# Customer Centric Marketing: Abordagem Analítica

*Rafael Oliveira Marques*

**Dissertação de Mestrado**

Orientador na FEUP: Prof. Nuno Soares

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

**Mestrado Integrado em Engenharia e Gestão Industrial**

2020-06-29

# Resumo

Com a proliferação de perfis de clientes cada vez mais distintos, é de crucial importância as empresas adotarem uma postura centrada no cliente de forma a que se mantenham competitivas. Tal implica a criação de uma cultura baseada no cliente sustentada por uma estrutura organizacional comprometida com a satisfação das necessidades particulares de cada cliente. Uma das funções que deve ser promotora desta nova "forma de estar" é o Marketing, daí a importância da incorporação das ideologias presentes no *Customer Centric Marketing* para ser estabelecida esta postura centrada no cliente na organização.

O objetivo do presente projeto é o desenvolvimento de um modelo analítico centrado no cliente que sustente a realização de uma campanha de marketing. A Sonae FS nos últimos anos tem adotado técnicas de *data mining* para selecionar os clientes mais propensos a receberem uma comunicação, sendo bastante frequente a utilização de modelos de propensão em conjunção com comunicações por *batch*. A dissertação visa o desenvolvimento de um modelo analítico diferente dos referidos- um modelo baseado num algoritmo de *Sequential Pattern Mining* (SPM) para a obtenção de um conjunto de padrões comportamentais interessantes que auxiliem uma comunicação com a qual se pretende aumentar o número de clientes registados na App Universo. Após a aplicação do modelo em questão foram realizadas comunicações para os clientes que não têm app e que realizaram padrões comportamentais obtidos pelo algoritmo. Deste modo, no projeto adotou-se uma combinação disruptiva de um modelo que capta padrões comportamentais típicos de clientes com uma comunicação *event-driven* com o intuito de aumentar a performance das campanhas a promover a App Universo.

A performance da campanha, calculada tendo em conta duas medidas de performance (*uplift* e taxa de conversão), foi superior à performance das comunicações anteriormentes realizadas que tinham o mesmo objetivo e eram baseadas em critérios empíricos. Para além disso, comparou-se a performance da campanha com outras duas em que foram realizadas comunicações *event-driven* e mais uma vez os resultados da campanha baseada no algortimo de SPM foram superiores. Por outro lado, ficou comprovada a aplicabilidade deste tipo de algoritmos na área das comunicações para os clientes visto que o melhor padrão sequencial do algoritmo de SPM apresentou a melhor performance na comunicação. Fica assim evidenciada a supremacia da abordagem adotada, e a relevância do modelo analítico baseado nas sequências de eventos.

Já existem alguns estudos que juntam as áreas do SPM e do marketing, contudo, este é o primeiro trabalho que incluí o *deployment* deste tipo de algoritmos, logo, o estudo realizado pode ser um ativo valioso para quem pretende explorar a aplicabilidade deste tipo de algoritmos nas comunicações com os clientes. Fica assim também demonstrada, a valia de técnicas de *data mining* na área do marketing para serem melhoradas as interações cliente-empresa e adotada uma estratégia mais *customer centric*.

# Abstract

With the proliferation of increasingly distinct customer profiles, it is of crucial importance for companies to adopt a customer-centric posture in order to remain competitive. Adopting such a posture implies the creation of a culture based on the client supported by an organizational structure committed in satisfying the particular needs of each client. One of the functions that should promote this new "way of being" is Marketing, hence the importance of incorporating the ideologies present in the Customer Centric Marketing to establish this customer-centric posture in the organization.

The objective of the current project is the development of an analytical model centered on the client that supports the implementation of a marketing campaign. In recent years, Sonae FS has adopted data mining techniques to select the most prone customers to receive a communication, with the use of propensity models in conjunction with batch communications. The dissertation aims to develop a different analytical model - a model based on a Sequential Pattern Mining (SPM) algorithm that obtains a set of interesting behavioural patterns that will support a communication in which the goal is to increase the number of customers registered in the App Universo. After the model application, communications were carried out for customers who do not have an app and who performed behavioural patterns obtained by the algorithm. Thus, the project adopted a disruptive combination of a model that captures typical behavioral patterns from customers with an event-driven communication in order to increase the performance of campaigns that promote the App Universo.

The campaign performance, calculated taking into account two performance measures (uplift and conversion rate), was higher than the performance of the comunications implemented that had the same objective and were based on empirical criteria. In addition, the performance of the campaign was compared with two other campaigns in which event-driven communications were carried out and once again the results of the campaign based in the sequences of events were better. On the other hand, the applicability of this type of algorithms in communications to the customers was supported, since the best sequential pattern of the SPM algorithm had the best performance in the communication. This highlights the supremacy of the approach adopted, and the relevance of the analytical model based on sequences of events.

There are already some studies that bring together the areas of SPM and marketing, however, this is the first work that includes the deployment of this type of algorithms, so the study carried out can be a valuable asset for those who want to explore the applicability of this type of algorithms in communications with customers. It is also demonstrated the value of data mining techniques in the marketing area to improve customer-company interactions and adopt a more customer-centric strategy.

# Agradecimentos

Queria aproveitar esta secção da dissertação para agradecer a um conjunto de pessoas que ao longo do meu trajeto foram essenciais.

Agradeço a todos os docentes do Mestrado Integrado de Engenharia Eletrotécnica e de Computadores e do Mestrado Integrado em Engenharia e Gestão Industrial, da Faculdade de Engenharia da Universidade do Porto, anos letivos 2015/2020, por contribuírem para o meu crescimento pessoal, profissional e académico.

Um obrigado à Isabel Horta, orientadora na empresa, pelo apoio ao longo do projeto de dissertação. Obrigado pela oportunidade de realização de estágio, orientação, disponibilidade e sugestões que acrescentaram sempre valor para o projeto.

Ao Professor Nuno Soares pela orientação prestada, apoio, e total disponibilidade para atender às minhas solicitações e dúvidas.

À Professora Vera Miguéis por me ter proporcionado esta oportunidade, por estar sempre disponível para me ajudar e por encarar a sua profissão com paixão e profissionalismo.

A todos os elementos da equipa de Customer & Analytics da Sonae FS pelo apoio prestado ao longo do periodo da realização do projeto.

Aos meus pais por todo o apoio, compreensão e orientação ao longo destes 5 anos.

Ao meu irmão gémeo por toda a cumplicidade, carinho e ajuda disponibilizada.

Aos meus familiares e amigos pelo apoio, motivação e preocupação. Sem vocês este caminho teria sido muito mais difícil.

A todos com quem partilhei estes 5 anos de universidade. Momentos que jamais serão esquecidos. Um obrigado especial a todos os meus colegas de curso do MIEGI que me acolheram no 2ºano quando realizei a mudança de curso.

*"In God we trust, all others must bring data."*

W.Edwards Deming

# Conteúdo


**1 Introdução** **1**
- 1.1 A empresa . . . 2
  - 1.1.1 A Sonae FS . . . 2
  - 1.1.2 O mercado . . . 3
- 1.2 Motivação . . . 3
- 1.3 Objetivos do projeto . . . 4
- 1.4 Método seguido no projeto . . . 4
- 1.5 Estrutura da dissertação . . . 5

**2 Revisão de literatura** **7**
- 2.1 *Mobile banking* . . . 7
- 2.2 *Customer Relationhip Management* (CRM) . . . 8
  - 2.2.1 Definição . . . 8
  - 2.2.2 Implementação eficaz do CRM . . . 9
  - 2.2.3 CRM e *data mining* . . . 9
  - 2.2.4 CRM e *business performance* . . . 10
- 2.3 *Customer Centric Marketing* . . . 11
- 2.4 *Marketing Analytics* . . . 13
- 2.5 *Sequential Pattern Mining* . . . 15
  - 2.5.1 Introdução . . . 15
  - 2.5.2 SPM e Marketing . . . 15
- 2.6 Considerações Finais . . . 17

**3 Descrição do Problema** **19**
- 3.1 App Universo . . . 19
- 3.2 Interações que promovem a App . . . 21

**4 Metodologia** **24**
- 4.1 *Sequential Pattern Mining* (SPM) . . . 25
- 4.2 Data Understanding and Preparation . . . 26
  - 4.2.1 *Customer Jourtney* da Sonae FS e Grupos de eventos selecionados . . . 26
  - 4.2.2 Data description and exploration . . . 28
  - 4.2.3 Seleção final dos eventos . . . 30
  - 4.2.4 *Data preparation* . . . 32
- 4.3 *Modelling and evaluation* . . . 33
  - 4.3.1 Medidas de performance . . . 33
  - 4.3.2 Restrições nos algoritmos de SPM . . . 34
  - 4.3.3 Algoritmos de SPM . . . 35
- 4.4 *Deployment* . . . 37

4.4.1 Seleção das subsequências para o envio da comunicação . . . . . 37
4.4.2 Plano do *deployment* . . . . . 38

**5 Resultados** **41**
5.1 Escolha dos valores das restrições e da *confidence* . . . . . 41
5.2 Padrões sequenciais sem gap . . . . . 42
5.3 Padrões sequenciais com gap 60 dias . . . . . 44
5.4 Seleção das sequências para o envio da comunicação . . . . . 44
5.5 Performance da Campanha Realizada . . . . . 47

**6 Conclusão** **51**
6.1 Conclusões do projeto . . . . . 51
6.2 Trabalhos futuros e outras possíveis abordagens . . . . . 52

**A Razões que levam à adoção de *m-banking*** **59**

**B Definições de CRM** **61**

**C Descrição da App** **62**

**D Rácio Eventos/cliente nos diferentes grupos de eventos** **64**

**E Taxas de conversão dos eventos *Call-center*** **65**

**F Taxas de conversão dos eventos *Web-form*** **66**

**G *Boxplots* do *Support* e da *Confidence* sem *gap*** **67**

**H Análise descritiva aos valores de *support* e *Confidence* com *gap*** **68**

**I 10 melhores padrões sequenciais de 2, 5 e 6 elementos sem gap** **69**

**J 10 melhores padrões sequenciais de 2, 5 e 6 elementos com gap de 60 dias** **71**

**K Indicadores das subsequências** **73**

# Abreviaturas e Símbolos

| | |
|---|---|
| ATM | *Automated Teller Machine* |
| BFS | *Breadth-first Search* |
| CCM | *Customer Centric Marketing* |
| CJ | *Customer Journey* |
| CRISP-DM | *Cross Industry Standard Process for Data Mining* |
| CRM | *Customer Relationship Management* |
| DFS | *Depth-first Search* |
| FDM | Fim do Mês |
| PG | *Pattern-Growth* |
| RFM | *Recency, Frequent and Monetary* |
| ROE | *Return on Equity* |
| ROI | *Return on Investment* |
| SPM | *Sequential Pattern Mining* |
| SQL | *Structured Query Language* |

# Lista de Figuras


1.1 Grupo Sonae, adaptado de SONAE SGPS (2020) . . . . . 2
1.2 Cartão Universo-modalidades de pagamento e utilização do cartão, adaptado de Sonae FS (2020) . . . . . 3
1.3 Metodologia CRISP-DM (Ner Clinton, 2000, p.10) . . . . . 5

2.1 Técnicas de Data Mining no CRM (Ngai et al., 2009, p.3) . . . . . 10
2.2 Antecedentes e consequências do CCM, adaptado de Sheth et al. (2000) . . . . . 12
2.3 Níveis de análise do Marketing Analytics, adaptado de Wedel e Kannan (2016) . 14
2.4 Marketing Anaylitics research framework, adaptado de Iacobucci et al. (2019) . . 14

3.1 Número clientes com App Universo vs Número de cartões . . . . . 20
3.2 Evolução da percentagem de clientes com App Universo . . . . . 20
3.3 Processo das campanhas na Sonae FS . . . . . 22

4.1 Metodologia CRISP-DM (Ner Clinton, 2000, p.10) . . . . . 24
4.2 Estrutura de dados adotada no projeto . . . . . 26
4.3 Ilustração da *Customer Journey* da Sonae FS . . . . . 27
4.4 Grupos de variáveis selecionados para o modelo analítico . . . . . 28
4.5 Evolução temporal do número de eventos . . . . . 30
4.6 Taxas de conversão nos eventos transacionais e contratuais . . . . . 31
4.7 Variáveis incluídas no modelo . . . . . 32
4.8 Representação vertical da base de dados . . . . . 36
4.9 Processo de seleção dos padrões sequenciais . . . . . 38
4.10 Plano do *deployment* . . . . . 39

5.1 Distribuições das subsequências ao longo dos valores do *support* e da *confidence* 43
5.2 10 melhores padrões sequenciais de 3 elementos . . . . . 43
5.3 10 melhores padrões sequenciais de 4 elementos . . . . . 44
5.4 10 melhores padrões sequenciais de 3 elementos com *gap* de 60 dias . . . . . 45
5.5 10 melhores padrões sequenciais de 4 elementos com *gap* de 60 dias . . . . . 45
5.6 *Confidence* dos padrões sequenciais de 2 elementos com *gap* de 3 dias . . . . . 47
5.7 Padrões sequenciais selecionados para o envio da comunicação . . . . . 47
5.8 Número de registos após ter sido recebida a comunicação por dia . . . . . 48
5.9 *Dashboard* da campanha baseada nos padrões sequenciais . . . . . 49
5.10 Comparação entre a *confidence* e as outras medidas de performance para cada padrão sequencial . . . . . 50

A.1 Razões que levam à adoção de *m-banking*, adaptado de Shaikh e Karjaluoto (2015) 60

C.1 Funcionalidades da App Universo . . . . . 63

E.1 Taxas de conversão dos eventos *Call-center* . . . . . 65

F.1 Taxas de conversão dos eventos *Web-Form* . . . . . 66

G.1 *Boxplots* do *support* e da *confidence* sem gap . . . . . 67

H.1 Distribuições das sequências ao longo dos valores do *support* e da *confidence* . . 68
H.2 *Boxplots* do *support* e da *confidence* com gap . . . . . 68

I.1 Padrões sequenciais de 2 elementos . . . . . 69
I.2 Padrões sequenciais de 5 elementos . . . . . 70
I.3 Padrões sequenciais de 6 elementos . . . . . 70

J.1 Padrões sequenciais de 2 elementos . . . . . 71
J.2 Padrões sequenciais de 5 elementos . . . . . 72
J.3 Padrões sequenciais de 6 elementos . . . . . 72

K.1 Indicadores das subsequências . . . . . 73

# Lista de Tabelas

4.1 Número de eventos e clientes por tipo de evento . . . . . 29

5.1 Valores das *restrições* e da *confidence* . . . . . 42
5.2 *Uplift* e taxa de adesão das diferentes campanhas . . . . . 50

D.1 Eventos/cliente nos diferentes grupos de eventos . . . . . 64

# Capítulo 1

# Introdução

*"We see our customers as invited guests to a party, and we are the hosts. It's our job every day to make every important aspect of the customer experience a little bit better."* Jeff Bezos

Num mundo cada vez mais competitivo e em constante mudança, é fundamental as empresas adotarem uma postura centrada no consumidor de forma a que se mantenham relevantes. Isto significa que a eficiência, a qualidade do serviço e a potenciação de competências base não são garantia de sucesso em ambientes que mudam muito rapidamente. É necessário também uma abordagem que coloque o cliente no centro de todos os processos do negócio (Bolton, 2004). Apesar disso, de acordo com um estudo da *CMO Council* (CMO, 2018), apenas 14% dos *marketeers* referem que a centralização no cliente faz parte da filosofia da empresa e apenas 11% acreditam que os seus clientes se revêm nessa filosofia. Os maiores desafios para a implementação eficaz desta abordagem são: volume, velocidade e variedade de dados de clientes; sistemas e tecnologias que tenham capacidade de segmentar adequadamente os clientes; a ineficácia operacional para personalizar comunicações e experiências e por último, e não menos importante, a ausência de uma cultura organizacional baseada no cliente (Yohn, 2018).

Aliado a esta perspetiva, surge o conceito de *Customer Centric Marketing* (CCM). No CCM pretende-se identificar de forma exímia as necessidades e vontades de cada consumidor em particular em vez das mesmas de mercados ou segmentos de mercado. O CCM ambiciona mudar a orientação de todas as ações de marketing, passando de uma orientação ao mercado para uma orientação ao cliente (Sheth et al., 2000). Deste modo, existe um foco na criação de relações sustentáveis com os clientes de forma a que seja maximizado o valor destes para a empresa ao longo de toda a sua jornada com a mesma.

Paralelo a essa realidade surge a área da *data mining* que permite através da análise de um grande conjunto de dados extrair conhecimento relevante para a tomada de decisão (Guo, 2017). A aplicação de algoritmos de *data mining* para auxiliar a tomada de decisões nas empresas é uma área em franco crescimento. Na verdade, estima-se que valor do mercado de *Big Data and Business Analytics Solutions* em 2023 seja de 274 mil milhões de dólares (IDC, 2019).

Por sua vez, é cada vez mais frequente a adoção de métodos analíticos para a criação de planos

de marketing eficazes centrados no cliente (Ravi e Sun, 2016). Aliás, muitas das aplicações com maior relevância prática e de maior interesse científico do *data mining* encontram-se na área do marketing (France e Ghose, 2019). Enquadrado neste prisma surge este projeto de dissertação.

## 1.1 A empresa

### 1.1.1 A Sonae FS

O projeto de dissertação foi realizado na Sonae Financial Services (Sonae FS), a qual constitui uma unidade de negócio do grupo Sonae. A Sonae é uma multinacional com sede em Portugal, que gere um portefólio diversificado de negócios nas áreas de retalho, serviços financeiros, tecnologia, centros comerciais e telecomunicações (Figura 1.1). Atualmente, a Sonae é uma das maiores empresas portuguesas, contabilizando um volume de negócios de 6435 milhões de euros (valores de 2019) e empregando perto de 54.000 colaboradores (SONAE SGPS, 2019).

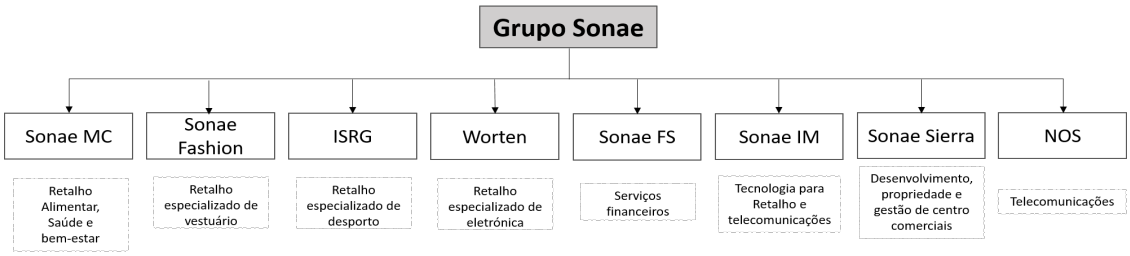



Figura 1.1: Grupo Sonae, adaptado de SONAE SGPS (2020)

Descrever a Sonae FS sem referir o grupo Sonae não seria idóneo tal é a dependência entre esta unidade de negócio e o seu parente. Na verdade, esta relação é tão próxima que o parente não cede apenas capital, intervém também na tomada de decisões estratégicas e operacionais. Para além disso, existem sinergias entre as diferentes unidades do grupo e a Sonae FS.

Fundada em 2015, a Sonae FS é a unidade de negócio do grupo responsável por oferecer soluções financeiras inclusivas e personalizadas a todos os clientes Sonae. O principal produto da Sonae FS é o cartão Universo. Este cartão é um produto único que combina o acesso a contas de débito e de crédito, bem como como a opção por diferentes modalidades escolhidas no momento do pagamento. Para além desta versatilidade no método de pagamento, o cartão Universo permite uma "utilização universal", ou seja, pode ser utilizado num conjunto alargado de serviços financeiros, entre os quais, compras, pagamentos de contas e serviços e transferências. Sob a marca Universo, para além do cartão, a Sonae FS tem outros 2 produtos financeiros: os seguros e o financiamento.

O projeto foi realizado na equipa de Customer & Analytics. A equipa de Customer & Analytics é a equipa da Sonae FS responsável pela interação com o cliente. Muitas das interações cliente-Sonae FS acontecem via comunicações para o cliente e neste âmbito, aquando da seleção de clientes, normalmente, são usados critérios empíricos ou modelos de propensão com o objetivo de serem aumentadas as taxas de conversão. A equipa de Customer & Analytics é uma

equipa transversal a todas as equipas/departamentos da Sonae FS e, interage, portanto com outras equipas/departamentos como por exemplo com a equipa de Product Development-Digital no que concerne comunicações relacionadas com a App Universo.

### 1.1.2 O mercado

A Sonae FS atua como um intermediário nos serviços financeiros realizados por um determinado cliente, ou seja, não é responsável pela cedência de crédito mas sim pela promoção do serviço e interação com o cliente. Neste mercado incluem-se todos os serviços financeiros realizados por um consumidor entre os quais: compras a crédito, transferências, pagamentos de serviços, entre outros (Figura 1.2).

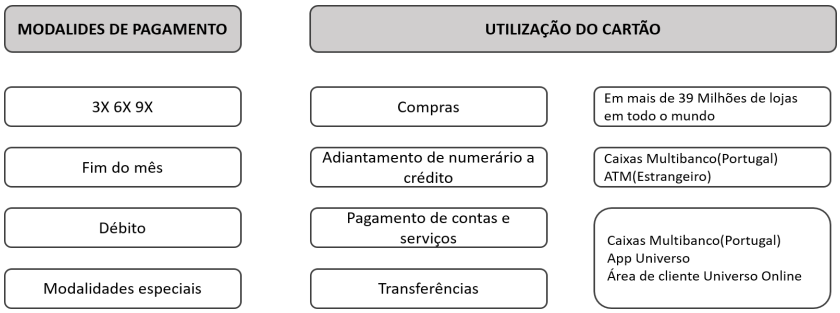



Figura 1.2: Cartão Universo-modalidades de pagamento e utilização do cartão, adaptado de Sonae FS (2020)

A performance deste mercado está altamente correlacionada com o crescimento económico pois o número de transações realizadas depende da confiança e expetativa dos consumidores. Nos últimos anos, verificou-se um aumento no número de transações (Euromonitor, 2019).

Como ao serviço oferecido pela Sonae FS são indexados programas de lealdade, promovidas ofertas convertíveis noutras empresas do grupo e para além disso, não são cobradas *fees* anuais, a empresa tem aumentado o número de cartões e o número de transações realizadas. Na verdade, a Sonae FS já é o nº2 no que diz respeito ao número de cartões emitidos em Portugal. Apesar disso, novos concorrentes com modelos de negócios similares surgiram no mercado o que poderá alterar o ambiente competitivo e, consequentemente, reduzir as margens de todos os incumbentes (Euromonitor, 2019).

## 1.2 Motivação

Depois de apresentado o contexto em que o projeto se insere, é fundamental perceber quais são os *drivers* deste projeto.

O projeto surge da necessidade de melhorar os resultados das interações com os clientes, concretamente, através de realização de uma campanha centrada nas vontades e necessidades destes. Já foi comprovada por Kumar et al. (2008) a relação positiva entre a realização de campanhas centradas no cliente e o aumento do *return on investment* (ROI) e do lucro. Posto isto, um dos

*drivers* do projeto é a melhoria da performance das ações promocionais da Sonae FS através do desenvolvimento de um modelo analítico em que o cliente desempenha o papel de ator principal.

Por outro lado, tem sido comum a adoção de técnicas de *data mining* para melhorar a interação cliente-empresa (France e Ghose, 2019). Uma dessas interações é realização de campanhas e, neste âmbito, é frequente o desenvolvimento de modelos de propensão que permitem identificar quais são os clientes mais propensos a interagirem com a empresa (Linoff e Berry, 2011). Neste projeto de dissertação, avaliar-se-á a exequibilidade de modelos *event-driven* na interação cliente-empresa. Através do cálculo de métricas de avaliação das campanhas pretende-se, com um caso de uso muito específico, perceber até que ponto este tipo de modelos pode ser útil na realização de campanhas que satisfaçam as vontades e necessidades do consumidor.

Concretizando, pretende-se através da realização de uma campanha centrada no cliente, promover a migração dos clientes para o digital com a adesão à App Universo. Esta digitalização encontra-se enquadrada com o plano estratégico do negócio mas também, com uma tendência que tenta reduzir o número de interações cliente-empresa.

Resumindo, pretende-se converter mais clientes para o digital através do envio de uma comunicação baseada num modelo *event-driven* centrado no cliente. Testar-se-á assim a aplicabilidade deste tipo de modelos disruptivos num caso de uso muito específico com o objetivo de melhorar a performance das comunicações da Sonae FS que promovem a conversão digital.

## 1.3 Objetivos do projeto

Tendo em conta os resultados esperados que se pretende atingir com a dissertação surgem os seguintes objetivos: (1) desenvolver um *Data Model System* que agregue todas as variáveis importantes para a realização do projeto e que seja adaptado ao modelo analítico a desenvolver, (2) selecionar e aplicar um algoritmo de *Sequential Pattern Mining* (SPM) para serem encontrados os comportamentos que poderão levar ao registo da app, (3) realizar uma campanha-teste através de uma plataforma de envio de comunicações e (4) monitorizar e avaliar os resultados da campanha através da criação de um *dashboard*.

Todos estes objetivos correspondem a *milestones* do projeto e estão enquadrados com a metodologia aplicada. O modelo analítico desenvolvido foi enquadrado num caso de estudo muito específico (aumentar o número de clientes registados na App Universo), no entanto, com o objetivo futuro de poder ser replicado noutros *use cases* igualmente relevantes para o negócio.

## 1.4 Método seguido no projeto

No projeto de dissertação, foi seguida a metodologia sugerida no *Cross Industry Standard Process for Data Mining* (CRISP-DM) (Ner Clinton, 2000). A metodologia descreve o ciclo de vida de um projeto de *data mining*, contendo as diferentes fases do projeto, as respetivas tarefas e as relações entre estas tarefas. Neste projeto de dissertação, passou-se pelas fases principais desta metodologia (Figura 1.3).

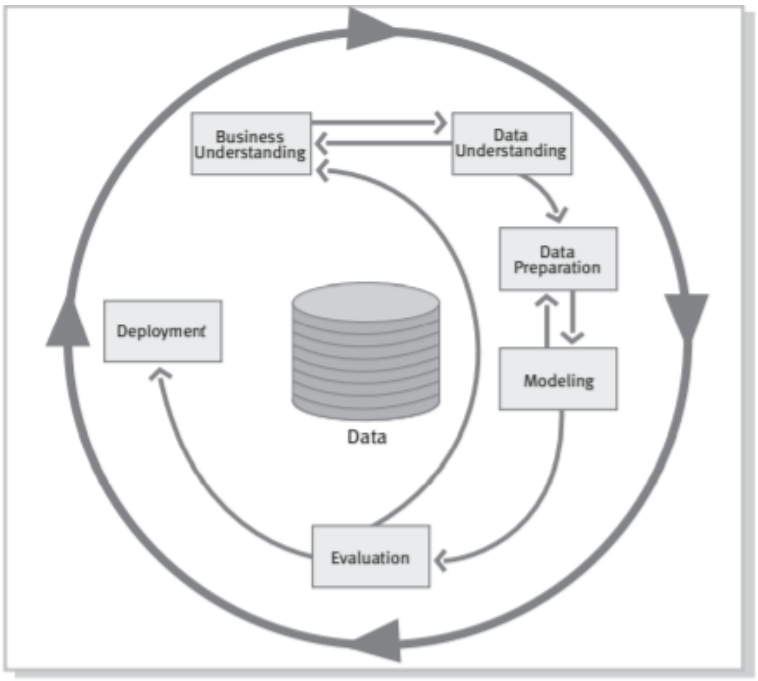



Figura 1.3: Metodologia CRISP-DM (Ner Clinton, 2000, p.10)

Inicialmente, tentou-se perceber o modelo de negócio da Sonae FS, mais concretamente como é que a empresa cria, captura e entrega valor. Após esta fase inicial, os dados foram analisados e foi preparada uma *Tabela Master* que incluí todas as variáveis importantes para a realização do projeto. Seguidamente, foi criado o modelo analítico de *Sequential Pattern Mining* (SPM). Depois de terem sido avaliados os outputs do modelo, efetuou-se o *deployment* deste com a criação de uma campanha baseada no comportamento dos clientes. De salientar também que após esta fase foi avaliada a performance da campanha com a criação de um *dashboard* que monitoriza os resultados desta. Este dashboard avaliará também a eficácia da metodologia seguida (CRISP-DM).

## 1.5 Estrutura da dissertação

A dissertação está organizada em 6 capítulos. No 1º capítulo, é apresentado o projeto de dissertação- incluí os seguintes tópicos: uma apresentação da empresa, da motivação e objetivos do projeto bem como uma descrição sucinta da metodologia seguida. O Capítulo 2 fornece o *background* teórico imprescindível para sustentar as diferentes decisões tomadas ao longo da dissertação. Numa primeira fase, é apresentado o *mobile banking*, depois, a literatura relativa a alguns dos conceitos abordados durante a dissertação tais como: *Customer Relationship Management*, *Customer Centric Marketing* e *Marketing Analytics*. Finalmente, será apresentado o *Sequential Pattern Mining*, a área onde se insere a abordagem analítica adotada no projeto.

De seguida, no Capítulo 3 é apresentado um enquadramento ao problema que se pretende resolver com o projeto. Começa-se com uma apresentação sucinta da App Universo e da sua importância para Sonae FS sendo depois apresentados e avaliados os resultados das interações

passadas que tinham como objetivo promover o registo na App Universo. Seguidamente, no Capítulo 4 descreve-se a metodologia seguida no projeto. Este capítulo é composto pelas diferentes fases da metodologia CRISP-DM. No Capítulo 5, são apresentados os resultados do modelo analítico - os padrões sequenciais que levam com maior probabilidade ao registo da App e os resultados da comunicação realizada baseada nos padrões sequenciais selecionados.

No fim da dissertação (Capítulo 6), conclui-se com uma síntese do trabalho desenvolvido. Apresentam-se as principais conclusões obtidas com o projeto, os contributos do mesmo para o conhecimento e os próximos passos nesta área.

# Capítulo 2

# Revisão de literatura

Este capítulo contextualiza os conceitos mais importantes que serão abordados durante a dissertação. Na revisão da literatura foi adotada uma abordagem em funil, isto quer dizer que se partiu de conceitos gerais para tópicos cada vez mais concretos. O primeiro tópico é o *mobile banking*, no qual se apresentam, entre outros conteúdos, os motivos que levam à adesão a um serviço deste tipo. Seguidamente, será apresentada a literatura relativa a alguns dos conceitos abordados durante a dissertação tais como: *Customer Relationship Management*, *Customer Centric Marketing* e *Marketing Analytics*. Por último, será apresentado o *Sequential Pattern Mining*, a área onde se insere a abordagem analítica adotada no projeto.

## 2.1 *Mobile banking*

De acordo com Shaikh e Karjaluoto (2015), *mobile banking* ou *m-banking* pode ser definido como: "Um produto ou serviço oferecido por um banco ou outro intermediário de serviços financeiros para realização de transações financeiras e não financeiras utilizando um dispositivo móvel, nomeadamente um telemóvel, smartphone, ou tablet" (Shaikh e Karjaluoto, 2015, p.2;tradução do autor). A utilização do serviço *mobile banking* tem sido cada vez mais uma tendência (em Portugal o número de utilizadores deste serviço ronda os 2 milhões 675 mil - 33,7% dos bancarizados, esse número era de 600 mil em 2014) (Marktest 2020; Marktest 2015).

Em Shaikh e Karjaluoto (2015) são apresentadas as razões que levam à adoção de *mobile banking* (Anexo A). Entre os antecedentes mais relevantes encontram-se a facilidade de utilização, a utilidade, a confiança no serviço e a influência social. Num estudo similar realizado por Cristino (2012) em que clientes de bancos portugueses responderam a um questionário sobre os fatores que levam à adesão ao *mobile banking* obtiveram-se como fatores mais preponderantes a vantagem relativa do serviço, a compatibilidade com o estilo de vida e a facilidade de uso.

Fatores demográficos também são apresentados na literatura como relevantes. Alafeef et al. (2012) apresentam a influência destes fatores na adoção de tecnologia, mais concretamente, do *mobile banking*. De acordo com os autores as pessoas do sexo masculino, salário elevado e com

um nível de educação superior tendem a usar mais este tipo de aplicações, já a idade é apresentada como um fator que influencia negativamente a adoção de *m-banking*.

Para além dos fatores demográficos a cultura também é uma variável relevante na adoção de serviços financeiros deste tipo (Hassan e Wood 2020; Baptista e Oliveira 2015). Na verdade, Hassan e Wood (2020) concluíram que a cultura influencia as perceções dos consumidores relativamente ao *mobile banking*. Com o conhecimento detalhado de quais são os fatores mais importantes que levam ao *m-banking*, os *marketers* poderão aumentar a eficácia das interações com os clientes e consequentemente, incrementar as taxas de adoção de *mobile banking* (Shaikh e Karjaluoto, 2015).

Em suma, o *mobile banking* tem sido cada vez mais uma tendência e espera-se um crescimento exponencial no número de utilizadores deste serviço. Os fatores mais preponderantes que levam à adesão a um serviço deste tipo são: a facilidade de utilização, a utilidade, a confiança no serviço e a influência social (Anexo A). Apesar disso fatores demográficos e culturais são também importantes.

## 2.2 *Customer Relationhip Management* (CRM)

### 2.2.1 Definição

Definir CRM não é uma tarefa fácil, isto porque são inúmeras e diversas as definições na literatura sobre este tópico. Payne e Frow (2005) recolheram um conjunto de definições de CRM (Anexo B) e constataram que na prática empresarial o CRM pode ser encarado mediante 3 perspetivas:

- **CRM como uma solução tecnológica particular**: O CRM corresponde à implementação de uma solução tecnológica específica num determinado projeto;
- **CRM como um conjunto diverso de projetos tecnológicos com foco no cliente**: O CRM, nesta perspetiva apresenta-se como um conjunto de soluções tecnológicas aplicadas em projetos diversos, sendo o cliente o centro desses projetos;
- **CRM como uma abordagem holística para gerir eficazmente as relações com o cliente**: Esta é a perspetiva mais estratégica que pretende colocar a relação com o cliente como um dos pontos mais relevantes no modelo de negócio da empresa.

De acordo com os autores, a perspetiva estratégica é aquela que poderá beneficiar mais a empresa isto porque encara-se o CRM não só como uma solução tecnológica para adquirir e reter mais clientes mas também como um *framework* que sustenta a criação de valor na organização. Obteve-se a seguinte definição: "O CRM é uma abordagem estratégica que tem como objetivo criar valor para os acionistas através do desenvolvimento de relações apropriadas com os principais clientes e segmentos de clientes. O CRM une o potencial das estratégias de *relationship marketing* e do IT para criar relações lucrativas e de longo prazo com clientes e outros *stakeholders*" (Payne e Frow, 2005, p.2;tradução do autor).

### 2.2.2 Implementação eficaz do CRM

Farhan et al. (2018) identificaram e categorizaram os fatores críticos de sucesso mais importantes na implementação do CRM. Entre os fatores mais relevantes encontram-se: o suporte da gestão de topo, a cultura organizacional e a capacidade técnica dos colaboradores. Outros aspetos que podem ter impacto no performance do CRM são: a natureza da organização e do produto, a definição clara e concisa dos objetivos organizacionais e a presença de processos automáticos na organização (Mendoza et al., 2007).

Contudo, a aplicação infrutífera desta filosofia tem sido bastante comum - 70% dos projetos de CRM não cumprem com os objetivos propostos (Farhan et al., 2018). De acordo com Kale (2004), existem 7 razões para a implementação ineficaz do CRM: (1) encarar o CRM como uma simples iniciativa tecnológica; (2) a falta de uma visão *customer-centric*; (3) não estruturar a estratégia tendo em conta o *customer lifetime value*; (4) suporte insuficiente da gestão de topo; (5) desvalorização do *change management*; (6) falhas na reengenharia dos processos de negócio e (7) desvalorização da importância da integração de dados e de técnicas de data mining.

Tal como referido em Bohling et al. (2006) é de primordial importância vincular à *marketing strategy* o CRM para que este seja implementado com sucesso. Isto quer dizer que ao atribuir a um nível funcional ou subfuncional a responsabilidade pelo CRM estamos a incorrer num erro que diminuirá a probabilidade de sucesso das iniciativas vinculadas a esta abordagem.

### 2.2.3 CRM e *data mining*

Efraim Turban et al. (2007) definiram o *data mining* como "o processo que utiliza técnicas estatísticas, matemáticas, de inteligência artificial e de *machine-learning* para extrair e identificar informações úteis e, posteriormente, adquirir conhecimentos a partir de grandes bases de dados"(p.305;tradução do autor). De acordo com Rygielski et al. (2002), *data warehousing*, *data mining* e software de gestão de campanhas tornaram o CRM uma área onde as empresas podem ganhar vantagem competitiva.

Na verdade, do ponto de vista arquitetónico o CRM pode ser classificado como operacional e analítico (He et al., 2004). CRM operacional refere-se à automatização dos processos de negócio, enquanto que no CRM analítico analisam-se as características e comportamentos dos clientes de forma a apoiar decisões que impactem a performance organizacional. Como tal, o CRM analítico pode ajudar uma organização a discriminar e alocar mais eficazmente os recursos para o grupo de clientes mais rentáveis. As ferramentas de *data mining* são um meio popular de análise de dados de clientes dentro da estrutura de CRM analítico (Ngai et al., 2009).

Ngai et al. (2009) obtiveram 4 dimensões do CRM que estão relacionadas com o *data mining* (*Customer Identification, Customer Attraction, Customer Retention and Customer Development*) e para além disso, categorizaram 7 funções do *data mining* relevantes para o CRM (Figura 2.1). É da interação entre estas dimensões e funções que surgem os projetos que acoplam as 2 áreas.

A aplicação de algoritmos de *data mining* em CRM é uma tendência emergente no mercado. De facto, analisar e compreender os comportamentos e características dos clientes é a base para

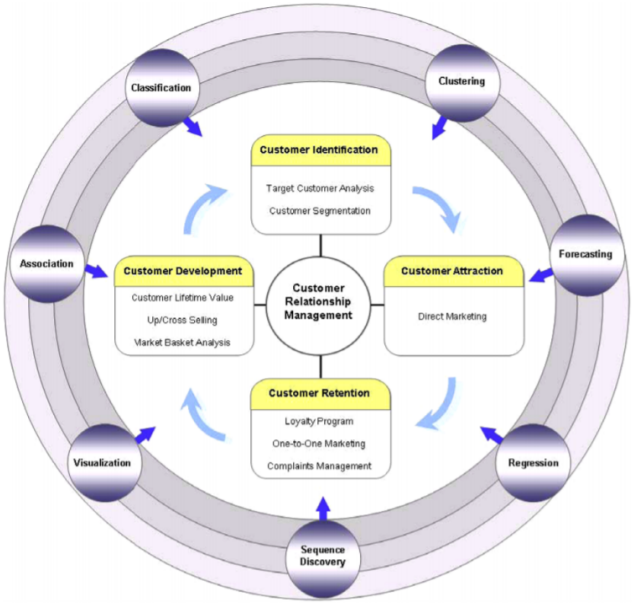



Figura 2.1: Técnicas de Data Mining no CRM (Ngai et al., 2009, p.3)

o desenvolvimento de uma estratégia de CRM competitiva, logo, a aplicação de técnicas de *data mining* no CRM é uma atividade profícua numa estratégia centrada no cliente (Ngai et al., 2009).

### 2.2.4 CRM e *business performance*

Chen e Wu (2016) obtiveram uma associação positiva entre a melhoria de performance e a implementação do CRM. Contudo, Krasnikov et al. (2009) verificou uma relação negativa entre a implementação do CRM e a eficiência ao nível dos custos, apesar disso, esta ineficiência não implicou uma menor rentabilidade para as empresas em estudo.

Jeong et al. (2014) mostraram que o CRM tem um papel fulcral como mediador entre a orientação ao cliente e a performance do negócio. Concluiu-se portanto que para melhorar o desempenho empresarial, não são apenas necessárias orientações estratégicas centradas no cliente , mas também o reforço da capacidade relacional através do CRM . Já Minami e Dawson (2008) obtiveram uma variação positiva do *return on equity* (ROE) após o investimento em CRM, apesar disso, essa relação é mais evidente em empresas que apresentam modelos de negócios centrados no cliente o que mais uma vez corrobora a importância desta orientação na eficácia das iniciativas do CRM.

Por outro lado, ao apresentar a relação positiva entre a adoção do CRM e a performance da empresa podemos estar a incorrer numa falácia pois o CRM pode ser um falhanço se for considerado apenas como uma perspetiva tecnológica ou se for aplicado de forma fragmentada (Payne e Frow, 2005). Para que o CRM propicie a criação de valor na organização deve ser integrado na estratégia corporativa e acoplado às ações de marketing da empresa (Chang et al., 2010).

Resumidamente, o CRM é uma abordagem estratégica que tem como objetivo criar valor para os acionistas através do desenvolvimento de relações com os principais clientes. A relação entre o

CRM e *data mining* tem sido cada vez mais próxima até porque analisar e compreender os comportamentos e características dos clientes é a base para o desenvolvimento de uma estratégia de CRM competitiva. Para que o CRM seja implementado com sucesso e, consequentemente, tenha mais impacto na melhoria da performance do negócio é necessária uma cultura organizacional baseada no cliente e uma ligação muito forte entre o CRM e a estratégia corporativa.

## 2.3 *Customer Centric Marketing*

De acordo com Cundari (2015), atualmente vivemos na *Age of the Customer*. Na prática, isto quer dizer que o cliente tornou-se o ator principal das transações económicas ditando os termos do *engagement* e assumindo um papel de liderança invisível no desenvolvimento de novos produtos. É da necessidade de adaptação a esta nova era que surge o CCM.

O CCM enfatiza a realização e satisfação das necessidades, desejos e vontades dos consumidores e clientes individuais, em vez das mesmas de segmentos de mercados. No CCM, o ponto inicial do processo de planeamento são as necessidades e vontades do cliente individual (Sheth et al., 2000).

Em Sheth et al. (2000) são apresentados os antecedentes e as consequências desta perspetiva. De acordo com os autores o CCM surge da necessidade de aumentar a produtividade do marketing, da necessidade de aplicar eficazmente novas tecnologias para a criação de valor na organização e do aumento da diversidade dos mercados.

De facto cada vez são mais comuns os relatos da inépcia da função do marketing na criação de valor para a organização - o CCM tenta aumentar a eficácia e a eficiência de todas as atividades desta função (Nwokah 2009; Sheth e Sisodia 2002). Por outro lado, o aumento da diversidade dos mercados através da proliferação de perfis de clientes cada vez mais distintos só pode ser endereçado com uma abordagem centrada no cliente (Shah et al., 2006). Finalmente, o crescimento exponencial de novas e poderosas tecnologias pode ser usado para serem encontradas de forma mais profícua as necessidades dos clientes (Figura 2.2) (Sheth et al., 2000).

O CCM terá impacto principalmente na função de marketing mas também um pouco por toda a estrutura organizacional. Com a adoção desta perspetiva espera-se que a empresa adote técnicas de *supply-chain management* para responder eficazmente às necessidades dos clientes, para além disso, a empresa terá tendência para realizar *outsourcing* de clientes que não são rentáveis através de parcerias com outras empresas. Com o CCM espera-se também que os clientes participem ativamente no design, desenvolvimento e produção do produto (Kumar et al., 2008). Outras consequências do CCM são o aumento da percentagem de custos fixos nos custos de Marketing e o desenvolvimento de uma organização com foco no cliente (Figura 2.2) (Sheth et al., 2000).

Muitas empresas ainda estão a lutar para se alinharem totalmente com este paradigma que coloca o cliente no centro de todas as ações de Marketing. De acordo com Shah et al. (2006), os principais desafios que as empresas enfrentam para se tornarem *customer-centric* são: (1) a cultura organizacional, (2) estrutura e processos organizacionais e (3) medidas financeiras inadequadas. Os autores propõem um *roadmap* que pretende potenciar a centralidade no cliente na organização.

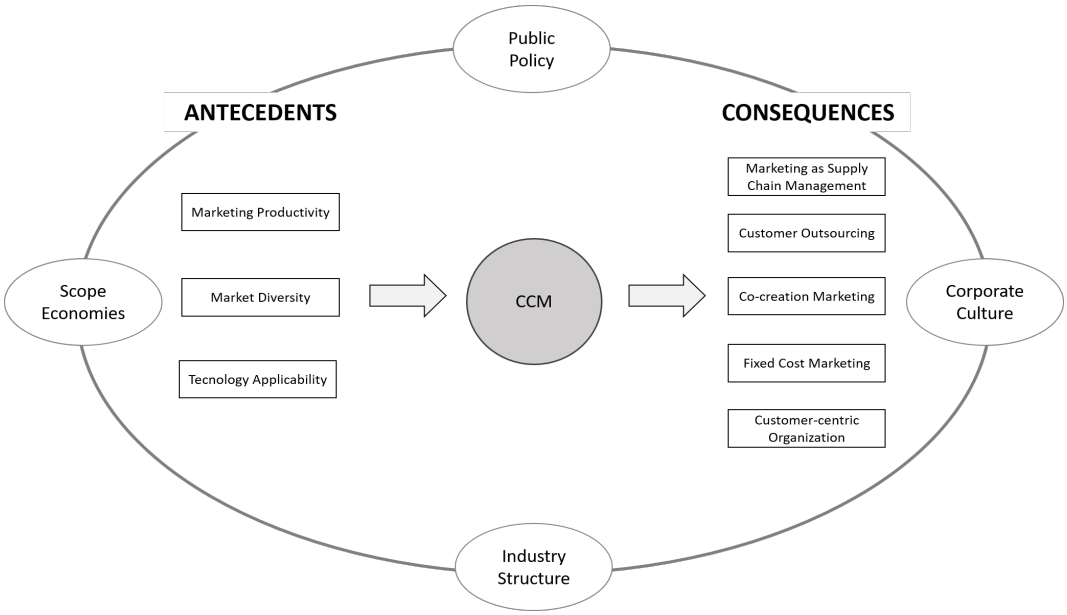



Figura 2.2: Antecedentes e consequências do CCM, adaptado de Sheth et al. (2000)

Para além disso, alertam para a necessidade de uma liderança comprometida com as vontades e necessidades dos clientes, identificam também como preponderante a criação de uma organização mais horizontal (menos hierárquica) que acople a presença de processos horizontais iniciados pelo cliente e sugerem ainda a adoção de medidas de performance *customer-centric* tais como lealdade e satisfação do cliente.

Para que o CCM seja aplicado convenientemente nas empresas Sheth e Sisodia (2012) criaram um *framework* (4A *framework*) que ajuda as empresas a evitar a *marketing myopia* (um foco excessivo no produto) e também *managerial myopia* (um foco excessivo no processo). Com este framework realiza-se a transição do *segmented-marketing* (4P's) para o *Customer Centric Marketing* (4 A's) (Dadzie et al., 2019). O *Framework* é constituído por 4 dimensões:

- ***Acceptability***: Em que medida o produto oferecido cumpre com ou excede as espectativas do cliente. Tem 2 dimensões: *functional acceptability* (performance objetiva dos atributos do produto) e *psycological acceptability* (performance subjetiva dos atributos do produto);

- ***Affordability***: Em que medida o cliente-alvo tem capacidade ou está disposto a pagar o preço pedido pelo produto. Tem 2 dimensões: a capacidade e a disponibilidade para pagar (a capacidade para pagar está relacionada com a disponibilidade económica para pagar pelo produto enquanto que a disponibilidade para pagar é determinada pela perceção de valor que o cliente obtém com o produto relativamente ao custo deste-*value for money*);

- ***Accessibility***: Em que medida os clientes conseguem facilmente adquirir e usar o produto. Tem 2 dimensões: disponibilidade (avalia a capacidade da empresa oferecer o produto tendo em conta a procura pelo mesmo) e a conveniência (quão fácil é para o cliente potencial adquirir e usar o produto);

- ***Awareness***: Em que medida os clientes são informados sobre as características do produto, persuadidos a experimentá-lo e, se aplicável, lembrados de o recomprar. Tem 2 dimensões: *brand awareness* (reconhecimento da marca pelos clientes potenciais) e *product knowledge* (conhecimento do produto por parte do cliente).

Apesar de o CCM ser reconhecido pela literatura como uma ideologia importante, existem autores, tal como Gummesson (2008), que acham a prática desta perspetiva ineficaz pois, apesar de o cliente ser relevante, em certos momentos outras áreas (como por exemplo gestão de operações e finanças) poderão ser prioritárias. Sendo assim, a adoção desta prática nas empresas não é consensual.

Em suma, o CCM tenta satisfazer as necessidades, desejos e vontades dos consumidores e clientes individuais, em vez das mesmas de segmentos de mercados. O CCM surge da necessidade de aumentar a produtividade do marketing, de aplicar eficazmente novas tecnologias e do aumento da diversidade dos mercados e impactará de diversas formas a performance organizacional. De facto, prevê-se que com a adoção do CCM aumente a percentagem dos custos fixos na área do marketing, se desenvolva uma organização com mais foco no cliente, sejam adotadas técnicas de *supply-chain management* para responder eficazmente às necessidades dos clientes, seja realizado o *outsourcing* de clientes não rentáveis e o *Co-creation marketing* seja uma realidade. Com o CCM o foco será o cliente por isso conceitos como *Acceptability*, *Affordability*, *Accessibility* e *Awareness* terão de ser incorporados na organização.

## 2.4 *Marketing Analytics*

*Marketing Analytics* é uma área do marketing que através da análise de uma grande quantidade de dados e da aplicação de ferramentas de modelação permite apoiar a tomada de decisão (Iacobucci et al., 2019). Esta área tem ganho cada vez mais importância principalmente, pelo facto de proporcionar vantagem competitiva sustentável para todas as empresas que a colocam como um dos pilares da organização. Apesar disso, muitas decisões na área do marketing não são apoiadas por modelos analíticos ou pela análise de uma grande quantidade de dados, verificando-se esta realidade, principalmente nas pequenas e médias empresas (Hanssens e Pauwels, 2016).

Um projeto na área de *marketing analytics* pode ter 4 níveis de análise (Figura 2.3) (Wedel e Kannan, 2016). A análise descritiva é na sua índole exploratória e permite a apresentação sumária do conjunto de dados analisado. Na análise de diagnóstico tenta-se obter relações entre as variáveis através de *explanatory models*. No 3º nível de análise é comum o desenvolvimento de modelos preditivos que permitem obter uma previsão do valor das variáveis em estudo. No último nível de análise (prescritivo) pretende-se apoiar a tomada de decisão com modelos de otimização usados para determinar níveis ótimos de variáveis preponderantes para a tomada de decisão.

De acordo com Wedel e Kannan (2016), níveis superiores apresentam maior complexidade e mais custos de implementação, contudo, acrescentam mais valor para a tomada de decisão. Nos diferentes níveis podem ser aplicadas diferentes técnicas ou modelos. Na verdade, esta área necessita de uma abordagem holística, uma combinação de técnicas idealmente com diferentes

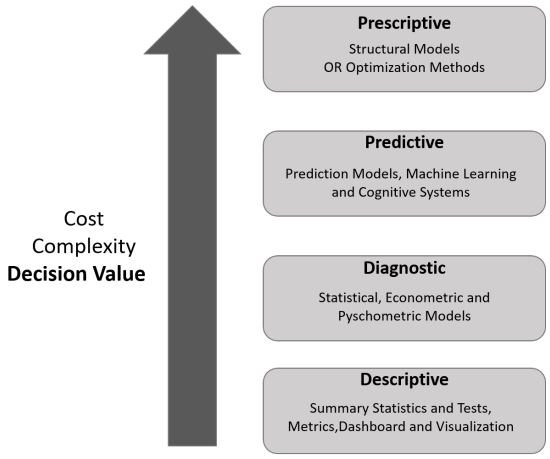



Figura 2.3: Níveis de análise do Marketing Analytics, adaptado de Wedel e Kannan (2016)

tipos de dados que vão de encontro ao perfil da empresa e que são interpretadas conjuntamente (Iacobucci et al., 2019).

O *Marketing Analytics* tem sido um suporte para a tomada de decisão mais informada nas diferentes subáreas do Marketing tal como comportamento do consumidor, *marketing strategy* e inovação. Para além disso, nos diferentes estudos explorados por Iacobucci et al. (2019), a aplicação proficiente de modelos analíticos permitiu por exemplo, aumentar o ROI (return on investment), maximizar a experiência do cliente e melhorar a capacidade preditiva da organização. Já as técnicas mais comuns foram a utilização de algoritmos de *data mining* e de modelos de regressão. Na Figura 2.4 encontram-se as diferentes subáreas, resultados e métodos desta área.

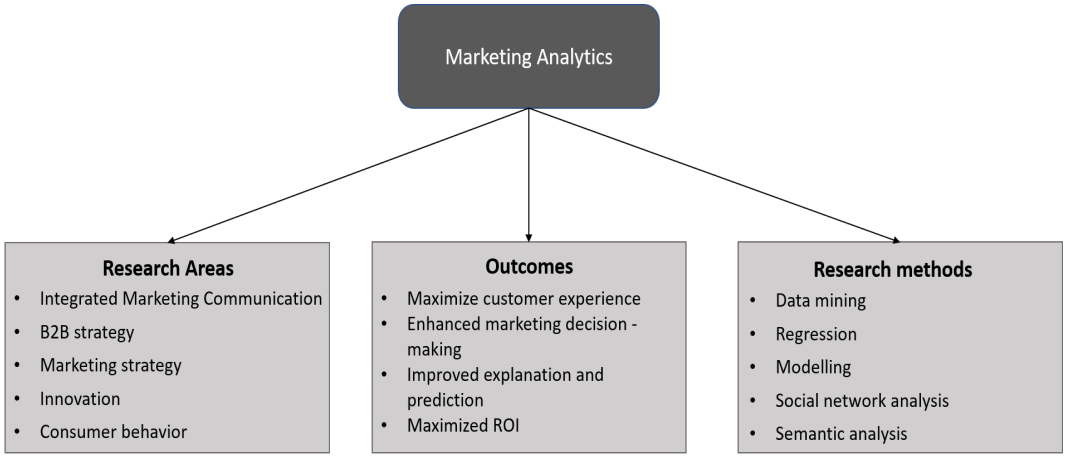



Figura 2.4: Marketing Anaylitics research framework, adaptado de Iacobucci et al. (2019)

O *Marketing Analytics* poderá melhorar a performance organizacional e consequentemente, ser fonte de vantagem competitiva se a sua integração na organização for adequada. Germann et al. (2013) concluíram que quanto mais intensa for a competição no mercado em questão e quanto mais rapidamente se alterarem as preferências do consumidor maior o impacto do *deployment* do *Marketing Analytics*. A análise realizada permitiu constatar também que o suporte da

gestão de topo, uma cultura que suporte o *Marketing Analytics*, dados apropriados, suporte do IT e competências analíticas na organização são "ingredientes" necessários para um *deployment* do *Marketing Analytics* eficaz.

Segundo Hauser (2007), com esta área pretende-se "mapear o DNA do cliente". Como um organismo vivo, este DNA continuará a evoluir e a adaptar-se ao seu ambiente ao longo de toda a vida do cliente. Os *marketeers* terão de usar eficazmente os *outputs* desta área para relacionarem-se com o cliente de forma exímia.

## 2.5 *Sequential Pattern Mining*

### 2.5.1 Introdução

A descoberta de padrões inesperados e úteis em bases de dados é uma tarefa fundamental na área do *data mining*. Nos últimos anos, uma tendência tem sido o desenho de algoritmos para descobrir padrões em dados sequenciais. Foi assim que surgiu o problema de *Sequential Pattern Mining* (SPM), proposto por Agrawal e Srikant (1995) e que pode ser definido como o problema de encontrar subsequências interessantes num conjunto de sequências. Embora tenha sido originalmente concebido para ser aplicado a sequências, também pode ser aplicado a séries temporais após a conversão de séries temporais em sequências utilizando técnicas de discretização (Fournier-Viger et al., 2017).

A área de *pattern mining* tem-se tornado muito popular devido às suas inúmeras aplicações em problemas reais. Entre as várias tipologias de *pattern mining* encontram-se o *association rule mining*, o *frequent itemset mining* e o SPM. O objetivo do *frequent itemset mining* é descobrir conjuntos de itens frequentes numa base de dados transacional. Já no *association rule mining* pretende-se encontrar regras de associação. O SPM distingue-se das restantes pelo facto de poder ser aplicado a dados em que se têm em conta as relações sequenciais entre eventos, não se ignorando portanto a variável tempo. Nestes últimos anos, foram desenvolvidas muitas soluções em SPM que procuram responder a problemas em diversos domínios tais como bioinformática (Wang et al., 2007), *e-learning* (Fournier-Viger et al., 2008), redução de energia em smarthomes (Schweizer et al., 2016) e *webpage click-stream analysis* (Fournier-Viger et al., 2012).

O interesse crescente nesta área está muito relacionado com o facto de se descobrirem padrões "escondidos" e inesperados em grandes conjuntos de dados que podem levar a tomadas de decisões mais informadas. Por exemplo, o padrão {leite, chocolate e bolachas, pão} pode ser usado para uma empresa de retalho tomar decisões operacionais mais eficazes e consequentemente, aumentar as vendas (Fournier-Viger et al., 2017).

### 2.5.2 SPM e Marketing

Os algoritmos de SPM não foram desenhados para resolver exclusivamente problemas de marketing, aliás, Meyer (2003) alertou para a necessidade de uma maior associação entre estas duas

áreas. Esta associação poderia ser realizada, por exemplo, com a criação de algoritmos com estruturas de dados que têm em conta diferentes tipos de eventos. No seu artigo, Meyer (2003) identificou como primordial a realização de estudos que acoplem uma área puramente técnica em que se tem desenvolvido algoritmos cada vez mais eficientes (SPM) com uma outra mais teórica em que se tentam criar relações de causa-efeito entre diferentes variáveis (Marketing). Só deste modo se poderá potenciar a criação de valor com estes algoritmos.

Diversos autores têm tentado explorar esta lacuna com a realização de estudos que liguem as duas áreas. Por exemplo, Desai e Ganatra (2015) utilizaram o algoritmo de SPM para recomendar a próxima compra de um determinado consumidor. O trabalho realizado focou-se na previsão de produtos que um determinado cliente tem maior probabilidade de comprar sabendo qual tem sido o comportamento desse consumidor. Para obter um sistema de recomendação fiável, Desai e Ganatra (2015) incorporaram medidas que têm em conta quão frequente é uma subsequência e que medem a probabilidade de uma determinada compra ocorrer sabendo o comportamento passado do cliente. Para além da inclusão destas medidas, teve-se em conta o fator tempo não só na consideração de um algoritmo de SPM que permite analisar as relações sequenciais entre eventos mas também com a inclusão de uma restrição de tempo que permite ter em conta apenas sequências que têm entre itens consecutivos um determinado espaço de tempo. Conseguiu-se desta forma recomendar sequências de curto e longo prazo. A adoção de um sistema deste tipo permite aumentar a probabilidade de compra de um determinado produto e, consequentemente, a eficácia das interações com os clientes (Desai e Ganatra, 2015).

Ben Baccar et al. (2019) utilizaram *sequential data-mining methods* para prever o movimento dos turistas com dados do Instagram. O trabalho desenvolvido permitiu determinar comportamentos comuns de um conjunto de turistas. Sabendo que o comportamento dos turistas é muito dependente da idade, país e classe social, ao aplicar este tipo de algoritmos a um conjunto de clientes que partilhem uma destas características podem-se obter padrões/sequências interessantes e assim customizar a oferta para esse conjunto de clientes. Algoritmos deste tipo podem-se tornar ativos valiosos para os atores deste ramo.

Chen et al. (2009) incorporaram o conceito de *Recency, Frequent and Monetary* (RFM) presente na literatura de marketing para definir *RFM sequential patterns* e desenvolver um algoritmo que gera este tipo de sequências a partir de dados de compras de consumidores. Considera-se sequência RFM- uma sequência recente, frequente e que envolva elevados valores monetários. Chen et al. (2009) demonstraram empiricamente os benefícios de adotar uma abordagem deste tipo quando se analisam dados transacionais de clientes. Na verdade, os algoritmos comuns de SPM consideram apenas a dimensão frequência, consequentemente, ao incorporar as outras duas dimensões nos padrões sequenciais, as empresas podem focar-se nas sequências de compras mais importantes. Para além disso, Chen et al. (2009) segmentaram as sequências tendo em conta os valores apresentados em cada uma das três dimensões. Com esta segmentação é possível identificar padrões que necessitam de especial atenção ou ação. Por exemplo, uma empresa pode encontrar padrões com grande valor monetário que apareceram recentemente. Nesse caso a empresa pode aplicar mais recursos na promoção dos produtos associados a esses padrões. De acordo com os au-

tores uma abordagem deste género pode ter muito impacto principalmente nas seguintes áreas do *e-commerce*: *cross-selling*, recomendação do produto, *e-catalog design*, marketing personalizado e *product bundle design* (Chen et al., 2009).

Sabendo que hoje em dia é possível registar cada interação com o consumidor na base de dados empresarial, o *marketeer* pode criar campanhas de marketing muito mais eficazes com a adoção de algoritmos de SPM. Foi essa abordagem que Sinha et al. (2015) implementaram através da identificação de sequências que podem levar à concretização de um objetivo de marketing. Sinha et al. (2015) aplicaram um algoritmo de SPM numa base de dados com interações cliente-empresa para obter as sequências de clientes com maior propensão em receber uma campanha de marketing com um objetivo definido.

Para além disso, Sinha et al. (2015) desenvolveram uma ferramenta que identifica as sequências ótimas para um determinado objetivo de marketing. Com esta ferramenta um *marketeer* facilmente cria uma campanha de marketing com um determinado objetivo. Este estudo revelou padrões ótimos bastante interessantes que pretendem aumentar a eficácia das comunicações de marketing, apesar disso é claro que existe uma necessidade de novos estudos que mostrem qual o impacto que este tipo de abordagem pode ter na melhoria dos resultados das campanhas de marketing.

Em suma, com o SPM pretende-se descobrir padrões inesperados em base de dados sequenciais. Os algoritmos de SPM não foram desenhados para resolver exclusivamente problemas de marketing, contudo, já existem alguns estudos que acoplam as duas áreas. Mesmo assim, é óbvio que a relação SPM-Marketing ainda não foi totalmente explorada, notando-se uma escassez de estudos em que tenham sido utilizados algoritmos de SPM para lançar campanhas de marketing.

## 2.6 Considerações Finais

Os conceitos explorados ao longo da revisão da literatura foram essenciais para desenvolver uma solução pertinente para o problema que a presente dissertação pretende resolver. Começando com o *mobile banking*, no presente capítulo exploraram-se os diferentes fatores que levam à adesão a um serviço deste tipo. Aquando da inclusão dos eventos no modelo teve-se em conta estes fatores/motivos pois pretende-se aumentar o número de clientes registados na App Universo.

Relativamente ao CRM e ao CCM, foram essenciais na medida em que se pretende realizar um modelo analítico que seja centrado nas necessidades e vontades do cliente, logo, conhecer em detalhe a relação cliente-empresa através do CRM e explorar quais são os antecedentes e consequências do CCM bem como quais são os fatores que definem o sucesso/insucesso desta perspetiva foi bastante importante para a realização de um estudo mais profícuo. A incorporação destes tópicos trouxe mais profundidade e relevância ao trabalho desenvolvido.

A área do *Marketing Analytics* permitiu uma escolha mais assertiva do método adotado no projeto. Com o estudo dos diferentes níveis de análise e a pesquisa dos diferentes estudos nesta área conseguiu-se definir claramente qual era a abordagem mais adequada para o problema que se pretendia resolver com o projeto.

Por fim, no presente capítulo foram também explorados diferentes estudos com problemas semelhantes que utilizavam algoritmos de SPM (a abordagem analítica adotada no projeto). Foi a partir desta exploração que se definiram quais os algoritmos, medidas de performance e restrições de SPM mais adequados para o projeto de dissertação.

# Capítulo 3

# Descrição do Problema

## 3.1 App Universo

Os modelos de negócio das empresas tiveram de adaptar-se a um mundo cada vez mais digital. As empresas de serviços financeiros não foram uma exceção neste âmbito e por isso mesmo, tiveram de alterar a sua postura de modo a melhorar a experiência dos clientes e de não perder pontos de contacto com estes nas operações do dia-a-dia. Inicialmente houve alguma relutância, por questões de segurança e de perda de relação com o cliente, contudo verificou-se que o cliente valoriza a conveniência e disponibilidade de este tipo de ofertas mais digitais. Deste modo, a transformação digital foi ocorrendo gradualmente- primeiro as ATM, depois a banca telefónica, *home banking* e mais recentemente o *mobile banking*.

A Sonae FS passou também por este processo gradual de transformação digital e desde a sua criação foi introduzindo ofertas enquadradas com este novo paradigma. À possibilidade de o cliente poder realizar um conjunto de serviços financeiros a partir de uma ATM, seguiu-se o *home banking* que permite consultar em tempo real o saldo, movimentos e descontos acumulados e também, realizar pagamentos de serviços e transferências. O *mobile banking* foi o próximo passo. Criada em Março de 2017, a App Universo permite por exemplo realizar pagamentos de serviços e transferências, criar cartões virtuais para realizar compras, consultar o saldo do cartão Universo e consultar o histórico do extrato (Anexo C).

Desde que a App foi criada em Março de 2017, o número de clientes Universo com App tem aumentado substancialmente. Na verdade, o número de clientes que já se registaram na App Universo tem crescido a uma média de 5%/mês desde Janeiro de 2019 (Figura 3.1). Através da análise desta figura conseguimos perceber que a evolução positiva deste indicador pode estar relacionada com o aumento do número de cartões que apesar de não ter sido tão expressivo como o aumento do número de registados pode impactar no resultado deste. Por outro lado, a adesão global a este tipo de serviços tem crescido exponencialmente tal como é constatado nos relatórios da *Marktest* (Marktest (2020),Marktest (2015)), ou seja, o crescimento elevado no número de utilizadores da App Universo pode ser fruto desta tendência digital global.

Mas mais importante do que perceber como tem sido a evolução do número de clientes com

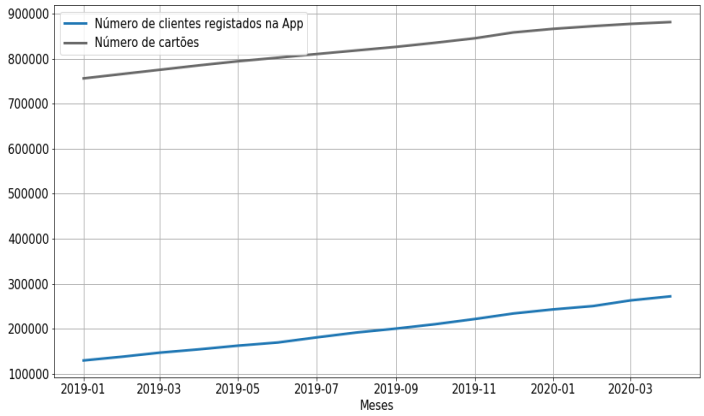



Figura 3.1: Número clientes com App Universo vs Número de cartões

App Universo, é necessário avaliar a percentagem de clientes Universo com app para se perceber se de facto existe potencial para aumentar o número de pessoas aderentes ao serviço através de uma comunicação. Essa percentagem tem crescido nos últimos meses tal como constatado na Figura 3.2, contudo, é de salientar o facto da percentagem de clientes com App Universo ser inferior à percentagem de clientes com *m-banking* em Portugal (34%) o que por si só é um indicador de que existe um conjunto de clientes Universo que não têm app e estariam propensos a tê-la visto que os serviços financeiros realizados com a App Universo são similares aos realizados com outras apps *m-banking* (Marktest, 2020).

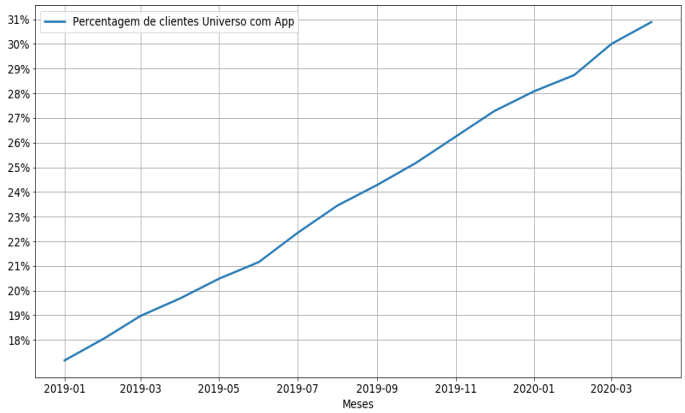



Figura 3.2: Evolução da percentagem de clientes com App Universo

Um cliente Universo ativo beneficiará com a App Universo em diversos contextos. Por exemplo, com a app consegue saber o seu saldo em cartão Universo e em cartão Continente, consultar o valor a pagar no extrato seguinte, e obter o histórico do extrato. Para além disso, consegue realizar pagamentos de serviço e ao estado, enviar/pedir dinheiro a outros clientes Universo ou alterar %

pagamento fim do mês (Anexo C). Muitas destas ações teriam de ser realizadas numa *Automated Teller Machine* (ATM) (os pagamentos e envio de dinheiro) ou com uma chamada ao *call-center* o que por vezes não é cómodo, prático e, para além disso envolve custos. Portanto, é óbvio que a Sonae FS ao tentar aumentar o número de clientes com App está a agir numa perspetiva centrada nas necessidades do cliente (*customer centric*) até porque o canal App é totalmente gratuito e existe a opção de o cliente realizar estas ações via outro canal se assim o preferir.

Por outro lado, para a Sonae FS um cliente digital é um cliente economicamente mais interessante do que um cliente não digital porque muitas das ações realizadas via *call-center* envolvem consumo de recursos que podem ser evitados se as mesmas forem realizadas via app (consultar Saldo, alterar % pagamento fim do mês, efetuar transferências, alterar a modalidade de pagamentos de compras já realizadas, entre outras). Tome-se como exemplo as chamadas *inbound* para alterar o pin (algo que consegue ser feito com a App), o número destas chamadas de clientes com App é muito inferior à dos clientes sem App o que por si só é uma prova de que existe um potencial de poupança com esta transformação digital. Para além disso, um cliente digital receberá o extrato via e-mail o que é economicamente mais viável do que o canal carta. Em suma, estima-se que cada cliente digital permite uma poupança total de 6,10€/ano (Equação 3.1)

$$Poupança\,total\ por\ cliente/ano = Poupança\,extrato\ por\ cliente/ano + Poupança\,call\text{-}center\ por\ cliente/ano \tag{3.1}$$

Apesar de todas estas vantagens económicas terão de se ter em conta os custos de desenvolvimento e manutenção da App que poderão ser elevados, mas que terão menos impacto na rentabilidade se o número de clientes registados na App aumentar. Em suma, prevê-se que para além de uma necessidade estratégica que advém do facto de a empresa pretender ser *customer centric* em todas as suas ações, existirão possíveis benefícios financeiros para a Sonae FS se um cliente aderir e utilizar frequentemente a App.

## 3.2 Interações que promovem a App

Como o *m-banking* é um dos pilares estratégicos da empresa, a Sonae FS tem interagido com os clientes de forma a que estes se registem na App Universo e sejam utilizadores deste serviço. Na verdade, em 2019, a Sonae FS realizou diversas comunicações para clientes não digitais a promover o registo na App.

Para avaliar a eficácia das campanhas a Sonae FS utiliza duas medidas: a percentagem de clientes que aderiram à campanha (taxa de adesão ou taxa de conversão) e o *uplift* (ou *lift*). Estas medidas são resultado de um processo bem definido em que os clientes são separados em 2 grupos (alvo e controlo) sendo que apenas são contactados os clientes do grupo alvo (Figura 3.3). A percentagem de clientes que aderiram à campanha não é mais do que a percentagem de clientes que responderam positivamente à campanha (3.2). Já o *uplift* é uma medida incremental que tenta avaliar quão eficaz foi a campanha (Lo, 2002). Na prática, é a diferença entre a percentagem de

clientes que aderiram à campanha no grupo alvo e essa mesma percentagem no grupo de controlo (3.3). Em baixo apresentam-se o processo de comunicação de campanhas da Sonae FS e as fórmulas para o cálculo do *uplift* e da percentagem de clientes que aderiram à campanha.




Figura 3.3: Processo das campanhas na Sonae FS

$$Taxa\ de\ adesão\ = \frac{Clientes\ que\ aderiram}{Total\ de\ clientes} \tag{3.2}$$

$$Uplift = Taxa\ de\ adesão\ do\ grupo\ alvo\ -\ Taxa\ de\ adesão\ do\ grupo\ de\ controlo \tag{3.3}$$

Em 2019, a Sonae FS realizou 7 campanhas mensais a promover a App Universo. Estas comunicações tinham como objetivo converter para o digital clientes pertencentes a segmentos específicos. As comunicações foram por *batch*, ou seja, selecionaram-se os clientes numa base mensal e depois uma mensagem informativa a promover a App foi enviada para cada um destes clientes selecionados. Os segmentos comunicados foram sendo alterados ao longo do ano quando a performance da comunicação do mês não era satisfatória.

Relativamente, aos resultados das comunicações é importante referir que os valores da taxa de adesão e do *uplift* não foram satisfatórios, isto porque noutras campanhas com mensagens informativas que pretendiam levar o cliente a realizar uma determinada ação (transação, ativação do cartão, transferências) foram obtidos valores superiores nestas duas medidas de performance.

De salientar também que nesta campanha definiram-se os clientes-alvo através de critérios empíricos. Esta seleção pode explicar os valores de *uplift* e da taxa de adesão. Tal como referido em Lo (2002) o uso de técnicas mais avançadas, tais como algoritmos de *data mining*, para encontrar os clientes mais propensos a receber uma determinada campanha tem maximizado o ROI através da melhoria da performance das comunicações. Deste modo, existe a necessidade de aplicar uma

abordagem deste género na Sonae FS para se melhorar a performance das campanhas a promover a app e consequentemente, converter mais clientes para o digital.

# Capítulo 4

# Metodologia

O objetivo do projeto é desenvolver um modelo analítico que auxilie uma comunicação de marketing, mais concretamente, que determine quais os clientes a contactar e em que momento contactá-los de forma a que se aumentem as taxas de adesão das campanhas a promover a App Universo. Deste modo, ao longo do projeto seguiu-se a metodologia CRISP-DM (Ner Clinton, 2000)- metodologia mais utilizada nos projetos de *data mining* pois pretende-se aplicar uma técnica de *data mining* para resolver um problema do negócio. Posto isto, neste capítulo começa-se com uma breve apresentação do *Sequential Pattern Mining* (SPM) e seguidamente, cobrem-se todas as fases da metodologia CRISP-DM (*Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation* e *Deployment*) (Figura 4.1).

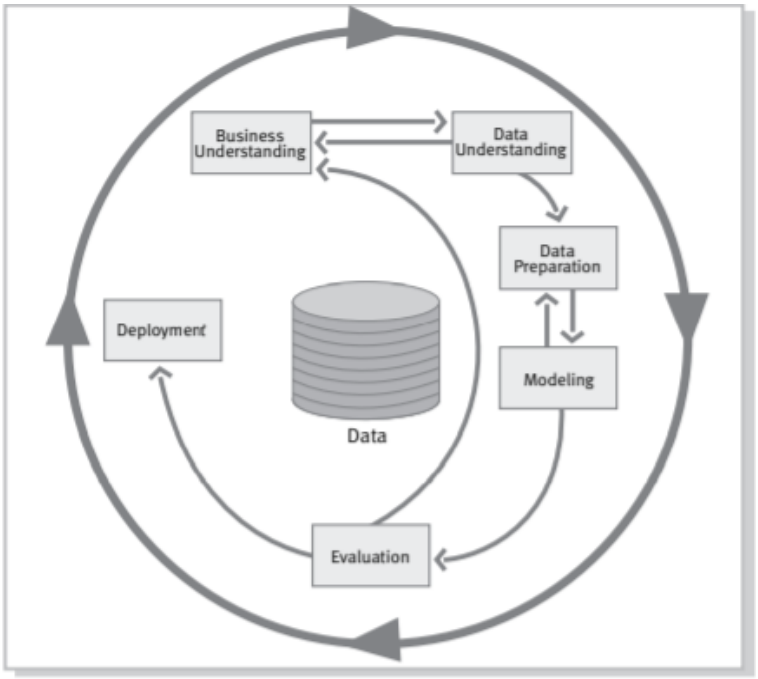



Figura 4.1: Metodologia CRISP-DM (Ner Clinton, 2000, p.10)

## 4.1 *Sequential Pattern Mining* (SPM)

Os algoritmos de SPM têm como principal objetivo encontrar padrões "escondidos" num conjunto de eventos sequenciais. Estes algoritmos procuram estes padrões sequenciais interessantes e inesperados (subsequências) numa base de dados com características e estrutura própria. No projeto, foram aplicados algoritmos deste tipo para serem encontrados padrões comportamentais que auxiliem a realização de uma campanha centrada nas vontades e necessidades do cliente.

Os elementos com maior granularidade nas bases de dados que servem de suporte aos algoritmos de SPM são os *items*. Consideremos então C={$I_1$,$I_2$...,$I_m$} um conjunto de *m* elementos distintos chamados *items*. Em C poderão existir *itemsets* definidos como conjuntos de *items*, sendo que a um conjunto de m elementos chama-se *m-itemset*. Uma sequência não é mais do que uma lista ordenada de *itemsets*, logo uma sequência S com comprimento l é definida como < $E_1$,$E_2$,...,$E_l$>, e cada elemento da sequência S ($E_i$) é um *itemset*. Assim, $E_i$ é um *itemset* que representa um conjunto de elementos que ocorrem no mesmo instante de tempo, logo a ordem desses elementos (*items*) dentro do *itemset* não é relevante (Fournier-Viger et al., 2017).

No caso concreto do projeto houve necessidade de categorizar alguns eventos. Definiu-se que cada evento representava um *itemset* e cada categoria um *item*. Consequentemente, se um evento tem várias categorias temos um *itemset* com vários *items*, por exemplo, uma compra no Continente na modalidade Fim do Mês (FDM) é um *itemset* com 2 *items* (<{Transação Continente, Transação FDM}>). Já um evento como a alteração do e-mail é um evento com uma categoria, logo, neste caso temos um *itemset* com um *item* (<{Alterar e-mail}>). Na secção seguinte serão apresentados os tipos de eventos e respetivas categorias que farão parte do modelo analítico.

O comprimento de uma sequência corresponde ao número de eventos dessa sequência. Uma sequência com comprimento l é uma *l-sequence*. Por exemplo, a sequência <{1,2}, {3}> tem um comprimento igual a 2, pelo que pode ser chamada de *2-sequence* (Fournier-Viger et al., 2017).

Uma sequência <$a_1$,$a_2$,$a_3$,...$a_n$> está contida numa outra sequência <$b_1$,$b_2$,$b_3$,...$b_n$> se $i_1$<$i_2$<...<$i_n$ tal que $a_1 \subseteq b_{i1}$ ,$a_2 \subseteq b_{i2}$ ,....$a_n \subseteq b_{in}$. Ou seja, a sequência <{3}, {6,7,9}, {7,9}> está contida na sequência <{2}, {3}, {6,7,8,9}, {7}, {7,9}> uma vez que {3} ⊆{3} , {6,7,9} ⊆{6,7,8,9} e {7,9} ⊆{7,9}. Contudo, a sequência <{2,3}> não está contida na sequência <{2}, {3}> pois na sequência anterior 3 é comprado depois de 2, enquanto que na sequência <{2,3}> os itens 2 e 3 são comprados em conjunto. Se uma sequência $s_a$ está contida noutra sequência $s_b$ então $s_a$ é uma subsequência de $s_b$. Uma sequência é uma *maximal sequence* se não estiver contida em mais nenhuma sequência (Zhao e Sourav S. Bhowmick, 2003).

Uma base de dados de sequências (SDB) é uma lista de sequências SDB = {$s_1$, $s_2$, ..., $s_p$} com identificadores de sequências (SID) 1, 2...p (Fournier-Viger et al., 2017). Na Figura 4.2 é apresentada uma base de dados exemplo de sequências SDB, que contém quatro sequências com os SID 1, 2, 3 e 4. Cada SID corresponde ao comportamento de um cliente e é constituído por inúmeros eventos/interações cliente-empresa. Esta será a estrutura de dados adotada no projeto.

Para a procura poder ser realizada o padrão sequencial tem de ter uma estrutura semelhante à sequência (ser constituído por um conjunto de *itemsets*) e para além disso, conter *items* presentes

| SID | Sequence |
|---|---|
| 1 | <{Transação Worten}, {Pagamento de serviços}, {App}> |
| 2 | <{Alterar e-mail}, {Alterar % revolving}, {Transação FDM}, {App}> |
| 3 | <{Aceder à App Universo-Call-center},{Transação Worten} ,{Transferência}, {App}> |
| 4 | <{Transação FDM,Transação Restauração}, {Transação FDM}> |

Figura 4.2: Estrutura de dados adotada no projeto

na base de dados. O interesse de cada padrão sequencial (subsequência) depende do seu valor nas medidas de performance adotadas. Com uma estrutura similar à apresentada, aplicar-se-á um algoritmo de SPM para se obter os padrões sequenciais mais interessantes (aqueles que levam com maior probabilidade ao *download* da app).

Estando apresentados o algoritmo de SPM, a sua estrutura de dados e a sua aplicabilidade no contexto do projeto, seguem-se as próximas secções da metodologia que cobrem as diferentes fases do CRISP-DM. As diferentes secções abordam os seguintes tópicos:

- **Data Understanding and Preparation**: São selecionados os grupos de eventos tendo em conta a *Customer Journey* (CJ) da Sonae FS, realizada a análise descritiva e exploratória aos dados e apresentados todos os passos efetuados na preparação da tabela final;
- **Modelling and Evaluation**: Apresentam-se os algoritmos de SPM, as suas medidas de performance e as restrições que podem ser aplicadas na procura pelas subsequências mais interessantes;
- **Deployment**: É apresentado o processo de seleção dos padrões sequenciais para o envio da comunicação e o plano do *deployment*;

## 4.2 Data Understanding and Preparation

### 4.2.1 *Customer Jourtney* da Sonae FS e Grupos de eventos selecionados

A seleção de eventos a incluir no modelo baseou-se na *Customer Journey* da Sonae FS. De acordo com Følstad e Kvale (2018), a *Customer Journey* (CJ) pode ser definida como "o processo, caminho ou sequência a partir do qual um cliente acede ou utiliza um serviço" (p.16;tradução do autor). De acordo com os mesmos autores, as diversas definições encontradas na literatura de CJ demonstram a importância da centralização no cliente nesta perspetiva. Desta forma, ao utilizar-se como ponto de partida a CJ da Sonae FS na seleção de variáveis a incluir no modelo pretende-se tornar o cliente o ator principal do modelo analítico e realizar uma campanha de marketing centrada nas necessidades e vontades deste.

A CJ da Sonae FS divide-se em 4 fases: Angariação, *Onboarding*, Desenvolvimento e Recuperação. Em cada uma destas fases o cliente Universo interage com a Sonae FS de diversas formas

sendo que é na fase de *Onboarding* e Desenvolvimento que se encontram o maior número de interações cliente-empresa. A fase de Angariação corresponde ao primeiro contacto do cliente com a empresa através da pesquisa do cartão Universo por parte do cliente e correspondente adesão ao serviço (é aqui que o cliente recebe o primeiro cartão Universo- um cartão temporário que já lhe permite realizar transações). O *Onboarding* inclui os primeiros passos da relação cliente-empresa, aqui fazem parte todas as interações transacionais, contratuais ou via *call-center*. O Desenvolvimento divide-se em 2 períodos: a utilização do cartão- interações relacionadas com o serviço prestado pela Sonae FS e o pagamento- receção do extrato e pagamento do mesmo. A última fase é a Recuperação. A Recuperação corresponde ao fim da relação entre o cliente e a empresa com o cancelamento do contrato ou a entrada do cliente em *churn* (o cliente reduz a frequência transacional e, posteriormente, abandona a empresa) (Figura 4.3).

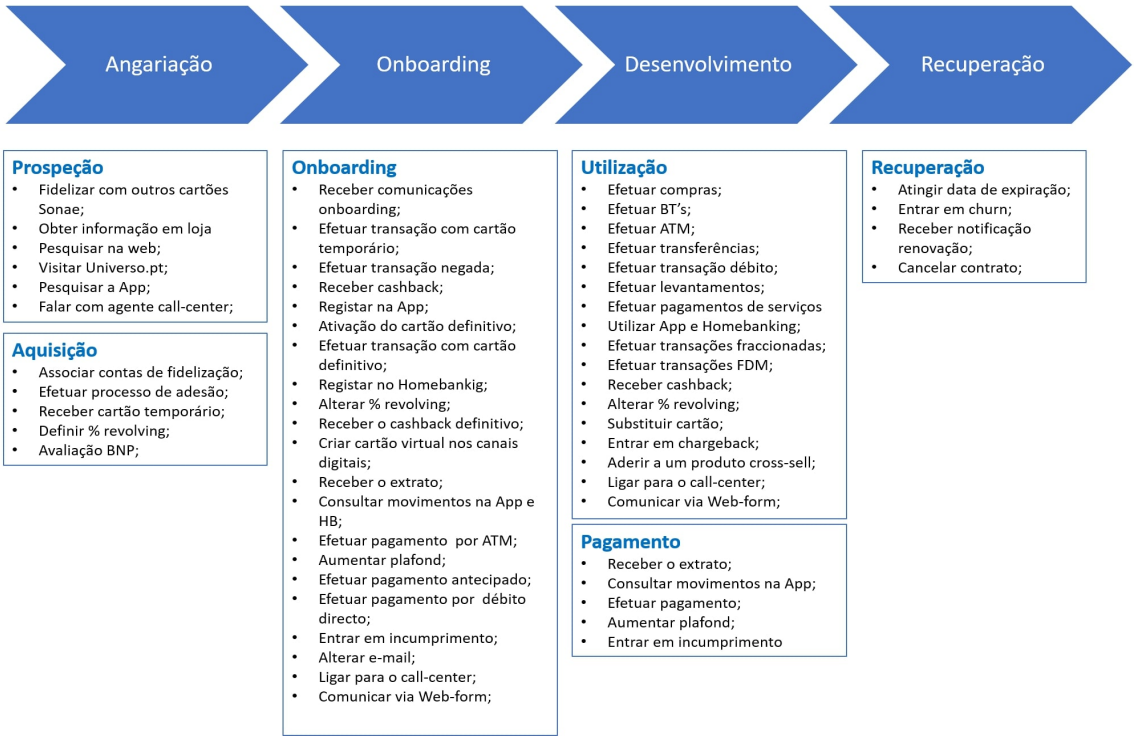



Figura 4.3: Ilustração da *Customer Journey* da Sonae FS

Para além da CJ da Sonae FS, na seleção de variáveis é necessário ter em conta o que é recomendado pela literatura como fatores determinantes na adoção ao *m-banking*, até porque pretende-se estudar quais os padrões sequenciais de eventos que levam com maior probabilidade ao *download* da App Universo. Deste modo, todas as variáveis a incluir no modelo analítico têm de estar em concordância com os fatores apresentados na revisão da literatura.

Posto isto, decidiu-se excluir da análise todos os eventos das fases de angariação e recuperação isto porque nestas fases o cliente ainda não estabeleceu uma relação sólida com a empresa (Angariação), ou a relação com a empresa chega ao fim (Recuperação). A inclusão de eventos destas fases não estaria em concordância com os seguintes fatores - confiança, utilidade e conveniência.

Para além desta exclusão removeram-se da análise outros eventos que não estavam em concordância com os fatores apresentados na revisão da literatura. Por exemplo, entrar em *chargeback* (processo complexo em que o comprador cancela a compra por diversos motivos) ou em incumprimento são reveladores de uma instabilidade na relação cliente-empresa e não seria conveniente propor a App num momento instável desta relação.

Após a realização das remoções referidas obtiveram-se 4 grupos principais de variáveis apresentados e explicados na Figura 4.4. Estes grupos correspondem às interações mais relevantes entre o cliente e a empresa e que tiveram como ponto de partida o cliente.

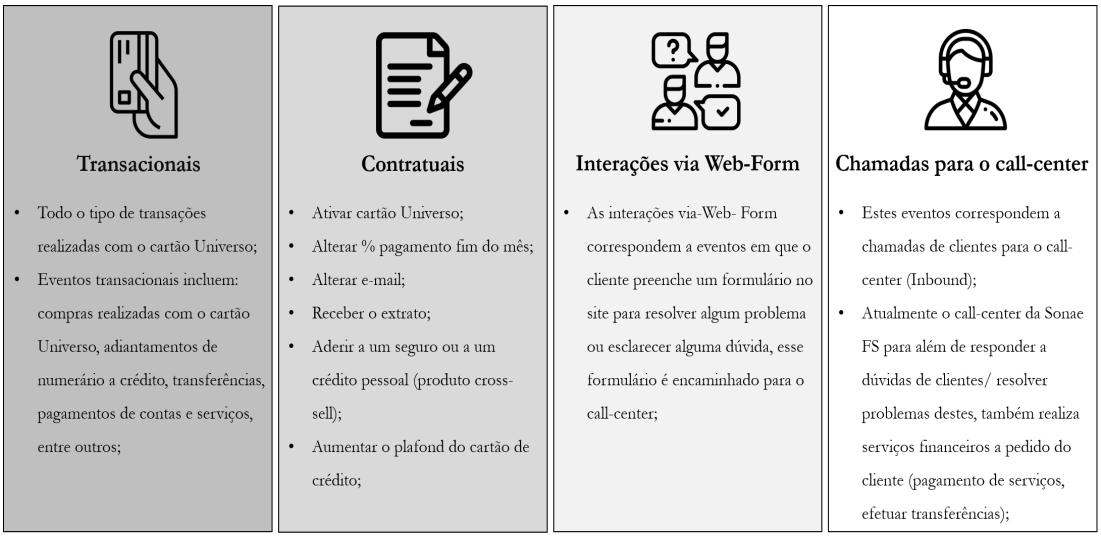



Figura 4.4: Grupos de variáveis selecionados para o modelo analítico

### 4.2.2 Data description and exploration

Após terem sido selecionados os eventos mais importantes é necessário realizar uma análise exploratória a estas variáveis. Com esta análise pretende-se perceber melhor os dados para serem tomadas decisões mais conscientes em fases posteriores do projeto.

Numa primeira fase, é necessário escolher os clientes que serão alvo da análise no projeto e como se pretende comunicar com os clientes para estes aderirem à App Universo foram apenas considerados os clientes atuais Universo, logo, os eventos de clientes inativos foram excluídos. Por outro lado, é necessário também escolher o limite temporal da análise e, neste tópico foram definidos os 2 últimos anos (desde abril de 2018 a abril de 2020).

Seguidamente, realizou-se uma análise descritiva tentando-se perceber qual era a frequência dos diferentes grupos de eventos apresentados na Figura 4.4. Como podemos constatar pela Tabela 4.1 o número de eventos transacionais é muito superior aos outros tipos de eventos. De salientar também o evento receber o extrato que sendo um evento periódico e comum a todos os clientes ativos apresenta-se também como bastante frequente.

Mas como no projeto estamos à procura de sequências de clientes interessantes e não de tipos de eventos frequentes, a análise dos tipos de eventos frequentes deve ser complementada com o

Tabela 4.1: Número de eventos e clientes por tipo de evento

| **Tipo de evento** | **Número de eventos** | **Número de clientes** |
|---|---|---|
| Transacionais | 62 888 143 | 799 581 |
| Chamadas *Inbound* | 773 548 | 273 676 |
| *Web-Form* | 118 824 | 71 606 |
| Alterar e-mail | 329 619 | 303 308 |
| Alterar % revolving | 9 492 | 8 385 |
| Receber o extrato | 12 533 838 | 842 931 |
| Aderir a um produto *cross sell* | 122 805 | 114 767 |
| Aumentar o *plafond* | 120 906 | 112 107 |
| Ativar cartão | 205 984 | 205 984 |

número de clientes que realizaram os diferentes tipos de eventos nos anos da análise. Através da análise da Tabela 4.1 podemos concluir que apesar de os eventos transacionais serem muito comuns, não são os que abrangem o maior número de clientes. Isto acontece porque o extrato é recebido por todos os clientes que tenham dívida com a Sonae FS e que podem ou não ter transacionado nos anos 2018, 2019 e 2020.

Relativamente aos outros eventos, percebe-se que apesar de serem muito menos frequentes que a receção do extrato e as transações, ainda abrangem um espectro relativamente considerável dos clientes Universo. Na prática isto acontece porque são realizados com menos frequência por um determinado cliente. Isto quer dizer que o rácio interações/cliente é inferior nestes eventos relativamente aos eventos transacionais e ao evento receber o extrato (Anexo D).

Outro fator bastante importante nos projetos de *data mining* que tentam captar comportamentos interessantes dos clientes é o fator tempo. Daí ser importante analisar a evolução dos eventos ao longo do tempo. Na Figura 4.5 podemos constatar a evolução positiva dos *downloads* da App e também uma evolução positiva do número de eventos em 2018 e 2019 contrastada com uma ligeira queda em 2020. Se existir algum mês em que o número de eventos for bastante díspar dos restantes é recomendável reformularmos os limites temporais da análise. Na verdade, essa realidade não se verifica logo, todos os meses podem ser incluídos.

Em suma, através da análise realizada foram retiradas algumas conclusões que, nos poderão auxiliar nas fases seguintes do projeto. As conclusões são as seguintes:

- Existe uma necessidade de criar categorias para os eventos transacionais, os eventos *Web-Form* e as chamadas *Inbound* (os eventos contratuais já se encontram categorizados na análise) isto porque se não realizarmos esta categorização, não conseguiremos extrair subsequências interessantes;

- Com o número total de eventos e clientes apresentado e sabendo da necessidade de categorizar os grupos referidos na alínea anterior, terão de ser eliminados mais eventos da análise pois a complexidade dos algoritmos de SPM aumenta exponencialmente com um incremento residual no número sequências (número de clientes) e número de elementos da

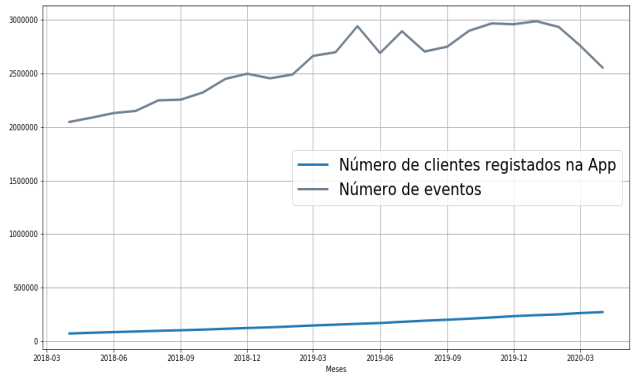



Figura 4.5: Evolução temporal do número de eventos

sequência (eventos) (Fournier-Viger et al., 2017). Na próxima subsecção serão selecionados os eventos finais a incluir no modelo;

- Com uma frequência tão elevada de eventos e um rácio de eventos/cliente também muito elevado os eventos transacionais terão muito provavelmente bastante relevância nas sequências de maior comprimento (sequências com um número elevado de eventos);
- Como o número de eventos é relativamente constante ao longo dos meses analisados existe confiança na realização de uma análise aos padrões sequencias de clientes entre abril de 2018 e abril de 2020.

### 4.2.3 Seleção final dos eventos

Chegados a esta fase, e sabendo da necessidade de criar diferentes categorias para as variáveis transacionais, interações *Web-form* e chamadas *Inbound* decidiu-se categorizar os eventos do seguinte modo:

- **Chamadas *Inbound***- A cada interação enquadrada nesta tipologia foi associada a razão para a chamada *Inbound*;
- **Interações *Web-Form***- A cada interação *Web-Form* foi associado um problema que se pretende resolver com essa interação;
- **Transações**- Foram criadas categorias transacionais que correspondem a diferentes modalidades de pagamento, diferentes locais onde é realizada a transação, diferentes tipo de transações (transferências, pagamento de serviços ou levantamentos) e diferentes montantes envolvidos na transação. Esta categorização não implica que cada transação tenha uma e só uma categoria, ou seja, a uma transação podem ser associadas 2 ou mais categorias desde que esta faça parte destas categorias (Por exemplo: uma compra realizada no Continente, acima de 100 euros, na modalidade Fim do Mês pertence a três categorias);

Como referido na subsecção anterior, os algoritmos de SPM têm um crescimento exponencial ao nível da sua complexidade e consequentemente, tempo computacional quando são introduzidos mais eventos sequenciais e mais sequências (Fournier-Viger et al., 2017). Deste modo, foi necessário proceder à remoção de mais eventos para ser plausível obter um conjunto de subsequências interessantes até ao final do projeto. Na verdade, só no grupo de eventos *Inbound* e *Web-Form* incluem-se 138 eventos pois cada razão dá origem a um evento diferente. A estes 138 juntam-se 23 categorias transacionais e contratuais. Fica assim comprovada a inevitabilidade de procedermos à eliminação de mais eventos pois se incluíssemos 161 eventos no nosso modelo não conseguíamos obter padrões sequenciais até ao final do projeto.

Para obtermos as variáveis/eventos finais, foi calculada, para cada evento, a percentagem de clientes não digitais que após 3 dias de terem realizado o evento X se tinham registado na App Universo. De facto, esta percentagem corresponde à taxa de conversão durante os 3 dias subsequentes à ocorrência do evento X. Seguidamente, selecionaram-se os 20 eventos transacionais e contratuais, os 10 eventos *Web-Form* e 10 chamadas *call-center* com maior taxa de conversão. (Figura 4.6, Anexo E e Anexo F)

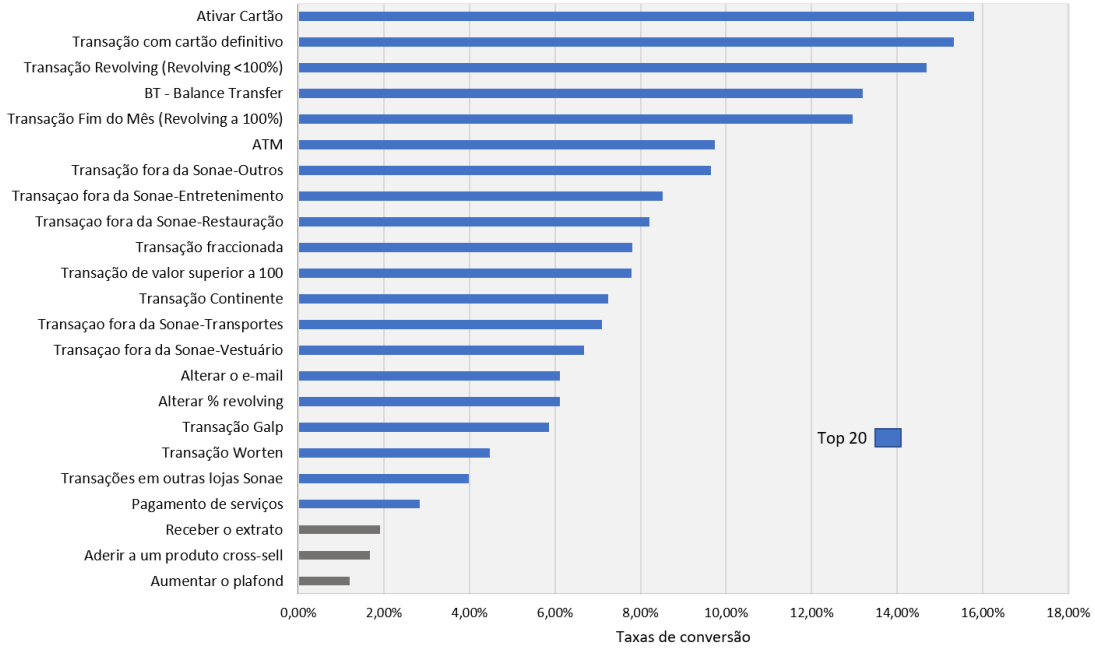



Figura 4.6: Taxas de conversão nos eventos transacionais e contratuais

Tal como seria de esperar, o evento receber o extrato não promove a realização do *download* da App e, consequentemente, não será incluído no modelo final. Isto acontece, pois é um evento frequente, síncrono e que não é uma iniciativa do cliente. Por outro lado, aderir a um produto *cross-sell* e aumentar o *plafond* também se revelaram eventos pouco estimulantes no que concerne o registo da App Universo e, por conseguinte, serão removidos da análise. Relativamente, aos eventos chamadas *Inbound* e interações *Web-Form* foi obtida a informação de que quando os clientes contactam para resolver certos problemas/esclarecer determinadas dúvidas o próprio

operador *Call-center* promove o registo na App o que por si só pode explicar os elevados valores obtidos nas taxas de conversão nestas tipologias de eventos.

Na figura abaixo são apresentados os 39 eventos incluídos no modelo analítico. De salientar, a exclusão do evento transação com cartão definitivo visto que mais tarde se verificou que devido ao seu pouco grau de granularidade poderia estar a afetar a procura pelas sequências mais interessantes.

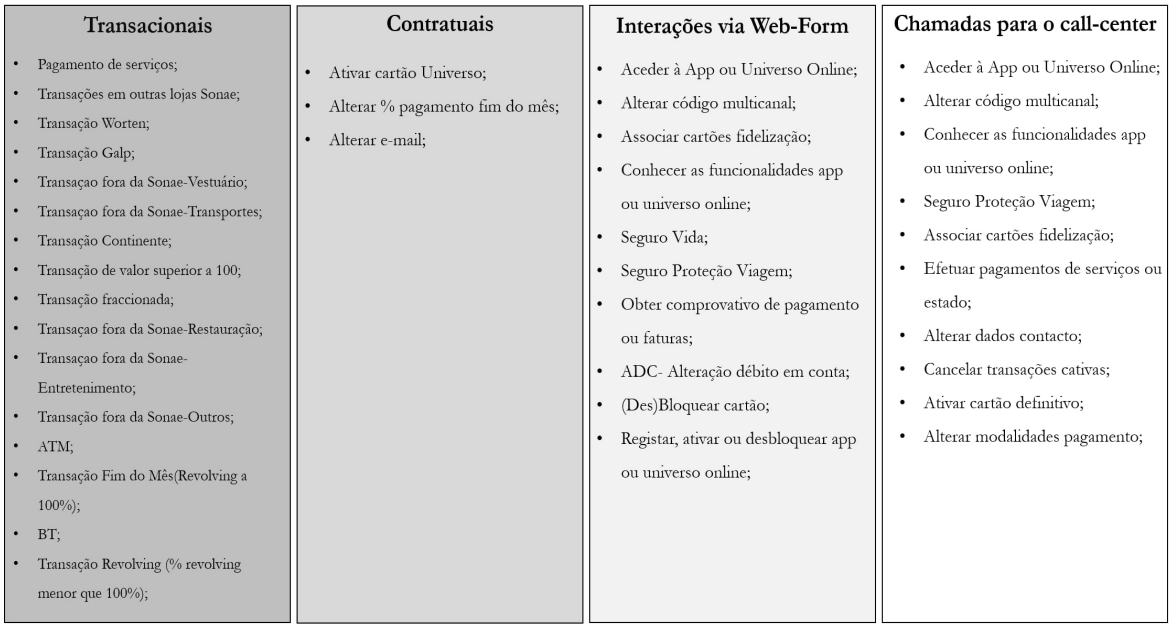



Figura 4.7: Variáveis incluídas no modelo

### 4.2.4 *Data preparation*

Depois de selecionadas as variáveis finais que farão parte do modelo e de ter sido realizada uma análise exploratória aos eventos, os dados foram preparados de forma a que cumpram com os requisitos exigidos pelo algoritmo de SPM. Nesta fase, foi criada uma Tabela Master adaptada ao problema (similar à Figura 4.2).

Foi definido nesta fase que só se pretendia estudar padrões sequenciais relativamente recentes, consequentemente, foram incluídos apenas 6 meses de eventos por cliente. Contudo, devido ao facto de estarmos apenas interessados nos eventos antes do registo da App para os clientes que fizeram o download, consideraram-se apenas os eventos antecedentes ao registo na App para este tipo de clientes. Sendo assim, a abrangência temporal dos eventos a serem incluídos na Tabela Master foi a seguinte:

- **Clientes que se registaram na App**: foram considerados os últimos 6 meses de eventos até ao registo da App, sendo que só se incluíram registos a partir de janeiro de 2019;
- **Clientes que não se registaram na App**: foram considerados os últimos 6 meses de eventos de clientes que tinham o último evento a partir de janeiro de 2019;

Para além disso, na inserção dos dados tiveram-se em conta fatores que podiam afetar a qualidade dos resultados obtidos com o algoritmo. Consequentemente, procedeu-se à eliminação de eventos duplicados e de clientes-teste (clientes criados artificialmente pela Sonae FS) pois a inclusão de eventos/clientes deste tipo podia enviesar a análise.

A remoção dos eventos duplicados ocorreu quando se verificavam uma das duas condições seguidamente apresentadas: transações realizadas pelo mesmo cliente no mesmo instante de tempo e eventos não transacionais realizados pelo mesmo cliente no mesmo dia. A distinção entre eventos não transacionais e transacionais deve-se ao facto da frequência dos eventos transacionais ser bastante superior tal como constatado na Tabela 4.1. Para além desta remoção, para os eventos *Web-Form* só foi considerado o primeiro *ticket* de um determinado caso.

## 4.3 *Modelling and evaluation*

### 4.3.1 Medidas de performance

O objetivo do SPM é descobrir subsequências interessantes numa base de dados de sequências, isto é, relações sequenciais entre itens relevantes para o utilizador. Várias medidas podem ser usadas para avaliar o quão interessante é uma subsequência. A medida tradicional para avaliar a exequibilidade de um padrão é o *support*(*sup*). O *support* de uma subsequência $s_a$ numa base de dados de sequências SDB é definido como o número de sequências que contem a subsequência $s_a$ a dividir pelo número total de sequências em SDB (4.1). Com o *support* pretende-se medir a frequência de uma determinada subsequência (Fournier-Viger et al., 2017).

Outra medida muito utilizada neste tipo de problemas é a *confidence*(*conf*). A *confidence* pode ser definida como a probabilidade de um determinado evento ocorrer sabendo que a subsequência antecedente ocorreu (a subsequência antecedente é constituída pelos *itemsets* antecedentes ao último evento). Efetivamente, a *confidence* é uma probabilidade condicional que pode ser calculada com a divisão entre o *support* da subsequência toda (a que incluí o evento final) e o *support* da subsequência antecedente (Meyer, 2003) (4.2).

$$\textbf{support(a)} = \frac{\textit{Sequências que contêm a}}{\textit{Total de sequências}} \tag{4.1}$$

$$\textbf{confidence}(\mathbf{a_1}\text{->}\mathbf{a_2}\text{->...->}\mathbf{a_q}) = \frac{sup(a_1 -> a_2 -> ... -> a_q)}{sup(a_1 -> a_2 -> ... -> a_{q\text{-}1})} \tag{4.2}$$

Na busca pelos padrões sequenciais mais relevantes adotou-se uma metodologia semelhante à realizada por Sinha et al. (2015) pois pretende-se obter as subsequências que levam com maior probabilidade a um objetivo de marketing, neste caso, o registo na App Universo. A metodologia é composta por 3 fases:

1. Na base de dados com sequências de clientes registados na App Universo obter todas as subsequências com *support* superior a um determinado *threshold*;

2. Na base de dados com sequências de clientes com e sem app, calcular a *confidence* das subsequências obtidas no passo 1;

3. Comunicar com os clientes que tenham feito uma subsequência (padrão sequencial) com uma *confidence* superior a um *threshold* definido;

Com esta metodologia temos em conta não só a frequência da subsequência no conjunto de sequências de clientes que se registaram na App (com o *support*) mas também a probabilidade desse padrão sequencial levar ao objetivo de marketing (com a *confidence*). Esta abordagem permite-nos encontrar as melhores subsequências suscetíveis de conduzir a um determinado objetivo, neste caso, o registo na App Universo.

### 4.3.2 Restrições nos algoritmos de SPM

Para reduzir o número de padrões sequenciais a procurar e tentar encontrar os padrões mais interessantes, diversos autores como Desai e Ganatra (2015) e Srikant e Agrawal (1996) integram restrições nos algoritmos de SPM. Uma restrição é um critério que o utilizador fornece para indicar com maior precisão os tipos de padrões a encontrar. Há duas formas de aplicar restrições. A primeira forma é aplicá-las como uma etapa de pós-processamento no conjunto de todos os padrões sequenciais para filtrar padrões desinteressantes. Contudo, um problema com esta abordagem é que a enumeração de todos os padrões sequenciais pode consumir muito tempo e requer uma enorme quantidade de memória. A segunda forma de abordar este problema é aplicando as restrições durante a procura de padrões para reduzir o *search space*. Os algoritmos que adotam este tipo de abordagem são muito mais rápidos e geram muito menos padrões do que os algoritmos tradicionais de extração de padrões sequenciais, por isso mesmo seguiu-se a segunda abordagem no projeto de dissertação. Dentro das várias tipologias de restrições encontram-se:

- ***Length constraints***: São impostos limites máximos e mínimos ao número de *itemsets* numa sequência;

- ***Support constraints***: Normalmente são impostos limites mínimos ao número de vezes que um determinado padrão tem de ocorrer nas sequências;

- ***Size constraints***: São impostos limites máximos e mínimos ao número de *items* num *itemset*;

- ***Gap constraints***: São restrições que limitam o tempo mínimo e máximo entre dois *itemsets* consecutivos;

Nos projetos de SPM, quem define o valor a atribuir para cada um dos parâmetros é o utilizador, visto que não seria profícuo utilizar valores teóricos para cada um dos parâmetros se cada projeto tem requisitos e objetivos diferentes. O utilizador tem sempre a hipótese de não incluir certas restrições na análise. Por exemplo, Desai e Ganatra (2015) incluíram na sua análise padrões sequenciais com e sem *gap*. Para além disso, o utilizador quando define os valores para cada um destes parâmetros deve ter em conta o tempo de computação, o consumo de memória, o número

de padrões sequenciais que pretende analisar, o comprimento pretendido para os padrões e *itemsets* e a distância temporal entre *itemsets* consecutivos. Na Secção 5.1 são apresentados os valores escolhidos para as restrições e medidas de performance.

### 4.3.3 Algoritmos de SPM

A tarefa de *Sequential Pattern Mining* é um problema de enumeração. O seu objetivo é enumerar todos os padrões (subsequências) que cumprem com as restrições impostas. Contudo, descobrir padrões sequenciais é uma tarefa árdua pois o número possível de subsequências numa base de dados sequencial é muito elevado, na verdade, uma sequência contendo q *items* pode ter 2q-1 subsequências distintas. Deste modo, aplicar uma abordagem *"naive"* acaba por ser infrutífero neste tipo de problemas e é com isto em mente que é necessário recorrer a algoritmos eficientes que evitam explorar todo o *search space* (Fournier-Viger et al., 2017).

Na realidade, todos os algoritmos de SPM têm os mesmos resultados se a base de dados utilizada e os parâmetros de entrada forem os mesmos. Consequentemente, o valor acrescentado de um determinado algoritmo relativamente a outro prende-se na eficiência desse mesmo algoritmo não no seu *output*. Normalmente, os algoritmos de SPM diferem em 4 pontos:

- Metodologia de pesquisa;
- Representação da base de dados;
- Na forma como geram os próximos padrões a serem pesquisados;
- Na forma como contam o *support*.

#### 4.3.3.1 Algoritmos *Breadth-first Search* (BFS)

Este tipo de algoritmos começa por procurar *1-sequences* frequentes (sequências de 1 só *item* que tem um *support* igual ou superior a determinado *threshold*), depois eles geram sequências com mais um item através de *s-extensions* e *i-extensions* das *1-sequences*, seguidamente *3-sequences* são geradas a partir das *2-sequences* frequentes e assim sucessivamente até não poderem ser geradas mais sequências (Fournier-Viger et al., 2017).

Os algoritmos *Breadth-first Search* mais comuns são o *GSP* e o *AproriAll*. Ambos usam bases de dados horizontais. Numa base de dados horizontal, cada linha corresponde a uma sequência (Figura 4.2). *AprioriAll* foi um dos primeiros algoritmos de SPM a ser desenvolvido e, ainda serve de base para a construção de algoritmos mais complexos. Este algoritmo baseia-se na *Apriori property* que determina que se uma sequência não é frequente todas as suas extensões não serão frequentes. Este tipo de algoritmos apresenta algumas desvantagens ao nível da eficiência e tempo de computação isto porque padrões candidatos não existentes na base de dados são gerados, todos os padrões candidatos são mantidos na memória e as bases de dados são lidas múltiplas vezes (Agrawal e Srikant, 1995).

O algoritmo *GSP* é uma versão melhorada do *AprioriAll* isto porque incluí uma gama de restrições que permitem realizar uma busca personalizada pelas sequências desejadas. Com o *GSP* é possível criar padrões que incluam *items* em diferentes níveis de taxonomia, flexibilizar o *support* de uma subsequência com a inclusão de *time-windows* e incluir restrições temporais que limitam o intervalo de tempo entre elementos adjacentes (Srikant e Agrawal, 1996).

#### 4.3.3.2 Algoritmos *Depth-first Search*(DFS)

Os algoritmos *Depth-first Search* tem uma metodologia de pesquisa diferente dos algoritmos *Breadth-first Search*. Em vez de serem pesquisados todos os "filhos" de um determinado nó (todas as sequências de um determinado comprimento) até podermos ir para o próximo nível, estes algoritmos começam por pesquisar padrões com sequências de 1 item e depois realizam extensões recursivas para apenas um desses padrões e só quando um determinado padrão não poder ser mais estendido é que se volta a explorar os outros padrões. A vantagem deste tipo de algoritmos é que consegue ser muito mais rápido a atingir sequências interessantes de grande comprimento (Chand et al., 2012).

Um dos algoritmos *Depth-first Search* é o *SPADE*. Este algoritmo foi criado para corrigir algumas desvantagens do *GSP*. O *SPADE* utiliza uma representação vertical da base de dados. Na prática, num formato deste tipo é criada uma lista para cada *item* com as sequências que o mesmo aparece e os *itemsets* em que esse *item* aparece nessa sequência (Figura 4.8). Para um determinado *item* esta lista chama-se *IDList* do *item*. A representação vertical permite um cálculo automático do *support*- o *support* de um padrão é simplesmente o número de sequências distintas da sua *IDList* e, para além disso, com esta representação, a base de dados só é pesquisada uma vez (a *IDList* de um determinado padrão $s_a$ obtido através de uma extensão entre uma sequência $s_b$ com um elemento i pode ser concebida com um *"join"* entre as *IDLists* de $s_b$ e i). Por estas razões a performance deste tipo de algoritmos é superior aos anteriormente apresentados (Zaki, 2001).

| a | |
|---|---|
| SID | *Itemsets* |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1,3,4 |

| b | |
|---|---|
| SID | *Itemsets* |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1,4 |

| c | |
|---|---|
| SID | *Itemsets* |
| 1 | 2 |
| 2 | 1 |
| 3 | 2 |
| 4 | 4 |

| d | |
|---|---|
| SID | *Itemsets* |
| 1 | |
| 2 | |
| 3 | |
| 4 | 2 |

| e | |
|---|---|
| SID | *Itemsets* |
| 1 | 5 |
| 2 | 2,5 |
| 3 | |
| 4 | |

| f | |
|---|---|
| SID | *Itemsets* |
| 1 | 3 |
| 2 | 3 |
| 3 | |
| 4 | |

| g | |
|---|---|
| SID | *Itemsets* |
| 1 | 3,4 |
| 2 | 4 |
| 3 | |
| 4 | |

Figura 4.8: Representação vertical da base de dados

#### 4.3.3.3 Algoritmos *Pattern-growth*

Para além dos algoritmos *DFS* e *BFS*, existem também os algoritmos *pattern-growth*-PG. Estes algoritmos resolvem um dos problemas dos algoritmos anteriormente apresentados que é o facto de serem geradas sequências candidatas que não existem na base de dados. Isto acontece porque os algoritmos *DFS* e *BFS* combinam padrões mais pequenos sem irem à base de dados.

Algoritmos *pattern-growth* evitam este problema através da busca recursiva na base de dados para encontrar padrões maiores (*divide-and-conquer approach*). Com esta abordagem eles têm apenas em conta os padrões que realmente aparecem na base de dados. A realização desta exploração pode, no entanto, ser dispendiosa. Para reduzir o custo dos algoritmos *pattern-growth*, foi criado o conceito de *project database* (Han et al., 2004), que visa reduzir o tamanho da base de dados à medida que padrões maiores são considerados pela *depth-first search*. Os algoritmos *pattern-growth* mais utilizados são: o PrefixSpan e o FreeSpan (Fournier-Viger et al., 2017).

#### 4.3.3.4 Seleção do algoritmo

No projeto selecionaram-se 2 algoritmos para procurar as subsequências mais interessantes. Os algoritmos selecionados foram o *SPADE* (Zaki, 2001) e uma versão melhorada do GSP em que é utilizada uma estrutura intermédia de dados mais eficiente (Masseglia et al., 1998). A seleção destes 2 algoritmos deveu-se a diversos fatores entre os quais a possibilidade de poder restringir a procura pelas sequências mais interessantes com os parâmetros apresentados na Subsecção 4.3.2, o tempo de computação, o consumo de memória e a possibilidade de procurar sequências com *itemsets* específicos. Os dois algoritmos foram implementados através de 2 *packages* no R - o *SPADE* com o *package arulessequences* (Buchta e Hahsler, 2020) e o GSP com o *TraMineR* (Gabadinho et al., 2020).

## 4.4 *Deployment*

### 4.4.1 Seleção das subsequências para o envio da comunicação

Não foi possível selecionar todas as subsequências obtidas com os algoritmos de SPM para comunicarmos com os clientes, isto porque o número de clientes passíveis de serem comunicados seria bastante elevado. Consequentemente, para selecionarmos os melhores padrões sequenciais foi seguido o processo apresentado na Figura 4.9.

Numa primeira fase foram obtidas as subsequências que cumprem com os *thresholds* definidos de *support*, *confidence* e de outras restrições impostas na procura pelas subsequências mais interessantes. Seguidamente, as subsequências foram divididas em grupos de 2, 3, 4, 5 e 6 elementos. O próximo passo foi selecionar a melhor medida de performance (*support* vs *confidence*) e depois disso foram escolhidas as melhores subsequências dentro de cada grupo de subsequências.

Para além do *support* e da *confidence*, aquando da seleção da melhor subsequência tiveram-se em conta mais dois indicadores de negócio que foram importantes na definição da melhor estratégia para o *deployment*. Esses indicadores são:

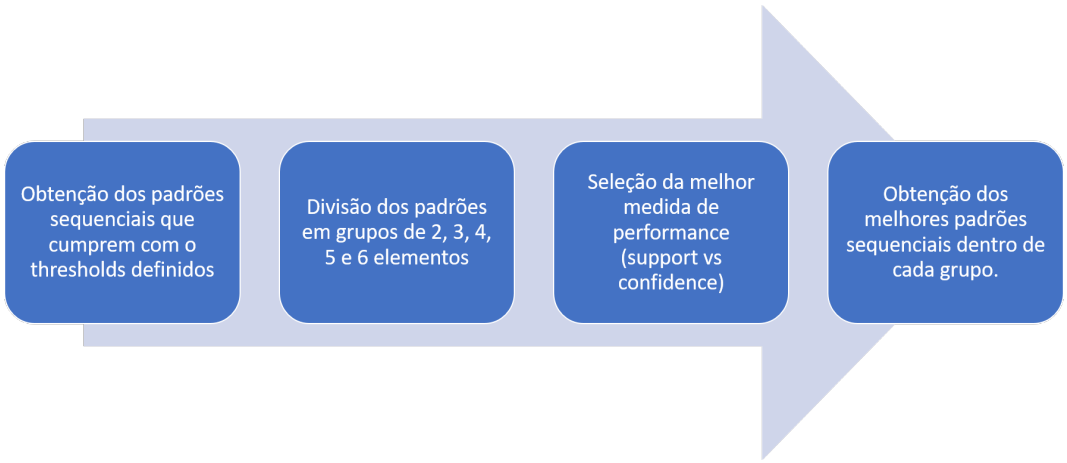



Figura 4.9: Processo de seleção dos padrões sequenciais

- **Número de clientes que realizaram a sequência (incluindo o registo na App)**: É um indicador que nos permite perceber o quão frequente foi o padrão sequencial no universo de clientes da Sonae FS nos últimos 6 meses;
- **Número de clientes potencial**: Corresponde ao número de clientes que realizaram o padrão sequencial até ao download da App, mas não realizaram o download da mesma, portanto, na prática, são os clientes potencialmente comunicáveis nos últimos 6 meses;

No projeto, realizou-se uma comunicação *event-driven* isto é comunicou-se com o cliente após este ter interagido com a Sonae FS. Sendo assim, a comunicação foi realizada após o cliente ter feito um evento antecedente ao download da App nas subsequências escolhidas. Sabendo disso, na seleção dos padrões sequenciais também se teve em conta a probabilidade do cliente registar-se na App após 3 dias sabendo que realizou o evento X (a *confidence* da sequência X->App com *timegap* de 3 dias).

### 4.4.2 Plano do *deployment*

Depois de selecionadas as subsequências mais interessantes chega a fase de planear a estratégia para o *deployment* do modelo. Como se pretendia comunicar com os clientes num espaço de tempo considerável (a empresa pretendia manter o modelo em produção pelo menos num espaço de 1 mês), a estratégia de monitorização e manutenção tem um papel crucial em evitar longos períodos de resultados incorretos de *data mining*. Sendo assim, foi necessário criar um plano de *deployment* que cumprisse com os requisitos estabelecidos pela empresa e que ao mesmo tempo tivesse em conta a manutenção e monitorização da comunicação.

O plano desenvolvido sustenta-se em 4 prismas fundamentais que pretendem aumentar a eficácia da comunicação são eles: comunicação com os clientes mais propensos, no momento certo, com uma mensagem apelativa e gerada automaticamente, numa base síncrona e diária. Só com a inclusão dos 4 prismas no plano se conseguiria explorar na totalidade todos os benefícios do modelo analítico e integrar os resultados deste no dia-a-dia do negócio.

O plano é constituído por 3 fases fundamentais: criação de um procedimento em *Structured Query Language* (SQL) que permite realizar uma atualização aos clientes elegíveis a receber a comunicação numa base diária (clientes que já realizaram a sequência até ao penúltimo evento e se fizerem o último evento estão aptos a receber a comunicação), obtenção do instante de tempo em que o cliente realiza a sequência (realiza o último evento) através do *Microsoft Azure* (*Azure Stream Analytics*) e envio da mensagem através da plataforma de envio de comunicações- *Marketing Cloud* da *SalesForce*.

Tiveram-se em conta os diferentes prismas apresentados, contudo, nem todas as comunicações puderam ser integralmente *event-driven* isto porque não foi possível obter o instante de tempo no caso dos eventos não transacionais, logo, a comunicação nesses casos não foi logo após o cliente ter realizado o evento X mas um dia após o evento. Sendo assim, os padrões sequenciais que acabam com eventos não-transacionais não passam pelo 2º passo. Nestes casos, os clientes comunicáveis são obtidos via SQL (o cliente já realizou a sequência toda tendo realizado o último evento no dia anterior) e daqui passa-se para o *Marketing Cloud*.

Relativamente à 1ª fase tiveram-se em conta diversos tópicos tais como: atualização dos clientes registados na App Universo, de clientes elegíveis a receber a comunicação por questões de pressão comercial, dos novos clientes e dos clientes que realizaram parte da sequência numa base diária. Na 2ª fase, obtêm-se os clientes que realizaram a transação X num determinado instante de tempo (é através do *Microsoft Azure* que se consegue realizar a comunicação *event-driven*). Na última fase, foi definido se o cliente pertence ao grupo de controlo ou ao grupo alvo (através da atribuição de uma distribuição de probabilidade ao cliente) e foi realizada a comunicação via SMS para o cliente (Figura 4.10).

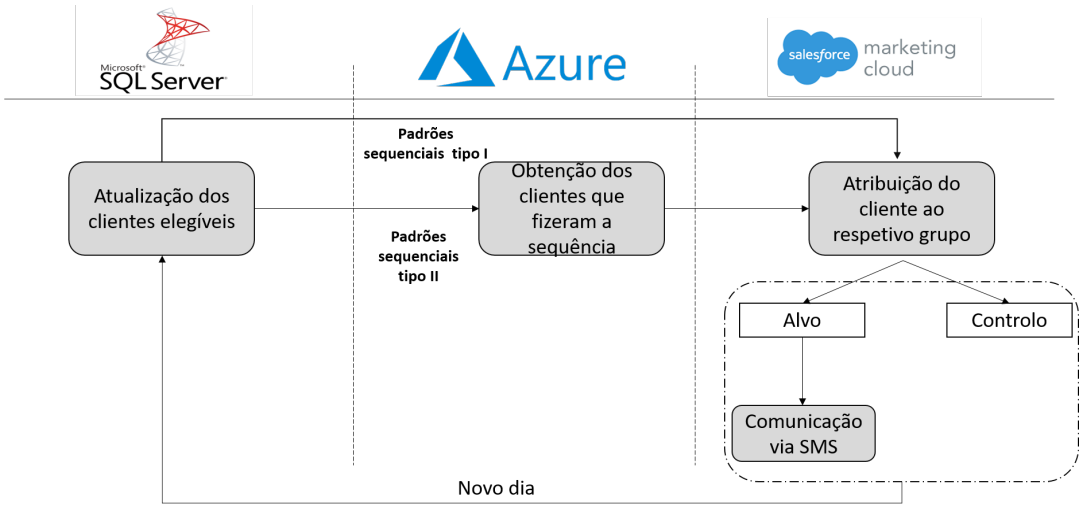



Figura 4.10: Plano do *deployment*

Para além deste plano foram definidas algumas restrições que pretendem tornar a campanha mais eficaz. As restrições foram as seguintes:

- Garantir que se um cliente está elegível para 2 ou mais padrões sequenciais no mesmo dia recebe apenas um SMS mesmo que realize os 2 ou mais padrões sequenciais;
- Excluir todos os clientes que realizaram algum padrão sequencial nos dias anteriores (tenham ou não esses clientes recebido a comunicação) ;

Adicionalmente, no que diz respeito os conteúdos das mensagens, de modo a tornar a mensagem mais adequada ao perfil do cliente que realizou a sequência, criaram-se conjuntos de padrões sequenciais similares (padrões realizados por um segmento de clientes com necessidades e vontades particulares) e atribuíram-se mensagens tipo a cada um destes conjuntos.

Finalmente, relativamente à mecânica da campanha é importante referir que cada padrão sequencial selecionado teve um grupo de controlo em que se tentou agrupar 10% dos clientes comunicáveis. Os resultados apresentados no capítulo seguinte correspondem a 7 dias de comunicação (desde o dia 2 de Junho ao dia 9 de Junho sendo que nos dias 6 e 7 não foram enviadas comunicações). Ao todo receberam uma mensagem informativa 3502 clientes (grupo-alvo) sendo que o grupo de controlo foi constituído por 358 clientes. De referir também que para efeitos de contagem dos registos na App Universo foram tidos em conta apenas os 3 dias subsequentes à realização do evento.

# Capítulo 5

# Resultados

Neste capítulo são apresentados os resultados do projeto. O capítulo divide-se em 5 secções. Numa primeira fase são apresentados os valores escolhidos para as restrições que tornam a procura pelas subsequências interessantes mais eficiente. Depois, apresentados os resultados com e sem *gap* e escolhidos os padrões sequenciais mais interessantes que sustentaram a comunicação. Finalmente, são apresentados os resultados da campanha realizada e comparadas as performances desta comunicação com outras campanhas com o mesmo objetivo de marketing.

## 5.1 Escolha dos valores das restrições e da *confidence*

Tal como referido na Subsecção 4.3.2, as restrições têm um papel crucial na procura pelos padrões sequenciais mais interessantes. Com a inclusão destas é possível tornar a procura mais eficiente e personalizada. As diferentes restrições impactam nos resultados obtidos com o algoritmo de SPM da seguinte forma:

- ***Length constraints***: São procuradas as subsequências com um determinado número máximo/mínimo de *itemsets*. Quanto menor o máximo e maior o mínimo mais eficiente é a procura;

- ***Support constraints***: É uma restrição que também é uma medida de performance, consequentemente, a definição de qual o valor mais apropriado para este parâmetro é de crucial importância para o estudo. Quanto menor o *support* menos eficiente é a procura;

- ***Size constraints***: Limita-se a procura pelas subsequências com um determinado número máximo/mínimo de *items* num *itemset*. Uma restrição mais exigente implica uma diminuição do número de padrões sequenciais obtidos;

- ***Gap constraints***: Foi incluído um *gap* máximo no nosso problema. A inclusão do *gap* máximo tem muito impacto no *support* e na *confidence* de sequências de 2 elementos. Isto porque o *sup*(A-App) sem *gap* > *sup*(A-App) com *gap*, e *sup*(A) é o mesmo com e sem *gap* logo a conf(A-App) com *gap* é inferior. O impacto também se faz sentir nas subsequências

com mais de 2 elementos mas em menor grau, principalmente na *confidence*. *Gaps* máximos mais pequenos implicam menos subsequências obtidas para o mesmo *support* e *confidence*;

Os valores para as restrições são definidos pelo utilizador pois dependem das características particulares do projeto. A Sonae FS não pretendia obter padrões sequenciais muito longos nem com *itemsets* com muitos *items*, logo considerou-se o valor máximo de 10 para estas restrições. Relativamente ao *gap*, tal como referido por Desai e Ganatra (2015) é interessante a sua inclusão para termos recomendações de curto e longo prazo, deste modo, foram obtidas subsequências com gap máximo de 60 dias e sem *gap*. Finalmente, o *support* e a *confidence* foram definidos de modo a que busca pelos padrões sequenciais fosse o mais eficiente possível com a garantia que se obtinha um número considerável de padrões sequenciais. Na Tabela 5.1 são apresentados os valores finais das restrições e da *confidence*.

Tabela 5.1: Valores das *restrições* e da *confidence*

| **Parâmetro** | **Sem gap** | **Com gap máximo de 60 dias** |
|---|---|---|
| *Length* | 10 | 10 |
| *Size* | 10 | 10 |
| *Support* | 0,08 | 0,07 |
| *Confidence* | 0,2 | 0,2 |

## 5.2 Padrões sequenciais sem gap

Com os valores das restrições apresentados na Tabela 5.1 e através do algoritmo SPADE, obtiveram-se um total de 1056 padrões sequenciais interessantes, o tamanho médio das subsequências foi de 5,55 *itemsets* e o número de *items* dentro de cada *itemset* não foi maior do que 3 em nenhum *itemset*.

Na Figura 5.1 encontram-se as frequências dos padrões sequenciais para diferentes valores de *support* e da *confidence*. Através do gráfico da direita é possível concluir que um grande número de padrões se encontra muito próximos do *support* mínimo. Já para a *confidence*, os padrões encontram-se mais dispersos ao longo dos diferentes valores desta medida de performance.

Como o número de padrões obtido é bastante elevado (1056) e sabendo da necessidade de escolher os melhores padrões dentro de cada grupo de 2, 3, 4, 5 e 6 elementos (Subsecção 4.4.1), temos de selecionar a melhor medida de performance de forma a que se obtenham as melhores subsequências dentro de cada grupo. Sinha et al. (2015) e Ben Baccar et al. (2019) selecionaram os padrões com maior *confidence* pois esta é uma medida robusta que mede a fiabilidade de um padrão e determina qual a probabilidade deste levar a um determinado objetivo. Para além disso, através da Figura 5.1 e dos anexos G e H, é possível constatar que os padrões se distribuem mais pelos diferentes valores da *confidence* do que com o *support*, logo esta é uma medida mais adequada para selecionar os melhores padrões.

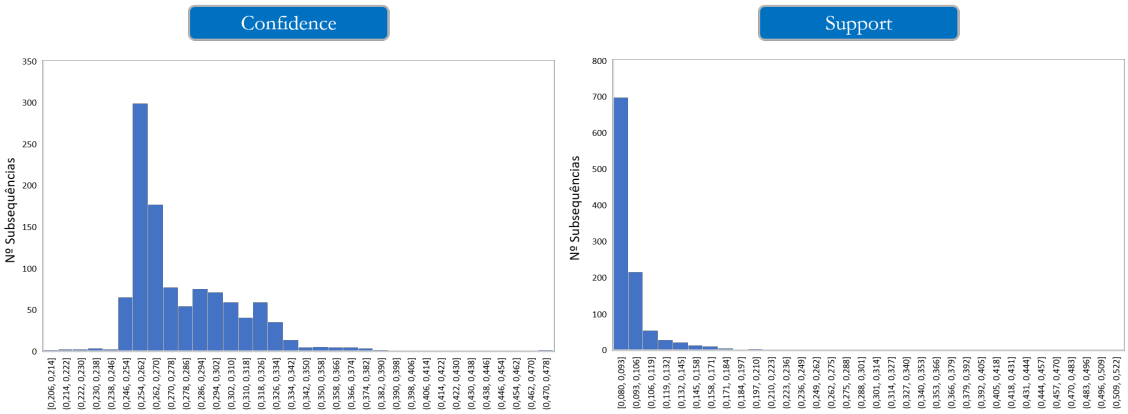



Figura 5.1: Distribuições das subsequências ao longo dos valores do *support* e da *confidence*

Depois de selecionada a melhor medida, foram obtidos os padrões sequenciais com melhor *confidence*. Nas Figuras 5.2 e 5.3 estão apresentados os 10 padrões sequenciais de 3 e 4 elementos com melhor *confidence*, no Anexo I encontra-se o top 10 para padrões sequenciais de 2, 5 e 6 elementos. Constatou-se que quase todos os padrões sequenciais incluem eventos transacionais. Para além disso, nas subsequências de 2, 3 e 4 elementos aparecem diversas categorias transacionais enquanto que nas de 5 e 6 elementos dominam as categorias D e F. Os padrões sequenciais apresentam valores similares da *confidence* dentro de cada grupo com exceção do padrão <{A},{B},{App}> que apresenta uma *confidence* bastante superior e que portanto é nesta fase o padrão sequencial mais interessante.

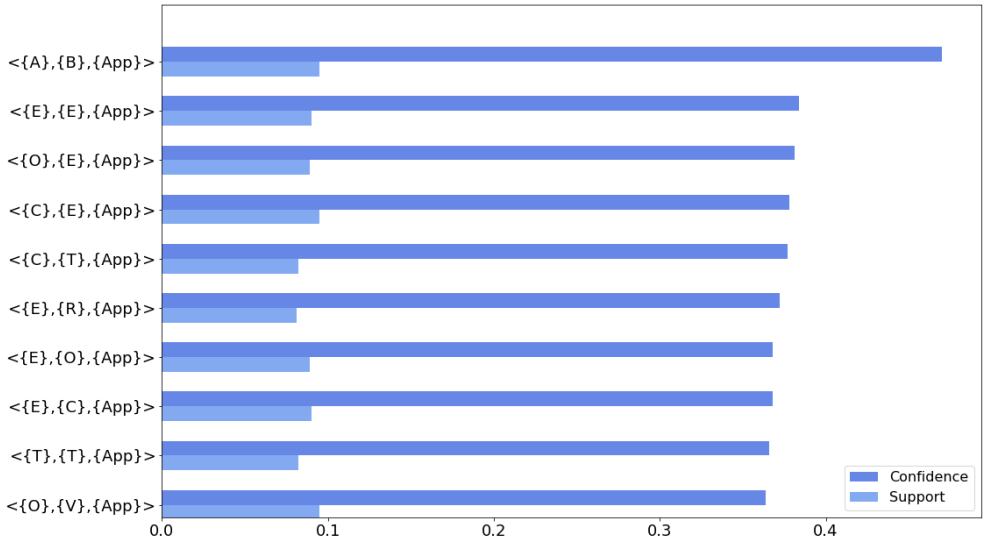



Figura 5.2: 10 melhores padrões sequenciais de 3 elementos

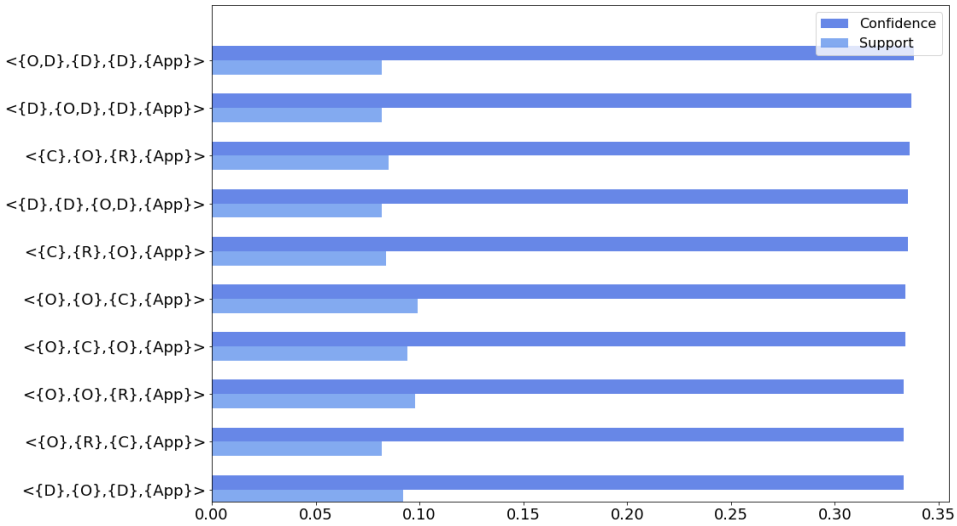



Figura 5.3: 10 melhores padrões sequenciais de 4 elementos

## 5.3 Padrões sequenciais com gap 60 dias

Como explicado na Secção 5.1, com a introdução do *gap*, o número de subsequências interessantes diminuí se os parâmetros do modelo mantiverem-se constantes. Consequentemente, para ser obtido um número de padrões sequenciais semelhante nas análises com e sem *gap* baixou-se o *support* para 0,07. Esta alteração pode ser justificada com as Figuras 5.1 e H.1. Na verdade, através da análise das figuras conseguimos perceber que o *support* é uma medida de performance que restringe bastante a procura pelos padrões mais interessantes visto que um grande número de padrões encontra-se no limite do mínimo desta restrição, logo, uma diminuição no valor de *support* mínimo aumenta o número de padrões sequenciais encontrados.

Com estas alterações obteve-se um total de 1277 padrões sequenciais sendo que as distribuições das subsequências pelos diferentes valores de *support* e da *confidence* são similares à análise sem gap o que corrobora a utilização da *confidence* como medida de performance mais adequada para selecionar o melhor padrão sequencial.

As subsequências obtidas com e sem *gap* são de certa forma similares, ou seja, são constituídas por muitos eventos transacionais. Contudo, com a introdução do *gap*, o evento F revelou-se bastante frequente nos padrões obtidos. Dentro dos grupos de padrões de 2, 3, 4, 5 e 6 elementos não existem padrões sequenciais com uma *confidence* bastante superior aos demais, contudo, é de destacar a presença das categorias transacionais F e O nos melhores padrões em todos os grupos (Figura 5.4 e 5.5 e Anexo J).

## 5.4 Seleção das sequências para o envio da comunicação

Chegados a esta fase foi necessário escolher as melhores subsequências para o *deployment*. Depois de analisar os resultados a Sonae FS pretendeu excluir os padrões sequenciais de 2, 5 e 6

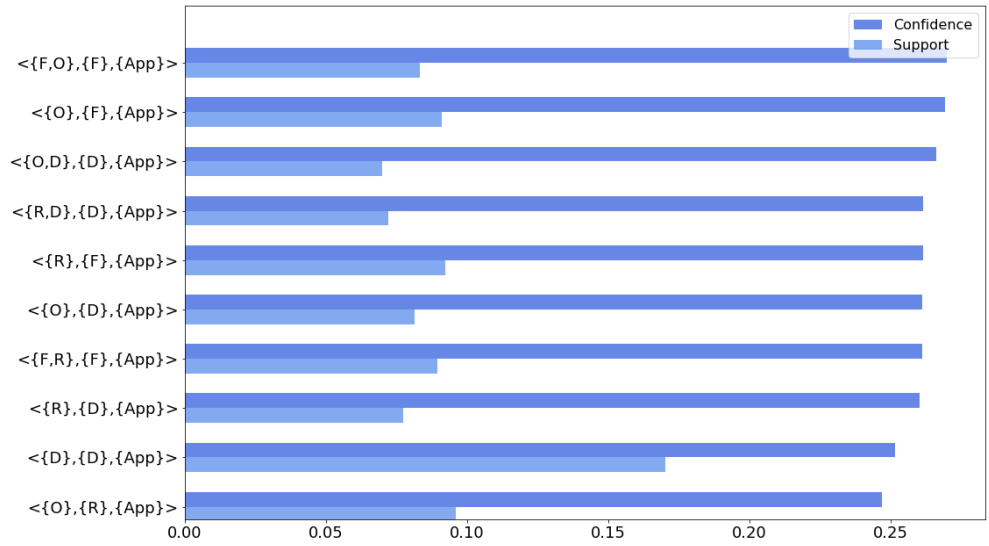



Figura 5.4: 10 melhores padrões sequenciais de 3 elementos com *gap* de 60 dias

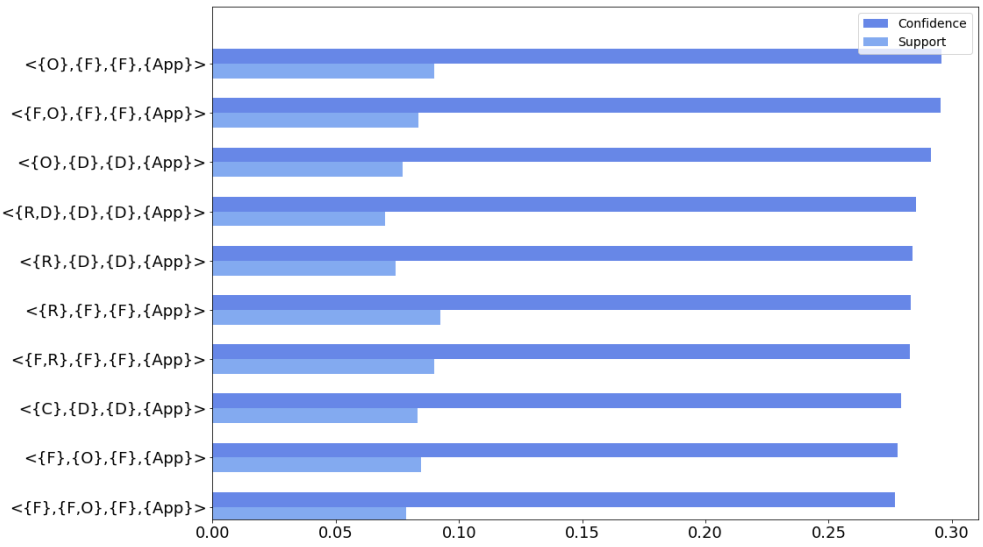



Figura 5.5: 10 melhores padrões sequenciais de 4 elementos com *gap* de 60 dias

elementos. Esta decisão foi uma decisão de negócio explicada também pelo facto de ter sido já realizada uma comunicação baseada em padrões sequenciais de 2 elementos que não foram *output* do modelo desenvolvido. Mais tarde (Secção 5.5) se comparará a performance dessa comunicação com a comunicação baseada no algoritmo de SPM.

Para se selecionar as melhores subsequências de 3 e 4 elementos tiveram-se em conta os pontos referidos na Subsecção 4.4.1. Numa primeira fase, foi preciso ter em conta os 2 indicadores referidos nessa subsecção: o número de clientes que realizaram o padrão sequencial e o número de clientes potencial. O número de clientes que realizaram a subsequência está dependente do *support* e é um indicador da relevância desse padrão nos meses de análise. A sua inclusão na análise prende-se com o facto de ser uma estimativa de quão frequente é aquele padrão no universo de clientes Sonae FS. O número de clientes potencial é um *proxy* para o número de clientes comunicáveis.

No Anexo K encontram-se os valores destes 2 indicadores para os padrões de 3 e 4 elementos com e sem *gap*. Escolher um padrão com mais clientes potenciais e que tenha sido realizado por muitos clientes não é necessariamente a melhor linha de ação até porque nas aplicações do SPM nesta área não foi encontrada uma melhor medida de performance do que a *confidence*. O que se pretende com a inclusão destes indicadores é garantir que o padrão escolhido tem um número considerável de clientes potenciais e foi realizado por alguns clientes Universo nos últimos 6 meses. Na escolha do melhor padrão sequencial teve-se em conta estes 2 indicadores.

Depois, foi realizada a análise referida na Subsecção 4.4.1 (com o algoritmo GSP). Com esta análise pretende-se encontrar padrões sequenciais que apesar de estarem presentes no top 10 podiam não ser a melhor opção pois tinham um último evento inadequado (a realização do último evento não levava com uma probabilidade elevada ao registo na App nos 3 dias subsequentes). Através da análise da Figura 5.6 é possível concluir que de facto alguns padrões não são interessantes nesta perspetiva. Por exemplo, os padrões que acabam com eventos transacionais e têm as categorias W, P e J não devem ser selecionados. Já padrões sequenciais que tenham como último evento I,H, F, D são mais indicados para comunicações em que se pretende converter o cliente no momento da comunicação.

Após terem sido realizadas as análises apresentadas, selecionaram-se os padrões sequenciais para o envio da comunicação (Figura 5.7). Escolher os melhores padrões sequenciais não é uma tarefa fácil (Fournier-Viger et al., 2017), contudo, tal como referido em Sinha et al. (2015) escolher os padrões com maior *confidence* faz sentido pois a campanha tem um determinado objetivo (registo na App) e a *confidence* dá-nos a probabilidade de o padrão atingir esse objetivo. Posto isto, foram selecionados os padrões com maior confidence de 3 e 4 elementos com e sem *gap*. Para além disso, decidiu-se selecionar os padrões <{E},{E}>, <{C},{T}> e <{E},{R}>. A escolha destes padrões sustenta-se em 2 razões:

- Estes padrões incluem eventos não encontrados noutros padrões. São por isso padrões inesperados pois não incluem as categorias D e F. De relembrar que um dos objetivos da aplicação do SPM é captar estes padrões "escondidos" e inesperados;

- Todos estes padrões apresentam uma boa probabilidade de levar à App após a realização do último evento, tal como constatado na Figura 5.6;

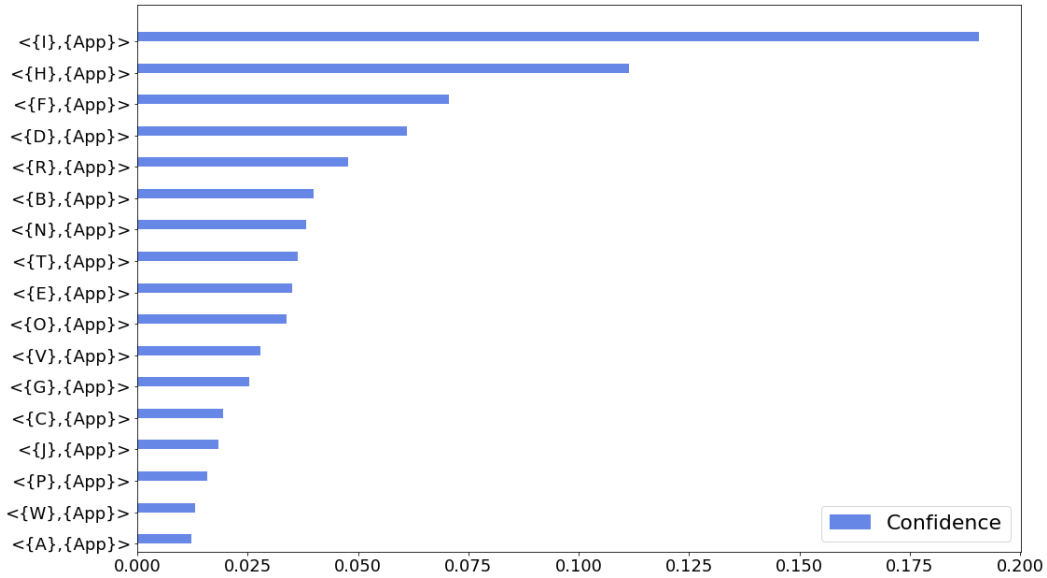



Figura 5.6: *Confidence* dos padrões sequenciais de 2 elementos com *gap* de 3 dias

| Padrões sequenciais selecionados | Gap de 60 dias | Sem gap |
|---|---|---|
| <{E},{E},{App}> | | x |
| <{C}, {T},{App}> | | x |
| <{E}, {R},{App}> | | x |
| <{O, D},{D},{D},{App}> | | x |
| <{F, O},{F},{App}> | x | |
| <{O},{F},{F},{App}> | x | |
| <{A},{B},{App}> | | x |

Figura 5.7: Padrões sequenciais selecionados para o envio da comunicação

## 5.5 Performance da Campanha Realizada

Após terem sido selecionados os padrões sequenciais, definiram-se os conteúdos das SMS a enviar para o cliente. O padrão sequencial <{A},{B}> destaca-se dos demais por não incluir eventos transacionais. Para além disso, muito provavelmente, o cliente que realizou este padrão estará numa fase do ciclo de vida diferente do que um cliente que tenha realizado uma transação e poderá ainda não conhecer todas as vantagens que a App Universo pode oferecer. Sendo assim,

foram criadas 2 SMS uma para o padrão sequencial <{A},{B}> e outra para os outros padrões sequenciais. As mensagens diferem na descrição das vantagens da App Universo. Na mensagem para o padrão sequencial <{A},{B}> tentou-se descrever vantagens genéricas da App Universo enquanto que a outra SMS descreve as vantagens da app para quem realiza muitas transações.

Nas campanhas realizadas no ano passado, tinham sido dados 15 dias para o cliente se registar na App Universo. Após uma análise cuidadosa da Figura 5.8 é possível constatar que o maior número de clientes regista-se nos primeiros 2/3 dias após receber o SMS sendo que o número de registos vai diminuindo à medida que nos afastamos do momento da comunicação pois o efeito da mesma começa a desvanecer-se. Sendo assim, foram dados 3 dias para o cliente se registar na App Universo. Este gráfico demonstra a valia da comunicação *event-driven*, isto porque se algo for oferecido no contexto mais adequado ao cliente certo a probabilidade de o cliente converter-se aumenta. Ou seja, para aumentarmos as taxas de conversão temos de selecionar o cliente mais propenso mas também escolher o momento mais adequado para comunicar o pretendido.

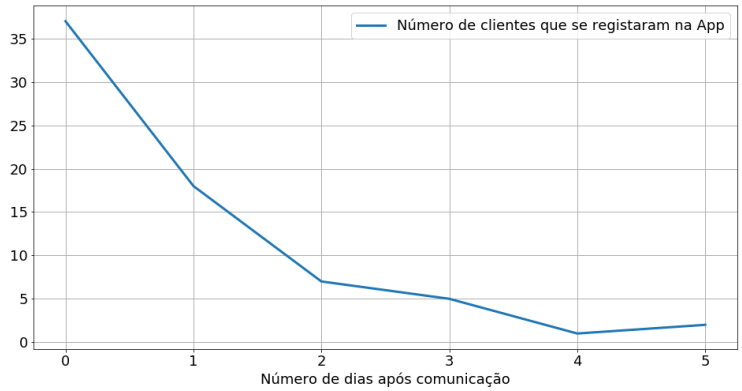



Figura 5.8: Número de registos após ter sido recebida a comunicação por dia

Na figura abaixo está apresentado o *dashboard* realizado com resultados de 7 dias de comunicação. Para cada subsequência tentou-se (com uma distribuição de probabilidade) agrupar 10% dos clientes no grupo de controlo. Isto foi realizado para que fosse possível obter o *uplift* por subsequência e, deste modo, comparar a performance entre os diferentes padrões sequenciais. Como podemos constatar pela análise do *dashboard*, todos os padrões sequenciais apresentaram um *uplift* positivo e uma taxa de conversão maior que 0.

De salientar também o facto de não ter havido significância estatística em todas as subsequência (com um nível de significância de 5%) para o teste Z (teste à diferença entre proporções binomiais) que compara as proporções de duas populações (Guimarães e Cabral, 2010). Com este teste pretendeu-se comparar as taxas de adesão do controlo e do alvo para cada padrão sequencial, com o objetivo de se averiguar se a taxa de adesão do grupo alvo era superior. O facto de não ter havido significância estatística pode estar relacionado com o tamanho da amostra (era necessário um maior número de clientes para tirarmos mais conclusões neste âmbito). No caso da campanha como um todo já houve significância estatística o que corrobora o facto de o tamanho das amostras

| | | Controlo | | | Alvo | | | Indicadores | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Padrão Sequencial | Canal | # Clientes | # Clientes sucesso campanha | [%] Taxa de adesão | # Clientes | # Clientes sucesso campanha | [%] Taxa de adesão | [p.p.] Uplift campanha | [%] p-value | Significância estatística |
| <{F, O},{F},{App}> | SMS | 7 | 0 | 0,000% | 61 | 1 | 1,639% | 1,639 | 36,64% | N |
| <{A},{B},{App}> | SMS | 21 | 0 | 0,000% | 182 | 18 | 9,890% | 9,890 | 6,56% | N |
| <{O},{F},{F},{App}> | SMS | 159 | 1 | 0,629% | 1606 | 22 | 1,370% | 0,741 | 21,60% | N |
| <{O, D},{D},{D},{App}> | SMS | 82 | 0 | 0,000% | 689 | 12 | 1,742% | 1,742 | 11,42% | N |
| <{E},{R},{App}> | SMS | 41 | 0 | 0,000% | 413 | 6 | 1,453% | 1,453 | 21,86% | N |
| <{E},{E},{App}> | SMS | 32 | 0 | 0,000% | 304 | 4 | 1,316% | 1,316 | 25,69% | N |
| <{C},{T},{App}> | SMS | 16 | 0 | 0,000% | 247 | 4 | 1,619% | 1,619 | 30,40% | N |
| **Total** | | 358 | 1 | 0,288% | 3502 | 67 | 1,913% | 1,625 | 1,26% | S |

Figura 5.9: *Dashboard* da campanha baseada nos padrões sequenciais

nos testes individuais poder estar a influenciar o resultado dos testes. Como este teste foi estatisticamente significativo podemos concluir que a campanha foi eficaz a selecionar os clientes alvo a receberem a comunicação.

Outro aspeto muito importante é a comparação da performance da comunicação entre os padrões sequenciais. Neste ponto, destaca-se o padrão sequencial <{A}, {B},{App}>. Os valores apresentados da taxa de adesão e de *uplift* deste padrão são bastante superiores aos mesmos dos outros padrões. Todos os outros padrões tiveram performances semelhantes com *uplifts* e taxas de conversão similares. Outro ponto importante é a comparação das sequências com e sem gap. Neste âmbito, a performance é similar se for excluído o padrão sequencial <{A},{B},{App}>. A performance superior do melhor padrão sequencial (<{A}, {B},{App}>) pode também estar relacionada com o conteúdo da mensagem (este padrão sequencial teve um conteúdo diferente dos demais). Para termos conclusões mais sólidas neste âmbito teriam de ter sido enviadas comunicações com diferentes SMS para cada padrão.

Foi também realizada uma comparação entre a *confidence* e as medidas de performance da campanha para cada padrão sequencial. Com esta análise pretende-se, de certa forma, validar a metodologia proposta, ou seja, perceber se de facto os padrões com maior *confidence* são aqueles com melhores resultados na comunicação. Como podemos visualizar na Figura 5.10, os padrões com maior *confidence* tiveram melhor performance na campanha. Veja-se, por exemplo, o padrão <{A},{B}>. Este padrão distingue-se pela sua *confidence* e acabou por revelar-se o padrão sequencial mais interessante na perspetiva de conversão digital após a receção da comunicação. Já o 2º padrão com a *confidence* mais baixa foi o que teve pior *uplift*. Fica assim comprovada a eficácia da metodologia proposta na busca pelos padrões sequenciais mais interessantes.

Mas mais importante que a comparação do desempenho dos diferentes padrões sequenciais é perceber se a campanha baseada nos padrões sequenciais apresentou uma performance superior a outras campanhas informativas com o mesmo objetivo de marketing. Na Tabela 5.2 encontram-se os *uplifts* e as taxas de adesão das 2 outras campanhas que tinham como objetivo promover a App Universo, os mesmos da campanha baseada no algoritmo de SPM e os *uplift* e taxa de adesão globais de todas as campanhas mensais realizadas em 2019. As campanhas 1 e 2 são campanhas *event-driven* em que a App é promovida após o cliente ter realizado determinados eventos. Estas campanhas começaram em abril de 2020 e atualmente ainda estão em produção, para além disso,

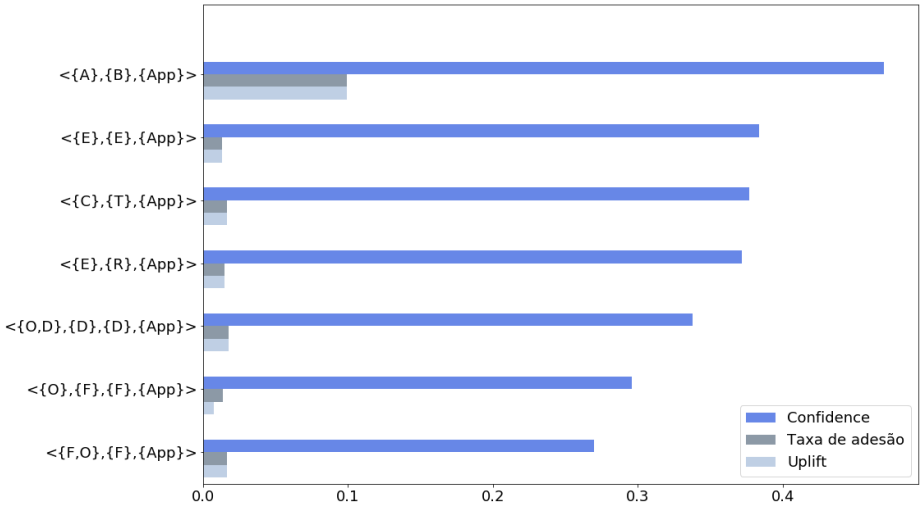



Figura 5.10: Comparação entre a *confidence* e as outras medidas de performance para cada padrão sequencial

nestas campanhas não foi criado grupo de controlo (não foi possível obter os *uplifts*). Os resultados das Campanhas 2019 correspondem a uma média ponderada do *uplift* e da taxa de adesão de todas as campanhas realizadas no ano de 2019. Nestas campanhas selecionaram-se os clientes *apriori* com critérios empíricos e depois enviaram-se SMS informativas para esses clientes (comunicação por *batch*).

Tabela 5.2: *Uplift* e taxa de adesão das diferentes campanhas

| **Medidas de Performance** | **Campanha 1** | **Campanha 2** | **Campanhas 2019** | **Campanha SPM** |
|---|---|---|---|---|
| *Uplift* (pontos percentuais) | - | - | 0,35 | 1,63 |
| Taxa de Adesão | 1,615% | 0,885% | 1,855% | 1,913% |

Como podemos constatar pela análise da Tabela 5.2, a campanha SPM foi a melhor campanha nas 2 medidas de performance. Podemos afirmar isto porque a campanha teve taxa de adesão e *uplift* superiores aos valores de 2019 e o *uplift* da campanha foi superior às taxas de adesão das campanhas 1 e 2 o que na prática quer dizer que em nenhuma circunstância as campanhas 1 e 2 teriam *uplifts* superiores à campanha baseada nos padrões sequenciais mesmo que fosse possível obter os *uplifts* destas campanhas. Por outro lado, é necessário referir que nas comunicações de 2019 foram dados 15 dias para os clientes se registarem na App o que pode explicar a aproximação entre as taxas de adesão destas comunicações e da Campanha SPM. Mesmo assim, e de forma a comprovar a supremacia da Campanha SPM calculou-se o valor da taxa de adesão se tivessem sido dados 7 dias para o cliente se registar na App e obteve-se o valor de 2,03% o qual é bastante superior à média ponderada da taxa de adesão de 2019, logo, podemos afirmar que a campanha baseada nos padrões sequenciais foi até este momento a melhor campanha da Sonae FS a promover a App Universo com mensagens informativas.

# Capítulo 6

# Conclusão

## 6.1 Conclusões do projeto

Selecionar o cliente mais propenso a receber uma comunicação já não é suficiente num mundo em que os clientes são "bombardeados" com comunicações de empresas a promover os seus produtos/serviços. É com isso em mente que surge este projeto de dissertação, no qual se pretende juntar as dimensões tempo e propensão através de uma abordagem analítica que ainda não tinha sido testada para se aumentarem as taxas de adesão das comunicações realizadas pela Sonae FS.

Aplicou-se esta abordagem num caso de uso muito específico - aumentar o número de clientes registados na App Universo. Este foi o caso de uso escolhido pois para além de cada conversão digital corresponder a uma poupança de 6,10€/ano, a empresa pretende agir numa perspetiva centrada no cliente e a App Universo acrescenta valor para um cliente Universo ativo de diversas formas. Para além disso, as comunicações a promover a App Universo no último ano não têm tido resultados satisfatórios e o número de clientes digitais apesar de estar a evoluir positivamente ainda não é suficientemente elevado.

Deste modo, e aplicando uma metodologia disruptiva desenvolveu-se um modelo analítico baseado num algoritmo de *Sequential Pattern Mining* que captasse padrões sequenciais interessantes (padrões que aumentem a probabilidade de levar ao registo da App) para posteriormente serem enviadas comunicações para os clientes que não têm App e tenham realizado estes padrões. A comunicação realizada foi uma comunicação *event-driven* para que no projeto se tivesse em conta a dimensão tempo. Ao longo do projeto passou-se pelas diferentes fases da metodologia CRISP-DM pois o objetivo era solucionar um problema de negócio com um algoritmo avançado de *data mining*. Para além disso, tentou-se sempre fazer a "ponte" entre o modelo analítico e a perspetiva *customer-centric* até porque se se pretende interagir de forma mais eficaz com os clientes temos de ter um modelo capaz de responder às necessidades e vontades de cada cliente.

Depois de ter sido desenvolvida a solução analítica e de se terem enviadas 1352 comunicações para os clientes-alvo, conseguiu-se numa semana aumentar em 67 o número de clientes com App Universo com um *uplift* de 1,63 pontos percentuais. Com o projeto foi obtida uma poupança baixa (408,7€/ano), contudo, estes dados correspondem apenas a 7 dias de comunicações e a campanha

ainda está a decorrer sendo que são esperados mais registos nas próximas semanas. A campanha realizada foi até ao momento a melhor campanha a promover a App através de uma mensagem informativa. Para além disso, o melhor padrão sequencial de acordo com o algoritmo de SPM apresentou uma performance excecional com um *uplift* de 9,89 pontos percentuais.

A abordagem utilizada ainda não tinha sido testada em clientes "reais", logo, este projeto contribuí para o conhecimento pois acrescentou a fase do *deployment* às metodologias semelhantes à adotada. Testou-se, desta forma, uma nova abordagem que junta uma área relativamente recente em que têm sido desenvolvidos algoritmos cada vez mais eficientes (SPM) com uma área que impacta a performance empresarial e as interações cliente-empresa (Marketing).

Em suma, este projeto apresentou resultados prometedores, principalmente pela performance brilhante da melhor subsequência proposta pelo algoritmo, contudo, os resultados devem ser analisados e interpretados com cautela pois o número de clientes comunicados não foi muito elevado. Para se comprovar a utilidade de algoritmos deste tipo com comunicações *event-driven* devem ser realizados mais estudos que utilizem abordagens semelhantes e que comuniquem com um maior número de clientes.

## 6.2 Trabalhos futuros e outras possíveis abordagens

Os algortimos de *data mining* que sustentam comunicações de marketing devem ser constantemente aplicados nos dados mais recentes dos clientes isto é verdade para os modelos de propensão mas também para outro tipo de abordagens tal como a seguida no projeto de dissertação. Isto acontece, pois, se queremos responder eficazmente às necessidades dos clientes temos de "mapear o DNA" destes de forma exímia. Tal como um organismo vivo, este DNA não é estático e continuará a evoluir e a adaptar-se ao seu ambiente ao longo de toda a vida do cliente, logo, utilizar dados imóveis para sustentar comunicações de marketing não é a abordagem mais correta. Deste modo, um dos próximos passos a realizar na Sonae FS é tornar o processo de obtenção dos padrões sequenciais mais interessantes um processo automático em que os padrões dos clientes são captados de forma síncrona e frequente. Quando a *confidence* de um padrão sequencial apresentar um valor elevado (como neste caso <{A},{B},{App}>) devem ser enviadas comunicações para os clientes que não têm App e realizaram tais padrões.

Na presente dissertação, aquando da seleção das variáveis/eventos a incluir no modelo analítico, foram eliminados vários eventos. Apesar do aumento nos tempos de computação e consumo de memória, a inclusão de tais eventos pode resultar na obtenção de padrões sequenciais mais interessantes. É um *tradeoff* entre eficiência e eficácia, mas, na prática, outras abordagens poderiam apresentar padrões mais interessantes.

Neste âmbito, é importante frisar também a categorização dos eventos transacionais. De facto, poderiam ter sido criadas outras categorias tais como a inclusão do canal (online, loja física) ou a inclusão de outros montantes gastos na transação (500, 1000 euros). Com a inclusão de tais categorias provavelmente iriam ser obtidos padrões com mais *items* num *itemset* o que poderia

impactar a seleção de clientes a enviar a comunicação e, consequentemente, a performance da campanha realizada.

Seria importante a realização de novos estudos que apliquem algoritmos de SPM na área de marketing. Esses estudos podiam incluir outras medidas de performance. Por exemplo, seria interessante incorporar o conceito RFM (*Recency, Frequent and Monetary*) presente na literatura de marketing para definir *RFM sequential patterns*. Considera-se sequência RFM- uma sequência recente, frequente e que envolva elevados valores monetários (Chen et al., 2009). A inclusão de tal conceito poderia afetar a procura pelos padrões sequenciais mais interessantes

Por outro lado, neste estudo explorou-se o tópico da procura das sequências ótimas para lançar uma determinada campanha de marketing. Tal abordagem pode ser considerada limitada pois não se têm em conta as características sociais e demográficas do cliente. Esse problema pode ser ultrapassado com a criação de um algoritmo de *Sequential Pattern Mining* para diferentes segmentos de cliente.

Seria também interessante realizar uma comparação entre a performance das comunicações *event-driven* baseadas nos algoritmos de SPM e a performance das comunicações por *batch* baseadas noutros modelos de propensão (por exemplo, algoritmos de classificação). Essa comparação teria de ser realizada para o mesmo caso de estudo (por exemplo, o registo na App Universo), sempre que possível no mesmo período temporal para os resultados não serem enviesados por outros fatores externos.

Em suma, com este projeto de dissertação explorou-se uma união nunca antes testada. Esta união está a começar a dar os primeiros passos. Novos estudos são necessários para diferentes hipóteses serem validadas e conclusões serem retiradas. No fim, na área do *Marketing Analytics* o objetivo será sempre o de realizar modelos cada vez mais eficazes que permitam melhorar a interação cliente-empresa e, consequentemente, a performance organizacional.

# Bibliografia


Agrawal, R. e Srikant, R. (1995). Mining sequential patterns. Em *International Conference on Data Engineering*, pgs. 3–14, Taipei, Taiwan. IEEE.

Alafeef, M., Singh, D., e Ahmad, K. (2012). The influence of demographic factors and user interface on mobile banking adoption: A review. *Journal of Applied Sciences*, 12(20):2082–2095.

Baptista, G. e Oliveira, T. (2015). Understanding mobile banking: The unified theory of acceptance and use of technology combined with cultural moderators. *Computers in Human Behavior*, 50:418–430.

Ben Baccar, L., Djebali, S., e Guérard, G. (2019). Tourist's Tour Prediction by Sequential Data Mining Approach. Em *Advanced Data Mining and Applications*, pgs. 681–695, Dalian, China. Springer.

Bohling, T., Bowman, D., LaValle, S., Mittal, V., Narayandas, D., Ramani, G., e Varadarajan, R. (2006). CRM Implementation. *Journal of Service Research*, 9(2):184–194.

Bolton, M. (2004). Customer centric business processing. *International Journal of Productivity and Performance Management*, 53(1):1741–0401.

Buchta, C. e Hahsler, M. (2020). Package arulesSequences. `https://cran.r-project.org/web/packages/arulesSequences/arulesSequences.pdf`. Consultado em 23 de maio de 2020.

Chand, C., Amit, T., e Ganatra, A. (2012). Sequential Pattern Mining: Survey and Current Research Challenges. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(1):185–193.

Chang, W., Park, J. E., e Chaiy, S. (2010). How does CRM technology transform into organizational performance? A mediating role of marketing capability. *Journal of Business Research*, 63(8):849–855.

Chen, J.-H. e Wu, S.-I. (2016). The impact of customer relationship management and internal marketing on business performance: a comparison of lodging industries. *Total Quality Management & Business Excellence*, 27(1-2):17–33.

Chen, Y. L., Kuo, M. H., Wu, S. Y., e Tang, K. (2009). Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. *Electronic Commerce Research and Applications*, 8(5):241–251.

CMO (2018). Mastering Adaptive Customer Engagements - Report - Thought Leadership - CMO Council™. Relatório técnico, CMO, San Jose.

Cristino, P. (2012). *Mobile Banking: fatores determinantes na adesão*. Dissertação de mestrado, Escola Superior de Comunicação Social, Lisboa.

Cundari, A. (2015). *Customer-centric marketing : build relationships, create advocates, and influence your customers*, volume 1. Wiley, New Jersey, 1 ed.

Dadzie, K., Dadzie, C., e Winston, E. (2019). The Transitioning of Marketing Practices from Segment to Customer-Centric Marketing in the African Business Context: Toward a Theoretical Research Framework. *Journal of African Business*, 20(2):206–223.

Desai, N. A. e Ganatra, A. (2015). Buying scenario and recommendation of purchase by constraint based sequential pattern mining from time stamp based sequential dataset. Em *International Conference on Advanced Computing Technologies and Applications (ICACTA)*, volume 45, pgs. 166–175, Mumbai, India. Elsevier B.V.

Efraim Turban, O. S. U., Efraim Turban, Jay E. Aronson, Ting-Peng Liang, e Ramesh Sharda (2007). *Decision Support and Business Intelligence Systems*, volume 1. Prentice Hall, London, 9 ed.

Euromonitor (2019). Credit Cards in Portugal | Market Research Report | Euromonitor. `https://www.euromonitor.com/credit-cards-in-portugal/report`. Consultado em 11 abril de 2020.

Farhan, M. S., Abed, A. H., e Ellatif, M. A. (2018). A systematic review for the determination and classification of the CRM critical success factors supporting with their metrics. *Future Computing and Informatics Journal*, 3(2):398–416.

Følstad, A. e Kvale, K. (2018). Customer journeys: a systematic literature review. *Journal of Service Theory and Practice*, 28(2):196–227.

Fournier-Viger, P., Chun, J., Lin, W., Kiran, R. U., Koh, Y. S., e Thomas, R. (2017). A Survey of Sequential Pattern Mining. *Data Science and Pattern Recognition*, 1(1):54–77.

Fournier-Viger, P., Gueniche, T., e Tseng, V. S. (2012). Using partially-ordered sequential rules to generate more accurate sequence prediction. Em *Advanced Data Mining and Applications*, pgs. 431–442, Nanjing, China. Springer.

Fournier-Viger, P., Nkambou, R., e Nguifo, E. M. (2008). A Knowledge Discovery Framework for Learning Task Models from User Interactions in Intelligent Tutoring Systems. Em *MICAI 2008: Advances in Artificial Intelligence*, pgs. 765–778, Atizapán de Zaragoza, México. Springer.

France, S. L. e Ghose, S. (2019). Marketing analytics: Methods, practice, implementation, and links to other fields. *Expert Systems with Applications*, 119(1):456–475.

Gabadinho, A., Studer, M., Müller, N., Bürgin, R., Fonta, P.-A., e Ritschard, G. (2020). Package TraMineR. `https://cran.r-project.org/web/packages/TraMineR/TraMineR.pdf`. Consultado em 23 de maio de 2020.

Germann, F., Lilien, G. L., e Rangaswamy, A. (2013). Performance implications of deploying marketing analytics. *International Journal of Research in Marketing*, 30(2):114–128.

Guimarães, R. C. e Cabral, J. A. (2010). *Estatistica*. Verlag, Lisboa, 2 ed.

Gummesson, E. (2008). Customer centricity: Reality or a wild goose chase? *European Business Review*, 20(4):315–330.

Guo, F. (2017). Data Mining Techniques For Customer Relationship Management. Em *International Conference on Communication and Electronic Information Engineering (CEIE 2016)*, volume 1, pgs. 61–70, Guangzhou, China. Atlantis Press.

Han, J., Pei, J., Yin, Y., e Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1):53–87.

Hanssens, D. M. e Pauwels, K. H. (2016). Demonstrating the Value of Marketing. *Journal of Marketing*, 80(6):173–190.

Hassan, H. E. e Wood, V. R. (2020). Does country culture influence consumers' perceptions toward mobile banking? A comparison between Egypt and the United States. *Telematics and Informatics*, 46.

Hauser, W. J. (2007). Marketing analytics: The evolution of marketing research in the twenty-first century. *Direct Marketing: An International Journal*, 1(1):38–54.

He, Z., Xu, X., Huang, J. Z., e Deng, S. (2004). Mining class outliers: concepts, algorithms and applications in CRM. *Expert Systems with Applications*, 27(4):681–697.

Iacobucci, D., Petrescu, M., Krishen, A., e Bendixen, M. (2019). The state of marketing analytics in research and practice. *Journal of Marketing Analytics*, 7(3):152–181.

IDC (2019). IDC Forecasts Revenues for Big Data and Business Analytics Solutions Will Reach $189.1 Billion This Year with Double-Digit Annual Growth Through 2022. `https://www.idc.com/getdoc.jsp?containerId=prUS44998419`. Consultado em 8 de março de 2020.

Jeong, D.-Y., Kim, S.-M., e Yoon, D.-J. (2014). Customer Orientation and Organizational Performance: Mediating Role of CRM. *Advanced Science and Technology Letters*, 57(1):35–39.

Kale, S. H. (2004). CRM Failure and the Seven Deadly Sins. *Marketing Management*, 13(5):42–46.

Krasnikov, A., Jayachandran, S., e Kumar, V. (2009). The Impact of Customer Relationship Management Implementation on Cost and Profit Efficiencies: Evidence from the U.S. Commercial Banking Industry. *Journal of Marketing*, 73(6):61–76.

Kumar, V., Venkatesan, R., e Reinartz, W. (2008). Performance Implications of Adopting a Customer-Focused Sales Campaign. *Journal of Marketing*, 72(5):50–68.

Linoff, G. e Berry, M. J. A. (2011). *Data mining techniques : for marketing, sales, and customer relationship management*, volume 1. Wiley, New Jersey, 3 ed.

Lo, V. S. Y. (2002). The true lift model: a novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter*, 4(2):78–86.

Marktest, G. (2015). 650 mil utilizadores de mobile banking. `https://www.marktest.com/wap/a/n/id{~}1f35.aspx`. Consultado em 4 de junho de 2020.

Marktest, G. (2020). Um em cada três possuidores de conta bancária usa mobile banking. `https://www.marktest.com/wap/a/n/id{~}25be.aspx`. Consultado em 22 de março de 2020.

Masseglia, F., Cathala, F., e Poncelet, P. (1998). The PSP approach for mining sequential patterns. Em *Principles of Data Mining and Knowledge Discovery*, volume 1510, pgs. 176–184, Nantes, França. Springer.

Mendoza, L. E., Marius, A., Pérez, M., e Grimán, A. C. (2007). Critical success factors for a customer relationship management strategy. *Information and Software Technology*, 49(8):913–945.

Meyer, M. (2003). Sequence Mining in Marketing. Em *Exploratory Data Analysis in Empirical Research*, pgs. 504–513, Munique, Alemanha. Springer.

Minami, C. e Dawson, J. (2008). The CRM process in retail and service sector firms in Japan: Loyalty development and financial return. *Journal of Retailing and Consumer Services*, 15(5):375–385.

Ner Clinton, J. (2000). Step-by-step data mining guide. Relatório técnico, SPSS, Chicago.

Ngai, E. W., Xiu, L., e Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2):2592–2602.

Nwokah, N. G. (2009). Customer-focus, competitor-focus and marketing performance. *Measuring Business Excellence*, 13(3):20–28.

Payne, A. e Frow, P. (2005). A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, 69(4):167–176.

Ravi, R. R. e Sun, B. (2016). *Customer-centric marketing : a pragmatic framework*, volume 1. The MIT Press, Cambridge, 1 ed.

Rygielski, C., Wang, J. C., e Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24(4):483–502.

Schweizer, D., Zehnder, M., Wache, H., Witschel, H. F., Zanatta, D., e Rodriguez, M. (2016). Using consumer behavior data to reduce energy consumption in smart homes: Applying machine learning to save energy without lowering comfort of inhabitants. Em *14th International Conference on Machine Learning and Applications, ICMLA 2015*, pgs. 1123–1129, Miami, Flórida,USA. Institute of Electrical and Electronics Engineers Inc.

Shah, D., Rust, R. T., Parasuraman, A., Staelin, R., e Day, G. S. (2006). The Path to Customer Centricity. *Journal of Service Research*, 9(2):113–124.

Shaikh, A. A. e Karjaluoto, H. (2015). Mobile banking adoption: A literature review. *Telematics and Informatics*, 32(1):129–142.

Sheth, J. e Sisodia, R. (2012). *The 4 A's of Marketing: Creating Value for Customer, Company and Society*, volume 1. Routledge, London, 1 ed.

Sheth, J. N. e Sisodia, R. S. (2002). Marketing productivity: Issues and analysis. *Journal of Business Research*, 55(5):349–362.

Sheth, J. N., Sisodia, R. S., e Sharma, A. (2000). The Antecedents and Consequences of Customer-Centric Marketing. *Journal of the Academy of Marketing Science*, 28(1):55–66.

Sinha, R., Mehta, S., Bohra, T., e Krishnan, A. (2015). Improving marketing interactions by mining sequences. Em *Web Information Systems Engineering – WISE 2015*, volume 9418, pgs. 277–292, Miami, Flórida, USA. Springer.

Sonae FS (2020). Universo - Infinitamente simples. `https://www.universo.pt/`. Consultado em 11 de abril de 2020.

SONAE SGPS (2019). RELATÓRIO e CONTAS 2019 SONAE SGPS. Relatório técnico, Sonae SGPS, Maia.

SONAE SGPS (2020). Portefólio - Investidores - Sonae. `https://sonae.pt/pt/investidores/portefolio/`. Consultado em 11 de abril de 2020.

Srikant, R. e Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. Em *Advances in Database Technology*, pgs. 3–17, Avignon, Franca. Springer.

Wang, J., Han, J., e Li, C. (2007). Frequent Closed Sequence Mining without Candidate Maintenance. *IEEE Transactions on Knowledge and Data Engineering*, 19(8):1042–1056.

Wedel, M. e Kannan, P. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing*, 80(6):97–121.

Yohn, D. (2018). 6 Ways to Build a Customer-Centric Culture. *Harvard Business Review*.

Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning*, 42(1):31–60.

Zhao, Q. e Sourav S. Bhowmick (2003). Sequential Pattern Mining: A Survey. Relatório técnico, Nayang Technological University Singapore, Singapore.

# Anexo A

# Razões que levam à adoção de *m-banking*

| No. | ID | Description | Frequency |
|---|---|---|---|
| 1 | PEOU | Perceived ease of use | 27 |
| 2 | PU | Perceived usefulness | 23 |
| 3 | TRU | Trust | 16 |
| 4 | SI | Social influence/ subjective norm | 15 |
| 5 | RIS | Perceived risk | 14 |
| 6 | SE | Self-efficacy | 11 |
| 7 | COM | Compatibility (lifestyle, device) | 10 |
| 8 | FC | Facilitation conditions | 9 |
| 9 | COS | Cost | 8 |
| 10 | CRE | Credibility | 8 |
| 11 | CUL | Culture | 8 |
| 12 | DEM | Demographic factors | 8 |
| 13 | SA | Structural assurance | 8 |
| 14 | RA | Relative advantage | 6 |
| 15 | ATT | Attitude toward use/m-banking | 5 |
| 16 | SYSQ | System quality | 5 |
| 17 | TRI | Trialability | 5 |
| 18 | EE | Effort expectancy | 4 |
| 19 | IQ | Information quality | 4 |
| 20 | SEC | Security | 4 |
| 21 | ACC | Accessibility | 3 |
| 22 | CPX | Complexity | 3 |
| 23 | CONV | Convenience | 3 |
| 24 | NI | Need for interaction | 3 |
| 25 | PB | Perceived benefit | 3 |
| 26 | PBC | Perceived behavioral control | 3 |
| 27 | PERI | Personal innovativeness | 3 |
| 28 | PERE | Performance expectancy | 3 |
| 29 | SAT | Satisfaction | 3 |
| 30 | SPE | Speed | 3 |
| 31 | UA | Uncertainty avoidance | 3 |
| 32 | CA | Consumer awareness | 2 |
| 33 | IMA | Image | 2 |
| 34 | PRI | Privacy | 2 |
| 35 | PEIN | Perceived innovativeness | 2 |
| 36 | ENJ | Perceived enjoyment | 2 |
| 37 | SQ | Service quality | 2 |
| 38 | UBI | Ubiquity | 2 |
| 39 | AFF | Affordability | 1 |
| 40 | ALE | Alertness | 1 |
| 41 | BAN | Banking needs | 1 |
| 42 | BEI | Behavioral introspection | 1 |
| 43 | BEN | Benevolence | 1 |
| 44 | COIN | Conventional interface | 1 |
| 45 | DEV | Device type/ features | 1 |
| 46 | ECOF | Economic factor | 1 |
| 47 | EXPEC | Expectations | 1 |

| | | | |
|---|---|---|---|
| 48 | EXP | Experience | 1 |
| 49 | EXPER | Expertise | 1 |
| 50 | EXTI | External influence | 1 |
| 51 | FREP | Firm reputation | 1 |
| 52 | FLOW | Flow (experience) | 1 |
| 53 | FU | Frequency of usage | 1 |
| 54 | FF | Functional factor | 1 |
| 55 | IP | Information presentation | 1 |
| 56 | INTEG | Integrity | 1 |
| 57 | MPE | Mobile phone efficacy | 1 |
| 58 | OF | Organizational factor | 1 |
| 59 | PCOMP | Perceived competence | 1 |
| 60 | PELIT | Perceived elitism | 1 |
| 61 | PLBC | Perception of latest banking channels | 1 |
| 62 | PINV | Personal involvement | 1 |
| 63 | PERS | Personalization | 1 |
| 64 | REPU | Reputation | 1 |
| 65 | RD | Results demonstrability | 1 |
| 66 | SITNO | Situational normality | 1 |
| 67 | SC | Service compatibility | 1 |
| 68 | SADO | Services adopted | 1 |
| 69 | SF | Strategic factor | 1 |
| 70 | TBI | Tag-based interface | 1 |
| 71 | TCHA | Task characteristics | 1 |
| 72 | TF | Technological factor | 1 |
| 73 | TA | Technology anxiety | 1 |
| 74 | TECHA | Technology characteristics | 1 |
| 75 | TR | Technology readiness | 1 |
| 76 | IM | Individual mobility | 1 |
| 77 | UE | Utility expectancy | 1 |
| 78 | VIS | Visibility | 1 |
| 79 | IR | Innovation resistance | 1 |
| 80 | PD | Power distance | 1 |
| 81 | TTF | Task technology fit | 1 |
| 82 | IND | Individualism | 1 |
| 83 | LTO | Long-term orientation | 1 |
| 84 | MAS | Masculinity | 1 |

Figura A.1: Razões que levam à adoção de *m-banking*, adaptado de Shaikh e Karjaluoto (2015)

# Anexo B

# Definições de CRM

- CRM is an e-commerce application
- CRM is a term for methodologies, technologies, and ecommerce capabilities used by companies to manage customer relationships
- CRM is an enterprisewide initiative that belongs in all areas of an organization
- CRM is a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer
- CRM is about the development and maintenance of longterm, mutually beneficial relationships with strategically significant customers
- CRM includes numerous aspects, but the basic theme is for the company to become more customer-centric. Methods are primarily Web-based tools and Internet presence
- CRM can be viewed as an application of one-to-one marketing and relationship marketing, responding to an individual customer on the basis of what the customer says and what else is known about that customer
- CRM is a management approach that enables organizations to identify, attract, and increase retention of profitable customers by managing relationships with them
- CRM involves using existing customer information to improve company profitability and customer service
- CRM attempts to provide a strategic bridge between information technology and marketing strategies aimed at building long-term relationships and profitability. This requires "information-intensive strategies"
- CRM is data-driven marketing
- CRM is an enterprise approach to understanding and influencing customer behavior through meaningful communication to improve customer acquisition, customer retention,customer loyalty, and customer profitability

# Anexo C

# Descrição da App

A App Universo – Mobile Banking, Financiamento e Seguros é uma aplicação gratuita que lhe dá o controlo do seu Cartão Universo, acesso a financiamento 100% online e à subscrição de um conjunto de seguros a qualquer momento e em qualquer lugar. As nossas vantagens são:

**Consultas em tempo real:**

- Saldo das suas contas e do Cartão Continente associado;
- Informação sobre a próxima mensalidade;
- Histórico de extratos;
- Descontos acumulados em cada compra;
- Analisar os seus Gastos por Categoria (alimentação, vestuário, entretenimento, etc.);
- Analisar os seus Gastos por Modalidade de pagamento (Fim do Mês, 3x com juros, Débito, etc.);

**Compras online:**

- Cartões Virtuais para compras online em segurança;
- Pagamentos em 3x, 6x ou 12x em todos os sites do mundo;

**Gestão do seu orçamento:**

- Realizar pagamentos de Serviços e ao Estado;
- Reembolso em 3x, 6x ou 12x;
- Realizar transferências para a sua conta à ordem ou para qualquer outra conta bancária;
- Realizar carregamentos de telemóveis;
- Alteração da percentagem de pagamento das suas compras efetuadas na modalidade Fim do Mês;
- Alterar a modalidade de pagamento de compras ou pagamentos já realizados;
- Agendamento de Transferências;
- Dividir uma conta com outros clientes Universo;

**Gestão personalizada do seu Cartão:**

- Total controlo e Segurança;
- Alertas de Utilização;
- Configuração de limites por gastos;
- Alterar o PIN do seu cartão;
- Substituição do novo cartão;
- Bloquear temporariamente e/ou substituir o seu cartão em caso de roubo ou perda;
- Associar cartões de fidelização (Cartão Continente e Worten Resolve);
- Definir controlos para alertar ou recusar transações especificas, como Internet, *Contactless* , no estrangeiro, entre outras e histórico de notificações;
- *Download* do comprovativo de pagamento para todas as suas transações;

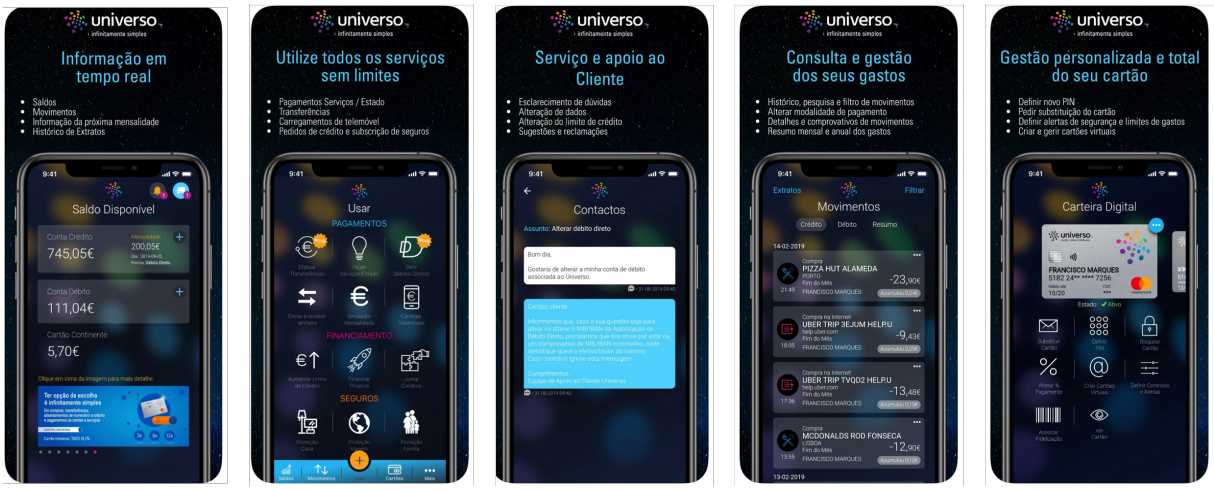



Figura C.1: Funcionalidades da App Universo

# Anexo D

# Rácio Eventos/cliente nos diferentes grupos de eventos

Tabela D.1: Eventos/cliente nos diferentes grupos de eventos

| | |
|---|---|
| Transacionais | 78,65 |
| Chamadas *Inbound* | 2,83 |
| *Web-Form* | 1,66 |
| Alterar e-mail | 1,09 |
| Alterar % revolving | 1,13 |
| Receber o extrato | 14,87 |
| Aderir a um produto *cross sell* | 1,07 |
| Aumentar o *plafond* | 1,08 |
| Ativar cartão | 1 |

# Anexo E

# Taxas de conversão dos eventos *Call-center*

| | | | |
|---|---|---|---|
| Aceder à App ou Universo Online | 38,37% | Alterar dados profissionais | 3,90% |
| Alterar código multicanal | 26,67% | (Des)Bloquear cartão | 3,78% |
| Conhecer as funcionalidades app ou universo online | 24,48% | Saber saldo em dívida | 3,62% |
| Seguro Proteção Viagem | 18,75% | Saber a mensalidade | 3,46% |
| Associar cartões fidelização | 15,60% | Saber o estado adesão | 3,42% |
| Efetuar pagamentos de serviços ou estado | 14,53% | Receber extratos | 3,42% |
| Alterar dados contacto | 10,72% | Esclarecer comissões ou encargos cobrados | 3,17% |
| Cancelar transações cativas | 10,45% | Obter reembolso de pagamentos duplicados | 3,01% |
| Ativar cartão definitivo | 8,81% | Seguro Crédito | 3,00% |
| Alterar modalidades pagamento | 8,55% | Perceber o extrato | 2,97% |
| Esclarecer campanhas promocionais | 8,47% | Obter cartão para 2º titular | 2,88% |
| Conhecer como funciona programa descontos | 7,43% | Efetuar pagamentos em atraso | 2,82% |
| Esclarecer transações não autorizadas por mim | 6,93% | Cancelar seguro | 2,77% |
| Alterar limite crédito | 6,87% | Conhecer oferta de crédito pessoal | 2,67% |
| Esclarecer serviços da linha de apoio universo | 6,28% | Mensalidade não cobrada | 2,63% |
| Obter reembolso devoluções compras | 6,28% | Saber data de receção do Cartão renovação | 2,62% |
| Alterar débito direto | 6,08% | Seguro Vida | 2,38% |
| Alterar % pagamento fim do mês | 5,59% | Alterar forma de envio do extrato | 2,37% |
| Esclarecer alertas transações suspeitas | 5,43% | Esclarecer linha crédito bloqueada | 2,30% |
| Esclarecer descontos | 5,27% | Seguro Protecção Casa | 2,12% |
| Sugestões e reclamações | 5,15% | Efetuar acordo de pagamento | 2,12% |
| Alterar PIN cartão | 5,06% | Alterar consentimentos marketing | 2,11% |
| Alterar dados identificação | 4,99% | Obter cartão substituição | 1,73% |
| Efetuar transferências | 4,81% | Aderir ao seguro | 1,65% |
| Esclarecer transações recusadas | 4,71% | Conhecer oferta de seguros | 1,55% |
| Saber data receção do cartão definitivo | 4,70% | Saber o motivo da recusa | 1,39% |
| Conhecer os ciclos do extrato | 4,66% | Conhecer as condições adesão | 1,17% |
| Conhecer funcionalidades do cartão | 4,59% | Ativar seguro | 0,82% |
| Obter declarações ou cópia contrato | 4,59% | Cancelar cartão | 0,38% |
| Obter comprovativo de pagamento ou faturas | 4,47% | Revogação de contrato | 0,00% |
| Saber saldo disponível | 4,42% | Diminuição ou reposição de linha de crédito | 0,00% |
| Efetuar pagamentos livres | 4,23% | Alteração de dados pessoais | 0,00% |
| Conhecer o preçário | 4,20% | Retirar titular da conta | 0,00% |
| NULL | 4,02% | Alteração de modalidade | 0,00% |
| Simular ou subscrever crédito pessoal | 3,91% | Ativar seguro | 0,00% |

Figura E.1: Taxas de conversão dos eventos *Call-center*

# Anexo F

# Taxas de conversão dos eventos *Web-form*

| | | | |
|---|---|---|---|
| Aceder à App ou Universo Online | 25,97% | Mensalidade não cobrada | 7,65% |
| Alterar código multicanal | 24,12% | Saber data receção do cartão definitivo | 7,63% |
| Associar cartões fidelização | 16,67% | Efetuar pagamentos em atraso | 7,31% |
| Conhecer as funcionalidades app ou universo online | 16,18% | Alterar débito direto | 7,21% |
| Seguro Vida | 15,79% | Conhecer oferta de crédito pessoal | 7,13% |
| Seguro Proteção Viagem | 15,00% | Esclarecer campanhas promocionais | 7,00% |
| Obter comprovativo de pagamento ou faturas | 14,95% | Saber o estado adesão | 6,93% |
| ADC | 14,29% | Esclarecer transações recusadas | 6,52% |
| (Des)Bloquear cartão | 13,40% | Conhecer como funciona programa descontos | 6,35% |
| Registar, ativar ou desbloquear app ou universo online | 12,50% | Obter reembolso de pagamentos duplicados | 6,21% |
| Alterar PIN cartão | 12,39% | Obter reembolso devoluções compras | 6,17% |
| Ativar cartão definitivo | 12,24% | Obter cartão para 2º titular | 6,14% |
| Alterar % pagamento fim do mês | 11,31% | Esclarecer descontos | 6,10% |
| Efetuar pagamentos de serviços ou estado | 11,25% | Conhecer o preçário | 6,06% |
| Seguro Protecção Casa | 10,45% | Alteração de dados pessoais | 5,99% |
| Simular ou subscrever crédito pessoal | 10,44% | Esclarecer linha crédito bloqueada | 5,81% |
| Saber a mensalidade | 10,36% | Cancelar transações cativas | 5,77% |
| Alterar dados contacto | 10,28% | Sugestões e reclamações | 5,49% |
| Receber extratos | 10,19% | Alterar dados profissionais | 5,43% |
| Conhecer funcionalidades do cartão | 10,12% | Cancelar seguro | 5,33% |
| Saber saldo disponível | 10,04% | Efetuar acordo de pagamento | 5,32% |
| NULL | 10,00% | Perceber o extrato | 5,26% |
| Aderir ao seguro | 9,76% | Alterar dados identificação | 5,19% |
| Alterar limite crédito | 9,62% | Obter cartão substituição | 5,00% |
| Efetuar transferências | 9,41% | Conhecer oferta de seguros | 5,00% |
| Obter declarações ou cópia contrato | 9,09% | Diminuição ou reposição de linha de crédito | 5,00% |
| Esclarecer serviços da linha de apoio universo | 8,60% | Esclarecer comissões ou encargos cobrados | 4,56% |
| Esclarecer transações não autorizadas por mim | 8,46% | Conhecer as condições adesão | 4,55% |
| Efetuar pagamentos livres | 8,45% | Outro | 4,27% |
| Alterar modalidades pagamento | 8,18% | Seguro Crédito | 4,13% |
| Esclarecer alertas transações suspeitas | 7,91% | Alterar forma de envio do extrato | 3,98% |
| Conhecer os ciclos do extrato | 7,89% | Saber data de receção do Cartão renovação | 3,92% |
| Saber o motivo da recusa | 7,85% | Revogação de contrato | 2,26% |
| Saber saldo em dívida | 7,70% | Cancelar cartão | 1,06% |
| Mensalidade não cobrada | 7,65% | Alterar consentimentos marketing | 0,00% |
| Saber data receção do cartão definitivo | 7,63% | Retirar titular da conta | 0,00% |
| Efetuar pagamentos em atraso | 7,31% | Alteração de modalidade | 0,00% |
| Alterar débito direto | 7,21% | Ativar seguro | 0,00% |

Figura F.1: Taxas de conversão dos eventos *Web-Form*

# Anexo G

# *Boxplots* do *Support* e da *Confidence* sem *gap*

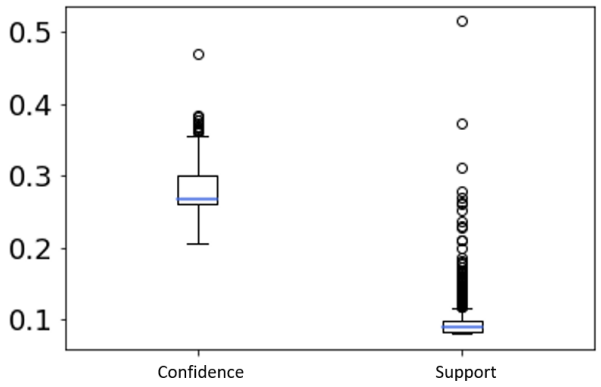



Figura G.1: *Boxplots* do *support* e da *confidence* sem gap

# Anexo H

# Análise descritiva aos valores de *support* e *Confidence* com *gap*

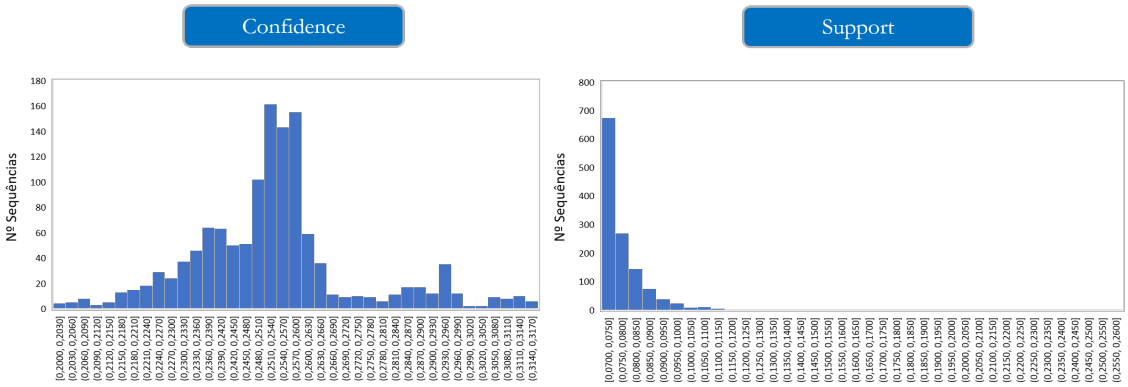



Figura H.1: Distribuições das sequências ao longo dos valores do *support* e da *confidence*

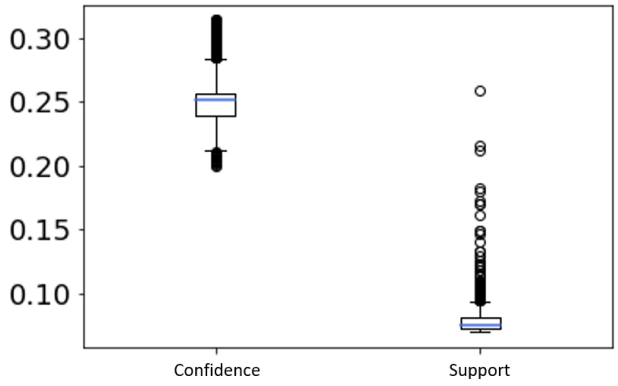



Figura H.2: *Boxplots* do *support* e da *confidence* com gap

# Anexo I

# 10 melhores padrões sequenciais de 2, 5 e 6 elementos sem gap

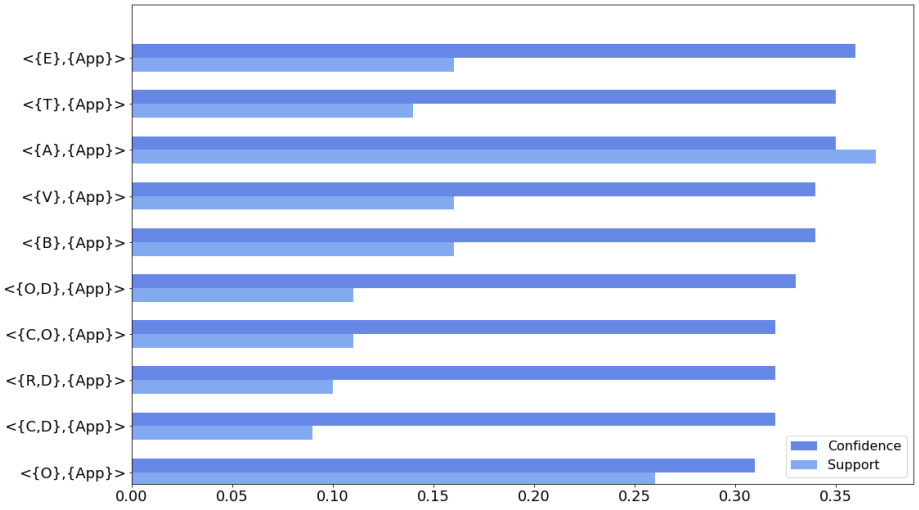



Figura I.1: Padrões sequenciais de 2 elementos

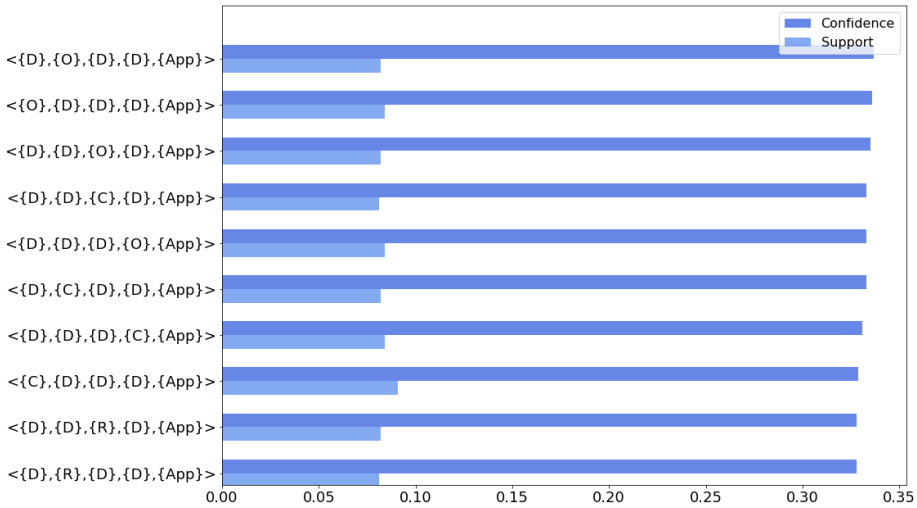



Figura I.2: Padrões sequenciais de 5 elementos

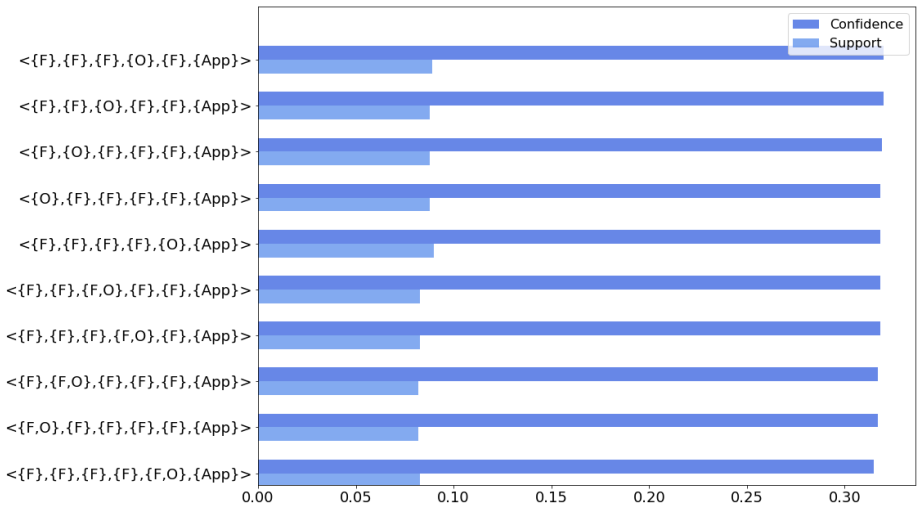



Figura I.3: Padrões sequenciais de 6 elementos

# Anexo J

# 10 melhores padrões sequenciais de 2, 5 e 6 elementos com gap de 60 dias

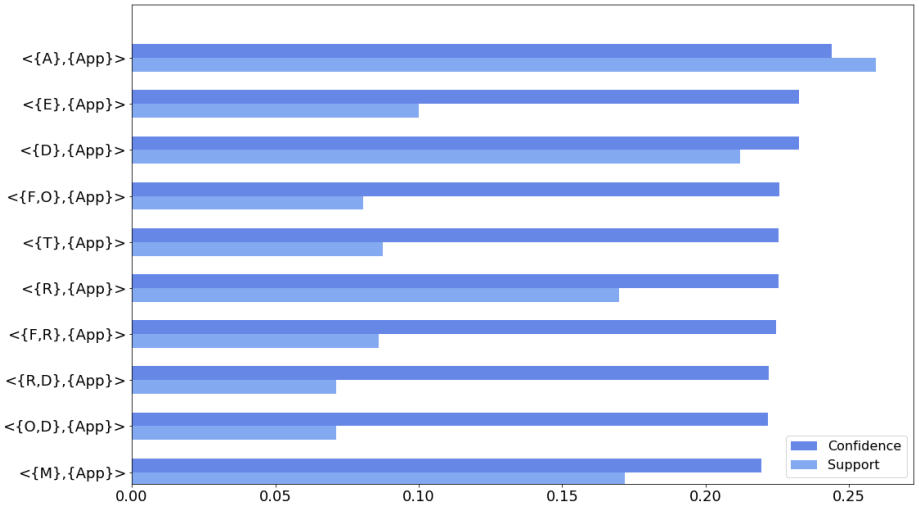



Figura J.1: Padrões sequenciais de 2 elementos

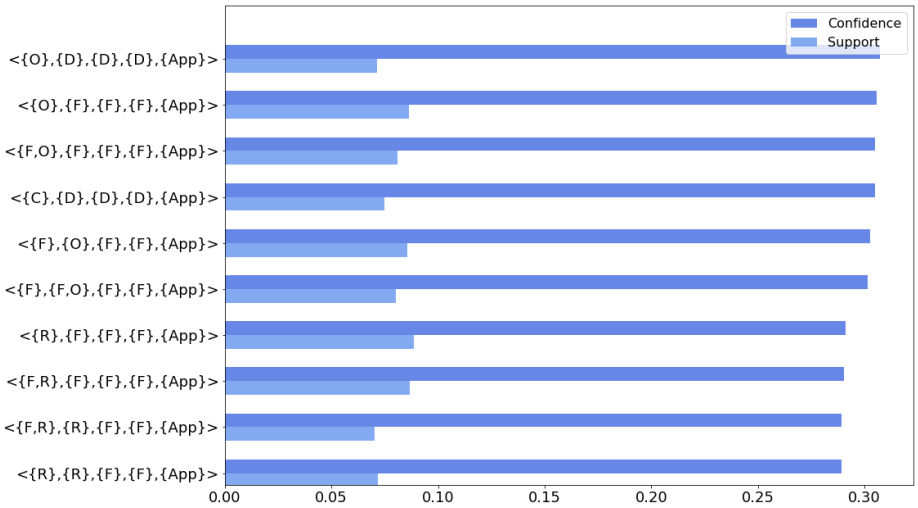



Figura J.2: Padrões sequenciais de 5 elementos

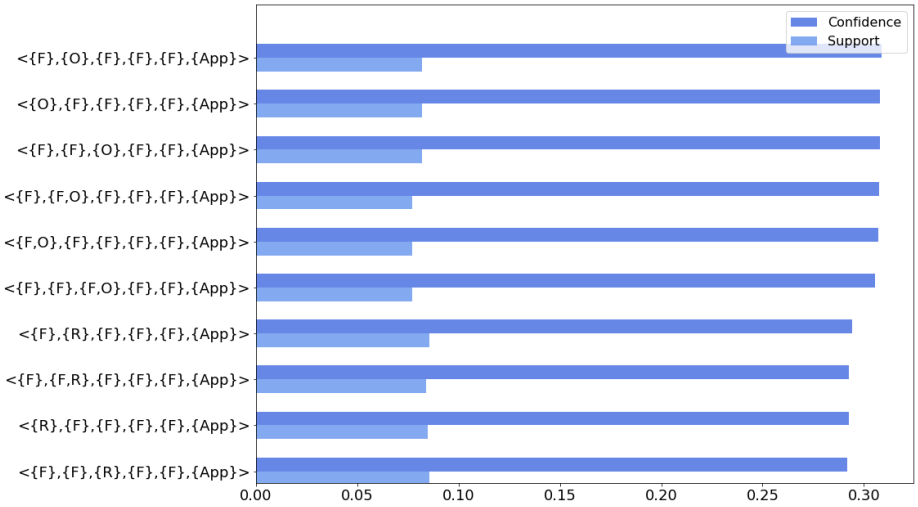



Figura J.3: Padrões sequenciais de 6 elementos

# Anexo K

# Indicadores das subsequências

| Padrões sequenciais sem gap | Nº clientes que fizeram a sequência | Nº clientes potenciais |
|---|---|---|
| <{A},{B},{App}> | 11 992 | 13 530 |
| <{E},{E},{App}> | 11 319 | 18 151 |
| <{O},{E},{App}> | 11 193 | 18 211 |
| <{C},{E},{App}> | 12 049 | 19 802 |
| <{C},{T},{App}> | 10 334 | 17 107 |
| <{E},{R},{App}> | 10 196 | 17 194 |
| <{E},{O},{App}> | 11 218 | 19 303 |
| <{E},{C},{App}> | 11 411 | 19 638 |
| <{T},{T},{App}> | 10 359 | 17 948 |
| <{O},{V},{App}> | 12 045 | 21 013 |
| <{O,D},{D},{D},{App}> | 10 304 | 20 180 |
| <{D},{O,D},{D},{App}> | 10 333 | 20 329 |
| <{C},{O},{R},{App}> | 10 781 | 21 290 |
| <{D},{D},{O,D},{App}> | 10 411 | 20 645 |
| <{C},{R},{O},{App}> | 10 668 | 21 204 |
| <{O},{O},{C},{App}> | 12 492 | 24 854 |
| <{O},{C},{O},{App}> | 11 847 | 23 632 |
| <{O},{O},{R},{App}> | 12 377 | 24 738 |
| <{O},{R},{C},{App}> | 10 335 | 20 666 |
| <{D},{O},{D},{App}> | 11 657 | 23 367 |

| Padrões sequenciais com gap | Nº clientes que fizeram a sequência | Nº clientes potenciais |
|---|---|---|
| <{F,O},{F},{App}> | 10531 | 28471 |
| <{O},{F},{App}> | 11502 | 31175 |
| <{O,D},{D},{App}> | 8861 | 24400 |
| <{R,D},{D},{App}> | 9107 | 25695 |
| <{R},{F},{App}> | 11682 | 32993 |
| <{O},{D},{App}> | 10317 | 29176 |
| <{F,R},{F},{App}> | 11328 | 32037 |
| <{R},{D},{App}> | 9793 | 27819 |
| <{D},{D},{App}> | 21517 | 64006 |
| <{O},{R},{App}> | 12143 | 37037 |
| <{O},{F},{F},{App}> | 11388 | 27108 |
| <{F,O},{F},{F},{App}> | 10576 | 25210 |
| <{O},{D},{D},{App}> | 9761 | 23715 |
| <{R,D},{D},{D},{App}> | 8869 | 22189 |
| <{R},{D},{D},{App}> | 9390 | 23664 |
| <{R},{F},{F},{App}> | 11666 | 29493 |
| <{F,R},{F},{F},{App}> | 11373 | 28840 |
| <{C},{D},{D},{App}> | 10518 | 27132 |
| <{F},{O},{F},{App}> | 10705 | 27789 |
| <{F},{F,O},{F},{App}> | 9951 | 25975 |

Figura K.1: Indicadores das subsequências