



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Modelo preditivo do comportamento de um doador: Caso de estudo Leigos para o Desenvolvimento

Frederico Miguel Oliveira Ladeira

Mestrado em Business Analytics

Orientadores:

Doutor Raul Manuel da Silva Laureano, Professor Associado,
Iscte Business School

Doutora Maria Madalena Eça Guimarães de Abreu, Professora
Adjunta, Coimbra Business School | ISCAC

Setembro, 2023



BUSINESS
SCHOOL

Departamento de Métodos Quantitativos para Gestão e
Economia

**Modelo preditivo do comportamento de um doador: Caso de
estudo Leigos para o Desenvolvimento**

Frederico Miguel Oliveira Ladeira

Mestrado em Business Analytics

Orientadores:

Doutor Raul Manuel da Silva Laureano, Professor Associado,
Iscte Business School

Doutora Maria Madalena Eça Guimarães de Abreu, Professora
Adjunta, Coimbra Business School | ISCAC

Setembro, 2023

*Dedico esta dissertação a todos os intervenientes
que acrescentaram valor à minha vida.*

Agradecimento

A realização desta dissertação descreveu-se numa experiência única e fascinante, da qual não era possível sem os contributos de alguns intervenientes.

Consequentemente, gostaria de aproveitar a oportunidade para agradecer ao Professor Doutor Raul Laureano e à Professora Doutora Madalena Eça de Abreu, pela sua sábia orientação, conhecimentos transmitidos e tempo despendido.

A minha gratidão alonga-se ao ISCTE e restantes professores da instituição, que a partir das suas práticas inovadoras de ensino contribuíram para o sucesso do meu percurso académico.

Um inevitável agradecimento à instituição não governamental, Leigos para o Desenvolvimento, pela sua disponibilidade e partilha dos seus dados.

Por fim, agradeço igualmente à minha família e amigos que demonstraram bastante compreensão perante a minha ausência, bem como à minha namorada pela motivação diária transmitida.

Um profundo obrigado!

Resumo

A presente dissertação procurou prever, a partir de dados históricos, qual o comportamento de um doador em campanhas de angariações de fundos num determinado ano. O principal objetivo deste estudo visa possibilitar a utilização de uma ferramenta analítica apta a prever com uma eficácia se um doador irá ou não doar para qualquer organização não governamental

A revisão sistemática da literatura incidiu, nomeadamente sobre variáveis, técnicas e métricas de avaliação utilizadas para prever o comportamento do doador.

A partir da metodologia CRISP-DM, foram desenvolvidos cinco modelos preditivos que possibilitam prever a intenção de doar. Além disso, para auxiliar a organização que cedeu os dados, foi desenhado um *dashboard* que responde às dúvidas sobre o comportamento dos seus doadores e das doações realizadas.

Finalizando, a técnica que obteve melhores resultados é um CART (93% de especificidade, 73,1% de sensibilidade, 82,8% de precisão, 0,7766 de medida F, 86,7% de *accuracy* e 0,901 de AUC), no qual, a partir das suas regras, é possível descrever com elevado grau de confiança três perfis propensos a doar e quatro propensos a não doar.

Palavras-chave: *Fundraising analytics*, Angariação de fundos, *Dashboard*, Organização sem fins lucrativos, ONGD, Modelo preditivo

Códigos JEL: M31, Y40

Abstract

This dissertation aimed to predict, based on historical data, the behavior of a donor in fundraising campaigns in a given year. The main objective of this study is to enable the use of an analytical tool capable of effectively predicting whether a donor will or will not donate to any non-governmental organization.

The systematic literature review focused on variables, techniques, and evaluation metrics used to predict donor behavior. Using the CRISP-DM methodology, it was developed five predictive models to predict donation intent. Additionally, to assist the organization that provided the data, it was designed a dashboard to answer questions about the behavior of their donors and the donations that were made.

In conclusion, the technique that achieved the best results was a CART (93% specificity, 73.1% sensitivity, 82.8% precision, 0.7766 F-measure, 86.7% accuracy, and 0.901 AUC), from which, based on its rules, it's possible to describe, with a high degree of confidence, three profiles inclined to donate and four profiles inclined not to donate.

Keywords: Fundraising analytics, Fundraising, Dashboard, Non-profit organization, NGO, Predictive model

JEL Codes: M31, Y40

Índice

Agradecimento.....	i
Resumo	iii
Abstract	iv
Índice.....	v
Índice Tabelas	ix
Índice de Figuras	xi
Abreviaturas, acrónimos e siglas	xiii
1 Introdução.....	1
1.1 Contexto e motivação	1
1.2 Questão de investigação	2
1.3 Objetivos e formas de validação.....	2
1.4 Contributos	3
1.5 Abordagem metodológica.....	3
1.6 Estrutura da dissertação.....	4
2 Revisão da literatura.....	7
2.1 Protocolo da revisão sistemática da literatura.....	9
2.2 Caracterização dos artigos	11
2.3 Síntese de conteúdos dos artigos.....	12
2.3.1 Contexto dos estudos	12
2.3.2 Técnicas e métricas de avaliação utilizadas	13
2.3.3 Variáveis preditoras do comportamento do doador.....	15
2.3.4 Contributos e limitações	16
2.4 Avaliação dos artigos da RSL.....	17
2.5 Sumário	17
3 Metodologia	21
3.1 Compreensão do Negócio	22
3.2 Compreensão dos dados.....	22
3.3 Preparação dos dados.....	25

3.3.1	Fluxo da preparação dos dados.....	26
3.3.2	Limpeza, transformação e criação de novas variáveis.....	27
3.3.3	Seleção dos dados	31
3.3.4	Criação do modelo de dados do <i>dashboard</i>	31
3.4	Modelação.....	33
3.4.1	Modelo preditivo	33
3.4.2	<i>Dashboard</i>	34
3.5	Avaliação.....	37
3.5.1	Modelo Preditivo.....	37
3.5.2	<i>Dashboard</i>	39
3.6	Implementação	39
4	Resultados e discussão.....	41
4.1	Modelos preditivos.....	41
4.1.1	Avaliação dos modelos preditivos.....	41
4.1.2	Importância das variáveis	43
4.1.3	Perfis de doadores com elevada e baixa propensão a doar.....	45
4.2	<i>Dashboard</i>	46
4.2.1	Descrição e avaliação do <i>dashboard</i> final.....	46
4.2.2	Caracterização do comportamento dos doadores e dos donativos.....	48
5	Conclusão.....	53
5.1	Sumário da investigação.....	53
5.2	Contributos	54
5.3	Limitações	55
5.4	Pistas futuras.....	55
	Bibliografia.....	57
	Anexos	61
	Anexo A: Dados solicitados.....	61
	Anexo B: Dicionário de dados	62
	Anexo C: Partes do código SQL.....	67

Anexo D: Preditores presentes na RSL.....	68
Anexo E: <i>Output</i> do CART-1	69

Índice Tabelas

Tabela 2.1 Critérios de exclusão e inclusão da RSL	10
Tabela 2.2 Critérios de avaliação dos artigos da RSL.....	11
Tabela 2.3 Artigos incluídos na RSL, publicados entre 2007 e 2022.....	12
Tabela 2.4 Sumarização do contexto do estudo dos artigos da RSL	13
Tabela 2.5 Sumarização das técnicas e métricas de avaliação utilizadas nos artigos da RSL	14
Tabela 2.6 Sumarização das variáveis utilizadas nos artigos da RSL.....	15
Tabela 2.7 Limitações e contributos dos artigos da RSL.....	16
Tabela 2.8 Avaliação dos artigos da RSL	17
Tabela 3.1 Intervalo de tempo das variáveis classificadas como data (conjunto de dados 1)	23
Tabela 3.2 Estatísticas descritivas das variáveis qualitativas (conjunto de dados 1)...	24
Tabela 3.3 Estatísticas descritivas das variáveis quantitativas (conjunto de dados 1).	25
Tabela 3.4 Limpeza e transformação de dados	28
Tabela 3.5 Variáveis independentes criadas no SPSS Modeler.....	29
Tabela 3.6 Variáveis criadas para a realização do dashboard	30
Tabela 3.7 Parametrizações dos algoritmos utilizados.....	34
Tabela 3.8 Medidas criadas para a realização do dashboard	35
Tabela 3.9 Visual escolhido escolhidos pelos LD.....	36
Tabela 3.10 Matriz de confusão	37
Tabela 4.1 Resultados dos modelos obtidos nas diferentes parametrizações	42
Tabela 4.2 Importância das variáveis independentes.....	43

Índice de Figuras

Figura 2.1 Síntese do processo de seleção de artigos.....	10
Figura 3.1 CRISP-DM	21
Figura 3.2 Fluxo de preparação de dados.....	27
Figura 3.3 Modelo de dados do dashboard.....	32
Figura 4.1 N.º de vezes que cada preditor surge	44
Figura 4.2 Dashboard sem drill down.....	47
Figura 4.3 Dashboard com drill down.....	47
Figura 4.4 Percentagem do n.º de doações por faixa etária e sexo	48
Figura 4.5 Evolução do n.º de doações e do valor médio doado por ano	49
Figura 4.6 Evolução do valor doado por ano.....	49
Figura 4.7 Evolução do n.º de doadores por ano	50
Figura 4.8 Evolução do n.º de doadores fidelizados por ano.....	50
Figura 4.9 Características demográficas dos doadores mais propensos a doar.....	51
Figura 4.10 Previsão do comportamento do doador no ano 2023.....	52
Figura 4.11 Previsão da intenção de doar.....	52

Abreviaturas, acrónimos e siglas

AC- Atas de conferências

AUC- Área sob a Curva ROC (*Receiver operator characteristic*)

CRISP-DM- *Cross Industry Standard Process for Data Mining*

CRM- *Customer Relationship Management*

DAX- *Data Analysis Expressions*

E.g.- *exempli gratia*

LD- Leigos para o Desenvolvimento

MP- Modelos Preditivos

ONG- Organização Não Governamental

ONGD- Organização Não Governamental para o Desenvolvimento

PBI- *Power BI*

RNA- Redes Neurais Artificiais

RSL- Revisão Sistemática da Literatura

SP- *Stored procedures*

YTD- *Year to Date*

1 Introdução

1.1 Contexto e motivação

O número das organizações não governamentais (ONG) sem fins lucrativos tem vindo a aumentar significativamente de ano para ano, levando a que o estudo do *fundraising* (Aldashev & Verdier, 2009) se encontre na ordem do dia num paradigma mundial (Laurett & Ferreira, 2018), especialmente em países como Estados Unidos da América e Inglaterra (Villar *et al.*, 2009).

A acompanhar essa tendência de crescimento, simultaneamente, tem surgido a importância da prática do *fundraising* (Lindahl & Conley, 2002) a partir do conhecimento já adquirido. Alinhar o conhecimento teórico com os resultados práticos é essencial para a aplicação de estratégias de angariações de fundos bem-sucedidas, uma vez que só assim uma organização se encontra preparada para enfrentar os desafios e as nuances do mundo real.

Outra tendência que tem despertado é a inserção de novos conceitos práticos provenientes do mundo tecnológico, especialmente juntando o conceito de *analytics* ao conceito de *fundraising*. Atualmente, já existem alguns autores a defender o *fundraising analytics* e a abordar tal junção como o futuro do *fundraising*, devido à prática da temática no quotidiano das organizações, permitir entender melhor o comportamento e as preferências dos doadores, a fim de informar e melhorar as estratégias de captação de recursos (Birkholz, 2008).

Autores como Verhaert e Van den Poel (2011), Laureano *et al.*, (2015), Hershberger e Hair (2022) defendem que os dados recolhidos pelas organizações, ao serem trabalhados, desempenham um papel crucial no *fundraising*. Tais dados fornecem informações valiosas, nomeadamente para a tomada de decisões baseadas na segmentação de doadores, previsão de doações, avaliação de desempenho e personalização das abordagens a doadores. O uso estratégico de dados pode também levar a uma captação de recursos mais eficaz, relacionamentos mais fortes com os doadores e melhores resultados financeiros para as organizações sem fins lucrativos.

Todavia, em Portugal, o tema de *fundraising* é ainda pouco explorado (Almendra, 2014), sendo que é exequível afirmar que o conceito de *fundraising analytics* nunca foi aplicado numa organização portuguesa. Tal constatação advém de as ONG não estarem preparadas para resolver problemas complexos que requerem um vasto conhecimento sobre elas próprias, assim como conhecimentos aprofundados sobre *analytics* (Laureano *et al.*, 2015; Lu *et al.*, 2020). Nesse sentido, é crucial as ONG possuírem um sistema que permite aplicar o *analytics*,

de forma a solucionar os problemas a partir dos dados históricos já existentes nas organizações.

Para realizar esta investigação, solicitou-se à organização Leigos para o Desenvolvimento (LD) um conjunto de dados históricos, visto que, além de ser a única Organização Não Governamental para o Desenvolvimento (ONGD) em Portugal que possui um registo abrangente e confiável de dados (segundo um especialista na área de *analytics*), não tem nenhum sistema ou pessoa capaz de aplicar o *fundraising analytics*, quer a nível teórico, quer a nível prático.

Por conseguinte, a presente dissertação vem colmatar a lacuna do *fundraising analytics* no contexto português. A sua principal motivação é proporcionar a uma ONG um sistema que permite aplicar o *analytics* nas suas campanhas de angariação de fundos, com principal enfoque nas campanhas de marketing direto. Assim, é possível conseguir um conjunto de vantagens na gestão do *fundraising* desta ONGD como aumentar a fidelidade dos doadores, reduzir significativamente os custos e proporcionar melhores resultados na obtenção de fundos (Kristoffersen & Singh, 2004; Power *et al.*, 2018; Cacciarelli & Boresta, 2022).

Detalhadamente, o presente estudo prevê, a partir de dados históricos, o comportamento de um doador em campanhas de angariações de fundos num determinado ano, considerando como “comportamento” se o doador é ou não recetivo a doar. Outro aspeto da investigação é que, apesar das contribuições dos doadores privados poderem ser obtidas através de voluntariado e/ou de um valor monetário (Srnrka *et al.*, 2003; Lee & Chang, 2007), unicamente são consideradas doações monetárias, isto é, somente são consideradas doações de dinheiro, cheques, cedência de cartões de crédito ou transferências de ações, títulos e bancárias (Van Lange *et al.*, 2007).

1.2 Questão de investigação

Considerando o referido no subcapítulo anterior, especialmente, a consideração de não haver qualquer aplicação teórico-prática de *fundraising analytics* em Portugal, assim como a necessidade de criar um sistema para os LD que possibilite tomadas de decisões mais informadas e eficientes, a questão geral de investigação apresenta-se da seguinte forma: Qual o comportamento do doador dos LD em futuras angariações de fundos?

1.3 Objetivos e formas de validação

Após a definição da questão de investigação, são definidos três objetivos que permitem responder à questão:

- Criar um modelo preditivo da intenção de doar no próximo ano.

Este objetivo é dado como cumprido quando o melhor modelo desenvolvido atingir no mínimo 70% em todas as métricas de avaliação que o estudo considera relevantes e que o modelo desenvolvido tenha somente preditores que a maioria das organizações possuem de forma explícita ou implícita.

- Identificar perfis de doadores que possuem alta e baixa propensão a doar no próximo ano.

Este objetivo é cumprido quando, a partir do melhor modelo preditivo, se conhecer, através de regras com suporte superior a 25 doadores e confiança superior a 80%, quais são os perfis de doadores mais e menos propensos a doar.

- Criar um *dashboard* para caracterizar o comportamento dos doadores e dos donativos realizados.

Este objetivo é cumprido quando a satisfação dos potenciais utilizadores do *dashboard* for elevada.

1.4 Contributos

Esta dissertação, especificamente considerando o domínio teórico, possibilita a apresentação e concretização do que poderá ser o conceito e prática do *fundraising analytics* na comunidade portuguesa. Simultaneamente, permite complementar o conhecimento já adquirido pela comunidade mundial.

Do ponto de vista prático, o presente estudo possibilita aos LD um sistema semiautomático apto a prever o comportamento dos doadores num determinado ano, bem como proporciona um *dashboard* sobre o comportamento dos seus doadores e dos donativos realizados. De outro ponto de vista, embora os preditores possam mudar de organização para organização, a investigação, ao utilizar variáveis presentes direta ou indiretamente nas bases de dados da maioria das organizações, possibilita a replicação do sistema preditivo a qualquer entidade interessada.

1.5 Abordagem metodológica

A presente investigação assume um cariz quantitativo, em que se aplica um conjunto de técnicas quantitativas de análise de dados para concretizar os seus objetivos. Por outro lado, assume-se como um estudo de caso, já que toda a investigação se enquadra no contexto de uma organização, os LD.

Os LD foram fundados em Portugal em 1986, tendo como inspiração os princípios da Companhia de Jesus (vulgarmente conhecidos como Jesuítas) e a missão de servir e promover a justiça social. Assume-se como uma organização não governamental para o

desenvolvimento (ONGD) com o objetivo de promover o desenvolvimento humano integral em países do continente africano e da América do Sul. Atua em projetos sociais, nos quais tenta melhorar as condições de vida de comunidades mais vulneráveis, por meio da cooperação, formação e educação. Complementarmente, em parceria com outras organizações locais e internacionais, visa capacitar as comunidades locais para que possam superar a pobreza, a exclusão social e outras desigualdades.

Os projetos desenvolvidos pelos LD abrangem diversas áreas, como educação, saúde, desenvolvimento comunitário, formação profissional, empreendedorismo e fortalecimento de capacidades locais, tendo sempre em consideração a cultura, os recursos disponíveis e os desafios enfrentados.

Além dos projetos e serviços desenvolvidos pelos LD, a organização promove o voluntariado internacional, permitindo aos voluntários oportunidades para que pessoas de diferentes origens e formações possam contribuir com o seu tempo e habilidades em projetos no exterior. Adicionalmente, possibilita aos voluntários vivenciar e compreender as diferentes realidades das comunidades onde a organização atua, promovendo também o intercâmbio cultural.

Atendendo ao âmbito do estudo, adotou-se a metodologia Cross-Industry Standard Process for Data Mining (CRISP-DM). O CRISP-DM visa uma interação acentuada das atividades analíticas realizadas com área da gestão e do negócio, o que justifica o permanente apoio e o acompanhamento dos LD e de dois especialistas, um da área de *fundraising* e outro de *analytics*. A metodologia inclui seis etapas. Na primeira etapa da metodologia, são definidos os objetivos e os critérios de sucesso. Nas etapas dois e três, compreensão e preparação dos dados, análise e preparação dos dados históricos dos donativos realizados no período de 1/01/2015 a 31/12/2022, assim como os dados sobre as características dos doadores. De seguida, na etapa quatro e cinco, criam-se e avaliam-se os modelos preditivos e o *dashboard* criado. Por fim, a última fase, a de implementação, restringe-se a esta dissertação e ao protótipo do *dashboard*.

1.6 Estrutura da dissertação

Esta dissertação está estruturada em cinco capítulos, incluindo esta introdução em que se apresenta o contexto do estudo, a questão de investigação e objetivos e a abordagem metodológica adotada.

No Capítulo 2, é apresentada a revisão sistemática da literatura (RSL), de forma a perceber o que já foi desenvolvido por outros autores sobre previsão de donativos. Assim, inicialmente, são descritos o protocolo e o processo de pesquisa utilizado, sendo demonstrado como é sistematizada a extração da informação importante. Posteriormente, é sintetizado o

conteúdo de cada artigo e esquematizadas as ideias relevantes a reter para o presente estudo. No final, é realizada a avaliação dos artigos presentes na RSL.

No Capítulo 3, é descrito toda a metodologia CRISP-DM e as tarefas executadas em cada uma das suas fases (compreensão do negócio, compreensão de dados, preparação de dados, modelação, avaliação e a implementação), dando-se um foco maior às fases de compreensão de dados, preparação de dados e modelação.

No Capítulo 4, são sintetizados todos os resultados da investigação e é feita a avaliação e discussão dos mesmos.

No Capítulo 5, são apresentadas as principais conclusões e contributos do estudo, as limitações sentidas ao longo da execução da investigação e ainda as sugestões para pesquisas futuras.

2 Revisão da literatura

As ONG são instituições sem fins lucrativos, orientadas por missões que cumprem várias tarefas e funções na sociedade, especialmente nas áreas de assistência social, bem-estar, saúde, arte e cultura, proteção ambiental e animal (Helmig & Boenigk, 2012). Muitas destas organizações têm como missão muito específica o apoio prioritário aos países em vias de desenvolvimento, focando os seus esforços no desenvolvimento internacional e no aumento do bem-estar dos cidadãos destes países (Werker & Ahmed, 2008). Como resultado deste vasto campo de atuação, tais organizações revelam-se cada vez mais uma parte integrante de todas as sociedades modernas, o que leva a que o número de ONG aumente significativamente de ano para ano, ao passo que o número de doações voluntárias permanece relativamente constante. Consequentemente, a tendência descrita tem gerado desafios para o funcionamento eficaz das ONG, uma vez que, com a inúmera concorrência existente, torna-se desafiante a captação de recursos financeiros que levam ao cumprimento das suas missões (Hommerová & Severová, 2019).

O *fundraising* desempenha um papel crucial na captação de fundos de qualquer ONG. Ao ser considerado uma atividade sistemática destinada a obter recursos financeiros que permitem à organização a realização das suas atividades, assume um papel fundamental na sustentabilidade financeira de uma ONG (Hommerová & Severová, 2019). Mais precisamente, o *fundraising* permite às instituições sem fins lucrativos posicionarem-se para a obtenção de subsídios e contribuições provenientes através do voluntariado, da venda de serviços, das taxas de sócios e das doações em dinheiro (Srunka *et al.*, 2003; Lee & Chang, 2007; Abreu 2012; Anheier, 2014).

Adicionando ao *fundraising* o valor do *analytics*, ou seja, a aplicação de processos e técnicas que transformam dados brutos em informações fundamentais para melhorar a tomada de decisões (Wilder & Ozgur, 2015), nomeadamente, a partir da análise descritiva, preditiva ou prescritiva (Gandomi & Haider, 2015; Bradlow *et al.*, 2017), surge o conceito de *fundraising analytics*.

Segundo Birkholz (2008), *fundraising analytics* pressupõe a utilização de técnicas de análise de dados para entender o comportamento e as preferências dos doadores, com o objetivo de informar e melhorar as estratégias de captação de recursos. Mais precisamente, o conceito abrange a recolha e a interpretação de dados sobre doações, doadores e atividades de angariações de fundos, através da utilização de técnicas estatísticas e de *machine learning*. Por outro lado, o objetivo passa por informar as ONG sobre as preferências dos seus doadores e futuros comportamentos, o que possibilita a personalização das campanhas de angariações de fundos consoante o perfil de cada doador.

Sucessivamente, as vantagens do *fundraising analytics* apresentam aspetos bastante convenientes, isto é, os *insights* que são gerados a partir do seu uso permitem compreender os doadores, otimizar processos internos, identificar reduções de custos e oportunidades de crescimento (Marr, 2016; Power *et al.*, 2018). O resultado da prática de *fundraising analytics* traduz-se numa criação incalculável de valor para a tomada de decisão, ou, melhor referindo, na criação de uma robusta vantagem competitiva perante um setor que cada vez apresenta mais concorrência, visto que existem mais organizações e o número de doações mantém-se constante (Hindle *et al.*, 2020; Kraus *et al.*, 2020).

Uma das aplicações da análise preditiva em *fundraising analytics* é a previsão do comportamento do doador. Por meio da análise preditiva, é exequível obter *insights* valiosos sobre quais os doadores mais propensos a contribuir numa determinada campanha de angariação de fundos (Birkholz, 2008), o que possibilita às organizações a oportunidade de uma redução acentuada de custos despendidos com o doador que não irá doar (Waller & Fawcett, 2013).

Similarmente, a previsão do comportamento do doador possibilita a identificação dos doadores em risco de abandono da organização. Deste modo, a organização possui um último momento para fidelizá-los novamente (Birkholz, 2008; Sargeant, 2010; Cacciarelli & Boresta, 2022).

O comportamento do doador (o valor, a forma e a frequência em que ocorre) depende, essencialmente, das suas características demográficas e sociodemográficas (Srnlka *et al.*, 2003). Consequentemente, o presente estudo, tendo como principal questão de investigação “qual é o comportamento do doador em futuras angariações de fundos” com foco na conceção num modelo preditivo, é imprescindível conhecer o que já foi realizado mundialmente, considerando variáveis demográficas e sociodemográficas utilizadas na criação de modelos preditivos sobre o tema. Torna-se igualmente importante conhecer as técnicas que possibilitam a previsão e as suas limitações.

Deste modo, este capítulo sintetiza, numa RSL, o conhecimento sobre a previsão do comportamento de um doador, contribuindo para a realização dos objetivos da dissertação e para a escolha da metodologia a adotar. Em concreto, a RSL visa responder à pergunta inicial: “Qual o melhor método e variáveis a usar na previsão do comportamento do doador em futuras angariações de fundos?” Para o efeito, definem-se quatro questões específicas, que possibilitam estruturar as conclusões obtidas:

- I. Qual o âmbito da investigação em que se utiliza previsão do comportamento do doador?
- II. Quais as técnicas e as métricas de avaliação utilizadas para prever o comportamento do doador?
- III. Quais as variáveis utilizadas para prever o comportamento do doador?

IV. Quais as limitações e os contributos do estudo?

De acordo com Tranfield *et al.* (2003), Siddaway *et al.* (2019) e Galvão e Ricarte (2019), a RSL é uma abordagem utilizada para identificar, avaliar e sintetizar a literatura. Nesta dissertação, o protocolo da RSL é baseado na metodologia PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*), de forma a reduzir o enviesamento de resultados durante o processo da revisão da literatura (Rethlefsen *et al.*, 2021), tornando-o mais transparente, completo e preciso (Liberati *et al.*, 2009; Page *et al.*, 2021).

O protocolo descreve minuciosamente o processo de uma revisão sistemática, iniciando com a definição dos objetivos e questões de pesquisa, já supramencionado. A partir dessas informações, é elaborada uma *query* e definidos critérios de inclusão e exclusão para selecionar os artigos pertinentes para as questões formuladas. Posteriormente, todos os artigos selecionados são analisados e resumidos numa tabela. Por fim, cada artigo é avaliado individualmente com o intuito de identificar os mais adequados ao tema em estudo, bem como os conteúdos mais e menos abordados na literatura.

A RSL, apesar de ser recente em estudos relacionados com a área de *fundraising*, é, em muitos aspetos, melhor do que as abordagens tradicionais (Kollmann *et al.*, 2017; Gazley, 2022).

2.1 Protocolo da revisão sistemática da literatura

Definidos o objetivo e as questões, segue-se um processo de seleção de artigos composto por três fases: identificação, exclusão e inclusão.

Na fase da identificação, são selecionados artigos de duas bases de dados científicas — a *Web of Science* e a *Scopus* —, uma vez que, apesar de a área do *fundraising* não ser consensual (Abreu, 2012), segundo Mongeon e Paul-Hus (2016), ambas adequam-se às possíveis áreas em que o *fundraising* está inserido.

A partir das bases de dados científicas referidas, selecionam-se os artigos que atendem no título à seguinte *query*, validada anteriormente por um especialista em *fundraising* e tendo como base na sua redação Santos *et al.* (2020): (*donor* or donation* or fundraising or fundraisers or "fund-raising" or "raise funds" or "request" or funding or journey or attitude* or behaviour or behavior or giving or give or solicitation*) and (*predict**) and (*"charities" or "charitable sector" or "charitable giv*" or "charity Fundraiser" or "non-profit" or nonprofit* or npo* or "nongovernmental sector" or "nongovernmental organizations" or ngo or "non-governmental organization*" or "third sector organization*" or "voluntary sector" or "voluntary organisation" or "voluntary organization"*).

Da aplicação da *query*, resultam 23 estudos provenientes da *Scopus* e 18 da *Web of Science*. Na segunda fase, é feita uma triagem segundo os critérios de exclusão, isto é, rejeitando-se artigos que cumprem os critérios de exclusão e, por fim, na última fase, tendo em conta os critérios de inclusão, selecionam-se os artigos a incluir na RSL.

Os critérios de inclusão e de exclusão estão representados na Tabela 2.1.

Tabela 2.1 Critérios de exclusão e inclusão da RSL

Critérios de exclusão
<ul style="list-style-type: none"> • Documentos que não sejam artigos publicado em <i>journals</i> ou atas de conferência • Artigos sem possibilidade de <i>download</i> • Artigos duplicados
Critérios de inclusão
<ul style="list-style-type: none"> • Artigos que auxiliam a criação de variáveis para a previsão do comportamento do doador • Artigos que abordam a previsão do comportamento do doador • Artigos que abordam a previsão do ato de doar ou da doação monetária

Detalhadamente, a partir dos critérios de exclusão, selecionam-se 18 artigos. Posteriormente, através da aplicação dos critérios de inclusão ao título e *abstract*, são selecionados novamente os 18 artigos, não excluindo nenhum já que todos se enquadravam nos critérios. Por fim, após a leitura na íntegra e a aplicação dos critérios de inclusão, são considerados nove artigos para serem analisados. Todo o processo de seleção está sintetizado na Figura 2.1, que demonstra o número de documentos após aplicado cada critério.

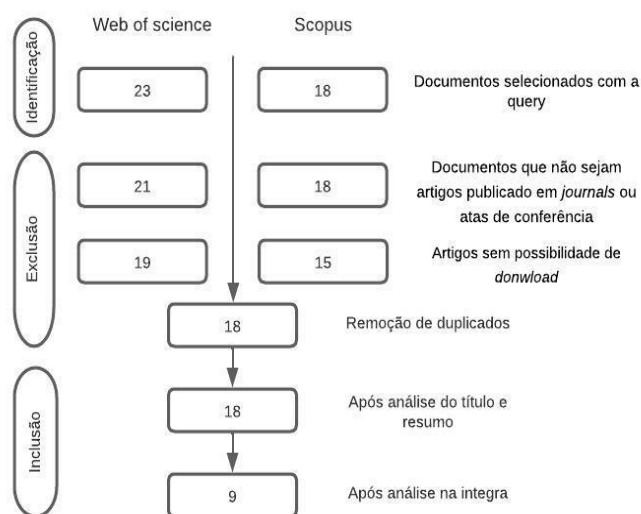


Figura 2.1 Síntese do processo de seleção de artigos

Posteriormente à leitura na íntegra dos artigos selecionados, procede-se a uma nova leitura com o objetivo de extrair a informação que responde a cada questão específica de pesquisa. Numa primeira fase, toda a informação relevante é inserida meticulosamente numa tabela para responder às questões levantadas. De seguida, agrupa-se a informação numa nova tabela, a fim de obter uma análise descritiva que possibilita identificar padrões, tendências e/ou exceções.

O último passo da RSL corresponde à avaliação da literatura selecionada. A cada questão da pesquisa é associado um conjunto de critérios (

Tabela 2.2) e operacionalizados na forma de questão. Atribui-se 1 ponto aos artigos que respondem totalmente ao critério; 0,5 pontos aos artigos que respondem parcialmente ao critério específico e 0 pontos aos artigos que não respondem ao critério específico.

A pontuação permite não só identificar o nível de qualidade e a importância dos artigos para a dissertação, mas também contribuir para clarificar a confiança dos resultados da RSL (Kitchenham & Brereton, 2013).

Salienta-se que o processo da RSL é inteiramente revisto por um especialista na área de *fundraising*, de forma a eliminar o enviesamento da interpretação individual de cada artigo.

Tabela 2.2 Critérios de avaliação dos artigos da RSL

Questão ⁱ	Critério para a avaliação dos artigos	
Q1	C1	Descreve profundamente o âmbito da investigação?
	C2	Descreve de forma esclarecedora aquilo que se propõe a prever?
	C3	Qual a origem dos dados?*
	C4	O tamanho da amostra foi específico?
Q2	C5	Descreve totalmente as técnicas utilizadas?
	C6	Descreve totalmente as métricas de avaliação das técnicas utilizadas?
	C7	Expõe os resultados das técnicas utilizadas?
Q3	C8	Descreve totalmente as variáveis utilizadas?
Q4	C9	Descreve as limitações do estudo?
	C10	Descreve os contributos do estudo?

* 1 – Dados provenientes de uma base de dados; 0,5 – Dados provenientes de questionários; 0 – N.E.

2.2 Caracterização dos artigos

Na Tabela 2.3, estão identificados os artigos selecionados, as suas características e objetivo. A partir da tabela referida, é possível observar que existem poucos artigos sobre a previsão do comportamento do doador, sendo que somente três artigos (Farrokhvar *et al.*, 2018; Eiland *et al.*, 2021; Cacciarelli & Boresta, 2022) de nove abordam exatamente a elaboração de modelos preditivos (33 %). Além disso, é de referir que, em 2021, se publicaram mais artigos

comparativamente aos outros anos (três artigos). Todos os artigos são escritos por autores diferentes, sendo que nenhum foi redigido por um único autor. Por fim, não existe nenhuma tendência de publicação em *journals* ou atas de conferências (AC).

Tabela 2.3 Artigos incluídos na RSL, publicados entre 2007 e 2022

ID	Ano	Título	Autores	Journal / AC	Objectivo
1	2007	Charitable giving: The effectiveness of a revised theory of planned behaviour model in predicting donating intentions and behaviour	Smith, JR; McSweeney, A	Journal of Community & Applied Social Psychology	Analisar os fatores psicológicos sociais que levam um indivíduo a realizar uma doação, tendo por base o TPB.
2	2018	Predictive models for charitable giving using machine learning techniques	Farrokhvar, L; Ansari, A; Kamali, B	Plos One	Desenvolvimento de modelos preditivos que prevê a intenção de doar.
3	2019	Donations to outgroup Charities, but not ingroup charities, predict helping intentions toward street-beggars in Sweden	Erlandsson, A; Nilsson, A; Tinghog, G; Andersson, D; Vastfjall, D	Nonprofit and Voluntary Sector Quarterly	Analisar a relação entre as intenções de doação e as intenções de ajudar migrantes na EU.
4	2019	What are the most powerful predictors of charitable giving to victims of typhoon Haiyan: Prosocial traits, socio-demographic variables, or eye cues?	Manesi, Z; Van Lange, PAM; Van Doesum, NJ; Pollet, TV	Personality and Individual Differences	Aumentar a compreensão das doações em desastres.
5	2020	What are the predictors of nonprofit association members' supportive behaviors?	Ki, EJ; Cho, MH	Southern Communication Journal	Determinar as intenções de doação e de recomendação de adesão.
6	2021	Charitable Giving in the Context of Unfamiliar Organizations: The Effectiveness of Construal Level Theory in Predicting Donating Intentions and Antecedents.	Bates, BR; Sherwani, SI; Grijalva, MJ	Systems and Information Engineering Design Symposium (IEEE Sieds 2021)	Determinar as intenções de doação a partir da teoria do nível de interpretação.
7	2021	Developing Models to Predict Giving Behavior of Nonprofit Donors	Eiland, J; Hammonds, CM; Ponos, SM; Weigand, SM; Scherer, WT	Decision Support Systems	Desenvolver modelos preditivos que preveem a intenção de doar e quantidade monetária a solicitar.
8	2021	Predicting donation behavior: Acquisition modeling in the nonprofit sector using Facebook data	Schetgen, L; Bogaert, M; Van den Poel, D	Journal of Philanthropy and Marketing	Demonstrar o valor dos dados do Facebook na previsão do comportamento de doação.
9	2022	What drives a donor? A machine learning-based approach for predicting responses of nonprofit direct marketing campaigns	Cacciarelli, D; Boresta, M	International Journal of Nonprofit and Voluntary Sector Marketing	Desenvolver modelos preditivos que preveem a intenção de doar.

2.3 Síntese de conteúdos dos artigos

2.3.1 Contexto dos estudos

Na Tabela 2.4, está identificado o que o artigo se propõe prever, a origem e o tamanho da amostra, o país de onde os dados são provenientes, assim como a que instituição o estudo é destinado.

A referida tabela demonstra que o âmbito da previsão mais estudado é a previsão da intenção de doação e que 56% dos estudos realizados foram elaborados nos Estados Unidos da América (Farrokhvar *et al.*, 2018; Manesi *et al.*, 2019; Ki & Cho, 2020; Sherwani *et al.*, 2021; Eiland *et al.*, 2021). Por outro lado, somente dois artigos preveem a intenção de doação baseada em dados históricos das doações da própria organização (Ki & Cho, 2020; Cacciarelli & Boresta, 2022).

Desta forma, a área do *fundraising analytics* representa uma oportunidade de investigação que deve ser aproveitada. Todavia, o facto de não haver muitos estudos que preveem a intenção de doar, tendo por base dados históricos de uma organização, também poderá ser considerado uma limitação, visto que não existe uma base substancial de trabalhos que permita uma fácil comparação de resultados entre as diferentes análises preditivas realizadas.

Tabela 2.4 Sumarização do contexto do estudo dos artigos da RSL

Id	Âmbito da previsão	Fonte	Amostra	País	Destino
1	Intenção de doar de um doador	Dois questionários	227 inquéritos	Austrália	Geral
2	Intenção de doar de uma região	Censos e formulários fiscais dos EUA*	N.E.	EUA *	Geral
3	Intenção de doar de um doador	Questionário	1050 inquéritos	Suécia	Geral
4	Intenção de doar de um doador	Questionário	643 inquéritos	EUA *	N.E.
5	Intenção de doar e recomendação à adesão	Base de dados de 3 instituições diferentes	3,464 membros	EUA *	Associações de engenheiros
6	Intenção de doar de um doador	Questionário	447 inquéritos	EUA *	Healthy Homes for Healthy Living (HHHL)
7	Intenção de doar e a quantidade monetária	N.E.	N.E.	EUA *	The Children's Inn (The Inn)
8	Intenção de doar pela primeira vez	Dados provenientes do Facebook	96 perfis	N.E.	Geral
9	Intenção de doar de um doador	Dados provenientes de 16 campanhas de marketing direto	1 milhão de doações	Itália	World Wide Fund for Nature (WWF)

* EUA – Estados Unidos da América; N.E. – Não especificado

2.3.2 Técnicas e métricas de avaliação utilizadas

A Tabela 2.5 revela as técnicas utilizadas, o tipo de problema (regressão ou classificação) e as métricas de avaliação.

Detalhadamente, é possível observar que o tipo de problema mais comum corresponde a problemas de regressão (56%) (Smith & McSweeney, 2007; Farrokhvar *et al.*, 2018; Erlandsson *et al.*, 2019; Manesi *et al.*, 2019; Sherwani *et al.*, 2021), por outro lado, apenas

22% correspondem a problemas de classificação (Farrokhvar *et al.*, 2018; Cacciarelli & Boresta, 2022). Quanto às técnicas, a rede neural artificial (RNA) é a técnica mais recorrente em problemas de classificação, visto que é utilizada em três artigos (Farrokhvar *et al.*, 2018; Eiland *et al.*, 2021; Cacciarelli & Boresta, 2022), enquanto nas técnicas de regressão, a mais utilizada é a regressão hierárquica (Smith & McSweeney, 2007; Erlandsson *et al.*, 2019). No que se refere às métricas de avaliação, em problemas de classificação, a área sob a curva ROC (*Receiver operator characteristic*) é a métrica mais comum, sendo utilizada em três artigos (Eiland *et al.*, 2021; Schetgen *et al.*, 2021; Cacciarelli & Boresta, 2022). Já o R^2 assume o papel de métrica de avaliação mais aplicada em problemas de regressão (Smith & McSweeney, 2007; Sherwani *et al.*, 2021).

Tabela 2.5 Sumarização das técnicas e métricas de avaliação utilizadas nos artigos da RSL

Id	Tipo	Técnicas utilizadas		Métricas de avaliação	
		Classificação	Regressão	Classificação	Regressão
1	Regressão	-	Regressões hierárquicas múltiplas	-	R^2
2	Regressão	-	Regressão linear múltipla, redes neurais artificiais e regressão de vetor de suporte	-	Erro percentual médio absoluto simétrico, erro médio absoluto, raiz do erro quadrático médio, erro quadrático médio normalizado, variação inexplicável após o ajuste do modelo, erro máximo, correlação entre real e previsto
3	Regressão	-	Regressão hierárquica	-	R^2
4	Ambos	Random Forests	Cforest	N.E.	N.E.
5	Regressão	-	N.E.	-	R^2
6	Regressão	-	N.E.	-	N.E.
7	Ambos	Árvore de classificação e regressão	Árvore de classificação e regressão	AUC *	N.E.
8	Classificação	Regressão logística, k-nearest neighbors, bagged trees, random forest, adaboost, extreme gradient boosting e redes neurais artificiais	-	AUC *	-
9	Classificação	Árvore de decisão, RUSBoost, Balanced Random Forest e redes neurais artificiais	-	AUC *, desvio padrão e total médio de incremento líquido (\$);	-

*AUC – Área da curva ROC, N.E. – Não especificado

2.3.3 Variáveis preditoras do comportamento do doador

Na Tabela 2.6 estão sumarizadas as variáveis utilizadas nas investigações. A tabela demonstra que quase todos os artigos utilizam tanto variáveis demográficas¹, como variáveis sociodemográficas. O rendimento e a educação são as variáveis sociodemográficas mais utilizadas (Smith & McSweeney, 2007; Erlandsson *et al.*, 2019; Manesi *et al.*, 2019; Ki & Cho, 2020), enquanto a idade e o género (Smith & McSweeney, 2007; Farrokhvar *et al.*, 2018; Erlandsson *et al.*, 2019; Manesi *et al.*, 2019; Ki & Cho, 2020) são as que mais se destacam enquanto variáveis demográficas. Relativamente às variáveis classificadas como outras, existe bastante dispersão, pelo que se torna desafiante determinar qual a variável usada mais frequentemente. No entanto, destacam-se as seguintes variáveis: mês que ocorreu a campanha, há quanto tempo o doador apoia a organização, segmento dos doadores, quantidade de doações realizadas (Cacciarelli & Boresta, 2022).

Tabela 2.6 Sumarização das variáveis utilizadas nos artigos da RSL

Id	Variáveis		
	Sociodemográficas	Demográficas	Outras
1	Rendimento, educação	Idade, sexo, estado civil, número de filhos e religião	Se o doador tem controlo sobre o que doa, se para o indivíduo é importante expor que doa dinheiro, crença normativa, normas descritivas, normas morais, comportamento passado em relação ao ato de doação, se o doador revela intenção de doar.
2	% de desemprego, renda média habitacional, % de pessoas que vivem na/ou abaixo da linha de pobreza, % de graduados e % de habitações com 2 ou mais veículos	Idade, sexo, % de não brancos e n.º de indivíduos numa localidade.	Número de doações que as ONG receberam
3	Educação	Idade, sexo, religião, intenção política	Se o doador: doou no último ano, valor da doação, número de doações para instituições dentro e fora do grupo e número de doações para instituições ambientais
4	Educação	Idade, sexo, etnia, religião, intenção Política, se possui inglês nativo	SVO (<i>social value orientation</i>), SoMi (<i>social mindfulness</i>) e Imagens dos olhos
5	Rendimento	Idade Género	Benefícios pessoais e profissionais, solicitação, se doou no passado, se foi voluntário e longevidade
6	-	-	Atitude, norma social e descritiva, se o doador tem controlo o sobre o que doa, intenção de doar, intenção de recomendar
7	N.E.	N.E.	N.E.
8	N.E.	N.E.	N.E.
9	-	Idade, género e religião	Tema principal da campanha, âmbito da campanha, se a campanha é nacionalidade ou internacional, mês que ocorreu a campanha, há quanto tempo o doador apoia a organização, segmento dos doadores, quantidade de doações e de compras

N.E.- Não especificado

¹ É de notar que, apesar de alguns estudos considerarem as variáveis demográficas como variáveis sociodemográficas, a presente investigação, com auxílio de um especialista em *fundraising*, entendeu que, para proporcionar um maior detalhe sobre os preditores presentes na literatura, era necessário agrupar as variáveis afetas ao rendimento e às habilitações como variáveis sociodemográficas e as restantes como demográficas.

2.3.4 Contributos e limitações

A Tabela 2.7 sumariza as principais limitações e os contributos referidos em cada estudo.

As limitações apresentadas são, por exemplo, relacionadas com a aplicação restrita do estudo, isto é, aplicável apenas a situações de catástrofe (Manesi *et al.*, 2019) ou o sistema de apoio à decisão desenvolvido ser somente alimentado por dados provenientes do *Facebook* (Schetgen *et al.*, 2021).

Outra limitação encontrada reside no facto de o conjunto de dados ser reduzido, o período temporal de recolha de dados não ser aconselhável e a utilização de dados ser pouco fidedigna (Smith & McSweeney, 2007).

Relativamente aos contributos dos estudos, estes possibilitam: um fornecimento de um sistema de apoio à decisão à organização (Eiland *et al.*, 2021; Cacciarelli & Boresta, 2022); atualização da literatura sobre *fundraising* (Cacciarelli & Boresta, 2022); e determinação das considerações que as organizações devem ter ao abordar um doador (Smith & McSweeney, 2007).

Tabela 2.7 Limitações e contributos dos artigos da RSL

ID	Limitações	Contributos
1	Dados	Determina as considerações que as organizações devem ter ao abordar um doador
2	N.E.	Permite perceber a importância dos modelos preditivos em situações de catástrofe
3	Dados	Permite compreender as dinâmicas externas e internas de doação
4	Restrito a um estilo de aplicação	Permite perceber o perfil de doador em situações de catástrofe
5	Dados	Contribui para a literatura do <i>fundraising</i>
6	N.E.	Dota a organização de um sistema de apoio à decisão
7	Restrito a um estilo de aplicação	Dota a organização de um sistema de apoio à decisão
8	N.E.	Contribui para a literatura do <i>fundraising</i>
9	Dados	Dota a organização de um sistema de apoio à decisão

N.E. – Não especificado

2.4 Avaliação dos artigos da RSL

Na Tabela 2.8, está sintetizada a avaliação realizada aos nove artigos da RSL, no qual as pontuações para cada critério de avaliação são 0, 0,5 ou 1.

Tabela 2.8 Avaliação dos artigos da RSL

ID	C										Total
	1	2	3	4	5	6	7	8	9	10	
1	1	0,5	0,5	1	1	1	1	1	1	1	9
2	1	1	0,5	0	1	1	1	1	0	1	7,5
3	1	1	0,5	1	1	1	1	1	1	1	9,5
4	1	1	0,5	1	0,5	0	1	1	1	1	8
5	1	0,5	1	1	0	1	1	1	1	1	8,5
6	1	1	0,5	1	0	0	1	1	0	1	6,5
7	1	1	0	0	1	0	1	0	1	1	6
8	1	1	1	1	1	1	1	0	0	1	8
9	1	1	1	1	1	1	1	1	1	1	10
Total	9	8	5,5	7	7,5	7	9	7	6	8	

É observável que o artigo mais bem pontuado é o número 9 (Cacciarelli & Boresta, 2022). Este possibilita compreender toda a envolvimento da criação de um modelo preditivo. Contrariamente, o artigo menos pontuado é o 7, sendo que apenas permite perceber as técnicas utilizadas (Eiland *et al.*, 2021).

O artigo 8 proporciona a percepção da importância de analisar os dados das redes sociais, nomeadamente na rede social *Facebook* (Schetgen *et al.*, 2021). Já os artigos 6, 5, 4, 3, 2, 1 (Smith & McSweeney, 2007; Farrokhvar *et al.*, 2018; Erlandsson *et al.*, 2019; Manesi *et al.*, 2019; Ki & Cho, 2020; Sherwani *et al.*, 2021) contribuem para o conhecimento das variáveis a utilizar, enquanto o artigo 2 e 4 (Farrokhvar *et al.*, 2018; Manesi *et al.*, 2019) têm uma posição de destaque, possibilitando um conhecimento mais detalhado sobre doações em situações de catástrofe. Em relação, às técnicas utilizadas, todos contribuem para o conhecimento, à exceção do artigo 5 e 6 (Ki & Cho, 2020; Sherwani *et al.*, 2021).

Por fim, todos os artigos possibilitam identificar e fundamentar o *gap* da investigação.

2.5 Sumário

Grande parte dos estudos que abordam a previsão do comportamento do doador são baseados em dados provenientes de questionários, nos quais, perante o decorrer do seu desenvolvimento, enfrentam limitações bastantes similares e graves. Uma amostra reduzida de dados observados ou respostas falsas por parte de algum indivíduo são problemas que se traduzem logo à partida num modelo preditivo irrealista, podendo considerar que os dados,

estão adulterados e não representam a maioria dos doadores da organização (Smith & McSweeney, 2007; Erlandsson *et al.*, 2019; Manesi *et al.*, 2019; Ki & Cho, 2020).

Acontecimento este que não sucede quando as investigações são fundamentadas em dados históricos das organizações. Por exemplo, uma amostra de 227 indivíduos (Smith & McSweeney, 2007) não pode ser comparada a um conjunto de dados históricos de um milhão de registos sobre doações (Cacciarelli & Boresta, 2022), isto porque existem mais dados para analisar, assim como para treinar o modelo, de forma que tenha sucesso no ato da previsão. Além disso, é importante destacar que o registo de dados provenientes de campanhas não está dependente da sinceridade do indivíduo (Farrokhvar *et al.*, 2018; Cacciarelli & Boresta, 2022), mas sim da capacidade de inserção de informações de uma máquina numa base de dados.

Deste modo, reforça-se, ainda mais, a relevância da utilização de dados históricos organizacionais em detrimento de questionários feitos aos doadores, bem como a escolha de a investigação se basear em dados provenientes de uma base de dados ao invés de questionários.

É essencial frisar que os estudos que têm como base questionários são importantes para complementar os estudos que utilizam dados provenientes das bases de dados das organizações, uma vez que permitem fundamentar a escolha de algumas variáveis para a previsão do comportamento do doador.

Sucessivamente, certas variáveis sociodemográficas e demográficas, como educação, idade, sexo e religião, assumem um papel de preditor significativo (Farrokhvar *et al.*, 2018; Erlandsson *et al.*, 2019; Manesi *et al.*, 2019; Ki & Cho, 2020; Sherwani *et al.*, 2021; Eiland *et al.*, 2021; Cacciarelli & Boresta, 2022). Outras variáveis como a frequência em que um doador faz a doação, o valor monetário da doação e se realizou uma doação recentemente e o mês em que aconteceu determinam significativamente o ato de doar (Erlandsson *et al.*, 2019; Manesi *et al.*, 2019; Ki & Cho, 2020; Cacciarelli & Boresta, 2022).

Em relação às técnicas de previsão aplicadas à previsão da intenção de doar, a que apresenta melhores resultados são as RNN (Farrokhvar *et al.*, 2018; Cacciarelli & Boresta, 2022), obtendo inclusive 0.85 na Área sob a Curva ROC (Cacciarelli & Boresta, 2022). Por outro lado, devido à escassa literatura existente, é impossível determinar qual a melhor técnica para prever o valor monetário, visto que os autores Eiland *et al.* (2021) foram os únicos a ter como objetivo antecipar o valor doado e não expuseram as métricas nem os resultados obtidos no problema de regressão que utilizaram.

Concluindo, tendo como base os resultados da RSL realizada, torna-se possível responder às questões de pesquisa:

- I. Qual é o âmbito da investigação em que se utiliza a previsão do comportamento do doador?

O âmbito mais usual da previsão do comportamento do doador é a intenção de doar para uma determinada organização. Além disso, grande parte dos estudos utiliza dados provenientes de inquiridos, ao invés de utilizar dados históricos sobre os doadores e as doações que realizaram.

- II. Quais as técnicas e as métricas de avaliação utilizadas para prever o comportamento do doador?

Existem várias técnicas utilizadas para prever o comportamento do doador, como por exemplo: Regressão logística, *k-nearest neighbors*, *bagged trees*, *random forest*, *adaboost*, *extreme gradient boosting* ou redes neurais artificiais.

Assim como também existem algumas métricas de avaliação, como é o caso da AUC, para problemas de classificação, ou do R^2 , para problemas de regressão.

- III. Quais foram as variáveis utilizadas para prever o comportamento do doador?

Ao realizar a RSL, foi possível constatar três grandes grupos de variáveis: sociodemográficas, demográficas e classificadas como outras. Por exemplo, o rendimento, habilitações académicas ou o número de doações podem contribuir para prever o comportamento de um doador.

- IV. Quais as limitações e os contributos do estudo?

Todos os artigos revelam os seus contributos para a comunidade científica ou para a organização em que se inseriu o estudo. No que diz respeito às limitações, agrupam-se em dois grupos: limitação dos dados e limitação de contexto, isto é, os resultados não podem ser generalizados a vários ambientes.

3 Metodologia

Para o presente estudo, adotou-se o CRISP-DM por ser das metodologias mais utilizadas na resolução de problemas que envolvem modelos preditivos (Chapman *et al.*, 2000).

O CRISP-DM foi concebido sobretudo para categorizar e orientar a execução das várias etapas necessárias em estudos de *data mining* (Sharma *et al.*, 2017; Martinez-Plumed *et al.*, 2021), sendo composto por seis etapas totalmente interativas (Figura 3.1).

Outra característica do CRISP-DM é que o retorno a uma etapa anterior é absolutamente normal (Chapman *et al.*, 2000; Maimon & Rokach, 2010; Schröer *et al.*, 2021), fazendo com que a metodologia seja mais flexível em comparação com abordagens mais tradicionais. Ainda de salientar que tal fator se evidencia extremamente importante, visto que, segundo Chapman *et al.* (2000), um estudo de *data mining* (em que o *fundraising analytics* está inserido) não deve ser de natureza sequencial e requer movimentação entre as diferentes fases.

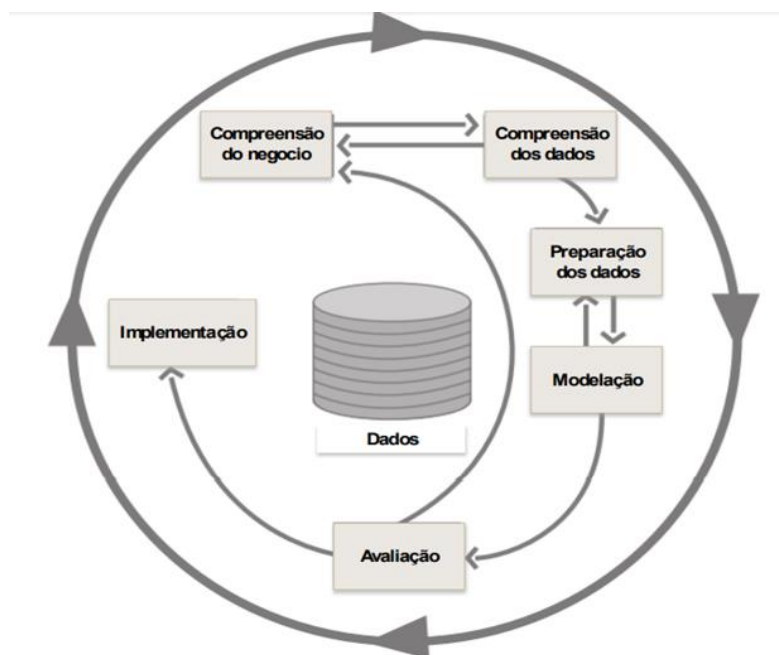


Figura 3.1 CRISP-DM

Fonte: Peixoto (2015)

3.1 Compreensão do Negócio

A compreensão do negócio corresponde à fase inicial de cada projeto. Nesta fase, o principal objetivo é não só conhecer e caracterizar a empresa, como também, simultaneamente, entender o problema e definir os objetivos do estudo.

Na presente dissertação, esta fase corresponde ao que foi apresentado nos capítulos de introdução e revisão da literatura.

Adicionalmente, o objetivo de negócio visa tornar as campanhas de angariação de fundos mais eficazes e eficientes. Este objetivo é dado por cumprido quando, a partir da utilização do modelo preditivo, os LD diminuem os custos provenientes das suas campanhas de angariação de fundos. É de salientar que não foi definido uma meta, visto que os períodos de seis meses de investigação não permitiram a implementação do modelo e retirar conclusões sobre os resultados.

Por outro lado, foi solicitado pelos LD um *dashboard* que possibilita suprimir as dúvidas que a organização tem sobre o comportamento dos seus doadores e dos donativos realizados.

3.2 Compreensão dos dados

Na segunda etapa do CRISP-DM, procede-se à compreensão total dos dados solicitados à organização LD.

Nesta fase, é desenvolvido um conjunto de tarefas que permitem perceber toda a conjuntura dos dados. As principais tarefas desenvolvidas são a identificação dos dados disponíveis que a organização possui, a determinação da sua origem, a interpretação dos dados provenientes da organização e a análise da qualidade dos mesmos. Além do referido, é esboçado um dicionário de dados com o intuito de documentar o que foi recolhido e a ação a realizar na próxima etapa do CRISP-DM.

Deste modo, solicitou-se aos LD o conjunto de variáveis². Enfatiza-se que o pedido aos LD teve o objetivo de perceber quais os dados que a organização possui, que, de acordo com a literatura, permitem a elaboração de um modelo preditivo sobre o comportamento do doador.

Posteriormente, a organização recolheu do seu *Customer Relationship Management* (CRM) dois conjuntos de dados estruturados, em formato .csv. Detalhadamente, o primeiro conjunto corresponde a dados de 21 219 doações ocorridas no período de 2015 a 2022, tendo algum detalhe sobre os doadores (e.g., forma de tratamento, cidade de correspondência). O segundo conjunto possui somente informação sobre os 3 327 doadores que fizeram doações.

² Consultar no anexo A: Tabela - Dados solicitados

A tarefa da análise da qualidade dos dados é uma das mais cruciais antes da “preparação de dados”. É através desta que se identificam lacunas nos dados existentes, de forma a transformá-los em dados mais consistentes e precisos, assim como menos redundantes e ruidosos (Mendes, 2022).

Com o intuito de realizar a última tarefa descrita, análise à qualidade dos dados, procede-se, a partir do *Power Query*, à realização de uma análise descritiva aos conjuntos de dados provenientes do CRM dos LD.

Constata-se que, no conjunto de dados número 2, na coluna “Identificação”, numa população de 3327 doadores, 1784 registos estão duplicados, enquanto 100 doadores não têm registo de doações na tabela 1. Por outro lado, a coluna “Data de nascimento” não possui 141 registos preenchidos. Ainda de referir, que a data de nascimento mais antiga corresponde a 25/12/1906 e a data mais recente é 09/10/2012.

Relativamente ao conjunto de dados número 1, as variáveis classificadas pelo estudo como datas não apresentam anomalias. Além disso, a partir da Tabela 3.1, é observável o intervalo de tempo de tais variáveis, nomeadamente a data mais antiga registada e a data mais recente (por exemplo, a data mais recente do primeiro donativo realizado por um doador corresponde a 31/12/2022, enquanto a data mais antiga corresponde a 31/05/1996).

Tabela 3.1 Intervalo de tempo das variáveis classificadas como data (conjunto de dados 1)

Conjunto de dados	Variáveis	Mais recente	Mais antiga
1	Data do primeiro donativo	31/12/2022	31/05/1996
1	Data do último donativo	31/12/2022	05/01/2015
1	<i>Last Donation Date</i>	31/12/2022	05/01/2015
1	Data de fechamento	31/12/2022	05/01/2015

Nota: n=21219 doações

Em contrapartida, analisando a qualidade de dados das variáveis classificadas pelo estudo como qualitativas, a variável “Tratamento” apresenta cinco categorias incoerentes, uma vez que existem várias formas de descrever a mesma categoria. Além disso, na mesma variável, no registo de duas doações, não existe qualquer tipo de dado, ou seja, o valor encontra-se em branco. A variável “Cidade de correspondência” possui dados relativos a cidades, freguesias ou concelhos, contudo, não apresenta dados uniformizados, sendo ainda de referir que, em 514 doações, não existem dados sobre o sítio de correspondência do doador. A variável “classificação” apresenta a categoria “particular”, indicando que os dados enviados pelos LD apenas correspondem a doações de particulares. As variáveis “Subclassificação 1”, “Identificação”, “Número do recibo” e “Origem da campanha principal” não possuem anomalias no que consta à qualidade de dados. A variável “Estado do doador” apresenta quatro categorias incoerentes, enquanto a variável “Forma de pagamento” apresenta incoerência em

13 categorias, visto que existem várias formas de descrever o mesmo dado. Por fim, a variável “Profissão” e a variável “Habilitações” apresentam inconsistência nos dados, devido à ausência de uniformização, sendo que o mesmo dado é descrito da mesma forma inúmeras vezes. Além do referido, as duas últimas variáveis descritas correspondem a mais de 50% de valores em branco.

Na Tabela 3.2, é possível ver a categoria mais frequente, menos frequente, os valores omissos e o número de categorias de cada variável qualitativa. É de destacar que as variáveis “Cidade de correspondência”, “Origem da campanha principal”, “Identificação”, “Profissão” e “Habilitações” não apresentam só uma categoria de menos frequente. A variável “Número do recibo” exclusivamente possui valores únicos, devido à correspondência ao recibo providenciado ao doador no momento da doação.

Tabela 3.2 Estatísticas descritivas das variáveis qualitativas (conjunto de dados 1)

Conjunto de dados	Variáveis	Categoria mais frequente		Categoria menos frequente		Nº Categorias	Valores em vazio
		Categoria	%	Categoria	%		
1	Tratamento	Exma. Sra.	61%	Revº Senhor	0,02%	5	2
1	Cidade de correspondência	Lisboa	73%	-	-	468	514
1	Estatuto Postal	Normal	82%	Não Contactar – Não Queremos Enviar	0,03%	11	0
1	Classificação 1	Particular	100%	-	0%	1	0
1	Subclassificação 1	Outra	40%	Jesuíta; Religioso; Ancião	0,004%	21	0
1	Estado Doador	Doador Ativo	60%	Doador Inativo	13%	4	0
1	Origem da campanha principal	Apadrinhamento Voluntários	49%	-	-	74	30
1	Forma de pagamento	Transferência bancária	36%	Cartão de crédito	0,09%	13	0
1	Profissão	Professora	3,68%	-	-	568	6 128
1	Habilitações	Direito	1%	-	-	274	15 591
1	Identificação	19486	0,5%	-	-	2078	0
1	Número do recibo	-	-	-	-	21 219	0

Nota: n=21219 doações

Por fim, analisou-se a qualidade de dados das variáveis classificadas pelo estudo como quantitativas. A variável “Valor do último donativo” apresenta 109 valores em branco, ao passo que a variável “Valor” (dado) possui 54 valores omissos. Por outro lado, o valor médio de

doação é cerca de 36,79 euros, podendo o intervalo de doações oscilar entre 1 e 3 000 euros (Tabela 3.3).

Tabela 3.3 Estatísticas descritivas das variáveis quantitativas (conjunto de dados 1)

Conjunto de dados	Variáveis	Média	Desvio Padrão	Mínimo	Máximo	Valores em branco	Percentis		
							25	50	75
1	Valor do último donativo	38,4	79,51	1	3 000	109	10	10	30
1	Valor	36,79	55,57	1	3 000	54	6,5	10	27,5

Nota: n=21219 doações

Em suma, o conjunto de dados 2 apenas apresenta falta de completude, nomeadamente valores omissos e valores duplicados. Enquanto, o conjunto de dados 1 apresenta as seguintes limitações:

- Inconsistências — a informação está escrita de várias formas (e.g., professor/docente, ensino superior/licenciatura);
- Ambiguidade — certas colunas possuem informação repetida;
- Falta de completude — existem valores omissos;
- Informação redundante — apresentam-se dados desnecessários para a realização do modelo, bem como informação escrita de forma implícita (ex., sr./sra.).

As variáveis pertencentes a cada conjunto, a sua tipologia e a ação sobre o próximo passo a realizar são apresentadas no dicionário de dados³.

3.3 Preparação dos dados

A terceira fase subdivide-se em quatro grandes etapas: limpeza e transformação dos dados; criação de variáveis pertinentes para a dissertação e integrando-as nas tabelas já existentes; seleção dos dados a integrar nos modelos preditivos; e criação do modelo de dados, para utilização no *dashboard*.

³ Consultar no anexo B: Tabela – Dicionário de dados *provenientes do CRM dos LD*

3.3.1 Fluxo da preparação dos dados

Este subcapítulo pretende ilustrar o fluxo de preparação de dados (Figura 3.2), tendo-se realizado em três plataformas diferentes:

- I. *SQL Studio*, para realizar a limpeza e transformação dos dados cedidos pelos LD.
- II. *SPSS Modeler*, para a criação de novas variáveis pertinentes à criação do modelo preditivo, seleção dos dados que vão incorporar tal modelo, assim como a criação de um conjunto de dados para treino e teste.
- III. *Power BI* (PBI) para a criação de novas variáveis e modelo de dados para a criação do *dashboard*.

A escolha do *SQL Studio* para o desempenho das tarefas atribuídas pela investigação deve-se às seguintes razões:

- Rapidez de processamento de preparação de dados comparativamente ao Excel ou *SPSS Modeler*: o *SQL Studio* é conhecido pela sua capacidade de processar grandes volumes de dados de forma rápida e eficiente.
- Facilidade de automação: através da criação de *jobs*, é possível automatizar o processo ETL de forma intuitiva, eliminando, por sua vez, a dependência humana.
- Integração com outras ferramentas: o *SQL Studio* permite a fácil integração com outras ferramentas e sistemas, o que se revela vantajoso para a organização, pois permite a incorporação dos dados já preparados noutros processos (e.g., *dashboards* para monitorização em tempo real).

Relativamente à escolha do *SPSS Modeler* para o desempenho de todas as funções associadas ao desenvolvimento do modelo preditivo, considera-se que o *software* facilita o acompanhamento total do fluxo de trabalho numa só plataforma.

Por fim, optou-se pela utilização do *Power BI* por quatro razões:

- Facilidade de integração com outras ferramentas Microsoft, nomeadamente *SQL Server* e Excel.
- A partir da linguagem *DAX* (*Data Analysis Expressions*), é possível criar cálculos complexos e realizar análises mais avançadas.
- Permite a importação e a consolidação de várias fontes de dados (e.g., bancos de dados, arquivos locais, serviços em nuvem e API), o que facilita a criação de relatórios que reúnem informações de diversas fontes.
- A plataforma permite criar visualizações interativas e dinâmicas.

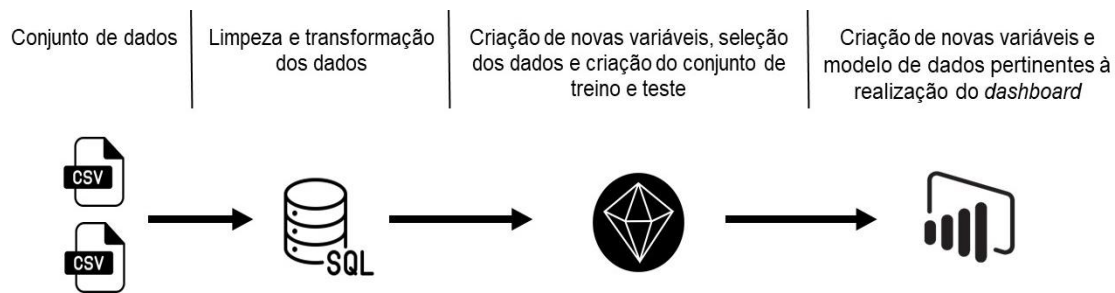


Figura 3.2 Fluxo de preparação de dados

3.3.2 Limpeza, transformação e criação de novas variáveis

Nesta etapa, é realizada a limpeza e correção dos erros e inconsistências encontrados nos registos resultantes da etapa “compreensão de dados”, isto é, são tratados os dados que apresentam anomalias.

Como mencionado, a limpeza e transformação dos dados realiza-se a partir do *SQL Studio*, mais precisamente a partir de uma *stored procedures* (SP) desenvolvida para o efeito (SP_ETL_DOAÇÕES).

A SP_ETL_DOAÇÕES possibilita, inicialmente, a extração de dados provenientes de ficheiros .csv, formato em que o conjunto de dados se encontra. Sucessivamente, a partir de outra *query* complexa, a SP realiza a limpeza e transformação de dados, assim como reorganiza a informação nas tabelas, isto é, a informação relativa ao doador é conduzida para a tabela “doadores”, enquanto a informação relativa às doações fica na tabela “doações”. Por fim, a SP está habilitada a enviar um *email* com a descrição do passo em que o processo falhou, caso aconteça.

Além da criação de uma SP para o efeito de transformação de variáveis, a partir de um *step*, a SP_ETL_DOAÇÕES é inserida num *job* criado no “SQL Server Agent” do *SQL Studio*. O referido *job* é composto por outro *step* que realiza a extração das tabelas em formato .csv e insere-as na máquina local, permitindo, conseqüentemente, a automatização do processo de limpeza e transformação de dados, eliminando, naturalmente, a necessidade de intervenção humana.

A Tabela 3.4 demonstra o processo de limpeza/transformação aplicável a cada variável (coluna “Limpeza de dados e/ou transformações realizadas”) e a razão de se ter realizado o processo de limpeza/transformação descrito (coluna “Razão”). Por fim, a Tabela 3.4 também demonstra a tabela em que cada variável foi inserida, com vista a reorganizar o conjunto de dados (coluna “Tabela destino”).

Tabela 3.4 Limpeza e transformação de dados

Tabela origem	Variáveis	Razão	Limpeza de dados e/ou transformações realizadas	Tabela destino
1	Identificação	Por se tratar de uma chave composta, não necessita de tratamento.	-	Doações
1	Tratamento	Sendo crucial saber o género do doador e não a forma de tratamento, é necessário transformar os dados em Masculino/Feminino, assim como eliminar os valores vazios.	Alteração do nome da coluna para "Género" e transformação do Exmo. Sr./Exma. Sra. em Masculino/Feminino. Eliminação dos valores omissos	Doadores
1	Cidade de correspondência	Não existe padronização de dados. Na respetiva coluna, é possível observar tanto freguesias como concelhos ou até distritos. Deste modo, é necessário categorizar a coluna num único tipo de informação.	Alteração do nome coluna para 'Distrito' e categorização dos dados em distrito (eg: Águeda para Aveiro). Optou-se por se manter os valores omissos	Doadores
1	Estatuto Postal	Existe inúmeras categorias com informação contraditória (ex: normal não contactar), assim sendo, é necessário categorizar os dados corretamente.	Alteração do nome coluna para "Estatuto" e categorização dos dados (eg: "normal sem boletim" ou "normal ab" para "normal")	Doadores
1	Classificação 1	A informação apresenta-se irrelevante para a criação de um modelo preditivo.	Exclusão da coluna	-
1	Subclassificação 1	A informação apresenta-se irrelevante para a criação de um modelo preditivo.	Exclusão da coluna	-
1	Estado Doador	Sem problemas de qualidade da informação.	-	Doadores
1	Data do primeiro donativo	Sem problemas de qualidade da informação.	-	Doadores
1	Data do último donativo	Sem problemas de qualidade da informação.	-	Doadores
1	Data da última doação	Trata-se de uma coluna repetida (data do último donativo) com um nome diferente.	Exclusão da coluna	-
1	Valor do último donativo	Irá ser realizado a técnica RFM pela investigação.	Exclusão da coluna	-
1	Data de fechamento	O nome da coluna poderá induzir em erro, uma vez que os dados presentes na coluna correspondem à data do último donativo.	Alteração do nome da coluna para "Data donativo"	Doações
1	Valor	É crucial à investigação e aos LD saber sempre o valor da doação.	Eliminação das linhas vazias	Doações
1	Número do recibo	A informação apresenta-se irrelevante para a criação de um modelo preditivo.	Exclusão da coluna	-
1	Origem da campanha principal	O nome da coluna não se encontra de forma clara, assim como existem valores omissos.	Alteração do nome da coluna para 'Nome campanha', eliminação das colunas que possuem valores omissos	Doações
1	Forma de pagamento	Não existe padronização de dados, isto é, por exemplo "MB Way" está escrito de várias formas.	Categorização dos dados (eg: "MB Way Site" para "MB Way").	Doações
1	Número total de donativos	A variável irá ser calculada pela investigação.	Exclusão da coluna	-
1	Profissão	O uso dos dados da coluna passa por corrigir valores omissos na coluna habilitações. (ex: Um doador tem a coluna habilitações em nulo e é Médico, é inserido o dado "Ensino Superior").	Exclusão da coluna	-
1	Habilitações	Não existe padronização de dados, isto é, por exemplo, "licenciatura" está descrito como engenharia, psicologia, etc.	Categorização dos dados em: 1.º, 2.º, 3.º ciclo, Secundário, Ensino Superior	Doadores
2	Identificação	Ao se tratar de uma chave composta primária, é necessário eliminar os duplicados. Também é necessário eliminar os doadores que não possuem registo de doações.	Eliminação de duplicados e de valores omissos	Doadores
2	Data de nascimento do doador	É necessário à investigação saber a data de nascimento do doador para conhecer a sua idade.	Eliminação das linhas que estão vazias	Doadores

Ainda nesta etapa, são criadas variáveis que podem fazer diferença na concepção de um modelo preditivo, ou até mesmo delimitar o *target* do modelo (criação da variável dependente).

Deste modo, tendo por base a opinião de um especialista e utilizando o *SPSS Modeler*, na tabela “doadores” são criadas as variáveis independentes presentes na Tabela 3.5 (a coluna “variáveis” indica o nome da variável, a coluna “descrição” descreve em que consiste a variável, a coluna “razão” justifica o motivo de se criar a variável e a coluna “tipo” menciona a tipologia da variável).

Tabela 3.5 Variáveis independentes criadas no *SPSS Modeler*

Variáveis	Descrição	Razão	Tipo
Idade no ano N-1	Idade do doador ao realizar a última doação	A literatura menciona que a idade é um preditor importante	Q.C.
Faixa etária n-1	Classificação da idade n-1 em: Adolescente (até aos 21 anos) Jovem Adulto (entre os 21 e os 34 anos) Adulto (entre os 35 e os 64 anos) Idoso (Superior a 65 anos)	Existe técnicas de modelação que preveem melhor a partir de variáveis qualitativas	Q.O.
Senioridade n-1	Diferença de anos entre primeira doação do doador e a última (Excluindo o ano de 2022), ou seja, antiguidade.	A literatura menciona que a senioridade é um preditor importante	Q.C.
Senioridade classificação	Classificação da senioridade em: Baby (até 12 meses) Child (entre os 1 e 4 anos) Teen (entre os 5 e os 8 anos) Adult (entre os 9 e os 18 anos) Old (superior a 18 anos)	Existe técnicas de modelação que preveem melhor a partir de variáveis qualitativas	Q.O.
Nº de anos que não doa	Número de anos em que um doador não realiza uma doação	O especialista considerou a variável interessante de explorar como preditor	Q.D.
Deu em quantos anos?	Número de anos que o doador realizou uma doação	O especialista considerou a variável interessante de explorar como preditor	Q.D.
Nº de doações	Contagem de todas as doações realizadas (exceto no ano de 2022)	É necessário para o desenvolvimento do RFM	Q.D.
Valor monetário acumulado	Valor monetário total doado (ano de 2022 excluído)	É necessário para o desenvolvimento do RFM	Q.C.
Valor médio de doação	Valor médio total doado (ano de 2022 excluído)	É necessário para o desenvolvimento do RFM	Q.C.
Data em que ocorreu a última doação antes do ano n	Data em que ocorreu a última doação antes de 2022	É necessário para o desenvolvimento do RFM	Data
Mês em que ocorreu a última doação	Mês em que ocorreu a última doação antes de 2022	A literatura menciona que a variável é um preditor importante	Q.D.
Estado n-1	Estado do doador: Ativo (a última doação ocorreu no ano de 2021) Adormecido (a última doação ocorreu no ano de 2020) Inativo (a última doação ocorreu até ao ano de 2019)	A literatura menciona que o estado é um preditor importante	Q.O.
Recency Score	Atribuição de um <i>score</i> à recência das doações 1 – A última doação realizada aconteceu até 02/08/2017 2 – A última doação realiza aconteceu entre 03/08/2017 e 31/12/2018 3 – A última doação realiza aconteceu entre 01/01/2019 e 09/12/2020 4 – A última doação realiza aconteceu entre 10/12/2020 e 01/12/2021 5 – A última doação realiza aconteceu depois de 02/12/2021	É uma componente do RFM	Q.O.

Tabela 3.5 Variáveis independentes criadas no *SPSS Modeler*

(conclusão)

Variáveis	Descrição	Razão	Tipo
<i>Frequency Score</i>	Atribuição de um <i>score</i> à frequência das doações: 1 – Realizou uma doação 2 – Realizou entre 2 doações e 5 doações 3 – Realizou entre 6 doações e 30 doações 4 – Realizou entre 31 doações e 60 doações 5 – Realizou mais que 61 doações	É uma componente do RFM	Q.O.
<i>Monetary Score</i>	Atribuição de um <i>score</i> ao valor médio das doações: 1 – O Valor médio de doação até 10 euros 2 – O Valor médio de doação entre 11 e 30 euros 3 – O Valor médio de doação entre 31 e 60 euros 4 – O Valor médio de doação entre 61 e 150 euros 5 – O Valor médio de doação maior que 151 euros	É uma componente do RFM	Q.O.
<i>RFM Score</i>	Junção do <i>Recency Score</i> , <i>Frequency Score</i> , <i>Monetary Score</i>	É uma componente do RFM	Q.D.

Q.N. –Qualitativa Nominal, Q.O – Qualitativa Ordinal, Q.D. – Quantitativa Discreta, Q.C. – Quantitativa Contínua, RFM- Recency, Frequency, Monetary⁴

Adicionalmente, na tabela “doadores”, é criada a variável dependente/*target* do estudo (“Doou no ano n”). A variável consiste em se o doador realizou pelo menos uma doação em campanhas do ano de 2022 é atribuído o número 1, caso não tenha realizado nenhuma doação é atribuído o número 0. De realçar que, no universo de 1543 doadores (valor de doadores resultantes da limpeza e tratamento de dados), apenas 35,52% (n=548) dos doadores realizaram pelo menos uma doação em 2022.

Por fim, utilizando o *Power BI*, importa-se a tabela de doações do *SQL Studio* e a tabela doadores do *SPSS Modeler*. Utilizando a linguagem *DAX*, criam-se as colunas presentes na coluna “Variáveis” da Tabela 3.6 (que inclui também a descrição da variável, em que tabela a mesma está inserida e a tipologia). A tabela “calendário” é criada com o intuito de inserir a coluna “Date”, “Mês” e “Dia”.

Tabela 3.6 Variáveis criadas para a realização do *dashboard*

Variáveis	Descrição	Tabela	Tipo
Idade	Idade atual do doador	Doadores	Quantitativa contínua
Faixa etária	Classificação da idade em: Adolescente (até aos 21 anos) Jovem Adulto (entre os 21 e os 34 anos) Adulto (entre os 35 e os 64 anos) Idoso (Superior a 65 anos)	Doadores	Qualitativa ordinal
Senioridade atual	Diferença de anos entre primeira doação do doador e a última	Doadores	Quantitativa contínua
Senioridade atual-classificação	Classificação da senioridade em: Baby (até 12 meses) Child (entre os 1 e 4 anos) Teen (entre os 5 e os 8 anos) Adult (entre os 9 e os 18 anos) Old (superior a 18 anos)	Doadores	Qualitativa ordinal
Nº de anos que não doa (atual)	Número de anos em que um doador não realiza uma doação	Doadores	Quantitativa discreta
Nº de doações (atual)	Contagem de todas as doações realizadas	Doadores	Quantitativa discreta

⁴ A análise RFM é usada para classificar e agrupar quantitativamente os doadores com base na recência, frequência e valor monetário médio de doações, identificando, por sua vez, os melhores doadores a contactar em campanhas de marketing direto (Pereira, 2015).

Tabela 3.6 Variáveis criadas para a realização do *dashboard*

(conclusão)

Variáveis	Descrição	Tabela	Tipo
Valor monetário acumulado (atual)	Valor monetário total doado	Doadores	Quantitativa contínua
Valor médio de doação (atual)	Valor médio total doado	Doadores	Quantitativa contínua
Mês em que ocorreu a última doação (atual)	Mês em que ocorreu a última doação	Doadores	Quantitativa discreta
Perfis de alta propensão a doar ou não	A partir dos perfis resultantes do modelo preditivo, indica se o doador é proporcional a doar ou não	Doadores	Qualitativa nominal
Intervalos de idades	Intervalo de 5 em 5 anos no Ano N-1	Doadores	Qualitativa nominal
Date	Todos os dias no período de 01/01/2015 a 31/01/2022	Calendário	Data
Mês	Mês da date	Calendário	Quantitativa discreta
Dia	Dia da date	Calendário	Quantitativa discreta

3.3.3 Seleção dos dados

A etapa “seleção dos dados” visa selecionar as variáveis a utilizar na próxima etapa do CRISP-DM, mais precisamente na criação dos modelos preditivos.

Nesse sentido, tendo como base a RSL e o consenso entre dois especialistas, um de *analytics* e outro de *fundraising*, de que tais variáveis são recorrentes nos bancos de dados da maioria das organizações (explicitamente ou implicitamente), são selecionadas as seguintes variáveis independentes: “Idade no ano n-1”, “Faixa etária n-1”, “Senioridade n-1”, “Senioridade classificação”, “N.º de anos que não doa”, “N.º de doações”, “Mês em que ocorreu a última doação”, “Estado n-1”, “Recency Score”, “Frequency Score”, “Monetary Score”, “RFM Score”.

É de salientar que, embora tenha sido realizado um tratamento de dados detalhado à variável “Habilitações”, por apresentar 56% de valores omissos, é descartado o seu uso em prol de obter melhores resultados nos modelos preditivos.

Por outro lado, o *target* (variável dependente) da investigação corresponde à variável binária “Doou no ano n”.

3.3.4 Criação do modelo de dados do *dashboard*

Segundo Han *et al.*, (2012), o modelo de dados mais utilizado nas organizações é o modelo multidimensional, que pode ser construído em forma de estrela, floco de neve ou constelação de factos. Nesse sentido, desenvolve-se um modelo em floco de neve como se pode observar na Figura 3.3.

Detalhadamente, a tabela “calendário” estabelece uma ligação um para muitos com a tabela “doação”, tendo como intersector o campo “date” e “data da doação”, respetivamente. De outra forma, a tabela “doadores” estabelece uma ligação um para muitos com a tabela “doações”, a partir do campo “identificação” presente nas duas tabelas.

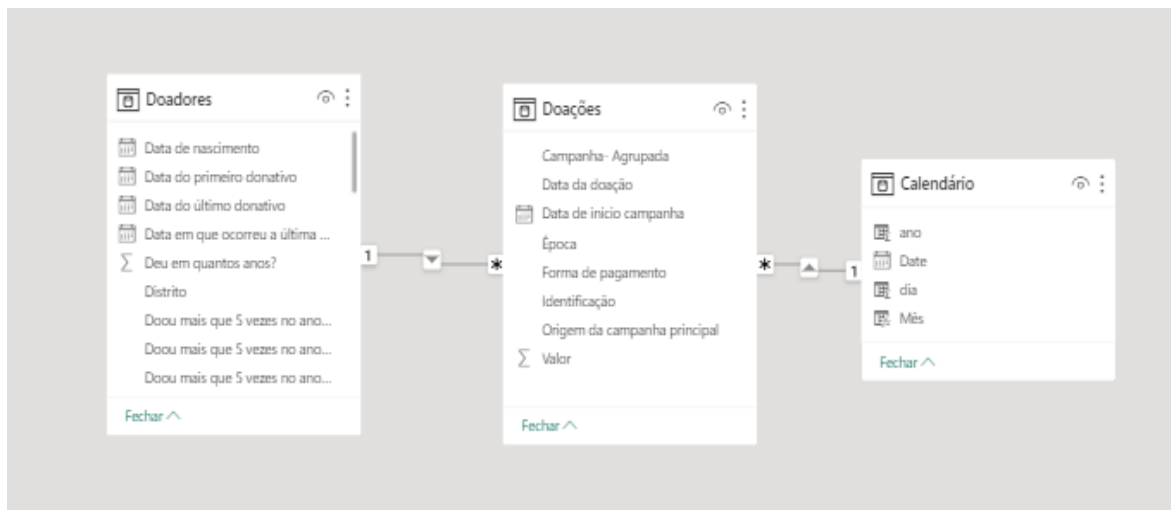


Figura 3.3 Modelo de dados do *dashboard*

3.4 Modelação

3.4.1 Modelo preditivo

Na fase da modelação, inicia-se a primeira iteração com a fase de avaliação. Selecionam-se as técnicas de modelação a utilizar tendo em conta os objetivos iniciais da investigação. Simultaneamente, delinea-se o procedimento de avaliação e produzem-se vários modelos para que sejam avaliados e, eventualmente, revistos. Por outro lado, dada a natureza iterativa do CRISP-DM, mantém-se em aberto a possibilidade de refazer algum passo ou até mesmo de voltar à etapa anterior para melhorar ou criar alguma variável.

No decorrer do desenvolvimento do modelo preditivo, testam-se as inúmeras técnicas de modelação identificadas na literatura, assim como outras técnicas recomendadas por um especialista da área de *analytics*. As técnicas que obtiveram melhores resultados são as seguintes:

- Redes Neurais Artificiais (RNA) — técnica inspirada em simular ações do cérebro humano. Possuem a versatilidade de serem utilizadas em problemas de regressão e de classificação, sendo que são praticamente incompreensíveis e dificultam a explicação ou justificação das previsões geradas (Maimon & Rokach, 2010).
- QUEST- — técnica proposta por Loh e Shihem (1997). A velocidade de cálculo do QUEST é maior do que noutros métodos baseados em árvores de decisão. O algoritmo é o mais adequado para variáveis dependentes de categoria múltipla e realiza também partições binárias (Lin & Fan, 2019).
- C5 — técnica resultante da atualização do algoritmo C4.5 (extensão de Quinlan de seu próprio algoritmo ID3). As divisões não são necessariamente binárias e, para variáveis categóricas, a técnica produz ramificações para cada categoria, que, conseqüentemente, o C5 produz uma árvore de formato mais variável (Larose & Larose, 2015).
- CHAID — significa Deteção Automática de Interação Qui-Quadrado projetado por Kass (1980). É uma técnica especialmente popular por se basear num dos testes de significância estatística mais usados — o teste Qui-Quadrado de Pearson (Mendes, 2022).

Adicionalmente, de forma a melhorar a qualidade dos modelos desenvolvidos, recorre-se ao uso de técnicas de *boosting* e de *bagging* (Zacarias, 2022):

- *Boosting* — é um método usado para reduzir erros na análise preditiva de dados. Um único modelo pode ter erros de previsão, dependendo da precisão do conjunto de dados de treino. O *boosting* tenta superar esse problema, treinando vários modelos sequencialmente para melhorar a precisão geral do modelo na fase de treino.

- *Bagging* — também conhecido como *Bootstrap Aggregating*, é um método usado para melhorar a capacidade de generalização de um modelo preditivo, isto é, ao utilizar-se o *bagging*, a estabilidade de um modelo preditivo aumenta (Abrantes, 2020).

Por outro lado, com o intuito de anular a complexidade dos modelos, opta-se por utilizar apenas uma profundidade de cinco níveis nas técnicas de modelação C5, Quest e CHAID, visto que se trata da profundidade padrão no *SPSS Modeler*. Por isso mesmo, opta-se por definir as referidas técnicas impondo, como mínimo, dois casos de nó pai e um caso de nó filho.

É de referir que, à medida que os modelos vão sendo realizados, são comparados com o melhor modelo até então com a mesma parametrização. Caso os resultados obtidos sejam inferiores, o modelo desenvolvido é eliminado. Em contrapartida, caso os resultados se apresentem realmente aperfeiçoados, o melhor modelo até então é substituído por este último. A Tabela 3.7 demonstra o melhor modelo desenvolvido, tendo em consideração as parametrizações utilizadas:

Tabela 3.7 Parametrizações dos algoritmos utilizados

Algoritmo	Redes Neurais Artificiais		
	Default	Boosting (n= 15)	Bagging (n= 15)
Modelo	RNN-1	RNN-2	RNN-3
Algoritmo	QUEST		
	Default	Boosting (n= 10)	Bagging (n= 10)
Modelo	QUEST-1	QUEST-2	QUEST-3
Algoritmo	C5		
	Default	Boosting (n= 15)	Bagging
Modelo	C5-1	C5-2	-
Algoritmo	CHAID		
	Default	Boosting (n= 15)	Bagging (n= 15)
Modelo	CHAID-1	CHAID-2	CHAID-3

3.4.2 Dashboard

Após a modelação de dados, procede-se ao desenvolvimento de um *dashboard* tático que permite entender o comportamento dos doadores a médio e longo prazo. O processo de criação resulta de três interações com os LD.

Na primeira interação, a responsável pelas angariações de fundos da organização explica que é fundamental conhecer as respostas às seguintes questões:

1. Qual a faixa etária mais usual de realizar doações?
2. Qual o género mais proeminente a realizar doações?

3. Qual a evolução anual e mensal de donativos realizados?
4. Qual a evolução anual e mensal do valor médio de doação?
5. Qual a evolução anual e mensal do valor doado?
6. Qual a evolução anual dos doadores que estão fidelizados (doadores que doam mais do que cinco vezes num ano)?
7. Qual a evolução anual do número de doadores?
8. Quais as características demográficas dos doadores mais propensos a doar e de não doar?

Deste modo, após a reunião, é desenhado um rascunho com inúmeros gráficos que possuem a informação solicitada, criando as medidas demonstradas Tabela 3.8 com o intuito de produzir certos gráficos (a coluna “Medidas” indica o nome dado à medida; a coluna “descrição” descreve em que consiste a medida).

Tabela 3.8 Medidas criadas para a realização do *dashboard*

Medidas	Descrição
#Count 2022	N.º de doações realizadas em 2022
#Count 2021	N.º de doações realizadas em 2021
#Count 2020	N.º de doações realizadas em 2020
Crescimento #	Comparação do número de doações do presente período com o período anterior (análise YTD)*
Crescimento do valor acumulado	Comparação do valor médio do presente período com o período anterior (análise YTD)*
Crescimento do valor médio	Comparação do número do valor acumulado do presente período com o período anterior (análise YTD)*
Valor acumulado 2022	Valor monetário total doado em 2022
Valor acumulado 2021	Valor monetário total doado em 2021
Valor acumulado 2020	Valor monetário total doado em 2020
Valor médio 2022	Valor médio doado em 2022
Valor médio 2021	Valor médio doado em 2021
Valor médio 2020	Valor médio doado em 2020

*Year to date⁵

Adicionalmente, são desenvolvidos filtros que permitem aumentar o nível de granularidade e interatividade do *dashboard*.

Segundo Aspin (2022), os filtros são uma parte fundamental de qualquer *dashboard*, uma vez que auxiliam o fornecimento de respostas de forma organizada e estruturada, permitindo, por sua vez, aos utilizadores aplicar uma restrição de resultados a um conjunto de valores. Assim, cria-se filtros que permitem ao utilizador selecionar um intervalo temporal (ano, mês,

⁵ A análise *Year to Date* (YTD) é uma métrica utilizada para acompanhar o progresso ou o desempenho ao longo de um período que começa no primeiro dia do ano e termina na data atual, independentemente de quando for essa data.

dia), o género do doador e a faixa etária, atendendo às características que o utilizador do *dashboard* procura para realizar a análise.

Na segunda interação, a partir do rascunho realizado, os LD escolhem os gráficos presentes na Tabela 3.9 (a coluna “Visual escolhido” indica os tipos de gráficos escolhidos; a coluna “questão que pretende responder” indica as questões que esses tipos de gráficos respondem). Após conhecer a disposição de *layout* que os LD pretendiam, seleciona-se uma disposição horizontal. Relativamente à paleta de cores a utilizar na conceção do *dashboard*, a opinião é consensual: as cores presentes no *website* da organização.

Tabela 3.9 Visual escolhido escolhidos pelos LD

Visual escolhido	Questão que pretende responder
Cartões	3.Qual a evolução anual e mensal de donativos realizados? 4.Qual a evolução anual e mensal do valor médio de doação? 5.Qual a evolução anual e mensal do valor doado?
<i>Waffle chart</i>	1.Qual a faixa etária mais usual de realizar doações?
Gráfico de barras	3.Qual a evolução anual e mensal de donativos realizados? 4.Qual a evolução anual e mensal do valor médio de doação? 5.Qual a evolução anual e mensal do valor doado?
Gráfico de área	6.Qual a evolução anual dos doadores que estão fidelizados (doadores que doam mais do que 5 vezes num ano)?
Gráfico de área	7.Qual a evolução anual do número de doadores?
Matriz	8.Quais são as características demográficas dos doadores mais propensos a doar e de não doar?
Tabela	3.Qual a evolução anual e mensal de donativos realizados? 4.Qual a evolução anual e mensal do valor médio de doação? 5.Qual a evolução anual e mensal do valor doado?

Assim sendo, posteriormente à segunda interação, é inserida uma disposição horizontal na única folha do *dashboard*. De outra forma, recorre-se ao *powerpoint* para criar o *background* da referida folha, no qual se insere o resultado final como imagem de fundo. Por fim, introduzem-se os visuais e os filtros pretendidos pela organização, utilizando sempre a paleta de cores selecionada.

A terceira interação serve para confirmar se estava tudo em conformidade e de acordo com as expectativas da organização.

3.5 Avaliação

3.5.1 Modelo Preditivo

A etapa de avaliação está subdividida em várias fases, em que algumas delas começam ainda durante a etapa de modelação, mais precisamente a avaliação dos modelos através de métricas específicas.

Atendendo ao objetivo de criar um modelo preditivo com enfoque em técnicas de classificação, isto é, se o doador doou ou não numa campanha do ano de 2022, e na dimensão da amostra ($n=1\ 543$), aplica-se a técnica *holdout* para avaliar os modelos. Assim, utiliza-se o critério de divisão aleatória de 70% dos dados para treino e os restantes 30% para teste, visto que, segundo Su *et al.* (2021), é o critério mais frequente em estudos com a mesma quantidade de dados.

Para cada modelo, analisa-se a matriz confusão (Tabela 3.10) em que as linhas representam as classificações reais e as colunas demonstram os valores previstos (Berthold *et al.*, 2020). Desta forma, calcula-se as métricas de avaliação, tanto para o conjunto de treino como para o conjunto de dados para teste.

Tabela 3.10 Matriz de confusão

		Previsão	
		0 (não doa)	1 (doa)
Real	0 (não doa)	Verdadeiros Negativos (VN)	Falsos Positivos (FP)
	1 (doa)	Falsos Negativos (FN)	Verdadeiros Positivos (VP)

As métricas a considerar na avaliação são:

- *Accuracy* — também conhecida como exatidão ou percentagem de casos corretamente classificados (PCCC) — mede a taxa de acerto total de um modelo preditivo, no qual, se o resultado for mais próximo de 1, mais semelhantes são as previsões dos valores reais. A *accuracy* (3.1) utiliza todas as classes da matriz de confusão no seu cálculo (Berthold *et al.*, 2020) e é dada pela expressão:

$$Exatidão = \frac{VN+VP}{VN+VP+FN+FP} \quad (3.1)$$

- Sensibilidade (3.2) ou *recall* — mede a capacidade de o modelo reconhecer corretamente a classe positiva, mais precisamente, indica se o modelo consegue classificar corretamente os verdadeiros positivos (Berthold *et al.*, 2020). A métrica é dada pela expressão:

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \quad (3.2)$$

- Especificidade (3.3) — mede a capacidade de o modelo reconhecer o que não pertence à classe positiva, ou seja, indica se o modelo consegue classificar corretamente os verdadeiros negativos (Berthold *et al.*, 2020). A métrica é dada pela expressão:

$$\text{Especificidade} = \frac{VN}{VN+FP} \quad (3.3)$$

- Precisão (3.4) — traduz o rácio entre o número de registos classificados corretamente como positivos e o número total de registos classificados, incorreta ou corretamente como positivos (Berthold *et al.*, 2020). A precisão é dada pela expressão:

$$\text{Precisão} = \frac{VP}{VP+FP} \quad (3.4)$$

- Medida-F (3.5) ou *F-score* — pode ser interpretada como uma média ponderada entre a precisão e a sensibilidade. Atinge o seu melhor valor quando é 1 e o pior quando é 0 (Berthold *et al.*, 2020). A sua expressão de cálculo é:

$$\text{Medida F} = 2 * \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (3.5)$$

Para a avaliação dos modelos, é ainda calculada a AUC (Área sob a curva ROC), sendo que quanto maior a área por baixo da curva ROC melhor é a solução para um problema de classificação (Berthold *et al.*, 2020).

É essencial salientar que se utiliza um grande conjunto de métricas com o intuito de controlar, da melhor forma, os erros de cada modelo. Em certos casos, a forma de cálculo de cada uma das métricas pode induzir em erro. Por exemplo, apesar de a exatidão ser uma das medidas mais utilizadas, pode ser facilmente enviesada se existir uma enorme disparidade na quantidade de casos de cada classe alvo (Mendes, 2022).

Deste modo, apesar de todas as métricas serem consideradas, a medida-F e a especificidade são as que têm um maior peso na seleção do melhor modelo. A razão para a

escolha da medida-F deve-se ao facto de englobar a precisão e a sensibilidade, mais precisamente, de permitir o equilíbrio entre a capacidade do modelo de identificar corretamente os verdadeiros positivos (doador provável de doar) e evitar falsos positivos (doador identificado incorretamente como provável de doar), enquanto a preferência pela especificidade, deve-se a esta se focar na classe que não doou.

3.5.2 Dashboard

De forma a avaliar o *dashboard*, é realizada uma entrevista via *teams* com a coordenadora de angariações de fundos dos LD, e principal utilizadora desta ferramenta de visualização, durante a qual são realizadas as seguintes perguntas:

1. O *dashboard* revela-se útil na realização das suas funções de trabalho?
2. Está de acordo com a informação exposta no *dashboard*?
3. Numa escala de 1 a 5, sendo um, péssimo e cinco, excelente, como classifica a usabilidade do *dashboard*?
4. Numa escala de 1 a 5, sendo 1 um, péssimo e cinco, excelente, como classifica geralmente o *dashboard*?

Considera-se o *dashboard* finalizado quando a satisfação for “sim” na primeira e segunda questão, e for atribuída a classificação “5” nas perguntas 3 e 4.

3.6 Implementação

Atendendo à duração do projeto, a implementação do modelo preditivo e do *dashboard* não é possível. Desta forma, apenas são facultados protótipos do modelo e da ferramenta de visualização, que permitem à organização testar estas duas novas ferramentas.

4 Resultados e discussão

4.1 Modelos preditivos

4.1.1 Avaliação dos modelos preditivos

Na medida em que a intenção de doação é influenciada por um conjunto de preditores e que à partida não há nenhum algoritmo e respetiva parametrização que se possa afirmar que leva a melhores previsões, realizam-se inúmeras experiências de desenvolvimento. Todas as técnicas foram submetidas às diferentes parametrizações (*default*, *boosting* e *bagging*), com o objetivo de aperfeiçoar os resultados obtidos. Similarmente, considerou-se a problemática do sobreajustamento, no qual tentou reduzir-se a complexidade, impossibilitando que a aprendizagem fosse tão detalhada no treino para se conseguir obter melhores resultados no teste. Por fim, embora tenha sido mencionado na secção 3.5.1, as métricas medida-F e a especificidade possuem maior peso na apresentação do melhor modelo comparativamente às outras.

Posto isto, a Tabela 4.1 apresenta os resultados das métricas de avaliação obtidos por cada algoritmo nas várias parametrizações. Ao analisar detalhadamente cada técnica de modelação, o modelo RNN2 exhibe indícios de sobreajustamento, visto que ocorre uma enorme disparidade entre os resultados do conjunto de treino e de teste. Por exemplo, enquanto o conjunto de treino regista uma sensibilidade (99,7%) e precisão (100%) substancialmente elevadas, os valores correspondentes no conjunto de teste são reduzidos, com valores de 75,2% e 75,7%, respetivamente. Similarmente, a mesma situação acontece no modelo RNN3, isto é, ao analisar a medida F, a métrica passa dos 0,9511 no conjunto de treino para 0,7746 no teste.

Tal situação repete-se no algoritmo CHAID-2, embora não seja tão acentuada como sucede nas RNN mencionadas. Por exemplo, o algoritmo apresenta 95,4% de sensibilidade no treino comparativamente a 74,5% no teste.

Ainda ao analisar a Tabela 4.1, observa-se que o melhor modelo produzido com o algoritmo RNN corresponde ao RNN1, apresentando uma especificidade de 91,7% e um valor de F1 de 0,7653 no teste. Por outro lado, o modelo mais proeminente que utiliza a técnica de modelação QUEST é o QUEST-3, com uma especificidade de 93,7% e um valor de F1 de 0,7442 no teste (sendo a especificidade maior no QUEST-1 e a medida F melhor no QUEST-3, recorre-se às restantes métricas para seleccionar como melhor o QUEST-3). Relativamente ao modelo criado com a técnica de modelação C5, destaca-se o C5-2 com uma especificidade de 91,4% e 0,7536 de medida F. Por fim, o melhor modelo criado com o algoritmo CHAID corresponde ao CHAID 1, demonstrando no teste uma especificidade de 93% e de medida F 0,7766.

Tabela 4.1 Resultados dos modelos obtidos nas diferentes parametrizações

Treino							
Técnica	Parametrização	Especificidade	Sensibilidade	Precisão	Medida F	Accuracy	AUC
RNN-1	Default	93,4%	79,1%	86,0%	0,8239	86%	0,937
RNN-2	<i>Boosting</i>	100%	99,7%	100%	0,9986	99,9%	1
RNN-3	<i>Bagging</i>	97,7%	94,8%	95,4%	0,9511	96,7%	0,989
QUEST-1	<i>Default</i>	94,4%	69,6%	86,8%	0,7727	86,1%	0,894
QUEST-2	<i>Boosting</i>	92,7%	78,8%	84,6%	0,8160	88,0%	0,903
QUEST-3	Bagging	93,7%	74,5%	85,8%	0,7975	87,2%	0,907
C5-1	<i>Default</i>	94%	75,6%	86,6%	0,8073	87,8%	0,873
C5-2	Boosting	93,4%	81,9%	86,4%	0,8412	89,5%	0,938
CHAID-1	Default	94,6%	77,1%	87,9%	0,8214	88,7%	0,939
CHAID-2	<i>Boosting</i>	94,4%	95,4%	96,74%	0,9610	97,4%	0,996
CHAID-3	<i>Bagging</i>	93%	80,2%	85,9%	0,8272	88,7%	0,973
Teste							
Técnica	Parametrização	Especificidade	Sensibilidade	Precisão	Medida F	Accuracy	AUC
RNN-1	Default	91,7%	73,1%	80,3%	0,7653	85,8%	0,919
RNN-2	<i>Boosting</i>	88,8%	75,2%	75,7%	0,7543	84,5%	0,869
RNN-3	<i>Bagging</i>	90,7%	75,9%	79,1%	0,7746	86,0%	0,897
QUEST-1	<i>Default</i>	94,6%	66,2%	85%	0,7442	85,6%	0,893
QUEST-2	<i>Boosting</i>	90,1%	71,0%	76,9%	0,7384	84,1%	0,893
QUEST-3	Bagging	93,6%	69,7%	83,5%	0,7594	86%	0,898
C5-1	<i>Default</i>	92,7%	69%	81,3%	0,7463	85,2%	0,862
C5-2	Boosting	91,4%	71,7%	79,4%	0,7536	85,2%	0,905
CHAID-1	Default	93%	73,1%	82,8%	0,7766	86,7%	0,901
CHAID-2	<i>Boosting</i>	89,1%	74,5%	76,1%	0,7526	84,6%	0,903
CHAID-3	<i>Bagging</i>	91,4%	72,4%	79,5%	0,7581	85,4%	0,908

Comparando os melhores modelos de cada algoritmo, conclui-se que o CHAID-1 é o que obtém melhor resultado nas métricas de avaliação. Este resultado contraria as expectativas, uma vez que o algoritmo RNN é identificado ao realizar a RSL como a técnica de modelação com melhores resultados (Farrokhvar *et al.*, 2018; Cacciarelli & Boresta, 2022).

Deste modo, o melhor modelo demonstra possuir a capacidade de prever as pessoas que não doam (especificidade de 93%) e as que doam (Medida F- 0,7766).

Ao comparar os resultados com os de Cacciarelli e Boresta (2022), a AUC revela-se superior neste estudo com uma AUC de 0,901, em comparação com uma AUC de 0,85. Todavia, ao comparar-se a mesma técnica de modelação, RNN, o presente estudo consegue

igualmente ser superior, uma vez que a AUC da melhor RNN (RNN 1) é de 0,919 em oposição à AUC 0,85 da RNN do estudo de Cacciarelli e Boresta (2022).

4.1.2 Importância das variáveis

Considerando os melhores modelos preditivos de cada algoritmo, torna-se relevante conhecer a importância que cada variável independente possui nos modelos desenvolvidos.

A Tabela 4.2 Importância das variáveis independentes apresenta os preditores utilizados em cada algoritmo e a sua importância para a previsão. De facto, é notável que a *Recency Score* se destaca por duas vezes como preditor mais relevante, nomeadamente nos modelos QUEST 3 e CHAID-1, e, inclusive, no último modelo referenciado, a variável assume uma importância de previsão bastante elevada (0,69). Sob outra perspetiva, ao analisar a quantidade de variáveis independentes que cada modelo utiliza, o RNN-1 é aquele que possui preditores em maior número (9), contrariamente ao que acontece com o CHAID-1 e o C5-2, uma vez que apenas possuem seis preditores.

Tabela 4.2 Importância das variáveis independentes

Variáveis Independentes							
RNN-1	Importância	QUEST-3	Importância	C5-2	Importância	CHAID-1	Importância
N.º de anos que não doa	0,22	<i>Recency Score</i>	0,31	Estado n-1	0,20	<i>Recency Score</i>	0,69
Senioridade	0,19	Estado n-1	0,25	Mês em que ocorreu a última doação	0,19	Deu em quantos anos?	0,12
<i>Recency Score</i>	0,17	N.º de anos que não doa	0,20	<i>Frequency Score</i>	0,16	Mês em que ocorreu a última doação	0,06
Deu em quantos anos?	0,12	Mês em que ocorreu a última doação	0,09	<i>Monetary Score</i>	0,15	Senioridade	0,06
Estado n-1	0,11	<i>Frequency Score</i>	0,07	Idade no ano n-1	0,15	N.º de doações	0,05
<i>Frequency Score</i>	0,06	Senioridade Classificação	0,03	Senioridade	0,15	Idade no ano n-1	0,02
Idade no ano n-1	0,06	Faixa etária n-1	0,01	-	-	-	-
<i>Monetary Score</i>	0,04	<i>Monetary Score</i>	0,01	-	-	-	-
Sexo	0,04	-	-	-	-	-	-

No que respeita ao número de vezes que cada preditor é usado, a figura 4.1 evidencia que os seguintes preditores surgem três vezes: “Senioridade”; “Recency Score”; “Estado N-1”, “Frequency Score”, “Idade no ano n-1”, “Monetary Score”, “Mês em que ocorreu a última doação” surgem por três vezes. Por outro lado, as variáveis “Sexo”, “Senioridade Classificação”, “Faixa etária”, “Nº de doações” surgem uma única vez.

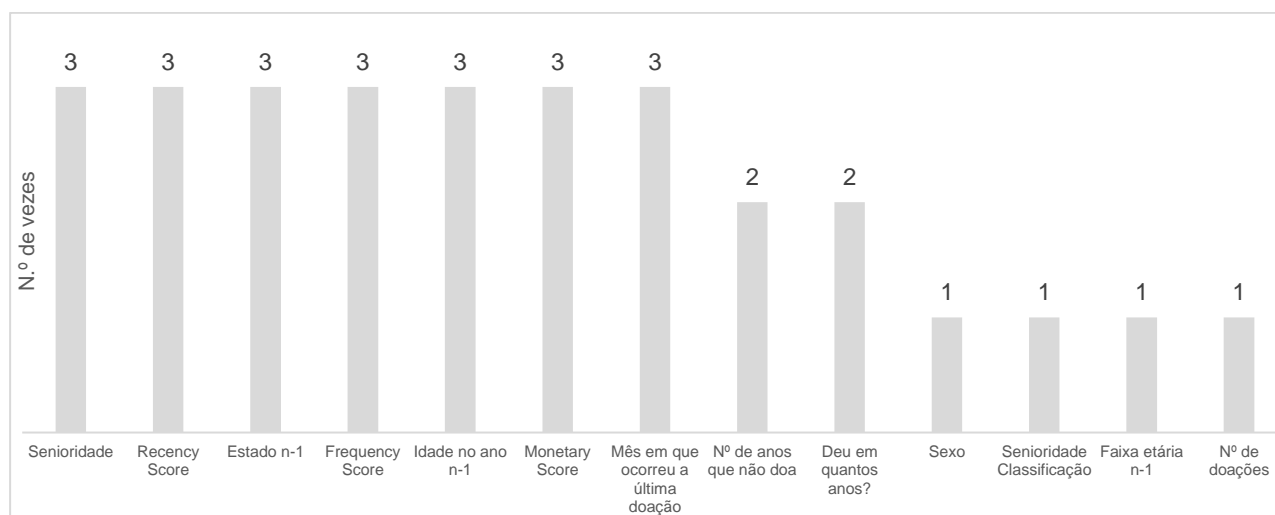


Figura 4.1 N.º de vezes que cada preditor surge

A figura 4.1 também demonstra que as variáveis em estudo presentes na literatura⁶ são realmente importantes na previsão da doação por parte de um doador. De outro modo, as variáveis sugeridas pelo especialista na área de *fundraising*, “Deu em quantos anos?”, “Nº de anos que não doa”, “Recency Score”, “Frequency Score”, “Monetary Score”, assumem-se igualmente como preditores. Por fim, era expectável que a variável “Sexo” aparecesse mais vezes como um preditor nas técnicas de modelação desenvolvidas, uma vez que é bastante referenciada na literatura.

Relativamente ao modelo seleccionado, o que obteve melhores avaliações (CHAID-1) possui os seguintes preditores: “Recency Score” (com uma importância de 0,69); “Deu em quantos anos?” (com uma importância de 0,12); “Mês em que ocorreu a última doação” (com uma importância de 0,06); “Senioridade” (com uma importância de 0,06); “N.º de doações” (0,05 de importância); e “Idade no ano n-1” (0,02 de importância).

⁶ Consultar no anexo D: Tabela - Preditores presentes na RSL

4.1.3 Perfis de doadores com elevada e baixa propensão a doar

Com base nas regras provenientes do modelo CHAID-1⁷, mais precisamente, a partir das regras que possuem uma maior confiança e um maior número de doadores de suporte, são descritos sete perfis que permitem, de forma mais clara, conhecer a previsão da combinação de características que provocam a realização de um donativo.

Neste sentido, os perfis de doadores que possuem uma baixa propensão a doar são:

- Com 98,6% de confiança e um suporte de 284 doadores, doadores que possuem *Recency Score* de 1 ou de 2 e senioridade igual ou inferior a 9 anos;
- Com 97% de confiança e um suporte de 72 doadores, os doadores que possuem *Recency Score* de 1 ou de 2, senioridade superior a 9 anos e idade igual ou superior a 56 anos;
- Com 95% de confiança e um suporte de 79 doadores, os doadores que possuem *Recency Score* de 3 e senioridade igual ou inferior a 5 anos;
- Com 95% de confiança e um suporte de 79 doadores, os doadores que possuem *Recency Score* de 4 e que doaram no período temporal de 1 ou 2 anos.

Por outro lado, os perfis de doadores com alta propensão a doar caracterizam-se da seguinte forma:

- Com 93,1% de confiança e um suporte de 27 doadores, os doadores que possuem *Recency Score* de 4 e que doaram em 7 ou mais anos;
- Com 94% de confiança e um suporte de 157 doadores, os doadores que possuem *Recency Score* de 5 e que doaram em 2 anos ou mais;
- Com 96,8% de confiança e um suporte de 30 doadores, os doadores que possuem *Recency Score* de 5, que doaram em 1 ano ou menos e mais do que 7 vezes.

Desta forma, observa-se que doadores que fizeram a última doação até 31/12/2018 (*Recency Score* de 1), geralmente, não possuem a intenção de doar novamente, especialmente, se a primeira doação deles ocorreu nos últimos nove anos. Além disso, quando a primeira doação aconteceu há mais de nove anos, também se observa que doadores com mais de cinquenta e cinco anos exibem uma propensão baixa para realizar novas doações.

Ao examinar um período diferente, mais precisamente, quando a última doação ocorreu entre 01/01/2019 e 09/12/2020 (*Recency Score* de 2), observa-se que, quando a primeira

⁷ Consultar no anexo E: Figura - *Output* do CART-1

doação foi realizada há menos de cinco anos, os doadores tendem novamente a ter uma menor propensão para fazer novas doações.

Analisando o cenário em que o doador efetuou a sua última doação entre 10/12/2020 e 01/12/2021 (*Recency Score* de 4), surge uma tendência: se o doador doou pelo menos em sete anos, é mais provável que continue a demonstrar intenção de doar. No entanto, a situação muda quando o doador fez apenas uma ou duas doações. Nesse caso, é mais provável que opte por não realizar uma nova doação.

Por fim, ao observar o período subjacente a 01/12/2021 (*Recency Score* de 5), doadores que fizeram doações com intervalos superiores a dois anos parecem mais propensos a retomar as suas doações. Além disso, é interessante observar que, no caso de doadores que deram uma única doação em um ano ou menos, mas que já realizaram mais de sete doações ao longo do tempo, também apresentam uma propensão a retomar as suas contribuições.

4.2 Dashboard

4.2.1 Descrição e avaliação do *dashboard* final

O *dashboard* desenvolvido (Figura 4.2), embora já referido, permite fornecer aos LD informações sobre o comportamento dos seus doadores. É composto por um único ecrã (ou folha), seguindo uma estrutura horizontal. Além disso, diferencia-se por ser totalmente interativo, uma vez que, a partir dos filtros criados, possibilita ao utilizador selecionar somente informação correspondente a um ano ou mês, assim como a faixa etária do doador ou género que quer analisar.

É de salientar que os gráficos “Evolução do n.º de doações por ano e mês”, “Evolução do valor médio de doações por ano e mês” e a “Evolução do valor doado por ano e mês”, assim como a matriz “Características Demográficas (MP)”⁸ possuem a capacidade de realizar um *drill down* (Figura 4.3).

⁸ Modelos Preditivos (MP)

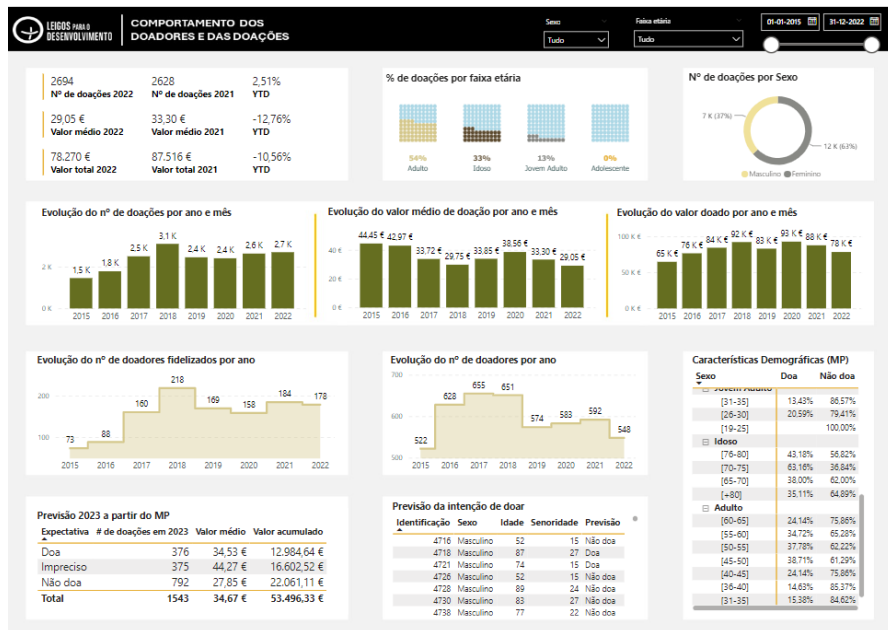


Figura 4.2 Dashboard sem drill down

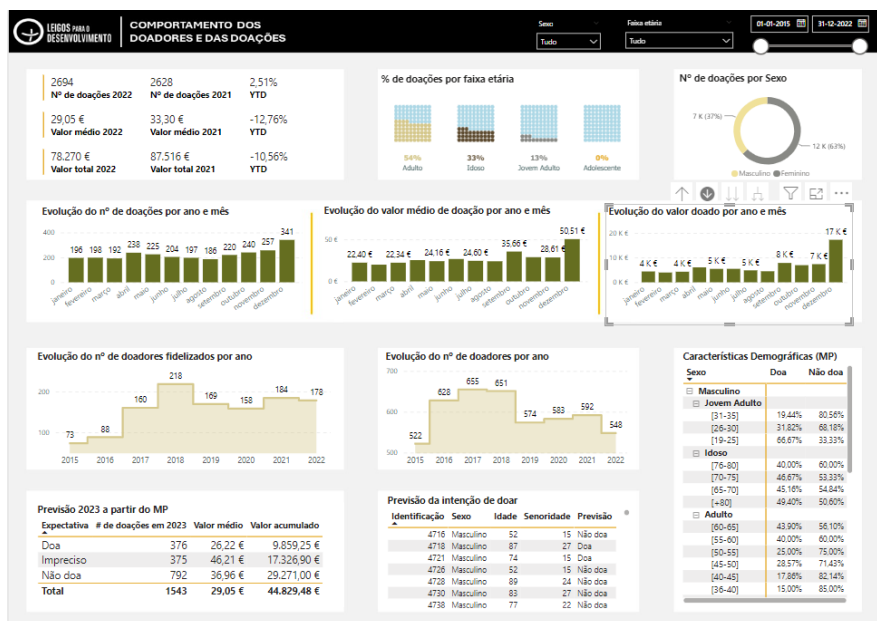


Figura 4.3 Dashboard com drill down

Relativamente à avaliação do *dashboard*, como referido na seção 3.5.2, realizou-se uma entrevista na qual as questões apresentadas obtiveram as seguintes respostas (transcrição completa):

Entrevistador — O *dashboard* revela-se útil na realização das suas funções de trabalho?

Entrevistada — Sim, definitivamente.

Entrevistador — Está de acordo com a informação que está exposta no *dashboard*?

Entrevistada — Estou de acordo, está um trabalho muito bem feito que vamos poder utilizar e aprofundar no futuro para uma maior eficácia e qualidade das nossas campanhas.

Entrevistador — Numa escala de 1 a 5, sendo um, péssimo e cinco, excelente, como classificaria a usabilidade do *dashboard*?

Entrevistada — 5

Entrevistador — Numa escala de 1 a 5, sendo um, péssimo e cinco, excelente, como classifica o *dashboard*?

Entrevistada — 5

De forma geral, é possível verificar que a organização está bastante satisfeita com o trabalho que foi realizado.

4.2.2 Caracterização do comportamento dos doadores e dos donativos

Analisando o *dashboard* ao detalhe, é possível caracterizar o comportamento dos doadores, assim como dos donativos.

No intervalo temporal de 2015 a 2022, a faixa etária que se destaca por ter realizado mais doações corresponde a “Adulto”, tendo cerca de 54% de doações feitas, como demonstra a Figura 4.4. Por outro lado, é possível observar que a faixa etária “adolescente” nem chega a 1% das doações realizadas (só existem 15 doações feitas). Relativamente ao sexo que realiza mais doações, os doadores femininos (63%) destacam-se comparativamente aos masculinos (37%).

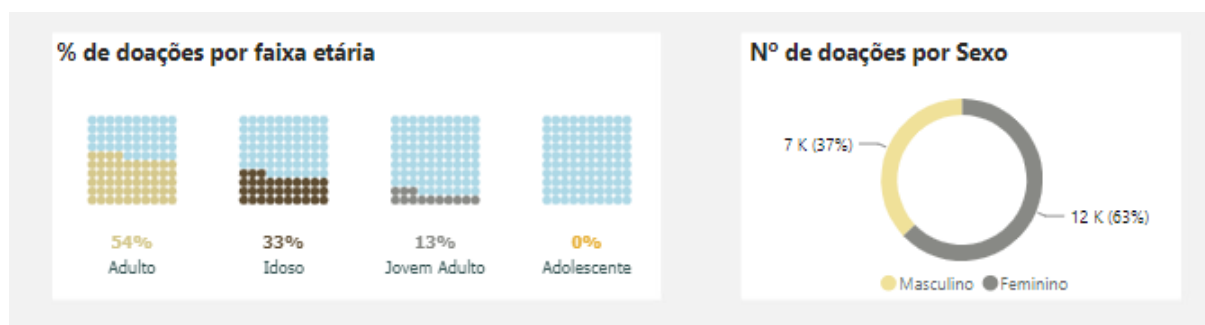


Figura 4.4 Percentagem do n.º de doações por faixa etária e sexo

Analisando a evolução do n.º de doações, é possível observar na Figura 4.5 que o ano de 2018 registou o maior número de doações, com um total de 3093. No entanto, após esse desempenho positivo, as doações diminuíram nos anos de 2019 e 2020. Em contrapartida a partir do ano de 2021, é notável um aumento gradual na quantidade de doações realizadas até ao ano 2022.

Ainda a partir da Figura 4.5, verifica-se que o valor médio de doação mais elevado ocorreu em 2015 (44,45€), enquanto o valor médio de doação mínimo registado aconteceu em 2022 (29,05€). De facto, o valor médio de doação de 2022 sugere que, embora as pessoas tenham gradualmente retomado as suas vidas e o número de doações tenha aumentado após o período intenso da pandemia COVID-19, os doadores dos LD sofreram o impacto da inflação vivida em Portugal (subida de 7,8%, segundo o Instituto Nacional de Estatística) (Neto, 2023).

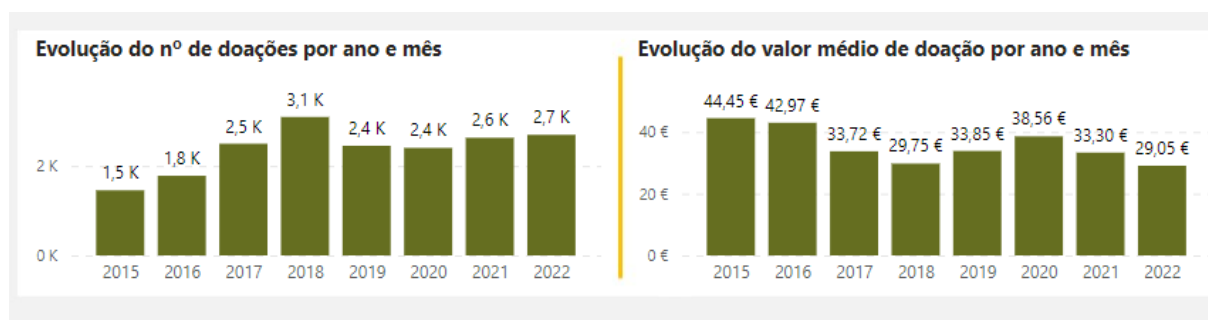


Figura 4.5 Evolução do n.º de doações e do valor médio doado por ano

É ainda de salientar que os doadores do sexo masculino são muito mais generosos comparativamente ao sexo feminino, apresentando um valor médio de doação maior, principalmente se estiverem inseridos na faixa etária “idoso”.

No que diz respeito ao valor total que foi doado por ano, a Figura 4.6 evidencia um aumento considerável de 2015 a 2018 que pode ser atribuído ao aumento do número de doações durante esse período. O ano de 2019, por outro lado, apresentou uma diminuição no valor doado, o que pode ser explicado pela redução no número de doações.

O valor de 93 mil euros registado em 2020 (o maior montante durante o período de 2015 a 2022) é resultado do valor médio de doação de 38,56 euros. No entanto, o declínio nos valores arrecadados em 2021 e 2022 pode ser diretamente atribuído à diminuição média das doações de cerca de 10 euros nesse período.



Figura 4.6 Evolução do valor doado por ano

Ainda de constatar que o mês de dezembro é o mês que apresenta maiores valores de doações, o valor médio de doação mais elevado e o maior valor acumulado, sendo que tal tendência é registada em todos os anos do período de 2015 a 2022.

A partir da Figura 4.7, é observável que o número de doadores aumentou de 2015 para 2016, passando de 512 doadores para 628, respetivamente. No ano de 2017, a organização registou o maior número de doadores, atingindo 655 doadores ativos. No entanto, tal número voltou a diminuir nos anos de 2018 (651 doadores) e 2019 (574 doadores). Em relação aos anos de 2020 e 2021, houve novamente um aumento de doadores ativos, com 2020 a registar um número de 583 doadores e 2021 a registar 592. Finalmente, em 2022, ocorreu uma redução para 548 doadores.

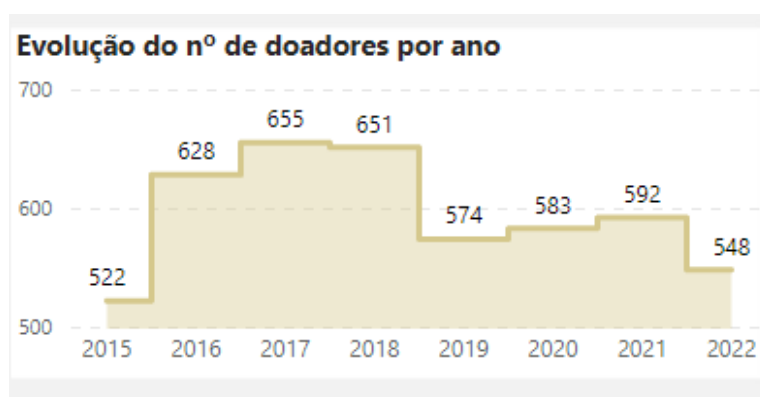


Figura 4.7 Evolução do n.º de doadores por ano

Analisando a Figura 4.8, note-se que o ano em que a organização teve mais doadores fidelizados (doadores que fizeram mais de 5 doações anuais) foi 2018 (218 doadores fidelizados). Em 2019 e 2020, tal número diminuiu para 169 e 158 doadores, respetivamente. No ano de 2021, o valor voltou a aumentar, registando 184 doadores fidelizados. Por outro lado, em 2022, o número de doadores fidelizados diminuiu novamente, totalizando 178 doadores que realizam mais que 5 doações anuais.



Figura 4.8 Evolução do n.º de doadores fidelizados por ano

A Figura 4.9 demonstra as características demográficas mais presentes em indivíduos que apresentam uma maior probabilidade de realizar uma doação, segundo os 7 perfis alcançados na criação do modelo preditivo. De forma geral, pessoas do sexo feminino com idades compreendidas entre os 19 e 25 anos (100% dos doadores encaixam nos perfis propensos a não doar) e os 31 e 40 anos (mais de 80% dos doadores encaixam nos perfis propensos a não doar) tendem a não realizar doações. Este fator repete-se em doadores do sexo masculino, mas apenas têm idades compreendidas entre 31 e 45 anos (mais de 80% dos doadores encaixam nos perfis propensos a não doar).

Contrariamente, um doador masculino que tenha entre 19 e 25 anos tende a doar (66,67% encaixam nos perfis propensos a doar); simultaneamente, um doador feminino, que tenha entre 70 a 75 anos, tende a ter a mesma intenção (63,16% encaixam nos perfis propensos a doar).

É de referir que devido à falta de dados demográficos não foi possível construir uma matriz com mais detalhe.

Sexo	Doa	Não doa
Masculino		
Jovem Adulto		
[19-25]	66,67%	33,33%
[26-30]	31,82%	68,18%
[31-35]	19,44%	80,56%
Idoso		
[+80]	49,40%	50,60%
[65-70]	45,16%	54,84%
[70-75]	46,67%	53,33%
[76-80]	40,00%	60,00%
Adulto		
[31-35]	15,38%	84,62%
[36-40]	15,00%	85,00%
[40-45]	17,86%	82,14%
[45-50]	28,57%	71,43%
[50-55]	25,00%	75,00%
[55-60]	40,00%	60,00%
[60-65]	43,90%	56,10%
Feminino		
Jovem Adulto		
[19-25]		100,00%
[26-30]	20,59%	79,41%
[31-35]	13,43%	86,57%
Idoso		
[+80]	35,11%	64,89%
[65-70]	38,00%	62,00%
[70-75]	63,16%	36,84%
[76-80]	43,18%	56,82%
Adulto		
[31-35]	15,38%	84,62%
[36-40]	14,63%	85,37%
[40-45]	24,14%	75,86%
[45-50]	38,71%	61,29%
[50-55]	37,78%	62,22%

Figura 4.9 Características demográficas dos doadores mais propensos a doar

Analisando a Figura 4.10, é possível observar a previsão para o ano de 2023, segundo os perfis alcançados na criação do modelo preditivo. Sucessivamente, no ano em questão, espera-se, com uma confiança superior a 80%, que 376 doadores realizem uma doação com um valor médio de 34,53€ que resultará em 12 984, 64€ de receita para os LD. Por outro lado, novamente com uma confiança superior a 80%, espera-se que 792 não realizem uma doação

no ano em questão. Ainda de salientar que também se espera que o número de doadores imprecisos seja de 375 doadores, uma vez que os perfis em que os doadores estão inseridos não possuem uma confiança alta para afirmar que vão realizar uma doação.

Expectativa	# de doações em 2023	Valor médio	Valor acumulado
Doa	376	34,53 €	12.984,64 €
Impreciso	375	44,27 €	16.602,52 €
Não doa	792	27,85 €	22.061,11 €
Total	1543	34,67 €	53.496,33 €

Figura 4.10 Previsão do comportamento do doador no ano 2023

Por fim, a Figura 4.11 demonstra, tendo por base a previsão realizada pelo modelo preditivo criado, se um doador irá realizar uma doação no próximo ano.

Identificação	Sexo	Idade	Senoridade	Previsão
4716	Masculino	52	15	Não doa
4718	Masculino	87	27	Doa
4721	Masculino	74	15	Doa
4726	Masculino	52	15	Não doa
4728	Masculino	89	24	Não doa
4730	Masculino	83	27	Não doa
4738	Masculino	77	22	Não doa

Figura 4.11 Previsão da intenção de doar

5 Conclusão

5.1 Sumário da investigação

Esta dissertação procura responder à questão de investigação “Qual o comportamento do doador dos LD em futuras angariações de fundos?”, no qual o termo "comportamento do doador" é definido pela intenção de realizar uma doação.

Nesse sentido, são estipulados três objetivos de forma a responder a tal questão, nomeadamente a criação de um modelo preditivo capaz de prever a intenção de doar no próximo ano, utilizando, como fonte de dados, o histórico de doações e as características dos doadores; a identificação dos perfis de doadores que possuem alta ou baixa propensão a doar; e a criação de um *dashboard* para caracterizar o comportamento dos doadores e dos donativos realizados.

Por conseguinte, realizou-se uma RSL com o intuito de conhecer os preditores já utilizados, as técnicas de modelação que permitem chegar à previsão, assim como entender as métricas de avaliação outrora usadas.

Posteriormente, a partir de um conjunto de dados solicitados aos LD, composto por algumas características dos seus doadores e pelo histórico de doações, criou-se uma SP apta a realizar o tratamento de dados provenientes do CRM dos LD, de forma automática. Após a criação de novas variáveis pertinentes à realização do modelo e de se recorrer à utilização do *holdout* para dividir aleatoriamente os dados em 70% para o conjunto de treino e 30% para o conjunto de teste, foram criados vários modelos preditivos utilizando todas as técnicas de modelação mencionadas na literatura e consideradas relevantes por um especialista na área de *analytics*.

O CHAID-1 destacou-se como sendo o melhor modelo (93% de especificidade, 73,1% de sensibilidade, 82,8% de precisão, 0,7766 de medida F, 86,7% de *accuracy* e 0,901 de AUC), no qual os preditores que caracterizam o modelo podem ser encontrados nas bases de dados da maioria das ONG de forma direta ou indireta, uma vez que qualquer organização possui condições de calcular a *Recency Score*, obter o mês em que ocorreu a última doação por parte do doador, saber a antiguidade, assim como a idade e o número de doações que realizou.

Por outro lado, o CHAID-1, ao possuir como output uma árvore de decisão, possibilita traçar, com uma confiança superior a 80% e um suporte acima de 25 doadores, sete perfis de doações, nomeadamente quatro perfis propícios a não doar e três a doar.

Por fim, foi realizado um *dashboard* que permite responder, de forma sintetizada e numa única folha, todas as questões que os LD tinham sobre a caracterização do comportamento dos seus doadores e das doações realizadas.

Sendo assim, o objetivo de investigação “Criar um modelo preditivo da intenção de doar no próximo ano” é dado por cumprido, visto que o CHAID-1 alcança no mínimo, pelo menos, 70% em todas as métricas que o estudo considera importante e é constituído por preditores que podem ser encontrados nas bases de dados da maioria das organizações de forma direta ou indireta. Sob outra perspetiva, o output do modelo ao proporcionar sete perfis de doadores com uma confiança superior a 80% e um suporte acima de 25 doadores, possibilita dar por cumprido o objetivo de identificar perfis de doadores que possuem uma alta e baixa propensão a doar. Por último, o facto de a coordenadora de angariações de fundos dos LD dar uma avaliação elevada ao *dashboard* possibilita igualmente dar por cumprido o objetivo “criar um *dashboard* para caracterizar o comportamento dos doadores e dos donativos realizados”.

Concluindo, o cumprimento dos objetivos torna possível responder à questão principal da investigação “Qual o comportamento do doador dos LD em futuras angariações de fundos?”, uma vez que os resultados do modelo preditivo, inseridos no *dashboard*, identificam os doadores que irão realizar uma doação, os perfis mais propensos a doar e a não doar, bem como, as características demográficas que levam um doador a realizar uma doação.

5.2 Contributos

Atualmente, como acontece com inúmeras ONG, os LD enfrentam cada vez mais desafios do mundo moderno. Todavia, a consciencialização e o aproveitamento do conhecimento proveniente dos dados que cada organização gera torna exequível uma fácil superação das problemáticas que advém de tais desafios. A necessidade de estudar o ato de doar e os atributos que melhor explicam essas doações assumem, sucessivamente, um papel de relevo na sustentabilidade financeira de uma ONG.

Deste modo, esta dissertação contribui para que os LD tenham, por alguns anos, um sistema apto a prever o comportamento de um doador num determinado ano, bem como um *dashboard* que permite analisar a caracterização do comportamento dos doadores e das doações.

Detalhadamente, este estudo vai ajudar a organização a crescer analiticamente e possibilitar que redirecione o valor que está a despender ao contactar doadores que não vão doar. Similarmente, a dissertação permite que os LD observem, ao longo do tempo, quais as características demográficas mais propensas que levam um doador a realizar uma doação. Tal conhecimento contribui para que a organização direcione as suas campanhas de angariação de novos doadores para indivíduos que possuem as características identificadas, aumentando assim o leque de doadores fidelizados, já que essas pessoas tendem a fazer doações de forma regular.

Ainda de referir que o *dashboard* também contribui para a resposta às outras questões que os LD possuem, possibilitando que a responsável de fundos tome decisões que permitam uma maior eficácia das campanhas da organização.

Noutra perspetiva, o presente estudo também promove o conhecimento sobre *fundraising analytics* junto da comunidade científica mundial e introduz o conceito no paradigma português. Além disso, possibilita que qualquer ONG, que tenha uma base de dados com informações dos seus doadores e doações, possa também ter um modelo apto a prever o comportamento dos seus doadores, visto que o presente estudo somente utilizou, como preditores, variáveis que qualquer organização possui direta ou indiretamente.

5.3 Limitações

As limitações sentidas durante a investigação estão principalmente associadas a problemas com dados e ausência de adesão do doador ao contacto realizado pelos LD.

Cada organização possui os seus próprios dados e, apesar de o estudo ter realizado um modelo com variáveis independentes que considera generalistas, em alguns casos, certas organizações podem não recolher tais informações. Assim sendo, o modelo desenvolvido e os resultados apresentados podem não ser tão amplos quanto o desejável.

Ainda de referir que tarefas distintas, como seleção de dados, transformação de atributos e tratamento de valores omissos, podem variar muito de uma organização para outra, afetando conseqüentemente o desempenho dos resultados obtidos.

Além disso, alguns preditores que certos estudos consideravam cruciais não foram encontrados nos bancos de dados dos LD, nomeadamente o rendimento do doador.

Por fim, o número reduzido de campanhas realizadas e o facto de a maioria dos doadores contactados não responderem positivamente aos apelos da organização pode ter condicionado os resultados obtidos.

5.4 Pistas futuras

Apesar de ter sido exequível determinar qual o comportamento de um doador num determinado ano, assim como a realização de um *dashboard* que permite sistematizar a caracterização do comportamento dos doadores e das doações, é crucial indicar pistas futuras sobre o tema em estudo.

Conseqüentemente, é importante implementar um sistema que possibilite automaticamente o envio de campanhas aos doadores que o modelo considera que vão doar.

Outro trabalho futuro seria estender o estudo a outras ONG, de forma a comparar os resultados dos modelos obtidos e a concluir que os preditores que o estudo utilizou estão presentes nas suas bases de dados.

Por fim, seria importante possibilitar aos LD um modelo que previsse a quantidade monetária doada por parte de um doador.

Bibliografia

- Abrantes, C. (2020). Os modelos preditivos do sucesso de candidaturas a fundos europeus: o papel da manipulação de resultados. [Dissertação de mestrado]. Repositório do Iscte – Instituto Universitário de Lisboa. <http://hdl.handle.net/10071/21827>
- Abreu, M. (2012). Drivers of donations practices: altruism and religiosity revisited [Tese de doutoramento]. Repositório do Iscte – Instituto Universitário de Lisboa. <http://hdl.handle.net/10071/6203>
- Aldashev, G., & Verdier, T. (2009). When NGOs go global: Competition on international markets for development donations. *Journal of International Economics*, 79(2), 198–210. <https://doi.org/10.1016/j.jinteco.2009.07.007>
- Almendra, T. (2014). A prática de Fundraising nas IPSS: (estudo do caso “Unidos, Une-te a Nós!”). [Dissertação de mestrado, Universidade do Minho]. RepositoriUM – Repositório da Universidade do Minho. <https://hdl.handle.net/1822/30564>
- Anheier, H. K. (2014). *Nonprofit Organizations Theory, Management, Policy* (3rd ed.). Routledge
- Aspin, A. (2022). Slicers and Cross-Filtering. In *Pro Power BI Dashboard Creation* (pp. 351–392). Apress. https://doi.org/10.1007/978-1-4842-8227-4_14
- Berthold, M. R., Borgelt, C., Höppner, F., Klawonn, F., & Silipo, R. (2020). *Guide to Intelligent Data Science*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-45574-3>
- Birkholz, J. (2008). *Fundraising Analytics: Using Data to Guide Strategy Fundraising Analytics* (1st ed.). John Wiley & Sons.
- Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The Role of Big Data and Predictive Analytics in Retailing. *Journal of Retailing*, 93(1), 79–95. <https://doi.org/10.1016/j.jretai.2016.12.004>
- Cacciarelli, D., & Boresta, M. (2022). What drives a donor? A machine learning-based approach for predicting responses of nonprofit direct marketing campaigns. *Journal of Philanthropy and Marketing*, 27(2), e 1724. <https://doi.org/10.1002/nvsm.1724>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T. P., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*.
- Eiland, J., Hammonds, C. M., Ponos, S. M., Weigand, S. M., & Scherer, W. T. (2021). Developing Models to Predict Giving Behavior of Nonprofit Donors. *2021 IEEE Systems and Information Engineering Design Symposium, SIEDS 2021*. <https://doi.org/10.1109/SIEDS52267.2021.9483771>
- Erlandsson, A., Nilsson, A., Tinghög, G., Andersson, D., & Västfjäll, D. (2019). Donations to Outgroup Charities, but Not Ingroup Charities, Predict Helping Intentions Toward Street-Beggars in Sweden. *Nonprofit and Voluntary Sector Quarterly*, 48(4), 814–838. <https://doi.org/10.1177/0899764018819872>
- Farrokhvar, L., Ansari, A., & Kamali, B. (2018). Predictive models for charitable giving using machine learning techniques. *PLoS ONE*, 13(10) 1–14. <https://doi.org/10.1371/journal.pone.0203928>
- Galvão, M. C. B., & Ricarte, I. L. M. (2019). Revisão sistemática da literatura: conceituação, produção e publicação. *Logeion: Filosofia Da Informação*, 6(1), 57–73. <https://doi.org/10.21728/logeion.2019v6n1.p57-73>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gazley, B. (2022). The Systematic Literature Review: Advantages and Applications in Nonprofit Scholarship. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 33(6), 1256–1262. <https://doi.org/10.1007/s11266-021-00410-1>
- Han, J., Pei, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques*. (3rd ed.). Elsevier.

- Helmig, B., & Boenigk, S. (2012). *Nonprofit Management*. (2nd ed.). Vahlen.
- Hershberger, E., & Hair, M. (2022). The role of request precision and perceived efficacy in increasing charitable donations. *Journal of Philanthropy and Marketing*, 27(4). <https://doi.org/10.1002/nvsm.1746>
- Hindle, G., Kunc, M., Mortensen, M., Oztekin, A., & Vidgen, R. (2020). Business analytics: Defining the field and identifying a research agenda. *European Journal of Operational Research*, 281(3), 483–490. <https://doi.org/10.1016/j.ejor.2019.10.001>
- Hommerová, D., & Severová, L. (2019). Fundraising of Nonprofit Organizations: Specifics and New Possibilities. *Journal of Social Service Research*, 45(2), 181–192. <https://doi.org/10.1080/01488376.2018.1479678>
- Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29(2), 119–127. <https://doi.org/10.2307/2986296>
- Ki, E. J., & Cho, M. (2020). What are the predictors of nonprofit association members' supportive behaviors? *International Journal of Nonprofit and Voluntary Sector Marketing*, 1–11. <https://doi.org/10.1002/nvsm.1693>
- Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. In *Information and Software Technology*, 55(12), 2049–2075. Elsevier B.V. <https://doi.org/10.1016/j.infsof.2013.07.010>
- Kollmann, T., Stöckmann, C., & Kensbock, J. M. (2017). Fear of failure as a mediator of the relationship between obstacles and nascent entrepreneurial activity – An experimental approach. *Journal of Business Venturing*, 32(3), 280–301. <https://doi.org/10.1016/j.jbusvent.2017.02.002>
- Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*, 281(3), 628–641. <https://doi.org/10.1016/j.ejor.2019.09.018>
- Kristoffersen, L., & Singh, S. (2004). Successful Application of a Customer Relationship Management Program in a Nonprofit Organization. *Journal of Marketing Theory and Practice*, 12(2), 28–42. <https://doi.org/10.1080/10696679.2004.11658517>
- Larose, D., & Larose, C. (2015). *Data Mining and Predictive Analytics (Wiley Series on Methods and Applications in Data Mining)* (2nd ed.). John Wiley & Sons.
- Laureano, R. M. S., Vinhas da Silva, R., Pereira, J., Martins, S., & Rodrigues, D. (2015). Leigos para o Desenvolvimento: donations optimization through private segmentation. *Revista de Casos de Marketing Público y No Lucrativo*, 2, 45–56. Recuperado a partir de <https://casos-aimpn.org/index.php/casos/article/view/49>.
- Laurett, R., & Ferreira, J. J. (2018). Strategy in Nonprofit Organisations: A Systematic Literature Review and Agenda for Future Research. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 29(5), 881–897. <https://doi.org/10.1007/s11266-017-9933-2>
- Lee, Y. K., & Chang, C. T. (2007). Who gives what to charity? Characteristics affecting donation behavior. *Social Behavior and Personality*, 35(9), 1173–1180. <https://doi.org/10.2224/sbp.2007.35.9.1173>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ (Clinical Research Ed.)*, 339. <https://doi.org/10.1136/bmj.b2700>
- Lin, C.-L., & Fan, C.-L. (2019). Evaluation of CART, CHAID, and QUEST algorithms: a case study of construction defects in Taiwan. *Journal of Asian Architecture and Building Engineering*, 18(6), 539–553. <https://doi.org/10.1080/13467581.2019.1696203>
- Lindahl, W. E., & Conley, A. T. (2002). Literature Review: Philanthropic Fundraising. *Nonprofit Management and Leadership*, 13(1), 91–112. <https://doi.org/10.1002/nml.13108>
- Loh, W.-Y., & Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica Sinica*, 7(4), 815–840. <http://www.jstor.org/stable/24306157>

- Lu, J., Cairns, L., & Smith, L. (2020). Data science in the business environment: customer analytics case studies in SMEs. *Journal of Modelling in Management*, 16(2), 689–713. <https://doi.org/10.1108/JM2-11-2019-0274>
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (2nd ed.) Springer New York. <https://doi.org/10.1007/978-0-387-09823-4>
- Manesi, Z., Van Lange, P. A. M., Van Doesum, N. J., & Pollet, T. V. (2019). What are the most powerful predictors of charitable giving to victims of typhoon Haiyan: Prosocial traits, socio-demographic variables, or eye cues? *Personality and Individual Differences*, 146, 217–225. <https://doi.org/10.1016/j.paid.2018.03.024>
- Marr, B. (2016). *Key Business Analytics: The 60+ Tools Every Manager Needs To Turn Data Into Insights* (1st ed.). FT Publishing International.
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- Mendes, D. (2022). *A Importância das Imagens de Alojamentos Turísticos nas Plataformas de Reservas Online: Uma Análise com Dados Reais da Feels Like Home*. [Dissertação de mestrado]. Repositório do Iscte – Instituto Universitário de Lisboa. <http://hdl.handle.net/10071/26489>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Neto, R. (2023, 11 de janeiro). *INE confirma inflação média anual de 7,8% em 2022, um máximo de 30 anos*. ECO. <https://eco.sapo.pt/2023/01/11/ine-confirma-inflacao-media-anual-de-78-em-2022-um-maximo-de-30-anos/>
- Page M J, Moher D, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D *et al.* (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n160>
- Ferreira Peixoto, R. D. (2015) Pervasive data mining engine. [Dissertação de mestrado, Universidade do Minho]. RepositórioUM – Repositório da Universidade do Minho. <https://hdl.handle.net/1822/40323>
- Pereira, J. (2015). *Técnicas de Data Mining aplicadas ao fundraising: Estudo de caso numa Organização Não Governamental para o Desenvolvimento*. [Dissertação de mestrado]. Repositório do Iscte – Instituto Universitário de Lisboa. <http://hdl.handle.net/10071/10950>
- Rethlefsen, M.L., Kirtley, S., Waffenschmidt, S. *et al.* (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, 10(39), 1–19. <https://doi.org/10.1186/s13643-020-01542-z>
- Santos, M. R. C., Laureano, R. M. S., & Moro, S. (2020). Unveiling Research Trends for Organizational Reputation in the Nonprofit Sector. *Voluntas*, 31(1), 56–70. <https://doi.org/10.1007/s11266-018-00055-7>
- Schetgen, L., Bogaert, M., & Van den Poel, D. (2021). Predicting donation behavior: Acquisition modeling in the nonprofit sector using Facebook data. *Decision Support Systems*, 141. <https://doi.org/10.1016/j.dss.2020.113446>
- Sharma, V., Stranieri, A., Ugon, J., Vamplew, P., & Martin, L. (2017). An agile group aware process beyond CRISP-DM: A hospital data mining case study. *ACM International Conference Proceeding Series, Part F130280*, 109–113. <https://doi.org/10.1145/3093241.3093273>
- Sherwani, S. I., Bates, B. R., & Grijalva, M. J. (2021). Charitable Giving in the Context of Unfamiliar Organizations: The Effectiveness of Construal Level Theory in Predicting Donating Intentions and Antecedents. *Southern Communication Journal*, 86(5), 472–486. <https://doi.org/10.1080/1041794X.2021.1929434>

- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annu. Rev. Psychol*, 70, 747–770. <https://doi.org/10.1146/annurev-psych-010418>
- Smith, J. R., & McSweeney, A. (2007). Charitable giving: The effectiveness of a revised theory of planned behaviour model in predicting donating intentions and behaviour. *Journal of Community and Applied Social Psychology*, 17(5), 363–386. <https://doi.org/10.1002/casp.906>
- Srnka, K. J., Grohs, R., & Eckler, I. (2003). Increasing Fundraising Efficiency by Segmenting Donors. *Australasian Marketing Journal*, 11(1), 70–86. [https://doi.org/10.1016/S1441-3582\(03\)70119-0](https://doi.org/10.1016/S1441-3582(03)70119-0)
- Su, T., Li, H., & An, Y. (2021). A BIM and machine learning integration framework for automated property valuation. *Journal of Building Engineering*, 44, 1–25. <https://doi.org/10.1016/j.jobbe.2021.102636>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. In *British Journal of Management*. 14(3), 207–222). <https://doi.org/10.1111/1467-8551.00375>
- Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M., & Vugt, M. Van. (2007). From Games to Giving: Social Value Orientation Predicts Donations to Noble Causes. *Basic and Applied Social Psychology*, 29(4), 375–384. <https://doi.org/10.1080/01973530701665223>
- Verhaert, G. A., & Van den Poel, D. (2011). Empathy as added value in predicting donation behavior. *Journal of Business Research*, 64(12), 1288–1295. <https://doi.org/10.1016/j.jbusres.2010.12.024>
- Villar, B. G., Padilla, R. A., Simó, M. M., & Carpio, M. O. (2009). Factores de influencia en la captación de recursos privados en las ONGD españolas. *Estudios de Economía Aplicada*, 27(3), 1–25. ISSN: 1133-3197. Recuperado a partir de: <https://www.redalyc.org/articulo.oa?id=30117059016>
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84. <https://doi.org/10.1111/jbl.12010>
- Werker, E., & Ahmed, F. Z. (2008). What Do Nongovernmental Organizations Do? *Journal of Economic Perspectives*, 22(2), 73–92. <https://doi.org/10.1257/jep.22.2.73>
- Wilder, C. R., & Ozgur, C. O. (2015). Business Analytics Curriculum for Undergraduate Majors. *INFORMS Transactions on Education*, 15(2), 180–187. <https://doi.org/10.1287/ited.2014.0134>
- Zacarias, J. M. P. (2022). Modelos automatizados de avaliação de imóveis: Aplicação à Área Metropolitana de Lisboa [Dissertação de mestrado]. Repositório do Iscte- Instituto Universitário de Lisboa. <http://hdl.handle.net/10071/26808>

Anexos

Anexo A: Dados solicitados

Tabela- Dados solicitados aos LD

Variável	LD possuem?	Está presente na literatura?
Nome Campanha	Sim	-
Periodicidade da campanha	Sim	-
Data de início da campanha	Sim	Cacciarelli & Boresta, (2022)
Data fim da campanha	Sim	Cacciarelli & Boresta, (2022)
Causa da campanha	Sim	Cacciarelli & Boresta, (2022)
A campanha solicitava dinheiro?	Não	-
Qual o valor monetário solicitado?	Não	-
A campanha incluía mensagem?	Não	-
Mensagem	Não	-
Sexo do doador	Sim	Smith & McSweeney, (2007); Farrokhvar <i>et al.</i> , (2018); Erlandsson <i>et al.</i> , (2019); Manesi <i>et al.</i> , (2019); Ki & Cho, (2020); Cacciarelli & Boresta, (2022)
Idade do doador	Sim	Smith & McSweeney, (2007); Farrokhvar <i>et al.</i> , (2018); Erlandsson <i>et al.</i> , (2019); Manesi <i>et al.</i> , (2019); Ki & Cho, (2020); Cacciarelli & Boresta, (2022)
Localidade do doador	Sim	Cacciarelli & Boresta, (2022)
Rendimento familiar do doador	Não	Smith & McSweeney, (2007); Ki & Cho, (2020)
Profissão	Sim	-
Escolaridade	Sim	Smith & McSweeney, (2007); Farrokhvar <i>et al.</i> , (2018); Erlandsson <i>et al.</i> , (2019)
Caraterização do doador	Sim	Manesi <i>et al.</i> , (2019)
Estado do doador	Sim	Cacciarelli & Boresta, (2022)
Quantia doada	Sim	Não se aplica

A coluna “variáveis” corresponde às variáveis solicitadas aos LD, a coluna “LD possuem?” menciona se os LD possuem tais variáveis nas suas bases de dados e a coluna “está presente na literatura?” diz respeito aos autores que utilizaram as variáveis nos seus estudos.

Anexo B: Dicionário de dados

Tabela- Dicionário de dados provenientes do CRM dos LD

Conjunto de dados	Variáveis	Descrição	Tipo	Ação
1	Identificação	Id da doação	Q.D.	Preparação de dados
1	Tratamento	Forma de corresponder à pessoa (Sr./Sra.)	Q.N.	Preparação de dados
1	Cidade de correspondência	Localidade do doador (freguesia ou concelho ou cidade)	Q.N.	Preparação de dados
1	Estatuto Postal	Se o doador deve ser contactado	Q.N.	Preparação de dados
1	Classificação 1	Tipo de doador (empresa ou particular)	Q.N.	Preparação de dados (excluir)
1	Subclassificação 1	Tipo de classificação atribuída ao doador privado	Q.N.	Preparação de dados (excluir)
1	Estado Doador	Se o doador se encontra ativo ou inativo	Q.N.	Preparação de dados
1	Data do primeiro donativo	Data em que ocorreu o primeiro donativo	Data	Preparação de dados
1	Data do último donativo	Data em que ocorreu o último donativo	Data	Preparação de dados
1	Last Donation Date	Data em que ocorreu o último donativo	Data	Preparação de dados (excluir)
1	Valor do último donativo	Valor monetário do último donativo	Q.C.	Preparação de dados (excluir)
1	Data de fechamento	Data em que ocorreu o donativo	Data	Preparação de dados
1	Valor	Valor doado	Q.C.	Preparação de dados
1	Número do recibo	Número do recibo	Q.D.	Preparação de dados (excluir)
1	Origem da campanha principal	Nome Campanha	Q.N.	Preparação de dados
1	Forma de pagamento	Forma de pagamento da doação	Q.N.	Preparação de dados
1	Número total de donativos	Número totais de donativos realizados	Q.D.	Preparação de dados (excluir)
1	Profissão	Profissão	Q.N.	Preparação de dados (excluir)
1	Habilitações	Habilitações académicas	Q.N.	Preparação de dados
2	Identificação	Id da doação	Q.D.	Preparação de dados
2	Data de nascimento do doador	Data de nascimento do doador	Data	Preparação de dados

Q.N.-Qualitativa Nominal, Q.O- Qualitativa Ordinal Q.D.-Quantitativa Discreta, Q.C.- Quantitativa Contínua

Tabela- *Dicionário de dados após o tratamento de dados em SQL Studio*

Conjunto de dados	Variáveis	Descrição	Tipo
Doações	Identificação	Id da doação	Q.D.
Doações	Data do último donativo	Data em que ocorreu o donativo	Data
Doações	Valor	Valor doado	Q.C.
Doações	Nome da campanha	Nome Campanha	Q.N.
Doações	Forma de pagamento	Forma de pagamento da doação	Q.N.
Doadores	Sexo	Sexo do doador	Q.N.
Doadores	Distrito	Distrito de correspondência do doador	Q.N.
Doadores	Estatuto	Se o doador deve ser contactado	Q.N.
Doadores	Estado Doador	Se o doador se encontra ativo ou inativo	Q.N.
Doadores	Data do primeiro donativo	Data em que ocorreu o primeiro donativo	Data
Doadores	Data do último donativo	Data em que ocorreu o último donativo	Data
Doadores	Habilitações	Habilitações académicas	Q.N.
Doadores	Identificação	Id da doação	Q.D.
Doadores	Data de nascimento do doador	Data de nascimento do doador	Data

Q.N.-Qualitativa Nominal, Q.O- Qualitativa Ordinal Q.D.-Quantitativa Discreta, Q.C.- Quantitativa Contínua

Tabela- Dicionário de dados após a criação de novas variáveis em *SPPS modeler*

Conjunto de dados	Variáveis	Descrição	Tipo
Doações	Identificação	Id da doação	Q.D.
Doadores	Sexo	Sexo o doador	Q.N.
Doadores	Distrito	Distrito de correspondência do doador	Q.N.
Doadores	Estatuto	Se o doador deve ser contactado	Q.N.
Doadores	Estado Doador	Se o doador se encontra ativo ou inativo	Q.N.
Doadores	Data do primeiro donativo	Data em que ocorreu o primeiro donativo	Data
Doadores	Data do último donativo	Data em que ocorreu o último donativo	Data
Doações	Data do último donativo	Data em que ocorreu o donativo	Data
Doações	Valor	Valor doado	Q.C.
Doações	Nome da campanha	Nome Campanha	Q.N.
Doações	Forma de pagamento	Forma de pagamento da doação	Q.N.
Doadores	Habilitações	Habilitações académicas	Q.N.
Doadores	Identificação	Id da doação	Q.D.
Doadores	Data de nascimento do doador	Data de nascimento do doador	Data
Doadores	Idade no ano N-1	Idade do doador ao realizar a última doação	Q.C.
Doadores	Faixa etária n-1	Classificação da idade n-1 em: Adolescente (até aos 21 anos) Jovem Adulto (entre os 21 anos e os 34) Adulto (entre os 35 e os 64) Idoso (Superior a 65 anos)	Q.O.
Doadores	Senioridade n-1	Diferença de anos entre primeira doação do doador e a última (Excluindo o ano de 2022), ou seja, antiguidade.	Q.C.
Doadores	Senioridade classificação	Classificação da senioridade em: Baby (até 12 meses) Child (entre os 1 e 4 anos) Teen (entre os 5 e os 8 anos) Adult (entre os 9 e os 18 anos) Old (superior a 18 anos)	Q.O.
Doadores	Nº de anos que não doa	Número de anos em que um doador não realiza uma doação	Q.D.
Doadores	Deu em quantos anos?	Nº de anos que o doador realizou uma doação	Q.D.
Doadores	Nº de doações	Contagem de todas as doações realizadas (exceto no ano de 2022)	Q.D.
Doadores	Valor monetário acumulado	Valor monetário total doado (ano de 2022 excluído)	Q.C.
Doadores	Valor médio de doação	Valor Médio total doado (ano de 2022 excluído)	Q.C.
Doadores	Data em que ocorreu a última doação antes do ano n	Data em que ocorreu a última doação antes de 2022	Data
Doadores	Mês em que ocorreu a última doação	Mês em que ocorreu a última doação antes de 2022.	Q.D.
Doadores	Estado n-1	Estado do doador: Ativo (a última doação ocorreu no ano de 2021) Adormecido (a última doação ocorreu no ano de 2020) Inativo (a última doação ocorreu até ao ano de 2019)	Q.N.
Doadores	Recency Score	Atribuição de um score à recencia das doações 1- A última doação realizada aconteceu até 02-08-2017 2- A última doação realiza aconteceu entre 03-08-2017 e 31-12-2018 3- A última doação realiza aconteceu entre 01-01-2019 e 09-12-2020 4- A última doação realiza aconteceu entre 10-12-2020 e 01-12-2021 5- A última doação realiza aconteceu depois de 02-12-2021	Q.O.
Doadores	Frequency Score	Atribuição de um score à frequência das doações: 1- Realizou uma doação 2- Realizou entre 2 doações e 5 doações 3- Realizou entre 6 doações e 30 doações 4- Realizou entre 31 doações e 60 doações 5- Realizou mais que 61 doações	Q.D.
Doadores	Monetary Score	Atribuição de um score ao valor médio das doações: 1- O Valor médio de doação até 10 euros 2- O Valor médio de doação entre 11 e 30 euros 3- O Valor médio de doação entre 31 e 60 euros 4- O Valor médio de doação entre 61 e 150 euros 5- O Valor médio de doação maior que 151 euros	Q.D.
Doadores	RFM Score	Junção do Recency Score, Frequency Score, Monetary Score	Q.D.

Q.N.-Qualitativa Nominal, Q.O- Qualitativa Ordinal Q.D.-Quantitativa Discreta, Q.C.- Quantitativa Contínua

Tabela- Dicionário de dados após a criação de novas variáveis em Power BI

Conjunto de dados	Variáveis	Descrição	Tipo
Doações	Identificação	Id da doação	Q.D.
Doadores	Sexo	Sexo o doador	Q.N.
Doadores	Distrito	Distrito de correspondência do doador	Q.N.
Doadores	Estatuto	Se o doador deve ser contactado	Q.N.
Doadores	Estado Doador	Se o doador se encontra ativo ou inativo	Q.N.
Doadores	Data do primeiro donativo	Data em que ocorreu o primeiro donativo	Data
Doadores	Data do último donativo	Data em que ocorreu o último donativo	Data
Doações	Data do último donativo	Data em que ocorreu o donativo	Data
Doações	Valor	Valor doado	Q.C.
Doações	Nome da campanha	Nome Campanha	Q.N.
Doações	Forma de pagamento	Forma de pagamento da doação	Q.N.
Doadores	Habilitações	Habilitações académicas	Q.N.
Doadores	Identificação	Id da doação	Q.D.
Doadores	Data de nascimento do doador	Data de nascimento do doador	Data
Doadores	Idade no ano N-1	Idade do doador ao realizar a última doação	Q.C.
Doadores	Faixa etária n-1	Classificação da idade n-1 em: Adolescente (até aos 21 anos) Jovem Adulto (entre os 21 anos e os 34) Adulto (entre os 35 e os 64) Idoso (Superior a 65 anos)	Q.O.
Doadores	Senioridade n-1	Diferença de anos entre primeira doação do doador e a última (Excluindo o ano de 2022), ou seja, antiguidade.	Q.C.
Doadores	Senioridade classificação	Classificação da senioridade em: Baby (até 12 meses) Child (entre os 1 e 4 anos) Teen (entre os 5 e os 8 anos) Adult (entre os 9 e os 18 anos) Old (superior a 18 anos)	Q.O.
Doadores	Nº de anos que não doa	Número de anos em que um doador não realiza uma doação	Q.D.
Doadores	Deu em quantos anos?	Nº de anos que o doador realizou uma doação	Q.D.
Doadores	Nº de doações	Contagem de todas as doações realizadas (exceto no ano de 2022)	Q.D.
Doadores	Valor monetário acumulado	Valor monetário total doado (ano de 2022 excluído)	Q.C.
Doadores	Valor médio de doação	Valor Médio total doado (ano de 2022 excluído)	Q.C.
Doadores	Data em que ocorreu a última doação antes do ano n	Data em que ocorreu a última doação antes de 2022	Data
Doadores	Mês em que ocorreu a última doação	Mês em que ocorreu a última doação antes de 2022.	Q.D.
Doadores	Estado n-1	Estado do doador: Ativo (a última doação ocorreu no ano de 2021) Adormecido (a última doação ocorreu no ano de 2020) Inativo (a última doação ocorreu até ao ano de 2019)	Q.N.
Doadores	Recency Score	Atribuição de um score à recência das doações 1- A última doação realizada aconteceu até 02-08-2017 2- A última doação realiza aconteceu entre 03-08-2017 e 31-12-2018 3- A última doação realiza aconteceu entre 01-01-2019 e 09-12-2020 4- A última doação realiza aconteceu entre 10-12-2020 e 01-12-2021 5- A última doação realiza aconteceu depois de 02-12-2021	Q.D.
Doadores	Frequency Score	Atribuição de um score à frequência das doações: 1- Realizou uma doação 2- Realizou entre 2 doações e 5 doações 3- Realizou entre 6 doações e 30 doações 4- Realizou entre 31 doações e 60 doações 5- Realizou mais que 61 doações	Q.D.
Doadores	Monetary Score	Atribuição de um score ao valor médio das doações: 1- O Valor médio de doação até 10 euros 2- O Valor médio de doação entre 11 e 30 euros 3- O Valor médio de doação entre 31 e 60 euros 4- O Valor médio de doação entre 61 e 150 euros 5- O Valor médio de doação maior que 151 euros	Q.D.
Doadores	RFM Score	Junção do Recency Score, Frequency Score, Monetary Score	Q.D.
Doadores	Idade	Idade atual do doador	Q.C.
Doadores	Faixa etária	Classificação da idade em: Adolescente (até aos 21 anos) Jovem Adulto (entre os 21 anos e os 34) Adulto (entre os 35 e os 64) Idoso (Superior a 65 anos)	Q.O.

Conjunto de dados	Variáveis	Descrição	Tipo
Doadores	Senioridade atual	Diferença de anos entre primeira doação do doador e a última	Q.C.
Doadores	Senioridade atual- classificação	Classificação da senioridade em: Baby (até 12 meses) Child (entre os 1 e 4 anos) Teen (entre os 5 e os 8 anos) Adult (entre os 9 e os 18 anos) Old (superior a 18 anos)	Q.O.
Doadores	Nº de anos que não doa (atual)	Número de anos em que um doador não realiza uma doação	Q.D.
Doadores	Nº de doações (atual)	Contagem de todas as doações realizadas	Q.D.
Doadores	Valor monetário acumulado (atual)	Valor monetário total doado	Q.C.
Doadores	Valor médio de doação (atual)	Valor Médio total doado	Q.C.
Doadores	Mês em que ocorreu a última doação (atual)	Mês em que ocorreu a última doação	Q.D.
Doadores	Perfis de alta propensão a doar ou não	A partir dos perfis resultantes do modelo preditivo, indica se o doador é proporcional a doar ou não	Q.N.
Doadores	Intervalos de idades	Intervalo de 5 em 5 anos da idade no Ano N-1	Q.N.
Calendário	Date	Todos os dias no período de 1-01-2015 a 31-01-2022	Data
Calendário	Mês	Mês da date	Q.D.
Calendário	Dia	Dia da date	Q.D.

Q.N.-Qualitativa Nominal, Q.O- Qualitativa Ordinal Q.D.-Quantitativa Discreta, Q.C.- Quantitativa Contínua

Anexo C: Partes do código SQL

Exportar as tabelas para a máquina local (exemplo)

```
CREATE PROCEDURE [dbo].[SP_EXPORT_TABLE_TO_CSV] (
    @db_name      nvarchar(max)='LD',      -- nome da base de dados
    @table_name   nvarchar(max)='Doadores', -- Nome da tabela
    @dir_output   nvarchar(max)='D:\Outpust\', -- diretoria do output (terminar com
um backslash)
    @delim        nvarchar(10)='\t'      -- De acordo com BCP ('\t',';')
)
AS
BEGIN

DECLARE @cmd VARCHAR(8000);
DECLARE @sql nvarchar(max)

-- exportação do cabeçalho da tabela
IF (@delim='\t')
    set @sql='DECLARE @colnames VARCHAR(max); SELECT @colnames =
COALESCE(@colnames+char(9),''') + column_name FROM
'+@db_name+'.INFORMATION_SCHEMA.COLUMNS WHERE TABLE_NAME=''+@table_name+''; SELECT
@colnames;'

IF (@delim<>'\t')
    set @sql='DECLARE @colnames VARCHAR(max); SELECT @colnames =
COALESCE(@colnames+''+@delim+'',''') + column_name FROM
'+@db_name+'.INFORMATION_SCHEMA.COLUMNS WHERE TABLE_NAME=''+@table_name+''; SELECT
@colnames;'

select @sql
set @cmd='bcp "'+@sql+'" queryout '+@dir_output+@table_name+'.hdr -C ACP -c -t '+@delim+'
-T -d '+@db_name
select @cmd
exec xp_cmdshell @cmd

-- exportação dos dados da tabela
set @cmd='bcp dbo.'+@table_name+' out '+@dir_output+@table_name+'.dat -C ACP -c -t
'+@delim+' -T -d '+@db_name
select @cmd
exec xp_cmdshell @cmd

-- Juntar o cabeçalho da tabela com os dados que estão no csv
set @cmd='copy /B '+@dir_output+@table_name+'.hdr + '+@dir_output+@table_name+'.dat
'+@dir_output+@table_name+'.csv'
-- select @cmd
exec xp_cmdshell @cmd

-- eliminar o ficheiro do cabeçalho
set @cmd='del '+@dir_output+@table_name+'.hdr'
-- select @cmd
exec xp_cmdshell @cmd

set @cmd='del '+@dir_output+@table_name+'.dat'
-- select @cmd
exec xp_cmdshell @cmd

end
```

Anexo D: Preditores presentes na RSL

Tabela- Preditores presentes na RSL

Variáveis	Artigo
Senioridade	Cacciarelli & Boresta, (2022)
Recency Score	-*
Estado n-1	Cacciarelli & Boresta, (2022)
Frequency Score	-*
Idade no ano n-1	Smith & McSweeney, (2007); Farrokhvar <i>et al.</i> , (2018); Erlandsson <i>et al.</i> , (2019); Manesi <i>et al.</i> , (2019); Ki & Cho, (2020); Cacciarelli & Boresta, (2022)
Monetary Score	-*
Mês em que ocorreu a última doação	Cacciarelli & Boresta, (2022)
Nº de anos que não doa	-*
Deu em quantos anos?	-*
Sexo	Smith & McSweeney, (2007); Farrokhvar <i>et al.</i> , (2018); Erlandsson <i>et al.</i> , (2019); Manesi <i>et al.</i> , (2019); Ki & Cho, (2020); Cacciarelli & Boresta, (2022)
Senioridade Classificação	Cacciarelli & Boresta, (2022)
Faixa etária n-1	Smith & McSweeney, (2007); Farrokhvar <i>et al.</i> , (2018); Erlandsson <i>et al.</i> , (2019); Manesi <i>et al.</i> , (2019); Ki & Cho, (2020); Cacciarelli & Boresta, (2022)
Nº de doações	Erlandsson <i>et al.</i> , (2019)

*Preditor recomendado por um especialista

Anexo E: Output do CART-1

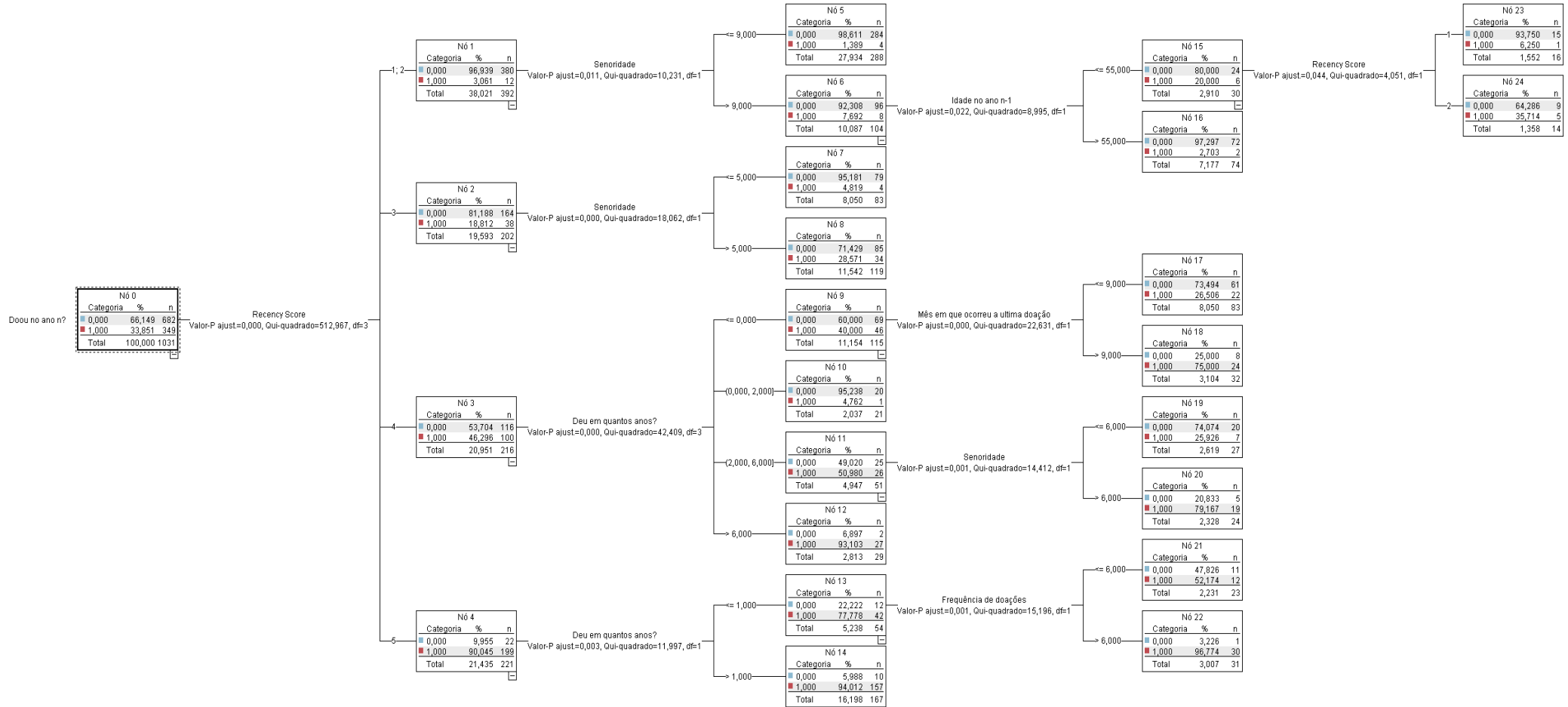


Figura- Ouput do CART-1
