Improving Supply Chain efficiency in e-tail by Redirecting Returns: a Machine Learning approach

André Meneses Rodrigues de Oliveira Alves

Master's Dissertation

Supervisor: Prof. Dr. Gonçalo Figueira

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Integrated Master's in Mechanical Engineering

2018-07-02

Abstract

E-commerce is growing over the years and with this growth new and demanding challenges present themselves to companies who make this business its core. In particular, supply chain management and cost efficiency in returns play a big role in a company's profit and loss. And now, more than never, decisions are driven by big data, which continues to pave its way into modern businesses. This represents an opportunity to better ground business analysis, sustain models, and take better decisions that can have crucial impacts in daily operational tasks.

The current dissertation focuses on studying a solution where customer returns would be redirected to new clients, therefore potentially reducing costs in return management. The project is divided in two main steps. The first step consists in portraying the global framework of the online marketplace for luxury fashion. A selection of the top markets for the proposed solution is filtered and a study of the impacts is conducted in order to predict the range of potential savings. The second part of this project is focused on building an algorithm through Machine Learning. In order to do so, a dataset containing information about returned products is gathered and an exploratory analysis is performed with the objective of classifying a return as resalable within the next 15 days from the return date. Three different machine learning algorithms are tested. *Random Forest, Logistic Regression* and *Support Vector Machine* are assessed and tuned in order to find the best classifier.

The top conclusions are that redirecting returns is shown to have the best opportunity through returns Via UK. And finally, the Random Forest algorithm demonstrated the best results using the features selected for the returned products dataset, which are shown to be important.

Resumo

O *E-commerce* está constantemente a crescer e, com este crescimento, novos desafios surgem nas empresas que fazem deste negócio o seu núcleo. Em particular, a gestão da cadeia de abastecimento e a gestão das devoluções assumem papéis fundamentais no balanço de uma empresa. E agora, mais do que nunca, decisões estão a ser guiadas por *Big Data* que continua a crescer dentro das empresas modernas, ajudando a sustentar análises, a construir modelos e a melhorar as tarefas operacionais do quotidiano.

Esta dissertação foca-se em estudar uma solução para a gestão de devoluções em que estas são redirecionadas para novos clientes sem que tenham de regressar à *boutique*, potenciando assim uma redução de custos. O projeto foi dividido em duas etapas. Primeiro, obtém-se uma visão geral da oferta e da procura, naquilo que é o mercado de retalho de moda de luxo. Filtra-se os possíveis candidatos para a implementação da solução e calcula-se o intervalo de redução de custos. A segunda etapa do projeto consiste na construção de um algoritmo através de *Machine Learning*. Para tal efeito, um *dataset* contendo informações acerca dos produtos devolvidos é extraído e uma análise exploratória às variáveis é realizada com o intuito de classificar uma devolução como passível de ter uma revenda nos 15 dias após à data de criação da devolução. São utilizados três algoritmos de *Machine Learning*. *Random Forest, Logistic Regression, Support Vector Machine* são testados e otimizados para escolher o melhor classificador.

As principais conclusões são que o redirecionamento das devoluções mostrou ter a melhor oportunidade nas devoluções que seguem Via UK. E finalmente, o algoritmo *Random Forest* mostrou ser o mais versátil e com o melhor potencial na avaliação das devoluções usando as características escolhidas que demonstraram ser importantes na criação do algoritmo.

Agradecimentos

Muitas pessoas interagiram comigo neste percurso de 5 anos. Apesar de todas terem contribuído, em pequenas partes, para desenvolver a minha personalidade e o meu caráter enquanto pessoa e jovem adulto, sinto a necessidade de destacar algumas pessoas. Em primeiro lugar, gostaria de agradecer aos meus pais, por terem sido os grandes impulsionadores da minha educação e por sempre me motivarem a dar o meu melhor. Um agradecimento especial à minha irmã, Cláudia, que sempre me guiou com uma palavra no jeito de conselho e me ajudou naquilo que foram 5 anos de constantes efemeridades. Quero ainda direcionar uma palavra especial ao meu "irmão por lei", Tiago, que também sempre tinha um conselho pronto para me dar.

Gostaria agora de agradecer à equipa de Delivery Development da Farfetch. Obrigado por me receberam tão bem e fazerem-me sentir bem a trabalhar, não estava à espera de receber um ambiente assim. À Ana, Diana, Hugo, Isabelle, Rita e Wilson um muito obrigado. Em particular, um agradecimento ao meu manager Pedro Bastos, também meu supervisor, por me ter dado a oportunidade de aprender com ele e de poder integrar esta equipa. Um agradecimento especial ao Tomás e à Inês por me ouvirem constantemente com as minhas dúvidas e hesitações, sem a experiência deles não teria conseguido desenvolver tão bem esta dissertação. E ao Diogo, por ter sido quem sempre me acompanhou nesta aventura na Farfetch, pela sua extraordinária capacidade de "educated guessing" e de aconselhamento.

Aos meus amigos, aos novos que formei durante este período de 5 anos (com um carinho especial aos de Erasmus) e aos que já me acompanham desde há mais anos, um agradecimento sentido por sempre me fazerem rir e nunca alguma vez me deixarem sem apoio.

Por último, agradecer ao meu orientador na FEUP, o Prof. Dr. Gonçalo Figueira, pelo seu guiamento e intuição, fazendo com que esta dissertação se desenvolvesse de uma maneira eficaz e frutífera.

"But an inner voice tells me that it is not yet the real thing. The theory says a lot, but does not really bring us any closer to the secret of the "old one." I, at any rate, am convinced that He does not throw dice."

Albert Einstein in a letter to Max Born, 1926

Contents

Acronyms and Symbols

3PL	Third-Party Logistics
APV	Average Product Value
ATV	Actual Transacted Value
AWB	Air Waybill
BDA	Big Data Analytics
D&T	Duties & Taxes
FN	False Negative
FP	False Positve
GMV	Gross Merchandise Value
KPI	Key Performance Indicator
LTS	Lead Time Sales
LR	Logistic Regression
ML	Machine Learning
OS	One Size
RF	Random Forest
RL	Reverse Logistics
ROC	Receiver Operating Characteristic
RSCM	Reverse Supply Chain Management
RTO	Return to Origin
RR	Return Rate
SLA	Service Level Agreement
SVM	Support Vector Machine
SRS	Sale Return Sale
TN	True Negative
TP	True Positive
VAT	Value Added Tax
WCO	World Customs Organization

List of Figures

List of Tables

Chapter 1

Introduction

As online e-commerce continues to grow, the need to be as efficient and cost effective as possible has never been so crucial. Globalized trade has allowed a massive upscale for almost every industry and for the online e-commerce in particular it represented an enormous uplift.

Personal luxury goods is one of the main sectors in the industry of fashion, having reached a record high of €262 billion in 2017 (Bain & Company, 2017). Moreover, this market is predicted to grow, in 2018, 4% to 5% as McKinsey and Business of Fashion settle in their annual State of Fashion report (Business of Fashion & Mckinsey, 2018). On top of that, taking into account that 60% of fashion executives are investing in e-commerce and omni-channel integration for 2018 (Business of Fashion & Mckinsey, 2018) and the fact that between 2016 and 2017 online shopping for personal luxury goods rose 24% (Bain & Altagamma, 2017), these represent strong indicators of this market's potential.

The luxury fashion e-commerce sales are expected to be 25% (Bain & Company, 2017) of the total personal luxury goods sector by 2025 so this translates into a major opportunity for fashion e-tailers. The need to have an efficient supply chain management and logistic solutions in such a demanding, heterogeneous and global market becomes a key component to stay competitive and profitable.

Returns and return management represent a major role in logistics profit and losses and it is also a main driver for customer satisfaction (Rogers et al., 1999). For that reason, projects aimed at reducing costs in returns and improving the customer's experience become a priority, not only for e-commerce, but for companies thriving for success in today's global market.

Due to its added complexity, return flow is often costlier than the forward distribution of goods. Not only retailers are not able to sell the item while it is being processed, returned items often need to be recycled, dismantled or perform other activity in order to recapture its value. This may often result in expensive processes that fail to bring value to the company.

Considering all the arguments aforementioned, this dissertation embraces a new point. The underlying research problem is to analyze the return flow of the products and study a solution where returned products are redirected to new customers without being sent to the original partners whilst service and cost levels objectives are met. This thesis is conducted using a case-study of a large e-tailer: Farfetch.

1.1 Luxury E-Commerce and Farfetch

Farfetch is an e-commerce platform that connects consumers to carefully selected products from a large network of more than 700 boutiques and 200 brands (South China Morning Post, 2016). Since it was launched in 2008, the company's vision has always been to be the global platform for luxury. This is the vision of the founder and CEO José Neves that, back in 2007, spotted an opportunity in merging luxury fashion with online e-commerce.

Farfetch's business model is based on a commission per sale and it distinguishes itself from its competitors in the fact that it does not possess any stock it sells. Due to this fact and the consequent complexity of its supply chain, Farfetch operates in a drop-shipping model, which means it serves as an intermediary between boutiques and consumers, using third party logistics (3PL) partners to fulfill the delivery process.

This company offers a premium service to customers by delivering to their home luxury products worldwide and providing boutiques with access to an e-commerce platform, which acts as a sales uplift, where payment processing, customer service, marketing, fraud and customer service is handled by Farfetch.

Farfetch currently has offices in 11 sites including Porto, Lisbon, Guimarães, Braga, São Paulo, Hong Kong, Moscow, Los Angeles, New York, London and Tokyo. It employs more than 2000 people in its offices worldwide and has been growing at a pace of 70% YoY. It was one of the first of its kind to achieve the "unicorn" status which labels "start-up's" valued above of $1 billion. In Figure 1.1, the key markets are portrayed. As of 2017 the list contains the United States, United Kingdom, Hong Kong, Russia, Australia, Germany and China.

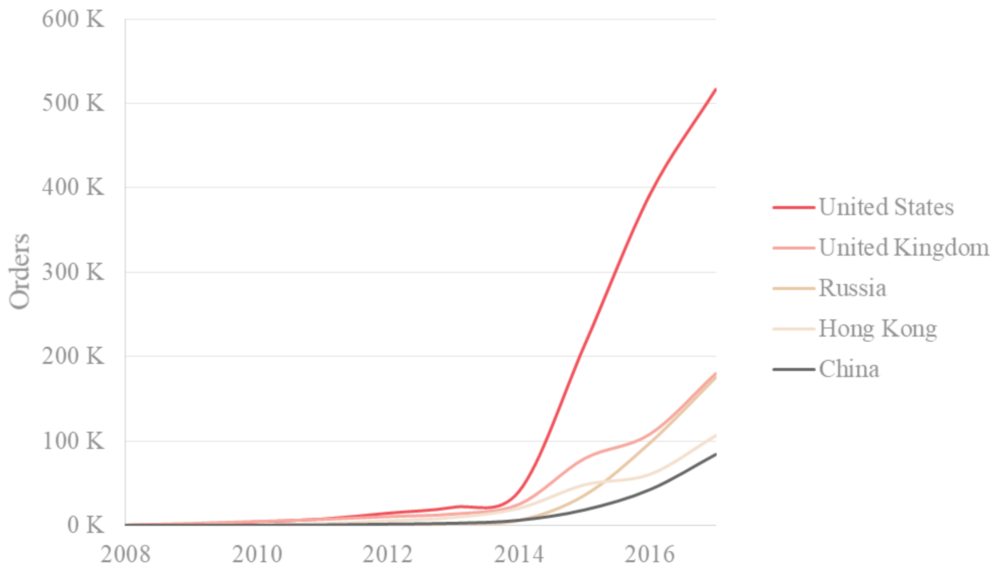


Figure 1.1: Growth of the main markets of Farfetch.

In the beginning of 2018 news on Farfetch teaming up with the Chalhoub group, one of the biggest distributors of fashion and luxury goods in the Middle East, started to appear (Business of Fashion, 2018). This is aligned with the company plans to grow its presence in this market and to match its leading competitor Yoox Net-a-Porter Group who already launched offices and distribution centers in Dubai at the end of 2017 (Fashionista, 2018). More recently, Farfetch has been partnering with bankers from JP Morgan and Goldman Sachs planning an eventual IPO with a valuation of possibly $5 billion (Financial Times, 2018). These business plans are supported by the last round of investment in the company led by the Chinese e-commerce giants JD.com (TechCrunch, 2017).

1.2 Project

In Farfetch, there is currently an average 20% return rate (RR). Each product is shipped from the boutique to the consumer, and then backwards, from the consumer to the boutique in what it is called reverse logistics. As one of the main values of Farfetch is to "Amaze Customers", the company is very lenient in its return policy easily offering free returns driven by customer requests.

This translates into relatively high costs for returns, approximately £ 11 M in 2017. As stated beforehand, one of the main strategic objectives of Farfetch for the next years is to consolidate operations and reduce costs. The dissertation has two main objectives. The first focuses on finding the best conditions to redirect returns based mainly on simulating the expected savings with the proposed solution against the existing process where products flow across the entire supply chain. The second consists in selecting which items should be redirected and develop an algorithm, based on machine learning, to predict such selection, considering the costs of a wrongful prediction. This will have in sight a possible operational tool that would decide which products to maintain and which products to return to the boutique.

A series of steps are presented in an attempt to structure the approach and perform a deconstruction of the complex problem presented beforehand:

1. Supply chain mapping – A geographical overview of supply and demand as well as a review of the ordering and return processes;

2. Scenario design – Identify opportunities for possible savings with the proposed solution and outline every possible situation where that can occur;

3. Candidate evaluation and first simulations – Assessment of savings for each scenario proposed and selection of the one which represents the best opportunity;

4. Dataset exploratory analysis – Using statistical tools, understand which features are the most important for the classifier;

5. Machine learning technique test and parameter tuning – With the right independent variables, test several machine learning techniques with the aim of finding the correct fit for this case;

6. Results and adjustment of savings – After having found the right learning technique, connect the results with the impacts predicted in the first steps and adjust the possible savings.

To better plan the distribution of work needed in order to successfully achieve the goals described in the section above, a *Gantt* diagram was constructed, as seen in Figure 1.2.

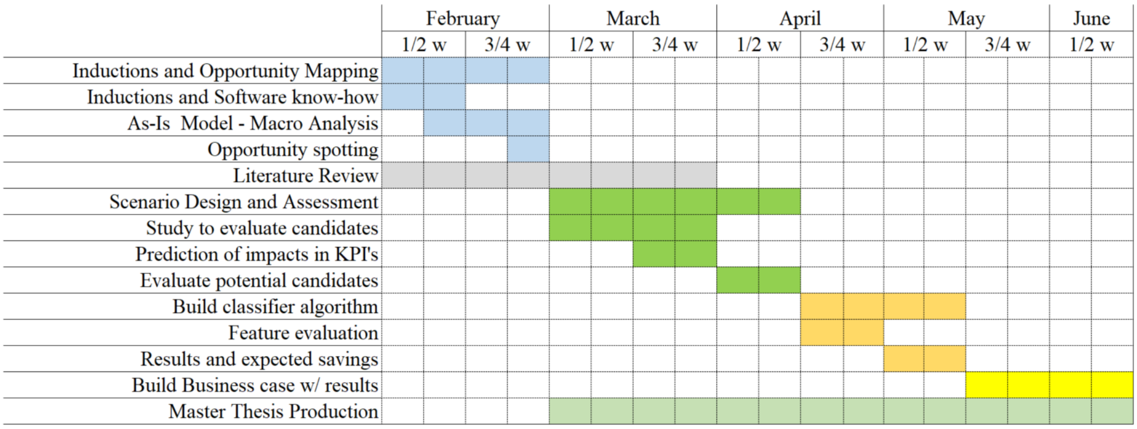


Figure 1.2: Gantt diagram with the different stages of the methodology.

From a certain point of view, the approach selected is similar to Deming's Plan Do Study Act (PDSA) cycle (Deming Institute, 2018). In this continuous improvement methodology, Deming developed a four-way model of implementing change in a new or in an existing process, as portrayed in Figure 1.3.

It is possible to draw an analogy between the Deming's cycle and this dissertation's proposed methodology. Being a twofold approach, some steps may not be strictly in numerical order. Steps 1, 2 and 4 correspond to the *Plan* phase where the solution is planned and analyzed. *Do* and *Study* phases are interconnected and steps 3 and 5 share the characteristics of those phases. Lastly, step 6 is the *Act* phase where results are adjusted and conclusions are drawn.

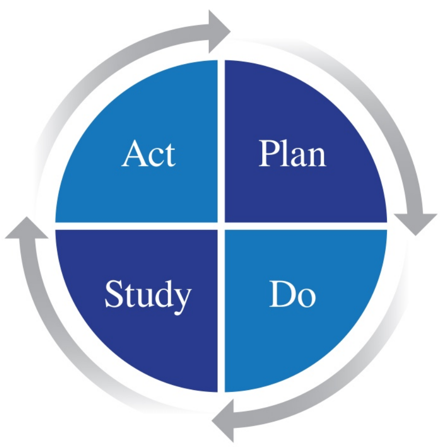


Figure 1.3: Deming PDSA Cycle, in (Deming Institute, 2018).

1.3 Structure

The structure of this dissertation is organized in the following way to give a complete scope of the project. Chapter 2 comprises a literature review on reverse logistics, data mining and machine learning theoretical principles, from which classification algorithms are further explored. Chapter 3 is the description of the current flow of products across Farfetch's supply chain, corresponding to step 1 of the methodology. The return process is mapped, providing insight of the opportunities for this project, as well as an analysis on the delivery cost structure. Chapter 4 is a study of the impacts and scenario design, corresponding to steps 2 and 3. Steps 4, 5 and 6 are accounted for in Chapter 5. This chapter focuses on a machine learning problem, that involves feature evaluation. The classifier and adjustment of the expected savings are also stated in Chapter 5. Finally, Chapter 6 is a conclusive reflection about the results and future work that would be complementary to the study developed over this dissertation.

Chapter 2

State of the Art

Throughout this chapter, the relevant concepts for this dissertation are explored. Firstly, the definition of Reverse Logistics, the drivers and current framework on that subject are detailed. On a different, but equally relevant matter, is Big Data and Machine Learning. The basics of Machine Learning are covered and with particular attention to Classification Algorithms.

2.1 Reverse Logistics

Reverse Logistics (RL) is generally defined as being the process of planning, implementing, and controlling the efficient, cost effective flow of raw materials, in-process inventory, finished goods and related information from the the point of consumption to the point of origin for the purpose of recapturing or creating value or proper disposal (Rogers et al., 1999). Other definitions may have a narrower perspective, i.e., more recycling oriented interpretations (Guiltinan and Nwokoye, 1974), which is not the case of this dissertation. For instance, RL activities include collection, disassembly, and processing of used products, product parts as well as materials with the purposes of reusing the items or assigning an environmentally friendly recovery (Kokkinaki et al., 2000).

Other authors suggest a Reverse Supply Chain Management (RSCM) as an extension of traditional supply chains (Prahinski and Kocabasoglu, 2006). Although this perspective was common up until the late 90's, in the beginning of the new century, research showed that, in a retail environment, the reverse flow of items is not a mirror of the forward flow (Tibben-Lembke and Rogers, 2002), as seen in Figure 2.1.

The definition of RL has been changing over time, widening its scope with the increased interest of researchers (Agrawal et al., 2015). For (Pokharel and Mutha, 2009) one of the main drivers behind RL is the potential to recover value from used products. On top of that, consumer awareness, legislation, directives (Ravi and Shankar, 2005) and environmental responsibilities (Bloemhof-Ruwaard et al., 1999) are contributing to the interest in optimizing RL. In this case-study, the majority of products are in good conditions, so the focus is more about recovering its value rather than social or environmental responsibilities. (De Brito and Dekker, 2002) suggest that the drivers should be categorized into three distinct areas: Economics (direct and indirect), Legislation and Extended Responsibility.

Indirect Economics can be related with the recovery of spare parts or abating costs. Moreover, a company can get involved in RL as a marketing strategy or to prepare for future legislation. Extended responsibility is the case when a company is impelled to become responsibly engaged to RL due to a set of values or principles (De Brito and Dekker, 2002).

As depicted in Figure 2.1, a typical RL structure develops over multiple stages: product acquisition, collection, inspection, sorting and disposition. These stages are further discussed.

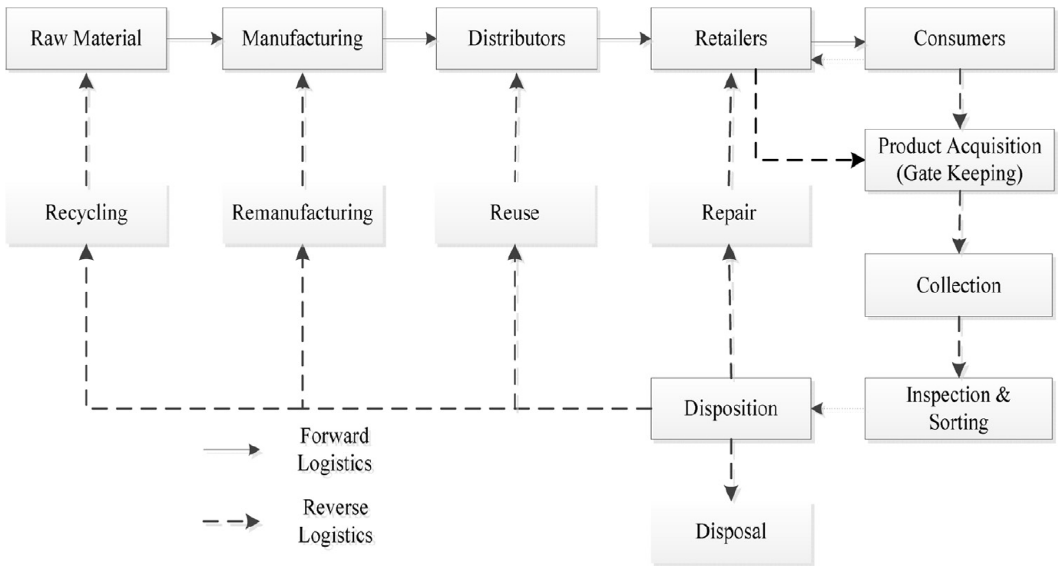


Figure 2.1: Forward and Reverse flow comparison, in (Agrawal et al., 2015).

Product acquisition

The first step in a RL structure is to secure the item, component or material from the customers for further processing. This step is of particular importance since products have an uncertain return date as well as an undefined quantity or quality (Fleischmann et al., 1997). Research has also shown that the acquisition of returns is of vital importance in order to have a profitable RL (Guide and Wassenhove, 2003).

Collection

After acquisition, products are collected in order to have a post processing phase which involves inspecting, sorting and disposing the item. This phase alludes to the action in which a firm claims ownership of the items (Fleischmann et al., 2003). As (Kumar and Putnam, 2008) have discussed, there are three fundamental kinds of collection strategies, that rely upon where the collection happens: from customers directly, through retailers or by means of an outsider 3PL company, portrayed in Figure 2.2.

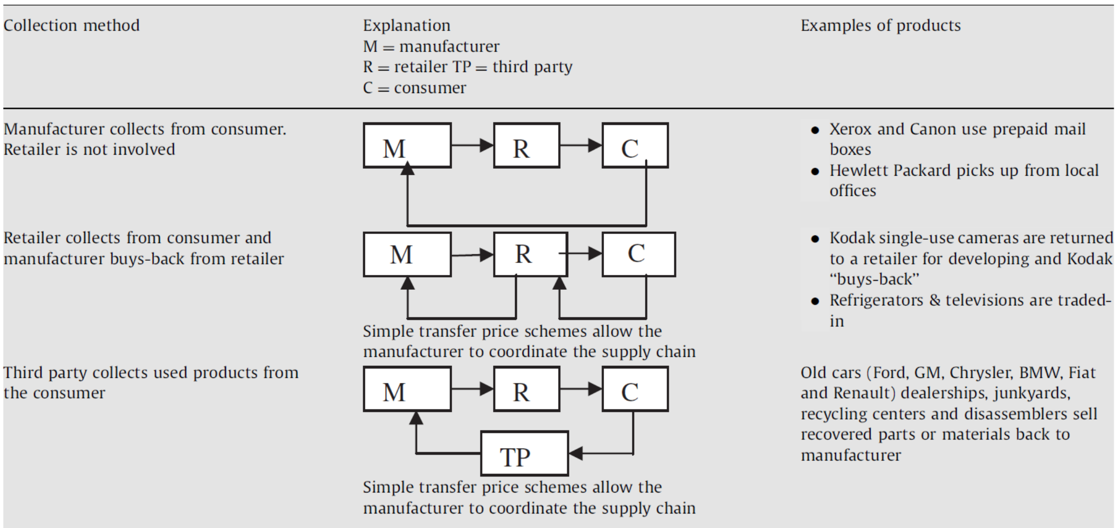


Figure 2.2: Types of collection methods, in (Kumar and Putnam, 2008).

Inspection and Sorting

(Rogers et al., 1999) shed light to the fact that a customer may return the items due to known or unknown reasons, and, on top of that, the state of the returned products may contrast greatly. For this reason, a dedicated inspection for each product is required for adequate sorting. The timing of this action depends on the transportation, disposal and disassembly cost, according to what (Zikopoulos and Tagaras, 2008) found. Research on optimal acquisition and sorting policies has been done and it was concluded that, for linear acquisition costs, such optimal policy does exist (Galbreth and Blackburn, 2006).

Disposition

The next stage is to decide how to dispose the item for further processing. Several authors have researched ways of disposing items. While (Thierry et al., 1995) categorized disposal into reuse, product recovery and waste management, (Krikke et al., 2003) and (Tibben-Lembke and Rogers, 2002) found more accurate the options reuse, product upgrade, material recovery and waste management. A more consensual way to classify disposition is through reuse, repair, remanufacturing, recycling and disposal (Thierry et al. (1995), De Brito and Dekker (2002), Fleischmann et al. (1997)).

2.1.1 Cost Reduction in Reverse Logistics

Cost reduction through an improved RL structure has been an underestimated subject in most of the companies (Stock and Mulki, 2009). In the same study, Stock claims that a good returns process can uplift profitability through cost reductions and higher product recovery. Through an extensive analysis of RL practices, large retailers have estimated savings of up to $6 million per $1 billion in retail sales (Jedd, 2000). On top of that, a good RL management not only reduced costs but also improved profitability (Rogers and Tibben-Lembke, 2001), as previously mentioned.

(Tibben-Lembke, 1998) highlighted the importance of including disposal and end-of-life costs in any total cost of ownership a company sets to calculate. Total cost of ownership being an approach of determining all costs associated with acquisition and subsequent use of an item or service.

On a more empirical approach, (Stuart et al., 2005) aimed to build an algorithm that allowed for an improvement in the returns process. The algorithm was focused on choosing the best option to dispose an item depending on inventory level, demand pattern, cost and lead time. Costs and lead time were reduced by more than 20%.

2.2 Big Data and Machine Learning

The growing interest in *Big Data* in the last years led to several authors trying to define what big data is. Tech America Foundation defines big data as "a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information” (TechAmerica Foundation’s Federal Big Data Commission, 2012).

Big Data Analytics (BDA) is becoming a standard for companies that thrive for competitiveness in the 21st century. A study showed that e-commerce firms who integrate BDA into their value chain experience 5% - 6% higher productivity than their competitors (McAfee et al., 2012). Another study states that BDA contributes to at least 10% of the growth for 56% of companies (Columbus, 2014). This translates in 91% of Fortune 1000 organizations allocating resources to BDA ventures, an increase of 85% from the previous year (Kiron et al., 2014).

One of the fields of computer science which plays a big role in BDA and derives from *Data Mining* is *Machine Learning* (ML), as depicted in Figure 2.3. ML is defined as computational methods using *experience* to improve performance or make accurate predictions (Mohri et al., 2012). The term *experience* refers to the data that is previously available to "learn" from. The data can come from digitized human-labeled training sets, or other types of information collected via interaction with the environment. There can be many types of applications associated with ML, from document classification, fraud detection to medical diagnosis, etc.

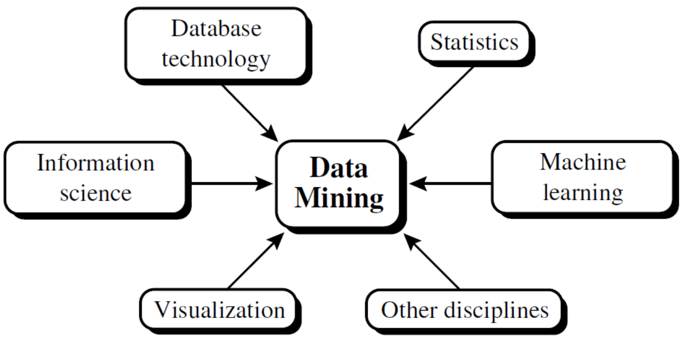


Figure 2.3: Origin of Machine Learning from Data Mining, in (Han et al., 2011).

The aforementioned examples of ML applications fit into a list of distinct ML algorithms that adapt to each individual situation. From those typical categories the following are detailed (Mohri et al., 2012).

Classification - To assign a category for each item through a binary output. E.g., image classification as a portrait, a landscape or other.

Regression - Predict a real value for each item. Applicable when predicting stock values or doing a sales forecast. In this case, the penalty for an incorrect prediction depends on the magnitude of the difference in contrast to a classification algorithm.

Clustering - Partition items into homogeneous groups. Usually associated with large data sets. E.g., for social network analysis, to identify groups of people with similar likes and behaviors within a vast group of people.

One of the core subjects of ML is the concept of *supervised* and *unsupervised* learning. Supervised learning is when the learner receives a set of labeled examples as training data and makes predictions for all unseen points. This the most recurrent scenario in classification, regression and ranking problems. Unsupervised learning occurs when the learner receives unlabeled training data and attempts to make predictions for all unseen points. This is more frequent in clustering and dimensionality reduction problems and it can be quite difficult to assess the performance of the learner (Mohri et al., 2012). Portrayed in Figure 2.4 is a summary of the described ML categories, among others.

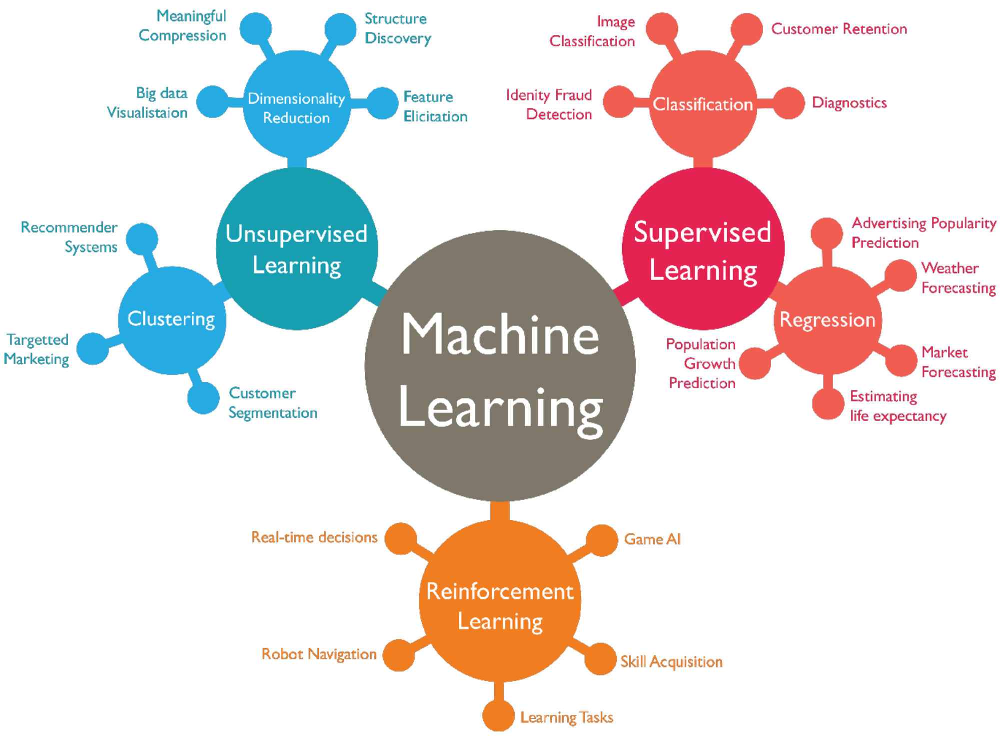


Figure 2.4: Machine Learning problems and common applications, in (Vibhor Agarwal, 2017).

A common process of model construction is to split the data into a *training* and a *testing* set, as depicted in Figure 2.5. In the same figure, the technique is also described. It involves processing data from past observations, training the model by understanding the relationship between the features selected and its labels (output). Lastly, perform a validation using the testing set and assessing the performance.

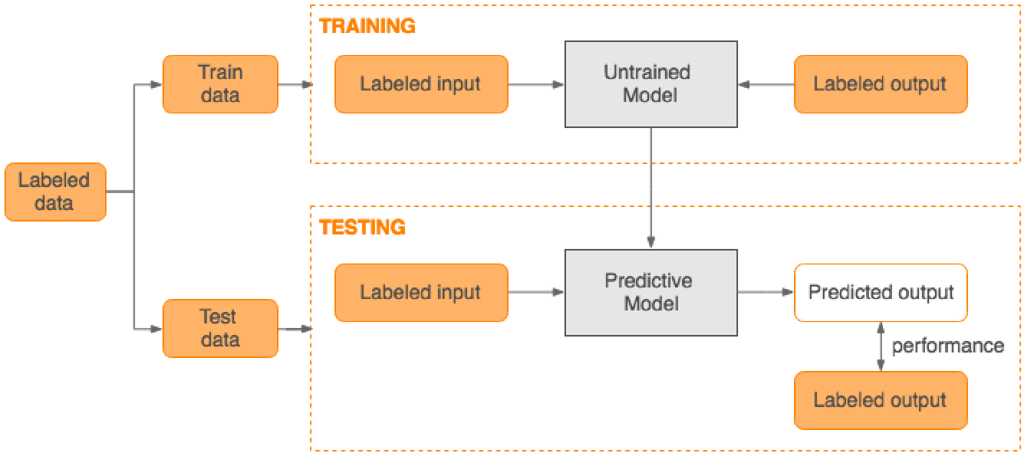


Figure 2.5: Common model construction in Machine Learning, in (EBC, 2017).

This process of validating a trained set with a testing a set is called *cross-validation* (Refaeilzadeh et al., 2009). According to the literature, there are two widely used cross-validation techniques (Reitermanova, 2010).

- Hold-out cross-validation: mostly used due to its efficiency and easiness. The data set is divided into three mutually disjoint subsets - training, validation and testing. The model is trained on the training test and has its performance measured using the validation test. When the performance is good enough or when it stops improving, the testing subset is used to obtain a fare estimate of the model's performance;

- K-fold cross-validation: this method uses a combination of more tests in order to gain a stable estimate of the model error. The data set is divided into k parts of the same size. One part is validation subset and the other parts are the training subset. The process is repeated for each part of the data and finally the mean for each iteration is performed;

ML is typically a key piece in a broader data mining project, which will be the focus of the second part of this dissertation. As such, it follows the life cycle of those ventures. This process is known as CRISP-DM which stands for *Cross Industry Standard for Data Mining* and is displayed in Figure 2.6. It contains all the relationships between the tasks and the development of the different phases (Chapman et al., 2000).

As it will be further explained, the second step of the dissertation consist of a classification problem. Therefore, the following section focuses on a more detailed review of classification problems and algorithms for a better understanding of the developed classifier.

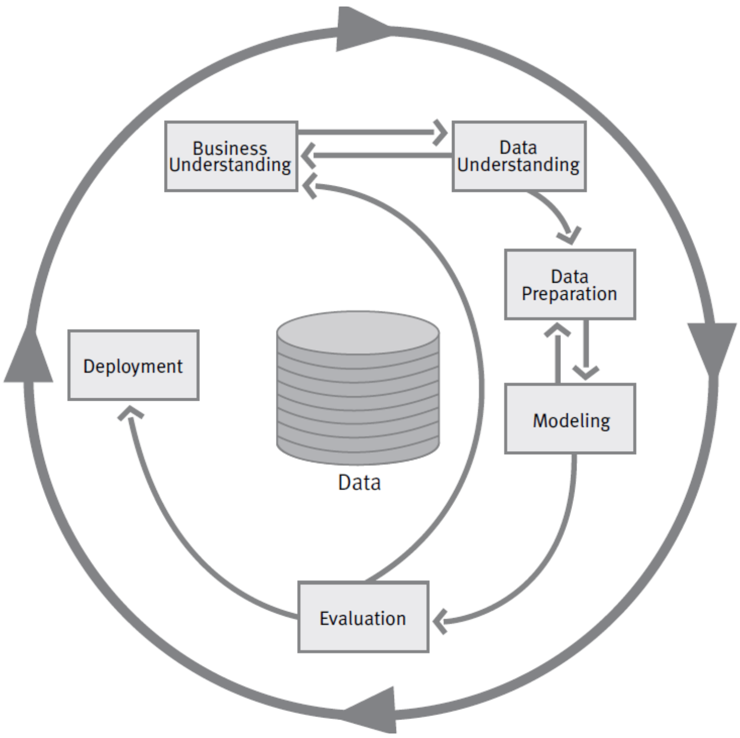


Figure 2.6: CRISP-DM approach in data mining projects, in (Chapman et al., 2000).

2.2.1 Classification Algorithms

A natural question that arises with a classification problem is what algorithms are best for that instance regarding effectiveness and accuracy. A study conducted by D.H. Wolpert opposed the two perspectives about having *a priori* distinctions between algorithms (Wolpert, 1996) but the *No Free Lunch* theorems states there are no distinctions (Wolpert and Macready, 1997).

However, some authors who opted to perform an empirical comparison of the performance of several learning algorithms have demonstrated that for some metrics such as accuracy, there are algorithms which outperform others (Caruana and Niculescu-Mizil, 2004). In that study, boosted trees and support vector machines (SVM) surpassed others in accuracy optimization due to their simplified learning process. *Random Forest*, *Support Vector Machine* and *Logistic Regression* are further detailed to promote a full grasp of the developed model.

Random Forest (RF)

Random Forest is classified as an ensemble learning technique and is a combination of decision trees predictors with a random vector sampled independently and with the same distribution for all trees in the forest, as exemplified in Figure 2.7. This way, the output of the algorithm is the mode of their predictions (in classification) or their mean (in regression) (Breiman, 2001). Besides being computationally demanding, one of the negative aspects of this learning technique is that it acts like a black-box, giving no interpretability to its outcome. On the other hand, it tends to overfit (over adapt to the training data leading to a wrong performance) less than individual decision trees.

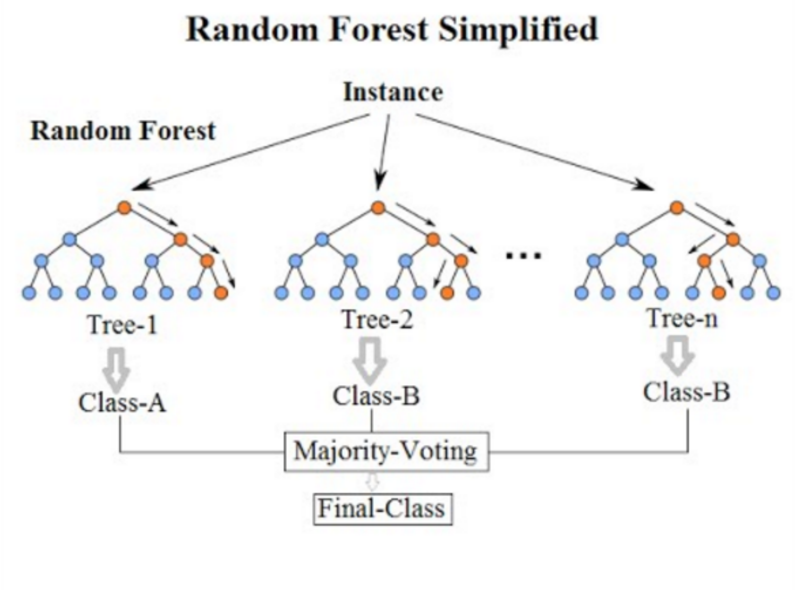


Figure 2.7: Example of Random Forest learning technique, in (William Koehrsen, 2017).

Hyperparameter tuning is when a data scientist searches for the best "settings" to optimize performance. While a model *parameters* are learned during the training, hyperparameters must be set *a priori* by the data scientist. According to (Tavish Srivastava, 2015), the hyperparameters responsible for making predictions more accurate are the following:

- Number of Trees - This is the number of trees chosen before taking the maximum voting or averages of the predictions. This is not a tuning parameter *per se* because more trees will improve accuracy, however, while more trees does not increase the risk of overfitting, it does reduce variance, demanding more computational power.
- Maximum number of features - This controls the maximum number of features an individual tree is allowed to collect. Generally, the performance of the model increases with more features being considered for a tree but the speed of training decreases.
- Minimum sample size of a leaf - A leaf is the end node of a decision tree and this controls the minimum amount of data samples to consider in that leaf. Usually, less samples tend to capture more noise in the train data.

Support Vector Machine (SVM)

Support Vector Machine is a supervised learning model with associated learning algorithms that break down data used for classification and regression. An SVM model constructs a feature space and maps data as points in that space, in a way that examples of separate categories are divided by a clear gap to outline the differences between categories (Hearst et al., 1998). As new data points are are fed to the model it gets mapped into that space and predicted to belong to a category according to which space it fell.

SVMs can perform linear and non-linear classification by the use of a *kernel trick* which implicitly projects the data into a high-dimensional feature space while performing the calculations in the input space (Hearst et al., 1998). This learning technique is widely used in pattern recognition and one of the major downsides is the computational demand of the algorithm (Cortes and Vapnik, 1995).

The hyperparameter possibly subjected to tuning in this dissertation is the type of *kernel* used. A *kernel* is the function used to transform the data and it can be linear, radial, polynomial, among others. The choice depends on the type of problem considered as one type of kernel can be a better fit in one situation and inadequate in others (Cortes and Vapnik, 1995).

Logistic Regression (LR)

A Logistic Regression is a multivariable analysis method used to predict a single categorical outcome based on multiple variables. This learning technique explores the relation between two, or more, independent variables with one dependent variable, the outcome variable. The predicted value assumes the sum of the multiplication of each independent variable value and coefficient (Park, 2013). Thus, ranking the impact of the independent variables on the output variable. The differences between a logistic model and a linear one are portrayed in Figure 2.8.

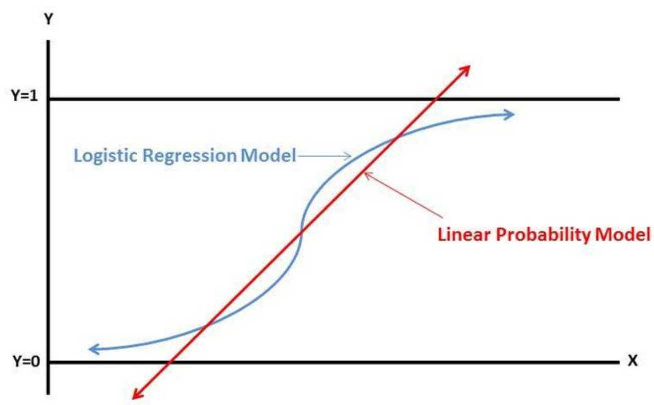


Figure 2.8: Differences between a Linear model and a Logistic Regression Model, in (Issa et al., 2017).

In a classification problem, the output of the LR can be interpreted as the probability of a given data point belonging to a certain class. One of the main advantages is the low computational power required to train the model and it is a widely used algorithm in clinical researches, e.g., to predict a diagnosis (Zhang et al., 2018). The hyperparameter that may be subject to tuning is the regularization term C which controls how big a penalty should be applied for increasing the magnitude of value coefficients in order to reduce overfitting.

2.2.2 Performance Metrics

After going through an overview of the chosen ML techniques in this dissertation, the performance metrics that decided which path to follow are detailed in this section.

Confusion Matrix

For a classification problem, such as the one presented further in Chapter 5, the concept of a *Confusion Matrix* introduces several metrics that can derive from it. In Table 2.1, an example of a confusion matrix is displayed, in the context of a binary classification problem (Powers, 2011).

Table 2.1: Confusion Matrix adapted to a classification problem.

		True Value		
		1	0	Total
Prediction	1	TP	FP	$\overline{P}$
	0	FN	TN	$\overline{N}$
	Total	P	N	

For every new data entry, there is an associated real value, either 0 or 1. *True Positives* (TP) are predictions which were correctly classified as 1, while *True Negatives* (TN) were correctly classified as 0. The two other categories correspond to the type I and type II errors in statistics. A *False Positive* (FP), or a type I error, occurs when a data point is classified with 1 but it is actually a 0. While *False Negatives* (FN), or type II errors, were wrong predictions of the class 0.

These concepts introduce the metric *Accuracy*. Accuracy measures the total number of correct predictions over the total number of observations. While this is a major indicator of a model's performance, it can be misleading (Provost et al., 1997). If a data set contains 70% of a class 0, a model trained with that data and having an accuracy of 70% could mean that it is always predicting 0 and not learning at all the relationships between the variables. For that reason, other metrics were considered to evaluate the classifier.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Positive Predictive Value (known as } Precision\text{)} = \frac{TP}{TP+FP}$$

$$\text{True Positive Rate (known as } Recall\text{)} = \frac{TP}{TP+FN}$$

$$\text{F1-Score} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Precision is an estimator of the classifier's ability to not label an entry as positive if it is negative. For cases where finding the true values is more important than general accuracy this indicator is a good estimator of the model's performance. *Recall* is the ability of the classifier of finding all the positive samples. It can be interpreted as how much is the algorithm "taking a chance" in predicting a data point as 1, which is also important in assessing a model's performance with the same objective of highlighting the class 1. The *F1-Score* is the weighted harmonic mean of precision and recall. It varies between 0 and 1 and returns an overview of the predictor, by favoring strong classifiers both in precision and recall. These metrics, among others, are common in assessing a classifier's performance (Powers, 2011).

Receiver Operating Characteristic (ROC)

The *Receiver Operating Characteristic* curve plots the true positive rate against the false positive rate. It is a way of visualizing trade-offs between benefits (true positives) and costs (false positives) (Fawcett, 2006). Since 1989, ROC curves have been used to evaluate and compare algorithms as it became clear its usefulness with skewed class distributions and unequal classification error costs. In Figure 2.9 an example of an ROC curve for two classification models is depicted. It is not uncommon to represent a diagonal line which indicates the performance of a random classification model.

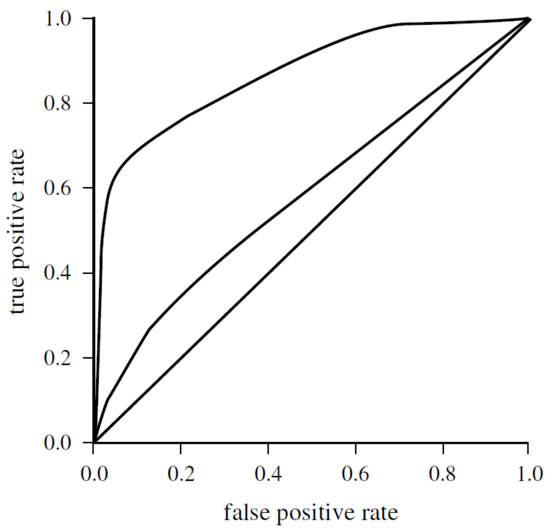


Figure 2.9: ROC curve of two classification models, in (Han et al., 2011).

The point (0,1) would correspond to a perfect classifier, which means it will never wrongly classify as positive a sample if it was negative. On the other hand, the point (1,0) is the classifier with the worst performance, for the inverse reason of the best classifier. It would never classify correctly a positive data point. A common indicator used as a general measure of a classifier performance is the *Area Under the Curve* (AUC-ROC). The AUC is shown to be, in a medical context, the probability that a random pair of normal and abnormal images will be correctly ranked as to their disease state (Hanley and McNeil, 1982). This means that the AUC highlights the probability of the classifier giving a higher score to a randomly chosen observation of class 1.

Chapter 3

Current Status

3.1 Global context for Farfetch

In this chapter, an extensive overview of this dissertation's case-study is presented. The current ordering and return processes at Farfetch are dissected, the return policy and cost/revenue structure detailed as well as the proposed To-Be solution.

In order to gain perspective into Farfetch's top markets, the first step is to understand the geographical distribution of both supply and demand. As Europe still represents the core of luxury fashion designers and brands it is relatively clear that most of the supply comes from countries such as Italy and the United Kingdom, as seen in Figure 3.1.

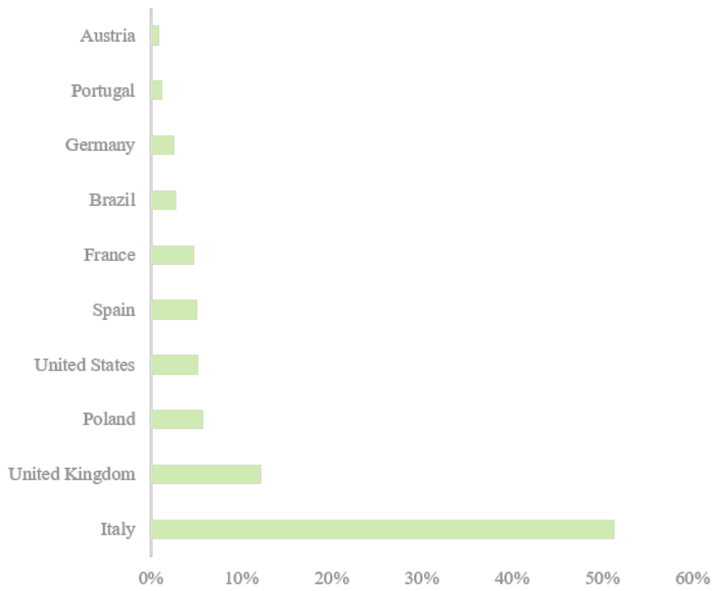


Figure 3.1: Farfetch supply in % of orders as of 2017.

On the other hand, the demand has been led by one country, the United States. Since 2008 that the North American market has been the most important pillar in the Gross Merchandise Value (GMV) of Farfetch.

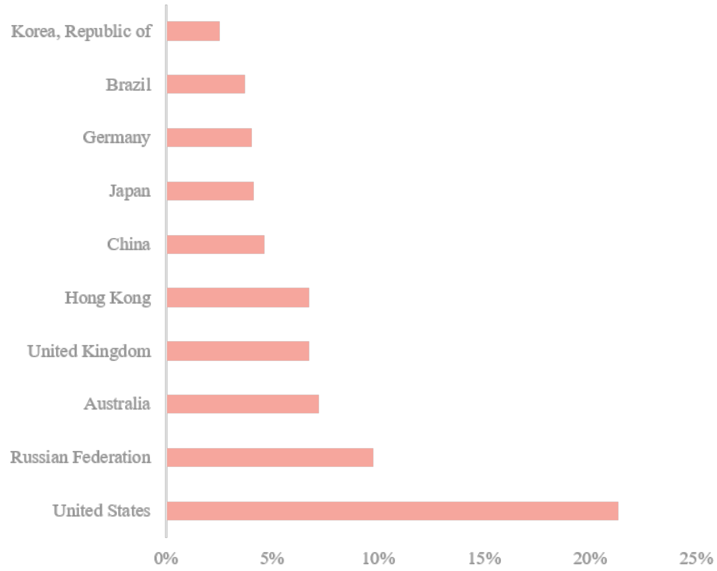


Figure 3.2: Farfetch demand in % of orders as of 2017.

One of the most interesting market changes in the last years, has been the near exponential growth of the Asian Pacific region. As seen in Figure 3.2, Australia, China, Hong Kong, Japan and the Republic of Korea are included in the top 10 markets podium. This translates into opportunities in these countries and, as it will be explained afterwards, China is one of the selected countries for which the impacts are to be further explored in the proposed To-Be solution.

3.1.1 Order processing

Before reviewing the return process, the ordering process should be detailed to give a more fulfilling overview of the company. The ordering process was one of the baselines that José Neves, the CEO and founder of Farfetch, first devised and which the company revolves around. When an order is placed, also called a Portal Order, it is divided into one or more Boutique Orders, as explained in Figure 3.3.

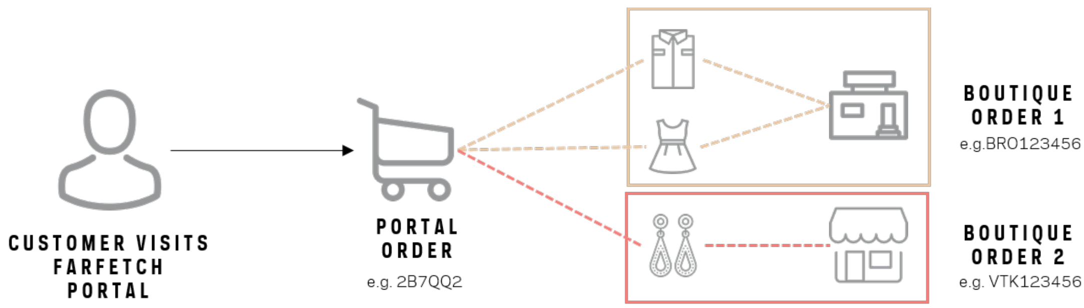


Figure 3.3: Division of a Portal Order into two Boutique Orders.

Orders must then go through a series of six steps:

- Step 1. Boutique checks for stock;
- Step 2. Payment approval;
- Step 3. Packaging decision made by Farfetch partner;
- Step 4. Create shipping label or an Air waybill (AWB);
- Step 5. Send parcel;
- Step 6. Parcel in transit, until delivered.

Almost every team in Farfetch is involved, from Payments to Customer Service, Delivery Support etc, as detailed below in Figure 3.4.

Before an item is placed online for sale a Production process occurs in which boutiques send packages (called slots) up to 50 items they wish to sell in Farfetch's platform. This process is represented in Figure 3.4 along with the entire ordering process.

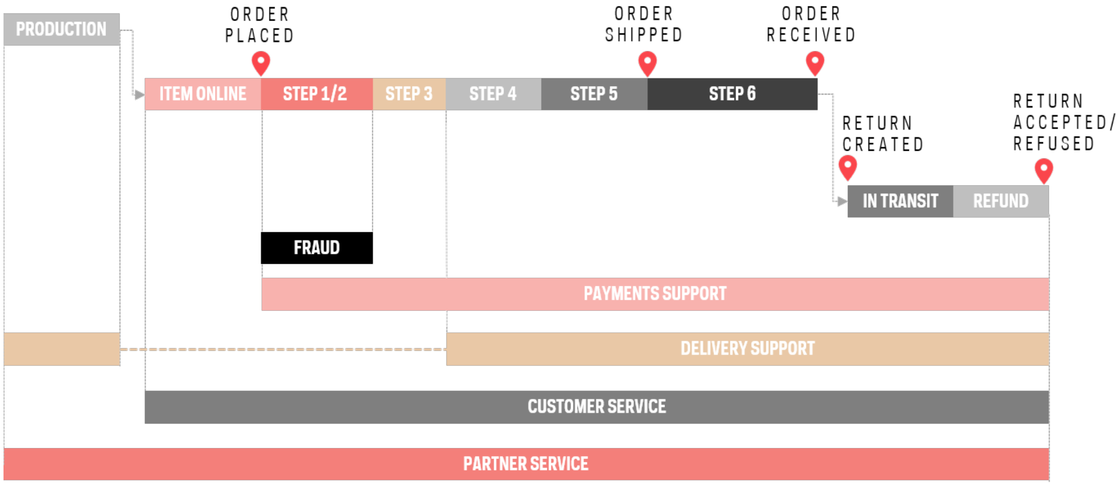


Figure 3.4: Main processes and teams involved.

The delivery department is divided into two distinct teams: Delivery Support and Delivery Development. While the Support team handles day-to-day issues like shipping problems, AWB creation problems or dealing with unexpected situations with custom authorities, the Delivery Development team is responsible for creating and analyzing solutions that can enhance the shipping experience while reducing costs and manual labour, thus not appearing in Figure 3.4.

3.2 Returns at Farfetch

Following the previous overview, as the scope of this dissertation is indeed the return process, in this section, returns at Farfetch are dissected and an accurate mapping of the return process is outlined. All the analysis was performed based on the year of 2017, as this provided clean, complete data and, on top of that, it was a year of natural growth and with no unexpected issues.

Similar to what happens with the ordering process, the return process is a series of stages that involve as many teams as the ones involved in ordering a product. It can be triggered by many factors that can be categorized in a Customer post-delivery Request or a Return to Origin.

3.2.1 Customer post-delivery Request

Returns created by the user after delivery are the ones that have the most impact on the company. These types of returns are characterized by the desire of the final customer to return the item after it has been delivered to his/her address. This can happen for several reason being sizing and fitting the most common, as portrayed in Figure 3.5.

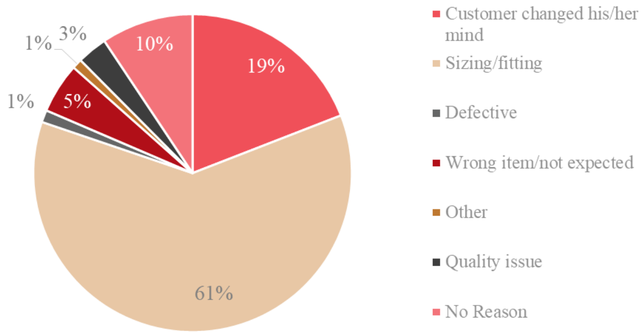


Figure 3.5: Distribution of return reasons.

This is easily understandable since most products at Farfetch are wearable and, as such, were not tried on prior to being purchased. These returns that are created through a customer request after the delivery are the ones this dissertation focuses on and where opportunities for cost improvement are emphasized.

3.2.2 Return to Origin - RTO

A "Return to Origin" or a RTO is, by definition, a cancellation of the shipping of an order before the item reaches the customer. For a RTO to happen, it must have a post-order trigger that flags the order as heading to the wrong customer or with the wrong product. There can be several reasons for that.

- RTO by Fraud - This situation happens when the Fraud team has already approved the payment for that order but it is indeed a fraudulent payment and it must return to the boutique.

- Customer Request - This RTO is triggered by a desire manifested by the client to Customer Services to cancel the order. The Delivery Support team then takes care of tracking the order and cancelling it.

- RTO by Boutique - Sometimes the boutiques, due to a wrongful identification of the item and consequent realization only after the order has been shipped, contact Partner Services who then reach Delivery Support team to request a RTO.

Regardless of the reason why there is a RTO, sometimes the lag between the trigger and parcel identification is too long leading to the order being wrongfully delivered. Despite this type of returns being a demanding problem for Farfetch when it comes to resources and operational excellence they are a minor ocurrence when compared to Customer post delivery Request.

3.2.3 Return Policy

As many e-commerce companies, Farfetch offers its customers returns free of charge and without a mandatory request for motive. This flexible policy is shown to incentive purchases rather than increasing returns (Janakiraman et al., 2016). The order includes the return AWB alongside billing documentation and so, the customer passes through a very easy process to return an item. The following steps should be respected in order to successfully create a return:

1. Access 'My Orders & Returns' in Farfetch website account and select the order;

2. Choose return 'By courier' and select the item(s) to be returned;

3. Confirm the pick up address and a convenient collection time;

4. Prepare the package for return, meaning that a signed return invoice and AWB are attached to the exterior of the box, which should be left open in order for the courier to check its contents.

If there is a boutique near the client region, Farfetch also offers the option to return in store. This return policy states that a customer should create a return within **14** days of receiving the item. As for the state in which the item must be, Farfetch has a set of conditions, e.g., an item must be returned undamaged, unused and unworn, with all the tags and original packaging; beauty and cosmetic products must be returned unopened and unused, with the seals of any packaging still intact. These guidelines are of an extreme importance because if an item that does not comply with these standards arrives at the boutique there is a possibility that it might contest the return. Due to partnership quality between Farfetch and the boutiques, a negotiation takes place by Partner Services in which there is an attempt to split the cost of the item if the damage does not affect its sale entirely. When the item is stained and there is an attempt to recover the initial condition, this comes at Farfetch's expenses. In the unfortunate event that the damage fully compromises the sale, Farfetch takes responsibility for it and assumes the entire cost of the item.

As for Farfetch's compromise to the boutiques, there is an agreement that, in the case of a return, the item should not arrive to the boutique later than **28** days upon the item arrival's at the customer address. After this period, a late return can be contested by the boutiques legitimately.

3.2.4 Process Mapping

Considering all the detailed aspects of the return process in the previous sections it is now important to summarize and map the entire process. As it happens with the ordering process, the return process is sequential and follows the steps below in Figure 3.6.

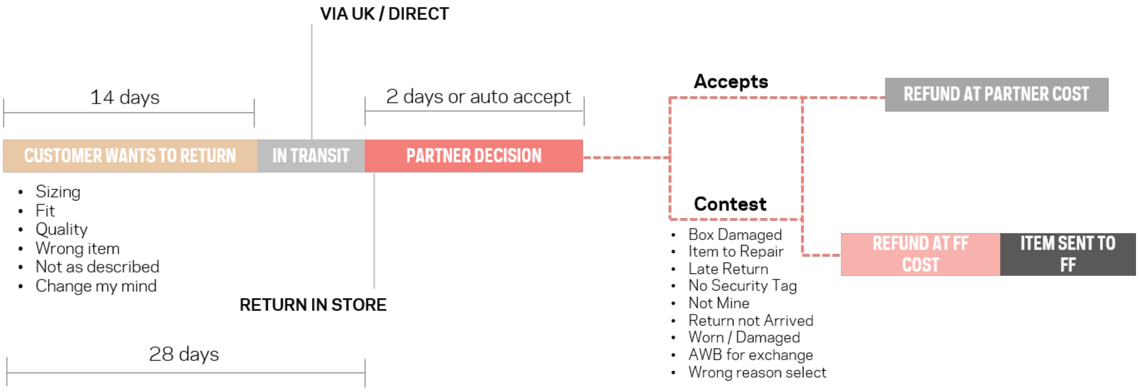


Figure 3.6: Map of the Return process.

<u>Customer wants to return</u> - After the client receives the item and plans to return it, he/she creates the return and ensures it complies with Farfetch return policy, during a period of no more than 14 days after receiving it.

<u>In transit</u> - The return is shipped to the boutique via three ways. Direct returns, as the name implies, go directly from the customer to the boutique through a courier. Indirect returns have two legs of shipping, meaning they go from the customer country to the United Kingdom and then finally to the boutique. This scenario will be explored further in Section 3.5. Finally, In Store returns are specific to some regions and for some boutiques, where the client returns the item directly to the store.

<u>Partner Decision</u> - The boutique now proceeds with the assessment of the item, including the verification of all security tags still attached, and may choose to accept the return or contest it. The partner has a maximum of 2 days to state its final decision with the consequence of a return auto-accept if the due date is violated.

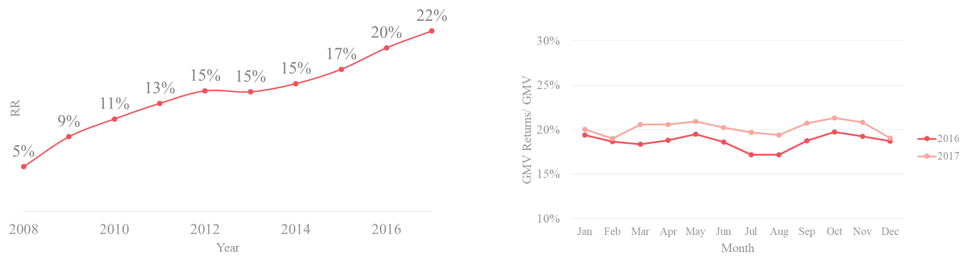
There can be two possible outcomes from the partner decision.

<u>Return Accepted</u> - If everything is verified, the boutique accepts the return or, due to lack of an answer, the return is automatically accepted.

<u>Return Contested</u> - In case the item is irreparably damaged, the return is contested and, following a negotiation between Partner Service and the boutiques, if the item is indeed damaged, it is shipped to Farfetch's office in Lionesa - Porto, Portugal.

3.3 Returns in-depth review

Farfetch had a RR of **22%** meaning that for 100 boutique orders, approximately 22 boutique orders had, at least, one product being returned to the store. As the business suffers its natural growth and starts penetrating more diverse markets, the RR also comes impacted and this translates into the climbing slope of Figure 3.8a.



(a) (b)

Figure 3.7: RR growth and GMV.

From Figure 3.8b we can deduct that the value of the returned products is aligned with the global GMV for each year. This means that the price tag of each returned product is neither more expensive nor cheaper than average product value.

Regarding the RR by country, an interesting perspective is to calculate the domestic RR, meaning for all inbound orders of a country, which percentage resulted in a return. This analysis is performed for eight countries of the top markets according to Figure 3.2 and summarized in Table 3.1.

Table 3.1: Domestic RR for top 8 countries.

Country	**Return rate**
Germany	29,9%
United Kingdom	22,2%
United States	21,1%
Hong Kong	16,2%
China	14,3%
Australia	13,6%
Russian Federation	13,4%
Japan	10,7%

According to a study performed in Europe about RR from Ecommerce Europe (2016), northern and western European countries tend to have higher RR than southern and eastern European fellow countries. This is explained due to the fact countries like Germany and Denmark have wealthier buyers who are more prone to do impulse shopping, while the other countries tend to behave more carefully in proceeding to the shopping cart.

From the company's point of view, this may indicate that developed e-commerce markets like Germany and United Kingdom have, indeed, wealthier buyers, therefore higher domestic RR. This can possibly be a good indicator of the stability and maturity of each market.

3.3.1 Redirecting Returns: To-Be model

In order to fully understand how redirecting returns is going to impact the current return process, a more simplified and graphic timeline of a return and its next sale is mapped in Figure 3.8.

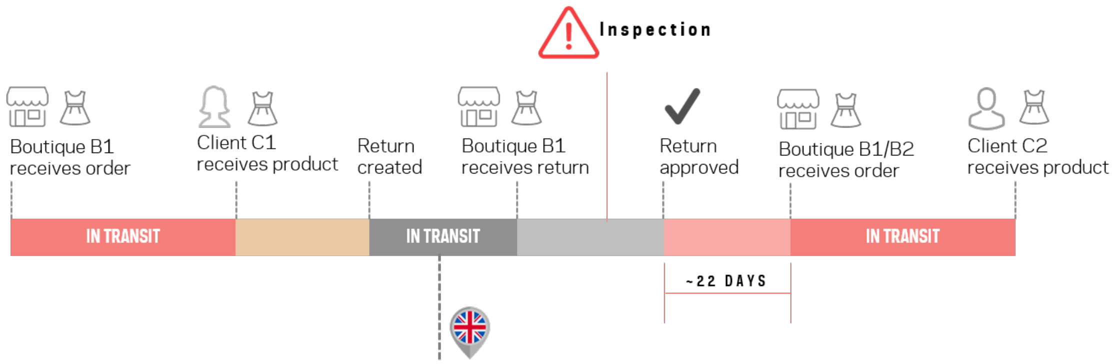


Figure 3.8: As-Is model for a Return and its next sale.

The As-Is model is explained in the following terms. There is an order for boutique B1 of a product which then is shipped to client C1. Client C1 wants to return that item and does so according to the guidelines established before. As such, it goes back to boutique B1 which then accepts it and inserts it back into the supply chain. After an average of 22 days from the supply chain insertion date, there is a second order for either boutique B1 or B2 from a client C2 for the same product. And the order fulfillment process starts.

The proposed To-Be model starts exactly when the return is created and aims to improve costs and business KPI's by performing changes in the next steps, following the return creation date.

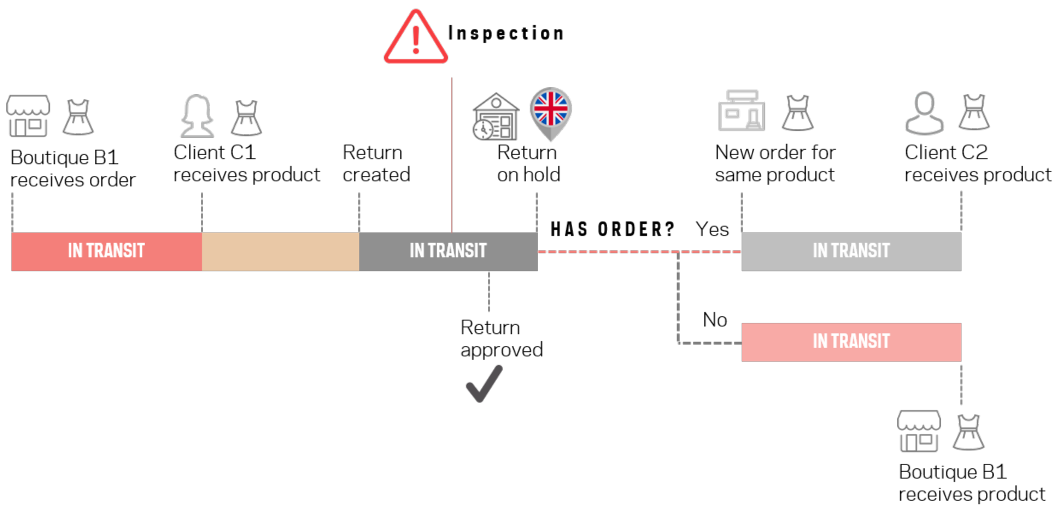


Figure 3.9: To-Be model for a return and its next sale.

The proposed model starts with the assumption that a returned product has a sale after its return. This is a pattern called Sale-Return-Sale (SRS). In order to study the life of a returned product, the following analysis is performed.

For every product that was returned in 2017, there was a **77,1%** chance of it having at least one sale afterwards. This, of course, is over a period of more than 100 days from its original return. The cumulative distribution is as follows in Figure 3.10.

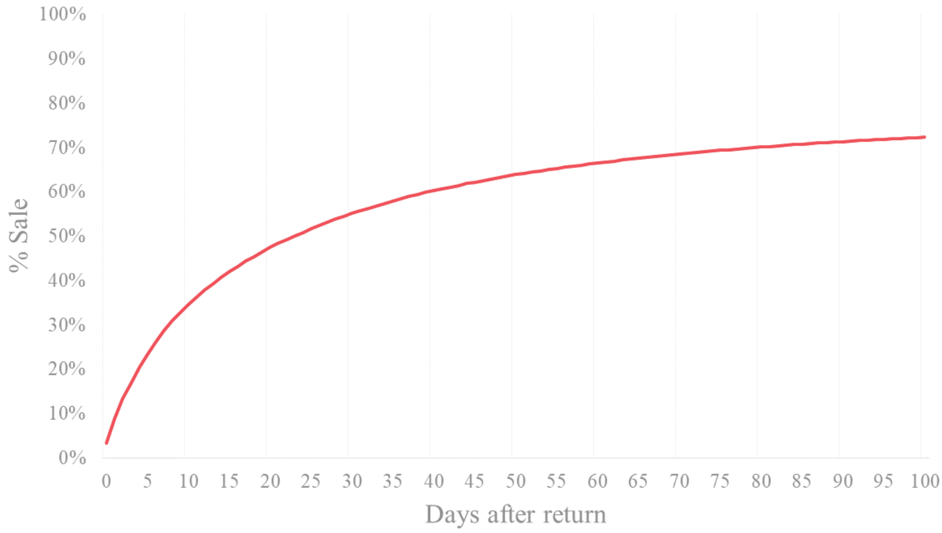


Figure 3.10: Cumulative distribution for the next sale of a returned product.

If a geographical context is added to the analysis, it sheds light to the fact that almost **33%** of next sales were going to the same region of the original return and more than **13%** to the same country.

In Table 3.1, the domestic RR was analyzed. An interesting perspective is to frame the domestic RR together with order volume and percentage of next sale for each country. Figure 3.11 pinpoints exactly that scenario.

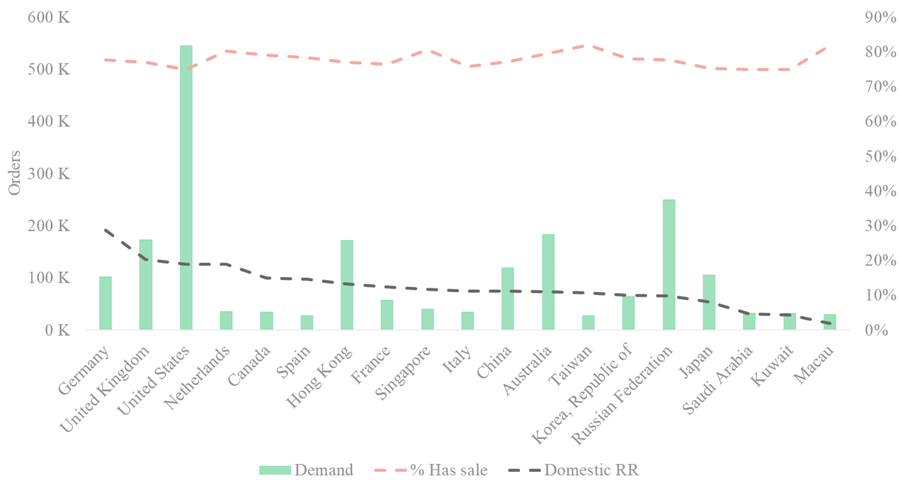


Figure 3.11: SRS percentage, order volume and domestic RR by country.

3.4 Cost/Revenue structure

3.4.1 Costs

Once the return process was analyzed, to understand the expected savings context, an overview of the cost structure for delivery is presented. Essentially, delivery and return costs can be unfolded into four types: shipping charges, duties, taxes and extra charges.

Shipping Charges - the biggest chunk of costs are shipping charges which represent the cost of moving a product from one place to another. It varies with volume/weight, it depends on the origin/destination and service type (Express or Standard).

Duties - tax on incoming products as stated by the destination country customs authority. This cost depends strongly on the type of product, its value and of the country of entrance. The European Union currently is a free trade zone, meaning no Duties neither Taxes are applied to products moving across borders inside the European Union, as it happens with domestic orders. On the other hand, when a product comes from the United States, there are rates applied which depend on the Harmonized System of product classification, as predicted by the World Customs Organization (WCO).

Tax - on top of all duties customs can charge over a product, it can also be eligible for the Value Added Tax (VAT) of the country, depending mainly on the value of the product.

Extra Charges - sometimes when a courier has to perform an action that escapes the regular span of events such as delivering in a remote area, fuel surcharges or heavier packages than usual, it charges Farfetch with additional charges.

A distribution of the costs is represented in Figure 3.12. The scope of this dissertation will be focused on Shipping Charges, Duties and Taxes since these costs represented almost **93%** from a total of £ 98 M in 2017.

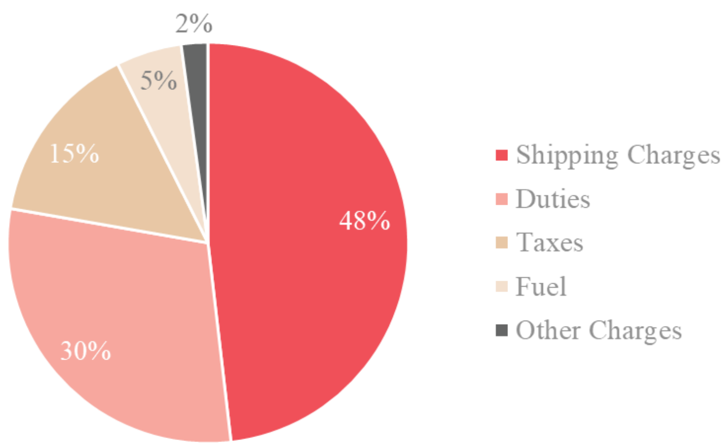


Figure 3.12: Average charge type in 2017.

3.4.2 Revenues

To sustain its business, Farfetch operates in a commission type of revenue, where for each boutique order Farfetch takes a percentage of the value of the transaction. Regarding shipping, the Supply team developed a program called "Service 3.0". This involves a Service Level Agreement (SLA) between the boutiques and Farfetch that, if the boutiques exceed the threshold values of No Stock (when a boutique lately notifies Farfetch that does not possess an item a client has already selected and paid) and Speed of Sending (speed, in days, for a boutique to prepare the item for shipping), it has to pay 1.5% of the Actual Transacted Value (ATV) of that boutique.

ATV is a common KPI in e-commerce and it is calculated by subtracting cancellations, "No Stock" occurrences and refused payments from total sales. This metric helps Farfetch finance the aforementioned return costs.

3.5 Shipping Services

To better understand the reverse flow of items, it is now essential to obtain insights about the outbound of items and how they are shipped. Farfetch has a long-term relationship built with DHL and, as a consequence, it is the mainly used courier. It represents 80% of the shipping volume and, together with UPS, comprises 97% of shipments. The remaining percentage belongs to small, more local carriers, that are used due to extraordinary needs like VIP services or when special domestic services are required.

For the same carrier, there can be different price ranges, depending on the speed and type of transport used. The types of service most used can be summarized into 6 types:

- Standard: delivery within 5-7 days;
- Express for most of Europe and USA: delivery within 2-4 days;
- Express for Rest of the World: delivery within 3-7 days;
- Next Day: delivery in the day after the purchase ;
- Same Day: delivery in the same day of the purchase;
- 90 MD: delivery in 90 minutes after order placement.

The most common service is Express for Europe and USA since supply and demand are concentrated here, respectively. The use of Standard is reserved for routes with terrestrial access.

3.6 Returns Via UK

Duties & Taxes (D&T) account for 45% of total shipping costs and occur when a parcel crosses a border and enters a country which customs law predicts such charges. In order to try to minimize those types of charges in returns, Farfetch has studied and implemented a solution where returns have a transshipment point in the United Kingdom. This results in return shipments with two legs, one from the customer country to the UK and another one from the UK to the boutique country.

Figure 3.13: Example of an order with a Return Via UK.

As the European Union represents a customs union, products are freely traded among the constituting countries without Duties being charged. However, the same does not apply to VAT, that may be charged. On top of that, returning items are often seen as imported goods, and, according to that status, D&T may be charged.

The UK has a different perspective on returning items, identifying them as "exported goods returning to origin", therefore facilitating the entrance of Returns into the European Union. In addition to the fact that Farfetch's main account for DHL is in the UK and the relatively frictionless way to solve problems in UK customs led to conclude this would be a cost-effective way of making returns enter the European Union.

In order for this to happen, Farfetch has a partner in the UK - Norsk - who does the cross-docking. It interrupts the normal path of the return and dispatches it to the boutique. This process involves two AWBs, therefore two legs of shipping. The first AWB is relative to the shipment from the customer country to the UK (the printed paper that is inside the packaging) and the second AWB is generated by Norsk and takes the return to its final destination, the boutique.

This two leg return shipment, also called "Returns Via UK", is only used for certain countries and not for all boutiques. The countries are the following: Belgium, Bulgaria, Cyprus, Finland, France, Greece, Italy, Latvia and Spain. In 2017, more than 40% of returns were processed as "Returns Via UK".

Chapter 4

Impacts Study

After an initial overview of Farfetch's main markets, having the return process mapped and dissected and having pinpointed the opportunities for the project, this chapter discusses scenario design and exploration with the goal of finding the best conditions and candidates, i.e., countries, to implement the proposed solution. This is achieved through simulating the different scenarios using historical data.

First, the method of comparison is explained, meaning how the assessment of cost savings was performed together with a short description of the window of comparison between the As-Is and To-Be models. The following section lists all the conditions and assumptions made throughout the impacts study of the To-Be model. Finally, the remaining sections describe the entire process of filtering down the best scenarios and, at last, the results and takeaways.

4.1 Method of Comparison

4.1.1 Cost Assessment

In order to have a consistent and accurate way of comparing both models regarding return cost savings, after a study of the cost variables that would make sense to include in the model, the following types of costs and lead time sales, as seen in Figure 4.1, are demonstrated to completely showcase the cost assessment.

Shipping Costs - Probably the first and immediate driver behind this entire study, possible savings in terms of shipping costs are mandatory to consider. The possibility of saving "legs" of shipping is a factor to acknowledge as avoiding the journey back to the boutique is the main purpose of the model.

Duties and Taxes - As described in Section 3.4.1, D&T represent 45% of shipping costs for Farfetch, so they also must be considered. Although, for returns and for most boutiques, D&T are charged to the boutique, in the sale afterwards, D&T can be charged again if the item travels back to the same country, e.g., from Italy to USA, then back to Italy and then finally back to USA (next sale). Thus showcasing a cost saving potential.

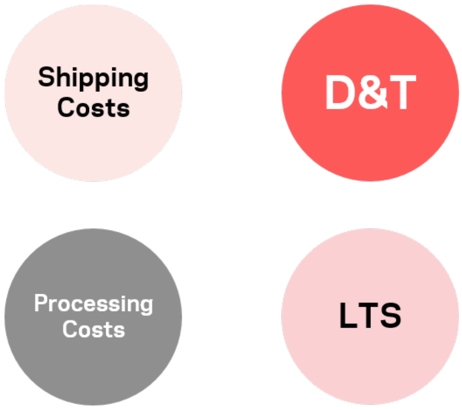


Figure 4.1: Types of costs and lead time sales considered in the assessment.

Processing Costs - For an item to be inspected and kept at Farfetch's responsibility for the days following its return, a 3PL must be used to fulfill those needs. For that, costs must be considered.

Lead Time Sales - Lead Time Sales, or LTS, represent extra sales that redirecting returns will create. Since products will be inserted sooner into the supply chain, it is expected to have more stock available for more time. This may lead to an increase in sales. Equation 4.1 translates the shortened lead time into expected sales.

$$E(LTS) = \delta P_{res} \times APV \times N_{Returns} \tag{4.1}$$

, where APV is the Average Product Value, $N_{Returns}$ is the number of returns in the sample and δP_{res} is the variation of the probability of a returned item having a sale afterwards, therefore Resale Probability. the latter is calculated by computing, for each day after product reinsertion in the supply chain, the number of products that had a next sale, divided by the number of total products reinserted in that day. Thus, by evaluating some days earlier, it becomes possible to assess the change in the probability.

4.1.2 Comparison Overview

To better interpret the impacts predicted in the present chapter, understanding where exactly the As-Is and To-Be model differ reveals to be of vital importance. Having said that, Figure 4.2 shows a graphic explanation of where they differ.

Figure 4.2: Differences between the two models.

Every time there is an order placement, the path to the first client is the same in both models. Only when the return is created, the two models start to diverge. On one hand, the return is redirected to a 3PL and then awaits for a second sale or a journey back to the boutique, where, even then, might have a second sale. On the other hand, the established path for returns, after its creation, it goes back to the boutique, either Via UK or directly. E.g., for a product that was returned in the USA, the As-Is model computes the costs of shipping the item to the UK, followed by the journey back to the boutique. If this item in particular had a next sale outside the EU, the costs of shipping it to the new client would be summed. For the To-Be model, the costs involved shipping to the 3PL facilities and then sending it to the new client, alongside with the processing costs concerned. Finally, the costs from both models are summed for all returns and then compared. The analysis was made with data from 2017, as previously mentioned.

4.2 Conditions and Assumptions

To perform the simulations and obtain results as close to reality as possible, some assumptions had to be made. The following list of conditions allowed for a simplification of the simulations making them possible to execute in the project timeline while guaranteeing accurate projections.

4.2.1 Legal Context

To redirect returns from a 3PL to a new client, some documents such as invoices and AWBs have to follow different legal standards than they do now. For a normal return, attached to the packaging, comes the return AWB with the boutique address. Then, if there is a second sale for that item, a new AWB and invoice are printed in the boutique for the new customer.

To redirect returns, changes to current standard would be needed. The return AWB would have to be directed to the 3PL "warehouse" address. Then, one out of three scenarios would have to happen.

1. There is a subsequent sale and it is in the boutique country or region (e.g., EU). In this case the invoice of the order and the second AWB will have to be printed in the boutique country or region;

2. There is a sale afterwards and it is in the Rest of the World or a domestic order (for the returning country). Here, the invoice and second AWB may be printed in the 3PL facilities;

3. There is no next sale and the second leg AWB is printed in the returning country with the boutique as addressee.

It is assumed throughout this dissertation that it is legally possible to follow the steps described above to successfully redirect returns.

4.2.2 Operational Conditions

Obtaining accurate results from the simulations implies that processing costs are well estimated. For that purpose, after discussions with other team members in the Delivery Development team, it is assumed that the price of **£ 3** per article, independently of how many days the 3PL holds the item, is a reasonable estimate of those costs.

Another important detail is the decision of how many days should the 3PL hold the item, waiting for a next sale. After observation of Figure 3.10, a slight decrease of the slope is noticeable in the region of $x = 15$ days. In that range, there is a 35%-40% chance a returned item has a subsequent sale, and, for that reason, **15** days was the standard selected to perform the simulations.

The last operational assumption will have to be that the algorithm behind Farfetch's attribution of orders would assign the subsequent sale to this solution and not to a boutique, as the current process predicts.

4.2.3 Boutique Agreement

The last assumptions made to simplify the analysis regard the fact that holding items that do not belong to Farfetch have an impact in a boutique's physical sales, stock and late returns. For that reason, the following two conditions had to be considered:

1. Boutiques accept late returns with 35 days (14+3+15+3). This is the most extreme case, where an item waits for 15 days with the 3PL and then takes an average of 3 days to travel into the warehouse and back to the boutique. Although the algorithm will minimize this precise case, boutiques will sometimes have to accept longer than usual late returns. However, they will have the upsides of less returns to handle and a higher probability of selling them.

2. Boutiques also accept that the quality check for an item must be done by a third party. This may be a difficult assumption to be made as some brands are very strict in their quality guidelines. Nevertheless, the upsides mentioned above will try to act as countermeasures.

4.3 Simulations Overview

To obtain significant results when assessing the different scenarios, three metrics were developed.

Score - This metric is a ratio between the costs of the As-Is model and the To-Be model. It was designed to immediately signal and flag if the proposed model will do better than the current solution for returns. E.g., if the score of scenario 1 for the USA was 1.2 it means that costs with the proposed model are 20% cheaper. The expression is detailed in Equation 4.2.

$$Score = \frac{S_1 + DT_1}{S_2 + DT_2 + P_2 - LTS_2} \tag{4.2}$$

S represents the Shipping Costs, DT represents the costs associated with Duties & Taxes, the Processing Costs are represented by P and, finally, Lead Time Sales by LTS. The subscript 1 refers to the As-Is model and 2 to the To-Be model.

Savings - As the name implies, this metric is simply the difference between the two costs. The expression is as follows.

$$Savings = (S_1 + DT_1) - (S_2 + DT_2 + P_2 - LTS_2) \tag{4.3}$$

Savings per Return - In order to shed light over countries that have a low volume of returns, savings divided by the amount of returns from that country is an attempt to show precisely that, the potential each one might have and to have a global benchmark for all scenarios. Equation 4.4 translates this metric.

$$Savings\ per\ Return = \frac{(S_1 + DT_1) - (S_2 + DT_2 + P_2 - LTS_2)}{N_{ret}} \tag{4.4}$$

As previously explained, this dissertation is a two-step approach. The algorithm developed in Chapter 5 will have direct impact in the savings predicted in the impacts study. Thus, to have a lower limit and a upper limit of possible savings two outlines were considered:

- Best Case. The algorithm is always correct, predicts with an accuracy of 100% if a return will have a next sale in the following 15 days. This leads to only consider in the simulations the sample which **only** includes the selling returns;
- Worst Case. Without an algorithm, the solution would be to consider **all** the returns. This will include the ones that had a next sale and the ones which did not.

Having detailed how each scenario is assessed, the remaining of the chapter is focused on explaining the actual process of simulating the savings and filtering out candidates, concluding with the results and final takeaways.

4.4 Scenario Design

4.4.1 First Stage Filtering

The first stage of assessing the impacts of redirecting returns was to select important countries that showed promising potential in being the ones whose returns gave origin to more subsequent sales, therefore, to more savings. On a preliminary analysis the countries in the upper band of Figure 4.3 were considered.

This selection was merely a starting point in order to start considering scenarios. Those were chosen mainly because either they are important markets of Farfetch, have high RR or are big suppliers, in this case, Italy, or even the three cases, e.g., the United Kingdom.

Quickly it became clear that only three countries were worth investigating: the United States, China and the United Kingdom. The reasons behind this selection were mainly connected to Farfetch's company guidelines and business strategy. The United States is, by far, the strongest market for luxury fashion, at least for Farfetch. Having that in mind, it is perfectly reasonable to lead this country into further analysis.

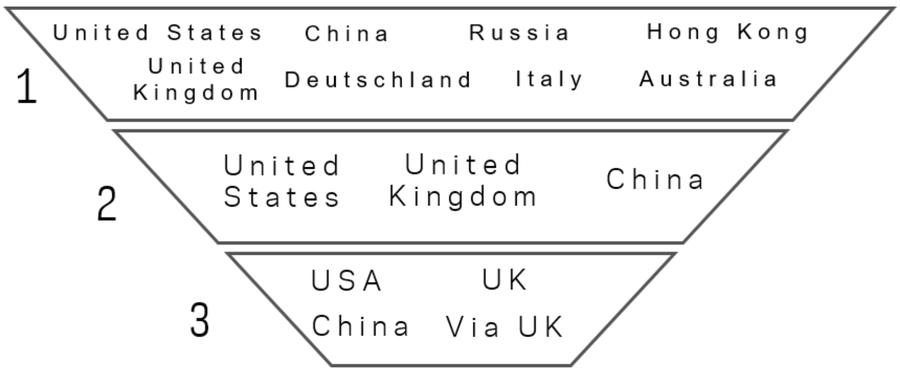


Figure 4.3: First stages of candidates for simulations.

China is becoming a very important bet for Farfetch (TechCrunch, 2017). The luxury fashion market in Asia Pacific region is growing insanely fast, as in 2018, more than half of the apparel and footwear sales will originate outside of Europe and North America for the first time (Business of Fashion & Mckinsey, 2018). Moreover, Farfetch is currently planning to grow its influence in China, therefore making this country one to consider.

The United Kingdom has long been a "partner" in Farfetch's *modus operandi* since the relationship with customs is one of the smoothest, returns have their transshipment point here and it is the second most significant country in supply and the fourth in demand. For these reasons, it completes the list of three countries to perform the simulations on.

At this point, the flow of products was considered to be **unrestricted**, i.e., an item that has been returned in the USA or China can have a domestic, a Rest of the World or a EU next sale. Quickly after some calculations, it became impossible to respect the unrestricted flow of products. This happened because orders that, in the As-Is model are intra-EU, were now being redirected from the USA or China to the EU, hence aggravating D&T costs.

The first action decided was that, for returns that were redirected from the USA or China, they would have to be behave like Returns Via UK and then proceed to the client. This originated the scenario "Orders Via UK" for USA and China. The results are summarized in Appendix A.

Only after analyzing the results it became clear that a new candidate should be considered. As previously stated, more than 40% of returns in 2017 were Via UK. Since Farfetch already works with Norsk to complete the transshipment of returns, the scenario of supplying orders from there must be contemplated as well. This led to the bottom band of Figure 4.3.

4.4.2 Scenario Tree

In Figure 4.4, a scheme with all contemplated scenarios is presented. After reviewing where a return may be redirected, three options were sketched: Domestic, European Union (EU) and Rest of the World (RoW). Outlined in **blue** are the considered scenarios for the simulations and in **red** the countries (or Via UK) from where the returns are coming and where the 3PL facilities would be.

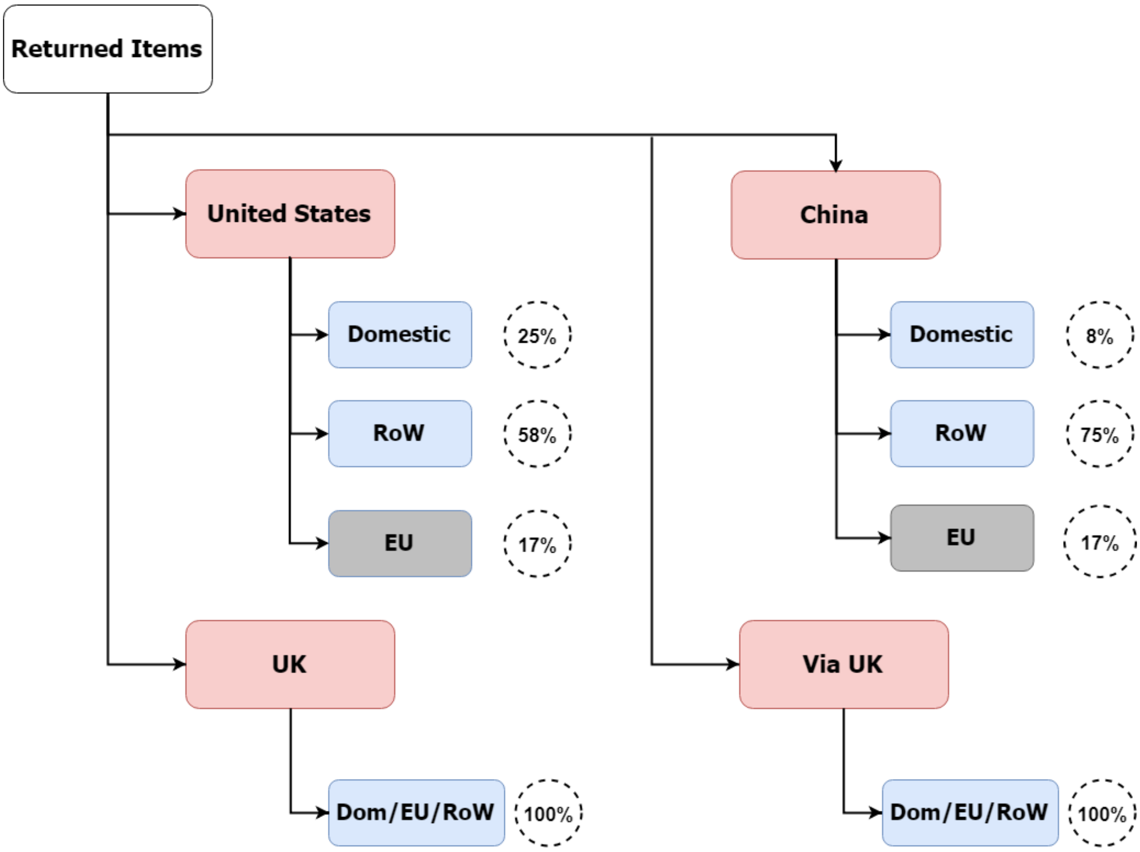


Figure 4.4: Scenario tree and distribution of returns.

Withholding the fact that more than 40% of returns in 2017 were "Via UK" (these may include items from the USA or China, for example), the distribution of returns from the USA, China and UK is represented in Figure 4.5.

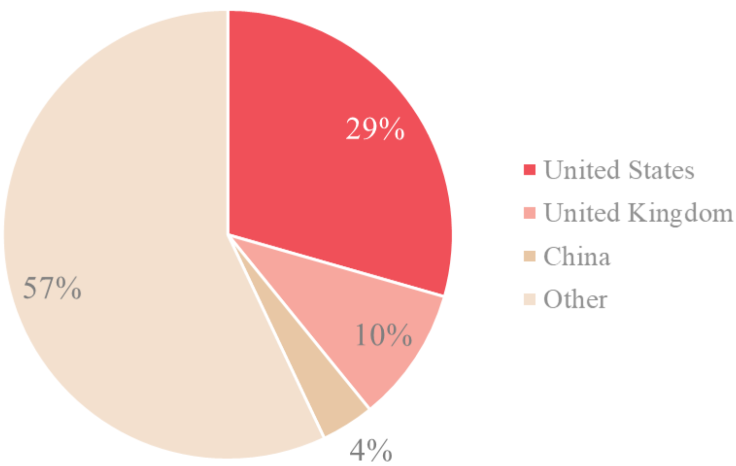


Figure 4.5: Distribution of the USA, China and UK returns.

4.5 Results

Having described the simulation process, scenarios and assessment metrics, it is now appropriate to exhibit the results for each and every one of the candidates: USA, UK, China and Via UK. As detailed in Section 4.3, the Worst Case and Best Case are represented by **WC** and **BC**, respectively.

Table 4.1: Simulation results by scenario and indicator

		Score		**Savings**		**Savings/Return**	
	Scenario	WC	BC	WC	BC	WC	BC
USA	Domestic	1.01	1.74	£ 7 247	£ 253 742	£ 0.2	£ 18.3
	Rest of the World	0.87	1.26	-£ 579 915	£ 246 242	-£ 7.3	£ 13.6
UK	Dom/EU/RoW	0.95	1.18	-£ 79 194	£ 99 769	-£ 1.7	£ 6.1
China	Domestic	1.41	4.99	£ 51 025	£ 56 809	£ 37.6	£ 95.3
	Rest of the World	0.96	1.65	-£ 22 288	£ 78 845	-£ 1.7	£ 20.5
Via UK	Dom/EU/RoW	1.18	1.28	£ 450 028	£ 799 585	£ 2.5	£ 13.1

One of the first bounded objectives of this dissertation and the closure of one of two steps, was to find where was the best opportunity in redirecting returns to new customers. To give a short answer to this far-reaching end, taking advantage of *Returns Via UK* is the safest and most profitable scenario to implement a *Return Redirection* solution.

It is by far the solution that gathers the best conditions. On one hand, there is a strong relationship with the UK Customs and a 3PL partnership is already a reality. On the other hand, the savings are maximized, even without any classification algorithm.

The USA has revealed to be a very sensitive scenario regarding algorithm accuracy, at least for Rest of the World. For only a Domestic solution, it still represents the second best choice. Yet, its vulnerability to shipping costs, particularly the case where a domestic shipping is considered without a next sale, is very penalizing towards the cost in the Worst Case.

A proof that the metrics developed reflected well up and coming countries/scenarios is China. A possible saving of £ 95.3 per item and a fivefold cost cut is extraordinarily good. Due to China predicted growth and the fact that would be more supply in Asia (3PL facilities would be located in China), the Domestic solution for China could be the third best choice.

A question that easily pops up is why are not scenarios UK and Via UK joined. In the first place, it made sense to evaluate individually the impacts from each scenario. After the results from Table 4.1, it was unclear if adding UK to Via UK was going to be advantageous. Depending on the algorithm, it may be a good strategy when it comes close to the Best Case, therefore uplifting savings and possibly reducing operational costs by economies of scale.

Having concluded the simulations and finding the ultimate best candidate, Figure 4.3 is completed with the last band, *Via UK*.

Chapter 5

Classification Algorithm

This chapter is comprised of the several stages predicted in the CRISP-DM methodology, described in Section 2.2. The first two sections describe data understanding and preparation, including how the data was retrieved, a description of the dataset and an exploratory feature analysis to pinpoint important independent variables. The last section includes the modelling and evaluation phases, encapsulating the machine learning technique choice and final results.

The dataset was retrieved through the use of the programming language SQL (Date and Darwen, 1989), through Microsoft SQL Server Management Studio, version 12.0. All data handling was performed with the programming language Python (Van Rossum and Drake Jr, 1995), version 3.6.4., through Jupyter Notebook interface, available at https://www.jupyter.org. The used modules in Python were Pandas data structures (McKinney et al., 2010), Scikit-Learn (Pedregosa et al., 2011), NumPy (Walt et al., 2011) and the Seaborn package for data visualization, version 0.8.1, from https://seaborn.pydata.org/index.html.

5.1 Data Understanding

5.1.1 Data Collection

In Chapter 4, *Returns Via UK* were selected as being the ones where a return redirection solution showed its best potential. Pursuing the full extent of the expected saving implies that only items that have a high resale probability within the time window of 15 days would be considered.

The first decision was related to how the data was to be structured in each row of the dataset. Being the algorithm a classifier that divides returns into resalable or not, it made sense that each row represented an item that had been returned with the corresponding label of an afterwards sale.

Following that structure, each row represents a return that happened in 2017. Each column is an attribute of that returned item. In Table 5.1, the summary of the input variables is described. The label "Hot_Seller" represents the predictor's outcome.

Table 5.1: Summary of the dataset's variables.

Name	Type	Description
Stsize_L	Categorical	Items of size "L"
...	...	(other 9 sizes)
Brand_ADIDAS	Categorical	Products with brand "ADIDAS"
...	...	(other brands: top 20)
Family_Clothing	Categorical	Products with family "Clothing"
...	...	(other families: top 10)
Category_Dresses	Categorical	Items which category falls into "Dresses"
...	...	(other categories: top 20)
Gender_WOMEN	Categorical	Products for gender "Women"
...	...	(other gender categories: Men, Kids, Unisex)
Pageviews	Numerical	Views from customers on Farfetch's platform for that item
Main_Color_BLACK	Categorical	Items with color "BLACK"
...	...	(other 13 colors)
Carryover	Numerical	If item was transposed from one season to another
CurrentSS	Numerical	If item is from the current Spring Summer season
CurrentAW	Numerical	If item is from the current Autumn Winter season
Retcountry_Canada	Categorical	Country of the original return
...	...	(other categories: top 10)
Sales_per_DaysOnl	Numerical	Sales of the item divided by the number of days since it has been online
Orders_per_DaysOnl	Numerical	Number of Orders of the item divided by the number of days since it has been online
Price	Numerical	Price of the product in £
Season_timing	Numerical	Item timing regarding its season
Hot_Seller	Categorical	Label of an afterwards sale within 15 days

Labelling of a return as a "Hot_Seller" is straightforward. If, for that item, a new order request was approved and fell within the time frame of 15 days after the initial return it is labelled as a "Hot_Seller". The process is depicted in Figure 5.1.

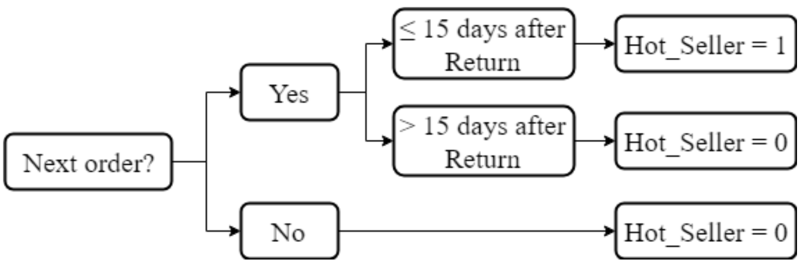


Figure 5.1: Criteria for labelling a return as a *Hot_Seller*.

5.1.2 Exploratory analysis on data

The analysis performed in this section was done with the entire dataset. Before advancing into the classification learning techniques, it is important to understand which variables will have the most impact in predicting a resalable item. For that purpose, an exploratory analysis on the dataset is performed.

In Tables 5.2 and 5.3, a summary of the non-binary and categorical variables is presented.

Table 5.2: Statistical summary of the non-binary variables.

Name	Count	Mean	Std	Skewness	Kurtosis
Pageviews	158017	1689.93	3190.18	14.14	468.05
Sales_per_daysOnl	158017	61.44	262.72	17.72	592.69
Orders_per_daysOnl	158017	17.11	62.31	17.07	609.21
Price	158017	356.10	365.25	4.48	50.61
Season_Timing	158017	0.46	0.47	-0.91	11.35

Table 5.3: Statistical summary of the categorical variables.

Name	Count	Unique	Top	Freq
Stsize	158017	9	M	39690
Retcountry	158017	99	United States	66865
Brand	158017	1670	DOLCE & GABBANA	5577
Family	158017	29	Clothing	89257
Category	158017	146	Trainers	17114
Gender	158017	4	WOMEN	96341
Main_Color	158017	13	BLACK	56206

The first thing to notice is how every variable has a count of 158.017 observations. This happened due to a pre-processing of the variables that was performed during the retrieval phase, through SQL. The initial dataset was composed of nearly 200.000 observations which included all the *Returns Via UK*, but it was reduced to 158.017 data points in order to have a complete dataset without *null* values.

In Table 5.2 high values of standard deviation (at the same level as the mean) are common. This may indicate that returned items share a lot of diversity and heterogeneity, which may may make it difficult for the algorithm to identify patterns. High kurtosis values indicate a long tail of extreme values, meaning a strong influence from outliers. And finally, some variables' distributions are highly skewed, which is a sign that those distributions are asymmetric.

The following subsections are dedicated to investigating every non-binary and categorical variable as well as the predictor output, "Hot_Seller". Some variables such as Brand and Family have been handled to only include the top 10 or 20 of the dataset. This manipulation will be covered in Section 5.2.

Hot_Seller

The first important variable to get an overview is the output variable, "Hot_Seller". It is composed of 100.892 ***0***'s and 57.125 ***1***'s. In Table 5.4, means of other numerical variables grouped by "Hot_Seller" categories are displayed.

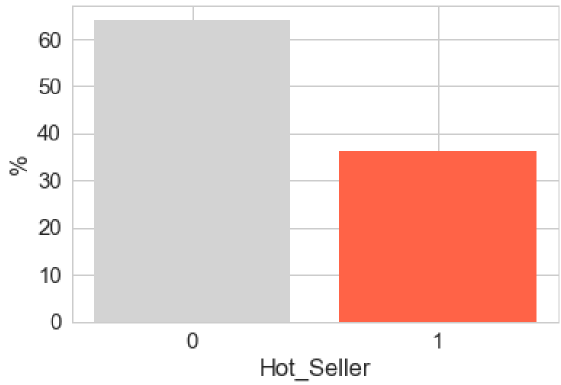


Figure 5.2: Relative frequency plot for *Hot_Seller*.

Table 5.4: Means of numerical variables grouped by *Hot_Seller* categories.

Hot_Seller	Pageviews	CurrentSS	CurrentAW	Carryover	Sales per_daysOnl	Orders per_daysOnl	Price	Season_timing
0	1283.23	0.41	0.34	0.11	46.82	11.65	381.55	0.48
1	2408.20	0.41	0.36	0.13	87.26	26.74	311.14	0.42

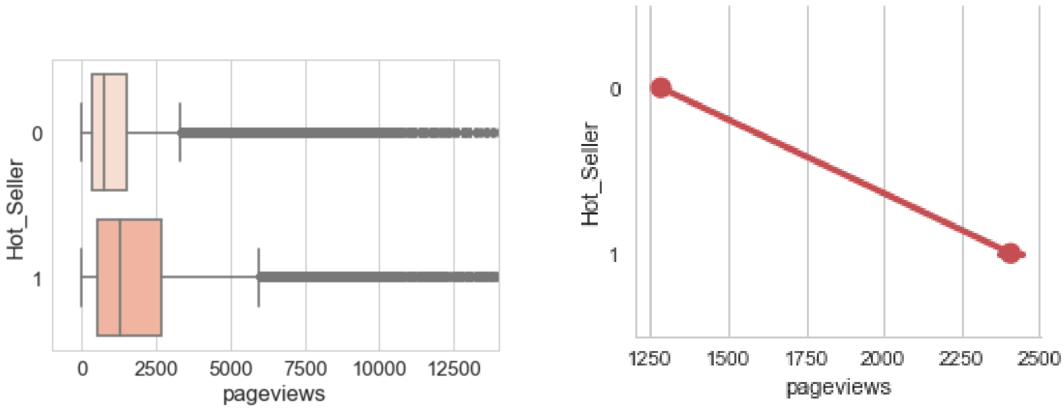
After observing the grouped means, it seems to indicate that Pageviews, Orders_per_daysOnl and Sales_per_daysOnl will have the most impact in the predictions due to their fairly different means. Although Price and Season Timing also demonstrate a difference in the mean. A further analysis of these variables is conducted.

Pageviews and Price

Figures 5.3 and 5.4 depict how the two variables are distributed by displaying a boxplot of each one and the difference of the grouped mean by "Hot_seller" category.

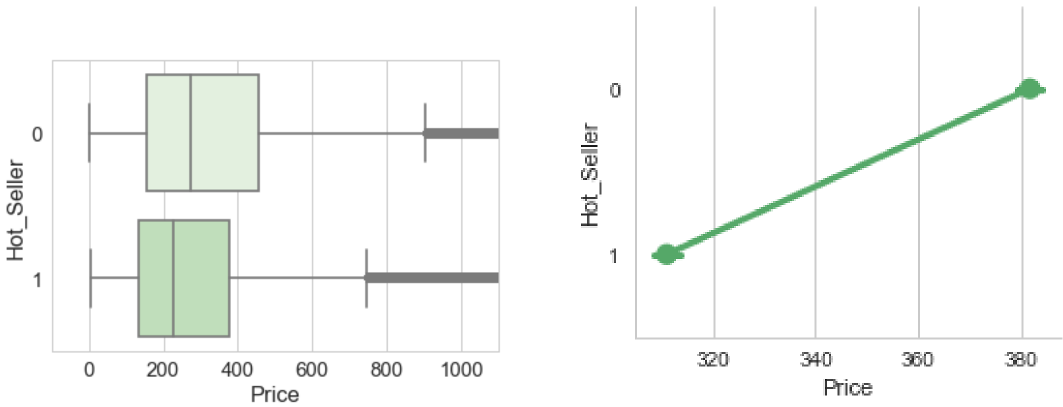
As previously mentioned, Pageviews is the number of views an item obtains during the established period of analysis. Each view is unique, meaning a customer who repeatedly checks the item only accounts for 1 view. By Figure 5.4a it is possible to observe the great amount of outliers present in the sample. Taking a look at the mean values for Pageviews one can conclude that there is a higher rate of page viewing for resalable items in the class 1 of "Hot_Seller".

Regarding Price, the conclusions are similar, although with inverse reasoning. The Price of products in class 1 seem to be lower when compared to the ones of class 0. This can possibly indicate that items with a lower price tag are prone to have a higher buying cadence.



(a) (b)

Figure 5.3: Pageviews boxplot and mean difference for *Hot_seller*.



(a) (b)

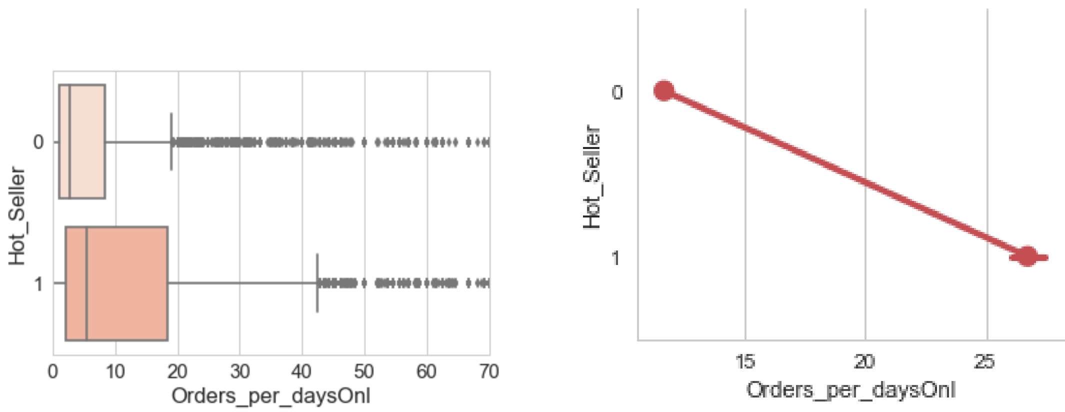
Figure 5.4: Price boxplot and mean difference for *Hot_seller*.

Season Timing, Orders per days Online and Sales per days Online

These variables are called *engineered* variables as they result of abstraction and combination of previous variables. The purpose of creating them was to obtain more significant variables to train the model with. **Sales_per_daysOnl** is simply the number of sales of that item in the established period for analysis divided by the number of days the product has been online. A similar calculation was performed for **Orders_per_daysOnl** but with the number of orders for that item.

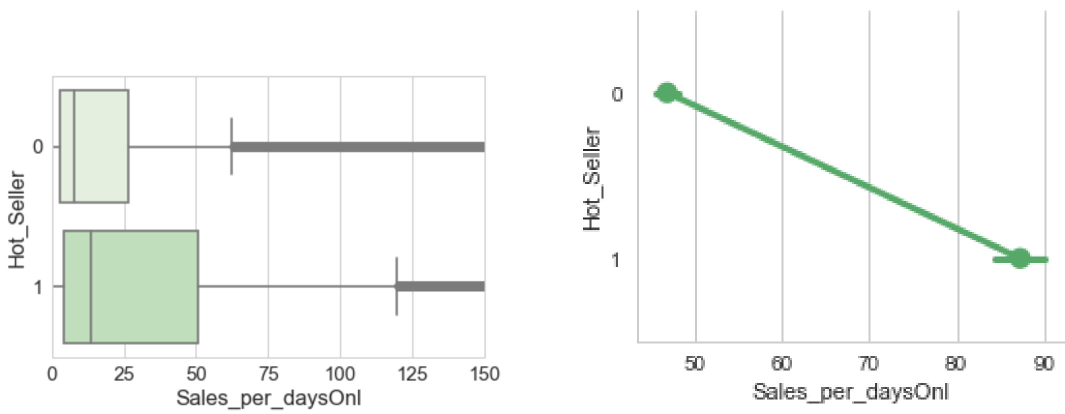
In Figures 5.5 and 5.6, boxplots of orders and sales per days online are depicted, as well as the respective grouped means. One can conclude that items that have higher values of these variables may have a higher probability of having an afterwards sale.

Season Timing was created in order to have a feature that allowed the algorithm to understand the evolution of a return throughout its season and how the probability of it having a sale afterwards changes during the months that comprise a season.



(a) (b)

Figure 5.5: Orders boxplot and mean difference for *Hot_seller*.



(a) (b)

Figure 5.6: Sales boxplot and mean difference for *Hot_seller*.

In Figure 5.7 a representation of the order volume for Autumn Winter and Summer Spring seasons of two consecutive years is depicted. As an item that is returned in the end of a season tends to have less popularity, this feature attempts to demonstrate that effect during the season by attributing a value between *0* and *1*. *0* corresponds to an item that has been returned in the very beginning of the season while *1* on the very end. Equation 5.1 deconstructs the calculation behind the engineered variable. Figure 5.8 portrays the distribution of the variable and its grouped mean.

$$S_t = \frac{Return_{date} - Season_{begin}}{Season_{end} - Season_{begin}}; \qquad S_t \in]-\infty; +\infty[\tag{5.1}$$

Most of the returned items follow the typical curve of its season. The long tail is explained by the items that are returned either before the beginning of the season or on a period after. An interesting insight is to note that items of class 1 tend to have a lower Season_timing, possible meaning these products are quickly more returned in the season.

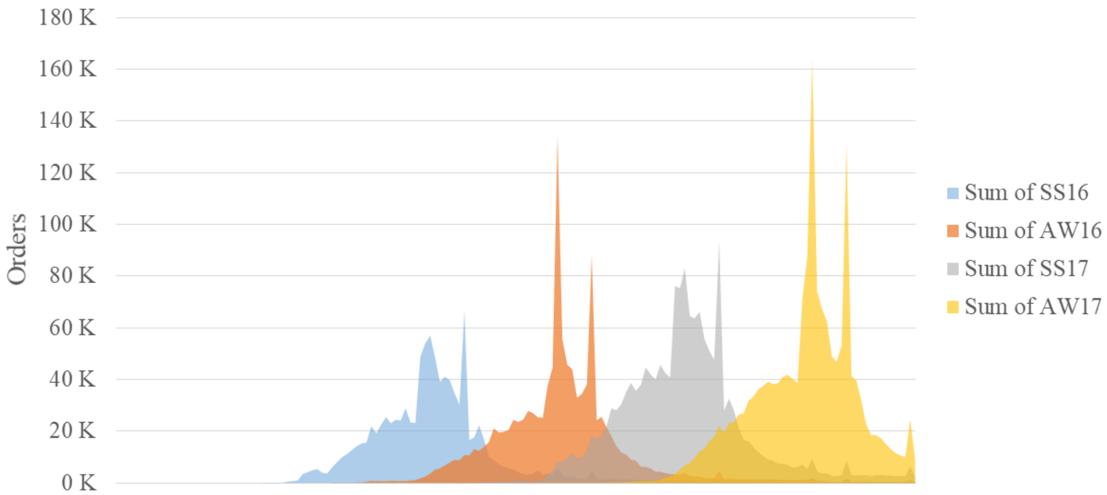
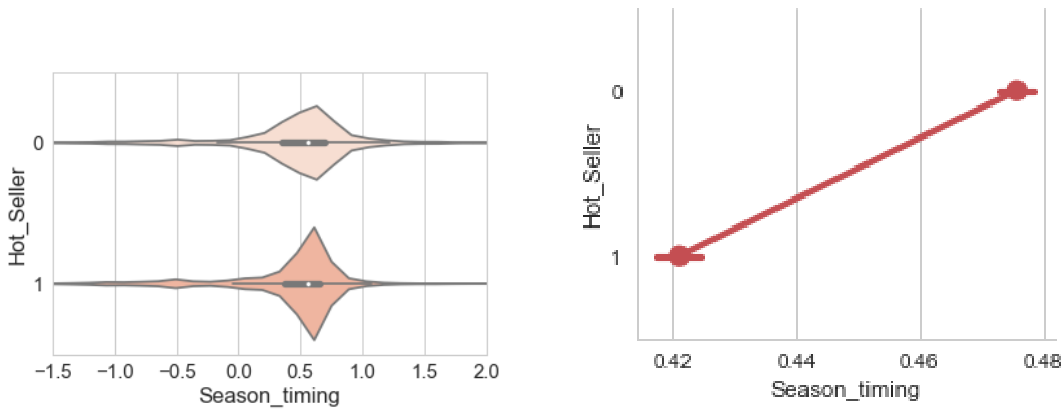


Figure 5.7: Distribution of order volume for four consecutive seasons.



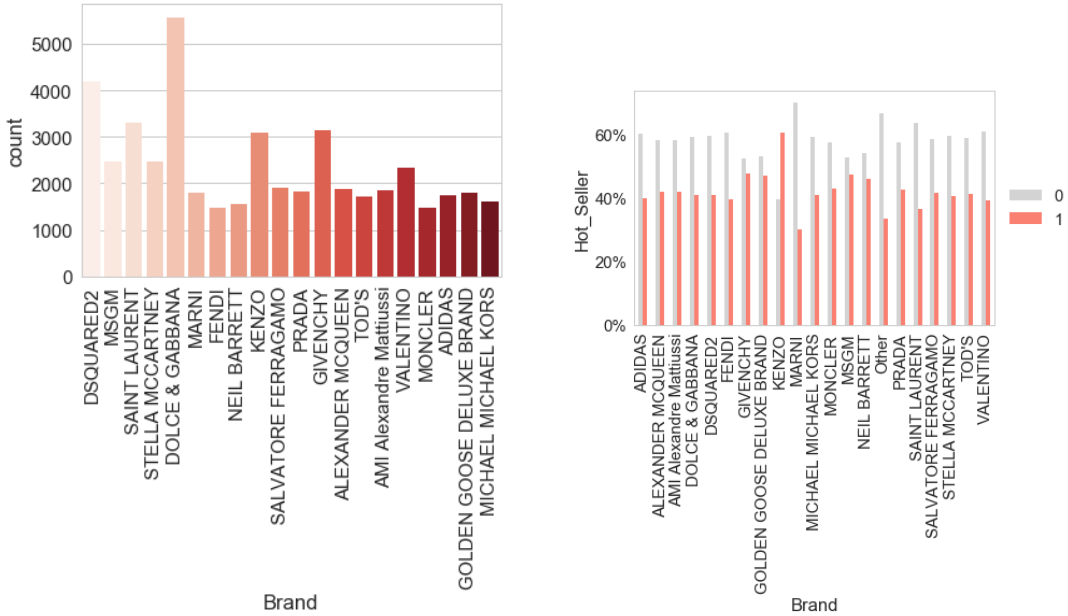
(a) (b)

Figure 5.8: Season timing distribution and mean difference for *Hot_seller*.

Brand, Family and Category

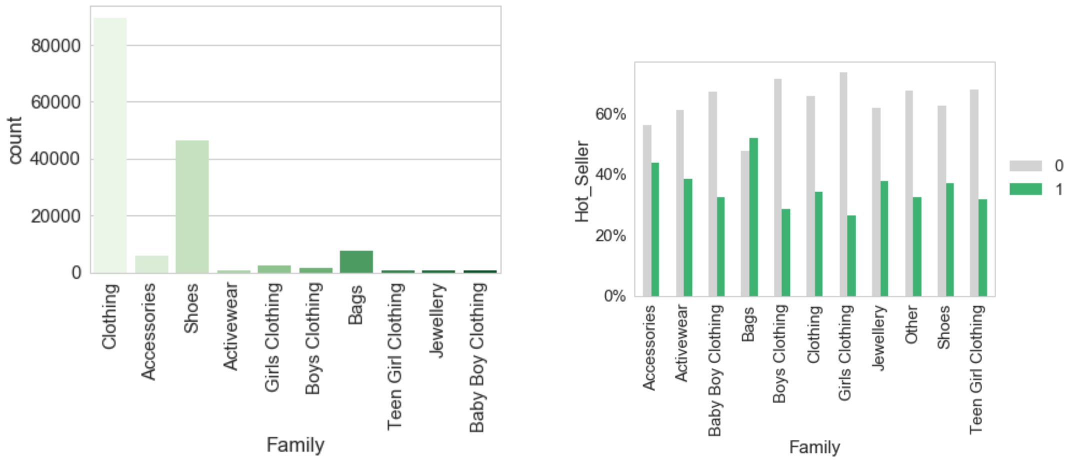
Brand, Family and Category are categorical variables that may represent an important factor in predicting if an item will have a subsequent sale in the 15 day period after its return. A certain brand or category may present this pattern. Figures 5.9, 5.10 and 5.11 depict the frequency plot and "Hot_Seller" percentage for each variable.

Taking a look at Figure 5.10b, there is evidence that some brands behave differently from others. E.g., while OFF-WHITE is a brand that has over 60% of resalable items, MARNI has a high rate of products that do not have a sale afterwards, approximately 70%. The same conclusions can be deducted for the Family and Category variables. Bags are certainly items that are being constantly requested while Girls Clothing is a more particular family of items, therefore behaving differently. For categories, T-Shirts and Skirts share that same contrast. This type of analysis potentially indicates the features which the algorithm will base its prediction on and these three categorical variables seem vital for that output.



(a) (b)

Figure 5.9: Brand frequency count and difference for *Hot_seller*.



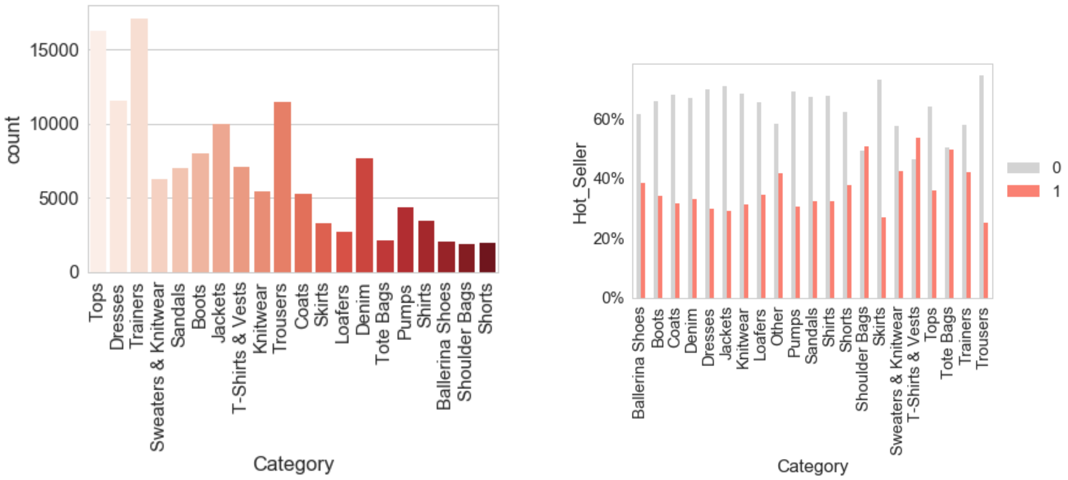
(a) (b)

Figure 5.10: Family frequency count and difference for *Hot_seller*.

Gender, Color and Size

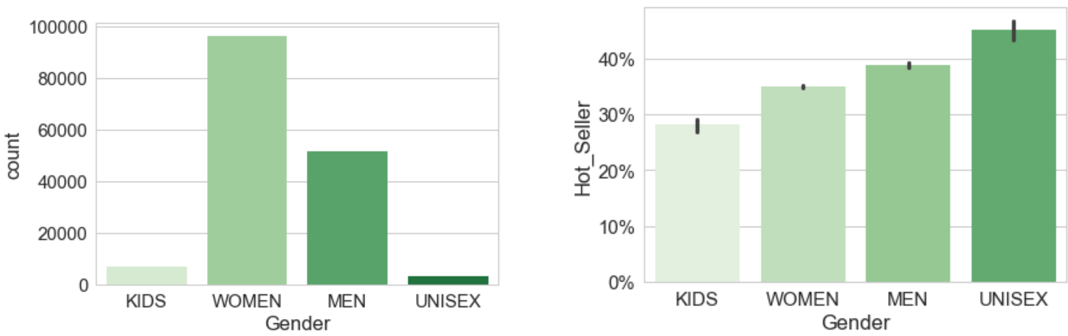
Gender, Color and Size are categorical variables that can also have impact in the predictor´s output, thus a similar analysis is performed. Figures 5.12, 5.13 and 5.14 depict the frequency plot and "Hot_Seller" percentage for each variable.

Regarding Gender, interesting insights are extracted from Figure 5.15b where UNISEX items have an advantage over the others genders in demand. A different conclusion is drawn for Colors, where probably it will not impact the output as much as other features. For Sizes, ONESIZE (OS) items have a 50% of being a return with a subsequent sale, while others like XXS have a lower probability of 30%, thus highlighting the importance of this variable.



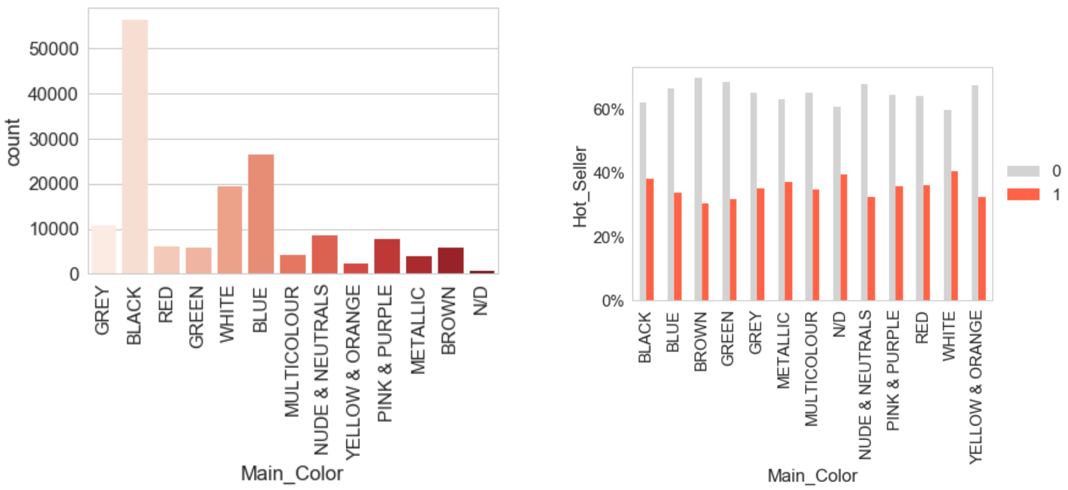
(a) (b)

Figure 5.11: Category frequency count and difference for *Hot_seller*.



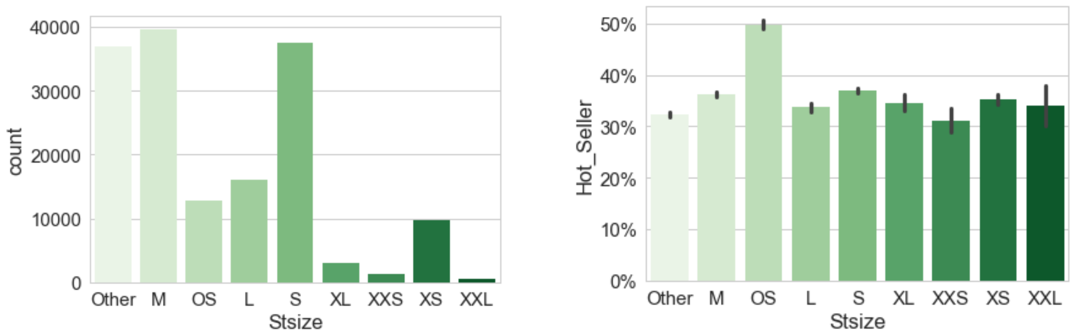
(a) (b)

Figure 5.12: Gender frequency count and difference for *Hot_seller*.



(a) (b)

Figure 5.13: Main Color frequency count and difference for *Hot_seller*.



(a) (b)

Figure 5.14: Size frequency count and difference for *Hot_seller*.

Returning Country

Every country has a distinct marketplace by which the sample of returned items is strongly heterogeneous. To further investigate this hypothesis, Figure 5.15 explores the geographical distribution of returns and Figure 5.16 the frequency of the countries grouped by "Hot_Seller".

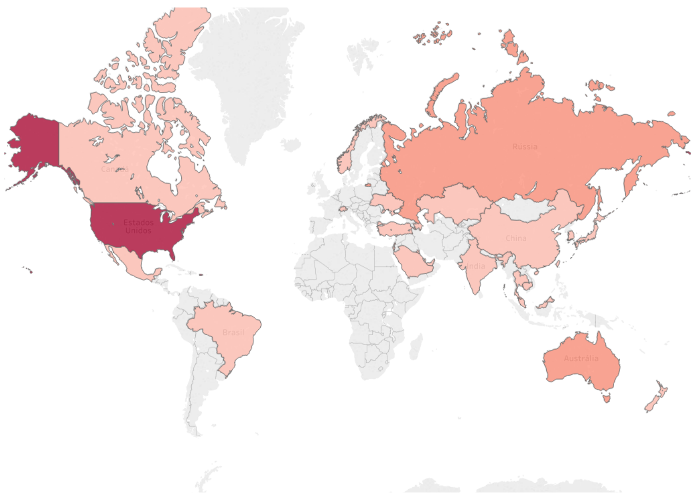


Figure 5.15: Geographical distribution of returns.

Darker colors in Figure 5.15 represent countries with a bigger share of returns. After observing the geographical plot, the United States stand out for its volume of returns, therefore the North American market will be representative of the item's nature. Nevertheless, in Figure 5.16, variations of "Hot_Seller" for the different countries does not seem willing to give impact in the predictor, with every returning country remaining between 30% and 40% of "Hot_Sellers".

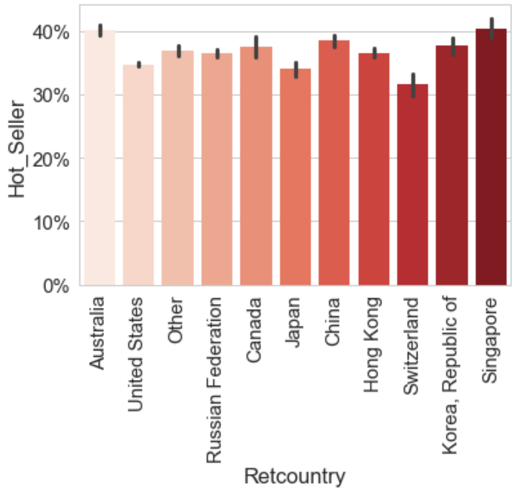


Figure 5.16: Grouped frequency difference for *Hot_seller*.

5.2 Data Preparation

The previous section offered insights about the relative importance of each variable to the predictor outcome. This section, on the other hand, includes the preparation methods used in order to successfully apply the machine learning techniques chosen: Random Forest, Support Vector Machine and Logistic Regression.

There were two data processing techniques used. One-hot-encoding using dummy variables to transform categorical variables into numerical ones, and the Min-Max method to adjust and match the numerical scale for every variable. Support Vector Machine, for example, requires data input to be standardized as aforementioned (Nisbet et al., 2009).

5.2.1 One-hot-encoding for categorical variables

One-hot-encoding consists in creating a set of dummy variables that represent a categorical variable through a binary one. Features such as Brand, Family and Category were filtered down initially to a top 10 or 20, as previously mentioned. After this stage, dummy variables were created. Tables 5.5 and 5.6 display an example of the process.

Table 5.5: Before one-hot-encoding

Item	Brand
1	Brand1
2	Brand2
3	Brand3
4	Brand4
5	Brand2

Table 5.6: Four dummy variables created after one-hot-encoding

Item	Brand1	Brand2	Brand3	Brand4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	0	1	0	0

By doing this, categorical variables are transformed to numerical, enabling a Random Forest or a SVM to analyze the data.

5.2.2 Min-Max method

The last method used to standardize the variables was the *Min-Max method*. Despite having more standardization methods to choose from, such as the widely used Z-values, the Min-Max method was proven to be the most adequate for data mining problems (Al Shalabi et al., 2006).

As numerical variables have different scales and units, these were standardized to values between *0* and *1*, according to Equation 5.2.

$$X^{'} = \frac{X - X_{min}}{X_{max} - X_{min}}; \qquad X^{'} \in [0;1] \tag{5.2}$$

5.3 Modelling and Evaluation

Throughout this chapter, the standard process for a data mining project was followed. From business and data understanding, to data handling and transformation. In this last section, the modelling phase begins and the chosen algorithms are tested and tuned to achieve the best global performance.

5.3.1 Splitting method

In this phase the dataset was split into a training, validation and testing set. An extensively used division is to select 20% to 40% of the data to form the testing set. Considering the size of the dataset, it was decided that 20% of the data was for testing and 80% for training and validating. The dataset was finally divided into a random training/validating set, 126.413 data points, and a random testing set, 31.604 data points, totalling 158.017 observations. The training set was then used for the algorithm to learn the intrinsic relationships between the variables and the validation set to obtain unbiased results of the performance metrics. In order to do so, the *K-fold cross-validation* technique has been chosen. This splitting method allows for a stable estimate of the model's performance metrics and error. An example of the prior process is displayed in Figure 5.17.

In this dissertation a ten-fold cross-validation was chosen, as this represents a common practice in machine learning problems (Reitermanova, 2010). As for the testing set, this was later used to perform the final assessment and parameter tuning of the selected algorithm.

5.3.2 Performance assessment

In Section 2.2.2, the common metrics for a model's performance were reviewed. As such, global metrics including Accuracy, Precision, Recall, F1-Score and AUC-ROC were selected to evaluate each model. However, taking into account the purpose which this algorithm was designed for, it is possible to build a metric based on the expected savings per return simulated in Section 4.3.

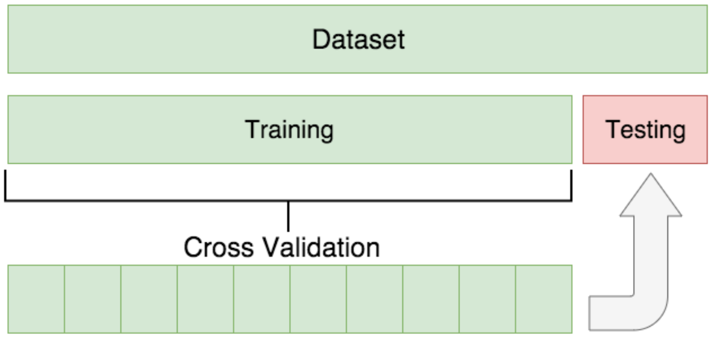


Figure 5.17: Division of the dataset using a ten-fold cross-validation technique, adapted from (Adi Bronshtein, 2017).

For every item that is classified with the label *1*, there can be two possible outcomes:

1. The item has a sale afterwards and, in that case, it was correctly classified as a "Hot_Seller", *True Positive*;
2. No subsequent sale happened and the item incurs in the processing costs of handling and storage of the 3PL, *False Positive*;

Having explored the options behind a predictor's labelling of *1*, a new metric to assess the model's performance was formulated. Considering the expression of foreseen savings per return, simulated in Section 4.5, Equation 5.3 outcomes the predictor's expected savings, E_{sav}. The simulated average savings per item and the estimated processing costs are summarized in Table 5.7.

$$E_{sav} = \left(\frac{TP \times Sav_{ret} - FP \times Proc_{ret}}{N_{sample}} \right) \times N_{ret} \tag{5.3}$$

Table 5.7: Values for expected savings and processing cost per item.

	Expression	**Expected Value**
Sav_{ret}	$S_{ret} + DT_{ret} - Proc_{ret} + LTS_{ret}$	£ 13
S_{ret}	-	£ 15
DT_{ret}	-	£ 0
$Proc_{ret}$	-	£ 3
LTS_{ret}	-	£ 1
N_{sample}	-	31604
N_{ret}	-	198027

5.3.3 Model selection

The selected algorithms to test and evaluate the best fit for this particular case were Random Forest, Logistic Regression and Support Vector Machine. The goal of this section was to take the training subset and, through cross-validation, measure the performance of each model. In Table 5.8 a summary of the standard hyperparameters used for model selection is displayed (Pedregosa et al., 2011).

Table 5.8: Initial parameters per algorithm.

	Algorithm		
Hyperparameter	**RF**	**SVM**	**LR**
Nr_Trees	200	-	-
Max_features	Auto	-	-
Min_samples_leaf	1	-	-
Kernel	-	Linear	-
C	-	-	1

From Table 5.9, it is possible to conclude that the best performance is achieved by the Random Forest algorithm. Support Vector Machine and Logistic Regression are similar in performance and achieve worse results especially in recall and AUC-ROC. Taking these conclusions into account, the Random Forest algorithm is selected to be tested and tuned.

Table 5.9: Results of cross-validation.

	Algorithm		
Performance metric	RF	SVM	LR
Accuracy	0.715	0.660	0.680
Precision	0.675	0.617	0.646
Recall	0.408	0.156	0.254
F1-Score	0.508	0.249	0.364
AUC-ROC	0.648	0.550	0.587

5.3.4 Testing and tuning

Finally, the testing set, which was separated from the entire dataset in Section 5.3.1, is now used to tune the Random Forest classifier and obtain the final results of the model. The hyperparameters subject to tuning are *Nr_Trees*, *Max_number_features* and *Min_samples_leaf*.

As stated earlier in this dissertation, the dataset is **imbalanced**. There is approximately 35% of "Hot_Sellers" in the dataset and this causes a misleading model's performance. One hyperparameter that can also be changed when unbalanced datasets are processed is *Class_weight*. When no input is given, the algorithm assumes a balanced dataset. In this case, class_weight was also explored in order to achieve the most accurate fit for the model.

Tables 5.10 through 5.13 reflect grid searches on algorithm parameters with three metrics chosen to perform the evaluation: E_{sav}, F1-Score and AUC-ROC.

Table 5.10: Grid search for maximum number of features.

Maximum number of features							
			Class_weight				
Max_features	Min_samples_leaf	Nr_Trees	*0*	*1*	E_{sav}	F1_Score	AUC-ROC
Sqrt (= Auto)	1	200	1	1	£ 331 653	0.502	0.646
Log2	1	200	1	1	£ 312 831	0.484	0.638
98 (dataset columns)	**1**	**200**	**1**	**1**	**£ 371 059**	**0.535**	**0.659**

Table 5.11: Grid search for number of trees.

Number of trees							
			Class_weight				
Max_features	Min_samples_leaf	Nr_Trees	*0*	*1*	E_{sav}	F1_Score	AUC-ROC
Auto	1	200	1	1	£ 330 067	0.501	0.644
Auto	1	300	1	1	£ 332 423	0.503	0.647
Auto	1	600	1	1	£ 331 546	0.503	0.646
Auto	1	900	1	1	£ 333 482	0.501	0.646
Auto	1	1200	1	1	£ 332 035	0.503	0.646
Auto	**1**	**1500**	**1**	**1**	**£ 333 558**	**0.504**	**0.646**

Table 5.12: Grid search for minimum sample of a leaf.

Minimum sample of a leaf							
			Class_weight				
Max_features	Min_samples_leaf	Nr_Trees	*0*	*1*	E_{sav}	F1_Score	AUC-ROC
Auto	**1**	**200**	**1**	**1**	**£ 328 833**	**0.500**	**0.644**
Auto	3	200	1	1	£ 322 185	0.494	0.642
Auto	5	200	1	1	£ 314 390	0.486	0.641
Auto	7	200	1	1	£ 306 840	0.482	0.636
Auto	9	200	1	1	£ 304 503	0.478	0.632
Auto	11	200	1	1	£ 299 791	0.471	0.631
Auto	13	200	1	1	£ 292 460	0.463	0.628

The hyperparameters that enabled the algorithm to output the best results in the context of this problem are highlighted in bold. Class_weight had the most impact in the performance metrics due to the fact that, as mentioned beforehand, the dataset was imbalanced. The maximum savings was not the criteria in this hyperparameter due to the considerable loss in AUC-ROC. A graphical representation of the tuned hyperparameters are shown in Appendix B. In Table 5.14, the final testing results are summarized. The ROC curve is displayed in Figure 5.18 and the confusion matrix is presented in Table 5.15. On top of that, the Random Forest algorithm makes it possible to check the relative importance of each feature for its predictions. In Figure 5.19 a bar plot with that information is displayed. The results are discussed in the following Section.

Table 5.13: Grid search for optimized class weight.

Class weight							
			Class_weight				
Max_features	Min_samples_leaf	Nr_Trees	*0*	*1*	E_{sav}	F1_Score	AUC-ROC
Auto	1	200	1	1	£ 331 552	0.502	0.645
Auto	1	200	1	2	£ 482 222	0.599	0.682
Auto	1	200	1	3	£ 545 890	0.615	0.680
Auto	**1**	**200**	**1**	**4**	**£ 571 993**	**0.613**	**0.669**
Auto	1	200	1	5	£ 582 150	0.606	0.655
Auto	1	200	1	6	£ 587 658	0.600	0.643
Auto	1	200	1	7	£ 590 960	0.596	0.634
Auto	1	200	1	8	£ 591 737	0.591	0.625
Auto	1	200	1	9	£ 591 793	0.587	0.618

Figure 5.18: ROC curve of testing results with Random Forest algorithm.

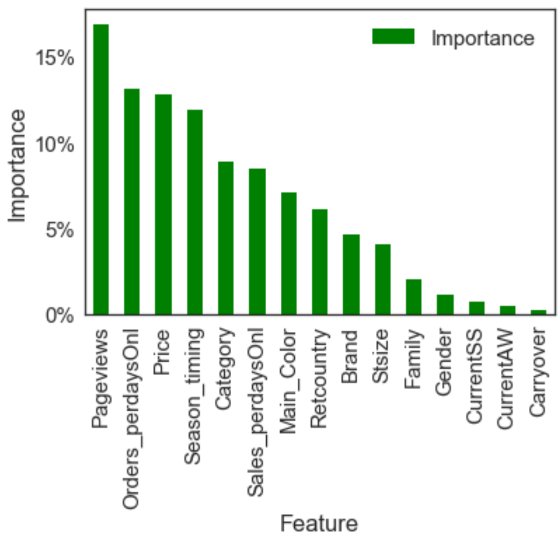


Figure 5.19: Relative importance of the features.

Table 5.14: Final testing results.

Classifier	E_{sav}	F1-Score	AUC-ROC	Precision	Recall
Random Forest	£ 510807	0.602	0.674	0.538	0.682

Table 5.15: Confusion matrix of results.

	Hot_Seller (True)	Not Hot_Seller (True)	Total
Hot_Seller (Predict)	7821 (24,75%)	6717 (21,25%)	14538
Not Hot_Seller (Predict)	3641 (11,52%)	13425 (42,48%)	17066
Total	11462	20142	31604

5.3.5 Discussion of results

Before proceeding to the discussion, it is important to recall the context and meaning of the classifier built in this section. To achieve meaningful savings by redirecting returns, only the items with potential to be resold should be considered. This meant that the challenge presented was one of the most demanding in a retail environment: predict what characteristics of an item will succeed in its market. Having said that, it was not clear if the data would show a pattern clear enough for any Machine Learning technique to apprehend and on such a complex market as the one of luxury fashion.

The performance in cross-validation and tuning is often a more optimistic bias, for which the results were expected to be lower in the testing phase. Despite that, the final outcome can be considered satisfactory. Perhaps one of the most important metrics is the confusion matrix, in Table 5.15. It shows that 42.48% of returns were correctly classified as not having a subsequent sale, and only 24.75% as having one. The reason behind this is the fact that the dataset is so unbalanced that it is "easier" for the predictor to identify items that do not have next sales. However, from all the positive samples, having correctly classified 68.2% (recall) of them is very good.

From cross-validation to testing, the recall value rose from 40.8% to 68.2% and the driver behind this growth was that, in this problem's context, it made sense to "force" the algorithm into finding more positive samples. By doing this, the precision was lowered (from 67.5% to 53.8%), but true positives have a higher valuation than the false positives penalties. This growth was mainly manifested through the optimization of the weight of class 1 and transmitted through the value of expected savings.

An opportunity was missed with the 11.52% of false negatives in the sample, and this reveals to be of vital importance when the expected savings potential was £ 800 000. These items were not identified as "Hot_Sellers", although having a subsequent sale in the next 15 days. The key in reducing this error may come from more engineered variables, specially about the sales profile of an item. They represented potential savings and its case should be minimized in further work.

About Figure 5.19 and the relative importance of the features, the variables pageviews, price and the engineered variables number of orders and season timing were among the most relevant features to the algorithm, ranging from 10% to 17% in relative importance. This verifies some of the predictions settled in Section 5.1.2, mostly from the aforementioned variables. Moreover, variables carryover, currentAW and currentSS were estimated to have low impact, which was also confirmed. On the other hand, variables such as family and gender did not have the expected impact. This may be due to the concentration of items from the "Clothing" family and "Women" gender that tampered with the sample balancing making it difficult for the algorithm to retrieve valuable insights from these variables. A more specific division of those categories such as Clothing_Summer, for example, could aid the algorithm in the future.

Chapter 6

Conclusions and future work

This dissertation had a demanding purpose that more and more companies are focusing on, which is supply chain efficiency and cost saving, more precisely in reverse logistics. For Farfetch, supply chain plays a main role in the company's profit and loss. As it operates through a drop-shipping model, the delivery experience is the "product" Farfetch sells the most. In this particular case, the dissertation focused on studying the impacts of redirecting returns on the main business KPIs. After an initial assessment, it became clear that the approach would have to be two-folded: a first step where the impacts of redirecting returns regarding potential cost savings through different scenarios were explored; and a second and final step by which taking advantage of Farfetch being a data driven company, a Machine Learning algorithm to classify returns as "Hot_Sellers" was built.

The premise behind redirecting returns is straightforward. After a customer creates a return of an item, it gets shipped to a new customer, instead of heading back to the boutique, expecting to save costs along the way. However, some questions derive from this solution. Which markets or countries are the most adequate to implement this? Should every returned item be considered? What impacts can this have on the boutique's perspective? Is the customer's viewpoint not important? In order to provide answers to these questions, simulations involving redirecting returns were performed with historical data from 2017. The analysis was structured into initially obtaining an overview of the Farfetch's top markets and return process. Afterwards, each specific market and expected impacts were assessed. A set of assumptions had to be made in order to make the savings estimate calculation possible.

The filtering process led to four possible candidates: the United States, China, UK and Returns Via UK. Each one had its own scenarios that were individually assessed. After developing specific metrics to evaluate each scenario, *Returns Via UK* was shown to demonstrate the best potential. As already stated, Farfetch has a transshipment point in the UK, through which some returns enter the EU without D&T being charged. Not only Farfetch already has a 3PL partner (Norsk) in the country but also no additional "legs" of shipping are added.

During the cost simulations, two cases were considered: the best case, where we only targeted the best selling returns (with a 15 days subsequent sale) into the solution and the worst case where every return was considered, either having an afterwards sale or not.

With the aim of achieving the full potential behind redirecting returns through Via UK a Machine Learning classifier was proposed. This was the second step in the dissertation and it followed the cross industry standard for data mining projects methodology. The objective was to identify every item that had a sale following its return creation date. To do that, a dataset with several features concerning the item characteristics was retrieved.

After going through an extensive exploratory analysis on the variables, three Machine Learning techniques were tested. Random Forest, Support Vector Machine and Logistic Regression were assessed and, through cross-validation, the Random Forest showed the most promising potential. It was essential to always recall the "physical" meaning of this classifier for Farfetch. Its goal was to try and find all the "Hot_Sellers", or finding all the positive samples in the dataset. Consequently, the hyperparameter tuning phase revealed to be of an extreme importance. It allowed the algorithm to "take chances" in finding all the "1's" and, by doing this, improving the final results. Conclusively, the classifier had a 67% chance of giving a higher score to a "Hot_Seller".

This dissertation was a very extensive project, regarding the level of granularity one can consider when assessing scenarios and performing simulations. The work developed had to consider some assumptions and most of the future suggestions are related to exploring the conditions exposed in Section 4.2 and doing sensitivity analysis.

The first suggestion has to do with extending the impacts of having a return redirection solution to other company's performance indicators. The approach would be to evaluate the changes in refused returns, RTO, retention rate (probability of a customer buying again at Farfetch) and in the net promoter score (customer feedback after order/return). Studying the impacts in these areas would have extensively enriched the whole project.

Another suggested study is to perform a sensibility analysis regarding the 3PL processing cost (£ 3 per article) and the number of days an item is being kept at a 3PL facility. This would help to identify the expected savings confidence interval and provide more information to base operational decisions.

Finally, concerning the algorithm developed in the second step, a deeper variable study should be conducted, with more engineered variables to improve the algorithm. Also, the deployment phase at Farfetch should be studied. This would allow to establish how the algorithm is trained, how frequently and with which data. After implementation, a KPI to monitor which items are successfully being classified as "Hot_Sellers" should be created.

Bibliography

Adi Bronshtein (2017). Train/test split and cross validation in python.

Agrawal, S., Singh, R. K., and Murtaza, Q. (2015). A literature review and perspectives in reverse logistics. *Resources, Conservation and Recycling*, 97:76–92.

Al Shalabi, L., Shaaban, Z., and Kasasbeh, B. (2006). Data mining: A preprocessing engine. *Journal of Computer Science*, 2(9):735–739.

Bain & Altagamma (2017). Altagamma 2017 worldwide luxury market monitor.

Bain & Company (2017). Luxury goods worldwide market study, fall - winter 2017.

Bloemhof-Ruwaard, J. M., Fleischmann, M., and Van Nunen, J. A. (1999). Reviewing distribution issues in reverse logistics. In *New trends in distribution logistics*, pages 23–44. Springer.

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

Business of Fashion (2018). On road to autumn ipo, farfetch inks middle east deal.

Business of Fashion & Mckinsey (2018). State of fashion 2018.

Caruana, R. and Niculescu-Mizil, A. (2004). Data mining in metric space: an empirical analysis of supervised learning performance criteria. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 69–78. ACM.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). Crisp-dm 1.0 step-by-step data mining guide.

Columbus, L. (2014). Making analytics accountable: 56% of executives expect analytics to contribute to 10% or more growth in 2014.

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

Date, C. J. and Darwen, H. (1989). *A guide to the SQL Standard: a user's guide to the standard relational language SQL*. Addison-Wesley.

De Brito, M. P. and Dekker, R. (2002). Reverse logistics-a framework. Technical report, Erasmus School of Economics (ESE).

Deming Institute (2018). Plan do study act by w. edwards deming.

EBC (2017). Predicting with labeled data.

Ecommerce Europe (2016). Ecommerce benchmark retail report 2016.

Fashionista (2018). Farfetch is touching down in the middle east.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.

Financial Times (2018). London-based online retailer farfetch eyes us public offering.

Fleischmann, M., Bloemhof-Ruwaard, J. M., Dekker, R., Van der Laan, E., Van Nunen, J. A., and Van Wassenhove, L. N. (1997). Quantitative models for reverse logistics: A review. *European journal of operational research*, 103(1):1–17.

Fleischmann, M., Van Nunen, J. A., and Gräve, B. (2003). Integrating closed-loop supply chains and spare-parts management at ibm. *Interfaces*, 33(6):44–56.

Galbreth, M. R. and Blackburn, J. D. (2006). Optimal acquisition and sorting policies for remanufacturing. *Production and Operations Management*, 15(3):384–392.

Guide, V. and Wassenhove, L. (2003). Full cycle supply chains.

Guiltinan, J. and Nwokoye, N. (1974). Reverse channels for recycling: an analysis for alternatives and public policy implications. In *American Marketing Association Proceedings*, volume 36, pages 341–46. Spring and Fall.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28.

Issa, M. B., Daraghmeh, M., Jararweh, Y., Al-Ayyoub, M., Alsmirat, M., and Benkhelifa, E. (2017). Using logistic regression to improve virtual machines management in cloud computing systems. In *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pages 628–635. IEEE.

Janakiraman, N., Syrdal, H. A., and Freling, R. (2016). The effect of return policy leniency on consumer purchase and return decisions: A meta-analytic review. *Journal of Retailing*, 92(2):226 – 235.

Jedd, M. (2000). Returns happen: reverse logistics online. *Inbound Logistics*.

Kiron, D., Prentice, P. K., and Ferguson, R. B. (2014). The analytics mandate. *MIT Sloan management review*, 55(4):1.

Kokkinaki, A. I., Dekker, R., Van Nunen, J., and Pappis, C. (2000). An exploratory study on electronic commerce for reverse logistics. In *Supply chain forum: An international Journal*, volume 1, pages 10–17. Taylor & Francis.

Krikke, H., Bloemhof-Ruwaard, J., and Van Wassenhove, L. (2003). Concurrent product and closed-loop supply chain design with an application to refrigerators. *International journal of production research*, 41(16):3689–3719.

Kumar, S. and Putnam, V. (2008). Cradle to cradle: Reverse logistics strategies and opportunities across three industry sectors. *International Journal of Production Economics*, 115(2):305–315.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., and Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10):60–68.

McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.

Nisbet, R., Elder, J., and Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press.

Park, H. (2013). An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2):154–164.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Pokharel, S. and Mutha, A. (2009). Perspectives in reverse logistics: a review. *Resources, Conservation and Recycling*, 53(4):175–182.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.

Prahinski, C. and Kocabasoglu, C. (2006). Empirical research opportunities in reverse supply chains. *Omega*, 34(6):519–532.

Provost, F. J., Fawcett, T., et al. (1997). Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. In *KDD*, volume 97, pages 43–48.

Ravi, V. and Shankar, R. (2005). Analysis of interactions among the barriers of reverse logistics. *Technological Forecasting and Social Change*, 72(8):1011–1029.

Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross validation, encyclopedia of database systems (edbs). *Arizona State University, Springer*, page 6.

Reitermanova, Z. (2010). Data splitting. In *WDS*, volume 10, pages 31–36.

Rogers, D. S. and Tibben-Lembke, R. (2001). An examination of reverse logistics practices. *Journal of business logistics*, 22(2):129–148.

Rogers, D. S., Tibben-Lembke, R. S., et al. (1999). *Going backwards: reverse logistics trends and practices*, volume 2.

South China Morning Post (2016). Online fashion curator farfetch, which links shops to customers, grows in china.

Stock, J. R. and Mulki, J. P. (2009). Product returns processing: an examination of practices of manufacturers, wholesalers/distributors, and retailers. *Journal of business logistics*, 30(1):33–62.

Stuart, J. A., Bonawi-tan, W., Loehr, S., and Gates, J. (2005). Reducing costs through improved returns processing. *International Journal of Physical Distribution & Logistics Management*, 35(7):468–480.

Tavish Srivastava (2015). Tuning the parameters of your random forest model.

TechAmerica Foundation's Federal Big Data Commission (2012). Demystifying big data: A practical guide to transforming the business of government.

TechCrunch (2017). Jd.com invests $397m into luxury marketplace farfetch as part of a new strategic partnership.

Thierry, M., Salomon, M., Van Nunen, J., and Van Wassenhove, L. (1995). Strategic issues in product recovery management. *California management review*, 37(2):114–136.

Tibben-Lembke, R. S. (1998). The impact of reverse logistics on the total cost of ownership. *Journal of Marketing Theory and Practice*, 6(4):51–60.

Tibben-Lembke, R. S. and Rogers, D. S. (2002). Differences between forward and reverse logistics in a retail environment. *Supply Chain Management: An International Journal*, 7(5):271–282.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vibhor Agarwal (2017). Getting started with machine learning using sklearn-python.

Walt, S. v. d., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30.

William Koehrsen (2017). Random forest simple explanation.

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

Wolpert, D. H. and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82.

Zhang, Z., Trevino, V., Hoseini, S. S., Belciug, S., Boopathi, A. M., Zhang, P., Gorunescu, F., Subha, V., and Dai, S. (2018). Variable selection in logistic regression model with genetic algorithm. *Annals of translational medicine*, 6(3).

Zikopoulos, C. and Tagaras, G. (2008). On the attractiveness of sorting before disassembly in remanufacturing. *IIE Transactions*, 40(3):313–323.

Appendix A

Impacts study first stage results

The first approach to simulate the impacts of redirecting returns was performed as described in Section 4.4.1. This involved considering that new orders that originated from the EU and for which the returns would be redirected, would have to follow the path of "Returns Via UK". This allowed the exemption of D&T upon the entrance in the EU. Quickly it became clear this was not a solution to consider. The results are displayed in Table A.1.

Table A.1: Simulation results by scenario and indicator

		Score		**Savings**		**Savings/Return**	
	Scenario	WC	BC	WC	BC	WC	BC
USA	Via UK orders	0.961	1.416	-£ 258 210	£ 646 760	-£ 6.1	£ 15.3
	No Via UK Orders	0.829	0.909	-£ 1 297 276	-£ 221 208	-£ 30.8	-£ 5.2
UK	N/A	0.95	1.184	-£ 79 194	£ 99 769	-£ 1.7	£ 6.1
China	Via UK orders	0.984	1.301	-£ 13 960	£ 74 959	-£ 2.5	£ 13.2
	No Via UK orders	0.96	1.65	-£ 22 288	£ 78 845	-£ 1.7	£ 20.5

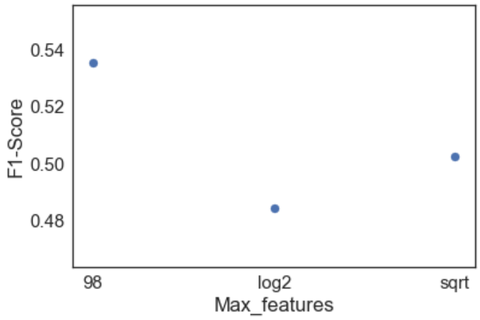
Appendix B

Tuning hyperparameters

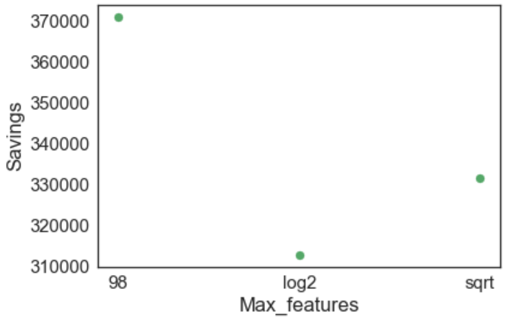
After the cross-validation results the Random Forest algorithm was chosen to be the model to predict if a return is to be classified as a "Hot_Seller" or not. In order to achieve the maximum potential of the algorithm, refined hyperparameter tuning is needed. As described in Section 5.3.4, the hyperparameters to optimize are the following:

- Maximum number of features;
- Minimum samples of a leaf;
- Number of estimators (trees);
- Weight of Class 1.

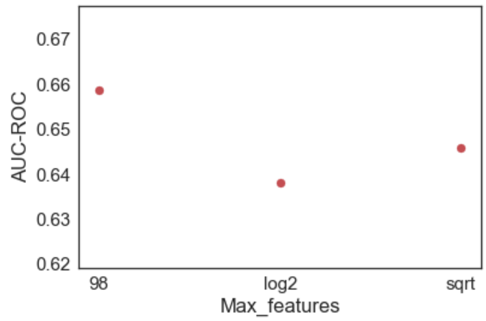
Figures B.1 through B.4 depict a graphical representation of the hyperparameter tuning.



(a)

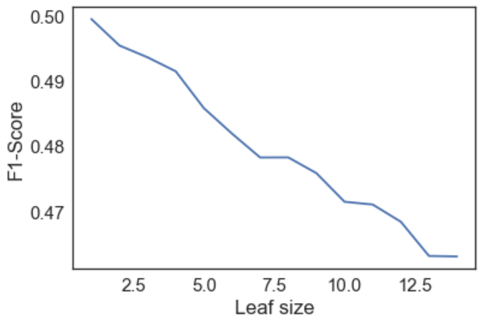


(b)

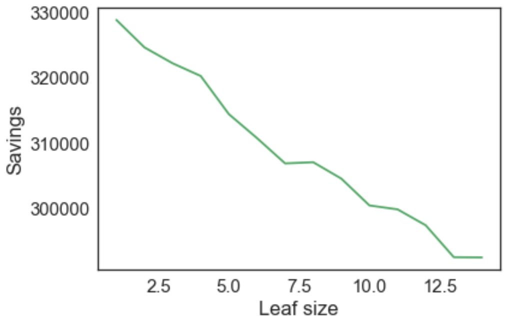


(c)

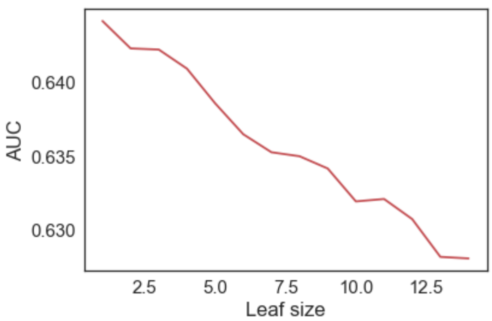
Figure B.1: Exhaustive search for optimum maximum number of features.



(a)

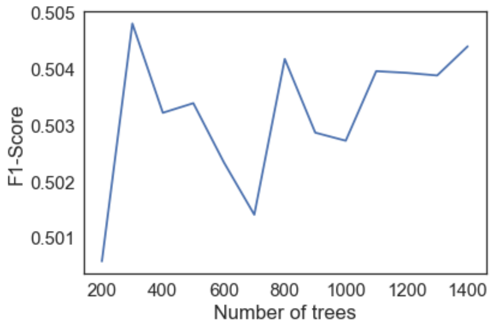


(b)

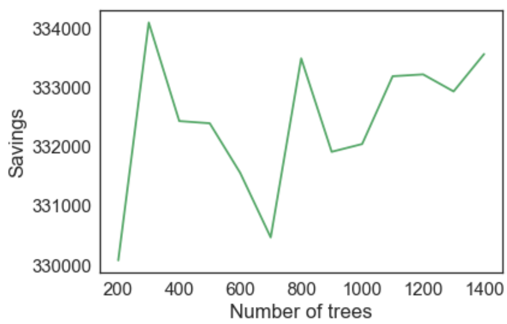


(c)

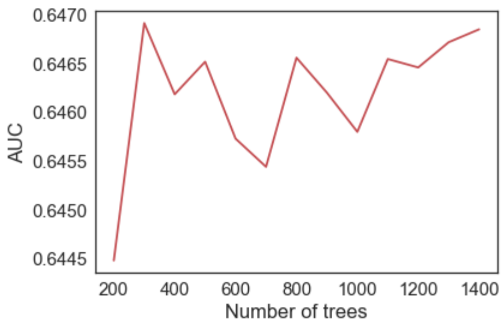
Figure B.2: Exhaustive search for optimum size of a leaf.



(a)

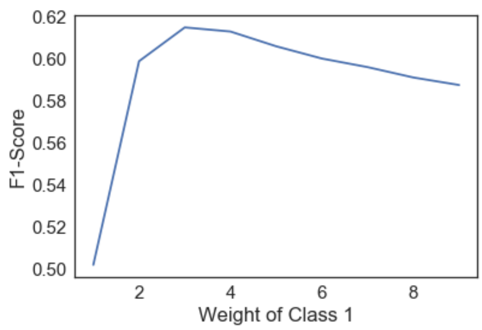


(b)

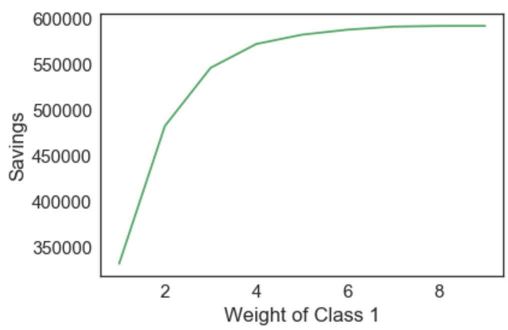


(c)

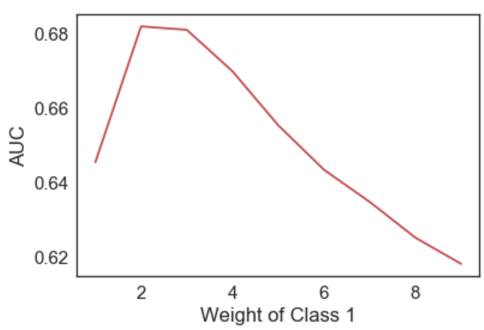
Figure B.3: Exhaustive search for optimized number of estimators.



(a)



(b)



(c)

Figure B.4: Exhaustive search for optimum weight of Class 1.