A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

EVALUATING THE EFFICACY OF GRADIENT BOOSTING ALGORITHMS IN RETAIL DEMAND FORECASTING:

A CASE STUDY OF TRIUMPH INTERNATIONAL

MONICA SCHIRRIPA

Work project carried out under the supervision of:

Rodrigo Belo

19/12/2023

Abstract

This thesis is the result of my internship at The Data Cooks, an Amsterdam-based data agency, and aims to enhance demand forecasting for Triumph International, a leading lingerie manufacturer. The research is focused on addressing the complex challenge of disaggregated forecasting in the fashion retail industry, often characterized by a large number of stores and products. Central to this research is the application of gradient boosting algorithms, a cutting-edge approach in machine learning. By focusing on methods such as LightGBM and XGBoost, this study delves into the effectiveness of ensemble learning in handling complex, time-sensitive data.

Keywords

Demand Forecasting, Machine Learning, Time-Series Forecasting, Fashion Retail, Master Thesis, Gradient Boosting

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Table of Contents

1. Introduction

The ongoing digital revolution is dramatically reshaping industries across the board, and the retail sector is no exception. A central pillar supporting this transformation is machine learning (ML), a subfield of artificial intelligence that enables computers to learn from data. One of the most potent applications of ML in retail is in the area of demand forecasting, a long-standing challenge for retailers seeking to align their operations with market needs. This study focuses on Triumph International, a multinational lingerie manufacturer grappling with the complexities of demand forecasting to improve replenishment planning. Guided by the overarching research question "How can Gradient Boosting algorithms enhance disaggregated demand forecasting for Triumph International?, this study aims to explore, analyze, and provide empirical evidence on this topic.

1.1. The critical role of demand forecasting in the fashion retail industry

Demand forecasting refers to the systematic process of estimating future customer demand for a product or a service (Ingle et al., 2021). This practice is widespread across industries, playing a crucial role in sectors as diverse as healthcare, energy, automotive, and finance. In each of these fields, demand forecasting assists in optimizing operational efficiencies, from inventory management to production scheduling. However, its application becomes particularly intricate in the fashion retail industry.

Fashion represents one of the most challenging categories for forecasting demand (Swaminathan & Venkitasubramony, 2023). Due to the high degree of uncertainty, the fashion industry faces unique challenges such as rapid changes in consumer preferences, high levels of seasonality, and short product life cycles. Such factors contribute to the complexity of making accurate forecasts (Giustiniano et al., 2013). Another important aspect is its significant level of granularity in the product offer and distribution channels, which further complicates the forecasting process (Andrade & Cunha, 2023).

Accurate demand forecasting is not just a technical necessity but a strategic imperative for retailers. One of the most evident impacts is seen in supply chain management. When demand forecasts are precise, the supply chain can operate at its optimum level. Retailers can coordinate better with suppliers, ensuring that products are manufactured and delivered in the right quantities and at the right times.

Furthermore, retailers with a clear understanding of future demand can make informed decisions regarding stock levels in their inventory management. Overstocking and understocking, both of which can be costly missteps, are avoided. Overstocking leads to increased holding costs, potential obsolescence, and eventual markdowns, while understocking can result in missed sales opportunities and potential damage to brand reputation as customers face out-of-stock scenarios (Andrade & Cunha, 2023).

The ripple effect of accurate demand forecasting also extends to customer service. When products are available and when consumers want them, it ensures a seamless shopping experience, leading to increased customer loyalty and satisfaction. On the flip side, inaccuracies in demand prediction can lead to scenarios where customers are left disappointed, having to either wait for products to be restocked or, worse, turning to competitors. Such instances not only result in immediate revenue loss but can also tarnish a brand's reputation in the long run. Lastly, the implications of accurate forecasting touch the very core of retail – profitability. Efficient supply chain management and inventory control, coupled with enhanced customer satisfaction, culminate in increased sales and reduced operational costs. The balance achieved between demand and supply, driven by precise forecasting, translates to a healthy bottom line, ensuring sustainable growth and long-term success for retailers (Ren et al, 2016).

1.2 Thesis structure

The structure of this thesis has been designed to provide a comprehensive understanding of the research objectives and findings.

Section 1 serves as the introduction and provides essential background information on demand forecasting, particularly its critical role in the fashion industry. Moreover, it gives an overview of the thesis structure.

Section 2 offers a detailed overview of Triumph International, the company at the focus of this study. This section will explore the company's history and the integral role that demand forecasting plays in its business strategy.

Section 3 includes an exhaustive literature review offering a theoretical framework, emphasizing the existing methodologies, technologies and trends in demand forecasting within the domain of the fashion retail sector.

Section 4 delves into the methodological choices made for this research. This includes data collection, preparation, and modeling. I explain why specific machine learning models were chosen and what validation techniques were employed.

Section 5 interprets the empirical results, highlighting their implications for Triumph International. Here findings are synthesized, and practical, actionable insights are offered for both Triumph International and the broader fashion retail industry.

Section 6 reflects on the study's strengths and weaknesses, and suggests future research directions.

Finally, Section 7 culminates with the conclusion and encapsulates the overall significance of the study within the context of the fashion retail industry.

2. Company Overview

2.1. History & Background

Founded in 1886, Triumph International has evolved from a small corset factory in Germany into a globally recognized leader in the lingerie industry. Today, Triumph is present in over 120 countries, offering a diverse range of products that include bras, panties, shapewear, sleepwear, and swimwear. Through its flagship brands – Triumph and Sloggi – the company caters to varied consumer preferences, emphasizing quality, innovation, and comfort. Its market reach extends dominantly across Europe and Asia, with significant operations in Japan, Germany and, especially, Austria - our target country of this research, as it represents the most characteristic and representative market for examining Triumph International's demand dynamics.

2.2. Demand Forecasting in Triumph's Business Strategy

Demand planning and forecasting is a crucial part of Triumph's business strategy, directly influencing production planning, warehouse inventory management, sales forecasting, and financial planning. Achieving precision in demand forecasting is instrumental for ensuring product availability precisely when and where consumers require it, which, in turn, elevates customer satisfaction and fortifies brand loyalty.

In the last couple of years, Triumph's approach to demand forecasting has been undergoing a notable transformation. The company is in the process of shifting from a primarily sales-driven forecasting approach to a more data-centric model, facilitated by its Integrated Business Planning (IBP) SAP system.

The internal forecasting process at Triumph, though comprehensive and multi-layered, tends to be convoluted, involving a cyclic sequence of intricate steps. This process encompasses everything from refining historical data to analyzing current order trends, and then extends to employing a range of statistical models for future projections. This sequence typically culminates in manual adjustments to these models, adding a layer of human judgment to the analytical rigor.

While Triumph's internal IBP forecasting model marks a significant stride towards a more integrated and analytical methodology, the complex and layered nature of this process presents its own sets of challenges. It primarily employs traditional statistical models for forecasting, which often rely on simple trend and seasonality analyses. They are built upon the assumption that historical orders patterns are reliable indicators of future demand, which, in the rapidly evolving retail landscape, can be a limiting perspective. Furthermore, the high level of granularity, particularly in terms of product availability and distribution channels, introduces additional layers of complexity that these traditional models might not adequately address.

Recognizing the dynamic nature of the fashion retail sector, where consumer preferences and market trends can shift rapidly, Triumph acknowledges the need to enhance its forecasting capabilities with more advanced techniques. This is where the integration of machine learning (ML) into their forecasting process becomes crucial.

3. Literature Review

After a comprehensive understanding of the thesis's objectives and background, this chapter aims to provide a well-rounded overview of demand forecasting methodologies, from their foundational concepts to the latest advancements. This comprehensive view is particularly essential for navigating the multifaceted challenges presented by the fashion retail industry, setting a foundation for the research and analysis that follows.

3.1. Forecasting methods

3.1.1. Traditional statistical approach

Historically, the fashion industry has predominantly relied on statistical methods for demand forecasting due to their ease of implementation, computational speed, and the ability to neatly integrate them into other business processes like inventory management (Ren et al., 2019). By leveraging historical data, this practice assumes that previous patterns in the past, whether seasonal, trend, cyclical, or irregular, will reappear in the future (Chopra & Meindl, 2016).

Time-series forecasting models, such as naive forecasts, moving averages, exponential smoothing and Autoregressive Integrated Moving Average (ARIMA), are particularly well-suited for products with stable demand, and they often serve as baseline models against which more advanced techniques can be compared. For example, naive forecasts are commonly used as benchmarks for predicting new item demand and daily SKU-level demand (Singh et al., 2019; Spiliotis et al., 2020). Furthermore, due to the intrinsic volatility of the fashion industry, ARIMA and its seasonal variant, SARIMA, have gained popularity for delivering quick, intuitive forecasts (Liu et al., 2013; Box et al., 2015; Ren et al., 2016).

However, despite their widespread use and convenience, traditional statistical methods are not without limitations. Most notably, they struggle to capture non-linear patterns and are sensitive to external factors, such as promotions, holidays, and weather conditions, which can serve as noise and consequently reduce forecast accuracy (Brownlee, 2017; Hyndman & Athanasopoulos, 2021). These shortcomings are especially magnified in the fashion industry, which is characterized by high fluctuations and complexities in demand patterns.

3.1.2. Machine Learning models

Given that the fashion sector is also characterized by disaggregated demand patterns across various product options (style, color and size) and different distribution channels, traditional forecasting methods often fall short in capturing the nuanced dynamics of consumer behavior. This industry-specific challenge has led to a shift toward disaggregated forecasting approaches, which aim to refine predictions at a granular level, such as per product attribute or SKU. In response, machine learning strategies have been developed to address the limitations hold by old statistical methods. These techniques are adept at uncovering complex, non-linear patterns in datasets, a capability that is particularly pertinent in the fashion industry where demand pattern can change unpredictably and swiftly. Research has extensively explored a range of machine learning models, from foundational to more advanced algorithms, to cater the intricate forecasting requirements of the fashion sector (Ren et al., 2016, Ren et al., 2020).

However, aiming to provide a balance between model complexity and interpretability for this thesis, we narrow this section only on foundational machine learning models.

The most common algorithms used in this specific field are Decision Trees. These algorithms segment data into smaller subsets and construct localized predictions within each segment. This methodology forms the basis for more sophisticated ensemble methods like Random Forest and Gradient Boosting Machines (GBM), which integrate numerous decision trees to enhance predictive accuracy and robustness. Among the gradient boosting algorithms, two variants stand out for their efficiency and effectiveness in the context of fashion retail forecasting: Light Gradient Boosting Machine (LGBM) and eXtreme Gradient Boosting (XGBoost). One of the key advantages of LGBM is its speed and lower memory usage compared to traditional gradient boosting methods. This efficiency is crucial in the fashion industry, where the volume and variety of data can be substantial. This is achieved through its innovative handling of categorical features and implementation of histogram-based algorithms for faster training and reduced memory consumption.

XGBoost, on the other hand, has been consistently placed among the top contenders in Kaggle competitions as highlighted by Chen and Guestrin (2016). XGBoost provides a robust platform for developing predictive models, offering features like built-in cross-validation and regularized boosting techniques to prevent overfitting. It is designed to be highly efficient, flexible, and portable, make it a powerful tool for predictive modeling in the fashion industry.

The utility of these models in the fashion industry is further supported by studies such as those by Andrade and Cunha (2023) and Saha et al. (2022), which demonstrate how XGBoost and LGBM can effectively capture the nuances of consumer behavior and trends in fashion.

However, these models are not without drawbacks. Firstly, the implementation of these models necessitates significant financial investment and a deep understanding of the technology. Additionally, GBMs require careful tuning of parameters to avoid overfitting, and their computational complexity can be a concern when dealing with extremely large datasets.

4. Methodology

To systematically and efficiently address the complexities of the project, the entire machine learning project was divided into five distinct phases:

1. Data Collection & Preparation
2. Exploratory Data Analysis (EDA)
3. Features Selection & Engineering
4. Model Selection & Training
5. Model Evaluation

The main programming language used to conduct the analysis was Python, leveraging a suite of powerful software packages (Table 4) to facilitate various aspects of data processing, analysis, and modeling. Structured Query Language (SQL) was mainly used for data collection.

4.1. Data Collection & Preparation

4.1.1. Data Gathering

The initial stage of this research involved gathering raw, historical data stored in Triumph's data warehouse, a system meticulously designed, managed and maintained by The Data Cooks. The method employed to access this data was through Microsoft's Azure Data Studio, leveraging SQL for precise querying and data extraction.

To align with the company's specific operational objectives and requirements, the data collection process was strategically focused on three fundamental categories: product attributes, store metadata, and order details. Each category was initially represented as an individual table within the data warehouse. By executing a series of well-crafted SQL queries, the tables were sequentially joined on common key identifiers to create a singular, unified table.

This final dataset spans a scattered daily time frame from January 2019 to August 2023.

4.1.2. Data Filtering & Cleaning

In order to maintain the accuracy, consistency, and relevance of the data used for this research, an extensive data filtering and cleaning process was implemented. The following table outlines this critical phase:

Criteria	Details	Action Taken
Brands	Triumph, Sloggi	Filtered to include only these brands.
Order Reason	Replenishment	Retained records signifying only replenishment orders
Product Lifecycle Status	Running	Selected only products in the 'Running' Phase. Product always in stock.
Country	Austria	Ensured data consistency by focusing on a single most representative market.

Returns	Negative values	Excluded to captures demand data without reverse logistics
NaNs in Product Type	24 records	Labeled as 'Not Assigned'
NaNs in Style Type	289 records	Labeled as 'Not Assigned'
NaNs in Sub Ops Channel	46 records	Removed these entries as they represent non-traditional channel.
NaNs in Cup	363,605 records	Labeled as 'Not applicable' - not all products have cups size

As a result of this meticulous process, the initial dataset containing 584,275 records was refined down to 564,505 records. This streamlined dataset, now free from any irrelevant entries and inconsistencies, served as the backbone for the subsequent stages of the research.

4.2. Exploratory Data Analysis (EDA)

The EDA phase of the research played a critical role in shaping the subsequent modelling efforts. The dataset presented a complex array of variables, which includes 17 categorical variables detailing various product attributes, 1 datetime variable for daily timestamps, and 1 numerical variable representing the quantity of orders (Table 1).

Adding to the complexity, the dataset encompassed data from 378 different stores, spanning 8 distinct distribution channels, and covered over 1100 unique options. Here, an 'option' is defined as a specific combination of material, color, and cup size of the product, a common categorization in the fashion industry, particularly in lingerie retail. This level of granularity highlighted the diverse and multi-dimensional nature of the data, posing a significant challenge in terms of both data management and predictive modeling.

Moreover, the EDA revealed a marked weekly pattern in the dataset. This cyclical trend in demand behaviour led to the strategic decision to aggregate data on a weekly basis. Weekly aggregation was chosen to effectively capture these recurring temporal patterns, while also addressing the challenge of missing days discovered in the dataset. These gaps represents lack of entries for days with zero demand. To address this issue, since the model should also predict zero (no demand), a specific approach was adopted (Read 4.5.3 Zero Handling).

4.2.1. Covid Years Handling

The retail industry was the most impacted during the pandemic years, characterized by lockdowns, shifts in consumer spending towards essentials and distribution in supply chain. These factors were extraordinary and, hopefully, not recurrent in the same form.

In the context of this research, the handling of data from the COVID years necessitated careful consideration, given their potential to skew the forecasting model towards anomalies rather than standard market behaviours. After thorough analysis and model testing, I decided to exclude the data from 2020 and 2021 from the final forecasting model. This decision was anchored in the objective of developing a robust model that reflects and adapts to standard market conditions rather than adapting it to the significant but atypical disruptions caused by the pandemic.

Supporting this decision was the insight gained from the EDA, where I observed a notable similarity in the order trends in 2019 and 2022. This confirms the continuity of market behavior in the pre and post-pandemic periods.

In light of these considerations, the model training was conducted on data from 2019 and 2022, providing a comprehensive view of typical market conditions, while the model testing was then carried out on data from 2023.

4.2.2. Outliers Handling

There has been much debate in the literature regarding what defines an outlier as whether to remove them or not, as they can sometimes hold valuable information (Osborne & Overbay, 2019). Outliers are extreme data points that deviate significantly from the majority of data. They can arise due to various reasons, such as human error, measurement inaccuracies, or they may represent legitimate occurrences in the dataset.

A key discovery during the EDA was the identification of long-tail distributions in the dataset. This finding indicated that extreme values are indeed present.

In this research, initial explorations into outlier removal showed promising results. By eliminating these extreme values, the MAPE (Read 6.1. Performance Metrics) of the initial model improved significantly, dropping from 73% to 56%. This improvement indicated the potential impact of outliers on the forecasting accuracy. However, the data set encompasses various operational channels with differing characteristics. For instance, small retail channels typically have smaller order sizes compared to larger stores or wholesale channels, where bulk purchases are more common. Such bulk orders, though appearing as outliers, are in fact legitimate and important for understanding the demand patterns of larger stores.

Understanding this, I've made the decision to not remove these extreme data points. Instead, to better manage and interpret these points, a more tailored approach was later adopted. I created separate models for each distribution channel group. By doing this, the models now are able to take into consideration the reasonable variations in order sizes that occur across various retail channel types, guaranteeing that the forecasting is accurate and indicative of the wide range of retail environments in which Triumph International works. This approach embraces the potential of outliers in a regulated and planned way, while acknowledging the complexity and unpredictability present in retail data.

4.3. Features Selection & Engineering

In the domain of machine learning, feature selection and engineering are crucial steps that significantly influence the performance of predictive models. The process of identifying relevant features and removing irrelevant, redundant, or noisy data is called feature selection, also known as variable selection, attribute selection, or variable subset (Kumar et al., 2014). On the other hand, the process of extracting features from raw data and transforming them into formats that are suitable for the machine learning models is known as feature engineering (Casari & Zheng, 2018).

4.3.1. Features Selection

To ensure the relevance and contribution of each feature towards the prediction target, I used a correlation matrix as a quantitative tool to evaluate the relationships between variables. Features that exhibited correlation coefficients lower than 0.10 with the target variable ("Orders") were considered weak predictors and were consequently removed from the dataset (Table 2):

- Product Line
- Style Type
- Product Family

4.3.2. Features Engineering

Date-related features are generated based on the datetime attribute of the dataset. By accessing the 'dt' accessor in pandas, I was able to decompose the datetime feature into several more granular components. Specifically, I extracted the following features:

- pandas.Series.dt.year : The year of the datetime.
- pandas.Series.dt.month: The month as January=1, December=12.
- pandas.Series.dt.isocalendar().week: The week ordinal of the year.

Next, to deepen the analysis and enhance the model's predictive capabilities, I decided to introduce a numerical feature to my predominantly categorical dataset: "Order Value" - the monetary amount of each transaction.

I then proceeded to derive a more granular metric, 'Value per Piece', by calculating the unit price for each item in an order. To further refine this feature and capture underlying trends, I computed a rolling average of the 'Value per Piece' over a span of 12 months.

By computing the rolling average of the 'Value per Piece', I aimed to smooth out the price fluctuations that are not just a result of seasonal consumer demand, but also of the ordering patterns and discounting policies associated with different channels. (Final dataset in Table 3)

4.4. Model Selection & Training

In addressing the time series forecasting challenge presented by Triumph International's demand, I approached the problem through the lens of regression analysis, specifically utilizing gradient boosting methods. This decision was rooted in the recognition that the nature of time series data in the retail sector, particularly for demand forecasting, often exhibits complex and disaggregated patterns and relationships that traditional time series models, such as Arima, may not capture effectively.

Gradient Boosting is an algorithm that improves its predictions by combining multiple simple models (weak learners) into a more accurate one (strong learner). It does this iteratively, where each new model focuses on correcting the errors made by the previous ones. This process repeats until it meets a specific stopping condition (Bentéjac et al., 2020).

Gradient boosting machines are a family of powerful machine-learning techniques that have shown considerable success in a wide range of practical applications. In this particular study, we will focus on the two variant of gradient boosting: XGBoost and LightGBM.

4.4.1 Data preprocessing

A comprehensive data preprocessing pipeline was developed to ensure that the dataset was optimally prepared for the subsequent modeling stages. The pipeline was structured into several key components, each designed to address specific types of features within the dataset.

This pipeline is a testament to the importance of proper data treatment in the realm of machine learning, particularly in handling diverse types of data that are typical in retail datasets.

Firstly, I utilized RobustScaler, an algorithm that scales features using statistics that are robust to outliers (Ahsan et al., 2021). For categorical variables, the pipeline incorporated two distinct approaches: Label Encoding and Ordinal Encoding. The label encoding was applied to products and store features. This transformation is crucial as it converts categorical data into a machine-readable numerical format (Dahouda & Joe, 2021). Ordinal encoding was employed for features where the order matters, such as 'Year', 'Month', 'Week', and 'Season'.

4.4.2. Hyperparameters tuning

The process of hyperparameters tuning is a critical step in optimizing the performance of machine learning models. Hyperparameters are the configurations that govern the overall behaviour of machine learning algorithms. Unlike model parameters, which are learned from the data, hyperparameters are set prior to the training process and can have a significant impact on the efficiency and effectiveness of the model. The challenge in hyperparameter tuning lies in finding the optimal combination of these settings that results in the best model performance (Shekhar at al., 2021).

4.4.2.1 Optuna

For this study, Optuna, an open-sourced hyperparameter optimization framework, is employed for tuning the hyperparameters of the gradient boosting machines (Srinivas & Katarya, 2022). The conventional approach to hyperparameter tuning, such as grid search and random search, often involves selecting parameters manually, a process that can be both time-consuming and energy-intensive. Both LGBM and XGBoost are characterized by an abundance of hyperparameters making their optimal configuration a challenging task. This is where Optuna offers a transformative solution, it streamlines the tuning process by automating it.

Its define-by-run API facilitates a dynamic and flexible approach to constructing the parameter search space, allowing for a more intuitive and adaptable optimization process. Additionally, Optuna's pruning strategy efficiently identifies and terminates unpromising trials, saving valuable computational resources. Finally, its capacity for both relational and independent sampling, coupled with distributed computing capabilities, makes Optuna an excellent tool at handling the intricate hyperparameter optimization needs of sophisticated models. (Shekhar et al., 2021)

For a detailed view of the Optuna optimization code used in this study, see Code 1 in the Appendix.

4.5. Model Evaluation

4.5.1. Performance Metrics

In order to quantify different aspects of the model performance, it is often essential to use a variety of metrics. The metrics employed in this thesis include:

- *Root mean square error (RMSE)*

  It represents the square root of the average of the squared differences between forecast and actual values (Formula 1). It provides an indication of the typical error magnitude in the system's predictions, assigning greater significance to larger errors.

- *Mean absolute error (MAE)*

  It calculates the average absolute difference between the predicted values and the actual values, offering a clear representation of the prediction accuracy (Formula 2).

- *Mean absolute percentage error (MAPE)*

  It represents the average of the absolute percentage errors (Formula 3). MAPE is simple to comprehend and explain because it is expressed as a percentage error measure. Furthermore, it gives model performance context. As a result, it serves as the main goal of several comparisons in my analysis.

(St-Aubin et al., 2022, Hyndman, & Koehler, 2006).

4.5.2. Results

In this section, we delve into the intricate dynamics of model performance across various operational channels, discuss the implications of feature importance on forecasting efficacy and present the solution to the current non-zero dataset.

Distribution Channel	**Metrics**	**LGBM**	**XGBOOST**
Own Retail	RMSE	4.93	2.95
	MAE	1.07	1.13
	MAPE	24%	28%
Department	RMSE	2.98	4.94
	MAE	1.12	1.17
	MAPE	25%	32%
Wholesale	RMSE	2.08	1.97
	MAE	1.18	1.16
	MAPE	33%	39%
Field	RMSE	6.88	6.67
	MAE	2.07	2.06
	MAPE	34%	41%
Key Account Hyperstore	RMSE	66.72	66.83
	MAE	16.08	16.35
	MAPE	63%	67%

For the **Own Retail** channel, Triumph's branded stores and outlets, the preferred model is LGBM. It has a lower MAPE (24% vs. 28%), suggesting that its predictions are proportionally closer to the actual values, which is crucial for managing diverse product lines and preventing stock discrepancies in individual stores. For **Department**, LGBM outperforms in all metrics, including the lowest MAPE (25% vs. 32%). The XGBoost model is preferred for the **Wholesale** channel. Although its MAPE is higher (39% vs. 33%), it has a marginally better RMSE and MAE. This suggests that while the LGBM model is more accurate in percentage terms, the XGBOOST model may be better at predicting the actual demand volumes that are critical in wholesale transactions where large quantities are involved. **Field** channels require a delicate balance because they include boutique stores where demand can be volatile and less predictable. Here, the LGBM model's significantly lower MAPE is advantageous. Despite XGBOOST's marginally better RMSE and MAE, the lower MAPE of LGBM suggests it will provide forecasts that are more proportionally accurate relative to the actual sales, which is crucial for maintaining operational efficiency in such varied environments. Finally, for **Key Account Hyperstore** channel, the preferred model is LGBM if we look at the (63% vs 67%). In the context of hyperstores, where large volumes and a wide array of products are managed, this proportional accuracy is crucial for maintaining inventory levels and minimizing overstocking or understocking risks. However, the high error rate in this specific channel suggests that both models require further refinement.

To summarize, each distribution channel presents unique challenges and demands specific to its operations, suggesting that there is no one-size-fits-all model.

4.5.3 Zero Handling

Acknowledging the significance of zero demand entries, I implemented a specific approach. This strategy involved creating a 'zero table' through SQL queries, encompassing every combination of product attributes – specifically Material, Color, and Cup features. The aim of this table was to systematically represent days with no recorded demand, effectively filling the gaps in the original dataset. Given the considerable volume of the resulted dataset, approximately 750 million rows, the analysis was strategically streamlined. The focus was confined to the 'Own Retail' channel, which had exhibited promising outcomes in early model iterations. Within this segment, the dataset was further narrowed to include a select sample: the top 10 products with the highest demand, the bottom 10 with the least demand, and 10 randomly selected products. This approach of creating a small, diverse subset was not aimed at comprehensive model training but rather at exploring a potential solution to the zero-demand issue. In line with this exploratory approach, a two-stage modeling process was employed. Initially, a classification model (LightGBM Classifier) was used to predict the likelihood of a product being ordered. Notably, the classification model achieved a high accuracy score of 94%, indicating its effectiveness in distinguishing between zero and non-zero demand days. However, the second step, which employed a regression model to estimate the quantity for those products anticipated to be ordered, did not yield as promising results. The performance of this approach, as indicated by the MAPE (40%), highlighted the need for further improvements. The implementation of this zero handling strategy, particularly towards the project's conclusion, was primarily designed to demonstrate a methodological approach to address the issue of missing zero-demand data. The insights gained from this exercise are invaluable for future projects, providing a foundation for more robust and accurate demand forecasting models in the retail sector.

5. Discussion and Practical Implications

The comparative analysis of the LGBM and XGBoost models across Triumph's distribution channels reveals significant opportunities, even considering the potential for improvement in their accuracy. These models substantially contribute to various business aspects, notably in inventory management, customer satisfaction, supply chain management, and operational efficiency.

In inventory management, the implementation of accurate demand forecasting through these models is invaluable. By maintaining optimal inventory levels, Triumph effectively reduces the need for markdowns to clear excess stock, adhering to its strategy of minimal discounting. This approach not only aligns with the brand's philosophy of providing 'the right product at the right time in the right store' but also ensures efficient use of storage space, a critical cost factor. Moreover, streamlined inventory mitigates risks like safety hazards and merchandise damage, fostering a more organized and safer store environment, which is essential for compliance with health and safety regulations. For customer satisfaction, the ability of these models to predict the right product quantities directly enhances the shopping experience. Precise meeting of customer needs boosts satisfaction levels and reduces the workload on staff managing inventory, improving the in-store experience for both customers and employees. When it comes to supply chain management, enhanced demand forecasting provides Triumph with a stronger position in supplier negotiations. This knowledge allows for strategic budget usage, including leveraging volume discounts without over-purchasing risks, leading to more informed purchasing decisions and maximizing net sales. Operational efficiency is another critical area where these models make a significant impact. By streamlining tasks like inventory management, ordering, and stock placement, they translate into considerable time savings and cost reductions, freeing staff to focus on customer service and other value-added activities.

Having said that, the LGBM and XGBoost models, despite the current limitations in accuracy, hold substantial promise in refining Triumph's operational strategies and enhancing customer engagement. This forms a foundation for sustained growth and adaptability in the ever-evolving landscape of fashion retail.

This thesis, through its in-depth analysis and practical implications, aims to serve a broad spectrum of stakeholders. Triumph International stands to benefit directly as it seeks to revolutionize its demand forecasting practices. The empirical evidence provided here can inform their decision-making processes, leading to improvements in inventory management and overall operational efficiency. For the academic community, this work offers a robust framework and valuable insights for future studies in retail demand forecasting. Moreover, the retail industry at large can learn from this study, which showcases both the potential and the challenges of implementing machine learning in retail operations. The insights gained are an invaluable resource for other retail organizations contemplating a shift towards more technologically advanced forecasting methods. Furthermore, The Data Cooks, the data agency overseeing this research and under whose I undertook this internship, also stand to gain significantly. This study serves as a practical case study for future machine learning projects with other clients, providing validated methodologies and best practices that can be adapted to different sectors within the retail industry.

6. Strengths, Limitations and Future Work

This thesis represents a significant undertaking in the realm of demand forecasting for Triumph International. A key strength of this project lies in its practical application to a real business scenario. Working with Triumph International, a client with no prior experience in advanced forecasting techniques, provided a valuable opportunity to apply theoretical knowledge to actual business needs. Another strong point of my research is the integration and application of sophisticated data preprocessing and hyperparameters optimization techniques that contributed to the robustness and reliability of this research.

However, forecasting is a very difficult task, and my background in data science and machine learning is relatively limited. Another notable challenge was the interruption of this project for several months since the start of my internship. While it is a very common scenario in real-time project work, it presented difficulties in maintaining continuity in the research. Given these circumstances, the project inevitably faced its share of limitations. One notable weakness is the nature of the dataset. The lack of zero-order entries may lead to an overly optimistic forecast, assuming a continuous flow of orders, while scattered days compromise the model's ability to accurately capture seasonal variations and cyclic behaviors. In addition, the exclusion of pandemic-affected years, while necessary to preserve data normality, resulted in a reduction of the dataset size. This means less data points for the model to learn from. Moreover, the project's scope, being limited to a single country, may restrict the generalizability of the findings to other markets with varied consumer behaviors and economic conditions. Another limitation is the lack of external factors such as weather, promotion, and socio-economic features. I acknowledge that these factors, if not captured in the dataset, can lead to discrepancies between forecasted and actual demand. Another significant limitation, was the restricted computational resources in-house, which limited the handling of extensive datasets. Due to this constraint, the research could not address the zero-order table, which encompassed around 750 million rows.

Looking ahead, to address these limitations and build on the current research, several promising areas can be explored. For example, to enhance the robustness of these forecasting models, it would be interesting to apply more advanced data collection techniques, such as web scraping and API integrations, to gather relevant external data mentioned above. Moreover, sentiment analysis on social media and consumer trends could offer additional layers of predictive power.

Further advancements could include the application of neural networks or hybrid models. These techniques have shown remarkable success in handling complex patterns and large datasets, making them well-suited for the nuanced demands of fashion retail forecasting (Frank et al. 2003). Another area for exploration is the integration of clustering techniques for the products offered. Grouping similar products could provide more insights into consumer behavior and preferences. This approach would allow for a more segmented analysis, where demand predictions could be tailored to specific product clusters, enhancing the granularity and accuracy of the forecasts. Moreover, with the rapid advancement in technology, especially in the era of big data, investing in improved computational resources is imperative. Enhanced computational capabilities would enable the handling of larger datasets and more complex models, thus breaking new ground in predictive analytics.

7. Conclusion

This thesis embarked on the journey to utilize advanced machine learning techniques to enhance demand forecasting in the fashion retail sector, with a specific focus on Triumph International. Throughout this research, I have thoroughly investigated the capabilities of sophisticated predictive models, particularly emphasizing the role of gradient boosting algorithms like LightGBM and XGBoost. One of the pivotal discoveries of this study is the realization that there is no 'one-size-fits-all' model in demand forecasting for a multifaceted enterprise like Triumph International. By implementing distinct models for each distribution channel, it became evident that each channel possesses unique characteristics and demands a tailored approach. This insight reinforces the importance of customization in predictive modelling, ensuring that forecasts are not just data-driven but also contextually aligned with specific operational needs.

Reflecting on the journey, this project offered an invaluable opportunity for me to spearhead a complete machine learning project, encompassing everything from data collection to model evaluation. Our data science team, though small and collectively new to this realm, provided invaluable support and collaboration, pivotal in overcoming the challenges and achieving the milestones of this project.

In conclusion, this research has not only laid a robust foundation for future explorations in demand forecasting but also marks a significant personal and professional milestone.

Appendix

Table 1

Initial dataset

Variable Name	Description	Data type
O_d_date	Date of order placement	datetime
O_d_store	Identifier of the store where the order was placed for.	object
Orders	Quanity of the order	int64
P_MaterialId	Unique identifier for the style of the product	object
P_ColourId	Unique identifier for the color of the product	object
P_CupId	Unique identifier for the cup size of the product	object
P_BrandCode	Code for the brand of the product	object
P_SubBrandCode	Code for sub-brand categorization	object
P_GlobalSeriesCode	Code series identification for matching product sets.	object
P_GlobalSegmentCode	Code to identify the 'occasion' for a product (e.g. Curves, Everyday, Fashion)	object
P_ProductLineCode	Code for line categorization (e.g., home wear, intimate apparel)	object
P_ProductGroupCode	Specific group categorization (e.g., Brief, Corsetry)	object
P_ProductTypeCode	Type categorization within product groups	object
P_StyleTypeCode	Code for seasonal or basic style.	object

P_SeasonCode	Code for the product's intended season.	object
P_ProductFamily	Code for specific product families with variations	object
P_PLCStatusCode	Status of the product in its product life cycle (PLC)	object
ST_CountryCode	Country code of the store	object
ST_Ops_Sub_Channel_ID	Code for distribution channel	object

Table 2

Correlation matrix

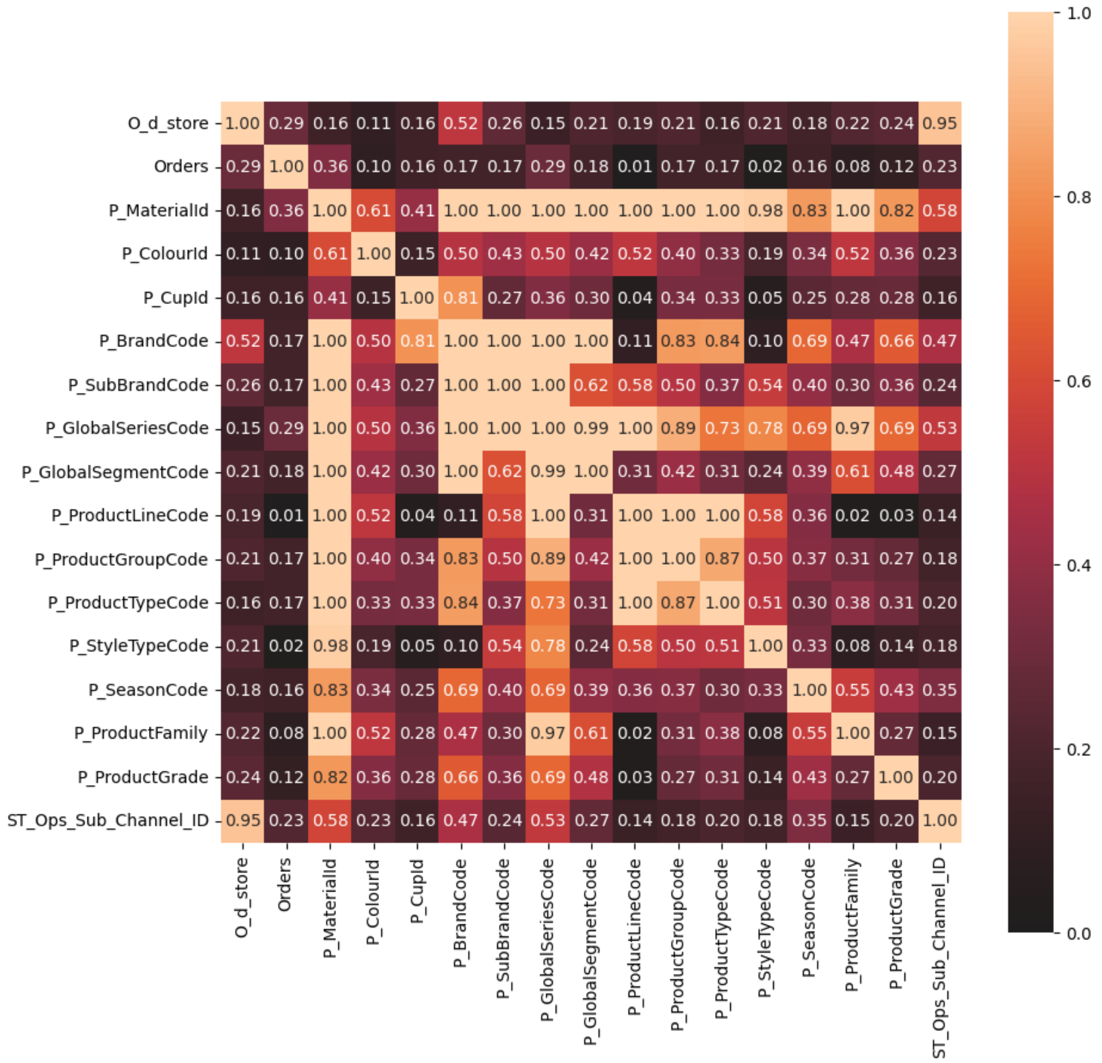

Table 3

Final dataset

Variable Name	Description	Data type
Year	Year of order placement	int64
Month	Month of order placement	int64
Week	Week of order placement	int64
O_d_store	Identifier of the store where the order was placed for.	object
Orders	Quanity of the order	int64
value_per_piece_mov_avg	Moving average of the value per individual product piece	Float64
P_MaterialId	Unique identifier for the style of the product	object
P_ColourId	Unique identifier for the color of the product	object
P_CupId	Unique identifier for the cup size of the product	object
P_BrandCode	Code for the brand of the product	object
P_SubBrandCode	Code for sub-brand categorization	object
P_GlobalSeriesCode	Code series identification for matching product sets.	object
P_GlobalSegmentCode	Code to identify the ‘occasion’ for a product (e.g. Curves, Everyday, Fashion)	object
P_ProductGroupCode	Specific group categorization (e.g., Brief, Corsetry)	object
P_ProductTypeCode	Type categorization within product groups	object
P_SeasonCode	Code for the product's intended season.	object

Table 4

Python packages used in the analysis

Packages	Purpose
pandas	Data manipulation and analysis
numpy	Scientific computing, handling arrays
seaborn	High-level statistical graphics
plotly	Interactive and aesthetic data visualization
pyodbc	Accessing ODBC databases
pickle	Saving models
optuna	Hyperparameters optimization
sklearn	Machine Learning tools and algorithms
lightgbm	Light Gradient Boosting Machine library
xgboost	Extreme Gradient Boosting library

Code 1

Optuna for LGBM

```python
def objective(trial, X_train_prepared, y_train, X_test_prepared, y_test, metric="mape"):
    param_grid = {
        "random_state": 42,
        "verbosity": 0,
        "n_estimators": trial.suggest_int("n_estimators", 50, 500),
        "lambda_l1": trial.suggest_float("lambda_l1", 0, 100),
        "lambda_l2": trial.suggest_float("lambda_l2", 0, 100),
        "max_bin": trial.suggest_int("max_bin", 32, 255),
        "max_depth": trial.suggest_int("max_depth", 3, 12),
        "min_gain_to_split": trial.suggest_float("min_gain_to_split", 0, 15),
        "feature_fraction": trial.suggest_float("feature_fraction", 0.1, 1.0),
        "learning_rate": trial.suggest_float("learning_rate", 1e-3, 1e-1, log=True),
        "num_leaves": trial.suggest_int("num_leaves", 2, 256),
        "min_data_in_leaf": trial.suggest_int("min_data_in_leaf", 1, 100),
        "bagging_fraction": trial.suggest_float("bagging_fraction", 0.1, 1.0),
        "colsample_bytree": trial.suggest_float("colsample_bytree", 0.1, 1.0),
        "boosting_type": trial.suggest_categorical("boosting_type", ["gbdt", "dart"]),
        "objective": trial.suggest_categorical("objective", ["regression", "regression_l1", "mape"])
    }

    model = LGBMRegressor(**param_grid)
    model.fit(X_train_prepared, y_train)

    y_pred = model.predict(X_test_prepared)

    if metric == "mae":
        error = mean_absolute_error(y_test, y_pred)
    elif metric == "mape":
        error = mean_absolute_percentage_error(y_test, y_pred)

    return error
```

Formula 1

RMSE

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

Formula 2

MAE

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}.$$

Formula 3

MAPE

$$\mathrm{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

References

Ahsan, Md, M. Mahmud, Pritom Saha, Kishor Gupta, and Zahed Siddique. 2021. "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance." Technologies 9 (3): 52. https://doi.org/10.3390/technologies9030052.

Andrade, Luiz Augusto C. G., and Claudio B. Cunha. 2023. "Disaggregated Retail Forecasting: A Gradient Boosting Approach." Applied Soft Computing 141 (110283): 110283. https://doi.org/10.1016/j.asoc.2023.110283.

Bentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. 2021. "A Comparative Analysis of Gradient Boosting Algorithms." Artificial Intelligence Review 54 (3): 1937–67. https://doi.org/10.1007/s10462-020-09896-5.

Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. Time Series Analysis: Forecasting and Control, 5th Edition | Wiley. New York: Wiley.

Brownlee, J. 2017. "Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future." Machine Learning Mastery.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM.

Chopra, Sunil, and Peter Meindl. 2016. Supply Chain Management: Strategy, Planning, and Operation, Global Edition. 6th ed. London, England: Pearson Education.

Dahouda, Mwamba Kasongo, and Inwhee Joe. 2021. “A Deep-Learned Embedding Technique for Categorical Features Encoding.” IEEE Access: Practical Innovations, Open Solutions 9: 114381–91. https://doi.org/10.1109/access.2021.3104357.

Frank, Celia, Ashish Garg, Les Sztandera, and Amar Raheja. 2003. “Forecasting Women’s Apparel Sales Using Mathematical Modeling.” International Journal of Clothing Science and Technology 15 (2): 107–25. https://doi.org/10.1108/09556220310470097.

Hyndman, Rob J., and George Athanasopoulos. 2021. Forecasting: Principles and Practice. 3rd ed. Australia: OTexts.

Hyndman, Rob J., and Anne B. Koehler. 2006. “Another Look at Measures of Forecast Accuracy.” International Journal of Forecasting 22 (4): 679–88. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Ingle, Chaitanya, Dev Bakliwal, Jayesh Jain, Preeyesh Singh, Preeti Kale, and Vaibhav Chhajed. 2021. “Demand Forecasting : Literature Review on Various Methodologies.” In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE.

Kumar, Vipin. 2014. “Feature Selection: A Literature Review.” Thesmart Computing Review 4 (3). https://doi.org/10.6029/smartcr.2014.03.007.

Liu, Na, Shuyun Ren, Tsan-Ming Choi, Chi-Leung Hui, and Sau-Fun Ng. 2013. “Sales

Forecasting for Fashion Retailing Service Industry: A Review." Mathematical Problems in Engineering 2013: 1–9. https://doi.org/10.1155/2013/738675.

Nenni, Maria Elena, Luca Giustiniano, and Luca Pirolo. 2013. "Demand Forecasting in the Fashion Industry: A Review." International Journal of Engineering Business Management 5: 37. https://doi.org/10.5772/56840.

Osborne, Jason W., and Amy Overbay. 2004. "The Power of Outliers (and Why Researchers Should ALWAYS Check for Them)." University of Massachusetts Amherst. https://doi.org/10.7275/QF69-7K43.

Ren, Shuyun, Hau-Ling Chan, and Pratibha Ram. 2017. "A Comparative Study on Fashion Demand Forecasting Models with Multiple Sources of Uncertainty." Annals of Operations Research 257 (1–2): 335–55. https://doi.org/10.1007/s10479-016-2204-6.

Ren, Shuyun, Hau-Ling Chan, and Tana Siqin. 2020. "Demand Forecasting in Retail Operations for Fashionable Products: Methods, Practices, and Real Case Study." Annals of Operations Research 291 (1–2): 761–77. https://doi.org/10.1007/s10479-019-03148-8.

Saha, Priyam, Nitesh Gudheniya, Rony Mitra, Dyutimoy Das, Sushmita Narayana, and Manoj K. Tiwari. 2022. "Demand Forecasting of a Multinational Retail Company Using Deep Learning Frameworks." IFAC-PapersOnLine 55 (10): 395–99. https://doi.org/10.1016/j.ifacol.2022.09.425.

Shekhar, Shashank, Adesh Bansode, and Asif Salim. 2021. "A Comparative Study of Hyper-

Parameter Optimization Tools.” In 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE.

Singh, Pawan Kumar, Yadunath Gupta, Nilpa Jha, and Aruna Rajan. 2019. “Fashion Retail: Forecasting Demand for New Items.” arXiv [Cs.OH]. http://arxiv.org/abs/1907.01960.

Spiliotis, Evangelos, Spyros Makridakis, Artemios-Anargyros Semenoglou, and Vassilios Assimakopoulos. 2022. “Comparison of Statistical and Machine Learning Methods for Daily SKU Demand Forecasting.” Operational Research 22 (3): 3037–61. https://doi.org/10.1007/s12351-020-00605-2.

Srinivas, Polipireddy, and Rahul Katarya. 2022. “hyOPTXg: OPTUNA Hyper-Parameter Optimization Framework for Predicting Cardiovascular Disease Using XGBoost.” Biomedical Signal Processing and Control 73 (103456): 103456. https://doi.org/10.1016/j.bspc.2021.103456.

St-Aubin, Philippe, and Bruno Agard. 2022. “Precision and Reliability of Forecasts Performance Metrics.” Forecasting 4 (4): 882–903. https://doi.org/10.3390/forecast4040048.

Swaminathan, Kritika, and Rakesh Venkitasubramony. 2024. “Demand Forecasting for Fashion Products: A Systematic Review.” International Journal of Forecasting 40 (1): 247–67. https://doi.org/10.1016/j.ijforecast.2023.02.005.

Zheng, A., and A. Casari. 2018. Feature Engineering for Machine Learning: Principles and

Techniques for Data Scientists.