



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Apostas desportivas *online* em Portugal: análise de variáveis com influência no lucro obtido através de *data mining*

Guilherme dos Santos Pereira

Mestrado em Informática e Gestão

Orientador:

Doutora Joana Martinho de Almeida Costa Pardal, Professora Auxiliar Convidada
Iscte - Instituto Universitário de Lisboa

Coorientador:

Doutor Nuno Miguel da Conceição António, Professor Auxiliar Convidado
Nova IMS, Universidade Nova de Lisboa

Novembro, 2020



TECNOLOGIAS
E ARQUITETURA

Apostas desportivas *online* em Portugal: análise de variáveis com influência no lucro obtido através de *data mining*

Guilherme dos Santos Pereira

Mestrado em Informática e Gestão

Orientador:

Doutora Joana Martinho de Almeida Costa Pardal, Professora Auxiliar Convidada
Iscte - Instituto Universitário de Lisboa

Coorientador:

Doutor Nuno Miguel da Conceição António, Professor Auxiliar Convidado
Nova IMS, Universidade Nova de Lisboa

Novembro, 2020

Agradecimentos

O meu sincero agradecimento a todos aqueles que, direta ou indiretamente, contribuíram para a conclusão desta etapa da minha vida.

Em primeiro lugar, um agradecimento especial à Professora Doutora Joana Costa e ao Professor Doutor Nuno António por toda a disponibilidade, ajuda e partilha de conhecimento com vista a alcançar os objetivos desta dissertação.

À minha família, em especial aos meus pais, por tudo. Desde os valores e princípios que sempre me transmitiram ao apoio, motivação e investimento na minha formação académica.

À Inês, minha namorada, por toda a cumplicidade ao longo de todo o meu percurso académico, por estar sempre presente em todas as etapas da minha vida, por todo o apoio incondicional e por todos os sorrisos que recompensaram todo o esforço.

Ao Bruno Martinho, um amigo que o futebol me deu, por me ter ajudado na escolha do tema para esta dissertação.

Ao apostador profissional, por se ter disponibilizado para ajudar na validação do questionário.

Por fim, a todos os amigos que me ajudaram a difundir os questionários e a todos os que contribuíram com o seu preenchimento.

Resumo

O mercado dos jogos *online*, nomeadamente o das apostas desportivas, tem tido um enorme crescimento nos últimos anos e estima-se que continue a aumentar. Este estudo procura aplicar técnicas de *data mining* para detetar padrões nos dados obtidos, desenvolvendo um modelo preditivo que permite antecipar o nível de lucro com base nos comportamentos dos apostadores. Estes comportamentos foram obtidos com recurso a um questionário online, desenvolvido e validado para esta investigação. O modelo com melhor desempenho, baseado numa árvore de decisão, apresentou uma exatidão (*accuracy*), no global das várias classes (níveis) de lucro, de 48%. Os resultados obtidos parecem indicar que a frequência com que se obtém lucro e a quantia apostada desempenham um papel importante na obtenção de lucro por parte dos apostadores. Esta dissertação permitiu colmatar algumas das lacunas identificadas, nomeadamente a análise de características relativas às apostas que podem ser controladas pelos apostadores, e também contribuir para que estes possam apostar de forma mais responsável e consciente, possibilitando um aumento do seu lucro.

Palavras-chave: Apostas desportivas *online*; Investigação por Questionário; *Data mining*; Modelo preditivo; Árvore de decisão; Lucro.

Abstract

The online games market, namely sports betting, has grown tremendously in recent years and is expected to continue to grow. This study aims to apply data mining techniques to identify patterns in the obtained data, developing a predictive model that anticipates the level of profit based on bettor's behaviors. These behaviors were obtained using an online survey, developed and validated for this investigation. The model with the best performance, based on a decision tree, presented an accuracy, regarding all profit levels, of 48%. The obtained results seem to reveal that the frequency with which profit is obtained and the amount bet play an important role in the profitability of bettors. This dissertation allowed to fill some of the identified gaps, namely the analysis of bet-related characteristics that can be controlled by bettors, and also to contribute to more responsible and conscious betting behavior, allowing an increase in their profit.

Keywords: Online sports betting; Survey research; Data mining; Predictive model; Decision Tree; Profit.

Índice

Agradecimentos.....	i
Resumo.....	iii
Abstract	v
Índice.....	vii
Índice de Figuras	ix
Índice de Tabelas.....	xi
Lista de Acrónimos	xiii
1. Introdução.....	1
1.1. Enquadramento geral.....	1
1.2. Motivações e Objetivos.....	2
1.3. Abordagem metodológica	3
1.4. Estrutura e organização do documento.....	3
2. Revisão da Literatura.....	5
2.1. Revisão Sistemática.....	5
2.1.1. Critérios de inclusão e exclusão	6
2.2. Enquadramento Concetual	6
2.2.1. Conceitos e definições relacionados com o jogo e apostas desportivas	6
2.2.1.1. Tipos de Jogos	6
2.2.1.2. Tipos de Apostas e Modalidades.....	7
2.2.1.3. Outros	8
2.2.2. Mercado das apostas desportivas no mundo.....	9
2.2.3. Mercado do jogo em Portugal	13
2.2.3.1. Dados relativos à atividade de jogo <i>online</i>	13
2.3. Investigação sobre apostas desportivas	16
2.3.1. Previsão de resultados de eventos desportivos.....	16
2.3.2. Perfil e comportamentos dos apostadores	19
2.4. Contextualização das variáveis utilizadas	20
2.5. <i>Research Gap</i>	21
3. Metodologia.....	23
3.1. Recolha dos dados.....	23
3.1.1. População e Amostra.....	23
3.1.2. Instrumento utilizado.....	24
3.1.2.1. Construção do questionário	24
3.1.2.2. Testes piloto	27
3.1.2.3. Versão final do questionário.....	27

3.1.2.4.	Avaliação da validade e fiabilidade da versão final do questionário.....	28
3.1.3.	Procedimento de recolha dos dados	34
3.2.	Desenvolvimento dos modelos preditivos.....	35
3.2.1.	Compreensão dos dados	35
3.2.2.	Preparação dos dados	38
3.2.2.1.	Análise descritiva	39
3.2.2.2.	Verificação das correlações entre as variáveis	43
3.2.3.	Modelação	45
4.	Análise e discussão dos resultados	49
4.1.	Avaliação.....	49
4.1.1.	Matriz de confusão e <i>accuracy</i> do modelo.....	50
4.1.2.	<i>Precision</i> e <i>recall</i> de cada uma das classes	51
4.1.3.	Árvore de decisão.....	52
5.	Conclusões e recomendações	55
5.1.	Principais conclusões	55
5.2.	Limitações do estudo.....	57
5.3.	Propostas de investigação futura	58
Referências	61
Apêndices	65
Apêndice A.....		65
Apêndice B.....		73
Apêndice C		75
Apêndice D.....		85
Apêndice E.....		87
Apêndice F		99
Apêndice G.....		101
Apêndice H.....		103
Apêndice I.....		109
Apêndice J.....		113
Apêndice K.....		117
Apêndice L		121
Apêndice M.....		125

Índice de Figuras

Figura 2.1: Tipos de Jogo existentes	7
Figura 2.2: Volume de negócios e GGR global das apostas desportivas em 2017 (US\$bn).....	9
Figura 2.3: Top 10 dos países com mais receitas brutas <i>online</i> vs. <i>offline</i> em 2017 (US\$bn)	11
Figura 2.4: Top 10 dos países com mais receitas <i>online onshore</i> vs. <i>offshore</i> em 2017 (US\$bn)	11
Figura 2.5: Receita bruta e IEJO dos jogos e apostas <i>online</i> , em 2018.....	14
Figura 2.6: Receita bruta das apostas desportivas à cota <i>online</i> , em 2018.....	14
Figura 2.7: Apostas desportivas à cota por modalidade, em 2018	14
Figura 2.8: Distribuição dos jogadores por estrutura etária, em 2018.....	15
Figura 3.1: Escalão etário dos inquiridos (Frequências relativas).....	39
Figura 3.2: Habilitações literárias dos inquiridos (Frequências relativas)	40
Figura 3.3: Regiões de residência dos inquiridos (Frequências relativas)	41
Figura 3.4: Dinheiro gasto pelos inquiridos (Frequências relativas).....	43
Figura 3.5: Lucro obtido pelos inquiridos (Frequências relativas).....	43
Figura 4.1: Árvore de decisão	53

Índice de Tabelas

Tabela 2.1: Apostas desportivas globais em 2017 (US\$bn)	10
Tabela 2.2: GGR das apostas desportivas globais (2012-2017) (US\$bn)	10
Tabela 2.3: GGR esperado das apostas desportivas globais (2017-2022e) (US\$bn)	10
Tabela 2.4: GGR observado (12-17) e esperado (17-22e) das apostas desportivas <i>online</i>	12
Tabela 2.5: GGR observado (12-17) e esperado (17-22e) das apostas desportivas <i>onshore</i>	12
Tabela 2.6: GGR observado (12-17) e esperado (17-22e) das apostas desportivas por telemóvel	12
Tabela 2.7: GGR observado (12-17) e esperado (17-22e) das apostas desportivas ao vivo	13
Tabela 3.1: Variáveis em estudo por questão	25
Tabela 3.2: Valores de referência do alfa de <i>Cronbach</i>	31
Tabela 3.3: Alfa de <i>Cronbach</i> de cada dimensão	31
Tabela 3.4: Alfa de <i>Cronbach</i> se o item for excluído (dimensão Estratégia Técnica)	33
Tabela 3.5: Alfa de <i>Cronbach</i> se o item for excluído (dimensão Estratégia Desportiva)	34
Tabela 3.6: Associação entre as questões do questionário e o nome das variáveis	36
Tabela 3.7: Valores de referência do Coeficiente de correlação	44
Tabela 3.8: Número de observações e frequências relativas pré e pós transformação	46
Tabela 4.1: Matriz de confusão	50
Tabela 4.2: Número de observações e frequências relativas por classe	51
Tabela 4.3: <i>Precision</i> e <i>recall</i> das classes da variável Lucro no <i>dataset</i> de teste	52

Lista de Acrónimos

CAGR - Compound Annual Growth Rate

CRISP-DM - Cross Industry Standard Process for Data Mining

DL - Decreto-Lei

EGBA - European Gaming & Betting Association

GGR - Gross Gaming Revenue

H2GC - H2 Gambling Capital

SRIJ - Serviço de Regulação e Inspeção de Jogos

1. Introdução

O objetivo deste capítulo é apresentar a temática desta investigação, através de um enquadramento sumário, seguido dos objetivos a alcançar e, por fim, apresentar a estrutura do documento.

1.1. Enquadramento geral

O jogo é um conceito que gerou desde sempre interesse nos seres humanos, sendo atualmente uma realidade cada vez mais presente na sociedade (European Commission, 2012).

Apesar de a indústria do jogo já ser antiga (espaços físicos como casinos já existem há muitos anos), com o aumento do uso da Internet, o jogo tem-se tornado cada vez mais uma atividade *online* (European Gaming & Betting Association [EGBA], 2018). A modernização da sociedade e a inovação tecnológica dos últimos anos permite agora que os jogadores europeus possam jogar e apostar *online* anonimamente e em privado, sem qualquer restrição no horário e localização. Graças a esta mudança tecnológica, o mercado dos jogos *online* tem aumentado nos últimos anos, nomeadamente o das apostas desportivas (EGBA, 2018).

Estudos indicam que o lucro bruto do mercado global do jogo *online* na Europa passou dos 10,48 biliões de euros em 2010 para 19,6 biliões de euros em 2017 (EGBA, 2018). Por sua vez, a atividade de jogos e apostas *online* em Portugal gerou, em 2018, cerca de 152,1 milhões de euros de receita bruta, valor superior em 29,5 milhões de euros face ao ano anterior, crescendo assim 24% (Serviço de Regulação e Inspeção de Jogos [SRIJ], 2019). No que diz respeito às apostas desportivas à cota *online*, em Portugal, a receita bruta das entidades exploradoras registou, em 2018, o valor de 78,9 milhões de euros (51,8% do total), registando-se um aumento de 10,8 milhões de euros face ao ano anterior. Por sua vez, no mesmo ano, apuraram-se 380,1 mil novos registos de jogadores (SRIJ, 2019). É, portanto, possível compreender que tanto o mercado do jogo *online* global como o mercado das apostas desportivas à cota *online* têm crescido bastante nos últimos anos, tanto em Portugal como no resto da Europa, atraindo cada vez mais adeptos.

Assim, as apostas desportivas começaram a ser vistas como uma forma alternativa de investimento, em que a tomada de decisões é baseada em números e probabilidades e não na intuição. Paulo Rebelo, apostador profissional há mais de 15 anos e presidente da Associação Nacional de Apostadores *Online* (ANAon), afirma que a análise de dados e as estatísticas são ferramentas úteis para prever os resultados dos eventos desportivos, e que é possível obter retornos positivos nas apostas fazendo apenas uma análise quantitativa das duas equipas envolvidas num determinado jogo. Ainda assim, refere também que uma análise qualitativa (interpretação subjetiva das notícias e fatores que não são fáceis de quantificar numericamente, como lesões ou expulsões) acrescenta muito mais valor e, em caso de conflito, deve sobrepor-se à análise quantitativa (Rebelo, 2012). No entanto, muitos são os apostadores que apostam baseando-se na sua intuição ou na interpretação errada das estatísticas disponíveis. Para além disso, dada a elevada acessibilidade das apostas desportivas *online* e a sua associação ao

entretenimento, os apostadores tornam-se mais vulneráveis, nomeadamente os mais novos, realizando apostas mais frequentes e mais impulsivas.

1.2. Motivações e Objetivos

Têm sido realizados vários estudos que pretendem caracterizar os apostadores e os seus comportamentos para que estes não se tornem jogadores patológicos (por exemplo, Labrador e Vallejo-Achón, 2019). Têm sido também desenvolvidos vários modelos referentes à previsão de resultados de jogos com o objetivo de auxiliar na obtenção de lucro por parte dos apostadores (por exemplo, Stübinger e Knoll, 2018). No entanto, não foi encontrado nenhum estudo que procurasse associar os fatores relacionados com o conjunto de decisões que um indivíduo toma relativamente a uma aposta desportiva *online* com o lucro obtido na mesma. Assim, esta dissertação incide nos fatores supracitados.

Para atingir o objetivo proposto, os dados necessários são recolhidos através de um questionário online, que é previamente validado, e são utilizadas técnicas de *data mining*, nomeadamente árvores de decisão, de forma a desenvolver um modelo preditivo que permite antecipar o nível de lucro com base em determinados comportamentos dos apostadores. A utilização de árvores de decisão permitirá detetar os padrões por níveis de lucro e, assim, identificar os comportamentos que desempenham um papel importante na obtenção de lucro por parte dos apostadores.

Nesta dissertação é utilizada a metodologia *CRoss-Industry Standard Process for Data Mining* (CRISP-DM). Esta metodologia define um projeto como um processo cíclico e aplica uma sequência de seis etapas principais (Chapman et al., 2000). Este estudo trata-se de um problema de Aprendizagem Supervisionada do tipo Classificação (multi-classificação), no qual se pretende construir um modelo preditivo que, a partir de um conjunto de exemplos (instâncias), classifica novos registos numa das classes predefinidas da variável referente ao lucro (estas classes são apresentadas na secção 3.2.1.). Considera-se como sucesso se o modelo criado prever com alguma precisão o lucro obtido pelos apostadores com base nos comportamentos dos mesmos e se permitir identificar as variáveis mais importantes para a obtenção de lucro.

Pretende-se com esta dissertação colmatar algumas das lacunas identificadas na literatura, nomeadamente a análise de características relativas às apostas que podem ser controladas pelos apostadores nos estudos realizados, e também dar a conhecer o mercado das apostas desportivas e a forma como algumas variáveis (como a frequência das apostas, o tipo de aposta, a modalidade desportiva em que mais apostam, entre outras) se relacionam com a obtenção de lucro. Deste modo, os apostadores poderão apostar de forma mais responsável e consciente, possibilitando um aumento do seu lucro.

Para responder ao objetivo proposto, esta dissertação propõe-se a:

1. Analisar o mercado das apostas desportivas online em Portugal;
2. Recolher informação sobre o perfil e comportamentos dos apostadores portugueses no que diz respeito à forma de apostar;

3. Caracterizar demograficamente os apostadores;
4. Construir um modelo preditivo que possibilite a identificação dos comportamentos dos apostadores que conseguem obter retornos positivos de forma consistente a longo prazo (os vários modelos desenvolvidos são comparados e é escolhido o modelo com melhor performance preditiva).

1.3. Abordagem metodológica

Como foi referido na secção anterior, neste estudo foi adotada a metodologia CRISP-DM, que consiste numa sequência iterativa de seis fases (compreensão do negócio/problema, compreensão dos dados, preparação dos dados, modelação, avaliação, implantação) incluídas num ciclo. Cada uma das fases é composta por várias tarefas, sendo algumas delas dependentes umas das outras. As seis fases tendem a ser seguidas, no entanto, esta metodologia é flexível, podendo as tarefas ser executadas numa ordem diferente sendo, muitas vezes, necessário retornar a fases anteriores e repetir determinadas ações. (Chapman et al., 2000). O CRISP-DM fornece assim uma visão geral do ciclo de vida de um projeto de *data mining*.

Para manter a estrutura normal de uma dissertação, as iterações do CRISP-DM não são descritas por ordem, mas sim englobadas na estrutura típica de uma dissertação. Sendo assim, os capítulos da Introdução e Revisão de literatura correspondem ao que em CRISP-DM se designa pela fase de Compreensão do negócio/problema e o capítulo da Metodologia inclui as fases de Compreensão dos dados, Preparação dos dados e Modelação. O capítulo da análise e discussão dos resultados corresponde à fase de Avaliação do CRISP-DM. Note-se que a sexta e última fase, denominada Implantação, é a fase que indica o fim do projeto, onde se coloca o modelo em produção para que possa ser utilizado. Tal como é comum em projetos de caracterização em *data mining*, uma vez que o modelo não é um modelo para colocar em produção, apresenta-se apenas uma discussão da implementação do modelo no capítulo 5 (Han et al., 2011).

1.4. Estrutura e organização do documento

O presente estudo está organizado em cinco capítulos que pretendem refletir as diferentes fases até à sua conclusão.

O primeiro capítulo introduz o tema da investigação e objetivos da mesma, bem como uma breve descrição da estrutura deste documento. Pretende-se, assim, apresentar uma perspetiva global do presente estudo.

O segundo capítulo reflete o enquadramento teórico, designado por Revisão da literatura, onde é apresentado um enquadramento concetual e alguns estudos tidos em consideração nesta investigação.

O terceiro capítulo é dedicado à metodologia utilizada no processo de recolha e tratamento de dados bem como as técnicas utilizadas para desenvolver os modelos.

O quarto capítulo apresenta a análise dos resultados obtidos, de acordo com a metodologia definida no capítulo anterior.

No quinto e último capítulo apresentam-se as conclusões deste estudo bem como as limitações, recomendações e trabalhos futuros.

No final do documento encontram-se as referências e os apêndices.

2. Revisão da Literatura

Este capítulo apresenta o enquadramento teórico desta investigação. Em primeiro lugar, são mencionados alguns conceitos relacionados com as apostas desportivas, de seguida são apresentados alguns dados referentes ao mercado das apostas desportivas no mundo e em Portugal e, por último, são referidos alguns estudos tidos em consideração nesta investigação.

2.1. Revisão Sistemática

Para a elaboração da Revisão da Literatura foram consultados, desde outubro de 2019 a janeiro de 2020, diversos repositórios de pesquisa científica e educacional e alguns *sites*. Entre os utilizados destacam-se:

- ✓ Elsevier ScienceDirect;
- ✓ Springer Link;
- ✓ IEEE Xplore Digital Library;
- ✓ Wiley Online Library;
- ✓ Google Scholar;
- ✓ Repositório do ISCTE-Instituto Universitário de Lisboa e de outras universidades;
- ✓ Diário da República Eletrónico;
- ✓ SRIJ - Regulação e Inspeção de Jogos;
- ✓ European Gaming & Betting Association.

A pesquisa efetuada teve como base as seguintes palavras-chave/expressões:

- ✓ *online "sports (betting OR gambling)" AND "data mining" patterns*
- ✓ *online "sports (betting OR gambling)" AND "data mining" patterns AND "Portugal"*
- ✓ *online "sports (betting OR gambling)" AND "Portugal"*
- ✓ *online "sports (betting OR gambling)" success*
- ✓ *online betting success*
- ✓ *online betting profile*
- ✓ *online sports betting profile*
- ✓ *online betting patterns*
- ✓ *online betting game-play patterns*
- ✓ *game-play patterns of Internet gamblers*
- ✓ *online betting gamblers profile*
- ✓ *online sports betting gamblers profile*

Começou-se por utilizar expressões mais completas, mas rapidamente se percebeu que os resultados obtidos não eram os esperados, uma vez que surgiram muitos artigos em que o tema estava relacionado

com as apostas desportivas ou com o jogo numa perspetiva geral, mas poucos destes apresentavam conteúdo relevante. Assim, optou-se por utilizar *strings* mais gerais, obtendo assim artigos mais relevantes, contrariamente ao esperado.

Para seleccionar os artigos a referir neste capítulo foram considerados fatores como:

- ✓ O ano em que o artigo foi publicado;
- ✓ A reputação dos jornais, revistas, conferências, etc. onde os artigos foram publicados (consultou-se o *Scimago Journal & Country Rank* e verificou-se qual o *SCImago Journal Rank Indicator* de cada jornal, revista, etc.);
- ✓ A reputação das bibliotecas digitais onde os artigos foram colocados.

2.1.1. Critérios de inclusão e exclusão

Foram apenas incluídos estudos escritos em português e em inglês, efetuados após o ano de 2000 (exceto um caso que serviu como comparação com artigos mais recentes), uma vez que a situação atual do mercado de apostas desportivas se apresenta bastante diferente da dos estudos anteriores ao ano 2000.

Entre os estudos que entraram na revisão sistemática após a aplicação dos critérios de inclusão, foram excluídos os que apresentavam pelo menos uma das seguintes características:

- ✓ Quando o tema do estudo não é o jogo ou as apostas desportivas;
- ✓ Quando os estudos referentes à previsão dos resultados de eventos desportivos não testam os modelos desenvolvidos (pelo próprio estudo) e não apresentam retornos positivos, uma vez que o foco desta dissertação é, através do desenvolvimento de um modelo preditivo, perceber quais os fatores mais preponderantes para a obtenção de lucro.

2.2. Enquadramento Concetual

2.2.1. Conceitos e definições relacionados com o jogo e apostas desportivas

Esta secção é constituída por um conjunto de conceitos e definições relevantes para esta investigação e que serão mencionados ao longo da dissertação.

O Jogo pode ser definido como qualquer comportamento que envolva o risco de perder dinheiro ou bens valiosos em relação ao resultado de um jogo, concurso ou outro evento, no qual este resultado seja inteiro ou parcialmente determinado pela sorte (Whelan, Meyers & Steenbergh, 2007).

2.2.1.1. Tipos de Jogos

Segundo o Decreto-Lei (DL) n.º 66/2015 e a Portaria n.º 173/2015, existem os tipos de jogo apresentados na Figura 2.1.

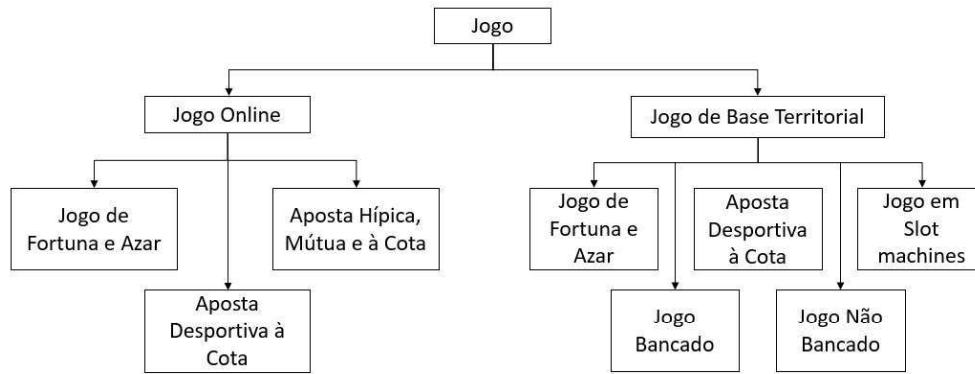


Figura 2.1: Tipos de Jogo existentes

De acordo com o artigo 4.º do capítulo I do DL n.º 66/2015, uma Aposta Desportiva é aquela através da qual se coloca uma quantia em dinheiro associada a um prognóstico sobre um determinado tipo de resultado de uma competição ou prova desportiva previamente identificada, cujo desfecho é incerto e não dependente da vontade dos participantes. Uma Aposta à Cota é aquela em que o apostador joga contra a entidade exploradora, com base num valor igual ou superior a 1,00 (cota) associado a cada um dos prognósticos possíveis para cada aposta em função da probabilidade de ocorrência de um determinado tipo de resultado (Ministério da Economia, 2015).

Este estudo cinge-se às apostas desportivas à cota *online*, que se distinguem das apostas desportivas à cota de base territorial, na medida em que estas últimas são efetuadas diretamente num local físico (*land-based*) (Ministério da Economia, 2015).

2.2.1.2. Tipos de Apostas e Modalidades

De acordo com o artigo 2.º do Anexo, alínea n), da Portaria n.º 173/2015, o Tipo de Aposta é “a pergunta colocada ao apostador, sobre factos que ocorrem no decurso de determinado período de jogo de um ou vários eventos desportivos”. Segundo o artigo 6.º do Anexo, alínea 1, da Portaria n.º 173/2015, existem inúmeros tipos de apostas, nomeadamente aposta ‘1 X 2’, ‘1 X 2 Intervalo’, ‘Intervalo/Final’, ‘Dupla Possibilidade’ e ‘Quem marca o primeiro golo’ (Ministério da Solidariedade, Emprego e Segurança Social, 2015).

De acordo com os artigos 7.º e 8.º do Anexo, da Portaria n.º 173/2015, a participação no jogo obedece a uma das seguintes modalidades¹ (Ministério da Solidariedade, Emprego e Segurança Social, 2015):

- ✓ Simples - Apostas em que o apostador seleciona um só prognóstico, constituindo cada prognóstico uma combinação autónoma;

¹ No Placard, os nomes das modalidades são os apresentados acima, no entanto, nas casas de apostas *online* (incluindo o placard.pt) a nomenclatura utilizada é a contrária, ou seja, as apostas múltiplas chamam-se combinadas e vice-versa.

- ✓ Combinada - Apostas em que o apostador seleciona entre dois a oito prognósticos no mesmo bilhete de aposta, constituindo o conjunto dos prognósticos selecionados uma única combinação e uma única aposta;
- ✓ Múltipla - Apostas em que o apostador seleciona entre três a cinco prognósticos no mesmo bilhete de aposta e seleciona entre um mínimo de três e um máximo de dez possíveis combinações autónomas, consoante o número de prognósticos efetuados.

2.2.1.3. *Outros*

A H2 *Gambling Capital* (H2GC, 2018), consultora internacional líder do mercado da indústria de jogos que fornece estudos de mercado relativos à indústria do jogo (nomeadamente apostas desportivas, bingo, casino, jogos *online*, póquer e lotarias), utiliza a métrica ‘*gross win*’ (ou ‘*gross gaming revenue*’) para avaliar o setor dos jogos, por considerar que esta fornece uma medida consistente para comparação em todo o setor. Como o nome indica, esta métrica corresponde ao ganho bruto ou receita bruta do jogo, ou seja, à quantia que as casas de apostas (também chamadas de *bookmakers*) ganham antes de as suas despesas operacionais serem calculadas. Esta métrica será utilizada na secção seguinte.

A H2GC categoriza a atividade do setor de apostas desportivas em três mercados (H2 *Gambling Capital* [H2GC], 2018):

- ✓ Mercado branco - Apostas em que o operador está licenciado *onshore* na mesma jurisdição em que o jogador está localizado;
- ✓ Mercado cinza - Apostas onde o operador está licenciado *offshore* numa jurisdição diferente;
- ✓ Mercado negro - Apostas em que o operador é completamente desregulado ou ilegal.

Estes mercados podem ser agrupados em:

- ✓ Apostas regulamentadas - Inclui os mercados branco e cinza;
- ✓ Apostas não regulamentadas - Inclui o mercado negro.

As apostas desportivas podem ainda ser classificadas como (H2GC, 2018):

- ✓ Apostas desportivas *online onshore* - Quando um operador de apostas desportivas *online* é licenciado na jurisdição em que a aposta é realizada;
- ✓ Apostas desportivas *online offshore* - Quando um operador de apostas desportivas *online* é licenciado numa determinada jurisdição, mas não especificamente licenciado na jurisdição em que a aposta é realizada.

Teixeira (2017), *trader* na bolsa desportiva Betfair desde 2009, afirma que existem três sistemas de apostas desportivas, sendo eles o sistema de *Stake Fixa*, o sistema de *Stake* por Níveis e o sistema *Martingale*. O sistema de *Stake Fixa* consiste em apostar sempre o mesmo valor em qualquer aposta efetuada e este deve ser uma percentagem do montante total da banca de cada apostador. O sistema de

Stake por Níveis consiste em estabelecer valores diferentes (em percentagem), adequados a situações diferentes. Desta forma, podem ser criados os seguintes níveis: nível 2,5%, adequado em apostas com risco elevado; nível 5%, o valor da *stake* fixa; e nível 7,5%, adequado em apostas cujo grau de confiança seja elevado. Por fim, o sistema *Martingale*, que consiste em duplicar o valor da aposta sempre que se perde. Caso não seja utilizado qualquer tipo de sistema de apostas desportivas, as apostas são chamadas de apostas livres.

2.2.2. Mercado das apostas desportivas no mundo

Segundo a H2GC (2018), e de acordo com a Figura 2.2, o mercado global de apostas desportivas (incluindo as corridas de cavalos) apresentou, em 2017, um *Gross Gaming Revenue* (GGR) de 67,5 biliões de dólares (US\$bn) e um volume de negócios de 593 biliões de dólares. Se se excluir as corridas de cavalos, obtém-se um valor de 39,3 biliões de dólares para o GGR e 370 biliões de dólares para o volume de negócios. É também possível observar que em qualquer um dos métodos, *online* ou base territorial, a receita bruta e o volume de negócios das apostas desportivas são superiores aos das corridas de cavalos.

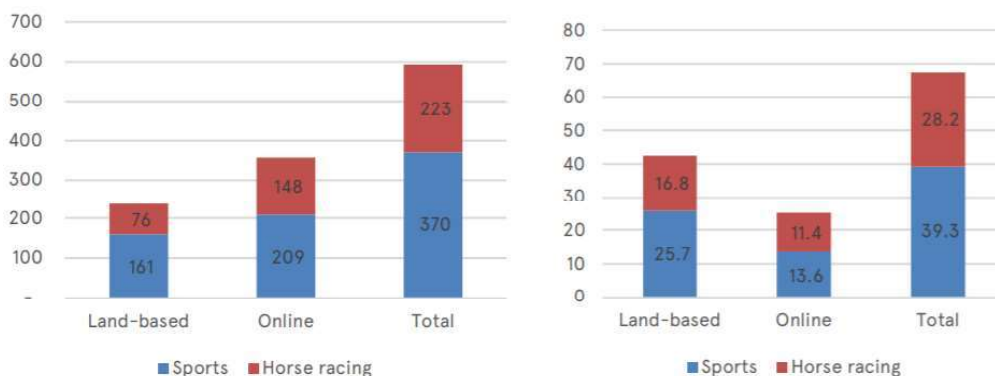


Figura 2.2: Volume de negócios e GGR global das apostas desportivas em 2017 (US\$bn)
Fonte: H2GC, 2018

De acordo com a H2GC (2018), e em termos do GGR, o mercado não regulamentado apresentou em 2017 um valor ligeiramente superior ao do mercado regulamentado. No entanto, o mercado não regulamentado possuiu uma margem GGR significativamente menor, apresentando, em termos de volume de negócios, um valor 2,5 vezes superior ao do mercado regulamentado. A margem é calculada através da divisão do GGR pelo volume de negócios. Desta forma, ao incluir as corridas de cavalos e o mercado não regulamentado, é possível calcular os valores totais do volume de negócios e da receita bruta (*online* e *offline*) do mercado das apostas desportivas (2107 e 143,2 biliões de dólares respetivamente). Estes dados podem ser observados na Tabela 2.1.

Tabela 2.1: Apostas desportivas globais em 2017 (US\$bn)

2017 (US\$bn)	Mercado Regulamentado	Mercado não Regulamentado	Mercado Total
Volume de negócios	593	1513	2107
GGR	67.5	75.7	143.2
Margem GGR (%)	11.4%	5.0%	6.8%

Adaptado de: H2GC, 2018

A partir deste momento, as informações e valores apresentados excluem o mercado não regulamentado.

Através da análise da Tabela 2.2, é possível verificar que o mercado total de apostas desportivas (excluindo as corridas de cavalos) cresceu a uma taxa de crescimento anual composta (*Compound Annual Growth Rate* [CAGR]) de 5 anos (2012-2017) de 10%. No entanto, o crescimento das apostas desportivas *online* de 13% ultrapassou o crescimento das apostas de base territorial de 8% (H2GC, 2018). Esta taxa é calculada com base nos valores do GGR dos 5 anos mencionados.

Tabela 2.2: GGR das apostas desportivas globais (2012-2017) (US\$bn)

GGR (US\$bn)	2012	2013	2014	2015	2016	2017	2012-17 CAGR
Base territorial	17.4	18.4	21.6	22.0	24.2	25.7	8%
<i>Online</i>	7.3	8.1	9.4	10.8	12.1	13.6	13%
Total	24.7	26.5	31.0	32.8	36.3	39.3	10%

Adaptado de: H2GC, 2018

De acordo com a H2GC (2018), e como é possível observar na Tabela 2.3, no futuro, prevê-se que o mercado das apostas desportivas continue a crescer, no entanto, a uma taxa menor. Excluindo as corridas de cavalos, prevê-se um CAGR de 5 anos (2017-2022e) de 8%, com as apostas desportivas *online* a continuarem a crescer mais do que as de base territorial (CAGR de 10%).

Tabela 2.3: GGR esperado das apostas desportivas globais (2017-2022e) (US\$bn)

GGR (US\$bn)	2017	2018e	2019e	2020e	2021e	2022e	2017-22e CAGR
Base territorial	25.7	28.5	29.5	32.3	34.3	36.1	7%
<i>Online</i>	13.6	15.5	16.8	18.4	19.6	21.5	10%
Total	39.3	44.0	46.3	50.7	53.9	57.6	8%

Adaptado de: H2GC, 2018

De seguida, é possível observar o top 10 dos países com mais receitas brutas provenientes das apostas desportivas (incluindo as corridas de cavalos) em 2017, divididas entre apostas *online* e *offline*. Para realizar esta análise, a H2GC optou por agrupar os países da região nórdica. É possível constatar

que os países que apresentam uma maior receita bruta são o Japão e a China, em larga escala e que na maioria dos países esta receita é superior nos jogos *offline* (H2GC, 2018). Estes dados podem ser observados na Figura 2.3.

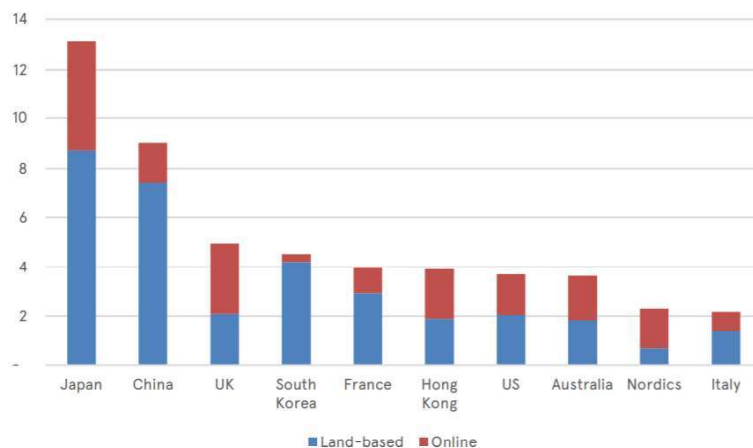


Figura 2.3: Top 10 dos países com mais receitas brutas online vs. offline em 2017 (US\$bn)
Fonte: H2GC, 2018

Também é possível observar o top 10 dos países com mais receitas provenientes das apostas desportivas (incluindo as corridas de cavalos) em 2017, divididas entre apostas *onshore* e *offshore*. Verifica-se que os países que apresentam uma maior receita bruta são o Japão e o Reino Unido e que na maioria dos países esta receita é superior nos jogos *online onshore* (H2GC, 2018). Estes dados podem ser observados na Figura 2.4.

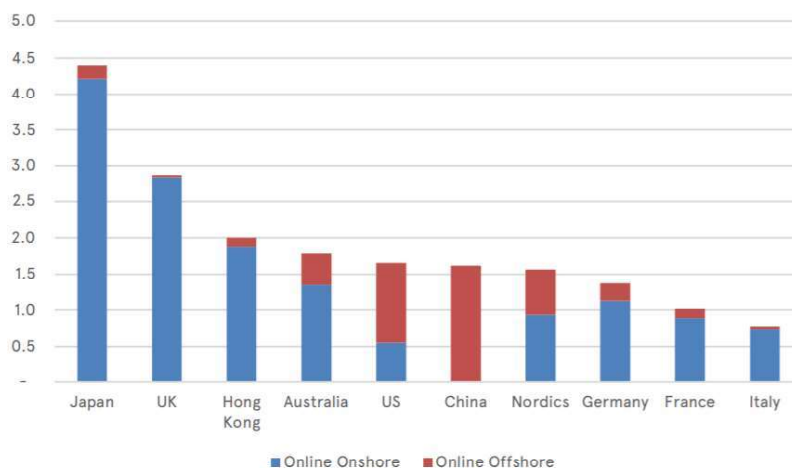


Figura 2.4: Top 10 dos países com mais receitas online onshore vs. offshore em 2017 (US\$bn)
Fonte: H2GC, 2018

Através da análise da Tabela 2.4, é possível verificar que o setor *online* foi o setor com mais influência no forte crescimento das apostas desportivas nos últimos anos, algo que a H2GC prevê que continue a acontecer. Apesar de as receitas brutas das apostas desportivas de base territorial também

terem aumentado, o setor *online* foi o que apresentou um maior crescimento, aumentando de 30% (GGR) em 2012 para 37% em 2017 (H2GC, 2018). A H2GC prevê que se atinja os 42% até 2022.

Tabela 2.4: GGR observado (12-17) e esperado (17-22e) das apostas desportivas *online*

Quota <i>online</i> GGR (%)	2012	2013	2014	2015	2016	2017	2018e	2019e	2020e	2021e	2022e
Apenas desporto	29%	31%	30%	33%	33%	35%	35%	36%	36%	36%	37%
Corridas de cavalos	30%	32%	34%	35%	38%	40%	42%	44%	46%	47%	50%
Total Apostas desportivas	30%	31%	32%	34%	35%	37%	38%	39%	40%	40%	42%

Adaptado de: H2GC, 2018

De acordo com a H2GC (2018), e como é possível observar na Tabela 2.5, o crescimento das apostas desportivas *online* foi impulsionado pelas apostas desportivas *onshore*. Prevê-se que o mercado *onshore* continue a crescer mais rapidamente do que o mercado *offshore*, em parte impulsionado por haver cada vez mais jurisdições que regulam os seus mercados de apostas desportivas *online* (principalmente nos EUA). Estima-se que o mercado *onshore* irá representar quase 75% do total de apostas desportivas *online* até 2022. Em 2012 este valor era apenas de 58% (H2GC, 2018).

Tabela 2.5: GGR observado (12-17) e esperado (17-22e) das apostas desportivas *onshore*

GGR (US\$bn)	2012	2013	2014	2015	2016	2017	2018e	2019e	2020e	2021e	2022e
Apostas desportivas <i>onshore</i>	8.6	9.5	10.9	12.4	14.1	16.8	19.1	21.2	23.4	25.1	27.6
Total de apostas desportivas <i>online</i>	14.8	16.2	18.2	20.0	22.4	25.0	27.8	29.9	32.6	34.4	37.5
<i>Onshore</i> %	58%	59%	60%	62%	63%	67%	69%	71%	72%	73%	74%

Adaptado de: H2GC, 2018

No que diz respeito ao uso do telemóvel como meio para realizar apostas desportivas, este tem sido um dos principais fatores de crescimento das apostas desportivas *online*, dada a maior disponibilidade oferecida aos apostadores. A receita bruta do jogo através do telemóvel aumentou de 29% (GGR) em 2012 para 50% em 2017 e prevê-se que atinja os 60% até 2022 (H2GC, 2018). Estes dados podem ser observados na Tabela 2.6.

Tabela 2.6: GGR observado (12-17) e esperado (17-22e) das apostas desportivas por telemóvel

GGR (US\$bn)	2012	2013	2014	2015	2016	2017	2018e	2019e	2020e	2021e	2022e
Apostas desportivas por telemóvel	4.3	5.7	7.6	9.0	10.7	12.5	14.4	16.2	18.3	20.1	22.7
Total de apostas desportivas <i>online</i>	14.8	16.2	18.2	20.0	22.4	25.0	27.8	29.9	32.6	34.4	37.5
Telemóvel %	29%	35%	42%	45%	48%	50%	52%	54%	56%	58%	60%

Adaptado de: H2GC, 2018

Através da análise da Tabela 2.7, é possível verificar que a proliferação dos mercados de apostas ao vivo (o chamado ‘apostar em live’, ou seja, apostas efetuadas durante a realização de um encontro) também tem sido um fator importante para o crescimento das apostas desportivas *online*, aumentando significativamente o número de mercados disponíveis para os apostadores. Embora a maioria dos mercados de apostas ao vivo também esteja disponível antes do jogo, as apostas ao vivo permitem que os apostadores apostem durante mais tempo. Através da Tabela 2.7, é possível verificar que o GGR das apostas desportivas *online* ao vivo aumentou de 32% em 2012 para 39% em 2017 e prevê-se que atinja os 57% até 2022 (H2GC, 2018).

Tabela 2.7: GGR observado (12-17) e esperado (17-22e) das apostas desportivas ao vivo

GGR (US\$bn)	2012	2013	2014	2015	2016	2017	2018e	2019e	2020e	2021e	2022e
Apostas desportivas ao vivo	2.4	2.7	3.5	3.9	4.7	5.3	6.5	7.7	9.2	10.5	12.3
Total de apostas desportivas <i>online</i> (excluindo corridas de cavalos)	7.3	8.1	9.4	10.8	12.1	13.6	15.5	16.8	18.4	19.6	21.5
Ao vivo %	32%	33%	37%	36%	39%	39%	42%	46%	50%	53%	57%

Adaptado de: H2GC, 2018

A H2GC (2018) também revela que, apesar de o futebol continuar a dominar o mercado das apostas desportivas *online*, tem havido uma mudança significativa em relação a outros desportos, além de mudanças nas competições de níveis mais baixos em cada desporto. Desta forma, o ténis e o basquetebol têm sido dois dos segmentos de apostas desportivas que mais têm crescido, tanto nos mercados regulamentados como nos não regulamentados.

2.2.3. Mercado do jogo em Portugal

2.2.3.1. Dados relativos à atividade de jogo *online*

Segundo o SRIJ (2019), em 2018, os portugueses apostaram ao todo 2,4 mil milhões de euros, um valor que representa mais de 1% do PIB. A atividade de jogos e apostas *online* gerou em 2018 cerca de 152,1 milhões de euros de receita bruta, valor superior em 29,5 milhões de euros face ao ano anterior, crescendo assim 24%. Relativamente à tributação, o valor total do Imposto Especial de Jogo *Online* neste ano foi de 66,5 milhões de euros, tendo aumentado cerca de 23% face ao ano anterior (SRIJ, 2019). Estes dados podem ser observados na Figura 2.5.



Figura 2.5: Receita bruta e IEJO dos jogos e apostas online, em 2018
 Fonte: SRIJ, 2019

Através da análise da Figura 2.6, é possível verificar que a receita bruta das entidades exploradoras de apostas desportivas à cota *online* registou, em 2018, o valor de 78,9 milhões de euros (51,8% do total), registando-se um aumento de 10,8 milhões de euros face ao ano anterior (SRIJ, 2019).



Figura 2.6: Receita bruta das apostas desportivas à cota online, em 2018
 Fonte: SRIJ, 2019

De acordo com o SRIJ (2019), em 2018, o Futebol continuou a ser a modalidade desportiva onde se regista o maior volume de apostas, representando aproximadamente 73,8% do total de apostas desportivas efetuadas. No seu conjunto, as modalidades de Ténis, Basquetebol e Hóquei no Gelo, representaram aproximadamente 23% do total das apostas desportivas (11,9%, 7,4% e 3,6%, respetivamente) (Figura 2.7).

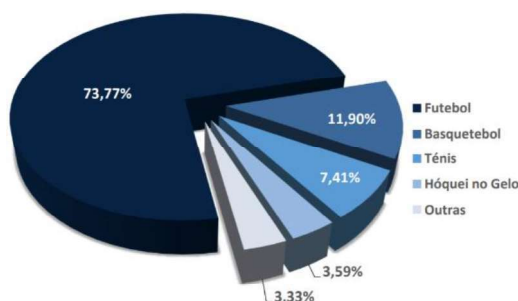


Figura 2.7: Apostas desportivas à cota por modalidade, em 2018
 Fonte: SRIJ, 2019

Segundo o SRIJ (2019), em 2018, e no conjunto das nove entidades exploradoras, apuraram-se 380,1 mil novos registos de jogadores, 25,6 mil autoexclusões e 11,8 mil términos de autoexclusões.

Através da observação da Figura 2.8, é possível verificar que, em 2018, os jogadores com idades compreendidas entre os 25 e os 34 anos representavam 38,8% do total de jogadores registados, tendo sido também este o grupo predominante no ano anterior (SRIJ, 2019).

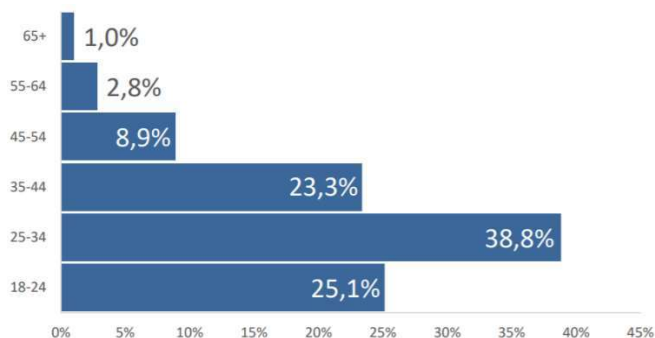


Figura 2.8: Distribuição dos jogadores por estrutura etária, em 2018
Fonte: SRIJ, 2019

De acordo com o SRIJ (2019), o litoral de Portugal continua a apresentar a maior concentração de jogadores, com mais de metade dos jogadores registados a residirem nos distritos do Porto, de Lisboa e de Braga (21,7%, 19,6% e 9,5%, respetivamente). Os jogadores residentes nos distritos de Aveiro e Setúbal representam, no seu conjunto, cerca de 15,9% do total de jogadores.

Com a expansão da Internet e com o aumento da utilização desta por parte da população portuguesa (possível de verificar através de estudos do Instituto Nacional de Estatística), surgiu a necessidade de criar locais virtuais de apostas com o intuito de oferecer aos apostadores a possibilidade de jogarem *online*. Assim, começaram a surgir as casas de apostas *online*, entidades licenciadas que aceitam apostas dos apostadores na previsão de um certo acontecimento. No entanto, só no final de maio de 2016 foi emitida a primeira licença pelo SRIJ.

A 31 de dezembro de 2018, encontravam-se nove entidades autorizadas a exercer a atividade de exploração de jogos e apostas *online* em Portugal, mais duas do que as autorizadas no período homólogo de 2017. Estas entidades são detentoras de 15 licenças, sete para exploração de apostas desportivas à cota e oito para exploração de jogos de fortuna ou azar (SRIJ, 2019).

Como é possível verificar, a procura por apostas desportivas *online* em Portugal tem sido cada vez maior e são cada vez mais as casas que providenciam este serviço aos apostadores portugueses, tornando este mercado cada vez mais competitivo. Por outro lado, a cada vez maior disponibilidade dos serviços *online* destas entidades leva ao crescimento do mercado de apostas desportivas.

2.3. Investigação sobre apostas desportivas

2.3.1. Previsão de resultados de eventos desportivos

A procura de lucro no mercado de apostas desportivas tende a ser feita através da previsão de resultados dos eventos desportivos. Desde há algumas décadas atrás que vários estudos começaram a ser realizados neste sentido. É o caso do estudo realizado por Dixon e Coles (1997). Estes autores basearam-se no modelo desenvolvido por Maher (1982) que assume distribuições de *Poisson* independentes para o número de golos marcados por cada equipa num jogo de futebol, com recurso a dados específicos do desempenho passado das mesmas. No entanto, este modelo não servia como um modelo previsional para os resultados finais dos jogos, sendo apenas utilizado após a conclusão dos mesmos. Desta forma, Dixon e Coles (1997) desenvolveram um modelo capaz de prever o número de golos de um jogo, baseando-se na distribuição de *Poisson* bivariada e ultrapassando a barreira da previsão *ex post* (após o término dos jogos) para *ex ante* (antes do início dos jogos). O modelo desenvolvido forneceu melhores estimativas das *odds*² dos jogos do que as estimativas subjetivas atribuídas pelos *bookmakers*. Desta forma, as probabilidades de cada partida eram calculadas e comparadas com as *odds* dos *bookmakers*. Os autores conseguiram assim elaborar uma estratégia de apostas que, utilizando dados históricos, permite a obtenção de retornos positivos.

Outros estudos foram posteriormente realizados, existindo alguns mais recentes, apresentados abaixo.

Miljković et al. (2010) apresentaram um sistema que utiliza técnicas de *data mining* para prever quais os vencedores dos jogos de basquetebol da liga de NBA (National Basketball Association). O problema de previsão de resultados foi formalizado neste estudo como um problema de classificação, no qual várias técnicas foram testadas (como por exemplo árvores de decisão, *Naive Bayes*, máquinas de vetores de suporte) e onde o método *Naive Bayes* foi o que apresentou melhores resultados. Para avaliar o sistema realizado, os autores utilizaram um conjunto de dados reais referentes a 778 jogos da época 2009/2010 e utilizaram a técnica *10-fold cross validation*. Os autores concluíram que o sistema desenvolvido prevê corretamente os vencedores de cerca de 67% dos jogos.

Leung e Joseph (2014) apresentaram uma abordagem de *data mining* aplicado ao desporto que ajuda a prever os resultados dos jogos de futebol americano universitário. Os autores afirmaram que nos anos anteriores ao do lançamento deste artigo foram aplicados alguns algoritmos estatísticos simples aos resultados de jogos passados, com o objetivo de prever os resultados de jogos futuros. Estes algoritmos geralmente consideravam as duas equipas de um jogo e comparavam as suas estatísticas, pontos fortes e pontos fracos para fazer a previsão. No entanto, estas previsões não eram muito precisas quando as equipas envolvidas nunca tinham jogado uma contra a outra ou quando existia pouca informação sobre os jogadores. Desta forma, os autores desenvolveram um modelo que não se concentra diretamente em

² O termo *odd* (ou probabilidade) representa o valor potencial que um apostador pode receber por cada unidade monetária investida (neste caso, o €).

calcular qual das duas equipas tem maior probabilidade de ganhar, mas sim em analisar as equipas mais semelhantes a cada uma das equipas do jogo, recolher os resultados dos jogos entre essas equipas e utilizar estes resultados para prever o resultado do jogo em questão. Para isso, os autores realizaram uma combinação de quatro estatísticas diferentes (índice percentual de classificações, número de ‘*pythagorean wins*’, estratégia ofensiva e diferencial de rotatividade) e somaram os valores de cada uma delas. A equipa com a maior soma seria a equipa mais semelhante e a utilizada no algoritmo de previsão. A aplicação dos resultados obtidos em dados reais das épocas 2005/2006 a 2013/2014 revelou que esta abordagem permite prever com uma precisão relativamente alta os resultados dos jogos.

Stübinger e Knoll (2018) construíram uma estratégia de apostas com o objetivo de prever o resultado dos jogos das cinco principais ligas europeias de futebol (Inglaterra, França, Alemanha, Itália e Espanha), utilizando uma base de dados com 8082 jogos das temporadas 2013/2014 a 2017/2018. Os autores compararam diferentes abordagens (sete ao todo) para descobrir se abordagens mais complexas, como algoritmos de *machine learning* (por exemplo, *random forest*, *boosting* e máquinas de vetores de suporte), permitiriam obter melhores resultados. Estas estratégias foram assim utilizadas para prever a diferença de golos entre a equipa da casa e a equipa visitante (variável dependente) com base nos dados de jogos passados (variáveis independentes). No que diz respeito à qualidade da previsão, o método *random forest* foi o que apresentou melhores resultados, ou seja, a maior precisão, o menor RMSE (*Root Mean Squared Error*) e o menor desvio absoluto médio (MAD), seguindo-se os outros dois algoritmos de *machine learning*. Os autores verificaram assim que através de algoritmos de *machine learning* é possível obter retornos económicos significativos no que diz respeito a apostas desportivas. Além disso, verificou-se que estes resultados não podem ser alcançados com um modelo de regressão linear ou com estratégias simples de apostas, como por exemplo apostar sempre na equipa que joga em casa.

Constantinou (2019) criou o Dolores, um modelo que prevê os resultados de jogos de futebol num país tendo em conta dados de jogos de futebol de outros países. O modelo consiste numa junção de dois métodos: um sistema de classificação dinâmico, que tem em consideração as discrepâncias de golos observadas em cada jogo, e redes bayesianas híbridas, que utilizam as classificações resultantes do método anterior para prever os resultados dos jogos. O autor afirmou que os modelos anteriores tendem a concentrar-se apenas numa única liga ou torneio, com previsões calculadas utilizando modelos estatísticos, *machine learning*, modelos gráficos probabilísticos e sistemas de classificação. O modelo desenvolvido possui assim um conjunto de dados de treino que incorpora os resultados de 216743 jogos de 52 ligas de futebol de 35 países diferentes. O conjunto de dados de teste possui 206 resultados de jogos de 26 ligas diferentes, disputados de 31 de março a 9 de abril de 2017. Este artigo adotou uma abordagem anterior do autor, onde as classificações das equipas são baseadas em resultados de jogos passados e onde as previsões dos jogos são baseadas em resultados de jogos de equipas diferentes. Esta abordagem estende este modelo a divisões e países diferentes. Para avaliar a precisão do modelo desenvolvido foi utilizada a função RPS (*Rank Probability Score*). Este estudo revelou que, apesar de o desempenho passado de cada equipa ajudar a maximizar a precisão preditiva, é possível fazer uma boa

previsão do resultado de um jogo mesmo quando esta deriva de dados passados de equipas não participantes nesse jogo.

Domingues et al. (2019) afirmaram que normalmente as previsões dos resultados dos jogos são efetuadas através da utilização de modelos estatísticos. Sendo assim, optaram por desenvolver um modelo de aprendizagem incremental para prever os resultados dos jogos de futebol das equipas da primeira divisão da Liga Portuguesa. Mais concretamente, desenvolveram um modelo de regressão logística multinominal para prever o resultado dos jogos, formulando-o como um problema de classificação com três classes: vitória da equipa da casa, empate ou vitória da equipa visitante. Este estudo é uma extensão de um estudo do mesmo ano que consistia num modelo preditivo jornada a jornada onde o conjunto de dados de treino era constituído pelos jogos da jornada 1 a $n-1$ e o conjunto de dados de teste pelos jogos da jornada n , para cada jornada n em N , onde N é o número total de jornadas da competição. Os autores optaram então por utilizar, neste novo modelo, as primeiras n jornadas para treinar o modelo e as jornadas $N - n$ para testá-lo. No final de cada jornada, o n é incrementado e o modelo é de novo “treinado” com os novos dados. Desta forma, ao invés de utilizar um modelo de redes neurais artificiais, os autores desenvolveram um modelo de regressão logística multinominal. Foi também aplicada análise fatorial para tentar extrair as características mais discriminatórias. O conjunto de dados utilizado é constituído por informação pública relativa a 306 jogos de futebol da época 2017/2018.

Apesar de apenas serem apresentados os exemplos acima, muitos são os modelos referentes à previsão de resultados de jogos das mais variadas modalidades desportivas, sendo que são utilizadas técnicas que vão desde métodos estatísticos a *data mining*. Todos estes se relacionam direta ou indiretamente com o mercado das apostas desportivas, no entanto, apenas alguns referem que o objetivo dos mesmos é auxiliar na obtenção de lucro por parte dos apostadores no que diz respeito à realização de apostas desportivas (o objetivo desta dissertação), sendo este o motivo para apenas serem considerados relevantes os estudos apresentados acima.

Como é possível constatar, as variáveis utilizadas na maioria destes modelos são variáveis baseadas em estatísticas de jogos passados, não podendo os apostadores modificar os seus comportamentos para obter mais lucro. A verdade é que as apostas desportivas fazem parte dos chamados jogos de “azar”, e por isso, são baseadas em cálculos de probabilidades. Por esse motivo, um matemático poderá afirmar que qualquer regra identificada neste sentido é falaciosa. Este é um fator pelo qual esta dissertação se pretende diferenciar, uma vez que as variáveis a utilizar podem ser controladas pelo apostador (não sendo baseadas em cálculos de probabilidades), ajustando este as suas decisões com base nos padrões identificados entre determinados comportamentos e o lucro obtido. É de salientar também que os modelos apresentados acima se baseiam na previsão dos resultados de jogos, tendo, portanto, de ser executados para cada jogo em que um apostador pretende apostar, enquanto que o modelo desta dissertação pode ser adotado durante um longo período de tempo, abrangendo um maior número de jogos.

Através da pesquisa efetuada, foi possível perceber que a maioria dos modelos desenvolvidos dizem respeito à previsão de resultados de jogos de futebol, uma vez que este é, como mencionado nas secções anteriores, a modalidade desportiva onde se regista o maior número de apostas desportivas.

2.3.2. Perfil e comportamentos dos apostadores

São também muitos os estudos desenvolvidos no sentido de perceber qual o perfil e comportamento dos apostadores.

LaBrie et al. (2007) desenvolveram o primeiro estudo longitudinal prospetivo relativo ao comportamento dos apostadores *online* de 85 países, obtendo os dados necessários da casa de apostas *online* Bwin Interactive Entertainment AG. Este estudo teve a duração de oito meses e procurou analisar o comportamento dos apostadores perante dois tipos de apostas desportivas: as apostas ao vivo (as chamadas apostas em 'live') e as apostas efetuadas antes do início dos jogos. Dos 40499 apostadores, 91,6% eram do sexo masculino, sendo a idade média de 31 anos. Os autores avaliaram os seguintes comportamentos: frequência das apostas desportivas, número de apostas realizadas por dia, quantia apostada e quantia ganha. O estudo revelou em primeiro lugar que os apostadores do sexo feminino apresentam um comportamento semelhante aos do sexo masculino. Revelou também que os apostadores que apostam antes do início dos jogos fazem-no durante um período de tempo mais longo e efetuam, em média, mais de quatro apostas por dia. O valor total apostado pelos apostadores que apostam ao vivo é substancialmente menor do que nas apostas efetuadas antes do início dos jogos.

McBride e Derevensky (2009) desenvolveram um estudo que examinou o comportamento do jogo na Internet numa amostra de jogadores *online* (não se focando apenas em apostas desportivas). Os 563 participantes foram recrutados através de um anúncio colocado no *site* casinocity.com, dos quais 382 eram do sexo masculino, sendo a grande maioria dos EUA. Apesar desta investigação ter como objetivo o estudo dos problemas relacionados com o jogo *online*, apresenta também conteúdos relevantes no que diz respeito ao perfil e comportamentos dos jogadores. Os autores avaliaram assim características demográficas, como a idade, o género, o país de residência, o estado civil, etc., e características referentes ao jogo, como a frequência e duração do jogo, a quantia gasta, se depois de perderem dinheiro voltam a jogar para o tentar recuperar, o tipo de jogo preferido, entre outras. Também avaliaram o motivo para os participantes terem optado por jogar *online*. O estudo revelou que 65,2% dos participantes joga em dois a cinco *sites*, que 59,3% joga entre 1 a 2h por sessão e que 54,8% gasta abaixo de 25 dólares por sessão. No caso do tipo de jogo mais jogado, as apostas desportivas apresentaram um valor baixo em comparação com outros tipos de jogos, como as *slot machines* por exemplo. O estudo revelou também que os principais motivos para jogar *online* são a sua conveniência, não ter de sair de casa para jogar e os participantes terem acesso aos jogos 24h por dia.

Labrador e Vallejo-Achón (2019) desenvolveram um estudo exploratório com o objetivo de perceber qual a prevalência das apostas desportivas numa população de jovens estudantes de Madrid (incluindo as suas características), uma vez que estudos anteriores revelaram que existe uma elevada

prevalência de problemas de jogo em Espanha. Assim, os autores desenvolveram um questionário que foi entregue a uma amostra de 735 estudantes dos 18 aos 25 anos, dos quais 50,2% são do sexo feminino. Neste estudo, foram analisadas algumas características sociodemográficas dos apostadores, como a idade, o género, o estado civil, etc., e características referentes às apostas efetuadas, como a frequência das apostas, a quantia apostada, o dinheiro gasto, o tipo de aposta (onde foram analisados 12 tipos de apostas diferentes), o desporto escolhido, entre outras. O estudo revelou, assim, que os jovens do sexo masculino apostam com mais frequência e gastam mais dinheiro e mais tempo com as apostas desportivas do que os do sexo feminino. Também revelou que os estudantes que apostam *online* fazem-no com mais frequência e gastam mais dinheiro e mais tempo do que quem aposta *offline*. O desporto onde são efetuadas mais apostas é o futebol, em grande parte devido aos meios de comunicação como a televisão, e as apostas mais frequentes são as mais simples, ou seja, apostar no vencedor de um jogo.

Apesar de apenas serem apresentados os exemplos acima, muitos são os estudos realizados no sentido de perceber qual o perfil e comportamento dos apostadores perante as apostas desportivas. Todos estes se relacionam direta ou indiretamente com o mercado das apostas desportivas, mas apenas alguns têm em conta o tipo de variáveis comportamentais consideradas neste estudo, sendo este o motivo para apenas serem considerados relevantes os estudos apresentados acima. É, no entanto, possível verificar que todos estes estudos têm apenas como objetivo a caracterização dos apostadores e dos seus comportamentos, no sentido de identificar o seu perfil e a sua relação com as apostas desportivas. Este é um fator pelo qual esta dissertação também se diferencia, uma vez que o objetivo passa por analisar os comportamentos dos apostadores e verificar quais são mais influentes na obtenção de lucro.

2.4. Contextualização das variáveis utilizadas

Como é possível perceber pela literatura acima referenciada, são vários os estudos onde são analisadas variáveis importantes relacionadas com o comportamento dos apostadores no que diz respeito ao jogo ou às apostas desportivas. É o caso da frequência com que os indivíduos jogam/apostam, se apostam *online* ou não, qual o tipo de aposta mais escolhido, qual a modalidade desportiva em que mais apostam, se apostam ao vivo ou antes do início dos jogos, etc. Contudo, não foi encontrado nenhum estudo que utilizasse estas variáveis para tentar prever quais contribuem para uma maior obtenção de lucro. Desta forma, para este estudo, as variáveis acima referidas foram também consideradas, no entanto foram também tidas em conta novas variáveis não presentes nos estudos referidos acima, como por exemplo o sistema de apostas utilizado, a quantidade de jogos por aposta, se apostam sempre nas mesmas equipas ou nos habituais vencedores das competições, se utilizam algum *site* ou aplicação como apoio para a realização de apostas desportivas, entre outras.

2.5. *Research Gap*

Apesar da crescente investigação acerca do mercado de apostas desportivas *online*, a maioria dos estudos realizados foca-se na previsão de resultados de jogos para maximizar o lucro ou nos padrões dos jogadores apenas para perceber as suas características e comportamentos, não tendo, portanto, como objetivo, relacionar estes padrões com a obtenção de lucro.

À luz desta lacuna nos estudos apresentados na revisão de literatura, a presente investigação tem como objetivo identificar os fatores relacionados com o conjunto de decisões que um indivíduo toma relativamente a uma aposta desportiva *online* que são mais preponderantes para a obtenção de lucro por parte do mesmo.

3. Metodologia

A ausência de dados disponíveis para realizar este estudo motivou a necessidade de desenvolver e validar um instrumento que permitisse recolher os dados necessários, de forma viável. Sendo assim, este capítulo encontra-se dividido em duas secções. A primeira diz respeito à criação e validação do questionário, onde é apresentada a amostra em estudo, os métodos utilizados para a recolha dos dados junto dos participantes, a construção do instrumento e as análises conduzidas para verificar a sua fiabilidade e validade. Na segunda secção são apresentados os métodos utilizados para o tratamento dos dados e as técnicas utilizadas para desenvolver os modelos implementados neste estudo.

3.1. Recolha dos dados

Para auxiliar nas validações do questionário utilizou-se o Microsoft Excel e para realizar as validações utilizou-se o programa informático *Statistical Package for the Social Sciences* (SPSS 19).

3.1.1. População e Amostra

A população alvo desta investigação corresponde a todos os indivíduos com idade igual ou superior a 18 anos que efetuam as suas apostas no mercado de apostas desportivas *online* e que residem em Portugal.

Para selecionar a amostra foi utilizada uma técnica de amostragem não probabilística, uma vez que não é possível ter acesso a todos os indivíduos que formam a população. Optou-se então por uma amostragem por conveniência, selecionando uma amostra da população que fosse acessível. Ou seja, os participantes foram selecionados por estarem disponíveis, por terem acesso à Internet e por conseguirem aceder ao *link* do questionário, e não por meio de um critério estatístico. Houve, assim, uma maior facilidade operacional e um baixo custo de amostragem.

Desta forma, a amostra desta investigação é composta por um conjunto de apostadores desportivos registados no fórum de apostas Apostaganha (www.apostaganha.pt) e um conjunto de apostadores dos grupos ‘Aqueles Tips’, ‘Grão a Grão by Ricardo Simões’, ‘NOP RED Apostas Desportivas’, ‘ASES Tips’, ‘NSOUSATIPS’ e ‘ISCTE - IUL’ da rede social Facebook. Para além disso, fazem também parte da amostra alguns apostadores conhecidos que foram contactados por mensagem direta (com recurso às redes sociais Facebook, Instagram e WhatsApp).

Estabeleceram-se como critérios de inclusão na amostra: residir em Portugal; ter idade igual ou superior a 18 anos; efetuar as suas apostas no mercado de apostas desportivas *online* (mais concretamente ter realizado pelo menos uma aposta desportiva à cota em casas de apostas *online* no último ano); e o preenchimento adequado do instrumento utilizado.

Participaram nesta investigação 833 indivíduos, de forma voluntária, no entanto as respostas de 367 inquiridos foram removidas da amostra por não terem cumpridos os critérios de inclusão na mesma (restando apenas 466).

3.1.2. Instrumento utilizado

Em primeiro lugar, procurou-se obter dados reais através do contacto com instituições e empresas ligadas ao jogo, nomeadamente algumas casas de apostas *online* e o SRIJ. No entanto, como o mercado das apostas desportivas é um mercado onde as informações são privadas e não devem ser partilhadas, não foi possível obter dados junto das mesmas (todas as respostas foram negativas ou inexistentes).

Ponderou-se também a realização de entrevistas para a obtenção dos dados necessários mas sendo este um tema com alguma sensibilidade, não se considerou ser a melhor opção. De facto, estudos realizados confirmaram que as entrevistas não são o melhor meio para a obtenção de dados neste tipo de estudos. Por exemplo, Wood e Williams (2007) desenvolveram um estudo referente ao jogo *online* onde referem que, quando se trata de comportamentos possivelmente ilegais ou sensíveis, tais como o jogo *online*, os inquiridos podem ter a tendência para distorcer as suas respostas.

Desta forma, e tendo em conta os dados necessários para a realização desta investigação (alguns deles que só o próprio inquirido pode responder, não podendo ser obtidos com recurso às casas de apostas), foi utilizado um método de investigação quantitativo, obtendo-se as informações necessárias por meio de um inquérito por questionário (instrumento). Optou-se por este método por acarretar poucos custos financeiros e de tempo, permitindo recolher a informação desejada de uma forma mais prática e rápida.

Após ser tomada a decisão de que se iria utilizar um questionário como técnica de recolha de dados, foi necessário decidir se se iria utilizar um questionário já existente ou um novo. Uma vez que não foi encontrado nenhum estudo com um questionário semelhante ao desejado nesta investigação e que satisfizesse os objetivos da mesma, decidiu-se criar um novo.

3.1.2.1. Construção do questionário

Todas as versões do questionário (incluindo a versão final) encontram-se divididas em duas secções. A primeira secção inclui questões demográficas (como o género, escalão etário, habilitações literárias, etc.) e a segunda inclui questões específicas desta investigação.

Apesar de o questionário ter sido criado de raiz, para a formulação dos vários itens do mesmo foram analisados alguns estudos já efetuados. Esta análise foi um requisito fundamental para a identificação das variáveis a considerar para dar resposta ao objetivo da investigação.

Na Tabela 3.1, apresentam-se as variáveis correspondentes a cada questão do questionário.

Tabela 3.1: Variáveis em estudo por questão

Número da questão	Variável em estudo
2	Género
3	Escalão etário
4	Habilitações literárias
5	Ocupação
6	Setor de atividade
7	Região de residência
8	Distrito de residência (Norte)
9	Distrito de residência (Centro)
10	Distrito de residência (Sul)
11	Casa de apostas
12	Frequência com que se aposta
13	Número de apostas
14	Modalidade de aposta
15	Sistema de apostas desportivas
16	Tipo de aposta
17	Lucro obtido com tipo de aposta mais utilizado
18	Modalidade desportiva
19	Estratégia
20	Apostar nas mesmas equipas
21	Colocar os habituais vencedores das nas apostas
22	Utilizar <i>site</i> ou aplicação como apoio
23	Frequência com que se obtém lucro
24	Dinheiro gasto
25	Lucro obtido

As questões demográficas (2 a 10) baseiam-se, assim, em algumas questões do questionário da dissertação de mestrado desenvolvida por Magalhães³ (2011), mas com algumas adaptações:

- ✓ Para além das classes ‘Masculino’ e ‘Feminino’, o item referente ao Género passou a ter também a classe ‘Outro’ (questão 2);
- ✓ O item referente ao Ano de Nascimento foi transformado em Escalão etário, possuindo seis classes (questão 3). Estas estão de acordo com os escalões apresentados nos relatórios de estatísticas do SRIJ (SRIJ, 2019);
- ✓ O item referente ao Grau de instrução foi mantido com as mesmas classes, mas com a junção das classes ‘Ensino secundário (12.º ano)’ e ‘Curso profissional’ e com a separação de ‘Pós-Graduação, Mestrado e Doutoramento’ em três classes distintas. Este item foi renomeado para Habilitações literárias (questão 4). As classes estão de acordo com o definido pela DGAEP - Direção-Geral da Administração e do Emprego Público;
- ✓ As classes do item referente à Ocupação foram alteradas, mantendo-se apenas as classes ‘Desempregado’, ‘Estudante’, ‘Reformado’ e ‘Outra’ e acrescentando-se ‘Desempregado à procura

³ O questionário desta dissertação foi adaptado da secção *Demographics* do PPGM (*Problem & Pathological Gambling Measure*) do estudo desenvolvido por Wood e Williams (2009).

do primeiro emprego’, ‘Empregado por conta própria’ e ‘Empregado por conta de outrem’ (questão 5);

- ✓ Foi criado um novo item referente ao setor de atividade em que o inquirido trabalha (questão 6). Como auxílio, foi utilizado o *site* Pordata (Pordata, 2020). As classes deste item correspondem aos vários setores de atividade económica considerados na base de estatísticas certificadas mencionada acima. Foram apenas acrescentadas as seguintes classes: ‘Desporto’, ‘Direito’ e ‘Tecnologias de Informação e Comunicação’;
- ✓ O item referente à Região de origem foi transformado em Região de residência, possuindo seis classes (questão 7);
- ✓ Foram criadas três novas questões referentes aos distritos de cada uma das regiões de residência (questões 8, 9 e 10). Uma vez que alguns distritos podem ser considerados em duas regiões diferentes (exemplo: ‘Aveiro’, ‘Viseu’ e ‘Guarda’ podem ser considerados como norte ou centro do país), optou-se pelo que pareceu mais correto.

Para construir as questões específicas do estudo em questão, foram tidas em conta algumas variáveis utilizadas em estudos anteriores (mencionados no capítulo 2), mas também outras variáveis não presentes nesses estudos (escolhidas com base em experiência prévia na temática e com a ajuda de um apostador). Estas questões foram sofrendo alterações à medida que foram sendo realizados os testes piloto. As informações referentes aos testes piloto encontram-se descritas na secção seguinte.

Para definir as opções de resposta de algumas das questões apresentadas nesta secção, foi necessário recorrer à literatura mencionada no capítulo 2. Na questão 11, as casas de apostas apresentadas como opções de resposta são as únicas entidades com licença para a exploração de apostas desportivas à cota *online* em Portugal (SRIJ, 2019). As restantes casas de apostas estão incluídas na opção ‘Outra’ e dizem respeito a casas de apostas não legalizadas em Portugal. Na questão 14, as modalidades de aposta foram retiradas do Artigo 7.º do Anexo, alínea 1, da Portaria n.º 173/2015 (Ministério da Solidariedade, Emprego e Segurança Social, 2015). Na questão 15, as opções de resposta correspondem aos sistemas de apostas desportivas apresentados por Teixeira (2017). Na questão 16, os tipos de apostas foram retirados do Artigo 6.º do Anexo, alínea 1, da Portaria n.º 173/2015 (Ministério da Solidariedade, Emprego e Segurança Social, 2015).

Note-se que, na versão final, em nove dos quinze itens desta secção foi utilizada uma escala de *Likert* (ou item de *Likert*) simétrica com 5 alternativas e com níveis equidistantes, ou seja, as distâncias entre os níveis são iguais. A escala utilizada foi a seguinte: ‘Nunca’, ‘Poucas vezes’, ‘Algumas vezes’, ‘Muitas vezes’ e ‘Sempre’. É possível verificar que os extremos são nitidamente opostos e que o nível ‘Algumas vezes’ é o ponto intermédio (ponto neutro).

3.1.2.2. Testes piloto

Neste estudo foram realizados dois testes piloto que contribuíram para a melhoria do questionário desenvolvido.

A primeira versão do questionário (ver Apêndice A) foi distribuída por 10 apostadores (primeiro teste piloto), com o objetivo de o validar e identificar inconsistências nas questões e por forma a poder corrigir alguns aspetos de forma a não suscitar qualquer tipo de dúvidas aos inquiridos. Estes 10 indivíduos, todos do sexo masculino, foram contactados por mensagem direta no WhatsApp no dia 27 de fevereiro de 2020, tendo-se conseguido obter todas as respostas até ao dia 29 de fevereiro de 2020. Os participantes foram esclarecidos sobre as instruções de preenchimento e foram esclarecidas algumas dúvidas. É importante referir que as respostas destes participantes não entraram em nenhuma das fases de recolha de dados seguintes, ou seja, só foram utilizadas neste teste piloto. Os resultados deste teste piloto foram analisados e foi efetuada uma primeira validação do questionário. Mediante esta validação e mediante o *feedback* dos 10 inquiridos, foram efetuadas algumas alterações ao questionário, dando origem a uma segunda versão do mesmo. Foram efetuadas as alterações apresentadas no Apêndice B.

Tendo em conta todas estas alterações, decidiu-se efetuar um segundo teste piloto.

A segunda versão do questionário (ver Apêndice C) foi distribuída por 25 apostadores (segundo teste piloto), com o objetivo de obter informação que permitisse aperfeiçoar o instrumento. Estes 25 indivíduos (24 do sexo masculino e 1 do sexo feminino) foram contactados por mensagem direta no dia 16 de abril de 2020 (com recurso às redes sociais Facebook, Instagram e WhatsApp), tendo-se conseguido obter todas as respostas até ao dia 22 de abril de 2020. Desta vez não foi necessário qualquer tipo de esclarecimento de dúvidas referentes a questões do questionário, uma vez que foram adicionadas pequenas notas em algumas questões. É importante referir que as respostas destes participantes não entraram na fase de recolha de dados final, ou seja, só foram utilizadas neste teste piloto. Os resultados deste segundo teste piloto foram analisados e foi efetuada uma nova validação do questionário. Mediante esta validação, foram novamente efetuadas algumas alterações ao questionário, dando origem a uma terceira e última versão do mesmo. Foram assim efetuadas as alterações apresentadas no Apêndice D.

3.1.2.3. Versão final do questionário

Como foi referido na secção 2.1., a versão final do questionário encontra-se dividida em duas secções, num total de 24 questões (ver Apêndice E). Fora destas secções, encontra-se uma questão inicial para garantir que apenas responderiam ao questionário os apostadores que tivessem realizado pelo menos uma aposta desportiva à cota em casas de apostas *online* no último ano. Caso o inquirido respondesse ‘Não’ a esta questão, ficaria automaticamente excluído do estudo e não responderia às restantes questões do questionário.

Apesar de as questões abertas permitirem investigações mais precisas e profundas, porque o inquirido tem mais liberdade de resposta, podem provocar alguma falta de motivação ao mesmo, não respondendo às questões com frieza e podendo fornecer informações falsas (por razões conscientes ou

não). Sendo assim, o questionário desenvolvido neste estudo é composto apenas por questões fechadas, por serem mais diretas, simples e rápidas de responder e por apresentarem menores dificuldades no tratamento estatístico dos dados (Costa & Miranda, 2017; Gil, 2008).

A primeira secção destina-se a fazer a caracterização dos apostadores, incluindo questões demográficas (tais como o género, o escalão etário, as habilitações literárias, a ocupação, o setor de atividade, a região de residência e o distrito de residência) e é composta por nove questões (questão 2 a 10). No entanto, os inquiridos apenas tiveram de responder a uma das questões 8, 9 e 10, consoante a sua região de residência (seleccionada na questão 7).

A segunda secção diz respeito aos comportamentos dos apostadores, incluindo questões específicas do estudo em questão e é composta por quinze questões (questão 11 a 25).

3.1.2.4. Avaliação da validade e fiabilidade da versão final do questionário

Para assegurar que um determinado questionário é eficaz, e pode ser utilizado como instrumento de investigação, devem ser consideradas duas características de medição no seu desenvolvimento: a validade e a fiabilidade (Freitas & Rodrigues, 2005; Souza et al., 2017). Só assim é possível assegurar a qualidade do instrumento utilizado e dos seus resultados.

A validade de um questionário refere-se à propriedade de um instrumento medir exatamente o que se propõe a medir. A validade pode ser comprovada através da validade de conteúdo, de critério, de construto, entre outros. Nesta investigação foi considerado o tipo de validade de conteúdo, que corresponde ao grau em que um instrumento inclui todos os itens necessários para representar o conceito que está a ser medido (Souza et al., 2017). Neste caso, para um questionário ser considerado válido, normalmente utiliza-se uma abordagem qualitativa, por meio da avaliação de um comité de especialistas (para que estes verifiquem se os itens são ou não adequados para atingir os objetivos propostos), e uma abordagem quantitativa, onde se calcula o índice de validade de conteúdo (Souza et al., 2017).

Para satisfazer este ponto, procurou-se contactar alguns especialistas na área das apostas desportivas para que fizessem parte do comité, no entanto, apenas se obteve uma resposta positiva. Desta forma, optou-se por submeter o questionário ao escrutínio deste perito (um apostador profissional de uma empresa especializada em apostas desportivas *online*), que procedeu à avaliação dos itens da segunda secção do mesmo quanto à sua relevância e clareza. Foi solicitado ao perito que avaliasse cada item como um todo, ou seja, apenas uma pontuação para a questão e para as opções de resposta. Para isso foram enviados dois documentos: um com as questões do questionário completas (incluindo as opções de resposta de cada questão) e outro com uma tabela semelhante à do Apêndice F, onde o perito poderia classificar cada item. Os itens foram classificados numa escala de 5 pontos (1 = Nada relevante; 2 = Pouco relevante; 3 = Relevante; 4 = Muito relevante; 5 = Extremamente relevante). A mesma escala foi utilizada para a clareza, substituindo a palavra relevante por claro. A média das pontuações atribuídas pelo perito foi elevada nos dois aspetos em consideração (Relevância: M = 4.4, Min. = 2, Max. = 5; Clareza: M = 5, Min. = 5, Max. = 5). Na tabela apresentada no Apêndice F, é possível verificar a cotação

atribuída pelo perito a cada um dos itens. No que diz respeito à relevância, apenas houve um item cotado abaixo de 4 (questão 24). Isto indicaria que esta questão deveria ser repensada e melhorada ou até mesmo removida, uma vez que foi classificada como ‘Pouco relevante’. No entanto, como não se conseguiu obter a percepção de mais peritos, decidiu-se mantê-la com o objetivo de confirmar se o dinheiro gasto por um apostador influencia o lucro. Desta forma, não foi eliminado nenhum dos itens do questionário. No que diz respeito à clareza, todos os itens foram avaliados pelo perito como extremamente claros, chegando-se assim à conclusão de que estavam bem construídos e perceptíveis, não sendo necessário proceder a nenhuma alteração. Após esta análise procedeu-se à distribuição do questionário final.

A fiabilidade (ou confiabilidade) refere-se à capacidade de um instrumento produzir resultados consistentes, no tempo e no espaço, ou a partir de observadores diferentes, permitindo avaliar a pertinência ou não de determinados itens do instrumento (Souza et al., 2017). A fiabilidade de um questionário pode ser comprovada através de vários critérios, entre eles a estabilidade, a consistência interna e a equivalência. Este último critério corresponde ao grau de concordância entre dois ou mais observadores quanto aos scores de um instrumento. A forma mais comum de avaliar a equivalência é a fiabilidade interobservadores, que envolve a participação independente de dois ou mais avaliadores e que, para variáveis categóricas, é medida através do coeficiente Kappa (Souza et al., 2017). Uma vez que, tal como foi referido acima, apenas se obteve uma resposta positiva de um especialista na área das apostas desportivas, este critério não foi considerado. Sendo assim, nesta investigação foram tidos em conta os dois primeiros critérios, ou seja, o critério de estabilidade e o de consistência interna.

A estabilidade corresponde ao grau em que resultados similares são obtidos em dois momentos distintos. A avaliação da estabilidade (temporal) é normalmente realizada através do método de teste-reteste, que consiste na aplicação de uma mesma medida em dois momentos distintos e ao mesmo grupo de participantes. O intervalo de tempo entre as medições influencia a interpretação da fiabilidade do teste-reteste portanto considera-se adequado um intervalo de 10 a 14 dias entre o teste e o reteste (Souza et al., 2017). O objetivo deste método consiste então em verificar se os resultados obtidos são semelhantes nos dois momentos distintos e, se o teste for bem realizado e se não ocorrer nenhuma alteração significativa entretanto, devem sê-lo. O uso deste método requer que o fator a ser medido permaneça o mesmo nos dois momentos e que qualquer alteração na pontuação obtida pode ser causada por erros aleatórios. Este método estava previsto ser implementado aquando da realização do segundo teste-piloto, no entanto, devido à COVID-19 e ao cancelamento de praticamente todos os jogos das várias modalidades, considerou-se não justificável. Dificilmente, em circunstâncias normais, se consegue obter uma fiabilidade teste-reteste perfeita com correlação igual a 1 pois existem inúmeros fatores que influenciam as pontuações, mas o objetivo é sempre de obter um valor próximo de 1 (Daniel et al., 2015). Neste caso, as pontuações poderiam estar longe do real, uma vez que houve uma modificação das circunstâncias durante o intervalo entre os dois momentos distintos previstos. 2020 foi um ano atípico, com um grande impacto em todas as áreas, e esta modificação (cancelamento de praticamente todos os jogos das várias modalidades) poderia fazer com que fossem obtidas pontuações

que indiciam que o questionário não foi corretamente construído. Desta forma, optou-se por não aplicar este método, sendo utilizado apenas o critério de consistência interna.

A consistência interna, ou homogeneidade, avalia se todas as subpartes de um instrumento medem a mesma característica. A consistência interna dos instrumentos normalmente é avaliada por meio de um procedimento estatístico muito aceite e utilizado na comunidade científica, o coeficiente alfa de *Cronbach* (Souza et al., 2017). Este coeficiente reflete o grau de covariância entre os itens de uma escala, sendo que, quanto menor a soma da variância dos itens, mais consistente é o instrumento. Optou-se então por utilizar esta medida neste estudo.

Como foi referido na secção 3.1.1., participaram nesta investigação 833 indivíduos, dos quais apenas 597 foram válidos (*a priori*) e considerados no estudo. Destas 597 respostas foram retiradas trinta para realizar esta última validação, sendo elas as primeiras quinze e as últimas quinze respostas do questionário, para conferir alguma aleatoriedade (após efetuar algumas correções apresentadas na secção 3.2.2.). Optou-se por utilizar 30 respostas nesta validação para que se mantivesse aproximadamente o mesmo número de respostas da validação anterior e para conseguir guardar o máximo possível de respostas para a análise dos dados.

Das 25 questões do questionário, optou-se por excluir as questões da primeira secção (e também questão inicial e a questão 11) do cálculo do alfa de *Cronbach* (α), uma vez que estas questões serão apenas utilizadas para caracterizar a amostra. Das 15 questões restantes (da segunda secção), foi necessário analisar o que cada uma estava a medir, pois todos os itens devem ser representativos do mesmo conceito que está a ser medido. Rapidamente se percebeu que o questionário mede mais do que uma variável, ou seja, há itens que estão relacionados com a estratégia e outros com o lucro. Sendo assim, foram definidas inicialmente duas variáveis latentes (dimensões ou domínios), sendo elas a Estratégia e o Lucro. De seguida, foi necessário definir quais os itens que medem cada variável latente para depois se testar o alfa de *Cronbach* para cada grupo de itens (analisar cada dimensão em separado). Ao passar por este processo, concluiu-se que a dimensão Estratégia deveria ser dividida em duas dimensões distintas: a Estratégia Técnica e a Estratégia Desportiva. A primeira diz respeito à parte técnica, ou seja, ao conhecimento do apostador em relação ao funcionamento das apostas desportivas e do método de apostar, e a segunda diz respeito à parte mais desportiva, ou seja, ao conhecimento do apostador em relação às modalidades desportivas, às equipas, etc. Desta forma, a dimensão Estratégia Técnica é composta pelas questões 14, 15, 16 e 22, a dimensão Estratégia Desportiva pelas questões 18, 19, 20 e 21 e a dimensão Lucro é composta pelas questões 17, 23, 24 e 25.

Para medir a fiabilidade das várias dimensões, foram utilizados como referência os valores apresentados na Tabela 3.2, propostos por Freitas e Rodrigues (2005).

Tabela 3.2: Valores de referência do alfa de Cronbach

Fiabilidade	Valor do Alfa de Cronbach
Muito Alta	$\alpha > 0,90$
Alta	$0,75 < \alpha \leq 0,90$
Moderada	$0,60 < \alpha \leq 0,75$
Baixa	$0,30 < \alpha \leq 0,60$
Muito Baixa	$\alpha \leq 0,30$

Fonte: Freitas e Rodrigues (2005)

Conforme apresentado na Tabela 3.3, e de acordo com os valores de referência da Tabela 3.2, os valores do alfa revelam que a fiabilidade do questionário para a dimensão Estratégia Técnica é Alta (alfa $\approx 0,8$) e para a dimensão Estratégia Desportiva é Moderada (alfa $\approx 0,7$). Por outro lado, revelam também que a fiabilidade do questionário para a dimensão Lucro é Baixa (alfa $\approx 0,5$).

A dimensão Estratégia Técnica apresenta uma fiabilidade Alta possivelmente por ser composta por um elevado número de itens (este é um dos fatores que influenciam a fiabilidade). Todos os itens desta dimensão são construídos a partir da mesma escala de *Likert*. As questões 12 e 13 seriam incluídas na dimensão Estratégia Técnica, no entanto, por serem de escalas diferentes e por baixarem ligeiramente o valor do alfa desta dimensão (ver Tabela 3.3), foram excluídas do cálculo do mesmo, mas foram mantidas por se considerar que são importantes nesta investigação. A dimensão Estratégia Desportiva apresenta uma fiabilidade Moderada, que é considerada aceitável segundo alguns autores (Freitas & Rodrigues, 2005). Por último, a dimensão Lucro apresenta uma fiabilidade Baixa possivelmente por ser composta por um número de itens mais reduzido e de escalas diferentes. No entanto, estudos indicam que para uma previsão para grupos de 25 a 50 indivíduos, o alfa considerado aceitável é acima de 0.5 (Maroco & Garcia-Marques, 2006).

É importante ressaltar que não é possível afirmar que a dimensão Estratégia Técnica tem maior fiabilidade que a dimensão Lucro, ou seja, não é possível comparar o valor do alfa entre as dimensões (visto que cada dimensão possui um número de itens diferente).

Tabela 3.3: Alfa de Cronbach de cada dimensão

Dimensões	Alfa de Cronbach
Estratégia Técnica (Q14, Q15, Q16 e Q22)	0,827
* Estratégia Técnica (com Q12 e Q13)	0,818
Estratégia Desportiva (Q18, Q19, Q20 e Q21)	0,698
Lucro (Q17 e Q23 a Q25)	0,510
Todos os itens (Q14 a Q25)	0,797

Apesar de ainda não haver um consenso em relação à interpretação da fiabilidade de um instrumento obtida através do valor do coeficiente alfa de *Cronbach*, considera-se satisfatório um questionário que obtenha alfa $\geq 0,60$ (Freitas & Rodrigues, 2005). Considerando as 30 respostas, e como verificado

anteriormente, o questionário apresenta fiabilidade distinta consoante as diferentes dimensões. Geralmente, quanto maior o número de itens de um questionário, maior é o valor do alfa de *Cronbach*. No entanto, todos esses itens devem ser representativos do mesmo conceito que está a ser medido. Neste caso, o cálculo de uma fiabilidade global do questionário (envolvendo todos os itens) não se justifica, visto que se iriam estar a agrupar itens que medem conceitos distintos e que o valor do alfa global é maior do que o de algumas das dimensões. No entanto, apresenta-se na Tabela 3.3 o valor do alfa global, que apesar destas particularidades, revela uma fiabilidade Alta (alfa $\approx 0,8$).

Por fim, nas duas dimensões referentes à Estratégia, por serem compostas por um maior número de itens, optou-se por calcular o alfa de *Cronbach* se cada um deles for excluído. Se ao excluir determinado item o valor do alfa diminuir muito, então significa que esse item influencia significativamente, e de forma positiva, o alfa. Caso contrário, se ao excluir um item o valor do alfa aumentar muito, então deve ser repensado se esse item deve ou não ser mantido na dimensão em questão (ou até mesmo no questionário).

Como é possível verificar na Tabela 3.4, referente à dimensão Estratégia Técnica, o alfa mantém-se sempre acima dos 0,80 quando qualquer um dos itens é excluído. Sendo assim, todos os itens são importantes.

Tabela 3.4: Alfa de Cronbach se o item for excluído (dimensão Estratégia Técnica)

Item	Alfa de Cronbach se o item for excluído
Q14_1	0,838
Q14_2	0,818
Q14_3	0,817
Q14_4	0,819
Q14_5	0,822
Q15_1	0,832
Q15_2	0,835
Q15_3	0,834
Q15_4	0,817
Q16_1	0,814
Q16_2	0,811
Q16_3	0,819
Q16_4	0,822
Q16_5	0,810
Q16_6	0,822
Q16_7	0,817
Q16_8	0,816
Q16_9	0,813
Q16_10	0,814
Q16_11	0,821
Q16_12	0,816
Q16_13	0,820
Q16_14	0,817
Q16_15	0,820
Q16_16	0,820
Q16_17	0,813
Q22	0,853

Já no caso da dimensão Estratégia Desportiva, na maioria dos casos, quando um item é excluído, o alfa mantém-se acima dos 0,60 e abaixo dos 0,70. No entanto, em alguns casos, se esses itens fossem excluídos o alfa subiria para valores superiores a 0,70 (Tabela 3.5). À primeira vista, isto indica que esses itens possivelmente poderiam ser removidos, obtendo-se assim um maior alfa, mas como o valor do alfa não iria aumentar muito, optou-se por mantê-los.

Tabela 3.5: Alfa de Cronbach se o item for excluído (dimensão Estratégia Desportiva)

Item	Alfa de Cronbach se o item for excluído
Q18_1	0,661
Q18_2	0,666
Q18_3	0,691
Q18_4	0,681
Q18_5	0,638
Q18_6	0,653
Q18_7	0,670
Q18_8	0,676
Q18_9	0,661
Q19_1	0,710
Q19_2	0,679
Q19_3	0,724
Q20	0,710
Q21	0,712

Tendo em conta a avaliação da validade e fiabilidade apresentadas acima, considera-se que o questionário se encontra validado e pode ser utilizado como instrumento de investigação.

Nota: Mediante a validação final efetuada (após o cálculo dos alfas de *Cronbach*), optou-se por remover deste estudo a questão ‘Caso tenha escolhido a opção ‘Mais que uma vez por semana’ na questão 11, qual(ais) o(s) dia(s) da semana em que costuma apostar?’ por se considerar que foi mal construída, não abrangendo todos os cenários possíveis.

3.1.3. Procedimento de recolha dos dados

Para ser possível efetuar a recolha dos dados necessários para alcançar os objetivos deste estudo, foi utilizada a plataforma *online* Google Forms. Todas as versões do questionário foram, assim, transpostas para a versão *online* utilizando este serviço, por ser gratuito e pela facilidade de utilização e de armazenamento dos dados.

A recolha dos dados nesta investigação decorreu em nove etapas, sendo elas:

1. Construção e desenho do questionário;
2. Distribuição do questionário por 10 indivíduos (1º teste piloto);
3. Validação do questionário inicial (foram detetadas algumas situações que podiam gerar confusão, levando a que fosse necessário realizar um novo teste);
4. Alteração do questionário inicial;
5. Distribuição do novo questionário por 25 indivíduos (2º teste piloto);
6. Validação do novo questionário;
7. Alteração do novo questionário;
8. Distribuição do questionário final;

9. Validação do questionário final (para garantir a sua validade e fiabilidade).

O questionário final foi então partilhado no fórum de apostas Apostaganha, na rede social Facebook (nos grupos já mencionados na secção 3.1.1.) e com alguns apostadores conhecidos, desde o dia 25 de junho de 2020 até ao dia 16 de julho de 2020 (três semanas). Este instrumento foi distribuído através da divulgação de um *link* com acesso direto ao questionário, juntamente com uma explicação do objetivo do estudo e a pertinência da colaboração do participante.

As questões do questionário encontram-se antecedidas por um conjunto de informações referentes ao mesmo, nomeadamente o tempo de resposta estimado (aproximadamente 10 minutos), a possibilidade de solicitar informações adicionais via email e um esclarecimento de que as informações recolhidas se destinam única e exclusivamente ao suporte da dissertação, sendo garantido o anonimato e a confidencialidade das mesmas (assim consegue-se minimizar a probabilidade de serem fornecidas respostas falsas por parte dos inquiridos).

Na sua versão *online*, o questionário foi construído de forma a que praticamente todos os itens fossem de resposta obrigatória e com a impossibilidade de avançar para a secção seguinte sem o correto preenchimento dos itens da secção anterior. O mesmo foi construído à base de botões e caixas de seleção de opções, sendo possível responder ao questionário através de simples cliques.

Após o término da recolha de respostas, estas foram extraídas para uma folha de cálculo do Microsoft Excel, dando origem a um documento que serviu de base para a continuação do estudo.

3.2. Desenvolvimento dos modelos preditivos

Nesta secção, são detalhadas as principais tarefas implementadas no âmbito das fases Compreensão dos Dados, Preparação dos dados e Modelação do CRISP-DM. Para auxiliar na análise das respostas ao questionário utilizou-se o Microsoft Excel e para o tratamento e análise dos dados utilizou-se o programa informático *Statistical Package for the Social Sciences* (SPSS 19). Para o desenvolvimento dos modelos preditivos utilizou-se o R (R Core Team, 2020), no ambiente RStudio.

3.2.1. Compreensão dos dados

Como já foi possível perceber, o conjunto de dados⁴ a utilizar foi obtido através de um questionário construído especificamente para este estudo (e que foi validado). A primeira parte desta etapa do CRISP-DM referente aos métodos de obtenção dos dados encontra-se, assim, apresentada na secção 3.1.

O modelo preditivo a desenvolver poderia incidir sobre a caracterização dos jogadores (questão 2 a 11 do questionário), os potenciais fatores que influenciam o lucro (questão 12 a 25 do questionário) ou em ambos. Apesar de todas as questões poderem ser importantes para a componente preditiva, optou-se

⁴ Normalmente é utilizado o termo *dataset* para se referir ao conjunto de dados utilizado. Sendo assim, e para que seja mais claro, será utilizado o termo em inglês.

por construir o modelo preditivo utilizando apenas as relacionadas com os comportamentos dos apostadores (questão 12 a 25). A questão 17 foi excluída das modelações por se considerar que, pela forma como foi construída, não possui grande poder preditivo.

Uma vez que as questões foram construídas de forma a conseguir recolher a informação pretendida, as respetivas variáveis necessárias para desenvolver o modelo preditivo já estavam compreendidas e selecionadas. Na Tabela 3.6, apresentam-se as questões do questionário e os respetivos nomes e classes das variáveis (que serão utilizadas na construção dos modelos preditivos).

Tabela 3.6: Associação entre as questões do questionário e o nome das variáveis

Identificação da questão	Nome da variável	Classes da variável
Q12	Frequencia	Diariamente; Mais que uma vez por semana; Semanalmente; Quinzenalmente; Mensalmente; Menos que mensalmente
Q13	NumeroApostas	Até 4 apostas; Entre 5 e 9 apostas; Entre 10 e 14 apostas; Entre 15 e 19 apostas; Entre 20 e 24 apostas; Mais de 25 apostas
Q14_1	Modalidade_Simples	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q14_2	Modalidade_Combinada_ UmTresJogos	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q14_3	Modalidade_Combinada_ QuatroSeisJogos	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q14_4	Modalidade_Combinada_ SeteOitoJogos	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q14_5	Modalidade_Multipla	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q15_1	Sistema_StakeFixa	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q15_2	Sistema_StakeNiveis	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q15_3	Sistema_Martingale	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q15_4	Sistema_ApostasLivres	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_1	TipoAposta_ UmXDois	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_2	TipoAposta_ UmXDois Intervalo	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_3	TipoAposta_ Ambas Marcam	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_4	TipoAposta_ Cantos	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_5	TipoAposta_ Dupla Possibilidade	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_6	TipoAposta_ Equipa MarcaAmbasPartes	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_7	TipoAposta_ Golos Equipa	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_8	TipoAposta_ Intervalo Final	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_9	TipoAposta_ Mais MenosGolos	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_10	TipoAposta_ Mais MenosGolosIntervalo	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_11	TipoAposta_ Jogador VenceSet	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_12	TipoAposta_ Primeira EquipaMarcar	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_13	TipoAposta_ Primeira EquipaMarcarXPontosSet	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre

Identificação da questão	Nome da variável	Classes da variável
Q16_14	TipoAposta_QuemMarca Golo	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_15	TipoAposta_QuemMarca Ensaio	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_16	TipoAposta_Resultado Exato	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q16_17	TipoAposta_Vencedor	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_1	Modalidade_Andebol	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_2	Modalidade_Basquetebol	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_3	Modalidade_Futebol	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_4	Modalidade_Futsal	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_5	Modalidade_Hoquei	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_6	Modalidade_Rugby	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_7	Modalidade_Tenis	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_8	Modalidade_TenisMesa	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q18_9	Modalidade_Voleibol	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q19_1	Estrategia_JogosProprio DiaAoVivo	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q19_2	Estrategia_JogosProprio DiaAntesInicio	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q19_3	Estrategia_JogosDias Diferentes	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q20	MesmasEquipas	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q21	HabituaisVencedores	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q22	SiteOuAppApoio	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q23	FrequenciaLucro	Nunca; Poucas vezes; Algumas vezes; Muitas vezes; Sempre
Q24	DinheiroGasto	Até 49€; Entre 50€ e 99€; Entre 100€ e 499€; Entre 500€ e 999€; Entre 1000€ e 4999€; Mais de 5000€
Q25	Lucro	Até 49€; Entre 50€ e 99€; Entre 100€ e 499€; Entre 500€ e 999€; Entre 1000€ e 4999€; Mais de 5000€

Através da análise da Tabela 3.6, verifica-se que as variáveis Frequencia, NumeroApostas, DinheiroGasto e Lucro são claramente ordinais. No entanto, as restantes, como são construídas a partir da escala de *Likert* mencionada anteriormente, podem ser classificadas tanto como variáveis ordinais como de intervalo. Esta decisão tem originado algumas discussões entre investigadores, sendo que, Ray Poynter, Diretor Administrativo da TFP (*The Future Place*), afirma que esta decisão cabe ao investigador, mas que deve ser justificada (Poynter, 2010). Neste estudo, optou-se por considerar estas variáveis como ordinais, uma vez que não fará sentido calcular a média de um dos itens e esta dar 3,3, por exemplo. Isto apenas nos indicaria que as respostas a esse item estão entre ‘Algumas vezes’ e ‘Muitas vezes’. Considera-se que estes dados não acrescentariam qualquer informação útil.

Além disso, foi efetuada uma breve análise exploratória dos dados. No que diz respeito aos *outliers*, procurou-se criar um instrumento com variáveis ordinais (e algumas construídas a partir de escala de *Likert*) propositadamente, para que não houvesse a possibilidade de haver *outliers*. Foi ainda possível verificar que algumas das questões do questionário possuíam valores omissos, cujo tratamento é apresentado na secção seguinte.

3.2.2. Preparação dos dados

A fase de Preparação dos dados requer a tomada de decisões acerca do *dataset* final, que serve como base para a fase de modelação. Sendo assim, deve-se proceder à seleção, limpeza, transformação e integração dos dados, caso seja necessário. Nesta fase devem também ser efetuadas algumas análises aos dados obtidos, entre elas: averiguar a existência de informação incompleta, *outliers*, atributos pouco importantes ou irrelevantes, demasiadas classes nas variáveis ou distribuições enviesadas e não uniformes. No final desta etapa, o *dataset* encontrar-se-á preparado para a modelação (no entanto poderá ser necessário regressar a esta etapa para efetuar alterações no sentido de melhorar a performance do modelo).

Uma vez que os dados foram obtidos com base num questionário construído especificamente para este estudo, não foi necessário criar, agregar ou eliminar variáveis, uma vez que todas estão de acordo com o desejado e são relevantes. No entanto, foi necessária uma seleção das instâncias do *dataset*, tendo sido eliminados alguns registos por uma questão de qualidade dos dados. Estas eliminações encontram-se apresentadas de seguida.

Dos 833 indivíduos que participaram nesta investigação, 236 responderam ‘Não’ à primeira questão do questionário, ficando automaticamente excluídos do estudo (não responderam às restantes questões do mesmo). Sendo assim, apenas 597 respostas foram consideradas, a priori, como válidas. No entanto, foi necessário analisar os dados e verificar se todos eles eram válidos.

Apenas alguns inquiridos deveriam ter respondido à questão 6 da versão final do questionário, consoante a sua resposta à questão anterior. Verificou-se que alguns deles não responderam por lapso (valores omissos). Por outro lado, alguns inquiridos responderam a esta questão quando não deveriam. Perante o sucedido, há duas opções que podem ser tomadas. A primeira é manter as respostas desses inquiridos e considerar que foi apenas um lapso. A segunda, e regra habitual, é remover a sua participação total do estudo pois assume-se que, se não respondeu quando devia ou respondeu quando não devia, foi porque não estava a responder ao questionário com total atenção e as suas respostas poderão enviesar os resultados. Desta forma, optou-se pela segunda opção, tendo sido removidas as respostas a todas as questões dos dez inquiridos que responderam à questão 6 quando não deviam e dos 19 inquiridos que não responderam à questão 6 quando deviam, restando 568 respostas.

Verificou-se também que 18 inquiridos residem no estrangeiro. Uma vez que este estudo se aplica apenas a Portugal e que noutros países, para além de poderem ser utilizadas casas de apostas distintas, os comportamentos podem também ser distintos, optou-se por remover as respostas destes inquiridos, restando assim 550 respostas.

Por fim, foram retiradas as 30 respostas utilizadas na validação final. Foram assim retiradas 47 respostas que aparentavam ser válidas (para além das 30 utilizadas na validação final), no entanto, considerou-se ser necessário para que estas respostas não influenciassem os resultados do estudo.

O *dataset* final é assim composto por um total de 520 registos (observações) e um total de 46 variáveis referentes aos comportamentos dos apostadores.

3.2.2.1. *Análise descritiva*

Esta secção apresenta os resultados após a remoção das respostas não válidas, sendo o número total de respostas igual a 520.

3.2.2.1.1. *Caraterização demográfica dos inquiridos*

Como é possível verificar através da análise da tabela apresentada no Apêndice G (e das figuras apresentadas abaixo), a maioria dos inquiridos é do sexo masculino (93,3%), tem entre os 18 e os 24 anos (29,8%), possui como habilitação literária ‘Ensino secundário (12.º ano) ou Curso profissional’ (45,2%), trabalha por conta de outrem (67,3%) e reside no sul de Portugal (50,8%).

Analisando mais detalhadamente a tabela mencionada e a Figura 3.1, no que diz respeito ao escalão etário, é possível constatar que mais de metade dos inquiridos (56,2%) tem entre 18 e 34 anos. Como apresentado na secção 2.2.3.1., em 2018, os jogadores com idades compreendidas entre os 18 e os 34 anos representavam mais de metade do total de jogadores registados. Sendo assim, é perceptível que a maioria dos inquiridos neste estudo tenha entre 18 e 34 anos, visto corresponder à maioria dos apostadores portugueses. No entanto, também há um elevado número de inquiridos que possui idades compreendidas entre os 35 e os 44 anos (26,2%).

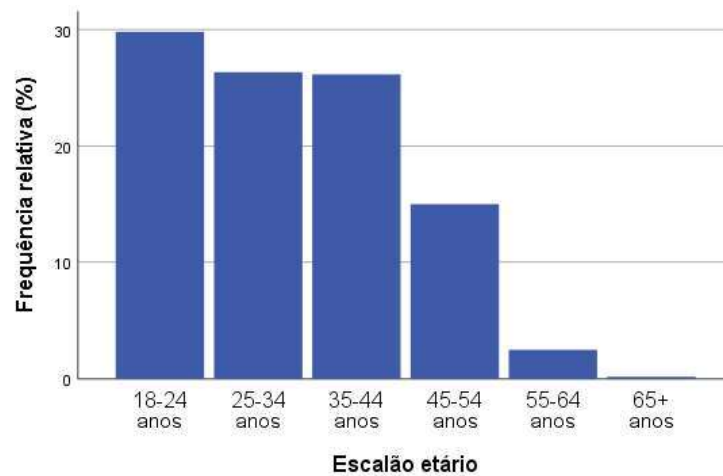


Figura 3.1: Escalão etário dos inquiridos (Frequências relativas)

Como foi referido acima e como é possível verificar na Figura 3.2, a maioria dos inquiridos possui como habilitação literária o ‘Ensino secundário (12.º ano) ou Curso profissional’. A esta classe segue-se a ‘Licenciatura’ com 23,5% e o ‘3.º ciclo do ensino básico (9.º ano)’ com 17,5%. É também possível verificar que mais de metade dos inquiridos (66,9%) possui uma habilitação literária igual ou inferior a ‘Ensino secundário (12.º ano) ou Curso profissional’.

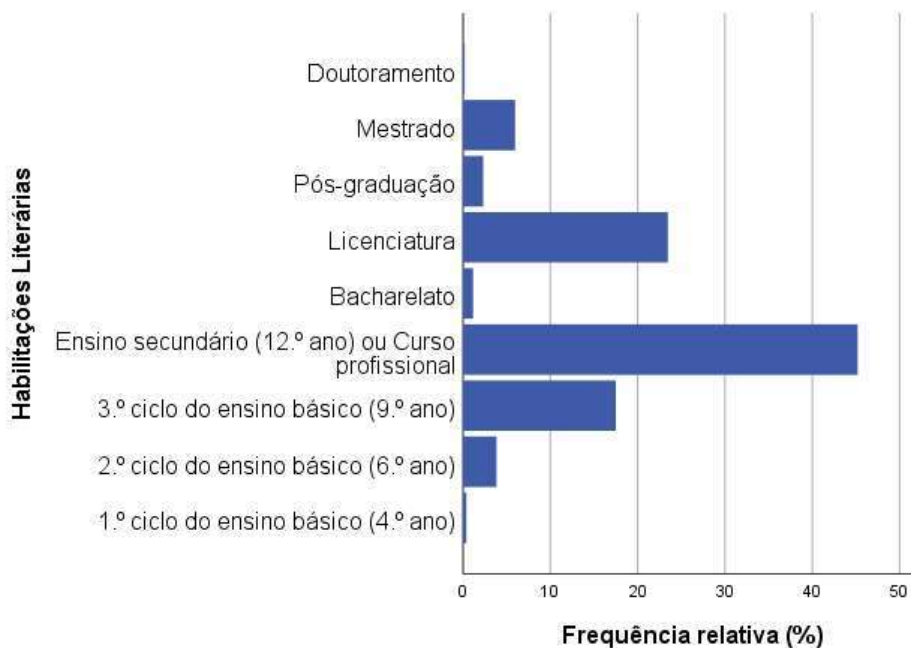


Figura 3.2: Habilitações literárias dos inquiridos (Frequências relativas)

No que diz respeito à ocupação, constata-se que a grande maioria dos inquiridos trabalha por conta de outrem. A segunda maior ocupação é ser ‘Estudante’ e corresponde a apenas 15,8%. Note-se que, nesta variável, o total é superior ao número de respostas pois alguns inquiridos selecionaram mais do que uma opção na questão 5 referente à Ocupação. Há, portanto, 18 inquiridos a estar em duas situações distintas, como por exemplo Estudante e Empregado por conta de outrem.

Relativamente ao setor de atividade, note-se que 109 inquiridos (21%) não responderam à questão 6 que se refere ao setor de atividade, por não serem trabalhadores por conta de outrem ou por conta própria (sobrando assim 411 inquiridos). No entanto, é possível verificar na tabela apresentada no Apêndice G que 419 inquiridos trabalham por conta de outrem ou por conta própria. Isto acontece porque há oito inquiridos que trabalham por conta de outrem e por conta própria, em simultâneo. Para além destas informações, não é possível retirar outras ilações, uma vez que a maioria dos inquiridos (23,3%) trabalha em setores não apresentados nas opções de resposta da questão 6. Os restantes inquiridos encontram-se espalhados pelos vários setores apresentados, sendo que a maioria deles trabalha no setor do ‘Comércio por grosso e a retalho’ (10,2%).

No que diz respeito à região de residência, procurou-se recolher dados das várias regiões do país, por forma a tentar evitar algum enviesamento cultural devido a possíveis hábitos e comportamentos específicos de uma só região. Como é possível verificar pela análise da tabela mencionada e da Figura 3.3, a maioria dos inquiridos reside no sul de Portugal (50,8%), seguindo-se o centro com 23,7% e o norte com 19,8%. As ilhas correspondem a apenas 5,8% dos inquiridos.

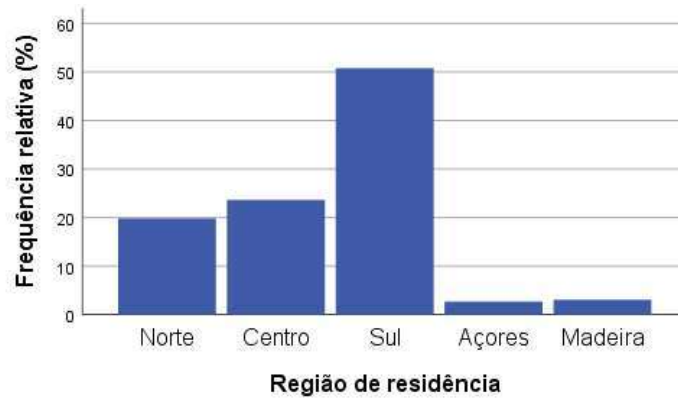


Figura 3.3: Regiões de residência dos inquiridos (Frequências relativas)

Relativamente ao distrito de residência, e através da análise da tabela apresentada no Apêndice G, considera-se que a amostra poderia estar mais bem distribuída, uma vez que a maioria dos inquiridos reside em Lisboa (35,8%), seguindo-se o Porto como segundo distrito onde residem 12,9% dos inquiridos. No entanto, uma vez que Lisboa é o distrito com maior número de habitantes, é perceptível que a maioria dos inquiridos neste estudo residam em Lisboa. É possível verificar que 30 inquiridos não responderam às questões referentes aos distritos de residência. Isto deve-se ao facto de estes inquiridos residirem nas ilhas (Madeira e Açores).

É apresentada na tabela do Apêndice G a frequência relativa acumulada apenas para as duas variáveis qualitativas ordinais (Escala etária e Habilitações literárias).

3.2.2.1.2. Caracterização dos comportamentos dos inquiridos

Como foi referido na secção 3.1.2.1., na questão 11 da versão final do questionário, e por sua vez na variável referente às casas de apostas, as casas apresentadas são as únicas com licença para a exploração de apostas desportivas à cota *online* em Portugal. Sendo assim, e através da análise da tabela apresentada no Apêndice H, é possível verificar que 22,1% dos inquiridos aposta em casas de apostas não legalizadas em Portugal. Para além disso, constata-se também que a maioria dos inquiridos efetua apostas na casa de apostas ‘Betclíc’ (27,7%), seguindo-se o ‘Placard.pt’ com 24%.

Em relação à frequência com que os inquiridos apostam, é possível comprovar que mais de metade dos inquiridos efetua apostas desportivas ‘Diariamente’ ou ‘Mais que uma vez por semana’ (67,9%).

No que diz respeito ao número de apostas efetuadas, é possível verificar que mais de metade dos inquiridos efetua ‘Até 4 apostas’ (53,1%).

Relativamente à modalidade de aposta, é possível constatar que a modalidade ‘Combinada (entre um e três prognósticos)’ é a modalidade de aposta mais escolhida (89,8%), isto é, 89,8% dos inquiridos efetua apostas combinadas (entre um e três prognósticos), podendo parte deles efetuar também apostas segundo outras modalidades. Por outro lado, a modalidade ‘Múltipla’ é a menos escolhida pelos inquiridos (52,5%).

É também possível verificar na tabela do apêndice mencionado que o sistema ‘*Stake por Níveis*’ é o sistema de apostas desportivas mais escolhido (76,9%), isto é, 76,9% dos inquiridos efetua apostas utilizando o sistema ‘*Stake por Níveis*’, podendo parte deles efetuar também apostas com recurso a outros sistemas. Por outro lado, o sistema ‘*Martingale*’ é o menos utilizado pelos inquiridos (34,2%).

Em relação ao tipo de aposta, é possível comprovar que a aposta ‘1 X 2’ é a aposta mais escolhida (92,3%), isto é, 92,3% dos inquiridos efetua apostas do tipo ‘1 X 2’, podendo parte deles utilizar também outro tipo de aposta nas suas apostas desportivas. Por outro lado, o tipo de aposta ‘Quem marca o X.º ensaio’ é o menos escolhido pelos inquiridos (20,4%).

No que diz respeito à obtenção de lucro com o tipo de aposta mais utilizado, 45,6% dos inquiridos afirmam que sim, sendo que 18,1% dos inquiridos não tem a perceção se consegue ou não obter lucro com o tipo de aposta mais utilizado.

Relativamente à modalidade desportiva, é possível constatar que o ‘Futebol’, como seria de esperar, é a modalidade mais escolhida (99,0%), isto é, 99,0% dos inquiridos efetua apostas em jogos de futebol, podendo parte deles apostar também noutras modalidades. Por outro lado, o ‘Ténis de mesa’ é a modalidade desportiva menos escolhida pelos inquiridos (14,6%).

É também possível verificar na tabela do apêndice mencionado que apostar ‘Apenas nos jogos do próprio dia e antes do início dos encontros desportivos’ é a estratégia mais utilizada (94,6%), isto é, 94,6% dos inquiridos aposta apenas nos jogos do próprio dia e antes destes começarem podendo parte deles utilizar também outra estratégia. Por outro lado, ‘Apostar em jogos de dias diferentes’ é a estratégia menos utilizada pelos inquiridos (82,3%).

Em relação à frequência com que os inquiridos apostam nas mesmas equipas, 44,4% afirma que aposta ‘Muitas vezes’, sendo que apenas 0,8% afirma que nunca o faz.

No que diz respeito à frequência com que os inquiridos colocam os habituais vencedores das competições nas suas apostas, 46,7% afirma que o faz ‘Muitas vezes’, sendo que 6,0% afirma que nunca o faz.

Relativamente à frequência com que os inquiridos utilizam algum *site* ou aplicação como apoio, 32,3% afirma que o faz ‘Algumas vezes’, sendo que 9,6% afirma que nunca o faz.

No que concerne à frequência com que os inquiridos obtêm lucro nas suas apostas desportivas, e mais uma vez de acordo com a tabela apresentada no Apêndice H, 53,5% afirma que consegue obter lucro ‘Algumas vezes’, sendo que 2,5% afirma que nunca consegue obter lucro.

Em relação ao dinheiro gasto em apostas desportivas, como é possível verificar pela análise da Figura 3.4, mais de metade dos inquiridos gasta até 99€ (60,4%), sendo que a maioria gasta ‘Até 49€’ (33,5%).

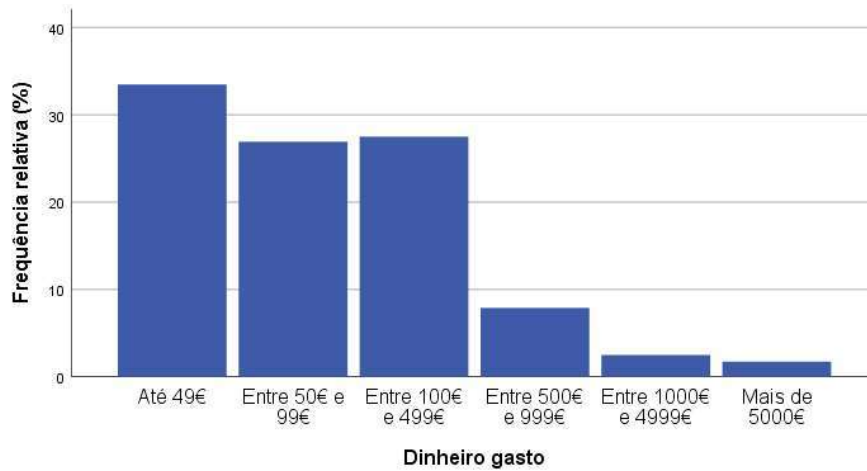


Figura 3.4: Dinheiro gasto pelos inquiridos (Frequências relativas)

Por fim, no que diz respeito ao lucro obtido, e de acordo com a Figura 3.5, mais de metade dos inquiridos consegue obter até 499€ de lucro (75%), sendo que a maioria consegue obter ‘Até 49€’ de lucro (27,7%).

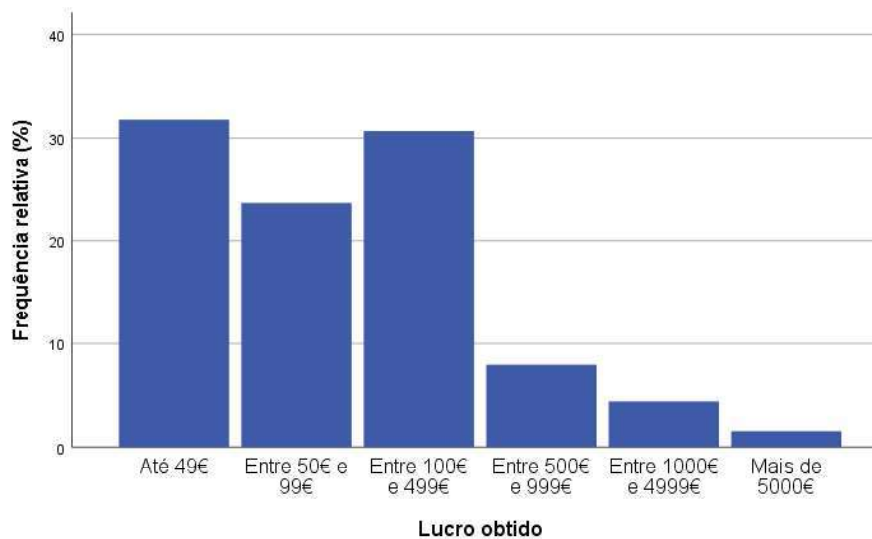


Figura 3.5: Lucro obtido pelos inquiridos (Frequências relativas)

3.2.2.2. Verificação das correlações entre as variáveis

Previamente à construção dos modelos preditivos, é importante verificar a correlação entre todas as variáveis pois caso existam duas variáveis altamente correlacionadas, não devem entrar ambas no modelo (Guyon & Elisseeff, 2003; RekhaMolala, 2019). Para que se possa decidir qual o teste estatístico a utilizar neste sentido, é necessário identificar o tipo das variáveis. Como foi referido na secção 3.2.1., todas as variáveis a considerar na construção dos modelos preditivos foram classificadas como ordinais. De acordo com Marôco (2018), a correlação de *Spearman* (teste estatístico não paramétrico) é a mais indicada quando os dados violam as suposições paramétricas, tais como dados não normais, ou quando

os dados são ordinais. Sendo assim, nesta fase, foi utilizado o coeficiente de correlação de *Spearman* para verificar a correlação entre as variáveis. Foi utilizada como referência a classificação da correlação proposta por Bisquerra et al. (2004), apresentada na Tabela 3.7.

Tabela 3.7: Valores de referência do Coeficiente de correlação

Correlação	Valor do Coeficiente de correlação
Nula	0,00
Positiva muito baixa] 0,00 ; 0,20]
Positiva baixa] 0,20 ; 0,40]
Positiva moderada] 0,40 ; 0,60]
Positiva alta] 0,60 ; 0,80]
Positiva muito alta] 0,80 ; 1,00 [
Positiva perfeita	1,00

Fonte: Bisquera et al. (2004)

As mesmas interpretações serão utilizadas para os coeficientes negativos.

Para efetuar esta verificação, foram definidas as seguintes hipóteses:

- ✓ Hipótese nula (H0): Não existe uma relação entre as duas variáveis (uma variável não influencia a outra);
- ✓ Hipótese alternativa (H1): Existe relação entre as duas variáveis (uma variável influencia a outra).

Nas tabelas apresentadas nos Apêndices I, J, K, L e M, é possível encontrar os coeficientes de correlação e os valores de significância. Estes valores foram divididos em várias tabelas para simplificar a sua leitura.

Para analisar a correlação entre as várias variáveis, é necessário analisar tanto os coeficientes de correlação como os valores de significância. Começando pelo valor de significância (p-value), este indica-nos se se deve aceitar ou rejeitar a hipótese nula. Para isso, é necessário comparar este valor com o nível de significância. O nível de significância deve ser definido pelo investigador, sendo que, neste caso, foi definido como 0,05. Como é possível verificar nas cinco tabelas dos apêndices mencionados, em alguns casos o valor de significância é menor que 0,05 (p-value < 0,05 com um nível de confiança de 95%), o que nos leva a concluir que existe uma relação estatisticamente significativa entre as duas variáveis (rejeita-se H0). No entanto, para compreender se esta relação é alta ou baixa, é necessário analisar o coeficiente de correlação. Verifica-se, assim, que, apesar de p-value < 0,05 em alguns casos, a correlação nunca é muito alta (acima de 0,8 em módulo). Sendo assim, não é necessário remover nenhuma variável. As únicas células onde a correlação é superior a 0,8 (e igual a 1,00 (correlação perfeita)) são as que comparam a mesma variável.

3.2.3. Modelação

Esta é a fase na qual se constrói os modelos com base no algoritmo definido. Normalmente, existem várias técnicas de modelação para o mesmo tipo de problema de *data mining*, sendo que, neste estudo, como já foi referido, optou-se pelas árvores de decisão. As árvores de decisão são algoritmos que pretendem modelar uma variável dependente a partir de um conjunto de variáveis explicativas (independentes) que vão sendo divididas sequencialmente dando origem a ramos que melhor discriminam as classes de um problema. Apesar de não ser uma técnica que produza os melhores resultados a nível preditivo, optou-se por utilizar esta técnica por poder ser diretamente visualizada e interpretada pelo ser humano, sendo fácil de identificar as variáveis mais importantes para a discriminação das classes. Para além disso, permite identificar um conjunto de regras interpretáveis relacionadas com estas variáveis e é uma técnica não paramétrica, ou seja, não necessita que as variáveis estejam distribuídas de acordo com uma distribuição específica (Nguyen, 2020).

Para isso, foi decidido proceder à partição do *dataset* utilizando a técnica *Holdout*, onde se dividiu o mesmo em conjuntos mutuamente exclusivos. Esta técnica consiste na separação do *dataset* em conjuntos de dados de treino e de teste, sendo que se constrói o modelo no primeiro conjunto e se testa a sua qualidade no segundo. O *dataset* pode ou não ser separado em quantidades iguais. O mais comum é considerar 2/3 ou 80% dos dados para treino e o 1/3 ou 20% restante para teste (Monteiro et al., 2016; Nurhayati et al., 2014). Foram testadas as duas opções, para que fosse possível compará-las.

Existem casos em que uma árvore de decisão pode ficar com uma quantidade de arestas muito grande, aumentando a sua complexidade. Isto pode gerar um problema conhecido como *overfitting* e que é bastante usual quando se trata de árvores de decisão. O *overfitting* ocorre quando um modelo estatístico se ajusta muito bem ao *dataset* através do qual foi criado, mas se mostra ineficaz para prever novos resultados (Bramer, 2007; Dietterich, 1995). Para resolver este tipo de situação, podem ser efetuados dois procedimentos distintos. O primeiro consiste em deixar a árvore crescer totalmente e depois efetuar a poda da mesma e o segundo consiste em ajustar os parâmetros do modelo, na expectativa de se conseguir obter um melhor desempenho. Neste estudo, foi então utilizado o segundo método. Os parâmetros que foram ajustados nos vários modelos foram o *minsplit* (número mínimo de casos num nó antes da divisão), o *maxdepth* (profundidade máxima da árvore), o *minbucket* (número mínimo de casos em cada nó final, ou seja, nas folhas) e o *cp* (para cada nível de profundidade da árvore é calculado o *cp*. Se este valor for inferior ao valor definido como *threshold*, os nós nesse nível não são mais divididos). Como os valores destes parâmetros não podem ser obtidos diretamente através da análise dos dados, são chamados de hiperparâmetros. Foi utilizada a abordagem *random search* para ajustar os hiperparâmetros dos vários modelos.

Para cada método de partição e *dataset*, foram efetuadas duas modelações, uma primeira em que o modelo foi desenvolvido com os hiperparâmetros *default* e uma segunda em que estes foram ajustados para valores ideais, na tentativa de afinar o modelo. Desta forma, foi possível comparar as várias modelações (com e sem ajuste dos parâmetros) e ver qual a que permitia obter um melhor desempenho.

Foram assim desenvolvidas dezenas de iterações, tendo-se completado apenas 16, correspondentes a 16 modelos⁵.

Nas duas primeiras iterações foi utilizado o método de partição *Holdout* 80-20, no entanto, na segunda procedeu-se ao ajuste dos hiperparâmetros do modelo. O *dataset* para estas duas iterações foi o apresentado na secção 3.2.2. (520 registos e 46 variáveis).

Nas iterações 3 e 4 foi utilizado o método de partição *Holdout* 2/3, no entanto, na quarta iteração procedeu-se ao ajuste dos hiperparâmetros do modelo. O *dataset* para estas duas iterações foi o apresentado na secção 3.2.2. (520 registos e 46 variáveis).

Após as primeiras quatro iterações, procedeu-se a uma pequena revisão da variável de saída (Lucro). Uma vez que esta variável não se encontrava balanceada, optou-se por agrupar algumas classes da mesma, com o objetivo de a tornar mais balanceada (regressando-se assim à fase de Preparação dos dados). Foi definido um *threshold* de 20%, ou seja, se uma classe tiver mais de 20% das observações então mantém-se, caso contrário é agrupada com outras classes contíguas, se possível. Tendo isto em conta, foram agrupadas as últimas três classes da variável Lucro ('Entre 500€ e 999€', 'Entre 1000€ e 4999€' e 'Mais de 5000€'), dando origem a uma única classe com o nome 'Mais de 500€'. Na Tabela 3.8, é possível observar o número de observações e a frequência relativa de cada classe da variável Lucro, antes e depois da transformação mencionada.

Tabela 3.8: Número de observações e frequências relativas pré e pós transformação

Classe	Pré transformação		Pós transformação	
	Número de observações	Frequência relativa (%)	Número de observações	Frequência relativa (%)
Até 49€	144	27,7	144	27,7
Entre 50€ e 99€	107	20,6	107	20,6
Entre 100€ e 499€	139	26,7	139	26,7
Entre 500€ e 999€	36	6,9	-	-
Entre 1000€ e 4999€	20	3,8	-	-
Mais de 5000€	7	1,3	-	-
Não respondido	67	12,9	67	12,9
Mais de 500€	-	-	63	12,1
Total	520	100	520	100

Tendo isto em conta, nas iterações 5, 6, 7 e 8 foi utilizado este novo *dataset*, no qual se mantiveram os 520 registos e as 46 variáveis, no entanto, com a variável de saída a sofrer as alterações mencionadas. Nas iterações 5 e 6 foi utilizado o método de partição *Holdout* 80-20, no entanto, na sexta iteração procedeu-se ao ajuste dos hiperparâmetros do modelo. Nas iterações 7 e 8 foi utilizado o método de partição *Holdout* 2/3, no entanto, na oitava iteração procedeu-se ao ajuste dos hiperparâmetros do modelo.

⁵ Apenas serão consideradas as 16 iterações que foram totalmente completadas.

Após a oitava iteração, foram detetadas algumas anomalias no *dataset* que necessitaram de ser revistas. Apenas os inquiridos que obtiveram lucro nos últimos três meses deveriam ter respondido à questão 25. Desta forma, esta questão deveria ter apenas 13 valores omissos, que corresponde ao número de inquiridos que responderam ‘Nunca’ na questão 23, ou seja, que nunca têm lucro nas suas apostas desportivas. No entanto, verificou-se que a variável Lucro continha 67 valores omissos. Ou seja, 54 inquiridos deveriam ter respondido à questão 25 e não o fizeram. Existe a possibilidade de não o terem feito porque não sabiam bem o lucro obtido ou simplesmente por lapso. Seguindo a mesma lógica apresentada na secção 3.2.2., optou-se por remover estes registos do *dataset* inicial (regressando-se assim mais uma vez à fase de Preparação dos dados).

Tendo isto em conta, nas iterações 9, 10, 11 e 12 foi utilizado este novo *dataset*, no qual se mantiveram as 46 variáveis (tendo a variável Lucro novamente as sete classes iniciais) mas com apenas 466 registos. Nas iterações 9 e 10 foi utilizado o método de partição *Holdout* 80-20, no entanto, na décima iteração procedeu-se ao ajuste dos hiperparâmetros do modelo. Nas iterações 11 e 12 foi utilizado o método de partição *Holdout* 2/3, no entanto, na décima segunda iteração procedeu-se ao ajuste dos hiperparâmetros do modelo.

Nas últimas quatro iterações procedeu-se à junção das duas alterações mencionadas acima, ou seja, criou-se um novo *dataset* sem os 54 registos errados e com a variável lucro com as três classes agrupadas. Nas iterações 13 e 14 foi utilizado o método de partição *Holdout* 80-20, no entanto, na décima quarta iteração procedeu-se ao ajuste dos hiperparâmetros do modelo. Nas iterações 15 e 16 foi utilizado o método de partição *Holdout* 2/3, no entanto, na décima sexta iteração procedeu-se ao ajuste dos hiperparâmetros do modelo.

Para efetuar estas modelações, foi utilizado o package ‘*rpart*’ que fornece a função *rpart* (*Recursive Partitioning and Regression Trees*) (Therneau & Atkinson, 2019). A árvore de decisão do modelo com melhor performance (apresentada no capítulo seguinte) foi construída através da função *rpart.plot*.

4. Análise e discussão dos resultados

O objetivo deste capítulo é apresentar a análise dos resultados obtidos, de acordo com a metodologia definida no capítulo anterior.

4.1. Avaliação

É nesta fase onde se efetua a avaliação dos modelos desenvolvidos e se verifica se o modelo final permitiu atingir os objetivos da forma mais adequada. Existem várias métricas para avaliar o desempenho de um modelo preditivo com uma variável de saída binária, entre elas a *accuracy*⁶, especificidade, sensibilidade, curva ROC, área abaixo da curva ROC, entre outras. No entanto, em problemas de classificação multi-classe como é o caso deste estudo, existem algumas métricas que não se adequam. De acordo com Brownlee (2016), a *accuracy* é mais útil em problemas de classificação binária do que em problemas de classificação multi-classe porque pode não ser claro como é que a *accuracy* se divide entre as classes da variável de saída. Ou seja, é possível obter uma *accuracy* alta, mas não se sabe se é por todas as classes estarem a ser igualmente bem previstas ou se algumas classes estão a ser negligenciadas pelo modelo. Sendo assim, devem ser utilizadas métricas que avaliem cada classe em separado.

De acordo com Bleik e Gauher (2016), dois *data scientists* da Microsoft, para avaliar o desempenho de um modelo em relação a cada classe, é comum calcular-se a *precision*, *recall* e *F1-score*. Estas métricas são particularmente úteis quando as classes não estão balanceadas. Nesse caso, a *accuracy* pode ser enganadora, uma vez que o modelo pode estar a prever muito bem as observações de uma classe, obtendo-se assim um valor muito alto, mas apresentar uma *precision* ou *recall* muito baixo para as outras classes. Posto isto, são aconselháveis outras métricas para além da *accuracy*, nomeadamente a *precision*, *recall*, *F1-Score*, entre outras. Estas métricas podem ser calculadas a partir da matriz de confusão. Segundo Shmueli (2019), codiretor do laboratório de pesquisa do Instituto Rigsum, a métrica *F1-Score* (ou qualquer uma das suas variantes) deve ser utilizada com muito cuidado pois atribui pesos iguais à *precision* e ao *recall*, o que não é correto. Sendo assim, neste estudo foi utilizada a métrica *accuracy* para avaliar o modelo no global mas também as métricas *precision* e *recall* para avaliar cada classe em separado. Note-se que o valor das três métricas mencionadas deve ser o mais próximo possível de 100%.

A metodologia CRISP-DM, como já foi referido, é iterativa, sendo o resultado final o culminar de um conjunto de iterações através das quais os modelos vão sendo aprimorados até se obter o melhor modelo. Sendo assim, apesar de terem sido completados 16 modelos preditivos diferentes, nesta secção, que no âmbito do CRISP-DM se poderia designar como sendo a fase de *Evaluation*, apenas se apresentam os resultados do último modelo efetuado (o modelo desenvolvido com base no *dataset* sem

⁶ Normalmente é utilizado o termo *accuracy* para se referir à exatidão e o termo *precision* para se referir à precisão. Sendo assim, e para que seja mais claro, serão utilizados os termos em inglês. As métricas *recall* e *F1-score* serão também mencionadas em inglês.

os 54 registos errados, onde se agrupou três classes da variável de saída e onde se utilizou o método de partição *Holdout 2/3*). Este modelo foi o que revelou um melhor poder preditivo, pelo que não se justifica apresentar as métricas obtidas nos restantes modelos, uma vez que tornaria esta secção demasiado exaustiva.

4.1.1. Matriz de confusão e *accuracy* do modelo

Para avaliar o modelo final segundo as três métricas mencionadas, é necessário analisar a matriz de confusão do modelo (Tabela 4.1). Esta matriz apresenta o número de observações (do conjunto de dados de teste) que o modelo conseguiu prever corretamente ou incorretamente, consoante as classes da variável final (Lucro). A diagonal apresenta as previsões corretas e os restantes valores dizem respeito às previsões incorretas.

Tabela 4.1: Matriz de confusão

		Previsto				
		Até 49€	Entre 50€ e 99€	Entre 100€ e 499€	Mais de 500€	Não respondido
Real	Até 49€	33	7	7	1	0
	Entre 50€ e 99€	10	9	15	2	0
	Entre 100€ e 499€	6	10	30	0	0
	Mais de 500€	0	0	19	2	0
	Não respondido	2	0	1	1	0

Através da matriz apresentada, é possível obter a *accuracy* do modelo, dividindo o número de previsões corretas (diagonal) pelo número total de previsões (todos os valores da matriz). Desta forma, é possível verificar que a *accuracy* deste modelo é de 47,7%, ou seja, o modelo conseguiu prever corretamente a classe de 47,7% das observações (do conjunto de dados de teste).

Através de uma tabela semelhante à Tabela 4.1, foi possível obter a *accuracy* do modelo para o conjunto de dados de treino. Este valor foi de 56,3%.

Analisando a *accuracy* do modelo tanto para o conjunto de dados de treino como de teste, pode-se considerar que neste modelo não ocorreu *overfitting* (ao contrário de alguns dos modelos desenvolvidos anteriormente). Caso tivesse ocorrido *overfitting*, o valor da *accuracy* para o conjunto de dados de treino seria bastante elevado e bem maior do que o valor da *accuracy* para o conjunto de dados de teste (pelos motivos explicados na secção 3.2.3.), o que não se verifica. No entanto, o valor para o conjunto de dados de treino é logicamente superior ao valor para o conjunto de dados de teste, uma vez que o modelo foi criado tendo como base os dados de treino. Sendo assim, de todas as previsões efetuadas, o modelo conseguiu prever corretamente apenas cerca de 48%. A razão para este valor ser relativamente baixo é explicada de seguida.

Analisando a Tabela 4.1 mais pormenorizadamente, é possível obter algumas informações importantes relativamente às previsões efetuadas. Somando cada linha da matriz, obtém-se o número de

observações por classe no conjunto de dados de teste. Estes valores são apresentados na Tabela 4.2. Das 48 observações da classe ‘Até 49€’ (ver Tabela 4.2), o modelo conseguiu prever corretamente 33 (68,8%). Das 36 observações da classe ‘Entre 50€ e 99€’ (ver Tabela 4.2), o modelo conseguiu prever corretamente apenas 9 (25,0%). Das 46 observações da classe ‘Entre 100€ e 499€’ (ver Tabela 4.2), o modelo conseguiu prever corretamente 30 (65,2%). E das 21 observações da classe ‘Mais de 500€’ (ver Tabela 4.2), o modelo conseguiu prever corretamente apenas 2 (9,52%).

Os registos em que não houve resposta por parte dos inquiridos à questão do Lucro foram mantidos para se tentar perceber se existe algum padrão para não se conseguir obter lucro. No entanto, uma vez que havia poucas observações desta classe (apenas 9 registos para treino (ver Tabela 4.2)), o modelo não conseguiu identificar nenhum padrão, não conseguindo por sua vez efetuar nenhuma previsão correta para esta classe chamada ‘Não respondido’ (das 4 observações, o modelo não conseguiu prever corretamente nenhuma).

É assim possível perceber que o modelo conseguiu prever relativamente bem as observações das classes ‘Até 49€’ e ‘Entre 100€ e 499€’. No entanto, para as restantes classes, este já não funcionou tão bem. Estes valores são examinados ao pormenor de seguida, através da análise das métricas *precision* e *recall*.

Tabela 4.2: Número de observações e frequências relativas por classe

Classe	Teste		Treino		Total	
	Número de observações	Frequência relativa (%)	Número de observações	Frequência relativa (%)	Número de observações	Frequência relativa (%)
Até 49€	48	30,9	96	30,9	144	30,9
Entre 50€ e 99€	36	23,2	71	22,8	107	22,9
Entre 100€ e 499€	46	29,7	93	29,9	139	29,8
Mais de 500€	21	13,5	42	13,5	63	13,5
Não respondido	4	2,6	9	2,9	13	2,8
Total	155	100	311	100	466	100

4.1.2. *Precision* e *recall* de cada uma das classes

Como foi referido anteriormente, num problema de classificação multi-classe a análise da *accuracy* pode não ser suficiente. Desta forma, foram também analisadas a *precision* e o *recall* de cada uma das classes da variável de saída. Estes valores apresentam-se na Tabela 4.3.

Tabela 4.3: *Precision e recall das classes da variável Lucro no dataset de teste*

Classe	<i>Precision (%)</i>	<i>Recall (%)</i>
Até 49€	64,7	68,8
Entre 50€ e 99€	34,6	25,0
Entre 100€ e 499€	41,7	65,2
Mais de 500€	33,3	9,5
Não respondido	NaN ⁷	0,0

A título de exemplo, e no que diz respeito à classe ‘Até 49€’, a *precision* corresponde ao número de previsões ‘Até 49€’ corretas (33) a dividir pelo número de previsões ‘Até 49€’ totais (51)⁸. Apenas 64,7% das observações que o modelo classifica como ‘Até 49€’ são realmente ‘Até 49€’, ou seja, que foram corretamente previstas. Por sua vez, e de acordo com a Tabela 4.3, o *recall* corresponde ao número de previsões ‘Até 49€’ corretas (33) a dividir pelo número real de observações ‘Até 49€’ (48). Apenas 68,8% das observações que são realmente ‘Até 49€’ foram classificadas pelo modelo como ‘Até 49€’. A mesma análise pode ser efetuada para as restantes classes.

É, portanto, possível perceber que em algumas classes os valores das métricas são mais elevados. Isto acontece porque a variável de saída (Lucro) não é balanceada. Como é possível observar na Tabela 4.2, as classes ‘Até 49€’ e ‘Entre 100€ e 499€’ são as que possuem mais observações. Desta forma, foram obtidas métricas melhores nestas classes uma vez que, quando o modelo foi construído, foi possível analisar mais casos das mesmas.

Sobressai também da Tabela 4.3 o valor bastante baixo do *recall* da classe ‘Mais de 500€’. Através da análise da matriz de confusão (Tabela 4.1), é possível verificar que o modelo previu a maioria dos registos desta classe como ‘Entre 100€ e 499€’, daí o valor do *recall* ser mais baixo. Como foi referido aquando da análise da matriz de confusão, o modelo não conseguiu prever corretamente nenhuma observação da classe ‘Não respondido’.

Após todas estas análises efetuadas, compreende-se agora a razão para a *accuracy* possuir um valor tão baixo, uma vez que o modelo prevê corretamente as observações das classes com mais registos mas não o consegue fazer para as classes com menos registos.

4.1.3. Árvore de decisão

Após o cálculo da *accuracy*, *precision* e *recall*, procedeu-se à visualização da árvore de decisão do modelo final. A árvore de decisão tem um carácter explicativo, ou seja, explica quais as variáveis mais importantes para a discriminação das classes da variável de saída (neste caso, o Lucro). Sendo assim, a árvore não tem um carácter preditivo, para isso são utilizadas as métricas já mencionadas.

⁷ NaN é o termo utilizado nas linguagens de programação para representar um valor numérico indefinido ou irrepresentável (neste caso a tabela possui este termo pois no cálculo da *precision* para a classe ‘Não respondido’, o denominador é igual a 0).

⁸ Este valor pode ser obtido através da matriz de confusão, somando todos os valores da coluna da classe ‘Até 49€’.

Foi utilizado o coeficiente de Gini para dividir os nós, ou seja, para escolher quais os atributos que iriam servir para fazer crescer a árvore. O coeficiente de Gini mede a pureza de cada um dos nós da árvore de decisão. Sendo assim, a árvore faz a seleção das variáveis, ficando apenas as que influenciam diretamente a variável de saída (neste caso, o Lucro).

Tal como descrito na secção anterior, para que o modelo final tivesse um poder preditivo mais elevado, foram ajustados os seus hiperparâmetros. Desta forma, a árvore acabou por ficar com poucos níveis (bastante mais simples), não sendo possível identificar muitas regras na mesma, ao contrário do que se pretendia. Ao ficar com poucos níveis, acabam por ser poucas as variáveis identificadas pelo modelo como mais preponderantes para a obtenção de lucro. Uma vez que o modelo não apresentou métricas elevadas, talvez seja arriscado afirmar que as restantes variáveis não influenciam o lucro.

Na Figura 4.1, é possível observar a árvore de decisão obtida.

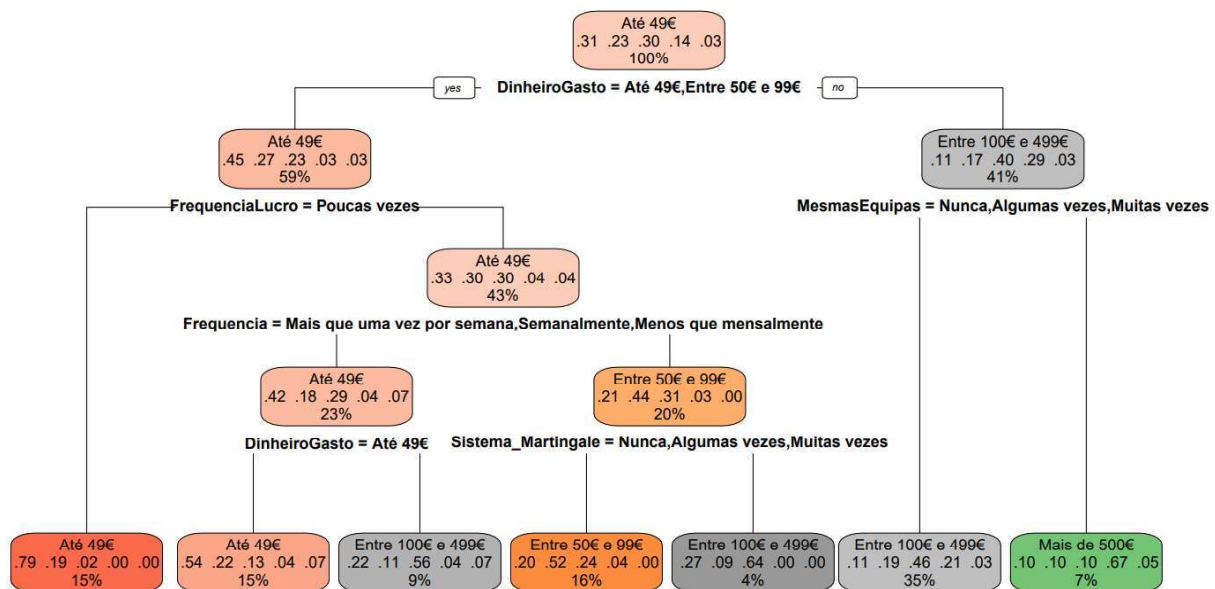


Figura 4.1: Árvore de decisão

A árvore de decisão apresenta as variáveis necessárias para discriminar as cinco classes da variável Lucro. É composta por nós (as caixas coloridas) que em conjunto formam vários caminhos até chegar às folhas (nós finais). Cada nó tem sempre dois nós filhos (tirando os nós finais). O “nome” de cada nó corresponde à classe que tem maior probabilidade. No entanto, em cada nó encontram-se as probabilidades das observações que caem nesse nó de pertencerem a cada uma das classes da variável Lucro. Por exemplo, no nó final mais à esquerda, o .79 corresponde à probabilidade das observações que caem nesse nó de pertencerem à primeira classe da variável Lucro, ou seja, ‘Até 49€’. O que se deseja é que uma das probabilidades de cada nó seja substancialmente superior a todas as outras. Por sua vez, a percentagem apresentada nos nós corresponde à percentagem de observações que caem nesses mesmos nós. Note-se que a soma de todas as percentagens dos nós finais dá 100%.

Tal como já foi referido, o modelo não conseguiu prever nenhum caso da classe ‘Não respondido’, sendo esse o motivo para esta classe não aparecer na árvore.

O primeiro nó no topo da árvore apresenta as probabilidades totais de cada classe, ou seja, mostra a proporção de apostadores que pertencem a cada classe. É possível verificar que, tal como já foi referido, as duas classes com maior número de observações são as classes ‘Até 49€’ (31%) e ‘Entre 100€ e 499€’ (30%). Continuando a analisar a árvore de cima para baixo, é possível verificar que 59% dos inquiridos gasta até 100€. Desses, a probabilidade de ganharem ‘Até 49€’ é de 45%. Esta informação é retirada do nó filho da raiz do lado esquerdo. Por sua vez, o nó final mais à esquerda revela que 15% dos inquiridos gasta até 100€ e obtém lucro ‘Poucas vezes’. Desses, a probabilidade de ganharem ‘Até 49€’ é de 79%.

Uma boa árvore de decisão permitiria identificar mais regras deste género que permitissem descobrir quais as variáveis que influenciam o lucro. No entanto, a árvore obtida não permite tirar grandes conclusões.

É, portanto, possível concluir através da análise da árvore de decisão apresentada na Figura 4.1, e talvez até pela *accuracy* global do modelo preditivo, que com estas variáveis o modelo não conseguiu identificar grandes padrões nos dados, não conseguindo, portanto, criar caminhos claros e interpretáveis. No entanto, é possível identificar algumas regras interessantes:

- ✓ Quem obtém mais lucro (‘Mais de 500€’), gasta ‘Acima de 100€’;
- ✓ Quem gasta ‘Abaixo de 49€’ e obtém lucro ‘Algumas vezes’, ‘Muitas vezes’ ou ‘Sempre’, obtém o lucro mais baixo (‘Até 49€’);
- ✓ Quem gasta ‘Entre 50€ e 99€’ e obtém lucro ‘Algumas vezes’, ‘Muitas vezes’ ou ‘Sempre’, já consegue obter um lucro considerável (‘Entre 100€ e 499€’);
- ✓ Quem gasta ‘Abaixo de 100€’ e obtém lucro ‘Poucas vezes’, obtém o lucro mais baixo (‘Até 49€’).

Estes resultados parecem indicar que a frequência com que se obtém lucro e a quantia apostada desempenham um papel importante na obtenção de lucro por parte dos apostadores.

5. Conclusões e recomendações

O objetivo deste capítulo é sintetizar as conclusões mais importantes desta investigação. Em primeiro lugar, são apresentados os resultados obtidos e algumas notas sobre os mesmos, de seguida são mencionadas as dificuldades encontradas e as limitações deste estudo e, por último, são apresentadas algumas sugestões para futuras investigações.

5.1. Principais conclusões

O mercado das apostas desportivas *online* é um mercado que tem crescido cada vez mais nos últimos anos. O desenvolvimento do presente estudo teve assim como objetivo principal a identificação dos fatores relacionados com o conjunto de decisões que um indivíduo toma relativamente a uma aposta desportiva *online* que são mais preponderantes para a obtenção de lucro por parte do mesmo. Através da concretização deste objetivo, os apostadores poderiam apostar de forma mais responsável e consciente, possibilitando um aumento do seu lucro. Desta forma, este estudo é orientado para o consumidor do mercado de apostas desportivas.

Os dados necessários para a realização deste estudo foram recolhidos através de um questionário *online*, construído especificamente para este estudo, e que foi previamente validado para garantir que poderia ser utilizado como instrumento de investigação. O questionário procurou recolher informações acerca das características demográficas dos apostadores e acerca dos seus comportamentos no que diz respeito à forma de apostar.

Para se proceder ao desenvolvimento dos modelos preditivos, foi utilizada como técnica de modelação a árvore de decisão e como metodologia o CRISP-DM, sendo que foram concluídas 16 iterações. Para cada método de partição e *dataset*, foram efetuadas duas modelações, uma primeira em que o modelo foi desenvolvido com os hiperparâmetros *default* e uma segunda em que estes foram ajustados para valores ideais. Foi possível verificar que, em todos os modelos nos quais os parâmetros foram ajustados, foram obtidas métricas mais elevadas do que nos respetivos modelos sem ajuste dos parâmetros. Foi também possível comprovar que a *accuracy* de um modelo não depende apenas do *dataset* utilizado, mas também do método de partição selecionado, tendo-se obtido valores diferentes para o método *Holdout 2/3* e para o *Holdout 80-20* (valores superiores para o *Holdout 2/3*).

O modelo no qual se conseguiu obter uma melhor *accuracy* nos dados de teste (48%) foi o modelo da última iteração, no qual se utilizou o método de partição *Holdout 2/3* e onde se procedeu ao ajuste dos hiperparâmetros do modelo. Este valor foi relativamente baixo devido ao facto de a variável de saída não ser balanceada (mesmo após terem sido agrupadas algumas classes). Verificou-se que o modelo previu corretamente as observações das classes com mais registos mas não o conseguiu fazer para as classes com menos registos. A própria *accuracy* para o conjunto de dados de treino acabou por não ser muito alta, o que revela que os dados talvez não estivessem em perfeitas condições para serem utilizados

nos modelos preditivos. Para além do motivo mencionado, podem existir outras razões para justificar o reduzido valor da *accuracy*, como por exemplo:

- ✓ As restantes variáveis do *dataset* também não serem balanceadas (grande dispersão dos valores das várias classes);
- ✓ As variáveis de entrada escolhidas talvez não tenham sido as mais indicadas para o problema em questão. Desta forma, há a possibilidade de existirem outras variáveis que podem ser mais preponderantes para a obtenção de lucro do que as consideradas neste estudo;
- ✓ Os métodos de partição do *dataset* escolhidos ou a própria técnica de modelação selecionada podem não ter sido os mais adequados para os dados recolhidos.

Em problemas de classificação binária, uma vez que a variável de saída apenas possui duas classes distintas, torna-se mais fácil de desenvolver modelos com um poder preditivo mais elevado. No entanto, em problemas de classificação multi-classe este processo é mais complexo uma vez que o modelo pode detetar padrões em algumas classes mas noutras não ter essa capacidade (que foi o que aconteceu neste estudo). Note-se que, como não foram encontrados estudos semelhantes, não é possível comparar a *accuracy* do modelo obtido com outros estudos.

Para além dos problemas já identificados, a amostra de análise deste estudo foi muito reduzida tendo em conta o número de apostadores em Portugal pelo que é importante reconhecer as limitações que isto acarreta. Ainda assim, os resultados obtidos através da árvore de decisão parecem indicar que a frequência com que se obtém lucro e a quantia apostada desempenham um papel importante na obtenção de lucro por parte dos apostadores. Verificou-se também que a árvore obtida não permitiu identificar regras mais concretas que permitam tirar melhores conclusões, como se pretendia.

É possível perceber que uma parte significativa do esforço depositado neste estudo consistiu em tarefas relacionadas com a qualidade do questionário, que passou por todo um processo de validação antes de ser utilizado, e dos dados recolhidos. No entanto, perdeu-se algum poder preditivo pelos motivos mencionados acima, acabando por não se conseguir obter o modelo desejado.

Em suma, pode-se concluir que, apesar das várias tentativas de obter um modelo com um bom desempenho, não se conseguiu atingir o objetivo proposto, ou seja, desenvolver um modelo preditivo com *accuracy* elevada que permita identificar algumas regras que revelem quais os comportamentos mais importantes para a obtenção de lucro. Apesar disso, o questionário desenvolvido ou até mesmo o *dataset*, podem ser utilizados em estudos futuros por outros investigadores da área de Tecnologias de informação ou de apostas desportivas.

Sabia-se desde o início que este estudo era ambicioso e que iria ser um desafio, pois se a previsão de lucro fosse assim tão fácil, todos os apostadores conseguiriam obter lucro. Sendo assim, e tendo em conta o trabalho desenvolvido neste estudo, conclui-se que procurar prever o lucro com base em determinados comportamentos dos apostadores é uma tarefa difícil, mas que deve continuar a ser estudada.

5.2. Limitações do estudo

Tal como em qualquer estudo, existem algumas limitações da presente investigação, entre elas:

- ✓ Alguns autores consideram que a recolha de dados através de questionários *online* pode levar à obtenção de respostas repetidas e a pouca diversidade demográfica (Gosling et al., 2004). Apesar de ainda haver alguma controvérsia em relação a este tema, por mais que se valide um questionário, há sempre a possibilidade de o mesmo não ser 100% fidedigno, o que pode ter influenciado o resultado deste estudo (exemplo: as questões podem não ser interpretadas de igual forma por todos os inquiridos). O ideal teria sido, portanto, a utilização de alguns dados reais (de algumas casas de apostas) para o desenvolvimento deste estudo;
- ✓ A seleção da amostra deste estudo teve por base uma técnica de amostragem não probabilística por conveniência. Desta forma, não é possível generalizar (com rigor estatístico) os resultados obtidos para a população;
- ✓ Os itens da segunda parte do questionário não foram todos construídos a partir de escalas iguais, o que tornou a sua análise em conjunto mais difícil;
- ✓ O questionário foi distribuído num período em que os campeonatos estavam interrompidos (devido à COVID-19), o que pode ter influenciado o resultado deste estudo pois os apostadores talvez não se lembrassem a 100% quanto gastam/ganham em média. O ideal teria sido partilhar o questionário numa altura em que os campeonatos não estivessem interrompidos, para que as respostas fossem mais fidedignas;
- ✓ Em termos da fiabilidade do instrumento utilizado, não foi possível aplicar o teste-reteste. Desta forma, o critério da estabilidade não foi satisfeito (o que teria sido uma mais valia);
- ✓ Procurou-se obter o máximo de respostas possível ao questionário para desenvolver um trabalho fundamentado, no entanto, por uma questão de tempo, o mesmo apenas ficou disponível durante três semanas. Para além disso, na fase de Preparação dos dados foram removidos alguns registos para que os dados fossem mais consistentes e coerentes. Isto levou a que a amostra deste estudo tenha sido muito reduzida tendo em conta o número de apostadores em Portugal, acabando esta por não ser representativa da população;
- ✓ A maioria das variáveis do *dataset* utilizado para construir os modelos preditivos não se encontrava balanceada, existindo uma grande dispersão dos valores das várias classes, o que pode ter influenciado o poder preditivo dos modelos;
- ✓ A quase inexistência de estudos semelhantes a este, tornou possível a realização desta investigação, no entanto, não permitiu efetuar comparações entre os resultados obtidos neste estudo e outros.

5.3. Propostas de investigação futura

Para futuros trabalhos, e atendendo a que existem poucos trabalhos semelhantes nesta área, apresentam-se as seguintes propostas:

- ✓ Desenvolver um estudo semelhante, mas onde se recorram a métodos de amostragem aleatórios, por exemplo, e a maiores amostras, para que seja possível extrapolar os resultados obtidos para a população. Deverá procurar-se obter, através dos próprios promotores de jogo, dados reais dos seus clientes, por exemplo;
- ✓ Desenvolver um estudo semelhante, mas tendo em conta outro tipo de variáveis que possam ser mais adequadas para a previsão do lucro;
- ✓ Desenvolver um estudo semelhante, mas baseado nouro tipo de jogos, como por exemplo os jogos de fortuna e azar *online* ou os jogos de base territorial. Caso seja desenvolvido um estudo do mesmo género, mas para apostas desportivas à cota de base territorial, poderá comparar-se os dois estudos para se ter uma ideia das diferenças nos comportamentos dos apostadores *online* vs. *offline*;
- ✓ Desenvolver um estudo semelhante, mas distinguindo os comportamentos dos jogadores patológicos dos não patológicos ou em risco;
- ✓ Desenvolver estudos semelhantes com dados referentes a outros países, para que seja possível comparar os comportamentos dos apostadores de cada país (comparar os resultados obtidos para Portugal com os resultados doutros países);
- ✓ Desenvolver um estudo semelhante, mas realizado pós COVID-19, para que se consiga verificar se os comportamentos que desempenham um papel importante na obtenção de lucro por parte dos apostadores atualmente são os mesmos após este período atribulado que o país atravessa;
- ✓ Neste estudo optou-se por utilizar uma técnica de modelação que pode ser diretamente visualizada e interpretada pelo ser humano (árvore de decisão). No entanto, em estudos futuros, podem ser desenvolvidos modelos preditivos semelhantes ao desenvolvido neste estudo, mas utilizando técnicas que podem permitir obter uma melhor performance preditiva, como por exemplo máquinas de vetores de suporte, redes neuronais, *random forest*, etc.;
- ✓ Desenvolver um estudo semelhante, mas utilizando outros métodos de partição do conjunto de dados, como por exemplo o método *K-fold cross-validation*, que funciona como o *Holdout* mas utilizando K partições e rotações (em cada rotação treina-se um modelo). Assim, poderá ser possível comparar as duas técnicas de partição para ver qual a mais adequada;
- ✓ O modelo preditivo foi apenas testado no *dataset* obtido. Sendo assim, poderá ser testado noutros *datasets* para verificar se as conclusões são consistentes ou se apenas se confirmam para este conjunto de dados;
- ✓ Os problemas de classificação não balanceados geram problemas em muitos algoritmos. Estes problemas são caracterizados pela proporção desigual de casos de cada classe do problema. Uma vez que o *dataset* utilizado contém algumas variáveis não balanceadas, poderão ser aplicadas técnicas de balanceamento nas classes que têm poucas observações (classe ‘Mais de 500€’ da variável Lucro,

por exemplo), e que não foram aplicadas neste estudo por falta de tempo. Uma função que poderá ser utilizada é a função SMOTE. Esta função pode ser utilizada para criar artificialmente novos exemplos das classes minoritárias utilizando os vizinhos mais próximos desses casos;

- ✓ Este estudo foi desenvolvido no sentido de analisar o impacto de determinados comportamentos dos apostadores na obtenção de lucro. No entanto, esta análise também pode ser feita no sentido de analisar a relação desses comportamentos (e outros) com o dinheiro gasto pelos apostadores. Assim, talvez se consiga perceber o que leva os apostadores a gastarem mais dinheiro para que possam ser implementados alguns fatores preventivos para que estes não se tornem jogadores patológicos. Dado o aumento do número destes jogadores no mundo (Calado et al., 2017; Hing et al., 2016; Lopez-Gonzalez et al., 2019), surge a necessidade de criar novos programas de prevenção para apostadores mais vulneráveis (com o objetivo de minimizar os danos causados pelo jogo *online*). Pode-se, então, utilizar este estudo como ponto de partida para novas investigações tanto na área do jogo *online* como na área da saúde.

Referências

- Bisquerra Alzina, R., Sarriera, J. C., & Martínez-Olmo, F. (2004). *Introdução à Estatística: Enfoque informático com o pacote estatístico SPSS*. Artmed.
https://www.researchgate.net/publication/236885258_Introducao_a_estadistica_enfoque_informatico_com_o_pacote_estadistico_SPSS
- Bleik, S., & Gauher, S. (2016). *Computing Classification Evaluation Metrics in R*. *Revolutions*. Accessed on 20/11/2020.
https://blog.revolutionanalytics.com/2016/03/com_class_eval_metrics_r.html
- Bramer, M. (2007). Avoiding Overfitting of Decision Trees. In *Principles of Data Mining* (pp. 119–134). Springer London. https://doi.org/10.1007/978-1-84628-766-4_8
- Brownlee, J. (2016). *Machine Learning Evaluation Metrics in R*. Machine Learning Mastery. Accessed on 20/11/2020. <https://machinelearningmastery.com/machine-learning-evaluation-metrics-in-r/>
- Calado, F., Alexandre, J., & Griffiths, M. D. (2017). Prevalence of Adolescent Problem Gambling: A Systematic Review of Recent Research. *Journal of Gambling Studies*, 33(2), 397–424.
<https://doi.org/10.1007/s10899-016-9627-5>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 - Step-by-step data mining guide. *CRISP-DM Consortium*.
- Constantinou, A. C. (2019). Dolores: a model that predicts football match outcomes from all over the world. *Machine Learning*, 108(1), 49–75. <https://doi.org/10.1007/s10994-018-5703-7>
- Costa, J. M., & Miranda, G. L. (2017). Desenvolvimento e validação de uma prova de avaliação das competências iniciais de programação. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 25, 66–81. <https://doi.org/10.17013/risti.25.66-81>
- Daniel, F., Gomes, A., & Ferreira, P. L. (2015). Contributo para a Discussão da Avaliação da Fiabilidade de um Instrumento de Medição. *Revista de Enfermagem Referência*, 4(7), 129–137.
- Dietterich, T. (1995). Overfitting and Undercomputing in Machine Learning. *ACM Computing Surveys (CSUR)*, 27(3), 326–327. <https://doi.org/10.1145/212094.212114>
- Dixon, M. J., & Coles, S. G. (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2), 265–280. <https://doi.org/10.1111/1467-9876.00065>
- Domingues, J., Lopes, B., Mihaylova, P., & Georgieva, P. (2019). Incremental Learning for Football Match Outcomes Prediction. In: Morales A., Fierrez J., Sánchez J., Ribeiro B. (eds) *Pattern Recognition and Image Analysis. IbPRIA 2019. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11868 LNCS, 217–228. https://doi.org/10.1007/978-3-030-31321-0_19
- European Commission. (2012). *Online gambling in the Internal Market. Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions Towards a comprehensive framework for online gambling*. http://ec.europa.eu/internal_market/gambling/docs/121023_online-gambling-staff-working-paper_en.pdf
- European Gaming & Betting Association. (2018). *Statistics - EGBA*. Accessed on 12/12/2019. <https://www.egba.eu/resources/statistics/>
- Freitas, A. L. P., & Rodrigues, S. G. (2005). A avaliação da confiabilidade de questionários: uma análise utilizando o coeficiente alfa de Cronbach. *XII Simpósio de Engenharia de Produção*, 1–12.
https://www.researchgate.net/publication/236036099_A_avaliacao_da_confabilidade_de_questio_narios_uma_analise_utilizando_o_coeficiente_alfa_de_Cronbach
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social* (6th ed.). Atlas S.A.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, 59(2), 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1157–1182.
- H2 Gambling Capital. (2018). *Tennis Betting Integrity Report - Specialist Data Analysis*. Independent

- Report. <https://ibia.bet/wp-content/uploads/2019/06/H2-ESSA-Sportradar-Tennis-Betting-Integrity-Report-Analysis-FINAL-VERSION-UPDATED-17-AUG.pdf>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques* (3rd ed.). Elsevier Inc. <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
- Hing, N., Russell, A. M. T., Vitartas, P., & Lamont, M. (2016). Demographic, behavioural and normative risk factors for gambling problems amongst sports bettors. *Journal of Gambling Studies*, 32(2), 625–641. <https://doi.org/10.1007/s10899-015-9571-9>
- Labrador, F. J., & Vallejo-Achón, M. (2019). Prevalence and Characteristics of Sports Betting in a Population of Young Students in Madrid. *Journal of Gambling Studies*, 36, 297–318. <https://doi.org/10.1007/s10899-019-09863-y>
- LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., & Shaffer, H. J. (2007). Assessing the playing field: A prospective longitudinal study of Internet sports gambling behavior. *Journal of Gambling Studies*, 23(3), 347–362. <https://doi.org/10.1007/s10899-007-9067-3>
- Leung, C. K., & Joseph, K. W. (2014). Sports Data Mining: Predicting Results for the College Football Games. *Procedia Computer Science*, 35(C), 710–719. <https://doi.org/10.1016/j.procs.2014.08.153>
- Lopez-Gonzalez, H., Estévez, A., & Griffiths, M. D. (2019). Internet-Based Structural Characteristics of Sports Betting and Problem Gambling Severity: Is There a Relationship?. *International Journal of Mental Health and Addiction*, 17(6), 1360–1373. <https://doi.org/10.1007/s11469-018-9876-x>
- Magalhães, R. M. (2011). *Jogo a dinheiro na internet: estudo exploratório para a caracterização do consumo de jogo online, especializado nas populações de estudantes universitários de ciências tecnológicas e ciências sociais* [Master Thesis, ISCTE - Instituto Universitário de Lisboa, Portugal]. <http://hdl.handle.net/10071/4224>
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3), 109–118. <http://www.90minut.pl/misc/maher.pdf>
- Marôco, J. (2018). *Análise Estatística com o SPSS Statistics* (6th ed.). ReportNumber, Lda. https://books.google.pt/books?hl=pt-PT&lr=&id=Ki5gDwAAQBAJ&oi=fnd&pg=PP1&dq=Análise+Estatística+com+o+SPSS+Statistics&ots=zNoopzE8iG&sig=0E6I8jzaLJ_MwgjTkP2HJYm86VY&redir_esc=y#v=onepage&q&f=false
- Maroco, J., & Garcia-Marques, T. (2006). Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? *Laboratório de Psicologia*, 4(1), 65–90.
- McBride, J., & Derevensky, J. (2009). Internet Gambling Behavior in a Sample of Online Gamblers. *International Journal of Mental Health and Addiction*, 7(1), 149–167. <https://doi.org/10.1007/s11469-008-9169-x>
- Miljković, D., Gajić, L., Kovačević, A., & Konjović, Z. (2010). The use of data mining for basketball matches outcomes prediction. *SIISY 2010 - 8th IEEE International Symposium on Intelligent Systems and Informatics*, 309–312. <https://doi.org/10.1109/SISY.2010.5647440>
- Ministério da Economia. (2015). *Diário da República, 1.ª série - N.º 83 - 29 de abril de 2015. Decreto-Lei n.º 66/2015*. 2139–2170. <https://dre.pt/application/file/67096527>
- Ministério da Solidariedade Emprego e Segurança Social. (2015). *Portaria n.º 173/2015 - Diário da República n.º 110/2015, Série I de 2015-06-08*. 1–16. https://dre.pt/web/guest/legislacao-consolidada-/lc/107738429/201909061540/73451700/exportPdf/normal/1/cacheLevelPage?_LegislacaoConsolidada_WAR_drefrontofficeportlet_rp=diploma
- Monteiro, J. M., Rao, A., Shawe-Taylor, J., & Mourão-Miranda, J. (2016). A multiple hold-out framework for Sparse Partial Least Squares. *Journal of Neuroscience Methods*, 271, 182–194. <https://doi.org/10.1016/j.jneumeth.2016.06.011>
- Nguyen, H. (2020). *Classifying with Trees - Decision Trees*. RPubs. Accessed on 21/11/2020. <https://rpubs.com/Haibiostat/DecisionTreesSimpleShowCase>
- Nurhayati, Soekarno, I., Hadihardaja, I. K., & Cahyono, M. (2014). A Study of Hold-Out and K-Fold Cross Validation for Accuracy of Groundwater Modeling in Tidal Lowland Reclamation Using Extreme Learning Machine. *2nd International Conference on Technology, Informatics,*

- Management, Engineering & Environment Bandung*.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7011623>
- Pordata. (2020). *PORDATA - Empresas: total e por sector de actividade económica*. Accessed on 19/02/2020.
<https://www.pordata.pt/Portugal/Empresas+total+e+por+sector+de+actividade+económica-2856>
- Poynter, R. (2010). *The Likert Scale - TARSK 14 (Things All Researchers Should Know)*. The Future Place. Accessed on 07/11/2020. https://thefutureplace.typepad.com/the_future_place/2010/09/the-likert-scale-tarsk-14-things-all-researchers-should-know.html
- R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Accessed on 21/11/2020. <https://www.r-project.org/>
- Rebelo, P. (2012). Ganhar com as Apostas Desportivas. In *Collection "Marcador Gestão e Informática."*
- RekhaMolala. (2019). *Correlation analysis and Collinearity - How they can make or break a model*. Clairvoyant. Accessed on 20/11/2020. <https://blog.clairvoyantsoft.com/correlation-and-collinearity-how-they-can-make-or-break-a-model-9135fbc6936a>
- Serviço de Regulação e Inspeção de Jogos. (2019). *SRIJ - Regulação e Inspeção de Jogos*. Accessed on 12/12/2019. <https://www.srij.turismodeportugal.pt/pt/>
- Shmueli, B. (2019). *Multi-Class Metrics Made Simple, Part II: the F1-score*. Towards Data Science. Accessed on 20/11/2020. <https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1>
- Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saude : Revista Do Sistema Unico de Saude Do Brasil*, 26(3), 649–659.
<https://doi.org/10.5123/S1679-49742017000300022>
- Stübinger, J., & Knoll, J. (2018). Beat the Bookmaker - Winning Football Bets with Machine Learning (Best Application Paper). In: Bramer M., Petridis M. (eds) *Artificial Intelligence XXXV. SGAI 2018. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11311 LNAI, 219–233.
https://doi.org/10.1007/978-3-030-04191-5_21
- Teixeira, R. (2017). *Apostas online: Gestão de Banca*. Accessed on 12/12/2019.
<https://pplware.sapo.pt/informacao/apostas-online-gestao-de-banca/>
- Therneau, T. M., & Atkinson, E. J. (2019). rpart: Recursive Partitioning and Regression Trees. *R Package Version 4*. Accessed on 25/11/2020., 1–15. <https://cran.r-project.org/package=rpart>
- Whelan, J. P., Meyers, A. W., Steenbergh, T. A. (2007). *Problem and Pathological Gambling (Vol. 8)*. Hogrefe Publishing. https://books.google.pt/books?hl=pt-PT&lr=&id=RkRfAgAAQBAJ&oi=fnd&pg=PT15&dq=Problem+and+pathological+gambling+Whelan&ots=6eAzv4Si73&sig=cseVAO7kZpo9NfbNWDlsDmisllk&redir_esc=y#v=onepage&q=Problem+and+pathological+gambling+Whelan&f=false
- Wood, R. T., & Williams, R. J. (2007). *Internet Gambling: Past, Present and Future*. In: Smith, G., Hodgins, D., & Williams, R.J. (eds.), *Research and Measurement Issues in Gambling Studies*. San Diego, California: Elsevier Publishing. 491–514.
<http://opus.uleth.ca/bitstream/handle/10133/422/Internet-RMIGS-2007.pdf?sequence=1&isAllowed=y>
- Wood, R. T., & Williams, R. J. (2009). *Internet gambling: Prevalence, patterns, problems, and policy options. Final Report prepared for the Ontario Problem Gambling Research Centre; Guelph, Ontario*.
https://www.researchgate.net/publication/328107914_Internet_Gambling_Prevalence_Patterns_Problems_and_Policy_Options

Apêndices

Apêndice A Questionário do 1º Teste Piloto

Estudo universitário sobre apostas desportivas online em Portugal (Validação 1)

O questionário abaixo apresentado surge no âmbito da elaboração de uma dissertação de mestrado em Informática e Gestão no Iscte - Instituto Universitário de Lisboa, cujo principal objetivo passa por analisar a influência de algumas variáveis no lucro obtido por parte dos apostadores.

As informações recolhidas destinam-se única e exclusivamente ao suporte desta dissertação, sendo garantido o anonimato e a confidencialidade das mesmas.

O tempo de resposta estimado é de 7 minutos.

Qualquer questão relativa a este questionário poderá ser colocada via email: gspas1@iscte-iul.pt.

A sua colaboração é essencial para o sucesso desta investigação, pelo que, desde já, agradeço a sua participação.

***Obrigatório**

Realizou pelo menos uma aposta desportiva à cota em casas de apostas online no último ano? *

- Sim
- Não

Questões demográficas

Género: *

- Masculino
- Feminino
- Outro

Escalão etário: *

- 18-24 anos
- 25-34 anos
- 35-44 anos
- 45-54 anos
- 55-64 anos
- 65+ anos

Habilitações literárias: *

- Sem escolaridade
- 1.º ciclo do ensino básico (4.º ano)
- 2.º ciclo do ensino básico (6.º ano)
- 3.º ciclo do ensino básico (9.º ano)
- Ensino secundário (12.º ano) ou Curso profissional
- Bacharelato
- Licenciatura
- Pós-graduação
- Mestrado
- Doutoramento

Ocupação: *

- Desempregado
- Desempregado à procura do primeiro emprego
- Empregado por conta própria
- Empregado por conta de outrem
- Estudante
- Reformado
- Outra

Caso tenha escolhido a opção 'Empregado por conta própria' ou 'Empregado por conta de outrem' na questão anterior, qual o setor de atividade em que trabalha?

- Agricultura, produção animal, caça, silvicultura e pesca
- Alojamento, restauração e similares
- Atividades de saúde humana e apoio social
- Atividades financeiras e de seguros
- Atividades imobiliárias
- Comércio por grosso e a retalho
- Construção
- Desporto
- Direito
- Educação
- Eletricidade, gás e água
- Indústrias extrativas
- Indústrias transformadoras
- Tecnologias de Informação e Comunicação
- Transporte e armazenagem
- Outros setores

Região de residência: *

- Norte
- Centro
- Sul
- Açores
- Madeira
- Estrangeiro

Distrito de Residência: *

- Braga
- Bragança
- Porto
- Viana do Castelo
- Vila Real

Apostas desportivas *online* em Portugal

Distrito de Residência: *

- Aveiro
- Castelo Branco
- Coimbra
- Guarda
- Leiria
- Santarém
- Viseu

Distrito de Residência: *

- Beja
- Évora
- Faro
- Lisboa
- Portalegre
- Setúbal

Questões específicas do estudo em questão

Qual a casa de apostas que utiliza (maioritariamente) para colocar as suas apostas desportivas à cota? *

- Bet
- Betano
- Betclik
- Estoril Sol Casino (ESC) Online
- Luckia
- Mooshi
- Nossa Aposta
- Placard.pt
- Outra

Qual a frequência com que costuma apostar? *

- Diariamente
- Mais que uma vez por semana
- Semanalmente
- Quinzenalmente
- Mensalmente
- Menos que mensalmente

Caso tenha escolhido a opção 'Diariamente', quantas apostas desportivas realiza por dia?

A sua resposta _____

Caso tenha escolhido a opção 'Mais que uma vez por semana', quantas vezes por semana costuma apostar?

A sua resposta _____

Caso tenha escolhido a opção 'Mais que uma vez por semana', qual(ais) o(s) dia(s) da semana em que costuma apostar?

- Segunda-feira
- Terça-feira
- Quarta-feira
- Quinta-feira
- Sexta-feira
- Sábado
- Domingo

Qual a modalidade que mais utiliza nas suas apostas desportivas? *

- Simples (apenas um prognóstico, numa única aposta)
- Combinada (entre dois a oito prognósticos e uma única combinação, numa única aposta)
- Múltipla (entre três a cinco prognósticos e várias combinações possíveis, numa única aposta)

Caso tenha escolhido uma das duas últimas opções na questão anterior, quantos jogos coloca, em média, nas suas apostas desportivas?

- 1-3
- 4-6
- 7-8

Apostas desportivas *online* em Portugal

Qual o sistema de apostas desportivas que mais utiliza nas suas apostas desportivas? *

- Stake Fixa (aposta sempre o mesmo valor em qualquer aposta efetuada)
- Stake por Níveis (estabelece valores diferentes de acordo com os encontros, consoante o risco e o nível de confiança)
- Sistema Martingale (duplica o valor da aposta sempre que perde)
- Apostas livres (não utiliza qualquer tipo de sistema de apostas desportivas)

Qual o tipo de aposta que mais utiliza nas suas apostas desportivas? *

- 1 X 2
- 1 X 2 Intervalo
- Ambas as equipas marcam
- Cantos
- Dupla Possibilidade
- Equipa marca em ambas as partes
- Golos por equipa
- Intervalo/Final
- Mais/Menos golos
- Mais/Menos golos Intervalo
- O jogador X vencerá pelo menos um set
- Primeira equipa a marcar
- Primeira equipa a marcar X pontos num set
- Quem marca o X.º golo
- Quem marca o X.º ensaio
- Resultado exato
- Vencedor (equipa/competidor/piloto)
- Outra: _____

Considera que obtém lucro com o tipo de aposta que mais utiliza? *

- Sim
- Não
- Não sei

Qual a modalidade desportiva na qual mais aposta? *

- Basquetebol
- Futebol
- Futsal
- Hóquei no gelo
- Ténis
- Voleibol
- Outra: _____

Qual o período em que costuma apostar mais frequentemente? *

- Ao vivo (durante a realização dos encontros desportivos)
- Antes do início dos encontros desportivos

Qual a estratégia que mais utiliza nas suas apostas desportivas? *

- Apostar apenas nos jogos do próprio dia
- Apostar em jogos de dias diferentes

Costuma apostar sempre (ou quase sempre) nas mesmas equipas? *

- Sim
- Não

Utiliza algum site ou aplicação como apoio para a realização de apostas desportivas, nomeadamente os que efetuam estudos às equipas? *

- Sim
- Não

Considera que consegue obter lucro com as apostas desportivas? *

- Sim
- Não
- Não sei

Apostas desportivas *online* em Portugal

Aproximadamente quanto dinheiro gastou em apostas desportivas online nos últimos 3 meses? *

- Até 49€
- Entre 50€ e 99€
- Entre 100€ e 499€
- Entre 500€ e 999€
- Entre 1000€ e 4999€
- Mais de 5000€

Caso tenha escolhido a opção 'Sim', aproximadamente quanto dinheiro ganhou (lucro) em apostas desportivas nos últimos 3 meses?

- Até 49€
- Entre 50€ e 99€
- Entre 100€ e 499€
- Entre 500€ e 999€
- Entre 1000€ e 4999€
- Mais de 5000€

Apêndice B

Alterações à primeira versão do questionário

Após a primeira validação do questionário e mediante o *feedback* dos 10 inquiridos, foram efetuadas as seguintes alterações ao mesmo:

- ✓ O tempo de preenchimento do questionário foi aumentado de sete para 10 minutos;
- ✓ Alguns dos inquiridos não perceberam o significado de ‘à cota’ na primeira questão do questionário, por não estarem familiarizados com este termo. Sendo assim, optou-se por acrescentar uma pequena nota a explicar o que é uma aposta desportiva à cota;
- ✓ Nos casos em que a questão começa com ‘Caso tenha respondido X na questão anterior...’, pode não ser claro qual a questão a que essa questão se refere. Por exemplo, na última questão (lucro dos últimos três meses) só deve responder quem considera ter lucro com as apostas desportivas e houve um inquirido a responder mesmo tendo respondido que não tem lucro. Sendo assim, optou-se por numerar todas as questões e fazer referência ao número da questão onde fosse necessário;
- ✓ Foi acrescentada uma nota com uma explicação referente aos distritos de residência nas questões 8, 9 e 10 (da segunda versão do questionário);
- ✓ A questão 'Qual a casa de apostas que utiliza (maioritariamente) para colocar as suas apostas desportivas à cota?' foi eliminada pois se se concluisse que há uma casa de apostas onde o lucro é maior, esta poderia estar a ser publicitada (por ter *odds* mais elevadas, por exemplo);
- ✓ Foram acrescentadas várias opções de resposta na questão 'Caso tenha escolhido a opção ‘Diariamente’, quantas apostas desportivas realiza por dia’, deixando esta de ser uma questão de resposta aberta (dando origem à questão 12 da segunda versão do questionário);
- ✓ A questão 'Caso tenha escolhido a opção ‘Mais que uma vez por semana’, quantas vezes por semana costuma apostar?' foi eliminada por se considerar que foi mal construída;
- ✓ A questão 'Qual a modalidade que mais utiliza nas suas apostas desportivas?' foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 14 da segunda versão do questionário;
- ✓ A questão 'Caso tenha escolhido uma das duas últimas opções na questão anterior, quantos jogos coloca, em média, nas suas apostas desportivas?' foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 15 da segunda versão do questionário;
- ✓ A questão 'Qual o sistema de apostas desportivas que mais utiliza nas suas apostas desportivas?' foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 16 da segunda versão do questionário;
- ✓ A questão 'Qual o tipo de aposta que mais utiliza nas suas apostas desportivas?' foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 17 da segunda versão do questionário;

- ✓ A questão 'Qual a modalidade desportiva na qual mais aposta?' foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 19 da segunda versão do questionário;
- ✓ Foi acrescentada uma nova questão, por sugestão de um dos inquiridos (a questão 23 da segunda versão do questionário);
- ✓ A questão 'Considera que consegue obter lucro com as apostas desportivas?' foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 25 da segunda versão do questionário;
- ✓ Foi acrescentada uma nota referente à COVID-19 nas questões 26 e 27 da segunda versão do questionário.

Apêndice C

Questionário do 2º Teste Piloto

Estudo universitário sobre apostas desportivas online em Portugal

O presente questionário surge no âmbito da elaboração de uma dissertação de Mestrado em Informática e Gestão no Iscte - Instituto Universitário de Lisboa, cujo principal objetivo passa por analisar a influência de algumas variáveis no lucro obtido por parte dos apostadores.

As informações recolhidas destinam-se única e exclusivamente ao suporte desta dissertação, sendo garantido o anonimato e a confidencialidade das mesmas.

O tempo de resposta estimado é de 10 minutos.

Qualquer questão relativa a este questionário poderá ser colocada via email: gspas1@iscte-iul.pt.

A sua colaboração é essencial para o sucesso desta investigação, pelo que, desde já, agradeço a sua participação.

***Obrigatório**

1. Realizou pelo menos uma aposta desportiva à cota em casas de apostas online no último ano? *

Nota: uma aposta desportiva à cota é aquela em que o apostador joga contra a entidade exploradora, organizadora da aposta, com base num valor igual ou superior a 1,00 (cota), comportando até duas casas decimais, previamente definido ou convencionalmente fixado, valor esse associado a cada um dos prognósticos possíveis para cada aposta em função da probabilidade de ocorrência de um determinado tipo de resultado.

- Sim
- Não

Questões demográficas

2. Género: *

- Masculino
- Feminino
- Outro

3. Escalão etário: *

- 18-24 anos
- 25-34 anos
- 35-44 anos
- 45-54 anos
- 55-64 anos
- 65+ anos

4. Habilitações literárias: *

- Sem escolaridade
- 1.º ciclo do ensino básico (4.º ano)
- 2.º ciclo do ensino básico (6.º ano)
- 3.º ciclo do ensino básico (9.º ano)
- Ensino secundário (12.º ano) ou Curso profissional
- Bacharelato
- Licenciatura
- Pós-graduação
- Mestrado
- Doutoramento

5. Ocupação: *

- Desempregado
- Desempregado à procura do primeiro emprego
- Empregado por conta própria
- Empregado por conta de outrem
- Estudante
- Reformado
- Outra

6. Caso tenha escolhido a opção 'Empregado por conta própria' ou 'Empregado por conta de outrem' na questão 5, qual o setor de atividade em que trabalha?

- Agricultura, produção animal, caça, silvicultura e pesca
- Alojamento, restauração e similares
- Atividades de saúde humana e apoio social
- Atividades financeiras e de seguros
- Atividades imobiliárias
- Comércio por grosso e a retalho
- Construção
- Desporto
- Direito
- Educação
- Eletricidade, gás e água
- Indústrias extrativas
- Indústrias transformadoras
- Tecnologias de Informação e Comunicação
- Transporte e armazenagem
- Outros setores

7. Região de residência: *

- Norte
- Centro
- Sul
- Açores
- Madeira
- Estrangeiro

8. Distrito de Residência: *

Alguns distritos podem ser considerados como pertencentes a duas regiões de residência diferentes. Caso não encontre aqui o seu distrito de residência, retorne à secção anterior pois encontra-se numa das outras regiões.

- Braga
- Bragança
- Porto
- Viana do Castelo
- Vila Real

Apostas desportivas *online* em Portugal

9. Distrito de Residência: *

Alguns distritos podem ser considerados como pertencentes a duas regiões de residência diferentes. Caso não encontre aqui o seu distrito de residência, retorne à secção anterior pois encontra-se numa das outras regiões.

- Aveiro
- Castelo Branco
- Coimbra
- Guarda
- Leiria
- Santarém
- Viseu

10. Distrito de Residência: *

Alguns distritos podem ser considerados como pertencentes a duas regiões de residência diferentes. Caso não encontre aqui o seu distrito de residência, retorne à secção anterior pois encontra-se numa das outras regiões.

- Beja
- Évora
- Faro
- Lisboa
- Portalegre
- Setúbal

Questões específicas do estudo em questão

11. Qual a frequência com que costuma apostar? *

- Diariamente
- Mais que uma vez por semana
- Semanalmente
- Quinzenalmente
- Mensalmente
- Menos que mensalmente

12. Caso tenha escolhido a opção 'Diariamente' na questão 11, em média quantas apostas desportivas realiza por dia (live e pré-live)?

- Até 4 apostas
- Entre 5 e 9 apostas
- Entre 10 e 14 apostas
- Entre 15 e 19 apostas
- Entre 20 e 24 apostas
- Mais de 25 apostas

13. Caso tenha escolhido a opção 'Mais que uma vez por semana' na questão 11, qual(ais) o(s) dia(s) da semana em que costuma apostar?

- Segunda-feira
- Terça-feira
- Quarta-feira
- Quinta-feira
- Sexta-feira
- Sábado
- Domingo

14. Com que frequência utiliza cada uma das seguintes modalidades nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
Simple (apenas um prognóstico, numa única aposta)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Combinada (entre dois a oito prognósticos e uma única combinação, numa única aposta)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Múltipla (entre três a cinco prognósticos e várias combinações possíveis, numa única aposta)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Apostas desportivas *online* em Portugal

15. Com que frequência coloca o seguinte número de jogos nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
1 a 3 jogos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4 a 6 jogos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7 ou 8 jogos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16. Com que frequência utiliza cada um dos seguintes sistemas de apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
Stake Fixa (aposta sempre o mesmo valor em qualquer aposta efetuada)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stake por Níveis (estabelece valores diferentes de acordo com os encontros, consoante o risco e o nível de confiança)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sistema Martingale (duplica o valor da aposta sempre que perde)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Apostas livres (não utiliza qualquer tipo de sistema de apostas desportivas)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

17. Com que frequência utiliza cada um dos seguintes tipos de aposta nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
1 X 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1 X 2 Intervalo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ambas as equipas marcam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cantos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dupla Possibilidade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Equipa marca em ambas as partes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Golos por equipa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intervalo/Final	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mais/Menos golos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mais/Menos golos Intervalo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
O jogador X vencerá pelo menos um set	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Primeira equipa a marcar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Primeira equipa a marcar X pontos num set	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quem marca o X.º golo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quem marca o X.º ensaio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resultado exato	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vencedor (equipa/competidor/piloto)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

18. Considera que obtém lucro com o tipo de aposta que mais utiliza? *

- Sim
- Não
- Não sei

Apostas desportivas *online* em Portugal

19. Com que frequência utiliza cada uma das seguintes modalidades nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
Andebol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Basquetebol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Futebol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Futsal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hóquei no gelo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rugby (ou futebol americano)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ténis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ténis de mesa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Voleibol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

20. Qual o período em que costuma apostar mais frequentemente? *

- Ao vivo (durante a realização dos encontros desportivos)
- Antes do início dos encontros desportivos

21. Qual a estratégia que mais utiliza nas suas apostas desportivas? *

- Apostar apenas nos jogos do próprio dia
- Apostar em jogos de dias diferentes

22. Costuma apostar sempre (ou quase sempre) nas mesmas equipas? *

- Sim
- Não

23. Costuma apostar sempre (ou quase sempre) nos normais vencedores das competições (equipas nos primeiros lugares da tabela)? *

- Sim
- Não

24. Utiliza algum site ou aplicação como apoio para a realização de apostas desportivas, nomeadamente os que efetuam estudos às equipas? *

- Sim
- Não

25. Com que frequência obtém lucro nas suas apostas desportivas? *

- Nunca
- Poucas vezes
- Algumas vezes
- Muitas vezes
- Sempre

26. Aproximadamente quanto dinheiro gastou em apostas desportivas online nos últimos 3 meses? *

Caso não tenha continuado a realizar apostas desportivas devido à COVID-19, responda a esta questão tendo como base os últimos 3 meses em que o fez.

- Até 49€
- Entre 50€ e 99€
- Entre 100€ e 499€
- Entre 500€ e 999€
- Entre 1000€ e 4999€
- Mais de 5000€

27. Caso tenha obtido lucro nos últimos 3 meses, aproximadamente quanto dinheiro ganhou (lucro) em apostas desportivas nesse período?

Caso não tenha continuado a realizar apostas desportivas devido à COVID-19, responda a esta questão tendo como base os últimos 3 meses em que o fez.

- Até 49€
- Entre 50€ e 99€
- Entre 100€ e 499€
- Entre 500€ e 999€
- Entre 1000€ e 4999€
- Mais de 5000€

Apêndice D

Alterações à segunda versão do questionário

Após a segunda validação do questionário, foram efetuadas as seguintes alterações ao mesmo:

- ✓ A questão 'Qual a casa de apostas que utiliza (maioritariamente) para colocar as suas apostas desportivas à cota?' foi novamente acrescentada para ser utilizada apenas para caracterizar a amostra (dando origem à questão 11 da versão final do questionário);
- ✓ A questão 11 da segunda versão do questionário passou a 12 na versão final do questionário;
- ✓ A questão 12 da segunda versão do questionário passou a 13 na versão final do questionário (com uma pequena alteração);
- ✓ As questões 14 e 15 da segunda versão do questionário foram agrupadas e substituídas pela questão 14 na versão final do questionário;
- ✓ A questão 16 da segunda versão do questionário passou a 15 na versão final do questionário;
- ✓ A questão 17 da segunda versão do questionário passou a 16 na versão final do questionário;
- ✓ A questão 18 da segunda versão do questionário passou a 17 na versão final do questionário;
- ✓ A questão 19 da segunda versão do questionário passou a 18 na versão final do questionário;
- ✓ As questões 20 e 21 da segunda versão do questionário foram agrupadas e substituídas pela questão 19 na versão final do questionário;
- ✓ A questão 22 foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 20 da segunda versão do questionário;
- ✓ A questão 23 foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 21 da segunda versão do questionário;
- ✓ A questão 24 foi transformada numa questão construída a partir da escala de *Likert* mencionada acima, dando origem à questão 22 da segunda versão do questionário;
- ✓ A questão 25 da segunda versão do questionário passou a 23 na versão final do questionário;
- ✓ A questão 26 da segunda versão do questionário passou a 24 na versão final do questionário;
- ✓ A questão 27 da segunda versão do questionário passou a 25 na versão final do questionário.

Apêndice E

Questionário Final

Estudo universitário sobre apostas desportivas online em Portugal

O presente questionário surge no âmbito da elaboração de uma dissertação de Mestrado em Informática e Gestão no Iscte - Instituto Universitário de Lisboa, cujo principal objetivo passa por analisar a influência de algumas variáveis no lucro obtido por parte dos apostadores.

As informações recolhidas destinam-se única e exclusivamente ao suporte desta dissertação, sendo garantido o anonimato e a confidencialidade das mesmas.

O tempo de resposta estimado é de 10 minutos.

Qualquer questão relativa a este questionário poderá ser colocada via email: gspas1@iscte-iul.pt.

A sua colaboração é essencial para o sucesso desta investigação, pelo que, desde já, agradeço a sua participação.

***Obrigatório**

1. Realizou pelo menos uma aposta desportiva à cota em casas de apostas online no último ano? *

Nota: uma aposta desportiva à cota é aquela em que o apostador joga contra a entidade exploradora, organizadora da aposta, com base num valor igual ou superior a 1,00 (cota), comportando até duas casas decimais, previamente definido ou convencionalmente fixado, valor esse associado a cada um dos prognósticos possíveis para cada aposta em função da probabilidade de ocorrência de um determinado tipo de resultado.

Sim

Não

Questões demográficas

2. Género: *

Masculino

Feminino

Outro

3. Escalão etário: *

18-24 anos

25-34 anos

35-44 anos

45-54 anos

55-64 anos

65+ anos

4. Habilitações literárias: *

- Sem escolaridade
- 1.º ciclo do ensino básico (4.º ano)
- 2.º ciclo do ensino básico (6.º ano)
- 3.º ciclo do ensino básico (9.º ano)
- Ensino secundário (12.º ano) ou Curso profissional
- Bacharelato
- Licenciatura
- Pós-graduação
- Mestrado
- Doutoramento

5. Ocupação: *

- Desempregado
- Desempregado à procura do primeiro emprego
- Empregado por conta própria
- Empregado por conta de outrem
- Estudante
- Reformado
- Outra

6. Caso tenha escolhido a opção 'Empregado por conta própria' ou 'Empregado por conta de outrem' na questão 5, qual o setor de atividade em que trabalha?

- Agricultura, produção animal, caça, silvicultura e pesca
- Alojamento, restauração e similares
- Atividades de saúde humana e apoio social
- Atividades financeiras e de seguros
- Atividades imobiliárias
- Comércio por grosso e a retalho
- Construção
- Desporto
- Direito
- Educação
- Eletricidade, gás e água
- Indústrias extrativas
- Indústrias transformadoras
- Tecnologias de Informação e Comunicação
- Transporte e armazenagem
- Outros setores

7. Região de residência: *

- Norte
- Centro
- Sul
- Açores
- Madeira
- Estrangeiro

8. Distrito de Residência: *

Alguns distritos podem ser considerados como pertencentes a duas regiões de residência diferentes. Caso não encontre aqui o seu distrito de residência, retorne à secção anterior pois encontra-se numa das outras regiões.

- Braga
- Bragança
- Porto
- Viana do Castelo
- Vila Real

Apostas desportivas *online* em Portugal

9. Distrito de Residência: *

Alguns distritos podem ser considerados como pertencentes a duas regiões de residência diferentes. Caso não encontre aqui o seu distrito de residência, retorne à secção anterior pois encontra-se numa das outras regiões.

- Aveiro
- Castelo Branco
- Coimbra
- Guarda
- Leiria
- Santarém
- Viseu

10. Distrito de Residência: *

Alguns distritos podem ser considerados como pertencentes a duas regiões de residência diferentes. Caso não encontre aqui o seu distrito de residência, retorne à secção anterior pois encontra-se numa das outras regiões.

- Beja
- Évora
- Faro
- Lisboa
- Portalegre
- Setúbal

Questões específicas do estudo em questão

11. Qual a casa de apostas que utiliza (maioritariamente) para colocar as suas apostas desportivas à cota? *

- Bet
- Betano
- Betclik
- Estoril Sol Casino (ESC) Online
- Luckia
- Moosh
- Nossa Aposta
- Placard.pt
- Outra

12. Qual a frequência com que costuma apostar? *

- Diariamente
- Mais que uma vez por semana
- Semanalmente
- Quinzenalmente
- Mensalmente
- Menos que mensalmente

13. Tendo em conta a frequência escolhida na questão 12, em média quantas apostas desportivas realiza (live e pré-live)? *

Exemplo: Se apostar semanalmente, indique quantas apostas efetua por semana (em média).

- Até 4 apostas
- Entre 5 e 9 apostas
- Entre 10 e 14 apostas
- Entre 15 e 19 apostas
- Entre 20 e 24 apostas
- Mais de 25 apostas

Apostas desportivas *online* em Portugal

14. Com que frequência utiliza cada uma das seguintes modalidades de aposta nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
Simple (apenas um prognóstico)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Combinada (entre um e três prognósticos e uma única combinação)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Combinada (entre quatro e seis prognósticos e uma única combinação)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Combinada (sete ou oito prognósticos e uma única combinação)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Múltipla (entre três e cinco prognósticos e várias combinações possíveis)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15. Com que frequência utiliza cada um dos seguintes sistemas de apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
Stake Fixa (aposta sempre o mesmo valor em qualquer aposta efetuada)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stake por Níveis (estabelece valores diferentes de acordo com os encontros, consoante o risco e o nível de confiança)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sistema Martingale (duplica o valor da aposta sempre que perde)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Apostas livres (não utiliza qualquer tipo de sistema de apostas desportivas)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Apostas desportivas *online* em Portugal

16. Com que frequência utiliza cada um dos seguintes tipos de aposta nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
1 X 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1 X 2 Intervalo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ambas as equipas marcam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cantos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dupla Possibilidade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Equipa marca em ambas as partes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Golos por equipa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intervalo/Final	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mais/Menos golos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mais/Menos golos Intervalo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
O jogador X vencerá pelo menos um set	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Primeira equipa a marcar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Primeira equipa a marcar X pontos num set	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quem marca o X.º golo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quem marca o X.º ensaio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resultado exato	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vencedor (equipa/competidor/piloto)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

17. Considera que obtém lucro com o tipo de aposta que mais utiliza? *

- Sim
- Não
- Não sei

18. Com que frequência utiliza cada uma das seguintes modalidades nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
Andebol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Basquetebol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Futebol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Futsal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hóquei no gelo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rugby (ou futebol americano)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ténis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ténis de mesa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Voleibol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

19. Com que frequência utiliza cada uma das seguintes estratégias nas suas apostas desportivas? *

	Nunca	Poucas vezes	Algumas vezes	Muitas vezes	Sempre
Apostar apenas nos jogos do próprio dia e ao vivo (durante a realização dos encontros desportivos)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Apostar apenas nos jogos do próprio dia e antes do início dos encontros desportivos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Apostar em jogos de dias diferentes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Apostas desportivas *online* em Portugal

20. Com que frequência aposta nas mesmas equipas? *

- Nunca
- Poucas vezes
- Algumas vezes
- Muitas vezes
- Sempre

21. Com que frequência coloca os habituais vencedores das competições (equipas nos primeiros lugares da tabela) nas apostas como vencedores? *

- Nunca
- Poucas vezes
- Algumas vezes
- Muitas vezes
- Sempre

22. Com que frequência utiliza algum site ou aplicação como apoio para a realização de apostas desportivas, nomeadamente os que efetuam estudos às equipas? *

- Nunca
- Poucas vezes
- Algumas vezes
- Muitas vezes
- Sempre

23. Com que frequência obtém lucro nas suas apostas desportivas? *

- Nunca
- Poucas vezes
- Algumas vezes
- Muitas vezes
- Sempre

24. Aproximadamente quanto dinheiro gastou em apostas desportivas online nos últimos 3 meses? *

Caso não tenha continuado a realizar apostas desportivas devido à COVID-19, responda a esta questão tendo como base os últimos 3 meses em que o fez.

- Até 49€
- Entre 50€ e 99€
- Entre 100€ e 499€
- Entre 500€ e 999€
- Entre 1000€ e 4999€
- Mais de 5000€

25. Caso tenha obtido lucro nos últimos 3 meses, aproximadamente quanto dinheiro ganhou (lucro) em apostas desportivas nesse período?

Caso não tenha continuado a realizar apostas desportivas devido à COVID-19, responda a esta questão tendo como base os últimos 3 meses em que o fez.

- Até 49€
- Entre 50€ e 99€
- Entre 100€ e 499€
- Entre 500€ e 999€
- Entre 1000€ e 4999€
- Mais de 5000€

Apêndice F
Pontuação do perito quanto à relevância e clareza dos itens

Itens	Relevância do item (1 = Nada relevante; 2 = Pouco relevante; 3 = Relevante; 4 = Muito relevante; 5 = Extremamente relevante)	Clareza do item (1 = Nada claro; 2 = Pouco claro; 3 = Claro; 4 = Muito claro; 5 = Extremamente claro)
11. Qual a casa de apostas que utiliza (maioritariamente) para colocar as suas apostas desportivas à cota?	4	5
12. Qual a frequência com que costuma apostar?	4	5
13. Tendo em conta a frequência escolhida na questão 12, em média quantas apostas desportivas realiza (live e pré-live)?	4	5
14. Com que frequência utiliza cada uma das seguintes modalidades de aposta nas suas apostas desportivas?	4	5
15. Com que frequência utiliza cada um dos seguintes sistemas de apostas desportivas?	4	5
16. Com que frequência utiliza cada um dos seguintes tipos de aposta nas suas apostas desportivas?	4	5
17. Considera que obtém lucro com o tipo de aposta que mais utiliza?	5	5
18. Com que frequência utiliza cada uma das seguintes modalidades nas suas apostas desportivas?	5	5
19. Com que frequência utiliza cada uma das seguintes estratégias nas suas apostas desportivas?	5	5
20. Com que frequência aposta nas mesmas equipas?	5	5
21. Com que frequência coloca os habituais vencedores das competições (equipas nos primeiros lugares da tabela) nas apostas como vencedores?	5	5
22. Com que frequência utiliza algum site ou aplicação como apoio para a realização de apostas desportivas, nomeadamente os que efetuam estudos às equipas?	5	5
23. Com que frequência obtém lucro nas suas apostas desportivas?	5	5
24. Aproximadamente quanto dinheiro gastou em apostas desportivas online nos últimos 3 meses?	2	5

Apostas desportivas *online* em Portugal

25. Caso tenha obtido lucro nos últimos 3 meses, aproximadamente quanto dinheiro ganhou em apostas desportivas nesse período?	5	5
Média	4.4	5
Máx.	2	5
Min.	5	5

Apêndice G

Caraterização demográfica dos inquiridos

Variável	Frequência Absoluta (N)	Frequência Relativa (%)	Frequência Relativa Acumulada (%)
Género			
Masculino	485	93,3	-
Feminino	35	6,7	-
Total	520	100	-
Escalão etário			
18-24 anos	155	29,8	29,8
25-34 anos	137	26,3	56,2
35-44 anos	136	26,2	82,3
45-54 anos	78	15,0	97,3
55-64 anos	13	2,5	99,8
65+ anos	1	0,2	100
Total	520	100	
Habilitações literárias			
Sem escolaridade	0	0,0	0,0
1.º ciclo do ensino básico (4.º ano)	2	0,4	0,4
2.º ciclo do ensino básico (6.º ano)	20	3,8	4,2
3.º ciclo do ensino básico (9.º ano)	91	17,5	21,7
Ensino secundário (12.º ano) ou Curso profissional	235	45,2	66,9
Bacharelato	6	1,2	68,1
Licenciatura	122	23,5	91,5
Pós-graduação	12	2,3	93,8
Mestrado	31	6,0	99,8
Doutoramento	1	0,2	100
Total	520	100	
Ocupação			
Desempregado	23	4,3	-
Desempregado à procura do primeiro emprego	5	0,9	-
Empregado por conta própria	57	10,6	-
Empregado por conta de outrem	362	67,3	-
Estudante	85	15,8	-
Reformado	2	0,4	-
Outra	4	0,7	-
Total	538	100	-
Setor de atividade			
Agricultura, produção animal, caça, silvicultura e pesca	3	0,6	-
Alojamento, restauração e similares	31	6,0	-
Atividades de saúde humana e apoio social	17	3,3	-
Atividades financeiras e de seguros	14	2,7	-
Atividades imobiliárias	3	0,6	-
Comércio por grosso e a retalho	53	10,2	-
Construção	23	4,4	-
Desporto	17	3,3	-
Direito	2	0,4	-
Educação	19	3,7	-
Eletricidade, gás e água	8	1,5	-
Indústrias extrativas	6	1,2	-
Indústrias transformadoras	23	4,4	-
Tecnologias de Informação e Comunicação	38	7,3	-
Transporte e armazenagem	33	6,3	-
Outros setores	121	23,3	-
Não respondido	109	21,0	-
Total	520	100	-
Região de residência			
Norte	103	19,8	-

Apostas desportivas *online* em Portugal

Centro	123	23,7	-
Sul	264	50,8	-
Açores	14	2,7	-
Madeira	16	3,1	-
Total	520	100	-
Distrito de residência			
Aveiro	24	4,6	-
Beja	7	1,3	-
Braga	25	4,8	-
Bragança	1	0,2	-
Castelo Branco	10	1,9	-
Coimbra	27	5,2	-
Évora	9	1,7	-
Faro	14	2,7	-
Guarda	9	1,7	-
Leiria	24	4,6	-
Lisboa	186	35,8	-
Portalegre	4	0,8	-
Porto	67	12,9	-
Santarém	20	3,8	-
Setúbal	44	8,5	-
Viana do Castelo	4	0,8	-
Vila Real	6	1,2	-
Viseu	9	1,7	-
Não respondido	30	5,8	-
Total	520	100	-

Apêndice H

Caraterização dos comportamentos dos inquiridos

Variável	Frequência Absoluta (N)	Frequência Relativa (%)
Casa de apostas		
Bet	56	10,8
Betano	54	10,4
Betclic	144	27,7
Estoril Sol Casino (ESC) Online	22	4,2
Luckia	1	0,2
Moosh	0	0,0
Nossa Aposta	3	0,6
Placard.pt	125	24,0
Outra	115	22,1
Total	520	100
Frequência		
Diariamente	220	42,3
Mais que uma vez por semana	133	25,6
Semanalmente	58	11,2
Quinzenalmente	22	4,2
Mensalmente	31	6,0
Menos que mensalmente	56	10,8
Total	520	100
Número de apostas desportivas		
Até 4 apostas	276	53,1
Entre 5 e 9 apostas	166	31,9
Entre 10 e 14 apostas	39	7,5
Entre 15 e 19 apostas	16	3,1
Entre 20 e 24 apostas	6	1,2
Mais de 25 apostas	17	3,3
Total	520	100
Modalidades de aposta		
Simples		
Nunca	106	20,4
Poucas vezes	186	35,8
Algumas vezes	130	25,0
Muitas vezes	60	11,5
Sempre	38	7,3
Total	520	100
Combinada (entre um e três prognósticos)		
Nunca	53	10,2
Poucas vezes	108	20,8
Algumas vezes	212	40,8
Muitas vezes	124	23,8
Sempre	23	4,4
Total	520	100
Combinada (entre quatro e seis prognósticos)		
Nunca	74	14,2
Poucas vezes	151	29,0
Algumas vezes	191	36,7
Muitas vezes	95	18,3
Sempre	9	1,7
Total	520	100
Combinada (sete ou oito prognósticos)		
Nunca	161	31,0
Poucas vezes	150	28,8
Algumas vezes	140	26,9
Muitas vezes	58	11,2

Sempre	11	2,1
Total	520	100
Múltipla		
Nunca	247	47,5
Poucas vezes	167	32,1
Algumas vezes	78	15,0
Muitas vezes	24	4,6
Sempre	4	0,8
Total	520	100
Sistemas de apostas desportivas		
<i>Stake Fixa</i>		
Nunca	148	28,5
Poucas vezes	160	30,8
Algumas vezes	127	24,4
Muitas vezes	55	10,6
Sempre	30	5,8
Total	520	100
<i>Stake por Níveis</i>		
Nunca	120	23,1
Poucas vezes	127	24,4
Algumas vezes	165	31,7
Muitas vezes	87	16,7
Sempre	21	4,0
Total	520	100
<i>Sistema Martingale</i>		
Nunca	342	65,8
Poucas vezes	115	22,1
Algumas vezes	44	8,5
Muitas vezes	15	2,9
Sempre	4	0,8
Total	520	100
<i>Apostas livres</i>		
Nunca	181	34,8
Poucas vezes	150	28,8
Algumas vezes	95	18,3
Muitas vezes	61	11,7
Sempre	33	6,3
Total	520	100
Tipos de aposta		
1 X 2		
Nunca	40	7,7
Poucas vezes	157	30,2
Algumas vezes	90	17,3
Muitas vezes	171	32,9
Sempre	62	11,9
Total	520	100
1 X 2 Intervalo		
Nunca	94	18,1
Poucas vezes	233	44,8
Algumas vezes	125	24,0
Muitas vezes	63	12,1
Sempre	5	1,0
Total	520	100
<i>Ambas as equipas marcam</i>		
Nunca	52	10,0
Poucas vezes	205	39,4
Algumas vezes	137	26,3
Muitas vezes	118	22,7
Sempre	8	1,5
Total	520	100

Cantos		
Nunca	325	62,5
Poucas vezes	119	22,9
Algumas vezes	46	8,8
Muitas vezes	28	5,4
Sempre	2	0,4
Total	520	100
Dupla Possibilidade		
Nunca	88	16,9
Poucas vezes	210	40,4
Algumas vezes	113	21,7
Muitas vezes	93	17,9
Sempre	16	3,1
Total	520	100
Cantos		
Nunca	325	62,5
Poucas vezes	119	22,9
Algumas vezes	46	8,8
Muitas vezes	28	5,4
Sempre	2	0,4
Total	520	100
Dupla Possibilidade		
Nunca	88	16,9
Poucas vezes	210	40,4
Algumas vezes	113	21,7
Muitas vezes	93	17,9
Sempre	16	3,1
Total	520	100
Intervalo/Final		
Nunca	182	35,0
Poucas vezes	232	44,6
Algumas vezes	72	13,8
Muitas vezes	28	5,4
Sempre	6	1,2
Total	520	100
Mais/Menos golos		
Nunca	57	11,0
Poucas vezes	181	34,8
Algumas vezes	129	24,8
Muitas vezes	131	25,2
Sempre	22	4,2
Total	520	100
Mais/Menos golos Intervalo		
Nunca	243	46,7
Poucas vezes	162	31,2
Algumas vezes	64	12,3
Muitas vezes	43	8,3
Sempre	8	1,5
Total	520	100
O jogador X vencerá pelo menos um set		
Nunca	347	66,7
Poucas vezes	116	22,3
Algumas vezes	34	6,5
Muitas vezes	19	3,7
Sempre	4	0,8
Total	520	100
Primeira equipa a marcar		
Nunca	241	46,3
Poucas vezes	193	37,1
Algumas vezes	62	11,9

Apostas desportivas *online* em Portugal

Muitas vezes	19	3,7
Sempre	5	1,0
Total	520	100
Primeira equipa a marcar X pontos num set		
Nunca	369	71,0
Poucas vezes	124	23,8
Algumas vezes	18	3,5
Muitas vezes	7	1,3
Sempre	2	0,4
Total	520	100
Quem marca o X.º golo		
Nunca	350	67,3
Poucas vezes	125	24,0
Algumas vezes	30	5,8
Muitas vezes	13	2,5
Sempre	2	0,4
Total	520	100
Quem marca o X.º ensaio		
Nunca	414	79,6
Poucas vezes	84	16,2
Algumas vezes	16	3,1
Muitas vezes	4	0,8
Sempre	2	0,4
Total	520	100
Resultado exato		
Nunca	277	53,3
Poucas vezes	184	35,4
Algumas vezes	38	7,3
Muitas vezes	16	3,1
Sempre	5	1,0
Total	520	100
Vencedor (equipa/competidor/piloto)		
Nunca	231	44,4
Poucas vezes	147	28,3
Algumas vezes	60	11,5
Muitas vezes	60	11,5
Sempre	22	4,2
Total	520	100
Modalidade		
Lucro com o tipo de aposta mais utilizado		
Não	189	36,3
Sim	237	45,6
Não sei	94	18,1
Total	520	100
Modalidade desportiva		
Andebol		
Nunca	357	68,7
Poucas vezes	133	25,6
Algumas vezes	23	4,4
Muitas vezes	5	1,0
Sempre	2	0,4
Total	520	100
Basquetebol		
Nunca	118	22,7
Poucas vezes	190	36,5
Algumas vezes	154	29,6
Muitas vezes	52	10,0
Sempre	6	1,2
Total	520	100
Futebol		

Nunca	5	1,0
Poucas vezes	5	1,0
Algumas vezes	156	30,0
Muitas vezes	139	26,7
Sempre	215	41,3
Total	520	100
Futsal		
Nunca	196	37,7
Poucas vezes	173	33,3
Algumas vezes	119	22,9
Muitas vezes	25	4,8
Sempre	7	1,3
Total	520	100
Hóquei no gelo		
Nunca	225	43,3
Poucas vezes	148	28,5
Algumas vezes	117	22,5
Muitas vezes	26	5,0
Sempre	4	0,8
Total	520	100
Rugby (ou futebol americano)		
Nunca	361	69,4
Poucas vezes	120	23,1
Algumas vezes	33	6,3
Muitas vezes	5	1,0
Sempre	1	0,2
Total	520	100
Ténis		
Nunca	194	37,3
Poucas vezes	172	33,1
Algumas vezes	106	20,4
Muitas vezes	39	7,5
Sempre	9	1,7
Total	520	100
Ténis de mesa		
Nunca	444	85,4
Poucas vezes	59	11,3
Algumas vezes	16	3,1
Muitas vezes	1	0,2
Sempre	0	0,0
Total	520	100
Voleibol		
Nunca	379	72,9
Poucas vezes	109	21,0
Algumas vezes	27	5,2
Muitas vezes	4	0,8
Sempre	1	0,2
Total	520	100
Estratégia		
Apostar nos jogos do próprio dia e ao vivo		
Nunca	49	9,4
Poucas vezes	124	23,8
Algumas vezes	174	33,5
Muitas vezes	116	22,3
Sempre	57	11,0
Total	520	100
Apostar nos jogos do próprio dia e antes		
Nunca	28	5,4
Poucas vezes	78	15,0
Algumas vezes	196	37,7

Muitas vezes	164	31,5
Sempre	54	10,4
Total	520	100
Apostar em jogos de dias diferentes		
Nunca	92	17,7
Poucas vezes	165	31,7
Algumas vezes	170	32,7
Muitas vezes	77	14,8
Sempre	16	3,1
Total	520	100
Apostar nas mesmas equipas		
Nunca	4	0,8
Poucas vezes	51	9,8
Algumas vezes	221	42,5
Muitas vezes	231	44,4
Sempre	13	2,5
Total	520	100
Colocar habituais vencedores		
Nunca	31	6,0
Poucas vezes	49	9,4
Algumas vezes	187	36,0
Muitas vezes	243	46,7
Sempre	10	1,9
Total	520	100
Utilizar site ou app como apoio		
Nunca	50	9,6
Poucas vezes	85	16,3
Algumas vezes	168	32,3
Muitas vezes	130	25,0
Sempre	87	16,7
Total	520	100
Frequência com que se obtém lucro		
Nunca	13	2,5
Poucas vezes	146	28,1
Algumas vezes	278	53,5
Muitas vezes	74	14,2
Sempre	9	1,7
Total	520	100
Dinheiro gasto		
Até 49€	174	33,5
Entre 50€ e 99€	140	26,9
Entre 100€ e 499€	143	27,5
Entre 500€ e 999€	41	7,9
Entre 1000€ e 4999€	13	2,5
Mais de 5000€	9	1,7
Total	520	100
Lucro obtido		
Até 49€	144	27,7
Entre 50€ e 99€	107	20,6
Entre 100€ e 499€	139	26,7
Entre 500€ e 999€	36	6,9
Entre 1000€ e 4999€	20	3,8
Mais de 5000€	7	1,3
Não respondido	67	12,9
Total	520	100

Apêndice I

Correlação entre as variáveis - Parte I

	Frequencia	Numero Apostas	Modalidade _Simples	Modalidade Combinada_UmTresJogos	Modalidade Combinada_QuatroSeisJogos	Modalidade Combinada e SeteOitoJogos	Modalidade _Multipla	Sistema_StakeFixa	Sistema_StakeNiveis	Sistema_Martingale
Frequencia										
Coefficiente de Correlação	1,000	,236	,141	-0,062	-0,049	,159	-0,006	-0,082	,180	-0,005
P-value		0,000	0,001	0,159	0,265	0,000	0,885	0,063	0,000	0,904
NumeroApostas										
Coefficiente de Correlação	-236	1,000	,104	-0,015	-0,008	0,012	0,003	,144	0,042	-0,005
P-value	0,000	0,017	0,741	0,852	0,777	0,000	0,938	0,001	0,343	0,905
Modalidade_Simples										
Coefficiente de Correlação	,141	1,000	,109	-0,368	,382	,126	,208	,195	,112	,112
P-value	0,001	0,017	0,013	0,000	0,000	0,004	0,000	0,000	0,000	0,011
Modalidade_Combinada_UmTresJogos										
Coefficiente de Correlação	-0,062	-0,015	1,000	,347	0,084	,104	0,055	,342	,185	,185
P-value	0,159	0,741	0,013	0,000	0,056	0,018	0,211	0,000	0,000	0,000
Modalidade_Combinada_QuatroSeisJogos										
Coefficiente de Correlação	-0,049	-0,008	,347	1,000	,593	,221	0,014	0,081	0,071	0,071
P-value	0,265	0,265	0,000	0,000	0,000	0,000	0,746	0,064	0,105	0,105
Modalidade_Combinada_SeteeOitoJogos										
Coefficiente de Correlação	-159	0,012	0,084	0,056	1,000	,344	-0,032	0,058	,088	,088
P-value	0,000	0,777	0,018	0,056	0,000	0,000	0,473	0,183	0,046	0,046
Modalidade_Multipla										
Coefficiente de Correlação	-0,006	0,003	,104	,104	,344	1,000	-0,010	-0,002	,181	,181
P-value	0,885	0,938	0,018	0,018	0,000	0,000	0,814	0,964	0,000	0,000
Sistema_StakeFixa										
Coefficiente de Correlação	-0,082	,144	0,055	0,014	-0,032	-0,010	1,000	,183	0,073	0,073
P-value	0,063	0,001	0,211	0,746	0,473	0,814	0,000	0,000	0,099	0,099
Sistema_StakeNiveis										
Coefficiente de Correlação	,180	0,042	,342	0,081	0,058	-0,002	,183	1,000	,161	,161
P-value	0,000	0,343	0,000	0,064	0,183	0,964	0,000	0,000	0,000	0,000
Sistema_Martingale										
Coefficiente de Correlação	-0,005	-0,005	,185	0,071	,088	,181	0,073	,161	1,000	1,000
P-value	0,904	0,905	0,000	0,105	0,046	0,000	0,099	0,000	0,000	0,000
Sistema_ApostasLivres										
Coefficiente de Correlação	,141	-121	,113	,105	,135	,137	-218	0,003	,184	,184
P-value	0,001	0,006	0,010	0,016	0,002	0,002	0,000	0,952	0,000	0,000
TipoAposta_UmXDois										
Coefficiente de Correlação	-0,045	0,032	,275	,186	0,047	-0,081	,118	,223	-0,009	-0,009

	Frequencia	Numero Apostas	Modalidade _Simples	Modalidade Combinada_UmTresJogos	Modalidade Combinada_QuatroSeisJogos	Modalidade e_Combinada_SetecOitoJogos	Modalidade _Multipla	Sistema_StackeFixa	Sistema_StackeNiveis	Sistema_Marketingale
P-value	0,302	0,460	0,000	0,000	0,000	0,288	0,065	0,007	0,000	0,838
TipoAposta_UmXDoisIntervalo										
Coefficiente de Correlação	-,149	0,071	0,082	,236	,204	,231	,109	0,045	,194	,156
P-value	0,001	0,108	0,061	0,000	0,000	0,000	0,013	0,307	0,000	0,000
TipoAposta_AmbasMarcam										
Coefficiente de Correlação	-,149	,095	,110	,282	,231	,201	0,054	,112	,246	,115
P-value	0,001	0,030	0,012	0,000	0,000	0,000	0,219	0,011	0,000	0,009
TipoAposta_Cantos										
Coefficiente de Correlação	-,260	,211	,272	,176	0,007	-0,021	0,019	,160	,160	,148
P-value	0,000	0,000	0,000	0,000	0,877	0,631	0,667	0,000	0,000	0,001
TipoAposta_DuplaPossibilidade										
Coefficiente de Correlação	-,214	0,049	,180	,339	,165	,091	0,031	,115	,263	0,085
P-value	0,000	0,268	0,000	0,000	0,000	0,037	0,482	0,009	0,000	0,052
TipoAposta_EquipaMarcaAmbasPartes										
Coefficiente de Correlação	-,108	,090	,110	,134	0,080	,095	,168	0,021	,120	,195
P-value	0,014	0,040	0,012	0,002	0,068	0,030	0,000	0,630	0,006	0,000
TipoAposta_GolosEquipa										
Coefficiente de Correlação	-,159	,112	,200	,191	0,042	0,076	0,082	0,076	,161	,165
P-value	0,000	0,010	0,000	0,000	0,336	0,082	0,063	0,084	0,000	0,000
TipoAposta_IntervaloFinal										
Coefficiente de Correlação	-0,029	0,009	-0,033	,147	,181	,239	,231	0,005	,092	,184
P-value	0,515	0,841	0,458	0,001	0,000	0,000	0,000	0,901	0,036	0,000
TipoAposta_MaisMenosGolos										
Coefficiente de Correlação	-,209	0,083	,183	,276	,168	,151	-0,029	,140	,294	0,025
P-value	0,000	0,059	0,000	0,000	0,000	0,001	0,509	0,001	0,000	0,566
TipoAposta_MaisMenosGolosIntervalo										
Coefficiente de Correlação	-,125	,135	,238	,152	0,007	-0,016	0,078	0,067	,162	,141
P-value	0,004	0,002	0,000	0,001	0,881	0,720	0,074	0,129	0,000	0,001
TipoAposta_JogadorVenceSet										
Coefficiente de Correlação	-,118	0,064	,146	,168	0,041	0,079	,099	0,049	0,055	,211
P-value	0,007	0,146	0,001	0,000	0,346	0,073	0,024	0,266	0,213	0,000
TipoAposta_PrimeiraEquipaMarcar										
Coefficiente de Correlação	-0,064	-0,007	0,076	,209	,132	,124	,158	0,041	,124	,225
P-value	0,143	0,875	0,081	0,000	0,003	0,005	0,000	0,347	0,005	0,000
TipoAposta_PrimeiraEquipaMarcarXPont										
Coefficiente de Correlação	-0,037	0,048	0,044	,137	0,042	,098	,177	-0,015	0,045	,263
P-value	0,396	0,275	0,317	0,002	0,335	0,026	0,000	0,729	0,305	0,000

	Frequencia	Numero Apostas	Modalidade _Simples	Modalidade Combinada_3 Jogos	Modalidade Combinada_4 Jogos	Modalidade Combinada_5 Jogos	Modalidade Multipla	Sistema_StackFixa	Sistema_StackNiveis	Sistema_Marketingale
TipoAposta_QuemMarcaGolo										
Coefficiente de Correlação	-0,051	0,038	,099	,167	0,065	,118	,126	0,049	,127	,205
P-value	0,248	0,390	0,024	0,000	0,136	0,007	0,004	0,267	0,004	0,000
TipoAposta_QuemMarcaEnsaio										
Coefficiente de Correlação	-0,013	0,043	0,060	,121	0,044	,087	,202	0,037	0,054	,257
P-value	0,776	0,326	0,174	0,006	0,312	0,048	0,000	0,403	0,223	0,000
TipoAposta_ResultadoExato										
Coefficiente de Correlação	-0,006	0,019	-0,085	0,039	,217	,259	,233	-0,041	0,039	,193
P-value	0,896	0,668	0,053	0,377	0,000	0,000	0,000	0,350	0,379	0,000
TipoAposta_Vencedor										
Coefficiente de Correlação	0,059	-0,037	0,032	,183	0,041	0,076	0,086	-0,078	,095	,135
P-value	0,179	0,399	0,461	0,000	0,349	0,083	0,051	0,077	0,031	0,002
Modalidade_Andebol										
Coefficiente de Correlação	-,184	0,049	-0,023	0,086	,117	,222	,133	-0,001	-0,018	,142
P-value	0,000	0,266	0,593	0,051	0,008	0,000	0,002	0,974	0,679	0,001
Modalidade_Basquetebol										
Coefficiente de Correlação	-,179	0,050	-0,038	,238	,228	,200	0,072	0,063	,138	,093
P-value	0,000	0,260	0,385	0,000	0,000	0,000	0,101	0,154	0,002	0,034
Modalidade_Futebol										
Coefficiente de Correlação	-0,063	0,044	,142	,175	,109	0,024	-0,097	,173	,244	-0,021
P-value	0,150	0,316	0,001	0,000	0,013	0,593	0,027	0,000	0,000	0,631
Modalidade_Futsal										
Coefficiente de Correlação	-,153	0,045	-0,107	,142	,230	,231	0,079	-0,008	0,030	,169
P-value	0,000	0,301	0,015	0,001	0,000	0,000	0,073	0,853	0,497	0,000
Modalidade_Hoquei										
Coefficiente de Correlação	-,213	0,063	-0,081	,087	,127	,251	0,079	0,010	0,070	,100
P-value	0,000	0,149	0,066	0,048	0,004	0,000	0,073	0,812	0,109	0,023
Modalidade_Rugby										
Coefficiente de Correlação	-,092	0,045	-0,011	,103	,098	,200	,170	0,022	0,021	,118
P-value	0,036	0,301	0,802	0,019	0,025	0,000	0,000	0,616	0,628	0,007
Modalidade_Tenis										
Coefficiente de Correlação	-,160	0,085	,121	,200	0,086	,152	0,083	0,045	,170	,117
P-value	0,000	0,054	0,006	0,000	0,050	0,000	0,057	0,308	0,000	0,008
Modalidade_TenisMesa										
Coefficiente de Correlação	-,144	0,079	,103	0,086	0,012	0,060	,127	0,031	-0,006	,168
P-value	0,001	0,071	0,019	0,050	0,777	0,169	0,004	0,481	0,889	0,000
Modalidade_Voleibol										
Coefficiente de Correlação	-,203	,110	0,029	,098	0,068	,187	,181	0,046	0,041	,183
P-value	0,000	0,012	0,515	0,025	0,120	0,000	0,000	0,292	0,355	0,000

	Frequencia	Numero Apostas	Modalidade _Simples	Modalidade Combinada_UmTresJogos	Modalidade Combinada_QuatroSeisJogos	Modalidade e_Combinada_SetecOitoJogos	Modalidade _Multipla	Sistema_StakeFixa	Sistema_StakeNiveis	Sistema_Martingale
Estrategia_JogosProprioDiaAoVivo										
Coefficiente de Correlação	-0,207	0,093	0,302	0,165	-0,093	-0,070	-0,069	0,187	0,267	0,062
P-value	0,000	0,033	0,000	0,000	0,034	0,113	0,114	0,000	0,000	0,156
Estrategia_JogosProprioDiaAntesInicio										
Coefficiente de Correlação	-0,122	0,078	0,081	0,285	0,202	0,115	0,002	0,107	0,195	0,035
P-value	0,005	0,074	0,064	0,000	0,000	0,008	0,955	0,015	0,000	0,426
Estrategia_JogosDiferentes										
Coefficiente de Correlação	0,065	-0,079	-0,041	0,221	0,258	0,237	0,210	0,040	0,039	0,172
P-value	0,136	0,071	0,350	0,000	0,000	0,000	0,000	0,362	0,379	0,000
MesmasEquipas										
Coefficiente de Correlação	-0,097	0,026	0,052	0,120	0,076	0,117	0,001	0,072	0,111	0,063
P-value	0,027	0,561	0,235	0,006	0,085	0,007	0,978	0,103	0,011	0,149
HabituaisVencedores										
Coefficiente de Correlação	0,000	-0,060	-0,063	0,092	0,110	0,098	-0,022	-0,024	0,090	0,018
P-value	0,994	0,170	0,152	0,036	0,012	0,025	0,618	0,580	0,040	0,684
SiteOuAppApoio										
Coefficiente de Correlação	-0,159	0,136	0,103	0,026	-0,005	0,001	-0,151	0,097	0,159	-0,045
P-value	0,000	0,002	0,019	0,560	0,907	0,981	0,001	0,027	0,000	0,302
FrequenciaLucro										
Coefficiente de Correlação	-0,105	-0,023	0,265	0,087	-0,155	-0,239	-0,133	0,133	0,087	0,013
P-value	0,017	0,595	0,000	0,047	0,000	0,000	0,002	0,002	0,047	0,768
DinheiroGasto										
Coefficiente de Correlação	-0,448	0,266	0,052	0,002	0,015	0,103	0,034	0,053	0,019	0,076
P-value	0,000	0,000	0,240	0,972	0,733	0,018	0,441	0,225	0,661	0,082
Lucro*										
Coefficiente de Correlação	-0,318	0,150	0,180	-0,016	-0,110	-0,084	-0,038	0,040	0,045	0,063
P-value	0,000	0,001	0,000	0,740	0,019	0,075	0,417	0,394	0,334	0,183

* Ao contrário das restantes variáveis que possuem N=520, a variável Lucro possui apenas N=453.

Apêndice J

Correlação entre as variáveis - Parte II

	Sistema_Apostas_Livres	TipoAposta_XDois	TipoAposta_Um	TipoAposta_XDoisUm	TipoAposta_Ambas	TipoAposta_Cantos	TipoAposta_Dupla	TipoAposta_Possibilidade	TipoAposta_Marca	TipoAposta_Partes	TipoAposta_MarcaA	TipoAposta_EstaEquipa	TipoAposta_Golo	TipoAposta_Intervalo	TipoAposta_MaisMenos	TipoAposta_Golos
Frequencia	,141 0,001	-0,045 0,302	,149 0,001	,149 0,001	,149 0,001	-260 0,000	-214 0,000	-214 0,000	,108 0,014	,108 0,014	,159 0,000	-0,029 0,515	-209 0,000			
NumeroApostas	,121 0,006	0,032 0,460	0,071 0,108	0,095 0,030	0,095 0,030	,211 0,000	0,049 0,268	0,049 0,268	,090 0,040	,090 0,040	,112 0,010	0,009 0,841	0,083 0,059			
Modalidade_Simples	-0,005 0,902	,156 0,000	0,082 0,061	,110 0,012	,110 0,012	,272 0,000	,180 0,000	,180 0,000	,110 0,012	,110 0,012	,200 0,000	-0,033 0,458	,183 0,000			
Modalidade_Combinada_UmTresJogos	,113 0,010	,275 0,000	,236 0,000	,282 0,000	,282 0,000	,176 0,000	,339 0,000	,339 0,000	,134 0,002	,134 0,002	,191 0,000	,147 0,001	,276 0,000			
Modalidade_Combinada_QuatroSeisJogos	,105 0,016	,186 0,000	,204 0,000	,231 0,000	,231 0,000	0,007 0,877	,165 0,000	,165 0,000	0,080 0,068	0,080 0,068	0,042 0,336	,181 0,000	,168 0,000			
Modalidade_Combinada_SeteeOitoJogos	,135 0,002	0,047 0,288	,231 0,000	,201 0,000	,201 0,000	-0,021 0,631	,091 0,037	,091 0,037	,095 0,030	,095 0,030	0,076 0,082	,239 0,000	,151 0,001			
Modalidade_Multipla	,137 0,002	-0,081 0,065	,109 0,013	0,054 0,219	0,054 0,219	0,019 0,667	0,031 0,482	0,031 0,482	,168 0,000	,168 0,000	0,082 0,063	,231 0,000	-0,029 0,509			
Sistema_StakeFixa	-218 0,000	,118 0,007	0,045 0,307	,112 0,011	,112 0,011	,160 0,000	,115 0,009	,115 0,009	0,021 0,630	0,021 0,630	0,076 0,084	0,005 0,901	,140 0,001			
Sistema_StakeNiveis	0,003 0,952	,223 0,000	,194 0,000	,246 0,000	,246 0,000	,160 0,000	,263 0,000	,263 0,000	,120 0,006	,120 0,006	,161 0,000	,092 0,036	,294 0,000			
Sistema_Martingale	,184 0,000	-0,009 0,838	,156 0,000	,115 0,009	,115 0,009	,148 0,001	0,085 0,052	0,085 0,052	,195 0,000	,195 0,000	,165 0,000	,184 0,000	0,025 0,566			
Sistema_ApostasLivres	1,000	,124 0,004	0,073 0,098	,107 0,015	,107 0,015	-0,009 0,832	,130 0,003	,130 0,003	0,031 0,474	0,031 0,474	,088 0,044	,106 0,016	0,042 0,334			
TipoAposta_UmXDois	,124 0,004	1,000	,414 0,000	,477 0,000	,477 0,000	0,073 0,097	,423 0,000	,423 0,000	0,069 0,116	0,069 0,116	,098 0,025	0,061 0,165	,406 0,000			

	Sistema_Apostas Livres	TipoAposta_UmXDois	TipoAposta_UmXDoisIntervalo	TipoAposta_Ambas Marcam	TipoAposta_Cantos	TipoAposta_DuplaPossibilidade	TipoAposta_EquipaMaisPartes	TipoAposta_Golos	TipoAposta_MaisMenosGolos
TipoAposta_UmXDoisIntervalo									
Coefficiente de Correlação	0,073	,414	1,000	,491	,207	,298	,369	,311	,430
P-value	0,098	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_AmbasMarcam									
Coefficiente de Correlação	,107	,477	,491	1,000	,218	,448	,305	,369	,624
P-value	0,015	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_Cantos									
Coefficiente de Correlação	-0,009	0,073	,207	,218	1,000	,257	,330	,479	,300
P-value	0,832	0,097	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_DuplaPossibilidade									
Coefficiente de Correlação	,130	,423	,298	,448	,257	1,000	,213	,321	,506
P-value	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_EquipaMarcaAmbasPartes									
Coefficiente de Correlação	0,031	0,069	,369	,305	,330	,213	1,000	,471	,267
P-value	0,474	0,116	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_GolosEquipa									
Coefficiente de Correlação	,088	,098	,311	,369	,479	,321	,471	1,000	,417
P-value	0,044	0,025	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_IntervaloFinal									
Coefficiente de Correlação	,106	0,061	,477	,298	,215	,191	,436	,378	,300
P-value	0,016	0,165	0,000	0,000	0,000	0,000	0,000	1,000	0,000
TipoAposta_MaisMenosGolos									
Coefficiente de Correlação	0,042	,406	,430	,624	,300	,506	,267	,417	1,000
P-value	0,334	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_MaisMenosGolosIntervalo									
Coefficiente de Correlação	0,033	0,074	,372	,363	,450	,236	,478	,515	,461
P-value	0,449	0,092	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_JogadorVenceSet									
Coefficiente de Correlação	,104	0,051	,219	,224	,397	,281	,348	,407	,220
P-value	0,018	0,246	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_PrimeiraEquipaMarcar									
Coefficiente de Correlação	0,068	0,037	,360	,286	,236	,225	,366	,370	,277
P-value	0,123	0,397	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_PrimeiraEquipaMarcarXPontosSet									
Coefficiente de Correlação	0,032	-0,034	,234	,157	,332	,223	,435	,385	,148
P-value	0,467	0,433	0,000	0,000	0,000	0,000	0,000	0,000	0,001
TipoAposta_QuemMarcaGolo									
Coefficiente de Correlação	0,064	0,016	,268	,182	,300	,211	,390	,388	,171
P-value	0,148	0,715	0,000	0,000	0,000	0,000	0,000	0,000	0,000

	Sistema_Apostas Livres	TipoAposta_UmXD	TipoAposta_Intervalo	TipoAposta_Ambas Marcas	TipoAposta_Cantos	TipoAposta_Dupla Possibilidade	TipoAposta_Quilpa mbasPartes	TipoAposta_Esta_Quilpa	TipoAposta_Intervalo Final	TipoAposta_MaisMenos Gotos
TipoAposta_QuemMarcaEnsaio										
Coefficiente de Correlação	0,020	-0,052	,211	,160	,331	,174	,440	,350	,325	,112
P-value	0,648	0,234	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,010
TipoAposta_ResultadoExato										
Coefficiente de Correlação	,146	-0,011	,270	,174	,135	0,057	,316	,251	,432	0,079
P-value	0,001	0,801	0,000	0,000	0,002	0,192	0,000	0,000	0,000	0,073
TipoAposta_Vencedor										
Coefficiente de Correlação	,195	0,067	,199	,229	,203	,186	,227	,286	,259	,216
P-value	0,000	0,126	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Modalidade_Andebol										
Coefficiente de Correlação	0,027	-0,007	,211	,187	,137	,140	,153	,192	,255	,147
P-value	0,536	0,872	0,000	0,000	0,002	0,001	0,000	0,000	0,000	0,001
Modalidade_Basquetebol										
Coefficiente de Correlação	0,041	,101	,162	,268	,107	,299	,100	,139	,158	,221
P-value	0,346	0,022	0,000	0,000	0,015	0,000	0,023	0,002	0,000	0,000
Modalidade_Futebol										
Coefficiente de Correlação	0,027	,596	,285	,480	,118	,327	,129	,140	,124	,514
P-value	0,535	0,000	0,000	0,000	0,007	0,000	0,003	0,001	0,005	0,000
Modalidade_Futsal										
Coefficiente de Correlação	0,030	0,015	,171	,226	,154	,227	,219	,259	,268	,205
P-value	0,500	0,727	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Modalidade_Hoquei										
Coefficiente de Correlação	-0,018	-0,003	,153	,171	,124	,178	,095	,159	,182	,181
P-value	0,682	0,954	0,000	0,000	0,004	0,000	0,030	0,000	0,000	0,000
Modalidade_Rugby										
Coefficiente de Correlação	0,064	-0,013	,143	,155	,167	,131	,140	,200	,286	,132
P-value	0,146	0,764	0,001	0,000	0,000	0,003	0,001	0,000	0,000	0,003
Modalidade_Tenis										
Coefficiente de Correlação	,126	0,074	,188	,266	,222	,309	,172	,330	,241	,309
P-value	0,004	0,093	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Modalidade_TenisMesa										
Coefficiente de Correlação	0,046	-0,032	,132	,126	,260	,156	,237	,210	,195	0,070
P-value	0,290	0,463	0,003	0,004	0,000	0,000	0,000	0,000	0,000	0,112
Modalidade_Voleibol										
Coefficiente de Correlação	0,050	-0,070	,117	,120	,188	,168	,178	,234	,261	0,085
P-value	0,258	0,112	0,008	0,006	0,000	0,000	0,000	0,000	0,000	0,051
Estrategia_JogosProprioDiaAoVivo										
Coefficiente de Correlação	-0,051	,215	,142	,241	,269	,259	,154	,192	,132	,349
P-value	0,250	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,003	0,000

	Sistema_ Apostas Livres	TipoApo sta_ Um XDois	TipoApost a_ UmXDo isIntervalo	TipoApost a_ Ambas Marcam	TipoApost a_ Cantos	TipoApost a_ DuplaPo ssibilidade	TipoApost quipaPobasPartes	TipoApo sta_ Golo sEquipa s	TipoApost a_ Interval oFinal	TipoAposta_ MaisMenos Gotos
Estrategia_JogosProprioDiaAntesInicio										
Coefficiente de Correlação	,177	,397	,252	,340	,108	,335	,102	,193	,117	,401
P-value	0,000	0,000	0,000	0,000	0,014	0,000	0,020	0,000	0,008	0,000
Estrategia_JogosDiasDiferentes										
Coefficiente de Correlação	,202	,220	,177	,160	-0,029	,189	0,061	0,081	,186	,138
P-value	0,000	0,000	0,000	0,000	0,516	0,000	0,166	0,066	0,000	0,002
MesmasEquipas										
Coefficiente de Correlação	0,028	,133	,090	,106	0,065	,115	0,059	,103	0,038	,098
P-value	0,528	0,002	0,041	0,015	0,137	0,009	0,183	0,019	0,391	0,026
HabituaisVencedores										
Coefficiente de Correlação	0,085	,128	0,040	0,073	-0,001	0,076	-0,019	0,011	0,037	0,054
P-value	0,054	0,003	0,367	0,098	0,987	0,082	0,673	0,809	0,400	0,218
SiteOuAppApoio										
Coefficiente de Correlação	-0,036	,107	0,062	,097	,105	,151	-0,059	0,079	0,033	,202
P-value	0,415	0,015	0,161	0,027	0,016	0,001	0,182	0,071	0,457	0,000
FrequencialLucro										
Coefficiente de Correlação	-,121	,152	0,014	0,054	,124	0,079	0,044	,118	0,014	0,076
P-value	0,006	0,000	0,747	0,217	0,005	0,074	0,317	0,007	0,758	0,082
DinheiroGasto										
Coefficiente de Correlação	-0,019	0,004	,144	0,074	,227	,098	0,085	,133	,109	,110
P-value	0,662	0,929	0,001	0,092	0,000	0,026	0,052	0,002	0,013	0,012
Lucro*										
Coefficiente de Correlação	-0,068	0,038	0,082	0,051	,179	0,035	0,074	,147	0,040	0,088
P-value	0,148	0,423	0,081	0,279	0,000	0,457	0,116	0,002	0,397	0,060

* Ao contrário das restantes variáveis que possuem N=520, a variável Lucro possui apenas N=453.

Apêndice K

Correlação entre as variáveis - Parte III

	TipoAposta_ MaisMenosGolsIntervalo	TipoAposta_ JogadorVeniceSet	TipoAposta_ PrimeiraEquipaMarcar	TipoAposta_ Pri meiraEquipaMarcarXPontosSet	TipoAposta_ QuemMarca Golo	TipoAposta_ QuemMarca Ensaio
Frequencia						
Coefficiente de Correlação	-,125	-,118	-0,064	-0,037	-0,051	-0,013
P-value	0,004	0,007	0,143	0,396	0,248	0,776
NumeroApostas						
Coefficiente de Correlação	,135	0,064	-0,007	0,048	0,038	0,043
P-value	0,002	0,146	0,875	0,275	0,390	0,326
Modalidade_Simples						
Coefficiente de Correlação	,238	,146	0,076	0,044	,099	0,060
P-value	0,000	0,001	0,081	0,317	0,024	0,174
Modalidade_Combinada_ UmTresJogos						
Coefficiente de Correlação	,152	,168	,209	,137	,167	,121
P-value	0,001	0,000	0,000	0,002	0,000	0,006
Modalidade_Combinada_ QuatroSeisJogos						
Coefficiente de Correlação	0,007	0,041	,132	0,042	0,065	0,044
P-value	0,881	0,346	0,003	0,335	0,136	0,312
Modalidade_Combinada_ SeteOitoJogos						
Coefficiente de Correlação	-0,016	0,079	,124	,098	,118	,087
P-value	0,720	0,073	0,005	0,026	0,007	0,048
Modalidade_Multipla						
Coefficiente de Correlação	0,078	,099	,158	,177	,126	,202
P-value	0,074	0,024	0,000	0,000	0,004	0,000
Sistema_ StakeFixa						
Coefficiente de Correlação	0,067	0,049	0,041	-0,015	0,049	0,037
P-value	0,129	0,266	0,347	0,729	0,267	0,403
Sistema_ StakeNiveis						
Coefficiente de Correlação	,162	0,055	,124	0,045	,127	0,054
P-value	0,000	0,213	0,005	0,305	0,004	0,223
Sistema_ Martingale						
Coefficiente de Correlação	,141	,211	,225	,263	,205	,257
P-value	0,001	0,000	0,000	0,000	0,000	0,000
Sistema_ ApostasLivres						
Coefficiente de Correlação	0,033	,104	0,068	0,032	0,064	0,020
P-value	0,449	0,018	0,123	0,467	0,148	0,648
TipoAposta_ UmXDois						
Coefficiente de Correlação	0,074	0,051	0,037	-0,034	0,016	-0,052
P-value	0,092	0,246	0,397	0,433	0,715	0,234

	TipoAposta_ MaisMenosGolosIntervalo	TipoAposta_ JogadorVen	TipoAposta_ PrimeiraEq	TipoAposta_ Pri	TipoAposta_ QuemMarca	TipoAposta_ QuemMarca
	losIntervalo	cesSet	uiPaMarcar	meiraEquipaMa	Golo	Ensaio
TipoAposta_ UmXDoisIntervalo						
Coefficiente de Correlação	,372	,219	,360	,234	,268	,211
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ AmbasMarcam						
Coefficiente de Correlação	,363	,224	,286	,157	,182	,160
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ Cantos						
Coefficiente de Correlação	,450	,397	,236	,332	,300	,331
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ DuplaPossibilidade						
Coefficiente de Correlação	,236	,281	,225	,223	,211	,174
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ EquipaMarcaAmbasPartes						
Coefficiente de Correlação	,478	,348	,366	,435	,390	,440
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ GolosEquipa						
Coefficiente de Correlação	,515	,407	,370	,385	,388	,350
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ IntervaloFinal						
Coefficiente de Correlação	,359	,263	,405	,378	,314	,325
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ MaisMenosGolos						
Coefficiente de Correlação	,461	,220	,277	,148	,171	,112
P-value	0,000	0,000	0,000	0,001	0,000	0,010
TipoAposta_ MaisMenosGolosIntervalo						
Coefficiente de Correlação	1,000	,373	,410	,390	,323	,412
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ JogadorVenceSet						
Coefficiente de Correlação	,373	1,000	,402	,614	,430	,554
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ PrimeiraEquipaMarcar						
Coefficiente de Correlação	,410	,402	1,000	,464	,509	,447
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ PrimeiraEquipaMarcarXPontosSet						
Coefficiente de Correlação	,390	,614	,464	1,000	,592	,730
P-value	0,000	0,000	0,000	0,000	0,000	0,000
TipoAposta_ QuemMarcaGolo						
Coefficiente de Correlação	,323	,430	,509	,592	1,000	,678
P-value	0,000	0,000	0,000	0,000	0,000	0,000

	TipoAposta_ MaisMenosGolosIntervalo	TipoAposta_ JogadorVen	TipoAposta_ PrimeiraEq	TipoAposta_ meiraEquipaMa	TipoAposta_ Pri	TipoAposta_ QuemMarca	TipoAposta_ QuemMarca	TipoAposta_ Ensaio
TipoAposta_ QuemMarcaEnsaio								
Coefficiente de Correlação	,412	,554	,447	,730	,678	1,000		
P-value	0,000	0,000	0,000	0,000	0,000			
TipoAposta_ ResultadoExato								
Coefficiente de Correlação	,279	,246	,333	,399	,320	,389		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
TipoAposta_ Vencedor								
Coefficiente de Correlação	,247	,341	,271	,306	,278	,316		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
Modalidade_ Andebol								
Coefficiente de Correlação	,216	,213	,251	,270	,217	,245		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
Modalidade_ Basquetebol								
Coefficiente de Correlação	,097	,226	,194	,167	,133	,126		
P-value	0,028	0,000	0,000	0,000	0,002	0,004		
Modalidade_ Futebol								
Coefficiente de Correlação	,147	0,034	0,044	-0,021	0,035	-0,047		
P-value	0,001	0,434	0,322	0,630	0,431	0,288		
Modalidade_ Futsal								
Coefficiente de Correlação	,173	,238	,307	,234	,184	,191		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
Modalidade_ Hoquei								
Coefficiente de Correlação	,149	,183	,263	,148	,111	,089		
P-value	0,001	0,000	0,000	0,001	0,011	0,043		
Modalidade_ Rugby								
Coefficiente de Correlação	,184	,260	,273	,265	,217	,238		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
Modalidade_ Tenis								
Coefficiente de Correlação	,246	,475	,303	,289	,198	,191		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
Modalidade_ TenisMesa								
Coefficiente de Correlação	,268	,323	,278	,308	,277	,352		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
Modalidade_ Voleibol								
Coefficiente de Correlação	,230	,320	,284	,294	,211	,272		
P-value	0,000	0,000	0,000	0,000	0,000	0,000		
Estrategia_ JogosProprioDiaAoVivo								
Coefficiente de Correlação	,221	,103	,130	,113	,108	0,055		
P-value	0,000	0,018	0,003	0,010	0,013	0,211		

	TipoAposta_ MaisMenosGolosIntervalo	TipoAposta_ JogadorVen	TipoAposta_ PrimeiraEq	TipoAposta_ meiraEquiparcarXPontosSet	TipoAposta_ Pri QuemMarca Golo	TipoAposta_ QuemMarca Ensalo
Estrategia_JogosProprioDiaAntesInicio						
Coefficiente de Correlação	,091	,087	0,083	0,022	0,013	-0,030
P-value	0,038	0,048	0,058	0,614	0,772	0,493
Estrategia_JogosDiasDiferentes						
Coefficiente de Correlação	0,028	,126	,182	,102	,140	,097
P-value	0,526	0,004	0,000	0,020	0,001	0,026
MesmasEquipas						
Coefficiente de Correlação	-0,005	,119	0,076	0,048	,112	0,073
P-value	0,917	0,007	0,085	0,274	0,011	0,096
HabituaisVencedores						
Coefficiente de Correlação	-,129	-0,036	0,060	-0,007	0,020	-0,035
P-value	0,003	0,417	0,175	0,878	0,652	0,422
SiteOuAppApoio						
Coefficiente de Correlação	0,060	0,003	0,004	-0,015	0,001	-0,068
P-value	0,175	0,954	0,922	0,733	0,974	0,122
FrequenciaLucro						
Coefficiente de Correlação	0,049	,102	,114	0,054	0,081	0,047
P-value	0,265	0,020	0,009	0,220	0,065	0,284
DinheiroGasto						
Coefficiente de Correlação	,126	,104	0,079	,098	,093	,099
P-value	0,004	0,017	0,071	0,025	0,033	0,024
Lucro*						
Coefficiente de Correlação	,103	,160	0,084	,099	0,079	0,073
P-value	0,029	0,001	0,074	0,036	0,091	0,120

* Ao contrário das restantes variáveis que possuem N=520, a variável Lucro possui apenas N=453.

Apêndice L

Correlação entre as variáveis - Parte IV

	TipoAposta a_ Resulta doExato	TipoAposta Vencedor	Modalidad e_Andeбол	Modalidade_B asquetebol	Modalidad e_Futebol	Modalidade _Futsal	Modalidade _Hoquei	Modalidade _Rugby	Modalidade _Tenis	Modalidade _TenisMesa
Frequencia										
Coefficiente de Correlação	-0,006	0,059	-,184	-,179	-0,063	-,153	-,213	-,092	-,160	-,144
P-value	0,896	0,179	0,000	0,000	0,150	0,000	0,000	0,036	0,000	0,001
NumeroApostas										
Coefficiente de Correlação	0,019	-0,037	0,049	0,050	0,044	0,045	0,063	0,045	0,085	0,079
P-value	0,668	0,399	0,266	0,260	0,316	0,301	0,149	0,301	0,054	0,071
Modalidade_Simples										
Coefficiente de Correlação	-0,085	0,032	-0,023	-0,038	,142	-,107	-0,081	-0,011	,121	,103
P-value	0,053	0,461	0,593	0,385	0,001	0,015	0,066	0,802	0,006	0,019
Modalidade_Combinada_ UmTresJogos										
Coefficiente de Correlação	0,039	,183	0,086	,238	,175	,142	,087	,103	,200	0,086
P-value	0,377	0,000	0,051	0,000	0,000	0,001	0,048	0,019	0,000	0,050
Modalidade_Combinada_ QuatroSeisJogos										
Coefficiente de Correlação	,217	0,041	,117	,228	,109	,230	,127	,098	0,086	0,012
P-value	0,000	0,349	0,008	0,000	0,013	0,000	0,004	0,025	0,050	0,777
Modalidade_Combinada_ SeteOitoJogos										
Coefficiente de Correlação	,259	0,076	,222	,200	0,024	,231	,251	,200	,152	0,060
P-value	0,000	0,083	0,000	0,000	0,593	0,000	0,000	0,000	0,000	0,169
Modalidade_Multipla										
Coefficiente de Correlação	,233	0,086	,133	0,072	-,097	0,079	0,079	,170	0,083	,127
P-value	0,000	0,051	0,002	0,101	0,027	0,073	0,073	0,000	0,057	0,004
Sistema_ StakeFixa										
Coefficiente de Correlação	-0,041	-0,078	-0,001	0,063	,173	-0,008	0,010	0,022	0,045	0,031
P-value	0,350	0,077	0,974	0,154	0,000	0,853	0,812	0,616	0,308	0,481
Sistema_ StakeNiveis										
Coefficiente de Correlação	0,039	,095	-0,018	,138	,244	0,030	0,070	0,021	,170	-0,006
P-value	0,379	0,031	0,679	0,002	0,000	0,497	0,109	0,628	0,000	0,889
Sistema_ Martingale										
Coefficiente de Correlação	,193	,135	,142	,093	-0,021	,169	,100	,118	,117	,168
P-value	0,000	0,002	0,001	0,034	0,631	0,000	0,023	0,007	0,008	0,000
Sistema_ ApostasLivres										
Coefficiente de Correlação	,146	,195	0,027	0,041	0,027	0,030	-0,018	0,064	,126	0,046
P-value	0,001	0,000	0,536	0,346	0,535	0,500	0,682	0,146	0,004	0,290
TipoAposta_ UmXDois										
Coefficiente de Correlação	-0,011	0,067	-0,007	,101	,596	0,015	-0,003	-0,013	0,074	-0,032
P-value	0,801	0,126	0,872	0,022	0,000	0,727	0,954	0,764	0,093	0,463

Apostas desportivas *online* em Portugal

	TipoApost a_ Resulta doExato	TipoAposta_ Vencedor	Modalidad e_Andebo l	Modalidade_B asquetebo l	Modalidad e_Futebo l	Modalidade _Futsal	Modalidade _Hoquei	Modalidade _Rugby	Modalidade _Tenis	Modalidade _TenisMesa
TipoAposta_ UmXDoisIntervalo										
Coefficiente de Correlação	,270	,199	,211	,162	,285	,171	,153	,143	,188	,132
P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,003
TipoAposta_ AmbasMarcam										
Coefficiente de Correlação	,174	,229	,187	,268	,480	,226	,171	,155	,266	,126
P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,004
TipoAposta_ Cantos										
Coefficiente de Correlação	,135	,203	,137	,107	,118	,154	,124	,167	,222	,260
P-value	0,002	0,000	0,002	0,015	0,007	0,000	0,004	0,000	0,000	0,000
TipoAposta_ DuplaPossibilidade										
Coefficiente de Correlação	0,057	,186	,140	,299	,327	,227	,178	,131	,309	,156
P-value	0,192	0,000	0,001	0,000	0,000	0,000	0,000	0,003	0,000	0,000
TipoAposta_ EquipaMarcaAmbasPartes										
Coefficiente de Correlação	,316	,227	,153	,100	,129	,219	,095	,140	,172	,237
P-value	0,000	0,000	0,000	0,023	0,003	0,000	0,030	0,001	0,000	0,000
TipoAposta_ GolosEquipa										
Coefficiente de Correlação	,251	,286	,192	,139	,140	,259	,159	,200	,330	,210
P-value	0,000	0,000	0,000	0,002	0,001	0,000	0,000	0,000	0,000	0,000
TipoAposta_ IntervaloFinal										
Coefficiente de Correlação	,432	,259	,255	,158	,124	,268	,182	,286	,241	,195
P-value	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,000
TipoAposta_ MaisMenosGolos										
Coefficiente de Correlação	0,079	,216	,147	,221	,514	,205	,181	,132	,309	0,070
P-value	0,073	0,000	0,001	0,000	0,000	0,000	0,000	0,003	0,000	0,112
TipoAposta_ MaisMenosGolosIntervalo										
Coefficiente de Correlação	,279	,247	,216	,097	,147	,173	,149	,184	,246	,268
P-value	0,000	0,000	0,000	0,028	0,001	0,000	0,001	0,000	0,000	0,000
TipoAposta_ JogadorVenceSet										
Coefficiente de Correlação	,246	,341	,213	,226	0,034	,238	,183	,260	,475	,323
P-value	0,000	0,000	0,000	0,000	0,434	0,000	0,000	0,000	0,000	0,000
TipoAposta_ PrimeiraEquipaMarcar										
Coefficiente de Correlação	,333	,271	,251	,194	0,044	,307	,263	,273	,303	,278
P-value	0,000	0,000	0,000	0,000	0,322	0,000	0,000	0,000	0,000	0,000
TipoAposta_ PrimeiraEquipaMarcarXPontos										
Coefficiente de Correlação	,399	,306	,270	,167	-0,021	,234	,148	,265	,289	,308
P-value	0,000	0,000	0,000	0,000	0,630	0,000	0,001	0,000	0,000	0,000
TipoAposta_ QuemMarcaGolo										
Coefficiente de Correlação	,320	,278	,217	,133	0,035	,184	,111	,217	,198	,277
P-value	0,000	0,000	0,000	0,002	0,431	0,000	0,011	0,000	0,000	0,000

	TipoApost a_ Resulta doExato	TipoAposta_ Vencedor	Modalidad e_Andeбол	Modalidade_B asquetebol	Modalidad e_Futeбол	Modalidade _Futsal	Modalidade _Hoquei	Modalidade _Rugby	Modalidade _Tenis	Modalidade _TenisMesa
TipoAposta_ QuemMarcaEnsaio										
Coefficiente de Correlação	,389	,316	,245	,126	-0,047	,191	,089	,238	,191	,352
P-value	0,000	0,000	0,000	0,004	0,288	0,000	0,043	0,000	0,000	0,000
TipoAposta_ ResultadoExato										
Coefficiente de Correlação	1,000	,259	,205	,095	-0,015	,226	,140	,200	,115	,213
P-value	0,000	0,000	0,000	0,031	0,731	0,000	0,001	0,000	0,009	0,000
TipoAposta_ Vencedor										
Coefficiente de Correlação	,259	1,000	,112	,094	0,068	,152	0,005	,122	,187	,135
P-value	0,000	0,000	0,010	0,032	0,122	0,001	0,904	0,005	0,000	0,002
Modalidade_ Andebol										
Coefficiente de Correlação	,205	,112	1,000	,309	-0,022	,425	,422	,436	,336	,376
P-value	0,000	0,010	0,000	0,000	0,616	0,000	0,000	0,000	0,000	0,000
Modalidade_ Basquetebol										
Coefficiente de Correlação	,095	,094	,309	1,000	0,055	,391	,518	,280	,439	,163
P-value	0,031	0,032	0,000	0,000	0,210	0,000	0,000	0,000	0,000	0,000
Modalidade_ Futebol										
Coefficiente de Correlação	-0,015	0,068	-0,022	0,055	1,000	0,079	0,012	-0,003	0,081	-0,061
P-value	0,731	0,122	0,616	0,210	0,000	0,072	0,783	0,951	0,066	0,167
Modalidade_ Futsal										
Coefficiente de Correlação	,226	,152	,425	,391	0,079	1,000	,453	,329	,390	,291
P-value	0,000	0,001	0,000	0,000	0,072	0,000	0,000	0,000	0,000	0,000
Modalidade_ Hoquei										
Coefficiente de Correlação	,140	0,005	,422	,518	0,012	,453	1,000	,458	,403	,296
P-value	0,001	0,904	0,000	0,000	0,783	0,000	0,000	0,000	0,000	0,000
Modalidade_ Rugby										
Coefficiente de Correlação	,200	,122	,436	,280	-0,003	,329	,458	1,000	,363	,461
P-value	0,000	0,005	0,000	0,000	0,951	0,000	0,000	0,000	0,000	0,000
Modalidade_ Tenis										
Coefficiente de Correlação	,115	,187	,336	,439	0,081	,390	,403	,363	1,000	,325
P-value	0,009	0,000	0,000	0,000	0,066	0,000	0,000	0,000	0,000	0,000
Modalidade_ TenisMesa										
Coefficiente de Correlação	,213	,135	,376	,163	-0,061	,291	,296	,461	,325	1,000
P-value	0,000	0,002	0,000	0,000	0,167	0,000	0,000	0,000	0,000	0,000
Modalidade_ Voleibol										
Coefficiente de Correlação	,239	,093	,527	,272	-,110	,359	,401	,516	,473	,634
P-value	0,000	0,035	0,000	0,000	0,012	0,000	0,000	0,000	0,000	0,000
Estrategia_ JogosProprioDiaAoVivo										
Coefficiente de Correlação	-0,011	0,046	0,032	,164	,361	,098	0,040	-0,003	,141	,093
P-value	0,811	0,299	0,471	0,000	0,000	0,025	0,358	0,946	0,001	0,034

Apostas desportivas *online* em Portugal

	TipoApost a_Resulta doExato	TipoAposta_ Vencedor	Modalidad e_Andebo l	Modalidade_B asquetebo l	Modalidad e_Futebo l	Modalidade _Futsal	Modalidade _Hoquei	Modalidade _Rugby	Modalidade _Tenis	Modalidade _TenisMesa
Estrategia_JogosProprioDiaAntesInicio										
Coefficiente de Correlação	0,020	,146	0,056	,196	,434	0,079	0,082	0,029	,145	-0,016
P-value	0,644	0,001	0,205	0,000	0,000	0,071	0,062	0,516	0,001	0,714
Estrategia_JogosDiasDiferentes										
Coefficiente de Correlação	,150	,149	0,049	0,046	,132	,149	0,065	,091	,088	-0,015
P-value	0,001	0,001	0,265	0,296	0,002	0,001	0,140	0,037	0,044	0,735
MesmasEquipas										
Coefficiente de Correlação	,113	,088	0,003	0,070	,157	0,048	-0,015	-0,038	,095	-0,062
P-value	0,010	0,044	0,950	0,112	0,000	0,274	0,729	0,384	0,031	0,156
HabituaisVencedores										
Coefficiente de Correlação	0,054	0,084	-0,110	-0,018	,126	0,044	-0,044	-0,076	-0,009	-0,076
P-value	0,216	0,055	0,012	0,687	0,004	0,320	0,312	0,084	0,832	0,082
SiteOuAppApoio										
Coefficiente de Correlação	-0,047	0,005	-0,022	0,085	,136	0,069	0,045	0,026	,142	-0,030
P-value	0,281	0,905	0,622	0,053	0,002	0,114	0,303	0,560	0,001	0,490
FrequenciaLucro										
Coefficiente de Correlação	-0,021	,098	-0,014	,140	,129	0,037	-0,017	0,011	,110	0,074
P-value	0,632	0,026	0,754	0,001	0,003	0,404	0,698	0,811	0,012	0,091
DinheiroGasto										
Coefficiente de Correlação	,126	-0,022	,106	0,047	0,041	,119	,097	0,064	0,081	,132
P-value	0,004	0,622	0,016	0,282	0,346	0,006	0,027	0,144	0,066	0,003
Lucro*										
Coefficiente de Correlação	-0,026	0,075	,107	0,048	0,008	0,089	0,051	0,024	,117	,100
P-value	0,588	0,113	0,023	0,303	0,863	0,060	0,274	0,612	0,013	0,033

* Ao contrário das restantes variáveis que possuem N=520, a variável Lucro possui apenas N=453.

Apêndice M

Correlação entre as variáveis - Parte V

	Modalidade _Voleibol	Estrategia_Jo gosProprioDia AoVivo	Estrategia_Jo gosProprioDia AntesInicio	Estrategia_Jo gosDiasDifere ntes	Mesmas Equipas	HabituaisVe nedores	SiteOuAppA pto	Frequen ciaLucro	Dinheiro Gasto	Lucro
Frequencia										
Coefficiente de Correlação	-,203	-,207	-,122	0,065	-,097	0,000	-,159	-,105	-,448	-,318
P-value	0,000	0,000	0,005	0,136	0,027	0,994	0,000	0,017	0,000	0,000
NumeroApostas										
Coefficiente de Correlação	,110	,093	0,078	-0,079	0,026	-0,060	,136	-0,023	,266	,150
P-value	0,012	0,033	0,074	0,071	0,561	0,170	0,002	0,595	0,000	0,001
Modalidade_Simples										
Coefficiente de Correlação	0,029	,302	0,081	-0,041	0,052	-0,063	,103	,265	0,052	,180
P-value	0,515	0,000	0,064	0,350	0,235	0,152	0,019	0,000	0,240	0,000
Modalidade_Combinada_UmTresJogos										
Coefficiente de Correlação	,098	,165	,285	,221	,120	,092	0,026	,087	0,002	-0,016
P-value	0,025	0,000	0,000	0,000	0,006	0,036	0,560	0,047	0,972	0,740
Modalidade_Combinada_QuatroSeisJogos										
Coefficiente de Correlação	0,068	-,093	,202	,258	0,076	,110	-0,005	-,155	0,015	-,110
P-value	0,120	0,034	0,000	0,000	0,085	0,012	0,907	0,000	0,733	0,019
Modalidade_Combinada_SeteOitoJogos										
Coefficiente de Correlação	,187	-0,070	,115	,237	,117	,098	0,001	-,239	,103	-0,084
P-value	0,000	0,113	0,008	0,000	0,007	0,025	0,981	0,000	0,018	0,075
Modalidade_Multipla										
Coefficiente de Correlação	,181	-0,069	0,002	,210	0,001	-0,022	-,151	-,133	0,034	-0,038
P-value	0,000	0,114	0,955	0,000	0,978	0,618	0,001	0,002	0,441	0,417
Sistema_StakeFixa										
Coefficiente de Correlação	0,046	,187	,107	0,040	0,072	-0,024	,097	,133	0,053	0,040
P-value	0,292	0,000	0,015	0,362	0,103	0,580	0,027	0,002	0,225	0,394
Sistema_StakeNiveis										
Coefficiente de Correlação	0,041	,267	,195	0,039	,111	,090	,159	,087	0,019	0,045
P-value	0,355	0,000	0,000	0,379	0,011	0,040	0,000	0,047	0,661	0,334
Sistema_Martingale										
Coefficiente de Correlação	,183	0,062	0,035	,172	0,063	0,018	-0,045	0,013	0,076	0,063
P-value	0,000	0,156	0,426	0,000	0,149	0,684	0,302	0,768	0,082	0,183
Sistema_ApostasLivres										
Coefficiente de Correlação	0,050	-0,051	,177	,202	0,028	0,085	-0,036	-,121	-0,019	-0,068
P-value	0,258	0,250	0,000	0,000	0,528	0,054	0,415	0,006	0,662	0,148
TipoAposta_UmXDois										
Coefficiente de Correlação	-0,070	,215	,397	,220	,133	,128	,107	,152	0,004	0,038
P-value	0,112	0,000	0,000	0,000	0,002	0,003	0,015	0,000	0,929	0,423

	Modalidade _Voleibol	Estrategia_Jo gosProprioDia AoVivo	Estrategia_Jo gosProprioDia AntesInicio	Estrategia_Jo gosDiasDifere ntes	Mesmas Equipas	HabituaisVe nedores	SiteOuAppA poio	Frequen ciaLucro	Dinheiro Gasto	Lucro
TipoAposta_ UmXDoisIntervalo										
	,117	,142	,252	,177	,090	0,040	0,062	0,014	,144	0,082
Coefficiente de Correlação	0,008	0,001	0,000	0,000	0,041	0,367	0,161	0,747	0,001	0,081
P-value										
TipoAposta_ AmbasMarcam										
	,120	,241	,340	,160	,106	0,073	,097	0,054	0,074	0,051
Coefficiente de Correlação	0,006	0,000	0,000	0,000	0,015	0,098	0,027	0,217	0,092	0,279
P-value										
TipoAposta_ Cantos										
	,188	,269	,108	-0,029	0,065	-0,001	,105	,124	,227	,179
Coefficiente de Correlação	0,000	0,000	0,014	0,516	0,137	0,987	0,016	0,005	0,000	0,000
P-value										
TipoAposta_ DuplaPossibilidade										
	,168	,259	,335	,189	,115	0,076	,151	0,079	,098	0,035
Coefficiente de Correlação	0,000	0,000	0,000	0,000	0,009	0,082	0,001	0,074	0,026	0,457
P-value										
TipoAposta_ EquipaMarcaAmbasPartes										
	,178	,154	,102	0,061	0,059	-0,019	-0,059	0,044	0,085	0,074
Coefficiente de Correlação	0,000	0,000	0,020	0,166	0,183	0,673	0,182	0,317	0,052	0,116
P-value										
TipoAposta_ GolosEquipa										
	,234	,192	,193	0,081	,103	0,011	0,079	,118	,133	,147
Coefficiente de Correlação	0,000	0,000	0,000	0,066	0,019	0,809	0,071	0,007	0,002	0,002
P-value										
TipoAposta_ IntervaloFinal										
	,261	,132	,117	,186	0,038	0,037	0,033	0,014	,109	0,040
Coefficiente de Correlação	0,000	0,003	0,008	0,000	0,391	0,400	0,457	0,758	0,013	0,397
P-value										
TipoAposta_ MaisMenosGolos										
	0,085	,349	,401	,138	,098	0,054	,202	0,076	,110	0,088
Coefficiente de Correlação	0,051	0,000	0,000	0,002	0,026	0,218	0,000	0,082	0,012	0,060
P-value										
TipoAposta_ MaisMenosGolosIntervalo										
	,230	,221	,091	0,028	-0,005	-1,129	0,060	0,049	,126	,103
Coefficiente de Correlação	0,000	0,000	0,038	0,526	0,917	0,003	0,175	0,265	0,004	0,029
P-value										
TipoAposta_ JogadorVenceSet										
	,320	,103	,087	,126	,119	-0,036	0,003	,102	,104	,160
Coefficiente de Correlação	0,000	0,018	0,048	0,004	0,007	0,417	0,954	0,020	0,017	0,001
P-value										
TipoAposta_ PrimeiraEquipaMarcar										
	,284	,130	0,083	,182	0,076	0,060	0,004	,114	0,079	0,084
Coefficiente de Correlação	0,000	0,003	0,058	0,000	0,085	0,175	0,922	0,009	0,071	0,074
P-value										
TipoAposta_ PrimeiraEquipaMarcarXPontosSet										
	,294	,113	0,022	,102	0,048	-0,007	-0,015	0,054	,098	,099
Coefficiente de Correlação	0,000	0,010	0,614	0,020	0,274	0,878	0,733	0,220	0,025	0,036
P-value										
TipoAposta_ QuemMarcaGolo										
	,211	,108	0,013	,140	,112	0,020	0,001	0,081	,093	0,079
Coefficiente de Correlação	0,000	0,013	0,772	0,001	0,011	0,652	0,974	0,065	0,033	0,091
P-value										

	Modalidade_Voleibol	Estrategia_Jo gosProprioDia AoVivo	Estrategia_Jo gosProprioDia AntesInicio	Estrategia_Jo gosDiasDifere ntes	Mesmas Equipas	HabituaisVencedores	SiteOuApp	FrequenciaLucro	Dinheiro Gasto	Lucro
TipoAposta_QuemMarcaEnsaio										
Coefficiente de Correlação	,272	0,055	-0,030	,097	0,073	-0,035	-0,068	0,047	,099	0,073
P-value	0,000	0,211	0,493	0,026	0,096	0,422	0,122	0,284	0,024	0,120
TipoAposta_ResultadoExato										
Coefficiente de Correlação	,239	-0,011	0,020	,150	,113	0,054	-0,047	-0,021	,126	-0,026
P-value	0,000	0,811	0,644	0,001	0,010	0,216	0,281	0,632	0,004	0,588
TipoAposta_Vencedor										
Coefficiente de Correlação	,093	0,046	,146	,149	,088	0,084	0,005	,098	-0,022	0,075
P-value	0,035	0,299	0,001	0,001	0,044	0,055	0,905	0,026	0,622	0,113
Modalidade_Andebol										
Coefficiente de Correlação	,527	0,032	0,056	0,049	0,003	-1,110	-0,022	-0,014	,106	,107
P-value	0,000	0,471	0,205	0,265	0,950	0,012	0,622	0,754	0,016	0,023
Modalidade_Basquetebol										
Coefficiente de Correlação	,272	,164	,196	0,046	0,070	-0,018	0,085	,140	0,047	0,048
P-value	0,000	0,000	0,000	0,296	0,112	0,687	0,053	0,001	0,282	0,303
Modalidade_Futebol										
Coefficiente de Correlação	-,110	,361	,434	,132	,157	,126	,136	,129	0,041	0,008
P-value	0,012	0,000	0,000	0,002	0,000	0,004	0,002	0,003	0,346	0,863
Modalidade_Futsal										
Coefficiente de Correlação	,359	,098	0,079	,149	0,048	0,044	0,069	0,037	,119	0,089
P-value	0,000	0,025	0,071	0,001	0,274	0,320	0,114	0,404	0,006	0,060
Modalidade_Hoquei										
Coefficiente de Correlação	,401	0,040	0,082	0,065	-0,015	-0,044	0,045	-0,017	,097	0,051
P-value	0,000	0,358	0,062	0,140	0,729	0,312	0,303	0,698	0,027	0,274
Modalidade_Rugby										
Coefficiente de Correlação	,516	-0,003	0,029	,091	-0,038	-0,076	0,026	0,011	0,064	0,024
P-value	0,000	0,946	0,516	0,037	0,384	0,084	0,560	0,811	0,144	0,612
Modalidade_Tenis										
Coefficiente de Correlação	,473	,141	,145	,088	,095	-0,009	,142	,110	0,081	,117
P-value	0,000	0,001	0,001	0,044	0,031	0,832	0,001	0,012	0,066	0,013
Modalidade_TenisMesa										
Coefficiente de Correlação	,634	,093	-0,016	-0,015	-0,062	-0,076	-0,030	0,074	,132	,100
P-value	0,000	0,034	0,714	0,735	0,156	0,082	0,490	0,091	0,003	0,033
Modalidade_Voleibol										
Coefficiente de Correlação	1,000	0,052	0,020	0,026	-0,004	-0,081	0,016	0,044	,145	,112
P-value		0,233	0,654	0,560	0,924	0,065	0,722	0,321	0,001	0,017
Estrategia_JogosProprioDiaAoVivo										
Coefficiente de Correlação	0,052	1,000	,354	-,131	0,079	0,047	,184	,186	,159	,183
P-value	0,233		0,000	0,003	0,073	0,280	0,000	0,000	0,000	0,000

	Modalidade _Voleibol	Estrategia_Jo gosProprioDia AoVivo	Estrategia_Jo gosProprioDia AntesInicio	Estrategia_Jo gosDiasDifere ntes	Mesmas Equipas	HabituaisVe ncedores	SiteOuAppA poio	Frequen ciaLucro	Dinheiro Gasto	Lucro
Estrategia_JogosProprioDiaAntesInicio										
Coefficiente de Correlação	0,020	,354 0,000	1,000	,155 0,000	,165 0,000	,106 0,015	,142 0,001	0,054 0,217	0,025 0,573	0,034 0,468
P-value										
Estrategia_JogosDiasDiferentes										
Coefficiente de Correlação	0,026	-,131 0,003	,155 0,000	1,000	,198 0,000	,136 0,002	-0,069 0,117	-0,053 0,224	-0,041 0,350	-,146 0,002
P-value										
MesmasEquipas										
Coefficiente de Correlação	-0,004	0,079 0,073	,165 0,000	,198 0,000	1,000	,386 0,000	,137 0,002	,227 0,000	,139 0,001	0,079 0,095
P-value										
HabituaisVencedores										
Coefficiente de Correlação	-0,081	0,047 0,280	,106 0,015	,136 0,002	,386 0,000	1,000	,177 0,000	0,079 0,073	0,036 0,417	-0,070 0,139
P-value										
SiteOuAppApoio										
Coefficiente de Correlação	0,016	,184 0,000	,142 0,001	-0,069 0,117	,137 0,002	,177 0,000	1,000	,173 0,000	,127 0,004	,109 0,021
P-value										
FrequenciaLucro										
Coefficiente de Correlação	0,044	,186 0,000	0,054 0,217	-0,053 0,224	,227 0,000	0,079 0,073	,173 0,000	1,000	,096 0,028	,416 0,000
P-value										
DinheiroGasto										
Coefficiente de Correlação	,145	,159 0,000	0,025 0,573	-0,041 0,350	,139 0,001	0,036 0,417	,127 0,004	,096 0,028	1,000	,584 0,000
P-value										
Lucro*										
Coefficiente de Correlação	,112	,183 0,000	0,034 0,468	-,146 0,002	0,079 0,095	-0,070 0,139	,109 0,021	,416 0,000	,584 0,000	1,000
P-value										

* Ao contrário das restantes variáveis que possuem N=520, a variável Lucro possui apenas N=453.