Modeling the pricing strategy of an e-commerce luxury fashion retailer: A machine learning approach

Diogo Vieira da Silva

Master's Dissertation

Supervisor: Prof. Dr. Mário Amorim Lopes

U. PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Integrated Master in Industrial Engineering and Management

2018-06-15

Abstract

Pricing has always been a process of the utmost importance for retailers. The amount charged to a customer conditions the success of a transaction and in turn, the success of a business. In the past decades, e-commerce has completely revolutionized the way companies operate in an increasingly competitive and saturated marketspace. Due to the level of control and traceability that this new business model provides, retailers can now conveniently craft customized and creative pricing strategies that range from shipping subsidy fees to margins on other transaction related costs (eg: duties).

In this dissertation we propose a data mining approach to model the impact different components of the pricing structure of the luxury fashion e-tailer Farfetch have on the company's financial strength and customer satisfaction. A comprehensive study of the current pricing mechanisms is done. Data from the past two years regarding orders, customer feedback and other key indicators is collected and processed. Multiple machine learning algorithms are tested and tuned for ordinal classification problems. In addition, a new metric for model performance evaluation and comparison in ordinal classifiers is developed, and, the machine learning approach is compared to a personalized exploratory analysis of a real Farfetch case study in order to assess the viability of automated predictive tools in supporting strategic pricing decisions.

This study provides a framework for structuring a model for the evaluation of price sensitivity in several key business indicators for e-commerce retailers and how to assess its viability by comparing it to a manual personalized exploratory analysis.

For the machine learning approach, the results suggest that free shipping campaigns and sales events are the most influential variables in sales development, closely followed by the remaining pricing components. For the personalized study, multiple actionable insights were drawn: free shipping drove forth business growth in China but increased the country's return rate by 4 p.p due to the customers taking advantage of the permanent free shipping conditions; increasing China's shipping subsidy rate to 5.79% would offset all additional shipping costs resultant from the new shipping conditions.

The results suggest that there might be potential for machine learning to support e-commerce pricing decisions if there are enough pricing variations for the model to learn from, otherwise a deep-dive exploratory analysis might be a better option.

Resumo

A definição de preço sempre foi um processo de extrema importância para retalhistas. O valor cobrado a um cliente condiciona o sucesso de uma transação e, por sua vez, o sucesso de uma empresa. Nas últimas décadas, o comércio eletrónico revolucionou por completo a forma como as empresas operam num mercado progressivamente mais competitivo e saturado. Devido ao nível de controlo e rastreabilidade que este novo modelo de negócio oferece, os retalhistas podem agora criar estratégias de preço customizadas e altamente criativas que tanto contemplam taxas de subsídio de remessa até margens sobre outros custos relacionados a transações (Ex: Taxas alfandegárias).

Nesta dissertação, uma abordagem de *data mining* para modelar o impacto que diferentes componentes na estrutura de preço do *e-tailer* de moda de luxo Farfetch têm sobre a força financeira da empresa e a satisfação do cliente é proposta. Um estudo abrangente dos mecanismos de preços atuais é feito. Os dados dos últimos dois anos de encomendas, *feedback* de clientes e outros indicadores importantes são recolhidos e processados. Vários algoritmos de *machine learning* são testados e afinados para problemas de classificação ordinal. Para além disso, uma nova métrica para a avaliação e comparação do desempenho de modelos preditivos para classificação ordinal é desenvolvida. A abordagem de *machine learning* é comparada com uma análise exploratória personalizada de um caso de estudo real da Farfetch de modo a avaliar a viabilidade de ferramentas preditivas automatizadas no suporte a decisões estratégicas de definição de preço.

Este estudo desenvolve uma estrutura sobre como abordar e estruturar um modelo para a avaliação da sensibilidade a preço de vários indicadores-chave de negócios para retalhistas de comércio eletrónico e como avaliar sua viabilidade, comparando-os com uma análise exploratória personalizada manual.

Relativamente à abordagem de *machine learning*, os resultados revelam que as campanhas de entrega gratuita e os eventos promocionais são as variáveis que mais condicionam o desenvolvimento de vendas, seguidas das restantes componentes de preço. Para o estudo personalizado, vários *insights* acionáveis foram extraídos: a entrega gratuita impulsionou o crescimento de negócio na China, mas aumentou a taxa de devoluções do país em 4 p.p devido aos clientes explorarem a seu proveito das condições de entrega gratuita permanente; Aumentar a taxa de subsídio de transporte da China para 5.79 % compensaria todos os custos de envio adicionais resultantes das novas condições de entrega.

Os resultados sugerem que pode haver potencial para *machine learning* apoiar as decisões de definição de preço em comércio eletrónico se houver variações de preço suficientes para o modelo conseguir aprender a relação entre variação de preço e o seu respetivo impacto no negócio, caso contrário, uma análise exploratória de *deep-dive* poderá ser uma opção melhor.

Acknowledgements

I want to thank all of my colleagues from Delivery Development for their constant support, patience and, above all, friendship. Never had I dreamed that in such a short time this team could end up meaning so much to me and help me grow as an individual and as a professional. Diana, Hugo, Inês, Isabelle, Pedro, Tomás and Wilson - A most sincere thank you!

Ana Castro, more than a manager, you were a true friend. Thank you for guiding me both professionally and personally. I am looking forward to having a place to stay when I go visit you in your future endeavours. André, thank you for being an amazing "partner in crime", both in work and in life. I will never forget the great moments that we shared and can't wait for the ones that are yet to come.

Thank you Mário Amorim for helping me structure and see the right approach to follow. Your help was invaluable. Last but not least, I owe a big thank you to all of my closest friends and family for always having my back and unconditionally supporting me throughout my path - I would not be who I am without you all and for that I am very grateful.

More than a mere technical challenge, this project was an important learning experience. From it, I learned the value of teamwork and made incredible friends that I will take with me along the road that is life. Again, thank you all!

"I'm for anything that gets you through the night."

Frank Sinatra

Contents

Acronyms and Symbols

KPI	Key Performance Indicator
FS	Free Shipping
RRP	Recommended Retail Price
VAT	Value Added Tax
FX	Forex
D&T	Duties and Taxes
DDP	Delivered Duty Paid
DAP	Delivered At Place
GMV	Gross Merchandise Value
ATV	Actual Transaction Value
NTV	Net Transaction Value
AOV	Average Order Value
P&L	Profit and Loss
NPS	Net Promoter Score
CI	Confidence Interval
SVM	Support Vector Machines
MSE	Mean Squared Error
OLR	Ordinal Logistic Regression
NB	Naive Bayes Classifier
RF	Random Forests
WA	Window Accuracy
OAS	Overall Adjusted Sensitivity
YoY	Year over Year

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Pricing has long been a critical factor in the success of competitive strategies in most sectors of activity. According to a study from Stax Inc (Stax Inc, 2016), price remains the most relevant decision driver in most industries, with nearly half of the survey's participants electing it as one of their most important considerations during purchase decisions. With this in mind, it is fair to consider that the consumers' price sensitivity is a most valuable insight for any strategic or tactical plan.

With the rise of e-commerce channels, retailers are now presented with a wide range of tools that allow to explore patterns on the sales records and segment customers according to their specific characteristics (Vulkan, 2003), paving the way towards more effective price discrimination strategies (Vlasova, 2017). E-commerce's potential for smarter pricing policies has been a subject of interest for many researchers in the past years. (Archak et al., 2011) demonstrates the effectiveness of text mining in extracting important insights on the consumers' most valued product features in relation to the price paid. In addition, a report from the Executive Office of the President of the United States of America shows that personalized coupons campaigns can increase redemption rates as much as 25% (White House, 2015).

The range of possibilities is vast and companies are quickly recognizing the effectiveness of such analytical approaches to e-commerce pricing, showcasing growing tendencies towards price steering and price discrimination tactics online (Hannak et al., 2014). This study aims to contribute to this area of knowledge by exploring the effect of different price components in customer satisfaction and financial performance, in the context of a real luxury fashion e-tailer.

1.2 Farfetch

According to a report from Bain & Company (Bain & Company, 2018), the personal luxury goods market is expected to grow at a 5% compound annual rate over the next three years. At the same time, e-commerce is quickly driving forth the transition from offline to online channels for fashion companies, with China presenting the highest average growth rate of 18% (Statista, 2018). These insights support the existence of favorable market conditions in the high-end fashion industry, with prospects of growth attracting new players and raising up the competitiveness of the sector.

Farfetch is an e-commerce marketplace for luxury fashion that connects over 900 boutiques and brands to a customer base of over 2 million people around the world. The company was launched in 2007 by founder and CEO José Neves and has grown in the past years at an approximate year-over-year rate of 70%, quickly rising as an industry leader. Farfetch acts as an enabler for transactions performed between boutiques and customers, providing an online selling platform with worldwide availability that benefits both parties: the customer is presented with a diverse arrangement of luxury fashion products that are usually only available for offline trade, while the boutiques gain access to a bigger potential market due to the integration of an e-commerce sales channel.

One of the most differentiating characteristics of Farfetch's business model is that it employs a drop shipping model. The company does not keep stock in warehouses to supply demand but instead manages the delivery service provided by third-party logistics (3PL) providers that guarantee the transportation of an order from the boutique to the customer. This approach has both advantages and disadvantages. On the one hand, Farfetch is freed from the high holding costs related to luxury goods' stock keeping and is able to provide a varied assortment of products. On the other hand, the company is heavily dependent on the performance of external partners.

In order to sustain and promote further growth in such a competitive industry, Farfetch is constantly searching for ways to innovate and improve the service that it provides to its customers and partners. It is in this context that this dissertation was carried out.

1.3 Objectives and methodology

The project at hand was developed in Farfetch's E-commerce Operations department, more specifically in the Delivery Development team. In the past years, Farfetch has introduced and changed specific price components that affect the value paid by the customer. However, prior to this study, there had been no projects aimed at fully comprehending how these alterations impacted demand and customer satisfaction in different regions. With this in mind, the main goal of this dissertation was to develop models capable of predicting the impact of pricing alterations on the company's main business indicators.

Figure 1.1 summarizes the main steps of the project, including their duration and their respective order.

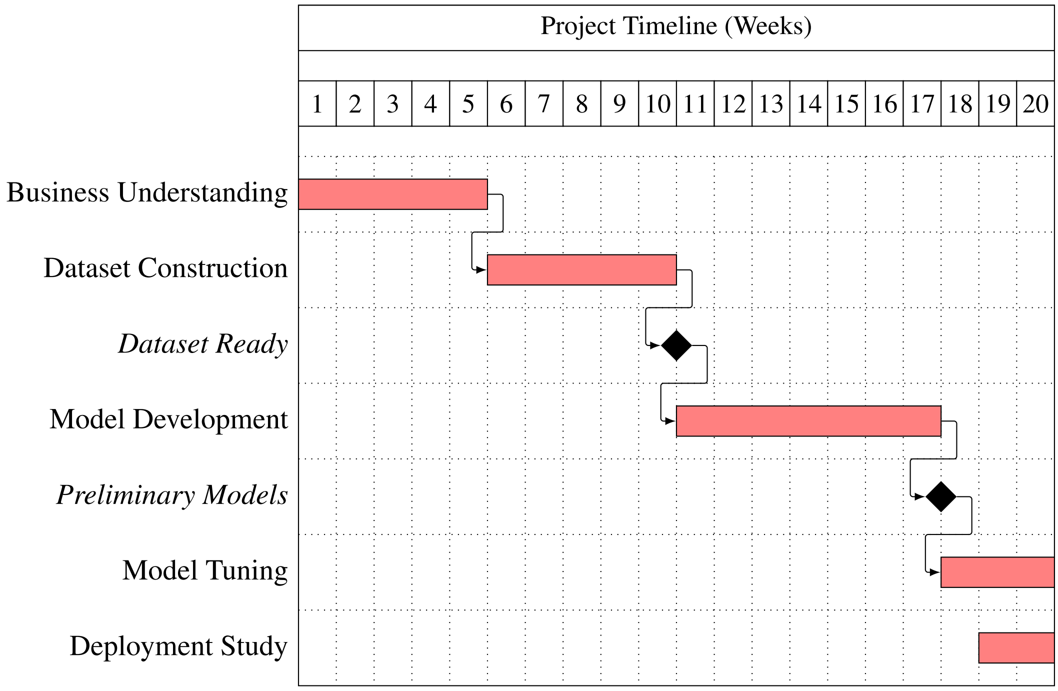


Figure 1.1: Gantt chart for the main project stages, based on the CRISP-DM Model (Shearer, 2000).

As an initial step of the project, a thorough study of the pricing mechanisms and database structure was performed (**Business Understanding**). Afterwards, datasets containing the most important target variables and relevant inputs for the priority countries was constructed, taking into consideration the feedback from more experienced members of the team (**Dataset Construction**). Several approaches to modeling the pricing components' impact were tested through classifiers and regression-based predictors. In this stage, the models developed for each business indicators were compared and evaluated through different performance metrics (**Model Development**). As a final step, the main hyperparameters of the selected models were tuned in order to maximize the predictions' accuracy. In addition, a study for the deployment of the tool was done (**Model Tuning** and **Deployment Study**).

1.4 Dissertation's structure

The remainder of this dissertation is structured as follows: in Chapter 2, the theoretical framework for this study is presented, exploring the most relevant concepts of machine learning and pricing, as well as introducing the market of luxury fashion e-commerce. Chapter 3 offers an overview of the pricing structure at Farfetch, introducing company-specific terminology for pricing components and contextualizing the problem at hand. Chapter 4 describes the definition of the dataset, the method used for processing and exploring initial data **and reveals the main insights extracted from the deep-dive exploratory study that was used for comparison**. Chapter 5 complements the previous one by presenting the variable transformation mechanisms used, the predictive algorithms applied and the performance evaluation metrics constructed for the problem at hand. Chapter 6 reports the results obtained for the selected models and for the personalized study, comparing them between each other and discussing the results from an overall perspective. Chapter 7 draws the conclusions of the project and paves the way for possible follow-up studies.

Chapter 2

Theoretical Background

2.1 E-commerce

2.1.1 Definition and Contextualization

The rise of the digital era revolutionized the way people communicate and companies operate. In the past decades, one of the most important innovations was the development of electronic commerce. In its essence, the term e-commerce reflects the trade of goods or services through digital networks (Laudon and Laudon, 2001). In comparison to traditional retail, e-tail has clear advantages for both the companies and the buyers: On the one hand, companies are able to market their products with nearly no geographical constraints and are unburdened of the heavy costs related to operating a wide chain of physical stores; On the other hand, customers are provided with a highly transparent marketplace that allows agile comparison between the offers of different competitors, through an intuitive interface (Khan, 2016). Table 2.1 presents the features of e-commerce that mostly drive its competitive potential.

Table 2.1: E-commerce features that enhance customer experience. Adapted from (Ahmed et al., 2002).

Web Element	Effect on consumers
Interactivity	Compensates for the lack of human presence
Fast Service	Saves time
Convenience	Provides goods and services at anytime and anyplace
Personalization	Empowers consumers to be co-creators
Customization	Gives a sense of individual recognition
Privacy	Associates the brand with ethics
Real-time communication	Shows excellent customer relationship management
Security	Breeds brand trust and loyalty
Instant product availability	Provides instant gratification
Low transaction costs	Saves mental energy and time
Additional features	Creates an enhanced experience with educational benefits

In their work, (Fruhling and Digm, 2000) support the potential of e-commerce by showcasing how it provides effective tools for companies to be more competitive on the main business-level strategies, such as: *differentiation*, *cost leadership*, *added-value*, and so on. Additionally, according to (Pasumarthy and Kumar, 2015), e-commerce is ranked as the third most popular internet activity, preceded only by e-mail usage and web browsing.

At the moment, the disruptive potential of electronic commerce is becoming more clear to companies. For example, online marketplaces are starting to become the primary retail channel in Asia, surpassing more traditional inventory-led companies (Deloitte, 2015). With the growing adoption of the internet, smartphone usage and increasing transactional trust, e-commerce is well set to play a vital role on how the current retail paradigm will develop on the upcoming years (Efendioglu et al., 2005).

Figure 2.1 presents how e-commerce has grown to impact more and more the way retail operates nowadays.

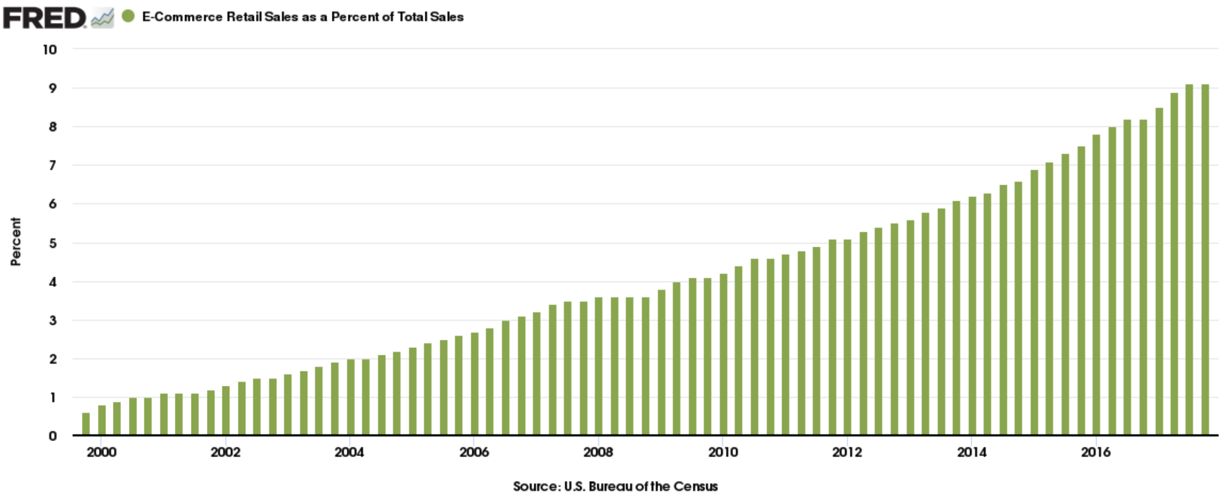


Figure 2.1: Growth of the e-commerce retail sales' contribution to total sales (U.S. Bureau of the Census, 2018).

2.1.2 Pricing Strategy and Concurrent Practices

Throughout history, price has always been a decisive factor in transactional decisions. Before the development of the 4Ps of the Marketing Mix (*price*, *product*, *place*, *promotion*), price was a critical factor in decision making, as seen in the microeconomics theory (Chong, 2003). Additionally, according to (Kurtz and Boone, 1987), price is viewed as the most significant "P" of the Marketing Mix on the perspective of most business owners.

In an e-tailer's perspective, price can, in some cases, be more critical to purchase decisions because of the low search costs related to price comparison (Dinerstein et al., 2014) and the possibilities that e-commerce creates for personalized pricing discrimination strategies (Vlasova, 2017).

In terms of research, (Hwang and Kim, 2006) present an algorithm for the dynamic optimization of the advertised price that is particularly useful for companies with a diverse assortment of products, in which the manual adjustment of the price can be costly and very time consuming. (Lewis, 2006) studies the impact of shipping fees in an online grocer and more specifically how customer retention and average expenditure is affected by the employment of shipping-related fees to the price paid. In his paper, (Shin, 2001) reflects on how e-tail allows for a more effective use of promotions and sales coupons as these customer engagement tools can be customized to meet their needs and wants, thus helping to drive a more profitable and loyal client base when compared with traditional target-free promotional efforts.

As demonstrated above, pricing strategies for electronic commerce have been a research topic of great importance in the past two decades. The flexibility that an online platform brings to pricing decisions is one of the main reasons that nowadays retailers are able to tailor their approach to each customer segment, paving the way towards more creative approaches to pricing (Jurievich, 2012).

2.1.3 Luxury Fashion

Luxury goods are not simple products, but status goods from which indirect social stratification can be derived (Husic and Cicic, 2009). They are consumed for the prestige that they carry more than for its practical utility (Doss and Robinson, 2013). Nowadays, luxury customers are not only more knowledgeable but also more demanding in regards to the products they acquire (Boston Consulting Group, 2014). Luxury fashion, a market segment that is highly dynamic, is significantly affected by this consumer behavior due to the high expectations and constant shifts in customers' demands regarding the products sold and the experience provided (Franzé, 2016).

By using the internet as a retail channel, high-fashion brands are able to capitalize on: access to new markets; means to enhance brand awareness and a platform that allows for the development of deeper relationships with consumers (Okonkwo, 2016).

In the literature, a wide variety of roadmaps for the integration of an online sales channel are presented. Despite their differences, a common aspect between them is the focus on *value development* and *operational integration* (Doherty and Ellis-Chadwick, 2010). Figure 2.3 showcases an approach that is representative of this framework.

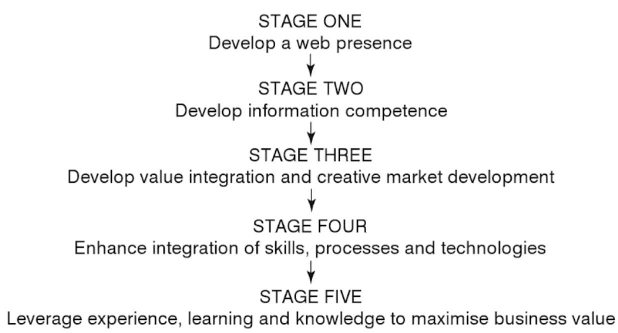


Figure 2.2: Main development stages for an e-tail platform (Hines and Bruce, 2007).

Regarding growth prospects, e-tail businesses on the luxury sector are quickly rising to unprecedented heights. According to a report from Deloitte, China's online sales growth hit the double-digit mark in 2016. Additionally, in the period ranging from 2011 to 2016, luxury e-commerce presented a year-on-year growth rate of approximately 12% - a noticeable difference when compared to the 2% year-on-year growth of traditional luxury retail sales (Deloitte, 2017). Additionally, a report from Mckinsey & Company states that for the luxury goods market, e-commerce's contribution may triple to the total brand revenue in the next decade.

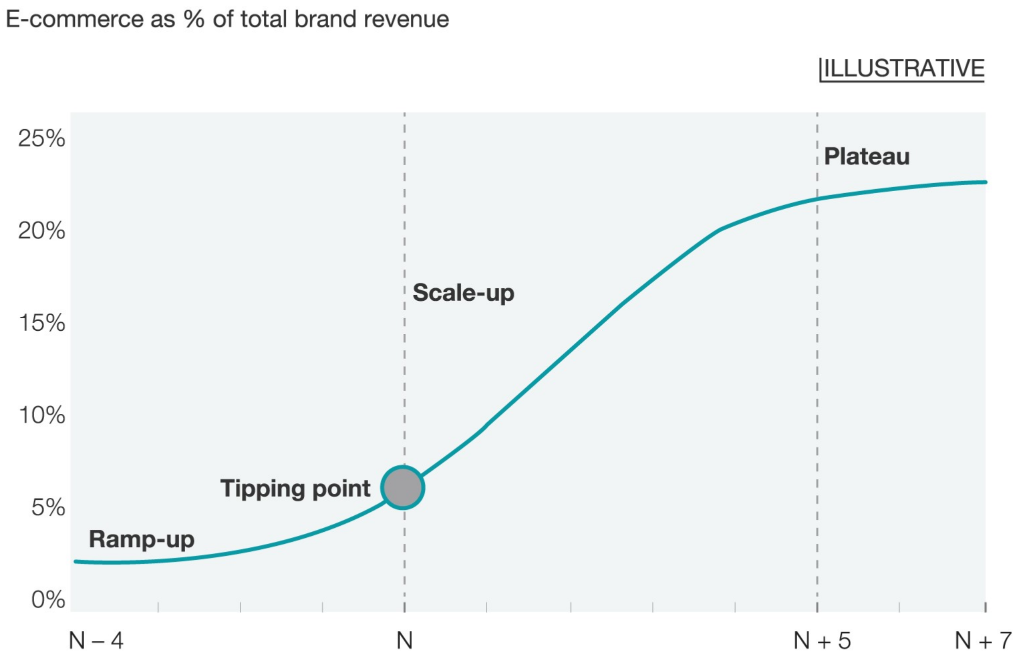


Figure 2.3: E-commerce's contribution to total revenue in the luxury goods market. (Catena et al., 2015).

According to Antonio Achille, Mckinsey's global head of Luxury, the future of online luxury fashion will favour companies that are the most scalable and technologically disruptive. The appearence and growth of more online marketplaces is also expected to occur, since this type of business model allows companies the promotion of a diverse assortment of luxury items without the need to carry inventory and with a worldwide presence (Marc Bain, 2018). At the moment, 30% of global luxury sales belong to Millenials and to Generation Z, who are currently driving 85% of this market's growth. These younger and affluent age groups have expectations that differ from the ones of past generations - they value more the purchase experience and service personalization than just the product per se (Samantha Woodworth, 2018).

The future for online luxury fashion is promising, with two main drivers leading the growth of this market: the value-adding potential of e-commerce and the growth of younger generations that are more tech-savvy and value a personalized and seamless experience.

2.2 Machine Learning

Machine learning is the field of computer science dedicated to the development of algorithms with the ability to "learn" and make accurate predictions based on past information available to the learner (Mohri et al., 2012). The importance of computational learning methods is clear on our current society, with many machine learning applications impacting our daily lives in the most various ways: preventative maintenance on motors and generators by analyzing past failure occurrences; sales and marketing forecasting for more efficient production planning; cybersecurity through pattern recognition in order to prevent potential intrusions; medical diagnosis by listing potential causes to a symptom; and others (Witten et al., 2016).

The techniques used in field of machine learning are vast. Figure 2.4 presents a summary of the methods used and their respective categorization.

Figure 2.4: Mindmap of the main machine learning techniques (Sridharan, 2015).

Another important concept of machine learning is the difference between *supervised* and *unsupervised* learning. Supervised learning is done with the knowledge of both the input and output variables. The goal of this technique is to construct a model that learns and best explains the relationship between a sample of data and its corresponding output values, so that when given a new observation, it would be able to produce an accurate prediction of the output variable. Unsupervised learning techniques only learn from a set of input data. Their goal is not to derive a prediction for an output variable, but instead to learn the relationship structure within a dataset, so that a new observation could be labeled according to its membership to the natural groupings of the dataset (Witten et al., 2016). Examples for these types of learning are Regression and Clustering, respectively.

2.2.1 Methods for Prediction

One of the most important applications of machine learning lies in the development of predictive analytical solutions. In general, Prediction is done when a set of input variables (*predictors*) is processed through a function or set of functions that return the corresponding output variable (*prediction*). The function responsible for modeling the relationship between the set of variables differs according to the technique used (Bishop, 2006).

The two main typologies of prediction problems in the field of supervised learning are Regression and Classification. Regression problems deal with the development of a real-valued output, given a certain set of input variables. Regression models weigh each variable according to its contribution to the output, thus modeling the relationship structure of the dataset. Classification problems deal with the labeling of a new observation in relation to its class membership (sub-population) from a set of all possible classes. Examples of Regression and Classification problems are demand forecasting and the assessing if a customer is a potential churner or not, respectively.

Figure 2.5 presents the general process of model construction for labeled data. In this process, the technique used (untrained model) processes information from past observations regarding the inputs (labeled input) and the corresponding output (labeled output) in order to learn their intrinsic relationship. Figure 2.5 also introduces the concept of *training* and *testing* data separation. Its purpose is mostly related to the assessment of model performance and it will be explained in the next subsection.

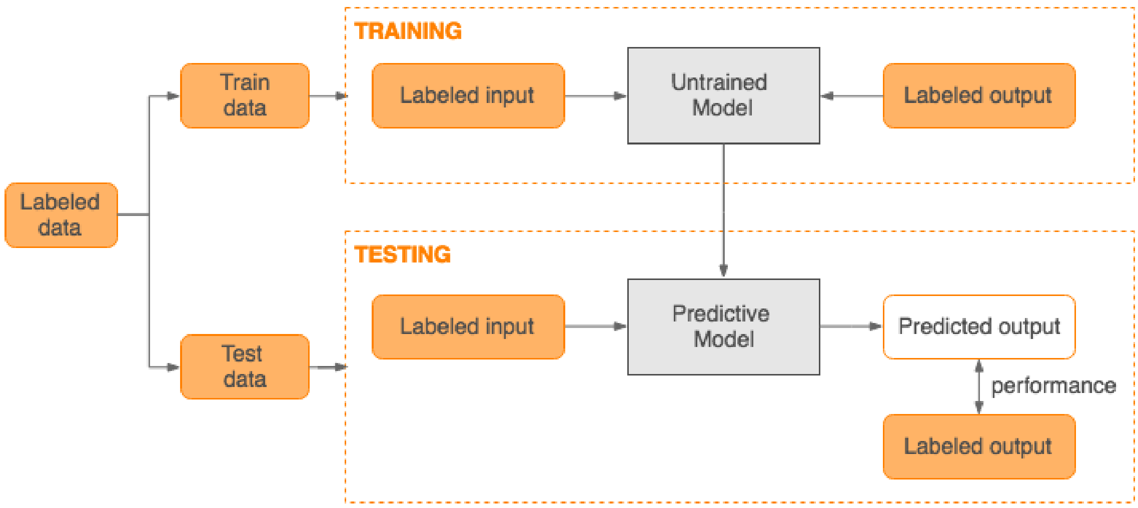


Figure 2.5: Construction of a predictive model for labeled data (EBC, 2017).

Regarding relative effectiveness, the "No Free Lunch" Theorem states that there are no *a priori* distinctions between learning algorithms, meaning that no technique will necessarily outperform another on all possible scenarios (Wolpert, 1996). Nonetheless, there are models that on average yield better results than others due to their ability to discriminate and learn more complex and deeper relationships in a data structure (LeCun et al., 2015).

(Caruana and Niculescu-Mizil, 2006) present a large-scale empirical comparison of the performance of several learning algorithms. In their study, techniques such as boosted decision trees, random forests and support vector machines (SVM) significantly outperform other methods that present less complex learning processes, such as the Naive Bayes classifier and regular decision trees. Also, a case study comparing the relative performance of different approaches to stock price forecasting concluded that a neural network model could perform better than other multivariate analytical techniques in terms of predictive power (Yoon and Swales, 1991).

In the next topics, the rational of the learning algorithms Random Forests, Logistic Regression and Support Vector Machines will be explained in order to contextualize the work developed in this dissertation.

Random Forests

Random Forests is classified as an ensemble learning technique and is commonly used for the development of predictive models. This method combines multiple randomly generated decision trees by outputting the mode of their predictions (in Classification) or their mean (in Regression) (Hastie et al., 2008). In comparison to individual decision trees, random forests overfit much less due to the Law of Large Numbers and present significantly less variance. Despite being a technique that generally yields good results in terms of predictive performance, Random Forests still tend to overfit data when compared to other algorithms and are also very computationally demanding. Another negative point is the fact that they work similarly to a black-box, with the output being given with no interpretability (Breiman, 2001). Figure 2.6 presents a visual representation of how Random Forests work in a classification problem.

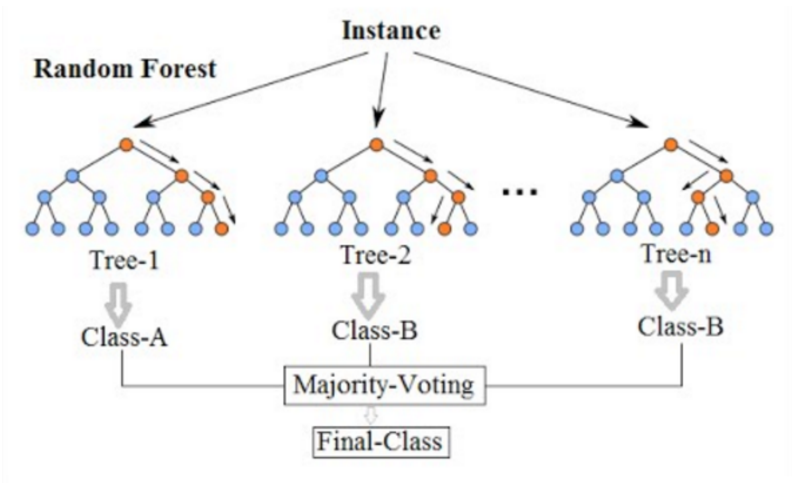


Figure 2.6: Random Forests model for a classification problem (Koehrsen, 2017).

The main hyperparameters of a Random Forests model that are often subject to tuning are (Breiman, 2001):

- **Number of splitting features per node:** this parameter refers to the number of randomly selected input variables that are compared during a node split decision. When the number of splitting variables increases, the likelihood of finding a better split also increases. However, increasing the number of variables also increases the correlation between trees, thus amplifying the variance in results.

- **Number of trees:** the number of trees used in the model does not contribute to overfitting and reduces the variance of the model. Although the use of more trees improves the model's accuracy, this happens at a diminishing rate and at the cost of significantly slower processing time.

- **Tree size:** larger trees are trees that encompass more splits. A larger tree is able to have more discriminatory ability and understand more complex data structures. However, models that use them are also more prone to overfitting the data.

Support Vector Machines

Support Vector Machines (SVM) is a supervised learning method commonly used in classification problems. It works by mapping the data as points in space, and setting a decision frontier (hyperplane) in which the records on each side belong to a different class. For efficient non-linear classification, the model can be adjusted by mapping the input observations to a high-dimensional space, through the use of a *kernel trick* (Cortes and Vapnik, 1995). Figure 2.7 presents the decision process for this type of learning algorithm.

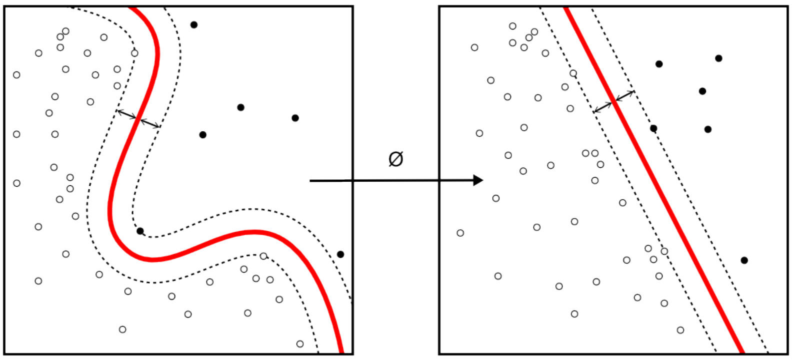


Figure 2.7: Application of a SVM model in a classification problem.

SVM models tend to yield good results in terms of predictive performance. They are computationally demanding and require the tuning of several hyperparameters in order to fully capitalize on their discriminatory ability (Blanchard et al., 2008). In this dissertation, the only SVM hyperparameter studied was the kernel type. Kernels are functions that enable the transformation of data. The kernel used can be linear, polynomial, radial, in others. The choice of which kernel to use depends on the type of relationship that the algorithm is intended to model. For example, a linear kernel will not learn the intrinsic relationships associated with non-linear problems, thus resulting in an inadequate model (Cortes and Vapnik, 1995).

Logistic Regression

Logistic Regression is a learning algorithm commonly used in classification problems. Similarly to a regular Regression, a logit model provides coefficients for each regressor variable. The main difference between the two models lies on how their output is generated, since a logistic function returns a real-valued output within the range [0;1]. In classification problems, the output of a logistic regression can be interpreted as the probability of a given observation belonging to the class that is to be predicted (Hosmer and Lemeshow, 2005).

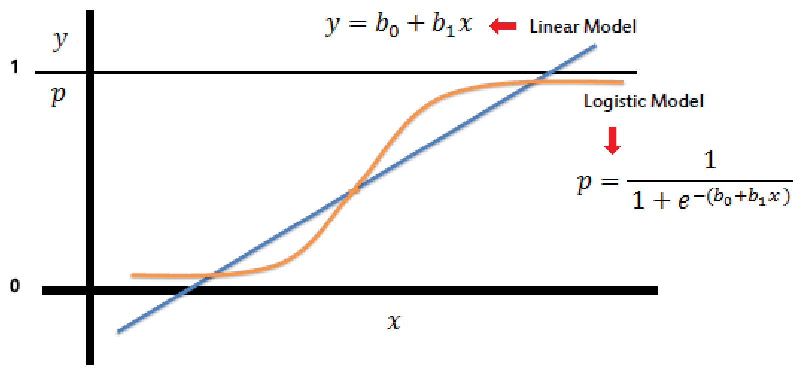


Figure 2.8: Logistic Regression Model (Sayad, 2018).

Two main advantages of this learning algorithm are that it is not as computationally demanding as the others mentioned above, and also the model holds interpretability due to the weight/contribution of each input variable that can be understood through its respective coefficient, thus enabling a clearer understanding of the relationship between the predictors and prediction variables (Park, 2013).

2.2.2 Performance Evaluation

In order to properly assess the performance of a predictive model, one should pay caution on how to interpret the data. If the model's performance is calculated with the data used to train it, then the results obtained will not constitute a reliable estimate of the model's generalization ability. This happens because the model will most likely reflect and mirror the specific patterns of the training sample and not of the whole population (Caruana and Niculescu-Mizil, 2004).

The most common way to train a learning algorithm and assess its predictive power is to split the dataset into a training and testing sample, so that the model learns from one portion of the data and is then evaluated with previously "unseen" observations, as shown in Figure 2.5. Some of the most commonly used types of data-splitting techniques are (Reitermanová, 2010):

- **Hold-out Method:** The dataset is divided in two samples: training and testing. A typical split proportion is 3:1, since training a model requires more data than to evaluate it. It is one of the simplest and less computationally demanding data-splitting techniques.

- **K-fold Cross Validation :** The dataset is split in *k* samples of the same size and the holdout method is repeated *k* times. This technique cycles through *k* iterations, in which models are trained using *k-1* samples and tested with the remaining sample. In the end, every data point is used for testing exactly once and *k-1* times for training. This technique significantly reduces performance variance. However, it requires the computation of *k* models, which can easily become a problem depending on the learning algorithm used, the dimension of the dataset and the number of folds considered.

- **Leave-one-out Cross Validation:** This method represents the most "extreme" case of the k-fold cross validation, since it is equivalent to having a number of folds equal to the number of observations. It is useful for applications containing a low number of data points and provides relatively unbiased results, though at the cost of high variance.

Lastly, if one plans to fine tune the hyperparameters of a model, then an additional portion of data should be kept "unseen". This sample is known as *validation* data. In order to avoid overfitting, the models are to be trained with the same training sample and have their performance measured with the validation subset, instead of the original testing sample (Ripley, 1996).

In the literature, the field of performance evaluation metrics is vast and presents many alternatives that complement one another for a richer interpretation of a model's predictive ability. For regression problems, metrics such as the Pearson's correlation coefficient, mean absolute deviation (MAD) and mean squared error (MSE) contribute in different ways to a more robust understanding of a model's ability to generalize.

Table 2.2: Examples of performance evaluation metrics for regression problems (Spüler et al., 2015).

Metric	Formula	Utility
Pearson's Correlation Coefficient	$\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\Sigma(y_i-\bar{y})^2}}$	A value of 1 means that the two series are perfectly and positively correlated. This is a good measure to track if predicted and real values share the same directional variations.
Mean Absolute Deviation	$\frac{\Sigma\lvert y_i-x_i\rvert}{N}$	Allows for the intuitive understanding of the average deviation that a prediction has to its corresponding real value (ground truth).
Mean Squared Error	$\frac{\Sigma(y_i-x_i)^2}{N}$	Useful for relative performance comparisons. The "squaring" of the deviations allows the metric to penalize bigger deviations between predictions and actuals.

For classification problems, the set of metrics used is different. Figure 2.3 introduces the concept of a confusion matrix, in the context of binary classification. From the confusion matrix, multiple metrics can be derived (Sokolova and Lapalme, 2009).

Accuracy is the metric that defines the ratio between the number of accurate predictions and the total number of observations ($\frac{TP+TN}{TP+TN+FP+FN}$). Although it is an important metric, its interpretation alone can lead to misleading results. For example, for an unbalanced dataset with 90% of observations belonging to the negative class, an accuracy of 90% might not be impressive, since a classifier can simply state that all observations belong to the negative class. For that reason, measures such as *sensitivity* ($\frac{TP}{TP+FN}$) and *specificity* ($\frac{TN}{TN+FP}$) prove to be useful and of complementary interpretation, since these represent the accuracy for each possible class. In the example mentioned above, the sensitivity and the specificity are represented by 0 and 1, respectively.

Table 2.3: Confusion matrix for a binary classification problem.

		Ground Truth		
		1	0	$\sum$
Prediction	1	TP	FP	$\bar{P}$
	0	FN	TN	$\bar{N}$
	$\sum$	P	N	

Another important metric for the evaluation of unbalanced datasets is Cohen's Kappa Coefficient. This indicator measures the level of agreement between two classifiers. Its value is given by $\frac{p_o-p_e}{1-p_e}$, where p_o represents the model's accuracy (observed agreement) and p_e is the expected accuracy of the model, that is, the likelihood of chance agreement between the two classifiers. Due to its definition, this metric is inherently robust to unbalanced data. Cohen's Kappa takes the value 1 when the two raters are in perfect agreement and can be lower than 0 if the agreement between the two raters is lower than chance probability (Hauser, 1993).

In the field of ordinal classification, the variety of performance evaluation metrics is not as big, since this type of problems are less common in practical applications. (Cardoso and Sousa, 2011) present a novel metric that is calculated directly from the confusion matrix, in which the performance of the model is directly related to the benefit path of the diagonal matrix. Additionally, (Gaudette and Japkowicz, 2009) suggest that measures such as the mean square and mean absolute errors are good performance metrics for this type of problems, even so when compared with plain accuracy.

2.2.3 Variable Significance

Wald Test

The Wald Test is a statistical test used to assess if the explanatory variables in a model are significant or not - that is, if their presence in a model contributes in any meaningful way to the resulting output (Wasserman, 2010). The Wald statistic and its respective hypothesis test are defined as follows:

$$W_i = \frac{(\hat{\beta}_i - \beta_0)^2}{Var(\beta_i)} \sim {\chi_1}^2 \tag{2.1}$$

Considering β_0=0, then:
H_0: $\hat{\beta}_i$ = β_0 (Coefficient i is not significant)
H_1: $\hat{\beta}_i \neq \beta_0$ (Coefficient i is significant)

$\hat{\beta}_i$ represents the coefficient associated with the explanatory variable i. If H_0 is true, than the attribute i is not significant in explaining the phenomenon being modeled. For a significance level of 5% (α), every p-value below that threshold will result in the rejection of H_0 - that is, the conclusion that the coefficient is significant.

Chapter 3

Farfetch Case Study

As explained in Chapter 2, one of the most distinguishing characteristics of e-commerce pricing is its flexibility and adaptiveness. This chapter aims to introduce how price is determined at Farfetch, the companies' main performance indicators for the evaluation of sales, consumer behavior and costs and explain the objective of this project in the context of the company's pricing strategy.

3.1 Price Structure

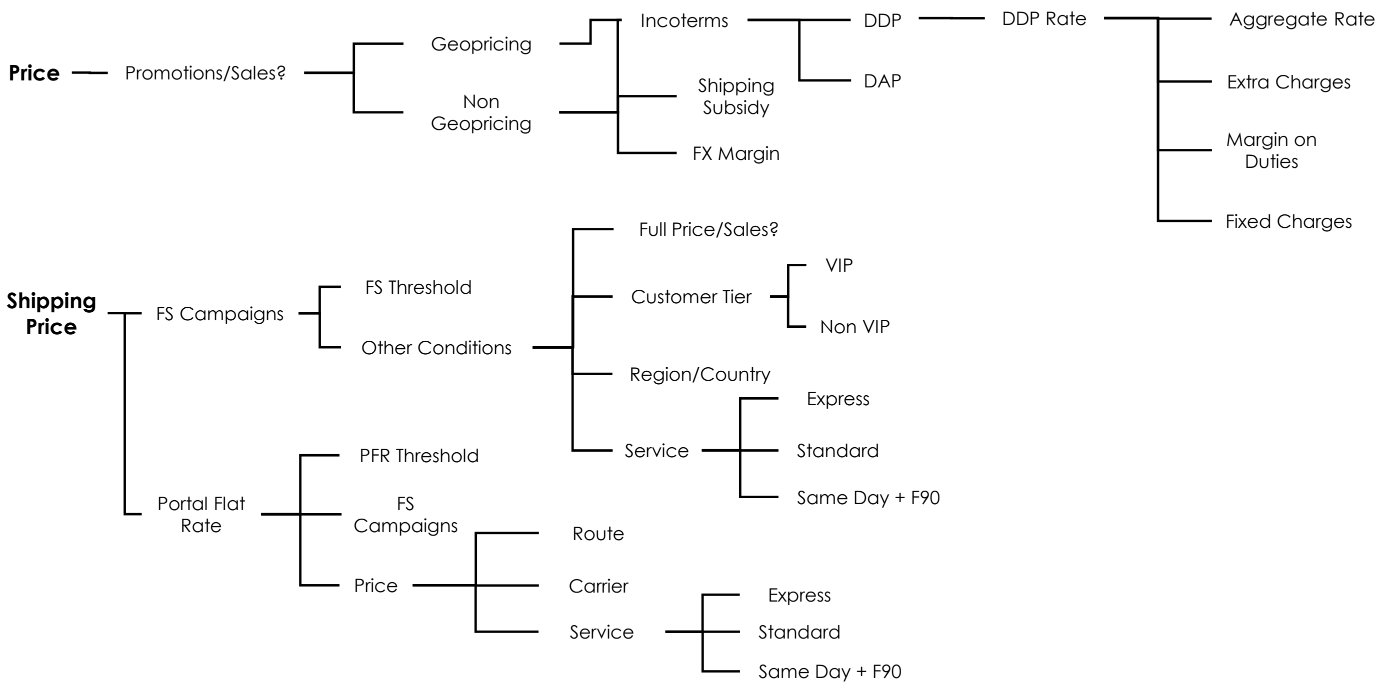


Figure 3.1: Farfetch's Pricing Structure Overview.

Figure 3.1 illustrates the main price components at Farfetch, alongside some of their interdependencies. The purpose of this overview is not to exhaustively list all components and their relationships, but to provide the necessary context for the following chapters.

3.1.1 Product Price

Sales and Promotions

One of the most influential price conditions that directly impact sales and other business-related KPIs is the occurrence or non-occurrence of promotional events. At Farfetch, there are two main global sales seasons, which lead to two distinctive sales peaks on the months of May and November. In addition to these events, more spontaneous and shorter promotions are done with the purpose of customer compensations and to stimulate purchases in periods of lower traffic.

The two main options for promotional price reduction are *sales discounts* and *promo codes*. The first option is done through a percentage discount applied over the RRP, whilst the latter offers a percentage discount over the final price. The distinction between retail price and final price will become clear after introducing the composition of all pricing components.

Geopricing

Geopricing occurs when the price of an item is defined by the brand, based on the competitiveness of the destination country. In e-marketplaces such restrictions are relatively common, since stores from different countries might offer different prices and such would lead to pricing inconsistencies that are not desired by major brands.

For non-geopriced items, the price is defined by the boutique/partner based on the typical retail price charged physically in store. The geopricing condition is set at the product and customer country level.

FX Margin

FX margin is a percentage value used to cover potential exchange rate fluctuations in the market. It is only applied on non-geopriced items and when the customer's currency differs from the boutique's. Each currency has its own FX margin. This price component is applied directly over the exchange rate.

Figure 3.2 depicts a representation of the exchange rate with and without this price component. The offset between the two series compensates for any possible deviations between the exchange rate used (stored in database) and the real exchange rate at the time of the transaction.

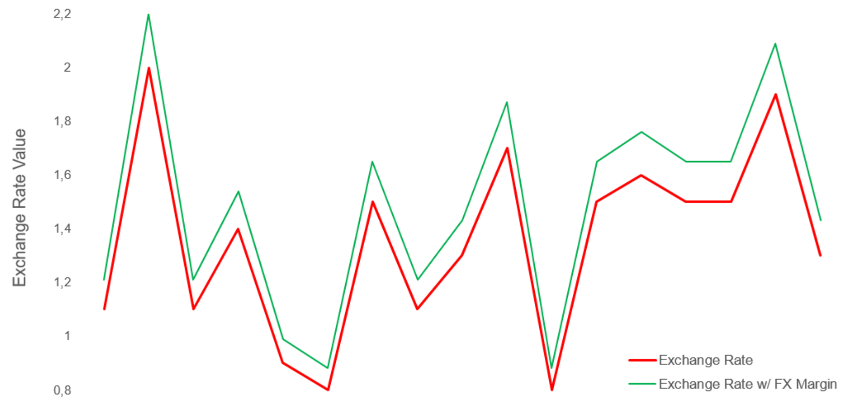


Figure 3.2: Exchange rate with and without a 10% FX margin.

Shipping Subsidy

Shipping subsidy is a percentage value used to cover the shipping costs resultant from permanent free shipping campaigns. This rate is defined by destination country, and is only applicable in non-domestic orders and in non-geopriced items. The shipping subsidy is applied over the converted price in customer's currency, with the FX margin already included.

Incoterms

Incoterms are a set of commercial terms of international commercial law (International Chamber of Commerce, 1980). Farfetch only operates over the Group D of the Incoterms 2010, more specifically the DDP and DAP incoterms:

- **DDP (Delivered Duty Paid):** The seller holds the responsibility, risk and costs related to the transportation of goods from himself to the buyer. Typically, the duties and taxes are charged within the price of the order.
- **DAP (Delivered At Place):** This incoterm requires the buyer to deal with all costs and risks associated with customs clearance in the importing country. This encompasses all customs duties and taxes.

If the destination country operates over the DDP incoterm then the *Final Rate*, that groups duties, taxes and other specific price fees is applied over the value paid by the customer for his order. If this is not the case, the customer will be required to proceed to payment in customs or pay an extra fee to have it delivered at home.

Final Rate

The Final Rate (DDP Rate) is the rate that is effectively used in the calculation of duties and taxes, for DDP countries. Equation 3.2 presents calculation method for this value.

$$Final\,Rate = [Aggregate\,Rate + Extra\,Charges + \frac{Fixed\,Charges}{Avg\,Price*} * (1 + Margin\,on\,Duties) - 1 \tag{3.1}$$

In which:

- **Aggregate Rate (%):** Represents the product between the Duty Rate and Sales Tax. Both of these values are dependent on the destination country and the category/classification of the product.

$$Aggregate\,Rate = [(1 + Duty\,Rate) * (1 + Sales\,Tax)] - 1 \tag{3.2}$$

- **Extra Charges (%):** Percentage that is used to cover small losses related to Farfetch's landed cost model.

- **Fixed Charges:** Besides the Duty Rate and Sales Tax, some countries charge additional fixed fees according to their import regulations. For example, Australia charges a merchandising fee of 83 $AUD per import. * - This fixed cost is then divided by the average price of the range that its order value belongs to, so that the percentage equivalent is obtained.

- **Margin on Duties (%):** Serves to cover the losses that Farfetch incurs when the client returns his order, since the duties charged are embedded in the order's value and are fully reimbursed at the time of the return. This rate is applied in countries operating within the DDP incoterm.

3.1.2 Product Price - Composition

The price components mentioned in the previous subsection come together to define the product's landed cost. This value represents the aggregation of price, customs duties, sales taxes and additional rates. The final price, that is, what the customer ultimately pays for an item, depends on whether the product is geopriced (fixed price) or not.

Non-Fixed Price Product:

$$Final\,Price = \frac{RRP}{1 + Store\,VAT} * (1 + Exchange\,Rate) * (1 + FX\,Margin) * (1 + Shipping\,Subsidy)] \\ *(1 + Final\,Rate) \tag{3.3}$$

If the order is Intra-EU then the Store VAT is not removed before applying the exchange rate. Additionally, the price is composed in the order presented, from left to right. Although this does not impact the final value paid by the customer, it has a significant effect on the individual value of each price component.

Fixed Price Product:

$$Final\,Price = Value\,imposed\,by\,Brand \tag{3.4}$$

$$D\&T\ Paid = Final\,Price - \frac{Final\,Price}{1 + Final\,Rate} \tag{3.5}$$

For geopriced items, the brands set the price paid by the customer. In this scenario, the customer does not pay the price components FX Margin and Shipping Subsidy. The only price component that remains is the Final Rate, from which the D&T are calculated, as shown in Equation 3.5.

3.1.3 Shipping Price

Portal Flat Rate

The calculation of the shipping price of an order is a complex process performed by a model called Portal Flat Rate. In this model, the origin countries (i.e: the boutiques' locations), destination country, order value and country-specific thresholds are used to determine the shipping price paid by the client. The prediction model developed in this dissertation does not contemplate this tool and so it will not be explained in more detail.

Free Shipping Campaigns

Free shipping campaigns can be either *promotional* or *permanent*. Promotional free shipping is temporary and can range from a single country, to a group of countries or can be global. It can be personalized to the customer's tier, to the platform used to access the website, among others. Similarly to sales periods, these type of campaigns is effective in achieving revenue uplifts at times of lower traffic and in keeping customers engaged throughout the year.

Permanent free shipping is implemented through two main criteria: Country and Customer Tier. Countries with permanent free shipping have a threshold that defines the minimum required order value for the order to have free shipping. Additionally, other constraints such as the order being bought at full price (i.e: without price discounts) might take place, depending on the country in question. Regarding customer tier, VIP clients are offered free shipping in all of their orders.

Lastly, the option to have free shipping can be overruled if the customer decides to upgrade his service (e.g: pick Same Day delivery over the base Standard option).

3.2 Key Performance Indicators

As almost any business nowadays, Farfetch has a wide variety of performance indicators that allow the company to track business growth and development throughout time. These indicators can be financial or customer-oriented. The three following categories of performance indicators were assessed: Sales, customer behavior and costs.

3.2.1 Sales

Farfetch analyzes its revenue mainly through three inter-connected metrics called GMV, ATV and NTV. GMV corresponds to the sum of the value paid by the customer, excluding its shipping price. Its a metric that encompasses the gross merchandise value transacted at Farfetch. Figure 3.3 showcases the general differences between these metrics.

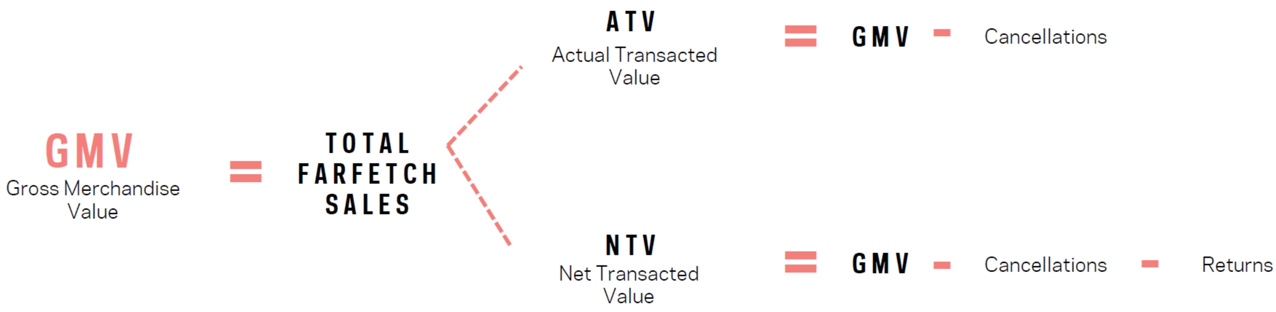


Figure 3.3: Farfetch's main sales-related KPIs.

There are situations in which a conjoint analysis of these metrics is more appropriate. For example, the implementation of a permanent free shipping campaign will expectedly boost sales. However, this initiative can also increase the return rate. In this situation, one could deduce that the NTV growth will be lower than the GMV's, thus better reflecting the campaign's actual impact.

Another important metric for the evaluation of sales and consumer behavior is the AOV (Average Order Value). This metric allows the company to understand how the customer's profile is evolving. For example, some countries might hold "higher spending" customers when compared against others whose customers opt for orders of lower individual value.

3.2.2 Customer Behavior

Farfetch has multiple customer-oriented metrics that allow for a comprehensive study on what drives the customer and how he reacts to new initiatives. For example, the NPS (Net Promoter Score) is a metric very commonly used to measure customer satisfaction.

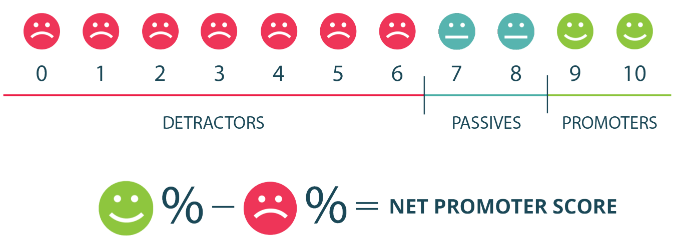


Figure 3.4: Net Promoter Score (Wootric, 2018).

Firstly, this metric attributes the title of: detractors to customers who rate the service provided with a score below or equal to 6; promoters to customers who rate service provided with a score equal to 9 or 10; and passives to the remaining. Promoters represent the customers that will most likely recommend the service to their peers and detractors are the customers who will most likely not return or not recommend the service. The NPS value is calculated as shown in the following Equation:

$$NPS = \%Promoters - \%Detractors] * 100 \tag{3.6}$$

A metric used to track consumers' patterns during their visits is the Conversion Rate. The conversion rate is a very valuable tool for finding improvement opportunities on the platform and understanding what deters a potential customer from buying the company's products. Figure 3.5 depicts the common funneling effect of customer conversion. High discrepancies between two levels of the pyramid might suggest the need to improve the way the company presents its information.

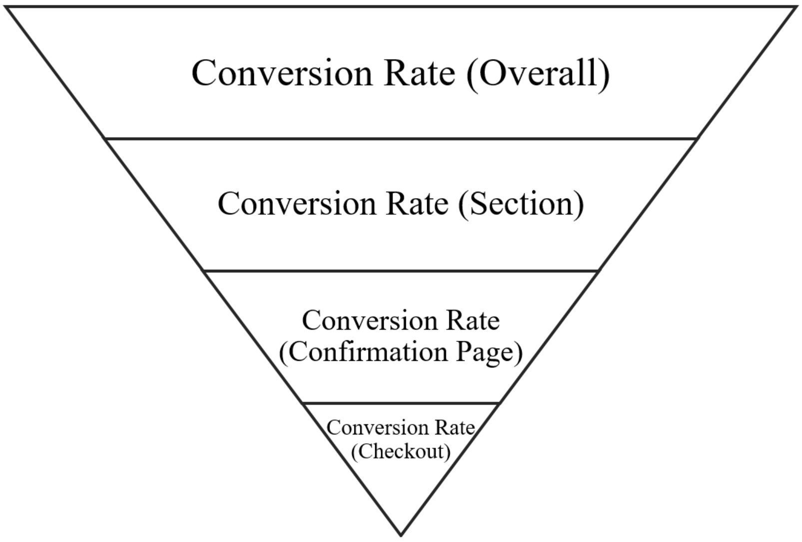


Figure 3.5: Conversion Rate Funneling effect.

The Return Rate is a metric that represents the proportion of returned goods and is very commonly used to study differences between different customer segments, regions and to assess the impact of new initiatives. It can be calculated as the ratio between the quantity or value of returned items and the quantity or value of total transacted items. This metric can be adapted to other similar ones, such as the Partial Return Rate, among others.

Lastly, a metric that helps companies follow-up on their customers' satisfaction and willingness to return is the Customer Retention. This indicator measures the percentage of customer that buy another product of the company after having already purchased once before.

3.2.3 Costs

In the context of the Delivery department, two very important metrics for cost tracking are the Shipping Cost Coverage Ratio and the P&L as a percentage of Total ATV. The shipping cost coverage ratio measures the quotient between the shipping income and the company's total shipping costs. Shipping income does not come exclusively from the value that the customer pays to have the item delivered at his/her home, because the shipping subsidy component of the product's price is also a source of shipping income.

This metric is very useful to assess how the shipping subsidy rate and the current route prices charged are able to cover the shipping costs that the company has, given any particular country and region. Knowing this value constitutes a most valuable input, since it will narrow down the main areas to focus on regarding potential price component adjustments and overall balancing of the shipping cost coverage.

The P&L as percentage of total ATV allows the company to know what is its deviation in cost coverage, taking into consideration the dimension of the business. This metric is calculated as the quotient between the net Shipping Income and total ATV. If its value is below a certain limit than the current shipping conditions should be reviewed.

Chapter 4

Data Collection and Exploration

The purpose of this chapter is to present and detail the process behind the development of the dataset used and its initial exploration. Additionally, an exploratory deep-dive study of China's shipping conditions will be presented for further comparison between the two approaches.

4.1 Constructing the Dataset

The main purpose of this dissertation was to develop a model that could predict the impact of alterations to Farfetch's pricing components on the company's main business performance indicators. As explained in Chapter 3, the company controls the value of multiple rates that are added to the product's price, depending on multiple factors.

The construction of the dataset had to take into account the specificities of each price component in order to be actionable and generalizable. The first decision regarding the structure of the dataset was to apply a time-basis index to the records. Each record of the dataset represents a complete day, because all pricing component transitions are made on this level of granularity. Simultaneously, the pricing components differ according to the geographical location of the customer. Thus, for each of the main priority countries of Farfetch a model was developed with time-indexed records.

Since not all pricing components existed before 2016 and due to the constant evolution and shifts in consumer behavior, the data collected was limited from present time up until the start of 2016.

4.1.1 Engineered Variables

During the development of the dataset, some variables were aggregated and transformed into new, artificial variables that expectedly present a higher discriminatory potential than their base attributes.

$$Adjusted\,FX\,Margin = FX\,Margin * (1 - \%Same\,Currency\,Orders) \tag{4.1}$$

Equation 4.1 returns an adapted version of the Forex Margin. Since this rate is only applied on orders where the origin and destination country share different currencies, then applying the suggested product will expectedly yield better results since it also takes into account the variation of the sales' {Origin-Destination} through time.

$$Adjusted\,Shipping\,Subsidy = Shipping\,Subsidy * \%International\,Orders \tag{4.2}$$

The Shipping Subsidy Rate is not applied on domestic orders. Thus, the proposed adaptation helps in assessing the actual impact of this price component.

$$Adjusted\,Extra\,Charges = Extra\,Charges * \%Orders\,of\,value >= EC\,Threshold] \tag{4.3}$$

Extra Charges are only applicable on basket values above country-specific thresholds. Thus, it makes sense to review this rate paired with the percentage of orders made above this threshold.

$$Adjusted\,Margin\,on\,Duties = Margin\,on\,Duties * \%Orders\,of\,value >= MD\,Threshold] \tag{4.4}$$

The reasoning behind Equation 4.4 is the same as for the Extra Charges transformation.

$$Adjusted\,Free\,Shipping\,Threshold = 1 * \%Orders\,of\,value >= FS\,Threshold] \tag{4.5}$$

If there is a free shipping campaign occurring (permanent or temporary), then a value of 1 will be taken by the Free Shipping binary attribute. Equation 4.5 proposes the adaption of multiplying this binary value by the correspondent threshold value, since a higher threshold would most likely lead to a decreased impact of the free shipping campaign on the company's results.

$$Sales\,Peak = \begin{cases} 1, & \text{if } d_t - d_{t-1} > Threshold_i \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

The Sales Peak variable corresponds to the attribute used for the automatic identification of sales peaks in the sales time series. d_t and d_{t-1} correspond to the daily sales of two consecutive days. $Threshold_i$ corresponds to a value that defines the transition between typical sales fluctuations and other punctual occurrences such as sales or free shipping campaigns, for a given country i. This value depends on the size of the specific market and can be corrected by manually validating its results against the company's promotional calendars.

4.1.2 Dataset Structure

Table 4.1 summarizes the variables included in the final dataset. From this structure, the goal was to construct a model for each business indicator listed, considering the pricing variables in study -

normal and adjusted.

Table 4.1: Final structure of the dataset.

Variable	Type
Date	Numeric
GMV	Numeric
NTV	Numeric
Return Rate	Numeric
NPS	Numeric
AOV	Numeric
P&L	Numeric
Checkout Conversion Rate	Numeric
FX Margin	Categorical
Shipping Subsidy	Categorical
FX Adjusted	Numeric
Shipping Subsidy Adjusted	Numeric
Extra Charges Adjusted	Numeric
Margin on Duties Adjusted	Numeric
Free Shipping	Categorical
Free Shipping Threshold Adjusted	Numeric
Sales Peak	Categorical

4.2 Exploratory Analysis

As explained in the previous section, each dataset corresponds to a single priority country due to the nature of the pricing components. Since the structure of the analysis is mostly the same regardless the country, from this section forward the analysis will only reflect the results obtained for China - one of Farfetch's main markets.

4.2.1 Summary of the Variables

Table 4.2: Summary of the dataset's numeric variables.

Name	Count	Min	1st Qu.	Median	Mean	3rd Qu.	Max
GMV*	791	0	0.02	0.03	0.05	0.06	1
NTV*	791	0	0.02	0.03	0.05	0.06	1
Return Rate	760	0	0.11	0.14	0.139	0.166	0.337
NPS	772	64.1	75.4	77.8	77.6	89.7	90.1
AOV	791	259	403.7	471.9	477.3	541.3	972.2
P&L*	791	0	0.87	0.89	0.88	0.90	1
Checkout Conversion Rate	791	0.196	0.358	0.431	0.423	0.486	0.661
FX Adjusted	791	4.2	4.7	7.0	6.4	7.5	8.1
Shipping Subsidy Adjusted	791	0	0	0.166	0.239	0.352	2.00

Table 4.2 above summarizes a few descriptive statistics for the numerical variables of the dataset. The Extra Charges and Margin on Duties variables are not present since they are not used

for the China market. The variables marked with an asterisk were transformed through a Min-Max normalization to a [0,1] scale in order to preserve data confidentiality.

The Return Rate and NPS variables represent a lower of number of records since the most recent observations do not fully reflect the true values of these variables - the clients returns' might have not been processed or requested or the client might still provide his feedback through the NPS questionnaire. Both the GMV and NTV present very high maximum values due to the abnormally high order traffic during sales seasons. To complement this summary, a few plots were developed in order to understand the evolution of some variables and identify patterns in the data.

Figure 4.1 presents the gross sales evolution in China. The most relevant insights are that the market is growing at a fast rate with very prominent sales peaks whose occurrence matches the sales seasons calendars.

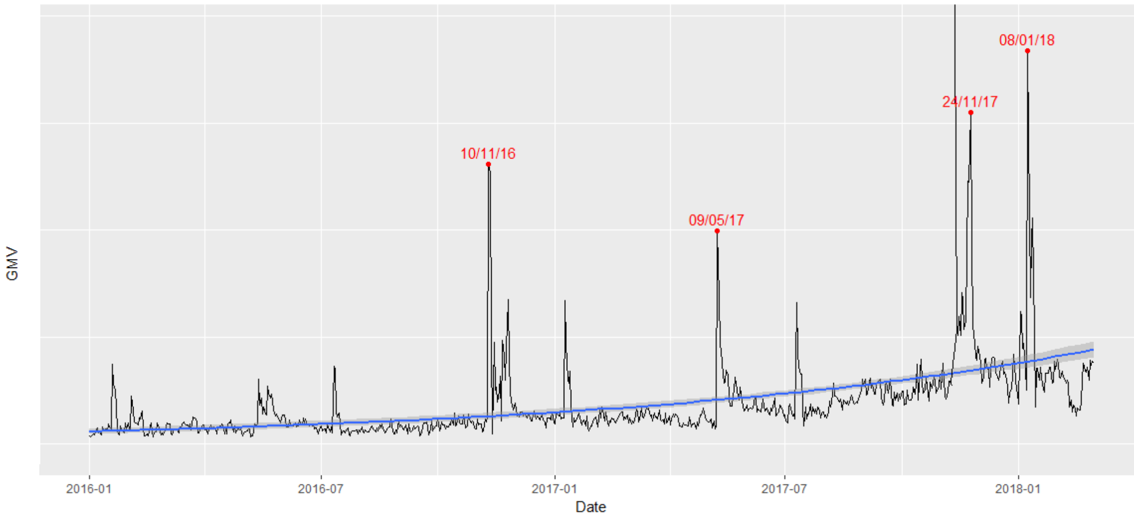


Figure 4.1: GMV growth in China, in 2016-2018.

To explicit the relationship between GMV and NTV, Figure 4.2 presents the difference between these two indicators. The filled area between the two series corresponds to the value lost due to cancellations and returns.

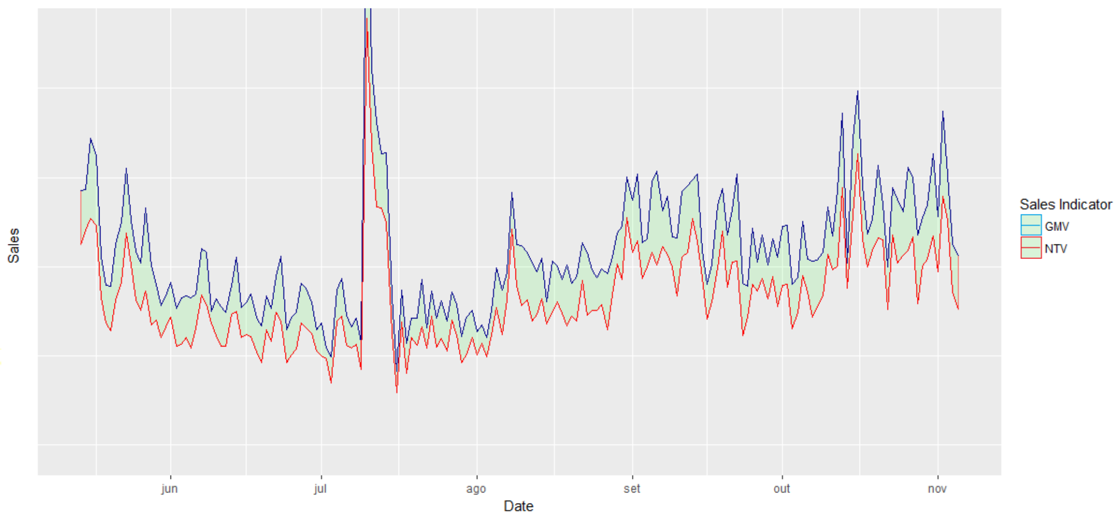


Figure 4.2: Comparison between GMV and NTV for China's market.

Additionally, Figure 4.3 presents a negative NPS trend that opens room for exploration on what is driving the deterioration of customer satisfaction while increasing sales.

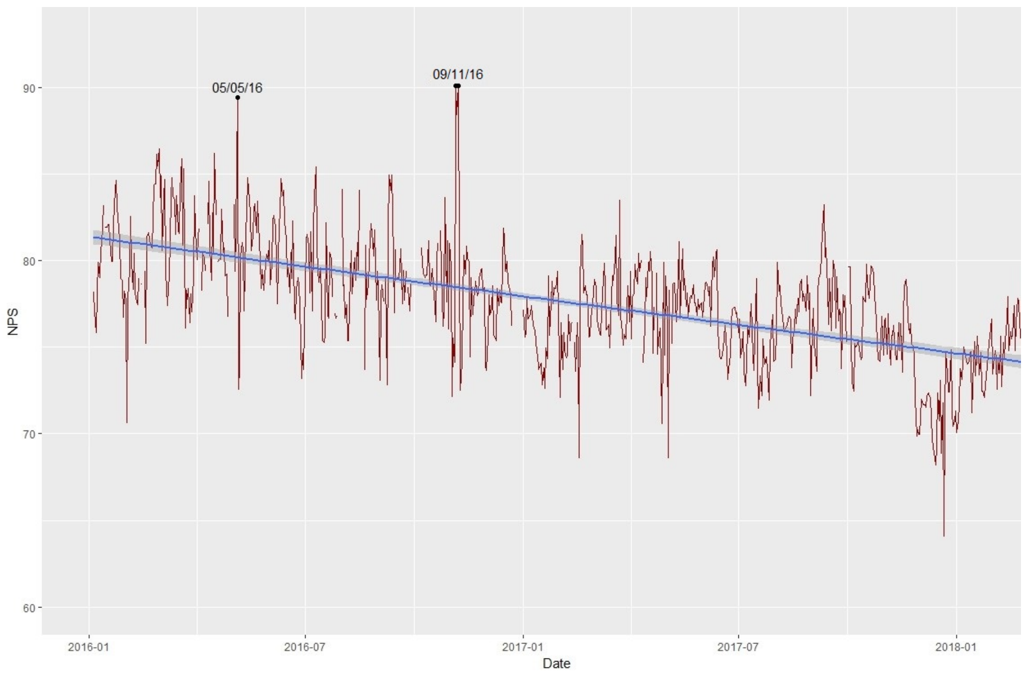


Figure 4.3: NPS Trend in China's market.

The following table summarizes some descriptive statistics regarding the categorical variables of the dataset.

Table 4.3: Summary of the dataset's categorical variables.

Name	**Count**	**Levels (*Frequency*)**
FX Margin	791	{4.7 (*179*), 5 (*79*), 7.5 (*472*), 8.3 (*61*)}
Free Shipping	791	{0 (*416*), 1 (*375*)}
Shipping Subsidy	791	{0 (*208*), 1.92 (*130*), 2.76 (*221*), 4 (*86*), 5.31 (*146*)}
Sales Peak	791	{0 (*701*), 1 (*90*)}

Frequency corresponds to the amount of days that each variable level lasted. According to Table 4.3, both the Shipping Subsidy and FX Margin went through significantly large changes in the past two years. Also, the Free Shipping binary variable presented a high percentage of days with free shipping because Farfetch launched its permanent free shipping campaign in China in the mid-February 2017.

4.2.2 Personalized Study

Although the pricing components under evaluation sustained large variations in value throughout the period of analysis, the total amount of variations is still relatively small in comparison to the amount of days analyzed. In order to grasp the feasibility of having a prediction model learning from such limited past pricing component variations, the results that were obtained through this approach were systematically compared with a personalized analysis that was done in parallel for the assessment of the impact of Free Shipping in China, regarding: NTV, return rate growth and shipping cost coverage.

The main specific questions that this study addressed were:

1. *Did permanent Free Shipping drive forth growth in China?*
2. *Did the distribution of sales by basket value change?*
3. *What was the impact on the country's return rate?*
4. *Were the changes to the shipping subsidy rate enough to cover the additional shipping costs?*

Business Growth

To determine if permanent free shipping brought growth to the chinese market two perspectives were analyzed: business growth for basket values below and above the free shipping threshold after the implementation of permanent free shipping, and comparison between growth trends before and after the implementation of permanent free shipping.

By taking into consideration the permanent free shipping threshold, the first perspective provides specific insights of how business growth occurred in China. If there is a difference between the growth rate below and above the threshold, then the hypothesis that free shipping helped to increase sales in China is corroborated. On the other hand, the second perspective provides a more general overview of if the growth rate changed before and after the implementation of free shipping. Although it is time-bound statistic, it does not consider other factors such as basket value. If a positive difference in growth rate exists then the results are indicative of the success of permanent free shipping in driving growth in China.

Distribution of Sales

In order to grasp how the distribution of sales in China was affected by the permanent free shipping campaign, a matrix based on the Equation 4.7 was constructed.

$$ATV\,Contribution\,(\%)_{ij} = \frac{ATV_{i,j,t}}{Total\,ATV_t} - \frac{ATV_{i,j,t-1}}{Total\,ATV_{t-1}} \tag{4.7}$$

Where i represents intervals of basket value and j represents the total number of products in the basket. t and $t-1$ represent the periods of one year after and before the start of permanent free shipping, respectively. These values open room for insights such as: how did the AOV change; have the customers started buying more products; how did the overall percentage contribution to total ATV change after the implementation of permanent free shipping.

Return Rate

The study of the impact of permanent free shipping on China's return rate consisted of two steps. First, the country's return rate 6 months before and after permanent free shipping was compared. If there was a significant difference between these two rates, then a plot with the relative growth values for the partial return rate* by basket value would likely clarify the reason for this increase. Equation 4.8 explicits the calculation process for the plot aforementioned.

$$Partial\,Return\,Rate\,Growth\,(\%)_i = \frac{\frac{Returned\,Value}{Value}_t - \frac{Returned\,Value}{Value}_{t-1}}{\frac{Returned\,Value}{Value}_{t-1}} \tag{4.8}$$

* - The partial return rate metric corresponds to the ratio between the returned value on orders where the number of returned items is lower than the basket's total and the overall value of orders.

Shipping Cost Coverage

China's shipping subsidy rate changed multiple times through the dataset's time horizon. Most of these changes were done in order to accommodate the additional shipping costs resultant from the country's permanent free shipping campaign. During the first sales season of 2017, the shipping subsidy rate was set to zero to promote price competitiveness during a high sales peak period. To determine the ideal shipping subsidy rate that would offset all of the additional shipping costs, the shipping subsidy income's coverage of the additional shipping costs in 2017 was determined and the correspondent ideal was calculated by equaling the additional shipping costs to the shipping subsidy rate variable multiplied by the total net sales of 2017.

Chapter 5

Model Development and Evaluation

At an initial stage, the models produced for China were aiming the output variables GMV, NPS and Return Rate - some of the priority indicators at the time of this study. However, due to time constraints and the total amount of models that would have to be constructed, only some prediction algorithms were tested. As discussed in Section 2, although there are no techniques that yield the best results for all possible scenarios, some algorithms generally perform better than others, on average. For that reason, the Ordinal Logistic Regression, Random Forests, Support Vector Machines and Naive Bayes were used and compared against each other in terms of their predictive performance. The software R Studio was used for the development of these machine learning models.

5.1 Data Processing

In order to study the models through a wider range of evaluation techniques and to reduce some of the noise that the time series presented, the prediction variables were converted to an ordinal categorical scale through value binning. Since the logic is the same for the rest of the output variables, the following subsection solely describes the variable transformation process for the GMV variable. The main steps of the data processing stage are the following:

1. Determine the business growth trend (only for the GMV variable)
2. Attribute a class bin to each output variable's data points
3. Standardize/Scale the output variables
4. Assess the impact of the input variables (eg: difference in mean of the output variables)

5.1.1 Transformation of the GMV variable

Firstly, the GMV variable was fitted through a polynomial least absolute deviation regression that minimizes the mean absolute error cost function presented in Table 2.2. This method was used because it is more robust against sporadic and very high sales peaks that overestimate the real business growth trend. Figure 5.1 presents the trend line obtained through this approach.

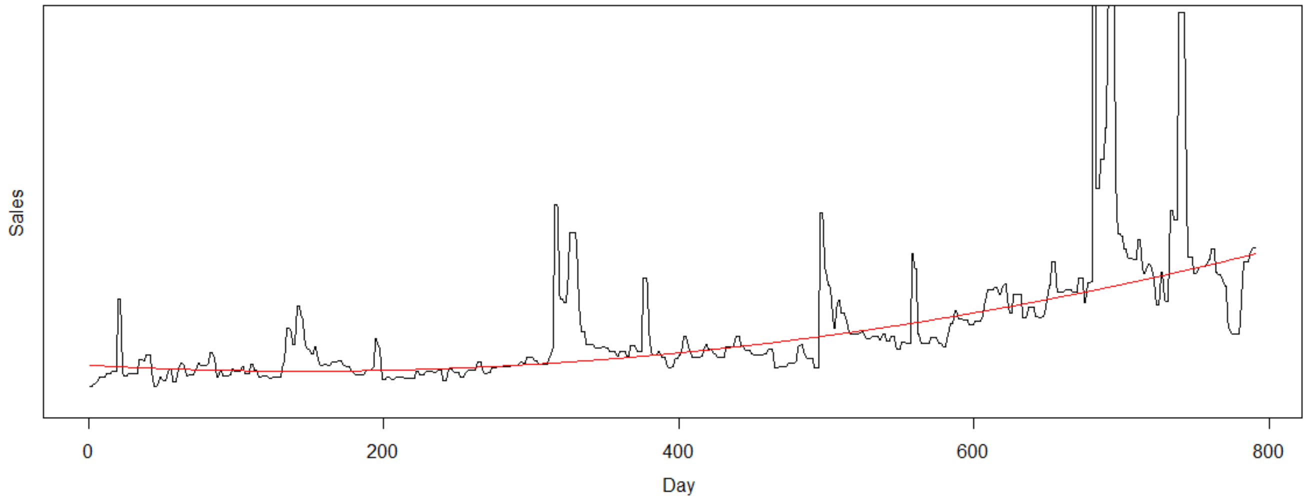


Figure 5.1: Fitting GMV through a Least Absolute Deviation regression.

After determining the trend, the relative distance of each observation was calculated through the following equation (d_t and $\bar{d}_t$ represent the real and fitted GMV values):

$$Relative\,Distance\,(\%) = \frac{d_t - \bar{d}_t}{\bar{d}_t} * 100 \tag{5.1}$$

After this calculation, each observation was associated with its respective ordered bin depending on its value. For example, a relative distance between 5-10% in GMV corresponds to class 1 in the new categorical variable. Appendix A presents the equivalence table for the transformation of the numerical GMV variable. Figure 5.2 depicts the transformation process explained in this section.

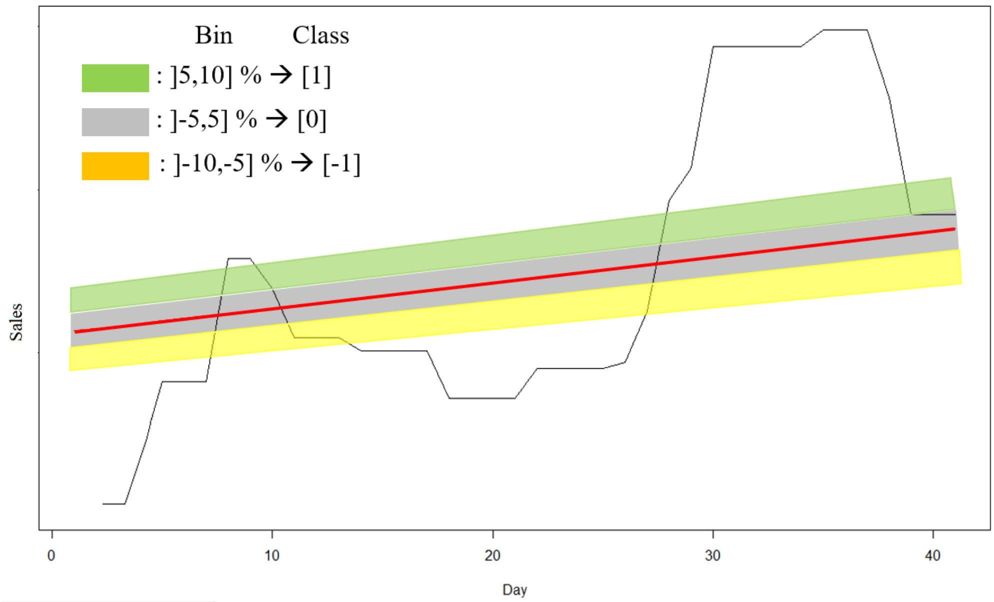


Figure 5.2: Illustration of the output variable transformation.

Since the values are now scaled relatively to the sales' time-dependent growth trend, the data is expectedly non-stationary. The following bar plot presents how the transformed output variable is now centered on the growth trend and most likely only reflects the effects provoked by other non-stationary business factors.

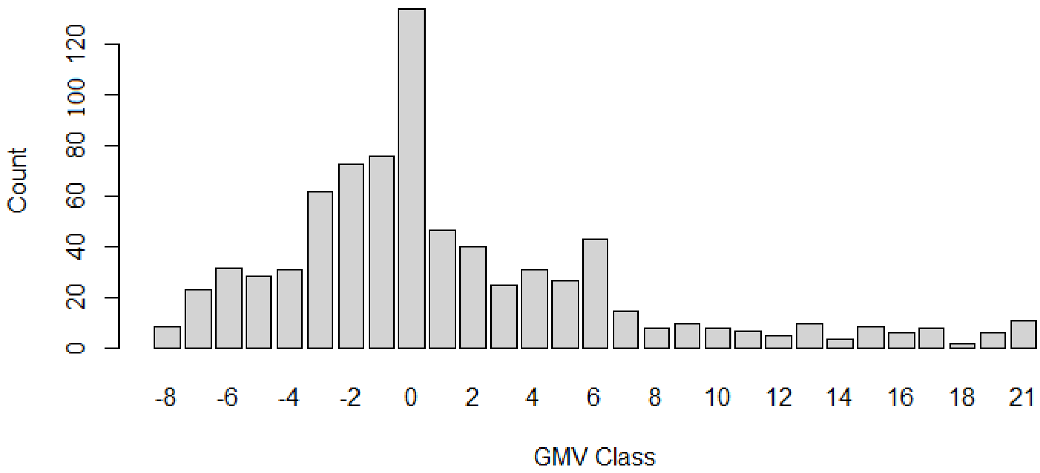


Figure 5.3: Bar plot of the transformed GMV variable.

Before developing the predictive models, an assessment of how the GMV variable varies depending on the values taken by the pricing input variables was performed, using Tukey's "Honest Significant Distance" (HSD) method. In this analysis, the confidence intervals for the difference in the means of the output variable were performed after fitting the data to an analysis of the variance model. Table 5.1 summarizes the results obtained.

Table 5.1: Summary of Tukey's HSD method results - Mean differences and p-values.

Pricing Variable	**Level Diff. (*)**	**Mean Diff.**	***P-value***
Sales Peak	1-0	11.50	*≈ 0*
Free Shipping	1-0	9.51	*≈ 0*
FX Margin	7.5-4.7	-1.22	*0.196*
	7.5-5	-1.1	*0.299*
	8.3-4.7	-3.55	*0.002*
	8.3-5	-3.33	*0.0006*
Shipping Subsidy	5.31-0	-2.07	*0.007*
	5.31-2.76	-3.60	*≈ 0*
	5.31-4	-3.55	*0.002*
	4-0	-1.52	*0.006*
	4-2.76	0.72	*0.04*

(*) - Some level differences were not included due to their redundancy in regards to the conclusions drawn by the rest.

The results obtained corroborate the effectiveness of Free Shipping in increasing sales and how increasing the value of the pricing components FX Margin and Shipping Subsidy seemingly has a negative impact on price competitiveness and, thus, sales. However, in some of the intervals the conclusions do not appear to be the same (Increasing a price component resulting in increased sales). This can happen since these are one factor studies and some of the levels presented by the input variables did not last long.

5.1.2 Standardization

Some prediction models require the variables to be scaled in order to provide good results. This happens when the learning algorithms follow a distance or similarity based criteria to produce their predictions, which makes them sensitive to differences in scale. For example, Height (meters) and Weight (kilograms) are two numeric variables that do not share a common scale.

Out of the four prediction algorithms mentioned in the beginning of the present chapter, only Support Vector Machines requires feature scaling prior to the training of the model. Since the code package used to implement this model also performed feature scaling, there was no need to scale each variable prior to the construction of the models.

5.2 Model Construction

The technique Leave-one-out cross validation presented in Section 2.2.2 was used to split the data into training and testing sets. For validation purposes, 10% of the records were kept "unseen" from this data division process. The main reason for choosing this approach to data splitting was the relatively low number of data points, since all information was grouped to a day level.

5.3 Evaluation Techniques

As presented in Section 2.2.2, there is a wide range of metrics that allow to measure the performance of a predictive model through different perspectives. In the context of this study, five different and complementary metrics were used to assess and compare the performance of the models developed. The metrics are: Mean Squared Error, Accuracy, Kappa, Overall Adjusted Sensitivity and Window Accuracy. Additionally, the plot of the differences between the ground truth and predictions was also reviewed.

The last two metrics were specifically constructed for this case study and are generalizable for most typical ordinal classification problems. Since the first three metrics were already described in the Theoretical Background, in this section the rationale behind the other metrics will be presented. Equation 5.2 describes the formula that defines Overall Adjusted Sensitivity.

$$Overall\,Adjusted\,Sensitivity = \sum_{i=1}^{N}(1 - Relative\,Frequency_i) * Sensitivity_i \quad (5.2)$$

- **N:** Number of classes of the output variable;
- **Relative Frequency:** Ratio between the number of observations in which the output variable's class i is present and the total number of observations;
- **Sensitivity:** The true positive rate of class i.

This metric is useful to identify cases of prediction bias, in which the algorithm gives preference to the majority class for the sake of achieving a higher accuracy (instead of actually modeling the intrinsic relationships of data). By weighing each class i sensitivity with $1 - RelativeFrequency$, the metric values more the correct identification of the less common classes.

Although the weights mentioned are not tuned and are relatively arbitrary, the Overall Adjusted Sensitivity still manages to provide a solid perspective on the occurrence of prediction bias. Table 5.2 exemplifies the use of this metric.

Table 5.2: Prediction example for the overall adjusted sensitivity metric.

(a) Example dataset of actual and biased prediction values.

Actuals	Predictions
1	4
4	4
4	4
4	4
4	4
4	4
4	4
4	4
4	4
4	4
4	4
12	4
15	4
18	4

(b) Auxiliary table for the calculation of the overall adjusted sensitivity.

Class	Rel. Freq.	Sensitivity
1	0.071	0
2	0	0
3	0	0
4	0.72	1
5	0	0
6	0	0
...	...	...
12	0.071	0
13	0	0
14	0	0
15	0.071	0
16	0	0
17	0	0
18	0.071	0

From Table 5.2a one can calculate the predictions' accuracy by dividing the number of "matches" over the total number of records. In this example, the accuracy is 71.4%. From Table 5.2b, one can determine the overall adjusted sensitivity, which in this case is 28.6%. Since the model always classifies records as the majority class, it is fair to consider this predictive model biased and that is reflected in the difference between the two metrics.

Another custom metric developed for this dissertation is the Window Accuracy. Equation 5.3 explicits its mathematical definition.

$$Window\,Accuracy = \sum_{k=0}^{M} \frac{Accuracy\,at\,k\,neighbours}{2k+1} \tag{5.3}$$

Accuracy at k neighbours corresponds to an adjusted metric of accuracy, where a positive or negative deviation equal or less than k consecutive classes still counts as a "match" towards the metric's accuracy measure. The standard accuracy measure corresponds to an accuracy at 0 neighbours. $2k+1$ represents the number of valid combinations between actuals and predictions for any predicted class, when considering k neighbours. M represents an user defined number that caps the maximum number of neighbours considered. Figure 5.4 explains the accuracy-matching process described above.

K = 0		
Actuals	Predictions	Match?
4	6	No
7	10	No
4	3	No
10	10	Yes
4	3	No
4	4	Yes
7	3	No
5	2	No
9	8	No

(a)

K = 1		
Actuals	Predictions	Match?
4	6	No
7	10	No
4	3	Yes
10	10	Yes
4	3	No
4	4	Yes
7	3	No
5	2	No
9	8	Yes

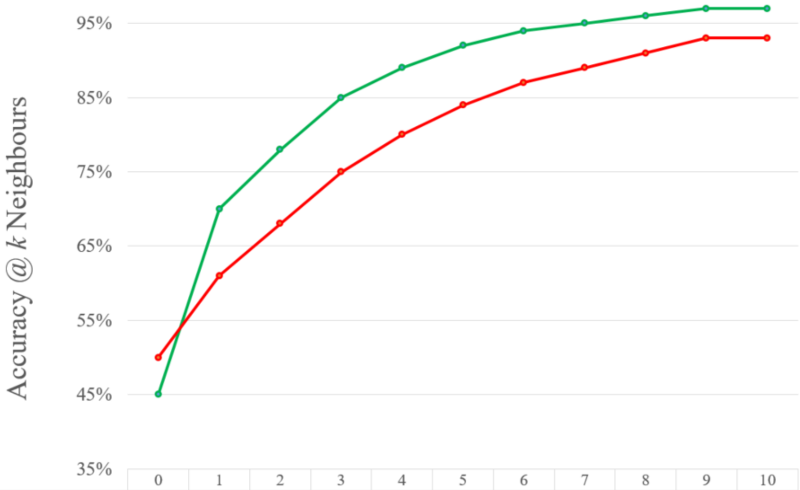
(b)

Figure 5.4: Comparison of the matching results for accuracy at 0 and 1 neighbours - (a) and (b), respectively.

The concept of accuracy at k neighbours is useful for ordinal classification problems, because when there is a very high number of classes, an error margin of a single class might be relatively negligible when compared to binary classification problems. The purpose of the window accuracy metric is to consolidate the values of accuracy at different numbers of neighbours into a single comparable metric. Figure 5.5 illustrates an example of the application of the window accuracy metric.

Regarding that example, despite Model 2 achieving a higher standard accuracy (50%), Model 1 consistently presents higher accuracies at k neighbours, for $k > 0$. The window accuracy metric is able to compile all those individual metrics by summing all accuracy values and weighing them down by their respective number of potential match combinations. For a maximum number of 3 neighbours ($M = 3$), the window accuracies of Model 1 and Model 2 are 0.96 and 0.95, respectively.

The main advantage of this metric is that it provides a single value that takes into account not only the standard accuracy of the predictive model but also the size of the prediction deviations.



	0	1	2	3	4	5	6	7	8	9	10
Accuracy [Model 1]	45%	70%	78%	85%	89%	92%	94%	95%	96%	97%	97%
Accuracy [Model 2]	50%	61%	68%	75%	80%	84%	87%	89%	91%	93%	93%

Figure 5.5: Comparison of two models - Accuracy at *k* neighbours' perspective.

Chapter 6

Results

In this chapter, the results obtained for China's GMV, Return Rate and NPS prediction models are compared in regards to each learning algorithm used and each performance evaluation metric considered. The relative importance of each variable towards the models' predictive performance is analyzed. Additionally, the results obtained for the personalized study are also presented and discussed. Lastly, a brief discussion regarding the viability of both approaches is done.

6.1 Machine Learning Approach

6.1.1 Variable Importance

In terms of input variable importance, the Random Forests' importance metric based on a mean decrease of accuracy principle and the Logistic Regression's statistical test for coefficient significance provide two different perspectives that help in understanding how each variable weighs on each model's prediction accuracy. The statistical test used for the evaluation of coefficient significance is the Wald Test presented in Chapter 2. Table 6.1 summarizes the results regarding variable importance.

Table 6.1: Results for input variable importance - different perspectives.

Variable	**Mean Decrease in Accuracy**	***p-value (Wald Test)***
Free Shipping	0.024	0.025
Sales Peak	0.022	<0.01
Adjusted Shipping Subsidy	0.017	0.047
Shipping Subsidy	0.016	0.094
FX Margin	0.015	0.04
Adjusted FX Margin	0.05	<0.01

The two indicators present concordant results for the majority of the variables. The results suggest that Free Shipping campaigns and Sales events are the variables that mostly drive GMV growth. The Shipping Subsidy and the FX Margin price components come next and present similar results between each other.

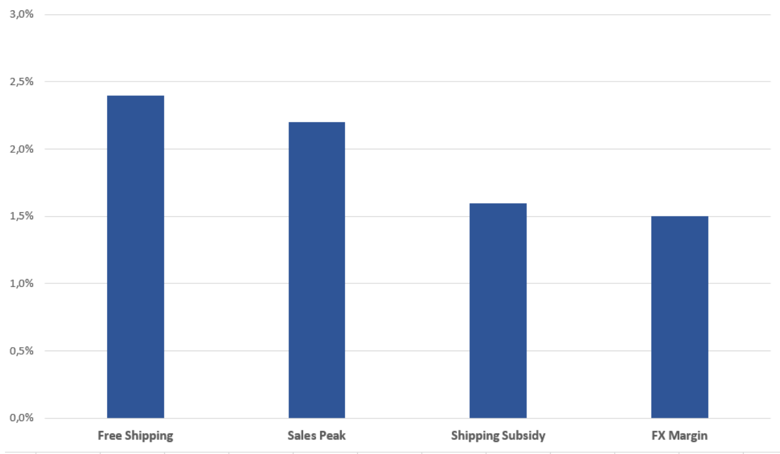


Figure 6.1: Random Forests' variable importance metric.

Figure 6.1 illustrates that although the Free Shipping and Sales Peak variables are the most important, the remaining pricing components also hold relative high importance. Since the changes to these price components led to price variations of up to 5%, it is expected that such would affect the price competitiveness in China's market and, thus, sales.

6.1.2 Model Performance

GMV Variable

Table 6.2: Results for the GMV models, by performance metric and learning algorithm.

Metrics	O.L.R	N.B	R.F
M.S.E	17.6	55.6	27.2
W.A @ 5 Neighbours	0.62	0.60	0.71
Accuracy	0.13	0.15	0.21
Kappa	0.06	0.09	0.20
O.A.S	2.64	4.76	2.21

Comparing each model relatively to one another, multiple insights can be drawn. In terms of model accuracy, the Random Forests model presents the best results, leading with a standard accuracy of 21% and with a window accuracy at five neighbours of 71%. Also, despite holding the lowest predictive accuracy out of all learning algorithms, the Ordinal Logistic Regression model presents the lowest Mean Squared Error, which means that, on average, this model's incorrect predictions are "less off" the actual value than the remaining models. Lastly, the Naive Bayes model presents the highest Overall Adjusted Sensitivity, suggesting that this model performs better in capturing the occurrence of minority classes. However, this comes at the expense of a relatively high Mean Squared Error, indicating that the Naive Bayes model tends to risk more in its predictions.

Figure 6.2 illustrates the accuracy at k neighbours for the Random Forests and Ordinal Logistic Regression's GMV prediction models.

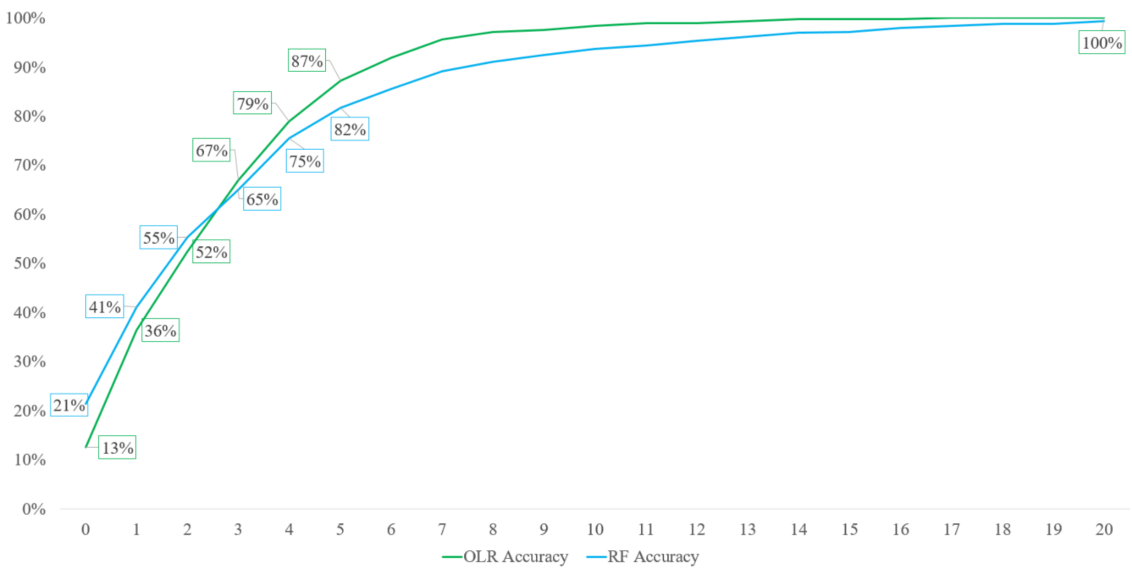


Figure 6.2: Results of the O.L.R and R.F models - Comparison of accuracy at *k* neighbours.

According to the results from Table 6.2, the Random Forests model was the one that yielded the best results in most of the metrics used. Figure 6.3 presents an overlay between the predicted classes and the correspondent actual GMV classes for the Random Forests model.

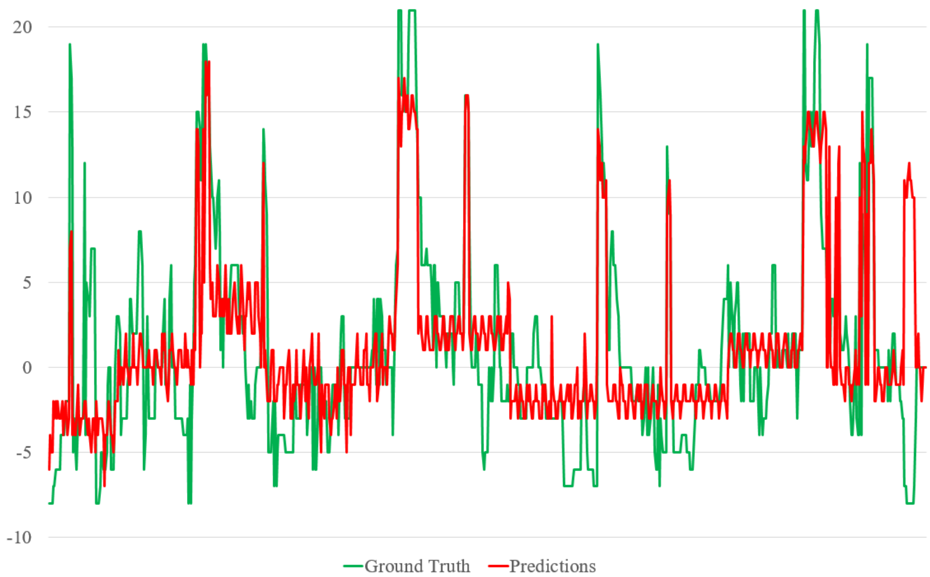


Figure 6.3: Comparison between predicted and actual classes for the GMV model.

Although the model is able to accurately predict some of the tendencies in Farfetch's GMV indicator, there are evident gaps in the model's learning capability which are represented by the oscillating prediction sequences.

NPS Variable

Table 6.3: Results for the NPS models, by performance metric and learning algorithm.

Metrics	O.L.R	N.B	R.F
M.S.E	1.18	8.1	0.94
W.A @ 5 Neighbours	0.97	0.91	0.94
Accuracy	0.47	0.16	0.32
Kappa	0.28	0.1	0.13
O.A.S	16.96	2.67	4.8

Regarding the outcomes for the NPS models, the Ordinal Logistic Regression model presented the best results, leading with the best values for most of the performance indicators.

Return Rate Variable

Table 6.4: Results for the Return Rate models, by performance metric and learning algorithm. [1/2]

Metrics	O.L.R	N.B	R.F	S.V.M (Linear)	S.V.M (Poly, N=2)	S.V.M (Poly, N=3)
M.S.E	19.6	120.5	34.2	26.3	31.6	33.7
W.A @ 5 Neighbours	0.45	0.18	0.51	0.45	0.43	0.48
Accuracy	0.06	0.06	0.14	0.10	0.09	0.12
Kappa	0.004	0.052	0.09	0.07	0.04	0.07
O.A.S	1.12	1.73	2.04	1.78	1.5	2.04

Table 6.5: Results for the Return Rate models, by performance metric and learning algorithm. [2/2]

Metrics	S.V.M (Poly, N=4)	S.V.M (Poly, N=5)	S.V.M (Radial)
M.S.E	27.6	32.8	27.7
W.A @ 5 Neighbours	0.53	0.49	0.49
Accuracy	0.15	0.12	0.12
Kappa	0.11	0.07	0.07
O.A.S	2.06	1.87	1.63

For the Return Rate model, the Support Vector Machines algorithm tuned with a polynomial kernel of $N = 4$ yielded the best results in comparison to the remaining models, presenting the best results for most of the evaluation metrics.

6.1.3 Deployment Test

The results presented above provided insights regarding which were the best models for each output variable. To grasp whether or not these models were generalizable for future alterations of pricing components the following tests were ran: the Shipping Subsidy and the FX Margin pricing variables were incremented in small values and the corresponding Random Forests' predictions regarding GMV variation were performed, *ceteris paribus*. Table 6.6 presents the results obtained from these tests. These results suggest that the model is not able to produce consistent predictions from the current set of past pricing variations.

Table 6.6: Results obtained during the deployment test for the GMV prediction model.

Input Variable	***p.p* Variation**	**Pred. GMV Class**
Shipping Subsidy	+0.25%	-0.53
-	+0.50%	-0.74
-	+0.75%	-0.98
-	**+1.00%**	**3.52**
-	**+1.10%**	**3.78**
-	+1.25%	-1.17
-	+1.50%	-1.42
-	+1.75%	-1.47
FX Margin	+1.00%	-1.11
-	+1.25%	-1.20
-	+1.50%	-1.37
-	**+1.75%**	**6.24**
-	**+2.00%**	**6.24**
-	**+2.25%**	**6.12**
-	+2.25%	-1.46
-	**+2.50%**	**6.80**

The results obtained reveal a lack of ability to generalize from past observations. This became evident due to the inconsistencies in the outputs and the lack of continuity for consecutive values of additional Shipping Subsidy and FX Margin.

6.2 Personalized Study Approach

For the personalized study on the impact of permanent free shipping in China, the key metrics under evaluation were NTV, Return Rate and Shipping Cost coverage.

NTV Variable

Multiple perspectives on business growth were prepared in order to quantify the increase in NTV growth caused by the implementation of permanent free shipping in China. Table 6.7 showcases the differences in business growth for the two "sides" of the of permanent free shipping threshold (500 USD).

Table 6.7: China's NTV growth on basket values below and above the permanent free shipping threshold.

Basket Value (USD)	NTV Growth
< 500	57.4%
>= 500	83.7%

The values above corroborate the effectiveness of permanent free shipping in driving growth in China's market.

Figure 6.4 showcases the difference in the business growth trend for the period before and after the implementation of permanent free shipping. The red cut lines correspond to peaks related to sales promotions and the green area corresponds to the gain between the two growth trends.

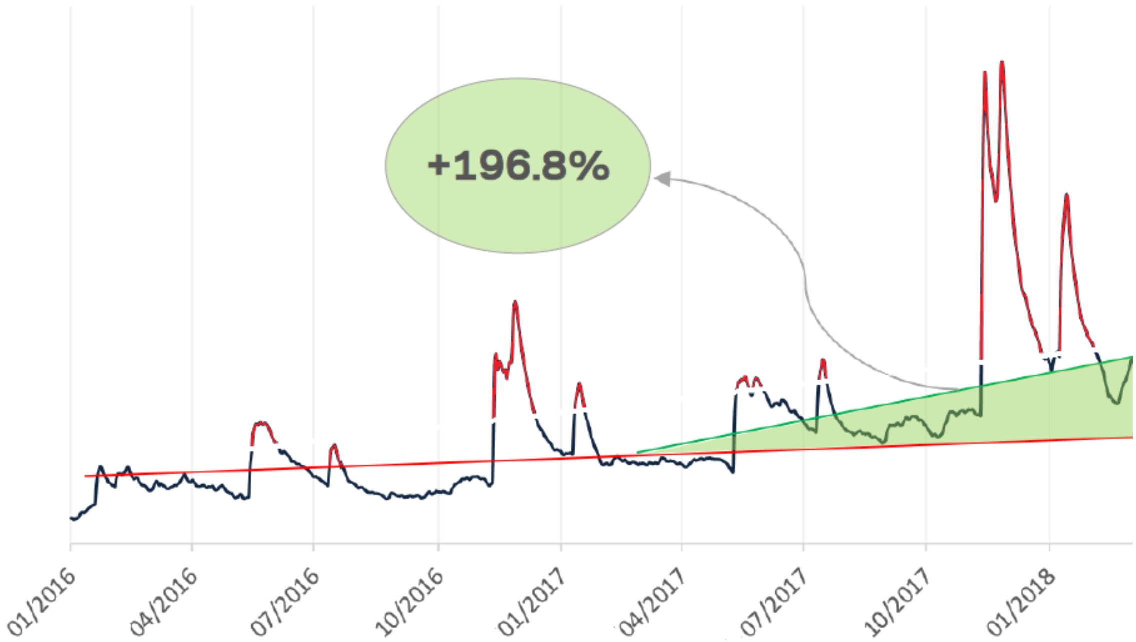


Figure 6.4: Increase of China's growth rate after the implementation of permanent free shipping.

Although this perspective only measures the overall growth for the period, the results still corroborate the hypothesis that free shipping helped drive forth growth in China.

In order to comprehend how NTV growth is distributed by basket value and by number of items purchased, the matrix in Figure 6.6 was developed. The value in each cell represents the Before$\rightarrow$After variation of the Number of Products-Basket Value pairings' contribution to total NTV. Two clear insights are the shift in demand towards baskets of value above the free shipping threshold and how this shift happened mostly for baskets of a single to three products.

The results show a shift in demand towards baskets of value above the free shipping threshold and how this shift happened mostly for baskets of a single to three products. The AOV increased for baskets containing 1 to 3 products.

% of ATV		Number of Products in Basket					
		1	2	3	4	5	>=6
Basket Value [USD]	0-100	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	100-200	-0,7%	-0,1%	0,0%	0,0%	0,0%	0,0%
	200-300	-1,1%	-0,2%	0,0%	0,0%	0,0%	0,0%
	300-400	-1,3%	-0,4%	-0,1%	0,0%	0,0%	0,0%
	400-500	-1,5%	-0,5%	-0,2%	0,0%	0,0%	0,0%
	500-600	0,2%	0,5%	0,4%	0,2%	0,1%	0,1%
	600-700	-0,1%	0,6%	0,3%	0,1%	0,0%	0,0%
	700-800	0,5%	0,5%	0,1%	0,1%	0,0%	0,0%
	800-900	0,5%	0,2%	0,1%	0,0%	0,0%	0,0%
	900-1000	0,3%	0,2%	0,0%	0,0%	0,0%	0,0%
	1000-1100	-0,1%	0,1%	0,0%	-0,1%	0,0%	0,0%
	1100-1200	0,1%	0,1%	0,0%	0,0%	0,0%	0,0%
	1200-1300	0,0%	0,1%	0,1%	0,0%	0,0%	0,0%
	>1300	0,3%	0,2%	0,3%	0,2%	-0,1%	0,0%

Figure 6.5: Variation of China's market NTV distribution, by basket value and number of items purchases, before and after permanent free shipping.

Return Rate Variable

Two perspectives on Return Rate were studied: regular Return Rate and partial Return Rate. Table 6.8 depicts the difference in Return Rate due to the implementation of Free Shipping.

Table 6.8: Return Rate growth after the implementation of permanent Free Shipping.

Period	Return Rate
20/08/2016 - 20/02/2017	14.4%
20/02/2017 - 20/08/2017	18.4%

The results reveal an increase of 4 p.p to the Return Rate. To understand what drove this increase the partial Return Rate metric was calculated. Figure 6.6 presents the relative growth of the partial Return Rate between the periods of one year before and one year after the implementation of permanent Free Shipping.

By revealing that the partial return rate, on average, improved for basket values below the free shipping threshold and skyrocketed for basket values immediately above the free shipping threshold, the chart corroborates the hypothesis that chinese customers are buying extra items in order to reach the Free Shipping threshold and returning them. This practice massively prejudices Farfetch because the customer does not cover the costs associated with a return.

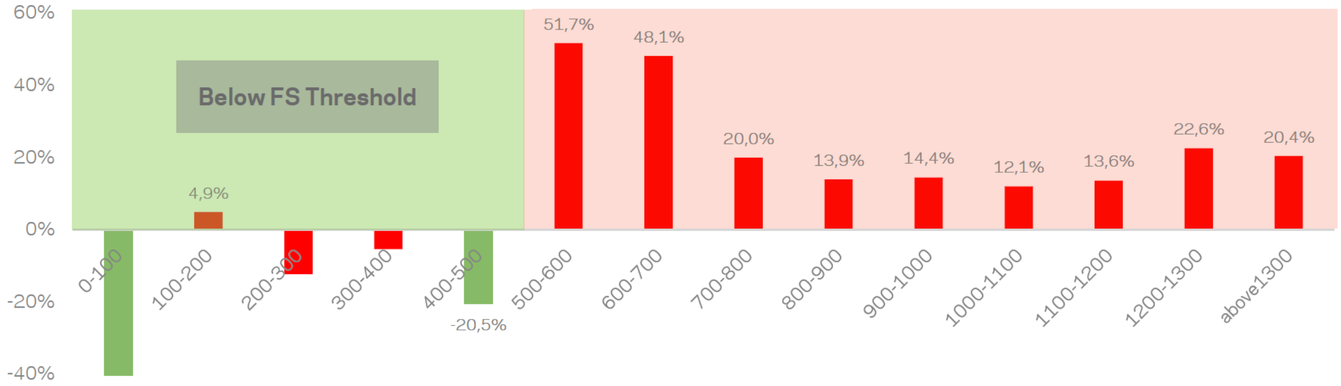


Figure 6.6: Relative growth of China's partial Return Rate, by basket value.

Shipping Cost Coverage

Figure 6.7 illustrates the difference in shipping cost coverage in 2017 by comparing the real shipping subsidy rates with an ideal constant rate.

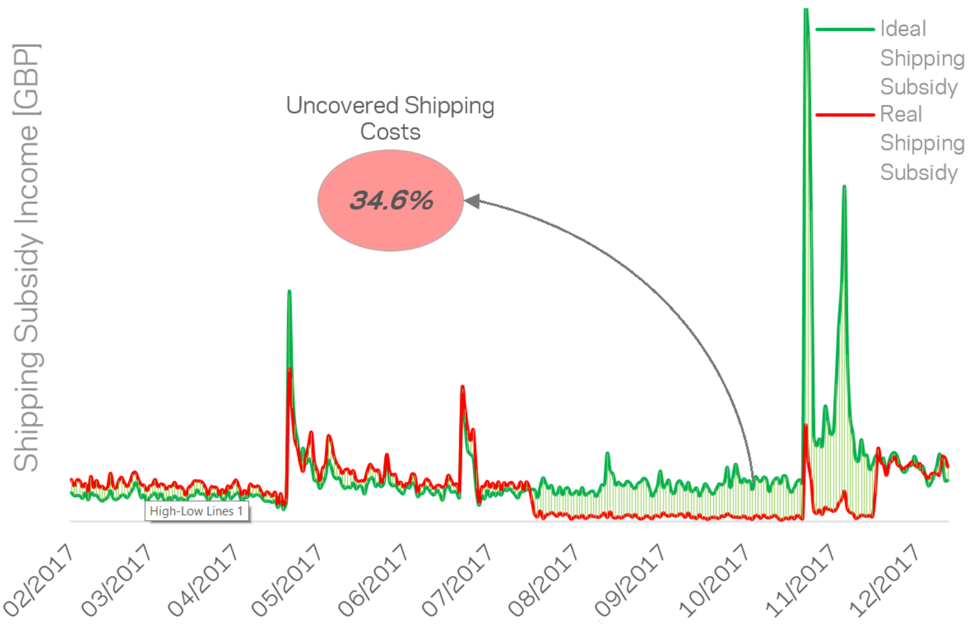


Figure 6.7: Comparison between shipping cost coverage for the real and ideal shipping subsidy rates.

6.3 Takeaways from both approaches

When comparing the results from the machine learning models with the deep-dive exploratory analysis, it became clear that each approach presented advantages and disadvantages. Table 6.9

lists the main advantages and disadvantages of the two approaches developed.

Table 6.9: Advantages and disadvantages of the two approaches.

	Machine Learning	**Personalized Study**
Advantages	Fast predictions of expected impact Accurate results Easy to adapt to different output indicators	Extract specific metrics or insights Useful when data available is limited
Disadvantages	Requires large amounts of data	Time consuming

The machine learning approach led to the conclusion that Free Shipping campaigns and Sales events were the most determining factors for GMV variation. Additionally, the Shipping Subsidy and FX Margin components presented an unexpectedly high impact when compared with the two variables aforementioned. A reason for this result might be that these variables sustained large variations in the period of analysis which in turn conditioned the company's price competitiveness in China. The inconsistent results from the deployment test suggest that the model will require more data to properly learn the full extent of the impact of pricing alterations. In time, the model is expected to have more information to learn from and improve its current effectiveness, since Farfetch is planning to start to leverage their control on price and test price sensitivity in multiple markets.

On the other hand, the personalized exploratory study brought valuable insights about the impact of permanent Free Shipping in China. The analysis revealed that although permanent Free Shipping generated growth in the chinese market, Farfetch's Return Rate and shipping cost coverage suffered significant damage. The Return Rate increased by 4 p.p mostly due to China's customers taking advantage of the Free Shipping Threshold. The shipping subsidy set for China was not able to cover all the additional shipping costs. According to this study, Farfetch would have to increase China's shipping subsidy rate to 5.79% to fully cover the current losses.

Chapter 7

Conclusions

This dissertation addressed the analytical problem of assessing the impact that Farfetch's pricing components had on the company's main sales and customer satisfaction indicators. The main approach followed in this study was based on machine learning algorithms. In parallel, a deep-dive exploratory analysis of China's permanent free shipping conditions was performed in order to compare the two approaches in terms of their viability and relative potential in assessing the impact of price alterations in Farfetch's main business indicators.

One of the main concerns was whether or not there was enough data for the learning algorithms to grasp the intrinsic relationships between the pricing variables and the respective output indicators. Initially, a preliminary exploratory study of China's dataset was performed in order to know what values each pricing component presented during the period of analysis and what was the mean difference in sales for different factor levels of each pricing component. The results suggested that higher pricing component rates typically leads to low sales' performance relative to the business growth trend, which was to be expected since increasing price directly impacts the company's price competitiveness in the market.

Along with this analysis, the deep-dive study was performed and led to multiple actionable insights. The implementation of a permanent free shipping campaign with threshold in China increased the company's partial return rate by over 30% on basket values above the free shipping threshold, which in turn resulted in unexpected losses due to the return of the orders' FX Margin, Shipping Subsidy and Final Rate, that help to finance the company's operational activity. Additionally, the shift in AOV was quantified and the business growth driven by this initiative was discriminated through multiple perspectives due to the difficulty of isolating the impact of simultaneous effects that promote business growth. Lastly, a shipping subsidy rate of 5.79% was proposed in order to fully cover all additional shipping costs resulting from the permanent free shipping campaign.

Regarding Data Processing and Model Evaluation, the variable transformation from numerical to ordinal categorical through value binning proved to be useful in reducing data noise and providing a proxy to analyze the results through a broader range of performance evaluation metrics. For the ordinal categorical classification problem, the Window Accuracy metric at *k* neighbours revealed significant potential in delivering a more fulfilling view of model accuracy, specially in problems with a high number of classes.

In terms of sales performance, free shipping and promotional events are the most impacting variables, closely followed by the remaining price components. For the GMV, NPS and Return Rate indicators, the Random Forests, Ordinal Logistic Regression and tuned Support Vector Machines models yielded the best results, respectively. However, the deployment test revealed inconsistent results in terms of the models' current ability to support pricing alterations.

For the machine learning approach to be feasible, the model will have to be fed with more data in order to fully capture the intrinsic relationships of the dataset. Two of Farfetch's plans for the upcoming year are to leverage results through the use of price and to study the price sensitivity of each market through price A/B testing. It is expected that with these future price component alterations the models' effectiveness will improve and provide more actionable results. Additionally, non-price related variables would have to be included to fill the "learning gap" that using only pricing-related variables causes. In case it is not possible to guarantee these two conditions, then a deep-dive analysis for a specific market will be the better option, as seen by the results of the study aforementioned.

Regarding future work, the addition of non-pricing related variables to the predictions models developed would be valuable to fill the "learning gap" that using only pricing-related variables causes and learn the impact of other business-related inputs. Additionally, a practical approach to this problem would be to prepare a simulation where different positive and negative scenarios (eg: 10% or 20% uplift in sales) would be tested for any variation of any pricing component and for any specific market. This type of sensitivity analysis would provide a solid base for future pricing decisions that paired with the business knowledge of an Operations' analyst could lead to faster and more better informed decisions regarding potential pricing alteration outcomes. Lastly, the price elasticity of demand could be analyzed by running short A/B tests on priority markets and using the pricing components controlled by the Operations department as a way to test the impact of specific pricing variations.

Bibliography

Ahmed, P., Rafiq, M., and of Marketing, C. I. (2002). *Internal Marketing: Tools and Concepts for Customer-focused Management*. Chartered Institute of Marketing. Butterworth-Heinemann.

Archak, N., Ghose, A., and Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8):1485–1509.

Bain & Company (2018). *Luxury Goods Worldwide Market Study, Fall–Winter 2017*. Business Insights.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer.

Blanchard, G., Bousquet, O., and Massart, P. (2008). Statistical performance of support vector machines. *Annals of Statistics*, 36:489–531.

Boston Consulting Group (2014). *Shock of the New Chic: Dealing with New Complexity in the Business of Luxury*.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Cardoso, J. S. and Sousa, R. G. (2011). Measuring the performance of ordinal classification. *IJPRAI*, 25:1173–1195.

Caruana, R. and Niculescu-Mizil, A. (2004). Data mining in metric space: An empirical analysis of supervised learning performance criteria. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 69–78, New York, NY, USA. ACM.

Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 161–168, New York, NY, USA. ACM.

Catena, M., Remy, N., , and Durand-Servoingt, B. (2015). Is luxury e-commerce nearing its tipping point?

Chong, K. (2003). *The Role of Pricing in Relationship Marketing: A Study of the Singapore Heavy Equipment Spare Parts Industry*. University of South Australia.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Deloitte (2015). *Future of e-Commerce: Uncovering Innovation*.

Deloitte (2017). China luxury: E-commerce whitebook.

Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2014). Consumer price search and platform design in internet commerce. Working Paper 20415, National Bureau of Economic Research.

Doherty, N. F. and Ellis-Chadwick, F. (2010). Internet retailing: the past, the present and the future. *International Journal of Retail & Distribution Management*, 38(11/12):943–965.

Doss, F. and Robinson, T. (2013). Luxury perceptions: luxury brand vs counterfeit for young us female consumers. *Journal of Fashion Marketing and Management: An International Journal*, 17(4):424–439.

EBC (2017). Predicting with labeled data.

Efendioglu, A. M., Yip, V. F., and Murray, W. L. (2005). E-commerce in developing countries: Issues and influences.

Franzé, G. E. (2016). Creating the ultimate luxury fashion customer experience.

Fruhling, A. L. and Digm, L. A. (2000). The impact of eelctronic commerce on business level strategies. *J. Electron. Commerce Res.*, 1(1):13–22.

Gaudette, L. and Japkowicz, N. (2009). Evaluation methods for ordinal classification. In Gao, Y. and Japkowicz, N., editors, *Advances in Artificial Intelligence*, pages 207–210, Berlin, Heidelberg. Springer Berlin Heidelberg.

Hannak, A., Soeller, G., Lazer, D., Mislove, A., and Wilson, C. (2014). Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, IMC '14, pages 305–318, New York, NY, USA. ACM.

Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition.

Hauser, G. (1993). *Galton and the Study of Fingerprints*, pages 144–157. Palgrave Macmillan UK, London.

Hines, T. and Bruce, M. (2007). *Fashion Marketing*. Taylor & Francis.

Hosmer, D. W. and Lemeshow, S. (2005). *Applied Logistic Regression*. Wiley-Blackwell.

Husic, M. and Cicic, M. (2009). Luxury consumption factors. *Journal of Fashion Marketing and Management: An International Journal*, 13(2):231–245.

Hwang, S. B. and Kim, S. (2006). Dynamic pricing algorithm for e-commerce. In Sobh, T. and Elleithy, K., editors, *Advances in Systems, Computing Sciences and Software Engineering*, pages 149–155, Dordrecht. Springer Netherlands.

International Chamber of Commerce (1980). *Incoterms*. International Chamber of Commerce.

Jurievich, P. A. (2012). Marketing plan for e-commerce start-up company.

Khan, A. G. (2016). Electronic commerce: A study on benefits and challenges in an emerging economy. *Global Journal of Management and Business Research*.

Koehrsen, W. (2017). Random forest simple explanation. *Medium*.

Kurtz, D. and Boone, L. (1987). *Marketing*. Dryden Press Series in Finance. Dryden Press.

Laudon, K. C. and Laudon, J. P. (2001). *Management Information Systems: Managing the Digital Firm*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 7th edition.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436—444.

Lewis, M. (2006). The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing*, 82(1):13 – 23.

Marc Bain (2018). 3 retail technology trends transforming luxury.

Mohri, M., Rostamizadeh, A., Talwalkar, A., and Bach, F. (2012). *Foundations of Machine Learning*. MIT Press.

Okonkwo, U. (2016). *Luxury Fashion Branding: Trends, Tactics, Techniques*. Palgrave Macmillan UK.

Park, H.-A. (2013). An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. College of Nursing and System Biomedical Informatics National Core Research Center, Seoul National University.

Pasumarthy, P. B. and Kumar, P. (2015). E-loyalty and e-satisfaction of e-commerce. *International Journal in Management and Social Science*.

Reitermanová, Z. (2010). Data splitting.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Samantha Woodworth (2018). The luxury online fashion market is set to more than triple by 2025.

Sayad, S. (2018). Logistic regression.

Shearer, C. (2000). The crisp-dm model: The new blueprint for data mining. *Journal of Data Warehousing*, 5(4).

Shin, N. (2001). Strategies for competitive advantage in electronic commerce. *Journal of Electronic Commerce Research*.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4):427 – 437.

Spüler, M., Sarasola Sanz, A., Birbaumer, N., Rosenstiel, W., and Ramos-Murguialday, A. (2015). Comparing metrics to evaluate performance of regression methods for decoding of neural signals.

Sridharan, M. (2015). Data mining: Practical machine learning tools and techniques.

Statista (2018). *e-Commerce: Fashion*.

Stax Inc (2016). Customer decision making criteria and the importance of price.

U.S. Bureau of the Census (2018). E-commerce retail sales [ecomsa]. retrieved by fred.

Vlasova, A. (2017). *Economic Power of E-retailers Via Price Discrimination in E-commerce: Price Discrimination's Impact on Consumers' Choices and Preferences and Its Position in Relation to Consumer Power*. LAP LAMBERT Academic Publishing.

Vulkan, N. (2003). *The Economics of E-commerce: A Strategic Guide to Understanding and Designing the Online Marketplace*. Princeton University Press.

Wasserman, L. (2010). *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated.

White House (2015). Big data and differential pricing.

Witten, I., Frank, E., Hall, M., and Pal, C. (2016). *Machine Learning Algorithms Mindmap*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, 8(7):1341–1390.

Wootric (2018).

Yoon, Y. and Swales, G. (1991). Predicting stock price performance: a neural network approach. In *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences*, volume iv, pages 156–162 vol.4.

Appendix A

Equivalence table for the GMV variable

Table A.1: Equivalence table for the GMV variable.

Relative Distance (%)	Class/Level
$]-\infty,-260]$	-21
$]-260,-240]$	-20
$]-240,-220]$	-19
$]-220,-200]$	-18
$]-200,-180]$	-17
$]-180,-160]$	-16
$]-160,-140]$	-15
$]-140,-120]$	-14
$]-120,-100]$	-13
$]-100,-90]$	-12
$]-90,-80]$	-11
$]-80,-70]$	-10
$]-70,-60]$	-9
$]-60,-50]$	-8
$]-50,-40]$	-7
$]-40,-30]$	-6
$]-30,-25]$	-5
$]-25,-20]$	-4
$]-20,-15]$	-3
$]-15,-10]$	-2
$]-10,-5]$	-1
$]-5,5]$	0
... (*symmetrical*)	...