterations) * 10

where mean percentage rewards are % of items labelled Positive by the human annotator from all items recommended, and 10 is used as a scaling factor which was chosen based on experiments.

Then, Sampling Rate for category $i = s_i =$ random sample from Beta(α, β)

Number of items to be recommended from category i = $n_i = N * s_i / \sum_i s_i$

This approach is called **concurrent-arm Beta sampling** since while it favors higher sampling rate for category with higher reward but does not strictly enforce that and still leaves room for uncertainty helping us explore other categories too. Even within a category, higher sampling rate only increases proportion of items selected from that category but still leaves room for uncertainty around exactly which items are recommended helping us explore further within a category as well.

In context of our application, one iteration of sampling may look something like as shown below:

Category	Team 1 Exploit	Team 2 Exploit	Team 3 Exploit	Team 1 Explore	Team 2 Explore	Team 3 Explore

Random Beta Sampling Output	0.6	0.2	0.7	0.9	0.4	0.3

Table 1: Table showing the sampling rates from the Beta Sampling model.

For Recommendation Cap of $N = 15$, number of items to be recommended from category 'Team 1 Exploit' is

$$15 * \frac{0.6}{0.6 + 0.2 + 0.7 + 0.9 + 0.4 + 0.3} = 2.9 \sim \text{about } 3$$

Results on simulated data prove that sampling rates from concurrent-arm Beta sampling balance both exploration and exploitation and converge to stable number of fewer iterations of learning loop making Classification Model improve faster and reach peak performance. This is equivalent of putting an Active Learning engine before recommending to human annotator.

The implementation of this solution and experimentation results show around 25% to 30% potential incremental lift in hit-rate for the detection of risky items by the supervised model using the active learning-based recommendations in a very short time.

3.8. Retail Waste Reduction using Prescriptive Machine Learning

Sustainability and waste reduction is a major focus for retail. However, there is no clear guidance or recommendations on efficient waste management. Thus, subjective decisions taken by store associates may not be optimal. Waste management remains a global challenge for retailers with huge chunks of items going to waste, especially from fresh departments. Among the unavoidable waste, we want to increase the recovery through optimal markdown sales using Customer Value Proposition (CVP). The objective is to reduce this wastage of claims items (which leave store unsold). To do this, we first understand the waste and markdown sales, by identifying the stores and items having highest waste contribution and lowest markdown sales. Then, we determine the root causes which affect waste and find points of high/low wastage with respect to sales and markdowns. Finally, we recommend data driven methods to reduce waste and increase markdown sales. To delve deeper, we analyzed waste metrices across items and stores in different departments, markets & countries, using descriptive analytics. To keep our analysis limited to relevant stores and items, we use multi-stage Isolation Forest as multivariate anomaly detection technique to identify anomalous stores and corresponding anomalous items, within anomalous stores, which have high waste and low markdowns. Then, we develop a "waste Risk Score" using factors like Total Loss, waste to sales ratio and waste to CVP Ratio and categorize the stores and items in different zones based upon the score percentiles instead of using traditional methods of univariate sorting. To understand the cause of waste we find out waste associated events like abrupt changes in waste and markdowns by utilizing Change Point detection algorithms, like Dynamic Programming & Binary Segmentation Search. Then, we explore drivers which are related to claims waste management process and utilize them to create causal discovery models, like Conditional Dependent Testing & Greedy Search of DAG Space etc. where our motive is to infer causal structure from data. Finally, from the selected waste drivers we develop an optimization engine to understand the impact of variations in drivers on waste and thus suggest optimal disposition actions to the store associates. With even one-fifth of the adoption of waste optimization recommendations from our methodology would save millions of dollars for the retail stores.

3.8.1. Need for Intelligent Waste Management system.

Waste management is a ubiquitous challenge for retailers globally, leading to $18.2 billion of loss annually, around 0.3% of total sales (ReFED,2018) with largest contributors being the Fresh Departments. Fresh departments consist of Strategic Business Units (SBUs) like Grocery, Dairy, Fresh Meat & Produce etc.

In retail, an item is treated as claim in the backroom when it is left unsold due to various reasons like customer return, overstock, product damage due to store negligence etc. (process diagram can be seen in Figure 4).

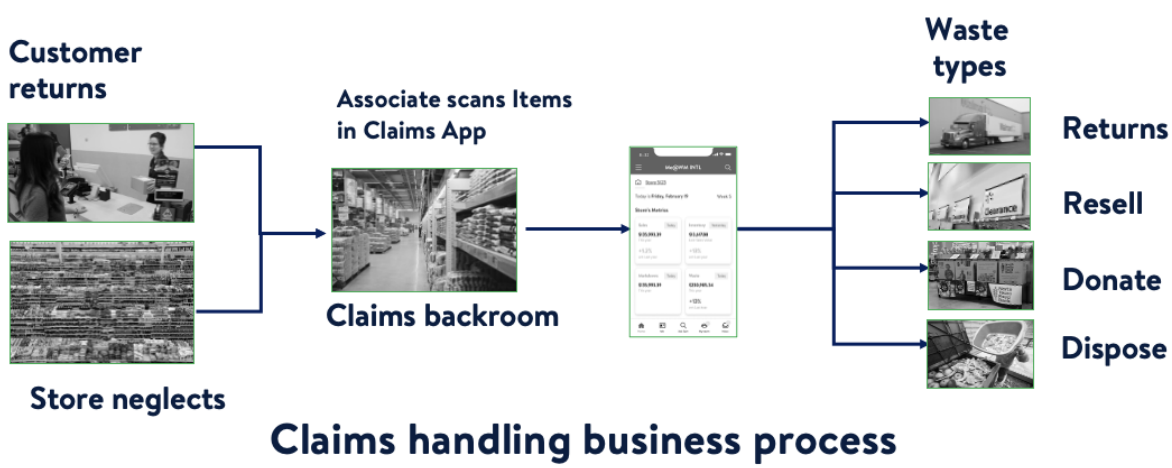


Figure 4: The overall process for Handling Claims in the stores by associates.

This claim can then be considered for resell by marking down the price or other disposition options dispose and return etc which is performed via store associates in the backroom. Thus, analysing claims and choosing the best possible disposition option remains one of the major unsolved problems. This can also be seen from low percentage of markdown sales using discounts(CVP).

But the problem with existing strategies is that most of them are isolated and are based on separate research. Thus, due to lack of proper guidelines or recommendations at region, store and item level, subjective decisions are being taken by store associates at day-to-day level. This has resulted in the significant growth of fresh waste over the years (11% YoY increase).

Various research has been done in the direction of waste reduction in the past. Some of them are theoretical, where the focus was to categorize the type of waste and find the broad factors (like logistics, demand prediction, shelf life etc.) affecting waste. For instance, Moraes et. al. (2021) explained various theoretical strategies for waste reduction like assortment selection process, improving demand forecast and ensuring on-shelf availability. While others took data analytics approaches to reduce waste. Ahmadzadeh et. al. (2023) explained the use of big data and machine learning for waste reduction like use of classification algorithms to classify quality of food items, while Sakoda et. al. (2019) demonstrated the method of demand estimation for stock management. Past research provides us with a good idea about waste areas and factors, which can be checked to have reduction in waste. But there exists no research for providing recommendations to retail stores and which can be adopted as standard methodology to define, understand, and control waste.

In the current solution, we have focused on waste in retail stores and attempted to provide a 360-degree solution to describe the waste causing areas via multiple waste KPIs, causal drivers of waste and finally data-driven recommendations which can be directly utilized by store associates to reduce waste.

The novelty of our solution lies in analyzing waste areas using multi-variate techniques, providing causal relationship between waste drivers and waste types, and finally utilizing causal relationships to provide prescriptive recommendations for waste reduction and profit increase.

3.8.2. Overall model Framework

To create a systematic approach for waste reduction, we started with the task of analyzing waste and metrices which can gauge waste and are of relevance to business. This task was done to develop a comparative study of various factors like waste types, waste to sales ratio, top and bottom stores having the waste etc. at weekly level in different market regions. To observe waste in high cardinality feature, we utilized the multi-variable anomaly detection technique to separate out the items which are anomalous not only in terms of waste but also showing poor performance in other business defined KPIs like waste to sales or CVP to waste ratios. Once relevant stores and items are identified via anomaly detection, we next focused on sorting stores and items by developing a multi-variate waste risk score and identified top and bottom stores and items using this score. This was done to identify stores and items which are performing worst from various fronts altogether rather than analyzing stores and items for each KPI individually.

There are several factors that can affect retail waste –

i. Product Packaging: One of the main contributors to retail waste is excessive packaging. Excessive packaging can lead to more waste and increase the carbon footprint of products. To reduce this type of waste, the retailer can implement various initiatives such as reducing packaging materials, using eco-friendly packaging materials, and promoting products with minimal packaging.

ii. Customer Behavior: Customer behavior can also play a significant role in retail waste. For example, customers may purchase more than they need or return items that are still in good condition. To reduce this type of waste, the retailer can implement policies such as limiting the number of items that can be purchased at one time and offering incentives for customers who bring their own reusable bags.

iii. Inventory Management: Inefficient inventory management can lead to waste due to overstocking or spoilage. To reduce this type of waste, the retailer can use advanced technology such as RFID tags and data analytics to optimize inventory management and reduce spoilage.

iv. Supply Chain Inefficiencies: Inefficiencies in the supply chain can also contribute to retail waste. For example, products may be damaged or expire during transportation, leading to waste. To reduce this type of waste, the retailer can work closely with its suppliers to optimize the supply chain and reduce damage and spoilage.

v. Store Operations: Store operations can also contribute to retail waste. For example, energy consumption, water usage, and waste disposal can all contribute to environmental waste. To reduce this type of waste, the retailer can implement various initiatives such as using energy-efficient lighting and HVAC systems, reducing water usage, and implementing recycling programs.

vi. Product Design: Product design can also play a role in retail waste. For example, products that are not designed to be repairable or recyclable can contribute to waste. To reduce this type of waste, the retailer can encourage suppliers to design products that are more sustainable and easier to recycle.

vii. Customer Education: Customer education is an important factor in reducing retail waste. By educating customers about the importance of sustainability and the impact of their purchasing decisions, the retailer can encourage customers to make more sustainable choices through programs which educates customers about sustainable products and practices.

viii. Employee Engagement: Employee engagement is also crucial in reducing retail waste. By educating and empowering employees to make sustainable choices, the retailer can reduce waste in stores and promote sustainability, such as the "Green Team" program, which encourages employees to take action on sustainability initiatives in their stores.

ix. Community Involvement: Community involvement is another important factor in reducing retail waste. By engaging with local communities and promoting sustainability initiatives, the retailer can reduce waste and promote sustainability, such as the "Community Garden" program, which encourages communities to grow their own food and reduce waste.

x. Technology: Technology can also play a role in reducing retail waste. By using advanced technologies such as RFID tags, data analytics, and artificial intelligence, the retailer can optimize inventory management, reduce waste, and promote sustainability. For example, the retailer can implement a "Smart Cart" program, which uses AI to optimize inventory management and reduce waste.

Prescriptive machine learning models use algorithms and techniques to analyze historical data and provide actionable insights and decisions. These models are often used in the field of supply chain and inventory management to reduce waste.

- Predictive Demand Forecasting: These models use historical sales data to predict future demand for products. By accurately predicting demand, stores can ensure that they only stock what they need, thus reducing waste.
- Inventory Optimization: These models use data on sales, inventory levels, and supply chain constraints to recommend optimal inventory levels. This can prevent overstocking, which can lead to waste if products expire or become obsolete before they are sold.

- Price Optimization: Prescriptive models can also recommend optimal pricing strategies. By dynamically adjusting prices based on demand, stores can sell more products before they expire, thus reducing waste.
- Product Life Cycle Management: These models can predict the life cycle of a product, including when it will become obsolete. This can help stores to manage their inventory more effectively and reduce waste.
- Waste Analytics: These models analyze waste data to identify patterns and trends. This can help stores to identify the causes of waste and take action to reduce it.

Causal Machine Learning is a subfield of machine learning that focuses on understanding the cause-effect relationships between variables. In the context of reducing waste in stores, Causal ML can be used to understand the factors that lead to waste. For example, a causal ML model might reveal that waste increases when certain products are overstocked, or when products are not priced correctly. This information can then be used to inform strategies for reducing waste, such as improving inventory management or adjusting pricing strategies.

Retail is an industry that has been greatly affected by digital transformation, with businesses leveraging technology to improve operations, reduce costs, and enhance customer experiences. One of these technologies is machine learning, which can be used to reduce waste in retail stores and omni-channel retail. It can be applied to both fresh waste and general merchandise to optimize inventory management, improve demand forecasting, and reduce overstocking and underselling.

a. Inventory Management

The traditional inventory management process involves a lot of guesswork, which often leads to overstocking or understocking. Overstocking results in waste as perishable goods may spoil before they're sold, while understocking can lead to lost sales opportunities and unsatisfied customers.

Machine learning can greatly enhance inventory management by learning patterns from past data and making accurate predictions on what will be needed in the future. It uses algorithms to analyze historical sales, seasonal trends, and other relevant data to provide accurate forecasts. This reduces waste by ensuring that only the necessary quantity of goods is ordered and stocked, thereby reducing fresh waste and overstocking of general merchandise.

Methodology: The machine learning models used for inventory management usually involve supervised learning algorithms such as linear regression, decision trees, and random forests. These algorithms are trained on historical data and then used to make predictions about future demand.

b. Demand Forecasting

Predicting customer demand is one of the most challenging aspects of retail. Traditional methods of demand forecasting often fall short, leading to overproduction and waste.

Machine learning algorithms can analyze large amounts of data, including historical sales data, market trends, promotional activities, and even external factors like weather and holidays, to make accurate predictions about customer demand. This allows retailers to produce and stock the right amount of products, reducing waste and improving customer satisfaction.

Methodology: Time series analysis and regression models are commonly used for demand forecasting. Deep learning models, such as recurrent neural networks (RNNs), have also proven effective in capturing complex patterns in time-series data.

c. Personalized Recommendations

Personalized recommendations can also help reduce retail waste by promoting products to customers who are most likely to purchase them. This not only improves sales but also helps to move inventory faster, reducing the chance of waste.

Machine learning algorithms can analyze customer behavior, purchase history, and browsing patterns to make personalized product recommendations.

Methodology: Collaborative filtering and content-based filtering are two common techniques used in recommendation systems. Collaborative filtering recommends products based on the behavior of similar customers, while content-based filtering recommends products based on the customer’s past behavior.

d. Dynamic Pricing

Dynamic pricing involves adjusting prices based on supply and demand. When there's a surplus of stock, prices can be lowered to encourage purchase and reduce waste. When stock is low, prices can be increased to maximize profit.

Machine learning can help establish dynamic pricing strategies by predicting demand and optimizing prices accordingly.

Methodology: Machine learning models for dynamic pricing often involve reinforcement learning, where the model learns to make decisions by receiving feedback on its actions.

e. Predictive Maintenance

Retailers can also reduce waste by using machine learning for predictive maintenance. This involves predicting when equipment, such as refrigeration units or delivery vehicles, is likely to fail and performing maintenance before it does. This not only prevents costly downtime but also reduces the risk of product spoilage.

Methodology: Predictive maintenance models often use anomaly detection algorithms to identify unusual patterns in sensor data that may indicate a potential failure.

f. Waste Analytics

Machine learning can also be used to analyze waste data and identify patterns and trends. This can help retailers understand where most of their waste is coming from and develop strategies to reduce it.

Methodology: Clustering algorithms, such as k-means, can be used to group similar waste items together and identify common characteristics. Regression models can also be used to predict future waste levels based on current trends.

In conclusion, machine learning offers a wide range of techniques that can be used to reduce waste in retail stores and omni-channel retail. By improving inventory management, enhancing demand forecasting, providing personalized recommendations, implementing dynamic pricing, predicting maintenance, and analysing waste, retailers can significantly reduce both fresh waste and general merchandise waste. This not only helps to save costs but also contributes to sustainability efforts.

After completion of waste descriptive analysis, we then identified the factors which can cause waste. We identified a list of controllable and non-controllable factors which can affect waste and then performed causal discovery analysis to establish the relationship between waste and its drivers. To explain waste in terms of its drivers in a more realistic manner, we identified the series of events using state-of-the-art change point detection techniques. Finally, by utilizing the relationships between waste and its drivers, we have developed a recommendation system to provide recommendations to store associates at various levels (like departments, finelines & items) to optimize waste.

Here can be some of the controllable factors that affect waste -

a. Inventory Management: The way a retailer manages its inventory can significantly impact waste levels. Overstocking can lead to waste, particularly with perishable items, while understocking can result in missed sales opportunities.
b. Procurement Practices: The choice of suppliers, the quality of the goods procured, and the timing of procurement can affect waste.
c. Product Life Cycle Management: How a retailer manages the lifecycle of its products, from introduction to discontinuation, can affect waste.

d. Packaging: The type and amount of packaging used can contribute to waste. More packaging typically means more waste, although some types of packaging can extend product shelf life and reduce waste.
e. Storage Conditions: The conditions in which products are stored can affect their shelf life and, therefore, waste levels. This includes temperature, humidity, and light exposure.
f. Staff Training: Employees who are not properly trained in inventory management, product handling, and waste reduction practices may contribute to higher waste levels.
g. Pricing Strategy: Pricing strategies can influence consumer purchasing behaviour and, thus, waste. For example, discounting perishable items close to their sell-by date can reduce waste.
h. Promotional Activities: Promotions can increase sales but can also lead to overproduction and waste if not properly managed.
i. Product Placement: The way products are displayed in the store can affect their likelihood of being sold before their expiration date.
j. Sell-by and Use-by Dates: Retailers have control over these dates, which can influence consumer purchasing behaviour and waste.
k. Customer Education: Educating customers about waste and how to reduce it can help decrease waste.
l. Waste Management Systems: The systems a retailer has in place to manage and reduce waste can significantly affect waste levels.

m. Product Selection and Range: The range of products a retailer carries can impact waste. A wider range may lead to more waste due to lower turnover of each product.
n. Supply Chain Management: The efficiency and effectiveness of a retailer's supply chain can impact waste levels.
o. Store Layout: Store layout and design can influence shopping behavior and, therefore, waste.
p. Technology and Equipment: The use of technology and equipment can affect waste. For example, refrigeration equipment that is not functioning optimally can lead to spoilage and waste.
q. Order Frequency: The frequency and quantity of orders placed by a retailer can impact waste. Frequent, smaller orders may reduce waste but increase delivery costs.
r. Return Policy: A retailer's return policy can influence consumer behaviour and waste.
s. Quality Control Measures: Effective quality control can reduce waste by ensuring that damaged or substandard goods are not sold.
t. Collaboration with Suppliers: How a retailer works with its suppliers can affect waste. For example, arranging for just-in-time deliveries can reduce storage time and potential waste.

There may be also some non-controllable factors like -

a. Consumer Behaviour: Consumer purchasing habits and preferences can influence waste levels. For example, consumers who prioritize fresh produce over canned or frozen options may contribute to higher waste levels.

b. Seasonal Demand: Demand for certain products can fluctuate seasonally, leading to potential overstocking and waste.
c. Market Competition: The level of competition in the market can influence a retailer's pricing and stocking strategies, which can impact waste.
d. Economic Conditions: The state of the economy can affect consumer spending and, therefore, waste.
e. Legal Regulations: Laws and regulations can limit a retailer's options for managing waste.
f. Environmental Factors: Weather, natural disasters, and other environmental factors can affect supply chains and demand, leading to potential waste.
g. Technological Changes: New technologies can disrupt retail operations and supply chains, potentially leading to waste.
h. Industry Trends: Trends such as the move towards organic or locally-sourced products can affect waste.
i. Supplier Reliability: If suppliers are not reliable, it can lead to stockouts or overstocking, both of which can result in waste.
j. Political Factors: Changes in government policy or political instability can disrupt supply chains and affect waste.
k. Cultural Trends: Changes in consumer preferences driven by cultural trends can lead to waste if retailers do not adjust their product offerings accordingly.

l. Demographic Changes: Changes in the population, such as aging or urbanization, can affect demand for certain products and, therefore, waste.
m. Changes in Disposable Income: Changes in consumers' disposable income can affect their purchasing behavior and, thus, waste.
n. Social Trends: Social movements, such as the push for sustainable and ethical consumption, can impact waste.
o. Public Health Situations: Health crises like the COVID-19 pandemic can dramatically affect consumer behavior and supply chains, leading to increased waste.
p. Climate and Weather Conditions: Changes in climate or extreme weather conditions can disrupt supply chains and affect product shelf life, leading to waste.
q. Geographical Location: The location of a store can affect its waste levels. For example, stores in urban areas may have different waste patterns than those in rural areas.

Performing causal discovery analysis involves using statistical and machine learning techniques to identify and quantify the relationships between various factors and waste in a retail setting. This analysis can help to identify which factors are causing waste, and to what extent they are contributing.

Methodology:

a. Data Collection: Gather data on waste and all potential driving factors. This can include sales data, inventory data, promotional data, economic indicators, and more.

b. Data Preprocessing: Clean and preprocess the data to prepare it for analysis. This can involve handling missing data, normalizing data, and encoding categorical variables.
c. Feature Selection: Use techniques like correlation analysis, mutual information, and feature importance from machine learning models to identify the most relevant factors.
d. Model Building: Build a causal discovery model using techniques like Bayesian networks, structural equation modeling, or causal inference algorithms. These models can help to identify the causal relationships between factors and waste.
e. Model Evaluation: Evaluate the model using techniques like cross-validation and by checking the statistical significance of the identified relationships.
f. Interpretation: Interpret the results to understand which factors are driving waste and how they can be controlled.

Through this analysis, retailers can gain insights into the main drivers of waste in their operations and take targeted measures to reduce it. For example, if inventory management practices are identified as a major cause of waste, retailers can implement machine learning models to optimize their inventory and reduce waste. Similarly, if consumer behavior is found to be a significant factor, retailers can develop strategies to influence consumer behavior and reduce waste.

The model framework adopted in our waste management system can be seen in the bock diagram of Figure 5. Our work has been divided into three phases where the initial part comprises of development of dashboard to create a comparative study of waste among various items, departments, stores, markets and regions. Our current research is focused on predictive and prescriptive analytics to understand waste via drivers and provide recommendations to the business and store associates for waste mitigation.

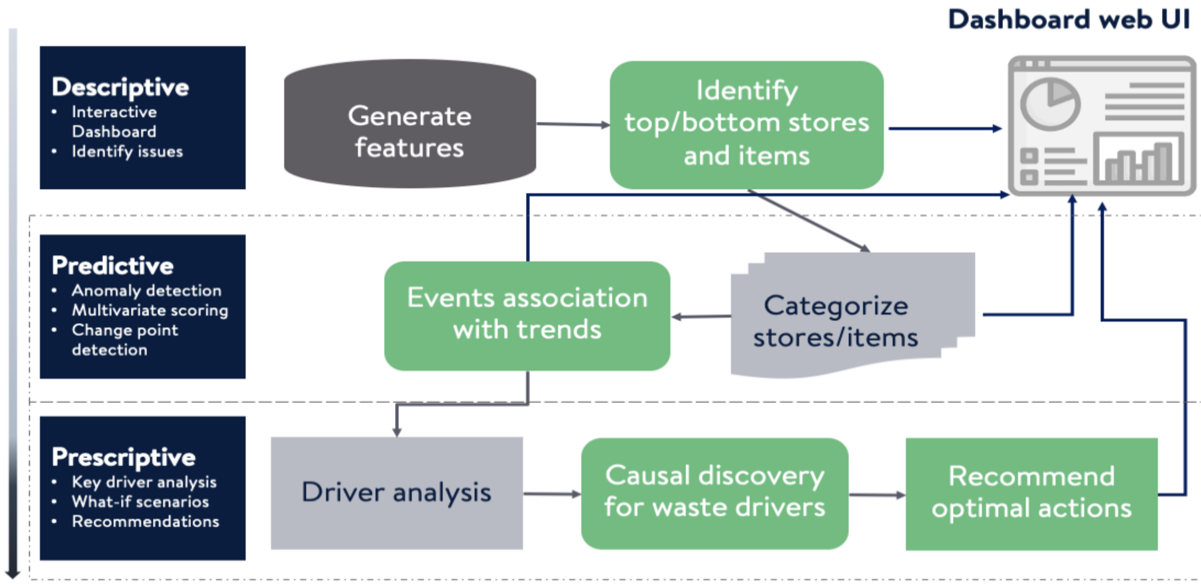


Figure 5: Model Framework

3.8.3. Model Details

To create efficient waste management methodology, we created numerous data driven models at various phases of research solving different areas of waste understanding.

3.8.3.1. Multi-Variate Anomaly Detection

A brute force approach to analyze waste at some levels is feasible due to the limited number of regions, markets, and areas. But this approach fails in the case of high cardinality of the feature (such as more than 20K+ items in each store). To tackle this, we utilized the multi-variable anomaly detection technique to separate out the items which are anomalous, not only in terms of waste but also showing poor performance in other business defined KPIs like waste to sales or CVP to waste ratios.

Anomaly detection, also known as outlier detection, is the identification of rare items, events, or observations that raise suspicions by differing significantly from the majority of the data. In the context of multivariate data, anomaly detection aims to find patterns in multidimensional datasets that do not conform to expected behaviour.

Several techniques are widely used in multivariate anomaly detection, including:

a. Statistical techniques – These methods are based on the assumption that the normal data follows a Gaussian distribution. Any data instance that doesn't follow this pattern is deemed an anomaly.
b. Machine Learning-based techniques - These techniques involve training a model to understand patterns in the data and then identifying any data points that do not fit the model. Some of these techniques include Neural Networks, Decision Trees, k-Nearest Neighbours, Clustering-based methods, etc.
c. Density-based techniques - These methods, such as Local Outlier Factor (LOF), are based on the density of data points. They work by calculating the local density of a data point and comparing it to the densities of its neighbours. If the density of a particular point is significantly different from its neighbours, it's considered an anomaly.

d. Distance-based techniques - These methods, such as k-nearest neighbours (k-NN), measure the distance between data points. If a data point is a certain distance from others, it's considered an anomaly.
e. Ensemble techniques – These methods combine the results of multiple anomaly detection algorithms to improve the detection performance.

The use of these techniques in retail stores for waste reduction opportunities can be highly beneficial.

a. Inventory Management: Multivariate anomaly detection can be used to identify unusual purchases or stock levels that could indicate waste. For instance, if a particular product is being overstocked consistently while its sales remain low, this could signify a waste of storage space and resources.
b. Fraud Detection: Anomaly detection can help in identifying fraudulent transactions or behaviours that result in financial waste. For instance, if there are unusual patterns in employee discounts or returns, these could be potential signs of fraud.
c. Energy Usage: Anomaly detection can also identify unusual patterns in energy usage, which could indicate inefficiencies or equipment malfunctions. Correcting these anomalies can lead to significant cost savings.
d. Customer Behaviour: Understanding customer behaviour patterns can also help in waste reduction. For example, if customers consistently avoid certain products, it may indicate that these products are not well-received and stocking them is a waste of resources.

e. Supply Chain: Anomaly detection can help identify issues in the supply chain that might be causing waste. For instance, if certain suppliers consistently deliver late or provide subpar products, these are wastes that can be reduced.

f. Pricing: Unusual pricing patterns can also be identified and corrected. For instance, if certain products are consistently over or underpriced, this could lead to financial waste.

In conclusion, multivariate anomaly detection can be a powerful tool in identifying waste reduction opportunities in retail stores. By identifying and correcting these anomalies, retailers can significantly reduce waste, improve efficiency, and increase profits.

We tried various anomaly detection techniques like Local Outlier Factor and Isolation Forests to identify subset of stores and items which have poor performance in not only absolute waste but even other waste related KPIs like waste to sales ratio etc. The results from Local Outlier factor are not encouraging but Isolation Forest algorithm provided good results, where the results from anomaly detection are evaluated by values of multiple KPIs in resulting set of stores and items. The multi-variate Isolation Forest methodology is explained in Figure 6. Here, instead of using a single Isolation Forest on all predictors, we have created two models on subset of predictors with each subset having a similar type of predictors. These subsets of predictors are based on:

i. Waste related factors like types of waste and waste to sales

ii. Markdown related factors like CVP to waste ratio

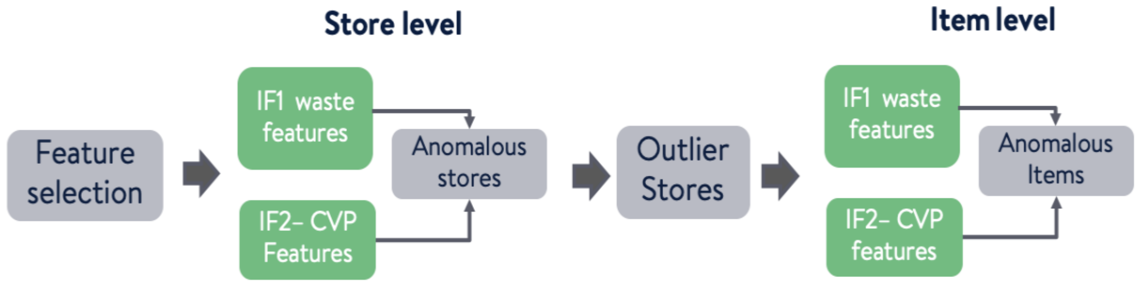


Figure 6: Multi-Variate Anomaly Detection

3.8.3.2. Waste Risk Score

The second approach was the development of waste risk score to sort stores and items. The methodology adopted for this multi-variate score can be seen in Figure 7. To develop this score, we have created features using both business KPIs and statistical techniques like binning and transformations (log, square root etc.). We also removed correlated features to choose only distinct set of features for scoring and estimated the coefficients of each predictor by fitting the data into normal distribution. From the descriptive analysis, we had observed that different regions have different distributions of waste and thus creation of single score for all regions was not feasible. Thus, we developed separate equations for each region which can further be tuned by re-run of the method as waste distribution for any region can be varied over a period of time. Parameters used to calculate final waste risk for different regions are –

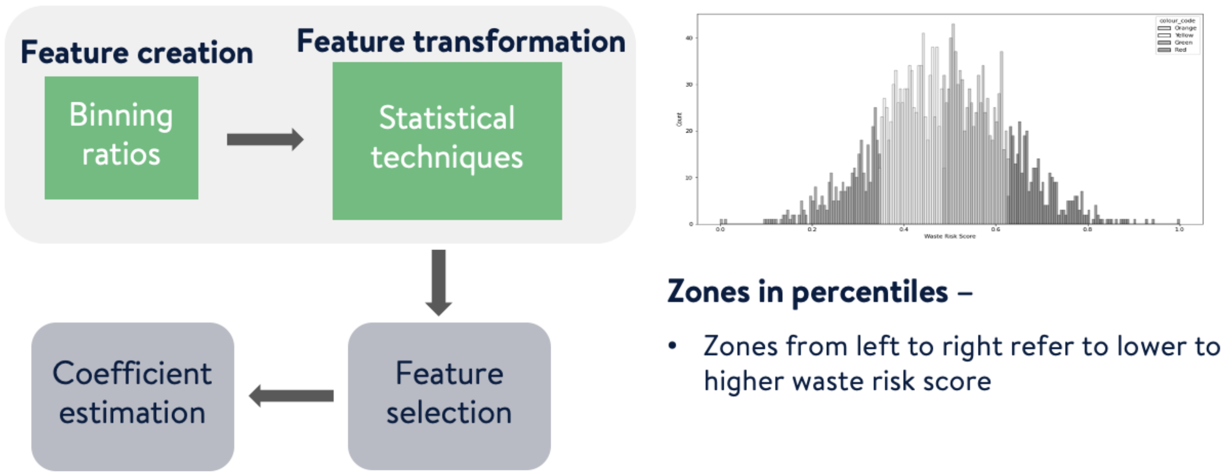


Figure 7: Waste Risk Score

i. For Canada and Mexico region, waste risk score is calculated based on waste, throwaway to sales, waste to CVP and waste to donation
ii. For Central America and Chile region waste risk score is calculated based on waste, waste to CVP and waste to sales

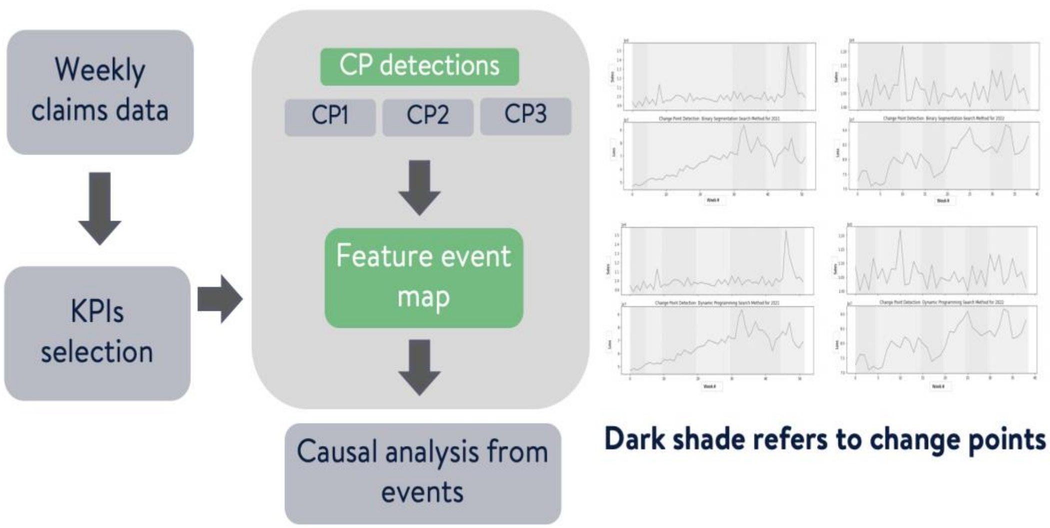


Figure 8: Change Point Detection

3.8.3.3. Change Point Detection

Listing down waste drivers from hit and trail would be an inefficient and time-consuming task. Thus, we adopted a better method of tracking down events related to waste KPIs and then to understand the reason for the change related to the event.

Also, due to varying behavior of various KPIs at different time intervals, we adopted a methodology shown in Figure 8, in which we have separately performed change point detection algorithms on each KPI. Finally, intersection of all change points was done to find such change points which have maximum impact from various waste perspectives.

Change point detection is a statistical technique used to identify points in a data sequence where the properties of the data significantly change. These points, known as 'change points', often denote a significant shift in the underlying process that generates the data.

In the context of waste management, change point detection can be incredibly useful. For example, it can help to identify sudden increases in waste generation, which might indicate a problem or inefficiency in a production process. This could be a shift in the amount of waste being generated, the type of waste, or in the pattern or frequency of waste generation. By identifying these change points, businesses can investigate and address the root causes of these shifts. This could lead to cost savings, improved efficiency, and reduced environmental impact. Furthermore, understanding these changes can also help in predicting future waste generation and planning appropriate waste management strategies.

There are several techniques used in change point detection, including:

a. Cumulative Sum (CUMSUM): This algorithm detects shifts in the mean of a process. It calculates the cumulative sum of the deviations of each data point from the mean, and a significant change in this sum can indicate a potential change point.
b. Bayesian Change Point Analysis: This technique uses Bayesian statistics to infer the probability of a change point at each position in the dataset. It can handle changes in mean, variance, and distribution of the data.
c. Sequential Analysis: This involves analysing data as it becomes available, and deciding whether to continue with the process or stop it due to a detected change.
d. Likelihood Ratio Test: This technique tests the likelihood of two models – one with a change point and one without. The change point is determined by comparing the likelihoods.
e. Bootstrapping: This is a resampling technique that can be used to estimate the variability of change point estimates.
f. Spectral Density Ratio Method: This method detects change points based on the ratio of spectral densities before and after each candidate change point.
g. Binary Segmentation: This method involves iteratively dividing the time series into binary segments and testing for changes within each segment.
h. Pruned Exact Linear Time (PELT): An optimal method for multiple change-point detection. It aims to minimize the computational cost while accurately detecting change points.

These techniques can be applied to detect changes in variance, mean, distribution, and other parameters depending on the specific needs of the analysis.

We have attempted various state-of-the-art change point detection algorithms like Dynamic Programming and Binary Segmentation, but Dynamic Programming Search gave most appropriate change point events when evaluated against actual events.

3.8.3.4. Waste Drivers & Causal Discovery

After getting events which are related to waste KPIs, we identified the drivers by performing exploratory data analysis and discussing with business for causality rather than just correlation. Following are the categories of variables which we have tried for causal analysis –

i. Item and store dimension features
ii. Markdown related features
iii. On-shelf Customer Availability (OSCA) and Perpetual Inventory (PI) related features
iv. Delivery/shipment features

To identify store level relationship, we have calculated these features based different item categories which we have calculated by clustering of items based on values like shelf life etc. Then we utilized these features to develop causal analysis between waste and its drivers. Baseline causal discovery has been performed initially using tree-based models but then we shifted to advanced graph based causal discovery models. Using these models, we can infer graph-based relationships between waste to sales KPI and drivers which is depicted in Figure. 9. Some examples are –

- Minimal CVP usage -> Throwaway
- Aggressive Markdowns -> CVP lost sales.
- Markdown very close to expiry -> CVP Throws

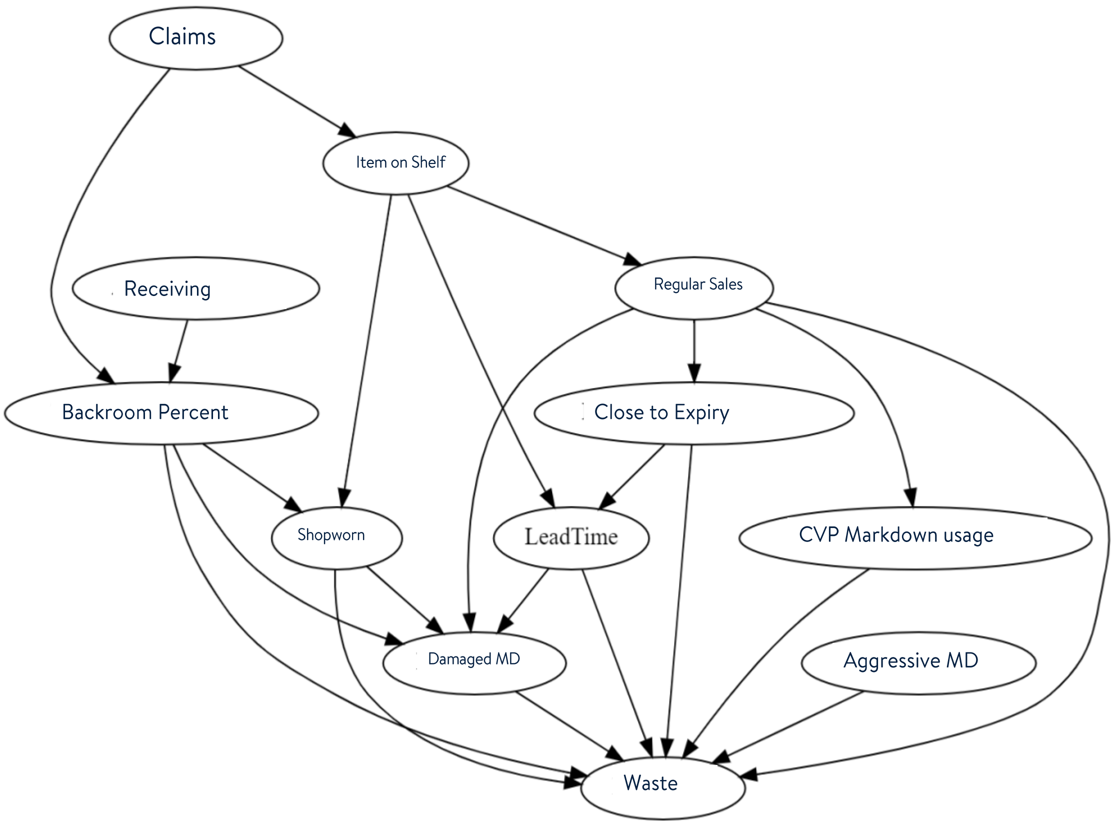


Figure 9: Causal Relationship among waste KPI and its drivers

Using these relationships, we have developed the recommendation engine using Causal ML techniques to provide acute pointers to store associates related to item handling, markdown patterns etc. which can result in substantial waste savings.

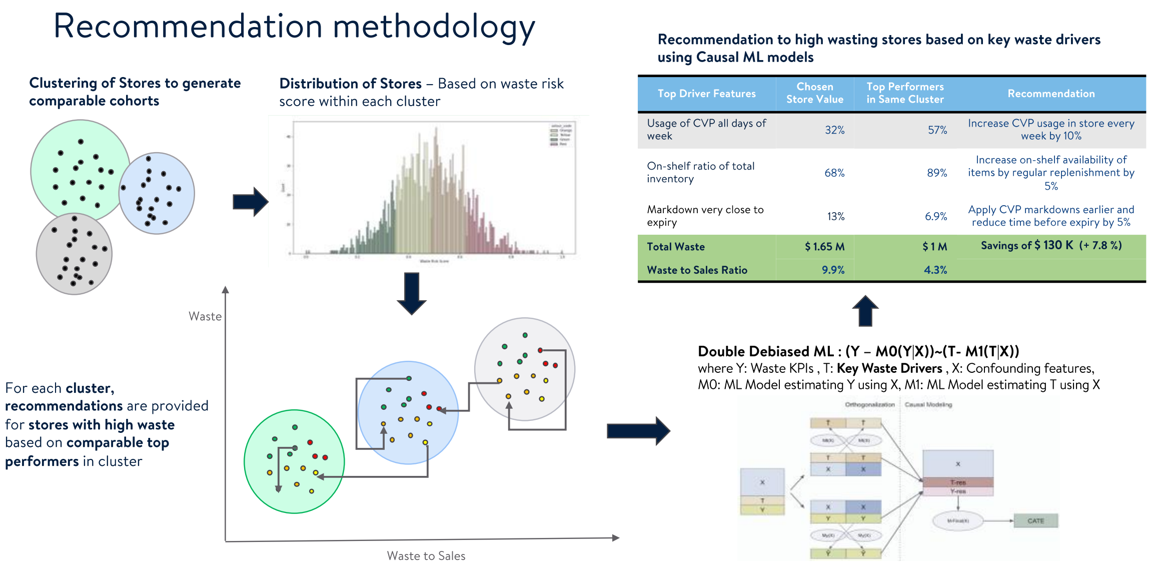


Figure 10: Methodology to identify critical recommendations to reduce waste.

We first cluster the stores into homogeneous groups for better comparison, next we overlay the waste risk score to bin the stores within each cluster to identify the top and bottom performers. The recommendations will be provided to bottom preforming stores by comparing with top performers in the same cluster. Also, since there will be a natural hierarchy of the clusters in terms of waste, we can recommend the top performers in a cluster to move up to the next higher cluster which is comparable. We use a double-debiased ML solution to understand predict the incremental impact of the key waste driving features on the outcome Waste KPIs. We can provide recommendations to the chosen store, comparing with top performers, like increasing the CVP usage by 10%, and at an overall level get the impact on the outcome, which has been illustrated in Figure. 10.

3.9. Early Detection of Out-of-Stock using Computer Vision

3.9.1. Need for early detection of Out-of-Stock

In any retail business, it is very important that the customers are able to find the required items on shelf. At retail stores, out-of-stock shelves inevitably reduce sales and customers and cannot be considered a temporary loss. Some survey results claim that in case of out-of-stock, in 31% cases the customers would purchase in a different store, and 9% customers don't purchase any products. Since checking stock is such an important task, stores generally increase the frequency of checking, but this also increases the time spent by staff. The challenge for store owners is how to create an efficient process for checking the shelves. For very large stores with millions of items, it becomes very difficult to manually keep track of all the items on shelf, and hence a computer vision-based solution can automate the task of determination of potential out-of-stock and an integrated system can be developed to replenish the items from the back room or raise alerts to the replenishment managers to deliver the next batch of products. In some situations, the customers can decide they no longer need a product, and can put it in a different store location than the original. This is a serious problem, since the product becomes unaccounted for, and might be lost or stolen. Usually, the back room of the stores would maintain inventory required to replenish the empty shelves, and such restocking usually happens across the stores during off-hours for all categories. There can be some high moving items which can get over in a short period of time, and customers would not be able to find the required product, even when it is available at the back room. As a retailer, the end goal is to increase sales, improve overall category profitability and become a store that caters for its customer so that they'll keep coming back. By ensuring that the shelves are well stocked, the retailer can create a culture of accountability in their stores and increase customer satisfaction by manifolds. With the advances in technology, computer vision and artificial intelligence can come to the rescue of the retailers. There are robots which can scan the shelves, or drones which fly overhead, and even static cameras which can take images at regular interval. The images from these devices are passed to a central database, where they are processed, stitched together to recreate the actual implementation in the store shelves. Our system utilizes computer vision to get real-time feed of the store shelves and can determine beforehand if a particular product is out-of-stock or is tending towards out-of-stock, and then generates an alert to the store associates to replenish the required products. In some cases, the replenishment process could also be automated using drones, or robots in the store. Usually, the store specific planogram details like product name, horizontal and vertical facing quantities, would be available beforehand in the central data storage. Now our system can utilize the pre-defined information, stitched store shelf image, to determine exact or partial out-of-stock scenarios at a real time. There has been some work in the domain of out-of-stock detection, however all such work is heavily dependent on supervised learning, with tagged data, and none of them consider partial out-of-stocks. It is very difficult and too expensive for a large retailer to tag millions of items and build a supervised learning framework. That is where our novel technique and system is very crucial.

3.9.2. Overview of the system

Our system is primarily developed as a real-time Artificial Intelligence system, which keeps capturing feeds of images from either drones, shelf-scanning robots or static cameras, stiches together the shelf layout, runs a series of models for out-of-stock detection and finally generates a report through automated alert to the responsible store manager or associates to replenish the items. The input to our system mainly consists of Shelf Image from Stores, Planogram details like product name, horizontal and vertical facing quantities. The target is to identify and detect the voids present in the Shelf image as Partial Void or Out of Stock. Here Partial void is referred to the configuration when some of the products of a given type are sold out. Out of Stock indicates that the product is completely sold out. The images are collected from the cameras or drones, then converted to an encoded string format, and then passed through an API to the central computing device hosted in cloud. The ensembled model is deployed as a service through the API, which takes as an input the actual store image in encoded format, then processes the image using pre-trained weights, and outputs the presence or absence of complete or partial out-of-stocks, along with the location co-ordinates of the void. The output image, with the demarcated voids are encoded and sent back to the store systems as alerts. The model training happens in a batch process at regular interval to keep updating the model parameters with the new data received. The overall architecture has been represented in Figure 11.

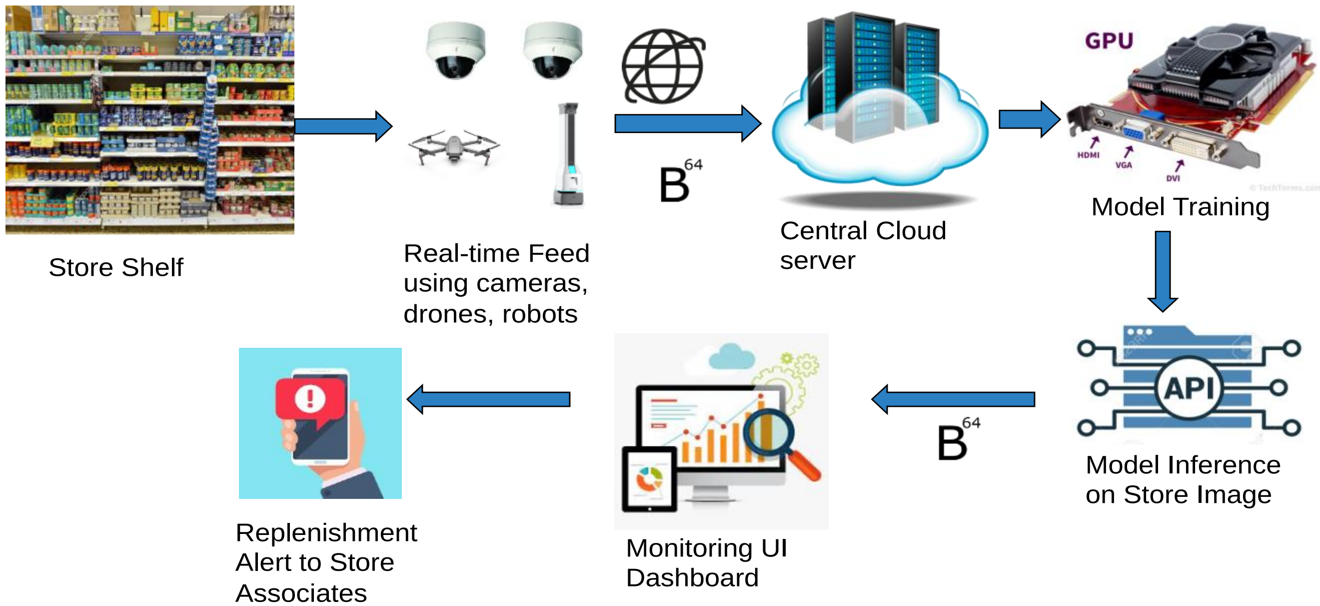


Figure 11. Computer Vision based Out-of-Stock detection system architecture.

The images from the drones or cameras, once stitched are passed on to the system mentioned above. For each image frame the steps are ran iteratively and then a comprehensive report on the shelf availability is generated for regular monitoring. If there are instances of partial or completed out-of-stock present in any particular aisle, then an alert would be sent across to the particular store associates or the store manager, on their mobile devices or through email, with details on which products need to be replenished. The system also cross-references with back-room inventory to decide if the supply chain manager needs to be alerted to expediate the next batch of shipment.

3.9.3. Ensembled Out of Stock Detection Framework

For the system to work very fast and be scalable, we need to optimize the performance. Also, we cannot rely upon only one variant of the model, since it needs to be generalized across all types of products in the store. We can use a variety of supervised techniques to create an Ensembled Out-of-Stock framework. We have the following main algorithms:

Region Based Convolutional Neural Networks: The goal of R-CNN is to correctly identify the regions of the main object in the image via proposing bounding boxes having objects and classifying them accurately. The approach proposed to apply high-capacity convolutional networks (CNNs) to bottom-up region proposals to localize and segment objects and when there is minimal supervision, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, boosts performance significantly [3].

Masked Region Based Convolutional Neural Networks: This solves Instance segmentation problems in a 2-stage framework. In the first stage, it detects the bounding boxes and in the second stage predicts the object class and generates a mask in pixel level for the object. Masked R-CNN presents a simple, flexible, and robust framework for object instance segmentation which efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance [4].

Single Shot Detection: SSD, need to take one single shot to detect multiple objects within the image and is much faster compared with two-shot RPN-based approaches. The approach presents an object detection framework using a single deep neural network by discretizing the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. During inference, the model generates scores for the presence of each object category in each default box and suggests robust and efficient adjustments to the box and enhance the mapping [5].

These steps are followed in sequence to properly identify the most probable region where the out-of-stock can be present. This ensemble framework ensures high accuracy of the model with very fast response time of the model within milliseconds.

3.9.4. Modelling & Inferencing Framework

In this section we would discuss in more details the actual model training and inferencing framework to determine out-of-stock scenarios. Broadly we follow the below-mentioned series of steps:

Dataset Preparation and Augmentation: For our case, the primary challenge is availability of good quality labeled images with complete and partial out-of-stock present. We overcome this challenge through a novel data augmentation technique. The primary task comprises of i) Dataset creation and formulation in an Object Detection framework, ii) Dataset Augmentation for Object detection. The augmentation strategies for object detection tasks are much more complex that in simple classification tasks as we must keep a track of the position of the object while rotation and translation of the image. We leverage concepts from [6] which helps us in learning and augmenting high quality data with limited features

Feature Extraction: When it comes to working with deep learning models there is no explicit need for extracting the features from the data to train the models as compared to traditional approaches. Convolutional Neural Networks acts as feature extraction layer and these features are used for downstream like in this case detecting void regions in planogram. Convolutional features are used for classification as well as localization for the task in hand and sometimes features from multiple layers are also used to make the network predict accurate outcomes irrespective of the object size.

Model calibration: The model calibration is done by minimizing on deviation: i) Where the object actually is (location loss), ii) What is the object (class loss)

These techniques help in increasing confidence of detection of out-of-stock. This is further validated with the information present in the planogram about the number of products planned to be present in the shelf to accurately determine the partial out-of-stock and predict beforehand the estimated time when the product will go out-of-stock based on the rate of purchase.

3.9.4. Advantages of Computer Vision powered Inventory Management System

We have faced multiple challenges while building our system, and we have made our system resilient to these. The primary advantages of the system are following:

- Semi-supervised Out of Stock Detection Methodology:
    - Very few examples of void images are fed as an input to the model.
    - Created an intelligent data augmentation framework to enhance the training set intelligently
    - Ensembled model approach to enhance the semi-supervised performance by capturing the complimentary information.

- Detection of Partial Void and Out of Stock Detection in very less time
    - High accuracy of the model helps in detecting both the Partial void and Out of Stock
    - It is very crucial for business to understand when there is an out of stock or partial voids so that they can take immediate actions and our model is efficient and can inference in very less time
    - It is scalable and can be implemented for any Shelf images and it will be able to identify the partial voids and out of stock.

Our system also works under different lighting conditions in the store and is robust to partial image presence, and approximate matching.

An extension of this solution is to get viewed and ultimately bought and item similarity scores at the store level. The following steps will be followed for the calculation of the store VUB scores:

a. For each shelf in the store sensors with IOT technology need to be placed at equidistant spaces.

b. The location of the sensors will be category specific based on average item dimensions.

c. The time spent by a customer in front of a particular product can be calculated by the position and the height of the eye-level of the customer.

d. The data recorded by adjacent sensors can be used to identify the list of items viewed and the maximum time spent by a customer in front of any particular product in a list before putting it in the cart.

Items in the same category which are viewed and bought immediately after a purchase can be used to obtain complimentary items and item similarity index. Based on the VUB scores we can identify the substitutes and based on similarity we can identify the complements. Based on the VUB scores and the set of attributes of the products on the viewed list we can determine the key driver attributes which drive the ultimate selection of product. The items with consistently undesirable attributes can be substituted with suitable products of better attributes which will increase the sales. The system will send automated item assortment recommendations to the store manager and category manager so that better assortment decisions can be made. The VUB scores can be correlated with POS data to determine the actual proportion of items bought among those which are picked from the shelf. This will help keep track of shrinkage in the stores.

Currently there is no existing process for calculating the viewed and bought scores at the store level. Decisions of complementary and substitute items are made based on just past transaction data and the customer buying decision is not properly captured. Inclusion of online sales data into store level assortment can be an alternative approach but the huge dissimilarity in the product sets in online and store data and also difference in customer base can make such analysis less fruitful.

Detect the items which are viewed before purchase: The first target of this system is to identify the list of items which are considered by a customer before making a purchase decision. For the particular category under consideration the average item dimension of the products can be calculated. The optical sensors can be placed on the shelves at equal distances based on average item dimension to differentiate between the products chosen.

Bluetooth technology can be used to communicate within adjacent sensors to avoid double counting of customers and handling data gathering in case large number of customers are present during peak hours. The time spent in front of each item will be recorded by the system using the sensor data. Based on a logistic curve the optimal time cut-off will be decided so that if the customer spends more time viewing that product than the cut-off time then that product will be considered in the list. Also the system will record the censored data above the cut-off period for each customer-product combination.

If an item is picked up by the customer and again returned to the shelf a higher importance will be given and that item will be surely considered in the list of substitutes. This part of the process will generate a list of items considered by the customer along-with the time spent viewing it.

Calculation of the VUB score: The proportion of customers who viewed the item B among those who purchased the item A will give a base VUB score of the item B with respect to A. The ratio of the average time spent by the customers in viewing the item B among those who purchased A to the average time spent in viewing all items in the category among the same set of customers, can be multiplied to the base score to get an improved VUB score at the store level. The items with the best VUB scores with respect to an item A are considered to be the most suitable substitute items for A.

Item similarity score calculation : After an item is added by the customer in the cart the immediate next item which is viewed by the customer for consideration will be identified, and it will be added to the list of similar items. The next item which is put in the cart immediately after putting one item in the cart is also added to the list, and it is giver a higher weightage. The proportion of customers who viewed the item B immediately after buying item A is the base “also viewed item similarity score” of B with respect to A. The proportion of customers who bought the item B immediately after buying item A is the base “also bought item similarity score” of B with respect to A. The base item similarity score of item B with respect to A is the weighted average of the respective “also viewed item similarity score and “also bought item similarity score”, with a higher weightage given to “also bought item similarity score”.

The ratio of the average time spent by the customers in viewing the item B immediately after purchasing A to the average time spent in viewing all items in the category among the same set of customers, can be multiplied to the base score to get an improved item similarity score at the store level. The items with the best item similarity score with respect to an item A are considered to be the most suitable complements for A.

Key item attribute analysis: The items in the list in previous sections are the list of suitable substitute items for any given product. The products in the category can be thought of as a function of several latent product attributes. The differentiating attributes among the products in the substitution list can be treated as dependent variables and the VUB score as the independent one. Then a suitable classification model like random forest can be run and the feature importance score obtained from the model will be used to identify the key driver attributes which lead to the buying decision. The magnitude and number of differences among the features of the items viewed and bought should decrease as the VUB score increases, thus creating a decision tree mimicking the customer's judgement process. Using this decision tree the items furthest away from the item bought can be identified as the one with most distinguished features from the bought item. Comparing the popularity of the products the least popular items can be substituted with the substitute products with more desirable features.

Comparing the attributes of the items in the complement list in part 3, we can understand the buying decision of the customer and place those complimentary items adjacently to increase cross-sales. The recommendations of substituting less desirable items and changing the placement of complimentary items will be sent at a monthly interval to the store manager and category managers through an alert in a mobile application, so that they can make better assortment decisions. The key attribute analysis will also be insightful in driving the pricing decisions. When all other attributes match between the bought item and its substitute, other than price then the price of the substitute item can be changed accordingly to increase its sales.

There is no current method of getting this level of item affinity and substitutability at the store level. This algorithm is a new feature which will enable better assortment and pricing decisions.

The system will help in increasing the revenue of the store by suggesting proper substitutes for poor performing items and increase cross-sales by identifying complements. The automated alerts to the store manager will decrease his manual labour of checking the situations of all the shelves in the store and will help him take assortment decisions easily. The VUB scores and the item similarity scores can be used in creating the Customer Behaviour Tree (CBT) which is a part of assortment planning. The VUB scores which are calculated at a regular interval also give the number of items of that category which are getting picked up from the shelves by customers as a byproduct. This measure can be used along with the actual POS data of that time period to find the proportion of items actually bought among those which are removed from the shelves. This can be adjusted with the return items to calculate the theft and loss of inventory. The system can be utilized for better understanding of the demand pattern of various categories. The system can be used in future for integration of online and stores shopping experience. It will improve customer experience and increase the overall revenue

Detection of Customer Interest in an Item : A system of sensors (could be optical sensors, weight sensors, RFID, camera and combination of these) detects and records the amount of time spent by the user in front of the item. The adjacent sensors communicate through Bluetooth to detect movement of customers from one item to another and to detect double counting.

Working Mechanism of the Sensor Network: Sensors are placed at equidistant intervals on the shelves so that one sensor system can monitor all the occurrences of a particular item. Now the order of the sensors determines the movement of a person through the isle.

When a person comes in the area captured by a particular sensor, it goes from an inactive state to semi-alert state and also sends the information through Bluetooth to the neighbouring sensors so that they may be in stand-by position from inactive.

Now the primary sensor fetches the data from the database about the optimal cut-off time for it to consider that the person is interested in the item.

If a person comes in the radar of the first sensor in the isle and then also captured by the neighbouring sensor, before the cut-off period of the first sensor is reached, then the person is considered to ignore the first item and consider the second one. Similarly, if the first isle is already under observation by a customer, the change in the area of focus helps in determining whether the next person is also interested in the first item or has moved on to the second one.

The order of the sensors can be considered in both ways, beginning from the ends to the middle of the isle.

The system of sensors behaves as a network in the sense that they communicate with neighbouring sensors via Bluetooth to make them semi-alert that they might be of focus next , as well as they communicate with the backend server to get the information of the optima cut-off time to consider the interest of a customer for an item.

Let time spent by customer i in front of item j be denoted by Tij.

The history of recorder times for each item and customer has been recorded by the server and whether that item was put in the cart by the customer or not.

Let Yij denote an indicator variable such that:

Yij = 1 ,if the customer i purchases the item j

= 0 ,otherwise

Then for the jth item a logistic regression is performed as follows:

Pij = Probability(Yij = 1)

log(Pij/(1-Pij)) =alpha + beta *Tij + eij

Pij = 1/(1+exp(-(alpha + beta *Tij))) + eij* ——(1)

where eij* ~ N(0,sigma)

Fitting a least squares solution by minimising the sum of squares of eij*'s, we obtain estimated values of the unknown parameters alpha and beta.

Using the estimated values of alpha and beta, we get predicted values of Pij from the observed Tij values in the equation (1).

Now let us start with any arbitrary cut-off value c0.

If predicted Pij>c0, then predicted Yij = 1. Now we define sensitivity = Proportion of times predicted Yij = actual Yij = 1

and specificity = Proportion of times predicted Yij = actual Yij = 0.

A plot of sensitivity or True Positive Rate and (1-specificity) or False Positive Rate is called the ROC curve. The Area under the ROC curve is defined as AUC.

Now AUC will be computed corresponding to all possible values of c0,(0<c0<1) and the optimal value of c0 = c1 is that for which the AUC is maximum.

In this example, the optimal value of c0 is corresponding to the blue curve as it has the maximum area under the curve(AUR).

Now substituting Pij by c1 and estimated values of alpha and beta, we get the optimal cut-off for Tij.

These optimal cut-off values are obtained for each item in the category and updated at a monthly interval.

Now once the $T_{ij} > c1$ for a given customer-item combination is observed, it indicates that the customer is interested in that item.

Calculation of the Viewed and Ultimately Bought Score:

Based on the past POS sales and customer data recorded from the server, we calculate the following:

$P_{b|a}$ = (number of customers who are interested in item b)/(number of customers who purchase item a)

$T_{b|a}$ = (average time spent by the customers,who purchase item a ,to consider item b)/(average time spent by the customers to consider item b)

Then the VUB score of item B relative to item A is given by $V(b|a) = P_{b|a} * T_{b|a}$

Best substitutes of item A are given by those with maximum values of $V(b|a)$

Calculation of Item Similarity score:

The list of items considered by customer i immediately after putting item A in the cart are taken into account.

The Also Bought Score of item B relative to item A = ABS(b|a) = (number of customers who add item B to the cart after adding item A)/(number of customers adding item A to cart)

The Also Viewed Score of item B relative to item A = AVS(b|a) = (number of customers who are interested in item B after adding item A to cart and before the next addition to cart)/(number of customers adding item A to cart)

Base Item similarity Score = BIS(b|a) = al*ABS(b|a) + AVS(b|a) , where al = (average number of items viewed by customers after purchasing item A)/(Average number of items purchased by customers after purchasing item A)

TIS(b|a) = (Average time spent by customers in viewing item B after purchasing item A)/(Average time spent by customers in viewing all items after purchasing item A before the next purchase)

Then Item Similarity Score of item B relative to item A ,I(b|a) = BIS(b|a)*TIS(b|a)

Best complements of item A are given by those with maximum values of I(b|a)

Item Attribution Analysis:

Using the VUB scores we obtain the list of best substitutes for each item.

Consider latent item attributes like price, weight, pack size, colour, flavour, nutrition index, etc depending upon the category. Now we run the following model: V(b|a) ~ l1,l2,…(Latent Attributes)

From the variable importance score of the Random Forest model, we get the most important distinguishable attributes L1,L2,L3,…

Now order the list of substitutes based on past sales. Comparing the important attributes L1,L2,.. between the highest and lowest selling items give the causal levels of the important attributes which cause the variation in item performance.

Item B1 is highest selling with the levels of attributes as L11,L21,… and Item B5 is lowest selling with attribute levels L15,L25,…

Low performing items can be substituted by those items which have high VUB scores and the causal levels of important attributes which cause high performance.

Then item B5 can be substituted by another item in the list say B7 with attribute levels L11,L21,L35,L45,…

If L1 = price,and remaining attributes are less important, then decreasing the price of the low performing item will lead to better sales.

Items with high similarity scores should be placed adjacently to increase the cross-sales. After computing the scores each month, the required changes in assortment and pricing are sent to the store managers.

Retail Shrinkage Management: During calculation of VUB scores, the server maintains a record of the items which were picked up from the shelves by the customers.

Similarly, the POS database also maintains a record of the number of purchases of the item. A correlation analysis of the number of items scanned (excluding return) in POS and number of items put in the cart by customers gives a idea of the amount of shrinkage from the store shelves, including theft and misplacement of items. Shelf regions can be identified where the correlation value is significantly lower than the average, which can be sent as an alert to store managers to decrease shrinkage.

Explanation of the Algorithm : Let us assume, for simplicity, that we have only five distinct products and a sensor mechanism (optical + weight) is placed at the center of each of the five sets (multiple facings) of products. Now the products are placed from the isle to the center of the shelf in order of 1,2,…5, so customers will have to cross these sensors in this mentioned order. Each of the sensors' status is monitored at each second. We have skipped the records where the entries remain unchanged. At initial state of store opening all the sensors are inactive except those on the isle which remain semi-active to detect the customers coming in. The sensor activation sequence helps in detecting the customer movements across the items. Every time the sensor goes ton active state; the sequence counter is incremented.

Now this feed is aggregated at each month on all the products to first calculate the cut-off time for an item to be considered by a customer as viewed. Once we calculate the initial cut-off times, we compute the VUB and Item similarity scores.

This helps us to get a better understanding of the customer behaviour and help in predictive replenishment of the shelves.

3.10. Fresh Production Demand Planning to reduce overproduction.

Forecasting demand for a particular item in each store can be a challenging task, as store and item level factors like location, customer demographics, weather, special events, socioeconomic factors and competition can all affect demand. Particularly, perishable fresh products have the added challenge of a short shelf life which can lead to increased waste if not properly managed.

This process begins by collecting historical demand data for the item in question at each store. This data is then used to select variables dynamically after going through feature engineering for every combination of store and item. After that, each combination goes through training and evaluation process for a suite of forecasting algorithms, including random forest, random forest with bootstrapped, moving average, Holt-Winters, and SARIMAX family models.

By dynamically selecting best model with highest accuracy in validation set for each store item combination can help to ensure that the solution is able to adapt to changing demand patterns over time and to take advantage of the strengths of different algorithms in different situations. This leads to more accurate and reliable forecasts, which in turn help to optimize production planning by reducing shrinkage.

Two challenges that arose during the implementation of the approach were dealing with the imputation of missing dates in the data and deciding on the appropriate back-testing method.

This approach uses walk-forward validation as an out-of-sample performance evaluation method. This involves dividing the data into training and testing sets, training the forecasting model on the training set, and evaluating its performance on the testing set. Repeating this process multiple times, comparing models' accuracy over time to determine the best model. The model that performs the best on most weeks is selected to use in the following week which can be different for different store item combinations. Imputing missing dates in forecasting solutions is important because missing data can introduce biases and errors into the forecast.

In today's fast-paced retail industry, being able to accurately forecast demand for fresh products is crucial for success in the retail market to thrive in the tremendous competition. Getting accurate forecasts is therefore more important in order to maintain the tradeoff of improving the user experience and minimizing the waste of the perishable products to save the cost of production and managing supply accordingly. With around 6000 distinct items across 2500 stores worldwide, the task of finding generalized solution can seem daunting. However, by utilizing a fine-tuned PySpark implementation, along with a variety of other techniques, this implementation was able to achieve significant improvements in forecasting accuracy, while keeping computing expenses low at less than $200 dollars daily. The accuracy on average is 85%, with several items achieving over 90% accuracy.

3.11. Causal Impact Measurement Framework in Retail Stores

In any technology driven industry, launch of a new business or new product features for an existing business to customers requires rigorous testing with certain confidence to ensure that the new feature will be beneficial for the business. One way to test is A/B testing. But a major requirement of A/B testing, is that the treatment should be randomly assigned. Now, the question is what if the treatment assignment is not random? What if, the control entities behave differently based on other factors? The answer to this is using a framework which uses Causal Inference to measure the impact of the new feature added.

3.11.1. Problem Overview

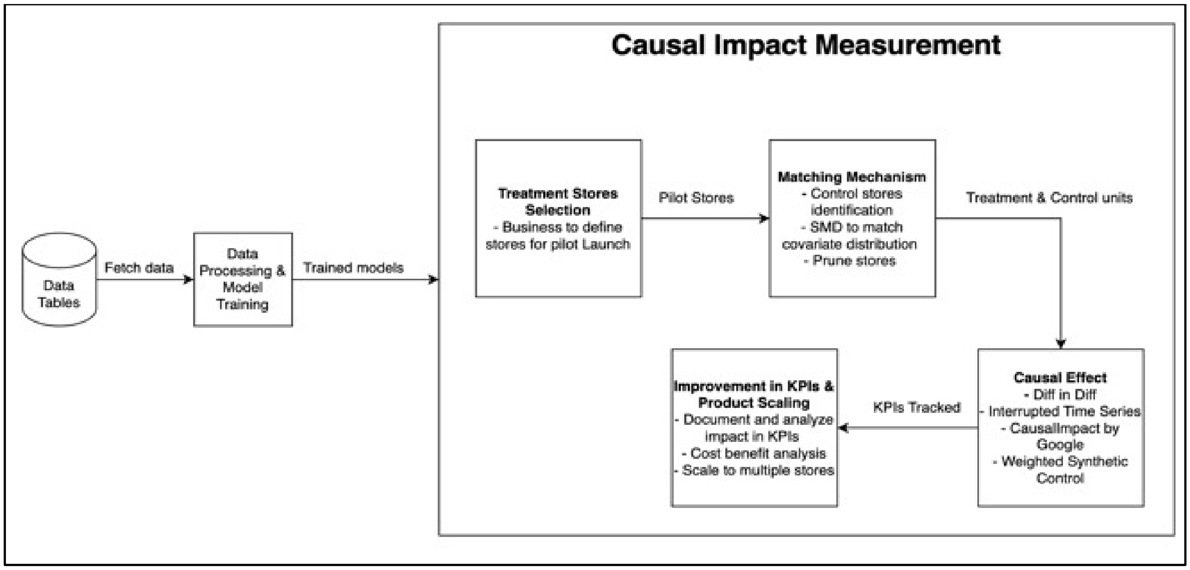


Figure 12. Proposed Causal Inference Framework

Measuring the improvement in KPIs due to a particular change is useful to decide whether i) change has produced the desired effect and whether the incremental benefit outweighs the cost to bring in the change at scale. ii) whether the effect generated is reliable (significant) and did not happen by chance. As highlighted in the previous section, randomized controlled trails cannot always be used to measure the impact in all cases, and we resort to observational techniques such as Causal ML. To be precise, we want to measure the effect due to an intervention (D) (ISA 2.0 rollout which is an ML driven inventory adjustment system) in a particular unit (store/item) i, i.e., D_i=1. Let $Y_{i.}$ be the observed outcome variable for unit i. Because we can only measure the potential outcome as we cannot observe the same unit with and without treatment, our individual treatment effect can be written as Y_{1i}-Y_{0i} where Y_{1i},Y_{0i} are the potential outcome for unit i with and without the treatment, respectively. In practice, we measure the Average Treatment Effect (ATE) and Conditional Average Treatment Effect (CATE) on the treated, defined as [1]

$$ATE = E[Y_1 - Y_0]$$

$$CATE = E[Y_1|X] - E[Y_0|X]$$

Multiple techniques (*see section 4.1 for more details)* are evaluated by conducting a comparison study of the store similarity results obtained using ~30 input attributes. Improvement in the KPI metrics between treatment and control units before and after the intervention (change) are measured through various techniques, each of them offering a different approach to calculate the impact with own set of assumptions, advantages, limitations, and situations where one is more ideal than the others. Our goal here is to analyze the results and provide insights into how, what, and why the different approaches worked or did not work for an ML solution that we plan to roll out in the Store Inventory Management initiative.

3.11.2. Methodologies

Estimating the effect of the intervention can be seen as a two-stage process. 1. To design and select the treatment and control units 2. Analyzing the outcome on the selected units.

Matching

Matching is a pre-processing step on the observational data for treatment and control effect estimation, as shown in Figure 13. It is done to avoid confounding bias by controlling for all common causes of the treatment and outcome.

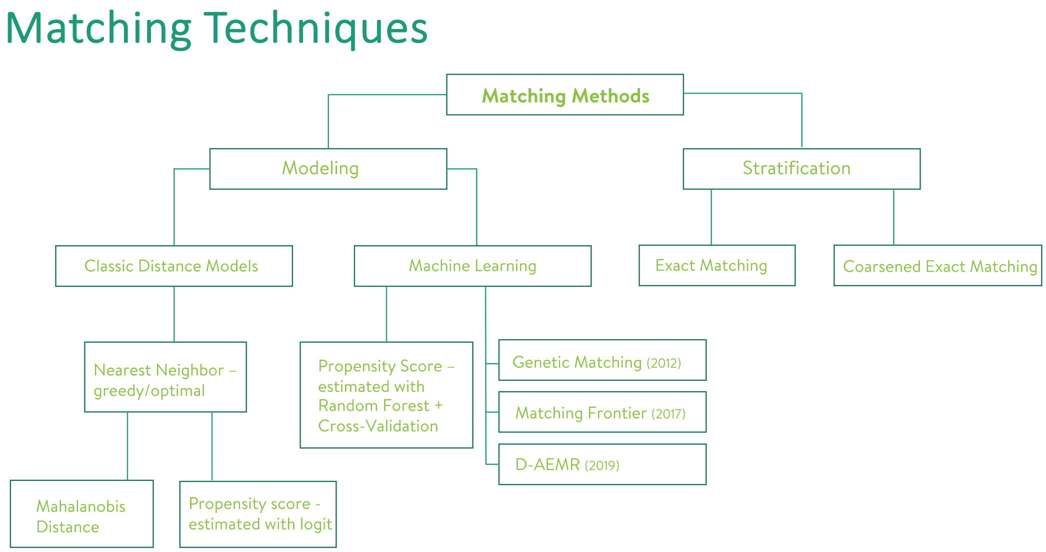


Figure 13. Various approaches for experimental impact measurement

This ensures our treatment and control units are fundamentally similar so that our ATE (Average Treatment Effect) observed is reliable. It becomes even more important to strategically select the units using matching techniques to balance the observed covariates distribution between treatment and control groups when we deal with a small sample size as RCTs (Randomized Control Trials) aren't effective.

Covariates are chosen to satisfy the assumption of ignorance by starting with a few covariates to match on and include additional ones that remain unbalanced after the matching. Various distance measures such as Mahalanobis Distance, Propensity scores are used to measure similarity of units. Later, the best set of similar units are chosen by minimizing the Standardized Mean Difference (SMD) of covariates between treatment and control units on average. We prune out any units with a caliper to avoid poor matches and perform a sensitivity analysis to minimize any hidden biases in the selection process. Once matched, difference in means in the outcome between treatment and control units becomes the unbiased estimate for ATE.

Causal Impact Measurement

Causal Inference effectively involves measuring the impact of a certain intervention, in terms of a KPI. KPIs can range from minor changes in personal habits to major changes in Sales patterns in retail giants.

Different impact measurement frameworks are relevant in different intervention scenarios. For example, if the intervention is a statistical experiment where test and control are divided at random without the knowledge of similarity, traditional A-B Testing, Bayesian Statistics, ANOVA etc. help measure the impact but if the setup is a quasi-experiment, then Differences-in-Differences, Matching, Regression Discontinuity help measure the impact. Currently, there are at least a dozen different methodologies to measure impact of an intervention through a Causal Inference framework, but this paper will concentrate on the ones relevant to the business problem and the data structure availability; namely Differences-in-Differences [1], Interrupted Time series [3], CausalImpact by Google [4,5] and Synthetic Control [6]

Differences-in-Differences

Diff-in-diff is an approach that has been used in longitudinal data to assess the effect of macro interventions that supposedly have a larger effect on the general population, like effect of immigration on unemployment, effect of smoking ban on public health, or as in this case effect of switching from a rule based engine to a Machine Learning engine on the inventory management system (especially recommended down adjustments for out of stock items).

Let D denote a treatment flag where 0 means the entity is not treated and 1 means it is. Let T denote time where 0 means pre-intervention and 1 means post-intervention. So, $Y_D(T)$ is the potential outcome for treatment D on period T. Ideally, the causal effect is the difference between the treatment effect after intervention and the same effect before intervention [1]

Eq 1: $ATE = E[Y(1)|D = 1] - E[Y(0)|D = 1]$

But this doesn't take into consideration any counterfactual treatment effects and any latent "incremental" lift that proves the effectiveness of the intervention. Hence, the classic estimator finds out the incremental lift between treatment and control, when difference between after and before intervention numbers are considered. Or,

Eq 2:

$$CATE = (E[Y(1)|D = 1] - E[Y(1)|D = 0]) - (E[Y(0)|D = 1] - E[Y(0)|D = 0])$$

This equation assumes that the treatment and control entities have a very similar trend for the measured KPI.

The next step is to measure the statistical significance of the CATE or Conditional Average Treatment Effect, which is done using a simple yet neat regression technique.

Eq 3:

$$Y_i = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 D * T + e_i$$

The coefficient of β_3 is same as the output of Eq 2 and the regression gives us its statistical significance in terms of p-value. Considering 95% confidence interval, if p-value is less than 0.05, the incremental lift β_3 is significant else the intervention was not significant.

Interrupted Time Series (ITS)

According to American Psychology Association, Interrupted Time Series (ITS) design is a quasi-experimental design in which the effects of an intervention are evaluated by comparing outcome measures obtained at several time intervals before and several time intervals after the intervention was introduced. Unlike traditional time-series designs, which make use of a continuous predictor variable, an interrupted time-series design uses a categorical predictor—the absence or presence of an intervention. [3]

ARIMA (or Seasonal ARIMA) is leveraged in this methodology to evaluate the significance of the said categorical predictor in the fitting of the model. A general ITS has the KPI as the dependent variable and three independent variables: a categorical intervention flag, a date variable (days since, week etc.) and a variable implying time since the intervention started. Autocorrelation and partial autocorrelation plots are used to ascertain the exact or range of values for AR (p, 0 to 6), and MA (q, 0 to 2) with a differencing value (d, 0 to 1) obtained from a significant Augmented Dickey-Fuller (ADF) test. A set of time series is run in a loop over the above range of values and the model with the lowest AIC (or MSE or RMSE) is marked as the trademark model. The significance of the categorical intervention predictor determines if the intervention was statistically successful, and the coefficient determines the extent or percentage change in the KPI due to the intervention.

Causal Impact

Brodersen et al., (2015), in his paper "Inferring Causal impact using Bayesian structural time-series models" [5], defines this method as a state-space model that predicts the counterfactual market response in a synthetic control that would have occurred had no intervention taken place. In contrast to classical difference-in-differences schemes, state-space models make it possible to infer the temporal evolution of attributable impact, incorporate empirical priors on the parameters in a fully Bayesian treatment and flexibly accommodate multiple sources of variation, including local trends, seasonality, and the time-varying influence of contemporaneous covariates.[4]

In this method, data is divided into two parts: the first part is the "pre-intervention" period, and the concept of Bayesian Structural Time Series is used to fit a model that best explains what has been observed. The fitted model is used in the second part of data ("post-intervention" period) to forecast what the response would look like had the intervention not taken place. The inferences are based on the differences between observed responses to the predicted one which yields the absolute and relative expected effect the intervention caused on data.

Synthetic Control

As the name suggests, synthetic control [6] is a weighted average of the actual control units and gives control on 1. relative contribution of each control unit to the counterfactual of interest 2. Similarities (or lack thereof) between the unit affected by the intervention and the synthetic control, in terms of preintervention outcomes and other predictors of postintervention outcomes. Synthetic Control can be handy in a brick-and-mortar setting where finding a similar control and treatment set of stores can be almost impossible due to lots of external factors that can't be controlled.

In this approach, the model extends the traditional linear panel data (diff-in-diff) framework, allowing that the effect of the confounding variables on the outcome vary over time. Weights to get the synthetic control from actual control units are derived by experimenting with various approaches such as a basic weight optimizer function, LASSO & Gradient Boosting Regression (GBR).

3.11.3. Experimentation Framework

The proposed experimentation framework can be used to perform observational inference for both customers and stores. A major benefit of having an offline experimentation platform is that it encapsulates the causal ML methods within it and takes away the requirement of manually running such experiments. In turn, this benefits democratization of experimentation to a wider business audience such as Product leaders, business leaders and analytical teams to turn ideas into experiments using plug and play without having to worry about technical details and get quick and reliable results. It has the following core components:

1. **Web UI:** To configure experiments, define hypothesis and showcase results with a verdict on rollout.
2. **Backend Services:** To fetch, process data required for the experiment and run Causal inference analysis. Figure 4 shows the data flow.
3. **Log, Maintain and Monitor:** The platform provides the necessary support for obtaining information on past and ongoing experiments and is seamlessly extendable for future needs.

Figure 14. The data flow within the experimentation framework.

Figure 14 shows the data flow in the CI experimentation framework. For the experimentation framework, typically the following information would be required:

1. Store or Customer level Features Data
2. Historical Experiment Data
3. Current KPIs

3.11.4. High-level design of Experimentation Framework

The high-level architecture for the experimentation framework uses Causal Inference based modeling and service from observational data, i.e., without having to set up an A/B test, which has been showcased in Fig. 15.

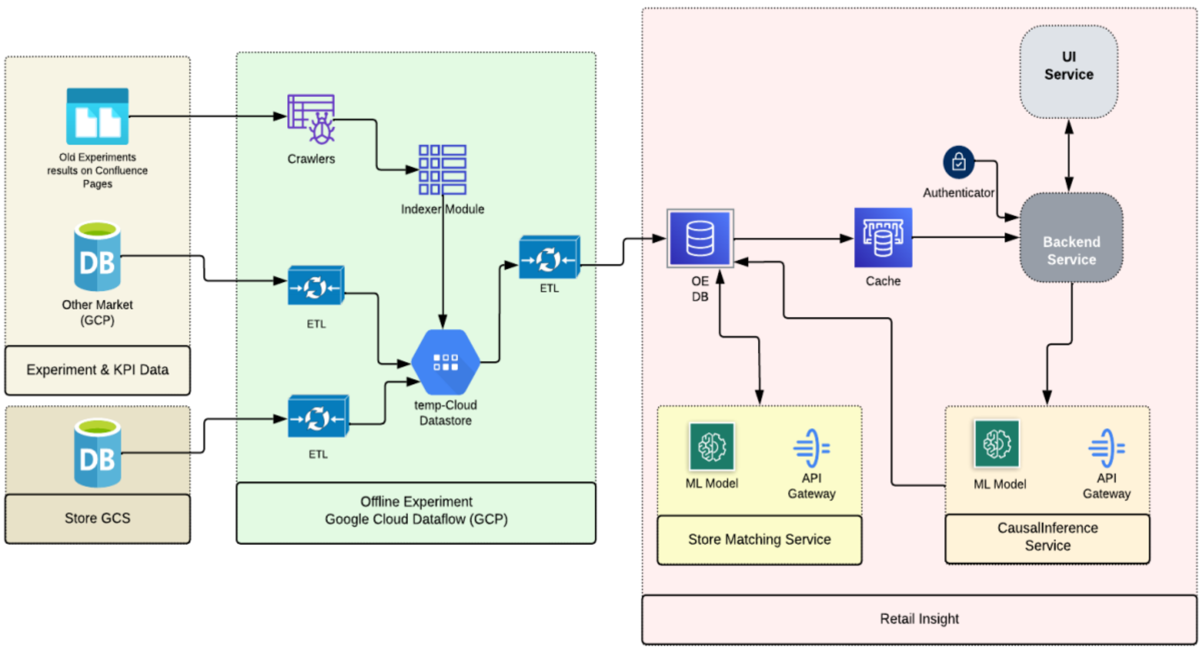


Figure 15: Experimentation Platform Architecture.

3.11.5. User-interface of Experimentation Framework

The user interface of the experimentation framework can comprise of three modules: The Landing Page, The Experiment Setup, and Result Readout.

The Landing Page

The Landing page (Figure 16) module allows users to view a list of existing experiments, including details such as experiment name, status, theme, results, and available actions.

Experiments can be in various stages, including ongoing, completed, yet to start, or in draft stage. For completed experiments, users can observe the experiment verdict, which can be categorized as Full rollout, partial rollout, or no rollout (Figure 8).

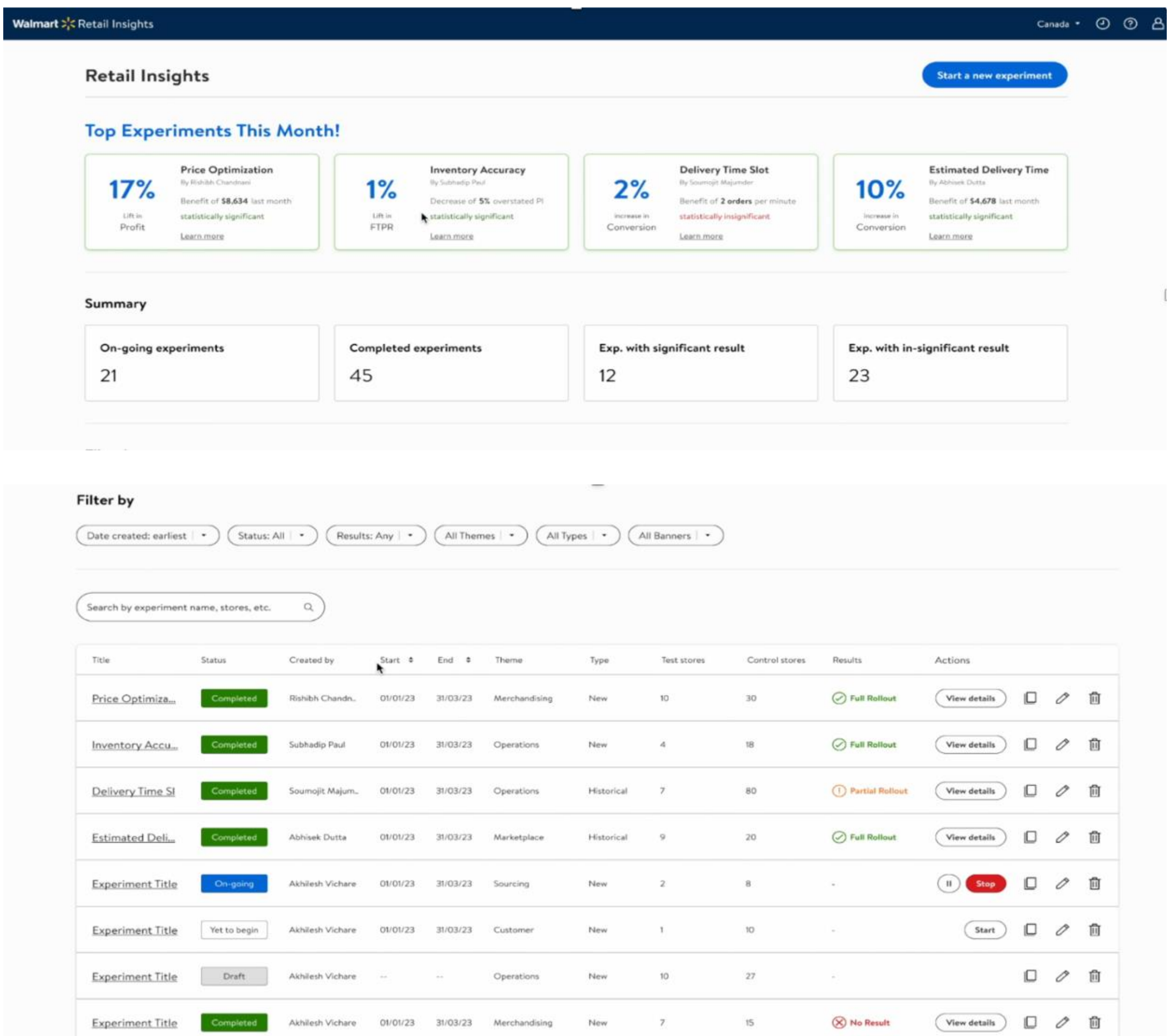
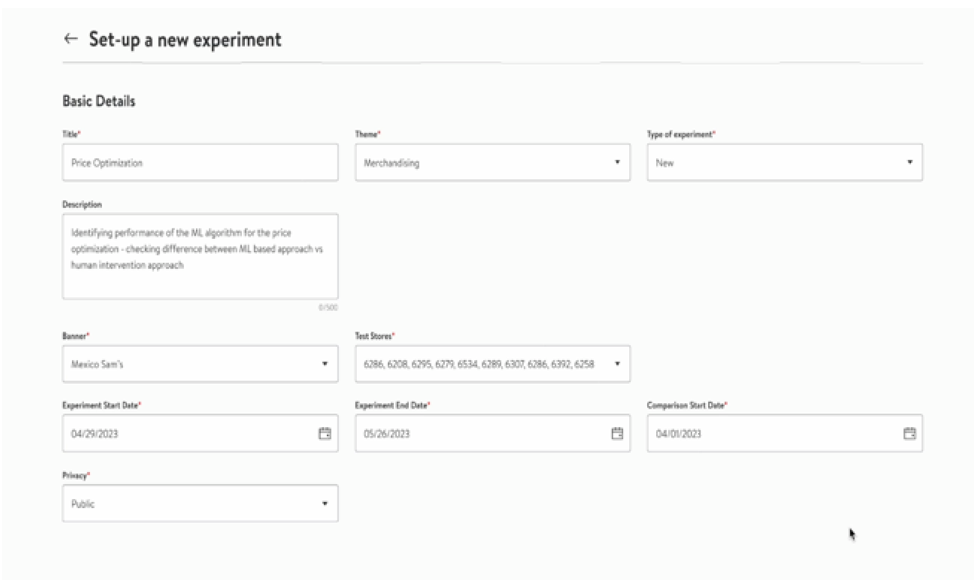
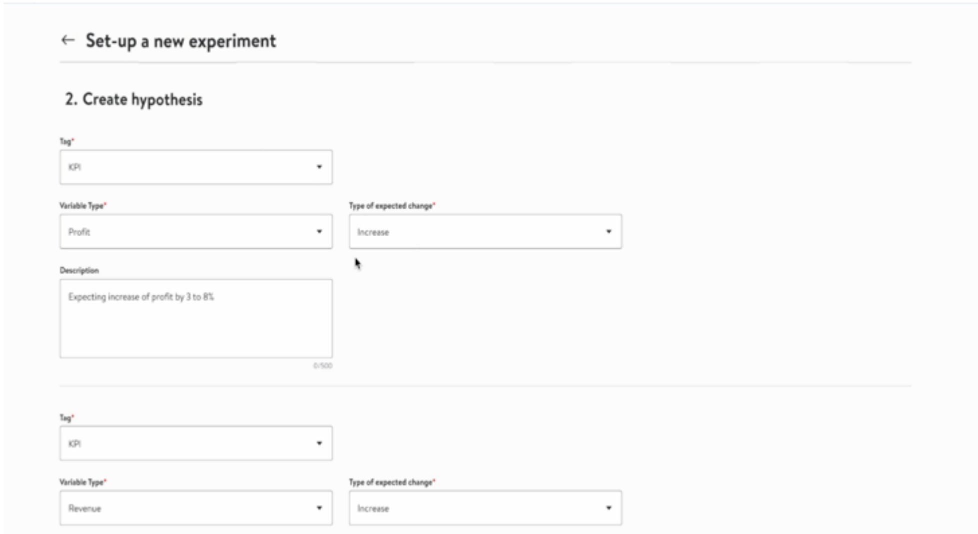


Figure 16: The screen for displaying relevant insights and overall usages and statistics pertaining to the platform itself.

The Experiment Setup

The Experiment Setup module provides users with a simple and intuitive way to configure experiments without requiring any technical expertise (Figure 17). The setup process consists of four steps: basic details, hypothesis creation, action setup, and selection of control stores.




← Set-up a new experiment
2. Create hypothesis
Tag*
KPI
Variable Type*
Profit
Type of expected change*
Increase
Description
Expecting increase of profit by 3 to 8%
0/500
Tag*
KPI
Variable Type*
Revenue
Type of expected change*
Increase

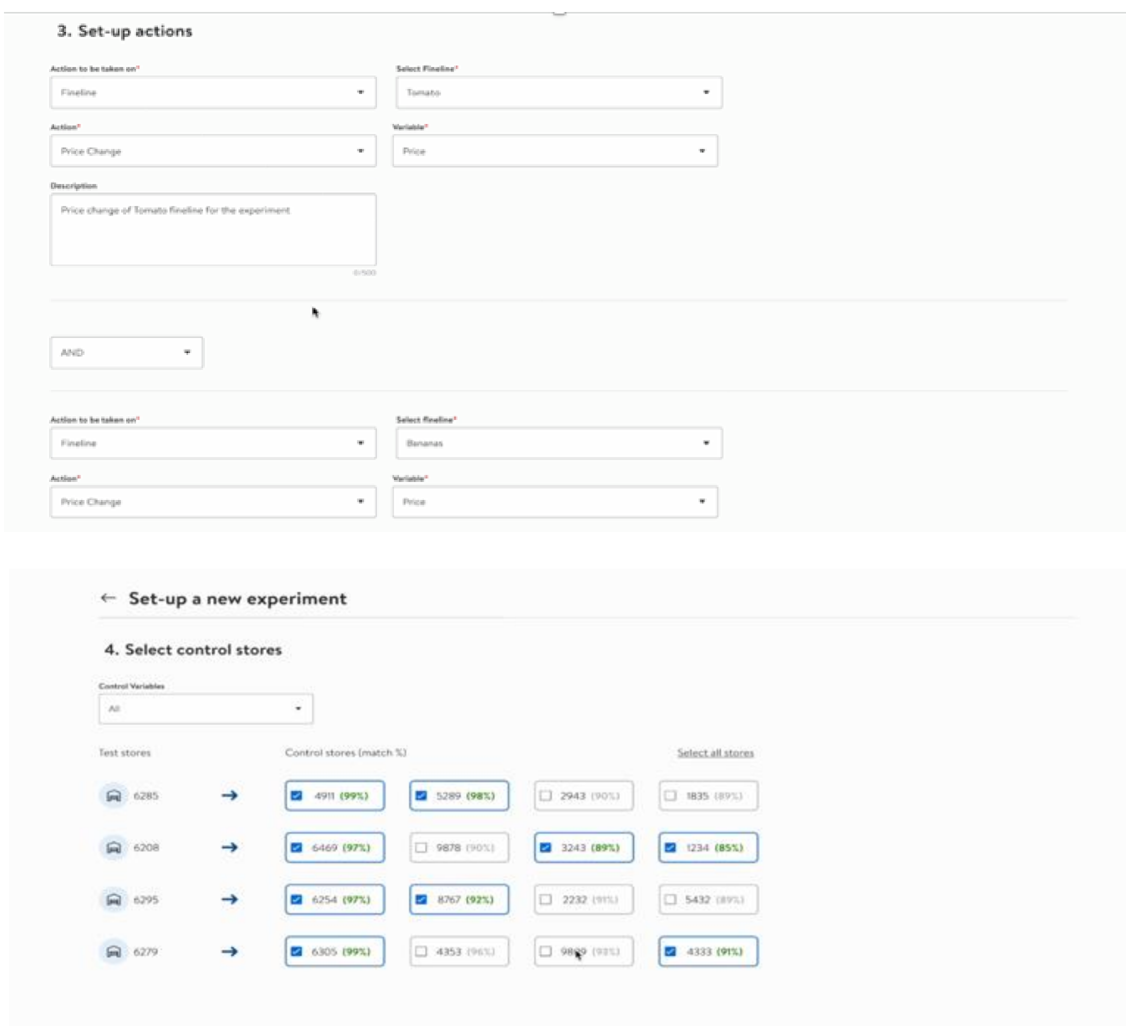


Figure 17: The experiment set up screen is used to set up experiment where the user provides the necessary details regarding the experiments.

Result Readout

The experiment Result Readout module enables users to comprehensively review the outcomes of their experiments (Figure 18). Users have the option to delve into detailed analyses by viewing trend charts of individual KPIs and Guardrail metrics (Figure 19).

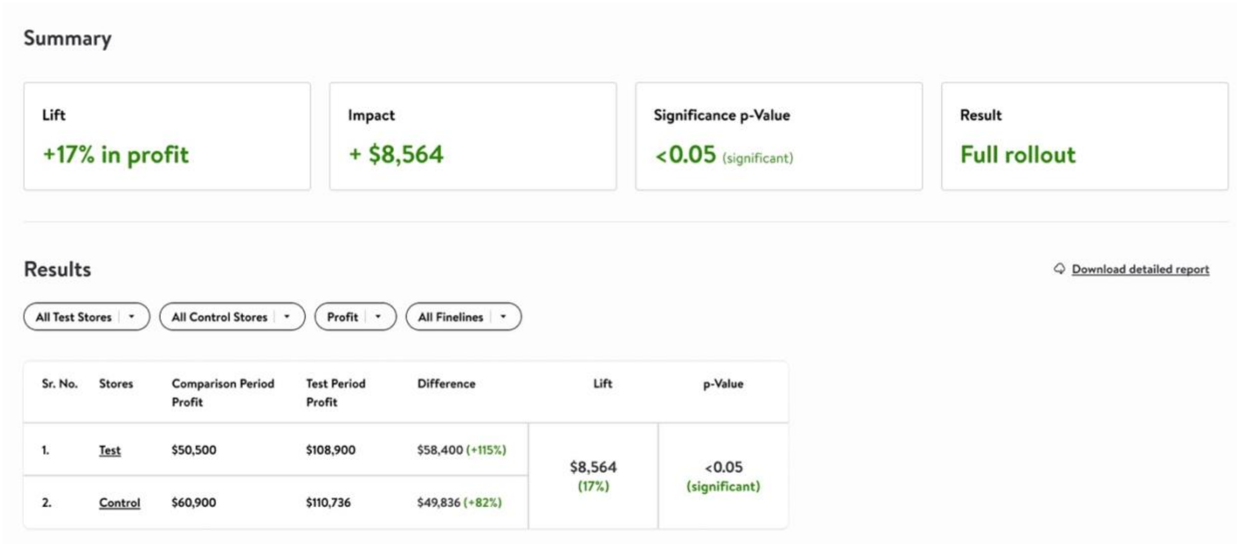


Figure 18: Output screen for reading the results of an experiment (CI run) and inferences.

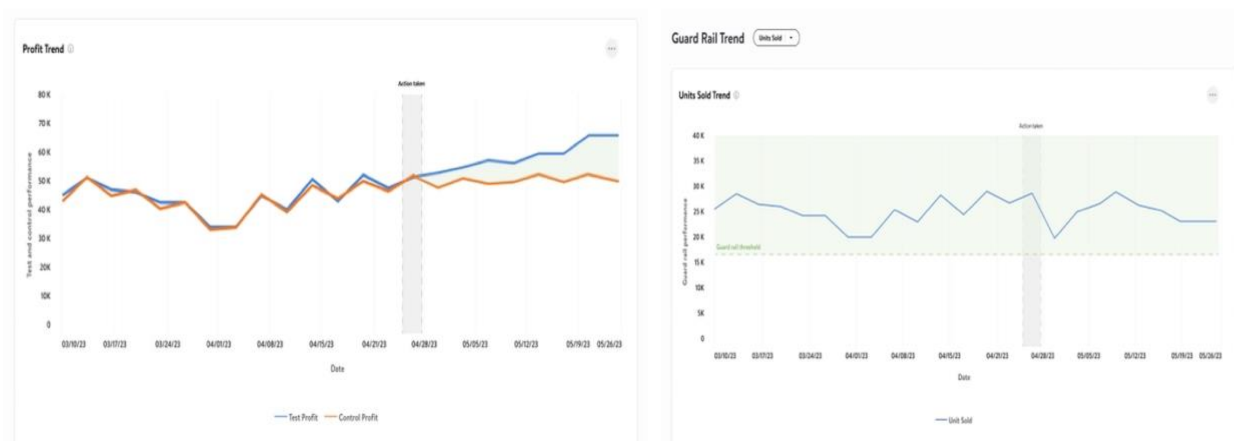


Figure 19: Change in the KPIs and guardrail metrics over the experimentation period.

3.11.6. Benefits

The Experimentation Platform can be used to run hundreds of experiments every single day since many experiments span from weeks to months and need continuous monitoring. Typically, the recommendation around full, partial or no rollout can be taken by considering the results coming out of the experimentation methodologies chosen supplemented by the recommendation offered by the platform.

Many companies like Netflix, Amazon, Microsoft, Meta, Booking.com run over thousands of experiments at any given moment. As companies grow in scale, the size of experiments should also grow to find truly impactful solutions. Experimentation, done right, can help in sorting the gold from the dross. There are multiple examples of companies who have used experimentation to accelerate growth around 5%-15%.

CHAPTER IV:

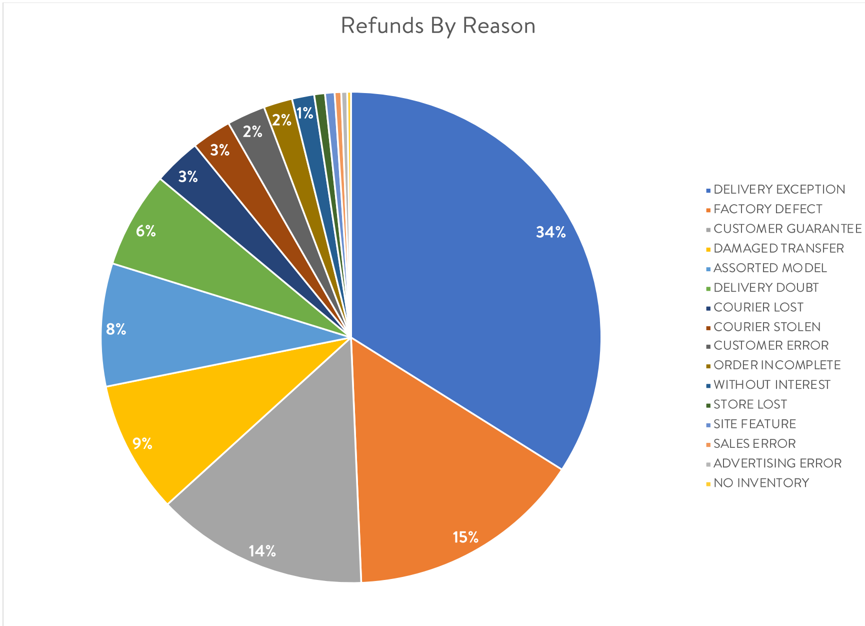
RESULTS

4.1 How Machine Learning can help in eliminating gaps between offline and online channels which lead to fraudulent activities?

The suite of ML models mentioned in the methodology section has been implemented in e-Commerce Total Loss Prevention. product which is an omni-channel fraud prevention system for all International Markets. The risky cases are highlighted in the eTLP tool in various sections like Customer Channel, Map View, Case Search section, which highlights risky customers, collusions, and process opportunities to the Asset Protection analysts. It showcases the risky customers, satisfying the anomaly detection thresholds, along with possible linked accounts, risk score, and interpretable reasons for being called out risky. The analysts can click on any customer to get details of risky locations, items, refunds, cancellations, and sales patterns over time, linked customer details. The analysts can also download the transaction details for building a case, and they can provide risk feedback, which would be used as training data for retraining our models. Currently this is in production for UK, Canada, and Mexico markets, and already has generated multi-million-dollar savings by identifying fraudulent cases.

The study has been conducted using both qualitative and quantitative methods where we can understand from the risk analysts what are the key pain points in the fraud detection and prevention journey, and also using the feedback from the recommended risky cases, develop a quantitative measurement of the benefits that can be reaped using the machine learning techniques.

Analysis of returns and cancellations have been performed to identify the total volume of loss, which has been highlighted in Figure 20.



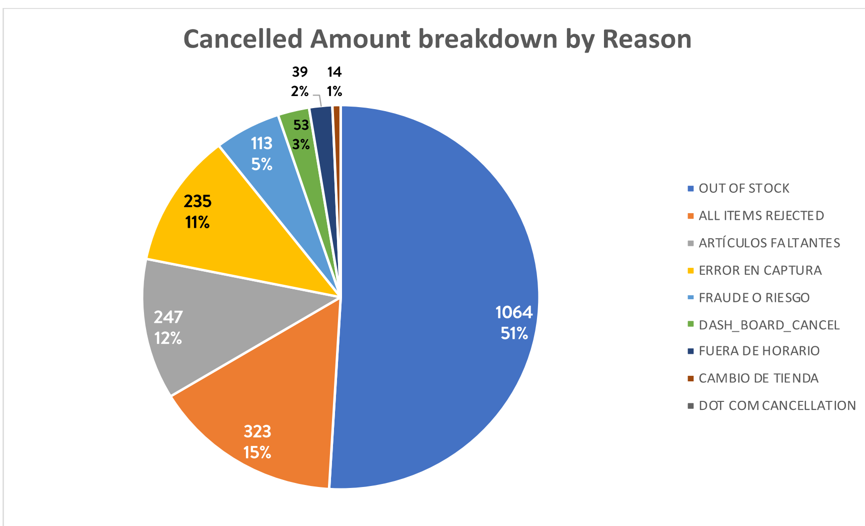


Figure 20 : Breakdown of Refunds and Cancellations by reasons

This illustrates that a huge portion of online returns happen due to Delivery exception and defective or damaged items being reported, which indicates potential collusion between customers and delivery drivers and store associates. Similarly, items being rejected or out-of-stock being reasons for cancellation can be traced to collusion with store associates.

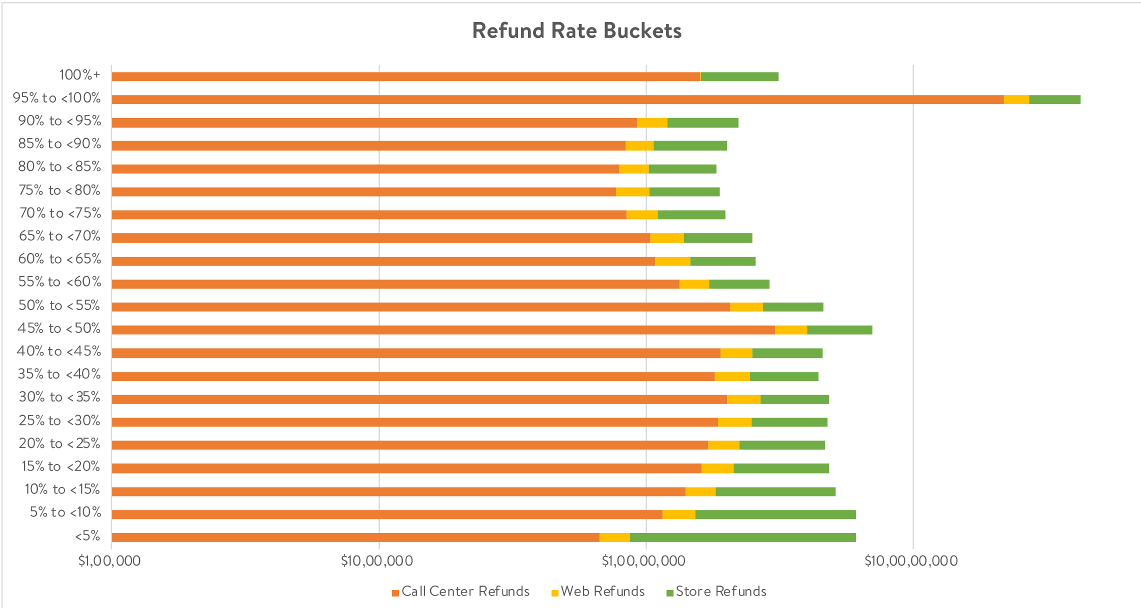


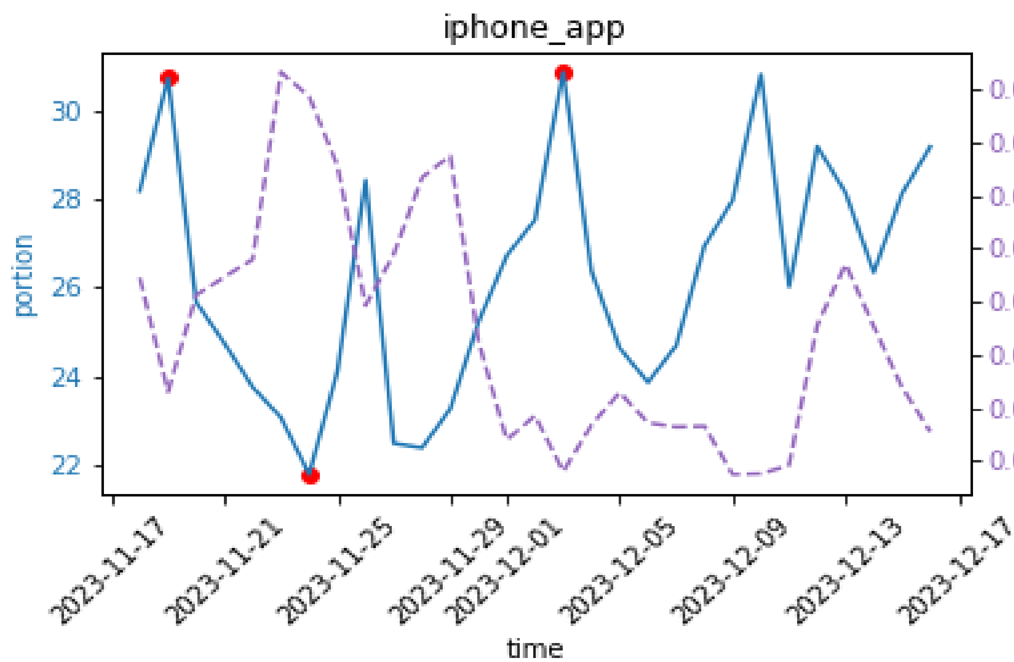
Figure 21 : Refunds by channels across different Refund-rate buckets

The refunds have been grouped by refund rate buckets, as shown in Figure 21, to identify that a huge proportion (around 56%) of total refunds happen with customers who return more than 95% of the orders that they purchase, indicating the pattern of repeated abuse customers.

Similar analysis reveals that Electronics department and items like “Macbook Air” have the highest chances of being refunded fraudulently. Among the refunds, a deep-dive into return reasons help to identify that more that 40% claim that items are missing, 31% claim damaged, and 11% claim that they got an unwanted substitute. These cases help to provide insight into the abuse patterns used in the fraudulent returns.

The Machine Learning solutions for Fraud prevention have been implemented in several phases. The first stage is using the Unsupervised Anomaly detection model.

When no or very minimal information is available about past fraudulent cases, this process helps to kick-start the fraud prevention process. Using multi-variate anomaly detection techniques like Isolation Forest and Time series-based anomaly detection, the system successfully detected several anomalous scenarios, which have been highlighted in Figure 22.



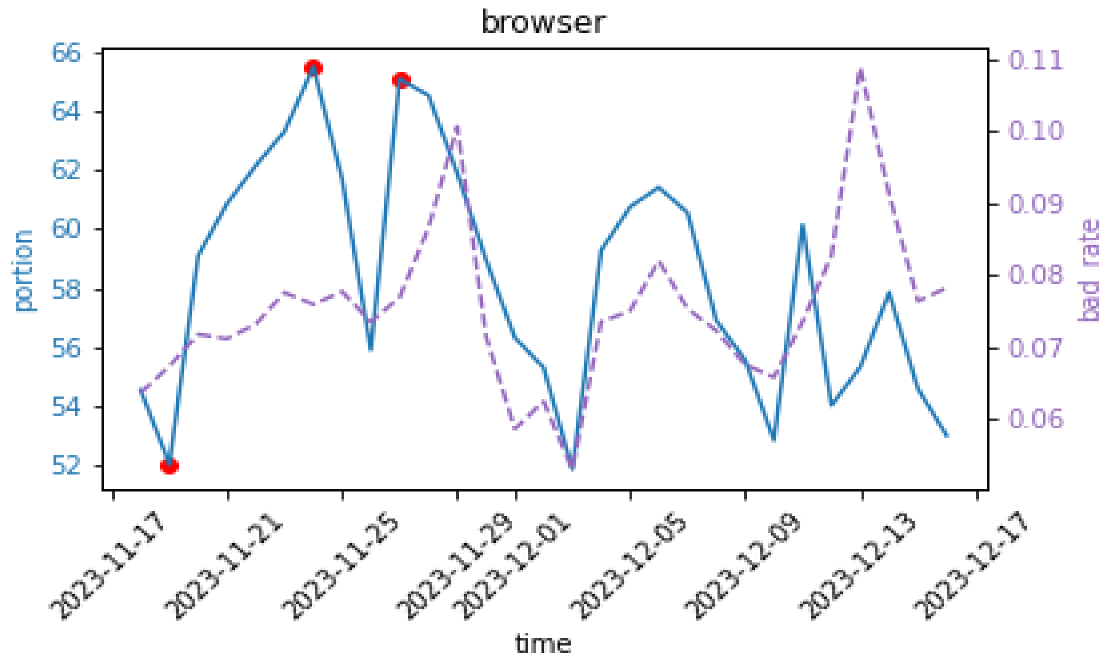


Figure 22: Anomalous transactions detected at Mobile and Web applications.

These anomalous cases have been shared with the business specialists on Loss prevention for further analysis. The challenges with the unsupervised techniques lie in a huge number of false positives, but this can be gradually resolved through user-feedback and introduction of more sophisticated supervised models. Based on feedback received from the end-users, several supervised and semi-supervised techniques have been applied for determination of a multivariate customer risk score, which takes into account several signals of fraudulent activities, and provides a holistic view of propensity of fraud for a customer. In the methodology section, we have proposed a unique combination of weighted average of unsupervised risk score based on precision and class proportion from an ensembled supervised risk score, where the weights are dynamically selected based on a validation dataset. The train data consisted of 6M cases, and it has been divided into 4M from training the model, 1 M for validation and hyper-parameter tuning and remaining 1M test data on which the results have been compared in Table 2.

Parameters	Decision Tree	MLP	kNN	SVM	xGBoost	LightGBM	Proposed ensemble semi-supervised model
Accuracy	75%	74%	77%	85%	88%	84%	91%
False Positive Rate	65%	67%	62%	71%	43%	31%	14%
False Negative rate	7%	12%	11%	9%	10%	5%	6%
F1 Score	54%	45%	56%	67%	72%	76%	82%

Table 2 : This table compares results of several supervised ML models and proposed ensemble model using both supervised and unsupervised techniques

Next we apply graph based techniques of linking different customer account to create a complete picture of multiple customers colluding with each other to create a ring of fraudsters, using a priority order of linkages, which has been highlighted in Figure 23.

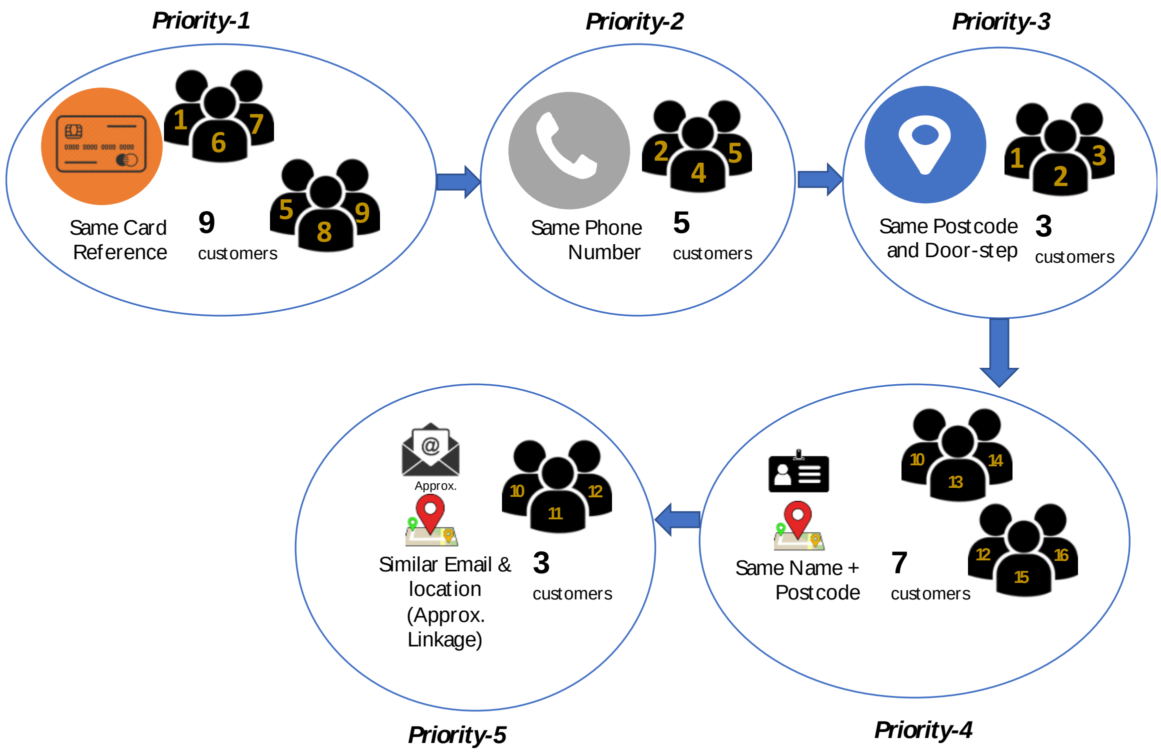


Figure 23: Priority based connected components used to create customer linkages.

The customer-pairs have been trained with both positive and negative examples of linkages using a Graph-based Deep learning model, and the results show around 0.864 AUC value representing the good performance of the trained model on testing phase, as highlighted in the Figure 24.

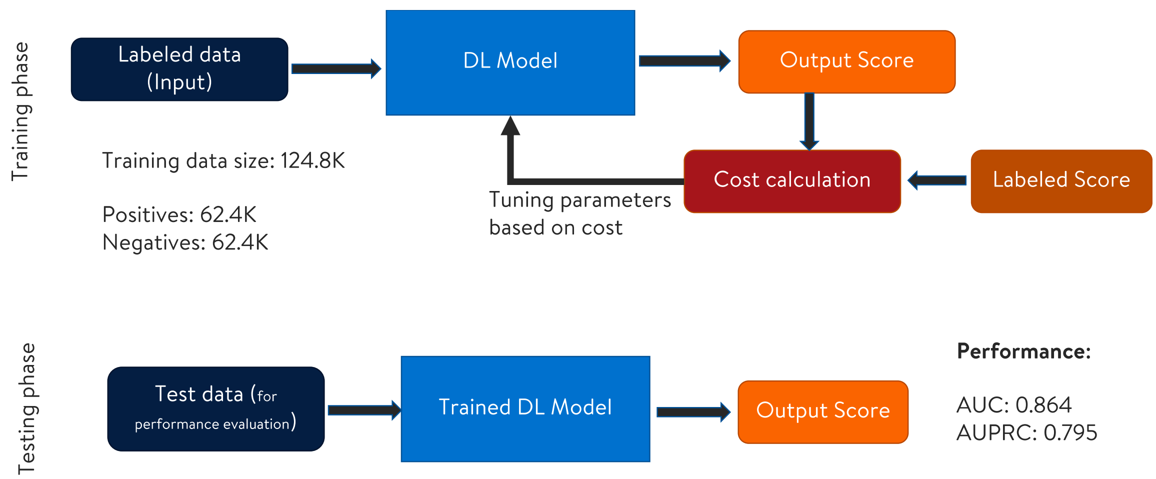


Figure 24: Performance of Graph Neural Network model.

At the next stage, we implemented the Active Learning technique of identifying the most crucial cases to be sent to the analysts for identification of fraud, in a human-in-the-loop fashion, to drastically improve the convergence rate of the underlying model.

Before this solution was implemented, Asset Protection associates achieved 13% hit-rate for recommended customers, i.e. 76 out 569 potentially risky customers identified on an average daily, were found to be actionable. Application of this innovative solution led to an incremental 27% lift in performance of recommendation system in just 5 weeks and just 75 additional feedback from human annotator is translated into $700k annually from identifying risky customers.

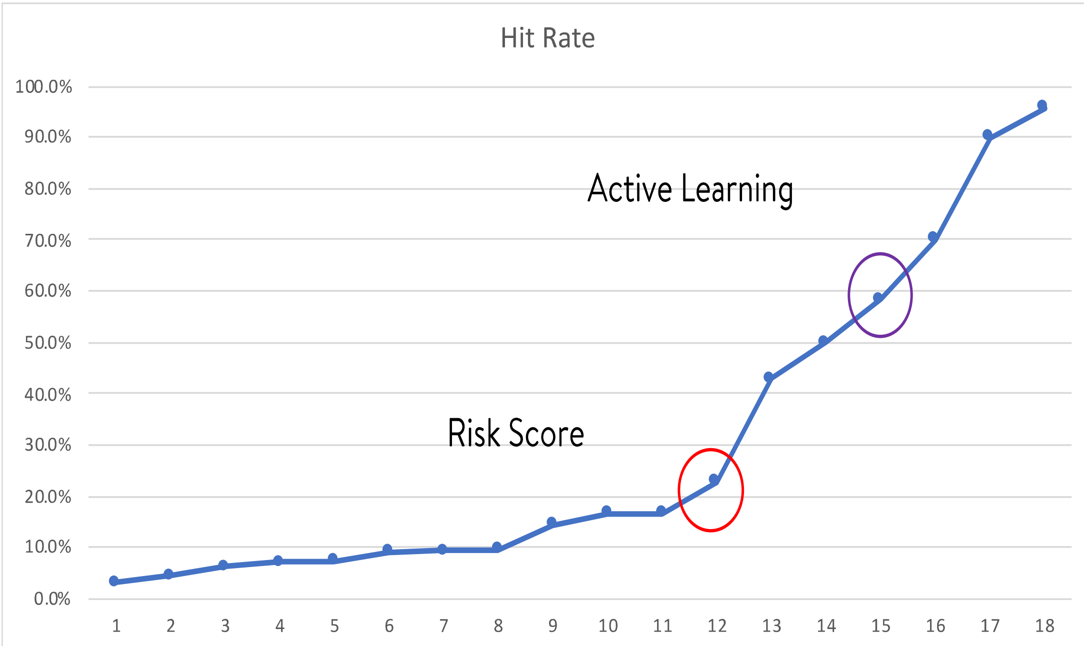


Figure 25. The growth of hit rate metric over the weeks

The Figure 25 above showcases how the hit rate of the fraud engine has increased significantly. Initially with only unsupervised learning it was close to 13%, and after implementation of risk score model increased to around 60%, but after the active learning-based recommendation system was implemented, the hit rate increased to almost 98% in a very short time. This proves that sampling rates from concurrent-arm Beta sampling balance both exploration and exploitation and converge to stable number of fewer iterations of learning loop making Classification Model improve faster and reach peak performance. This is equivalent of putting an Active Learning engine before recommending to human annotator.

4.2 What are the major causes of Retail Waste and how prescriptive Machine Learning recommendations can help to reduce waste?

Retail waste has been identified as a major concern across the world. We have implemented a multi-stage technique starting with descriptive and moving to predictive and prescriptive Machine Learning techniques to identify which stores and items are the highest contributors to waste, identified the leading factors for waste and then provided prescriptive recommendations to reduce waste.

The multi-variate anomaly detection models of this waste management system help by masking irrelevant or less important store and items and thus reduces the size of data to be analysed by more than 95%. This makes the waste management study more precise and allow business users to focus directly on the pain points. We have observed the KPI values for different stores and analysed that by multi-variate analysis we can focus on more relevant stores as compared to stores observed by univariate analysis (in this case absolute waste) which saves time and effort for the business.

Techniques like change point detection provided further streamline direction to waste management study by uncovering lurking variables which are used for causal discovery analysis.

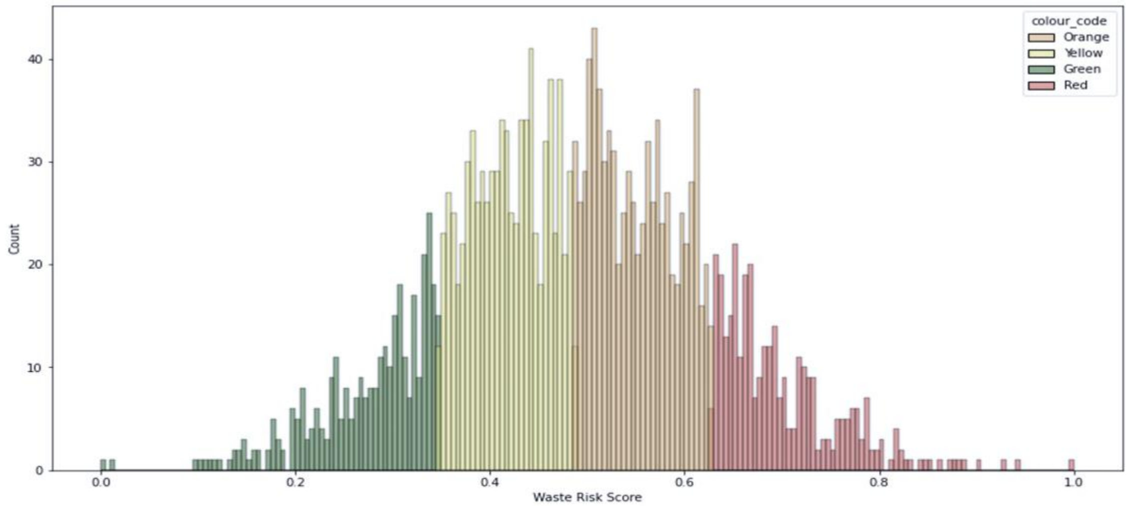


Figure 26. Waste Risk score-based percentile distribution of stores

The multi-variate waste risk score has been used to group the stores into different zones of importance as shown in Figure 26. Here the red zone has the highest waste risk score and needs to be investigated more diligently to reduce waste.

Comparison of multi-variate waste risk score with individual waste KPIs like total waste, CVP sales, etc help to identify the most prioritized cases to be looked into, as shown in Table 3, to highlight that store 5678 deserved more priority in terms of waste reduction than store 1234.

	Store 1234	Store 5678
Total Waste	$ 2.7 M	$ 380 K
Waste to Sales Ratio	9.8%	20.7%

CVP sales to Waste Ratio	0.9%	0%
Throwaway	49.8%	72.3%
Waste Risk Score	93%	100%

Table 3: Waste Risk score helps to prioritize important cases.

At the next stage, we identify the different potential factors that can affect the waste. The Causal discovery models have been applied to find the inter-relationships between the different factors, and feature importance from the bootstrapped Causal Forest model provides insights into which factors have most impact for driving waste. In Figure 27., we highlight the feature importance of the factors which have the most impact towards waste across all stores. The direction of the feature represents whether it impacts waste in a positive or negative manner, and the magnitude represents the impact of change. This can also be performed at each store level using the local importance factors using SHAP values.

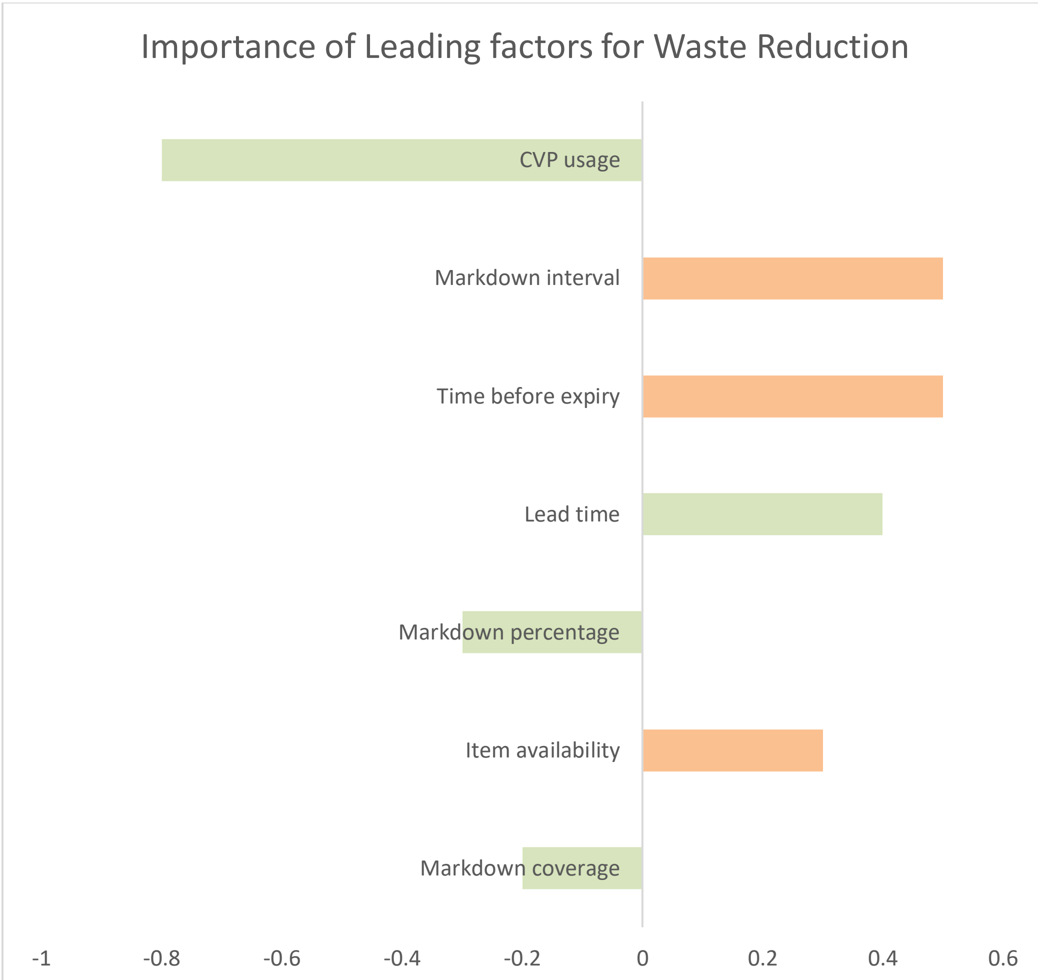


Figure 27: Causal Forest feature importance for Waste drivers

Precriptive recommendations are provided to the store associates in form of guided insights and actions in an embedded associate application. These also provide the Lost profit information if they are not acting upon these recommendations, as shown in Table 4. This highlights the top critical insights and recommendations privided for different types of items, regions, store and time level

SBU	Aggregation Level	Store	Market	Region	Date	Time Interval	Insight	Current Time Interval Total Loss (in $)	Current Time Interval Sales (in $)	Current Time Interval Waste %
Fresh	Region	All	All	3	2024-01-06	Week	High waste to sales ratio was reported 7.24% for Fresh departments this week in region 3	868778.18	12006833.82	7.24
Fresh	Store	3121	5	1	2024-01-06	Quarter	High waste to sales ratio was reported 25.94% for Fresh departments over last 12 weeks in store 3121	5111.23	19707.39	25.94
Fresh	Store	1046	7	1	2024-01-06	Month	High waste to sales ratio was reported 11.14% for Fresh departments over the last month in store 1046	17778.65	159646.47	11.14
Fresh	Store	3060	3	1	2024-01-06	Month	High waste to sales ratio was reported 11.67% for Fresh departments over the last month in store 3060	6745.49	57782.72	11.67
Fresh	Store	3191	1	1	2024-01-06	Week	High waste to sales ratio was reported 15.73% for Fresh departments this week in store 3191	1825.32	11601.49	15.73
Fresh	Store	1036	3	1	2024-01-06	Week	High waste to sales ratio was reported 9.11% for Fresh departments this week in store 1036	2534.37	27822.83	9.11
Fresh	Store	1022	10	1	2024-01-06	Month	High waste to sales ratio was reported 9.57% for Fresh departments over the last month in store 1022	66087.52	690811.16	9.57

Table 4: Critical Insights and recommendations provided to store associates.

Based on adoption of the recommendations provided by the store associates, over the course of 6 months, a reduction of wastage by 26% was noticed.

4.3 How can early Detection of Out-of-Stock using Computer Vision help in real-time Inventory management to reduce waste due to overstocking?

Accurate and early detection of Out-of-Stocks is a very crucial problem. While traditional methods exist where associates can identify out-of-stocks through regular weekly scanning of inventory and shelves, it may be already too late to identify, as well as requires a lot of manual effort. In this scenario Computer vision based out-of-stock detection can be very helpful for early real-time outs detection. We have tried out our models on many different categories of products. Overall accuracy of the model is around 90 % which is quite high, given that it is a semi-supervised model.

Figure 28 shows the model results for a few sample shelf-images:

Figure 28: Computer vision based Out-of-Stock detection results

Semi-Supervised Void Detection Results		
	No of Voids Detected by the model	Actual Void Count
HRes_DJI_0369.jpg_0000_orig	6	6
HRes_DJI_0369.jpg_0001_orig	1	1
HRes_DJI_0369.jpg_0002_orig	1	1
HRes_DJI_0369.jpg_0003_orig	5	6
HRes_DJI_0369.jpg_0004_orig	3	3
HRes_DJI_0369.jpg_0005_orig	2	3

Table 5: Comparison of Computer-vision based Out-of-Stocks and validation.

Table 5 shows the results of validation of our model accuracy on a variety of product images. We have implemented the solution for US and Canada with Store Shelf images provided by on-shelf and drone cameras in stores, Jan 2023 to Jan 2024.

Out-of-stock detection was compared between comparable stores where computer-vision techniques were used vs. not used. Figure 29 highlights that overall Computer vision enabled stores were able to detect 8 times more out-of-stocks than regular stores.

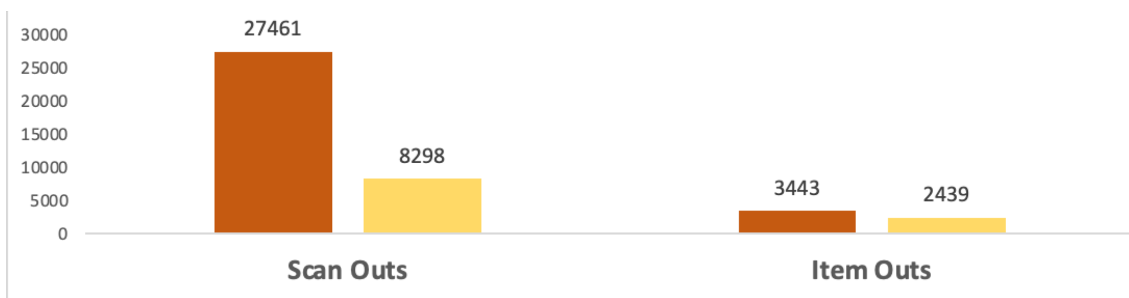


Figure 29: Comparion of outs detection in Computer vision based stores and others

The model helps in providing an automated alert to the store personnel whenever there is an out-of-stock scenario, so that the necessary steps can be taken else an out of stock in a shelf is responsible for customer dissatisfaction.It helps the business in merchandizing, replenishment & assortment decisions effectively.

This model will ensure better compliance of pre-emptive out-of-stock detection which will have significant uplift in incremental sales and improve customer experience. Estimates on optimal planogram showcases potential of around 10 % lift in incremental sales, from predicted demand models.

4.4 How can better demand planning for Fresh production help in reducing wastage due to over-production?

With increasing competition in the retail industry, accurately forecasting ever-changing demand for perishable items across multiple countries is a challenging task with high stakes. Over-forecasting leads to tremendous waste, while under forecasting results in bad customer experience. Fresh is currently facing approximately a billion dollars of waste across international countries. In this study, several forecasting algorithms were implemented and compared to identify which can accurately predict the demand without leading to over-forecasts or under-forecasts, that can have severe consequences on waste and lost sales opportunity. We used sample of 92 stores and 58 items for the comparison from store side. The stores and items were selected in different volume buckets for fair comparison. Overall MAPE came out to be 1.03%

Department	Units Produced	Units Sold	Units Forecasted	Error	Overall Accuracy
98	139,768,877.02	135,346,698.02	132,475,835.28	-2.12%	97.88%
93	18,925,194.90	18,530,082.90	19,737,562.67	6.52%	93.48%
81	1,107,367.43	1,022,503.43	1,089,088.27	6.51%	93.49%

Table 6: Demand forecasting results for Fresh production items.

This solution has helped decrease throws from 6% to 2.7% and also led to reduction of shrinkage, as shown in Figure 30.

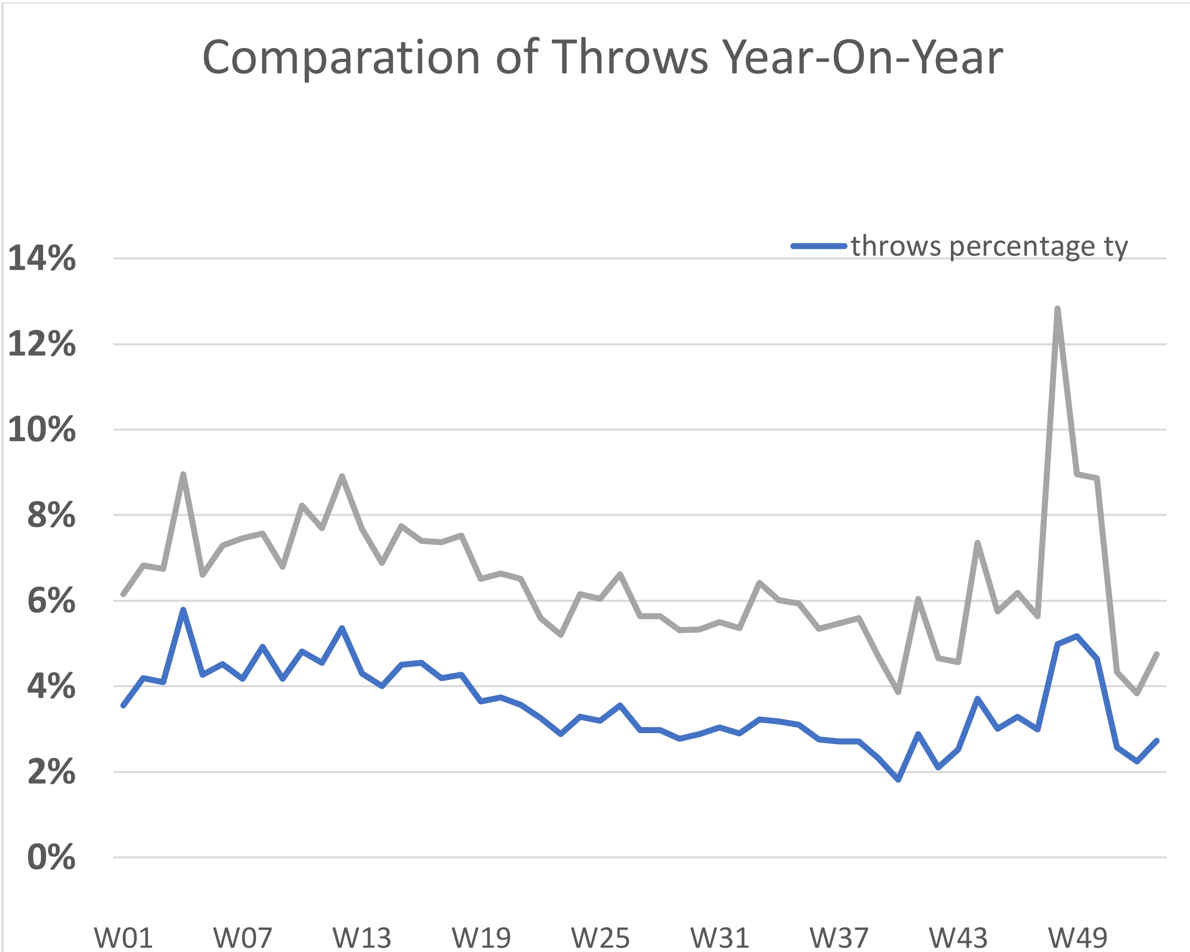


Figure 30: Comparion of waste and throws due to over-production

This shows that accurate forecasts of the demand for the fresh items can lead to significant reduction of wastage and throws.

CHAPTER V:

DISCUSSION

5.1 Discussion of results

The research has focused on identifying the top instances of losses that occur in omni-channel retail. Primarily this includes losses from fraudulent activities and malicious intent, as well as losses due to gaps in the processes. Some of the losses occur due to wastage, throwaways, lost sales opportunities due to out-of-stock, incorrect inventory information, and many others. The research has focused on how different Machine Learning techniques can be used to prevent such losses both from fraud and wastage. The following table represents the research questions that we started as our purpose, delves deeper into methodology and how measurement was performed, and the results of the research.

Research Question	Methodology and Measurement	Results
How Machine Learning can help in eliminating gaps between offline and online channels which lead to fraudulent activities?	Fraudulent transactions and customers are identified using multitude of supervised, unsupervised, graph ML based algorithms, and the recommendations are shared to group of risk	The system analyses returns and cancellations, identifying 56% of total refunds happen with customers who return more than 95% of their purchases. The ML based solution achieved 91%

	analysts for review. Based on feedback the model is improved, and better samples are shared till model convergance and sufficient accuracy is achieved.	accuracy. The system has improved the hit-rate for recommended customers from 13% to almost 98% in a short time span. This notable improvement, achieved by using the active learning-based recommendation system.
What are the major causes of Retail Waste and how prescriptive Machine Learning recommendations can help to reduce waste?	Step-by-step approach of descriptive, predictive and prescriptive models have been used for identification of waste, driving factors for waste using causal discovery, and provided prescriptive timely recommendations to thousands of store associates, using mobile devices. Recommendations are improved using user feedback. The ML driven recommendations are	The solution identified 200 risky and 30 high-risky stores with abnormally high waste, and about 20 factors driving waste. Based on the adoption of suggested recommendations by store associates, a reduction in wastage by 26% was seen over the course of six months.

	transformed to meaningful sentences using generative models.	
How can early Detection of Out-of-Stock using Computer Vision help in real-time Inventory management to reduce waste due to overstocking?	For 25 stores and 3 departments, on-shelf cameras are placed which can capture information on out-of-stock, and computer vision based object detection algorithms help in out-of-stock detection for real-time inventory adjustment. In addition, weekly inventory process by store associates, and unavailability of items during pick-up act as signals for out-of-stock. The reversal of inventory adjustments are treated as signals of false positives to improve the solution.	Stores using computer vision based solution were able to detect eight times more out-of-stocks than regular stores, when compared to group of matched control stores. Also, the time of adjustment of inventory was brought down from 15 days to less than 1 day.
How can better demand planning for Fresh	Store associates in bakery and meat areas, identify the	The forecasting algorithm utilising 10+ forecasting

production help in reducing wastage due to over-production?	right amount of items to produce. Due to short shelf-life any excess production would be wasted. Accurate short term demand forecasting algorithms across 50 stores and 200 fresh items have been recorded and shared to store associates on mobile applications to help with production of fresh items.The results of sales and wastage are compared with similar stores without ML driven forecasts.	algorithm ensemble models has around 5% SMAPE and has been used to provide accurate forecasts. When compared to similar group of control stores, the ML driven forecasting solution helped in reducing wastage by 55%.

Table 7: Summary of Research Question, methodology and results

The following sections cover a deeper discussion on the results on ML driven fraud prevention and waste reduction.

5.2 Discussion of ML driven fraud prevention

The use of Machine Learning models has produced significant results in the prevention of fraud, with the system already generating multi-million-dollar savings by identifying fraudulent cases.

The system analyses returns and cancellations, identifying a large proportion (56%) of total refunds happen with customers who return more than 95% of their purchases. These refunds are often due to delivery exceptions and defective items, indicating potential collusion between customers and delivery drivers or store associates. The machine learning solutions have detected several anomalous scenarios using multi-variate anomaly detection techniques. The methodology proposed a unique combination of weighted average of unsupervised risk score and an ensembled supervised risk score, leading to an accuracy of 91% and a false positive rate of 14%. The system has improved the hit-rate for recommended customers from 13% to almost 98% in a short time span. This notable improvement, achieved by using the active learning-based recommendation system, translates into an annual saving of hundreds of thousand dollars from identifying risky customers.

The research has uncovered gaps between the online and offline channels of retail that are being targeted by fraudsters to harm both retailers and customers. The ML driven approaches provide an easy way of highlighting and prioritising cases of immidiate importance to the fraud analysts with evidence-based recommendations to identify such scenarios of collusion early and prevent further losses.This extends far beyond traditional payments fraud systems which only identify credit card frauds, but also brings out a holistic picture of abuse and business process opportunities that can be mitigated.

5.3 Discussion of ML driven waste reduction

Waste is another major area of loss for retailers. A multi-stage machine learning technique was implemented to tackle the issue of retail waste, focusing on the identification of high waste contributing stores and items. The technique begins with descriptive analysis, moving onto predictive and then prescriptive methods to reduce waste. Anomaly detection models were used to reduce the size of data analysed by over 95%, making the study more precise. The waste risk score has been used to group stores based on importance, with stores in the highest risk score zone requiring more diligent investigation. Furthermore, the bootstrapped Causal Forest model provided insights into which factors have the most impact for driving waste. Based on the suggestions adopted by store associates, a reduction in wastage by 26% was seen over the course of six months. The waste reduction framework utilises several flavours of ML models including the advanced Causal ML models. The recommendations provided has been at various levels starting from department and store levels, and going to the granularity of individual items. Many retailers have applications that enable their staff to reduce the price of merchandise that is close to its expiration date, to sell it faster. Furthermore, there are apps that can send notifications to review fresh merchandise based on the entry and tracking of expiration dates. The Fresh Waste Reduction Lost Profit Notification app does not require any input from the user like an expiry date. Instead, it uses signals from store processes in a group of stores to determine when action is necessary and estimates the value of taking that action and provides detailed actions and recommendations to associates to reduce waste proactively. This results in user buy-in. As it's infeasible to have zero waste, our objective is to minimize waste as much as possible and ensure we meet our target goals set for the year on KPIs such as Waste to Sales Ratio, CVP Sales to Waste ratio etc. These goals serve as reference to indicate how good or poor a store is doing against its target and can also be used to compare and rank stores based on their performance. Lost Profit is nothing but the deviation in our KPIs performance from target goals calculated at department/SBU level before rolling it up at an entire store level. With the objective of minimizing lost profit value, particularly for Fresh departments, we developed a methodology to recommend specific actions to be taken by associates, review and measure the compliance of actions taken, all of which is integrated via a mini app. We provide recommendations to the associate at various levels such as store, dept, category, item etc. Recommendations can be distinct types such as goal-based set by business, similarity based such as recommend good performing store behavior to similar but poor performing stores based on different waste KPIs. Imagine we are on Day D6 and when a user logs into a store 5123, clicking on the recommendations on the lost profit card on home screen takes us to another screen shown Figure 2. Along with the lost profit value, savings due to actions taken on the previous day are also displayed. Below that is the section on recommendations, categorized as action required/completed/skipped. Usually, we limit the suggested actions to top 2/3 (in terms of opportunity savings) that are due. For example, our first recommendation is for tomatoes finelines: Apply 1st cvp markdown to reduce the % of items being donated, near to expiry. Second recommendation is for Formal Men’s Shirt: Return to Vendor. Along with the recommendation, estimated opportunity savings are also displayed. Clicking on either of the recommendations takes us to another screen with more details. For example, clicking on the first recommendation for tomatoes takes us to the screen on the left in Figure 3. Along with the original recommendations, users can see the top 3 high waste items within the Tomatoes fineline. Previous suggestions made on D1 through D5 along with the action taken by the associate (completed/skipped etc.) and the impact of the action are shown next to that. In this example, on D1, we ask the associate to increase on-shelf availability of tomatoes by 5% by replenishing from backroom. Estimated opportunity was $150-$200 but the actual impact was $50. On D4, we suggested to transfer excess inventory to Store 1182 but the associate skipped it. So, on D6, we suggest giving discounts to ensure we sell off at a lower price than original before it expires and avoid donations/throwaways when it's too late to dispose at zero price.

This journey of the item from the time of first recommendation to the last one can change based on what actions were taken and how the item status has changed (sell through rate, time to expiry etc.). Feedback is also collected from associates on whether the insights were helpful/not as well as whether they performed the actions we rolled out. Further, once the top 2 actions are addressed in a day, depending on the scenario, more actions are also suggested if opportunities are seen. This exercise also helps in following the optimal journey for all items that need to go through claims process such as returns, CVP, donation, throwaway etc. With such clear insights and suggested actions to take, this saves a lot of associates and regional/store managers time in identifying the pain points and implement specific changes in the stores in a short time and improve bottom line for retail stores. Our KPIs will support store level performance goals. For example, we will rank and compare stores to each other, then compare stores to their KPIs, Essentially, Lost Profit is the measurement (deviation) in our KPIs performance from target goals calculated at department/SBU level before rolling it up at an entire store level. The objective is to minimize lost profit value, particularly for Fresh departments. As a result, we developed a methodology to recommend specific actions to be taken by associates at the store level conveniently enabled via an integrated hand-held app.

The research also focused on early detection of Out-of-Stocks, which is a crucial problem. Computer vision-based out-of-stock detection has been implemented, providing 90% accuracy in real-time detection. The solution has been implemented for US and

Canada with store shelf images provided by on-shelf and drone cameras. Stores using computer vision were able to detect eight times more out-of-stocks than regular stores. This early detection leads to an automated alert for store personnel, which improves customer satisfaction and aids business merchandising, replenishment and assortment decisions. Furthermore, the study identified the importance of accurate demand forecasting for perishable and fresh items. Incorrect forecasting can lead to tremendous waste or bad customer experience. The implemented forecasting algorithms resulted in a reduction of waste from 6% to 2.7%.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

The retail sphere has undergone a significant transformation, evolving from simple neighborhood shops to large, complex omni-channel retail eco-systems. This evolution has been accompanied by the emergence of new challenges, particularly in the areas of fraud and waste. With thin margins, retailers are under constant pressure to innovate, with the goal of reducing losses stemming from various factors such as fraud, shrink, waste, and lost sales opportunities.

The research presented in this paper highlights the need to develop more robust fraud detection systems, improve the customer experience, and reduce returns. It explores the use of advanced predictive and prescriptive Machine Learning (ML) models to create tools and recommendations that can decrease losses and make the retail ecosystem safer from fraud and more sustainable against waste.

In the conducted study, data from omni-channel retail transactions over two years across multiple countries were analyzed. Feedback was gathered from loss analysts and store associates, and ML methods of supervised and unsupervised learning were used. The implementation of these ML techniques in retail total loss management has been shown to result in multi-million-dollar savings and improved fraud detection, inventory management, and waste management, with the performance of supervised models improving to nearly 98% over several months.

6.2 Implications

The findings of this research have very significant implications for the retail industry especially as more focus has been shifted in recent times to be data-driven and AI powered retail businesses. The results demonstrate that the implementation of Machine Learning models can lead to substantial savings, improved fraud detection, and waste reduction. Thus, the study shows that ML models are not only beneficial but also essential for retail businesses, considering the sector's razor-thin margins and the constant need for innovation.

The study also uncovers gaps between online and offline retail channels that are exploited by fraudsters. These insights can guide retailers in developing strategies and policies to close these gaps and prevent fraudulent activities. The innovative approach of prescriptive ML models to provide a guided workflow to store associates with recommendations on waste reduction and inventory management has brought on a digital transformation in the retail industry. Additionally, the research highlights the importance of accurate demand forecasting and early detection of out-of-stock items. Both of these factors greatly contribute to customer satisfaction and have a direct impact on a retailer's bottom line.

6.3 Recommendations for Future Research

While this research provides valuable insights and demonstrates the benefits of implementing ML models in retail, there are still areas that warrant further investigation. For instance, future research could explore how these models can be integrated more effectively and seamlessly into the existing systems of retailers to facilitate their adoption. There is opportunity of integration of the offline unsupervised models into real-time fraud prevention systems to better prevent a multitude of fraudulent activities during the customer journey as compared to post-purchase.

Additionally, the research could be extended to other facets of retail operations, such as customer service, marketing, and supply chain management. A focus on these areas could provide a more holistic view of the potential benefits of ML in retail.

Future research could also delve deeper into the issue of waste in retail, particularly focusing on strategies to reduce waste due to supply chain, sub-optimal pricing, and discounts. Research could also investigate how to sustainable packaging and renewal, remanufacturing and resale of the disposed products can further reduce retail waste. Future research can also focus on specific effects of in-store temperature, weather, malfunction of in-store devices and their impact on waste.

6.4 Conclusion

In conclusion, the research has demonstrated the significant potential of Machine Learning models in the retail industry, especially in terms of building customer's trust and reduction of losses in different ways. The use of these advanced techniques has been shown to result in considerable savings, improved fraud detection, enhanced inventory management, and waste reduction. The research underscores the importance of adopting such technologies in the retail sector, particularly in light of the sector's narrow margins and the ever-present need for innovation. The insights gained from this research can guide retailers in optimizing their operations, enhancing customer satisfaction, and ultimately, increasing their profitability.

REFERENCES

1. Adebanjo, D. and Mann, R., 2000. Identifying problems in forecasting consumer demand in the fast moving consumer goods sector. *Benchmarking: An international journal*, *7*(3), pp.223-230.
2. Abadie, A., Diamond, A. and Hainmueller, J., 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, *105*(490), pp.493-505.
3. Akturk, M.S., Ketzenberg, M. and Yıldız, B., 2021. Managing consumer returns with technology-enabled countermeasures. *Omega*, *102*, p.102337.
4. Allegra, D., Litrico, M., Spatafora, M.A.N., Stanco, F. and Farinella, G.M., 2021. Exploiting egocentric vision on shopping cart for out-of-stock detection in retail environments. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1735-1740).
5. Facure Alves, M., 2022. Causal Inference for The Brave and True. *Zenodo, online resorce*., pp.224-229.
6. Baeza-Yates, R. and Navarro, G., 1996, June. A faster algorithm for approximate string matching. In Annual Symposium on Combinatorial Pattern Matching (pp. 1-23). Berlin, Heidelberg: Springer Berlin Heidelberg.
7. Barratt, M., Kull, T.J. and Sodero, A.C., 2018. Inventory record inaccuracy dynamics and the role of employees within multi-channel distribution center inventory systems. Journal of Operations Management, 63, pp.6-24.
8. Beck, A. (2020). Total Retail Loss 2.0: Moving Beyond the Theory. [online] Loss Prevention Media. Available at: https://losspreventionmedia.com/total-retail-loss-2-0-moving-beyond-the-theory/ [Accessed 23 Apr. 2023].
9. Binyon, S., 2007. Reducing and managing waste in the food industry: food industry sustainability best practice workshop. *London, UK: Food and Drinks Federation*. ,pp.224-229.
10. Bolton, R.J. and Hand, D.J., 2002. Statistical fraud detection: A review. Statistical science, 17(3), pp.235-255
11. Borrello, M., Caracciolo, F., Lombardi, A., Pascucci, S. and Cembalo, L., 2017. Consumers' perspective on circular economy strategy for reducing food waste. Sustainability, 9(1), p.141.
12. Botelho, J. and Antunes, C., 2011. Combining Social Network Analysis with Semi-supervised Clustering: a case study on fraud detection. Proceeding of mining data semantics (MDS'2011) in conjunction with SIGKDD, pp.1-7.
13. Brandtner, P., Darbanian, F., Falatouri, T. and Udokwu, C., 2021. Impact of COVID-19 on the customer end of retail supply chains: A big data analysis of consumer satisfaction. Sustainability, 13(3), p.1464.
14. Brodersen, K.H., Gallusser, F., Koehler, J., Remy, N. and Scott, S.L., 2015. Inferring causal impact using Bayesian structural time-series models,pp.225-232.

15. Bury, K.V., 1999. Statistical distributions in engineering. Cambridge University Press, ,pp.105-121.
16. Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), pp.1-58.
17. Chen, J., Wang, S.L. and Lin, H.L., 2019, August. Out-of-stock detection based on deep learning. In International Conference on Intelligent Computing (pp. 228-237). Springer, Cham.
18. Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
19. Malefors, C., Secondi, L., Marchetti, S. and Eriksson, M., 2022. Food waste reduction and economic savings in times of crisis: The potential of machine learning methods to plan guest attendance in Swedish public catering during the Covid-19 pandemic. *Socio-Economic Planning Sciences*, *82*, p.101041.
20. Coppolino, L., D'Antonio, S., Formicola, V., Massei, C. and Romano, L., 2015. Use of the Dempster–Shafer theory to detect account takeovers in mobile money transfer services. Journal of Ambient Intelligence and Humanized Computing, 6(6), pp.753-762.
21. Cover, T.M., 1999. Elements of information theory. John Wiley & Sons.
22. da Cruz, N.F., Simões, P. and Marques, R.C., 2012. Economic cost recovery in the recycling of packaging waste: the case of Portugal. Journal of Cleaner Production, 37, pp.8-18.
23. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C. and Bontempi, G., 2015, July. Credit card fraud detection and concept-drift adaptation with delayed supervised information. In 2015 international joint conference on Neural networks (IJCNN) (pp. 1-8). IEEE.
24. De Brito, M.P., Carbone, V. and Blanquart, C.M., 2008. Towards a sustainable fashion retail supply chain in Europe: Organisation and performance. International journal of production economics, 114(2), pp.534-553.
25. De Souza, R., Goh, M., Lau, H.C., Ng, W.S. and Tan, P.S., 2014. Collaborative urban logistics–synchronizing the last mile a Singapore research perspective. Procedia-Social and Behavioral Sciences, 125, pp.422-431.
26. Doerfler, P., Thomas, K., Marincenko, M., Ranieri, J., Jiang, Y., Moscicki, A. and McCoy, D., 2019, May. Evaluating login challenges as adefense against account takeover. In The World Wide Web Conference (pp. 372-382).
27. Doganis, P., Alexandridis, A., Patrinos, P. and Sarimveis, H., 2006. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. Journal of Food Engineering, 75(2), pp.196-204.
28. El Kihal, S. and Shehu, E., 2022. It's not only what they buy, it's also what they keep: Linking marketing instruments to product returns. *Journal of Retailing*, *98*(3), pp.558-571.
29. Elshaar, S. and Sadaoui, S., 2020. Semi-supervised classification of fraud data in commercial auctions. Applied Artificial Intelligence, 34(1), pp.47-63.
30. Fao, F.A.O.S.T.A.T., 2018. Food and agriculture organization of the United Nations. *Rome, URL: http://faostat. fao. org*, pp.403-403.
31. Fink, D., 1997. A compendium of conjugate priors. *See http://www. people. cornell. edu/pages/df36/CONJINTRnew% 20TEX. pdf*, *46*.

32. Forgy, E.W., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. biometrics, 21, pp.768-769.
33. Frontoni, E., Marinelli, F., Rosetti, R. and Zingaretti, P., 2017. Shelf space re-allocation for out of stock reduction. Computers & Industrial Engineering, 106, pp.32-40.
34. Fullerton, D. and Kinnaman, T.C., 1995. Garbage, recycling, and illicit burning or dumping. Journal of environmental economics and management, 29(1), pp.78-91.
35. Gao, M., 2022, June. Account Takeover Detection on E-Commerce Platforms. In 2022 IEEE International Conference on Smart Computing (SMARTCOMP) (pp. 196-197). IEEE.
36. Garnett, T., 2011. Where are the best opportunities for reducing greenhouse gas emissions in the food system (including the food chain)?. Food policy, 36, pp.S23-S32.
37. Gill, M., 2018. The challenges to preventing losses in retailing: Views from retail loss prevention managers and directors. In Retail Crime (pp. 357-378). Palgrave Macmillan, Cham.
38. Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2015. Region-based convolutional networks for accurate object detection and segmentation. IEEE transactions on pattern analysis and machine intelligence, 38(1), pp.142-158.
39. Glatzel, C., Hopkins, M., Lange, T. and Weiss, U., 2016. The secret to smarter fresh-food replenishment? Machine learning. *McKinsey & Company*.
40. Gružauskas, V., Gimžauskienė, E. and Navickas, V., 2019. Forecasting accuracy influence on logistics clusters activities: The case of the food industry. Journal of Cleaner Production, 240, p.118225.
41. Hardgrave, B.C., Aloysius, J. and Goyal, S., 2009. Does RFID improve inventory accuracy? A preliminary analysis. International Journal of RF Technologies, 1(1), pp.44-56.
42. Hazewinkel, M., 2001. The algebra of quasi-symmetric functions is free over the integers. Advances in Mathematics, 164(2), pp.283-300.
43. Hollander, M., Wolfe, D.A. and Chicken, E., 2013. Nonparametric statistical methods. John Wiley & Sons.
44. Hollinger, R. (n.d.). 2002 National Retail Security Survey: Final Report. www.academia.edu. [online] Available at: https://www.academia.edu/18038890/2002_National_Retail_Security_Survey_Final_Report [Accessed 20 Nov. 2022].
45. Hollinger, R.C., Dabney, D.A. and Hayes, R., 1990. *National retail security survey*. Security Research Project, Department of Sociology, University of Florida.
46. Huang, I.Y., Manning, L., James, K.L., Grigoriadis, V., Millington, A., Wood, V. and Ward, S., 2021. Food waste management: A review of retailers' business practices and their implications for sustainable value. Journal of Cleaner Production, 285, p.125484.
47. Hudson, J., Fielding, S. and Ramsay, C.R., 2019. Methodology and reporting characteristics of studies using interrupted time series design in healthcare. BMC medical research methodology, 19, pp.1-7.
48. Hyde, K., Smith, A., Smith, M. and Henningsson, S., 2001. The challenge of waste minimisation in the food and drink industry: a demonstration project in East Anglia, UK. Journal of Cleaner Production, 9(1), pp.57-64.

49. Kantor, C., Rauby, B., Boussioux, L., Jehanno, E. and Talbot, H., 2020. Over-CAM: Gradient-Based Localization and Spatial Attention for Confidence Measure in Fine-Grained Recognition using Deep Neural Networks.
50. King, T., 2004. An analysis of the phenomenon of deshopping of garments in women's wear retailing. *Unpublished Doctoral Dissertation, Brunnel University, Brunel*.
51. Kondo, F.N. and Kitagawa, G., 2000. Time series analysis of daily scanner sales: extraction of trend, day-of-the-week effect and price promotion effect. Marketing Intelligence & Planning.
52. Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84-90.
53. Kulatilleke, G.K., 2022. Challenges and complexities in machine learning based credit card fraud detection. arXiv preprint arXiv:2208.10943.
54. Lemaire, A. and Limbourg, S., 2019. How can food loss and waste management achieve sustainable development goals?. Journal of Cleaner Production, 234, pp.1221-1234.
55. LexisNexis Risk Solutions. (n.d.). The True cost of FraudTM Study. [online] Available at: https://risk.lexisnexis.com/insights-resources/research/us-ca-true-cost-of-fraud-study. [Accessed 20 Nov. 2022].
56. Li, Z., 2016. Anomaly detection and predictive analytics for financial risk management (Doctoral dissertation, Rutgers University-Graduate School-Newark).
57. Liu, F.T., Ting, K.M. and Zhou, Z.H., 2012. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data (TKDD), 6(1), pp.1-39.
58. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016. Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.
59. Managa, M.G., Tinyani, P.P., Senyolo, G.M., Soundy, P., Sultanbawa, Y. and Sivakumar, D., 2018. Impact of transportation, storage, and retail shelf conditions on lettuce quality and phytonutrients losses in the supply chain. Food science & nutrition, 6(6), pp.1527-1536.
60. Lokanan, M.E. and Sharma, K., 2022. Fraud prediction using machine learning: The case of investment advisors in Canada. *Machine Learning with Applications, 8*, p.100269.
61. Marrucci, L., Marchi, M. and Daddi, T., 2020. Improving the carbon footprint of food and packaging waste management in a supermarket of the Italian retail sector. Waste Management, 105, pp.594-603.
62. Massarutto, A., De Carli, A. and Graffi, M., 2011. Material and energy recovery in integrated waste management systems: A life-cycle costing approach. Waste management, 31(9-10), pp.2102-2111.
63. Masud, M.M., Gao, J., Khan, L., Han, J. and Thuraisingham, B., 2008, December. A practical approach to classify evolving data streams: Training with limited amount of labeled data. In 2008 Eighth IEEE International Conference on Data Mining (pp. 929-934). IEEE.
64. Mena, C., Adenso-Diaz, B. and Yurt, O., 2011. The causes of food waste in the supplier–retailer interface: Evidences from the UK and Spain. Resources, Conservation and Recycling, 55(6), pp.648-658.
65. Milka, G., 2018. Anatomy of account takeover. In Enigma 2018 (Enigma 2018).

66. Nanduri, J., Jia, Y., Oka, A., Beaver, J. and Liu, Y.W., 2020. Microsoft uses machine learning and optimization to reduce e-commerce fraud. INFORMS Journal on Applied Analytics, 50(1), pp.64-79.
67. Niu, X., Wang, L. and Yang, X., 2019. A comparison study of credit card fraud detection: Supervised versus unsupervised. arXiv preprint arXiv:1904.10604.
68. NRF. (n.d.). National Retail Security Survey 2019. [online] Available at: https://nrf.com/research/national-retail-security-survey-2019 [Accessed 20 Nov. 2022].
69. NRF. (n.d.). NRF Reports Retail Shrink Nearly a $100B Problem. [online] Available at: https://nrf.com/media-center/press-releases/nrf-reports-retail-shrink-nearly-100b-problem#:~:text=Retailers%20are%20bolstering%20security%20efforts [Accessed 13 Jan. 2023].
70. Parfitt, J., Barthel, M. and Macnaughton, S., 2010. Food waste within food supply chains: quantification and potential for change to 2050. Philosophical transactions of the royal society B: biological sciences, 365(1554), pp.3065-3081.
71. Phau, I., Akintimehin, O.O., Shen, B. and Cheah, I., 2022. “Buy, wear, return, repeat”: Investigating Chinese consumers' attitude and intentions to engage in wardrobing. *Strategic Change*, *31*(3), pp.345-356.
72. Phua, C., Lee, V., Smith, K. and Gayler, R., 2010. A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.
73. Pingali, P., Mittra, B. and Rahman, A., 2017. The bumpy road from food to nutrition security–Slow evolution of India's food policy. Global food security, 15, pp.77-84.
74. Piron, F. and Young, M., 2001. Retail borrowing: definition and retailing implications. *Journal of Retailing and Consumer Services*, *8*(3), pp.121-125.
75. Post, A., Shah, V. and Mislove, A., 2011. Bazaar: Strengthening user reputations in online marketplaces. In 8th USENIX Symposium on Networked Systems Design and Implementation (NSDI 11).
76. Richter, B. and Bokelmann, W., 2016. Approaches of the German food industry for addressing the issue of food losses. Waste management, 48, pp.423-429.
77. Sadowski, G. and Rathle, P., 2014. Fraud detection: Discovering connections with graph databases. White Paper-Neo Technology-Graphs are Everywhere, 13.
78. Santarcangelo, V., Farinella, G.M., Furnari, A. and Battiato, S., 2018. Market basket analysis from egocentric videos. Pattern Recognition Letters, 112, pp.83-90.
79. Shang, G., Ferguson, M.E. and Galbreth, M.R., 2019. Where should I focus my return reduction efforts? Empirical guidance for retailers. Decision Sciences, 50(4), pp.877-909.
80. Shi, T. and Horvath, S., 2006. Unsupervised learning with random forest predictors. Journal of Computational and Graphical Statistics, 15(1), pp.118-138.
81. Shih, D.H., Huang, F.C., Chieh, C.Y., Shih, M.H. and Wu, T.W., 2021. Preventing return fraud in reverse logistics—A case study of ESPRES solution by Ethereum. Journal of Theoretical and Applied Electronic Commerce Research, 16(6), pp.2170-2191.
82. Shin, H.C., Lee, K.I. and Lee, C.E., 2020, February. Data augmentation method of object detection for deep learning in maritime image. In 2020 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 463-466). IEEE.
83. Solutions, L.R., 2018. True Cost of Fraud Study. Retrieved from.

84. Speights, D. and Hilinski, M., 2005. Return fraud and abuse: How to protect profits. *Retailing Issues Letter*, *17*(1), pp.1-6.
85. Stavarache, L.L., Narbutis, D., Suzumura, T., Harishankar, R. and Žaltauskas, A., 2019. Exploring Multi-Banking Customer-to-Customer Relations in AML Context with Poincar\'e Embeddings. arXiv preprint arXiv:1912.07701.
86. Thompson, W.R., 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika, 25(3-4), pp.285-294.
87. Tomaselli, D., 2019. Automated Recycling System Using Computer Vision. ECE-498: Capstone Design Project Advisor: Prof. Cotter November, 25.
88. van der Vorst, J.G., Beulens, A.J., De Wit, W. and van Beek, P., 1998. Supply chain management in food chains: Improving performance by reducing uncertainty. International Transactions in Operational Research, 5(6), pp.487-499.
89. Vanini, P., Rossi, S., Zvizdic, E. and Domenig, T., 2023. Online payment fraud: from anomaly detection to risk management. Financial Innovation, 9(1), p.66.
90. Wallny, F., 2022, January. False Positives in Credit Card Fraud Detection: Measurement and Mitigation. In HICSS (pp. 1-10).
91. Walsh, G., Albrecht, A.K., Kunz, W. and Hofacker, C.F., 2016. Relationship between online retailers' reputation and product returns. British Journal of Management, 27(1), pp.3-20.
92. Weber, F. and Schütte, R., 2019. A domain-oriented analysis of the impact of machine learning—the case of retailing. *Big Data and Cognitive Computing*, *3*(1), p.11.
93. Wedge, R., Kanter, J.M., Veeramachaneni, K., Rubio, S.M. and Perez, S.I., 2018, September. Solving the false positives problem in fraud prediction using automated feature engineering. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 372-388). Springer, Cham.
94. Winiarti, S., Prahara, A., Murinto, D.P.I. and Ismi, P., 2018. Pre-trained convolutional neural network for classification of tanning leather image. network (CNN), 9(1), p.9.
95. www.actenviro.com.,2022. Sustainable Retail Industry Waste Management. [online] Available at: https://www.actenviro.com/retail-industry-waste-management/.[Accessed 20 Nov. 2022].
96. Yamanishi, K., Takeuchi, J. and Maruyama, Y., 2005. Data mining for security. NEC journal of advanced technology, 2(1), pp.63-69.
97. Zhang, D., Frei, R., Senyo, P.K., Bayer, S., Gerding, E., Wills, G. and Beck, A., 2023. Understanding fraudulent returns and mitigation strategies in multichannel retailing. *Journal of retailing and consumer services, 70*, p.103145.