A Payments Routing Heuristic based on Machine Learning for High Volume E-Commerce Environments

Tiago Cardoso dos Santos Pinto de Sousa

Master's Dissertation

Supervisor: Prof. Mário Filipe Amorim Faria de Oliveira Lopes

U. PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia e Gestão Industrial

July 2018

Abstract

E-commerce is revolutionizing the way business is done. The possibility of reaching anyone in the world through the internet and the convenience of shopping online are crucial factors to its exponential growth. It is expected that e-commerce sales will increase by more than 110% worldwide, reaching a value of 4,88 trillion US dollars in 2021. The opportunities for improvement represent huge potential gains for companies.

Payments play a vital role in e-commerce. They are the facilitators of revenues to the companies, having a huge impact in its results, and a crucial element for customer satisfaction, enhancing (or not) its experience. Although the result of an e-payment is pretty simple, the payment either is accepted or rejected, the underlying system is quite complex as it involves several entities playing different roles in the system. Many times, these entities are not connected in the most optimal way, creating major inefficiencies in the payments processing system. This thesis arises in the context of a luxury fashion e-tail company with the main goal of reducing those inefficiencies. To tackle this problem, the current payment system was analyzed and mapped in detail. After detailing its limitations and opportunities, a payments routing solution was proposed in order to maximize the successful payments and to minimize processing costs.

To accomplish that, a payments routing heuristic was formulated, being composed of three parts: a success part, whose output is a probability of success for a given transaction in a specific payment route; a cost part, where the output is an estimation of the cost for a given transaction in a specific payment route; and a part that transforms the preferences of the stakeholders into weights that are incorporated in the heuristic. The final output is an ordered list of possible payment routes for a given transaction.

Due to the big quantity of data available in e-commerce companies, the success part is calculated resorting to a machine learning approach. Here, after choosing the correct variables and the proper training, validation and test sets, a model tuning, evaluation, and selection framework is created in order to choose the best technique for the problem at hand. The cost part is the simplest one, relying only on the estimation of the cost with the use of cost datasets. The part regarding the criteria weights is done using a multi-criteria decision analysis that assigns weights to defined criteria according to the stakeholder's interests. The formulated heuristic can be applied in any payments routing situation, with preference to high volume e-commerce environments.

After having formulated the heuristic, it is deployed in the company being studied. We start by describing how to properly tune and adjust the heuristic. Afterward, its integration within the company's systems and processes is explained and some examples are given to fully understand the heuristic and its implementation. As there was no time for implementation during the project, a simulation of the results was done, where the current payment system is compared to the proposed one. The results of using the heuristic are encouraging, but they should be analyzed with caution. Besides that, a close monitoring should be performed during its implementation. Finally, the main takeaway of this project is that making smart routing decisions can have tremendous gains in the company's results apart from increasing customer satisfaction.

Resumo

O *e-commerce* está a revolucionar a forma como os negócios são feitos. A possibilidade de alcançar qualquer pessoa no mundo através da *internet* e a conveniência de comprar *online* são fatores crucias para o seu crescimento exponencial. É esperado que as vendas de *e-commerce* aumentem mais de 110% em todo o mundo, atingindo um valor de 4,88 triliões de US dólares em 2021. As oportunidades de melhoria representam ganhos potencias consideráveis para as empresas.

Os pagamentos, por sua vez, têm um papel preponderante no *e-commerce*. Eles contribuem para as receitas das empresas, tendo um impacto enorme nos seus resultados, e um elemento importante para a satisfação dos clientes, melhorando (ou não) a sua experiência. Ainda que o resultado de um pagamento *online* seja bastante simples - o pagamento ou é aceite ou é rejeitado - o sistema subjacente é bastante complexo, pois envolve diversas entidades que desempenham diferentes funções no sistema. Muitas vezes estas entidades não estão ligadas da melhor maneira criando grandes ineficiências no sistema de pagamentos. Esta dissertação surge no contexto de um retalhista *online* de luxo com o objetivo de melhorar essas ineficiências. Para isso, o atual sistema de pagamentos é analisado em detalhe. Após clarificar as suas limitações e oportunidades de melhoria, é proposta uma solução para o roteamento de pagamentos com o objetivo de maximizar os pagamentos com sucesso e de minimizar os custos de processamento.

Assim, formulamos uma heurística para este roteamento constituída por três partes: a primeira referente ao sucesso da transação, cujo resultado é a probabilidade de sucesso de uma transação numa dada rota de pagamento; a segunda referente ao custo, cujo resultado é uma estimativa do custo da transação para a mesma rota; e uma parte que transforma as preferências dos *stakeholders* em pesos que são incorporados na heurística. O resultado final é uma lista ordenada de possíveis rotas de pagamento para uma dada transação.

Considerando a enorme quantidade de dados disponíveis em empresas de *e-commerce*, a parte referente ao sucesso é calculada recorrendo a técnicas de *machine learning*. Assim, depois de escolher as variáveis a usar e os corretos conjuntos de treino, validação e teste, um procedimento foi criado para selecionar a melhor técnica para o problema. A parte referente aos custos é a mais simples, dependendo apenas da estimativa de custos. A parte referente aos pesos é feita recorrendo a uma técnica de decisão multi-critério, que atribui pesos a diferentes critérios de acordo com os interesses dos *stakeholders*. A heuristíca pode ser aplicada em qualquer problema de roteamento de pagamentos, com preferência para ambientes de *e-commerce* com elevado volume transacional.

Depois de ter formulado a heurística, ela será implementada na empresa em estudo. Iniciamos pela descrição, calibração e ajuste da heurística, explicando posteriormente a sua integração nos sistemas e processos da empresa, onde alguns exemplos são dados de forma a perceber a sua implementação. Dada a falta de recursos para implementar a heurística, uma simulação foi feita onde o atual sistema é comparado com o sistema proposto. Os resultados são encorajadores mas estes devem ser analisados com cautela. Por fim, destaca-se o impacto que o roteamento inteligente de transações pode ter nos resultados das empresas e na melhoria da satisfação dos clientes.

Acknowledgements

I would like to begin by thanking Farfetch for giving me the opportunity to develop this project. I am particularly grateful for the assistance given by my supervisor and friend João Barros who guided me from day one, giving me constant support and important feedback. His patience in teaching me the intricate elements of this project, his availability to revise my work and his push for innovation were vital to the completion of this dissertation.

I would also like to thank Mark Hobbs for welcoming in the team, for making me feel one of them since the first day, for the counseling and patience during this time. A word of appreciation to Luis Ferreira for having readily clarified all my doubts and thoughts; to João Gomes and Hugo Pinto for the important technical inputs and ideas discussed with me; to the ones that started this journey with me and to all Farfetchers that helped me at some point.

This last five years were definitely essential to this accomplishment. I would like to appreciate all my professors for their role in this learning journey. My grateful thanks are also extended to my dissertation supervisor, Prof. Mario Amorim Lopes, whose help was invaluable for the completion of this document. The value of this dissertation is for sure majored by his guidance, helpful recommendations and constant availability. Special thanks to Prof. Vera Miguéis for the friendliness, constant willingness to help and important technical support.

To my friends, I express my gratitude for the countless memories, friendship and support throughout these years.

Last, but not least, I would like to thank my family, your education, friendship and love shaped the person I am today. A sincere thanks to my parents Angela and Paulo, my sister Inês, my aunt Paula, my grandfathers Mário and Francisco, and to Ana for the unconditional support, care and encouragement, without which none of this would be possible.

"If you want to be inventive - you have to be willing to fail"

Jeff Bezos

Contents

Acronyms and Symbols

AHP	Analytic Hierarchy Process
AMEX	American Express
API	Application Programming Interface
AR	Authorization Rate
AUC	Area Under Curve
B2B	Business-to-Business
B2C	Business-to-Customer
BIN	Bank Identification Number
BR	BRazil
C2C	Customer-to-Customer
CAGR	Compound Annual Growth Rate
CB	Carte Bancaire
CB	Japan Credit Bureau
CRISP-DM	CRoss Industry Standard Process for Data Mining
CUP	China Union Pay
DACH	Deutschland(D), Austria(A) and Switzerland (CH)
DEO	Director of E-commerce Operations
EDI	Electronic Data Interchange
EU	Europe
GBM	Gradient Boosting Machine
MCDA	Multi-Criteria Decision Analysis
ML	Machine Learning
POV	Point of View
PSP	Payment Service Provider
PTC	Payments Team - Costs
PTM	Payments Team Manager
PTP	Payments Team - Performance
ROC	Receiver Operating Characteristics
ROW	Rest Of the World
TID	Transaction IDentifier
US	United States

List of Figures

List of Tables

Chapter 1

Introduction

"To offer a quick, easy and reliable buying experience, at the right price to the company."

Payments team mission

1.1 Motivation

Retailing is experiencing a profound transformation with the rising dominance of e-commerce. Although e-commerce, as a concept, was created more than twenty years ago, its use in recent years has been increasing exponentially with no signs of slowing down. It has become the new standard in the way people purchase goods, being considered of major importance to the companies' strategy.

A market researcher (Statista, 2018) claims that from 2014 to 2017 there was an increase of more than 70% in global e-tail sales worldwide, with the estimation of growth of more than 110%, projecting a global sales value of 4,88 trillion US dollars in 2021. The humongous opportunities presented by e-commerce are accompanied by a fiercely increase in competition. E-tail companies have, all the more, to innovate and improve the efficiencies of its core processes.

This thesis arises in this context, of reducing inefficiencies and innovating the current processes of the online luxury fashion retailer, Farfetch, and consists in the development of a complete routing heuristic for online payments. E-payments are a crucial challenge for e-commerce business due to its critical impact on sales, and in turn on the profitability of a company. Moreover, the importance of e-payments is majored by its fragile role and the trust in them deposited by the customers.

1.2 Company description

Farfetch is a global luxury fashion e-commerce platform that links customers with an exquisite global network of boutiques and brands. It allows customers to buy from their favorite boutiques, from any place in the globe, that were inaccessible in the past. This universal approach of offering a streamlined online shopping technology to empower e-commerce is unrivaled, granting boutiques access to an e-commerce marketplace.

Farfetch differentiates from other fashion e-commerce companies through its business model. All the items sold through its website are owned and kept by the boutiques themselves. The company's service operates as a bridge between its partner boutiques and customers. Everything, from photographing the products for the catalog to post-sales customer service, is offered as part of the Farfetch service. Within the operational part of sales, boutiques only need to pack the items. Each product sold is collected by an external courier provider and directly delivered to the customer's address.

The company has grown at a rapid pace in its ten years of existence, becoming the first Portuguese "unicorn" start-up. Recently Farfetch raised almost 400 million dollars in a partnership with the online Chinese mall JD and signed an important partnership with the Chalhoub group, which operates in the luxury segment in the Middle East.

Farfetch has a complex business organization and operates in twelve offices spread around seven different countries. The dissertation was developed in the Operations department which is composed of the following main areas: E-Commerce Operations, Supply, Operations Strategy and Black and White Operations. E-Commerce Operations is subdivided into Delivery (Support and Development), Fraud, Payments and Premium Services. The Payments team main responsibilities are to thoroughly analyze and monitor the whole payment process, study different market opportunities and give support to other teams regarding payments. The team is also accountable for communicating with the different external entities involved in the payment process and to assess the impact of the several projects and actions in the results of the company.

1.3 The Project

Daily, thousands of customers place new orders at Farfetch. The order placement is, in practice, a transaction attempt that has two possible outcomes: the transaction is accepted and the order processed, or the transaction is rejected and the order canceled.

Although the final output is pretty straightforward, the process since the order is placed to when the final output is known is a very complex process involving several entities that exchange information among themselves in order to provide to the merchant (Farfetch in this case) the result of the transaction.

This kind of transactions is global due to the countless entities, from different countries, involved. The complexity generated by the deviation in the legislation and the inconsistency in the procedural practices is many times responsible for transaction failures. Additionally, there are several other reasons responsible for the payments failure such as customer issues, fraud suspicions, internal errors and other similar.

In January 2017 there were almost 300 thousand attempted transactions in the Farfetch platform. When looking at the same month in 2018, the value increases almost 50% to around 450 thousand transactions.

The immense amount of transactions suggests firstly that the current inefficiencies are going to be augmented, increasing the opportunities for improvement, and secondly, that there is a massive amount of meaningful data that can be used to take insights for the business.

Even though Farfetch actual system was built as a response to the company's necessities and strategic options, the system is not exploiting its full potential. Given that, this project aims to study the whole payment system and the way it is integrated within the company's processes, to understand what are its main inefficiencies and to find out the root cause of the refused transactions. Afterward, these analyses will support the development of a new system, efficient and dynamic, that is capable of improving the operational results of the company.

1.4 Goals and Methodology

The major goal of this dissertation is to formulate an online payments routing heuristic that is efficient, maximizing the authorized transactions and minimizing the costs; flexible and dynamic, coping with the different stakeholder's interests and being able to face constant improvements; and well-structured, allowing its use in the long-term.

To achieve this, a structured methodology has to be adopted. Initially, a robust and extensive analysis of the current payment system is performed. This analysis includes the mapping of the current workflow and the study of each entity involved and its impact on the whole system. After that, the identification of all the refused reasons received by Farfetch is done, in order to understand the kind of issues that have the most impact in the process. Having completed this analysis, the next step would comprise an assessment of all the costs involved in each transaction and how the different entities impact them. Finally, having collected and treated all the necessary data, a heuristic that is able to select the best possible payment route for each new transaction is going to be formulated. The project timeline can be observed in Figure 1.1.

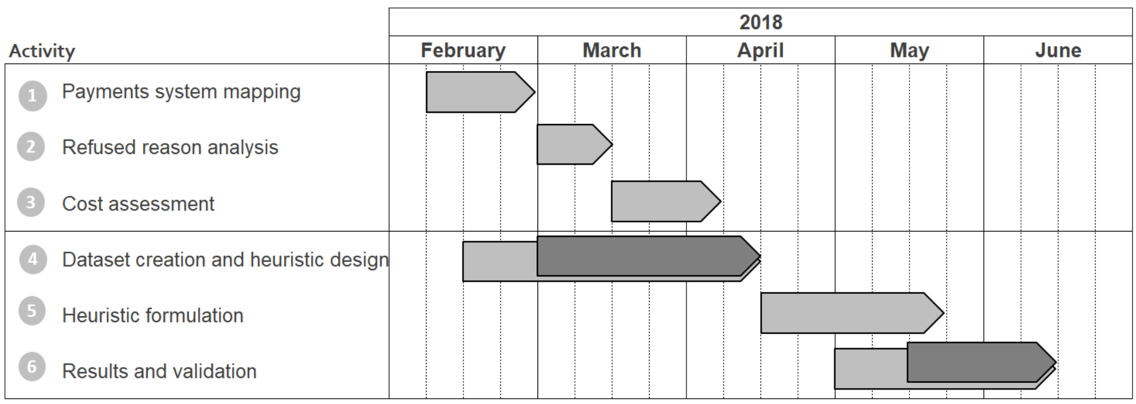

Figure 1.1: Project's timeline

1.5 Dissertation Structure

This thesis is organized into six chapters. Its structure is as follows:

Chapter 1 introduces the thesis, where the project is contextualized and the main objectives are defined.

Chapter 2 aims to provide a theoretical background on the pertinent subjects for this thesis. Firstly, the characteristics of the e-commerce retail and luxury markets are described. Secondly, a literature review of online payment systems is conducted. A study of data mining techniques and machine learning algorithms is done afterwards. Finally, a brief investigation on Multi-Criteria Decision Analysis (MCDA) techniques is performed.

Chapter 3 gives a detailed description of Farfetch's internal processes and how the current payment system was built, indicating its major limitations and opportunities for improvement.

Chapter 4 describes the methodology to follow. Firstly it focuses on the creation of a Machine Learning (ML) algorithm that can predict the best possible route for a specific transaction, identifying all the variables used, defining the datasets, tuning and evaluating the different models, concluding in the selection of the best model according to its performance. The second part targets the cost modeling, detailing the procedure to estimate the cost of a new transaction. The third part includes a brief explanation of the MCDA methodology that is used to transform the stakeholder's interests in criteria weights to be used in the final heuristic. Finally, the model formulation is described and the heuristic output is explained.

Chapter 5 describes the deployment of the proposed heuristic in the studied company. We start by calibrating the heuristic according to the problem at hand. After that, its implementation within the company's internal processes and systems is explained. Then, the focus is on the expected results and how one would use the heuristic in a correct way. Finally, a simulation is done in order to understand the impact of the proposed heuristic on the company.

Chapter 6 gives a small overview of all the conclusions taken along the dissertation and future enhancements to the work developed are discussed.

Chapter 2

Literature Review

The present chapter aims to provide a literature review on relevant topics within the scope of this dissertation. Firstly, the current position of luxury fashion industry towards e-business is explored. A detailed description of the way electronic payments are processed in this business and the wide range of concepts associated with it will be presented soon after. Afterward, a brief study on the current machine learning approaches and the key concepts with them associated is performed. Finally, an analysis of the MCDA approaches with a special focus on Analytical Hierarchy Process (AHP) is executed.

2.1 Luxury E-tail

2.1.1 E-commerce

Electronic commerce refers to the transaction of goods and services through electronic communications and has its origins in the implementation of EDI (Electronic Data Interchange) (Tian and Stewart, 2006). In fact, it dates back to the late ‘80s when the internet began to penetrate the lives of millions of users around the globe. Since then, internet and e-commerce had followed a similar road, with innovations in the field of internet technologies having important repercussions in the online business world (Mirescu et al., 2010).

The internet technologies and the knowledge on how to implement them in the business area is the starting point for the concept of e-business. As stated by IBM (1997), “e-business can be the key to transforming business processes using internet technologies”.

According to Niranjanamurthy et al. (2013), the use of mobile devices to conduct business transactions – known as m-commerce - is advancing step by step, alongside the classic instruments of the internet commerce. The main concepts related to e-commerce are mainly the business-to-business (B2B) and business-to-customer (B2C). New concepts like customer-to-customer (C2C) are also starting to become more important.

Regarding the advantages of e-commerce activities those include, from the customers’ point of view: quick shopping, fast delivery and high convenience, the possibility to perform a more effective comparison between different products and the possibility to buy/sell 24/7; and from the producers' point of view: no need of physical company set-up, the opportunity to attract new customers, the easy access to markets that were inaccessible otherwise and the ability to scale up rapidly (Mirescu et al., 2010). Regarding the disadvantages, from the customers' point of view, these are the inability to experience the product before purchasing it and the delay in receiving the goods. From the producers' point of view, these are mainly related to the security issues connected to transactional processes and the lack of the necessary infrastructures (Niranjanamurthy et al., 2013).

Another relevant aspect is related to the customer relationship with the merchant in online business. The customer became powerful, has easy access to information and exchanges it with other customers from other companies which makes the customer opinion on the company extremely important, turning customer satisfaction into a key figure for business success (Evans, 2001).

2.1.2 Luxury Fashion E-tail

The term luxury or luxury brand is hard to define. Because of its unique characteristics, it represents a very specific niche that is not recognized as a product or service, not even a concept but as "an identity, a philosophy and a culture" (Okonkwo, 2009). The literature largely defines luxury brands based on consumer perceptions and/or managerially determined dimensions such as marketing activities and product attributes. However, some characteristics are widely referred when talking about luxury brands such as high quality, rarity, premium pricing and high level of aesthetics.

"My Luxury", as defined by Kapferer (2012), has a different meaning, being referred to a small personal luxury purchase. This notion is an example of the well-known phenomenon whereby individuals purchase affordable luxuries as a substitute for more expensive items.

Finally, luxury can also be seen as a business model empirically fine-tuned over time by luxury brands that dominate worldwide, such as Louis Vuitton, Chanel, Gucci, Herme‘s, Ferrari, and Rolex. These business models run contrary to most present business models in any sector. It rests on strict principles that maintain the uniqueness of luxury and preserve the non-comparability of those luxury brands (Kapferer, 2012).

The luxury fashion industry deals with some paradoxes that should be treated carefully. The most relevant examples are the struggle between creating "desire and exclusivity" in a wider range of customers, maintaining the equity of the brand. Another important conceptual dispute lies in the wish of increasing sales volumes without the risk of overexposure. These are some of the arguments mostly used by skeptical brands to enter the online market, placing internet in the opposite position when referring to the core elements (Okonkwo, 2009).

According to Castillan et al. (2017), although the commercial strategies of a luxury brand are mainly the same as a regular brand, it is quite challenging to transfer the same atmosphere of a physical store into an online store. Luxury brands have many concerns such as keeping the exclusive image of the luxury brand, online luxury services and staying coherent with the brand image. All in all, luxury companies have to uphold a good balance between tradition and modernism of their brand to improve and satisfy consumers' shopping experience.

What online customers want is a website experience that is pleasant, interactive and engaging while maintaining the prestigious atmosphere of a boutique with its multi-sensory experience (Bjørn-Andersen and Hansen, 2011). The most required features for a successful online channel (from client POV) are undoubtedly aesthetics, communication by e-mail, information on products and easiness to navigate on the portal (Hines and Bruce, 2007). Besides that, studies show that customers tend to abandon firms that neglect the relation and not provide the required attention (Mosca, 2016).

According to Amed et al. (2017), in 2018 an important tipping point will be reached when, for the first time, more than half of apparel and footwear sales will originate outside of Europe and North America, being the emerging market countries across Asia-Pacific, Latin America, and other regions the main sources of growth. Until 2020 it is expected a Compound Annual Growth Rate (CAGR) of around 10% for growth in online sales of apparel and footwear globally.

2.2 Online Payment Systems

2.2.1 E-Payments

E-commerce is built upon e-payment systems and as it becomes a major component of business operations for many companies worldwide, e-payment has become one of the most critical issues for successful business and financial services (Kim et al., 2010).

E-payment is defined as the transfer of an electronic value of payment from a payer to a payee through an e-payment mechanism. E-payment services exist as web-based user-interfaces that allow customers to remotely access and manage their bank accounts and transactions (Weir et al., 2006). In comparison to the traditional payment methods, e-payment techniques have several favorable characteristics, including reliability, scalability, anonymity, acceptability, privacy, efficiency, and convenience (Kim et al., 2010).

Electronic payments can be divided into two main categories: cash-based and account-based systems. The first group includes electronic-cash solutions (such as PayPal), pre-paid cards, e-loyalty cards, reward cards, among others. These methods usually only involve three parts: the merchants, the client and the financial institution that owns the payment method and works as an intermediate between customers and merchants. As for the account-based systems, they comprehend credit and debit cards, smart cards and electronic checks (Kim et al., 2010). These payment methods are widely accepted by customers and merchants, being the most popular in the world. It works based on a cardholder account in a bank, where the customer has an account card for payments.

2.2.2 Trust and Security in E-payments

Customer trust and sense of security are critical in high volume e-commerce environments. Pursuant to Yao-Hua Tan (2000), people who participate in e-commerce transactions are willing to take a certain risk, but it must stand below the personal threshold. Furthermore, the majority of e-commerce transactions happen without any previous human contact, increasing the impression of security threat. In high volume e-commerce environments, all the processes involve collecting, storing and sending sensitive information from a large number of customers throughout all the internal and external entities. Therefore if any security breach occurs the damages to card users and merchants can be catastrophic (Liu et al., 2010).

Van Slyke and Belanger (2003) conclude that a secure e-payment system should provide security against fraudulent activities and must protect the privacy of consumers. To meet these market security requirements there are security protocols for e-commerce transactions (Khan et al., 2011). Initially, the card brands developed their own security programs tailored to their specific requirements, transforming the compliance programs among all unclear and disorganized. To tackle this issue and to prevent new security breaches, the major card brands created together a standardized group of security standards (Payment Card Industry Data Security Standard) (Liu et al., 2010). The compliance certificate - which implies endowing the payment system with high-security levels - is mandatory if the merchant wants to be able to process online transactions (Abdellaoui et al., 2011).

2.2.3 Transaction Flow

Even though the online payment system is a cumbersome procedure, there has been a great effort to standardize it. The most common is to have a solid basis involving 5 major entities: the customer (C), the merchant (M), the payment system provider (PSP) and the financial entities acquiring bank (A) and issuing bank (I) (Carbonell et al., 2008). Besides that, the card network also plays an important role in this system, creating, when needed, a link between the acquiring bank and the issuing bank.

An e-payment is a complex process, involving multiple steps and connections between all the entities mentioned before. After completing the order and confirming the payment, the authenticity of the transaction is verified. If the transaction is validated, the amount is eventually debited from the customer account (Liu et al., 2010).

Authorization, Capture and Settlement

The payment processing can be split into three distinct time-bounded parts: the authorization, the capture, and the settlement.

The authorization is a flow of information exchanged between all parties involved. It begins at the moment the client places the order. The payment request is then redirected through all the entities culminating at the issuing bank that verifies the account and transaction details (if the customer card details are correct or if the purchase amount is within the line of credit allowed to the customer, etc). Finally, the merchant receives an authorization code (Liu et al., 2010; Kossler, 2013). No money is actually transferred during this part of the process but an imaginary hold is placed on the customer account, protecting the amount of money the consumer spent (Perry, 2012).

The next step - capture - can happen immediately after the authorization is received or it can be triggered by some kind of action or time interval. This is when a merchant notifies the credit card company that the transaction has been completed between them and the customer. Now, money can be withdrawn from the customer's account.

The final step is the settlement. The merchant, in a defined periodic timeframe, sends the batch of transactions to its acquirer or payment processor. The batch is then divided by the card network being all the transactions sent to the corresponding issuing bank. Then the banks issue the purchase to the cardholder and the payment is sent to the acquirer. The acquirer then deposits the due amount to the account of the company minus the processing fees (Liu et al., 2010; Kossler, 2013). The process is outlined in Figure 2.1.

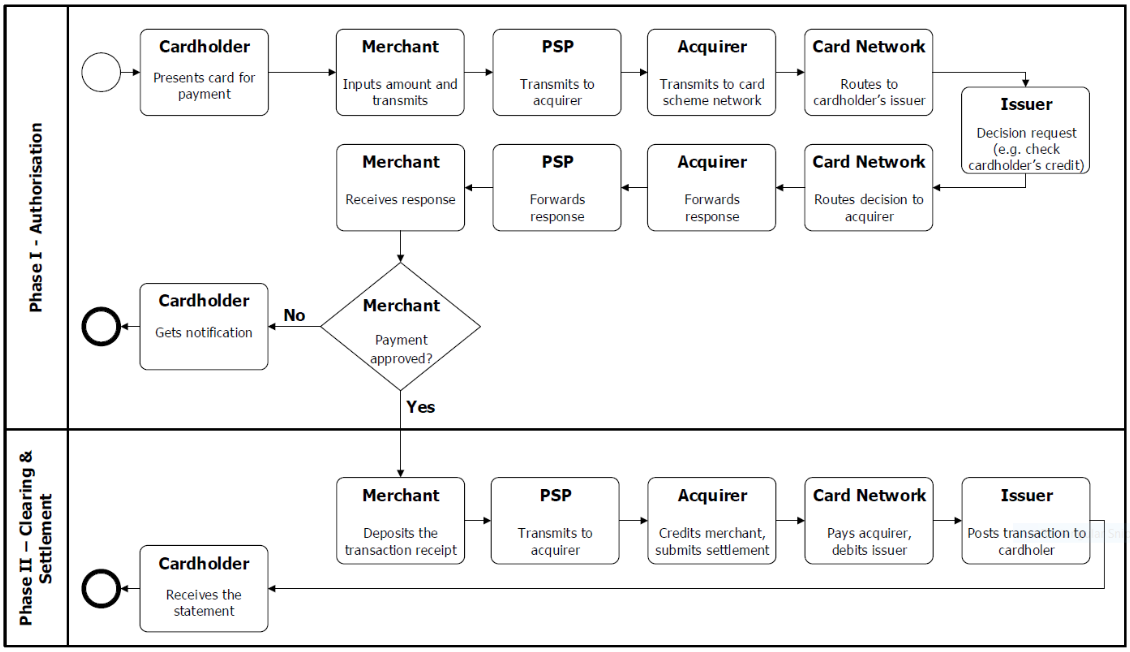

Figure 2.1: Payment Processing - Authorization and Settlement (Van der Valk, 2015)

2.2.4 Entities

Customer and Merchant

The customer triggers the whole process by requesting a service/product provided by the merchant and pays for that. The merchant is the entity accountable for providing products and services for which the customer actually paid (Carbonell et al., 2008).

Payment Service Provider

PSP is a pivotal element in the payment system. It is the entity that enables the connection between the merchant and the other parties. The provider is responsible for processing the payment, on behalf of the merchant, and for assuring the correct exchange of information (Carbonell et al., 2008). The provider usually integrates a wide range of payment methods (credit and debit cards, bank payment, e-banking or e-wallet), offers the possibility to work with multiple currencies and allows the merchant to capitalize main business issues like risk management, reporting or even fraud protection (Sellxed, 2016).

Acquirer
It is defined as the financial institution of the merchant. The acquirer's main responsibility is to communicate with the issuer and to assess the validity of the requested payments. It communicates directly with the PSP which in turn informs the merchant (Carbonell et al., 2008).

Card Network
Credit card networks provide all the credit cards as well as the services to them bounded and can be classified in two types: proprietary and open networks. Proprietary networks, such as American Express, operate as issuer, acquirer and network operator. Open networks are composed of member banks that can be issuers, acquirers or both. The main purpose of these organizations is to meet the needs of their members by providing a set of rules, infrastructures, and some level of research and development to improve their networks (Chakravorti, 2003).

Issuer
The issuer or issuing bank is the customer's bank and responsible for its account. It is the financial organization that checks the validity of the payment and accordingly transfers the funds to the acquirer (Carbonell et al., 2008).

2.2.5 Costs

Transaction Fees
In an e-payment transaction, there are numerous costs that should be taken into account. First of all, there are costs associated with payment method chosen and, within each payment method, there are costs related to the different entities and stages of the process. Independently of the success of the transaction there is always the processing fee, a fee that is charged every time an attempt is made, which is residual when compared to the transaction costs - those charged when a transaction is successful. These vary from payment method, but the most common are the Interchange++ and Commission based. The Commission based is a fee (normally a fixed percentage of the transaction value) that is negotiated with the different entities. The Interchange++ consists of 3 parts (Adyen, 2016):

- Scheme fees (or switch fee) – a fee that the card network collects from the acquirer and issuer;

- Markup (or merchant discount) – an amount charged by the acquirer to the merchant, which is the difference between the face value of the transaction and the amount the acquirer transfers to the merchant (Prager et al., 2009);

- Interchange fee – happens when the issuer and acquirer are different. The acquirer pays the issuer a fee, set collectively by the banks that belong to the system (Hunt, 2003).

Chargebacks

After completing a purchase several issues can still happen. The most common are refunds and chargebacks. Refunds happen when the customer requests directly the merchant for a refund regarding a purchase made. For chargebacks, rather than contacting the merchant for a refund, the customer asks the bank to forcibly take money from the merchant's account (Chargeback911, 2016). This occurs when the client alleges to have an unrecognized purchase on his card statement. The reasons can be many, though there is a rising trend of using chargeback as fraud. In these cases, customers deliberately steal from merchants by claiming that legitimate purchases are fraudulent. Merchants can always dispute illegitimate chargebacks and the final result will decide if the chargeback is reversed or not. Nonetheless, each time a customer files a chargeback, the merchant is charged a fee. Even if the customer later removes the chargeback, the merchants will still have to pay fees and administrative costs associated with the process. Payment system participants take chargebacks very seriously. There are extensive rules and regulations concerning chargebacks that acquirers and merchants must follow (Kossler, 2013).

2.2.6 BIN based Solutions

A bank identification number (BIN) is the first sequence in a payment card number (4-6 digits). It identifies the bank which issued the particular card, which is of major help to the merchants (Chargeback911, 2016). Besides that, the anatomy of the card number permits to retrieve some useful information. The first digit (major industry identifier) identifies the industry that issued the card; the two digits (the MII plus one) identify the brand (Visa - 4*; American Express - 35 or 37; Diners - 36; MasterCard - 51 or 55). These combined with two or four more digits identify the bank. The remaining numbers are the account identification number (Kossler, 2013).

BIN routing for payment processing

Fernandez (2009) presents a BIN routing method where a system configured to identify the card number exists, and then it routes the transaction according to the financial institution related to the BIN. The method provides e-payment optimization where the routing criteria for the optimization should be defined by the merchant. In this regard, the merchant can determine which transactions are routed to which of the banks based on defined parameters, including merchant or customer identification, device type used to process the transaction, product type (Visa, Mastercard, credit card, debit card) or currency.

2.3 Data Mining Approaches

2.3.1 The Concept

One of the earliest definitions of data mining is "the non-trivial extraction of implicit, previously unknown, and potentially useful information from data" (Fayyad, 2001).

Along with this, data mining algorithms are mechanisms for creating data mining models. In order to create a model, the algorithm first analyzes a set of data and looks for specific patterns and trends. The algorithm uses the results of this analysis to define the parameters of the mining model. These parameters are applied to the entire data set for extracting feasible patterns and detailed statistical information. Data mining is closely related to knowledge discovery. Knowledge discovery refers to the whole process of discovering useful knowledge from databases. It includes data selection, preprocessing, data transformation, data mining, schema interpretation and knowledge evaluation (Zhang et al., 2017).

Traditional statistical studies use past information to determine a future state of a system, whereas data mining studies use past information to construct patterns based not solely on the input data, but also on the logical consequences of those data. In summary, data mining can (1) provide a more complete understanding of data by finding patterns previously not seen and (2) make models that predict, thus enabling people to make better decisions, take action, and therefore mold future events.

Major data mining activities include the following general operations (Ham and Kostanic, 2000): exploratory data analysis; descriptive modeling; predictive modeling: classification and regression; discovering patterns and rules.

The CRISP-DM (Cross Industry Standard Process for Data Mining) format for expressing the data mining process is the most complete available. The process defines a hierarchy consisting of major phases, generic tasks, specialized tasks and process instances (Wirth and Hipp, 2000), as detailed in Figure 2.2.

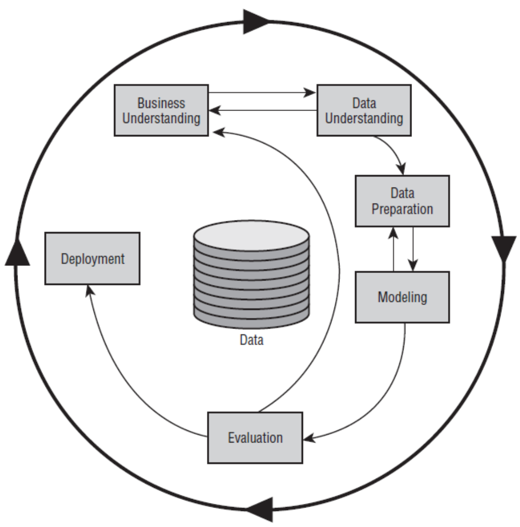

Figure 2.2: Data Mining process (CRISP-DM, 2006)

2.3.2 Classification

Classification is the process of separating several entities into different classes. Classification may be supervised - when it is based on a relationship between a known class and characteristics of the entity to be classified - or unsupervised - if no known examples of a class are available. The most common unsupervised classification method is clustering (Nisbet et al., 2009). For supervised classification, there are two general kinds of problem: (1) binary classification — only one target variable; and (2) multiple classification — involves assigning an object to one of several classes (Har-Peled et al., 2003). An example of a binary classification is a model to identify whether a transaction is fraudulent or not.

Any classification method uses a set of features to characterize each object, where these features should be relevant to the task at hand. We consider here methods for supervised classification, meaning that a human expert has both determined into what classes an object may be categorized and has also provided a set of sample objects with known classes (Nisbet et al., 2009).

There are many techniques used for classification in statistical analysis and data mining such as decision trees, random forest, logistic regression, naive Bayesian classifier, support vector machines, gradient boosting machines, among others. From these techniques some are solely machine learning algorithms, whereas others are combination - ensemble methods - of different machine learning algorithms. Below some of them are briefly explained.

2.3.2.1 Learning Algorithms

Logistic Regression

Regression methods have become an essential element of any data analysis concerned with describing the relationships between a response variable and one or more explanatory variables. Frequently, the outcome variable is discrete, taking on two or more possible values (Hosmer Jr et al., 2013).

Logistic regression is part of a larger class of algorithms known as generalized linear model. The logistic regression model is simply used to estimate the probability of a binary response based on one or more independent variables and does not perform any statistical classification. Still, by choosing a cutoff value it can be used to make a classifier, as the outputs with probability greater than the cutoff are classified as one class and below the cutoff as the other (Walker and Duncan, 1967; Cox, 1958).

Decision Tree

Decision trees are used to predict class or value target variables with the help of decision rules inferred from the data, by using a tree representation. In a decision tree, each leaf node is assigned a class label. The non-terminal nodes, which include the root and the internal nodes, contain attribute test conditions to separate records that have different characteristics (Tan et al., 2006).

The decision tree algorithm starts by setting the best attribute of the dataset at the root of the tree; the following step is to split the training set into subsets, divided in such a way that each subset contains data with the same value for the chosen attribute; and repeat the previous steps until every leaf node in all the branches of the tree is determined. Nevertheless, deciding the best attribute to use at the root or in the internal nodes is a complex task. For solving this attribute selection problem some criterion like information gain, Gini index, and so on, are used (Saxena, 2017).

2.3.2.2 Ensemble Methods

Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance (bagging), bias (boosting), or improve predictions (stacking) (Zhou, 2012; Smolyakov, 2017).

Bagging

Bagging, the abbreviation of bootstrap aggregating, involves having several models whose vote in the ensemble has equal weights. In order to promote model variance, bagging trains each model in the ensemble using a randomly drawn subset of the training set (Breiman, 1996).

Introduced by Breiman (2001), random forests are a type of ensemble built by combining the predictions of several trees, each of which is trained in isolation. The method then makes predictions by averaging over the predictions of several independent base models. In this kind of algorithms, instead of each node being allowed to split over all the features, it is only possible over a random subset of features. As a result, even though individual trees will be weaker predictors, they will be less correlated with all the other trees, improving the robustness and predictive power of the algorithm.

Boosting

Boosting involves incrementally constructing an ensemble by training each new model instance to emphasize the training instances that previous models misclassified. Unlike bagging, in the classical boosting, the subset creation is not random and depends upon the performance of the preceding models (Nagpal, 2017).

A Gradient Boosting Machine (GBM) is an example of a boosting ensemble and, as the random forests, it is also an ensemble of classification or regression trees. A GBM sequentially builds models that correct the mistakes that the previous ones made. Whereas random forests grow trees in parallel, GBM builds simpler trees and stacks them. GBM involves three elements: a loss function to be optimized, a weak learner to make predictions and an additive model to add weak learners to minimize the loss function (Brownlee, 2016).

Stacking

Model stacking is an efficient ensemble method structured in two steps. Firstly it uses several machine learning algorithms to make predictions that are going to be used in a second-layer learning algorithm. In the following step, the second-layer algorithm is trained to optimally combine the model predictions to form a new set of predictions (Güneş, 2017). According to Wolpert (1992), stacking typically generates better performance than any of the trained models.

2.3.3 Performance Metrics

The evaluation of the machine learning algorithm is an essential step in any project. Firstly, distinct performance metrics suit best to evaluate different machine learning algorithms. Secondly, the dataset to be used also impacts the choice of the different metrics. The focus will be on the ones used in classification problems.

By comparing the class prediction with the real class of the instance, one can draw a confusion matrix to assess the overall performance of the models, as shown in Table 2.1.

Table 2.1: Explanation of confusion matrix

		Real Class	
		0	1
Predicted	0	True Negative (TN)	False Negative (FN)
	1	False Positive (FP)	True Positive (TP)

With the confusion matrix several performance measures can be obtained:

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

$$\text{True Negative Rate (or } \textit{Specificity}\text{)} = \frac{\text{TN}}{\text{TN}+\text{FP}}$$

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

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

$$F_1 \text{ Score} = 2 \times \frac{1}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$$

Recall corresponds to the proportion of positive data points that are correctly considered as positive, with respect to all data points that are actually positive.

Specificity measures the proportion of negatives that are correctly identified as such, with the respect to all data points that are actually negative.

Precision measures the proportion of actual positives from all the data points classified as positives.

Accuracy measures the number of correct predictions made by the model over all the predictions made. Accuracy is a good measure when the dataset, with regards to the target variable, is nearly balanced.

F_1 **Score** represents the harmonic mean between precision and recall. A model with high precision but lower recall provides an extremely accurate model, although it misses a considerable number of instances of difficult classification. So F_1 Score tries to demonstrate how precise a classifier is, and at the same time give information about its robustness. The greater the F_1 Score, the better is the performance of our model.

ROC-AUC

In a binary classification problem, instances are classified according to a pre-defined threshold. Instances with a score under the threshold are classified as negative, while the ones above the threshold are classified as positive. A Receiver Operating Characteristics (ROC) curve plots the true positive rate and false positive rate for each threshold (Fawcett, 2006).

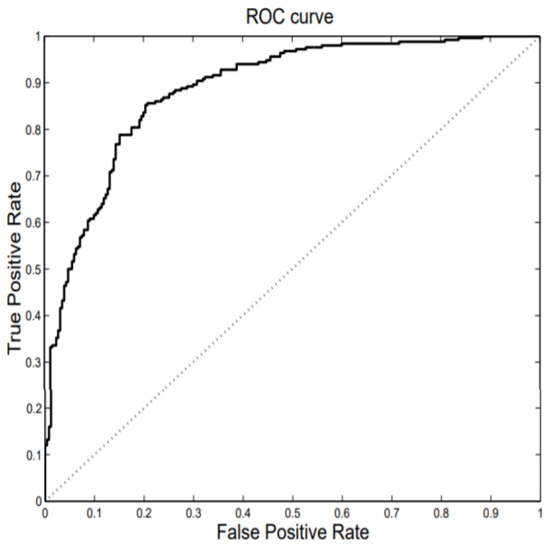

Figure 2.3: Example of ROC curve (Rakotomamonjy, 2004)

In Figure 2.3 a ROC curve of a given classifier can be observed. The diagonal line corresponds to the ROC curve of a classifier that randomly predicts the class. As the curve approaches the upper left corner of the plot the performance improves.

Even though the ROC curve can provide useful insights, the most frequently used performance measure associated with it is the value of the Area Under the Curve (AUC). When AUC is equal to 1, the classifier achieves perfect accuracy whereas a classifier that predicts the class at random has an AUC of 0.5.

As it is independent of the chosen threshold, AUC can describe a general behavior of the classifier (Rakotomamonjy, 2004). According to Fawcett (2006), the AUC performs very well in practice as a general measure of classifier performance.

2.3.4 Training, Test and Validation

As seen previously, a classifier is a function that maps an unlabeled instance to a class label using internal data structures, whereas an inducer builds a classifier from a given data set (Kohavi et al., 1995). Bearing these concepts in mind, an important step in machine learning problems is the estimation of the performance of a model created by the given inducer and the dataset. For this, it is common to have a training set (the actual dataset that is used to train the classifier), a validation set (used to evaluate and tune the model) and a test set (used to provide an unbiased evaluation of the final model with unseen instances). According to Bousquet et al. (2011), if the whole dataset is used both for training the classifier and for estimating its error, there is a serious threat of over-fitting the classifier to the training data. Kohavi et al. (1995) reviews the most common estimation methods and their comparative advantages.

The simplest is the holdout method which partitions the data into two mutually exclusive subsets called a training set and a test set, or holdout set. It is common to designate $2/3$ of the data set as the training set and the remaining $1/3$ as a test set. The training set is given to the inducer, and the inducer classifier is tested on the test set. Formally, let D_h, the hold-out set, be a subset of D of size h and let D_t, the training set, be $D \setminus D_h$. As only a fraction of the information is ever shown to the algorithm the resulting accuracy is a pessimistic estimator.

In cross-validation, also known as k-fold cross-validation, the dataset is randomly split into k mutually exclusive subsets (the folds) D_1, D_2, ..., D_k of approximately equal size. The inducer is trained and tested k times; each time $t \in \{1,2,...,k\}$, it is trained in $D \setminus D_t$, and tested on D_t. Calculating now the performance metric of each model, and averaging the total accuracy, a more reliable estimate can be obtained. This method circumvents biased estimates as all the data is analyzed both in the training and testing set. The confidence on the given estimate increases with k, at the expense of more computational effort.

2.3.5 Hyperparameters and Tuning

A machine learning model is a mathematical formulation with several internal parameters that must be learned from the data. The process of fitting the model to the data is done, as seen previously, through a process known as model training. By training a model with existing data, we are able to adjust the model parameters (Amatriain, 2016). Nonetheless, there is another type of parameters that cannot be directly learned from the normal training process, the hyperparameters. These parameters express high-level properties of the model such as its complexity or how fast it learns.

While the "normal" parameters are inferred via training the model, the hyperparameters are parameters whose value is set before the learning process.

Amatriain (2016) enumerates some examples of hyperparameters: number of leaves or depth of a tree; learning rate (in many models); the number of hidden layers in a deep neural network; the number of clusters in a k-means clustering.

Hyperparameter optimization or tuning is the problem of determining a set of optimal hyperparameters for a learning algorithm. In fact, there are no correctly defined optimization methods for choosing the best hyperparameters. Many users opt to use default values for hyperparameters or others prefer a trial search instead, according to Bergstra and Bengio (2012).

A very common method is grid search, which is simply an exhaustive searching through a specified subset of the hyperparameter space. Bergstra and Bengio (2012) propose a slightly different search method, where the levels searched are randomly picked from the subset. A comparison between the grid search and the random search can be seen in Figure 2.4:

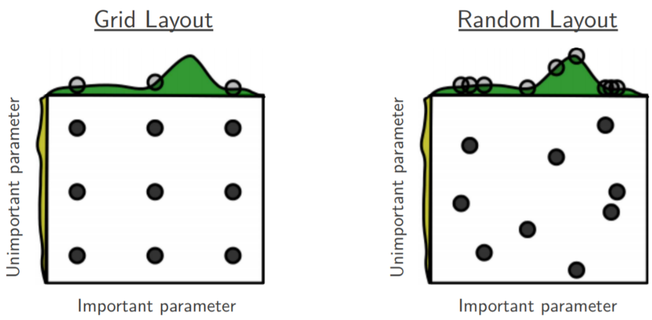

Figure 2.4: Grid and random search of nine trials for optimizing a specific function (Bergstra and Bengio, 2012)

2.4 Multi-Criteria Decision Analysis

Multi-Criteria Decision Making (MCDM) is a generic designation for all methods whose function is to support people making decisions according to their individual preferences, in circumstances where there is more than one conflicting criterion (Bogetoft and Pruzan, 1997).

Although Decision Makers (DMs) always try to determine the optimal solution, in most real decision situations basing a decision solely on one criterion is insufficient. Probably several conflicting and often non-commensurable objectives should be considered (Løken, 2007). Because of this, it is impossible to find a genuine optimal solution, a solution that is optimal for all DMs under each of the criteria considered (Gandibleux, 2006)

Another concept widely used is Multiple Criteria Decision Analysis (MCDA). The reason for using 'analysis' instead of 'making' is to highlight that these methods should aid DMs in making better decisions. (Belton and Stewart, 2002). As MCDM, MCDA problems usually involve a set of alternatives that are evaluated on the basis of conflicting and incommensurate criteria. (Malczewski, 1999).

The alternatives are oftentimes evaluated by a number of individuals (decision-makers, managers, stakeholders, interest groups) that are typically distinguished by unique preferences with respect to the relative importance of criteria (Malczewski, 2006).

There are many possible ways to classify the existing MCDA methods. According to Belton and Stewart (2002), there are three broad categories (or schools of thought): Goal, aspiration and reference level models; Outranking models; and Value measurement models.

Goal, aspiration and reference level models

Usually, GP is used as a common abbreviation for goal programming, the aspiration level, and the reference level methods. The purpose of using GP approaches is to determine the alternatives that somehow are the closest to achieve a specific goal or aspiration level (Belton and Stewart, 2002). Often the GP approach is used at the beginning of a multi-criteria process where there are many alternatives. In that case, GP is used to filter out the most unsuitable alternatives in an efficient way (Løken, 2007).

Outranking models

In outranking models, there is a pair-wise comparison between alternatives to verify which of them is preferred regarding each criterion. When aggregating the preference information for all the relevant criteria, the model determines to what extent one of the alternatives can be said to outrank another. We can say that an alternative a outranks an alternative b if there is enough evidence to conclude that a is at least as good as b when taking all criteria into account (Belton and Stewart, 2002). These methods are often called the French school and can be divided into two families of methods: ELECTRE and PROMETHEE.

Value measurement models

When using value measurement methods, a numerical score (or value) V is assigned to each alternative. These scores present, therefore, a preference order for the alternatives such that a is preferred over b if and only if $V(a) > V(b)$. When using this approach, the several criteria are given weights w that express their partial contribution to the overall score, based on how important this criterion is for the DMs. (Belton and Stewart, 2002; Stewart, 1992; Greening and Bernow, 2004).

The most commonly used approach is an additive value function (Multi-Attribute Value Theory (MAVT)):

$$V(a) = \sum_{i=1}^{m} w_i v_i(a) \tag{2.1}$$

where $v_i(a)$ is a partial value function reflecting alternative a's performance on criterion i. Using Equation 2.1, a total value score $V(a)$ is found for each alternative a. The alternative with the highest value score is preferred (Belton and Stewart, 2002). The Multi-Attribute Utility Theory (MAUT) originally proposed in detail by Jong and Stone (1976) can be said to be an extension of MAVT. MAUT is a more rigorous methodology for how to incorporate risk preferences and uncertainty into multi-criteria decision support methods (Belton and Stewart, 2002; Jong and Stone, 1976).

2.4.1 Analytical Hierarchy Process

The Analytical Hierarchy Process (AHP) was originally formulated by Saaty (1980) to provide a framework for solving different types of multi-criteria decision problems. AHP is a preference scoring model that relies on personal managerial inputs on the several criteria. These inputs are turned into scores that are used to evaluate each of the potential alternatives (Handfield et al., 2002). The AHP is a powerful management tool that has proven to be useful in structuring complex multi-person and multi-criteria decisions in business (Calantone et al., 1999).

The main characteristic of the AHP method is the application of pair-wise comparisons, which are used both to compare the alternatives with regards to the different criteria and to estimate criteria weights (Belton and Stewart, 2002; Greening and Bernow, 2004). In the pair-wise comparisons, a special ratio scale from 1 to 9 is used. The results from all the comparisons are put into matrices. From these matrices, an overall ranking of the alternatives can be aggregated. The alternative with the highest overall ranking is preferred to the others (Greening and Bernow, 2004). The methodology is further explained in Appendix A.

Chapter 3

Problem Description

The objective is, since the beginning, to improve the success and decrease the costs of the online payment system, by implementing a dynamic and flexible e-payments routing heuristic.

The routing procedure for online payments resumes itself to some predefined and rigid rules based on legal restrictions and know-how of the stakeholders. There is no rigorous and data-driven method to choose the best route for a specific payment. Besides that, the current system is not scalable, even more, when looking to Farfetch's expected growth.

3.1 Current Processes

The order processing starts when the customer confirms the order in the website portal. Before the confirmation page, the client must fill all required fields, like shipping details and payment details. Shipping details include the desired delivery address that should be error free to avoid operational problems in the following processes. Concerning payment details, the payment method needs to be chosen as well as the method correspondent details (card number and security code in credit card method). Once the customer confirms the payment, the payment system comes into operation giving, in a matter of seconds, a response to the customer. Even though a lot of information is exchanged in that process, the output is simply a positive or negative result: the authorization or rejection of the payment.

It is obvious that this simple result can have a major impact on the company, where both customer satisfaction and financial results are crucial aspects of the business. Given that, the payment system assumes a vital role in the Farfetch processes.

3.1.1 Payment Methods

Farfetch has the goal of implementing a global strategy with a country-specific approach. Besides the most common methods, adopted worldwide, like credit/debit card and PayPal, Farfetch strives to have innovative and flexible solutions that target, locally, the different markets.

Credit/debit card is a payment method used worldwide that integrates into the same system countless financial institutions and card networks providing a comfortable payment method for any user. At Farfetch this payment method can be used by any customer and is triggered as soon as the customer chooses this option on the checkout page. After choosing the credit/debit card option, the user has to introduce the card details: card number, cardholder name, and security code.

Paypal differs from credit/debit cards by operating itself as a third party payment processor, working as an intermediate between merchant and customer, in the exchange of a fee. It is recognized to be one of the most reliable payment methods and is present in most of the countries worldwide.

Local methods are payment methods only available in specific markets. The majority of these methods were developed based on the peculiarities of each market and some of them reached high levels of popularity. Therefore it became important to implement these methods in the corresponding countries. Alipay and WeChat in China, Boleto in Brazil, iDeal in the Netherlands or Sofort in the DACH region are examples of local methods currently in use.

3.1.2 Payment System Architecture

The payment system is complex and depends on the several specificities of the payment methods and the different markets where they are available. There is an enormous difference between the workflows of credit card payments and other payment methods (like Paypal or local methods). Even within the credit card workflows there are notable differences. For this reason, we can group them into AMEX, MasterCard+Visa, and Local cards workflow. Figure 3.1 shows the break down of the different existing workflows.

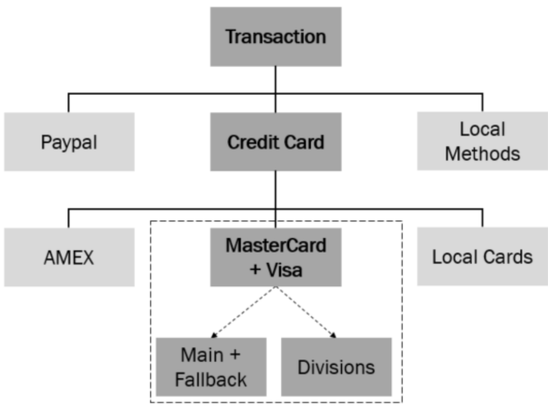

Figure 3.1: Break down of the different payment methods and the scope of this project

The other payment methods are intrinsically different, but at the same time share some common traits. Paypal is the only global method, all the other methods are local, having specific requirements that depend on the market, making them considerably different. However, the fact that they all need some kind of integration makes them similar. Figure 3.2 shows how these methods are routed.

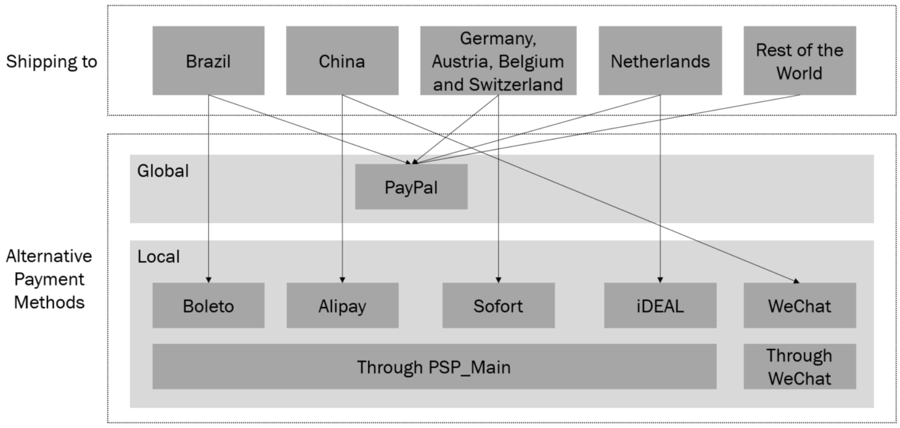

Figure 3.2: Workflow of the different local payment methods

Contrastingly, the credit card workflow encompasses all the transactions made with credit or debit card. Currently, the main card brands are Visa and MasterCard, but many more, like American Express (AMEX), Japan Credit Bureau (JCB), China Union Pay (CUP) or Carte Bancaire (CB), are also used. The system is complex and different workflows had to be created in order to cope with all the different requirements.

Parallel Workflows

Within the existing card networks available three main groups were created in order to simplify all the process, which can be verified in Figure 3.1. The first one is AMEX, which due to its special operating mode is treated as a separated workflow. Besides working as one of the main card networks, AMEX is also a big financial institution that can work as an integrated acquirer-issuer, working in a three-party card scheme, changing drastically the normal payment workflow. The second group includes the major card networks MasterCard and Visa. Those are the cards mostly used having a complex and many times redundant system. In the third group are included what we call the "local cards". These card networks work similarly to the second group, although due to their small transactional volume, their payment workflow is compact and simple.

Main and Fallback

The payment system currently in use is a result of continuous improvements since the company was created. One of the major flaws detected was the dependency on the main payment system, as there was no other solution for a failed transaction. To solve this issue, a background fallback model was introduced. The idea of this model is to use different providers and acquirers to process the transactions that failed through the main system. Thereby the dependency on the main flow can be diminished, improving the ability to process payment orders.

Divisions

The payment system is not universal. Legal requirements and strategic moves introduced variations in the original system. The system is now divided in three main parts. This division is based on the shipping country of an order, i.e., according to the country where the order will be delivered.

- **Europe and Rest of the World (EU & ROW):** this is the one with the widest geographic coverage. It is the simplest process because the main provider, PSP_Main, works also as acquirer (A_Main_ROW1). Even though they belong to the same institution they are considered as different entities. Due to the number of transactions made through this division and to decrease the risk of failure of A_Main_ROW1 there is also another acquirer, A_Main_ROW2. Regarding the fallback, it uses a different PSP and acquirer (PSP_Fallback _ROW and A_Fallback_ROW).
- **United States (US):** the American banking market is very selective and unique. The system is regulated by specific rules that many times are considerably different from the rest of the world, which in many cases results in incompatibilities hard to overcome when it comes to accepting a transaction. To cope with this, a solely US division was created. The most relevant difference from the first division is the introduction of American acquirers. Maintaining the same provider PSP_Main there are now two possible acquirers: A_Main_US1 and A_Main_US2. Regarding the fallback, the PSP is PSP_Fallback_US and the acquirer is A_Fallback_US.
- **Brazil (BR):** the way payment systems work in this country is remarkably different and it has some unique rules that often undermine business done with Brazilian customers. Besides that, it is common to pay with installments, a procedure not allowed worldwide. A third division was then created in order to fully explore this market's potential. Its structure is formed by the same main provider PSP_Main but with a Brazilian acquirer, already familiar to deal with installments and other market issues: A_Main_BR. For Fallback the PSP is PSP_Fallback_BR and the Acquirer is A_Fallback_BR.

Figure 3.3 depicts the relevant worflows in detail.

3.1.3 Routing Decision

One crucial aspect of the payment system is the way each transaction is routed through the system. The current routing decision is strongly connected with the divisions mentioned earlier. Those divisions are, in turn, intrinsically connected with the shipping country. Therewith, the routing decision is heavily dependent on the shipping country of a transaction. For instance, a transaction whose shipping country is the United States is going to be routed through the US Division, whereas a French or a Chinese transaction is going to be routed through the EU & ROW division.

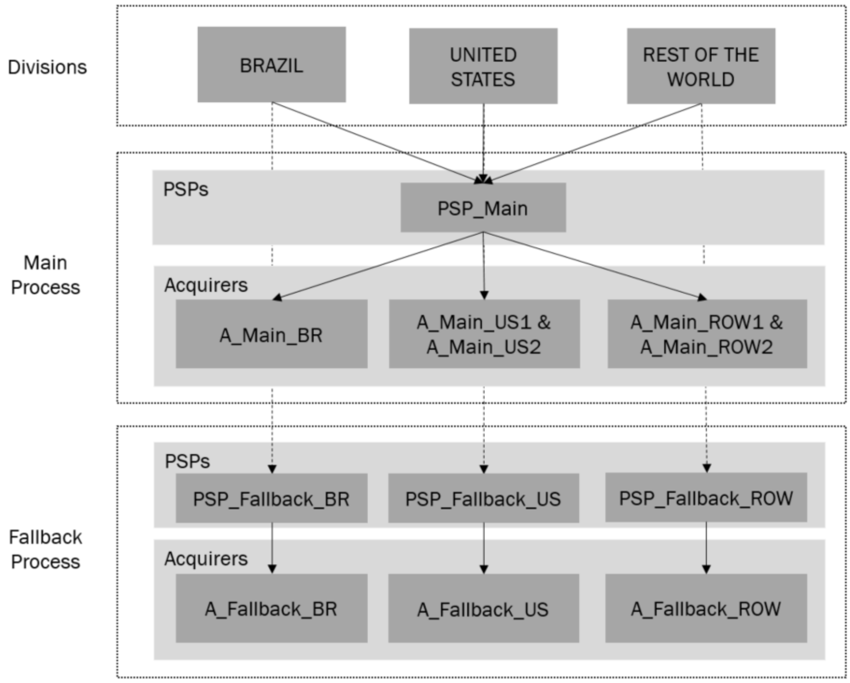

Figure 3.3: Workflow of the main and fallback models according to the division

Nonetheless, some decisions still have to be made: the PSP and the acquirer which are going to process the transaction. Regarding the PSP, the decision is straightforward, as in each division only one option is available (the PSP_Main). For the acquirer, with the exception of the Brazilian division (which has only one acquirer), the decision is not so simple. In those cases the transaction can be routed through two distinct acquirers after being processed by the PSP. After that, the decision depends heavily on the stakeholders' input. Many times A/B testings [1] are done in order to assess the response of the acquirers to specific transactions, e.g., an AB testing was performed for transactions with shipping country being Hong Kong. The result was that A_Main_ROW1 performed better than the other. With this conclusion, it was established that for Hong Kong 80% of the transactions were to be routed through there and the remaining through A_Main_ROW2.

Even though this is a good strategy to assess the best possible routes, it is a time-consuming procedure that requires the deployment of many resources. Moreover, it is not scalable neither dynamic (to cope with the constant adjustments in the system).

3.1.4 Cost Structure

Similar to the payment system, the cost structure is complex. As explained in the previous chapter, online payments have several costs that depend not only on the characteristics of each transaction but also on the entities involved. For example, for the same transaction, the cost of completing a transaction through A_Main_US1 or A_Main_ROW1 will be different.

[1] In web analytics, A/B is a way to compare two alternatives typically by testing a subject's response to variant A against variant B, and determining which of the two alternatives is more effective.

Currently, cost is not taken into consideration when deciding the best route for a transaction, however, the general payment costs are carefully controlled by the Payments team.

3.2 Opportunities

Even though the actual payment system is a result of numerous mutations and adaptation to the different requirements, the system still presents room for improvement. This section tries to indicate those limitations and highlight its opportunities.

As explained in the previous section, there are several payment methods, of which credit/debit cards are the most representatives, with a weight of almost 80% without any significant decreasing trend, even with the new local payments being introduced more recently. Figure 3.4 shows the distribution of the different payment methods in 2017 and 2018.

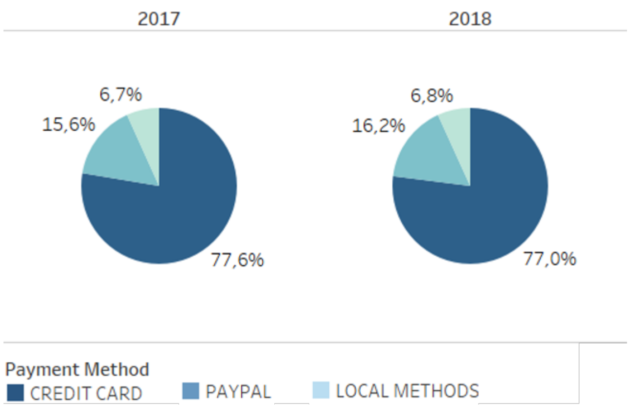

Figure 3.4: Distribution of the different payment methods in 2017 and 2018

Although Paypal and local methods represent more than 20% of all the transactions, due to their unique characteristics, they present small room for improvements. Besides a correct account configuration and a proper integration, little can be done to improve their performance. Credit/debit cards are, for this reason, the payment method that can cause the higher impact on this project. Figure 3.5 shows the weight of the different card brands in use.

Regarding the distribution of the card brands, it is clear that Visa and MasterCard play a vital role in the payment system, whilst AMEX and local cards represent around 20%. Therefore, Visa and MasterCard workflows, for being the most intricate ones and for representing more than 60% of all the online transactions, are those with the most opportunities for improvement.

An important component of the payment system is its performance. Many metrics are used to analyze the system, some simpler and others more complex. One of the most important is Authorization Rate (AR), which is widely used and understood in the business. It roughly indicates the percentage of transactions which have been successfully approved by the system.

This performance indicator can provide useful information to the business, specially to analyze the impact of the route (the acquirer) chosen. It is expected, for example, that a transaction

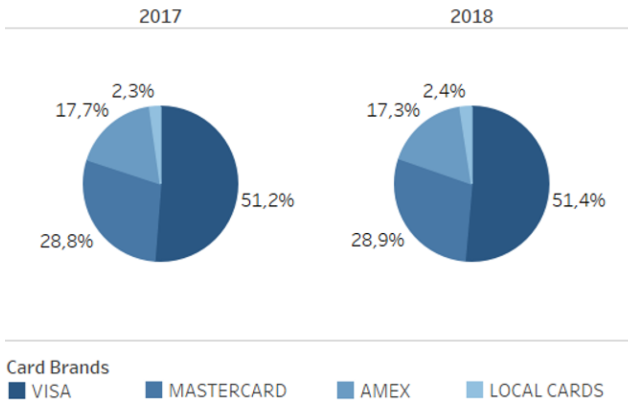

Figure 3.5: Distribution of the different card brands in 2017 and 2018

whose card is American to have higher chances of being successful if the acquirer is also American (because both financial institutions, issuer and acquirer, are from the same country). The same happens with costs: a national transaction is expected to be cheaper than an international (e.g. a Chinese card in a European acquirer). Figure 3.6 tries to show the impact in AR of the different routes for the same transaction.

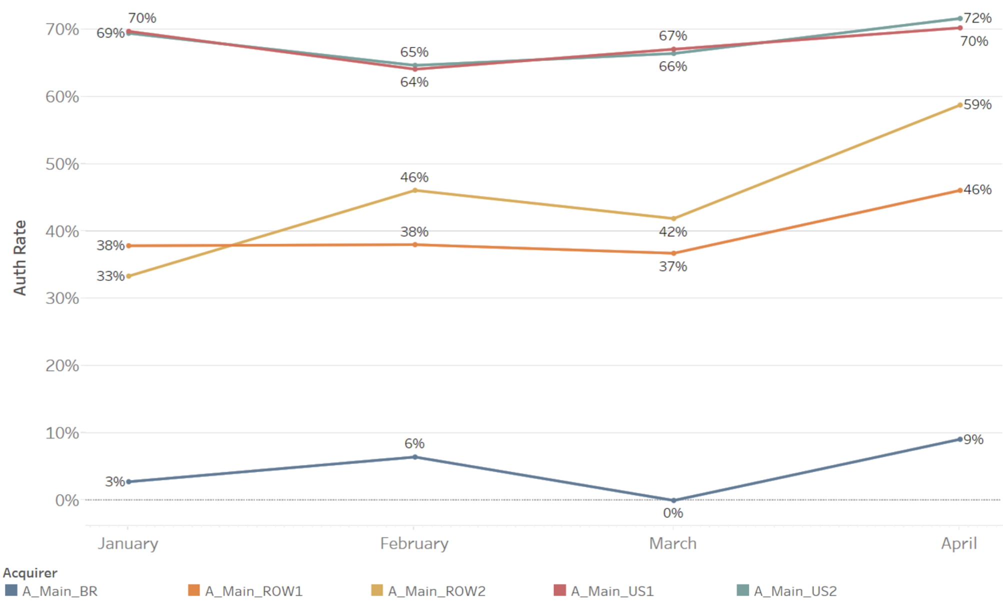

Figure 3.6: Authorization rates of the different acquirers for a transaction with the same characteristics (Card country:US ; CardBrand:MasterCard) in 2018

As we can see from Figure 3.6, for that specific transaction American acquirers are the best option, followed by European acquirers, being the Brazilian acquirer the worst option. Besides this, it is also important to assess the costs. For all the transactions routed through PSP_Main, the costs are recorded individually like is presented in Table 3.1.

Table 3.1: Simplified example of the transactions' caracteristics and the respective costs

TID	Acquirer	Card Information		Value ($)	Fees ($)		
		Country	Brand		Scheme	Interchange	Markup
1	A_Main_ROW1	Israel	Visa	402	0,19	11,27	1,19
2	**A_Main_US2**	US	MasterCard	106	**0,32**	**5,11**	0,32
3	**A_Main_ROW1**	US	MasterCard	106	**1,14**	**3,70**	0,32
4	A_Main_ROW2	Canada	Visa	322	4,14	9,03	0,96
5	A_Main_US1	Spain	MasterCard	3 954	4,98	20,76	11,77

Transactions 2 and 3 have the same value and card information but the fees vary (from 0,32 to 1,14 in Scheme Fees and from 5,11 to 3,70 in the Interchange, a total difference of around 0,6). Since everything is similar, what defines the cost of a transaction is the acquirer chosen. Thus, the selection of the best acquirer for a specific transaction allows not only a performance improvement, but also the opportunity to optimize costs.

This kind of analysis is feasible for a restricted number of markets and acquirers but is not reasonable for the entire business. First of all it is time-consuming, as one has to compare the AR and costs of different acquirers according to the transactions characteristics, but the foremost issue is the scalability. A procedure like this is not adaptable to the growth of the business and to the constants changes in these environments.

3.3 Proposed Solution

As it was seen in the last section, the opportunities for improvement are substantial. The proposed solution has to be precise, taking into account the right pieces of information and the correct assumptions; dynamic, being able to cover all the relevant criteria of the payment system; scalable, to cope with future changes in the business; complete, by considering all the relevant parts of the system; and feasible, to be easily implemented in the company.

Bearing in mind these requirements, the solution should comprise a section to improve the performance of the payment system (the only criteria being used at the moment) and a section to assess the costs (a critical issue to the company). Even though these two criteria are both important, they might have unequal importance to distinct stakeholders. For that reason, it would be worthwhile to add different weights to those criteria according to the stakeholders' inputs.

The proposed solution is a payments routing heuristic[2] that combines, with predefined weights, an element to improve performance (success) and another to optimize costs, with the final goal of ranking the different routing options for a specific transaction. With this ranking, the best route can be selected according to the company's strategy.

[2] A heuristic is a rule of thumb, strategy, trick, simplification, or any other kind of device which drastically limits search for solutions in large problem spaces. Heuristics do not guarantee optimal solutions; all that can be said for a useful heuristic is that it offers solutions which are good enough most of the time (Feigenbaum et al., 1963)

Chapter 4

Methodology

This is the chapter in which the methodologies to deploy the proposed solution are described. Although this methodology was developed to tackle the limitations and opportunities of the studied company, the heuristic can be implemented in almost every payments routing decision system in e-commerce companies. This routing heuristic is primarily composed of three parts: Success, Cost and Criteria Weights, as it is explained in the figure 4.1.

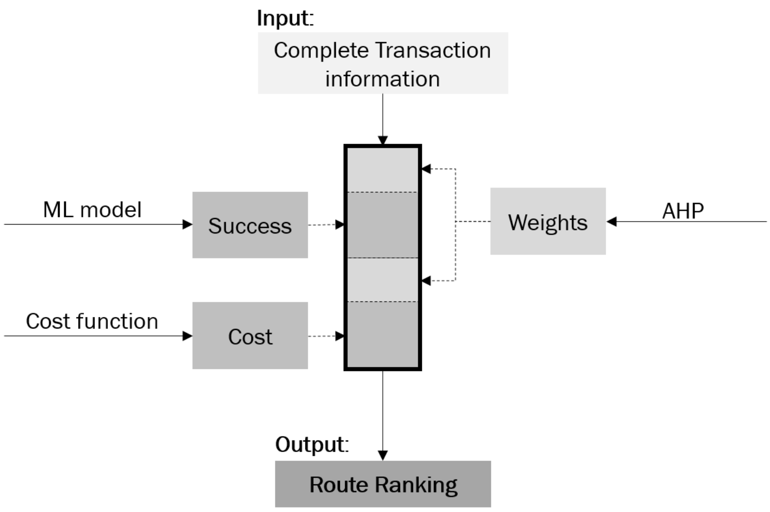

Figure 4.1: High level diagram of the proposed heuristic

4.1 Success Score

The goal of this section is to develop a machine learning approach to select the best possible route for a new transaction. Firstly a model is created to predict the result (success or failure) of a new transaction. The built model is then used to estimate the success probability of a transaction with specific characteristics. The output of this model is a list of all the possible routes for a transaction and the respective "Success Score".

To complete the model a methodology is followed. First of all, a selection of all the variables used in the model is done, accompanied by the definition of the train, validation and test sets. The next step is the selection of the most appropriate machine learning technique. As referred in Chapter 2, there are several techniques suitable to this problem. Using all the available techniques, a model tuning, evaluation and selection framework is performed in order select the best technique for the problem at hand. This methodology is further explained in the following subsections.

4.1.1 Input Data

The business implementation of the proposed heuristic is a real intention, so, besides the restriction of the variables to be used, its internal operation should be in accordance with the business working conditions. The starting point of the variable selection is all the data currently sent to the entities in the payment system. After understanding the business, the next step is to understand the data, what can be done through exploratory analysis. The last step involves data preparation. There are several methods to accomplish that, but a combination of feature selection (where, for example, the unnecessary variables are eliminated), feature engineering (the generation of new features by, for instance, combining existing variables), missing values handling (to ensure that the dataset is clean and complete), numerical standardization (standardize numerical values to the same range of values, so that they can be more easily compared) and grouping of high-dimensionality categorical variables (to reduce the dimensionality of the data, variables with many categories can be grouped in larger categories) can be used to prepare data for modeling.

4.1.2 Train, Test and Validation Sets

After having defined the variables, it is important to specify the dataset to be used. The dataset is split in two: the train and validation set, and the test set. The train/validation set is used to train the algorithm so that it can learn how to classify transactions, and, at the same time, to tune its hyperparameters. The test set is kept separate, to be used at the end of the modeling process in order to compare the different techniques and to chose the most appropriate one.

According to the literature, 20% to 40% of the observations should be used for testing and the remaining for training. After deciding the number of observation that should belong to each set, the criteria of division should also be defined. The goal of the testing set is to accurately replicate the performance of the model in a real situation. For this particular case, considering the most recent transactions for testing and the remaining for training and validation would be a good way to emulate new orders, as the patterns of performance, regarding the several entities, are continuously changing over time.

4.1.3 Tuning and Model Selection

A model tuning, evaluation, and selection framework was developed in order to choose the most appropriate machine learning technique for the heuristic. The model tuning, evaluation and selection framework is detailed in Algorithm 1.

input: transactions dataset;
initialization;
Split dataset in train/validation and test;
foreach *technique* **do**
  define: Hyperparameters range;
  generate: *n* Hyperparameters combinations;
  foreach *Hyperparameter combination h* **do**
    foreach *i in k folds* **do**
      Run *technique* with hyperparameters *h* holding fold *i* out;
      Compute score in fold *i*;
    end
    Average all score;
  end
  Select *Hyperparameter* combination with the best average score;
  Train *technique* in the entire train/validation set;
  Predict observations in test set and evaluate performance;
end
Select best *technique*;

Algorithm 1: Tuning and model selection framework

It starts by finding, for each technique, the optimal combination of hyperparameters. The range of the hyperparameters is defined based on the literature, taking also into consideration the computational effort needed. With a defined range for each technique, a random search is deployed: a $n \times h$ matrix of n combinations of h hyperparameters is randomly generated. Then, for a given combination of hyperparameters, a k-fold cross-validation is done and its performance recorded. Afterward, the combination of hyperparameters with the best performance is selected. Then, for each technique, the best hyperparameter combination is used to train the model with the training/validation set. Finally, the performance of the model is estimated with the test set.

4.2 Cost Estimation

Another crucial part of the routing model is the cost estimation. The information regarding each transaction is not only used to construct the predictive model but also to estimate, for each route, the expected cost of the transaction. The ideal would be to obtain the precise cost rate of a specific transaction in a given route. However, this value is sometimes difficult to obtain. In those cases, a cost rate approximation might be a solution.

4.3 Criteria Weight

The goal of this heuristic is to create a complete and flexible model that can respond properly according to different requirements. There is no optimal choice, but rather a trade-off between options (Cost and Success). By questioning the stakeholders, it is possible to have an idea of how each one of them weights each part. For instance, the finance department might only have interest on the cheapest route; on the other hand, the commercial department might only be interested in the success of the transaction in order to increase sales.

In order to transform the different stakeholders' interests, a multi-criteria decision approach was used to define precisely the way weights were assigned to each part of the model. Analytical Hierarchy Process, is able to assign weights to different criteria according to the different stakeholders' preferences.

In this AHP methodology, every stakeholder defines the preferred criterion (Success or Cost) and chooses a level of preference of that criterion over the other with a scale of 1 (equally important) to 9 (absolutely more important). For instance, a stakeholder chooses Success as the preferred criterion and assigns a scale of 9, the weights (according to the AHP methodology) are going to be 90% to Success and 10% to Cost. After gathering all the stakeholders' inputs and calculating each individual preference one can use the individual weighting choices or can determine the consolidated weights (that are calculated based on the individual weights and the deciding votes of each stakeholder).

4.4 Heuristic Formulation

The purpose of this section is to explain the assembly of all the elements earlier described. The proposed heuristic should be able to predict the best route for a new transaction according to the defined criteria. The final routing decision is made by comparing the expected value of each possible route, considering both revenue and costs of the transaction in that route. The part referring to Success is demonstrated in Equation 4.1.

$$E(revenue) = Score_{Success} \times Value_{transaction} \tag{4.1}$$

Therefore, the expected value in terms of revenue of a certain transaction can be obtained by multiplying the probabilistic score (the output of the machine learning model) with the value of the transaction. In order to have the expected cost of a transaction in a specific route, the cost rate estimation is multiplied by the $E(revenue)$ of the transaction, as it is demonstrated in Equation 4.2.

$$E(cost) = E(revenue) \times Cost_{rate} \tag{4.2}$$

Having formulated Equations 4.1 and 4.2, the expected value of a transaction is easily obtained by combining them with the weights previously defined. Equation 4.3 demonstrates it.

$$E(transaction) = W_{Success} \times E(revenue) - W_{Cost} \times E(cost) \tag{4.3}$$

As mentioned before, flexibility is a key issue in this heuristic, so that it can cope with different requirements and work properly in diverse situations. This can be attained by changing the weights in Equation 4.3. One might want to route a transaction solely grounded on its probability of success, whereas another might want to use cost and success equally, or even one might want to have into consideration all stakeholders' preferences. The heuristic allows the following different possibilities:

- Only success: $W_{Success}$ = 1 and W_{Cost} = 0

$$E(transaction) = E(revenue) \tag{4.4}$$

- Only cost: $W_{Success}$ = 0 and W_{Cost} = 1

$$E(transaction) = -E(cost) \tag{4.5}$$

- Real expected value: $W_{Success}$ = 1 and W_{Cost} = 1

$$E(transaction) = E(revenue) - E(cost) \tag{4.6}$$

- Taking into account the stakeholder's interests: $W_{Success} = S$ and $W_{Cost} = C$

$$E(transaction) = S \times E(revenue) - C \times E(cost) \tag{4.7}$$

With the heuristic fully formulated, the following methodology should be deployed to generate a routing decision:

1. Select a new transaction and its information (whenever a new transaction is attempted the heuristic is run);
2. Assess the possible routes for that transaction (according to legal restrictions or to the company's strategy);
3. Calculate the success score with the machine learning model for each possible route;
4. Calculate the cost rate with cost estimation algorithm for each possible route;
5. Define the criteria weights (the stakeholders' weights should be previously defined);
6. Calculate the expected value of the transaction for each possible route, according to the previous restrictions;
7. Rank the possible routes, according to their expected value;

This ranking is the final output of the heuristic, that besides being used to select the route, can also be used for further decisions (e.g. choosing, when needed, the fallback route).

Chapter 5

Case Study: a Large Fashion E-tailer

The purpose of this chapter is to analyze the execution of the heuristic in the studied company. We start by describing its calibration and implementation. After that some examples are given to understand its functioning. In the end, a simulation is done to assess the impact of the heuristic's deployment in the company.

5.1 Calibration of the Heuristic

This section explains the calibration of the proposed heuristic and follows the methodology described in the previous chapter. We start by focusing on the success score part, followed by the cost estimation and the definition of the criteria weights.

5.1.1 Success Score

Firstly we should define the variables to be used in this model. The starting point of the variable selection is all the data currently sent to the PSP when there is a transaction attempt, such as order (value, quantity of items,...) and transaction information (payment method and respective payment details, ...). The following paragraphs explain the procedure to define all the features used.

5.1.1.1 Variables

Feature Selection

From all the information that is sent, some variables do not have relevance in this particular situation. To assess the success of a transaction, information about the basket (products bought, quantity of items,...) or about the customer (i.e. email, phone number, etc) is not important. Moreover, some details, like the shipping address and billing address, don't have any meaning so even though they were considered for feature engineering, they were deleted.

Engineered Features

In order to have more significant features, new variables were engineered through the combination of variables and brainstorming. At first, some engineered features were constructed to represent the degree of similarity between variables that by themselves had no meaning. A binary variable was generated with 1 in case of match and 0 in the opposite case, in the following pair of variables {*BillingName*, *ShippingName*}, {*BillingCountry*, *ShippingCountry*}, {*BillingCity*, *ShippingCity*}. Afterward, several variables were created to summarize the customer's history. Variables like *Chargeback* – that describes the customer's fraud history; *NTransactions* – the total number of attempted transactions made by the customer until that moment; *TotalAmount* – the total amount spent by the customer in past orders; *PastSuccess* – a categorical variable that ranges from Very Low to Very High according to the ratio of successful transactions and total attempted transactions ; *FFAge* – time since the customer has made its first order; and *LastPurchase* – time since the customer's last order.

Data Pre-Processing

In order to have the best performance in machine learning algorithms, data must be clean and complete. However, and inherent to this kind of environments, often data is not filled as it should be, occurring missing values in several variables. The first rule to deal with missing data was to eliminate all the records with more than 5 missing variables. The reason to do so is that an entry with so many missing variables would most probably have been an error in the process of inserting data into the databases. Besides that, any entry with a missing value in *Acquirer* would also be deleted, as this is the variable used to calculate the final success score.

Furthermore, for categorical variables like *Bank*, *Card Type* or *Card Brand* the decision was to replace all the missing values by the same category – "Missing", with the purpose of preserving important data. A special case was the information regarding *Card Country* which is of major importance for routing decisions. This variable was, very often, null. On the other hand, the variable *Billing Country* was almost every time filled correctly, as it is of mandatory fill during the checkout. At the same time, business know-how indicates that those two variables (*Card Country* and *Billing Country*) are intrinsically related. To verify this relationship, the Cramér's V[1] test was used. According to Corbett and Le Roy (2003) the result of $0,691$, represents a strong association between the two variables, therefore, everytime *CardCountry* is null, it is replaced by *BillingCountry*.

A summary of all the variables used in the model can be observed in Appendix B.

5.1.1.2 Training, Test and Validation

Although there is data since 2013, the decision was only to consider data from the previous year, as information older than that might be outdated. Besides that, we decided to take into account the date of the transaction and consider the most recent 30% for testing.

[1] Cramér's V is a measure of association between two nominal variables, giving a value between 0 (no association) and 1 (complete association). It is based on Pearson's chi-squared statistic and was published by Cramer (1946)

The dataset size and distribution is represented in Table 5.1.

Table 5.1: Summary of the dataset split

	Period	**Observations**	**Weight (%)**
Train & Validation set	14/05/2017 - 31/01/2018	1 474 203	70
Test set	01/02/2018 - 30/03/2018	614 639	30
Total	14/05/2017 - 30/03/2018	2 088 842	100

5.1.1.3 Tuning and Model Selection

The first step to select the best technique for this particular case was to filter the best techniques from all the available. Without optimizing its hyperparameters, all the techniques are trained and their performance evaluated. Table 5.2 shows the results of all the techniques in the pre-selection step.

Table 5.2: Performance results of ML pre-selection

Technique	**ROC-AUC**	F_1 **Score**	**Accuracy (%)**	**Runtime (seconds)**
Naive Bayes	0,7610	0,7851	75,20	4,38
Logistic Regression	0,8025	0,8071	76,15	5,63
Random Forests	0,8310	0,8193	77,59	166,13
Gradient Boosting Machines	0,8257	0,8123	77,01	42,77
Neural Network	0,8271	0,8083	77,30	966,56

From all the techniques, the Naive Bayes is the one with the worst performance and the Neural Network, even though it presents on of the best performances, is the one that requires higher computational effort. Since the next step is the hyperparameter optimization, what requires computational resources, the techniques selected for further analysis are the Logistic Regression (good performance and excellent runtime), Random Forests (excellent results and average runtime) and Gradient Boosting Machines (good performance and good average time), because they present a good trade-off between runtime and performance. After that, the framework developed in the last chapter is executed. The first step is the selection of the hyperparameters to be used in each technique.

Logistic Regression: when using this kind of technique, a major issue is to control the overfitting phenomenon. In these cases, it is used a process called regularization that involves adding a penalty to a conventional error function (e.g. square loss) in order to control model complexity. A regression model that uses L_1 regularization is called Lasso Regression and model which uses L_2 is called Ridge Regression. The hyperparameters to be optimized in this technique are alpha and lambda. The alpha specifies the regularization distribution between L_1 and L_2 and the lambda specifies the regularization strength.

Random Forests: in this algorithm, the hyperparameters to be selected are the number of trees built; the leaf size, i.e., the minimum number of observations in each leaf; and the sample rate, i.e., the percentage of the training/validation dataset used to build each tree.

Gradient Boosting Machines: for this technique the hyperparameters used are the number of trees built; the sample rate; and the learn rate, i.e., a parameter that shrinks the contribution of each new tree as it controls the rate at which the loss function is minimized. Smaller values result in greater accuracy because with smaller steps, the optimization is more precise (however, it takes more time because more steps are required).

Table 5.3 summarizes the hyperparameters for each technique.

Table 5.3: Hyperparameters for each technique

Technique	**Hyperparameters**
Logistic Regression	lambda, alpha
Random Forests	number of trees, leaf size, sample rate
Gradient Boosting Machine	number of trees, learn rate, sample rate

With the help of the h2o R package, a random grid search was deployed in order to find the best hyperparameter combination for each technique. The grid search results for each technique can be observed in Table 5.4.

After selecting the best combination of hyperparameters, they are used to test each technique in the test set. The results are shown in Table 5.5 .

The highest performance is achieved by the stacked ensemble, with a performance near to the second highest - random forests. The main performance metric analyzed is the ROC-AUC, however, the $F_1 Score$ and *Accuracy* also give important information. Another important metric is the computational effort (runtime). For that metric the Stacked Ensemble is clearly the worst, as for the construction of that model the underlying models have to be created. Nevertheless this is the method selected for the rest of the heuristic.

5.1.2 Cost

Having access to internal databases where the all the costs per transaction are recorded, a dataset can be retrieved. The cost estimation algorithm receives as input the cost dataset, a list of possible routes to which the new transaction can be routed, and the information of the new transaction. The algorithm starts by restricting the cost dataset to a subset that only contains records with the same *Route* as the one proposed for the new transaction. The approach is to continuously narrow the dataset to have a dataset the closest possible to the new transaction information. With the most appropriate cost subset for the new transaction, all the costs are summed, generating an estimation of a cost rate for that specific transaction. This method is repeated for every route, creating a list of acquirers and the respective expected cost rate. An example of a cost estimation algorithm is described in Appendix C.

Table 5.4: Performance results of Hyperparameters tuning

Panel A: Grid search results for Logistic Regression

Model	Alpha	Lambda	ROC-AUC
1	0,60	5,0e-6	0,8493
2	0,70	1,0e-5	0,8585
3	0,80	1,0e-5	0,8484
4	0,7	1,5e-5	0.8481

Panel B: Grid search results for Random Forests

Model	# Trees	Leaf size	Sample Rate	ROC-AUC
1	28	5	0,6	0,8747
2	50	3	0,7	0,8730
3	100	3	0,5	0,8715
4	50	2	0,6	0,8693

Panel C: Grid search results for Gradient Boosting Machine

Model	# Trees	Learn rate	Sample Rate	ROC-AUC
1	157	0,7	0,6	0,8631
2	250	0,6	0,7	0,8620
3	29	0,7	0,6	0,8593
4	10	0,4	0,6	0,8494

Table 5.5: Performance results of the tuned models

Technique	ROC-AUC	F_1 Score	Accuracy (%)	Runtime (seconds)
Logistic Regression	**0,8143**	0,8149	76,50	6,76
Random Forests	**0,8367**	0,8262	78,06	81,41
Gradient Boosting Machines	**0,8311**	0,8233	77,32	61,50
Stacked Ensemble	**0,8387**	0,8262	78,09	1853,69

5.1.3 Criteria Weights

There are only 2 defined criteria: Success and Cost and 4 stakeholders whose opinion is important in this matter: the Payments team manager – PTM - (who has 4 deciding votes), the director of E-commerce operations – DEO - (who has 2 deciding votes), a Payments team member focused on costs - PTC - (who has 1 deciding vote) and another Payments team member focused on performance – PTP - (who has 1 deciding vote). All the stakeholders' preferences are recorded and the weights calculated, as it is demonstrated in Table 5.6.

All the calculations were done according to Goepel (2013).

Table 5.6: Summary of stakeholders' preferences and respective criteria weights (individual and consolidated)

Stakeholder	Votes	Preferred Criterion	Level of Preference	Weights (%)	
				Success	Cost
PTM	4	Success	9	90	10
DEO	2	Success	2	67	33
PTC	1	Cost	2	33	67
PTP	1	Success	6	86	14
		Consolidated Weights		**80%**	**20%**

5.2 Deployment at the Company

The current routing logic is rigid but well adjusted to the internal processes whereas the deployment of the created heuristic involves several issues, such as the way it will be updated and especially the integration within the information systems. The main challenge is its implementation in real time: a transaction should be routed and a message should be given to the customer in a matter of seconds. Figure 5.1 illustrates the integration of the routing heuristic with the comapany's internal processes and will be explained in the following paragraphs.

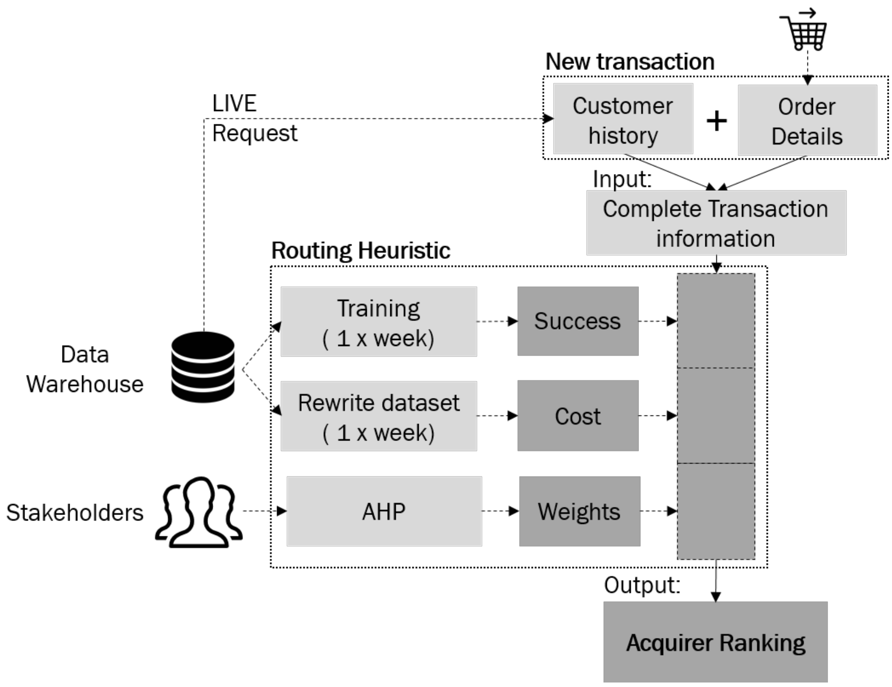

Figure 5.1: Diagram of the routing heuristic and its integration within Farfetch Processes

Success Score

In the order to maintain the relevance of the heuristic it is essential to update the data used for training. All transactions are recorded in databases that are continuously synchronized. Regarding the success score part of the heuristic it contains 2 parts:

1. Training the algorithm;

2. Calculating the success scores;

Training the algorithm happens once every week. As there is no need for a testing set, all the data is used in training. The output of this model is a classifier which will generate the success score.

The classification will run continuously. When there is a new transaction attempt, the details of the transaction are submitted through the Routing API (Application Programming Interface). At the same time the extra information, such as the customer history, is collected from the databases. After that, the necessary transformation of the data is done. With all the features in the right form, the classifier is called to calculate the success score for each one of the possible acquirers. The list of scores is saved to be used in the final heuristic.

Cost Estimation

As for the success score part, the cost estimation needs to be constantly updated. The cost breakdown is complex and volatile. In order to have the most precise estimation of the costs, the dataset used for the estimations uses the last 6 months of information (querying the whole dataset every time a transaction is attempted is not feasible). Given that, as for the score part, the goal is to have a dataset that is updated every week so that the cost calculation could be done in a matter of seconds.

Criteria Weight

From all the parts the criteria weight definition is the one with fewer restrictions. The criteria should be defined before the heuristic starts to be used. After that, the weights can be updated whenever the stakeholders want.

5.3 Heuristic Results

Consider the transactions summarized in Table 5.7:

Table 5.7: Example of new transactions and their main information

TID	Value ($)	Card Country	Card Brand	Success ?
1	161,26	US	Visa	Yes
2	1120,26	Hong Kong	Visa	Yes
3	119,65	France	MasterCard	Yes
4	1027,22	UK	MasterCard	No

Assuming that we want the real expected value (when both criteria weights are 1) we execute the heuristic for the transaction 2. Table 5.8 summarizes the results for each acquirer:

Table 5.8: Heuristic results for transaction 2 (TID=2 and value=1120,26 $)

Acquirer	**Success Score**	**Cost Rate**	**E(Revenue) ($)**	**E(Cost) ($)**	**E(Transaction) ($)**
A_Main_ROW2	0,8125	5,210%	910,237	47,427	862,811
A_Main_ROW1	0,7917	5,010%	886,911	44,438	842,473
A_Main_US2	0,7943	6,752%	889,871	60,080	829,791
A_Main_US1	0,7824	6,868%	876,504	60,202	816,302

A_Main_ROW2 is the selected acquirer according to the expected value of the transaction. In this case, the success score plays an important role, even without having the cheapest cost. However, this is the simplest strategy, yet the heuristic can be used in several ways. Table 5.9 summarizes the results of the different strategies for the same transaction.

Table 5.9: Different strategies results for transaction 3 (TID=3 and value=119,65 $)

Strategy	**Acquirer**	**Success Score**	**Cost Rate**	**E(Revenue) ($)**	**E(Cost) ($)**	**E(Transaction) ($)**
Success	A_Main_ROW2	0,8899	2,608%	106,475	2,776	106,475
Cost	A_Main_ROW1	0,8792	0,635%	105,195	0,669	-0,669
Success + Cost	A_Main_ROW1	0,8792	0,635%	105,195	0,669	104,526
With Weights	A_Main_ROW2	0,8899	2,608%	106,475	2,776	84,624

In this case, we can see that different strategies have different results. If the strategy is only to focus on Success or to take into account the stakeholders' preferences, the choice rests on A_Main_ROW2. If the strategy's focus is on the Cost or on the real expected value of the transaction, the choice is A_Main_ROW1.

5.4 Simulation Results

In order to evaluate the potential impact of the heuristic, a simulation is done. The goal of this simulation is to compare the current routing decision system with the proposed heuristic.

Basically, random transactions are picked from the test set and a comparison between the two systems is done. Currently, the number of daily transactions is rising exponentially (e.g. the daily number of Visa and MasterCard transactions for April 2018 was around 6800, whereas in 2017, for the same month, the number was down to 4600: an increase of almost 50%).

Even with this strong growth, a conservative approach is adopted. One day of transactions is simulated (only around 4600 transactions are considered). To analyze the current routing system, the value of the order, the success (or not) of the transaction and the respective cost are taken into account. To analyze the proposed heuristic, the only thing to take into account is the real expected value of the transaction (that already incorporates the success score and the cost). Table 5.10 summarizes the results of the simulation.

Table 5.10: Comparison beetween the results of the current routing system and the proposed heuristic

	Current Routing System	Routing Heuristic	**Potential Gain**
Success Rate (%)	70,91	72,19	**1,28**
Value per Order ($)	420	456	**36**

Assuming that the heuristic (and especially the underlying machine learning model) is well developed, for the same transactions, the success rate would increase 1,28 % (i.e. the number of successful transactions would increase 1,28% with the new payment system). Another interesting metric is that, for 60% of the transactions, the acquirer selected by the heuristic is different from the original one.

Sensitivity Analysis

The cost part the of the heuristic is an estimation, whose accuracy cannot be precisely assessed. On the other hand, the machine learning model is also not 100% accurate, though we can assess its performance through the metrics already explained. Given that, the heuristic's output can fluctuate according to the model's accuracy.

In order to assess the impact of the model's performance in the overall results, a sensitivity analysis was performed. The range of the sensitivity analysis was defined by the accuracy of the model. Maintaining all the remaining characteristics, the success core of the selected model was continuously changed by 1 p.p. (from -10% to +10%) and its impact assessed.

Performing this analysis, many times the second best acquirer outperformed the selected acquirer (due to the decrease/increase in the success score of the selected acquirer). There is an exponential increase in the total value when the value tends to +10%, whereas when the value tends to -10% the marginal decrease is smaller. Although this allows us to have an idea of the magnitude of the heuristic, the proper way to analyze its impact is to use the potential gain [2] per transaction. Figure 5.2 shows the variation in potential gain according to the change in the success score of the selected acquirer.

In the worst case scenario, this value is around 27$. This indicates that using the heuristic, Farfetch is expected to gain, on average, 27$ per order (just by making a better routing decision that increases the probability of success and decreases the cost of the transaction).

We can now use this approximate value to estimate the impact of the heuristic. For all the transactions made in 2017 using Visa and MasterCard (around 3,5 million) Farfetch would have a potential gain of around 50 million US dollars. It is a considerable amount but less relevant when compared to the total amount traded (more than 1 billion US dollars). Although that estimate seems incredibly encouraging, it should be analyzed prudently and in a conservative way.

[2] The true meaning of this value is the "non-lost value" instead of potential gain, i.e. if a transaction fails Farfetch will lose the correspondent amount, but if that same transaction was accepted, Farfetch would end up not losing that amount

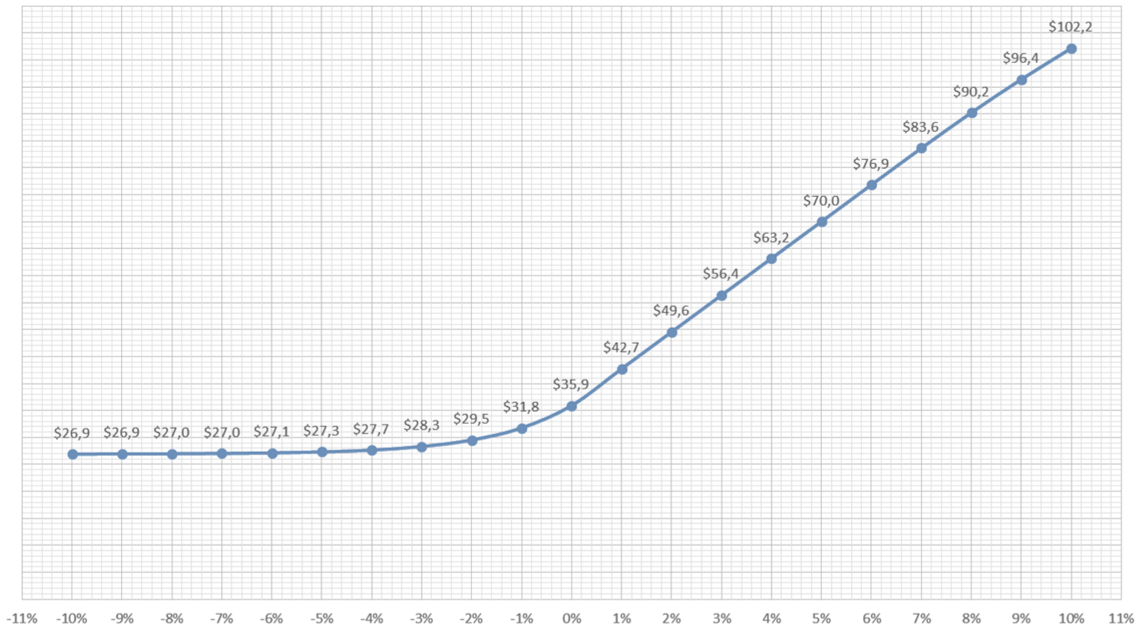

Figure 5.2: Variation in potential gain according to the success score of the selected acquirer

For the heuristic formulation no restrictions were taken into account, i.e. the heuristic was formulated having total freedom in the routing decision system. Besides that, only the monitoring of its implementation will bring meaningful insight into this analysis. Still, the preliminary results associated with the expected growth of the company are encouraging for this project.

Chapter 6

Conclusions and Future Work

This dissertation addresses the problem of payments in high volume e-commerce environments and details the many steps that need to be taken in order to develop a complete heuristic for e-payments routing. The benefits of this heuristic are not only to increase the potential gains of the company but also to improve customer experience and satisfaction, a critical issue in e-commerce environments. Although the heuristic is to be implemented, the methodology could be applied in all e-commerce situations, with preference to where big data is available.

6.1 Conclusions on the Heuristic

Even though there are several studies on machine learning approaches to e-commerce, the current state of the art in machine learning approaches for e-payments routing is poor. The concept of smart routing is increasing amongst e-commerce companies, but the innovation in this field is still scarce. Thus, the main takeaway from this project is that data-driven and machine learning techniques have an expected higher performance when compared to traditional decision methods.

Regarding the methodology, it can be verified that the model calibration might have a significant impact on the performance of the models. Many times, the different techniques come with default values for their hyperparameters but, as seen in this work, the algorithms are sensitive to the choice of hyperparameters. Therefore, tuning the hyperparameters is likely to improve results. With respect to model selection, testing several techniques is interesting from a knowledge discovery point of view. The goal of this step is to find the most accurate one from all the techniques. In the studied company, the best technique was the stacked ensemble but with a marginal difference to the second best technique, the random forests. From a business point of view, the deployment of a random forest instead of a stacked ensemble would be something to reflect upon, as the gain in using the best technique might not outweigh the computational effort of building a stacked ensemble.

A major advantage of the heuristic is that it is a scalable method. The business growth is inevitably related to an increase in the complexity of the payment system. With this heuristic, a decrease in routing process complexity is expected as well as an increase in its precision. We hope that this methodology might provide some useful e-payments insights in the perspective of merchants and that it could be a motivation for others to further explore the problem and present more innovative and valuable solutions.

6.2 Challenges, Limitations and Future Improvements

Despite the foreseeable improvements in the payments routing performance there are still opportunities for improvement. The fact is that both machine learning model and cost estimation can be more precise. A deeper analysis of all the variables that might impact the success of a transaction as well as more correct datasets can have a significant impact on the output of the heuristic. This is something that is out of the scope of this thesis since it implies a solid exchange of information with internal and external parties. A closer relationship with internal teams, such as the Business Intelligence and Data Science team, which can provide more data of quality, computational resources, and know-how; and trustworthy exchange of information with external entities that can provide us more meaningful and accurate information (such as more precise cost datasets).

Besides that, some assumptions were made during the development of the heuristic that might impact its final result. It was assumed that no legal restrictions would exist with regard to the e-payments routing, which is not completely true. Given that, in order to implement the heuristic several topics should be investigated in detail, such as legal, financial or data issues.

The proposed heuristic is the first step in a global strategy of e-payments optimization at Farfetch. The next step is to extend the heuristic so that the routing decision takes both provider and acquirer into account, being the final output the best combination of provider and acquirer for a given transaction.

A future extension to this project is the development of an API, where stakeholders can monitor the functioning of the heuristic and modify its internal settings (such as criteria weights, manual restrictions, etc).

Besides that, the idea is to continuously improve the API/heuristic adding fraud filters (that can, in advance, cancel orders according to their fraud results), fallback strategies (where the result of the heuristic can be combined with the information of the transaction failure, for a new routing decision) and more features that might be useful.

Bibliography

Abdellaoui, R., Pasquet, M., and Berthelier, O. (2011). Integration of new electronic payment systems into b2c internet commerce. In *Collaboration Technologies and Systems (CTS), 2011 International Conference on*, pages 484–491. IEEE.

Adyen (2016). Pricing. `https://www.adyen.com/pricing#faq`. [Online; accessed 19-April-2018].

Amatriain, X. (2016). What are hyperparameters in machine learning? `https://www.quora.com/What-are-hyperparameters-in-machine-learning`. [Online; accessed 31-May-2018].

Amed, I., Achim Berg, S. K., Saskia Hedrich, J. A., and Martine Drageset, R. Y. (2017). *The State of Fashion 2018*. McKinsey Global Institute and Business of Fashion.

Belton, V. and Stewart, T. (2002). *Multiple criteria decision analysis: an integrated approach*. Springer Science & Business Media.

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

Bjørn-Andersen, N. and Hansen, R. (2011). The adoption of web 2.0 by luxury fashion brands. *Proceedings of CONFIRM, Paper*, 34.

Bogetoft, P. and Pruzan, P. (1997). *Planning with multiple criteria: investigation, communication and choice*. Handelshøjskolens Forlag.

Bousquet, O., von Luxburg, U., and Rätsch, G. (2011). *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*, volume 3176. Springer.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

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

Brownlee, J. (2016). A gentle introduction to the gradient boosting algorithm for machine learning. `https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/`. [Online; accessed 2-June-2018].

Calantone, R. J., Di Benedetto, C. A., and Schmidt, J. B. (1999). Using the analytic hierarchy process in new product screening. *Journal of Product Innovation Management*, 16(1):65–76.

Carbonell, M., Torres, J., Izquierdo, A., and Suarez, D. (2008). New e-payment scenarios in an extended version of the traditional model. In *International Conference on Computational Science and Its Applications*, pages 514–525. Springer.

Castillan, L., Chheang, C., Denoux, C., Ferrenbach, C., Gérard, C., Hua, E., Julia, V., and Manière, I. C. (2017). Online luxury: The code breakers of a traditional sector. *Procedia Computer Science*, 122:579–586.

Chakravorti, S. (2003). Theory of credit card networks: A survey of the literature. *Review of network Economics*, 2(2).

Chargeback911 (2016). Chargebacks explained. `https://chargebacks911.com/chargebacks/`. [Online; accessed 19-May-2018].

Corbett, M. and Le Roy, M. K. (2003). *Research methods in political science: an introduction using MicroCase*. Wadsworth Pub Co.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 215–242.

Cramer, H. (1946). Mathematical methods of statistics princeton university press usa.

CRISP-DM (2006). Crisp-dm. `http://www.crisp-dm.org/`. [Online; accessed 19-April-2018].

Evans, C. (2001). An e-strategy for online e-business. *Information Systems Management*, 18(4):8–21.

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

Fayyad, U. (2001). Knowledge discovery in databases: An overview. In *Relational Data Mining*, pages 28–47. Springer.

Feigenbaum, E. A., Feldman, J., et al. (1963). *Computers and thought*. New York.

Fernandez, J. (2009). Bin routing for payment processing. *US Patent App. 12/634,472*.

Gandibleux, X. (2006). *Multiple criteria optimization: state of the art annotated bibliographic surveys*, volume 52. Springer Science & Business Media.

Goepel, K. D. (2013). Implementing the analytic hierarchy process as a standard method for multi-criteria decision making in corporate enterprises–a new ahp excel template with multiple inputs. In *Proceedings of the international symposium on the analytic hierarchy process*, volume 2013, pages 1–10.

Greening, L. A. and Bernow, S. (2004). Design of coordinated energy and environmental policies: use of multi-criteria decision-making. *Energy policy*, 32(6):721–735.

Güneş, F. (2017). Why do stacked ensemble models win data science competitions? `https://blogs.sas.com/content/subconsciousmusings/2017/05/18/stacked-ensemble-models-win-data-science-competitions/`. [Online; accessed 2-June-2018].

Ham, F. M. and Kostanic, I. (2000). *Principles of neurocomputing for science and engineering*. McGraw-Hill Higher Education.

Handfield, R., Walton, S. V., Sroufe, R., and Melnyk, S. A. (2002). Applying environmental criteria to supplier assessment: A study in the application of the analytical hierarchy process. *European Journal of Operational Research*, 141(1):70 – 87.

Har-Peled, S., Roth, D., and Zimak, D. (2003). Constraint classification for multiclass classification and ranking. In *Advances in neural information processing systems*, pages 809–816.

Hines, T. and Bruce, M. (2007). *Fashion marketing*. Routledge.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.

Hunt, R. M. (2003). An introduction to the economics of payment card networks. *Review of Network Economics*, 2(2).

IBM (1997). e-business. `http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/ebusiness/transform/`. [Online; accessed 18-April-2018].

Jong, N. K. and Stone, P. (1976). Keeney, rl &raiffa, h. decisions with multiple objectives: Preferences and value tradeoffs. In *In Proceedings of the ICML-06 Workshop on Kernel Methods in Reinforcement Learning*. Citeseer.

Kapferer, J.-N. (2012). Abundant rarity: The key to luxury growth. *Business Horizons*, 55(5):453–462.

Khan, D., Varshney, P., and Qadeer, M. A. (2011). E-commerce: From shopping carts to credit cards. In *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, pages 81–85. IEEE.

Kim, C., Tao, W., Shin, N., and Kim, K.-S. (2010). An empirical study of customers' perceptions of security and trust in e-payment systems. *Electronic commerce research and applications*, 9(1):84–95.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.

Kossler, T. (2013). *How credit card processing works,the payments system simplified. Technical report*. 2Checkout.com, Inc.

Liu, J., Xiao, Y., Chen, H., Ozdemir, S., Dodle, S., and Singh, V. (2010). A survey of payment card industry data security standard. *IEEE Communications Surveys & Tutorials*, 12(3):287–303.

Løken, E. (2007). Use of multicriteria decision analysis methods for energy planning problems. *Renewable and Sustainable Energy Reviews*, 11(7):1584–1595.

Malczewski, J. (1999). *GIS and multicriteria decision analysis*. John Wiley & Sons.

Malczewski, J. (2006). Gis-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20(7):703–726.

Mirescu, S. V. et al. (2010). The premises and the evolution of electronic commerce. *Journal of Knowledge Management, Economics and Information Technology*, 1(1):5–5.

Mosca, F. (2016). *Global marketing strategies for the promotion of luxury goods*. IGI Global.

Mu, E. and Pereyra-Rojas, M. (2018). Understanding the analytic hierarchy process. In *Practical Decision Making using Super Decisions v3*, pages 7–22. Springer.

Nagpal, A. (2017). Decision tree ensembles- bagging and boosting. `https://towardsdatascience.com/decision-tree-ensembles-bagging-and-boosting-266a8ba60fd9`. [Online; accessed 2-June-2018].

Niranjanamurthy, M., Kavyashree, N., Jagannath, S., and Chahar, D. (2013). Analysis of e-commerce and m-commerce: advantages, limitations and security issues. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(6).

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

Okonkwo, U. (2009). Sustaining the luxury brand on the internet. *Journal of brand management*, 16(5-6):302–310.

Perry, C. (2012). Authorization, capture & settlement. `https://www.totalmerchantconcepts.com/blog/posts/authorization-capture-and-settlement.html`. [Online; accessed 29-May-2018].

Prager, R. A., Manuszak, M., Kiser, E. K., Borzekowski, R., et al. (2009). Interchange fees and payment card networks: Economics, industry developments, and policy issues. *Finance and Economics Discussion Series*, 23.

Rakotomamonjy, A. (2004). Optimizing area under roc curve with svms. In *ROCAI*, pages 71–80.

Saaty, T. L. (1980). The analytic hierarchy process: planning, priority setting, resources allocation. *New York: McGraw*, 281.

Saxena, R. (2017). How decision tree algorithms work? `https://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/`. [Online; accessed 2-June-2018].

Sellxed (2016). What are the functions of the psp and the acquirer? `https://www.sellxed.com/en/functions-psp-and-acquirers`. [Online; accessed 19-April-2018].

Smolyakov, V. (2017). Ensemble learning to improve machine learning results. `https://blog.statsbot.co/ensemble-learning-d1dcd548e936`. [Online; accessed 2-June-2018].

Statista (2018). Retail e-commerce sales worldwide from 2014 to 2021 (in billion u.s. dollars). `https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/`. [Online; accessed 21-May-2018].

Stewart, T. J. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *Omega*, 20(5-6):569–586.

Tan, P.-N., Steinbach, M., and Kumar, V. (2006). Classification: basic concepts, decision trees, and model evaluation. *Introduction to data mining*, 1:145–205.

Tian, Y. and Stewart, C. (2006). History of e-commerce w: Encyclopedia of e-commerce. *E-government, and Mobile Commerce-M. Khosrow–Pour, London, Idea Group Reference*, pages 559–560.

Van der Valk, R. (2015). Why my payment got rejected: A method to mine payment refusal clues.

Van Slyke, C. and Belanger, F. (2003). E-business technologies. *New York*.

Walker, S. H. and Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54(1-2):167–179.

Weir, C. S., Anderson, J. N., and Jack, M. A. (2006). On the role of metaphor and language in design of third party payments in ebanking: Usability and quality. *International Journal of Human-Computer Studies*, 64(8):770–784.

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pages 29–39. Citeseer.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.

Yao-Hua Tan, W. T. (2000). Toward a generic model of trust for electronic commerce. *International journal of electronic commerce*, 5(2):61–74.

Zhang, J., Williams, S. O., and Wang, H. (2017). Intelligent computing system based on pattern recognition and data mining algorithms. *Sustainable Computing: Informatics and Systems*.

Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*. CRC press.

Appendix A

Analytical Hierarchy Process

While designing an AHP model several components have to be addressed. The DMs should clearly state the objective, define the relevant criteria, the relevant sub-criteria, if existent, and determine the different alternatives to be evaluated. According to Mu and Pereyra-Rojas (2018), the DM starts by considering each of the criteria pair-wise and assigning a relative importance to the criteria, usually on scale where 1 represents equal importance and 9 extreme importance, as it is displayed in Table A.1

Table A.1: Preferences table, constructed by Saaty (1980)

1	Equally preferred
3	Weak preference
5	Strong preference
7	Very strong preference
9	Extreme importance
2,4,6,8	Intermediate values

These relative importances are then used to construct a preference matrix, from which the weights for each criterion will be extracted. Note that the preference matrix is a square reciprocal matrix as it is demonstrated in Table A.2. (Handfield et al., 2002).

Table A.2: Example of preference matrix

	Criterion 1	Criterion 2	Criterion 3
Criterion 1	1	P_{12}	P_{13}
Criterion 2	$1 / P_{12}$	1	P_{23}
Criterion 3	$1 / P_{13}$	$1 / P_{23}$	1

Saaty (1980) recommends using a normalized eigenvector approach to calculate the weights for each criterion. This mathematical procedure is quite complex (more details can be found in Saaty (1980)) but in the end the expected result is to have normalized weights assigned to each criterion according to the individual input of each DM. With the individual criteria weights, it is also possible to calculate the overall criteria weights, by combining the individual weights and the votes of each DM (according to its importance) (Mu and Pereyra-Rojas, 2018).

Appendix B

Input Variables

Table B.1: Summary of input variables

Name	Type	Description	Engineered?
Success	Categorical	Class label that indicates the success(1) or failure(0)	No
Fallback	Categorical	If transaction is from main or fallback system	No
Currency	Categorical	Currency in which the transaction is made	No
OrderValue	Numeric	Value of the order	No
Provider	Categorical	Name of the provider	No
Acquirer	Categorical	Name of the acquirer	No
Bank	Categorical	Name of the customer's bank	No
BillingCountry	Categorical	Country indicated by the customer for billing	No
ShippingCountry	Categorical	Country indicated by the customer for shipping	No
Origin	Categorical	Origin of the transaction (Portal, App,..)	No
CardCountry	Categorical	Country where the card is from	No
CardBrand	Categorical	Card brand (VISA or MasterCard)	No
CardType	Categorical	Type of card (Debit, Credit, ...)	No
CardLevel	Categorical	Level of card (Premium, Standard, ...)	No
Tokenization	Categorical	If card data was saved by customer for re-use	Yes
CheckName	Categorical	If billing and shipping name are the same	Yes
CheckCity	Categorical	If billing and shipping city are the same	Yes
CheckCountry	Categorical	If billing and shipping country are the same	Yes
Weekday	Categorical	Weekday	Yes
Year	Categorical	Year	Yes
Month	Categorical	Month	Yes
Day	Categorical	Day	Yes
Hour	Categorical	Hour	Yes
Chargeback	Categorical	If the customer has a chargeback history	Yes
NTransactions	Numeric	Total number of attempts made by the customer	Yes
TotalAmount	Numeric	Total amount spent by the customer	Yes
PastSuccess	Categorical	Class defined by the customer's success history	Yes
FFAge	Numeric	Time since user's first order	Yes
LastPurchase	Numeric	Time since user's last order	Yes

Appendix C

Cost Estimation Algorithm

input : Cost dataset, $Costs$;
List of possible routes, $Routes$;
Transaction information, $Transaction$;
output: List of $CostRate$ estimations ;
initialization;
foreach $route$ ***in*** $Routes$ **do**
    create $Subset$ where $Costs_{route} = route$;
    if $Transaction_{CardBrand} \subset Subset$ **then**
        rewrite $Subset$ where $Subset_{CardBrand} = Transaction_{CardBrand}$;
        if $Transaction_{CardCountry} \subset Subset$ **then**
            rewrite $Subset$ where $Subset_{CardCountry} = Transaction_{CardCountry}$;
            if $Transaction_{CardType} \subset Subset$ **then**
                rewrite $Subset$ where $Subset_{CardType} = Transaction_{CardType}$;
                if $Transaction_{CardLevel} \subset Subset$ **then**
                    rewrite $Subset$ where $Subset_{CardLevel} = Transaction_{CardLevel}$;
                end
            end
        end
    end
    $TotalCost = \sum Subset_{Cost}$;
    $TotalTransactionValue = \sum Subset_{OrderValue}$;
    $CostRate_{route} = \frac{TotalCost}{TotalOrderValue}$;
end

Algorithm 2: Algorithm for estimating the cost rate per route of a new transaction