

Otimização das Quantidades de Compra

Jorge Emanuel Paulino e Pessoa

Dissertação de Mestrado

Orientador na FEUP: Prof. Vera Lúcia Miguéis Oliveira

Orientador na *Parfois*: Afonso Manuel da Silva Antunes Moreira



Mestrado Integrado em Engenharia Industrial e Gestão

2016-07-26

*Valeu a pena? Tudo vale a pena
Se a alma não é pequena.*
Fernando Pessoa

Resumo

No contexto do retalho, uma determinação errónea das quantidades a comprar de cada artigo aos fornecedores, quer por excesso quer por defeito, pode resultar em custos desnecessários de armazenamento ou perdas de vendas, respetivamente. Ambas as situações são de evitar para as empresas, advindo daí a necessidade de determinar esses valores eficientemente.

É neste âmbito que se insere o tema desta dissertação, cuja finalidade é a otimização das quantidades a comprar de cada artigo, recorrendo à análise do comportamento de artigos do histórico, com características similares, com vista a uma melhor previsão dos valores a adquirir em cada estação.

Para a consecução deste objetivo foram utilizadas ferramentas como o *Microsoft Excel*, fundamental na análise e preparação dos dados, assim como o programa de cálculo estatístico R, nomeadamente o pacote *random forest*.

Numa primeira fase, após definição dos dados a utilizar como histórico, procedeu-se a uma análise do comportamento dos artigos considerados durante o período normal da estação em causa, sobretudo ao nível das vendas perdidas devido a ruturas de inventário. Este estudo permitiu ajustar os valores das vendas reais, passando a ser tidos em linha de conta os valores de potenciais vendas que completam os anteriores com uma estimativa das vendas perdidas nos casos de rutura de *stock*.

Considerando os valores de vendas potenciais, que representam um aumento de cerca de 13% face aos valores iniciais, assim como as características definidas como relevantes dos artigos aos quais estão associadas, recorreu-se a uma técnica de *data mining*, denominada de *random forest*, que permitiu, através dos dados considerados, fazer uma previsão de valores de venda para novos artigos. O modelo desenvolvido permitiu obter um nível de variância explicada de 86,15% na determinação das vendas para os artigos.

Por fim, essas previsões de vendas dos novos artigos, obtidas com o modelo, são utilizadas para a determinação das quantidades de compra. À previsão de vendas em período normal é incrementado o valor de vendas esperado no período de saldos da estação, dado por uma percentagem objetivo de unidades a entrar em saldos, previamente definida pela empresa com base num conjunto de atributos dos artigos e lojas em que serão comercializados.

Todo este processo foi incluído numa ferramenta criada com base no programa *Microsoft Visual Basic* do *Microsoft Excel*, com o intuito de facilitar ao utilizador a obtenção do valor de quantidades de compra para os artigos pretendidos para a estação seguinte.

Abstract

In the retail context, an erroneous determination of the amounts to buy of each article from the suppliers, either by excess or defect, can result in unnecessary costs of storage or lost sales, respectively. Both situations should be avoided by companies, which promotes the need to determine these values efficiently.

It is in this scope that fits the theme of this dissertation, whose purpose is to optimize the amounts to buy of each item, using the analysis of the historical sales behaviour of items with similar characteristics, in order to improve the prediction of the values to be acquired in each season.

To achieve this goal, Microsoft Excel was used to analyze and prepare the data. R statistical calculation program, namely the random forest package, was also used to run the model.

Initially, after defining the data to be used as history, it was carried out an analysis of the considered items' behaviour during the normal period of the season in question, particularly in terms of lost sales due to inventory disruptions. This study allowed the adjustment of the actual sales figures, motivating the usage of the potential sales figures that complete the previous with an estimate of lost sales in the case of stockout.

Considering the potential sales figures, which represent an increase of about 13% compared to the baseline, as well as the relevant characteristics of the articles to which they are related, it was used a data mining technique, named random forests, which allowed, with the considered data, to predict sales values for new items. The created model enables to obtain a level of explained variance of 86.15% in the determination of sales for the items.

Finally, these sales predictions for the new products, obtained with the model, are used to determine the purchase amounts. To the prediction of sales in the normal period it is increased the sales value expected for the sales season, determined by a percentage goal of units to enter into sales, previously defined by the company based on a set of attributes of items and stores in which they will be sold.

This whole process is included in a tool created based on Microsoft Visual Basic program of the Microsoft Excel, in order to facilitate the obtainment of purchasing amounts for the intended items to the next season.

Agradecimentos

A conclusão desta etapa final do meu percurso académico só foi possível graças ao contributo de algumas pessoas, às quais pretendo mostrar o meu agradecimento.

À orientadora da Faculdade de Engenharia da Universidade do Porto, a professora Vera Miguéis, pela disponibilidade apresentada durante todo o projeto e pelos conhecimentos transmitidos.

Agradeço igualmente ao orientador do projeto na *Parfois*, Afonso Moreira, pelo auxílio na intergração na empresa e nos processos da mesma, assim como pelo tempo que despendeu para garantir que todas as informações necessárias para a execução do projeto me fossem transmitidas.

Aos elementos que compõem o departamento de Controlo de Gestão e Negócio, Catarina Morgado, Enrique Aparício, Francisca Marinho, Joana Mota, Maria Muñoz, Mónica Rocha e Vanja Jovanovic pelo ótimo ambiente proporcionado e pelos contributos dados nas múltiplas reuniões de acompanhamento do projeto.

Aos elementos do departamento de Compras, Juan Redondo, Miguel Rodrigues e Sofia Andrade pelo acompanhamento contínuo na melhoria da ferramenta desenvolvida.

À Inês Simões, Luís Calejo, Miguel Carvalho e Tiago Castro, colegas de curso também com projetos na *Parfois*, pelos bons momentos passados e pelas sessões de *brainstorming* coletivo.

Não posso deixar de agradecer também às pessoas que conheci e que me acompanharam durante esta nesta nova etapa que percorri, Ana Tavares, Emanuel Lopes, Inês Miroto, Inês Silva, Miguel Fernandes, Pedro Balsemão, Sara Palhares, Teresa Morim e Tiago Martins.

Agradeço também aos amigos que trouxe da minha estadia em Linköping, Bernardo Marçal, Marco Gentil e Ricardo Martins.

À minha irmã, mãe e pai por tudo aquilo que me proporcionaram e pelo apoio que me deram.

Por fim, mas não menos importante, ao meu avô, Fernando Pessoa, que me deixou um legado de perseverança, fazendo-me acreditar que, tal como o seu homónimo dizia, “tudo vale a pena se a alma não é pequena”.

Um bem-haja a todos,

Jorge Emanuel Paulino e Pessoa

Índice de Conteúdos

1	Introdução	1
1.1	Enquadramento.....	1
1.2	Apresentação da empresa	2
1.2.1	Visão, Missão e Valores	2
1.2.2	Modelo de Negócio	3
1.3	Objetivos do projeto	3
1.4	Metodologia.....	4
1.5	Estrutura da dissertação	5
2	Enquadramento Teórico.....	6
2.1	Retalho de Moda.....	6
2.1.1	Fast-Fashion.....	7
2.1.2	<i>Quick Response</i>	8
2.2	Descoberta de Conhecimento em Bases de Dados.....	8
2.2.1	Data Mining.....	9
2.2.1	Random Forest.....	10
3	Modelo atual de obtenção de Quantidades de Compra	15
3.1	Os produtos Parfois.....	15
3.2	As lojas Parfois	19
3.3	Modelo atual de obtenção de quantidades de compra.....	20
4	Modelo proposto de obtenção de Quantidades de Compra	23
4.1	Dados recolhidos	23
4.2	<i>Stockouts</i> e Vendas Potenciais.....	24
4.3	Modelo de previsão.....	30
5	Apresentação e análise de resultados	35
5.1	Análise de Resultados.....	35
5.2	Ferramenta de determinação de quantidades de compra.....	36
6	Conclusões e perspetivas de trabalho futuro.....	40
6.1	Satisfação dos objetivos.....	40
6.2	Trabalhos Futuros	41
	Referências	42
	Anexo A: Organograma da <i>Parfois</i>	44
	Anexo B: Cronograma do projeto.....	45
	Anexo C: Categorias de atributos usados no modelo.....	46

Índice de Figuras

Figura 1 - Vendas e Tempo de Reação ao Mercado.	7
Figura 2 - Etapas do Processo de Descoberta de Conhecimento em Bases de Dados	9
Figura 3 – Exemplo do funcionamento de uma árvore de decisão	11
Figura 4 - Processo de treino usando random forests	13
Figura 5 – Processo de regressão utilizando random forests	14
Figura 6 - Ilustração de produtos das diferentes gamas da Parfois	16
Figura 7 - Distribuição das Vendas por Gama, em Quantidade em 2015.....	16
Figura 8 - Distribuição das Vendas por Gama, em Valor em 2015.....	17
Figura 9 – Estrutura mercadológica dos artigos da Parfois	17
Figura 10 - Distribuição de artigos pelos níveis de aposta em cada gama	18
Figura 11 - Distribuição geográfica das lojas Parfois	19
Figura 12 – Prestação de um artigo em duas lojas, com base nas semanas do ano	24
Figura 13 – Prestação de um artigo em duas lojas, com base nas semanas de vida do produto	25
Figura 14 – Representação da prestação de um artigo numa certa loja.....	25
Figura 15 - Fluxograma dos processos de cálculo de vendas potenciais	26
Figura 16 - Representação das vendas reais e potenciais da gama Carteiras ao longo das semanas	29
Figura 17 - Relação entre a quantidade compra por artigo e o mês de entrada em loja	31
Figura 18 - Relação entre o erro out-of-bag e o número de árvores.....	33
Figura 19 - Relação entre o erro out-of-bag e o número de atributos	33
Figura 20 - Exemplo de uma das árvores de decisão criadas pelo modelo.....	34
Figura 21 – Resultados obtidos com a utilização da ferramenta <i>randomforest</i>	35
Figura 22 - Importância dos atributos escolhidos.....	36
Figura 23 - Janela inicial da ferramenta criada	37
Figura 24 - Janela para análise das características pretendidas no histórico.....	37
Figura 25 - Janela com a apresentação do artigo criado e respetivas quantidade de compra	38
Figura 26 - Janela com a apresentação de um ranking de outras possibilidades de cor do artigo criado e respetivas quantidade de compra	38
Figura 27 - Janela com a apresentação de artigos do histórico dentro de um certo grau de semelhança face ao artigo criado	39
Figura 28 – Restante parte da janela com a apresentação de artigos do histórico dentro de um certo grau de semelhança face ao artigo criado	39
Figura 29 - Organograma da Parfois	44
Figura 30 - Cronograma das etapas do projeto	45

Índice de Tabelas

Tabela 1 - Percentagem atuais de entradas em Saldos da <i>Parfois</i>	4
Tabela 2 - Relação entre tamanho das lojas e potencial de venda.....	20
Tabela 3 - Relação entre tipologia de loja e aglomerado em que se insere	20
Tabela 4 – Representação da tabela com as características dos artigos considerados.....	23
Tabela 5 – Tabela auxiliar de determinação dos pesos do dia da semana numa loja	27
Tabela 6 – Cálculo dos pesos a utilizar para a determinação das vendas potenciais	27
Tabela 7 - Cálculo das vendas a utilizar para a determinação das vendas potenciais	27
Tabela 8 – Determinação das vendas potenciais para situações de <i>stockout</i> entre 0% e 100% na semana.....	28
Tabela 9 – Determinação do nº de lojas e o valor das vendas do artigo no país/região.....	28
Tabela 10 - Determinação das vendas potenciais para situações de 100% de <i>stockout</i> na semana.....	28
Tabela 11 - Variação das vendas potenciais face às vendas reais e quantidade de <i>stockouts</i> no período considerado	29
Tabela 12 - Relação entre <i>Fashion Type</i> e o número de semanas em que se espera que o produto esteja em loja	30
Tabela 13 - Variação das vendas potenciais face às vendas reais e quantidade de <i>stockouts</i> para as restrições definidas	30
Tabela 14 - Exemplo da variável dependente e atributos usados no modelo.....	31
Tabela 15 - Exemplo da venda potencial e atributos usados no modelo, repartidos por tipologia.....	32
Tabela 16 - Número de lojas, por tipologia, em que os artigos estiveram expostos	32
Tabela 17 - Exemplo da venda potencial média e atributos usados no modelo, repartidos por tipologia	32
Tabela 18 - Categorias possíveis para os atributos “Família”, “SubFamília”, “Tipo Cor” e “Cor”	46
Tabela 19 - Categorias possíveis para os atributos “Fashion”, “Preço Base”, “Mês”, “Tipologia” e “Aposta”.....	47

1 Introdução

A presente dissertação foi realizada no âmbito do Mestrado Integrado em Engenharia Industrial e Gestão, da Faculdade de Engenharia da Universidade do Porto (FEUP).

O projeto foi desenvolvido na empresa *Parfois* – Barata & Ramilo, SA, mais especificamente no departamento de Controlo de Gestão e Negócio.

Neste capítulo, é feito o enquadramento do projeto no âmbito da empresa em que foi desenvolvido, assim como a apresentação da mesma. De seguida, apresentar-se-ão os objetivos do projeto, a metodologia utilizada e a estrutura da dissertação.

1.1 Enquadramento

A empresa *Parfois* é constituída por vários departamentos, apresentados, detalhadamente, no organograma (Anexo A).

Este projeto foi enquadrado na área de Desenvolvimento do departamento de Controlo de Gestão e Negócio, possuindo uma forte interligação com o departamento de Compras que usufruirá da ferramenta desenvolvida.

No departamento de Controlo de Gestão e Negócio, é desenvolvida uma vasta diversidade de atividades e ferramentas da empresa, tais como o desenvolvimento de orçamentos, a coordenação de atividades entre os diversos departamentos da organização, a comunicação e avaliação da informação de gestão e a criação, implementação e acompanhamento de ferramentas de controlo, tudo com o intuito de agilizar processos e otimizar o desempenho da empresa.

Por interagir com várias áreas de atividade da *Parfois*, o departamento de Controlo de Gestão e Negócio encontra-se subdividido em três áreas: Controlo de Compras, Controlo Comercial e Desenvolvimento.

No setor do Controlo de Compras, que está diretamente relacionado com o departamento de Compras, desenvolvem-se atividades de orçamentação, elaboração de planos de coleção e acompanhamento da produção e das vendas realizadas.

O Controlo Comercial dá maior enfoque a projetos relacionados com o aprovisionamento de produto, desenvolvendo e acompanhando ferramentas de controlo para o apoio a departamentos como os de Gestão de Mercado, Distribuição e Gestão de Produto.

Por sua vez, na área do Desenvolvimento, procede-se à criação de ferramentas de gestão, com um horizonte temporal de médio-longo prazo, que permitem um acompanhamento diário de várias atividades da empresa e, assim, agilizar a análise dos impactos e potenciais problemas.

1.2 Apresentação da empresa

Inspirada em modelos semelhantes já existentes no estrangeiro, Manuela Medeiros decide lançar, em Portugal, uma marca de acessórios de moda para mulheres, algo inexistente até à data. Assim, em 1994, nasce na Rua Santa Catarina, no Porto, a primeira loja *Parfois*.

A boa performance desta loja na baixa portuense fez o reconhecimento da marca aumentar e, aliando-se ao crescimento da rede de centros comerciais, a *Parfois* continuou na senda de crescimento, fazendo com que, no final dos anos 90, a marca já se encontrasse espalhada de Norte a Sul de Portugal, através de lojas próprias e *franchisadas*. Todo este sucesso em Portugal fez com que a marca se tornasse líder de mercado, o que acabou por catapultar a empresa para o processo de internacionalização.

Esse processo foi iniciado em 1999, expandindo horizontes para a vizinha Espanha, novamente com ótimos resultados. Desde então, a *Parfois* difundiu-se pelo mundo, tendo nos mercados emergentes a sua principal aposta, através da introdução de lojas *franchisadas* que implicam um risco menor, mas possibilitam uma avaliação que levará, ou não, a apostar nesse mercado.

A boa relação qualidade/preço, o excelente atendimento ao cliente e a capacidade de introduzir nas suas lojas, semanalmente, uma vasta diversidade de novos produtos, facilitou o processo de abertura de lojas um pouco por todo o Mundo, estando já presente em mais de 50 países, com cerca de 650 lojas próprias, *franchisadas* ou consignadas.

Atualmente, a estratégia da *Parfois* prende-se com a abertura de lojas próprias em mercados maduros ou com potencial de crescimento, entre os quais Portugal, Espanha, França, Polónia, Alemanha e Itália, sendo que a expansão para novos países está sujeita a seleção cuidada quer dos mercados quer dos parceiros nesses mesmo mercados (Canha 2014).

Para além do sucesso nas lojas físicas, a *Parfois* tem aumentado a sua aposta também no mundo *online*, tendo, no final de 2011, lançado a sua loja online. Em 2015, esta “passou a ser a melhor loja da insígnia, a faturar dois milhões de euros e com um crescimento de 112% face a 2014 e ainda assim tem uma margem de crescimento enorme” (Pimentel 2016).

1.2.1 Visão, Missão e Valores

A *Parfois* guia-se pela divisa de “Ser a melhor marca de acessórios de moda, onde quer que a *Parfois* decida operar” e, para isso, tem como missão “Ser a marca de referência das mulheres neo-tradicionais e *trendy* oferecendo uma gama variada e permanentemente renovada de acessórios de moda a preços justos e expetáveis, em espaços convidativos e dinâmicos onde as clientes sejam bem atendidas e gostem de comprar” (Parfois 2015).

A performance na *Parfois* rege-se pelos valores de Rigor, Ambição e Humildade:

- Rigor para ser eficiente desde o primeiro momento, fundamentar e tomar as melhores decisões, ter mais eficácia e cometer menos erros, conquistando o respeito dos outros;
- Ambição para uma melhoria e superação constantes;
- Humildade para reconhecer e aprender com pessoas e empresas, cujo desempenho é, até ao momento, mais eficiente.

1.2.2 Modelo de Negócio

Diretamente integrada no mundo da moda, a *Parfois* decide quais os produtos a expor ao cliente final, regendo-se por duas coleções principais: Primavera-Verão (*Spring-Summer*) e Outono-Inverno (*Fall-Winter*). No entanto, tendo em conta os estudos de mercado, existem ainda dois outros períodos importantes: *Pre-Fall* (Julho e Agosto) e *Resort/Cruise Collection* (Novembro e Dezembro) (Fury 2015).

Com o intuito de responder às necessidades do cliente e chegar atempadamente ao mesmo, o planeamento de cada coleção é iniciado com a análise das tendências de moda para a próxima estação, por parte quer dos *designers* quer de compradoras, de forma a definir quais e quantos produtos integrarão o plano de coleção.

Após a chegada dos artigos da coleção aos armazéns, estes são expedidos para as lojas, consoante as suas necessidades, definidas quer pela sua capacidade de exposição quer pelo seu fluxo de vendas.

Consoante a performance dos produtos ou necessidades demonstradas pelos clientes nas lojas, as equipas de *design* e compras podem desenvolver novos produtos ou repetir produções que entrarão para o mercado ainda no decorrer dessa estação.

Por apresentar um vasto leque de acessórios, a *Parfois* está direcionada às mulheres, dos 16 aos 80 anos, tendo, no entanto, como público-alvo a “mulher com idade entre os 25 e 35 anos, urbana, confiante e com gosto por moda. É uma mulher atual, informada, segura, que se veste para si mesma e que vê a forma de se arranjar e vestir como uma forma de expressão da sua individualidade” (Aparicio 2015).

Uma empresa com uma tão vasta variedade de artigos, por vezes expostos simultaneamente na loja, necessita proporcionar um ambiente de moda coordenado aos clientes, centrando a sua atenção na apresentação das equipas, num bom *visual merchandising*, num atendimento agradável e num ambiente acolhedor de forma a fortalecer a marca e, aliado a uma renovação semanal da montra e do *visual merchandising*, fomentar nos clientes a vontade de repetir, frequentemente, essa experiência.

1.3 Objetivos do projeto

Este projeto tem como principal objetivo a otimização das quantidades de compra dos produtos para posterior venda na loja. Trata-se de um projeto com bastante impacto para a empresa, pois com uma escolha mais acertada das quantidades a adquirir pretende-se não só reduzir os excessos que comportam custos de transporte, armazenamento e destruição do produto, como também evitar situações de falta de produto para venda antes do previamente assumido.

Para alcançar esse objetivo, pretende-se criar uma ferramenta que analise comportamentos passados de produtos com características similares e, com base na mesma, apresentar intervalos de valores que se encontrem ajustados às expectativas de vendas dos novos artigos. Este ajuste visa garantir uma diminuição da quantidade de artigos que são vendidos para além da fase normal da estação, ou seja, reduzir a percentagem de artigos que entra em saldos. A Tabela 1 permite observar o panorama atual da *Parfois* em termos de percentagens de artigos a entrar em saldos.

Tabela 1 - Percentagem atuais de entradas em Saldos da *Parfois*

Gamas	Nº Referências	Venda Média	Compra Média	%Entrada Saldos
Acessórios Cabelo	326	1 113	2 010	45%
Artigos Inverno	51	260	361	28%
Bijuteria	1 690	1 711	3 021	43%
Calçado	99	3 452	5 253	34%
Carteiras	408	3 273	4 415	26%
Carteiras Festa	102	968	1 479	35%
Carteiras Noite	235	1 711	2 193	22%
Chapéus	96	891	1 315	32%
Cintos	66	1 372	1 843	26%
Guarda Chuvas	22	1 617	2 500	35%
Lenços	246	2 249	2 933	23%
Óculos	83	1 402	1 917	27%
Pota-Moedas	317	1 689	2 411	30%
Relógios	113	1 983	2 605	24%
Vestuário	186	1 491	1 773	16%
Viagem	23	1 226	1 436	15%
Total	4 063	1 651	2 342	29%

1.4 Metodologia

A execução do projeto foi dividida em várias etapas, de forma a definir o problema de uma forma mais estruturada. Assim, as principais etapas foram:

- **Iniciação:** estudo da situação atual na empresa, nomeadamente dos departamentos mais diretamente relacionados com o projeto;
- **Análise:** processo que engloba a decisão da gama e períodos que servirão de base e posterior controlo dos resultados obtidos no projeto;
- **Levantamento de dados:** obtenção e tratamento de dados históricos por forma a observar padrões e/ou grupos de características;
- **Desenvolvimento:** compreende a fase de estudo de características *core* de produtos para a ferramenta, o processamento dos dados obtidos, recorrendo a técnicas de *data mining*, com o objetivo de extrair o máximo de informação útil possível, a criação da ferramenta e posterior fase de testes da mesma;
- **Implementação:** apresentação da ferramenta desenvolvida, do método de funcionamento e análise de resultados obtidos aos departamentos envolvidos.

No sentido de facilitar um maior contacto entre todas as partes interessadas na implementação do projeto, e para que houvesse um acompanhamento do mesmo, foram definidas metas quinzenais em que se apresentou a situação do projeto, nesse momento, e os passos seguintes, para além de um *brainstorming*, de modo a garantir-se uma evolução sustentada. No Anexo B é apresentado um cronograma com o planeamento das diferentes etapas do projeto, as durações previstas, assim como a calendarizações das diferentes reuniões.

1.5 Estrutura da dissertação

A presente dissertação, para além do capítulo introdutório, possui mais cinco capítulos.

No capítulo 2 é feita uma contextualização teórica do projeto, sendo exploradas, de forma genérica, as características do setor do retalho da moda e, em particular, do segmento de *fast fashion*. É ainda também apresentada tecnologia existente para tratamento de dados sendo dado maior enfoque à temática de *data mining* ou “prospecção de conhecimento em base de dados”.

No capítulo 3 é explicada a génese da ferramenta, nomeadamente o modelo atual de obtenção de quantidades de compra utilizado pela empresa, assim como a metodologia a ser aplicada e as características a considerar.

No capítulo 4 são expostas as técnicas desenvolvida e modelos utilizados para o cálculo das quantidades de compra.

No capítulo 5 é feita uma avaliação do desempenho do modelo criado, assim como apresentada a ferramenta criada que interliga o utilizador final ao modelo criado, auxiliando-o no processo de definição das quantidades de compra com base nas características pretendidas dos artigos.

Por fim, no capítulo 6, são apresentadas as conclusões tiradas com base nos dados obtidos, além dos principais obstáculos encontrados e feitas considerações sobre eventuais trabalhos que se venham a realizar, no futuro, identificando possíveis melhorias a aplicar à ferramenta.

2 Enquadramento Teórico

Neste capítulo é feita, numa primeira instância, uma contextualização do mercado em que a *Parfois* se encontra, o retalho da moda, apresentando as principais características e problemáticas que o diferenciam dos restantes mercados, assim como particularidades do mesmo.

Posteriormente será abordada a temática da descoberta de conhecimentos em base de dados, identificando-se as principais aplicações e técnicas existentes.

2.1 Retalho de Moda

O aumento galopante da concorrência, vinda de diferentes partes do globo e apresentando, por isso, diferentes valências, tem causado uma enorme alteração no mercado, conduzindo a uma cada vez maior quantidade e variedade de apresentação de produtos disponíveis ao consumidor final. Isto implica uma crescente flexibilidade por parte das empresas, desde o desenvolvimento mais célere de novos produtos à criação de cadeias de abastecimento mais responsivas e dinâmicas (Fisher et al. 1994).

Apesar de apresentar características que o diferenciam, o mercado da moda tem sido moldado, de forma categórica, por estas novas exigências, como referem Lo, Hong, and Jeng (2008), pelo que surgiu a necessidade de dar uma resposta mais rápida aos novos requisitos dos consumidores.

Segundo Christopher, Lowson, and Peck (2004), o retalho de moda pode ser definido de acordo com as seguintes características:

- Ciclos de vida curtos: por haver uma rápida variação de tendências, os produtos são concebidos em resposta a tendências que existam no momento, ou seja, só irão estar de acordo com os interesses dos consumidores durante um curto horizonte temporal;
- Elevada procura por impulso: no mercado da moda, as aquisições são efetuadas, maioritariamente, por impulso, quando confrontados com o produto numa loja, e não por necessidade, pelo que disponibilidade e visibilidade do produto nas lojas é de grande relevância;
- Alta volatilidade: este tipo de produtos têm uma grande instabilidade na procura, pois são usualmente afetados por condições externas, tais como as condições climáticas ou a utilização desses artigos por celebridades;
- Baixa previsibilidade: devido à elevada procura por impulso e à alta volatilidade, é muito difícil obter previsões semanais de procura, de forma satisfatória.

Estas características apresentadas foram catalisadoras para o aumento da exigência ao nível das cadeias de abastecimento, dando origem a um novo conceito de moda - o *fast fashion*.

2.1.1 Fast Fashion

O modelo de negócio *fast fashion*, que veio revolucionar drasticamente a indústria da moda a nível mundial, foi primeiramente introduzido na zona da Galiza, em Espanha, na década de 60.

Anteriormente à implementação deste novo conceito, a moda fluía, maioritariamente, de acordo com as produções de *designers* que, assim, definiam as tendências seguidas pelo público. No entanto, a introdução do *fast fashion* causou uma rutura de paradigmas e os clientes passaram a ser os principais influenciadores para as empresas que se regem por este conceito. Assim, os *designers* passaram a ajustar as suas criações à procura do mercado, visto ser uma forma efetiva de incentivar o consumo, através de constantes novidades apresentadas a um ritmo vertiginoso e de preços competitivos de forma a abrangerem uma maior percentagem de público (Crofton and Dopico 2012).

Sendo um negócio em constante transformação, segundo Bhardwaj and Fairhurst (2010), estas foram as mudanças mais significativas ocorridas desde a década de 90:

- Mudanças estruturais: alteração do foco das empresas do setor, que passou a funcionar de acordo com os gostos dos clientes ao invés do foco anterior no produto o que dava azo à criatividade dos *designers*;
- Estações fixas: a criação de novos produtos era feita para duas estações bastantes estáveis, *Spring-Summer* e *Fall-Winter*, não havendo alterações nas gamas no decorrer das estações. Todavia, a partir da década de 90, as coleções tornaram-se dinâmicas com constantes alterações ao longo das estações por forma a responder cabalmente às necessidades do mercado;
- Desaparecimento da produção em massa: a alteração do foco das empresas, assim como as mudanças de preferências dos consumidores ao longo das estações, impossibilitaram este tipo de produção que era alicerçada em estilos *standard* produzidos em larga escala, permitindo reduções significativas dos custos de produção.

Estas alterações permitem definir o *fast fashion* como um modelo caracterizado pela maior quantidade e variedade de artigos, produzidos em séries mais pequenas e com um período de vida útil mais reduzido, aumentando a rotatividade nas lojas (Caro and Gallien 2010). Esta rotatividade torna cruciais os *timings* de entrada em loja por forma a reduzir a perda de oportunidades e a quantidade de *stock* obsoleto (Figura 1) (Moreira 2011).

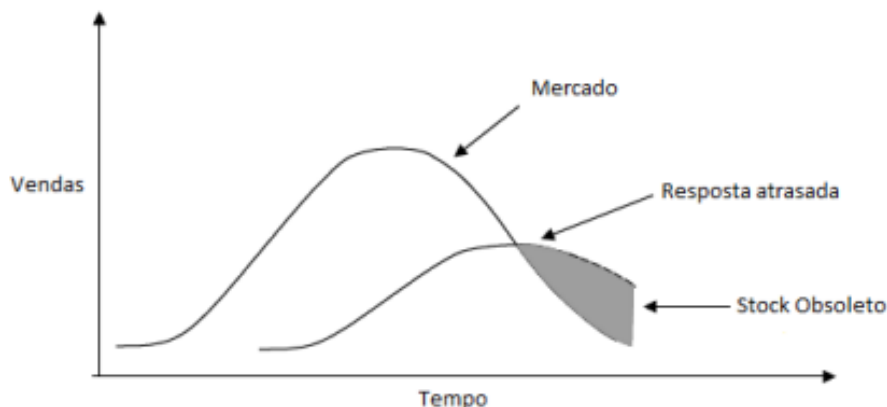


Figura 1 - Vendas e Tempo de Reação ao Mercado (Moreira 2011).

2.1.2 Quick Response

Com a mudança de foco, do produto para o cliente, as empresas passaram a estar bastante mais dependentes dos clientes e das tendências que estes irão seguir. Normalmente, por que essas tendências estão associadas a eventos de moda, é fácil antever quando é necessário iniciar o processo de definição de novos produtos para possibilitar uma maior margem temporal aos parceiros a montante.

No entanto, e como são várias as empresas a coabitarem no mesmo mercado, ocorre, por vezes, a necessidade de ajustar as coleções inicialmente definidas para incluírem novos produtos que colmatem uma carência causada pela concorrência, num certo tipo de artigo ou com determinada característica.

Para combater estes casos de forma mais eficiente, é necessário aproximar ao máximo os *timings* de compra do artigo ao fornecedor e o momento de compra por parte do consumidor final. Este método de resposta rápida e de produção em várias janelas temporais durante a coleção é conhecido como *Quick Response* (QR) (Patil, Avittathur, and Shah 2010).

Segundo Birtwistle, Siddiqui, and Fiorito (2003), este método tem como principais benefícios:

- Redução do risco de previsão errónea de necessidades quer ao nível de variedade de artigos, quer da quantidade dos mesmos;
- Redução de *stocks* excessivos, uma vez que as compras são efetuadas em resposta a necessidades apresentadas no mercado;
- Possibilidade de praticar preços mais acessíveis devido aos ganhos de eficiência em toda a cadeia como, por exemplo, a agilização da entrada em loja de novos artigos em resposta às necessidades observadas no mercado.

Apesar de apresentar bastantes melhorias em vários momentos da cadeia de abastecimento, sendo uma muito boa solução para produtos que surjam por necessidades já fora dos *timings* usuais de compra aos fornecedores, este método vai de certa forma encarecer nestes casos os custos de produção por levarem à perda das economias de escala, sendo, todavia, um custo de oportunidade que é facilmente compensado.

2.2 Descoberta de Conhecimento em Bases de Dados

Num mundo em constante expansão e em que os mercados estão cada vez mais saturados, a obtenção de novos conhecimentos é cada vez mais importante, como afirma Monte et al. (2013) quando refere que os dados que as empresas recolhem sobre os consumidores representam um dos seus mais poderosos ativos. A recolha de dados só é possível graças aos constantes avanços no mundo das Tecnologias de Informação que contribuíram para a obtenção, armazenamento e análise de grandes volumes de dados.

Contudo, apesar de ser benéfica a obtenção da informação referente aos consumidores, Monte et al. (2013) reitera que as empresas tendem a acumular quantidades massivas de dados nas suas bases de dados, entre dados úteis e desnecessários, necessitando posteriormente de passar por um processo de transformação complexo, lento e difícil.

Para que o tratamento da quantidade de informação disponível seja exequível, em geral as empresas recorrem ao processo de descoberta de conhecimento em base de dados (*Knowledge Discovery in Databases*). Este processo utiliza ferramentas de várias áreas, tais como visualização de dados, estatística, reconhecimento de padrões e inteligência artificial, que

auxiliam no processo de previsão de tendências e comportamento futuros, bem como no processo de identificação de padrões existentes nos dados (Monte et al. 2013).

Este processo de descoberta de conhecimento em base de dados é definido por Chen and Zhu (1998), como um “processo não trivial de identificação de informação válida, nova, potencialmente útil e com padrões compreensíveis”, sendo efetuado de forma interativa ao longo de várias etapas (Figura 2).

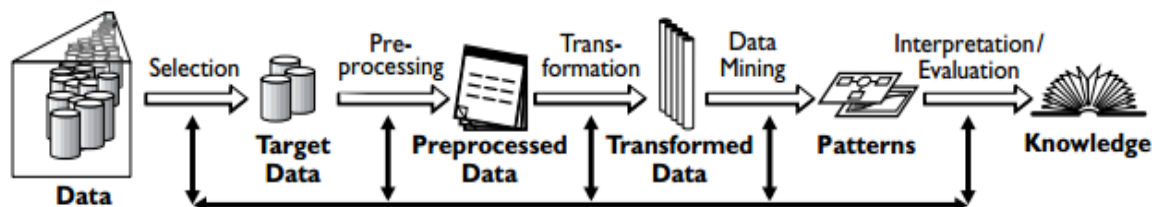


Figura 2 - Etapas do Processo de Descoberta de Conhecimento em Bases de Dados (Fayyad, Piatetsky-Shapiro, and Smyth 1996)

O processo representado na Figura 2 pode ser traduzido nas seguintes fases:

1. Contextualização do processo no seio da empresa, assim como a recolha de conhecimentos prévios que auxiliem na sua execução e identificação do seu objetivo final;
2. Criação da base de dados a estudar, através da definição dos dados a utilizar, variáveis a considerar, atributos mais relevantes e período temporal de execução;
3. Pré-processamento da base de dados de trabalho, através da remoção de ruído ou *outliers* presentes, além da obtenção de informação que permita modelar ou reconhecer ruído e definição de estratégias para tratamento de dados em falta;
4. Transformação dos dados, recorrendo à normalização, agregação ou discretização, de forma a poderem ser aplicados a ferramentas de *data mining*;
5. Definição da ferramenta de *data mining* a utilizar, tais como regressão, classificação ou *clustering*, para a descoberta de padrões nos dados trabalhados;
6. Interpretação dos padrões obtidos e avaliação da sua preponderância para o objetivo final.

2.2.1 Data Mining

Data Mining corresponde, de acordo com Linoff and Berry (2011), ao processo de exploração e análise de grandes quantidades de dados, com o intuito de encontrar regras e padrões que se provem importantes na resolução do problema. Fayyad, Piatetsky-Shapiro, and Smyth (1996) classificam o *data mining* como a etapa *core* do processo de descoberta de conhecimento em base de dados, pelo que lhe é dada uma maior atenção do que a qualquer outra das restantes etapas. A sua vasta utilidade e aplicabilidade conduzem à sua empregabilidade nas mais variadas áreas, tais como *marketing*, saúde ou setor financeiro (Brachman et al. 1996).

Apesar de ser uma das etapas do processo de descoberta de conhecimento em base de dados que requer um maior conhecimento prévio em termos de algoritmos a serem utilizados na análise, *data mining* é uma das etapas que consome menos tempo ao longo de todo o processo, representando, apenas, cerca de 10% (Cabena et al. 1998).

Segundo Ngai, Xiu, and Chau (2009), as principais técnicas de *data mining* podem ser classificadas da seguinte forma:

- Associação: técnicas que permitem estabelecer relações entre elementos existentes nas bases de dados;
- Classificação: técnicas que permitem categorizar os dados existentes em classes predefinidas;
- Clustering: técnicas que permitem agrupar um grupo heterogêneo de dados em *clusters* mais homogêneos. O processo de *Clustering* é análogo ao que acontece na classificação, sendo a principal diferença o facto de no caso do *Clustering* as classes serem determinadas a partir dos dados;
- Previsão: técnicas que permitem fazer estimativas de valores futuros de um certo atributo, com base nos seus padrões históricos;
- Regressão: técnica que, a partir da inferência estatística da relação entre duas ou mais variáveis, permite obter uma previsão para uma dada variável;
- Descoberta de Sequências: técnicas que permitem identificar relações entre os dados ao longo do período temporal considerado;
- Visualização: técnicas que permitem representar os dados para que os utilizadores possam observar padrões mais complexos.

Ao processo de *data mining* não está exclusivamente associada uma das técnicas apresentadas, podendo, por vezes, ser necessária a combinação de várias, com o objetivo de se extrair o maior conhecimento possível dos dados.

2.2.1 Random Forest

A técnica de *random forests* consiste na evolução de uma metodologia amplamente utilizada e reconhecida, as árvores de decisão, nomeadamente o algoritmo *CART* (*Classification and Regression Trees*) apresentado por Breiman et al. (1984).

As árvores de decisão, introduzidas por Morgan and Sonquist (1963), são ferramentas de classificação baseadas no conceito de dividir para conquistar. Partindo de um conjunto inicial de dados, os passos seguintes de um modelo de árvores de decisão consistem na sucessiva subdivisão do problema inicial em vários sub-problemas, sendo, a cada iteração, utilizado um subconjunto dos dados de dimensões. As subdivisões ocorridas durante o processo advêm de questões, como exemplificado na Figura 3, cujas respostas permitem a divisão em dois ou mais subconjuntos (Pang-Ning, Steinbach, and Kumar 2006).

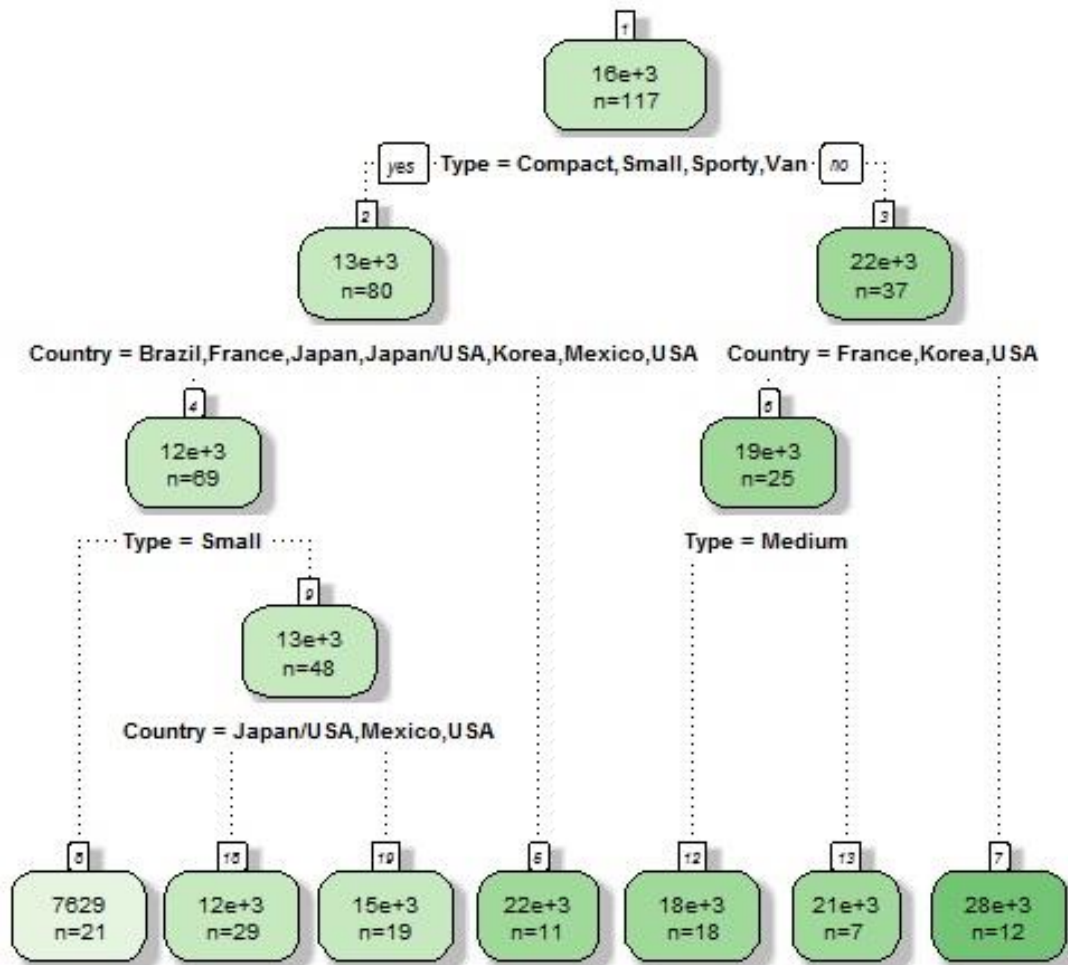


Figura 3 – Exemplo do funcionamento de uma árvore de decisão

Desenvolvida por Breiman (2001), a técnica das *random forests* pode ser interpretada como um método que aglomera árvores de decisão, o que permite melhorar a precisão das mesmas, contornando problemas como a alta sensibilidade a pequenas variações nos dados, o que resulta em inúmeras possibilidades de subdivisões dos conjuntos a cada iteração. (Friedman, Hastie, and Tibshirani 2001).

Podendo ser utilizadas como técnica de classificação ou regressão, as *random forests* são contruídas com base em diferentes metodologias, caracterizadas por Breiman (2001) do seguinte modo:

- *Bagging* – criação de um conjunto de subamostras, de forma aleatória, a partir do conjunto inicial de dados, selecionadas com reposição;
- *Boosting* – a partir de cada subamostra gerada, é construída uma árvore de decisão, explorando todas as possibilidades, ou seja, até ao ponto em que cada nó contenha apenas conjuntos de dados homogêneos;
- *Randomizing* – utilização de um subconjunto de variáveis, para cada árvore de decisão.

Conforme mencionam Maragoudakis and Loukis (2012), o funcionamento do algoritmo de *random forest* pode ser descrito através dos seguintes passos:

- Assumindo N como o número de árvores de decisão que se pretende criar, são escolhidos N subconjuntos aleatórios dos dados originais, com reposição (Figura 4);
- Para cada subconjunto N , dois terços dos dados são selecionados para treino sendo o restante deixado para teste da árvore de decisão criada (Breiman 2001);
- Definindo M como o número de características de *input*, é especificado um m , fixo para todo o modelo de *random forest*, para que, em cada nó da árvore de decisão sejam escolhidas m características aleatórias de M . De seguida, prossegue-se o desenvolvimento da árvore com a melhor divisão criada por uma dessas m características;
- Definidos os N subconjuntos e as m características a considerar, as árvores de decisão são expandidas até à máxima extensão possível.

Segundo Geurts, Ernst, and Wehenkel (2006), o número N de árvores a construir deve ser definido tão mais elevado quanto possível, pois permitirá obter modelos com elevada precisão. Assim, a quantidade de árvores é apenas restringida pela capacidade computacional da máquina, devendo-se optar por um valor que origine a melhor relação entre a precisão do modelo e capacidade do computador.

Através da especificação de N com um valor elevado evitam-se os problemas de sobreajustamento do modelo (*overfitting*) ao conjunto de dados. Segundo os fundamentos da Lei dos Grandes Números, nestas circunstâncias há sempre uma convergência do modelo, excluindo, então, a possibilidade de problemas de *overfitting* (Breiman 2001).

Quanto ao número de características a considerar, m , em cada nó das árvores de decisão, de acordo com Hastie, Tibshirani, and Friedman (2009) este deve variar caso o modelo seja utilizado para classificação ou para regressão. Sendo M o número de característica total, a quantidade m é o valor inteiro obtido por:

- Classificação: $m = \sqrt{M} \wedge m \geq 1$;
- Regressão: $m = M$.

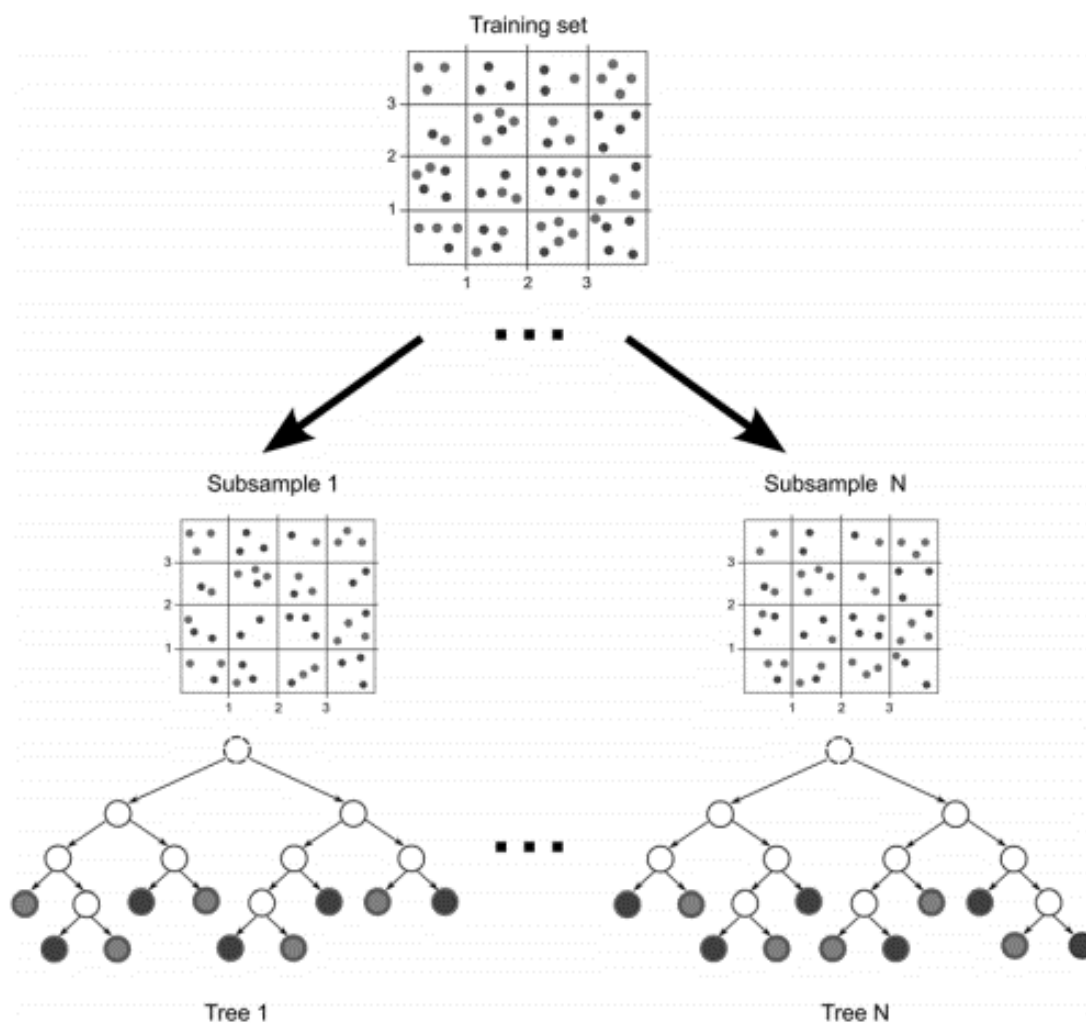


Figura 4 - Processo de treino usando random forests (Machado, Mendoza, and Corbellini 2015)

Após concluído o processo é determinado, para cada árvore criada e recorrendo à terça parte dos dados reservados para teste nessa árvore, um erro, denominado de erro *out-of-bag*. Esta medida de erro refere-se ao erro médio quadrado, que relaciona, para os dados de teste, o valor estimado pelo modelo para a variável dependente com o valor efetivamente observado. Com isto, deixa de haver a necessidade de um teste separado para obtenção de uma estimativa imparcial do erro de teste, servindo o erro *out-of-bag* como um bom estimador da precisão do modelo. (Breiman 1996).

Quando o modelo de *random forest* criado é utilizado para prever a variável dependente de um novo conjunto de dados, o processo decorre como o apresentado na Figura 5 em que cada novo dado é corrido pelas N árvores criadas, sendo o resultado obtido por agregação dos valores obtidos em cada árvore de decisão.

Para problemas de classificação, a agregação reflete a característica mais votada no nó terminal de cada árvore e, para problemas de regressão, a média da previsão obtida nesses mesmos nós (Liw and Wiener 2002).

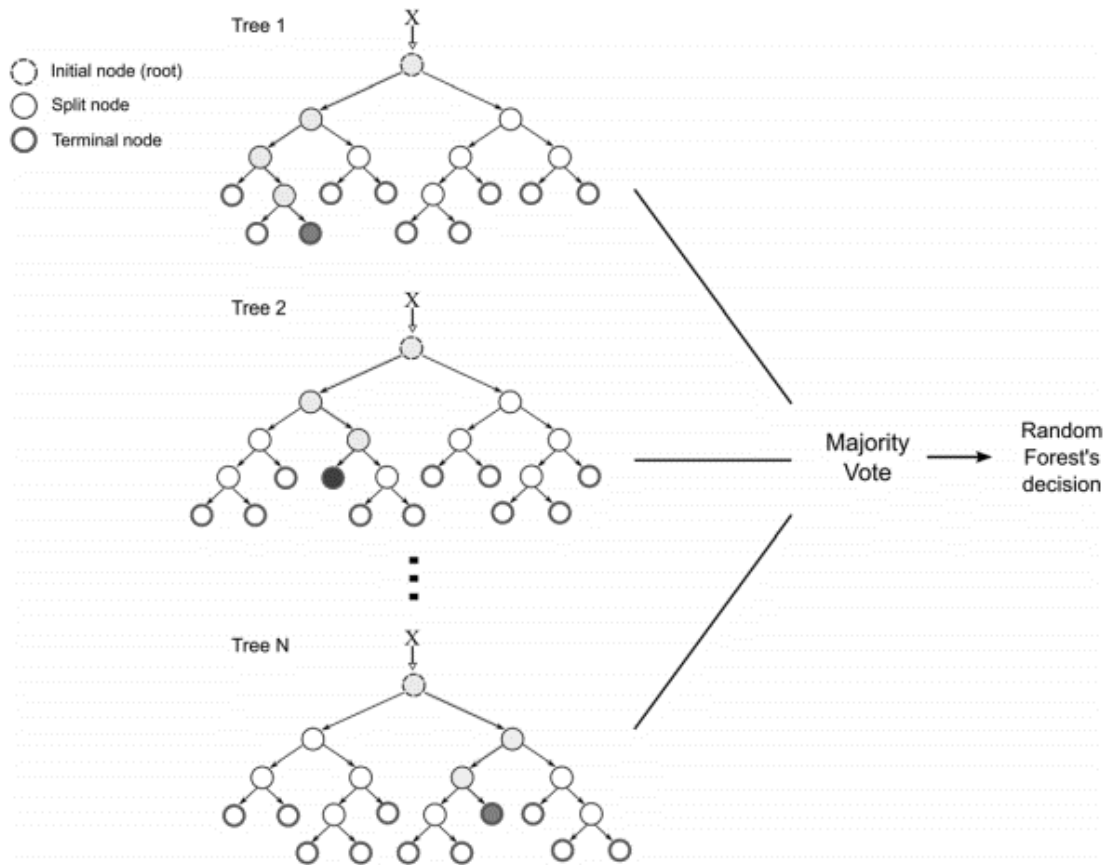


Figura 5 – Processo de regressão utilizando random forests (Machado, Mendoza, and Corbellini 2015)

3 Modelo atual de obtenção de Quantidades de Compra

Neste capítulo é descrito como se processa, neste momento, na *Parfois*, a determinação das quantidades de compra dos produtos que integrarão os futuros planos de coleção.

Visto tratar-se de um processo influenciado por uma grande quantidade de características específicas, identificar-se-ão, neste capítulo, as que apresentam maior relevância para o valor final das quantidades de compra dos artigos.

3.1 Os produtos Parfois

Sendo uma empresa integrada no mundo da indústria da moda, a *Parfois* aposta, de forma bastante intensa, na variedade de produtos que apresenta nas suas lojas, tendo por objetivo dispor, semanalmente, novos produtos, de modo a garantir que mesmo o consumidor mais frequente seja brindado com novidades a cada nova visita.

Para garantir que haja sempre novas entradas de artigos em loja, a *Parfois* introduz, a cada nova estação, em média 4000 novas referências, repartidas por uma grande variedade de gamas de produtos.

Atualmente, os produtos da *Parfois* são divididos em cinco grandes departamentos: acessórios têxteis, acessórios não-têxteis, bijuteria, carteiras e calçado. Posteriormente, estes departamentos são desagregados em várias gamas de produtos, nomeadamente:

Acessórios Têxteis	<ul style="list-style-type: none"> • Artigos de Inverno; • Lenços; • Vestuário; 	Bijuteria	<ul style="list-style-type: none"> • Bijuteria; • Acessórios de Cabelo;
Acessórios Não-Têxteis	<ul style="list-style-type: none"> • Chapéus; • Cintos; • Guarda-chuvas; • Óculos; • Relógios; 	Carteiras	<ul style="list-style-type: none"> • Carteiras; • Carteiras de Festa • Carteiras de Noite; • Porta-moedas; • Viagem;
		Calçado	<ul style="list-style-type: none"> • Calçado;

As gamas referidas encontram-se ilustradas na Figura 6.



Figura 6 - Ilustração de produtos das diferentes gamas da Parfois

Desde a sua criação, a *Parfois* destacou-se, maioritariamente, na gama de Bijuteria, sendo que esta sempre representou uma grande porção das vendas em valor e é a que apresenta uma maior variedade de referências, com cerca de 1500 em cada estação. No entanto, a partir do momento em que a *Parfois* iniciou a sua aposta na gama das Carteiras, esta tomou a liderança em termos de valor de vendas.

Nas Figura 7 e Figura 8, é apresentada a distribuição de vendas das seis principais gamas, em quantidade e valor, no ano de 2015, sendo as restantes aglomeradas na classe “Outros”.



Figura 7 - Distribuição das Vendas por Gama, em Quantidade em 2015

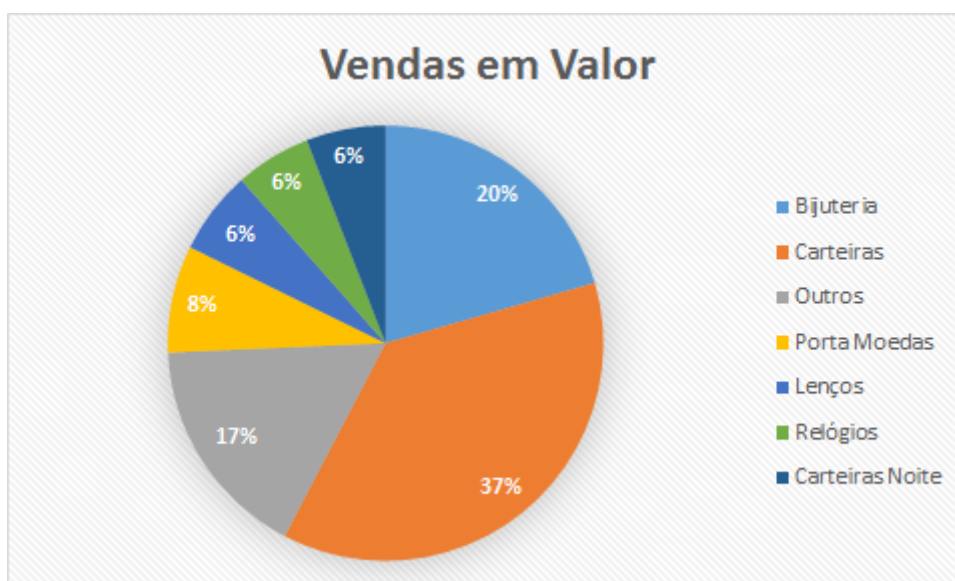


Figura 8 - Distribuição das Vendas por Gama, em Valor em 2015

Pelas figuras, observa-se a importância das gamas de Bijuteria e Carteiras para a *Parfois*, sendo que o seu conjunto representa, em média, cerca de 60% do volume de negócio, em termos de quantidade e de venda. Isto leva a que, em todos os processos da empresa, seja dedicada especial atenção a estas gamas.

O presente projeto de otimização das quantidades de compra baseia-se na gama de Carteiras, por ser a que apresenta, neste momento, maior retorno económico para a empresa. Todavia, apesar do foco nesta gama específica, o trabalho efetuado é realizado de modo a facilitar uma posterior aplicação nas restantes gamas.

Como referido anteriormente, a cada nova estação, a *Parfois* introduz no mercado cerca de 4000 novas referências. Todas estas referências têm características que as diferenciam, desde as mais generalistas que abarcam grupos mais alargados e heterogéneos de referências a outros mais específicos, com menos artigos.

Na *Parfois*, estas características descritivas são agrupadas segundo a estrutura mercadológica apresentada na Figura 9.

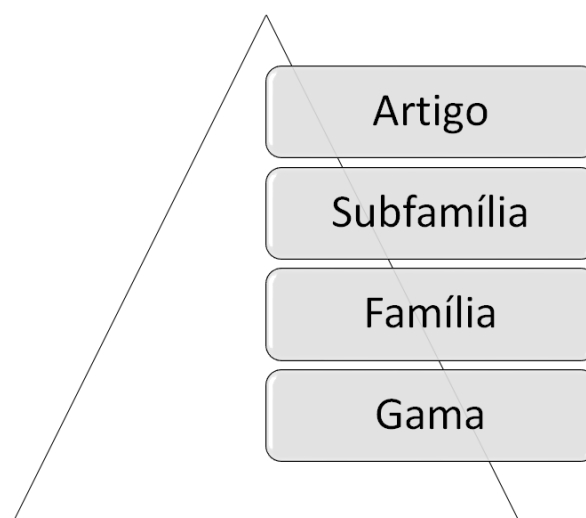


Figura 9 – Estrutura mercadológica dos artigos da Parfois

Para além destas características principais dos diferentes artigos, existem diferenciações ao nível dos seus atributos, destacando-se, para a gama de Carteiras, as seguintes:

- *Fashion Type*: característica subjetiva que enquadra os artigos numa mesma estratégia conjunta, afetando o seu tipo de exposição em loja e ao tempo de exposição na mesma;
- Tipo de Cor: associado aos padrões de cores do artigo, podendo ser de cor única caso haja uma que seja visivelmente predominante, ou multicolor, caso seja constituído por várias cores;
- Cor: característica que faz alusão à cor predominante do artigo;
- Mês: relacionado com o mês expectável de entrada do produto em loja, o que afetará a performance do produto;
- Preço Base: atributo que define o preço base a que se pretende que o produto seja vendido, em euros, durante o período normal da estação.

Existe ainda uma categorização dos produtos por nível de aposta, que pode variar entre alta, normal, média ou baixa. Esta categorização é definida pelo Departamento de Compras aquando da decisão de aquisição dos produtos e está associado à previsão de vendas do artigo e à sua venda média por loja. Assim sendo, um produto que se preveja que tenha uma venda média por loja elevada, por seguir uma tendência de mercado, será considerado uma aposta alta, o que levará a que a encomenda ao fornecedor seja em maior quantidade.

Tratando-se de uma avaliação qualitativa, há sempre possibilidade de o produto não se comportar como inicialmente previsto, resultando em excesso de *stock*, caso o produto tenha uma prestação inferior ao expectável, ou falta de *stock* (*stockout*) caso a venda do produto seja superior ao previsto.

Com o objetivo de tentar minorar as más avaliações da aposta dos produtos, foram definidos limites de quantidades de artigos para cada nível de aposta, em cada gama.

Na Figura 10 é apresentada uma aproximação da distribuição dos artigos consoante os níveis de aposta. Esta distribuição é obtida com base na quantidade de venda média da gama a considerar e o seu desvio padrão, estando os níveis de aposta balizados por incrementos ou decrementos de um percentual do desvio padrão face à média de vendas da gama.

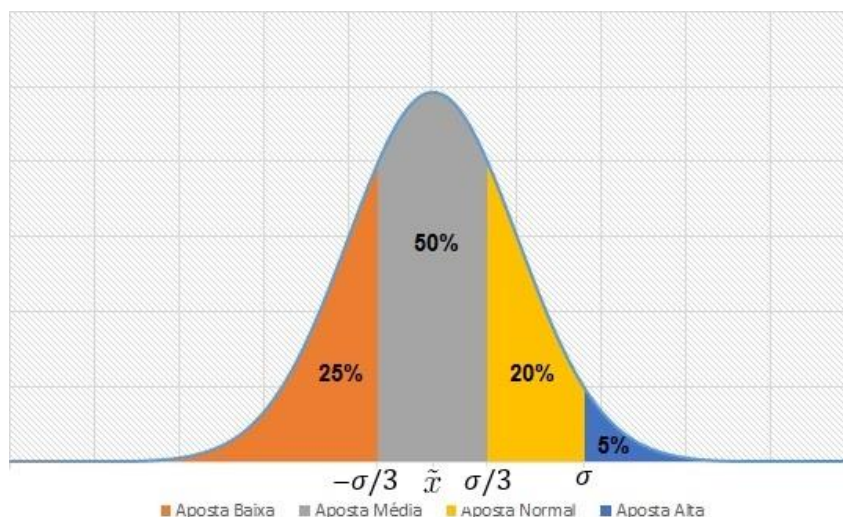


Figura 10 - Distribuição de artigos pelos níveis de aposta em cada gama

Tabela 2 - Relação entre tamanho das lojas e potencial de venda

Tamanho \ Potencial	1	2	3	4	5
Grande	A	A	A	B	C
Normal	A	A	B	B	C
Média	A	A	B	C	C
Pequena	B	B	C	C	D

Em resumo, uma loja Grande e com Potencial de Venda 1 será considerada como uma loja de tipologia A, o que faz com que receba a totalidade dos artigos de uma coleção, enquanto uma loja Pequena e com Potencial de Venda 5 irá receber uma menor variedade de artigos da coleção por ser uma loja de tipologia D.

Esta classificação por tipologias é utilizada como auxiliar na determinação das quantidades mínimas de cada artigo a serem compradas, sendo esse valor obtido pela agregação das quantidades a enviar para as lojas associadas a uma certa tipologia.

Como a decisão de compra dos artigos se baseia na tipologia de lojas em que será exposto, foram definidas aglomerações de tipologias para auxiliar o cálculo das quantidades de compra. Por exemplo, lojas definidas como tipologia D numa dada gama são lojas cujas expectativas de venda dessa gama são baixas e, por isso, não existem artigos que sejam comprados exclusivamente para esta tipologia. Isto significa que as lojas com tipologia D receberão apenas produtos que tenham sido comprados para abastecer todo o universo de lojas Parfois, fazendo com que a característica “Tipologia” desses artigos seja definida como DCBA.

Por outro lado, lojas classificadas como A podem, sem problemas, receber artigos específicos por terem capacidade para os escoar, ou seja, a característica “Tipologia” será definida apenas como A.

A Tabela 3 resume a relação entre a tipologia de loja e o aglomerado de tipologias em que se insere.

Tabela 3 - Relação entre tipologia de loja e aglomerado em que se insere

Tipologia	Aglomerado
A	A
B	BA
C	CBA
D	DCBA

3.3 Modelo atual de obtenção de quantidades de compra

Neste momento, para o cálculo das quantidades de compra são fulcrais os dois conceitos apresentados anteriormente: classificação das lojas por tipologia e distribuição dos artigos por níveis de aposta.

Numa fase inicial, é analisada a prestação dos produtos da gama na estação homóloga do ano anterior. É estudada, numa primeira instância, a velocidade de venda semanal de cada artigo, obtida através da equação 3.1, para cada combinação de tipologia de loja e nível de aposta de produto.

$$SLS\ Speed\ LY_{IJ} = SLS\ LY_{IJ} / (\#lojas_{ly_{IJ}} * \#week_{ly_{IJ}}) \quad (3.1)$$

Onde:

SLS Speed LY é a velocidade de venda de um artigo na estação homóloga do ano anterior
 SLS LY é o valor da venda de um artigo na estação homóloga do ano anterior
 I é a tipologia de loja
 J é o nível de aposta do produto
 #lojas_ly é o número de lojas em que o artigo esteve presente nessa estação
 #week_ly é o número de semanas em que o produto esteve em loja, durante o período normal da estação

Após determinado este valor para a estação do ano anterior e para o número de lojas em que esteve exposto, é calculada a venda semanal para a estação que se pretende analisar, ou seja, é feito o produto do valor obtido pela equação 3.1 pelo novo de número de lojas de cada tipologia em que se prevê que haja exposição do artigo até ao final da estação em análise. Assim, obtém-se, como se constata na equação 3.2 a venda semanal para a nova estação.

$$SLS\ week_{IJ} = SLS\ Speed\ LY_{IJ} * \#lojas_{IJ} \quad (3.2)$$

Onde:

SLS week é a venda semanal prevista de um artigo para a estação em estudo
 SLS Speed LY é a velocidade de venda na estação homóloga do ano anterior
 I é a tipologia de loja
 J é o nível de aposta do produto
 #lojas é o número de lojas em que se prevê que o artigo vai estar exposto

Para além do crescimento em termos de lojas, é também tido em consideração um desenvolvimento da performance de venda, dado pela equação 3.3. Esta evolução está associada ao crescimento da marca e à melhoria contínua dos restantes processos da empresa.

$$SLS\ INC_{IJ} = SLS\ week_{IJ} * (1 + \%inc) \quad (3.3)$$

Onde:

SLS INC é a velocidade de venda incrementada pelo crescimento expectado da empresa
 SLS week é a venda semanal prevista de um artigo para a estação em estudo
 I é a tipologia de loja
 J é o nível de aposta do produto
 %inc é o incremento expectado de performance da empresa

Por fim, e com base num *input* do número de semanas em que se espera que um dado produto, com um dado nível de aposta e que vai estar presente em lojas de uma certa tipologia, esteja exposto, é calculado o valor das vendas totais de cada referência durante o período normal, ou seja, excluindo o período de saldos (ver equação 3.4).

$$SLS_{IJ} = SLS\ INC_{IJ} * \#week_{IJ} \quad (3.4)$$

Onde:

SLS é o valor de venda de um artigo na estação em estudo
 SLS INC é a velocidade de venda incrementada pelo crescimento expectado da empresa
 I é a tipologia de loja
 J é o nível de aposta do produto
 #week é o número de semanas em que o produto vai estar em loja, durante o período normal da estação

O valor resultante da equação 3.4 apenas reflete as vendas que irão ocorrer durante o período normal da estação. Todavia, as estações são constituídas pelo período normal de vendas e por uma época de saldos em que são aplicados descontos sobre o preço base dos artigos.

Para esse período de saldos, a quantidade disponível para ser vendida é determinada com base numa percentagem de entrada em saldos definida como objetivo, previamente estabelecida de acordo com alguns critérios, tais como a tipologia da loja, nível de aposta do artigo e mês de entrada do mesmo em loja. Concretizando, com base neste valor pré-definido de entrada em saldos, a quantidade de compra é determinada pela equação 3.5.

$$\begin{aligned} PUR_{IJ} &= SLS_{IJ} + SLS_{ES_{IJ}} \\ &= SLS_{IJ} + (PUR_{IJ} * \%ES) \\ &= SLS_{IJ}/(1 - \%ES) \end{aligned} \quad (3.5)$$

Onde:

PUR é o valor de compra do artigo
 SLS é o valor de venda de um artigo na estação em estudo, durante o período normal
 SLS ES é o valor de venda de um artigo na estação em estudo, durante o período de saldos
 I é a tipologia de loja
 J é o nível de aposta do produto
 %ES é a percentagem de entrada em saldos definida como objetivo

Como no retalho existe sempre a possibilidade de extravio de produtos ou de os mesmos se danificarem, é ainda considerado um acréscimo na compra, afetado pelas eventuais quebras, de cerca de 2% face ao valor determinado pela equação 3.5.

Em suma, o valor de compra de um dado artigo é obtido através da equação 3.6.

$$\begin{aligned} PUR_{IJ} &= SLS LY_{IJ} * (\#lojas_{IJ}/\#lojas_{ly_{IJ}}) * (\#week_{IJ}/\#week_{ly_{IJ}}) * (1 + \%inc) * \\ &\quad * (1/(1 - \%ES)) * (1/(1 - \%Quebra)) \end{aligned} \quad (3.6)$$

Onde:

PUR é o valor de compra de um artigo na estação em estudo
 SLS LY é o valor da venda de um artigo na estação homóloga do ano anterior
 I é a tipologia de loja
 J é o nível de aposta do produto
 #lojas_ly é o número de lojas em que o artigo este presente nessa estação
 #lojas é o número de lojas em que se prevê que o artigo vai estar exposto
 #week_ly é o número de semanas em que o produto esteve em loja, durante o período normal da estação
 #week é o número de semanas em que o produto vai estar em loja, durante o período normal da estação
 %inc é o incremento esperado de performance da empresa
 %ES é a percentagem de entrada em saldos definida como objetivo
 %Quebra é a percentagem que se considera que haverá de quebras

4 Modelo proposto de obtenção de Quantidades de Compra

Como referido anteriormente, o objetivo do presente projeto passa pela otimização do processo utilizado na *Parfois*, para a obtenção de quantidades de compra, com o intuito de minorar situações de excesso de *stock* ou de *stock* insuficiente.

Neste capítulo são apresentados os dados usados assim como a metodologia utilizada com vista ao cumprimento do objetivo.

4.1 Dados recolhidos

Tal como exposto na secção 3.1, este projeto tem como foco os artigos da *Parfois* incluídos na gama de Carteiras, uma gama que contém cerca de 400 novas referências em cada estação. Por se pretender que a ferramenta sirva de apoio, numa fase inicial, à determinação das quantidades de compra para a estação Primavera-Verão 2017, foi utilizada para desenvolver o modelo a informação referente à estação Primavera-Verão de 2015. Usou-se esta informação devido à quantidade de dados associados e por ser a última estação homóloga completa.

Apesar de serem integrados no plano de coleção cerca de 400 novos artigos da gama de Carteiras em cada estação, nem sempre é conseguido que todos os artigos tenham um ritmo de vendas suficiente para escoar a quase totalidade da quantidade comprada. Este facto pode dever-se a múltiplos fatores como, por exemplo, entrada tardia no mercado ou baixo interesse dos consumidores no artigo.

Verificando-se então a situação de, no final do período normal da estação, ainda haver uma grande quantidade de *stock* referente a um dado artigo, este é reclassificado, ou seja, irá ter na estação seguinte uma nova oportunidade para escoar o seu *stock*.

Após excluídos os casos das carteiras da estação Primavera-Verão 2015 que foram reclassificados, restaram 384 referências para análise. Para esses artigos foram extraídas as características inerentes a esses produtos (ver Tabela 4), assim como os dados referentes às vendas e *stock* dos artigos nas lojas, durante o período normal da estação, ou seja, entre o início de novembro de 2015 e o final de junho de 2016.

Tabela 4 – Representação da tabela com as características dos artigos considerados

Artigo	Gama	Familia	SubFamilia	Tipo Cor	Cor	Fashion	Tipologia	Preço Base
Artigo 1	Carteiras	PVC Fantasia	Shopper	Cor Unica	Dourado	Moda	CBA	29,99
Artigo 2	Carteiras	Pastas	Falsa	Cor Unica	Camel	Básico Moda	DCBA	32,99
Artigo 3	Carteiras	PVC Fantasia	Tracar	Cor Unica	Cru	Básico Moda	BA	22,99
Artigo 4	Carteiras	PVC Estampado	Tracar	Multicolor	Bege	Moda	DCBA	22,99
...

4.2 Stockouts e Vendas Potenciais

Uma venda resulta da necessidade vivenciada por um cliente ou do facto de encontrar algum artigo do seu agrado, aquando da visita a uma loja. Em ambos os casos, essa venda está inteiramente dependente da existência do artigo nas lojas.

Quando ocorrem ruturas de *stock* (*stockout*), há um período de vendas perdidas, desde que o *stock* se esgotou, até que o mesmo é repostado, assumindo que ainda é suposto haver reposição. Assim sendo, é preciso analisar estes casos e estimar essas vendas potenciais que foram perdidas.

Para que esta análise se possa efetuar, e visto que não se consegue avaliar se um *stockout* ocorreu próximo ou longe da hora de fecho da loja, uma vez que apenas há acesso a dados diários agregados, foi necessário definir quais os casos a classificar como *stockout*, tomando as seguintes considerações para a relação artigo-loja:

- *Stockout* – casos em que não ocorram vendas e que o *stock* seja igual a zero;
- Não *stockout* – casos em que ocorram vendas, independentemente do valor de *stock*, ou casos em que este último seja positivo, mesmo que não haja vendas.

Definidos os casos de *stockout* para todos os casos loja-artigo-dia, a fase seguinte diz respeito à determinação da venda potencial que não se efetuou. Para isso, é tomado em consideração o comportamento dos artigos, para cada uma das lojas e em cada uma das semanas em que esteve exposto, durante o período normal da estação em estudo. No entanto, devido à localização bastante dispersa das lojas *Parfois*, os mesmos artigos não chegam a todas as lojas no mesmo momento, pelo que apresentam comportamentos desfasados quando avaliados com base nas semanas do ano, como se pode ver na Figura 12. Assim sendo, para que seja possível fazer uma análise coerente da prestação dos artigos, procedeu-se a um ajuste dessas semanas, passando-se a considerar, para cada par artigo-loja, que a semana 1 de vida do produto consiste na primeira semana com vendas do artigo. Isto significa que (ver Figura 12), na loja 1 a semana w50 – 2014 será a semana 1 de vida do produto e, para a loja 2, a mesma semana 1 será a semana w05 – 2015.

Este ajuste encontra-se representado na Figura 13.

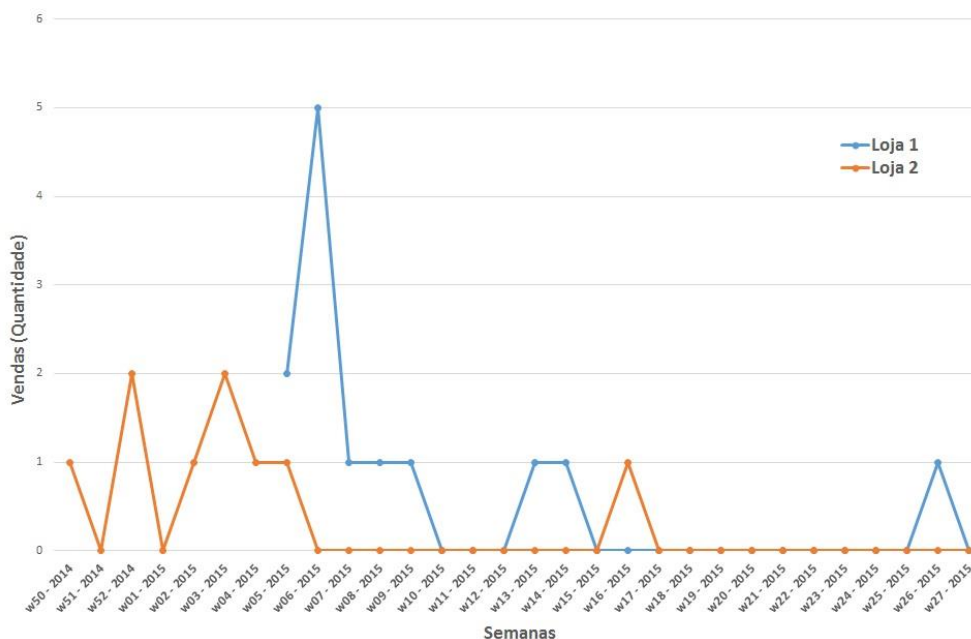


Figura 12 – Prestação de um artigo em duas lojas, com base nas semanas do ano

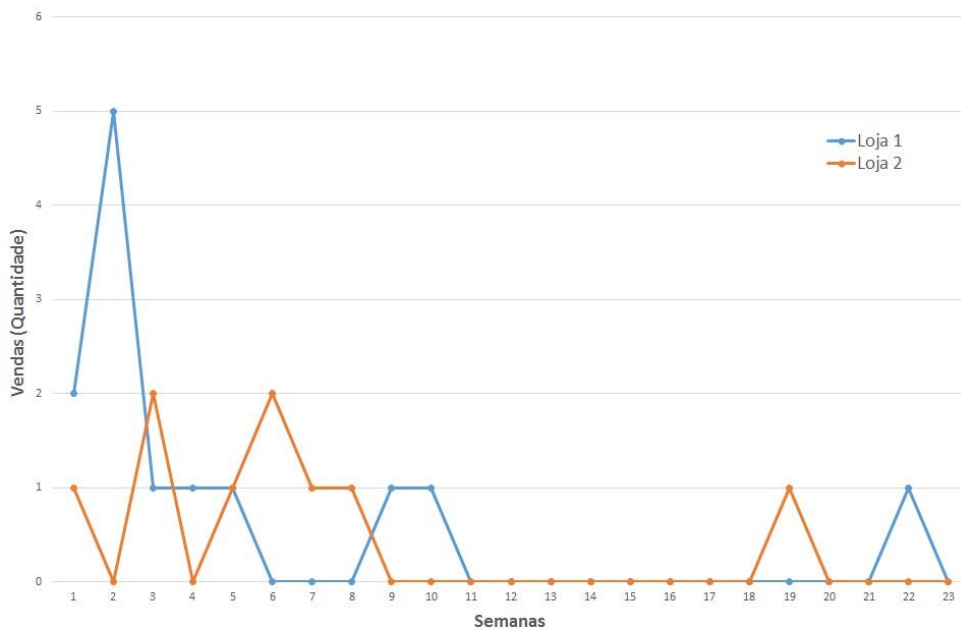


Figura 13 – Prestação de um artigo em duas lojas, com base nas semanas de vida do produto

Alinhadas as semanas dos artigos nas lojas, o passo seguinte consiste em perceber como efetivamente se desenrolaram as vendas dos artigos.

Na Figura 14 é feita uma apresentação do ciclo de vida de um artigo que é suposto estar exposto numa certa loja durante 6 semanas.

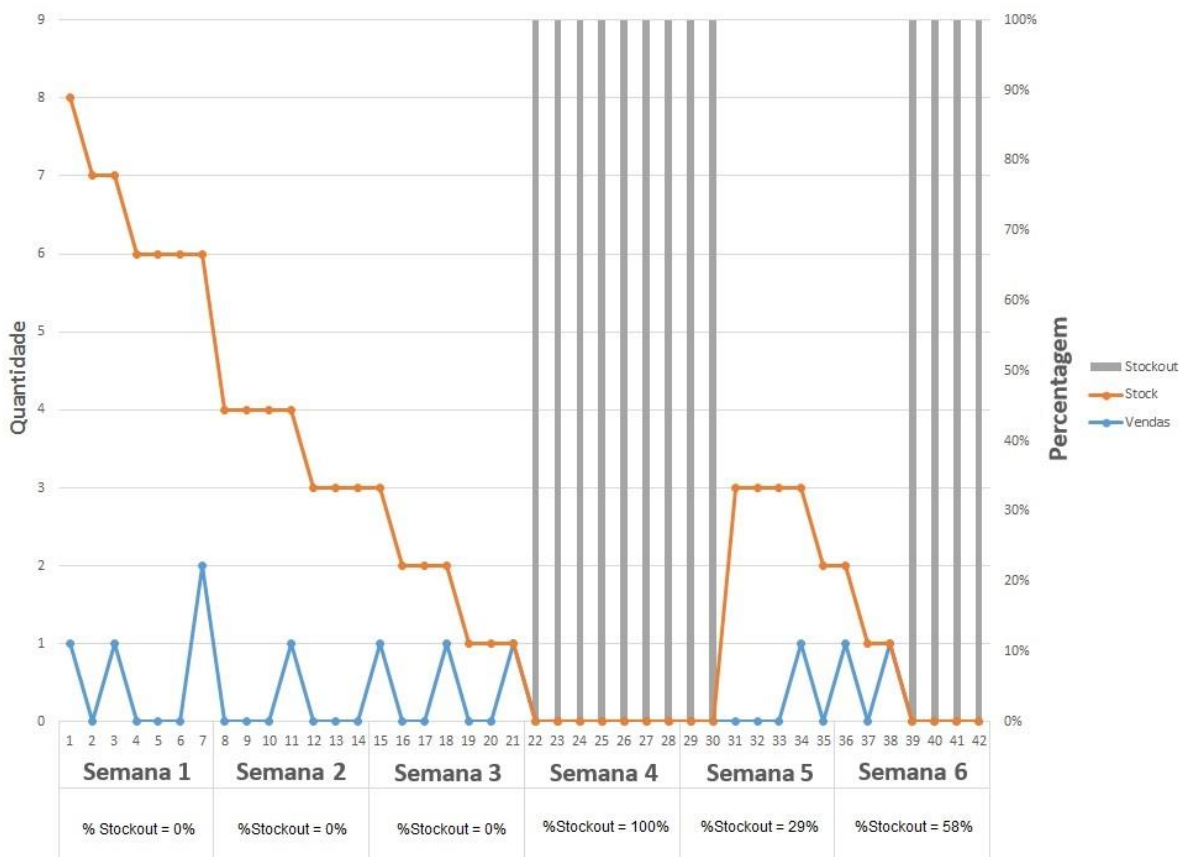


Figura 14 – Representação da prestação de um artigo numa certa loja

Como se pode constatar na Figura 14, no decorrer do ciclo de vida de um artigo numa certa loja, há três situações distintas a serem analisadas, com base na quantidade de *stockouts* por semana:

- *Stockout* = 0% – nestes casos não há vendas perdidas por falta de *stock*, ou seja, o valor das vendas não terá que sofrer qualquer alteração;
- $0\% < \text{Stockout} < 100\%$ – casos em que haja falta de *stock*, causadas ou por vendas de todo o *stock* (Semana 6 na Figura 14) ou por reposição de *stock* já no decorrer da semana (Semana 5 na Figura 14), o que significa perdas parciais de vendas na semana e, por isso, vendas potenciais que devem ser estimadas;
- *Stockout* = 100% - para estes casos, não ocorreu qualquer venda, pois a loja não possuía nenhum artigo em *stock*, obrigando a uma estimativa da totalidade da venda potencial dessa semana.

Estes casos exigem abordagens distintas para a obtenção das vendas potenciais, o que é representado na Figura 15.

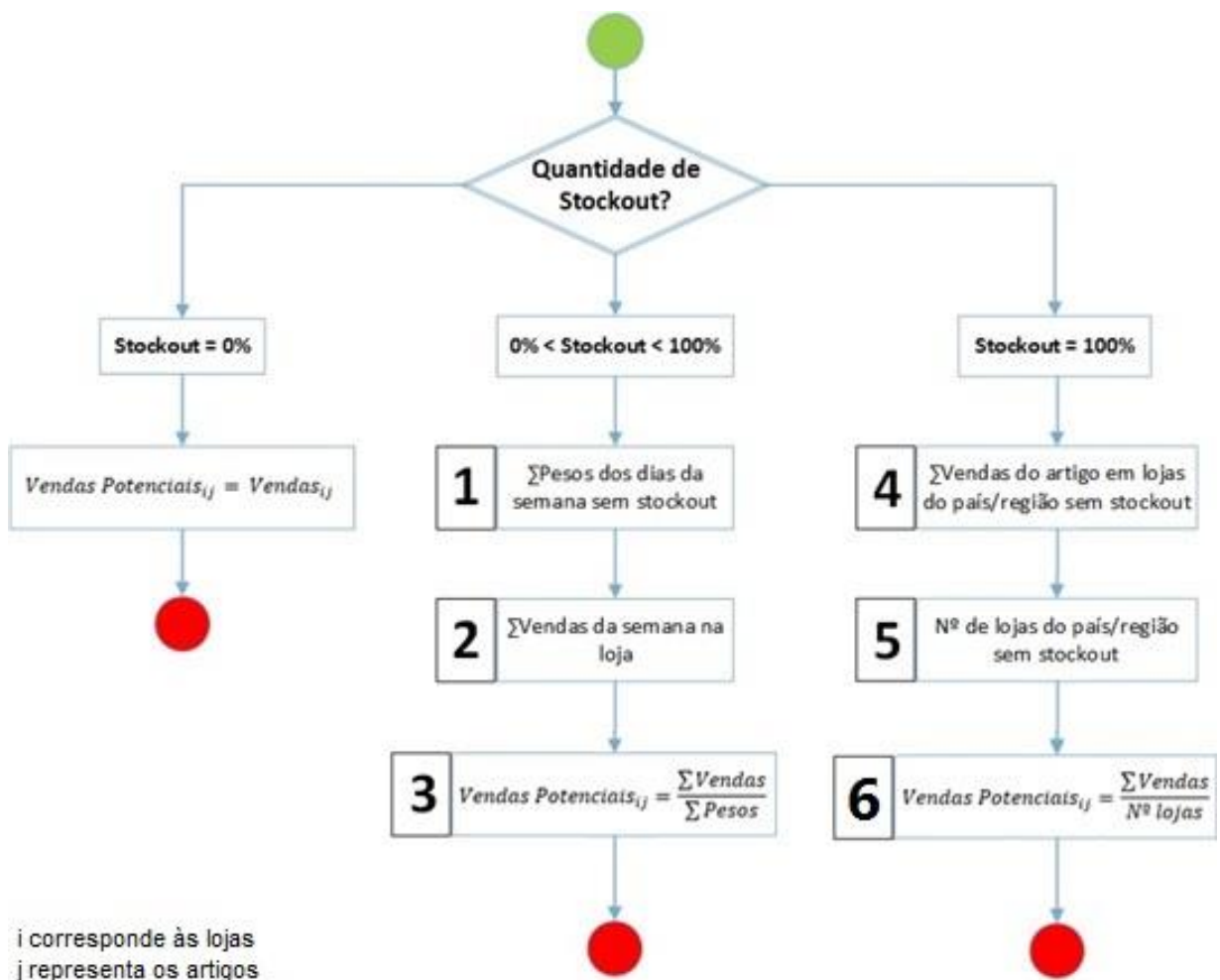


Figura 15 - Fluxograma dos processos de cálculo de vendas potenciais

Da Figura 15 há a salientar as seguintes etapas dos diferentes processos:

1. Σ Pesos dos dias da semana sem *stockout*: é determinado, para cada loja, o peso de cada dia da semana, através do quociente entre a soma das vendas totais da gama a considerar num determinado dia da semana, pelo total das vendas que a loja efetuou dessa gama (ver Tabela 5). Posteriormente é feita a adição dos pesos dos dias da semana em que não tenha ocorrido *stockout* (ver Tabela 6);

Tabela 5 – Tabela auxiliar de determinação dos pesos do dia da semana numa loja

Loja	Dia da Semana	Vendas	Pesos
Loja 1	2ª	252	10,87%
	3ª	259	11,17%
	4ª	292	12,59%
	5ª	294	12,68%
	6ª	332	14,32%
	Sábado	515	22,21%
	Domingo	375	16,17%
Total		2319	100,00%

Tabela 6 – Cálculo dos pesos a utilizar para a determinação das vendas potenciais

Loja	Semana	Dia da Semana	Vendas	Stockout	Pesos	Σ Pesos
Loja 1	5	2ª	0	1	10,87%	77,96%
	5	3ª	0	1	11,17%	
	5	4ª	1	0	12,59%	
	5	5ª	0	0	12,68%	
	5	6ª	2	0	14,32%	
	5	Sábado	1	0	22,21%	
	5	Domingo	0	0	16,17%	

2. Σ Vendas da semana: somatório das vendas do artigo, durante a semana pretendida do seu ciclo de vida, na loja em análise, nos dias em que não houve *stockout* (ver Tabela 7);

Tabela 7 - Cálculo das vendas a utilizar para a determinação das vendas potenciais

Loja	Semana	Dia da Semana	Vendas	Stockout	Σ Vendas
Loja 1	5	2ª	0	1	4
	5	3ª	0	1	
	5	4ª	1	0	
	5	5ª	0	0	
	5	6ª	2	0	
	5	Sábado	1	0	
	5	Domingo	0	0	

3. Vendas Potenciais, para este caso, calculadas pelo quociente entre os valores determinados anteriormente, como demonstrado na Tabela 8;

Tabela 8 – Determinação das vendas potenciais para situações de *stockout* entre 0% e 100% na semana

Loja	Semana	Dia da Semana	Vendas	Stockout	Pesos	Σ Vendas	Σ Pesos	Vendas Potenciais
Loja 1	5	2ª	0	1	10,87%	4	77,96%	5,1
	5	3ª	0	1	11,17%			
	5	4ª	1	0	12,59%			
	5	5ª	0	0	12,68%			
	5	6ª	2	0	14,32%			
	5	Sábado	1	0	22,21%			
	5	Domingo	0	0	16,17%			

4. Σ Vendas do artigo nas lojas do país/região sem *stockout*: para a semana a ter em consideração, determina-se o total das vendas do artigo nas restantes lojas do país, ou da região, caso seja a única loja do país, analisando apenas as lojas em que, nessa semana, não tenha havido *stockouts* (ver Tabela 9);
5. N° de lojas do país/região sem *stockout*: número de lojas do país, ou da região, caso seja a única loja do país, onde, para a semana a considerar, não houve *stockouts* do artigo (ver Tabela 9);

Tabela 9 – Determinação do n° de lojas e o valor das vendas do artigo no país/região

Artigo	Semana	Vendas	N° Lojas
Artigo X	1	2568	446
	2	4005	441
	3	2918	425
	4	1965	390
	5	870	309
	6	447	258

6. Vendas Potenciais, para este caso, calculadas pelo quociente entre os valores de vendas e número de lojas para a semana pretendida obtidos na tabela H, (ver Tabela 10).

Tabela 10 - Determinação das vendas potenciais para situações de 100% de *stockout* na semana

Loja	Semana	Dia da Semana	Vendas	Stockout	Σ Vendas País/Região	N° Lojas País/Região	Vendas Potenciais
Loja 1	4	2ª	0	0	1965	390	5,0
	4	3ª	0	0			
	4	4ª	0	0			
	4	5ª	0	0			
	4	6ª	0	0			
	4	Sábado	0	0			
	4	Domingo	0	0			

Determinado o valor das vendas potenciais é possível analisar o acréscimo que isso representa face aos valores reais de vendas ao longo das semanas. (Figura 16)

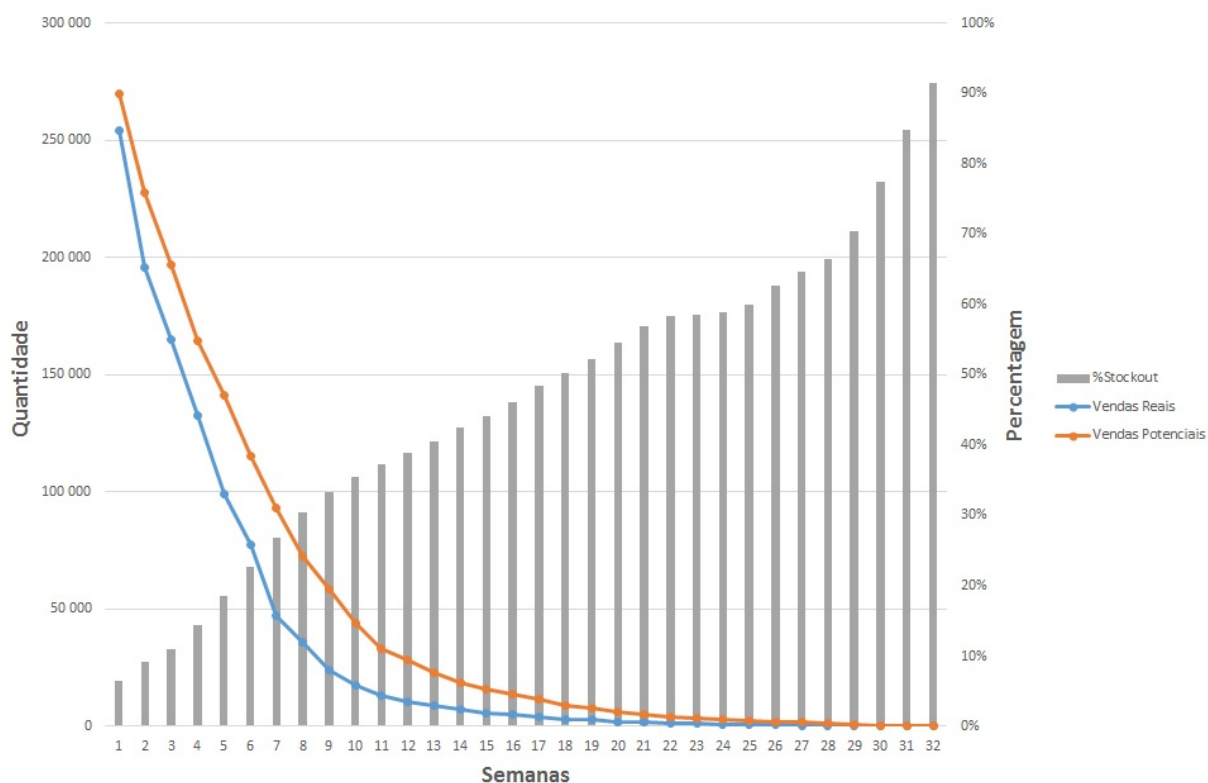


Figura 16 - Representação das vendas reais e potenciais da gama Carteiras ao longo das semanas

Como é visível na Figura 16, há um aumento da percentagem de *stockouts* à medida que o tempo de vida dos artigos nas lojas vai aumentando, o que se reflete num aumento considerável das vendas potenciais face às vendas reais. Esta percentagem de *stockouts* é obtida pelo quociente entre casos de *stockout* para o par artigo-loja-dia sobre todos os casos possíveis do par artigo-loja-dia.

Esse aumento encontra-se representado na Tabela 11.

Tabela 11 - Variação das vendas potenciais face às vendas reais e quantidade de stockouts no período considerado

Vendas Potenciais	Vendas Reais	Variação	%Stockout
1 575 889	1 217 182	29%	33%

Apesar de o valor de vendas potenciais ir de encontro ao expectável face à quantidade de *stockouts* que houve no período, foi definido, em conjunto com os agentes de decisão da *Parfois*, proceder a uma limitação deste valor, devido à quantidade de espaço de armazenamento que este aumento implicaria.

A restrição sugerida foi aplicada com base numa das características dos artigos, o *Fashion Type* que, como exposto anteriormente, está associado a um número de semanas em que se prevê que o produto esteja em loja.

A Tabela 12 apresenta a relação entre o *Fashion Type* e o número de semanas.

Tabela 12 - Relação entre *Fashion Type* e o número de semanas em que se espera que o produto esteja em loja

Fashion Type	Semanas
Moda	4
Básico Moda	6
Básico	8

Esta característica foi utilizada pois, como se verifica na Figura 16, havia artigos que chegavam a permanecer 33 semanas em loja, o que fazia com que o cálculo das vendas potenciais se prolongasse durante várias semanas. Para além disso, como já foi dito, há a intenção de introduzir produtos novos semanalmente, tornando-se impossível manter artigos durante tanto tempo em loja, por limitações de espaço.

Assim, o cálculo de vendas potenciais para cada artigo associado aos *fashion types* apresentados será estimado até a semana limite ser atingida.

Como se pode confirmar na Tabela 13, estas restrições reduziram para 8% a variação das vendas potenciais, face às vendas reais, um valor em consonância com a redução verificada também na percentagem de *stockouts* que passou para 13%.

Tabela 13 - Variação das vendas potenciais face às vendas reais e quantidade de *stockouts* para as restrições definidas

Vendas Potenciais	Vendas Reais	Variação	%Stockout
1 310 634	1 217 182	8%	13%

4.3 Modelo de previsão

Após determinados e avaliados os valores das vendas perdidas por falta de *stock* nas lojas, esse valor é utilizado como variável dependente de um modelo que visa determinar, com base num novo conjunto de variáveis, quais as quantidades de compra para os artigos da nova estação.

Esse modelo de previsão é apoiado por uma técnica de previsão, as *random forests*, apresentadas na secção 2.2.1. Optou-se por recorrer a esta técnica devido à sua extrema flexibilidade, uma vez que permite que sejam feitas análises para um grande número de atributos preditivos, independentemente do tamanho da base de dados em uso. As *random forests* destacam-se também por outros fatores, tal como a reduzida necessidade de otimização dos seus parâmetros e a capacidade de apresentar quais os atributos que foram mais importantes para o seu desempenho.

A implementação desta metodologia foi feita com recurso à ferramenta de cálculo estatístico R, mais especificamente a um dos muitos pacotes disponíveis de forma gratuita - o pacote “*randomforest*”.

Para que se possa utilizar esta função *randomforest* é necessário, para além da base de dados com os atributos e respetiva variável dependente, a definição dos parâmetros: número de árvores e número de características a considerar em cada nó.

A base de dados a ser usada consiste nas características das 384 carteiras consideradas, assim como a variável dependente a elas associada; o valor de vendas potenciais determinado anteriormente.

Em termos de características das carteiras a serem usadas no modelo, estas foram definidas após reuniões com os responsáveis pela gama visto serem conhecedores dos critérios utilizados para a criação de novos artigos, sendo assim mais fácil de perceber quais aqueles que teriam maior influência para o modelo. De entre os vários atributos associados a um artigo, aqueles que foram apresentados como mais importantes, por serem aqueles que melhor categorizam os artigos, foram os atributos “Família”, “Subfamília”, “Tipo de Cor”, “Cor” e “Preço Base”.

Verificou-se também que, para além das características obtidas com os responsáveis da gama, há uma relação entre a quantidade comprada por artigo e o seu mês de chegada às lojas como se pode ver na Figura 17. A quantidade vendida é menor quando a entrada do produto em loja é mais tardia.

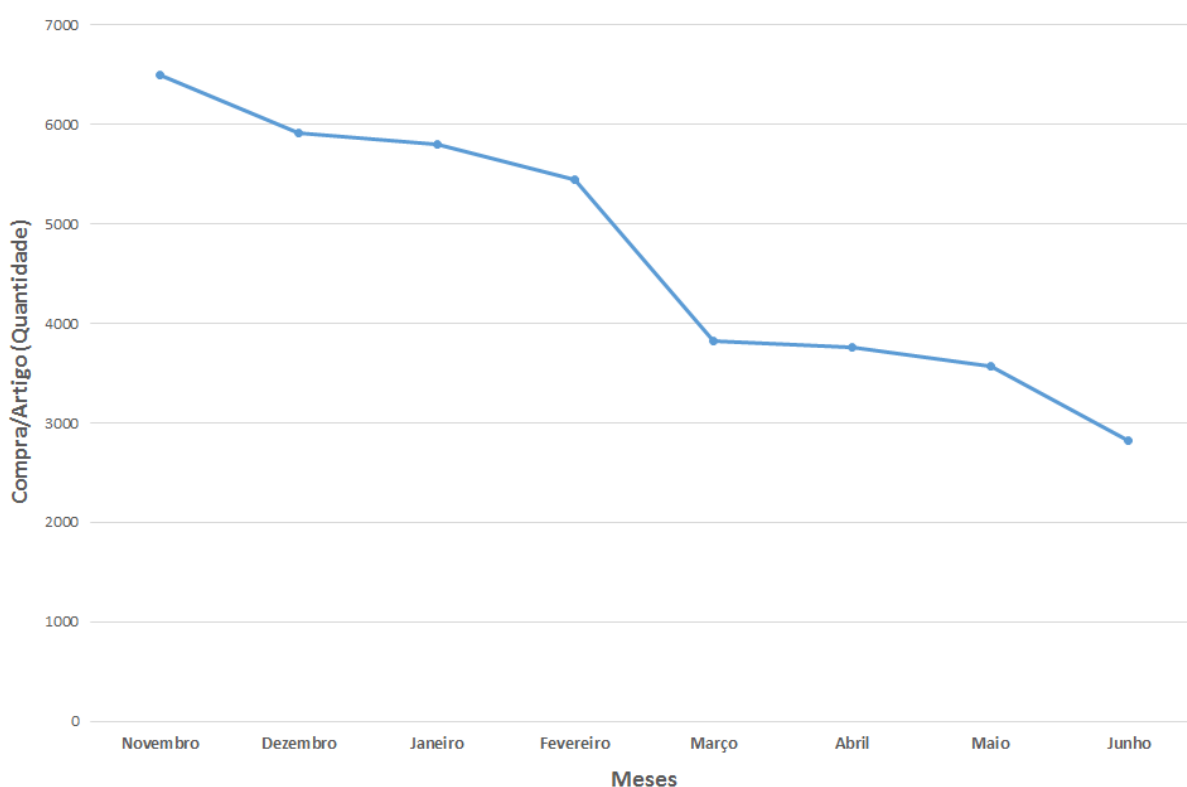


Figura 17 - Relação entre a quantidade compra por artigo e o mês de entrada em loja

Por fim, foram ainda incluídas as características “Aposta” e “Tipologia”, expostas respetivamente nas secções 3.1 e 3.2, e o “*Fashion Type*”.

Em suma, as variáveis a incluir no modelo apresentam-se na Tabela 14. O Anexo C apresenta as categorias de cada variável.

Tabela 14 - Exemplo da variável dependente e atributos usados no modelo

Artigo	Familia	SubFamilia	Típo Cor	Cor	Fashion	Mês	Preço Base	Tipologia	Aposta	Venda Potencial
Artigo1	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	DCBA	Alta	22 263
Artigo2	Pastas	A4	Cor Unica	Skin	Básico	MAI	29.99	BA	Alta	7 721

Apesar de o valor das vendas potenciais representar a quantidade que se poderia vender de um artigo, durante o período normal de vendas, na totalidade das lojas em que esteve exposto, o mesmo artigo pode apresentar velocidades de venda muito díspares, dependendo, por certo, da tipologia em que as lojas se enquadram. Por isso, para casos de artigos que foram enviados para aglomerados de tipologias, como o Artigo 1 da Tabela 14 que foi enviado para tipologias DCBA, há a necessidade de diferenciar a prestação do artigo nas diferentes tipologias, passando, assim, a ser considerada a venda potencial associada a cada uma das tipologias de forma diferenciada.

Este processo originou os valores apresentados na Tabela 15, tendo em conta a Tabela 14.

Tabela 15 - Exemplo da venda potencial e atributos usados no modelo, repartidos por tipologia

Artigo	Família	SubFamília	Tipo Cor	Cor	Fashion	Mês	Preço Base	Tipologia	Aposta	Venda Potencial
Artigo1 D	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	D	Alta	70
Artigo1 C	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	C	Alta	3 930
Artigo1 B	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	B	Alta	7 678
Artigo1 A	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	A	Alta	10 585
Artigo2 B	Pastas	A4	Cor Unica	Skin	Básico	MAI	29.99	B	Alta	2 623
Artigo2 A	Pastas	A4	Cor Unica	Skin	Básico	MAI	29.99	A	Alta	5 098

Por fim, e visto que há uma variação entre o número de lojas do histórico e o número de lojas a considerar para a próxima estação, é determinado o valor médio da venda potencial, para a quantidade de lojas associadas a cada tipologia (ver Tabela 16), que resultaram nos valores da Tabela 17.

Tabela 16 - Número de lojas, por tipologia, em que os artigos estiveram expostos

Tipologia	Nº Lojas
A	140
B	182
C	161
D	6

Tabela 17 - Exemplo da venda potencial média e atributos usados no modelo, repartidos por tipologia

Artigo	Família	SubFamília	Tipo Cor	Cor	Fashion	Mês	Preço Base	Tipologia	Aposta	Venda Potencial
Artigo1 D	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	D	Alta	11,74
Artigo1 C	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	C	Alta	24,41
Artigo1 B	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	B	Alta	42,18
Artigo1 A	PVC Basico	Shopper	Cor Unica	Camel	Moda	NOV	22.99	A	Alta	75,61
Artigo2 B	Pastas	A4	Cor Unica	Skin	Básico	MAI	29.99	B	Alta	14,41
Artigo2 A	Pastas	A4	Cor Unica	Skin	Básico	MAI	29.99	A	Alta	36,41

Estabelecida a base de dados a ser utilizada para a criação do modelo, com nove atributos e uma variável dependente, importa definir os parâmetros das *random forests* (números de árvores e de atributos a serem considerados em cada nó das árvores de decisão). A quantidade de árvores de decisão a ser criadas encontra-se apenas limitada pelas capacidades computacionais da máquina, pelo que se usou um valor de 2000 árvores. Optou-se por este valor por permitir obter um valor de erro *out-of-bag* reduzido (ver Figura 18) e por corresponder a tempos de execução aceitáveis.

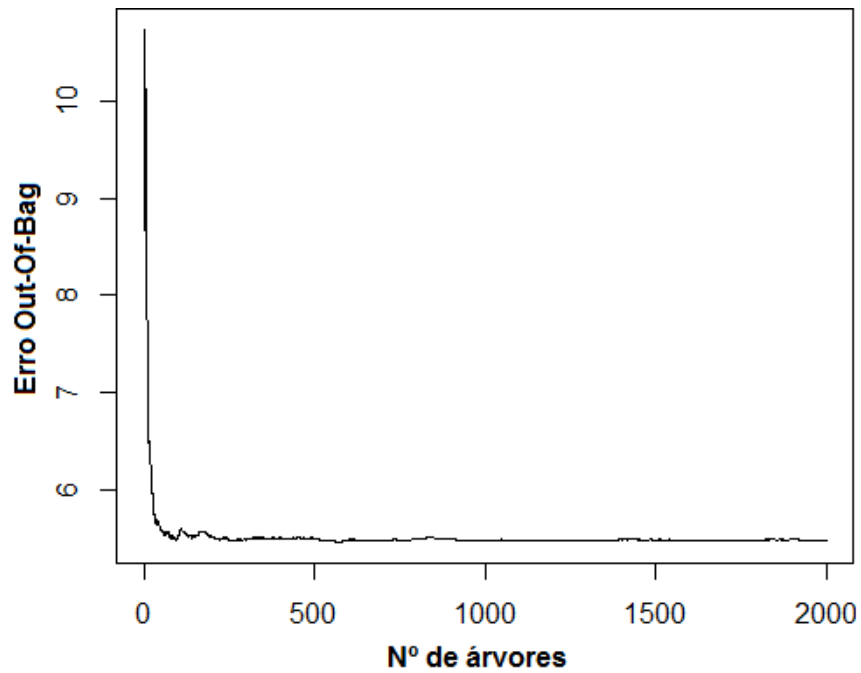


Figura 18 - Relação entre o erro out-of-bag e o número de árvores

Tendo em conta que o número de características a utilizar em cada nó das árvores de decisão deve ser igual ao número de atributos na base de dados, para este modelo o valor será de nove.

Na Figura 19 pode-se validar que é para esse valor que se obtém, efetivamente, o valor mais reduzido do erro *out-of-bag*.

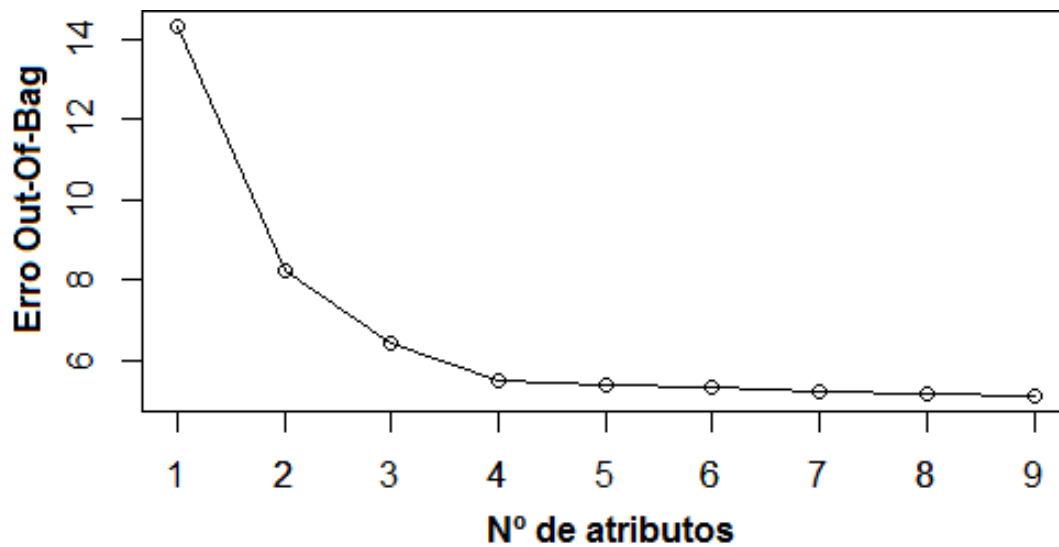


Figura 19 - Relação entre o erro out-of-bag e o número de atributos

Definidos todos os parâmetros necessários, é então executada a ferramenta para a criação do modelo, sendo criadas 2000 árvores de decisão como a apresentada na Figura 20.

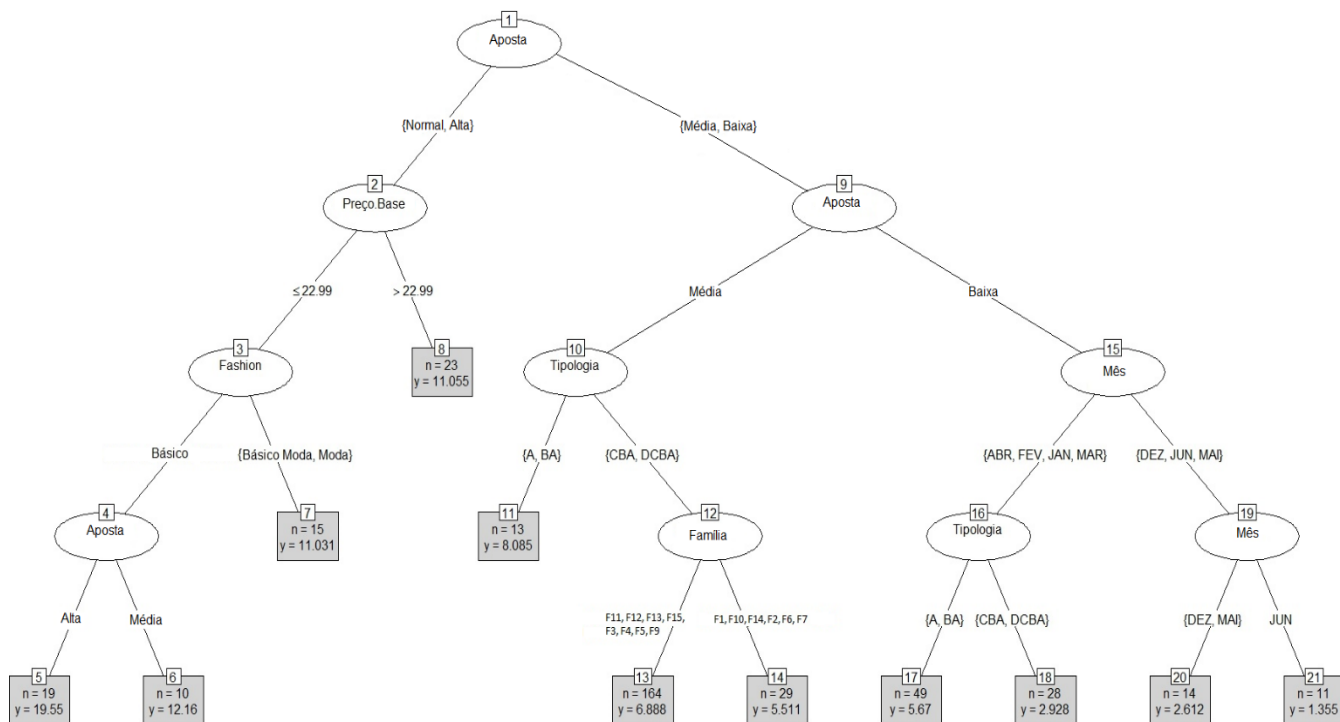


Figura 20 - Exemplo de uma das árvores de decisão criadas pelo modelo

Criado o modelo, a fase seguinte passa pela utilização do mesmo para efetuar previsões para novos artigos, por forma a determinar o valor de vendas médias por loja de cada tipologia. Em termos gerais, após se definirem os atributos desses novos artigos, estes são usados para percorrer cada uma das árvores criadas, sendo o valor de vendas para cada artigo obtido pela média dos diferentes valores resultantes das 2000 árvores criadas.

Obtido o valor de vendas médias por lojas de diferentes tipologias, é então determinado o valor total de venda, multiplicando os valores obtidos com as *random forests* pelo número de lojas de cada tipologia que se prevê que estejam em funcionamento até ao final da estação. Assim, encontra-se determinado aquele que se prevê ser o valor de vendas durante o período normal da estação em análise.

Obtidos os resultados para as previsões de quantidades vendidas durante o período normal da estação, o passo seguinte passa pela determinação das quantidades a comprar para venda durante o período de saldos da estação e as eventuais quebras.

Concretizando, a obtenção de quantidades de compra é feita através do quociente entre a quantidade vendida no período normal (valor obtido com o modelo proposto) e a percentagem de entrada em saldos, definida consoante a combinação de um conjunto de características, nomeadamente, mês de entrada do produto em loja, *fashion type*, tipologia de loja e nível de aposta do artigo. Ao valor obtido com a inclusão do período de saldos é, por fim, acrescentada uma percentagem correspondente a eventuais quebras.

A estimativa é obtida pela equação 4.1.

$$PUR = Vendas\ Potenciais * 1/(1 - \%ES) * 1/(1 - \%Quebra) \tag{4.1}$$

Onde:

- PUR é o valor de compra do artigo
- Vendas Potenciais é o valor de venda para o período normal obtido pelo modelo desenvolvido
- %ES é a percentagem de entrada em saldos definida como objetivo
- %Quebra é a percentagem que se considera que haverá de quebras

5 Apresentação e análise de resultados

Após se ter efetuado a descrição do modelo a utilizar e da forma de obtenção dos valores das quantidades de compra, neste capítulo será feita uma apresentação e análise dos valores obtidos.

É, também, apresentada a ferramenta desenvolvida com o objetivo de interligar o utilizador final com todo o processo de cálculo de quantidades de compra concebido.

5.1 Análise de Resultados

Como descrito na secção 2.2.1, ao contrário de outras metodologias em que é necessária a utilização de uma base de dados de teste para avaliar a performance do modelo, ao utilizar as *random forests* a base de dados é segmentada, para cada uma das árvores, entre dados de treino e teste, sendo determinado o erro *out-of-bag*, calculado automaticamente com base nos dados de teste e apresentado após a criação do modelo.

O valor do erro *out-of-bag* obtido é de cerca 5,11, apresentado na Figura 21 como “*Mean of squared residuals*”. No entanto, este resultado é de difícil interpretação pelo facto de não haver valores padrão do erro.

Outra medida extraída pelo modelo, que permite uma mais fácil interpretação da performance do modelo, é a percentagem de variância explicada, visto como a capacidade de explicar a variância da variável dependente. O modelo criado apresentou um valor da variância explicada de 86,15% como se pode certificar na Figura 21.

```

Type of random forest: regression
Number of trees: 2000
No. of variables tried at each split: 9

Mean of squared residuals: 5.112964
% Var explained: 86.15

```

Figura 21 – Resultados obtidos com a utilização da ferramenta *randomforest*

Outra das características da ferramenta aplicada é a possibilidade de perceber qual a importância dos atributos utilizados para determinar o valor final da previsão. Esta importância é medida pelo aumento percentual do erro, caso uma dada variável seja retirada do modelo. Assim, quanto maior for o incremento percentual do erro, mais importante é a variável para o modelo.

Como se pode atestar na Figura 22, dos atributos considerados para a criação do modelo destacam-se claramente os atributos “Tipologia” e “Aposta” como sendo os mais importantes para o mesmo.

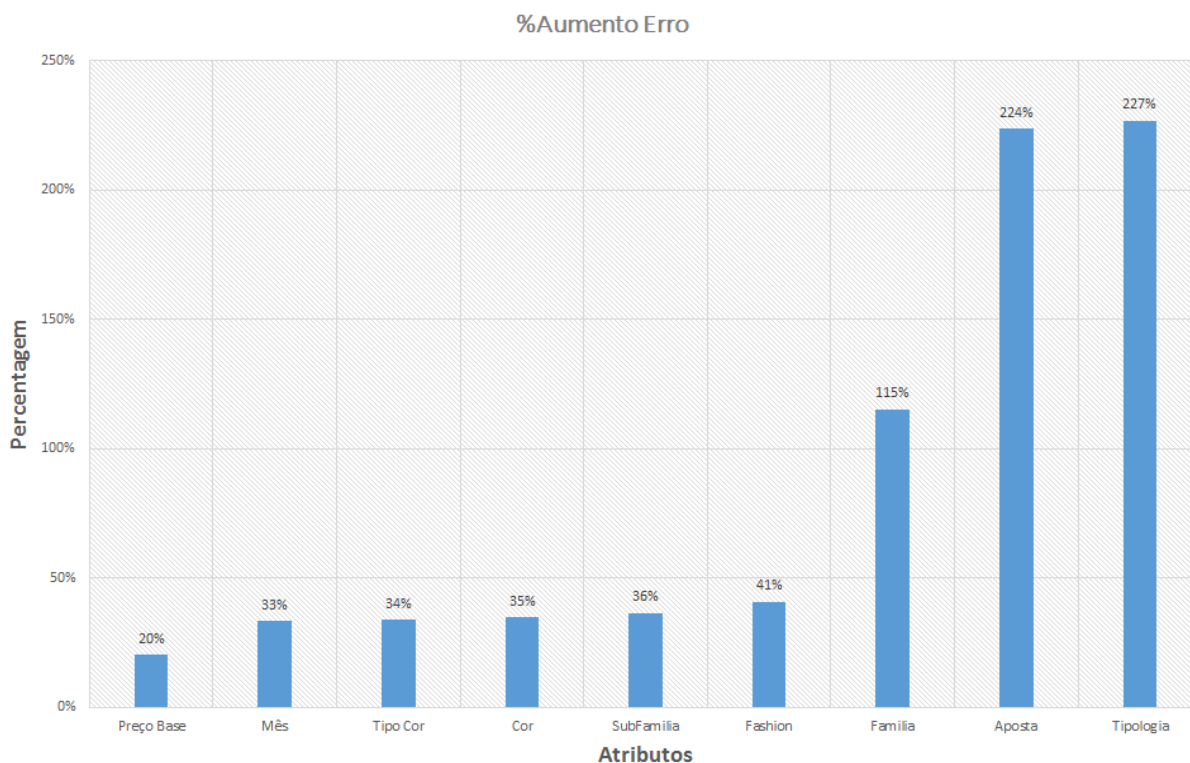


Figura 22 - Importância dos atributos escolhidos

A preponderância destes atributos era, à partida, expectável pois ambos permitem uma categorização mais específica do artigo quando comparado com os outros atributos que são bastante generalistas e que, por isso, não apresentam tanta relevância para o modelo.

Por último, faltaria analisar qual o real comportamento do modelo em relação a artigos de outras estações. No entanto, a criação de um indicador de performance que permitisse avaliar a capacidade preditiva para artigos de outra estação implicaria um tratamento prévio dos valores de vendas potenciais e de *stockout* desses artigos. Todo este processo implicaria acesso a dados difíceis de obter, revelando-se, por isso, uma hipótese inviável.

Para contornar este problema, foram realizadas reuniões com responsáveis do Departamento de Compras, com o objetivo de testar o modelo produzido. Os testes efetuados apresentaram valores que foram ao encontro das expectativas dos conhecedores da área.

5.2 Ferramenta de determinação de quantidades de compra

A ferramenta foi criada com recurso ao programa *Microsoft Visual Basic* do *Microsoft Excel* e consiste numa interface que permite ao utilizador a escolha de diversas funcionalidades, desde a determinação das quantidades de compra, consoante a combinação de características disponíveis, até à análise do comportamento dessas mesmas características nos dados históricos utilizados.

Numa primeira instância é exibida uma janela idêntica à apresentada na Figura 23, havendo a possibilidade de escolher a combinação de características que se pretende para os novos artigos.

The screenshot shows the 'PUR Quantity' tool interface. It features several dropdown menus and text input fields for filtering data. The fields are labeled as follows:

- * Família: [Dropdown]
- * Sub Família: [Dropdown]
- * Tipo Cor: [Dropdown]
- * Cor: [Dropdown]
- * Fashion: [Dropdown]
- * Mês: [Dropdown]
- * Preço Base: [Dropdown] [Text Input: ex: 16,99 ou 16,99]
- * Tipologia: [Dropdown]
- * Aposta: [Dropdown]

Below the fields, there is a red asterisk indicating mandatory fields: *** - Campos obrigatórios**. At the bottom, there are six buttons: Preview, Add, Go, Results, Clean, and Reset.

Figura 23 - Janela inicial da ferramenta criada

Antes de passar à determinação das suas quantidades de compra e depois de definidas as características do artigo, o utilizador tem a possibilidade de analisar as vendas no histórico, ou a percentagem de entrada em saldos, associadas às características escolhidas (ver Figura 24). Assim, esta funcionalidade auxilia numa escolha mais correta do nível de aposta a considerar para o artigo, pois permite uma análise do comportamento histórico dessa característica consoante os diferentes níveis de aposta.

The screenshot shows the 'PUR Quantity' tool interface with the analysis results. The left panel shows the selected filters:

- * Família: PVC Basico
- * Sub Família: Shopper
- * Tipo Cor: Cor Unica
- * Cor: Camel
- * Fashion: Básico
- * Mês: NOV
- * Preço Base: 22.99
- * Tipologia: DCBA
- * Aposta: [Dropdown]

The right panel shows the analysis results for different bet levels. The 'Preço Base' filter is selected, and the 'Vendas' radio button is checked. The results are displayed in a table:

Aposta	Min*	Média*	Max*	%SKU	#SKU
Baixa	300	1 360	2 223	25%	24
Média	2 223	4 191	6 932	59%	57
Normal	6 932	8 474	11 099	9%	9
Alta	11 099	15 280	19 900	6%	6
Total	300	4 578	19 900	100%	96

Below the table, there is a red asterisk indicating that the values correspond to potential sales before discounts: *** valores apresentados correspondem à venda potencial, antes de saldos**. A 'Refresh' button is located in the top right corner of the right panel.

Figura 24 - Janela para análise das características pretendidas no histórico

Finalizada a escolha dos atributos dos artigos, o processo continua com a determinação das quantidades de compra do(s) artigo(s) previamente definidos. Nesta etapa, há uma integração da ferramenta R e do modelo *randomforest* criado, que recebe como *input* os dados históricos e os dados referentes aos novos artigos e, após executar o modelo, retorna os valores obtidos. Por fim, a esses valores são acrescentadas as quebras e a quantidade a ser vendida nos saldos, com base na percentagem de entrada em saldos.

Após efetuados todos os cálculos necessários, é então apresentado, numa janela como a da Figura 25, o número de unidades a comprar para o(s) artigo(s) criado(s), a percentagem histórica de entrada em saldos, associada ao mesmo conjunto de características, assim como a percentagem atual que foi utilizada no cálculo.

Artigo	Familia	SubFamilia	Tipo Cor	Cor	Fashion	Month	Preco Base Iva	Tipologia	Aposta	%ES Hist	%ES	Potencial
Artigo 1	PVC Basico	Shopper	Cor Unica	Camel	Básico	NOV	22.99	DCBA	Alta	12%	8%	22.411

Figura 25 - Janela com a apresentação do artigo criado e respetivas quantidade de compra

Para além de serem determinadas as quantidades de compra para as características pretendidas pelo utilizador, são testadas outras combinações com todas as possibilidades de cores existentes no histórico. Tudo isto com o intuito de apresentar ao utilizador um *ranking* das cinco cores (ver Figura 26) que originaram os maiores valores de quantidades de compras, mantendo as restantes características pretendidas fixas.

Rank	Artigo	Familia	SubFamilia	Tipo Cor	Cor	Fashion	Mês	Preco Base	Tipologia	Aposta	%ES Hist	%ES	Potencial
1	Artigo 1	PVC Basico	Shopper	Cor Unica	Preto	Básico	NOV	22.99	DCBA	Alta	12%	8%	22.655
2	Artigo 1	PVC Basico	Shopper	Cor Unica	Camel	Básico	NOV	22.99	DCBA	Alta	12%	8%	22.411
3	Artigo 1	PVC Basico	Shopper	Cor Unica	Azul Marinho	Básico	NOV	22.99	DCBA	Alta	12%	8%	21.161
4	Artigo 1	PVC Basico	Shopper	Cor Unica	Rosa	Básico	NOV	22.99	DCBA	Alta	12%	8%	19.553
5	Artigo 1	PVC Basico	Shopper	Cor Unica	Bege	Básico	NOV	22.99	DCBA	Alta	12%	8%	19.371

Figura 26 - Janela com a apresentação de um ranking de outras possibilidades de cor do artigo criado e respetivas quantidade de compra

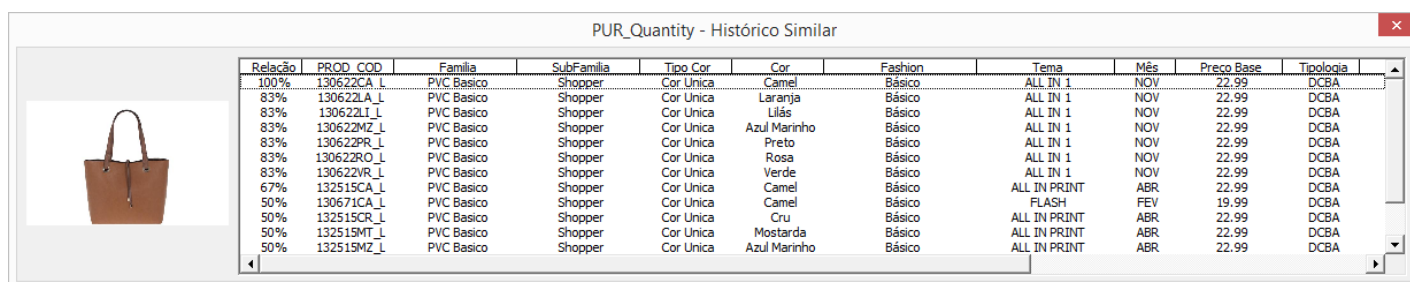
Através da janela da Figura 25 é ainda fornecida ao utilizador a possibilidade de analisar o comportamento histórico de artigos similares. Concretizando, enquanto anteriormente era apenas possível analisar um atributo de forma distinta, agora é possível ver, efetivamente, quais os artigos com características idênticas e como se comportaram.

Pelo facto de a variedade de artigos no histórico ser reduzida, não existiram muitos casos em que a combinação de características fosse exatamente idêntica. Por isso, optou-se por apresentar outros artigos até um certo grau de semelhança de algumas particularidades, definindo-se que as características “Família”, “SubFamília”, “Fashion Type” e “Tipologia” seriam fixas. Para as restantes características, os artigos seriam apresentados caso esses atributos se encontrassem dentro dos seguintes limites:

- Tipo de Cor e Cor – são apresentados artigos com outras combinações de tipo de cor e cor, sem restrições;
- Mês – são apresentados artigos independentemente do mês, sendo considerado um decréscimo na relação consoante o distanciamento face ao artigo original, ou seja, para um artigo que chegue às lojas em dezembro, é atribuída uma melhor relação a artigos de novembro e de janeiro, decrescendo gradualmente para os restantes meses;
- Preço Base – são apresentados artigos que tenham uma variação máxima, face ao artigo criado, de 20%; para um artigo com um preço base de 19.99€ são apresentados artigos que tenham preço entre 17.99€ e 22.99€;
- Aposta – são apresentados artigos com outros níveis de aposta, sem restrições.

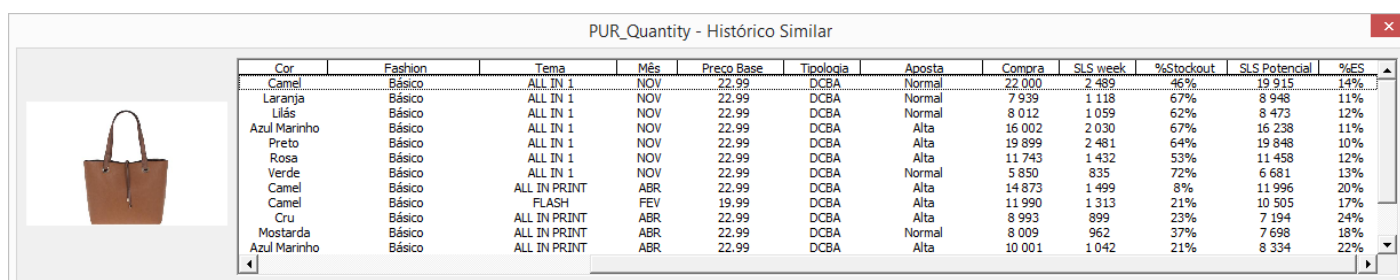
Outra das razões que justifica a versatilidade na apresentação dos artigos históricos, é o facto de os comportamentos dentro dos limites considerados serem comparáveis, nas características em que haja variação.

Estes artigos históricos são apresentados em janelas similares à das Figura 27 e Figura 28 em que, para além das características dos artigos históricos é apresentado o grau de semelhança, assim como uma imagem do artigo, o tema em que se insere, a quantidade de compra, a venda média por semana (SLS week), a percentagem de *stockouts*, vendas potenciais antes de saldos e a percentagem de entrada em saldos.



Relação	PROD_COD	Familia	SubFamilia	Tipo Cor	Cor	Fashion	Tema	Mês	Preço Base	Tipologia
100%	130622CA_L	PVC Basico	Shopper	Cor Unica	Camel	Básico	ALL IN 1	NOV	22.99	DCBA
83%	130622LA_L	PVC Basico	Shopper	Cor Unica	Laranja	Básico	ALL IN 1	NOV	22.99	DCBA
83%	130622LI_L	PVC Basico	Shopper	Cor Unica	Lilás	Básico	ALL IN 1	NOV	22.99	DCBA
83%	130622MZ_L	PVC Basico	Shopper	Cor Unica	Azul Marinho	Básico	ALL IN 1	NOV	22.99	DCBA
83%	130622PR_L	PVC Basico	Shopper	Cor Unica	Preto	Básico	ALL IN 1	NOV	22.99	DCBA
83%	130622RO_L	PVC Basico	Shopper	Cor Unica	Rosa	Básico	ALL IN 1	NOV	22.99	DCBA
83%	130622VR_L	PVC Basico	Shopper	Cor Unica	Verde	Básico	ALL IN 1	NOV	22.99	DCBA
67%	132515CA_L	PVC Basico	Shopper	Cor Unica	Camel	Básico	ALL IN PRINT	ABR	22.99	DCBA
50%	130671CA_L	PVC Basico	Shopper	Cor Unica	Camel	Básico	FLASH	FEV	19.99	DCBA
50%	132515CR_L	PVC Basico	Shopper	Cor Unica	Cru	Básico	ALL IN PRINT	ABR	22.99	DCBA
50%	132515MT_L	PVC Basico	Shopper	Cor Unica	Mostarda	Básico	ALL IN PRINT	ABR	22.99	DCBA
50%	132515MZ_L	PVC Basico	Shopper	Cor Unica	Azul Marinho	Básico	ALL IN PRINT	ABR	22.99	DCBA

Figura 27 - Janela com a apresentação de artigos do histórico dentro de um certo grau de semelhança face ao artigo criado



Cor	Fashion	Tema	Mês	Preço Base	Tipologia	Aposta	Compra	SLS week	%Stockout	SLS Potencial	%ES
Camel	Básico	ALL IN 1	NOV	22.99	DCBA	Normal	22 000	2 489	46%	19 915	14%
Laranja	Básico	ALL IN 1	NOV	22.99	DCBA	Normal	7 939	1 118	67%	8 948	11%
Lilás	Básico	ALL IN 1	NOV	22.99	DCBA	Normal	8 012	1 059	62%	8 473	12%
Azul Marinho	Básico	ALL IN 1	NOV	22.99	DCBA	Alta	16 002	2 030	67%	16 238	11%
Preto	Básico	ALL IN 1	NOV	22.99	DCBA	Alta	19 899	2 481	64%	19 848	10%
Rosa	Básico	ALL IN 1	NOV	22.99	DCBA	Alta	11 743	1 432	53%	11 458	12%
Verde	Básico	ALL IN 1	NOV	22.99	DCBA	Normal	5 890	835	72%	6 681	13%
Camel	Básico	ALL IN PRINT	ABR	22.99	DCBA	Alta	14 873	1 499	8%	11 996	20%
Camel	Básico	FLASH	FEV	19.99	DCBA	Alta	11 990	1 313	21%	10 505	17%
Cru	Básico	ALL IN PRINT	ABR	22.99	DCBA	Alta	8 993	899	23%	7 194	24%
Mostarda	Básico	ALL IN PRINT	ABR	22.99	DCBA	Normal	8 009	962	37%	7 698	18%
Azul Marinho	Básico	ALL IN PRINT	ABR	22.99	DCBA	Alta	10 001	1 042	21%	8 334	22%

Figura 28 – Restante parte da janela com a apresentação de artigos do histórico dentro de um certo grau de semelhança face ao artigo criado

6 Conclusões e perspetivas de trabalho futuro

Neste capítulo, é analisada a consecução dos objetivos propostos e, com base no trabalho elaborado e nos conhecimentos adquiridos, são apresentadas propostas de eventuais trabalhos futuros que permitam a continuação e eventual melhoria do estudo realizado no âmbito deste trabalho.

6.1 Satisfação dos objetivos

Tal como referido ao longo desta dissertação, os objetivos passavam pela otimização do processo atual de determinação de quantidades de compra de artigos na *Parfois*, utilizando, para isso, abordagens que até ao momento não tinham sido consideradas. Além disso, o projeto visava a implementação de uma ferramenta que permitisse ao utilizador, através de uma interface *user-friendly*, obter as quantidades de compras para uma combinação de características pretendidas.

Apesar de se ter restringido a quantidade de informação a utilizar apenas à gama de Carteiras da estação Primavera-Verão de 2015, uma vez que os registos são diários, a base de dados inclui mais de 30 milhões de registos, relativos a um período de oito meses, para 489 lojas e 384 artigos.

Esta quantidade colossal de dados originou vários problemas, desde a fase inicial em que se procedeu à extração dos mesmos dos servidores para uma aplicação que permitisse trabalhá-los, até à fase de cálculos sobre essa mesma base que, como seria de prever, resultou num tempo considerável de espera até que fossem efetuados.

Trabalhada a informação obtida e feita a análise ao comportamento dos artigos nas lojas, nomeadamente ao nível de *stockouts*, foi possível estimar as vendas perdidas nesses casos, resultando em vendas potenciais cerca de 13% superiores às reais.

Recorrendo às características dos artigos e às potenciais vendas associadas, foi criado um modelo de previsão suportado por uma técnica de *data mining*, as *random forests*, com o objetivo de prever as quantidades de compra de artigos para a nova estação com base em características definidas pelo utilizador. Os resultados obtidos revelam que o modelo proposto apresenta uma boa capacidade preditiva, uma vez que a variância explicada pelo modelo ronda os 86%.

Criado o modelo, este foi utilizado para fazer a estimativa das quantidades a comprar para a próxima estação. Uma vez que no momento não há informação disponível sobre as vendas efetivas para este período não foi possível estimar um índice de performance relativo às previsões para a nova estação. No entanto, o parecer dos elementos do Departamento de Compras, principais beneficiários da ferramenta, é bastante positivo acerca da metodologia e considerações utilizadas.

Por fim, foi criada a ferramenta que permite ao utilizador executar o modelo de forma automática e obter as quantidades de compra para os artigos que pretende. De forma a auxiliar a escolha dos atributos desses mesmos artigos, a ferramenta apresenta ainda a possibilidade de analisar o comportamento dos atributos e artigos presentes no histórico, que sejam similares aos do que se pretende conceber.

6.2 Trabalhos Futuros

Este projeto foi dirigido apenas à gama de Carteiras, uma das dezasseis gamas comercializadas pela *Parfois*. Os bons resultados obtidos abrem a porta a uma posterior adaptação às restantes gamas, sendo que, na criação desta ferramenta já foram tomadas algumas considerações com o objetivo de facilitar essa utilização nas outras gamas. Além disso, a ferramenta pode, ainda, ser otimizada, recorrendo à utilização de uma maior quantidade de dados para a criação do modelo, agregando, para isso, artigos referentes a outros anos e estações. No entanto, esta possibilidade poderá, eventualmente, acarretar alguma complexidade associada à quantidade de novos dados que seria necessário extrair para posterior análise.

Por fim, pelo facto de a ferramenta originar alterações nos processos atuais, haverá a necessidade de proceder à formação não só dos principais beneficiários da ferramenta, como também dos restantes intervenientes na cadeia de abastecimento, com o objetivo de se proceder a alguns ajustes para uma melhor interligação entre todas as etapas da cadeia.

Referências

- Aparicio, Enrique. 2015. "Case Study - Parfois."
- Bhardwaj, Vertica, and Ann Fairhurst. 2010. "Fast fashion: response to changes in the fashion industry." Accessed 1.
- Birtwistle, Grete, Noreen Siddiqui, and Susan S Fiorito. 2003. "Quick response: perceptions of UK fashion retailers." *International Journal of Retail & Distribution Management* 31 (2):118-128.
- Brachman, Ronald J, Tom Khabaza, Willi Kloesgen, Gregory Piatetsky-Shapiro, and Evangelos Simoudis. 1996. "Mining business databases." *Communications of the ACM* 39 (11):42-48.
- Breiman, Leo. 1996. Out-of-bag estimation. Citeseer.
- Breiman, Leo. 2001. "Random forests." *Machine learning* 45 (1):5-32.
- Breiman, Leo, Jerome Friedman, Charles J Stone, and Richard A Olshen. 1984. *Classification and regression trees*: CRC press.
- Cabena, Peter, Pablo Hadjinian, Rolf Stadler, Jaap Verhees, and Alessandro Zanasi. 1998. *Discovering data mining: from concept to implementation*: Prentice-Hall, Inc.
- Canha, Isabel. 2014. "Parfois – Espelho meu, espelho meu." UP Magazine. http://upmagazine-tap.com/pt_artigos/parfois-espelho-meu-espelho-meu/.
- Caro, Felipe, and Jérémie Gallien. 2010. "Inventory management of a fast-fashion retail network." *Operations Research* 58 (2):257-273.
- Chen, Zhengxin, and Qiuming Zhu. 1998. "Query construction for user-guided knowledge discovery in databases." *Information Sciences* 109 (1):49-64.
- Christopher, Martin, Robert Lowson, and Helen Peck. 2004. "Creating agile supply chains in the fashion industry." *International Journal of Retail & Distribution Management* 32 (8):367-376.
- Crofton, Stephanie O, and Luis G Dopico. 2012. "Zara-Inditex and the growth of fast fashion." *Essays in Economic & Business History* 25.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. "The KDD process for extracting useful knowledge from volumes of data." *Communications of the ACM* 39 (11):27-34.
- Fisher, Marshall L, Janice H Hammond, Walter R Obermeyer, and Ananth Raman. 1994. *Making supply meet demand in an uncertain world*: Graduate School of Business Administration, Harvard University Boston.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The elements of statistical learning*. Vol. 1: Springer series in statistics Springer, Berlin.
- Fury, Alexander. 2015. "Pre-collections: Fashion's new favourite thing." <http://www.independent.co.uk/arts-entertainment/art/news/pre-collections-fashion-s-new-favourite-thing-10303234.html>.
- Geurts, Pierre, Damien Ernst, and Louis Wehenkel. 2006. "Extremely randomized trees." *Machine learning* 63 (1):3-42.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *Unsupervised learning*: Springer.
- Liaw, Andy, and Matthew Wiener. 2002. "Classification and regression by randomForest." *R news* 2 (3):18-22.

- Linoff, Gordon S, and Michael JA Berry. 2011. *Data mining techniques: for marketing, sales, and customer relationship management*: John Wiley & Sons.
- Lo, Wei-Shuo, Tzung-Pei Hong, and Rong Jeng. 2008. "A framework of E-SCM multi-agent systems in the fashion industry." *International Journal of Production Economics* 114 (2):594-614.
- Machado, Gustavo, Mariana Recamonde Mendoza, and Luis Gustavo Corbellini. 2015. "What variables are important in predicting bovine viral diarrhea virus? A random forest approach." *Veterinary research* 46 (1):1-15.
- Maragoudakis, Manolis, and Euripides Loukis. 2012. "Using Ensemble Random Forests for the extraction and exploitation of knowledge on gas turbine blading faults identification." *OR insight* 25 (2):80-104.
- Monte, Ana, Carlos Soares, Pedro Brito, and Michel Byvoet. 2013. "Clustering for Decision Support in the Fashion Industry: A Case Study." In *Advances in Sustainable and Competitive Manufacturing Systems*, 997-1008. Springer.
- Moreira, Reinaldo Costa. 2011. "Estudo comparativo das Cadeias de Abastecimento na Indústria do Vestuário Sonae SR–Fashion Division." Universidade do Porto.
- Morgan, James N, and John A Sonquist. 1963. "Problems in the analysis of survey data, and a proposal." *Journal of the American statistical association* 58 (302):415-434.
- Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. 2009. "Application of data mining techniques in customer relationship management: A literature review and classification." *Expert systems with applications* 36 (2):2592-2602.
- Pang-Ning, Tan, Michael Steinbach, and Vipin Kumar. 2006. "Introduction to data mining." Library of congress.
- Parfois. 2015. "Manual de Acolhimento da Parfois."
- Patil, Rahul, Balram Avittathur, and Janat Shah. 2010. "Supply chain strategies based on recourse model for very short life cycle products." *International Journal of Production Economics* 128 (1):3-10.
- Pimentel, Ana. 2016. ""Quem se atrasar a entrar no comércio online, vai estar sempre atrasado". "<http://observador.pt/2016/01/20/atrasar-entrar-no-comercio-online-vai-estar-sempre-atrasado/>.

Anexo A: Organograma da *Parfois*

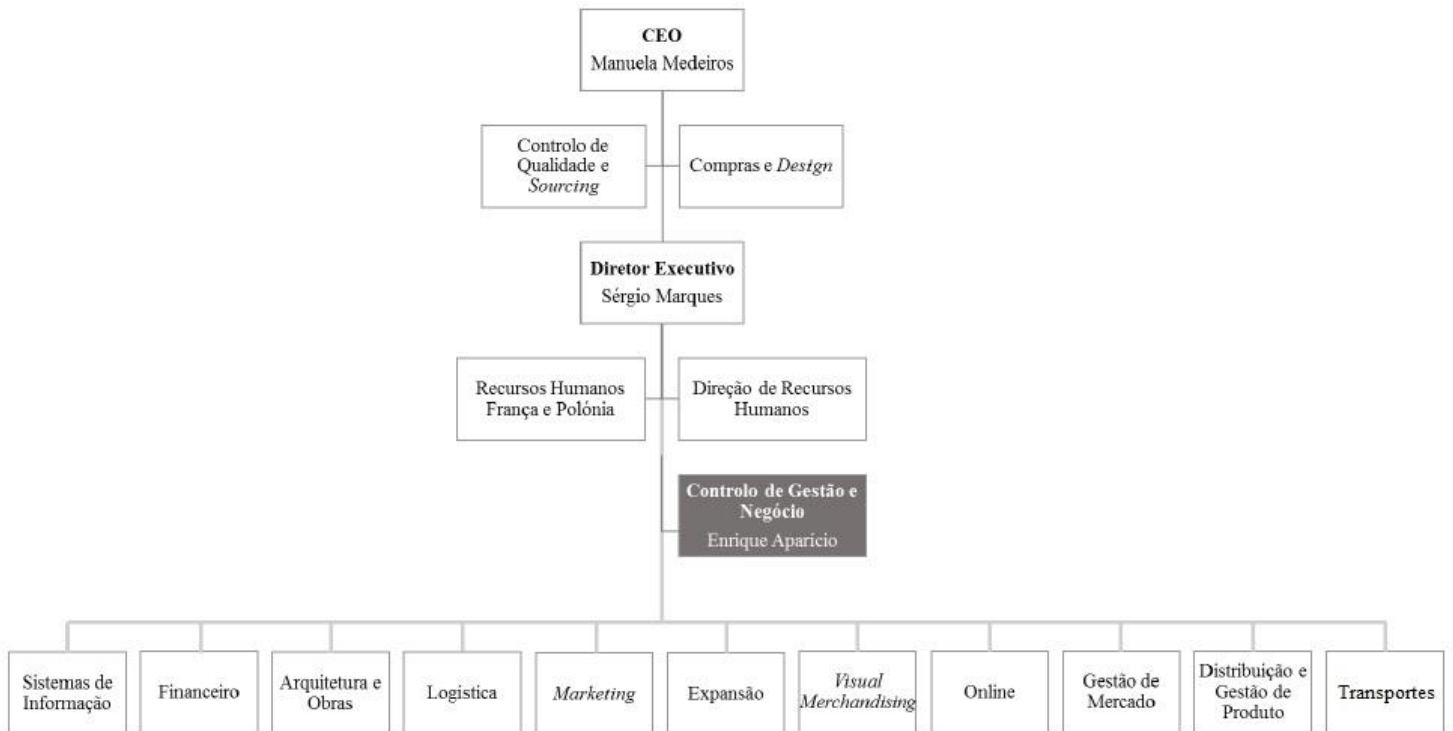


Figura 29 - Organograma da Parfois

Anexo B: Cronograma do projeto

Atividades	#Week	01-02	08-02	15-02	22-02	29-02	07-03	14-03	21-03	28-03	04-04	11-04	18-04	25-04	02-05	09-05	16-05	23-05	30-05	06-06	13-06	20-06	
		a 07-02	a 14-02	a 21-02	a 28-02	a 06-03	a 13-03	a 20-03	a 27-03	a 03-04	a 10-04	a 17-04	a 24-04	a 01-05	a 08-05	a 15-05	a 22-05	a 29-05	a 05-06	a 12-06	a 19-06	a 26-06	
1	Otimização Quantidades Compras																						
1.1	Obtenção dos dados			█	█	█	█	█	█	█													
1.2	Decisão de gama								█	█													
1.3	Análise de potenciais falhas									█	█												
1.4	Análise de Stockouts										█	█	█	█									
1.5	Estimação de Vendas Potenciais													█	█	█	█						
1.6	Influência de características																	█	█				
1.7	Random Forests																		█	█	█		
1.8	Criação da Ferramenta																				█	█	█
	Reuniões			█		█		█		█		█		█		█		█		█		█	

Figura 30 - Cronograma das etapas do projeto

Anexo C: Categorias de atributos usados no modelo

Tabela 18 - Categorias possíveis para os atributos “Família”, “SubFamília”, “Tipo Cor” e “Cor”

Família	SubFamília	Tipo Cor	Cor
Camurca Verdadeiro	A4	Cor Unica	Amarelo
Entrelacado	Bolinha	Multicolor	Aquamar
Estampado	Falsa		Azul
India	Lancheira		Azul Bebe
Palha	Malotes		Azul Marinho
Pastas	Mao		Azulão
Patchwork	Mochila		Bege
Pele Verdadeira	Sacos		Blue Jeans
Plastico	Shopper		Branco
Praia	Tracar		Burgundy
PVC Basico	Verdadeira		Camel
PVC Bicho			Caqui
PVC Estampado			Castanho
PVC Fantasia			Cinza
PVC Liso			Coral
Tecido Liso			Cru
Verniz			Dourado
Vintage			Fushia
			Laranja
			Lilás
			Lima
			Mostarda
			Peach
			Preto
			Rosa
			Skin
			Toupeira
			Turquesa
			Verde
			Vermelho

Tabela 19 - Categorias possíveis para os atributos “Fashion”, “Preço Base”, “Mês”, “Tipologia” e “Aposta”

Fashion	Preço Base	Mês	Tipologia	Aposta
Básico	14.99	Novembro	A	Alta
Básico Moda	16.99	Dezembro	BA	Normal
Moda	17.99	Janeiro	CBA	Média
	19.99	Fevereiro	DCBA	Baixa
	22.99	Março		
	24.99	Abril		
	26.99	Maio		
	29.99	Junho		
	32.99			
	34.99			
	36.99			
	39.99			
	44.99			
	54.99			
	59.99			