Enhancing the packaging allocation in a luxury fashion online marketplace

Martinho Sarmento e Cunha Abrunhosa de Brito

Master's Dissertation

Supervisor: Prof. José Luís Moura Borges

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Integrated Master in Mechanical Engineering

2018-07-26

Abstract

One of the major challenges of a marketplace industry is the lack of control. From the time a customer checks out on the platform until it receives the product, there are several crucial steps, during the order fulfilment process, performed by partners and carriers. The minimum quality of these processes should be contractually fixed and if a better quality adds value to the client, there should be established incentives and penalties. This should be conducted as an iterative approach as, with time, these relationships become stronger and customers' demands increase.

The problem addressed in this dissertation is one of these steps: choose the box for each product delivered in a marketplace company - Farfetch. Before this project, partners had total freedom in this process. This caused unpredictable and excessive packaging volumes being shipped from partners and Farfetch did not have a proper way to control them. This effect is particularly important during peak season when, not rarely, partners neglect packaging replenishment due to the amount of work. This is highly impacting in terms of sustainability, customer experience and shipping costs (which are not charged to partners).

A project team was set up with the main stakeholders and this work goes across all the phases any project should go through. The solution proposed was to benefit partners who use the fittest box to deliver products. The incentive created should be given depending on a packaging *KPI* which compares partners' used package and Farfetch suggestion for every order. The most accurate way to know which box was used on each order is using the carrier invoices and, with this information, processes were developed to build the incentive. The support to partners was also increased during this change with the creation of a channel which partners can use to alert Farfetch of wrong packaging recommendations. With the launch of the packaging incentive, the average volume shipped had decreased, comparing with last year's results, revealing that the project is showing some good results.

On the other side, the suggestion provided by the e-commerce company is not the best and some development must be made to maintain a good partner evaluation in this scope. The first step was to give visibility to the team responsible for these suggestions and try to establish guidelines to enhance them. The second step was to develop an algorithm which predicts the best box for a certain item given its characteristics. This prediction can help the team responsible for this suggestion or easily identify bad suggestions. The final algorithm is a machine learning algorithm that predicts boxes based on the smallest box used from past partners' decisions. It had substantially increased the knowledge about boxes, given that, this is the first time the company has this visibility.

Although the accuracy obtained is satisfactory, it is expected that, with some specific products' information (not available at the moment), the results would be more accurate. Apart from that, there is still work to do on attracting big partners which are still not very committed and to continue engaging medium and small partners. One of the most important points is reacting faster when a suggestion is not good.

Resumo

Dos maiores desafios de um *marketplace online* é a falta de controlo. Desde que o cliente faz a compra no *site* até receber o produto, grande parte dos passos considerados cruciais no processo de preparação e envio da encomenda são da responsabilidade de parceiros e transportadores. A mínima qualidade destes processos deve ser contratualmente estipulada, no entanto, se uma maior qualidade trouxer maior valor para o cliente, os responsáveis devem ser beneficiados ou penalizados em função desta. Idealmente, esta abordagem terá de ser iterativa uma vez que, com o avançar do tempo, os clientes tornam-se mais exigentes e relações com terceiros reforçam-se, tornando mais fácil colocar a fasquia mais alta.

O problema trabalhado ao longo deste projeto é um desses passos: escolher a caixa para cada produto enviado através de um *marketplace* - Farfetch. Antes deste projeto, os parceiros tinham total liberdade nesta seleção. Esta condição causava volumes de caixas imprevisíveis e excessivamente grandes para cada produto e a Farfetch não tinha forma de controlá-la. Este problema poderá implicar consequências em sustentabilidade, experiência do cliente e custos de expedição (que não são cobrados aos parceiros).

Uma equipa foi concebida com os maiores interessados na resolução deste problema. O presente trabalho acompanha todas as fases pelas quais passou este projeto até à sua conclusão. A solução proposta foi beneficiar os parceiros que usam a caixa mais pequena possível que não prejudicasse o produto ou a experiência do cliente. O incentivo criado deve ser dado dependente de um *KPI* que, para cada encomenda, compara a caixa que os parceiros usaram com a que a Farfetch sugere. Recorrendo às faturas do transportador é possível induzir a caixa usada e todos os processos do incentivo se baseiam nesta informação. Foram desenvolvidas formas específicas de ajudar os parceiros no processo de mudança, tais como um canal através do qual é possível dar feedback sobre más sugestões. Com o lançamento do incentivo, o volume médio enviado diminuiu comparando com os resultados do ano anterior, revelando que o projeto apresentou resultados positivos. Por outro lado, a sugestão que a empresa de comércio digital faz não é a melhor e algum desenvolvimento foi requerido para manter uma correta avaliação dos parceiros neste âmbito. O primeiro passo foi dar visibilidade à equipa responsável por fazer estas recomendações de caixas e tentar estabelecer algumas linhas orientadoras para as melhorar. O segundo passo foi desenvolver um algoritmo que prevê a melhor caixa para um produto dadas as suas caraterísticas. Esta previsão ajudará a supramencionada equipa no processo de decisão ou a facilmente identificar más sugestões. O algoritmo final é um algoritmo de inteligência artificial que prevê a melhor caixa baseada nas decisões anteriores de parceiros. Este ajudou a aumentar substancialmente o conhecimento sobre a relação entre produtos e caixas, uma vez que pela primeira vez, há alguma visibilidade sobre este assunto.

Apesar dos valores de precisão serem satisfatórios, é expectável que informações específicas sobre o produto (que não são recolhidas pela empresa) melhorem os resultados deste modelo. Além disso, ainda há algumas oportunidades de melhoria na atração de grandes parceiros, bem como, continuar a envolver boutiques pequenas e médias. Um dos pontos mais importantes para conseguir este objectivo será reagir cada vez mais rapidamente a más sugestões.

Acknowledgements

First of all, I would like to thank Farfetch for the opportunity of developing this project. To Carlos Gomes, my supervisor at this company for all the guidance and help provided during these months. To Inês Carvalho for the analytical thinking transmitted and to Fábio Moreira for showing me the direction to follow. To all my colleagues in Operations Strategy team for their friendship and support during this work. And, finally to everyone else at Farfetch who, somehow, helped me in this dissertation.

I would also like to appreciate my dissertation supervisor, Prof. José Luís Moura Borges whose feedback was invaluable for the completion of this document, for the constant availability and for his help.

Finally, but not less important, I would like to thank my parents and sisters whose support and strength was crucial to conclude this phase of my life. To my friends, for the unconditional help during this dissertation. And to Catarina for her positiveness and joy who encouraged me to make the best decisions.

Duke Wittenborn in his book *Fierce People* states that "We are the sum of all people we have ever met; you change the tribe and the tribe changes you.". Inspired by this quote, I would also want to thank everyone who somehow helped me not only during this dissertation nor during this five years of the Mechanical Engineering course but influenced me to become who I am today.

"You wouldn't abandon the ship in a storm just because you couldn't control the wind."

Thomas More

Contents

Acronyms

3PL	Third-Party Logistics
ATV	Actual Transaction Value
BAU	Business As Usual
BFC	Brand Family Category
CART	Classification and Regression Tree
CS	Customer Service
CV	Cross-Validation
ERP	Enterprise Resource Planning
ET	Extra Trees
GMV	Gross Merchandise Value
ISE	In Sample Error
KPI	Key Performance Indicator
LM	Linear Model
OSE	Out of Sample Error
PMBOK	Project Management Book Of Knowledge
PMI	Project Management Institute
PMO	Project Management Office
PS	Partner Service
RF	Random Forest
RoW	Rest of the World
SoS	Speed of Sending
USA	United States of America

List of Figures

List of Tables

Chapter 1

Introduction

This work was developed at Farfetch, a London-based luxury fashion online marketplace founded by a Portuguese economist in 2008. The company offers the possibility of selling luxury items from more than 900 different partners across the globe in a single website visited by more than 21 million people per month, according to Williams-Grut (2017). Most of this company's partners are boutiques - stores which sell clothes and accessories - however, brands and department stores are also starting partnering with Farfetch. Fraser (2016) states that this entrepreneurial venture, founded as a start-up, is valued today at more than one billion dollars - the first unicorn[1] coming from Portugal. Nowadays, it has more than 2000 employees in 11 different cities and is growing 70% every year, reaching over one billion pounds in sales last year (Williams-Grut, 2017).

Farfetch gathers luxury fashion products from different stores on a single online platform - a marketplace. It allows customers to buy products online and ensures their delivery. A customer pays a given price which includes product's price, Farfetch margin and a shipping fee which will cover shipping costs. The service of delivering orders to clients is outsourced to carriers which charge Farfetch for this service. It is Farfetch responsibility to manage all the information flow between this three entities: customer, partner and carrier.

On the one hand, as investigated by Armstrong (2017) and C.S.-W. (2013), for most of the partners (boutiques and brands), setting up an online retail website represents a big investment which is not worthy. Therefore, partnering with Farfetch appears as a very good opportunity to reinforce their market position, enabling to reach new clients without harming their bricks-and-mortar identity and without risking too much. This is a particularly charming deal for many small boutiques which do not have a sufficiently large client base, to be in a global storefront. It is also interesting that big stores and brands are recently betting on this partnership as an opportunity to expand their digital services, as it is shown by Armstrong (2018), White and Denis (2018) and Fernardez (2017). On the other hand, the e-commerce company allows high-end clients to buy luxury goods from wherever they are and receive them in any location.

[1]Unicorn is a term created by Aileen Lee and used in the finance world to call private startup companies valued at over one billion dollars. For more information about this nomenclature please refer to Rodriguez (2015)

There are other players in the market with a similar business, however, Farfetch is in a strategic position to be the number one in the world as it stands out by the number of products as well as their diversity. If from one side this e-commerce company has the most valuable shelf on the market, from the other it also has the biggest number of different brands, as represented in Figure 1.1. Additionally, being a marketplace distinguishes this company from other luxury fashion online retailers and turn it in a well-positioned competitor because it does not hold any stock or any transportation system. Everything sold on its web platform is stored in the partners' stock points. Thus, when an item is ordered, the product is prepared by the partner and is picked up by a 3rd party carrier who delivers to the final customer.

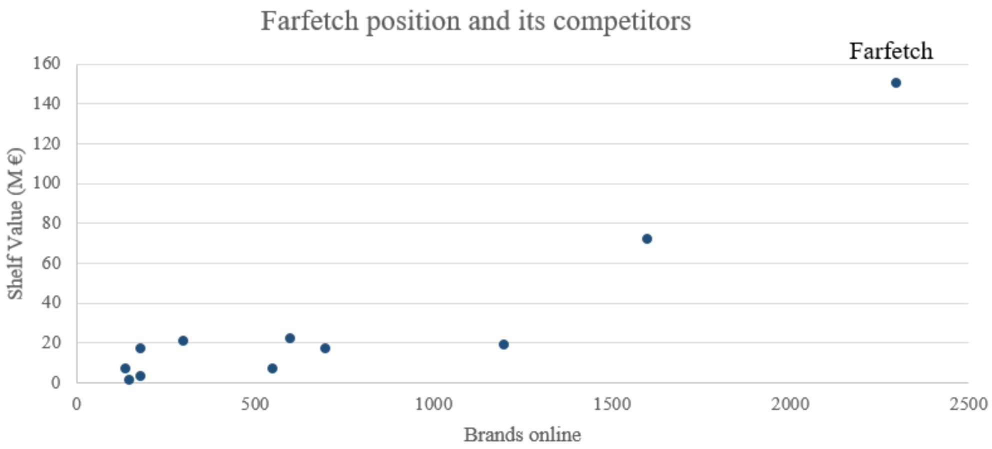

Figure 1.1: Farfetch position compared to other competitors

1.1 Project

After contextualising where the project took place, we should understand that every project is temporary in nature. As described in the diagram below, the team responsible for the project should create a specific product or service which will change the Business As Usual (BAU) operations into new and improved BAU process (see Figure 1.2).

When starting a project, it is important to answer some questions: when will it start and finish, who is interested and will execute it, what is the problem the project wants to tackle and which are the key goals. Although a very straightforward answer is expected to explain the key objectives, it might require a small framing of its context and why there is an opportunity.

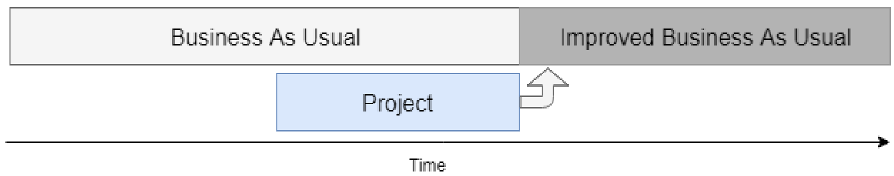

Figure 1.2: Project and Business as usual (BAU) relationship through time.

1.1.1 Problem

One of the greater opportunities in every e-commerce business area is related to decreasing delivery costs. Packaging selection has a big impact both on shipping costs and on customer experience. Therefore, it is very important to ensure control over the packaging selection and usage processes. Unlike other companies where such control is easily performed, since Farfetch is not responsible for preparing the order it does not control the stage of the order fulfilment where the packages are selected.

The partners are responsible for storing, folding and packaging the product, while Farfetch does not even see the product shipped on any step of the process. Farfetch offers a range of 9 different sizes of packages given the diversity of products. These packages are simple cardboard boxes which are used to physically protect the items from external factors, to make it easier to transport the product and to store the invoice or additional documentation. Moreover, as the first physical contact with Farfetch, it can be used as a marketing strategy as it will impact the unboxing experience. However, partners order these packages to Farfetch according to their needs and they pay a fixed price for each one, independent of their volume or its real cost to Farfetch (who connects directly with the packaging manufacturer). This means that, although Farfetch recommends a size for each product, there is not any incentive or awareness to reduce the amount of air shipped that results from delivering products with excessively big packages

1.1.2 Project goals

After this global perception of the project, it is now possible to fully define three goals this project aims to achieve.

- Map every packaging related process;
- Leverage the current processes and data to decrease delivery costs;
- Improve packaging supply chain sustainability and improve the customer experience through packaging.

1.1.3 Project owner and stakeholders

It is also important to describe the teams involved in the project to easily explain who will compose the project team and who are its main stakeholders.

The Operations Department is the section of the company responsible for ensuring that the service performed by Farfetch reaches the customer the best and fastest way possible. It is accountable for planning, managing, coordinating and controlling all daily activities related to the services provided by this company according to Sanders (2013). The Operations Strategy team is the one that acts all across this department, it is mainly composed of project managers with analytic skills who are responsible for every project involving more than one team inside Operations. To get a high-level overview, Farfetch's organisational chart is shown in Figure 1.3.

The current project was proposed by the Operations Strategy team with the main focus of finding a way to reduce shipping costs and improve customer experience. Apart from this team, there are two others which were directly integrated into the project: Delivery Development and Supply & Retail Operations (see Figure 1.3). The Delivery Development team is responsible for every concern related to the strategic side of delivering the products to the customers. Since this project can substantially control and reduce shipping costs, they are the top interested in this work. The Supply & Retail Operations team deals with packaging suppliers, boxes orders and communicates with the 3PL responsible for delivering them to the partners. As every project, there are stakeholders who compose the group of people inside the organisation which will or can be affected by a project. Apart from these two teams, as it will be further justified, there are other parties which should be involved in specific steps: Partner Service (communication with partners) and Production (product creation in the platform).

A project team was established with members from the three main teams involved: Operations Strategy, Delivery Development and Supply & Retail Operations. In the Figure 1.3 stakeholders, project team members and project managers are highlighted.

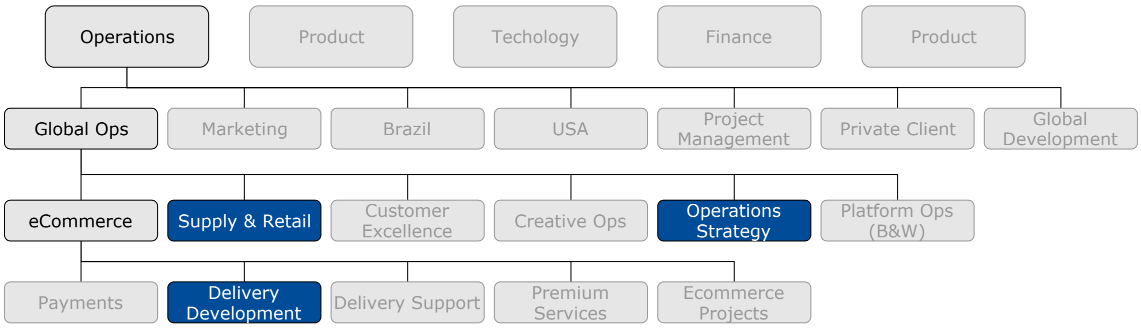

Figure 1.3: Summarised company's organisation chart with highlighted project members

1.1.4 Motivation

To assess the full potential of this project, estimations were made to evaluate the investment on the minimisation of the package volumes shipped. Assuming that 60% of partners follow Farfetch recommendations on 85% of the transactions they receive by the end of 2018 (in January 2018 just 20% of the partners accomplished), it is expected a return of 10 times the investment needed on an incentive to these partners. These estimations will be further explained in the section 4. Nonetheless, due to the nature of the online business, it's critical to monitor other KPIs which describe shipping costs and client experience. The average weight and the partners' accuracy on choosing the recommended box would be used to evaluate the costs, the experience can be evaluated using packaging rating given by customers.

1.2 Methodologies

To guide the work presented, a methodology based on the guidelines for project management within Project Management Body of Knowledge (PMBOK) was used. In this book, the American nonprofit professional organisation, Project Management Institute (PMI) developed a set of standards referring to the five process groups of project management: initiating, planning, executing, controlling, and closing. It is important to highlight that, in real life projects, it is crucial to adopt a circular or iterative approach. For instance, when a project is in the executing phase, it is important to go back to planning phase, since the knowledge about that specific area will increase and, with this knowledge, a better outcome will be certainly achieved (Project Management Institute, 2017). Although the specified sequence was not always rigorously followed, the current work will be structured going throughout these five groups of processes as phases. Inside each group, we will follow a step by step approach delineating all the work done during each one of the phases.

During the project initialisation and planning, different spheres of activity were identified:

- Refine partners' box usage:
  - Incentive plan for partners to use the best box;
  - Exception process for partners to feedback;
  - Increase control and support partners.
- Improve company's box suggestion:
  - Ongoing feedback and continuous improvement of Production Suggestion;
  - Develop Best Box Algorithm based on feedback from Partners.

Due to the nature and complexity of the work, the last point will be exhaustively described in this work. Thus, this dissertation will go through an in-depth and comprehensive review of the machine learning algorithms which will use historical data to predict the best box for a given product.

1.3 Thesis outline

This dissertation is organised in six chapters whose outline is the following:

Chapter 1 - Introduction of the present work, starting with the scope of this dissertation, going through its motivation, goals and how is the high-level structure used to achieve them.

Chapter 2 - Theoretical background to scientifically support the decisions and assumptions made during the thesis. This chapter will be divided into three different frameworks: luxury, project management and machine learning. Luxury will cover what is different from a traditional retailer and a luxury retailer as well as its conjunction with e-tail; Inside project management, we will cover the methodology and guidelines used to manage the present work; Machine learning section will briefly cover the important concepts to support what will be developed in this area.

Chapter 3 - Detailed description of how packaging related processes inside or outside the company work.

Chapter 4 - Going through how the work during the project life cycle was structured. Initiating on what motivated the development of the project and justifying why the project should be done. Afterwards, describing in detail the plan aligned to fulfil the objective and executing that plan. In this part, it is exhaustively presented the deliverables and changes made by the project team. A section about proper control and monitor is also mandatory to be sure the project keeps on its tracks. And lastly, the process of closing a project. A new section was added to describe the algorithm developed to predict the ideal box due to its complexity and importance to the project.

Chapter 5 - Final results to give an overview of the effects of the project, as well as, the obtained performances of the used models.

Chapter 6 - Conclusion of the current thesis summarising the main improvements achieved and the main knowledge acquired which opens doors to new projects.

Chapter 2

Theoretical background

This chapter summarises the most important concepts related to this project which were considered indispensable both to understand the world this work takes place and to theoretically support the work developed. Firstly, a brief conceptualization of luxury, e-commerce and its unique customers. Secondly, the project management methodology concepts used during this work who helped structure the approach. Lastly, the theoretical background required for understanding the classification model designed during the execution of the project.

2.1 Environment

2.1.1 Luxury

Coming from the Latin word *luxuria* or *luxus* which means "excess", luxury is described as "A state of great comfort or elegance, especially when involving great expense" according to the Oxford English Dictionary. Unanimously considered a concept hard to define, Vigneron and Johnson (2017) highlight the dependency on the recognition of value by others, dragging it into a totally subjective concept. A specific good could be considered a normal or a luxury product for two different people or even for the same person in two different situations. The famous fashion designer Coco Channel claimed "[Luxury] is the opposite of vulgarity" and Okonkwo (2009) goes even further defining it as "neither a product, an object, a service nor is it a concept or lifestyle. It is an identity, a philosophy and a culture.".

For a brand to be considered a luxury product seller, Vigneron and Johnson (2017) define five perceived dimensions: conspicuousness, uniqueness, quality, extended self and hedonism. The first three will lead to products which are easily noticed with a great focus on its aesthetics to the smallest detail, as exclusive as possible and made with the best materials and techniques which will automatically raise its price. The last two are personal dimensions and not directly related to the products. Dubois et al. (2001) also adds two important characteristics: history and superfluousness. Goods with some sort of tradition or personal history behind them like portraits or jewellery from a noble family are considered luxurious. On the other side, in order to show that a good is luxurious, it cannot be correlated with a necessity. Even a luxury product which will satisfy a need like food (i.e. caviar) or clothing (i.e. cashmere sweater) must stand themselves out from others through other characteristics like unique flavour and comfort, for example.

2.1.2 Luxury customers

After understanding the product it is also important to get to know the customers' expectations and demands. Okonkwo (2009) recognises that luxury retail is supported by an ancient and fundamental human need of distinction, admiration, recognition, appreciation and respect from others through his possessions. Holt (1995) also adds that it is important for these customers to identify with the wells they acquire.

All high-end customers are different and they are expecting a personalised treatment, however, it is possible to highlight three predominant characteristics from *Delloite*'s report (Arienti, 2017): perfectionism, high self-esteem and search for exclusivity. They are willing to pay more for a product they believe it is unique or it of greater quality. In contrast, it has interestingly been proved by Rao and Monroe (1989) that consumers associate higher prices with better quality even if it isn't the case. Given that one of the reasons for buying a luxury product is to draw the attention, the same happens before acquiring the product, during the buying experience. This way, it is very important for the client to feel important for the brand or for the boutique.

2.1.3 Luxury e-commerce

Luxury brands and internet are in many aspects contradictory. Exclusivity, emotions and desire are not particularly easy to create on the classless Internet world and probably that's why big luxury brands tend to resist to create e-commerce channels (Okonkwo, 2009). There are many disadvantages when buying online such as the non-existence of physical contact with the goods and a non-human experience with the brand. However, these drawbacks do not play a very important role when compared to the advantages: the fastness, the convenience, the availability anywhere and anytime and the lack of pressure and commitment when buying (Dauriz et al., 2014; Harris and Dennis, 2002).

Even that traditional luxury brands tend to slowly open e-commerce channels, there are some companies which started to bet on this niche and are growing fast. These companies are using three things as an advantage: digital resources to be flexible, big data and advanced analytics such as machine learning to recreate a more human and realistic experience and strategic partnerships between companies to enhance customer experience (Achille et al., 2018).

According to the same *McKinsey* report, by 2025, online luxury sales are expected to be more than the triple compared with 2016 sales, reaching a total of 74 million dollars (19% of all personal luxury sales). At that time, 40% of the customers will be millennials[1] according to D'Arpizio and Levato (2017).

[1]Millennials are defined as people born between 1980 and 1995

2.2 Project management methodology

"Project management is the application of knowledge, skills, tools, and techniques to a broad range of activities in order to meet the requirements of a particular project." This is the definition the "*A Guide to Project Management Book Of Knowledge* (PMBOK)" gives to project management which had been the book followed to manage this endeavour. The PMBOK Guide is the result of the work performed by the US nonprofit organisation for project management - Project Management Institute - and provides standards and guidelines to help managing individual projects.

The certified guide explains how to go through the different phases of the project. The uncertain nature and the potential changes can turn the balance between constraints such as budget, quality and schedule a difficult task as they strongly depend on each other. Such volatile environment doesn't mean that planning shouldn't be considered but, on the contrary, this action should be strongly performed, as an iterative activity with project's execution, throughout the project's lifecycle.

Structuring our approach, we can divide all projects' processes into five different groups: conception & initialisation, definition & planning, launch & execution, performance & control and close. As it is shown in Figure 2.1, the PMBOK Guide by Project Management Institute (2017) reinforces the overlap between different groups. On the chapter 4, the project will be divided into these five moments.

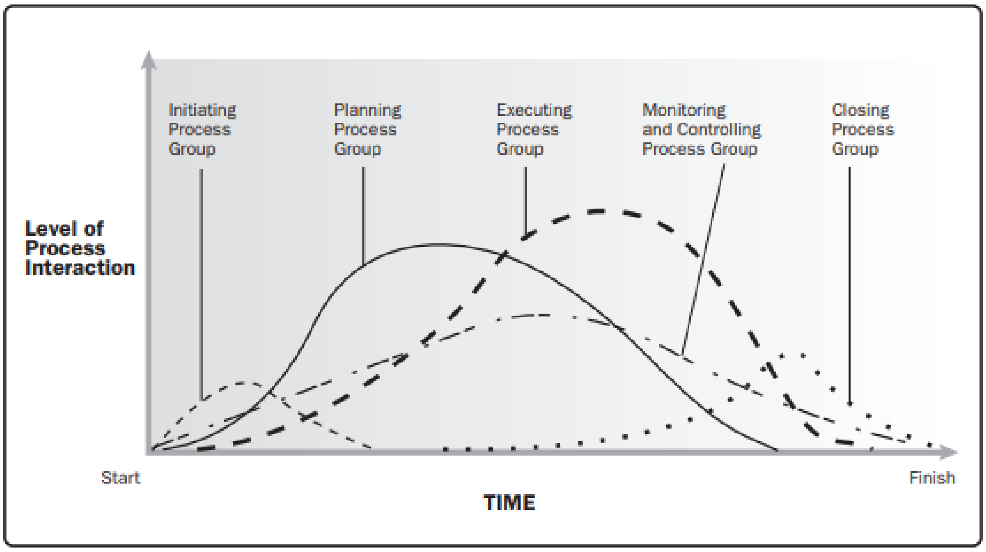

Figure 2.1: Level of process interaction during project's life cycle. Source: Project Management Institute (2017)

2.2.1 Initialisation Processes Group

Before a project or a phase starts it is necessary to obtain permission to go on with it. The purpose of this first Process Group is to adjust stakeholders' expectations with the project's goals and both make aware and motivate them about their roles during the project's next steps. Materialising these purposes it is important to identify internal and external stakeholders who will directly influence or be influenced by the result, the project team, the scope and necessary investment and expected outcome.

At the end of these first definitions, in order to increase stakeholders satisfaction and commitment, they should be asked to decide the project's feasibility and evaluate if it is worth going on or not. To get board's approval (or superiors', depending on the dimension), a business case should be developed to sum up all the information gathered. It should contain clear descriptions of the objectives, requirements, a forecast of resources needs (including time and budget) and stakeholders.

2.2.2 Planning Processes Group

After obtaining permission, a plan should be elaborated to guide everyone through the development of the project. Defining the strategy, tactics and the course of action will align everybody to successfully complete the project.

Usually, the first phase consists in the definition of SMART and CLEAR objectives. According to Doran (1981), SMART stands for Specific, Measurable, Attainable, Realistic and Time-bound whereas CLEAR are Collaborative, Limited, Emotional and Refinable goals in accordance with Adam Kreek cited in the article published by Economy (2015) in the magazine *Inc.*. It's also during this phase that a project management plan is developed. Based on the scope, schedule, and budget, a set of baselines or performance measures should be established in order to enable to assess how the project is going on. Roles and Responsibilities (R&R) are defined, so everyone involved knows what they are accountable for. To get a visual evidence of how the project will develop it is common to draw a Gantt Chart representing the main tasks and milestones.

2.2.3 Execution Processes Group

The execution group is when things are really performed, from paper to actions. Deliverables previously defined are developed and completed, therefore, it is at this time that a major part of the budget and project's time are spent. A kick-off meeting usually marks the start of the project execution phase where members involved are informed and discuss their responsibilities.

The aim here is to coordinate and assign people and resources, carry out with the project management plan previously defined and periodically discuss the status. It is important to reinforce that flexibility is crucial to achieving better results. Plan, execute and control should be integrated into an iterative loop.

2.2.4 Control Processes Group

As its name suggests, control processes measure the progress of the project to ensure everything is going as planned. The performance tracking must be done using key performance indicators - *KPIs* - which were previously aligned during the planning phase. It is considered a well-balanced decision to pick between two and five KPIs which should be able to:

- Measure how well project objectives are being achieved with an eye on resources and time management;
- Determine if tasks' deliverables are being done with the expected quality;
- Find if the project will be completed by the defined deadline and with the budget previously defined;
- Understand how changes and obstacles are faced during the realisation of the project and how fast they are addressed.

It is also necessary to create processes to identify which work-streams are not going in the right direction and decide to replan and recommend proactive or corrective actions to mitigate bad results.

2.2.5 Closure Processes Group

As a project is a temporary enterprise, there should be a time when you define an end. This includes all tasks needed to formally complete the project or phase. The project team should be dissolved, outsourced contracts done for this specific project are closed and valuable team members are recognised. A *post mortem* meeting can be held to evaluate what went wrong and right during the previous phases and therefore, learn from mistakes and good moves to retain, recording them in a lessons-learned document.

Project managers should make sure documents and deliverables are stored in the same place to make it easy to keep track of what was done. Obtaining acceptance by the sponsor to formally close the project will end up this process.

2.3 Machine Learning Algorithms

The inspiration behind this section relies on three machine learning books: *Pattern Recognition and Machine Learning* by Bishop (2006), *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner* by Kotu and Deshpande (2014), *Pattern Classification* by Duda et al. (2012). Therefore, everything not cited will be from this bibliography.

Predictive analytics is the branch of analytics in which historical records are used to make a prediction about an uncertain future. This includes a wide variety of different techniques such as data mining and machine learning, which are being carried by the big data revolution we are facing in the last years (McAfee et al., 2012). Machine learning grew out of computer science with the intention to be used as a tool to find patterns in big amounts of data.

It is possible to divide machine learning problems in three. In supervised learning, a model tries to learn patterns from examples, on the other side, in unsupervised learning it tries to discover underlying patterns in data. In reinforcement learning, the model interacts with an environment and receives rewards in case it behaves the right way.

Mathematically speaking, on supervised learning, the problem gives us the inputs (X) and the outputs (Y). The objective is to learn and optimise the black box function (f) between them as shown in the equation 2.1.

$$Y = f(X) \qquad X = \{x_1, x_2, ...x_n\} \tag{2.1}$$

In fact, using the inputs and outputs, an algorithm is trained to map a model h_θ which will be an approximation of the target function f. As in equation 2.2, this hypothesis will predict the target variables $\hat{Y}$ which are a representation of the reality.

$$\hat{Y} = h_\theta(X) \qquad X = \{x_1, x_2, ...x_n\} \tag{2.2}$$

Usually, real events are too complex to replicate so, this approximation is usually a simplification of the reality f. This simplification is achieved by making some assumptions of its form, for instance, the form represented on equation 2.3.

$$h_\theta(X) = \theta_0 + \sum_{i=1}^{n} \theta_i x_i \tag{2.3}$$

These underlying assumptions depend from algorithm to algorithm. Regardless of the amount of data and its quality, there is no single model that has a better performance for every problem. On real-world problems with real-world data, an error (e) must be taken into account when trying to predict. There are a lot of metrics to evaluate this error, however the most simple for classification problems is to directly compare the real output (Y) and the predicted output by the model ($\hat{Y}$) as it is demonstrated on equation 2.4.

$$e = \frac{COUNT(Y = \hat{Y})}{COUNT(Y)} \times 100(\%) \tag{2.4}$$

Most of this kind of problems can be divided into either classification and numeric prediction problems. In both cases, a model is supposed to predict an output based on historical data given an input or a group of inputs. Numeric prediction problems are those whose target variable is a continuous variable and the model is a function (e.g., Karandish and Šimůnek, 2016; Shahinfar and Kahn, 2018); Classification have discrete targets often labelled as classes and pattern recognition is used to predict data behaviour (e.g., Choi et al., 2015; Milosevic et al., 2017).

2.3.1 Evaluating

On any supervised machine learning algorithm, there are two types of errors: the in-sample error (*ISE*) and the out-of-sample error (*OSE*). Both errors are obtained when comparing the target variable predicted by the model with the real output. The difference remains on which data you use for each validation. The ISE or training error uses the same data used to train the model - *training set*-, the OSE or test error uses the unseen data by the model - *test set*. To evaluate and validate a model it is mandatory to use OSE because models can memorise training sets - this phenomenon is called *overfitting*. The ability of correctly categorising new sets of unseen data that are different from those used during the training phase is known as *generalisation*.

The generalisation error (OSE) could be decomposed into bias and variance. When a model has high variance it is usually because of overfitting, when it has a high bias, it is because of underfitting. The most common and simple analogy to these concepts is to compare them to darts. The model throws darts ($\hat{Y}$) and tries to hit the centre Y. The Figure 2.2 shows the relationship between bias and variance using this analogy.

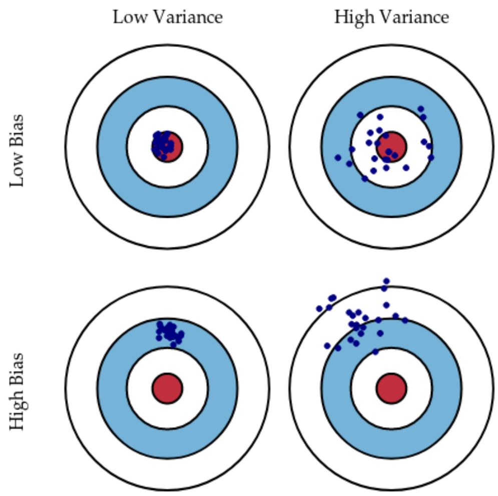

Figure 2.2: Dartboard showing high and low bias and variance. Source: Fortmann-Roe (2012)

Too simplistic assumptions and lack of malleability are a source of high bias. A model is learning the same wrong thing, therefore the predicted variables $\hat{Y}$ are pointing to a wrong direction. Bias is calculated as shown in the equation 2.5. To avoid biased models, you should add complexity to make it more flexible and adaptable.

$$Bias(\hat{Y}) = E[\hat{Y} - Y] \tag{2.5}$$

Too complex assumptions are a source of high variance. Extremely flexible models overfit the training data including the noise real-world data contain and end up memorising them. The mathematical definition of variance is on the equation 2.6 Decreasing the parameters will reduce variance from the model.

$$Variance(\hat{Y}) = E[(\hat{Y}(i) - E[\hat{Y}(i)])^2] \tag{2.6}$$

We end up in a tradeoff between complexity and simplicity, bias and variance. Figure 2.3 illustrates the point where both curves intersect is the optimal model. In theory, the perfect level of complexity is reached when the increase in bias is the same as the reduction in variance. Because of the random noise, in real problems, this point cannot be analytically found.

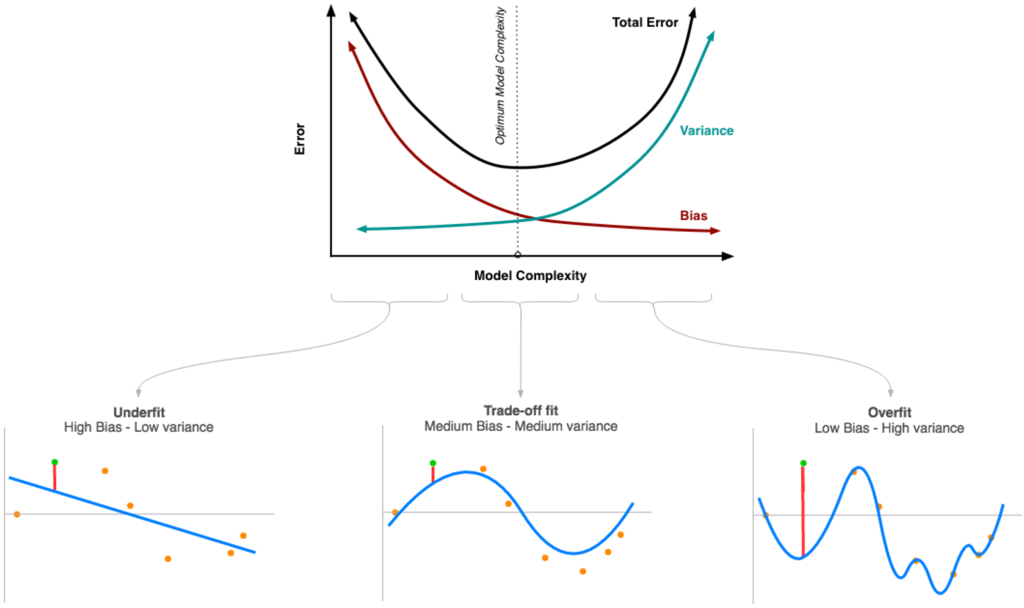

Figure 2.3: Relationship between model's error and complexity. Source: Fortmann-Roe (2012)

2.3.2 Validating

Given that data is not infinite, it is necessary to save a portion of the data to evaluate the OSE of the model. Splitting the available data on a training and a test set is the most simple split technique. With the development of algorithms, two more sophisticated techniques were invented and they will be further explained.

Train validation test split During the conception of the best algorithm to predict a target variable, there are many parameters that should be tuned until reaching the optimised result. To avoid the usage of the test set during this iterative process, it is common to make a split in three different sets: *train*, *validation* and *test*, as shown in Figure 2.4. Additionally, this method has another benefit, it can be used to double check if the model is overfitted. The graph of the Figure 2.5 shows the training (ISE) and validation (OSE) errors depending on the complexity of the model and their relationship with high bias and high variance

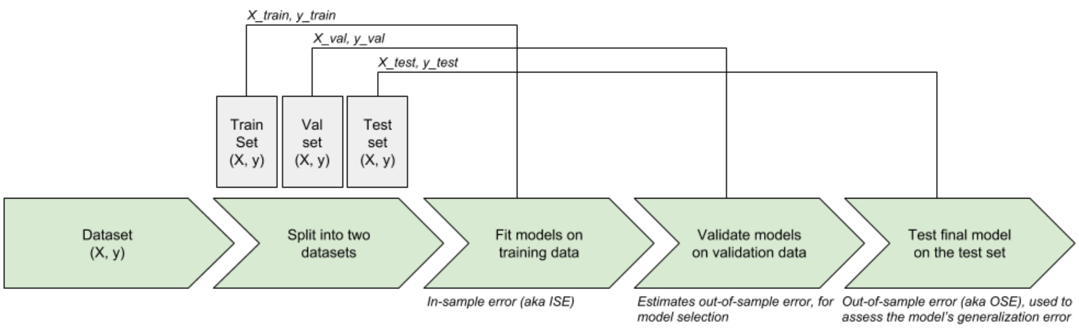

Figure 2.4: Pipeline of the method train validation test split

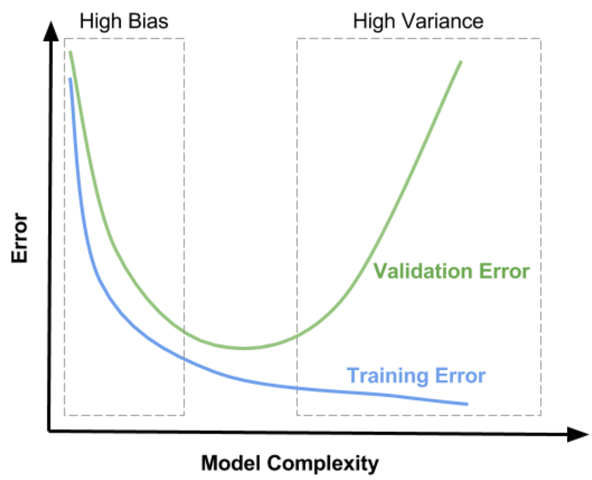

Figure 2.5: Relationship of train and validation error with model complexity

K-Fold cross validation Depending on how we split the data, test error results can be subject to great variability, especially for smaller datasets. Also, and quite obviously, holding out more data reduces the amount available for training, possibly leading us to overestimate the test error. To avoid this, another technique can be used - cross-validation.

In k-fold cross-validation, the data sample is randomly divided into k equally sized subsamples. At each iteration, one of these subsamples is left out, the model is trained with the other $k-1$ subsets and the OSE is obtained with the left out subsample. This process is repeated k times with the other parts as represented in Figure 2.6. In the end, a mean error and its variance are obtained, representing an estimation of what would be the *OSE*. Accordingly to Kohavi (1995), these estimations are reasonably good at $k = 10$ folds and at they are almost unbiased at $k = 20$.

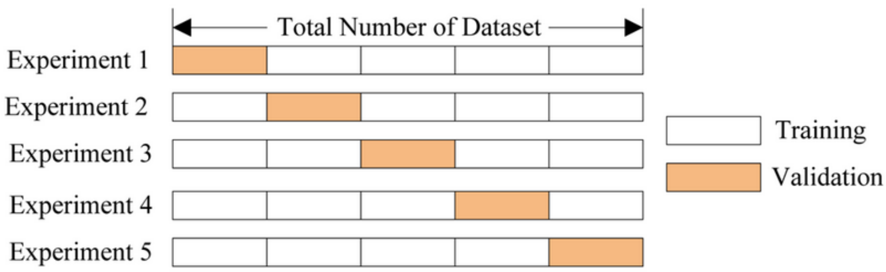

Figure 2.6: 5-Fold Cross Validation process. Source: Zhang et al. (2013)

2.3.3 Classification models

Algorithms are the group of procedures which run over the data to build the hypothesis h_θ. The model is this final hypothesis, which based on predictors (independent variables) predicts a number or a class (dependent variable). There are lots of algorithms but it is possible to divide them into five different schools of thought: Evolutionists, Connectionists, Symbolists, Bayesians and Analogists. For evolutionists, the natural selection is the mother of all learning; for connectionists, learning is the connections our cerebrum does; symbolists believe all knowledge can be achieved by manipulating symbol like mathematics; bayesians state that it is all about uncertainty and probabilities; for analogists it is necessary to find patterns in data to learn. We will cover the mathematics behind some algorithms to understand why they have different results. The algorithms are all from the symbolist school of thought and were the ones which revealed a better performance in the project.

Decision Tree Decision trees (Breiman et al., 1984) are one of the most simple and intuitive classification prediction techniques. It takes the form of an inverted tree or a pyramid, in each node a binary decision must be made until it reaches a leaf node where a target variable is predicted as shown in Figure 2.7. The mathematics behind this algorithm is explained, in detail, in the appendix A.

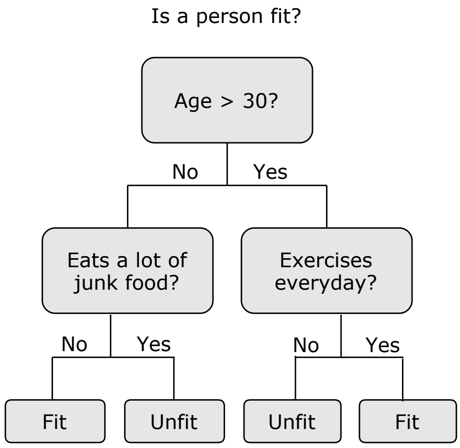

Figure 2.7: Decision tree diagram example

Random Forest There is a specific area of machine learning which focus on ensembled models. These models train different independent models on the same data, ask them which are their predictions and makes a decision based on their votes.

As a real forest, random forest (RF) is composed of a group of trees. This is one of the most simple ensemble models, like a democracy, a group of decision trees vote, with equal weights, on an output. In the end, the most voted class is the final prediction. Two techniques are used here to make decision tree different from each other and avoid overfitted models: bootstrap aggregating or bagging and a random subset of features. Bagging is done by selecting a random sample with replacement (the same record can appear more than once) of the data set and fitting different trees to these samples. A random subset of inputs is selected for each tree increasing the randomness and preventing models that perform poorly in unseen data (Ho, 2002).

Extra Trees Gashler et al. (2008) state that it is better to have ensemble models with less but heterogeneous elements rather than more and homogeneous. Based on this idea, Geurts et al. (2006) propose a new tree-based ensemble algorithm: Extremely Randomized Trees or simply Extra Trees (ET). The only difference between this and RF is that in this algorithm while splitting a tree node, it randomises strongly both attribute and cut-point choice. The final result is a random forest with very different trees. The algorithm performs well both on accuracy and on time, the computational efficiency is pretty good.

2.3.4 Feature engineering and selection

There is another technique that can lead to fewer errors and better computational performance. The idea is to modify the features or inputs the dataset to ease the work of the algorithm to find patterns and create a model.

Feature engineering is the process of creating better features using the knowledge of the area and the already existing features. This data treatment technique helps both models to take conclusions easier and user and researchers to interpret data. Feature selection is the process of eliminating features that do not contribute to the accuracy of the model. This has the advantage of reducing training times and enhancing generalisation by reducing over-fitting.

2.3.5 Hyperparameter tuning

In order to get the best equilibrium between bias and variance as pointed out in Figure 2.3, there are several methods to adjust model's complexity. As models act like black boxes it is difficult or almost impossible to optimise them through the analytical way. Therefore, it is necessary to try various combinations and choose the best one.

The simplest method to perform this is called *Grid Search*. With this method, we have different parameter values on a grid, a model is trained with each combination of parameters and it finds the best combination. Another simple method is *Randomized Search*. Instead of going through all the combinations, this method randomly chooses n previously defined combinations and outputs the best. Theoretically, this is not a good method, however, in practice even that their performance is slightly worse on a cross-validation this is a noise effect (Pedregosa et al., 2011). In a test validation, there is almost no difference between these methods with the added advantage that Randomized Search has a drastically lower run time. Bergstra and Bengio (2012) also adds that Random Search will much more effective than Grid Search because there some parameters are more important than other, so with the same amount of trials, it is possible to explore much more as it is illustrated in Figure 2.8.

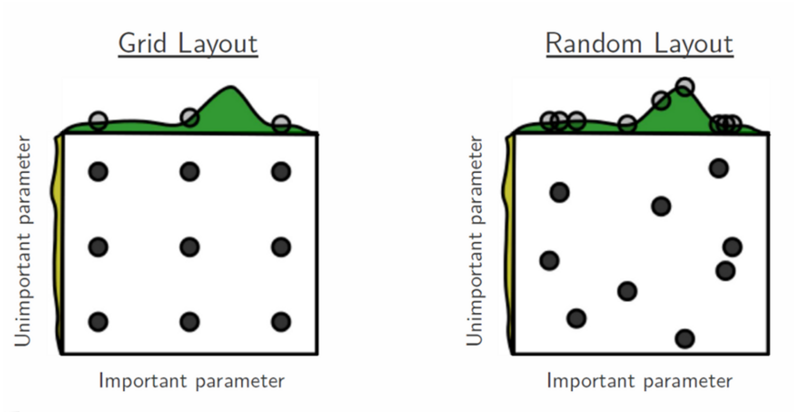

Figure 2.8: Grid *versus* Random Search diagram with 9 trials. Source: Bergstra and Bengio (2012)

Chapter 3

Problem description

The environment this project takes place has two peculiarities. Firstly, the company doesn't see the products it sells as it is an e-commerce marketplace without stock. And, secondly, the luxury market it operates in serves demanding customers that have exquisite habits. As previously mentioned, the only time a product passes through Farfetch Operations is before the beginning of a season. Every product which appears on Farfetch website is created by a partner (boutique or brand), a sample of the item should be sent to one of the company's offices called production centre (section 3.5) to be photographed and to add its information to the database which will appear on the website.

In the next sections, we will contextualise the key process in the organisation - order process, we will cover all processes related to packaging, we will introduce Farfetch packages and their attributes and, finally, the problem faced by this dissertation will be explained in detail.

3.1 Ordering process

It is important to get familiar with some processes and nomenclatures used in this document in order to fully understand the overall context of this work. When a customer adds more than one product to its basket on the website, although he doesn't have this perception, the products can come from different boutiques. This means that both products are being stored in different points and will most probably reach client's house at different times. Therefore, three different types of orders need to be defined: Portal order, Boutique order and Product order. In Figure 3.1, an example that illustrates the three concepts is given in which an order is made by a single customer of three different products. This portal order is composed of two different boutique orders, the first one with two products and the second with just one. Each one of the products is associated with an individual product order.

Once the order has been placed it needs to go through six different steps until it is delivered to the customer's house. All these steps are represented in the diagram in the Figure 3.2. At first, for each product order, the partner needs to confirm if it has stock of that product. In the case of not having, the boutique communicates with Farfetch and this will be taken into consideration in a no-stock indicator which could lead to a penalty. Farfetch tries to find the same product in a different

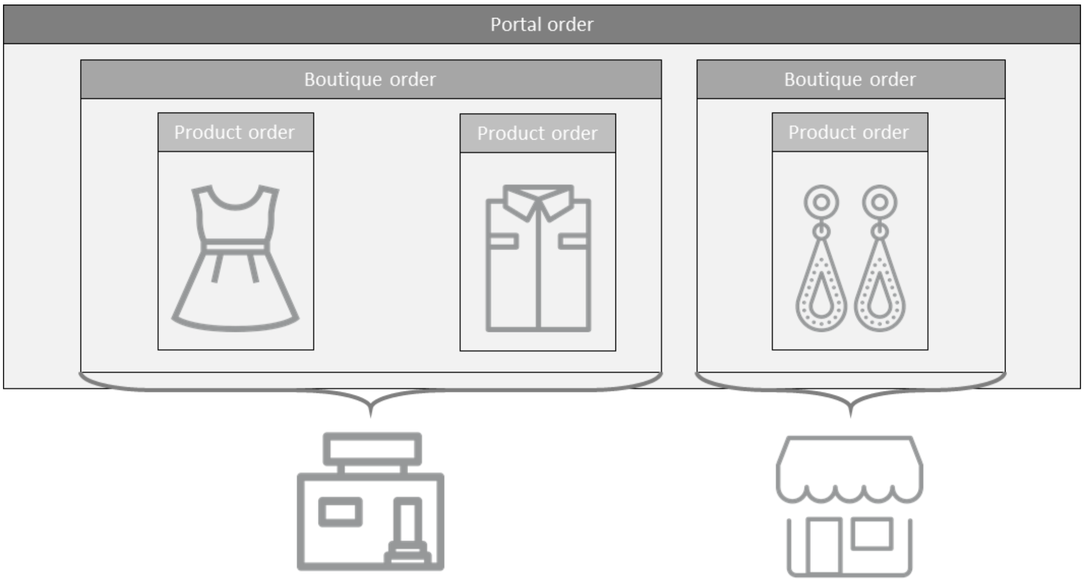

Figure 3.1: Diagram representing a portal order with 3 products from 2 different boutiques explaining different types of orders

store and proceed with the process. If the product ordered isn't available Farfetch contacts the customer, explains the situation, suggests similar products or performs a refund, depending on client's demands.

At the same time as the first, the second step begins: approve the order. Customers can be in the white-list, blacklist or none of them. Blacklist clients' orders are automatically cancelled, white-list are approved without manual intervention at all. The rest of the customers are evaluated and depending on their suspicion they are approved or go through a process called Under Investigation. Although this is called the step 2, since both steps are performed in parallel, 95% of the times the second one is faster than the first one.

The third step is a simple step where partners use a software called STORM to see Farfetch package recommendation and select the package(s) used for each order (this stage will be further mentioned in the section 3.3). The fourth is related to the creation of the shipping labels, which are indispensable for couriers to identify the delivery destination. The step 5 (Send Parcel) is when the package with the product is ready to send and the carrier (mainly DHL, UPS, Correios) is informed to pick-up the order. The sixth and last step starts when the courier collects the box and comprehends the stage when the parcel is in transit. It finishes when the order is delivered to the client. The carrier charges Farfetch a value based on the distance travelled and the volume/weight of the order, which will be further explained in the section 3.6. During this process, partners can create what it is commonly called exceptions in order to complain or inform about something that is not right (i.e. Courier failed to pick up the parcel, they do not have packages or faulty item)

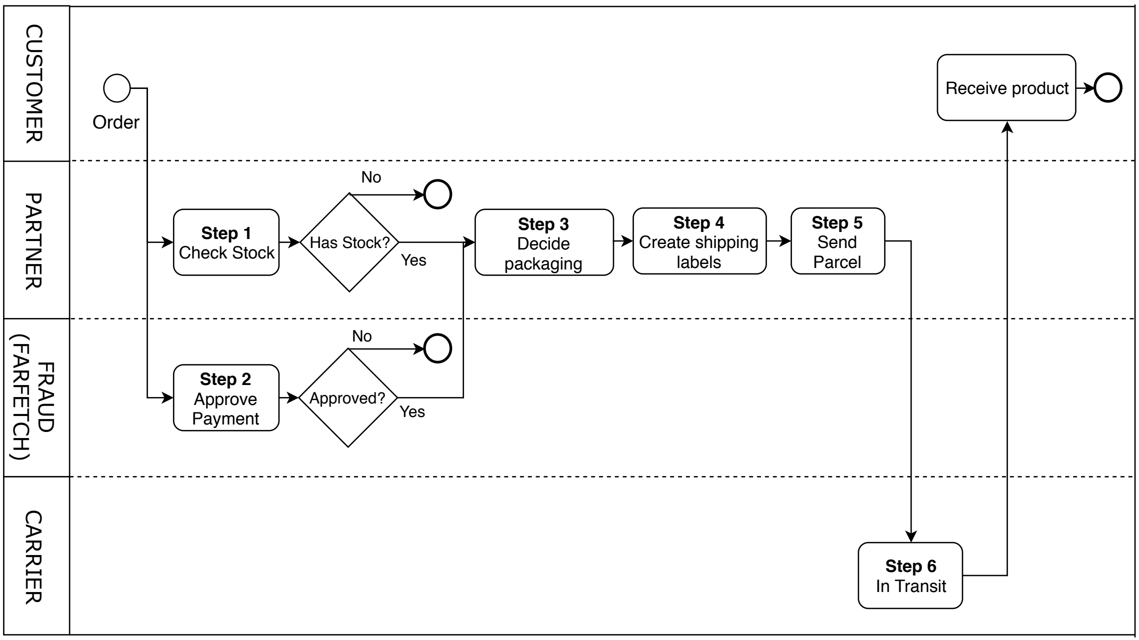

Figure 3.2: Steps between order and delivery

3.2 Farfetch packages

Farfetch sells lots of different kinds of fashion products - wallets, bags, dresses and boots - with very different dimensions. To maintain a consistent image, every order performed through Farfetch website is shipped with its company's package. Thus, the packages need to be versatile enough to enable packing every product Farfetch sells. This is the first important factor when thinking about this packaging, the second one is design. Most of the products Farfetch sells are expensive, therefore, the ability of quickly identifying its packages from the outside could be dangerous due to the risk of robbery. However, the client receiving the product is expecting a luxurious experience thereupon a too simple cardboard box is not good enough to achieve this.

Figure 3.3: Farfetch box. Source: Rockett (2016)

When opening the box, the experience should be as much as possible an amazing experience. The third one is the physical and functional part. Packages should be easy and fast to handle and to assembly (from partner's point of view), without sharp edges to avoid cuts and as light and robust as possible.

Given all these conditions, Farfetch box is the represented in Figure 3.3. A simple cardboard box from the outside but a white coating with the company logo from the inside. These packages must store the product, something which protects the product (wrapping paper or another box from the brand) and the invoice holder. Security cards and tags and stamps are mandatory to be on the package, as well. An interesting feature is the possibility to revert which is useful if the customer wants to keep it.

There are 3 different package systems inside the company: USA, Brazil and RoW. USA and Brazil packages are used for items produced in the United States and Brazil and mainly for domestic orders (inside the country). We will focus this work in the Rest of the World system as it represents 80% of all the orders and from now on everything will only focus on the orders covered by this system.

The RoW package system 9 different sizes of packages which are used based on article dimensions as shown below. This configuration was defined to get the adequate level of complexity. The Table 3.1 shows the different dimensions and the first type of products it was designed for. However, this category doesn't restrict the kind of products it transports.

Table 3.1: Packages names and respective sizes in centimeters

Box Number	Box Name	Dimensions (centimeters)
Box 15	Accessories/Jewellery	22.5x10x14
Box 16	T-Shirt	38x28x3
Box 17	Clothing	38x28x7
Box 3	Shoes	36x23x14
Box 5	Clothing/Boots	35x30x14
Box 6	Double Shoe	37x45x15
Box 7	Large Clothing	55x45x13
Box 13	Boots	70x40x15
Box 14	Large Box	60x45x25

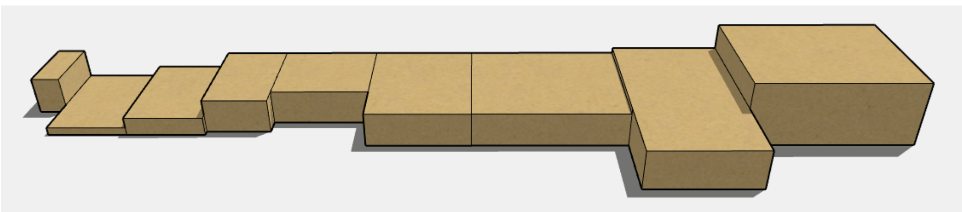

Figure 3.4: Digital representation of all packages. Image generated with *Google Sketchup*

3.3 Packaging Supply Chain

This section presents the packaging flow from end to end. There are four processes related to packages itself: box production, box ordering, box recommendation and box selection. In the Figure 3.5, it is represented a simple and high-level view of the first two processes. This was a necessary investigation in order to understand how the processes work and what will be the consequences of changing something in this flow.

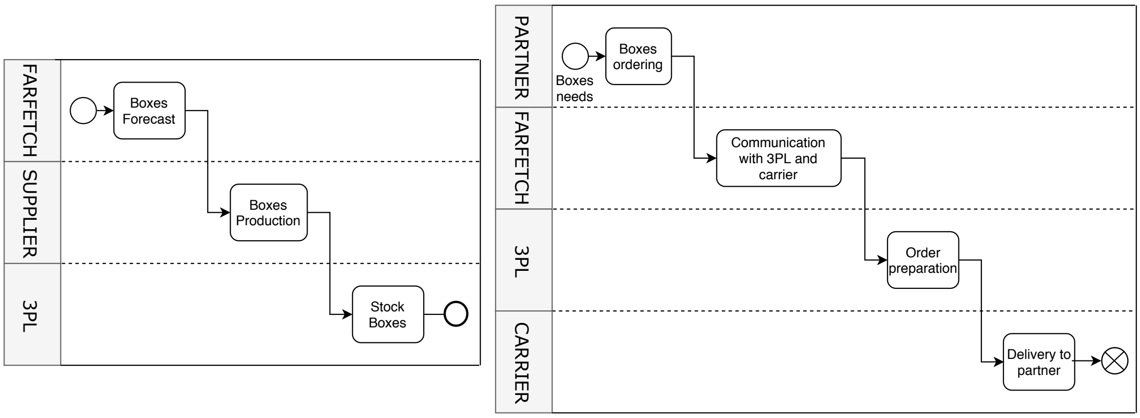

Figure 3.5: High level packaging processes: on the right it is represented the box production process and on the left it is represented the partner's box ordering process

Manufacturing and Stocking The packaging supply chain starts on the packaging supplier. During the project, there are two package suppliers both based near Porto, Portugal. Two reasons justify the use of different suppliers: freedom to choose the best price/quality and mitigate the risk of depending on only one supplier.

Based on a forecast, every month Farfetch orders the necessary boxes to these manufacturers. Each factory produces the unfolded cardboard boxes in lots of 10 and aligns with the Third-Party-Logistics (3PL) which is responsible for storing them. It also has its warehouses near the city of Porto which simplifies the communication and alignment with the suppliers.

Packaging Ordering Every time a partner needs Farfetch packages, it sends a *Microsoft Excel* file with a predefined template to the Supply & Retail Operations (S&RO) team. This template allows partners to choose the package size (and other packaging related elements), the quantity to order and type of service (standard or express). This team is responsible for inserting the orders into the ERP system which is connected to the 3PL system. The 3PL confirms it has stock through the ERP system, prepares the order for the carrier to deliver it to the client. The (S&RO) team is responsible for sending a report with this information to the Finance team in order to register the transaction and subsequently charge them. As referred in the section 1.1.1, the partner is charged a fixed price for every box purchased independently of the box volume.

Prepare order As described in the section 3.1, every order will go through 6 steps. For packaging, the most important one is the third where, using a platform called STORM, Farfetch recommends a box for each product of the order. As it is shown in Figure 3.6, partners can decide to follow the suggestion or to change it using a drop-down list. After filling this page, the order will proceed to step 4 where the shipping label is created.

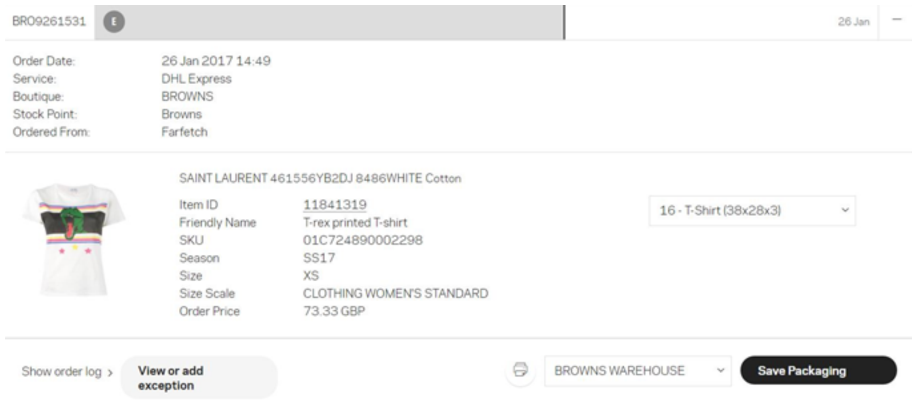

Figure 3.6: STORM in Step 3: Select Packaging

Unboxing Lastly, the consumer receives a cardboard box with the product he ordered. At this point, it is expected that the package doesn't show any signs of damaging as it will interfere in the customer experience. After opening the box, the disposition is totally decided by the boutique, however, the client associates this to Farfetch which means a very important part of this company's image is the partner's responsibility.

3.4 Incentive plan

High-end customers are used to receiving an excellent treatment when they are buying on boutiques, therefore they have high-quality standards for the service provided by Farfetch. An incentive plan was developed to reward those partners who are providing a good service and penalise those who are not. At the moment there are two metrics being analysed on a monthly basis: Speed of Sending (SoS) and No Stock. These metrics measure the time it takes for a customer to receive a product and the percentage of items ordered times the partner says it has no stock of a specific product when the process already is on step 1 (see section 3.1).

3.5 Production centres

In the fashion industry, a year is partitioned into two different seasons: Autumn-Winter and Spring-Summer. Brands introduce their collection at the beginning of these seasons and consequently, the amount of work on Farfetch is cyclic. This cycle is noticed particularly on two sides of the organisation: Production and Operations.

The production workflow has the final output of putting a product online. It starts when a partner creates a product on a platform called *SALES* and it is checked if that product had already been created by another partner (duplicate check). If not, a slot (box with certain insurances which will ship a group of products) is requested and the products are delivered to a Farfetch production centre. Afterwards, in this centre, the product is prepared and ironed to be photographed. Depending on the type of product, the photographs may defer. The last step before the product is shipped back to the partner is *Scan-out*. This process is where all relevant data is collected from the product such as washing instructions, composition and, also, the box recommended to ship the product. The product is then shipped back to the partner.

The period before the first days of a season is marked by the big amount of items received on the production centres. On the other side, the working peak of Operations team (this team's functions are described in the section 1.1.3) is reached when sale season starts. This creates an interesting pattern as shown in the graph of the Figure 3.7, where there are two clear peaks, happening twice a year with a difference of 2-3 months.

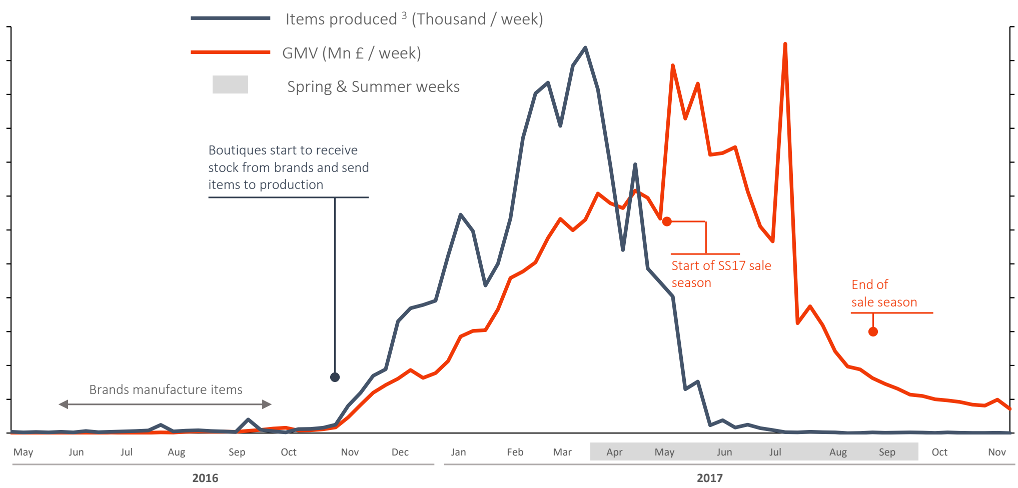

Figure 3.7: Sales and production peaks on Spring Summer 2017 collection. The peak on the left represents the amount of items produced on Farfetch production centres, the peak on the right the GMV.

3.6 Shipping invoices

Upon ordering at Farfetch the customer is charged a shipping fee to cover the delivery costs the company needs to pay. This fee varies depending on the size, weight and destination of the order and is based on the number of boutique orders. Notwithstanding, for basket values above the threshold of £ 100, 150 $ or 140 € a flat shipping fee is applied. The fee also varies depending on the type of delivery: Standard (5-7 days), Express (2-4 days), Same Day (less than 2 hours).

Although the e-tail company works with various carriers which deliver the products, it is *DHL* who deal with the major part of all volume. The bills charged to Farfetch are currently being automatic processed, charged and uploaded to the database. However, an important difference stands between invoices from this carrier and others. In their recipes, *DHL* gives more information about how much they pay for the service, using the term *Invoice weight*. This field is calculated using the biggest of two options: weight in kilograms of the product or using the volume in cm^3 with a correction factor as represented in the equation 3.1. Since Farfetch does not see what the package during the process of order fulfilment, it turns out that this field is the most accurate way of predicting which box was used to deliver a specific product.

$$Invoice\ weight = \max(weight; \frac{length * width * height}{5000}) \tag{3.1}$$

Even though the billing process is automatic, invoices are not sent instantaneously to Farfetch when the products are sent. In the graph of the Figure 3.8 is represented the number of days it takes to receive the invoice. It takes 30 days to receive approximately 99% of all the invoices.

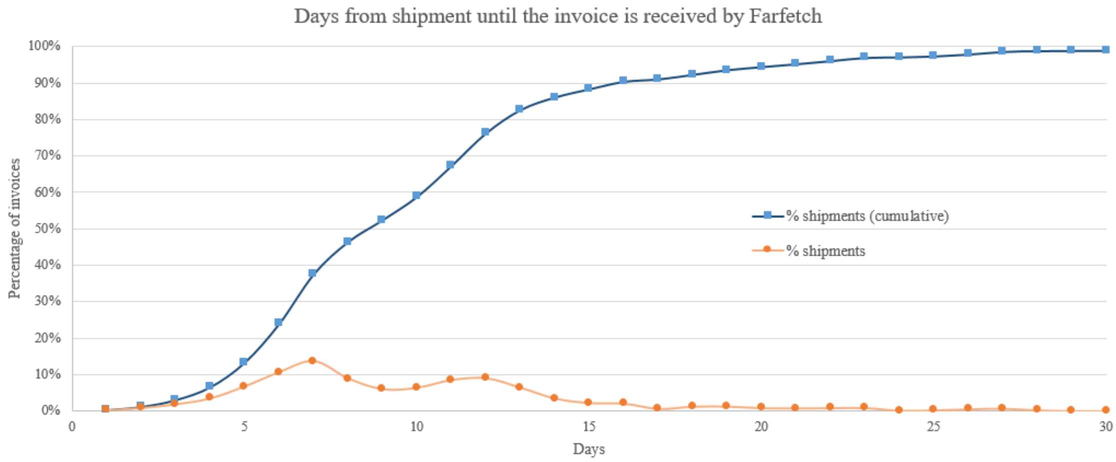

Figure 3.8: Percentage of *DHL* invoices received and its relationship with the the difference between shipment and invoice dates.

Chapter 4

Methodology

To successfully implement the project in the rapid growing environment of the company a project management methodology was followed. This approach helps identifying risks and limitations in order to mitigate them and improve the overall performance. As explained in the section 2.2, it was used an iterative approach as illustrated in Figure 4.1.

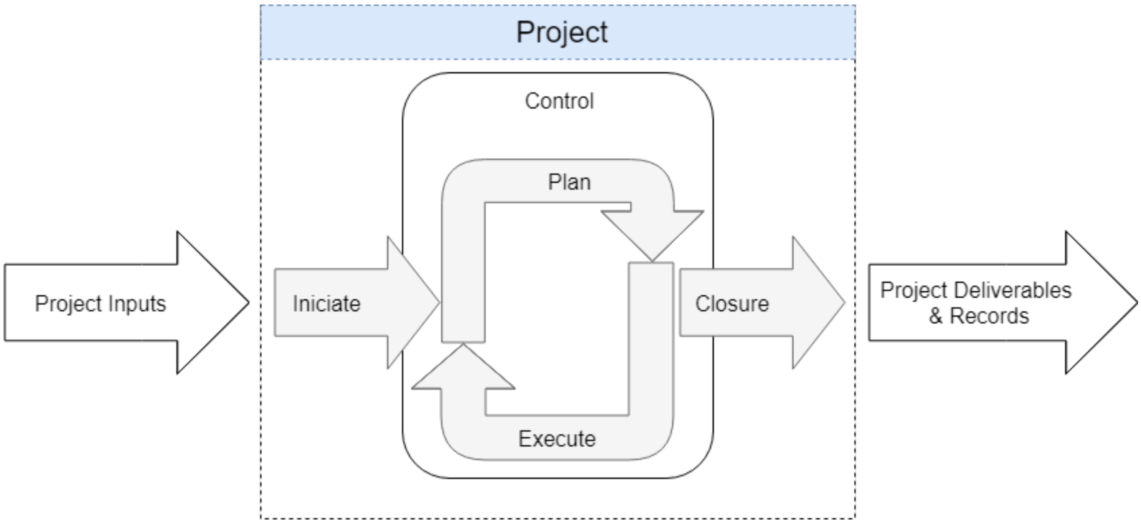

Figure 4.1: Project pipeline inspired on the PMBOK guidelines

4.1 Initialisation

The methodology proposed in the PMBOK by Project Management Institute (2017) states that, having defined objectives and motivation, every project should start with stakeholders identification, scope selection and an estimation of the investment needed.

After understanding the packaging supply chain in the chapter 3, it is noticeable that a decision with a considerable impact for Farfetch is on partners' hands: choosing the package for each product. Before this project, there was not any awareness on partners' side of using the smallest box possible. The partner is charged a fixed price per box ordered (section 3.3), the customer pays a fixed price to Farfetch depending on the type of delivery (section 3.6), and the e-commerce organisation is levied for the shipping costs by the carrier (section 3.6).

As noted on the section 1.1.3, this project will involve several parts of the Operations Department. The stakeholders were identified and the project team was built with some members of this teams as shown in Table 4.1.

Table 4.1: Stakeholders and project team members

Team	Role in the project	Justification
Operations Strategy	Project manager	Responsible for projects inside Operations
Delivery Development	Project team	Deal with delivery costs
Supply & Retail Logistics	Project team	Accountable for packaging suppliers and 3PL
Partner	External stakeholder	Liable for packaging selection
Partner Service	Stakeholder	Responsible for communicating with partners
Production	Stakeholder	Accountable for recommending packages

This way, the current project came up as an opportunity to build a path to optimise delivery costs, reduce the amount of air shipped contributing to a more sustainable supply chain and to enhance customer experience. Gathering all the inputs from the team, the main cause for the problem was identified: no partners' awareness upon selection a box. Afterwards, a solution was proposed: develop a packaging incentive plan who benefits those partners who are choosing suggested (or smaller) packages rather than bigger ones.

Aiming to increase the packaging accuracy, it is necessary to compare the box chosen with the recommendation. If from one side we already knew that partners' box selection was not the best possible, on the other side, Farfetch's box suggestion on each product is not accurate and must be improved.

Identifying the resources needed is also an important step to analyse whether the project should go on or not. Starting with the budget two things must be considered, working hours of project team members and the incentive given to partners. Given the complexity of processes and teams this project will involve, it was expected that the packaging incentive to be live on March 2018 and all the processes of improving our box suggestion on June.

This process culminates in a business case (appendix B) which was successfully approved by the board.

4.2 Plan

After the business case approval, it is time to start planning all the course of action of the project and redefine with precision what was defined during the initialisation. Using the division previously defined (section 1.1.4), the subsequent phases will cover the refinement of partners' box usage, starting with the packaging incentive plan, going through the channel created to manage exceptions, increasing the control and supporting partners. On the other workstream, the improvement of Farfetch recommendation will start with the mechanisms to support *Scan-out* team and ending with the development of an algorithm which tries to predict the best box for a given product.

4.2.1 Packaging Incentive Plan

Following the same line of reasoning that was used during the first project team planning meetings, we will start by explaining the main points of the incentive itself. Firstly, it was set up a packaging KPI per partner whose purpose was to evaluate their accuracy upon selecting the best possible box for each product order, as shown in the equation 4.1.

$$\text{KPI} = \frac{\text{number of orders using the recommended box or smaller}}{\text{number of orders}} \tag{4.1}$$

The proposal for this incentive was the following: boutiques and brands who accomplish, at least 85% on the packaging KPI would receive, approximately, 20% of their packaging expenses, based on their Actual Transacted Value (ATV). The threshold of 85% was defined because, before the incentive, 20% of the partners had already achieved it. Therefore, the threshold was a well-balanced trade-off between ambition and easiness. The incentive given to partners (20% of their packaging expenses) was specified given that a predefined percentage of the savings should go to partners. The incentive should be integrated into the current incentive plan (section 3.4) which is given every month, creating the least friction possible for partners and inside Farfetch. ATV is given by the equation 4.2 and Gross Merchandise Value (GMV) is represented in the equation 4.3.

$$\text{ATV} = \text{GMV} - \text{Cancellations} - \text{Payments Refused} \tag{4.2}$$

$$\text{GMV} = \text{Number of visits} \times \text{Conversion rate} \times \text{Average order value} \tag{4.3}$$

The packaging KPI will evaluate partners' package choices for the products delivered. However, given the available data, there are orders which are not eligible for this evaluation. The orders should match these criteria:

1. Boutique orders with just one product. It is possible that partners put two or more products in the same box and there are a huge amount of combinations. Farfetch only has information about the recommended package for each product;
2. Products shipped by DHL. Besides being the only carrier whose bills are automatically processed, it also communicates the invoice volume which is used to identify the box used;
3. Box recommended is from RoW package system (not USA or Brazil). In United States of America and Brazil different boxes sizes are used, therefore, partners who use these systems are not eligible for the incentive;
4. Orders charged by volume. There is also the possibility, however not very common, of products being charged by its weight instead of its volume and it is impossible to discover which box was used;

5. Orders actually charged. Although uncommon, there are orders which are not charged by any carrier due to a mistake.

6. Orders that partners created exceptions. Every time partners are not satisfied with our suggestion, they can use a channel to create an exception which will be further explained in the chapter 4.2.2;

It is represented in Table 4.2 the number of orders covered by the packaging incentive plan in March 2018 and the reason why they were not used.

Table 4.2: Percentage of March 2018 orders in the scope and out of scope of the packaging incentive evaluation (eligible partners)

	In scope	73%
Out of scope	Charged by weight	4%
	Multiple product order	4%
	Other carrier	13%
	Not charged	6%

4.2.2 Exceptions

As Farfetch employees are aware that their box suggestion (chosen by production team) is not very accurate, the decision of opening a channel to give partners the possibility to complain was made. Every time a partner is packing a product which has a wrong package suggestion, the partner can send a warning (exception) related to that order to Farfetch. Through this, it had been possible to both exclude these orders from the packaging KPI calculation and use this information as knowledge of the partners' point of view.

The number of exceptions which would be created was unknown and due to the need for a quick answer, it was decided to create two distinct phases. A first phase where partners create exceptions and can fill a free text field. In this phase, their management would be done manually and by the project team members. The guidelines for accepting or not these exceptions need to be defined and it was decided to develop them in parallel with the exceptions management. In the second phase, this channel would be revamped, a more work-requiring exceptions channel would be developed and their management would be integrated into the same platform used to manage exceptions related to SoS and No Stock.

4.2.3 Support and educate partners

After launching the incentive, given that package recommendations were not as accurate as desired, some partners' routines were influenced. Aware that these changes could create friction and lead to some confusion, two objectives should be explored:

- Maintain a fast and effective communication during this period of change where package suggestions are being tuned;

- Prevent any surprise and predict what can make partners neglect this incentive.

With these purposes in mind, Partner Service team should be always updated to help partners throughout this process. Additionally, it was noticed that during peak season there is an increase of exceptions related to packaging stock outs and as, at the beginning of the project we were heading to a sale season, it is important to support partners to avoid this which could decrease their packaging accuracy.

4.2.4 Production recommendation

Before the project, the box suggestion which appears in the commonly called "Step 3 - Packaging" on STORM (section 3.1) was only defined during the Scan-out process by the team with the same name. As stated in the section 3.5, in the production centres, a person decides which is the best package for each product. During this planning phase, several box suggestions were analysed and it was concluded that the decisions made by this team were not as good as desired.

Given this lack of accuracy, it is important to find what are the causes of these discrepancies and maintain them updated of their scores. This way, identifying the causes and creating guidelines to mitigate or eliminate this mistakes should be done, as well as, creating a routine with the purpose to create this awareness on the Scan-out team.

4.2.5 Best box algorithm

To guarantee a proper way to evaluate production suggestions it would be interesting to try to find patterns on packages selected on similar products. Given that it is possible to know accurately which box was used by partners, an algorithm should be developed to try to predict what is the best box for each product.

Two different algorithms were planned: in the first phase a simpler one where the products with similar characteristics were grouped and the algorithm memorises the best box for each of these groups. The second phase would consist in an algorithm based on machine learning models which try to independently find patterns between these characteristics and generalises the ideal box based on these patterns.

During these phases, it was also necessary to define what is the best box for a product. Using the information the first aforementioned algorithm gathered, it had been possible to understand that partners usually choose bigger boxes than needed. The smallest box used at least one time could be used, however, it is necessary to keep in mind three conditions:

- The same product can have several different sizes which could require different boxes;
- There could be mistakes on the information provided by DHL;
- A box too small for a product can harm the customer experience.

Using a conservative approach, 10 orders should be enough to have a representative sample to take conclusions. The fittest box used at least 20% of the times would be a good starting point.

To guarantee these 10 orders are achieved as fast as possible it is possible to group in *Designer ID* instead of *Product ID*. Products with the same *Designer ID* are similar products but with different colours.

4.2.6 Gantt chart

Finally, to maintain a good course of the project it is recommended that the project team have periodical meetings to discuss achievements, blockers, if schedule the execution is being followed as planned or if plans need to be changed. These meetings were decided to happen on a weekly basis. Given the workstreams defined in the previous sections, it was possible to build a Gantt chart (Figure 4.2) containing all the different parts of the projects in order to be on schedule and with the resources previously defined.

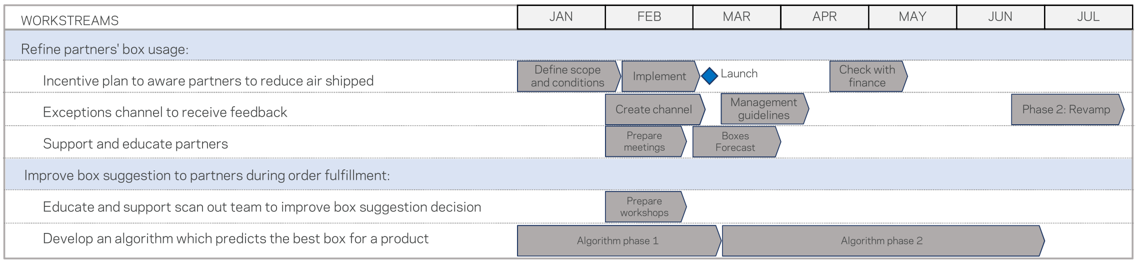

Figure 4.2: High level Gantt chart of the project

4.3 Execution and Control

Starting executing the project as planned, it is important to remember the Figure 4.1 where it is pictured the flexible and iterative approach a project should have. This chapter explains how the work was structured in each one of the workstreams previously defined, as well as, the control made to evaluate the actions performed.

4.3.1 Incentive Plan Implementation

In simple terms, the incentive will be given by comparing the box chosen by the partner and the box recommended by Farfetch for each product order. Partners' choices are given by the DHL invoices and Farfetch recommendation is defined by Scan-out team. Both types of information are on different tables in the company's database. One of the most important thing during the implementation of the packaging incentive is data storage. This process will support good communication inside Farfetch (between different departments) and outside (with partners). This is important for two main reasons: error traceability and data analysis are easier performed.

To make it easier to follow the evaluation and given that recommendations can change over time for a given product, both used and recommended box volumes should be registered on the database's table representing all product orders. Additionally, a third field should be added which will compare the two values and output *Yes* or *No*, based on whether the recommended box or a smaller one was used to pack each product order. This field can also take the value *Null* if the order is outside the scope outlined on the section 4.2.1, or *Exc* if an exception is created and accepted for that order.

To get the packaging KPI for a partner a simple operation should be performed which is represented in the equation 4.4. Every month, this calculation will be automatically inserted in a table which is used to make this information reach the Finance team. Through integration, it is added to a *Microsoft Excel* file called *Financial Report* for each partner. The diagram represented in Figure 4.3 illustrates the information flow for this incentive.

$$Packaging\ KPI_{partner,month} = \frac{\#Yes + \#Exc}{\#Yes + \#Exc + \#No} \tag{4.4}$$

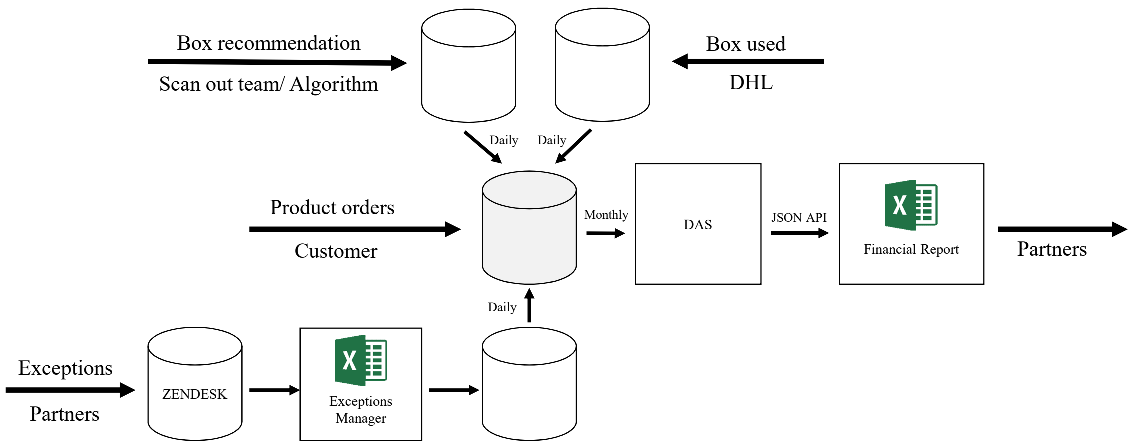

Figure 4.3: Information flow of the packaging incentive

After data storage, it is important to schedule all the important events that will happen monthly. The first restriction to take into consideration is that DHL bills could take one month to be sent as demonstrated in Figure 3.8. Thus, a partner's performance on month n will only have enough data to be assessable at the end of the month $n+1$ and included in the *Financial Report* sent on the month $n+2$. This financial routine is usually sent on the 10^{th} workday of each month and the benefits are transferred on the 20^{th}. Exceptions must be managed before these routines. The timeline of Figure C.1 of the appendix C summarises all the monthly routines related to this incentive.

4.3.2 Exceptions manager

The opening of the channel through which partners can complain was communicated in the middle of March. Its volume is very high compared to other exceptions highlighting the commitment of certain partners. Although these exceptions can become a very trustable source of knowledge, it is important to keep in mind some limitations:

1. Even after Farfetch send partners a newsletter informing them to use the exception only when box recommendations are too small, partners used them for every non-compliance (e.g. too big recommendation);

2. Some partners do not have all the packages or do not even know some sizes exist. They just order 3 or 4 boxes sizes to avoid increasing complexity;

3. Partners can lie using these exceptions just to receive the incentive when they do not have the recommended box;

4. Sometimes partners misunderstood these exceptions and use it to inform about packaging issues not related to this selection;

5. Some bigger boutiques use STORM and pack products at different times or these two processes could even be done by different employees. Firstly, one of the employees checks if the stock is available and go through step 3 and 4 just pressing the *Next* button on STORM. He prints all the papers needed for the order and sends them to the person responsible for packing. If this person spots a bad recommendation, he tells the first person who creates an exception on step 5.

Given this reasons, there should be created a mechanism with the purpose to filter just good exceptions. In a first phase, a tool made with *Microsoft Excel* was conceived to help the management of this exceptions. This tool would gather information from different *SQL* queries which connect to the database and shows the exceptions created, the box recommended, the box used, the most used box for that product and other important fields (shown in the Table 4.3) to make the decision of accepting it or not.

Table 4.3: Exceptions manager fields

Field Name	Description
Order Code ID	ID to identify order
Used Box	Name of the package used accordingly to DHL invoice
Recommended Box	Name of the package recommended by Farfetch
Is Recommended Box	Whether the package used is smaller or equal to the recommended
Product ID	ID of the product of the order
Designer ID	Product's Designer ID
Picture	Hyperlink with the picture of the product
Orders	Number of orders using each package for that Designer ID
Notes	Notes written by the partner when creating the exception
Recommended (% orders)	Percentage of orders using the package recommended
Invoice (% orders)	Percentage of orders using the package used
Answer (Y/N)	Y if the exception is accepted and N for not accepted

4.3.3 Support and educate partners

As planned, two actions were performed to help partners during the process: monthly meeting with PS and send a packaging forecast before sale season. In order to understand how partners were reacting to this incentive a monthly routine was created. This one-hour meeting at the end of the month between people working in Partner Service team and the project team was necessary to answer partners' doubts and give some feedback of how the project will develop in the future.

On the other hand, it was decided that the top 25 partners (in sales) should receive a packaging forecast based on the sales forecast already delivered by Farfetch and based on last year's boxes sizes ordered during this season.

4.3.4 Support and educate Scan-out team

The increasing volume of exceptions created by partners related to box recommendations, some other communications made by certain partners and the box recommendation procedures in the production centre just confirmed a strong hypothesis: Farfetch package recommendations were far from perfect. It was, then, necessary to develop a monthly meeting in the biggest production centre to aware them that their package suggestion has exponentially increased its importance. A visit to this production centre was done to identify certain reasons causing this lack of accuracy. These reasons are the following:

- Training. When a new worker is contracted, there is not any concern about training him to be proficient in this matter;
- Space perception. Most of the workers never had the packages on their hands. They are hanged in a wall 2 meters from the ground;
- Time. The worker has 2 seconds maximum to select the package for each product;
- Control. Until now, this step was being despised. No metrics were used to evaluate the worker performance;
- Items to production are different from items to clients. The items sent to Farfetch production centres do not have the packages used to send the items to customers, partners only need to send the right shoe to be photographed or Farfetch guidelines for folding specific types of clothes are different from partners.

From this causes, a monthly meeting was shaped. The first step was to create a dashboard showing how they were doing compared to the algorithm described in chapter 4.3.5. The results show that in 43% of the times, the packages didn't match. The meeting should create the need of looking at this as an important KPI to evaluate their work quality and to understand how, with training, the results can be improved.

4.3.5 Best Box Algorithm

The algorithms developed had the objective to predict the best box for a product. Using the definition of the section 4.2.5 both methods were performed with slightly different approaches.

On the first phase, six tables were created from specific to generic, each one with a different level, as it is represented in Figure 4.4. As previously mentioned, every item sold on Farfetch has a Product ID associated and similar items with different colours have the same Designer ID. BFC stands for Brand Family Category (i.e. Prada Tote Bags is a BFC), every product also has a Gender associated (Woman, Man, Unisex and Kids), a season (Spring/Summer, Autumn/Winter or Vintage) and a year corresponding to the collection the piece is integrated into. These tables show how many times each package was used for each combination. To compute the best box, the rule previously referred, is applied only for the first level (Designer ID): the fittest box used at least 20% of the orders. For all the other levels, as they are too generic, the best box is conservatively considered the most used one.

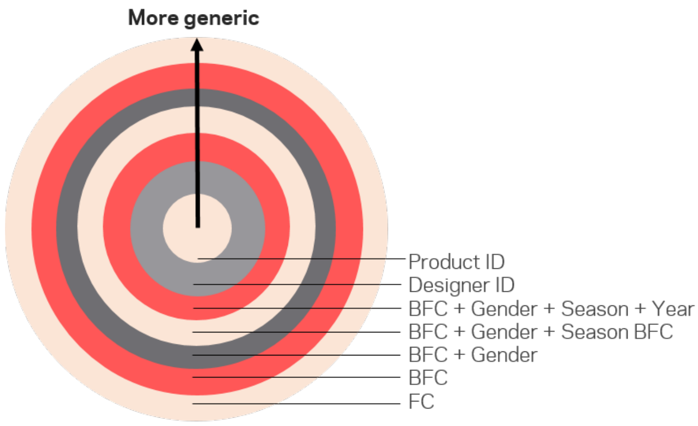

Figure 4.4: Levels used in the conception of the algorithm.

A test where some product's recommended packages were changed, using this algorithm first level, was performed to validate this information. The volume of exceptions created for that products was monitored and it turned out with great results - exceptions for the overwritten products decreased to 60% after the 11.000 products' overwrite.

However, it was necessary to highlight the following limitations and the suggested workaround:

- Too dependant on *DHL* errors. Analysing invoices, it is noticed that carriers can make mistakes and the invoice weight (volume) could not match the right package. As defined in the section 4.2.5, the best box is defined using a minimum of 20% threshold to avoid too avoid too small suggestions.

- Too dependant on boutiques which did not have the right packages. Bills collected by this algorithm are from 2017 until now, therefore, at that time partners did not have any awareness on choosing the fittest package. During peak season, it is common for boutiques to run out of specific packages and, in that case, bigger ones will be used to pack products. To avoid being biased by this cases, the rule above mentioned was used to the more specific level where this cases could have a greater impact;

- Box 15 and 16 have the same volume. The workaround developed for this was based on the 1st level category of the product;

- Strongly depends on orders. To predict the best box for a product which is launched online, the algorithm will probably need to go to higher levels which are very uncertain. Articles take an average of 4 months to reach 10 orders. This is the reason why it was created more than one level for this algorithm;

- It generalises blindly. This algorithm does not discover patterns between variables. To overcome this limitation a more complex algorithm was developed.

It is represented in Table 4.4 the percentage of products covered for each level of the previously defined algorithm and an idea of each level's accuracy when compared with level 0 (Designer ID).

Table 4.4: Number of products online on 18/05/2018 affected by each level of the algorithm and their accuracy comparing with level 0

Level	Name	# Products	%	Accuracy (level 0)
0	DesignerID	30132	8.5%	100%
1	BFC_G_SY	188355	53.1%	75,3%
2	BFC_G_S	54380	15.3%	74,6%
3	BFC_G	20889	5.9%	73,7%
4	BFC	5028	1.4%	72,7%
5	FC	55899	15.8%	64,5%

The approach for the development of the machine learning algorithm methodology, due to its complexity, is presented in its own section 4.5.

4.4 Closure

In this part of the project, a balance of the knowledge acquired was done and all documents and deliverables were kept for future reference. Some other ideas came out during the development of this project with the objective of increasing packaging recommendation accuracy which will be reported in the section 6.2.

4.5 Best box algorithm

One of the most important steps of this project was to develop an algorithm that would predict the smallest box that could be used for each product. As it can be a better recommendation, it can be used as a complimentary help when production is suggesting a box or to overwrite production suggestions every time these are not considered good. To mitigate the limitations the algorithm previously defined had, it was developed a machine learning algorithm which predicts the ideal box based on the characteristics of a product.

For this specific problem, it was used an approach adapted from the CRISP-DM[1] framework, as shown in Figure 4.5.

[1] Cross Industry Standard Process for Data Mining. Source: Chapman et al. (2000)

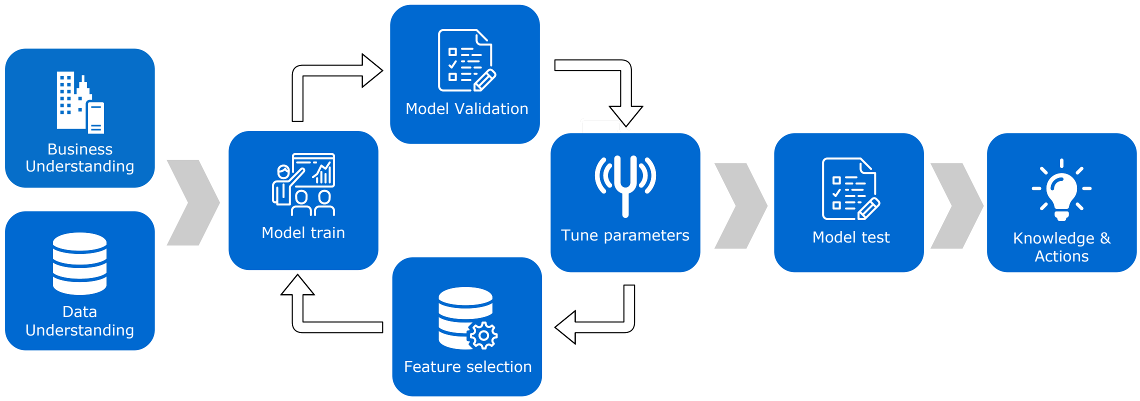

Figure 4.5: Framework used during the conception of the machine learning algorithm

4.5.1 Business and data understanding

After outlining the packaging processes, it is important to specifically point out the main drivers that affect box size. Partners are the only entity who deal with the real space required for a ready-to-send product, therefore their selection is the most accurate way we have to state which is the best box for a certain product. After some meetings with partners and Scan-out team, the inputs that affect the best box size are:

1. Product dimensions: measurements of the article;
2. Product density: type of materials used and *foldability*;
3. Brand's box: whether the brand of the product gives a box to deliver the product;
4. Protecting material used: quantity of paper or if a dust bag is used;
5. Extra partner's items: gifts, notes, discounts;
6. Farfetch must-have items: invoice and invoice holder.

From these drivers, meetings with some boutiques and with the Scan-out team emphasised that the ones that influence more the size of the packaging used to ship are product's dimensions, density and brand's box. However, as any real-life problem, we are limited to the data available. The first obstacle is that we do not have accurate measures of every product, their density nor if the product needs a box of the brand inside Farfetch package. To circumvent this problem, some features that could be intrinsically connected to these characteristics were selected to predict the package:

- Category 1st level - the high-level category i.e. Clothing, Shoes, Bags, Accessories;
- Category 2nd level - a subcategory that splits 1st category into more specific types of pieces i.e. for shoes categories: trainers, boots or loafers. Both of these categories can help predict the size of the product, as well as, whether it is possible or recommendable to fold them;

- Gender - the item was designed for Women, Men or Unisex. Even inside the same category men and women can have totally different sizes. For instance, woman's boots tend to be much bigger;
- Season - If the item is from a Spring/Summer, Autumn Winter or Vintage collection. There are several trends that some products follow from season to season as well as summer clothes are lighter and occupy less space compared with winter clothes;
- Brand - What is the brand of the product. This could help identify patterns on brands that have their own packages and trends some brands like to follow.

These features were also selected due to their trustability as they are used in the website and they are not free-text fields on the database. Before developing this machine learning algorithm, the best way we had to predict which is the best box for an item is the previously designed algorithm on Designer ID level (section 4.3.5).

It was used a dataset from the level where a Designer ID with the features aforementioned had an associated best box. This dataset was compared with all the orders to understand if it is a good representation of the reality as seen in Figure 4.6. It is noticed that there are not big deviations between these two. Furthermore, this analysis will give us the first estimation of how well the algorithm is working. Given that the most represented class is box 17 with 32,7% of the cases, an error of 67,3% would be our minimum. In other words, an algorithm that predicts all product's packages with the most used one (box 17) will have 67,3% of error.

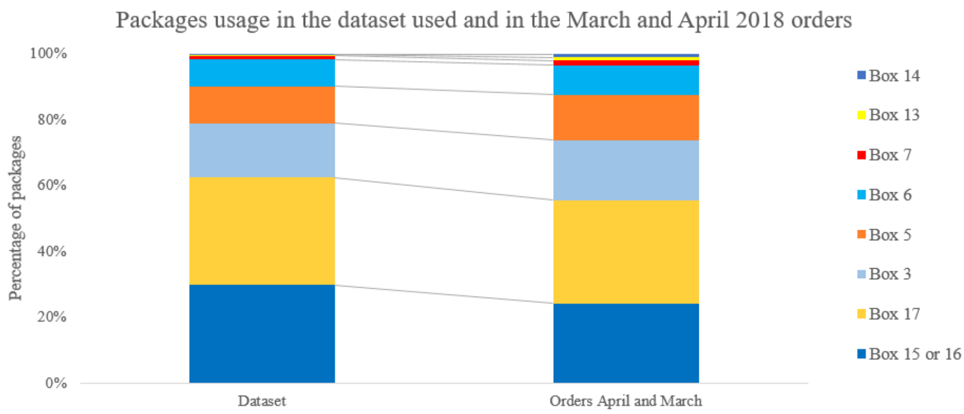

Figure 4.6: Packages used on March and April 2018 orders compared with the ones on the in the dataset used

After extracting the data from the database, it is necessary to clean it - remove all non-compliance or useless information like *null* or *other* fields. Afterwards, the dataset with more than 110.000 records and 1500 features was split into train and test sets (section 2.3.2 with the percentages of 80% and 20%, respectively. The amount of features is due to the dummy encoding which converts every categorical feature in a binary feature to allow algorithms to train models more easily.

All these procedures were developed in the computing environment *Jupyter Notebook* using the programming language *Python*. The libraries used to treat data and develop the model were: *SciPy* (Jones et al., 2001), *Pandas* (McKinney, 2010) and *Scikit-learn* (Pedregosa et al., 2011).

4.5.2 Modelling

There are a lot of machine learning algorithms but all of them have their restrictions and were designed to a specific application. Given that the aim is to predict nine categories, it is necessary a multi-class classification model. For this kind of problems, there are still several recommended algorithm families: logistic regression (*LR*), tree-based algorithms, artificial neural networks (*NN*) and one vs all support vector machines (*SVM*). One simple non-tuned algorithm from each family was tested doing a 10-Fold Cross-Validation using the train set with the performances presented in Table 4.5. The parameters used in the tested algorithms were recommended by the bibliography contained in the *scikit-learn* library.

Table 4.5: Algorithms tested and their Cross-Validation performance

Algorithm		10-Fold Cross-Validation		
Name	scikit-learn algorithm	Error: Mean	Variance	Runtime
Logistic Regression	linear_model.LogisticRegression	25,17%	0,16	9 seconds
Decision Tree	tree.DecisionTreeClassifier	19,65%	0,08	4 seconds
Neural Nets	neural_network.MLPClassifier	19,98%	0,02	3309 seconds
Support Vector Machines	svm.SVC	31,48%	0,12	7147 seconds

For bigger datasets, Neural Networks and Support Vector Machines algorithms usually have a better accuracy but with longer training times. In this case, as the dataset could be considered relatively small, tree-based algorithms have a good performance (even better than more complex algorithms in the aforementioned examples) with the advantage of being pretty fast. For every machine learning model development, it is important to discuss the tradeoff between accuracy and efficiency and it actually depends on each application.

In this particular application we have two possible ways: serve as a guide to the Scan-Out team when suggesting a box for a product or constantly overwriting wrong suggestions. It is also important to keep in mind that:

- Fashion trends change quickly, especially during the beginning of a season and this could influence box sizes for a given group of categories;
- Oscillations between package sizes for the same group of categories affects negatively the trustworthiness on Farfetch recommendations;
- It could take a month to receive some carrier's invoices, therefore, the algorithm could be one month delayed with reality.

With these three arguments, it is possible to understand that it will be necessary to train the algorithm with some periodicity. Therefore, the time taken to train the model should be considered as it cannot be too high. However, assuming that the training time is not more than a certain threshold, accuracy plays a more important role here. The approach to finding an ideal solution was to find good time performing algorithms and try to tune them until they reach a maximum of accuracy.

It is also important to be aware that these models are competing with a solution which takes 3 minutes minimum (strongly dependent on database traffic) with 25% of error (algorithm of the section 4.3.5). Apart from the training time, a machine learning algorithm uses data retrieved from a *SQL* table whose data take 50 seconds to update. Retrieving this data takes 5 seconds.

Firstly, it was explored the slower but usually more accurate algorithms with feature selection to understand if we could improve performance and maintain or increase accuracy. The most important features of the dataset were selected. For this purpose, it was used one of the most simple feature selection technique - removing features with low variance. It was decided to use this simple algorithm just to check if the time performance would be competitive. From the dataset but with just 62 features, the Artificial Neural Network still took 2440 seconds to train with 20,2% of mean error. Both performance and accuracy were worse than the tree-based algorithms, so this solution was not further explored.

Secondly, with the simpler algorithms, it was decided to select three different kinds of tree-based algorithms, each one with its advantages: Decision Tree, Random Forest and Extremely Randomized Trees. After optimising their parameters the best one will be chosen. After that, a feature selection analysis can be made to discuss the already known trade-off between time and accuracy.

Decision Tree The fastest solution with a very good error performance is the Decision Tree algorithm. Its path is designed from splitting data into sub-samples with less entropy, as explained in the section 2.3.3 and in the appendix A. The hyperparameters used in this algorithm were the maximum depth of the tree, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node, whether bootstrap samples are used when building trees and the criterion to measure the quality of a split: Gini index or entropy. The algorithm used is in the *scikit-learn* library and it is called *tree.DecisionTreeClassifier*

Random Forests A group of Decision Trees are grown using this algorithm and the most voted output will be the output of this model as explained in the section 2.3.3. The parameter to be tuned was the same as the Classification Tree plus the number of trees in the forest. The algorithm used is in the same library as the previous ones and it is called *ensemble.RandomForestClassifier*

Extremely Randomized Trees This algorithm is exactly the same as Random Forests however the split is selected totally or partially random, as explained in the section 2.3.3. The parameters to be tuned are exactly the same as Random Forests. The algorithm used is in the same library as the previous ones and it is called *ensemble.ExtraTreesClassifier*.

4.5.3 Tuning

Without tuning the three algorithms' error is shown in Figure 4.7. The next step is to optimise the hyperparameters of each model in order to give the best possible solution. A Random Search algorithm was used called *model_selection.RandomizedSearchCV* and the optimum parameters are shown in the Table 4.6.

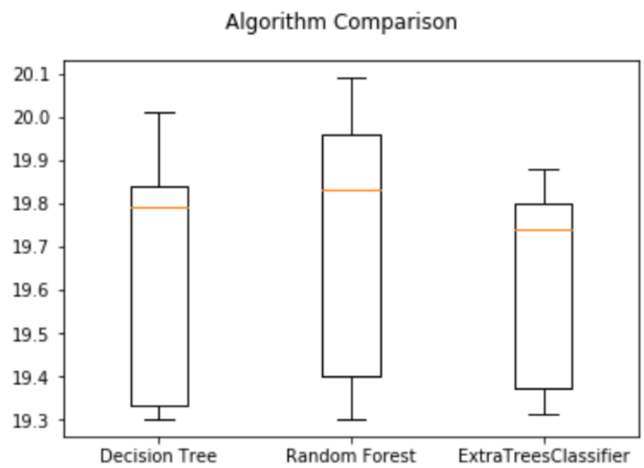

Figure 4.7: Tree-based algorithms comparison using 10-Fold Cross-Validation without tuning nor feature selection

Table 4.6: Optimal hyperparameters using Random Search and error results.

Algorithm	Iterations	Hyperparameters	10-Fold CV Mean Error
Decision Tree	30	criterion='Entropy', max_depth=None, max_features=531, min_samples_leaf=1, min_samples_split=14	20,6%
Random Forest	30	criterion='Gini', max_depth=300, max_features=41, min_samples_leaf=1, min_samples_split=6, num_estimators=930, bootstrap=True	19,4%
Extra Tree	30	criterion='gini', max_depth=300, max_features=1000, min_samples_leaf=1, min_samples_split=13, num_estimators=371, bootstrap=False	19,9%

It is concluded that this search, probably, did not turn out with the best results. The Random Search had worse results than without tuning. This probably happened because the parameters ranges used were too wide to find a good optimal point. To have a clear vision of what is happening, several analyses were done to the Random Forest algorithm to understand how the accuracy and time performance varies with some of the most important hyperparameters of this algorithm. The results shown in the Figures of the Appendix D can elucidate that the accuracy will only increase when there are more than 500 trees and this is where the minimum is. Relatively to the number of features used to build the trees, the accuracy tends to be optimal when it is used 150 features. Both the minimum number of samples required to split an internal node and the minimum number of samples required to be at a leaf node will reduce accuracy and make the algorithm performance worse.

4.5.4 Feature selection

At this time, due to the dummy encoding done to the variables, there are more than 1500 features. This could be a very high source of overfitting and result in bad computing performance. Using as an example the RF model and a feature selection based on the same algorithm, it is shown in Figure 4.8 the behaviour of the mean accuracy and the time it took to train depending on the number of features. It is interestingly noticed that until a threshold of $1E-4$, the accuracy almost does not change meaning that there are several features which are not influencing the algorithm.

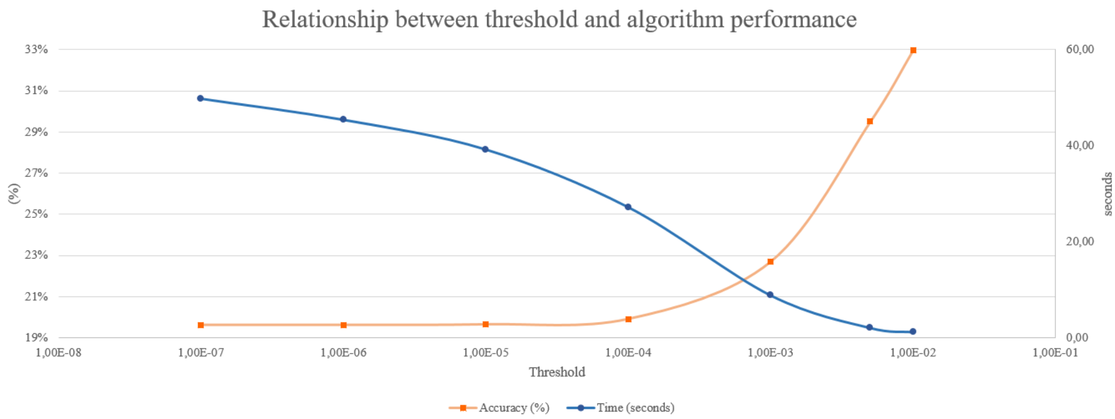

Figure 4.8: Relationship between accuracy and training time performance when changing threshold on a *Random Forest* algorithm

Feature selection was done using *feature_selection.SelectFromModel* and the algorithm selected is the same one used to build the model. Tuning parameters and selecting features were done iteratively until finding an *optimal* solution. On the chapter 5 it is presented the results obtained and conclusions about each algorithm.

4.5.5 Implementation

To understand the impact and confidence level of both algorithms, it was decided that the first algorithm on the Designer ID level should be used to overwrite wrong recommendations. The second step was not yet implemented due to the work required by the Tech team. This step consists in successively include the machine learning algorithm to help the Scan-out team to make decisions.

Chapter 5

Results

The results of this project are divided into two parts correspondent to the two workstreams developed. Firstly, everything related to partner's side will be explored and, afterwards, the focus will be on the refinement of the box suggestion with special emphasis on the results of the algorithms.

5.1 Refine partners' box usage

The current project is evaluated using several KPIs each one evaluating a different process. Firstly, the metrics which evaluate partner's commitment. It is possible to compare directly between the box Farfetch recommends and the real box used by partners in Figure 5.1. Despite seasonality, partner's used boxes and recommendations are converging which means the incentive is serving its purpose as partners are starting to be aware of this recommendation. There is also another factor that should be taken into consideration Farfetch recommendations are also getting better.

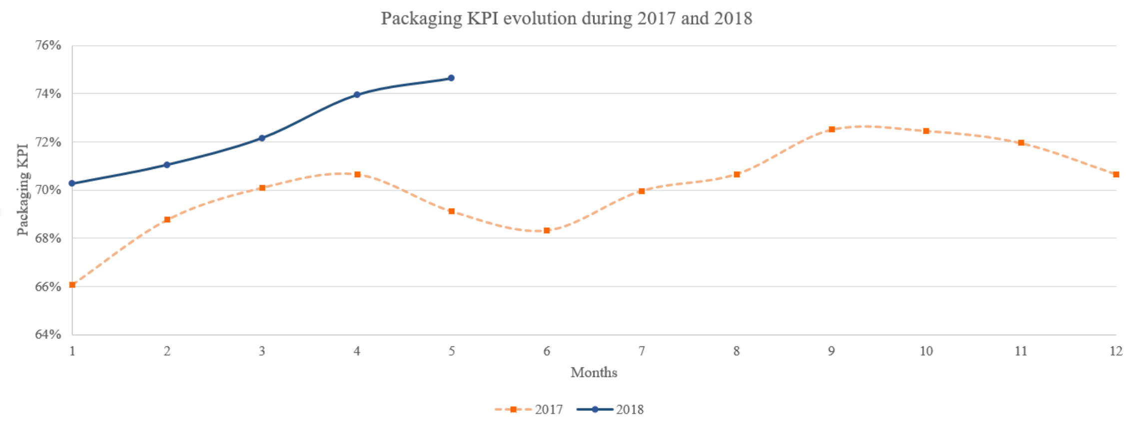

Figure 5.1: Packaging KPI evolution during 2017 and 2018 months.

Another way to evaluate partner's commitment and understand which partners are more committed to this project it is possible to evaluate exceptions' volume. As we can see in Figure 5.2, during the first weeks a peak was reached and, then, it started decreasing. This decrease in engagement happened probably because we started sale season and the amount of work boutiques face during this time is bigger. Thus, the box selection awareness is expectable to be pushed aside.

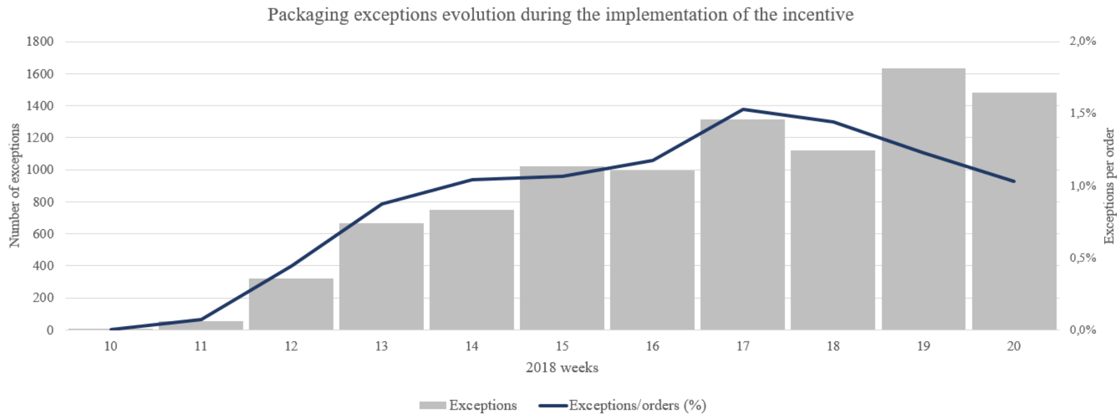

Figure 5.2: Volume of exceptions and percentage of orders with exceptions.

Although the positive results, it is important to highlight that the most committed partners are medium and small boutiques. For big boutiques, only 9% had reached the packaging KPI threshold - 85%.

Additionally, it is possible to measure the average weight (or volume) shipped which clearly shows if the packaging incentive is enhancing box selection. Although this average is decreasing, according to Figure 5.3 we can see, comparing with the results of the year 2017 that this decrease is due to seasonal factors - Spring Summer clothes are lighter than Autumn Winter ones, therefore, require smaller boxes. However, the difference between 2017 and 2018 is too small and it is still too early to take conclusions

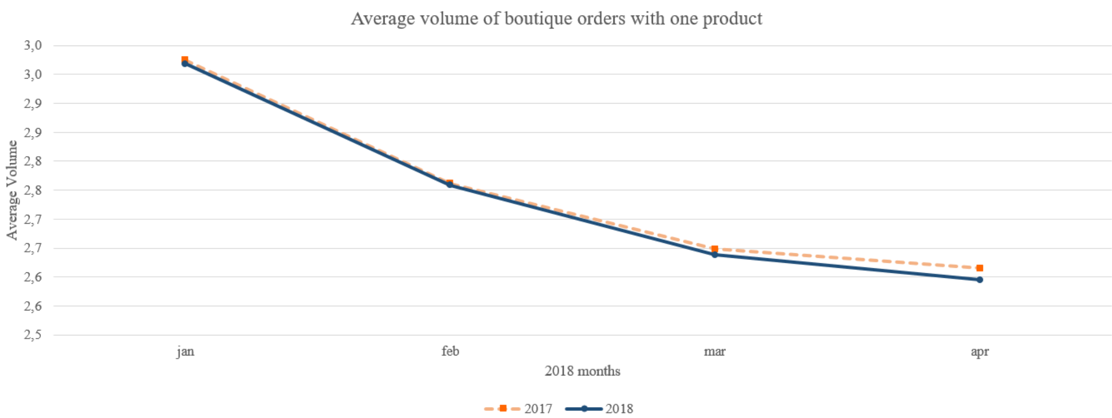

Figure 5.3: Average shipped volume during 2017 and 2018 months.

Apparently, no factors would influence this shipped volume, however after some analysis two insights were discovered:

1. The sales calendar differ slightly between 2017 and 2018 because this year it was launched a particular sale for Autumn Winter 2017 products which could bias the analysis;

2. Trainers orders (9% of Farfetch orders during 2018) have increased 0,25 *kg* on average since last year. After inquiring Marketing team about this big offset it was concluded that this year a new pair of big trainers were best sellers.

To mitigate the influence of these factors, orders from different seasons were split and the category *Trainers* was excluded. Analysing the graph shown in Figure 5.4, it is evident the decrease in the average shipped volume.

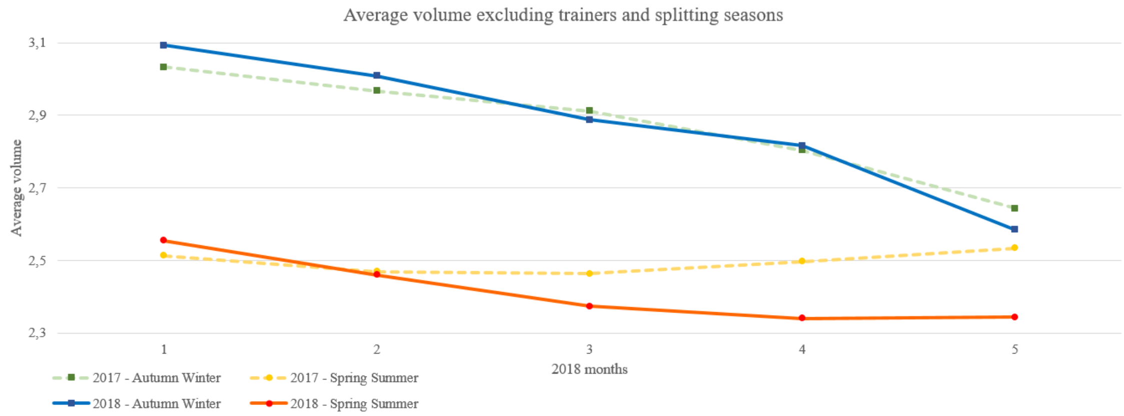

Figure 5.4: Average shipped volume during 2017 and 2018 months from Spring Summer orders (excluding Trainers).

It is also interesting to analyse that partners with a higher Packaging KPI are shipping smaller volumes as illustrated in Figure 5.5. Figure 5.6 shows that the most engaged partners (based on the exceptions raised) are reducing more the volume shipped than those which are less. These two conclusions denote that this incentive is really favouring the partners which are trying to reduce the volume of air shipped.

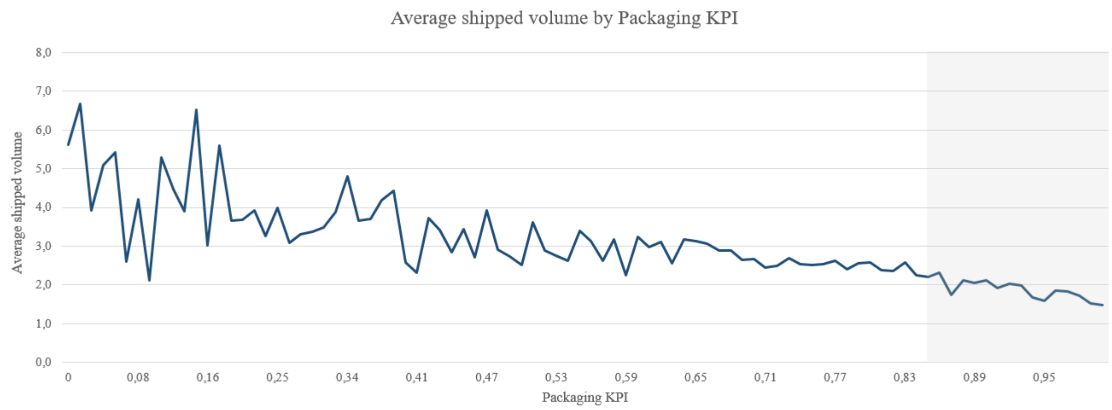

Figure 5.5: Relationship between the packaging KPI and the average volume from partners. Data from March 2018.

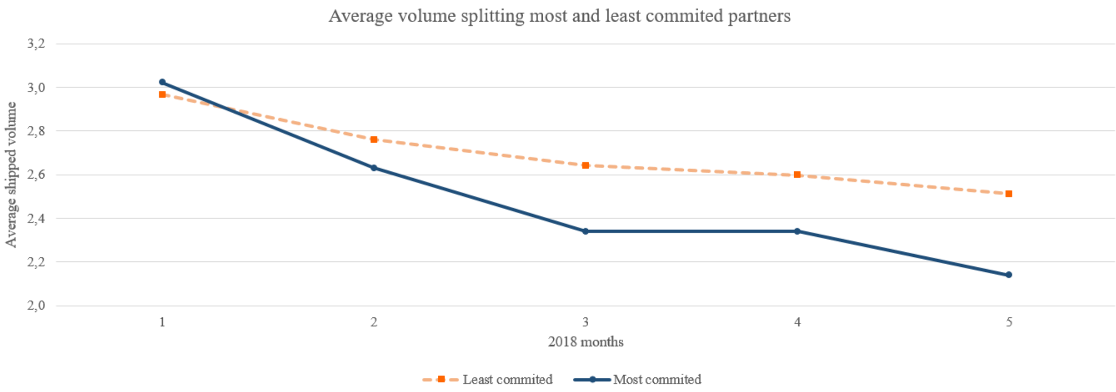

Figure 5.6: Average shipped volume on 2018 from most and less committed partners (based on exceptions raised).

5.2 Improve company's box suggestion

Following some preliminary results shown in chapter 4, we can say that from phase 1 to phase 2, the algorithm's results go from a model with 75% accuracy to 81% with the ability to easily generalise. This increase although not very high could be considered satisfactory.

In the Table 5.1, there are represented the final and best results for the different algorithms, the correspondent mean and variance error of a 10-Fold Cross-Validation, the test error (OSE) and the time performance. It is shown that the Random Forest algorithm with the parameters shown was the winning algorithm with a very good accuracy and a decent time performance.

Table 5.1: Optimal hyperparameters using random search and error results.

Algorithm	N. features	Criterion	max_features	max_depth	n_estimators	min_samples_split	Mean error	Variance	Test error	Runtime (seconds)
Decision Tree	1034	Gini	572	500	-	8	20,2%	0,09	19,8%	3,7
Random Forest	1057	Gini	500	300	930	6	19,1%	0,14	19,2 %	526,3
Extra Tree	1237	Gini	700	300	371	13	20,3%	0,15	23,9%	2360,7

Chapter 6

Conclusion

6.1 Project conclusions

One of the metrics e-commerce companies must be aware of is the amount of air shipped. This represents a waste of money and resources and does not add any value to the customer, quite the opposite, it will impact the unboxing experience. In this project, we covered two different ways of reducing volumes of air shipped: optimise boxes sizes and optimise box selection processes. The focus of this dissertation was in this second point.

To maintain a consistent brand image in a process totally performed and controlled by boutiques and brands, the solution proposed was simple: create an incentive plan which will reward partners who choose the best box for each order. Initiating the project with the solution in mind, the first step was to discover how can this evaluation be structured and define the main parameters which could allow maximise partners commitment and minimise cost. The conclusions were that this was a project with high potential even if parameters needed to be re-tuned with time.

Planning the approach with the main stakeholders was crucial, yet iterative with the execution. Despite the idea of the incentive being simple, the integration with an already consolidated system could be rather complex. The incentive was integrated into the package of incentives given to stores and it was necessary to work together with *Business Intelligence* and *Finance* teams which are responsible for data and financial transactions, respectively. To ensure that partners' interest in this incentive does not fade out, processes to support and hear their feedback were important. Thereat, it was created a channel to let them complaint when a box suggested by Farfetch is not possible to fit a product.

It is possible to conclude that this incentive shown good results on the average shipped weight indicting Farfetch is reducing its carbon footprint, standardising packaging choices and reducing shipping costs. However, on March 20% of the partners are still below 50% on the packaging accuracy and most of the top partners (responsible for the major part of the orders) did not achieve the threshold previously defined. Even though this dissertation finished with promising results, there are still work to do to improve this incentive: find ways to influence big partners and make exceptions process as easy as possible for partners and for internal analysis and processes.

The easiest way to understand whether a used box is the fittest, it is using the recommended box by Farfetch *Scan-out* team. Although it was difficult to measure the precise accuracy of this suggestion, analyses show that many times this is not the best decision. After a visit to the production centre, it was evident that this had been neglected. From this conclusions, it was developed a monthly routine to give them visibility of how they are doing compared to partners' best choices.

Additionally, algorithms were tested to build a consistent way of updating bad suggestions using partners' package choices and to find patterns on this selections. It possible to conclude that there are many factors Farfetch does not control, nor even is aware of them. It is possible to highlight three of them: understand more about brands' boxes which are mandatory to ship with Farfetch box; different folding standards from store to store; big partners do not follow the ordering process as it was expected (two different people pack the products and deal with the Farfetch software).

Using the boxes partners selected when shipping products, two different algorithms were developed: a simple one which grouped these box usage by several pre-selected categories and a fast and relatively simple machine learning algorithm which is based on the most specific and accurate level of the first algorithm and tries to generalise patterns. This second algorithm, a Random Forest, has a better accuracy (81%) than the first one (75%) and it generalises cleverly. Firstly, the first algorithm was used as a system to change wrong recommendations and to validate the changes were good. The next steps will comprehend the implementation of the more sophisticated algorithm as a guide during the choice of the recommendation and, afterwards, to overwrite as the first one.

A more exhaustive analysis using this last algorithm was carried out to understand its sensitiveness to hyperparameter tuning and the number of features selected. Two important factors were analysed and it was discovered that for many of these parameters we cannot achieve a solution where both are optimised, therefore, it requires a trade-off between them. These two factors are Accuracy and Time performance. In general, algorithms with better accuracy will have long training times, the ones with simpler parameters will be faster but their accuracy will not be so good. For this particular application, a compromise between accuracy and time favouring accuracy was decided justified by its application.

Summing up, the project has presented very promising results and partners accepted the packaging incentive better than expected. However, this path is still in its beginnings. It is necessary to raise awareness of all partners for this issues and maintain them committed, tune incentive parameters to optimise costs and commitment and change Farfetch suggestions faster.

6.2 Future Projects

During the realisation of this packaging project and the increase in the knowledge about the area, several opportunities show up as very promising projects for the companies and should be explored.

Firstly, with the exploration of the packaging replenishment process, an idea had arisen to improve the stock control and usage. Farfetch already registers which boxes are partners order because this is done through Farfetch. However, it does not accurately register which boxes are used. To overcome this problem it could be possible to print QR code, barcode or any similar simple system which communicates with the already existing software STORM and accurately records what is being sent. With this system, it would be possible to enhance packaging forecast and develop tools to help partners control their package usage.

Secondly, it could be possible to get more information from products which go to production related to dimensions which will be able to get better information for partners and to predict boxes sizes. This information could be retrieved from photographs with a scale factor or using a person who accurately measures the products and has strict guidelines to avoid non-standardised information.

Thirdly, more profound changes can be performed on the Scan-out team workstation to increase box recommendation accuracy. The proposal would be to mark the sizes of Farfetch packages in the desks used by this team. This way, it would be possible to fold the product and understand which is the fittest box.

All the knowledge acquired during this project was crucial to the development of an area which was almost unknown to the company. Probably, given the opportunities discovered, more projects will arise from this one.

Bibliography

Achille, A., Marchessou, S., and Remy, N. (2018). Luxury in the age of digital darwinism. *McKinsey & Company*.

Arienti, P. (2017). Global Powers of Luxury Goods 2017: The new luxury consumer. Technical report, Deloitte.

Armstrong, A. (2017). Farfetch boss José Neves: 'The magic of bricks-and-mortar shops will never die'. *Telegraph*.

Armstrong, A. (2018). Burberry to expand online reach with Farfetch tie-up. *Telegraph*.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305.

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

Breiman, L., Friedman, J., Stone, C., Olshen, R., and Stone, C. J. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis.

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. Technical report, The CRISP-DM Consortium.

Choi, J.-K., Won, Y., and Kim, J.-J. (2015). Classification prediction of the foot disease pattern using decision tree model. In *Information Science and Applications*, pages 785–791. Springer Berlin Heidelberg.

C.S.-W. (2013). Fetching far flung couture. *The Economist*.

Dauriz, L., Remy, N., and Sandri, N. (2014). Luxury shopping in the digital age. *Perspectives on retail and Consumers Goods. McKinsey & Company*.

Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*, 70:35–36.

Dubois, B., Laurent, G., and Czellar, S. (2001). Consumer rapport to luxury : Analyzing complex and ambivalent attitudes. Les Cahiers de Recherche 736, HEC Paris.

Duda, R., Hart, P., and Stork, D. (2012). *Pattern Classification*. Wiley & Sons.

D'Arpizio, C. and Levato, F. (2017). Luxury goods worldwide market study. millennial state of mind: the tailwind behind consumer behaviors and winning strategies. *Bain & Company*.

Economy, P. (2015). Forget SMART goals – try CLEAR goals instead | inc.com. `https://www.inc.com/peter-economy/forget-smart-goals-try-clear-goals-instead.html`. (Accessed on 05/09/2018).

Fernardez, C. (2017). Farfetch partners with JD.com to boost Chinese business. *Business of Fashion*.

Fortmann-Roe, S. (2012). Understanding the bias-variance tradeoff. `http://scott.fortmann-roe.com/docs/BiasVariance.html#fnref:1`. (Accessed on 06/07/2018).

Fraser, K. (2016). Farfetch valued at one billion dollars, hits one millionth customer. *Fashion United*.

Gashler, M., Giraud-Carrier, C., and Martinez, T. (2008). Decision tree ensemble: Small heterogeneous is better than large homogeneous. In *2008 Seventh International Conference on Machine Learning and Applications*, pages 900–905. IEEE.

Geurts, P., Geurts, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63:3–42.

Harris, L. and Dennis, C. (2002). *Marketing the E-business*. Routledge Textbooks in Ebusiness, 2. Routledge.

Ho, T. K. (2002). A data complexity analysis of comparative advantages of decision forest constructors. *Pattern Analysis & Applications*, 5:102–112.

Holt, D. B. (1995). How consumers consume: A typology of consumption practices. *Journal of Consumer Research*, 22(1):1–16.

Jones, E., Oliphant, T., and Peterson, P. (2001). Scipy: Open source scientific tools for python.

Karandish, F. and Šimůnek, J. (2016). A comparison of numerical and machine-learning modeling of soil water content with limited input data. *Journal of Hydrology*, 543:892 – 909.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*.

Kotu, V. and Deshpande, B. (2014). *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Elsevier Science.

McAfee, A., Brynjolfsson, E., Davenport, T., Patil, D., and Barton, D. (2012). Big data. *Harvard Business Review*, 90(10):61–67.

McKinney, W. (2010). Data structures for statistical computing in python. In van der Walt, S. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56.

Milosevic, N., Dehghantanha, A., and Choo, K.-K. (2017). Machine learning aided android malware classification. *Computers and Electrical Engineering*, 61:266–274.

Okonkwo, U. (2009). The luxury brand strategy challenge. *Journal of Brand Management*, 16(5):287–289.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Project Management Institute (2017). *Guide to the Project Management Body of Knowledge (PMBOK® Guide)–Sixth Edition*. PMBOK® Guide. Project Management Institute.

Rao, A. R. and Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of Marketing Research*, 26(3):351–357.

Rockett, J. (2016). A Farfetch splurge. `https://www.jayerockett.com/2016/07/a-farfetch-splurge.html`. (Accessed on 06/07/2018).

Rodriguez, S. (2015). The real reason everyone calls billion-dollar startups 'unicorns'. *International Business Times*.

Sanders, N. (2013). *The Definitive Guide to Manufacturing and Service Operations: Master the Strategies and Tactics for Planning, Organizing, and Managing How Products and Services Are Produced*. Council of Supply Chain Management Professionals. Pearson Education.

Shahinfar, S. and Kahn, L. (2018). Machine learning approaches for early prediction of adult wool growth and quality in Australian Merino sheep. *Computers and Electronics in Agriculture*, 148:72 – 81.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

Vigneron, F. and Johnson, L. W. (2017). *Measuring Perceptions of Brand Luxury*, pages 199–234. Springer International Publishing, Cham.

White, S. and Denis, P. (2018). Chanel, Farfetch pair up for digital push at fashion label's stores. *Reuters*.

Williams-Grut, O. (2017). Revenue jumps 70% at online fashion retailer Farfetch. *Business Insider*.

Zhang, Y.-D., Wang, S., and ji, G. (2013). A rule-based model for bankruptcy prediction based on an improved genetic ant colony algorithm. *Mathematical Problems in Engineering*, 2013:1–10.

Appendix A

Decision Tree

The idea behind Decision Tree algorithm is to split the data based on its homogeneity. By splitting the data the objective is to get purer and purer groups. Purity can be visualised by the percentage of records with the same output. The criteria used to decide if a node should be split or not is the purity a division will create, therefore, this purity or impurity must be mathematically defined with these two conditions:

(a) Impurity must be at its maximum when all outputs have the same probability of happening in given data set;

(b) Impurity must be at its minimum - zero - when only one output is represented in the dataset.

There are two measures used which match these two conditions: *entropy* and *Gini index*.

If there are T events with an equal probability of occurrence P, then $T = 1/P$. Entropy H is defined by Shannon (1948) as $log_2(1/P)$ or $-log_2(P)$. However, when the probability of all events is not identical, it is necessary to add a weighted expression as represented in the equation A.1. The k will take as many values as the number of output classes, p_k is the probability of the event k.

$$H = -\sum_{k=1}^{n} p_k log_2(p_k) \tag{A.1}$$

To validate the conditions above-mentioned, an example of a two target variable - *Yes* and *No* - with the same probability will give the maximum entropy $H = -[0.5log_2(0.5) + 0.5log_2(0.5)] = 1$. On the other side, if only one class on the data represented $H = -[1log_2(1) + 0log_2(0)] = 0$. All other combinations will give value between these two limits.

Using the same nomenclature to define the Gini index G we obtain the equation

$$G = 1 - \sum_{k=1}^{n} (p_k^2) \tag{A.2}$$

With the example used to validate the limits of entropy we can define the maximum of Gini index as $G = 1 - (0.5^2 + 0.5^2) = 0.5$ and the minimum as $G = 1 - (1^2 + 0^2) = 0$.

Now that we already have a mathematical way to evaluate a data split, it is important to explain how the algorithm decides when it will do the split and when will it stop splitting. Two important concepts need to be introduced *Information Gain* and *Information*. Information Gain IG is used to evaluate the reduction of entropy when you divide data. For a certain predictor, it is obtained computing the difference between the information contained on not split data $I_{pred,notpartitioned}$ and the information if you split using that predictor I_{pred} (equation A.3).

$$IG_{pred} = I_{pred,notpartitioned} - I_{pred} \tag{A.3}$$

The information before splitting data is given by the information entropy as shown in the equation A.4.

$$I_{pred,notpartitioned} = H_{pred} \tag{A.4}$$

The information after splitting using the predictor *pred* with n different values i (for example an example where the gender is a predictor it has two different values: male and female) is given by the weighted sum of the different entropy's H times its probability p.

$$I_{pred} = \sum_{i=1}^{n} p_{pred:i} * H_{pred:i} \tag{A.5}$$

For each independent variable, the model computes the IG and chooses the split which will obtain the bigger amount of information. This process is done at every node until we stop splitting: pruning. If the entropy of a node is zero we can terminate the divisions. Nonetheless, if this is the only criterion we will end up overfitting our model and, thus, there are three ways of preventing this:

1. Set a minimum information gain threshold;
2. Define a maximum depth;
3. Specify a minimum number of examples.

Appendix B

Project's Business Case

Business case: Packaging Accuracy KPI for Incentive Service

Product context and objectives			
Who is our customer? Farfetch Partners	**What problem are we trying to solve?** Data integration with Incentive Service	**Objective** Include Packaging Accuracy KPI in Incentive Service to add a new incentive to our partners	**Challenges/needs** Calculate Packaging KPI every month. Make it available to the Incentive Service through an API.

Main requirements identified
• Calculate Packaging Accuracy KPI every month • Make it available through an API for the incentive service to consume.
Business/technical areas impacted
Business: Delivery, Supply & Retail Operations, Partner Services **Technical:** BI, Finance
Key dates identified
• March 2018 before Incentive is launched

Main benefits
• Business case: xxx€ net savings in delivery costs (post-incentive) • Assumptions: improvement in KPI from 70 to 85% • Supporting strategic theme: Build our Brand (Because packaging will be improved with this savings)
How we will measure success
• Success will be improving the packaging accuracy KPI (currently 70%, goal: 85%)

FARFETCH

Figure B.1: Packaging Incentive Business Case

Appendix C

Routines

During the implementation of the current project, four main recurrent processes were created:

- Transfer packaging incentive;
- Update package suggestion based on the used boxes;
- Meeting with Partner Service;
- Meeting with Production.

In Figure C.1, a timeline shows how will these processes appear every month and what is the responsible team. This timeline is relative to the orders of month 0.

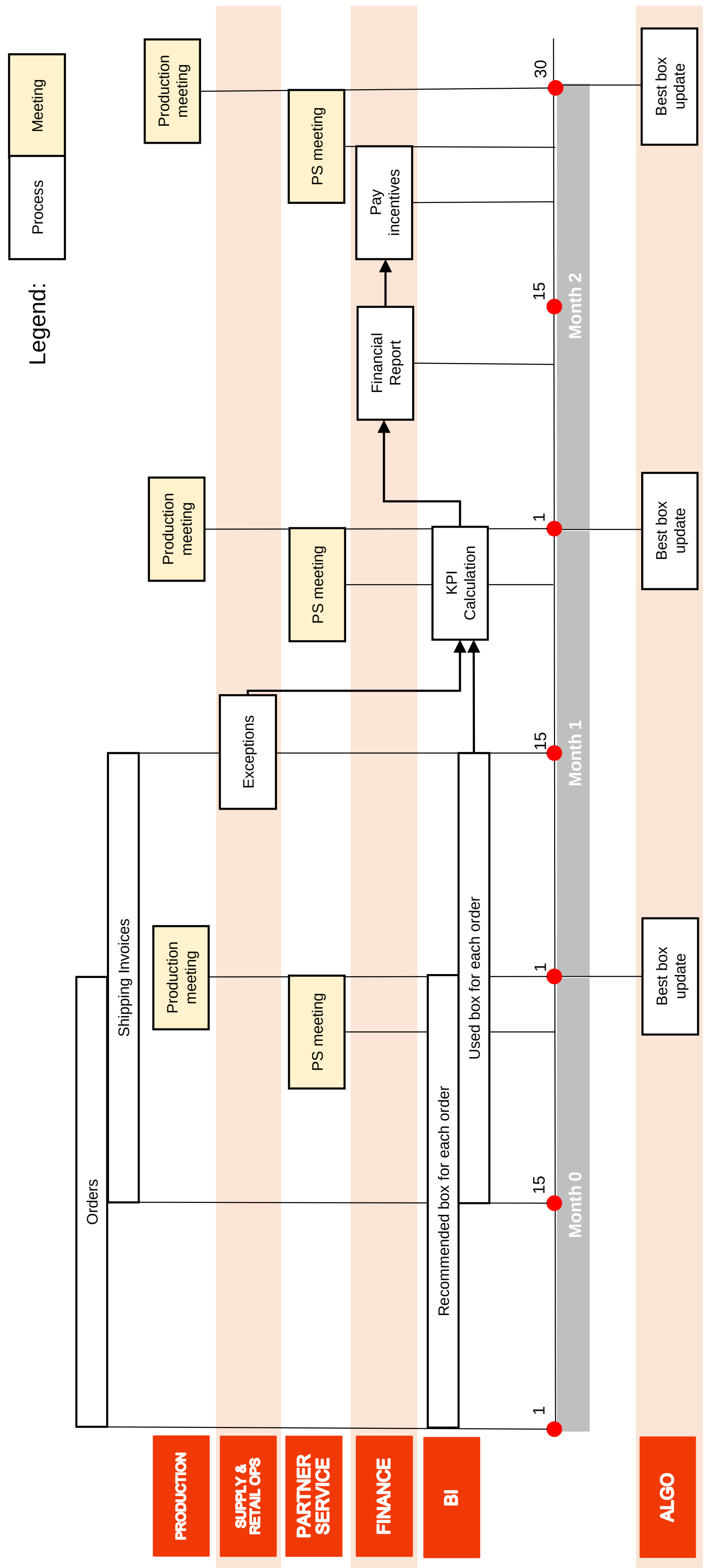

Figure C.1: Packaging monthly routines

Appendix D

Random Forest algorithm hyperparameters

When defining the ranges for the random search, several analyses were done using the most important hyperparameters of a Random Forest algorithm. These analyses aimed to discover which are the most promising ranges concerning time and accuracy performances. For each one of the analysis, the model was trained using all the default parameters, except the hyperparameters which were being analysed. The results are presented in the Figures D.1 to D.4.

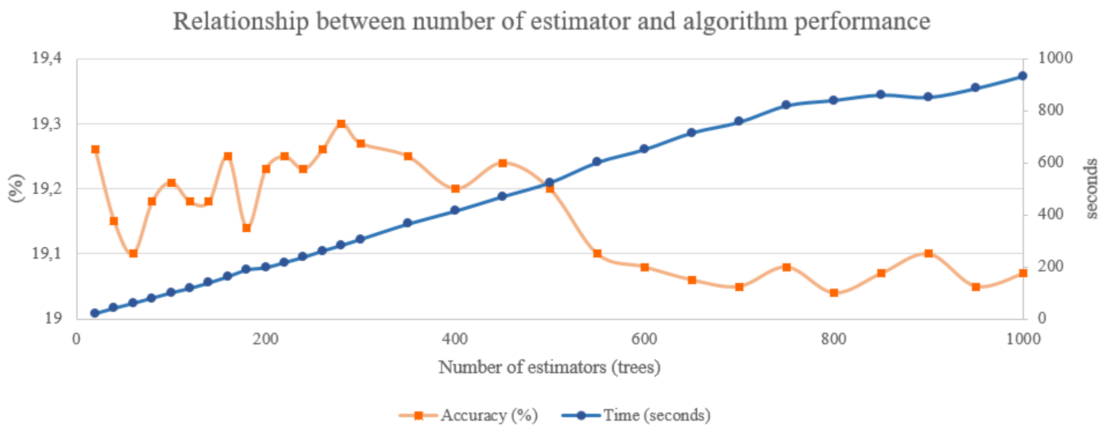

Figure D.1: Relationship between accuracy and training time performance when changing number of estimators on a *Random Forest* algorithm

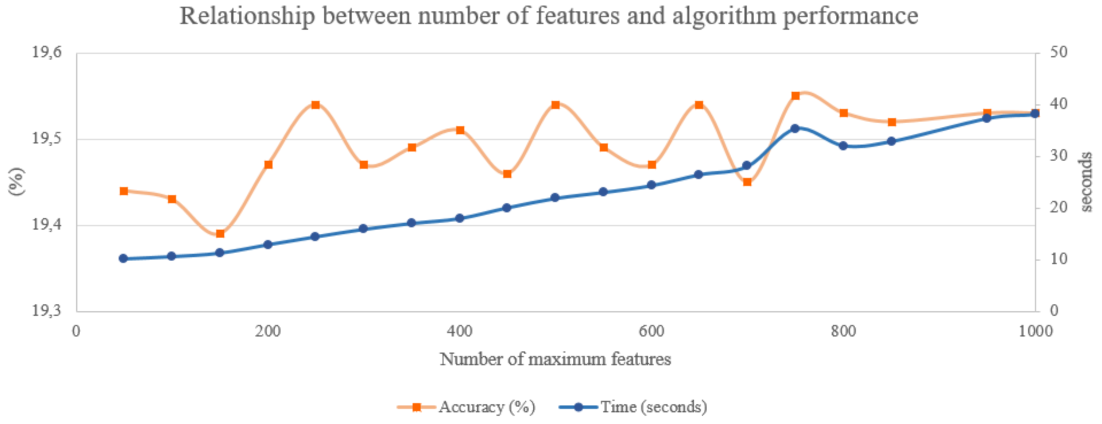

Figure D.2: Relationship between accuracy and training time performance when changing number of features on a *Random Forest* algorithm

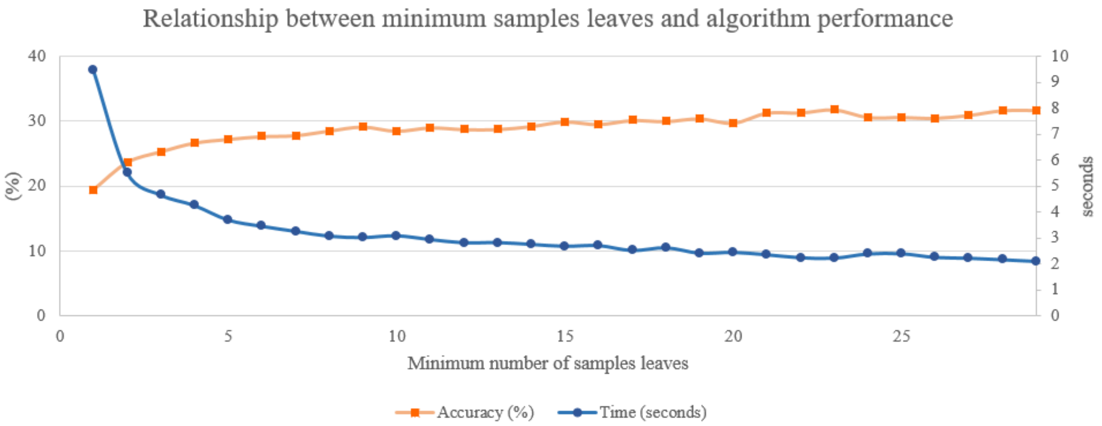

Figure D.3: Relationship between accuracy and training time performance when changing the minimum number of samples required to be at a leaf node on a *Random Forest* algorithm

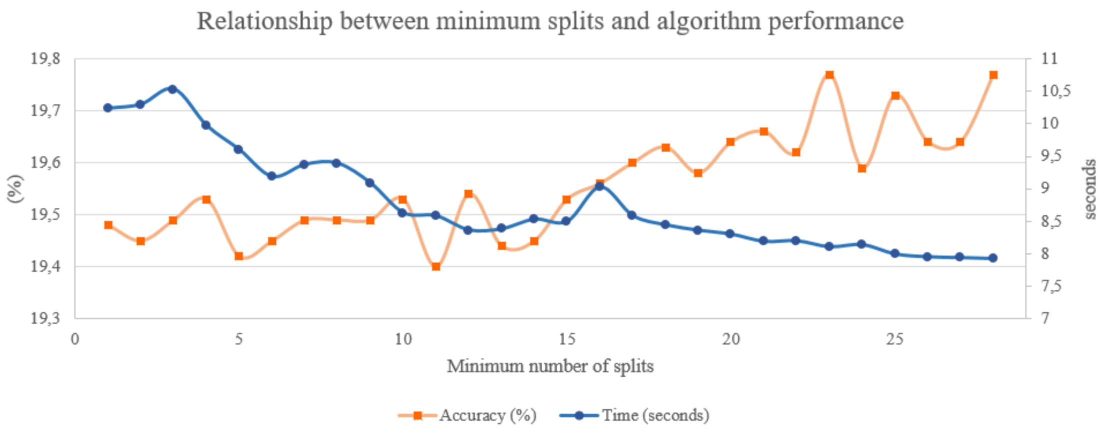

Figure D.4: Relationship between accuracy and training time performance when changing the minimum number of samples required to split an internal node on a *Random Forest* algorithm