

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**CURSO DE MESTRADO EM GEOTECNIA E TRANSPORTES**

Emília Guerra Dias

**ANÁLISE DO COMPORTAMENTO DO CONSUMIDOR DO COMÉRCIO  
ELETRÔNICO EM RELAÇÃO AOS PARÂMETROS DA ENTREGA**

Belo Horizonte  
2021

Emília Guerra Dias

**ANÁLISE DO COMPORTAMENTO DO CONSUMIDOR DO COMÉRCIO  
ELETRÔNICO EM RELAÇÃO AOS PARÂMETROS DA ENTREGA**

Dissertação apresentada ao Programa de Mestrado em Geotecnia e Transportes da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Geotecnia e Transportes.

Área de concentração: Transportes

Orientadora: Professora Dra. Leise Kelli de Oliveira

Belo Horizonte  
2021

D541a

Dias, Emília Guerra.

Análise do comportamento do consumidor do comércio eletrônico em relação aos parâmetros da entrega [recurso eletrônico] / Emília Guerra Dias. – 2021.

1 recurso online (102 f. : il., color.) : pdf.

Orientadora: Leise Kelli de Oliveira.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Apêndices: f. 93-102.

Bibliografia: f. 85-92.

Exigências do sistema: Adobe Acrobat Reader.

1. Transportes - Teses. 2. Transporte urbano - Teses. 3. Comércio eletrônico - Teses. 4. Comportamento do consumidor - Teses. 5. Entrega de mercadorias - Teses. 6. Redes neurais (Computação) - Teses. 7. Regressão logística - Teses. I. Oliveira, Leise Kelli. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.

CDU: 656(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE ENGENHARIA  
CURSO DE MESTRADO EM GEOTECNIA E TRANSPORTES

### FOLHA DE APROVAÇÃO

## ANÁLISE DO COMPORTAMENTO DO CONSUMIDOR DO COMÉRCIO ELETRÔNICO EM RELAÇÃO AOS PARÂMETROS DA ENTREGA

EMÍLIA GUERRA DIAS

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GEOTECNIA E TRANSPORTES, como requisito para obtenção do grau de Mestre em GEOTECNIA E TRANSPORTES, área de concentração TRANSPORTES.

Aprovada em 27 de agosto de 2021, pela banca constituída pelos membros:

Prof.<sup>a</sup>. Leise Kelli de Oliveira (Orientadora) -  
UFMG Prof. Cassiano Augusto Isler - USP  
Prof. Bruno Vieira Bertoncini – UFC

Belo Horizonte, 27 de agosto de 2021.



Documento assinado eletronicamente por Leise Kelli de Oliveira, Professora do Magistério Superior, em 27/08/2021, às 16:59, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Cassiano Augusto Isler, Usuário Externo, em 27/08/2021, às 17:00, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Bruno Vieira Bertoncini, Usuário Externo, em 27/08/2021, às 17:00, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador 0908084 e o código CRC 6640661A.

## **AGRADECIMENTOS**

Gostaria de agradecer e dedicar esta dissertação a Deus e à minha família, sobretudo à minha mãe Lizete, meu pai Nei e minha irmã Letícia, por serem a minha fortaleza. Agradeço a todos os amigos que se preocuparam comigo ao longo desse processo, por terem me cedido momentos de descontração e palavras de estímulo.

Agradeço, igualmente, à minha orientadora Leise Kelli de Oliveira, que acompanha minha trajetória acadêmica desde minha formação superior. Agradeço pela paciência, pelo conhecimento compartilhado e por me ensinar sobre resiliência, a qual me foi muito exigida durante a elaboração deste trabalho. Ademais, estendo meus agradecimentos aos professores e equipe técnica do curso de mestrado de Geotecnia e Transportes, sobretudo ao Leandro Cardoso e ao Elievam Bessa por suas contribuições ao longo da elaboração deste trabalho.

Em especial, dedico este trabalho ao meu marido Pedro Henrique, por ter estado ao meu lado em todas as etapas até aqui, passando por todas as emoções desta formação comigo, tanto as boas quanto as ruins. Te agradeço pela parceria, por testemunhar meus passos. Sinto as suas conquistas como minhas, então saiba que esta conquista também é sua.

*"A ignorância gera mais confiança do que o conhecimento: são os que sabem pouco, e não aqueles que sabem muito, que afirmam de uma forma tão categórica que este ou aquele problema nunca será resolvido pela ciência. "*

(Charles Darwin)

## RESUMO

DIAS, Emília Guerra. **Análise do Comportamento do Consumidor do Comércio Eletrônico em Relação aos Parâmetros da Entrega.** 102f. Dissertação (Mestrado em Geotecnia e Transportes) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

Este trabalho tem por objetivo analisar a influência de atributos do serviço de entrega no comportamento de compra do consumidor do comércio eletrônico. Impulsionada pela maior utilização da Internet nos últimos anos, a expansão do comércio eletrônico favoreceu o aumento da demanda por serviços de entrega domiciliares. Tal fenômeno agrava problemas já inerentes a grandes centros urbanos, como tráfego, degradação da qualidade do ar, ruído, aumento do risco de acidentes, redução da mobilidade urbana e saturação da infraestrutura existente. Pesquisas e análises de mercado mostram que os serviços de entrega influenciam na decisão do consumidor pela compra de um produto. Contudo, na análise de comportamento do consumidor, pouco enfoque é dado aos atributos do serviço de entrega, tais como: prazo, valor do frete e horário da entrega. Para tanto, a presente pesquisa propôs modelos baseados em Rede Neural Artificial (RNA), ferramenta de *machine learning*, com intuito de analisar os atributos que impactam na tomada de decisão do consumidor. A abordagem metodológica adotada iniciou-se pela coleta de dados junto aos consumidores de plataformas de comércio eletrônico, com proposição de questionário elaborado e aplicado via plataforma Google. A confiabilidade das respostas foi analisada pela aplicação do coeficiente alfa de Cronbach. Em seguida, prosseguiu-se com etapa de pré-processamento dos dados que incluiu a codificação e normalização dos dados, além do tratamento da multicolinearidade. Ao final do processo, definiu-se a arquitetura e treinamento da rede neural, seguida da validação e interpretação dos modelos comportamentais, a partir da ferramenta *Partial Dependence Plots* (PDP). Ademais, outra estimação dos modelos foi realizada por Regressão Logística (RL), com a finalidade de complementar a análise dos resultados por uma abordagem estatística. Os modelos estimados por RNA e RL permitiram aferir a relação entre as características do consumidor (seu perfil socioeconômico e hábitos de consumo), além de características do comércio eletrônico, com os atributos de entrega. Os resultados indicam que consumidores de menores faixas etárias, que comprem produtos eletrônicos, jogos, livros e afins são influenciados, sobretudo, pelo atributo valor do frete.

Palavras-Chaves: transporte urbano de carga; comércio eletrônico; comportamento do consumidor; serviço de entrega; rede neural artificial; regressão logística

## ABSTRACT

DIAS, Emília Guerra. **Análise do Comportamento do Consumidor do Comércio Eletrônico em Relação aos Parâmetros da Entrega.** 102f. Dissertação (Mestrado em Geotecnia e Transportes) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

This paper aims to analyze the influence of delivery service attributes on e-consumer purchase behavior. Driven by increased use of the Internet in recent years, the expansion of e-commerce has raised the demand for home delivery services. This phenomenon enhances problems already inherent in large urban centers, such as traffic, pollution, reduced urban mobility, and insufficient of existing infrastructure. Market research and analysis show that delivery services influence the consumer's decision to purchase a product. However, little focus is given to the attributes of the delivery service (such as delivery time, delivery fee, and delivery reception) in the analysis of consumer behavior. To this end, models based on Artificial Neural Networks (ANN), a machine learning tool, are proposed. The methodological approach adopted began with data collection, with the proposition of a questionnaire prepared and applied via Google form. The reliability of the answers was analyzed by applying Cronbach's alpha coefficient. Next, the data pre-processing stage proceeded, which included data coding and normalization, in addition to multicollinearity treatment. At the end of this process, the architecture and training of the ANN were defined, followed by the validation and interpretation of the behavioral models, using the Partial Dependence Plots (PDP) tool. Furthermore, another estimation of the models was performed by Logistic Regression (LR), with the purpose of complementing the analysis by adding statistical perspective. The models estimated by ANN and RL allowed to assess the relationship between consumer characteristics (their socioeconomic profile and consumption habits), as well as e-commerce characteristics, and delivery attributes. The results indicate that consumers in younger age groups who purchase electronic products, games, books and the like are influenced, above all, by delivery fee. Furthermore, the influence of delivery attributes is independent of the frequency of consumption, while privacy, product price and shopping opportunities are the e-commerce characteristics that show statistical significance, with specific effects per delivery attribute. This essay is justified by analyzing the purchase behavior of e- consumers from the perspective of delivery service, establishing relationships between consumer profile and e-commerce characteristics with the influence of delivery attributes.

Keywords: urban freight transport; e-commerce; consumer behaviour; B2C delivery; neural networks; logistics regression.

## LISTA DE FIGURAS

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 3.1 Interação de n neurônios e analogia com neurônio artificial simplificado Perceptron (Basheer e Hajmeer, 2000).....                                                                                                                                                                                                                                                                                                                                                      | 29 |
| Figura 3.2 Função de ativação ReLU - Unidade Linear Retificada.....                                                                                                                                                                                                                                                                                                                                                                                                                | 31 |
| Figura 4.1 Modelo Conceitual.....                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 37 |
| Figura 4.2 Estrutura de Multi-layer Perceptron (MLP). ....                                                                                                                                                                                                                                                                                                                                                                                                                         | 32 |
| Figura 4.3 Função Logística padrão.....                                                                                                                                                                                                                                                                                                                                                                                                                                            | 38 |
| Figura 5.1 Matriz de correlação Spearman – Variáveis independentes.....                                                                                                                                                                                                                                                                                                                                                                                                            | 45 |
| Figura 5.2 PDP – Análise da Influência do ‘valor do frete’ na efetivação da compra pela Internet .....                                                                                                                                                                                                                                                                                                                                                                             | 48 |
| Figura 5.3 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) dos consumidores (à esquerda)   Relação de dependência com o percentual a pagar pelo frete (V05) e com a renda familiar (V08) dos consumidores (à direita).....                                                                                                                                                                                                                      | 49 |
| Figura 5.4 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda)   Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio)   PDP – Relação de dependência com o percentual a pagar pelo frete (V05) e com o preço do produto (V13) (à direita) .....                                                                                                                                                        | 50 |
| Figura 5.5 PDP – Análise da Influência da ‘escolha do dia e da hora de entrega’ na efetivação da compra pela Internet.....                                                                                                                                                                                                                                                                                                                                                         | 51 |
| Figura 5.6 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda)   Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita) .....                                                                                                                                                                                                                              | 52 |
| Figura 5.7 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda)   PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior)  PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita)   PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior)..... | 53 |
| Figura 5.8 PDP – Análise da Importância do ‘prazo da entrega’ para a realização da compra pela Internet.....                                                                                                                                                                                                                                                                                                                                                                       | 54 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 5.9 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda)   PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita) .....                                                                                                                                                                                                                               | 55 |
| Figura 5.10 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda)   PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior)   PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita)   PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior).....      | 56 |
| Figura 5.11 PDP – Análise da Importância do ‘valor do frete’ para a realização da compra pela Internet .....                                                                                                                                                                                                                                                                                                                                                                              | 57 |
| Figura 5.12 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda)   PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita) .....                                                                                                                                                                                                                              | 58 |
| Figura 5.13 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda)   PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior)  PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita)  PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior).....        | 59 |
| Figura 5.14 PDP – Análise da Importância do ‘horário da entrega’ para a realização da compra pela Internet.....                                                                                                                                                                                                                                                                                                                                                                           | 60 |
| Figura 5.15 Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda)   PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita).....                                                                                                                                                                                                                                     | 61 |
| Figura 5.16 PDP – PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda)   PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior)  PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita)  PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior) ..... | 62 |
| Figura 8.1 Perfil socioeconômico dos respondentes.....                                                                                                                                                                                                                                                                                                                                                                                                                                    | 90 |

|                                                                        |    |
|------------------------------------------------------------------------|----|
| Figura 8.2 Perfil dos respondentes quanto aos hábitos de compras ..... | 91 |
| Figura 8.3 Histogramas e Curvas de Kernel – Variáveis V01 a V18.....   | 92 |

## LISTA DE TABELAS

|                                                                                                                                     |    |
|-------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 4.1 Construção do questionário.....                                                                                          | 38 |
| Tabela 4.2 Demonstração da técnica <i>One-Hot Encoding</i> para a variável “Faixa Etária”.....                                      | 28 |
| Tabela 4.3 Estruturação dos dados de entrada dos modelos .....                                                                      | 30 |
| Tabela 4.4 Exemplo de matriz de confusão 2x2.....                                                                                   | 34 |
| Tabela 5.1 Características sociodemográficas da amostra .....                                                                       | 42 |
| Tabela 5.2 Hábitos de compra da amostra .....                                                                                       | 43 |
| Tabela 5.3 Análise descritiva dos parâmetros de valor percebido.....                                                                | 44 |
| Tabela 5.4 Variáveis por modelo de análise.....                                                                                     | 45 |
| Tabela 5.5 Reclassificação de respostas neutras - Comparação da acurácia dos modelos .....                                          | 47 |
| Tabela 5.6 Efeito estimado da variável na influência dos atributos de entrega na efetivação da compra pelo comércio eletrônico..... | 62 |
| Tabela 5.7 Resultados da Regressão Logística.....                                                                                   | 65 |
| Tabela 5.8 Comparação da entre os modelos classificatórios .....                                                                    | 66 |
| Tabela 8.1 Métricas de qualidade RNA – Modelo 1 .....                                                                               | 93 |
| Tabela 8.2 Métricas de qualidade RNA – Modelo 2 .....                                                                               | 93 |
| Tabela 8.3 Métricas de qualidade RNA – Modelo 3 .....                                                                               | 93 |
| Tabela 8.4 Métricas de qualidade RNA – Modelo 4 .....                                                                               | 93 |
| Tabela 8.5 Métricas de qualidade RNA – Modelo 5 .....                                                                               | 93 |

## LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

ARIMA: *Autoregressive Integrated Moving Average*

CNN: *Convolutional Neural Network*

IBGE: Instituto Brasileiro de Geografia e Estatística

LSTM: *Long-Short-Term Memory network*

PDP: *Partial Dependence Plots*

PPD: Pesquisa de Preferência Declarada

ReLU: Unidade Linear Retificada

RL: Regressão Logística

RNA: Rede Neural Artificial

TUC: Transporte Urbano de Carga

UCC: *Urban Consolidation Centres*

## SUMÁRIO

|          |                                                                                        |           |
|----------|----------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>INTRODUÇÃO .....</b>                                                                | <b>15</b> |
| 1.1      | OBJETIVOS DO TRABALHO .....                                                            | 17        |
| 1.2      | ESTRUTURA DA DISSERTAÇÃO .....                                                         | 18        |
| <b>2</b> | <b>REVISÃO DA LITERATURA .....</b>                                                     | <b>19</b> |
| 2.1      | O CONSUMIDOR DO COMÉRCIO ELETRÔNICO .....                                              | 19        |
| 2.1.1    | <i>O comportamento do consumidor sob a perspectiva da entrega .....</i>                | <i>21</i> |
| 2.2      | MODELAGEM DO TRANSPORTE DE CARGA URBANA .....                                          | 24        |
| 2.3      | COMPARAÇÃO ENTRE MODELOS ESTATÍSTICOS E MODELOS DE APRENDIZAGEM .....                  | 25        |
| <b>3</b> | <b>MODELOS DE APRENDIZAGEM COM REDE NEURAL.....</b>                                    | <b>28</b> |
| 3.1      | APLICAÇÕES PRÁTICAS DE REDES NEURAIS ARTIFICIAIS.....                                  | 32        |
| 3.2      | ANÁLISE DE SENSIBILIDADE EM MODELOS DE APRENDIZAGEM.....                               | 33        |
| <b>4</b> | <b>METODOLOGIA .....</b>                                                               | <b>36</b> |
| 4.1      | QUESTIONÁRIO DE PESQUISA E COLETA DE DADOS .....                                       | 36        |
| 4.2      | PRÉ-PROCESSAMENTO DE DADOS .....                                                       | 26        |
| 4.2.1    | <i>Aplicação do coeficiente alfa de Cronbach.....</i>                                  | <i>26</i> |
| 4.2.2    | <i>Codificação e normalização dos dados.....</i>                                       | <i>27</i> |
| 4.2.3    | <i>Tratamento da multicolinearidade .....</i>                                          | <i>29</i> |
| 4.3      | ESTRUTURAÇÃO DAS VARIÁVEIS DE ENTRADA E DE SAÍDA.....                                  | 29        |
| 4.4      | REDE NEURAL: ESTRUTURA E TREINAMENTO .....                                             | 31        |
| 4.4.1    | <i>Validação dos modelos .....</i>                                                     | <i>34</i> |
| 4.4.2    | <i>Interpretação dos resultados.....</i>                                               | <i>36</i> |
| 4.5      | MODELO DE REGRESSÃO LOGÍSTICA .....                                                    | 37        |
| <b>5</b> | <b>RESULTADOS.....</b>                                                                 | <b>41</b> |
| 5.1      | DESCRIÇÃO DA AMOSTRA .....                                                             | 41        |
| 5.2      | DESCRIÇÃO DOS MODELOS .....                                                            | 44        |
| 5.3      | ESTIMAÇÃO DOS MODELOS POR RNA .....                                                    | 46        |
| 5.3.1    | <i>Influência do valor do frete na efetivação da compra .....</i>                      | <i>47</i> |
| 5.3.2    | <i>Influência da escolha do dia e horário de entrega na efetivação da compra .....</i> | <i>50</i> |
| 5.3.3    | <i>Importância do prazo de entrega na compra pela Internet.....</i>                    | <i>53</i> |
| 5.3.4    | <i>Importância do valor do frete na compra pela Internet.....</i>                      | <i>56</i> |
| 5.3.5    | <i>Importância do horário de entrega na compra pela Internet.....</i>                  | <i>59</i> |
| 5.3.6    | <i>Comparação dos modelos estimados por RNA.....</i>                                   | <i>62</i> |
| 5.4      | ESTIMAÇÃO DOS MODELOS POR REGRESSÃO LOGÍSTICA .....                                    | 63        |
| 5.5      | COMPARAÇÃO DOS MODELOS POR RNA E POR RL .....                                          | 66        |
| 5.6      | DISCUSSÃO DOS RESULTADOS .....                                                         | 67        |

|          |                                                            |           |
|----------|------------------------------------------------------------|-----------|
| 5.6.1    | <i>Implicação prática da RNA</i> .....                     | 69        |
| <b>6</b> | <b>CONSIDERAÇÕES FINAIS</b> .....                          | <b>71</b> |
| <b>7</b> | <b>REFERÊNCIAS</b> .....                                   | <b>75</b> |
| <b>8</b> | <b>APÊNDICE</b> .....                                      | <b>83</b> |
| 8.1      | QUESTIONÁRIO APLICADO.....                                 | 83        |
| 8.2      | ANÁLISE GRÁFICA: PERFIL DOS RESPONDENTES .....             | 90        |
| 8.3      | DISTRIBUIÇÃO DE FREQUÊNCIA DAS VARIÁVEIS DOS MODELOS ..... | 92        |
| 8.4      | MÉTRICAS DE QUALIDADE DOS MODELOS ESTIMADOS POR RNA.....   | 93        |

## 1 INTRODUÇÃO

Nos últimos anos, a expansão do comércio eletrônico, impulsionado pela maior disseminação do acesso à Internet, favoreceu o aumento das compras diretas, cada vez mais personalizadas pelos consumidores (Cunha *et al.*, 2013). Segundo Morganti *et al.* (2014), as compras *online* geram aumento da demanda por serviços de entrega que eleva a complexidade da logística do último quilômetro, afetando diretamente o planejamento e a operação dos sistemas de transporte urbanos.

As entregas domiciliares, preferidas pelos consumidores do comércio eletrônico, contribuem para a atomização do fluxo de mercadorias (Cárdenas *et al.*, 2017). A relevância desta discussão chama a atenção de pesquisadores, dos transportadores e de gestores públicos, uma vez que a entrega no último quilômetro se beneficia da alta demanda nas cidades (Cárdenas *et al.*, 2017). A consequência é o agravamento de problemas já inerentes a grandes meios urbanos, como tráfego, poluição, redução da mobilidade urbana e insuficiência da infraestrutura existente (Morganti *et al.*, 2014). Desse modo, Cárdenas *et al.* (2017) ressaltam que cabem aos *stakeholders* gerenciar as diferentes externalidades, próprias a cada região, que podem ressaltar os problemas associados à entrega no último quilômetro. Por exemplo, Dablanc e Rakotonarivo (2010) explicam que as emissões de CO<sub>2</sub>, em Paris, cresceram dramaticamente devido à dispersão geográfica das atividades do comércio eletrônico.

Pesquisas e análises de mercado mostram que os serviços de entrega são um dos fatores que influenciam na decisão do consumidor pela compra de um produto (Morganti *et al.*, 2014). Em geral, o enfoque dos estudos de comportamento do consumidor do comércio eletrônico é dado a atributos do produto, do site ou do preço. Aspectos sociodemográficos foram explorados em análises de aceitação de novas alternativas de entrega (Buldeo Rai *et al.*, 2021a; Yuen *et al.*, 2018). Em seu estudo, Liu *et al.* (2017) concluíram que as características sociodemográficas podem captar os efeitos indiretos dos atributos de entrega. Por sua vez, Lim e Dubinsky (2004) demonstraram que o tipo de produto e a frequência das compras influenciam a percepção do consumidor relativamente ao serviço de entrega. Ademais, algumas características do comércio eletrônico, como privacidade e promoções (Lim e Dubinsky, 2004), qualidade (Prebreza e Shala, 2021; Chincholkar e Sonwaney, 2017), preços (Prebreza e Shala, 2021; Chincholkar e Sonwaney, 2017; Lim e Dubinsky, 2004), conveniência (Prebreza e Shala, 2021; Lim *et al.*, 2009; Lim e Dubinsky, 2004), causam impacto positivo na satisfação do consumidor. Contudo,

poucos estudos consideraram a influência da entrega (no domicílio e em ponto de retirada) na efetivação da compra pelo comércio eletrônico (Prebreza e Shala, 2021; İnci *et al.*, 2019; Vasić *et al.*, 2019; Chincholkar e Sonwaney, 2017; Lim e Dubinsky, 2004).

Esses aspectos, juntamente com a otimização da operação de serviços logísticos no último quilômetro, despertam o interesse de estudar como atributos relacionados à entrega influenciam no comportamento do consumidor no momento da compra *online*. Tal análise se faz pertinente tanto do ponto de vista do consumidor, quanto do ponto de vista das empresas que vendem seus produtos, uma vez que problemas e custos de entrega preocupam a ambos (Morganti *et al.*, 2014). Assim, diante do exposto, cria-se a seguinte questão de pesquisa: *qual a influência de atributos relacionados à entrega na efetivação da compra realizada pelo consumidor do comércio eletrônico?*

Diante do problema científico e das buscas conduzidas na literatura, as seguintes hipóteses são indicadas:

- Hipóteses 1: A influência dos atributos de entrega na efetivação da compra está associada às características socioeconômicas do consumidor do comércio eletrônico;
- Hipótese 2: A influência dos atributos de entrega na efetivação da compra está associada aos hábitos de compra do consumidor do comércio eletrônico;
- Hipótese 3: A influência dos atributos de entrega na efetivação da compra está associada às características do consumo pelo comércio eletrônico, como comodidade, privacidade, preço do produto e ofertas de compra.

Os testes às hipóteses colocadas são endereçados a partir de estudo comportamental. Mitrevski e Hristoski (2011) afirmam que estudar o comportamento do consumidor, focando na interação desses com o sistema do comércio eletrônico, é a chave para satisfazer suas expectativas. A motivação para o estudo comportamental da demanda, para além da finalidade de servir ao público específico dos setores econômicos, está em orientar o planejamento, a implantação e a operação de serviços que envolvam alto aporte de investimento e/ou possam causar profundos impactos ao meio ambiente (Paiva Junior, 2006). Segundo Robertson, Zielinski e Ward (1984), a Ciência Comportamental, aliada a aspectos da Psicologia e da Sociologia, contribui ao processo de planejamento ao fornecer suporte metodológico para a análise, modelagem e compreensão do comportamento individual e coletivo. Compreende-se, portanto, que a

modelagem comportamental é uma ferramenta fundamental para o planejamento urbano e logístico, auxiliando na tomada de decisão e na elaboração de políticas públicas.

Nesse sentido, Farmer (2014) defende abordagens mais realistas, que estudem as regras de decisão dos indivíduos e as implementem diretamente em modelos computacionais que possam replicá-las. Sob essa perspectiva, propõe-se explorar o conceito das técnicas de aprendizagem de máquina (*machine learning*). Os mais recentes focos de aplicação dessas ferramentas têm sido na interpretação da tomada de decisão para auxiliar na elaboração de políticas públicas (Zhao *et al.*, 2019).

Por apresentarem altos níveis de precisão dos resultados, Zhao *et al.* (2019) afirmam que as técnicas de *machine learning* têm sido vastamente aplicadas na realização de previsões. Assim, com vista a analisar a sensibilidade do comportamento dos consumidores do comércio eletrônico a partir de abordagens mais realistas, tomar-se-á por base a aplicação da técnica de *machine learning* denominada Rede Neural Artificial (técnica antiga, de simples manipulação e entendimento) como ferramenta de modelagem. Essa técnica, que foi constantemente desenvolvida ao longo dos anos, é recorrentemente aplicada na literatura para análises classificatórias e preditivas (Basheer e Hajmeer, 2000).

Esse projeto se justifica, portanto, por sua relevância científica, ao analisar o comportamento de compra do consumidor do comércio eletrônico sob a ótica da entrega do produto, além de validar a aplicação de uma ferramenta de *machine learning* (a RNA) na modelagem do comportamento da demanda do comércio eletrônico.

### **1.1 Objetivos do Trabalho**

Este trabalho se propõe a analisar a influência dos atributos do serviço de entrega no comportamento de compra do consumidor do comércio eletrônico. Em complementação, são propostos os seguintes objetivos específicos: (1) determinar a relevância das características do consumidor (perfil socioeconômico e hábitos de compra) na influência dos atributos de entrega para efetivação da compra e (2) determinar a relevância dos hábitos e características do consumo pelo comércio eletrônico na influência dos atributos de entrega para efetivação da compra; e (2) modelar o comportamento dos consumidores do comércio eletrônico por meio da ferramenta de *machine learning* chamada Rede Neural Artificial, com vista a melhorar a interpretação dos aspectos da entrega que influenciam na tomada de decisão.

## **1.2 Estrutura da Dissertação**

A presente dissertação está estruturada de forma a apresentar uma revisão da literatura (capítulo 2) com abordagem sobre (1) as técnicas aplicadas à modelagem de demanda de carga e (2) o comportamento consumidor do comércio (3) estudado sob a perspectiva do serviço de entrega. O capítulo 3 segue com a apresentação de modelos de aprendizagem com rede neural, suas aplicações práticas e condução de análises de sensibilidade.

Na sequência, são descritas as etapas metodológicas seguidas para atingir o objetivo de pesquisa (capítulo 4). Os resultados da aplicação da metodologia propostas são analisados e discutidos no capítulo 5. Por fim, no capítulo 6 apresentam-se os comentários conclusivos em dessa dissertação.

As referências bibliográficas que embasaram o desenvolvimento desta dissertação, bem como a seção de apêndice, com resultados e análises complementares ao escopo, são abordados nos capítulos 7 e 8, respectivamente.

## 2 REVISÃO DA LITERATURA

Este capítulo se destina, a investigar a modelagem comportamental voltados à demanda do comércio eletrônico, com foco no comportamento do consumidor sob a ótica do serviço de entrega. Busca-se, também, introduzir os métodos de modelagem pelos quais a demanda do transporte de carga urbana têm sido abordada, para, por fim, propor uma comparação entre modelos estatísticos, que abrangem os clássicos métodos econométricos, e os modelos de aprendizagem de máquina.

### 2.1 O consumidor do comércio eletrônico

Como consequência do grandioso fenômeno do comércio eletrônico, surge uma nova categoria de consumidor denominado consumidor do comércio eletrônico, ou *e-consumer*, que apresenta as mesmas características do consumidor tradicional, mas que, porém, utiliza a Internet para adquirir produtos e serviços (Mitrevski e Hristoski, 2011).

Estudos empíricos realizados por Swinyard e Smith (2003) e por Cao *et al.* (2013) demonstraram que o público com maior probabilidade de comprar pela Internet é constituído sobretudo por homens, pessoas com maior nível de instrução e grupos sociais mais abastados. No cenário brasileiro, Oliveira *et al.* (2017) demonstram que o comércio eletrônico cresceu em 16% entre 2014 e 2015 e as categorias de produtos mais comercializados *online* são acessórios e produtos de moda, seguidos de cosméticos e produtos de cuidado pessoal (Oliveira *et al.*, 2017). Por sua vez, na Europa, destacam-se roupas e sapatos, seguidos de equipamentos eletrônicos (Morganti *et al.*, 2014).

Mitrevski e Hristoski (2011), por sua vez, estudaram o comportamento do consumidor do comércio eletrônico. Como benefícios da realização de estudos comportamentais para esse público, os autores listam: o aumento da receita da empresa de *e-commerce*, a otimização do design do produto, funcionalidade e interface da *homepage*, e a adequação às condições ideais de qualidade dos serviços (*quality of service* - QoS) do *e-commerce*. A influência da QoS na efetivação das compras *online* foi modelada por Ghavamipoor *et al.* (2017), que identificaram certa sensibilidade do comportamento dos *e-consumers* a esse fator.

Com foco no comportamento de compra do consumidor do comércio eletrônico, Laroche (2009) identificou a influência de fatores associados à percepção do indivíduo (como risco percebido e autoconfiança) na decisão de efetivar a compra *online*. O autor não fez menção à

entrega de mercadorias em domicílio (meio urbano) e o interesse de estudar o comportamento da demanda é realizado sob a ótica do mercado de *e-commerce*.

A análise comportamental do consumidor do comércio eletrônico é novamente focada na otimização da experiência de compra e no aumento de vendas para os comerciantes em estudo realizado por Zeng *et al.* (2018). Estes autores modelaram o padrão de compras móveis de dezenas de milhares de visitantes *online* durante evento festivo na China, com foco na influência de descontos e promoções no aumento do volume de compras. Também na China, Jin *et al.* (2016) analisaram a preferência de consumidores a fim de definir atributos que favoreçam a compra de produtos orgânicos via Internet.

Hsiao (2008) investigou como consumidores avaliam o atributo valor do tempo ao se depararem com a escolha de comprar em loja física ou *online* utilizando pesquisa de preferência declarada (PPD). As PPDs são utilizadas na previsão de demanda de mercado por incorporarem as escolhas individuais e permitirem estimar o modelo de escolha discreta (Louviere e Hensher, 1983). Quanto à escolha entre *e-shopping versus* compra em loja física, segundo Comi e Nuzzolo (2016), poucos estudos investigaram esse aspecto, o que demonstra a necessidade de mais pesquisas, sobretudo na ótica da logística urbana. O aumento de várias soluções de entrega gera mudanças no padrão de fluxos urbanos de carga e na movimentação de veículos nas cidades (Comi e Nuzzolo, 2016), assim as preferências dos compradores *online* devem ser objeto de investigação adicional na definição das operações de último quilômetro (Oliveira *et al.*, 2017).

Interessados no aumento expressivo de entregas urbanas oriundas de compras *online*, Dutra *et al.* (2006) abordam a problemática do último quilômetro e apresentam soluções praticadas em diversas partes do mundo. Tais soluções visam a redução do custo operacional no último quilômetro, por meio da diminuição de viagens improdutivas de entrega. Por sua vez, Cunha *et al.* (2013) abordaram a mesma problemática sobre a perspectiva dos impactos ambientais dos serviços de entregas logísticas em meio urbano. Através de uma investigação da literatura, Cunha *et al.* (2013) afirmam que, apesar de inúmeras alternativas logísticas terem sido implantadas e testadas pelo mundo, o tema mantinha-se pouco explorado no Brasil.

Complementarmente, Yuen *et al.* (2018) pontuam o *gap* existente na literatura em relação aos estudos teóricos que exploram a influência do comportamento dos consumidores, sua seletividade e suas escolhas quanto às entregas no último quilômetro.

### **2.1.1 O comportamento do consumidor sob a perspectiva da entrega**

Enquanto a maioria dos estudos de comportamento do consumidor do comércio eletrônico focam os atributos do produto, do site e do preço na efetivação de compras *online*, Morganti *et al.* (2014) apresentam dados europeus que afirmam que os problemas experimentados com as compras *online* estão principalmente relacionados com a entrega e não com o produto em si. Da perspectiva de consumidores europeus, as principais reclamações quanto aos serviços de entrega foram, em ordem de maior ocorrência: entrega realizada quando não havia ninguém em casa para receber o produto, atraso na entrega, alto custo do serviço de entrega, ausência de rastreabilidade do produto durante o processo de entrega e necessidade de coletar o produto em um local distante (Eurobarometer, 2013).

Os problemas relatados, juntamente dos custos de entrega, preocupam os operadores logísticos, além de chamar a atenção de gestores públicos para a problemática das entregas no último quilômetro (Morganti *et al.*, 2014). Na prática, a entrega no último quilômetro envolve muitas vezes um pacote por porta, o que não permite as economias de escala (Lin *et al.*, 2020). Ademais, esse processo pode resultar em viagens inúteis, caso o destinatário não se encontre na residência. O resultado é um processo de entrega dispendioso (Lin *et al.*, 2020) com impactos diretos no tráfego urbano, tema pouco explorado segundo Morganti *et al.* (2014).

Portanto, como abordar as preferências dos *e-consumers* é um tópico essencial para entender a entrega do comércio eletrônico. Em resposta, alternativas às entregas regulares ao domicílio têm chamado a atenção de operadores logísticos (Lin *et al.*, 2020). De acordo com Holguín-Veras *et al.* (2017), boas políticas de transporte urbano de mercadorias exigem um bom entendimento de como seus destinos, que neste caso se refere aos consumidores do comércio eletrônico, reagem a ela. Desse modo, na literatura recente, estudos comportamentais conduzidos junto aos consumidores do comércio eletrônico focam, principalmente, em analisar a aceitação a tais alternativas.

Entre as alternativas de entrega, Yuen *et al.* (2018) mencionam os *pick-up points*, que também podem ser denominados como estação de auto-coleta (*self-collection facilities*). Nestes

sistemas, o consumidor pode optar por coletar as suas encomendas de acordo com a sua conveniência dentro de uma determinada janela de tempo, o que reduz os custos de oportunidade associados as entregas domiciliares (Agatz *et al.*, 2011). Segundo Yuen *et al.* (2018), dado que a maioria das características identificadas para esse serviço são determinantes significativos da intenção dos consumidores do comércio eletrônico, estudos comportamentais podem ser ferramentas de melhoria para os *pick-up points*. Além disso, Choo (2016) sugere que a receptividade dos consumidores e a sua atitude em relação aos *pick-up points* é igualmente ou até mais importante do que a otimização das redes dessas instalações e do custo dos operadores logísticos, que têm sido o foco principal da literatura existente.

Contudo, Yuen *et al.* (2018) reforçam que, apesar das vantagens dos *pick-up points*, a entrega em domicílio continua a ser o modo mais popular no último quilômetro entre os consumidores de países como Singapura e França, onde, segundo Morganti *et al.* (2014), apenas 10% dos produtos adquiridos *online* são coletados em *pick-up points*. Lin *et al.* (2020) apontam ainda que, na prática, os consumidores podem preferir alternativas que não necessariamente são as mais convenientes para atender às suas demandas. Estudos empíricos e experimentais, como os realizados em 1994, por Hey e Orme, e em 1997, por Sippel, também mostram que os comportamentos de escolha individual são difíceis de capturar e são tipicamente interpretados de forma probabilística (Gul *et al.*, 2014).

Tais considerações reforçam a necessidade de avaliar como os estudos comportamentais têm sido conduzidos na literatura, além de quais parâmetros são utilizados para estudar as preferências do consumidor quanto à entrega do produto no último quilômetro.

No intuito de alinhar a análise comportamental do consumidor do comércio eletrônico com análise de adesão a novos serviços logísticos no último quilômetro, estudos exploratórios foram realizados em Belo Horizonte (Brasil) por Oliveira *et al.* (2017). As autoras aplicaram método de preferência declarada combinado com preferência revelada para investigar o comportamento do consumidor em relação a atributos do processo de recebimento do produto. Tais atributos foram localização, tempo de entrega, disponibilidade de informação e acompanhamento da entrega, custo de transporte, além da propensão dos *e-consumers* a aderir a um novo sistema de recebimento do produto: os *pick-up points*.

Por sua vez, Cherrett *et al.* (2017) conduziram estudos com estudantes residentes em moradias universitárias da Universidade de Southampton, no Reino Unido, com o objetivo de analisar a possibilidade de consolidação das entregas do comércio eletrônico destinadas ao *campus* universitário em centros de consolidação urbana (*Urban Consolidation Centres - UCCs*). Junto aos estudantes, foram obtidos dados sobre a frequência de compras *online*, tempo de entrega e janelas de horário mais adequadas, além da disponibilidade de aceitar diferentes opções de entrega frente a diferentes opções de tempo de recebimento e preço.

Ao separar os alunos entrevistados pelo critério de frequência com que estes realizam compras pela Internet, Cherrett *et al.* (2017) puderam relacionar as chances de sucesso das UCCs à sua aceitação por parte desse grupo de estudantes. Tal fato se deveu ao resultado de análises estatísticas que revelaram que compradores mais frequentes tinham maior probabilidade em utilizar uma gama de opções de entrega que reduzisse o atraso na gratificação do produto, sendo, portanto, menos propensos a aceitar a opção de consolidação das entregas em uma UCC. Os resultados indicaram que esse grupo estava mais preocupado com a velocidade de entrega, pois estavam mais dispostos a pagar para que essa fosse realizada no mesmo dia da compra.

Condições sociodemográficas, horário da viagem e modo de deslocamento foram critérios escolhidos por Liu *et al.* (2017) para estudar o comportamento de viagem de *e-consumers* que coletam seus produtos em *pick-up points* na Suécia. Por sua vez, Yuen *et al.* (2018) associaram paradigmas da psicologia, teoria motivacional, utilidade do consumidor e economia comportamental para estruturar um modelo que permitisse determinar a propensão de consumidores em utilizar *pick-up points*, os quais permitem a consolidação de entregas por locais agrupados (*clusters*).

Para além da identificação dos parâmetros de entrega estudados na literatura, quando as pesquisas não são direcionadas a uma parcela específica da população (como no trabalho de Cherrett *et al.*, 2017), observa-se a tendência de incorporar aspectos socioeconômicos (Yuen *et al.*, 2018 e Lin *et al.*, 2020), demográficos (Liu *et al.*, 2017) e psicológicos (Yuen *et al.*, 2018) nas análises comportamentais dos consumidores do comércio eletrônico. O emprego de parâmetros mais realistas em estudos comportamentais, leva cada vez mais pesquisadores a optar por outros modelos que não os tradicionais modelos de escolha discreta.

## 2.2 Modelagem do transporte urbano de carga

Modelos de demanda são as principais ferramentas do planejamento de transporte nos níveis estratégicos, tático e operacional, uma vez que auxiliam na previsão das necessidades futuras de transporte para pessoas e mercadorias, a fim de fornecer a infraestrutura e os recursos humanos que possibilitam esses deslocamentos (Regan e Garrido, 2002). Ademais, tais modelos desempenham um papel fundamental na proposição e na avaliação da performance e dos impactos de políticas públicas (Comi *et al.*, 2012).

De acordo com uma síntese das pesquisas nas áreas de demanda do transporte de carga e estudos de comportamento dos transportadores, publicada Regan e Garrido (2002), embora a modelagem do transporte de passageiros tenha se desenvolvido tanto como disciplina científica quanto como área de prática profissional, até aquele momento a abordagem de modelagem do transporte de mercadorias encontrava-se num estágio inicial. Em relação ao contexto urbano de carga, os autores afirmam que, até aquele momento, os poucos artigos disponíveis na literatura tratavam principalmente dos fluxos de veículos, especialmente os fluxos de caminhões. Assim, os modelos criados consideravam apenas uma dimensão do transporte de mercadorias, o que aportava inconsistências para as análises conduzidas, uma vez que os fluxos de carga representam a verdadeira demanda e as viagens representam decisões logísticas.

Nesta perspectiva, Comi *et al.* (2012) realizam uma revisão dos métodos e modelos empregados na simulação da demanda de transporte urbano de carga capaz de apoiar a avaliação de cenários de aplicação de soluções de logística urbana em curto prazo. Esses classificam os modelos de acordo com sua unidade de referência: modelos baseados no veículo, modelos baseados nas mercadorias e modelos baseados nas entregas.

No que diz respeito aos modelos baseados no veículo, Comi *et al.* (2012) concluem que, apesar da vantagem da facilidade de coleta de dados, esses são difíceis de usar para análise de previsão, pois não são capazes de contabilizar adequadamente as mudanças nos mecanismos subjacentes à demanda. Por sua vez, a abordagem baseada na mercadoria permite a simulação dos mecanismos fundamentais à procura subjacente por transporte de carga, no entanto uma deficiência desses modelos consiste na necessidade de obter dados em termos de fluxos quantitativos de mercadorias dos operadores de transportes e de logística (Comi *et al.*, 2012). Por fim, segundo Comi *et al.* (2012), os modelos baseados em entregas são mais interessantes, porque utilizam a mesma unidade de referência dos operadores de transporte e logística. Uma

limitação desses modelos diz respeito à mesma dificuldade de obtenção de dados por parte dos operadores do transporte de carga.

Modelos de demanda baseados nas entregas tornam-se ainda mais desafiadores dado o crescimento de serviços de entrega ao consumidor final relacionados com as compras pelo comércio eletrônico que resultam na crescente fragmentação das remessas no último quilômetro, a como é referido o segmento final na cadeia de abastecimento (Morganti *et al.*, 2014). Do ponto de vista urbano, as entregas ao domicílio constituem a solução mais problemática em termos de custos de serviço e organização, contudo esta opção é normalmente preferida pelos consumidores do comércio eletrônico, que procuram serviços expressos, organizados e confiáveis (Credoc, 2010).

A busca pelo aperfeiçoamento da modelagem de demanda se deve, em parte, à necessidade dos setores econômicos de prever o comportamento de seus clientes, além de realizar projeções de venda e consumo (Mitrevski e Hristoski, 2011). Farmer (2014) explica que os típicos modelos econométricos, tão amplamente empregados nesse contexto, assumem que os indivíduos agem de modo racional, tomam suas decisões de forma a maximizar suas preferências individuais e detêm de todas as informações necessárias para tanto. A partir daí, atribui-se a esses agentes uma dada função de utilidade, no intuito de maximizá-la através das decisões do indivíduo (Jeng e Fesenmaier, 2010).

Enquanto, historicamente, os modelos de demanda se baseiam na aplicação de funções de utilidade, Farmer (2014) defende abordagens mais realistas, como modelos computacionais a exemplo das técnicas de aprendizagem de máquina (*machine learning*). Sendo a econometria constituída da aplicação de métodos matemáticos e estatísticos à análise de conjuntos de dados econômicos (Frisch, 1933), o subcapítulo seguinte propõe comparação não exaustiva de modelos estatísticos e modelos de aprendizagem.

### **2.3 Comparação entre Modelos Estatísticos e Modelos de Aprendizagem**

Algumas similaridades podem ser apontadas entre modelos estatísticos (como modelos de escolha discreta) e modelos de aprendizagem de máquina (como as redes neurais artificiais). Ambas as técnicas se preocupam com o planejamento, a combinação de evidências e a assistência à tomada de decisões. Contudo, as principais diferenças são atribuídas à sua filosofia, aos seus objetivos e ao seu desenvolvimento (Karlaftis e Vlahogianni, 2011). Em

termos de sua filosofia, enquanto os modelos de aprendizagem enfatizam a implementação, o campo estatístico dá ênfase à inferência e à estimativa (Hand, 2000). Complementarmente, segundo Breiman (2001), as estatísticas consideram que a origem de dados se baseia em processos estocásticos. Por sua vez, o modelo de aprendizagem assume que os dados foram gerados por um mecanismo de dinâmica desconhecida.

A respeito dos objetivos de cada abordagem, a estatística visa fornecer um modelo - preditor ou classificador - com elementos autoexplicativos, que ofereça uma visão dos dados e da sua estrutura (Wild e Pfannkuch, 1999). Além disso, a estatística explica os fenômenos investigados através da interpretação dos efeitos e sinais marginais, estudando elasticidades e propriedades estimadoras. Por outro lado, a maioria das aplicações de modelos de aprendizagem não têm como alvo a interpretação, mas sim uma representação eficiente - em termos de precisão e tempo de desenvolvimento - das propriedades subjacentes dos dados, com vista a oferecer boas previsões para o fenômeno em estudo (Karlaftis e Vlahogianni, 2011).

Por fim, a diferença fundamental entre as técnicas é o processo de aprendizagem, o qual, independentemente do método de treinamento utilizado, resulta em mais de uma estrutura de modelo. Em outras palavras, a curva de aprendizagem pode ter vários mínimos locais e um modelo pode convergir para várias arquiteturas sequenciais que não estão necessariamente aninhadas (Ripley, 1996). Esta característica inerente torna tais modelos mais flexíveis do que a estatística, uma vez que a forma funcional é aproximada através da aprendizagem e não assumida, como é feito *a priori* na estatística (Warner e Misra, 1996). Assim, de maneira contrastante, os modelos estatísticos apresentam uma única estrutura final, cujo mecanismo de inferência é explícito e de fácil interpretação (Jiang *et al.*, 2016; Karlaftis e Vlahogianni, 2011).

O aumento do uso da autoaprendizagem para compor a análise de políticas públicas e auxiliar na tomada de decisões, têm favorecido a aplicação de modelos de aprendizagem de máquina em projetos de engenharia e científicos (Zhao *et al.*, 2019). Neste contexto, este estudo se propõe a modelar o comportamento do consumidor do comércio através de algoritmo de Rede Neural para representação mais precisa da influência da entrega na efetivação da compra pelo comércio eletrônico. Experiências com modelos de aprendizagem por RNA são abordadas no capítulo 3, a seguir.

Contudo, dadas as considerações anteriores sobre os mecanismos de inferência e fácil interpretação dos modelos estatísticos, propõe-se igualmente proceder com estimação do fenômeno por regressão logística, a fim de complementar e conferir maior robustez às análises de interpretação do comportamento do consumidor do comércio eletrônico.

### 3 MODELOS DE APRENDIZAGEM COM REDE NEURAL

No campo das análises estatísticas, Breiman (2001) contrasta duas culturas e seus diferentes objetivos: a cultura da modelagem de dados e a cultura da modelagem algorítmica. A primeira corresponde à aplicação de funções paramétricas, como modelos lineares, nos quais os parâmetros são facilmente interpretados. Por sua vez, os modelos algorítmicos, mais conhecidos como modelos de aprendizagem de máquina (*machine learning*), aproximam uma função a fim de maximizar a acurácia preditiva. Segundo Zhao e Hastie (2019), modelos de aprendizagem performam significativamente melhor que modelos paramétricos em termos de capacidade preditiva.

O conceito de *machine learning* é apresentado, por Abduljabbar *et al.* (2019), como uma subcategoria do campo da Inteligência Artificial (IA), cujo avanço tem proporcionado melhor desempenho para diferentes indústrias e empresas, incluindo o setor dos transportes. O *machine learning* implica em codificar computadores para se comportarem como um cérebro humano e aprender com base no acesso e tratamento a grandes bancos de dados (*Big Data*), extraindo deles características importantes para resolver problemas complexos (Abduljabbar *et al.*, 2019).

Entre as técnicas de *machine learning*, a Rede Neural Artificial (RNA), desenvolvida no início dos anos 1960 por McIlloach-Hopfield e descrita por Simpson (1990), equivale a um conjunto de nós (“neurônios”), que processam informações de entrada (“estímulos”), e a um conjunto de conexões (“sinapses”), ponderadas por pesos, que “memorizam” ou arquivam as informações. Desenvolvidas para simular o cérebro humano em termos de como o indivíduo armazena, acessa a informação e realiza julgamentos, as RNAs são dispositivos que aprendem através de exemplos e reagem a novas informações, produzindo resultados (Jeng e Fesenmaier, 2010).

A função das redes neurais não é replicar as operações do sistema biológico, mas sim reproduzir a funcionalidade biológica capaz de solucionar problemas complexos (Basheer e Hajmeer, 2000). Segundo Jain *et al.* (1996), a atratividade das RNAs é resultante principalmente da capacidade de processar dados não lineares, de alto paralelismo e robustez, além de apresentar tolerância a falhas e capacidade de aprendizagem, importantes no lido de informações imprecisas e difusas. Por fim, Jain *et al.* (1996) ressaltam a capacidade de generalização da qual as RNAs são dotadas.

Quanto à sua estrutura, cada “neurônio” (conjunto de nós) representa um *layer* da rede neural. Diz-se, assim, que uma rede neural de múltiplas camadas, ou *layers*, equivale a um conjunto de equações algébricas não lineares (Fishwick, 1989). A Figura 3.1 ilustra  $n$  neurônios biológicos com vários sinais de intensidade  $x$  e força (peso) sináptica  $w$  alimentando um neurônio com um limiar de  $b$  (Basheer e Hajmeer, 2000). Paralelamente, tal estrutura é comparada a uma rede neural artificial simplificada, denominada *Perceptron*.

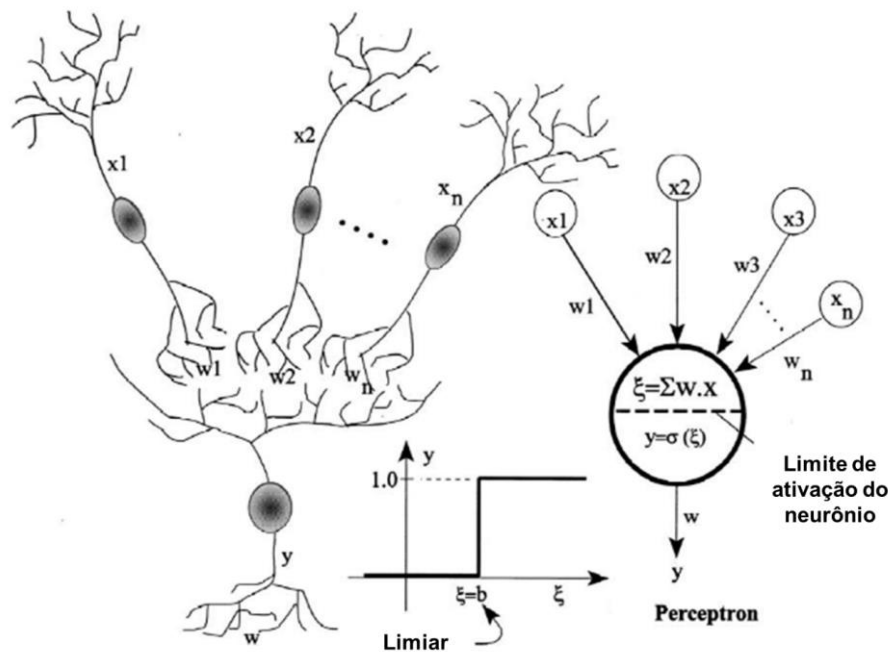


Figura 3.1 Interação de  $n$  neurônios e analogia com neurônio artificial simplificado Perceptron (Basheer e Hajmeer, 2000)

Fishwick (1989) explica que ao utilizar a rede neural para modelar um sistema, primeiro essa deve ser treinada com pares de dados de entrada e saída (*input* e *output*, respectivamente). A aprendizagem, segundo Basheer e Hajmeer (2000), é vista como o processo de atualização da configuração interna do sistema em resposta a estímulos externos, de modo a que este possa realizar uma tarefa específica. Tal reconfiguração inclui modificar a arquitetura da rede, o que implica em ajustar pesos, podar ou criar novas ligações e/ou alterar as regras de ativação dos neurônios (Schalkoff, 1997). Desse modo, a aprendizagem da RNA é consolidada através da experiência, ocorrendo de modo iterativo à medida que a rede é apresentada a exemplos durante uma etapa de treinamento. Uma vez treinada, a rede neural será utilizada para prever dados de saída com base em novos dados de entrada.

Chama-se *Feedforward Neural Network* aquelas RNA em que o processamento dos dados se dá em uma direção única, iniciando no *layer* de *input*, seguindo pelos *layers* intermediários (ocultos) até o *layer* de *output*. Esta categoria é a mais comum entre as redes neurais (Abduljabbar *et al.*, 2019), sendo o *Perceptron* a estrutura mais simplificada existente. Demais categorias de redes neurais são as *Convolutional Neural Network* (CNN), aplicadas ao processamento de imagens, e as *Recurrent Neural Network* (RNN), que processa uma sequência de dados sendo comumente aplicadas em reconhecimento textual, línguas e escrita. Essas redes neurais são frequentemente referidas como técnicas de *Deep Learning*, por apresentarem uma estrutura constituída por múltiplos *layers* ocultos (Abduljabbar *et al.*, 2019).

Nas redes *feedforward*, os dados são combinados em uma rede de *inputs* que são processados pela função de ativação e, na sequência, são transmitidos para o próximo neurônio. A transmissão do sinal entre os neurônios apenas ocorre quando o valor excede o limiar de cada neurônio. Diz-se, então que este foi ativado.

As funções de ativação desempenham um importante papel no treinamento de redes neurais (Ramachandran *et al.*, 2018). Atualmente, a função de ativação mais bem-sucedida e utilizada é a Unidade Linear Retificada (ReLU) (Hahnloser *et al.*, 2000). Essa determina um limiar igual a 0, ou seja, produz 0 quando  $x < 0$ , e inversamente, produz uma função linear, quando  $x \geq 0$  (Agarap, 2018). A ReLU é expressa segundo a Equação 1 e demonstrada na Figura 3.2.

$$f(x) = \max(x, 0) \quad (\text{Equação 1})$$

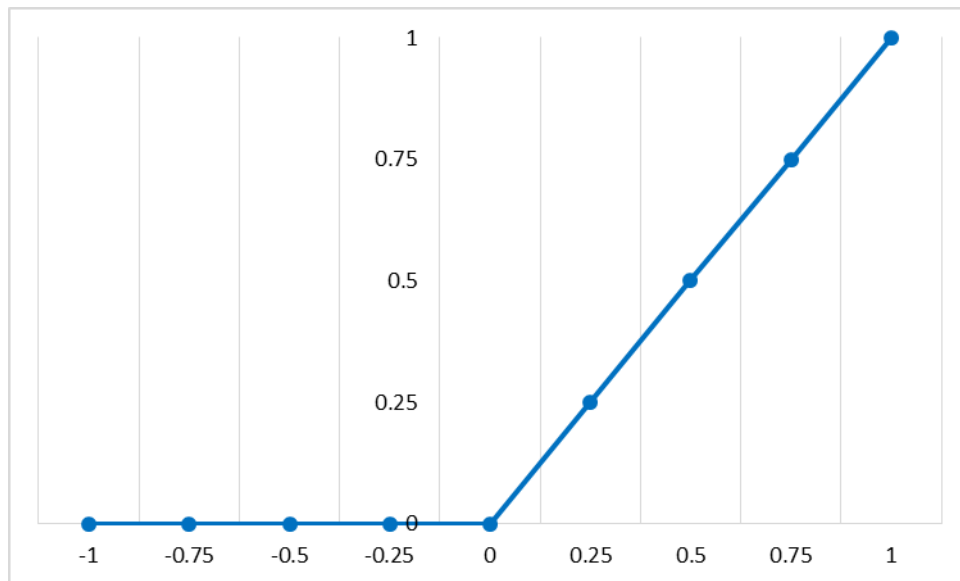


Figura 3.2 Função de ativação ReLU - Unidade Linear Retificada

Muitos pesquisadores têm privilegiado a simplicidade e a fiabilidade do ReLU, porque as melhorias de desempenho das outras funções de ativação tendem a ser inconsistentes entre diferentes modelos e conjuntos de dados (Ramachandran *et al.*, 2018). Outros modelos de funções de ativação utilizados no treinamento de redes neurais são a função de *Sigmoid* e funções tangenciais. Ramachandran *et al.* (2018) afirmam que a ReLU otimiza o processo de treinamento mais facilmente que as demais funções, tornando-se, desse modo, a função de ativação mais comumente aplicada entre a comunidade de estudo da *deep learning*.

No que tange a aplicação das RNA, essas são utilizadas na solução dos mais diversos problemas, entre eles classificatórios, a partir da detecção de padrões entre classes de dados; agrupamento de dados (*clustering*), a partir de similaridades e dissimilaridades entre padrões de dados de entrada com base em suas intercorrelações; otimização, que busca uma solução que maximize ou minimize uma função sujeita a uma série de restrições; além de modelos de projeção (Basheer e Hajmeer, 2000).

Para quaisquer aplicações dadas, é importante que, estatisticamente, possa ser inferida a relação entre as variáveis preditivas com as respectivas variáveis de “resposta”, de modo a compreender como variações em  $X$  podem afetar os valores obtidos em  $Y$ . Tais relações podem ser determinadas a partir de análises de sensibilidade que auxiliam na interpretação dos resultados obtidos por aplicação de *machine learning*.

### 3.1 Aplicações práticas de redes neurais artificiais

A problemática de questões relacionadas ao transporte se torna desafiadora quando o sistema e o comportamento dos usuários são muito complexos para modelar e prever (Abduljabbar *et al.*, 2019). Nesse contexto, a popularidade de redes neurais e demais técnicas de *machine learning* aumenta continuamente na área de transportes, com aplicações que variam desde engenharia de tráfego até projeções de demanda e modelagem de pavimentação (Zhao *et al.*, 2019).

Desde 1990, pesquisas foram conduzidas com a aplicação de redes neurais para o planejamento, *design* e modelagem de rodovias (Abduljabbar *et al.*, 2019). Um exemplo foi o trabalho de Rodrigue (1997), que modelou a relação espacial entre transporte e planejamento de uso do solo a partir de um sistema paralelo de rede neural. Em estudos mais recentes, realizados para Ankara, na Turquia, as RNA são aplicadas para modelagem de plano de gestão de segurança, que utilizou algoritmo genético conjuntamente (Akgungor e Dogan, 2009), e para a previsão de acidentes em vias altamente congestionadas (Dogan e Akgungor, 2013).

De maneira integrada a sistemas ITS (*Intelligent Transport Systems*), Gilmore e Abe (1995) desenvolveram dois sistemas de redes neurais para gerenciar uma rodovia de maneira mais eficiente com base em dados de simulação microscópica. O primeiro sistema era responsável pelo controle do sinal semafórico, enquanto o segundo realizava previsões de congestionamento do tráfego. Em seu trabalho, Nakatsuji e Kaku (1991) já haviam demonstrado a viabilidade de redes neurais no controle de tráfego, ao propor uma rede neural de múltiplas camadas, testada em três interseções distintas.

Em abordagens mais recentes, van Cranenburgh e Alwosheel (2019) utilizaram RNA para analisar a heterogeneidade das regras de decisão entre viajantes. Os resultados obtidos pela rede neural foram comparados com resultados de modelo de escolha discreta de classe única (estimado a partir de subconjuntos de dados) e modelo de escolha discreta de classe latente. Os autores concluem que as RNA podem ser aplicadas à investigação da heterogeneidade das regras de decisão e que corresponde a uma boa ferramenta de análise nesse contexto. De maneira similar, Aschwanden *et al.* (2019) aplicaram RNA na estruturação de um modelo de escolha modal para uma rede integrada de transporte. A rede neural foi treinada a partir de imagens georreferenciadas de viagens realizadas em um dia específico de 2013, que se iniciaram em 5.780 residências aleatoriamente selecionadas no estado de Victoria, na Austrália.

O resultado é um modelo capaz de inferir o compartilhamento do modo de transporte com alta precisão ( $R^2 = 0.58$ ) a partir de imagens de satélites apenas.

Demais aplicações de redes neurais foram feitas na previsão de movimentos de multidões em tempo real (Duives *et al.*, 2019), na projeção de consumo de energia para armazenar informações de banco de dados de um site de *e-commerce* (Ferreira *et al.*, 2019) e na estimação das preferências de curto prazo de usuários da Internet, a fim de propor conteúdo do interesse desses (Chen *et al.*, 2019).

No campo do comércio eletrônico, Li *et al.* (2018) compararam resultados de previsões do volume de vendas *online* de uma empresa chinesa realizadas por meio de redes neurais (não lineares e autorregressivas) e por meio de um modelo autorregressivo integrado de médias móveis (*Autoregressive Integrated Moving Average* - ARIMA). Os autores concluíram que a combinação do ajuste linear do método ARIMA com o mapeamento não linear realizado pela rede neural retornou os melhores resultados de previsão de vendas. Por sua vez, Bandara *et al.* (2019) aplicaram *Long-Short-Term Memory network* (LSTM) para realizar projeções de vendas com base em séries históricas de diversos grupos de produtos do Walmart.com. Segundo os autores, métodos tradicionais de projeção comportam apenas uma única série histórica, enquanto a LSTM permite incorporar em um único modelo dados de vários produtos correlacionáveis.

### 3.2 Análise de sensibilidade em modelos de aprendizagem

Um dos principais fatores que interferem no sucesso de um processo de modelagem é a habilidade de extrair informação a partir da estrutura do modelo e de determinar a relação entre os dados de entrada e os dados de saída obtidos. Para este fim, dispõe-se dos métodos de análise de sensibilidade. Essas não apenas são importantes para a otimização do processamento e validação do modelo (Hashem, 2003), mas também para o entendimento dos códigos científicos e estrutura do modelo (Rabitz, 1989), para a determinação das variáveis mais significantes (Hashem, 2003) e para a interpretação do comportamento e respostas obtidas (Razavi e Gupta, 2015; Borgonovo e Plischke, 2016; Jin *et al.*, 2018).

Asheghi *et al.* (2020) complementam que, em processos de tomada de decisão, a aplicação de análises de sensibilidade auxilia na identificação de variáveis efetivas a modelos de projeção, além de indicar variáveis críticas para as quais é necessário obter mais informação. Ademais,

análises de sensibilidade expõem estimativas inapropriadas e orientam a utilização de variáveis relevantes na tomada de decisão.

Uma aplicação de análise de sensibilidade em rede neural foi realizada por Asheghi *et al.* (2020), que atualizaram a estrutura de um modelo preditivo, baseado em rede neural, utilizando diferentes métodos de análise de sensibilidade: o *equation method* (EM), que analisa a sensibilidade dos pesos ajustados do modelo; o *weight magnitude analysis method* (WMAM); *variable perturbation method* (VPM); *partial derivative algorithm* (PaD); *profile method* (PM); *stepwise method* (SM); e *cosine amplitude method* (CAM). A proposta dos autores foi reduzir a complexidade do modelo original e, a partir da aplicação das ferramentas de sensibilidade, minimizaram o número de neurônios nos *layers* intermediários de 12/11 para 6/8. As ferramentas PaD e CAM são apontadas como as que apresentaram melhor performance nos estudos realizados.

Com o intuito similar de otimizar a estrutura do modelo, Zhang e Wallace (2016), realizaram uma análise de sensibilidade em modelo com base em algoritmo CNN para explorar o efeito dos componentes da arquitetura no desempenho do modelo. O objetivo foi distinguir entre decisões de concepção importantes e comparar seus efeitos para a classificação de sentenças linguísticas.

Aplicação com finalidade de determinar as variáveis mais significativas do modelo foi abordada por Wang *et al.* (2020), que empregaram uma análise de sensibilidade global para identificar fatores que influenciam na segurança de projetos de exploração de óleo e gás. O algoritmo de análise de sensibilidade global é capaz de executar uma análise F-dimensional tanto para tarefas de regressão como de classificação, embora com um elevado custo computacional (Cortez e Embrechts, 2013). Nesse sentido, a ferramenta foi aplicada, simultaneamente, a três modelos que simulam falhas no procedimento padrão.

A fim de auxiliar na interpretação de respostas obtidas em modelos de aprendizagem, métodos de visualização podem ser empregados (Craven e Shavlik, 1992; Cortez e Embrechts, 2013; Zhao e Hastie, 2019). A maioria destes métodos aborda aspectos relacionados com a multidimensionalidade dos dados e a utilização da visualização para modelos de aprendizagem com procedimentos complexos e difíceis de decifrar, também chamados modelos *black-box*, a exemplo das redes neurais (Cortez e Embrechts, 2013). Nesse contexto, alguns métodos

gráficos foram propostos, por exemplo, os diagramas de Hinton e Bond (Craven e Shavlik, 1992), que ajudam na interpretação dos métodos de classificação pela visualização dos pesos das redes neurais. Por sua vez, a abordagem de visualização de modelos de *data mining*, empregada por Cortez e Embrechts (2013), baseou-se em análise de sensibilidade. Estas, embora inicialmente propostas para as RNA, podem ser utilizadas com praticamente qualquer método de aprendizagem supervisionado (Cortez e Embrechts, 2013).

Análise com enfoque na visualização de relações entre variáveis, foi realizada por Zhao e Hastie (2019). Os autores aplicaram *Partial Dependence Plot* (PDP) para estabelecer as relações causais entre as variáveis explicativas e auxiliar na interpretação das respostas do modelo de classificação. Segundo os autores, trata-se da ferramenta mais aplicada à visualização em modelos *black-box*.

## 4 METODOLOGIA

A fim de responder à questão de pesquisa quanto à influência de atributos relacionados à entrega na efetivação da compra realizada pelo consumidor do comércio eletrônico, a metodologia adotada se dividiu segundo as etapas: coleta e pré-processamento dos dados, definição da estrutura de dados de entrada e de saída, estimação dos modelos por RNA (que envolveu treinamento do algoritmo, validação dos modelos e interpretação dos resultados) e, finalmente, estimação dos modelos por regressão logística (RL).

### 4.1 Questionário de pesquisa e coleta de dados

A coleta de dados teve por objetivo reunir as informações necessárias para treinar e simular uma rede neural. Os dados coletados orientaram a análise da influência dos parâmetros de entrega na efetivação da compra *online* e permitiram representar/modelar o comportamento do consumidor do comércio eletrônico.

A coleta foi formulada pela proposição de um questionário que se dividiu em três partes:

- Características do respondente;
- Grau de importância dos parâmetros do comércio eletrônico;
- Valor percebido dos parâmetros do comércio eletrônico pelo respondente.

Na primeira seção do questionário pretendeu-se investigar os hábitos de consumo do consumidor sobre o processo de compra pela Internet e o recebimento do produto. Além disso, foram coletados dados socioeconômicos e demográficos dos respondentes (gênero, idade, faixa de renda, entre outros).

As duas últimas seções do questionário se destinaram à compreensão do grau de importância de fatores relacionados à entrega do produto no ato de efetivação da compra *online*. O argumento central desta investigação é que os consumidores do comércio eletrônico têm maior intenção de oficializar as compras *online* se os recursos necessários e disponíveis se traduzirem em benefícios para si (Yuen *et al.*, 2018). Assim, com base na teoria dos recursos cognitivos, que associa a decisão de realizar uma tarefa aos recursos cognitivos do indivíduo (Pappas *et al.*, 2015), a segunda seção do questionário buscou determinar a relação de importância que é atribuída pelos respondentes a diferentes aspectos que constituem o processo de comprar *online*, como parâmetros do produto (preço), do serviço (comodidade, privacidade, segurança,

oportunidades de oferta e condições diferenciadas de pagamento) e da entrega (prazo, valor do frete, horário de entrega, dentre outros).

Ademais, a fim de determinar como os parâmetros de entrega influenciam na percepção de ganho do consumidor e, conseqüentemente, na intenção de efetivar a compra *online*, aplicou-se a teoria do valor percebido. Essa teoria postula que o valor percebido pelos consumidores deriva da análise dos benefícios e custos da compra de um produto ou serviço (Pappas *et al.*, 2015). O valor percebido refere-se, assim, à utilidade do consumidor. Desse modo, a terceira seção do questionário teve por objetivo obter dados para investigar como diferentes combinações entre parâmetros de entrega influenciam o valor percebido deste serviço pelos consumidores.

A Figura 4.1 apresenta o modelo conceitual deste estudo, que explica como as informações do consumidor, a importância atribuída aos parâmetros do *e-commerce* e o valor percebido se conectam à intenção de efetivar a compra *online*. Por sua vez, a Tabela 4.1 trata da composição e construção do questionário aplicado.

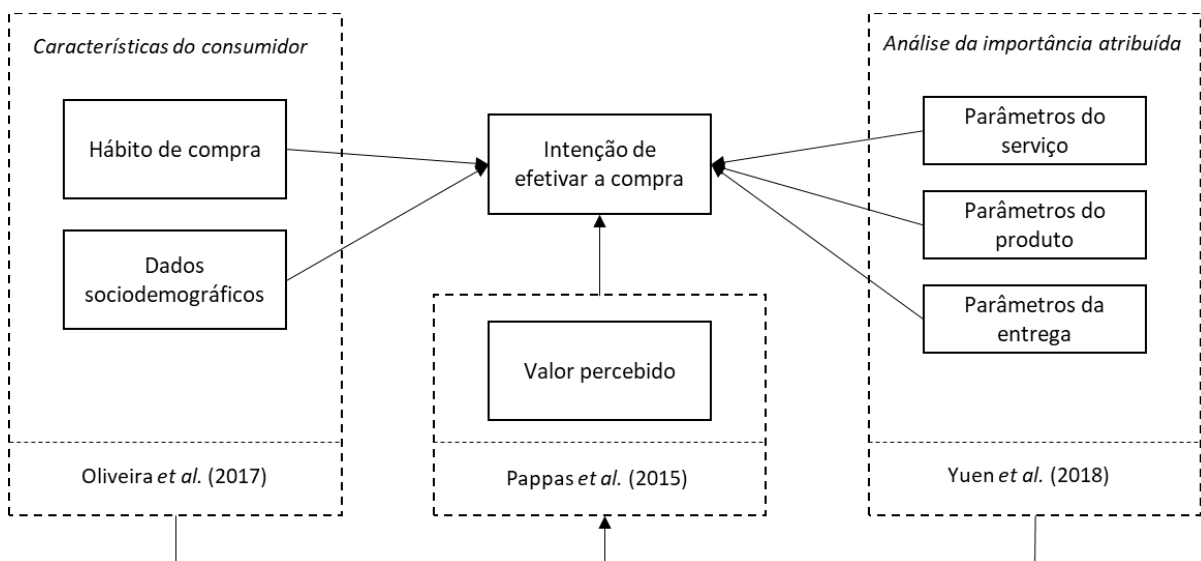


Figura 4.1 Modelo Conceitual

Tabela 4.1 Construção do questionário

| Campo                                | Variáveis de interesse                                                                                                                   | Tipo de resposta                       |
|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|
| 1ª seção:                            | Em relação aos produtos abaixo, quais você costuma comprar pela Internet?                                                                | Categórico (10 classes)                |
| Hábitos de compra                    | Qual o número de entregas que você recebe em casa de produtos adquiridos pela Internet?                                                  | Categórico (5 classes)                 |
|                                      | Para uma compra acima de R\$300, qual valor você aceita pagar pelo frete?                                                                | Contínuo                               |
|                                      | Qual é o período preferencial para recebimento das mercadorias compradas pela Internet?                                                  | Categórico (6 classes)                 |
| 1ª seção:                            | Faixa etária                                                                                                                             | Categórico (5 classes)                 |
| Dados socioeconômicos e demográficos | Renda familiar                                                                                                                           | Categórico (5 classes)                 |
|                                      | Gênero                                                                                                                                   | Binário                                |
|                                      | Em qual Estado você mora?                                                                                                                | Aberto                                 |
|                                      | Em qual cidade você mora?                                                                                                                | Aberto                                 |
|                                      | Qual o CEP da residência?                                                                                                                | Aberto                                 |
| 2ª seção:                            | Qual é o grau de importância da comodidade de comprar sem sair de casa na hora de comprar pela Internet?                                 | Categórico (5 classes – Escala Likert) |
| Parâmetros do serviço                | Qual é o grau de importância da privacidade na hora de comprar pela Internet?                                                            | Categórico (5 classes – Escala Likert) |
|                                      | Qual é o grau de importância das oportunidades de compra (aproveitar ofertas relâmpagos ou benefícios) na hora de comprar pela Internet? | Categórico (5 classes – Escala Likert) |
| 2ª seção:                            | Qual é o grau de importância preço do produto na hora de efetivar a compra pela Internet?                                                | Categórico (5 classes – Escala Likert) |
| Parâmetros do produto                | Qual é o grau de importância do prazo de entrega na hora de comprar pela Internet?                                                       | Categórico (5 classes – Escala Likert) |
| 2ª seção:                            | Qual é o grau de importância do valor do frete na hora de comprar pela Internet?                                                         | Categórico (5 classes – Escala Likert) |
| Parâmetros da entrega                | Qual é o grau de importância do frete grátis na hora de comprar pela Internet?                                                           | Categórico (5 classes – Escala Likert) |
|                                      | Qual é o grau de importância do horário da entrega na hora de comprar pela Internet?                                                     | Categórico (5 classes – Escala Likert) |
|                                      | Qual é o grau de importância do ponto de retirada na hora de comprar pela Internet?                                                      | Categórico (5 classes – Escala Likert) |
| 3ª seção:                            | Quais fatores influenciam o valor que está disposto a pagar pelo frete?                                                                  | Categórico (5 classes)                 |
| Valor percebido                      | Você estaria disposto a pagar mais pelo frete para escolher dia e horário da entrega.                                                    | Categórico (5 classes – Escala Likert) |
|                                      | Você estaria disposto a pagar mais pelo frete para receber um produto em um prazo menor.                                                 | Categórico (5 classes – Escala Likert) |
| 3ª seção:                            | O valor do frete é um fator determinante para você comprar pela Internet.                                                                | Categórico (5 classes – Escala Likert) |
| Intenção                             | A possibilidade de escolher o dia e o horário de entrega influência na decisão de realizar uma compra pela Internet.                     | Categórico (5 classes – Escala Likert) |

## 4.2 Pré-processamento de dados

Uma vez coletados, os dados foram pré-processados. Esta etapa consistiu, primeiramente, na análise de confiabilidade das respostas obtidas do questionário através da aplicação do coeficiente alfa de Cronbach. Num segundo momento, buscou-se preparar a estrutura de dados do questionário para alimentar a rede neural. Esse processo inclui uma etapa de codificação e normalização dos dados, além do tratamento de multicolinearidade entre eles.

### 4.2.1 Aplicação do coeficiente alfa de Cronbach

Dada a aplicação do questionário, as análises que se seguiram foram pautadas na validade e na confiabilidade dos resultados obtidos, a fim de evitar dúvidas e inconsistências acerca das conclusões extraídas. Para atestar a confiabilidade dos resultados foi utilizado o coeficiente alfa de Cronbach.

Segundo Almeida *et al.* (2010), um instrumento é válido quando mede o que se deseja e, para tanto, este deve ser confiável. Por sua vez, a confiabilidade pode ser definida como o grau em que o resultado medido reflete o resultado verdadeiro, em outras palavras, o quanto uma medida está livre da variância dos erros aleatórios (Hayes, 1998).

O estimador de confiabilidade a ser utilizado depende dos fatores particulares geradores de erros que o pesquisador busca identificar (Cronbach *et al.*, 1972). Assim, se fatores de erro associados ao uso de diferentes itens são de interesse (Almeida *et al.*, 2010), então o estimador de consistência, como o alfa de Cronbach, pode ser utilizado.

O coeficiente alfa de Cronbach, uma das ferramentas estatísticas mais importantes e difundidas em pesquisas envolvendo a construção de testes e sua aplicação (Cortina, 1993), leva em consideração a variância atribuída aos sujeitos e a variância atribuída à interação entre sujeitos e itens da pesquisa. Segundo Streiner (2003), o alfa de Cronbach é a média das correlações entre os itens que fazem parte de um instrumento e pode ser conceituado como a medida pela qual algum conceito ou fator medido está presente em cada item.

Assim, para se estimar o alfa de Cronbach, considerou-se a matriz  $X_{ij}$ , em que  $i$  correspondeu ao total de respostas quantificadas (595 linhas) e  $j$  corresponde à quantidade de itens categóricos em escala Likert (16 colunas) que compõem o questionário. O coeficiente foi calculado de acordo com a Equação 2 (Leontisis e Pagge, 2007):

$$\alpha = \frac{j}{j-1} \left[ \frac{\sigma_t^2 - \sum_{n=1}^j \sigma_n^2}{\sigma_t^2} \right] \quad (\text{Equação 2})$$

em que:

- $\sigma_n^2$  é a variância de cada coluna de X, ou seja, a variância relacionada à questão de X;
- $\sigma_t^2$  é a variância da soma de cada linha de X, ou seja, a variância da soma das respostas de cada sujeito.

Na Equação 1, a fração  $\frac{j}{j-1}$  é um fator de correção. Se há consistência nas respostas quantificadas, então o  $\alpha$  tenderá a 1 (Almeida *et al.*, 2010). Caso contrário, se as respostas forem randômicas, então  $\alpha$  tenderá a 0. O valor mínimo aceitável para o alfa é 0,70. Abaixo desse valor a consistência interna da escala utilizada é considerada baixa (Streiner, 2003). Em contrapartida, o valor máximo esperado é 0,90, uma vez que resultados acima desse valor podem apontar para a ocorrência de redundâncias ou duplicações. Portanto, resultados entre 0,80 e 0,90 são considerados satisfatórios (Streiner, 2003).

#### 4.2.2 Codificação e normalização dos dados

Uma vez testada a confiabilidade das respostas ao questionário, o passo seguinte foi normalizar os dados de entrada para que o algoritmo do RNA possa facilmente interpretá-los. Esta etapa pode melhorar o desempenho do modelo e fazê-lo convergir mais rapidamente (Mollalo *et al.*, 2019). Em geral, os algoritmos de aprendizagem se beneficiam da categorização do conjunto de dados. Aschwanden *et al.* (2019) relatam a experiência de preparação de seus dados de entrada, de modo a individualmente rotulá-los e classificá-los em grupos discretos, a fim de serem introduzidos na estrutura da RNA.

A primeira etapa desse processo consistiu na busca por lacunas nas respostas do questionário, que, quando detectadas, ocasionaram em sua eliminação. A codificação dos dados, que ocorreu a seguir, tratou de transformar as variáveis do modelo em contrapartes discretas com as quais o algoritmo da RNA saiba trabalhar. Entre suas limitações, as redes neurais têm dificuldades no processamento de valores categóricos. Portanto, diferentes técnicas foram empregadas para transformar variáveis categóricas em vetores numéricos.

Conceitualmente discretas e não contínuas, as variáveis categóricas podem ser classificadas em dois diferentes grupos: ordinais, aquelas que podem ser ordenadas (por exemplo, numeração de

roupa, como P, M e G), e nominais, aquelas para as quais não se pode definir relação de ordem (por exemplo, atribuição de cores, como verde, azul, etc). Assim, às variáveis ordinais foram atribuídos valores numéricos indicativos da ordem entre as opções de resposta. Esse processo de codificação foi realizado por emprego da ferramenta “*LabelEncoder*”, pertencente à biblioteca “*scikit learn*” (Pedregosa *et al.*, 2011) desenvolvida para a linguagem *Python*. Essa ferramenta é utilizada para codificar rótulos em valores entre 0 e  $n\_classes-1$ . Para tanto, é necessário definir a ordem (crescente ou decrescente) dos valores ordinais relacionados às classes (opções de respostas) da variável tratada.

Por sua vez, para variáveis nominais foi utilizada a técnica denominada *One-Hot Encoding*. De simples aplicação, essa técnica consiste em transformar as opções de resposta em variáveis Dummy, ou seja, variáveis binárias para as quais é atribuído o valor 1, em caso afirmativo, e 0, em caso negativo. Um exemplo da aplicação do *One-Hot Encoding* é demonstrado para a variável “Faixa Etária” na tabela a seguir.

Tabela 4.2 Demonstração da técnica *One-Hot Encoding* para a variável “Faixa Etária”

| Faixa Etária<br>Observações | 15 a 24 anos | 25 a 34 anos | 35 a 49 anos | 50 a 69 anos | Mais de 70 anos |
|-----------------------------|--------------|--------------|--------------|--------------|-----------------|
| 50 a 69 anos                | 0            | 0            | 0            | 1            | 0               |
| Mais de 70 anos             | 0            | 0            | 0            | 0            | 1               |
| 35 a 49 anos                | 0            | 0            | 1            | 0            | 0               |
| 15 a 24 anos                | 1            | 0            | 0            | 0            | 0               |

Uma vez finalizado o processo de codificação dos valores das variáveis do modelo, partiu-se para o processo de normalização dos dados. Esse processo consiste em escalar todos os dados numéricos para terem magnitudes mais ou menos semelhantes. Sem a etapa de normalização, o processo de aprendizado da rede neural pode ser dificultado, pois essa pode entender que variáveis de magnitude maiores têm mais peso que outras de menor ordem de grandeza.

Para prosseguir com a normalização dos dados do modelo, utilizou-se a ferramenta “*StandardScaler*” da biblioteca “*scikit learn*” (linguagem *Python*). Essa ferramenta padroniza os dados numéricos, calculando a média e escalonando pela variância da unidade. O resultado escalonado de uma amostra é calculado pela Equação 3.

$$z_i = \frac{(x_i - \bar{x})}{s} \quad (\text{Equação 3})$$

em que  $\bar{x}$  é a média dos elementos  $x_n$ , ou zero se não houver média, e  $s$  é o desvio-padrão da amostra ou 1 se não houver desvio padrão.

#### 4.2.3 Tratamento da multicolinearidade

Uma das consequências possíveis da aplicação da técnica de codificação *One-Hot Encoding*, é o surgimento da multicolinearidade entre o conjunto de dados, ou seja, forte relação de dependência entre variáveis. A correlação entre as variáveis independentes do modelo pode tornar tendenciosos os resultados, além de dificultar a análise da importância dessas através do vetor de pesos atribuídos pelo algoritmo da rede neural. Em outras palavras, a multicolinearidade impacta na interpretabilidade do modelo.

Alguns métodos gerais de eliminação da multicolinearidade incluem a exclusão de variáveis redundantes. Esse processo se dá a partir da observação da relação entre variáveis utilizando gráficos de dispersão, matriz de intercorrelação, além de técnicas diversas de pré-processamento de dados estatísticos (Garg e Tai, 2013).

#### 4.3 Estruturação das variáveis de entrada e de saída

As respostas ao questionário foram consolidadas e sistematizadas, segundo procedimento anterior. Num primeiro momento, para as variáveis categóricas com classes de respostas em escala Likert, considerou-se as cinco categorias: (1) ‘Discordo Totalmente’, (2) ‘Discordo’, (3) ‘Neutro’, (4) ‘Concordo’ e (5) ‘Concordo Totalmente’. Contudo, os primeiros testes com RNA apresentaram baixa performance de classificação para as respostas negativas (1 e 2) e repostas positivas (4 e 5). Para contornar este problema, essas foram agregadas, eliminando a intensidade da discordância/concordância e adotando a seguinte escala: (1) ‘Discordo’, (2) ‘Neutro’ e (3) ‘Concordo’.

Conforme a pertinência para análise da influência dos atributos de entrega na compra pelo comércio eletrônico, estabeleceram-se 13 variáveis explicativas ( $X_i$ ) e 5 variáveis dependentes ( $Y_j$ ), a saber: a importância dos parâmetros prazo de entrega, valor do frete e horário da entrega, além das variáveis categóricas ‘valor do frete é determinante para comprar pela Internet’ e ‘influência da possibilidade de escolher o dia e o horário de entrega na decisão de realizar uma

compra pela Internet’. Assim, os parâmetros de ‘valor do frete’ e ‘horário da entrega’ são investigados em duas abordagens distintas: pela análise do seu grau de importância na hora de comprar pela Internet e por sua influência direta na efetivação da compra.

A influência de cada parâmetro de entrega, em suas diferentes abordagens, foi modelada individualmente. Assim, 5 modelos foram propostos. Estes foram estruturados para permitir a compreensão de como o consumidor do comércio eletrônico se comporta e para determinar sua relação com características do consumidor, como seu perfil socioeconômico e seus hábitos de compra, além do grau de importância das características do comércio eletrônico.

Às variáveis trabalhadas foram atribuídos códigos de identificação, apresentados na Tabela 4.3, juntamente com descritivo e respostas codificadas. Ademais, no apêndice desta dissertação (item 8.3) apresenta-se a distribuição da frequência de respostas às variáveis modeladas.

Tabela 4.3 Estruturação dos dados de entrada dos modelos

| Código | Descrição                                                            | Respostas codificadas                                                           | Categoria das variáveis                         |
|--------|----------------------------------------------------------------------|---------------------------------------------------------------------------------|-------------------------------------------------|
| V01    | Compra produtos de beleza, vestuário ou afins                        | (1) Sim<br>(0) Não                                                              | Variáveis independentes (Hábitos de compra)     |
| V02    | Compra produtos de eletrônicos                                       | (1) Sim<br>(0) Não                                                              |                                                 |
| V03    | Compra produtos de cultura e lazer                                   | (1) Sim<br>(0) Não                                                              |                                                 |
| V04    | Compra móveis e utensílios domésticos                                | (1) Sim<br>(0) Não                                                              |                                                 |
| V05    | Percentual do preço do produto que está disposto a pagar pelo frete  | (1) Até 3%<br>(2) Até 7%<br>(2) Até 10%<br>(4) Acima de 10%                     |                                                 |
| V06    | Frequência de consumo                                                | (1) Uma entrega ou menos por mês<br>(2) Mais de uma entregas por mês            | Variáveis independentes (Dados socioeconômicos) |
| V07    | Faixa etária                                                         | (1) 15 a 24 anos<br>(2) 25 a 34 anos<br>(3) 35 a 49 anos<br>(4) Mais de 50 anos |                                                 |
| V08    | Renda familiar                                                       | (1) Até 1 SM<br>(2) 2 A 4 SM<br>(3) 4 a 10 SM<br>(4) Mais de 10 SM              |                                                 |
| V09    | Gênero                                                               | (2) Feminino<br>(1) Masculino                                                   |                                                 |
| V10    | Comodidade de comprar sem sair de casa                               | (1) Irrelevante<br>(2) Neutro<br>(3) Importante                                 |                                                 |
| V11    | Privacidade                                                          | (1) Irrelevante<br>(2) Neutro<br>(3) Importante                                 | Variáveis independentes (Parâmetros de compra)  |
| V12    | Oportunidade de compra (aproveitar ofertas relâmpagos ou benefícios) | (1) Irrelevante<br>(2) Neutro<br>(3) Importante                                 |                                                 |

| Código | Descrição                                                                                                            | Respostas codificadas | Categoria das variáveis                         |
|--------|----------------------------------------------------------------------------------------------------------------------|-----------------------|-------------------------------------------------|
| V13    | Preço do produto                                                                                                     | (1) Irrelevante       | Variáveis independentes (parâmetros do produto) |
|        |                                                                                                                      | (2) Neutro            |                                                 |
|        |                                                                                                                      | (3) Importante        |                                                 |
| V14    | Prazo da entrega                                                                                                     | (1) Irrelevante       |                                                 |
|        |                                                                                                                      | (2) Neutro            |                                                 |
|        |                                                                                                                      | (3) Importante        |                                                 |
| V15    | Valor do frete                                                                                                       | (1) Irrelevante       |                                                 |
|        |                                                                                                                      | (2) Neutro            |                                                 |
|        |                                                                                                                      | (3) Importante        |                                                 |
| V16    | Horário da entrega                                                                                                   | (1) Irrelevante       | Variáveis dependentes (parâmetros da entrega)   |
|        |                                                                                                                      | (2) Neutro            |                                                 |
|        |                                                                                                                      | (3) Importante        |                                                 |
| V17    | O valor do frete é um fator determinante para comprar pela Internet.                                                 | (1) Irrelevante       |                                                 |
|        |                                                                                                                      | (2) Neutro            |                                                 |
|        |                                                                                                                      | (3) Importante        |                                                 |
| V18    | A possibilidade de escolher o dia e o horário de entrega influencia na decisão de realizar uma compra pela Internet. | (1) Irrelevante       |                                                 |
|        |                                                                                                                      | (2) Neutro            |                                                 |
|        |                                                                                                                      | (3) Importante        |                                                 |

#### 4.4 Rede Neural: estrutura e treinamento

A capacidade de aprender é uma característica peculiar pertencente a sistemas inteligentes, biológicos ou não (Basheer e Hajmeer, 2000). Segundo Schalkoff (1997), em sistemas artificiais, a aprendizagem é vista como o processo de atualização da representação interna do sistema, em resposta a estímulos externos, para que ele possa realizar uma tarefa específica. Isso inclui modificar a arquitetura de rede, o que envolve ajustar os pesos dos links, podar ou criar alguns links de conexão e/ou alterar as regras de ativação dos neurônios individuais (Schalkoff, 1997).

À parte do pré-processamento dos dados de entrada, a definição da arquitetura da rede neural, ou seja, determinação da quantidade de nós e de *layers* intermediários, depende da robustez da amostra introduzida no modelo, além de outras ponderações. Quanto maior a quantidade de *layers*, maior a precisão dos resultados obtidos. Em contrapartida, uma grande quantidade desses, além de elevar o tempo de processamento, torna mais complexa classificação dos dados de entrada e a interpretação dos resultados obtidos. Desse modo, a estrutura da rede neural, além de se adequar ao tamanho da amostra de dados coletada, buscou alcançar o equilíbrio entre esses fatores mencionados.

O modelo de rede neural implementado foi o *Multi-layer Perceptron* (MLP), que apresenta mais de uma camada intermediária de neurônios. O MLP é um algoritmo de aprendizagem supervisionada, que aprende segundo a função  $f(X): R^n \rightarrow R^o$ , onde " $n$ " é o número de dimensões de entrada (*input*) e " $o$ " é o número de dimensões da saída (*output*). Tendo em conta

um conjunto de registros  $X = x_1, x_2, \dots, x_n$  e uma variável dependente  $y$ , o MLP pode aprender uma aproximação de função não-linear para realizar classificações ou regressões. A Figura 4.2 ilustra a estrutura de MLP utilizada no estudo conduzido.

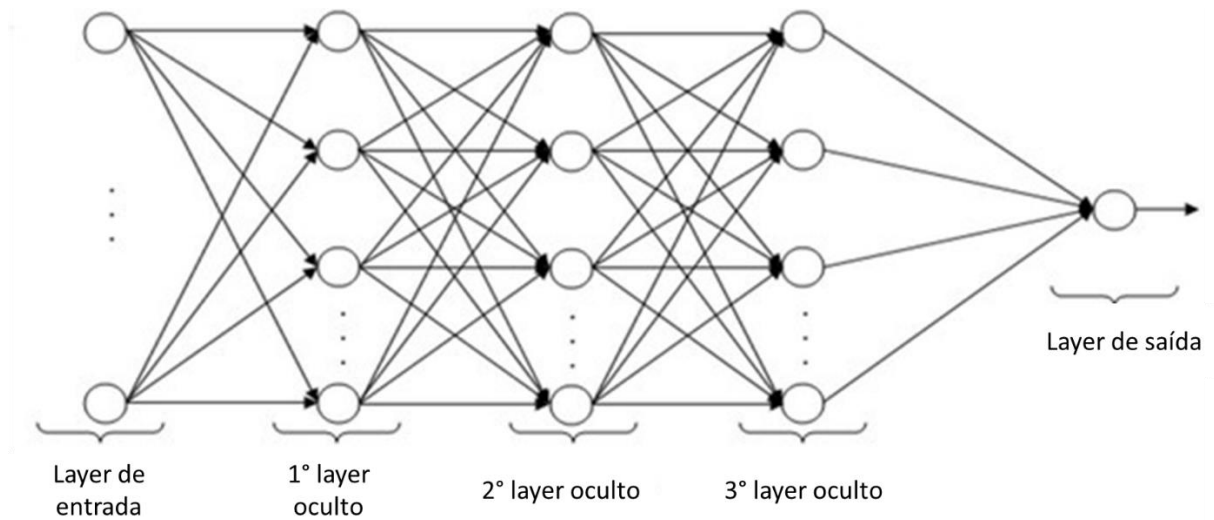


Figura 4.2 Estrutura de Multi-layer Perceptron (MLP).  
(Adaptado de Google Images)

O *layer*, ou camada, de entrada é constituído por um conjunto de neurônios representando os dados de entrada. Cada neurônio das camadas ocultas transforma os valores da camada anterior em uma soma linear ponderada  $w_1x_1 + w_2x_2 + \dots + w_nx_n$ , onde  $w$  corresponde aos pesos definidos para cada atributo das camadas, durante o processo de treinamento da rede neural. A sequência de neurônios nos *layers* é ativada segundo uma função pré-determinada de ativação não linear. A camada de saída recebe os valores do último *layer* oculto e transforma-os em valores de saída.

A rede neural foi implementada em linguagem *Python 3.7* em ambiente de desenvolvimento denominado *Spyder*. Para tanto, fez-se uso da biblioteca “*scikit learn*”, a qual dispõe de funções e métodos de pré-processamento e validação, além de ferramentas de classificação como a rede neural em si. Assim, o MLP foi aplicado a partir da ferramenta “*MLPClassifier*”, utilizada, majoritariamente, em sua configuração *default* por ser recomendada para o tamanho da amostra de dados.

Os parâmetros aplicados para o MLP foram:

- Quantidade de layers intermediários e quantidade de neurônios por layer: 3 e 11 (respectivamente);
- Função de otimização dos pesos: 'lbfgs', que é uma função da família dos métodos quasi-Newton. Essa função é indicada para pequenos conjuntos de dados, por apresentar melhor performance e convergência mais rápida.
- Máximo de iterações: 500;
- Função de ativação para as camadas ocultas: 'ReLU', função da unidade linear retificada (ver Equação 1).

Uma vez definida a arquitetura da rede neural, partiu-se para o processo de treinamento do modelo. Tipicamente, o conjunto de dados selecionados para o treinamento constitui a maior parcela da amostra total de dados coletados (Jeng e Fesenmaier, 2010). Desse modo, o treinamento da rede neural se deu com 85% dos dados coletados, sendo os demais 15% reservados para a validação do modelo. A seleção da amostra utilizada para treinamento é feita de modo aleatório a partir de ferramenta denominada “*train\_test\_split*” da biblioteca “*scikit learn*” do *Python* 3.7.

Ao processo de treinamento coube minimizar os erros intrínsecos aos dados de saída. Um sistema baseado em RNA pode ser considerado treinado (1) quando puder lidar com informações imprecisas, difusas, ruidosas e probabilísticas sem efeitos adversos perceptíveis na qualidade da resposta e (2) quando puder generalizar as tarefas que aprendeu às entradas de dados desconhecidos (Basheer e Hajmeer, 2000).

Aplicou-se, assim, o processo de aprendizagem/treinamento denominado *backpropagation*. Tal processo é definido por Hanson (1995) como uma aproximação estocástica multi-variável, não-linear e não-paramétrica, com extração/seleção de recursos dinâmicos, o que os torna capazes de aprender por mapeamento arbitrário. O treinamento por *backpropagation* é classificado como um processo de aprendizagem supervisionado, pois utiliza as variáveis de entrada e compara a saída modelada com rótulos/classificações pré-definidas para ajustar os pesos internos (Aschwanden *et al.*, 2019).

O treinamento ocorreu, assim, de modo iterativo. A cada etapa, de trás para frente na sequência da rede neural (*backpropagation*), os pesos da RNA eram ajustados a partir da minimização de um gradiente calculado pela função de perda de Entropia Cruzada (*Cross-Entropy loss*

*function*). O processo de treinamento minimiza o gradiente a partir de uma técnica denominada Gradiente Descendente Estocástico (*Stochastic Gradient Descent – SGD*). Essa é descrita pela Equação 4:

$$w \leftarrow w - \eta \left[ \alpha \frac{\partial R(w)}{\partial w} + \frac{\partial Loss}{\partial w} \right] \quad (\text{Equação 4})$$

onde  $\eta$  é a taxa de aprendizagem que controla o avanço na pesquisa pelos pesos  $w$ . Por sua vez, *Loss* representa a função de perda aplicada.

#### 4.4.1 Validação dos modelos

A essa etapa coube a determinação da confiabilidade dos resultados dos modelos. Para tanto, dispôs-se de 15% da amostra de dados coletados, tal como descrito no item anterior. Esta foi constatada a partir do cálculo de métricas comumente empregadas para aferir a qualidade de modelos de classificação: precisão (*precision*), revocação (*recall*), F1-score e acurácia.

Para explicar os conceitos citados acima, primeiramente, é importante apresentar a ferramenta denominada matriz de confusão. Essa oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas *versus* as classificações preditas para cada classe (Visa *et al.*, 2011). Assim, os resultados previstos para um exemplo de modelo de classificação binária (0 e 1) seriam expressos em uma matriz 2x2, tal como representado na tabela a seguir.

Tabela 4.4 Exemplo de matriz de confusão 2x2

|              | Positivo                   | Negativo                   |
|--------------|----------------------------|----------------------------|
| 1 (positivo) | Positivos verdadeiros (PV) | Positivos falsos (PF)      |
| 0 (negativo) | Negativos falsos (NF)      | Negativos verdadeiros (NV) |

Como observado, o número de acertos, para cada classe, se localiza na diagonal principal da matriz, enquanto os demais elementos representam erros na classificação.

A definição de precisão pode ser explicada segundo a Equação 5.

$$Precisão = \frac{PV}{PV + PF} \quad (\text{Equação 5})$$

Essa métrica expressa a frequência com a qual o modelo acerta suas classificações dentro do universo de respostas positivas. A precisão é, intuitivamente, a capacidade do classificador de não rotular como positivo uma amostra que seja negativa.

Por sua vez, o conceito de revocação corresponde à frequência com a qual o modelo encontra corretamente os exemplos de uma classe. A revocação é, intuitivamente, a capacidade do classificador de encontrar todas as amostras positivas e é obtido pela Equação 6.

$$Revocação = \frac{PV}{PV + NF} \quad (\text{Equação 6})$$

Indiretamente, a precisão pode ser entendida como o quanto o fator que se quer investigar através do modelo pode ser relacionado aos dados de entrada (resultados da pesquisa conduzida), enquanto a revocação representa o quão completos os dados de entrada estão e o respectivo impacto na capacidade de classificação do modelo. Numa circunstância hipotética, em que o modelo classificador não comete nenhum erro, então os resultados de precisão e *recall* seriam iguais a 1.

Geralmente, a precisão e a *recall* não são discutidas isoladamente. Em vez disso, ambos os parâmetros podem ser combinados em uma única medida, denominada F1-score. Essa métrica combina precisão e *recall* de modo a trazer um número único que indique a qualidade geral do seu modelo. Pode ser interpretada como uma média harmônica ponderada da precisão e da revocação, em que uma pontuação F1-score atinge o seu melhor valor em 1 e a pior em 0. Essa é obtida pela Equação 7.

$$F1score = 2 * \frac{precisão * revocação}{precisão + revocação} \quad (\text{Equação 7})$$

Por fim, a acurácia indica uma performance geral do modelo, com base no total de classificações que esse realiza corretamente. Sua obtenção se dá através da aplicação da Equação 8, a seguir.

$$Acurácia = \frac{PV + NV}{PV + PF + NV + NF} \quad (\text{Equação 8})$$

A função “*classification\_report*” da biblioteca “*scikit learn*” foi utilizada para obter os indicadores de validação do modelo.

#### 4.4.2 Interpretação dos resultados

Uma vez obtidos modelos comportamentais validados, capazes de determinar, com alta precisão, se os atributos de entrega influenciam no comportamento de compra do consumidor do comércio eletrônico, passou-se para a etapa de interpretação dos resultados, a fim de entender como a intenção de efetivar a compra pode estar relacionada aos atributos de entrega. Essa etapa referiu-se, portanto, à inferência das relações entre variáveis preditivas, ou explicativas, e as variáveis de “resposta” dos modelos.

Zhao e Hastie (2019) pontuam que algoritmos de *machine learning* são difíceis de interpretar por geralmente gerarem uma função não linear de alta dimensão. Contudo, os autores afirmam que esse é apenas um problema técnico e que o verdadeiro desafio está em determinar a causalidade no modelo: se alterarmos o valor de uma variável X, com as outras variáveis fixas, quanto o valor de Y mudará?

Uma das ferramentas mais utilizadas na visualização de modelos de aprendizagem (Zhao e Hastie, 2019) são os *Partial Dependence Plots* (PDP), proposta por Friedman (2001). Os PDP mostram o efeito marginal que uma ou duas variáveis têm sobre o resultado previsto de um modelo de aprendizagem (Friedman, 2001) e podem expressar se tal relação é linear, monótona ou mais complexa.

Os *Partial Dependence Plots* (PDP) podem ser definidas pela Equação 9 (Zhao e Hastie, 2019), em que o  $x_S$  são as variáveis para as quais a função PDP deve ser plotada e o  $x_C$  são as outras variáveis utilizadas no modelo de aprendizagem. As variáveis em S são aquelas para as quais se quer compreender a relação com os resultados do modelo. Os vetores  $x_S$  e  $x_C$  combinados correspondem a todo o conjunto de variáveis inseridos no modelo. Os PDP funcionam marginalizando a saída do modelo sobre a distribuição das características do conjunto C, de modo que a função mostra a relação entre as características do conjunto S (Zhao e Hastie, 2019).

Desse modo, obtemos uma função que depende apenas das características no conjunto  $S$ , incluindo as interações com outras características.

$$\hat{f}_{xs} = E_{x_c} [\hat{f}(x_s, x_c)] = \int \hat{f}(x_s, x_c) d\mathbb{P}(x_c) \quad (\text{Equação 9})$$

Zhao e Hastie (2019) afirmam que, na prática, os PDP são simplesmente estimados calculando a média da amostra de dados de treinamento do modelo  $\{X_i, i = 1, \dots, n\}$  com  $x_s$  fixado. A função  $\hat{f}_{xs}$  é apresentada na Equação 104:

$$\hat{f}_{xs}(x) = \frac{1}{n} \sum_{i=1}^n \hat{f}(x_s, x_c^{(i)}) \quad (\text{Equação 10})$$

Os PDP foram aplicados por meio de linguagem *Python*, utilizando a ferramenta “*plot\_partial\_dependence*” da biblioteca “*scikit learn*”. Uma vez que o conjunto de  $x_s$  pode contar apenas com duas variáveis, os *plots* 2D (gráficos) foram gerados aos pares de variáveis por vez. Esses demonstraram a relação entre as variáveis explicativas com cada uma das variáveis de ‘resposta’. A discussão proposta nesta dissertação, a respeito de como os atributos de entrega influenciam no comportamento de compra do consumidor do comércio eletrônico, foi baseada, portanto, na análise dos *plots* para cada variável explicativa do modelo.

#### 4.5 Modelo de Regressão Logística

Com a finalidade de complementar e conferir maior robustez às análises de interpretação do comportamento do consumidor do comércio eletrônico, os modelos propostos no item 4.3 foram estimados pela técnica estatística de Regressão Logística. Esta etapa agregou, assim, a análise de propriedades estatísticas dos resultados.

A regressão logística modela as probabilidades de predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias (Miller *et al.*, 1991). Num modelo de regressão logística binária, a variável dependente tem dois níveis categóricos. As saídas com mais de dois valores são modeladas por regressão logística multinomial e, se as múltiplas categorias forem ordenadas, por regressão logística ordenada.

O modelo de regressão logística se utiliza da função logística para obter a saída de uma equação linear entre 0 e 1. Trata-se de uma função sigmoide, que toma qualquer entrada real e produz um valor entre o intervalo  $[0,1]$  (Hosmer e Lemeshow, 2000). A unidade de medida da escala logarítmica das possibilidades (*log-odds*) é chamada *logit* (oriunda do termo **logistic unit**). Desse modo, a unidade *logit* corresponde ao logaritmo da razão das possibilidades  $\frac{p}{1-p}$ , onde  $p$  é a probabilidade (Cramer, 2003).

A função logística padrão é definida como:

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}} \quad (\text{Equação 11})$$

A representação gráfica da função logística é apresentada na figura x.

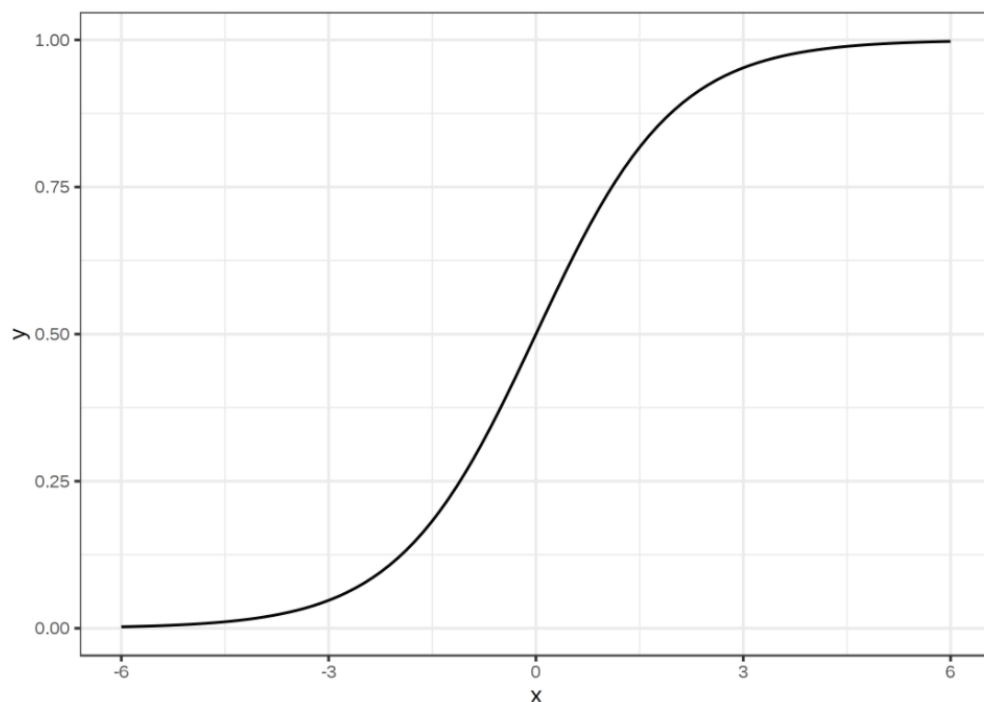


Figura 4.3 Função Logística padrão  
(Adaptado de Google Images)

A interpretação dos coeficientes da regressão logística é simplificada pela reformulação da equação, de modo a representar apenas termos lineares ao lado direito da fórmula. Para tanto, assume-se uma relação linear entre as variáveis preditoras e os *log-odds* (ou *logit*) do evento em que  $y = 1$ . Esta relação linear pode ser escrita na seguinte forma matemática (Hosmer e Lemeshow, 2000):

$$\log\left(\frac{P(y = 1)}{1 - P(y = 1)}\right) = \log\left(\frac{P(y = 1)}{P(y = 0)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (\text{Equação 12})$$

Para facilitar a interpretação do modelo de regressão logística como um modelo linear do logaritmo das possibilidades (*log-odds*), a função  $\exp()$  pode ser aplicada aos dois lados da equação acima:

$$\frac{P(y = 1)}{1 - P(y = 1)} = odds = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n) \quad (\text{Equação 13})$$

Entende-se, portanto, que a alteração de uma característica em uma unidade varia a razão de possibilidades (*odds ratio*) por um fator de  $\exp(\beta_n)$ . Em outras palavras, uma mudança de uma unidade em  $x_n$ , aumenta a razão da probabilidade logarítmica pelo valor do coeficiente correspondente ( $\beta_n$ ). Assim, para diferentes tipos de dados, as interpretações para o modelo de regressão logística são (Miller *et al.*, 1991; Cramer, 2003; Tolles e Meurer, 2016):

- Dado numérico: a alteração de uma unidade na variável  $x_n$ , varia a probabilidade estimada em  $\exp(\beta_n)$ ;
- Dado categórico binário: uma das categorias é referência (geralmente a categoria 0). Assim, a alteração da categoria de referência para outra categoria na variável  $x_n$ , varia a probabilidade estimada em  $\exp(\beta_n)$ ;
- Dado categórico com mais de duas categorias: uma solução para lidar com múltiplas categorias é a codificação a partir da técnica one-hot-encoding, o que significa que cada categoria tem a sua própria variável binária. Só são necessárias colunas L-1 para uma característica categórica com categorias L, caso contrário, esta é sobre-parametrizada. A categoria L-ésima é então a categoria de referência. Pode utilizar qualquer outra codificação que possa ser utilizada em regressão linear. A interpretação para cada categoria é então equivalente à interpretação das características binárias;
- Intercepto ( $\beta_0$ ): quando todas as variáveis numéricas são zero e as variáveis categóricas correspondem à categoria de referência, as probabilidades são estimadas por  $\exp(\beta_0)$ . A interpretação do peso da intercepção não é normalmente relevante.

Após obtenção do modelo, a primeira avaliação feita foi a respeito da significância estatística das variáveis. Este processo geralmente envolve a formulação e teste de hipóteses estatísticas para determinar se as variáveis independentes são significativas para determinação da resposta do modelo (Hosmer e Lemeshow, 2000). Assim, a significância das variáveis foi aferida por

análises de t-test e p-valor. A aplicação do modelo de regressão logística e a análise da significância das variáveis independentes foi feita no software R, através da função “glm”.

Adicionalmente, a acurácia do modelo foi medida a partir da obtenção da matriz de confusão para classificação das respostas nas categorias binárias 1 (concordo) ou 0 (discordo). O limiar de classificação estabelecido foi de 0,5, sendo 15% da amostra utilizada para teste da performance de classificação dos modelos (treinados com 85% dos dados).

## 5 RESULTADOS

Este capítulo é composto por 6 subseções. A primeira subseção se destina à descrição da amostra de informações coletadas pela aplicação do questionário elaborado para este estudo. Em seguida, a segunda subseção descreve a estrutura final dos modelos propostos para análise das hipóteses definidas para a questão de pesquisa. A terceira e a quarta subseções apresentam, respectivamente, a análise dos resultados obtidos para a estimação dos modelos por RNA e por regressão logística. Por sua vez, a quinta seção aborda uma análise comparativa dos resultados das duas técnicas aplicadas para modelagem do comportamento do consumidor do comércio eletrônico. Por fim, a sexta subseção consiste na discussão final dos resultados, com destaque às principais implicações dos resultados obtidos.

### 5.1 Descrição da amostra

O público-alvo da pesquisa foram os consumidores do comércio eletrônico, cuja compra gera a necessidade de uma entrega. Logo, especificou-se que produtos comprados para *download* não fariam objeto da pesquisa conduzida.

Os dados foram obtidos por uma pesquisa baseada na Internet disponibilizada na plataforma da Google (formulário Google) entre setembro e outubro de 2019. A amostra ideal estimada foi de 385 inquiridos (95% de nível de confiança e 5% de erro). Foram recebidas 615 respostas, para as quais verificou-se que parte dos respondentes não realizam compras *online*. Estes foram removidos da amostra, de forma que se trabalhou com um total de 595 respostas de consumidores do comércio eletrônico.

A análise de confiabilidade das respostas foi realizada pela aplicação do coeficiente alfa de Cronbach. Considerando todas as questões em escala Likert, obteve-se um valor de coeficiente equivalente a 0,79. O resultado encontrado é satisfatório (acima de 0,7), de modo que para as próximas análises será considerada a totalidade das 595 respostas.

As respostas obtidas tiveram abrangência em todo o território nacional. Os estados com maior representatividade na amostra coletada são, respectivamente, Minas Gerais (60%), São Paulo (8%), Rio Grande do Sul (6%) e Ceará (6%), sendo que 45% dos respondentes são de Belo Horizonte, 6ª maior cidade brasileira em termos de população e 4º PIB do país (IBGE, 2020).

No que tange o perfil dos consumidores do comércio eletrônico, 47% dos respondentes são do gênero masculino. A faixa etária predominante é entre 25 a 34 anos e, majoritariamente, os respondentes encontram-se em faixa de renda familiar entre 4 e 10 salários-mínimos. Estes dados estão detalhados na Tabela 5.1.

Tabela 5.1 Características sociodemográficas da amostra

| Características |                                                        | Frequência<br>(n = 595) | Porcentagem |
|-----------------|--------------------------------------------------------|-------------------------|-------------|
| Gênero          | Feminino                                               | 318                     | 53%         |
|                 | Masculino                                              | 277                     | 47%         |
| Faixa Etária    | 15 a 24 anos                                           | 123                     | 21%         |
|                 | 25 a 34 anos                                           | 275                     | 46%         |
|                 | 35 a 49                                                | 121                     | 20%         |
|                 | Mais de 50 anos                                        | 76                      | 13%         |
| Renda Familiar  | Até um salário-mínimo (R\$954,00)                      | 9                       | 2%          |
|                 | De 2 a 4 salários-mínimos (R\$955,00 a R\$3.816,00)    | 137                     | 23%         |
|                 | De 4 a 10 salários-mínimos (R\$3.817,00 a R\$9.540,00) | 235                     | 39%         |
|                 | Mais de 10 salários-mínimos (R\$19.081,00)             | 214                     | 36%         |

Quanto aos hábitos de compra, os produtos adquiridos são, principalmente, eletrônicos, produtos de beleza e vestuário. A maioria dos respondentes (69%) gera até uma entrega por mês. Quando perguntados a respeito do valor que se dispunham a pagar pelo frete de um produto no valor de R\$300,00, 6% das pessoas responderam que não pagam o frete. Contudo, a maior parcela dos respondentes se dispõe a pagar entre 6% e 10% do valor do produto no frete. Segundo os entrevistados, os fatores que mais influenciam no valor que estão dispostos a pagar pelo frete são o valor da compra (37%) e o prazo da entrega (29%). Além disso, 38% dos respondentes concordam em pagar mais pelo frete para receber o produto em um prazo menor.

Ademais, 30% dos entrevistados são indiferentes ao período da entrega. Contudo, preferencialmente, a escolha é por entregas realizadas em dias úteis (48%), de segunda a sexta, com distribuição equilibrada entre as opções de entrega em horário comercial (23%) e em período noturno (24%).

Os resultados sobre os hábitos de compra dos respondentes estão apresentados na Tabela 5.2. É importante ressaltar que tais questões no questionário permitiam a seleção de mais de uma classe de resposta, de modo que a soma das frequências pode superar o total de 595 respondentes.

Tabela 5.2 Hábitos de compra da amostra

| Características                                                   | Atributo                           | Frequência | Porcentagem |
|-------------------------------------------------------------------|------------------------------------|------------|-------------|
| Categoria de produtos consumidos                                  | Produtos de beleza e vestuário     | 374        | 25%         |
|                                                                   | Produtos eletrônicos               | 459        | 31%         |
|                                                                   | Produtos de cultura e lazer        | 322        | 21%         |
|                                                                   | Móveis e utensílios domésticos     | 30         | 2%          |
|                                                                   | Outros*                            | 313        | 20%         |
| Frequência de entrega de produtos adquiridos pela Internet        | Uma entrega ou menos por mês       | 410        | 69%         |
|                                                                   | De 2 a 4 entregas por mês          | 159        | 27%         |
|                                                                   | De 5 a 8 entregas por mês          | 22         | 4%          |
|                                                                   | Acima de 9 entregas por mês        | 4          | 1%          |
| Percentual do preço do produto que está disposto a pagar no frete | 0% (não paga o frete)              | 96         | 16%         |
|                                                                   | Menor ou igual a 5%                | 133        | 22%         |
|                                                                   | De 6% a 10%                        | 262        | 44%         |
|                                                                   | Acima de 10%                       | 104        | 17%         |
| Período preferencial para recebimento das mercadorias             | Segunda a sexta-feira, 8h às 18h   | 180        | 23%         |
|                                                                   | Segunda a sexta-feira, 18h às 20h  | 127        | 16%         |
|                                                                   | Segunda a sexta-feira, após às 20h | 65         | 8%          |
|                                                                   | Indiferente                        | 231        | 30%         |
|                                                                   | Sábado                             | 120        | 15%         |
|                                                                   | Domingo ou feriado                 | 55         | 7%          |

\*Alimentos e produtos de saúde, automobilísticos, esporte ou pet.

Na Tabela 5.3 é apresentada a análise descritiva das respostas ao grau de importância para os parâmetros de serviço, produto e entrega do comércio eletrônico, além do valor percebido dos atributos valor do frete e agendamento da entrega (possibilidade de escolha de dia e horário). Pela observação conjunta dos resultados de média, do 1º quartil (representando 25% dos respondentes) e do 3º quartil (75% dos respondentes), em geral, os respondentes atribuem maior importância ao preço do produto, à comodidade de comprar sem sair de casa e às oportunidades de compra pelo comércio eletrônico. Por sua vez, entre os parâmetros de entrega, o valor do frete é considerado um fator de grande importância na hora da compra *online*, seguido do prazo de entrega. O parâmetro de horário de entrega foi avaliado, em geral, como um aspecto neutro.

Tabela 5.3 Análise descritiva dos parâmetros de valor percebido

| Parâmetros de valor percebido                                                                                        | Alfa de Cronbach | Mín. | 1° quartil | Mediana | 3° quartil | Máx. |
|----------------------------------------------------------------------------------------------------------------------|------------------|------|------------|---------|------------|------|
| Importância da comodidade de comprar sem sair de casa                                                                | 0,783            | 1    | 3          | 3       | 3          | 3    |
| Importância da privacidade                                                                                           | 0,777            | 1    | 2          | 3       | 3          | 3    |
| Importância do preço do produto                                                                                      | 0,780            | 1    | 3          | 3       | 3          | 3    |
| Importância da oportunidade de compra (aproveitar ofertas relâmpagos ou benefícios)                                  | 0,778            | 1    | 3          | 3       | 3          | 3    |
| Importância do prazo de entrega                                                                                      | 0,767            | 1    | 2          | 3       | 3          | 3    |
| Importância do valor do frete                                                                                        | 0,777            | 1    | 3          | 3       | 3          | 3    |
| Importância do horário da entrega                                                                                    | 0,770            | 1    | 2          | 2       | 3          | 3    |
| O valor do frete é um fator determinante para você comprar pela Internet.                                            | 0,7994           | 1    | 3          | 3       | 3          | 3    |
| A possibilidade de escolher o dia e o horário de entrega influencia na decisão de realizar uma compra pela Internet. | 0,7900           | 1    | 1          | 2       | 3          | 3    |

A confiabilidade da amostra por parâmetro é atestada pelos resultados do alfa de Cronbach (superiores a 0,7). Desse modo, todos os itens foram utilizados nas análises posteriores.

## 5.2 Descrição dos modelos

A definição das variáveis independentes que compuseram cada modelo passou por um processo inicial de análise de multicolinearidade. A detecção de fortes relações de dependência orientou a seleção do conjunto de variáveis a ser aplicado a cada modelo, uma vez que tal aspecto pode dificultar a identificação dos efeitos isolados de cada variável independente ( $X_i$ ) em  $Y$ .

Assim, a partir da aplicação do coeficiente de Spearman, analisou-se a existência de relações monótonas entre as 13 variáveis explicativas. Os resultados, que variam entre -1 e 1, são apresentados no formato de matriz de correlação (Figura 5.1). Em geral, os resultados de correlação obtidos são baixos, situando-se entre o intervalo -0,5 e 0,5. Estes valores são indicativos de baixa relação de dependência entre variáveis.

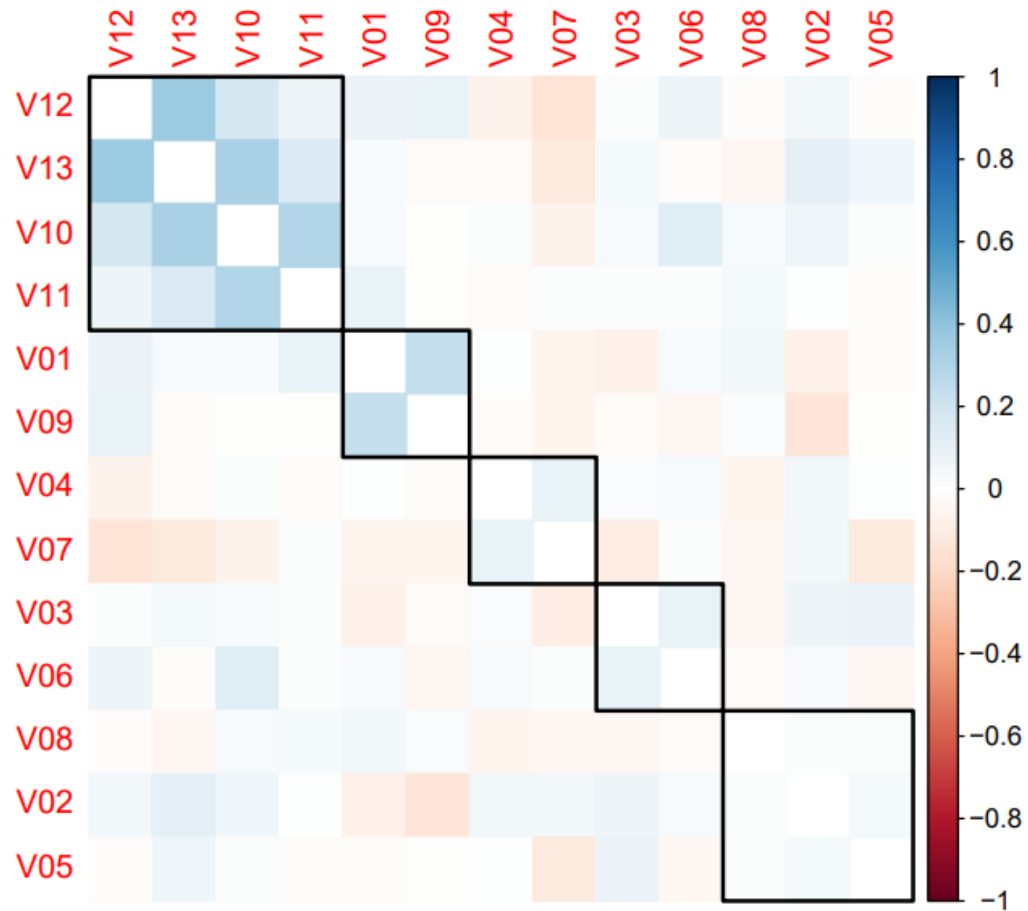


Figura 5.1 Matriz de correlação Spearman – Variáveis independentes

Constatada a baixa correlação entre as variáveis analisadas, todas foram consideradas na composição dos modelos propostos. Os conjuntos de variáveis por modelo, incluindo as variáveis ‘resposta’, são apresentados na Tabela 5.4.

Tabela 5.4 Variáveis por modelo de análise

|     |                                                                     | Modelo1                      | Modelo2                              | Modelo3                         | Modelo4                       | Modelo5                           |
|-----|---------------------------------------------------------------------|------------------------------|--------------------------------------|---------------------------------|-------------------------------|-----------------------------------|
|     |                                                                     | Influência do Valor do Frete | Influência do Agendamento da Entrega | Importância do Prazo de Entrega | Importância do Valor do Frete | Importância do Horário da Entrega |
| V01 | Compra produtos de beleza, vestuário ou afins                       | •                            | •                                    | •                               | •                             | •                                 |
| V02 | Compra produtos de eletrônicos                                      | •                            | •                                    | •                               | •                             | •                                 |
| V03 | Compra produtos de cultura e lazer                                  | •                            | •                                    | •                               | •                             | •                                 |
| V04 | Compra móveis e utensílios domésticos                               | •                            | •                                    | •                               | •                             | •                                 |
| V05 | Percentual do preço do produto que está disposto a pagar pelo frete | •                            | •                                    | •                               | •                             | •                                 |
| V06 | Frequência de consumo                                               | •                            | •                                    | •                               | •                             | •                                 |
| V07 | Faixa etária                                                        | •                            | •                                    | •                               | •                             | •                                 |
| V08 | Renda familiar                                                      | •                            | •                                    | •                               | •                             | •                                 |

|     |                                                                      |   |   |   |   |   |
|-----|----------------------------------------------------------------------|---|---|---|---|---|
| V09 | Gênero                                                               | • | • | • | • | • |
| V10 | Comodidade de comprar sem sair de casa                               | • | • | • | • | • |
| V11 | Privacidade                                                          | • | • | • | • | • |
| V12 | Oportunidade de compra (aproveitar ofertas relâmpagos ou benefícios) | • | • | • | • | • |
| V13 | Preço do produto                                                     | • | • | • | • | • |

### 5.3 Estimação dos Modelos por RNA

Os conjuntos de variáveis independentes selecionados para análise de cada parâmetro de entrega foram introduzidos à rede neural, estruturada e treinada. A fase de testes e validação do modelo utilizou 15% da amostra de dados (correspondente a 90 observações). Ao final de cada processamento na RNA, a qualidade dos modelos foi avaliada a partir da métrica de acurácia, que indica a capacidade de classificação correta das respostas.

Num primeiro momento, consideraram-se três categorias de resposta para as variáveis dependentes: (1) ‘Discordo’, (2) ‘Neutro’, (3) ‘Concordo’. Contudo, a RNA apresentou baixa performance para classificação das respostas neutras. Numa tentativa de melhorar a acurácia dos modelos, simplificaram-se as respostas para uma categoria binária: (0) Discordo e (1) Concordo.

Nesse formato, observou-se melhoria da acuracidade dos modelos na RNA. As respostas neutras foram reclassificadas como neutro positivo e neutro negativo. A comparação das métricas de qualidade dos modelos, ao testar as respostas neutras como positivas e como negativas, é apresentada na Tabela 5.5. Os resultados indicam que a RNA tem mais facilidade de classificar as respostas neutras quando essas são consideradas positivas. Compreende-se, assim, que aqueles respondentes que se posicionam como ‘neutros’ são susceptíveis à influência dos parâmetros de entrega ao comprar pela Internet. Desse modo, as análises prosseguiram a partir da reclassificação das respostas neutras como positivas. No apêndice 8.4, são apresentados os resultados completos das métricas de qualidade dos modelos estimados por rede neural.

Tabela 5.5 Reclassificação de respostas neutras - Comparação da acurácia dos modelos

| Codificação | Variáveis                                            | Acurácia (%)    |                 |
|-------------|------------------------------------------------------|-----------------|-----------------|
|             |                                                      | Neutro Positivo | Neutro Negativo |
| Modelo 1    | Influência do Valor do Frete na efetivação da compra | 88%             | 70%             |
| Modelo 2    | Influência do Agendamento na efetivação da compra    | 60%             | 57%             |
| Modelo 3    | Importância do Prazo da Entrega na hora da compra    | 82%             | 54%             |
| Modelo 4    | Importância do Valor do Frete na hora da compra      | 84%             | 70%             |
| Modelo 5    | Importância do Horário da Entrega na hora da compra  | 64%             | 50%             |

Considerando as respostas neutras positivas, pode-se proceder com a interpretação dos resultados dos modelos, a fim de verificar as hipóteses colocadas em resposta à questão de pesquisa deste trabalho. Para tanto, propôs-se análise das relações internas do algoritmo da RNA pela determinação dos *Partial Dependence Plots* (PDP). Os PDP mostram a dependência entre a variável ‘resposta’ e uma ou duas características (variáveis independentes), marginalizando os valores de todas as outras variáveis (as características ‘complemento’). Em outras palavras, os PDP permitem observar como uma alteração numa variável independente afeta a variável ‘resposta’, pela visualização gráfica bidimensional dos resultados, apresentados nas próximas subseções a seguir.

### 5.3.1 Influência do valor do frete na efetivação da compra

Nesta seção são discutidos os resultados do modelo que considera as características do consumidor e dos parâmetros do comércio eletrônico na influência do valor do frete na efetivação da compra. A Figura 5.2 apresenta as relações individuais dos PDP de cada variável explicativa com a variável ‘resposta’.

Os tipos de produto adquiridos pela Internet apresentam relação linear com a influência do valor do frete na realização da compra. Nota-se que os hábitos de comprar produtos de beleza e afins (V01), produtos eletrônicos (V02) e produtos de cultura e lazer (V03), como livros, colecionáveis e instrumentos musicais, apresentam relação positiva (curva crescente) com a variável ‘resposta’. Em contrapartida, ao adquirir móveis e utensílios domésticos (V04), o consumidor demonstra menor ponderação pelo valor do frete.

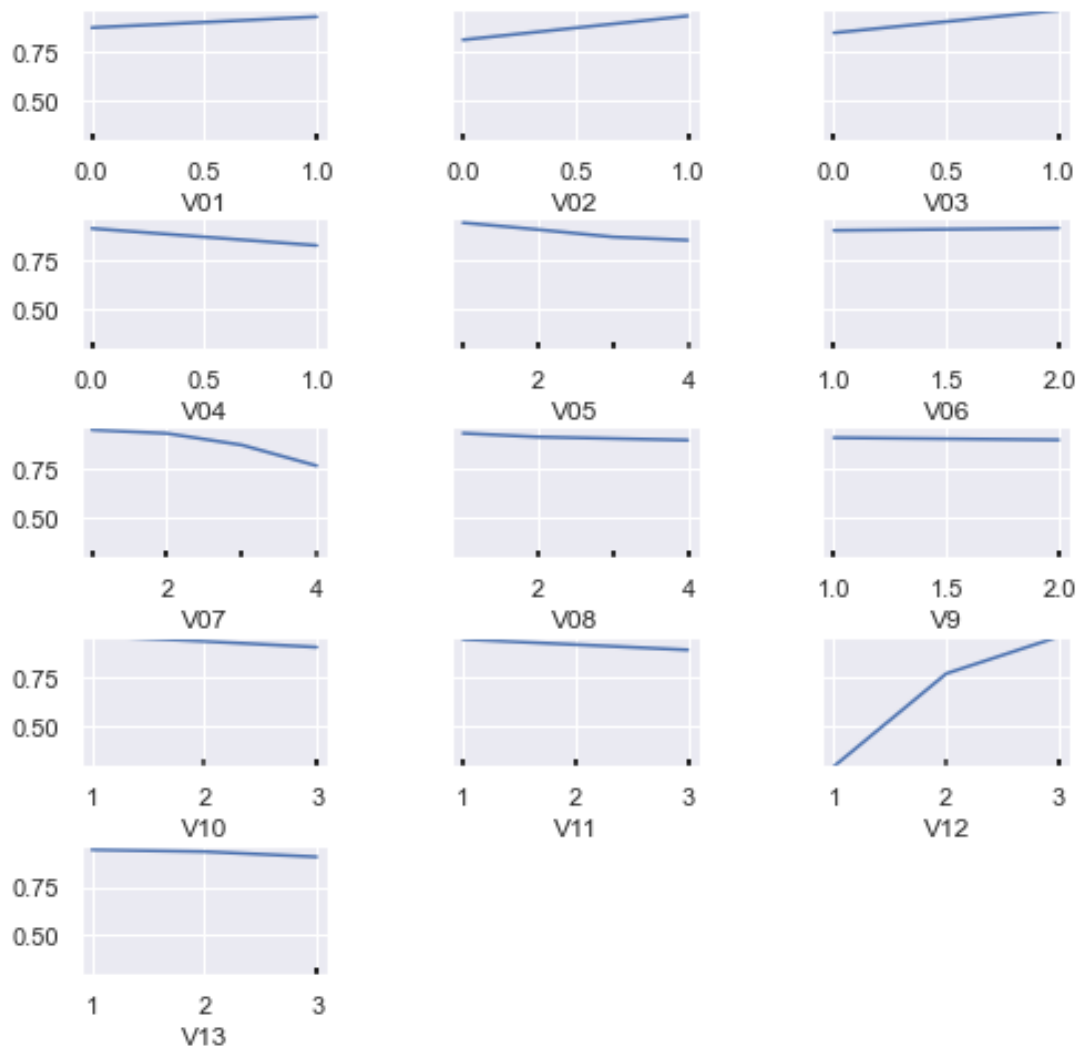


Figura 5.2 PDP – Análise da Influência do ‘valor do frete’ na efetivação da compra pela Internet

Ademais, cabe notar a relação constante com a variável frequência de consumo (V06), o que demonstra que a variabilidade dessa característica de consumo não altera o resultado da influência do valor do frete na efetivação da compra pelo comércio eletrônico. O mesmo pode ser dito da variável gênero (V09).

Quanto à relação de dependência com características sociodemográficas dos consumidores, a variável ‘resposta’ apresenta relação decrescente com a faixa etária (V07). Observa-se que consumidores de maiores idades são menos influenciados pelo valor a ser pago pelo frete. Assim como maiores faixas de renda familiar (V08), que demonstram menor influência pelo valor do frete na efetivação da compra. A Figura 5.3 demonstra a *Partial Dependence Plot* – PDP para as duas variáveis faixa etária (V07) e renda (V08). Percebe-se que grupos etários mais

jovens, mesmo os de maior renda, são os mais influenciados pelo valor do frete ao efetuar uma compra *online*. Na figura demonstra-se, também, a relação com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08). Observa-se que consumidores de maior renda, que se dispõem a pagar mais pelo frete, são menos influenciados por esse fator de entrega.

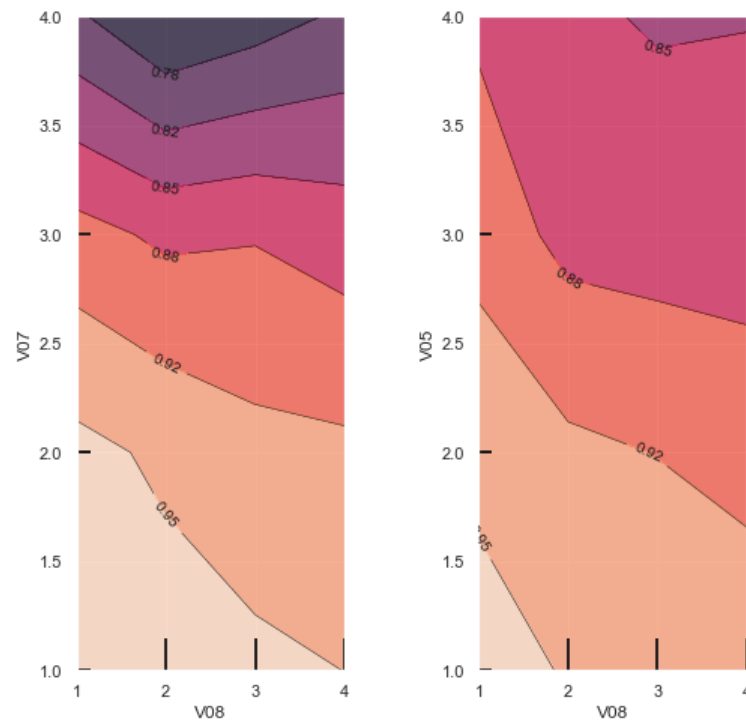


Figura 5.3 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) dos consumidores (à esquerda) | Relação de dependência com o percentual a pagar pelo frete (V05) e com a renda familiar (V08) dos consumidores (à direita)

No que tange aos parâmetros do comércio eletrônico, a relação com as variáveis comodidade de comprar sem sair de casa (V10), privacidade (V11) e preço do produto (V13) é decrescente, com pouca variação para a variável ‘resposta’. Por sua vez, quanto maior a importância atribuída às oportunidades de compra (V12), maior a influência exercida pelo valor do frete na efetivação da compra.

A relação de dependência com os parâmetros comodidade, privacidade e preço do produto é demonstrada na Figura 5.4. Observa-se que as três características demonstram efeitos semelhantes e de mesmo grau de intensidade sobre a variável ‘resposta’ (com curvas inclinadas a ângulos próximos a 45°). A Figura 5.4 também apresenta a dependência que a influência do valor do frete tem em relação ao preço do produto e ao percentual do valor do produto que o

consumidor está disposto a pagar pelo frete (V05). Novamente, ambas as variáveis demonstram efeito e grau de intensidade semelhantes sobre a variável ‘resposta’.

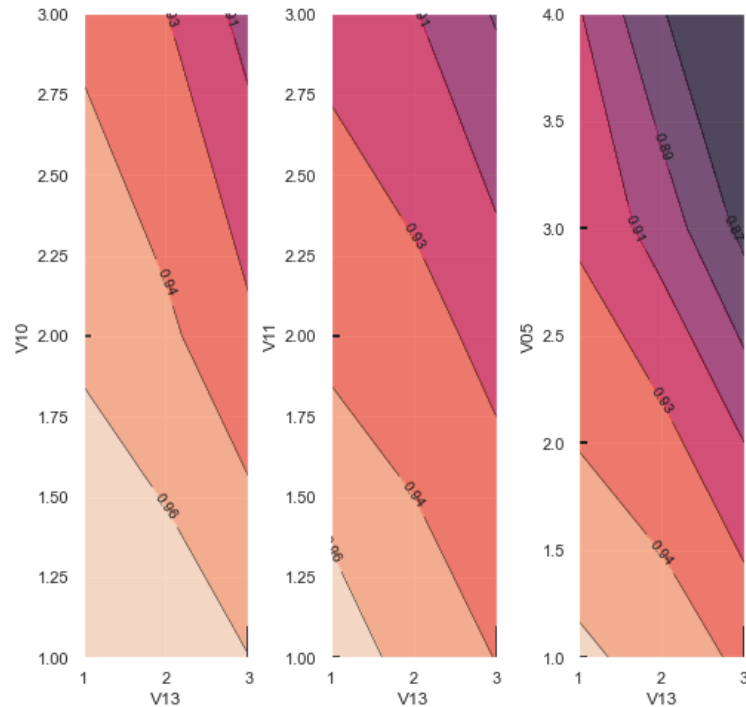


Figura 5.4 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda) | Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio) | PDP – Relação de dependência com o percentual a pagar pelo frete (V05) e com o preço do produto (V13) (à direita)

### 5.3.2 Influência da escolha do dia e horário de entrega na efetivação da compra

O segundo modelo propõe a análise das características do consumidor e dos parâmetros do comércio eletrônico e sua relação com a influência da possibilidade de agendar a entrega (escolher dia e horário) na efetivação da compra pela Internet. Os PDP gerados para análise de cada variável explicativa do modelo são apresentadas na Figura 5.5. Pela observação das características de consumo, entende-se que os compradores com hábito de adquirir produtos de moda, cosméticos e afins (V01) têm menor probabilidade de serem influenciados pela disponibilidade do agendamento da entrega ao comprar pela Internet. Tipicamente, por tratar-se de produtos de confecção pautada na demanda, podem exigir maiores prazos e, assim, sofrer menor influência da opção por escolha de dia e horário de entrega. Por outro lado, aqueles que consomem produtos eletrônicos (V02), produtos de cultura e lazer (V03) e móveis e utensílios domésticos (V04), podem considerar o agendamento como aspecto positivo para realizar a

compra. Ademais, quanto maior o valor do produto que se está disposto a pagar no frete (V05), menor a influência que o agendamento da entrega terá na efetivação da compra.

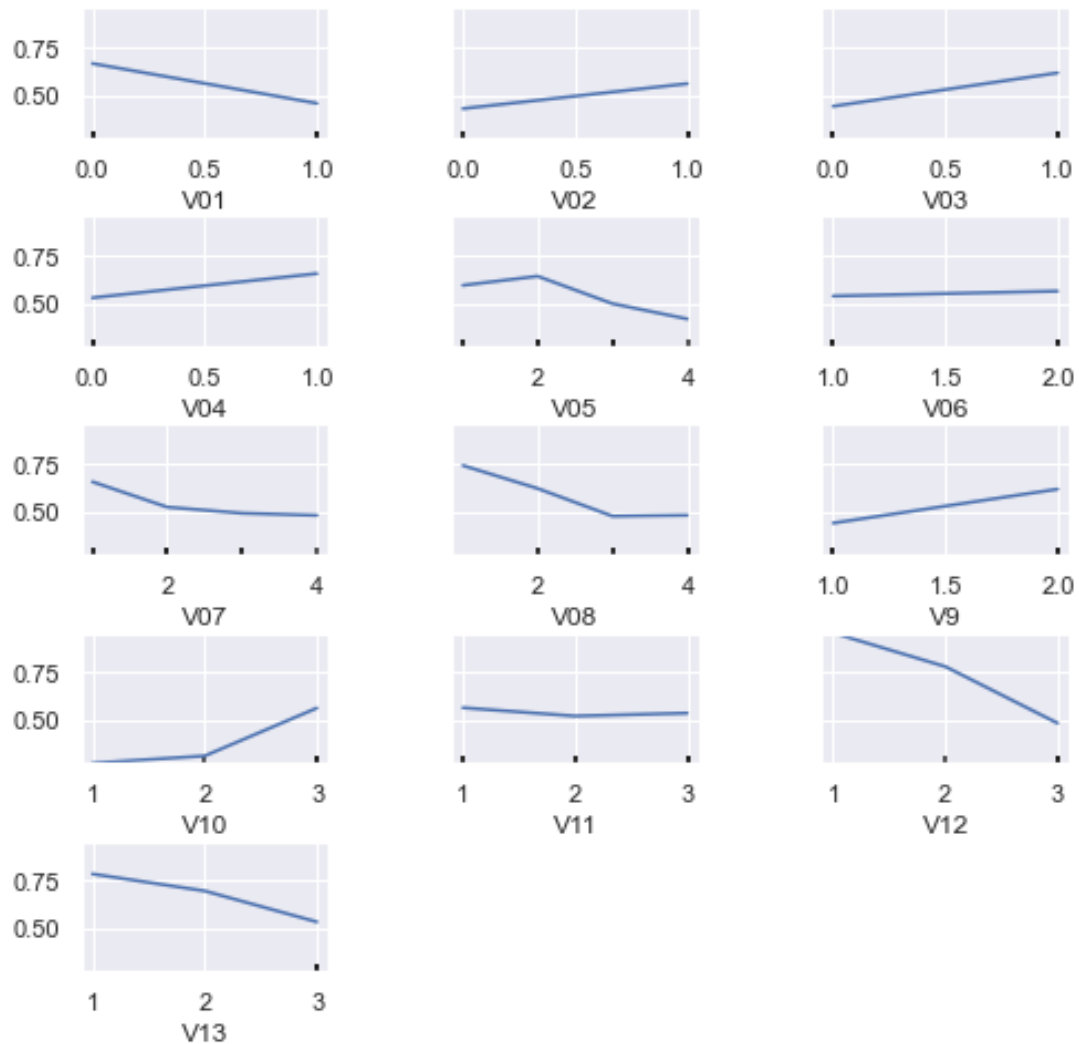


Figura 5.5 PDP – Análise da Influência da ‘escolha do dia e da hora de entrega’ na efetivação da compra pela Internet

A análise para o grupo de características sociodemográficas demonstrou que quanto maior a faixa etária (V07) do consumidor do comércio eletrônico, menos influenciado é pelo agendamento da entrega. Igualmente, quanto maior a renda familiar (V08) do consumidor, menor a influência exercida pelo atributo de entrega.

A Figura 5.6 apresenta a análise da dependência da variável ‘resposta’ com base no percentual do valor do produto que se dispõe a pagar no frete (V05), na faixa etária (V07) e na renda familiar (V08). Consumidores de renda mais baixa demonstram ser mais influenciados pelo agendamento da entrega, independentemente da faixa etária e do valor que se dispõem a pagar

no frete (pois as curvas dos diagramas são quase paralelas ao eixo y). Assim, entende-se que para esse perfil de consumidor, o agendamento da entrega agrega maior valor ao frete.

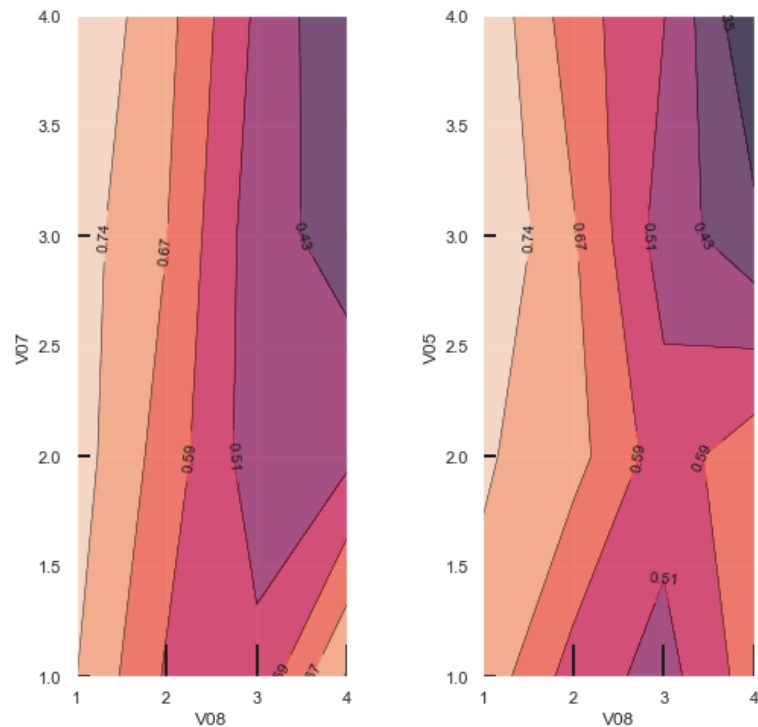


Figura 5.6 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda) | Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita)

No que tange os parâmetros do serviço do comércio eletrônico, cabe ressaltar a dependência que a variável ‘resposta’ demonstra em relação aos parâmetros comodidade de comprar sem sair de casa (V10), oportunidades de compra (V12) e preço do produto (V13). Quanto maior a importância de comprar sem sair de casa, maior será a influência do agendamento da entrega. Por sua vez, quanto maior a importância dada às promoções e preço do produto, menos o consumidor se preocupa em determinar dia e horário para receber o produto.

A Figura 5.7 apresenta a análise da dependência da variável ‘resposta’ com as características do comércio eletrônico e a variável V05. Observa-se que os atributos comodidade e preço do produto têm pesos equivalentes para efeitos opostos sob a variável ‘resposta’. Comparativamente, à variável oportunidades de compra, o preço do produto é pouco relevante à determinação da influência do agendamento da entrega, uma vez que as curvas no diagrama chegam a inclinações quase paralelas ao eixo de V13. Ademais, quanto maior a importância

atribuída ao preço do produto e maior o valor que se dispõe pagar pelo frete, menos o parâmetro agendamento influencia na efetivação da compra.

Esses resultados indicam que, em geral, a possibilidade de escolher o dia e horário para receber o produto não agrega valor ao serviço de entrega em si. Em outras palavras, para o consumidor, esse atributo não justifica pagar maior valor pelo frete.

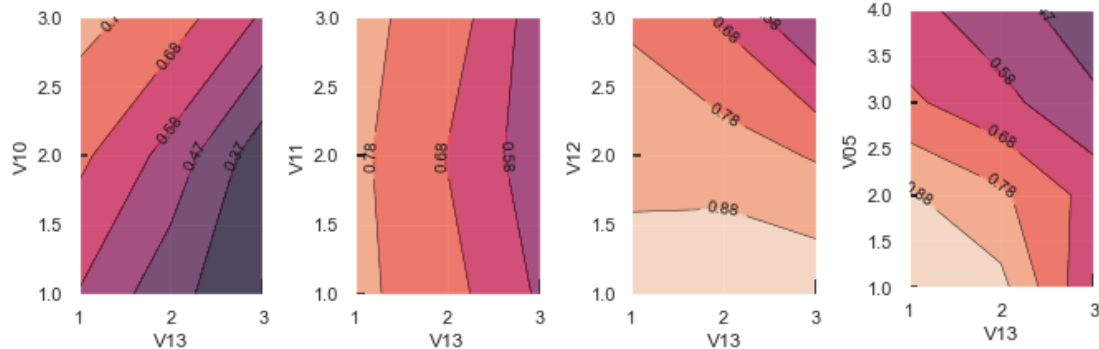


Figura 5.7 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda) | PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior) | PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita) | PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior)

### 5.3.3 Importância do prazo de entrega na compra pela Internet

Esse modelo buscou analisar a importância do prazo de entrega na realização da compra pelo comércio eletrônico. A Figura 5.8 apresenta os PDP para cada variável explicativa do modelo. As análises de dependência das características de consumo destacam a relação positiva com hábito de comprar produtos eletrônicos (V03). Contudo, com as demais variáveis, os *plots* demonstram relação quase constante com a variável ‘resposta’.

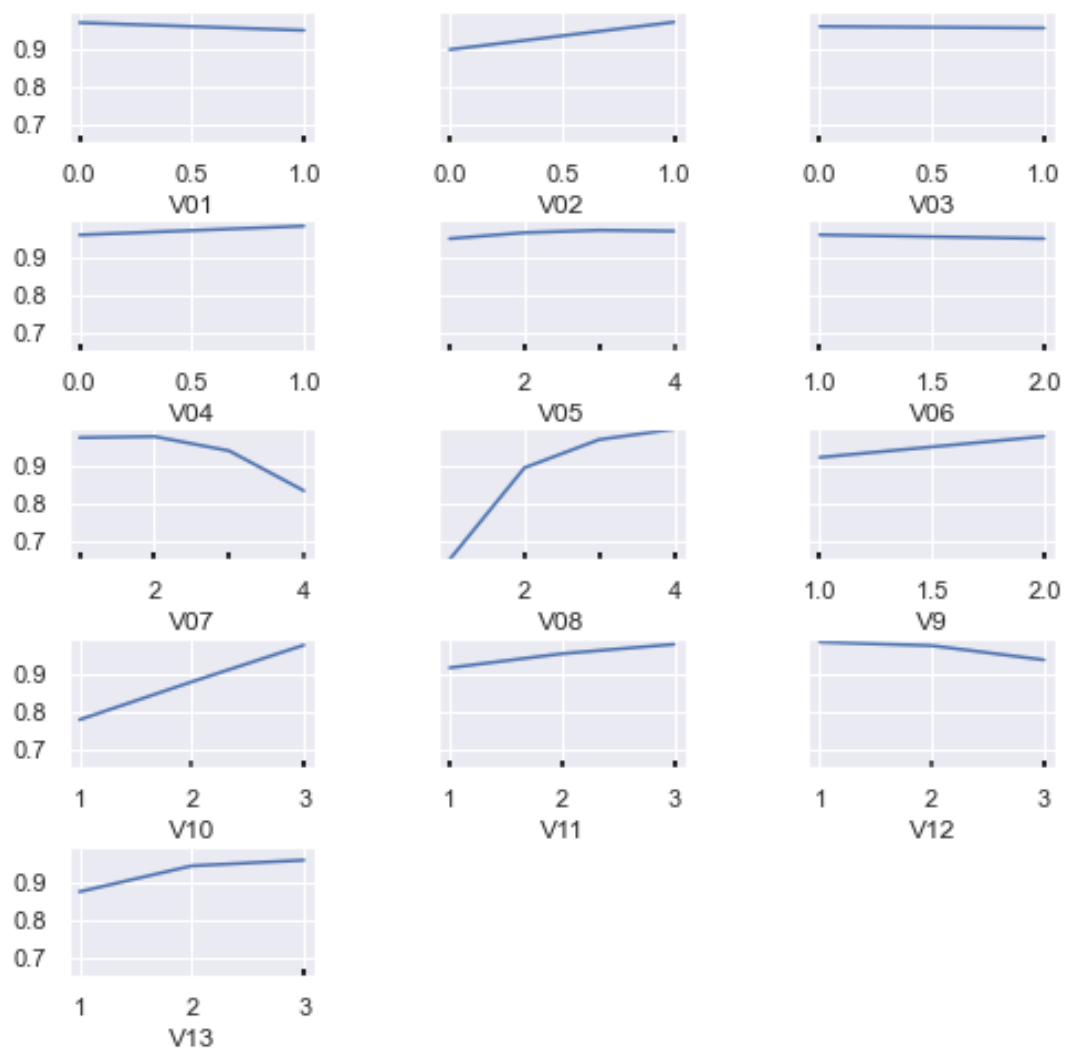


Figura 5.8 PDP – Análise da Importância do ‘prazo da entrega’ para a realização da compra pela Internet

Por sua vez, a variável ‘resposta’ demonstra, respectivamente, efeito decrescente e efeito crescente em relação às características sociodemográficas faixa etária (V07) e renda familiar (V08).

A Figura 5.9 apresenta a relação da variável ‘resposta’ com a faixa etária, a renda familiar e o percentual do valor do produto que se pagaria no frete (V05). Reforça-se que a influência do prazo de entrega é sobretudo orientada pela renda familiar, pois mesmo consumidores mais idade, quando pertencentes a maiores classes de renda, demonstram maior ponderação por esse atributo de entrega.

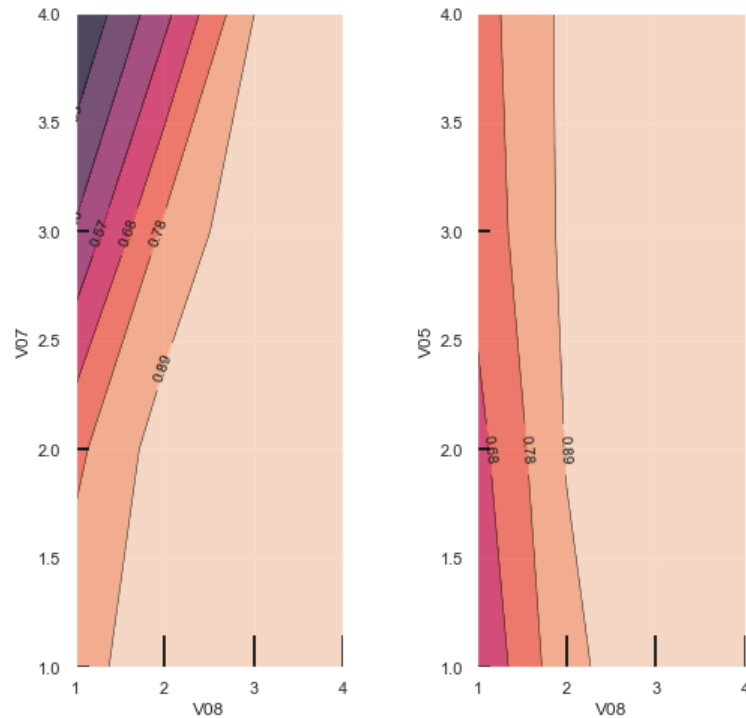


Figura 5.9 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda) | PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita)

No que tange as características do comércio eletrônico, a variável ‘resposta’ demonstra relação não-crescente com as oportunidades de compra (V12). Por outro lado, a relação com as demais características é crescente, com destaque para a comodidade de comprar sem sair de casa (V10).

A Figura 5.10 apresenta os PDP com a combinação entre essas variáveis e o preço do produto (V13). Nota-se que a importância do prazo de entrega para a realização da compra é mais dependente da comodidade e das oportunidades de compra do que do preço do produto em si. Por outro lado, o preço do produto e a privacidade demonstram pesos similares. Ademais, por maior que seja a disposição em pagar mais pelo frete (V05), a influência do prazo de entrega aumenta, sobretudo, em função da importância atribuída ao preço do produto.

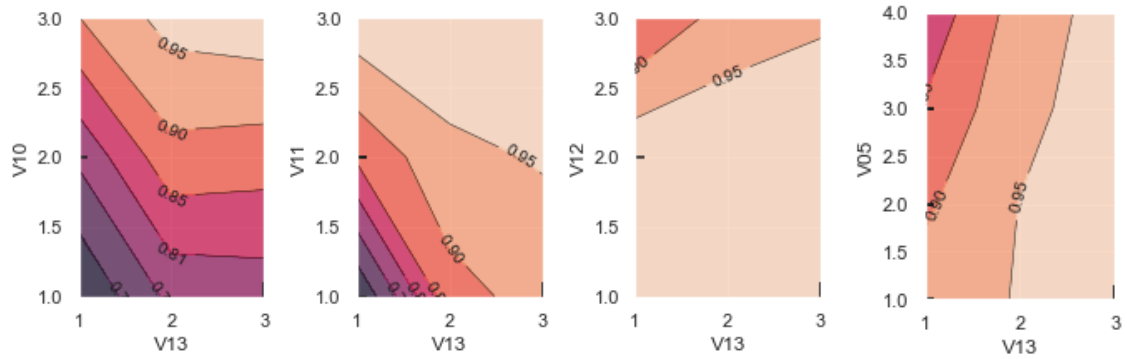


Figura 5.10 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda) | PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior) | PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita) | PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior)

#### 5.3.4 Importância do valor do frete na compra pela Internet

O modelo buscou capturar a importância do valor do frete na realização da compra *online*. A Figura 5.11 apresenta os PDP para cada variável explicativa do modelo. Os resultados de dependência com as características de consumo indicam pouca variação em relação às categorias de produtos adquiridos pela Internet e em relação à frequência de realização de compras (relação quase constante, com inclinações muito baixas). Ademais, observa-se que quanto mais disposto o consumidor está de pagar a mais pelo frete (V05), maior a importância percebida para esse serviço e maior sua influência na realização da compra.

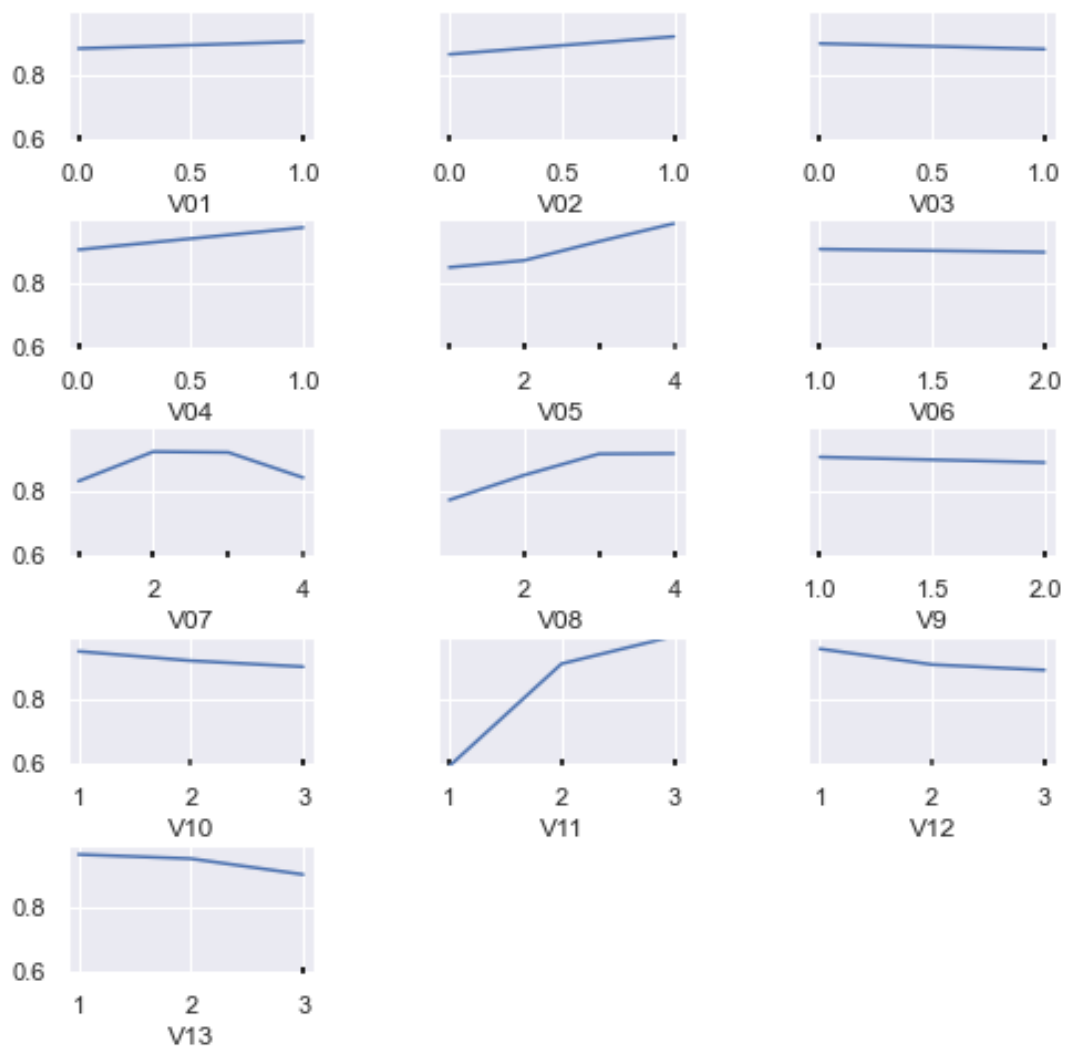


Figura 5.11 PDP – Análise da Importância do ‘valor do frete’ para a realização da compra pela Internet

No que tange as características sociodemográficas, a relação com a variável faixa etária (V07) demonstra curva com três padrões: crescente até 34 anos, constante entre 35 e 49 anos, decrescente para consumidores acima de 50 anos. Por sua vez, a relação com as variáveis renda familiar (V08) gênero (V09) é, respectivamente, crescente e constante.

A Figura 5.12 apresenta os PDP com cruzamento das variáveis faixa etária (V07) e valor que pagaria no frete (V05) com a renda familiar (V08). A análise simultânea da faixa etária com a renda familiar é complexa e de difícil interpretação. Esse fato pode sugerir que tal relação deva ser explorada sob outras perspectivas, com a incorporação de características como o nível de escolaridade do consumidor, que não foi abordada nesta dissertação. A figura também

demonstra que para consumidores de menor renda, o valor do frete é menos importante apesar da predisposição em pagar mais pelo frete.

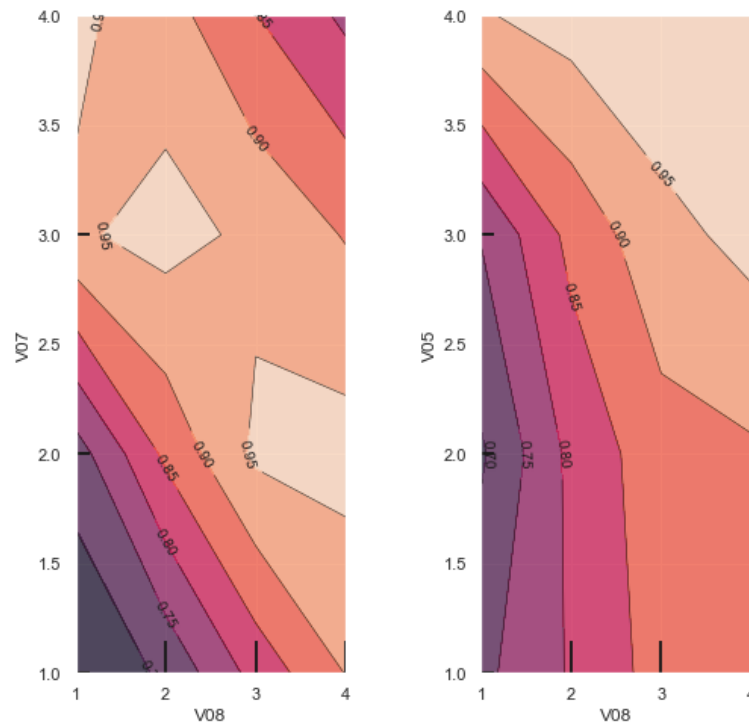


Figura 5.12 PDP – Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda) | PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita)

A análise das características do comércio eletrônico indica relação crescente com a privacidade (V11). Ou seja, quanto maior a importância da privacidade para o consumidor do comércio eletrônico, menor a importância do valor do frete para a realização da compra. Contudo, com os demais aspectos a relação demonstra-se decrescente, com baixa variabilidade para a variável ‘resposta’ (baixas inclinações das curvas).

A Figura 5.13 compara a dependência das características do comércio eletrônico, além do percentual do valor do produto que se pagaria no frete (V05), com o preço do produto (V13). Observa-se a preponderância da privacidade em relação à importância dada ao preço do produto (curvas quase paralelas ao eixo de V13). Complementarmente, a pré-disposição em pagar mais pelo frete demonstra a importância que o consumidor atribui ao valor do frete na hora de comprar pela Internet. Esse fato prevalece apesar da relação decrescente da variável ‘resposta’ com o preço do produto (tal como observado na Figura 5.13).

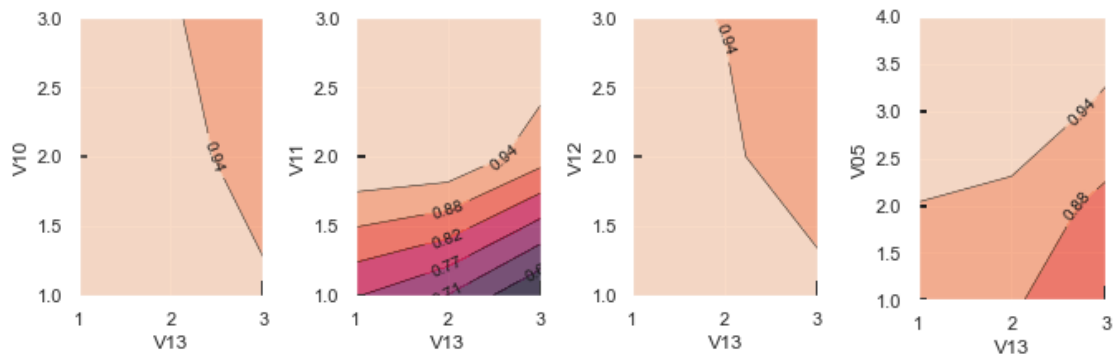


Figura 5.13 PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda) | PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior) | PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita) | PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior)

### 5.3.5 Importância do horário de entrega na compra pela Internet

O quinto e último modelo, por sua vez, buscou avaliar a importância do horário de entrega na compra no comércio eletrônico. A Figura 5.14 apresenta os PDP para cada variável explicativa do modelo, cujos resultados destacam a dependência da variável ‘resposta’ em relação à comodidade de comprar sem sair de casa (V10) e em relação ao preço do produto (V13). Quanto maior a importância atribuída a tais características do comércio eletrônico, maior a importância do horário da entrega na hora de comprar pela Internet. Com as demais variáveis, a relação demonstrada é praticamente constante (baixa inclinação das curvas dos *plots*).

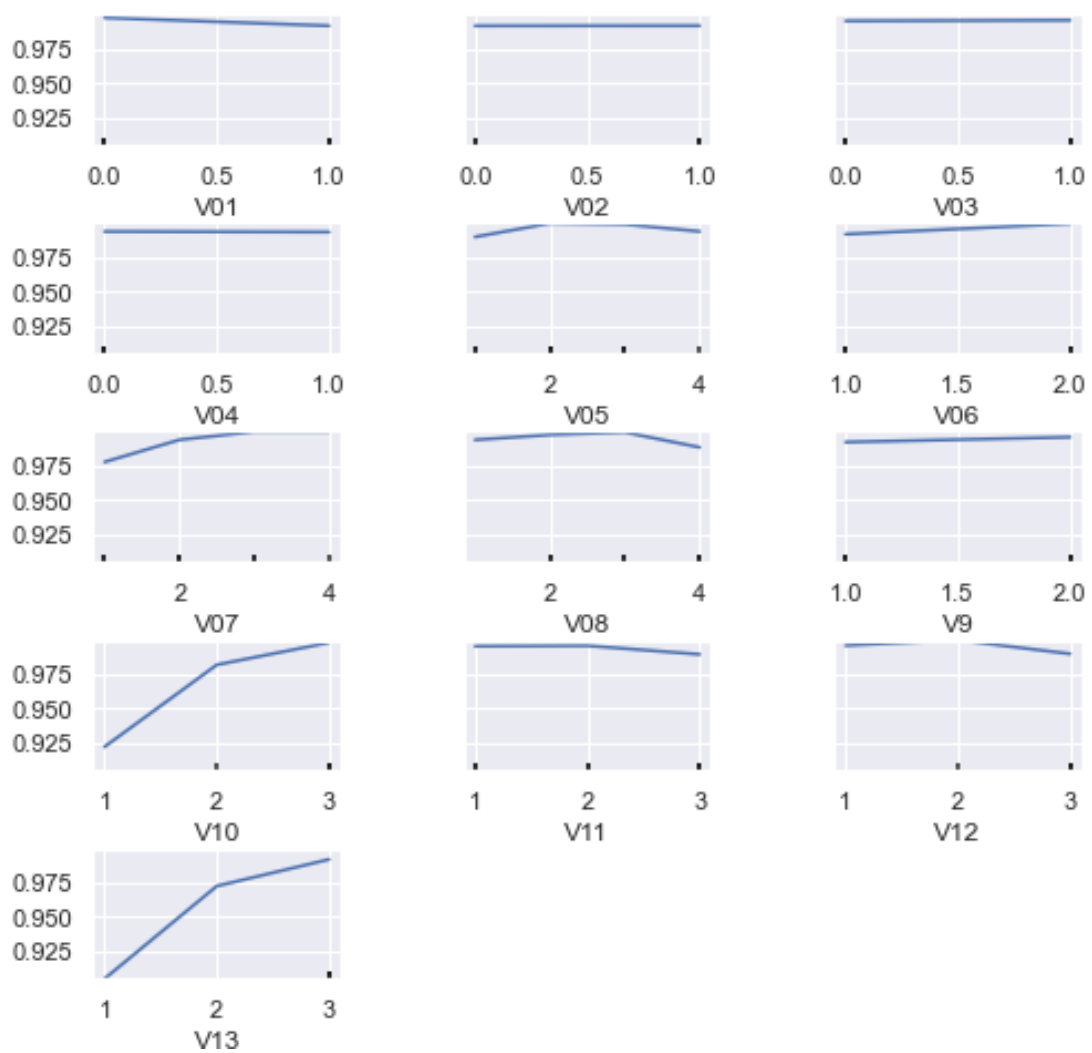


Figura 5.14 PDP – Análise da Importância do ‘horário da entrega’ para a realização da compra pela Internet

Na Figura 5.15 são apresentadas os PDP com comparação das variáveis percentual que se dispõe a pagar pelo frete (V05) e faixa etária (V07) com a renda familiar (V08). Apesar da complexidade das curvas, de modo geral compreende-se que os grupos etários têm maior peso do que a renda na determinação da importância do parâmetro horário da entrega para o consumidor do comércio eletrônico (curvas quase paralelas ao eixo de V08). Por outro lado, a renda familiar é mais relevante à análise do fenômeno do que o valor que se disporia a pagar no frete (curvas quase paralelas ao eixo de V05).

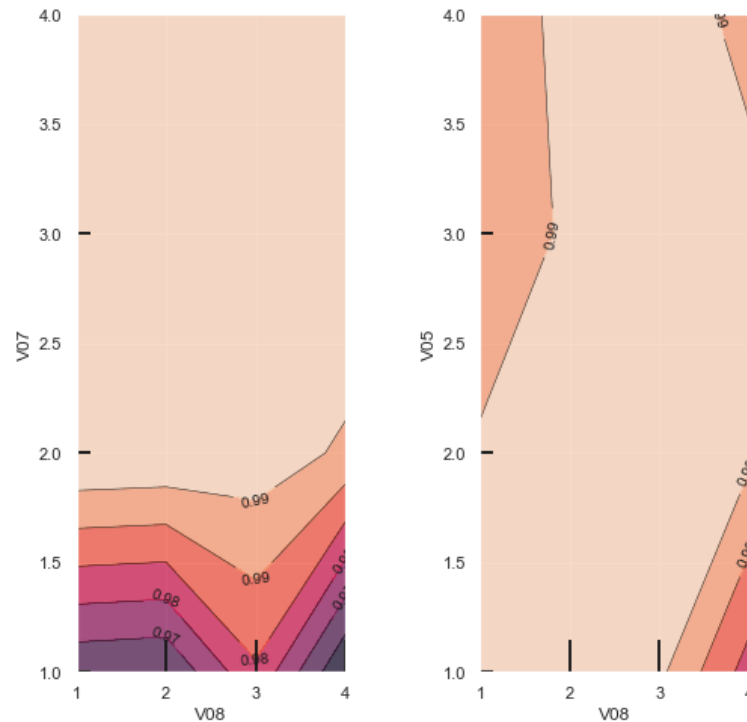


Figura 5.15 Relação de dependência com a faixa etária (V07) e com a renda familiar (V08) (à esquerda) | PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com a renda familiar (V08) (à direita)

Por sua vez, na Figura 5.16 são apresentadas os PDP com a combinação entre as variáveis comodidade (V10), privacidade (V11), oportunidades de compra (V12) e valor que se pagaria no frete (V05) com o preço do produto (V13). A comodidade e o preço do produto demonstram efeitos semelhantes e de mesmo grau de intensidade sobre a variável ‘resposta’ (com curvas inclinadas a ângulos próximos a 45°). Ademais, entende-se que a importância do fator horário da entrega ao comprar pela Internet independe do quanto o consumidor está disposto a pagar no frete. Ou seja, o valor que o consumidor percebe em relação ao serviço de entrega não tem relação com o horário de recepção do produto.

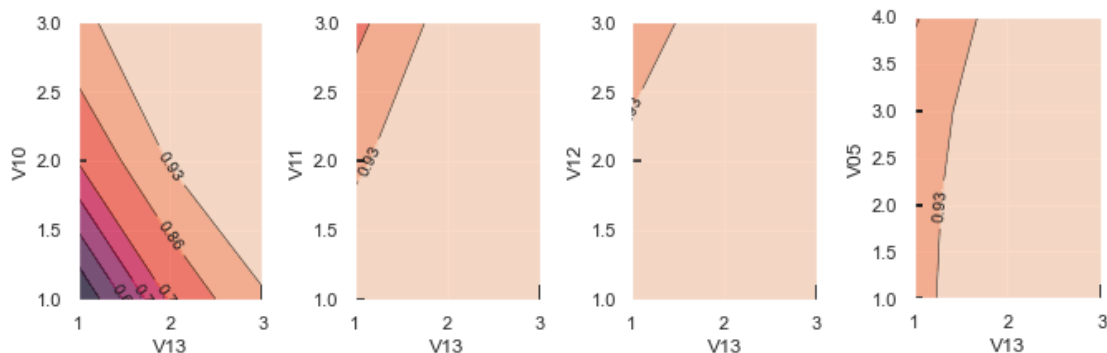


Figura 5.16 PDP – PDP – Relação de dependência com a comodidade (V10) e com o preço do produto (V13) (à esquerda) | PDP – Relação de dependência com a privacidade (V11) e com o preço do produto (V13) (meio superior) | PDP – Relação de dependência com a oportunidade de compra (V12) e com o preço do produto (V13) (à direita) | PDP – Relação de dependência com o percentual do valor do produto que se pagaria no frete (V05) e com o preço do produto (V13) (meio inferior)

### 5.3.6 Comparação dos modelos estimados por RNA

As análises anteriores demonstraram que uma mesma variável explicativa pode se relacionar diferentemente com cada parâmetro de entrega modelado. Pelos PDP foi possível aferir não apenas como, mas também o quanto a variável ‘resposta’ é dependente de cada variável explicativa.

Nesse sentido, a Tabela 5.6 reúne os resultados dos efeitos estimados das variáveis em cada modelo. O sentido das curvas dos *plots* foi representado pelos sinais ‘+’, para curvas crescentes, e ‘-’, para curvas decrescentes. Também são indicados efeitos constantes e complexos, sendo estes últimos representados por uma função côncava ou convexa. Por sua vez, o grau de dependência estimado para cada variável explicativa foi representado qualitativamente, com base na análise das inclinações das curvas pelas categorias baixa, média e alta. Ademais, na tabela também são apresentados os valores de acurácia de cada modelo.

Tabela 5.6 Efeito estimado da variável na influência dos atributos de entrega na efetivação da compra pelo comércio eletrônico

|     | Variável                                      | Modelo 1  | Modelo 2  | Modelo 3  | Modelo 4  | Modelo 5  |
|-----|-----------------------------------------------|-----------|-----------|-----------|-----------|-----------|
| V01 | Compra produtos de beleza, vestuário ou afins | Baixa (+) | Média (-) | Baixa (-) | Constante | Baixa (-) |
| V02 | Compra produtos de eletrônicos                | Baixa (+) | Baixa (+) | Baixa (+) | Baixa (+) | Constante |
| V03 | Compra produtos de cultura e lazer            | Baixa (+) | Baixa (+) | Constante | Constante | Constante |
| V04 | Compra móveis e utensílios domésticos         | Baixa (-) | Média (+) | Baixa (+) | Baixa (+) | Constante |

|     | Variável                                                            | Modelo 1   | Modelo 2   | Modelo 3   | Modelo 4           | Modelo 5   |
|-----|---------------------------------------------------------------------|------------|------------|------------|--------------------|------------|
| V05 | Percentual do preço do produto que está disposto a pagar pelo frete | Baixa (-)  | Média (-)  | Baixa (+)  | Média (+)          | Baixa (+)  |
| V06 | Frequência de consumo                                               | Constante  | Constante  | Constante  | Constante          | Baixa (+)  |
| V07 | Faixa etária                                                        | Média (-)  | Média (-)  | Média (-)  | Complexo (côncavo) | Baixa (+)  |
| V08 | Renda familiar                                                      | Baixa (-)  | Média (-)  | Alta (+)   | Média (+)          | Baixa (-)  |
| V09 | Gênero                                                              | Constante  | Média (+)  | Baixa (+)  | Constante          | Baixa (-)  |
| V10 | Comodidade de comprar sem sair de casa                              | Baixa (-)  | Alta (+)   | Média (+)  | Baixa (-)          | Alta (+)   |
| V11 | Privacidade                                                         | Baixa (-)  | Baixa (-)  | Baixa (+)  | Alta (+)           | Baixa (-)  |
| V12 | Oportunidade de compra                                              | Alta (+)   | Alta (-)   | Baixa (-)  | Baixa (-)          | Baixa (-)  |
| V13 | Preço do produto                                                    | Baixa (-)  | Média (-)  | Baixa (+)  | Baixa (-)          | Alta (+)   |
|     | <b>Acurácia</b>                                                     | <b>88%</b> | <b>60%</b> | <b>82%</b> | <b>84%</b>         | <b>64%</b> |

Nota-se que a variável frequência de consumo (V06) apresentou relação majoritariamente constante com os parâmetros de entrega modelados. Em relação aos produtos consumidos (V01, V02, V03 e V04), seus efeitos isolados distinguem-se para cada modelo. Contudo, de maneira geral, observa-se que esses apresentaram efeito baixo ou constante nas análises dos atributos da entrega. O mesmo pode ser dito para a variável gênero (V09), com a qual as variáveis ‘resposta’ apresentaram baixo grau de dependência, ou relação constante (com exceção apenas do Modelo 2).

As demais variáveis não apresentaram padrão entre os efeitos estimados em cada modelo.

#### 5.4 Estimação dos Modelos por Regressão Logística

Os mesmos parâmetros utilizados no modelo de rede neural foram estimados por um modelo de regressão logística (RL) para apoiar na inferência das relações das variáveis explicativas nas variáveis dependentes, além de permitir comparar os resultados obtidos por ambas as técnicas. Em cada modelo, foram analisadas as variáveis que retornaram nível de confiança maior ou igual a 95% ( $p\text{-valor} \leq 5\%$ ).

Os resultados da estimação para os cinco modelos são apresentados na Tabela 5.7. No modelo 1 (‘Influência do valor do frete na efetivação da compra’), apenas duas variáveis apresentaram significância estatística: produtos eletrônicos (V02) e faixa etária (V07). Pela regressão, a relação com a variável faixa etária indica que quanto maior a idade, menor a influência do valor

do frete para a realização da compra pela Internet. Além disso, demonstrou-se que o valor do frete é determinante para efetivação da compra pelos consumidores de produtos eletrônicos, como eletrodomésticos, celulares, tablets, entre outros.

No que tange o modelo 02 (‘Influência do agendamento para efetivação da compra’), a variável privacidade (V11) apresentou significância estatística para a categoria de consumidores que concordam que se trata de um importante aspecto no momento de comprar pela Internet. Logo, pelos coeficientes da regressão, demonstrou-se que esses consumidores são influenciados pela possibilidade de agendamento da entrega ao efetivar sua compra.

Quanto ao modelo 03 (‘Importância do prazo de entrega na hora da compra’), as variáveis relativas ao gênero (V09) e à privacidade (V11) e ao preço do produto (V13) – consumidores que concordam com a importância desses parâmetros - demonstraram significância estatística. Os coeficientes da regressão apontam que quanto maior a importância atribuída aos parâmetros privacidade e preço do produto, maior a relevância do prazo de entrega para o consumidor do comércio eletrônico. Por fim, os resultados indicam maior probabilidade de que consumidores do gênero feminino atribuam importância ao prazo de entrega ao comprar pela Internet.

Em relação ao modelo 04 (‘Importância do valor do frete na hora da compra’), o fenômeno pode ser explicado pelas variáveis oportunidades de compras (V12) e preço do produto (V13). Quanto mais relevantes esses parâmetros são para o consumidor, maior a importância atribuída ao valor do frete no ato de comprar pela Internet.

Finalmente, no modelo 05 (‘Importância do horário de entrega na hora da compra’), são significativas as variáveis: produtos de cultura e lazer (V03), frequência de consumo (V06), renda familiar (V08) – categoria entre 2 a 4 salários mínimos -, gênero (V09), privacidade (V11) e preço do produto (V13). A importância dos parâmetros privacidade e preço do produto associam-se positivamente ao parâmetro de entrega. Consumidores do gênero feminino e com renda entre 2 a 4 salários mínimos se importam mais com o horário da entrega, ao contrário de consumidores de produtos de cultura e lazer (instrumentos musicais, entre outros), que consomem mais de uma vez por mês.

Tabela 5.7 Resultados da Regressão Logística

| Variáveis                          | Influência do Valor do Frete |               | Influência do Agendamento da Entrega |               | Importância do Prazo de Entrega |               | Importância do Valor do Frete |               | Importância do Horário da Entrega |               |
|------------------------------------|------------------------------|---------------|--------------------------------------|---------------|---------------------------------|---------------|-------------------------------|---------------|-----------------------------------|---------------|
|                                    | Coeficiente                  | p-valor       | Coeficiente                          | p-valor       | Coeficiente                     | p-valor       | Coeficiente                   | p-valor       | Coeficiente                       | p-valor       |
| Intercepto                         | 18,7198                      | 0,9863        | 2,0550                               | 0,1396        | <b>-3,2087</b>                  | <b>0,0510</b> | -2,2378                       | 0,4193        | -2,0017                           | 0,0668        |
| V01 (Sim)                          | -0,0687                      | 0,8938        | -0,2233                              | 0,3396        | -0,2725                         | 0,4859        | -0,2411                       | 0,6565        | 0,1083                            | 0,6507        |
| V02 (Sim)                          | <b>1,9224</b>                | <b>0,0001</b> | -0,1277                              | 0,6318        | -0,7323                         | 0,1664        | -0,4182                       | 0,5000        | -0,3131                           | 0,2891        |
| V03 (Sim)                          | 0,0474                       | 0,9245        | -0,0335                              | 0,8781        | 0,2620                          | 0,4775        | <b>1,0320</b>                 | <b>0,0501</b> | <b>-0,4798</b>                    | <b>0,0424</b> |
| V04 (Sim)                          | 0,3936                       | 0,7199        | 0,2338                               | 0,6454        | -0,1805                         | 0,8230        | 1,9900                        | 0,3646        | -0,3505                           | 0,4718        |
| V05 (Até 7%)                       | -0,0934                      | 0,8877        | 0,2408                               | 0,3846        | -0,6917                         | 0,1928        | 0,2702                        | 0,7114        | -0,2927                           | 0,3309        |
| V05 (Até 10%)                      | -0,4760                      | 0,4506        | 0,4283                               | 0,1284        | -0,7000                         | 0,1813        | -0,2414                       | 0,7007        | -0,1459                           | 0,6306        |
| V05 (Acima de 10%)                 | -0,2863                      | 0,7227        | 0,6082                               | 0,0757        | -1,0693                         | 0,0604        | -0,8853                       | 0,1977        | 0,0723                            | 0,8384        |
| V06 (Acima de uma entrega por mês) | 0,4102                       | 0,4717        | -0,2694                              | 0,2448        | -0,1752                         | 0,6604        | 0,3555                        | 0,5158        | <b>-0,4842</b>                    | <b>0,0458</b> |
| V07 (25 a 34 anos)                 | -1,1414                      | 0,3086        | -0,1321                              | 0,6409        | -0,4825                         | 0,3455        | -0,6272                       | 0,3633        | -0,1780                           | 0,5558        |
| V07 (35 a 49 anos)                 | <b>-2,8655</b>               | <b>0,0104</b> | -0,2595                              | 0,4416        | -0,0889                         | 0,8859        | 0,3623                        | 0,6918        | 0,6453                            | 0,0953        |
| V07 (Mais de 50 anos)              | <b>-3,2076</b>               | <b>0,0059</b> | 0,7268                               | 0,1104        | -0,6437                         | 0,3156        | -0,5747                       | 0,5137        | -0,0984                           | 0,8113        |
| V08 (2 A 4 SM)                     | 1,2279                       | 0,4537        | -0,1108                              | 0,9051        | 0,6050                          | 0,6313        | -1,5808                       | 0,5731        | <b>0,4825</b>                     | <b>0,0000</b> |
| V08 (4 a 10 SM)                    | 1,2743                       | 0,4471        | -0,3621                              | 0,6952        | 0,8744                          | 0,4893        | -0,6873                       | 0,8028        | 0,7612                            | 0,3919        |
| V08 (Mais de 10 SM)                | 0,8582                       | 0,5932        | 0,3567                               | 0,6977        | 0,6368                          | 0,6032        | -1,5034                       | 0,5850        | 0,8767                            | 0,3185        |
| V09 (feminino)                     | 0,0420                       | 0,9324        | 0,0475                               | 0,8305        | <b>0,9201</b>                   | <b>0,0173</b> | 0,8949                        | 0,0895        | <b>0,5412</b>                     | <b>0,0213</b> |
| V10 (neutro)                       | 1,3931                       | 0,3618        | 0,5964                               | 0,4002        | 0,1903                          | 0,8380        | 0,2740                        | 0,8628        | 0,1044                            | 0,8891        |
| V10 (concordo)                     | 0,5469                       | 0,6077        | 0,4041                               | 0,4879        | 0,5222                          | 0,4605        | -0,2429                       | 0,8440        | -0,3334                           | 0,5709        |
| V11 (neutro)                       | 0,5213                       | 0,4971        | 0,1030                               | 0,7446        | 0,5479                          | 0,2388        | -0,5539                       | 0,3816        | <b>1,1942</b>                     | <b>0,0004</b> |
| V11 (concordo)                     | 0,3310                       | 0,5986        | <b>0,6683</b>                        | <b>0,0222</b> | <b>1,4177</b>                   | <b>0,0018</b> | 1,2045                        | 0,0852        | <b>1,2781</b>                     | <b>0,0000</b> |
| V12 (neutro)                       | -2,3970                      | 0,0743        | -0,3968                              | 0,5834        | 1,9654                          | 0,1238        | 18,1065                       | 0,9913        | -0,2782                           | 0,6771        |
| V12 (concordo)                     | -1,0588                      | 0,3749        | -0,2369                              | 0,6959        | 0,6467                          | 0,3693        | <b>1,6496</b>                 | <b>0,0349</b> | 0,2016                            | 0,7125        |
| V13 (neutro)                       | -15,9725                     | 0,9883        | -1,7201                              | 0,2344        | 19,5091                         | 0,9856        | 21,3156                       | 0,9940        | <b>2,5985</b>                     | <b>0,0355</b> |
| V13 (concordo)                     | -15,6161                     | 0,9885        | -1,5395                              | 0,2375        | <b>4,3660</b>                   | <b>0,0008</b> | <b>5,1268</b>                 | <b>0,0006</b> | <b>2,0856</b>                     | <b>0,0252</b> |
| <b>Acurácia</b>                    | <b>93%</b>                   |               | <b>73%</b>                           |               | <b>91%</b>                      |               | <b>94%</b>                    |               | <b>73%</b>                        |               |

### 5.5 Comparação dos modelos por RNA e por RL

A fim de consolidar o entendimento dos fenômenos estudados, nesta seção são comparados os resultados dos modelos, estimados pelas técnicas RNA e RL, bem como sua eficácia (capacidade de classificar corretamente as respostas). A regressão logística agrega robustez à análise ao trazer abordagem estatística para a interpretação dos resultados. Por sua vez, as redes neurais propõem solução de problemas classificatórios a partir da detecção de padrões entre dados de entrada, por uma abordagem de treinamento e aprendizagem, utilizando-se da própria amostragem de dados.

Os resultados da RL para os dois primeiros modelos (‘Influência do valor do frete na efetivação da compra’ e ‘Influência da escolha do dia e horário da entrega na efetivação da compra’) permitiriam análise limitada dos fenômenos estudados. O primeiro modelo apresentou apenas duas variáveis significativas, entre elas a faixa etária dos consumidores, que demonstrou relação decrescente com o fenômeno analisado por ambas as abordagens. No segundo modelo, por sua vez, apenas a variável apenas as respostas positivas à importância da privacidade apresentaram significância.

Na Tabela 5.8 são mostradas as análises de eficácia dos modelos a partir do cálculo da métrica da acurácia. De maneira geral, a acurácia obtida para as estimações por RL foram superiores, com exceção ao modelo 4, relativo à importância do valor do frete.

Tabela 5.8 Comparação da entre os modelos classificatórios

| Codificação | Modelos                                              | Acurácia    |                     |
|-------------|------------------------------------------------------|-------------|---------------------|
|             |                                                      | Rede Neural | Regressão Logística |
| Modelo 1    | Influência do Valor do Frete na efetivação da compra | 88%         | 94%                 |
| Modelo 2    | Influência do Agendamento na efetivação da compra    | 60%         | 73%                 |
| Modelo 3    | Importância do Prazo da Entrega na hora da compra    | 82%         | 93%                 |
| Modelo 4    | Importância do Valor do Frete na hora da compra      | 84%         | 73%                 |
| Modelo 5    | Importância do Horário da Entrega na hora da compra  | 64%         | 91%                 |

Os resultados do segundo modelo, pela RNA, demonstraram classificação correta de 60% das respostas, um tanto aquém quando comparado aos demais modelos. Por sua vez, a RL conseguiu convergir uma solução de maior acurácia para o modelo a partir do mesmo universo amostral (com 73% de eficácia na classificação das respostas).

O terceiro modelo, por sua vez, apresentou interpretações semelhantes para o fenômeno estudado em ambas as técnicas abordadas. Contudo, as variáveis com as quais o modelo da RNA demonstrou maior dependência, sendo elas faixa etária (V07), renda familiar (V08) e comodidade (V10), não apresentaram significância estatística na estimação pela RL. Infere-se, portanto, que a análise da influência do prazo de entrega deve ser explorada sob outras perspectivas, como a partir das distâncias do local de origem do produto e da residência do consumidor, que não foram incluídas neste estudo.

Em relação ao modelo 4, os resultados das diferentes técnicas apresentaram interpretações distintas para o comportamento do consumidor do comércio eletrônico. As variáveis significativas apresentam coeficientes ( $\beta_i$ ) positivos pela RL, enquanto seu efeito marginal é representado por funções decrescentes, com baixas grau de dependência, pela RNA. Ademais, a privacidade, com a qual a variável ‘resposta’ demonstrou alta dependência na análise dos PDP, não apresentou significância estatística na estimação por RL. Por tratar-se do único modelo no qual a RNA demonstrou melhor performance de classificação do que a RL, entende-se que a complexidade da relação entre as variáveis explicativas, representada pelos PDP, é melhor aproximada pelo algoritmo de *machine learning*. De toda forma, os resultados divergentes permitem inferir que existem outros fatores que influenciam no valor do frete.

Finalmente, para o quinto modelo, ambas as técnicas apresentaram efeito crescente para a variável preço do produto. Contudo, para o restante das variáveis com significância, a estimação por RNA apresentou relação constante ou com pouca sensibilidade a variações de categoria. A maior diferença entre as estimações, porém, diz respeito à comodidade (V10), com a qual a variável ‘resposta’ demonstrou forte relação de dependência pela RNA, mas não retornou significância estatística pela RL.

## 5.6 Discussão dos resultados

Endereçar os problemas gerados pelo aumento de entregas urbanas, oriundas do comércio eletrônico, requer captar a percepção do consumidor por esse serviço. As análises conduzidas nesta pesquisa aproximam o esclarecimento das relações entre os parâmetros do serviço de entrega e sua influência no comportamento do consumidor para efetivação da compra. Nesse sentido, associou-se características sociodemográficas, hábitos de compra e características de consumo pelo comércio eletrônico aos parâmetros do serviço de entrega.

As variáveis sociodemográficas analisadas nos modelos foram: faixa etária, renda familiar e gênero. De maneira geral, ao menos uma delas apresentou significância estatística para análise da influência do atributo valor do frete, e da importância dos atributos prazo e horário de entrega.

Os *Partial Dependence Plots* dos modelos estimados por RNA complementam a análise da relação causa x efeito das características sociodemográficas com a influência dos parâmetros de entrega. Compreende-se que consumidores de menores faixas etárias são mais influenciados pelos parâmetros de entrega, ponderando-os no momento de efetivar a compra. Esta relação é mais evidente ao tratar-se do parâmetro valor do frete. Em outras palavras, consumidores mais jovens são mais susceptíveis a rejeitar a compra se não concordarem com o valor cobrado pela entrega.

Os fenômenos investigados apontaram baixa sensibilidade ao gênero dos consumidores do comércio eletrônico. Contrariamente, as análises dos atributos de entrega demonstram relação de dependência com renda familiar do consumidor. No entanto, a variável não possui significância estatística na maioria dos modelos propostos, o que pode indicar a necessidade de incorporar outros aspectos socioeconômicos, demográficos e mesmo psicológicos às análises (a exemplo dos trabalhos de Xiao *et al.*, 2018; Yuen *et al.*, 2018; e Liu *et al.*, 2019 que analisaram nível de escolaridade, estado civil, estilo de vida, entre outros). Em suma, os PDP apontam que consumidores de maior renda são influenciados, sobretudo, pelo parâmetro prazo da entrega, independentemente do valor que estariam dispostos a pagar no frete.

Quanto aos padrões e hábitos de consumo pelo comércio eletrônico, foi analisada a aquisição por produtos do tipo beleza e vestuário, produtos eletrônicos, produtos de cultura e lazer, além de móveis e eletrodomésticos. Também foram investigados a frequência de realização de compras pela Internet e o percentual do valor do produto que se pagaria pelo frete. Entre essas variáveis, demonstraram significância estatística apenas a compra por produtos eletrônicos e de cultura e lazer, além da frequência de consumo.

A análise dos efeitos marginais demonstrou a baixa dependência dos atributos de entrega em relação aos tipos de produto adquiridos pelos consumidores. De toda maneira, alguns padrões puderam ser identificados. Consumidores de produtos eletrônicos e de cultura e lazer (como

jogos, livros, instrumentos musicais e afins) demonstraram-se sensíveis, sobretudo, ao atributo valor do frete.

Notavelmente, a variável frequência de consumo apresentou relação quase constante com os atributos de entrega. Assim, os resultados apontam que a influência exercida pela entrega na efetivação da compra independe da quantidade de vezes que o consumidor compra por mês. Ademais, quanto maior o valor percebido pela entrega (refletido no percentual do preço do produto que o consumidor está disposto a pagar pelo frete), maior a importância atribuída aos parâmetros desse serviço. Como reflexo disso, menos o consumidor pesará o valor do frete ou a possibilidade de agendamento no momento de efetivar a compra.

Finalmente, as características do consumo pelo comércio eletrônico investigadas foram: a comodidade de comprar sem sair de casa, a privacidade, as oportunidades e ofertas de compra e o preço do produto. Para análise da influência dos parâmetros de entrega, as três últimas características demonstraram significância estatística.

Pela análise dos efeitos marginais, consumidores que valorizam ter privacidade ao comprar pela Internet atribuem mais importância ao valor do frete e ponderam menos o agendamento e horário da entrega. Quanto à importância dada às ofertas de compra, essa impacta, majoritariamente, na redução da influência dos parâmetros de entrega, sobretudo da influência da possibilidade de agendamento. Por fim, consumidores que valorizam o preço do produto se preocupam, principalmente, com o horário da entrega. Esses também demonstram se preocupar menos com o valor do frete. Compreende-se, assim, que quanto mais o preço do produto se traduzir como um benefício para o consumidor, mais tolerante esse será a valores mais caros de frete.

### **5.6.1 Implicação prática da RNA**

Os modelos desenvolvidos em rede neural apresentaram boa performance de classificação da influência que os atributos de entrega exercem na efetivação da compra pelo comércio eletrônico. A principal vantagem observada nos modelos estimados por RNA é a sua grande capacidade de detectar relações não-lineares entre as variáveis, o que permitiu trabalhar com modelos de maior complexidade e multidimensionalidade de dados.

Outra vantagem associada a modelos de aprendizagem consiste em não requererem suposições relativas à estrutura dos dados. Contudo, cabe ressaltar que, para essa dissertação, optou-se por

assumir a categoria de respostas neutras como respostas positivas para as variáveis dependentes dos modelos. Essa abordagem demonstrou melhor performance de classificação pela RNA.

Finalmente, a aplicação dos *Partial Dependence Plots* – PDP permitiu a análise das relações causa x efeito entre as variáveis independentes e a variável dependente dos modelos estimados por RNA. Eliminou-se, portanto, o principal receio na aplicação de modelos de aprendizagem de máquinas, também chamados *black-box*: a visualização e interpretação das relações estabelecidas entre os dados nas camadas intermediárias dos algoritmos. Pelos PDP, foram aferidos não apenas os efeitos, mas também o grau de dependência que os resultados dos modelos têm das variáveis explicativas. Esse fato permitiu aprofundar a análise das relações entre as características do consumidor (sociodemográficas ou hábitos de consumo), além de características do próprio comércio eletrônico (preço do produto, ofertas de compra, comodidade, entre outros), com a influência dos parâmetros de entrega na efetivação da compra.

Entende-se, portanto, que a aplicação das redes neurais foi relevante às análises propostas, pois trouxe uma abordagem com foco na representação precisa das propriedades subjacentes dos dados, considerada como mais realista por Farmer (2014) e Zhao *et al.* (2019).

## 6 CONSIDERAÇÕES FINAIS

Este trabalho foi desenvolvido com o propósito de analisar a influência de parâmetros do serviço de entrega no comportamento de compra do consumidor do comércio eletrônico. Dado o problema científico e as oportunidades de pesquisa identificadas na literatura, três hipóteses de pesquisa foram formuladas. Estas assumiram que a influência dos atributos de entrega na efetivação da compra (1) varia conforme as características sociodemográficas do consumidor do comércio eletrônico, (2) está associada aos hábitos de compra do consumidor do comércio eletrônico e (3) associada às características do consumo pelo comércio eletrônico, como comodidade, privacidade, preço do produto e ofertas de compra.

Ademais, com vista a adotar abordagens mais realistas para análise do comportamento da demanda do comércio eletrônico, tomou-se por base estimar os modelos por meio da técnica de *machine learning* denominada Rede Neural Artificial. Assim, os objetivos específicos desta dissertação consistiram em propor modelo comportamental por meio de RNA e analisar os efeitos das características do consumidor (perfil socioeconômico e hábitos de compra) e do consumo pelo comércio eletrônico na influência dos atributos de entrega para efetivação da compra pela Internet.

A fim de testar as hipóteses colocadas nesta dissertação, foi proposta coleta de dados por pesquisa baseada na Internet, com aplicação de questionários disponibilizados na plataforma da Google (formulário Google). Os instrumentos de coleta consistiram no levantamento de dados de socioeconômicos, informação dos hábitos de compra do consumidor, além da determinação do nível de importância e valor percebido dos atributos do comércio eletrônico e da entrega, utilizando escala Likert. Ao final do período de aplicação dos questionários, obteve-se total satisfatório de 615 respostas, das quais 595 foram utilizadas para modelagem do comportamento do consumidor (a confiabilidade das respostas foi validada por alfa de Cronbach equivalente a 0,79).

Em seguida, pelo tratamento dos dados coletados, foram formulados 5 modelos que investigaram (1) a influência do valor do frete e (2) da possibilidade de agendar a entrega na efetivação da compra pela Internet, além da (3) a importância do prazo de entrega (4) do valor do frete e (5) do horário de entrega na realização da compra pelo comércio eletrônico. Os modelos foram estimados por rede neural com classificação binária das respostas (0 – discordo ou 1 – concordo) para melhor performance do algoritmo. Nesse intuito, as respostas definidas

como neutras, pela escala Likert, foram reclassificadas como positivas, o que resultou em melhor acurácia dos modelos. Tal fato indica que aqueles que se posicionam como ‘neutros’ podem ser susceptíveis à influência dos parâmetros de entrega ao comprar pela Internet. Na sequência, os *Partial Dependence Plot* foram aplicados para análise da relação causa x efeito entre as variáveis explicativas e variável dependente de cada modelo. Adicionalmente, os modelos mencionados foram também estimados por regressão logística, com finalidade de conferir maior robustez às análises a partir da interpretação de resultados estatísticos.

Assim, os modelos testaram as três hipóteses colocadas. Em relação à primeira hipótese, os resultados indicam que consumidores de menor faixa etária são mais influenciados pelos parâmetros de entrega, principalmente pelo valor do frete, ao efetivar a compra. Ademais, os testes para a segunda hipótese estabelecem que consumidores de produtos eletrônicos e de cultura e lazer (como jogos, livros, instrumentos musicais e afins) são influenciados, sobretudo, pelo parâmetro valor do frete. Por sua vez, a influência exercida pela entrega independe da frequência de consumo. Complementarmente, a disposição em pagar pelo frete demonstra valorização dos parâmetros de entrega pelo consumidor. Como consequência, o consumidor pesa menos o valor do frete e a possibilidade de agendamento no momento de efetivar a compra. Quanto às análises para a terceira hipótese, essas indicam que consumidores que valorizam a privacidade ao comprar pela Internet, se preocupam menos com prazo, agendamento e horário da entrega, porém atribuem maior importância ao valor do frete. Contrariamente, consumidores que valorizam o preço do produto, se preocupam mais com o horário da entrega e são mais tolerantes a valores mais caros de frete. Por fim, quanto maior a importância das ofertas de compra para o consumidor do comércio eletrônico, menor será a influência dos parâmetros de entrega na efetivação da compra.

Este estudo, contudo, apresenta algumas limitações que poderiam ser exploradas em trabalhos futuros. A saber:

- Os modelos estimados contêm fragilidades na análise de aspectos como a renda familiar e frequência de consumo, que demonstraram pouco significância estatística pelas regressões logísticas. Por tratar-se de características associadas ao perfil do consumidor, entende-se que a associação de outras variáveis sociodemográficas, como nível de escolaridade, além de variáveis demográficas, pode melhorar a performance dos modelos e facilitar sua interpretação. Por exemplo, Xiao *et al.* (2018) identificaram que

a frequência de consumo é positivamente influenciada pelo gênero feminino, pelo estado civil e pelo nível educacional dos consumidores, enquanto Irawan e Wirza (2015) descobriram que as características demográficas e sociodemográficas influenciam a frequência das compras pelo comércio eletrônico.

- A influência exercida pelo valor do frete e pelo agendamento da entrega na efetivação da compra foi pouco aprofundada pelos modelos propostos. Isto porque, apesar da análise dos efeitos marginais causados pelas variáveis independentes (relação causa x efeito) proposto pelos PDP, poucas variáveis foram significantes nas estimações por regressões logísticas. Em trabalhos futuros, propõe-se explorar esses fenômenos a partir da inserção de variáveis explicativas que investiguem como diferentes combinações entre parâmetros de entrega influenciam o valor percebido deste serviço pelos consumidores, segundo proposto em seção do questionário de pesquisa desenvolvido para este trabalho.
- A acurácia dos modelos de influência do agendamento e de importância do horário de entrega, estimados por rede neural, ficou aquém em comparação à dos demais. Não coincidentemente, esses atributos são os que mais dividem opiniões entre os consumidores do comércio eletrônico, uma vez que ao menos 50% da amostra discorda ou é neutra em relação à sua influência/importância. A inserção de novos dados às análises poderia auxiliar à detecção de padrões de comportamento do consumidor em relação a esses parâmetros. Além da melhora na performance da rede neural, buscar-se-ia aprofundar a compreensão da percepção do agendamento e do horário de entrega pelo consumidor.
- As estimativas por RL e RNA foram, por vezes, divergentes. De modo geral, as diferenças encontradas nos resultados sugerem a necessidade de inserir novos fatores às análises, para melhoria tanto da performance dos modelos comportamentais quanto da interpretação das causas dos fenômenos.

Ademais, as análises conduzidas permitiram modelar o comportamento do consumidor do comércio eletrônico. Como demonstrado pela associação com as variáveis sociodemográficas, esse varia segundo contexto geográfico e social. Contudo, cabe ressaltar o impacto do contexto histórico da pandemia do Covid-19 (decretada oficialmente pela OMS em março de 2020) nas compras pelo comércio eletrônico, e, conseqüentemente, no comportamento do consumidor. Uma vez que a coleta de dados se deu antes da pandemia, em estudo futuro poderia propor-se

a reaplicação do questionário e metodologia de pesquisa para comparação da percepção dos consumidores pelos atributos do serviço de entrega.

Em suma, os resultados deste estudo têm relevantes implicações para consumidores do comércio eletrônico, que se beneficiam ao terem suas percepções e preferências compreendidas; para empresas que comercializam seus produtos pela Internet, que buscam se adaptar segundo preferências do consumidor; para gestores urbanos, que conciliam a ascensão dos fluxos de entrega domiciliar e a preservação da mobilidade urbana, através da proposição de políticas públicas; e para transportadores logísticos, que se beneficiam da solução de problemas que custem a mais à entrega. Por fim, as análises desenvolvidas nesta dissertação reafirmam a modelagem comportamental como importante ferramenta de auxílio à tomada de decisão no planejamento urbano e logístico.

## 7 REFERÊNCIAS

- Abduljabbar, R.; Dia, H.; Liyanage, S.; Bagloee, S.A. (2019). Applications of Artificial Intelligence in Transport: An Overview. *Sustainability*, 11(1), 189. DOI: 10.3390/su11010189
- Agatz, N.; Campbell, A.; Fleischmann, M.; Savelsbergh, M. (2011). Time slot management in attended home delivery. *Transportation Science*, 45(3), 435-449
- Akgüngör, A.P.; Dogan, E. (2009). An artificial intelligent approach to traffic accident estimation: Model development and application. *Transport*, 24, 135-142. DOI: 10.3846/1648-4142.2009.24.135-142
- Almeida, D.; Santos, M.; Costa, A. (2010). Aplicação do coeficiente alfa um questionário para avaliação de desempenho da saúde pública. *Encontro Nacional de Engenharia de Produção – ENEGEP*; São Carlos, SP, Brasil.
- Aschwanden, G. D. P. A.; Wijnands, J. S.; Thompson, J.; Nice, K. A.; Zhao, H.; Stevenson, M. (2019). Learning to walk: Modeling transportation mode choice distribution through neural networks. *Urban Analytics and City Science*, 1-14. DOI: 10.1177/2399808319862571
- Asheghi, R.; Hosseini, S. A.; Mojtaba, S.; Shahri, A. A. (2020) Updating the neural network sediment load models using different sensitivity analysis methods: a regional application. *Journal of Hydroinformatics*, 562-577. DOI: 10.2166/hydro.2020.098
- Bandara, K.; Shi, P.; Bergmeir, C.; Hewamalage, H.; Tran, Q.; Seaman, B. (2019). *Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology*. Faculty of Information Technology Monash University, Melbourne, Australia.
- Basheer, I. A.; Hajmeer, M. (2000). Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *Journal of Microbiological Methods*, 43, 3-31. DOI: S0167-7012(00)00201-3.
- Borgonovo, E.; Plischke, E. (2016). Sensitivity analysis: a review of recent advances. *European Journal of Operational Research*, 248(3), 869-887. DOI: 10.1016/j.ejor.2015.06.032
- Breiman, L. (2001). Statistical modeling: the two cultures. *Statistical Science*, 16, 199-231. DOI: 10.1214/ss/1009213726
- Buldeo Rai, H.; Verlinde, S.; Macharis, C. (2021a). Who is interested in a crowdsourced last mile? A segmentation of attitudinal profiles. *Travel Behaviour and Society*, 22, 22-31. DOI: 10.1016/j.tbs.2020.08.004
- Cádernas, I.; Beckers, J.; Vanelslander, T. (2017). E-commerce last-mile in Belgium: Developing an external cost delivery index. *Research in Transportation Business & Management*, 24, 123-129. DOI: 10.1016/j.rtbm.2017.07.006
- Cao, X; Chen, Q.; Choo, S. (2013). Geographical distribution of e-shopping: An application of structural equations models in the twin cities. *Proceedings of the 92nd Transportation*

- Research Board annual meeting*. Washington, DC, USA: Transportation Research Board of the National Academies. DOI: 10.3141/2383-03
- Chen, W.; Cai, F.; Chen, H.; de Rijke, M. (2019). A Dynamic Co-attention Network for Session-based Recommendation. In: *28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, Beijing, China. DOI: 10.1145/3357384.3357964
- Cherrett, T.; Dickinson, J.; McLeod, F.; Sit, J.; Bailey, G.; Whittle, G. (2017). Logistics impacts of student online shopping – evaluating delivery consolidation to halls of residence. *Transportation Research part C: Emerging Technologies*, 78, 111-118. DOI: 10.1016/j.trc.2017.02.021
- Chincholkar, S.; Sonwaney, V. (2017). Website Attributes and its Impact on Online Consumer Buying Behaviour: An Empirical Study of Online Consumers in Mumbai Region. *Indian Journal of Science and Technology*, 10(47), 1-9. DOI: 10.17485/ijst/2017/v10i47/119973
- Choo, C. (2016). *Impact of a Delivery Point Network for Urban E-Commerce Deliveries*. Singapore University of Technology and Design, Singapore. p. 57
- Comi, A.; Nuzzolo, A. (2016). Exploring the relationships between e-shopping attitudes and urban freight transport. *Transportation Research Procedia*, 399-412. *The 9th International Conference on City Logistics*, Tenerife, Canary Islands (Spain), 17-19 June 2015.
- Comi, A.; Site, P.D.; Filippi, F.; Nuzzolo, A. (2012). Urban Freight Transport Demand Modelling: a State of the Art. *European Transport \ Trasporti Europei*, 51, Paper N° 7.
- Cortez, P.; Embrechts, M.J. (2012). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Science*, 225, 1-17. DOI: 10.1016/j.ins.2012.10.039
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104. DOI: 10.1037/0021-9010.78.1.98
- Cramer, J. S. (2003). *The origins and development of the logit model*. Cambridge UP.
- Craven, M.W.; Shavlik, J.W. (1992). Visualizing learning and computation in artificial neural networks. *International Journal on Artificial Intelligence*, 1(3), 399-425. DOI: 10.1142/S0218213092000260
- CREDOC (2010). Le profil des acheteurs à distance et en ligne, FEVAD, La Poste, Reed CCI.
- Cronbach, L. J.; Gleser, G. C.; Nanda, H.; Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory for generalizability of scores and profiles*. New York: John Wiley.
- Cunha, R.N.; Galo, N.R.; Vieira, J.G.V.; Moris, V.A.S. (2013). Logística Urbana e o Desafio das Megacidades. In *XXXIII Encontro Nacional de Engenharia de Produção – ENEGEP* (Brasil).

- Dablanc, L., Rakotonarivo, D. (2010). The impacts of logistics sprawl: How does the location of parcel transport terminals affect the energy efficiency of goods' movements in Paris and what can we do about it? *Procedia Social and Behavioral Sciences*, 2, 6087-6096. DOI: 10.1016/j.sbspro.2010.04.021
- Dogan, E.; Akgüngör, A.P. (2013). Forecasting highway casualties under the effect of railway development policy in Turkey using artificial neural networks. *Neural Computing and Applications*, 22, 869–877. DOI: 10.1007/s00521-011-0778-0
- Duives, D. C.; Wang, G.; Kim, J. (2019). Forecasting Pedestrian Movements Using Recurrent Neural Networks: An Application of Crowd Monitoring Data. *Sensors*, 19, 382. DOI: 10.3390/s19020382
- Dutra N. G. S.; Novaes, A. G. N.; Moreira M. E. P.; (2006) Novos Conceitos e Tecnologias na Distribuição Urbana de Mercadorias. In: *XXVI ENEGEP*, Fortaleza (Brasil).
- Eurobarometer (2013). *Consumer attitude towards cross-border trade and consumer protection*. Flash Eurobarometer 358. Disponível em: [https://ec.europa.eu/commfrontoffice/publicopinion/flash/fl\\_358\\_en.pdf](https://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_358_en.pdf). Acessado: 10/07/2020
- Farmer, J. D. (2014). Hypotheses non fingo: Problems with the scientific method in economics, *Journal of Economic Methodology*, 20(4), 377-385. DOI: 10.1080/1350178X.2013.859408
- Ferreira, J.; Callou, G.; Josua, A.; Tutsch, D.; Maciel, P. (2019). An Artificial Neural Network Approach to Forecast the Environmental Impact of Data Centers. *Information*, 10, 113. DOI: 10.3390/info10030113
- Fishwick, P. A. (1989). Neural Network Models in Simulation: A Comparison with Traditional Modeling Approaches. In: *Proceedings of the 1989 Winter Simulation Conference*. Gainesville, Florida, USA.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29 (5) 1189 – 1232. DOI: 10.1214/aos/1013203451
- Frisch R. (1933). Editor's Note, *Econometrica*, 63(4), 755-765. DOI: 10.2307/2171799
- Garg, A.; Tai, K. (2013). Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *International Journal of Modelling, Identification and Control*. v. 18(4) p. 295 – 312. DOI: 10.1504/IJMIC.2013.053535
- Ghavamipoor, H.; Golpayegani, S. A. H.; Shahpasand, M. (2017). A QoS-sensitive model for e-commerce customer behavior. *Journal of Research in Interactive Marketing*, 11(4)2040-7122. DOI: 10.1108/JRIM-08-2016-0080
- Gilmore, N.; Abe, J.F. (1995). Neural network models for traffic control and congestion prediction. *Journal Intell. Transp. Syst.*, 2, 231-252. DOI: 10.1080/10248079508903828

- Gul, F.; Natenzon, P.; Pesendorfer, W. (2014). Random choice as behavioral optimization. *Econometrica*, 82(5), 1873-1912. DOI: 10.3982/ECTA10621
- Hahnloser, R. H.R.; Sarpeshkar, R.; Mahowald, M. A; Douglas R. J; Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789), 947. DOI: 10.1038/35016072
- Hand, D.J. (2000). Data mining, new challenges for statisticians. *Social Science Computer Review*, 18(4), 442-449. DOI: 10.1177/089443930001800407
- Hanson, S.J. (1995). Backpropagation: some comments and variations. In: Rumelhart, D. E.,Yves, C.(Eds.), *Backpropagation: Theory, Architecture, and Applications*. Lawrence Erlbaum, 237-271.
- Hashem, S. (2003). Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. *International Joint Conference on Neural Networks*, 419-424. DOI: 10.1109/ijcnn.1992.287175
- Hayes, B. E. (1998). *Measuring Customer Satisfaction: Survey design, use, and statistical analysis methods*. Milwaukee, Wisconsin: ASQC Quality Press.
- Hey, J.; Orme, C.D. (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica*, 62(6), 1291-1326. DOI:10.2307/2951750
- Hosmer, D. W; Lemeshow, S. (2000). *Applied Logistic Regression*. Local: Wiley.
- Hsiao, M. H. (2008). Shopping mode choice: Physical store shopping versus e-shopping. *Transportation Research Part E*, 45, 86-95. DOI: 10.1016/j.tre.2008.06.002
- Irawan, M.Z. e Wirza, E. (2015). Understanding the Effect of Online Shopping Behavior on Shopping Travel Demand through Structural Equation Modeling. *Journal of the Eastern Asia Society for Transportation Studies*, 11, 614-625. DOI: 10.11175/easts.11.614
- IBGE, (2020). Informação. Brasília: Instituto Brasileiro de Geografia e Estatística. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html>Acessado em: 07/2021
- Jain, A.K.; Mao J., Mohiuddin, K.M. (1996). Artificial neural networks: a tutorial. *Computer*, 29, 31-44. DOI: 10.1109/2.485891
- Jeng, J.M; Fesenmaier, D.R. (2010). A Neural Network Approach to Discrete Choice Modeling. *Journal of Travel & Tourism Marketing*, 5(1-2), 119-144. DOI: 10.1300/J073v05n01\_08
- Jiang, H., Zou, Y., Zhang, S., Tang, J., Wang, Y. (2016). Short-term speed prediction using remote microwave sensor data: machine learning versus statistical model. *Mathematical Problems in Engineering*, 1-13. DOI: 10.1155/2016/9236156
- Jin, S.; Li, H.; Li, Y. (2016). Preferences of Chinese consumers for the attributes of fresh produce portfolios in an e-commerce environment. *British Food Journal*, 119(4), 817-829. DOI: 10.1108/BFJ-09-2016-0424

- Jin, S.; Liu, H.; Ding, W.; Shang, H.; Wang, G. (2018). Sensitivity analysis for the inverted siphon in a long distance water transfer project: an integrated system modeling perspective. *Water*, 10(3), p292. DOI: 10.3390/w10030292.
- Karlaftis, M., Vlahogianni, E. (2011). Statistical methods versus neural networks in transportation research: differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), 387-399. DOI:10.1016/j.trc.2010.10.004
- Laroche, M. (2009). New developments in modeling Internet consumer behavior: Introduction to the special issue. *Journal of Business Research*, 63(9-10), 915-918. DOI: 10.1016/j.jbusres.2008.12.013
- Leontisis, A.; Pagge, J. A. (2007). Simulation approach on Cronbach's alpha statistical significance. *Mathematics and Computers in Simulation*, 73, 336-340. DOI: 10.1016/j.matcom.2006.08.001
- Li, M.; Ji, S.; Liu, G. (2018). Forecasting of Chinese E-Commerce Sales: An Empirical Comparison of ARIMA, Nonlinear Autoregressive Neural Network, and a Combined ARIMA-NARNN Model. *Mathematical Problems in Engineering*, 6924960. DOI: 10.1155/2018/6924960
- Lim, H.; Dubinsky, A.J. (2004). Consumers' perceptions of e-shopping characteristics: an expectancy-value approach. *Journal of Services Marketing*, 18(7), 500-513. DOI: 10.1108/08876040410561839
- Lim, H., Widdows, R.; Hooker, N.H. (2009). Web content analysis of e-grocery retailers: a longitudinal study, *International Journal of Retail & Distribution Management*, 37(10), 839-851. DOI: 10.1108/09590550910988020
- Lin, Y.; He, D.; Wang, Y.; Lee, L. (2020). Last-mile delivery: optimal locker location under multinomial logit choice model. *Transportation Research Part E: Logistics and Transportation Review*, 142(C). DOI: 10.1016/j.tre.2020.102059
- Liu, C.; Wang, Q.; Susilo, Y. (2017). Assessing the impacts of collection-delivery points to individual's activity-travel patterns: A greener last mile alternative? *Transportation Research Part E: Logistics and Transportation Review*, 121, 84-99. DOI: 10.1016/j.tre.2017.08.007
- Liu, H., Huang, Y., Wang, Z., Liu, K., Hu, X., & Wang, W. (2019). Personality or Value: A Comparative Study of Psychographic Segmentation Based on an Online Review. *Enhanced Recommender System. Applied Sciences*, 9(10), 1992. DOI: 10.3390/app9101992
- Louviere, J. J.; Hensher, D. A. (1983). Using discrete choice model with experimental design data to forecast consumer demand for a unique cultural event. *Journal of Consumer Research*, 10, 348-361. DOI: 10.1086/208974
- Miller, M.E.; Hui, S.L.; Tierney, W.M. (1991). Validation Techniques for Logistic Regression Models. *Statistics in Medicine*, 10, 1213-1226. DOI:10.1002/sim.4780100805

- Mitrevski, P.; Hristoski, I. (2011). *e-Consumer Online Behavior: A Basis for Obtaining e-Commerce Performance Metrics*. Macedonia. ICT Innovations 2010, CCIS 83, pp. 142-151
- Mollalo, A.; Mao, L.; Rashidi, P.; Glass, G.E. (2019). A GIS-Based Artificial Neural Network Model for Spatial Distribution of Tuberculosis across the Continental United States. *International journal of environmental research and public health*, 16, 1-17. DOI: 10.3390/ijerph16010157
- Morganti, E.; Seidel, S.; Blanquart, C.; Dablanc, L.; Lenz, B. (2014). The impact of e-commerce on final deliveries: alternative parcel delivery services in France and Germany. *Transportation Research Procedia*, 4, 178-190. DOI: 10.1016/j.trpro.2014.11.014
- Nakatsuji, T.; Kaku, T. (1991). Development of a Self-Organizing Traffic Control System Using Neural Network Models. *Transportation Research Record*, 1324, 137-145.
- Oliveira, L.K.; Morganti, E.; Dablanc, L.; Diniz, Oliveira, R.L.M. (2017). Analysis of the potential demand of automated delivery stations for e-commerce deliveries in Belo Horizonte, Brazil. *Research in Transportation Economics*, 65, 34-43. DOI: 10.1016/j.retrec.2017.09.003
- Paiva Junior, H. (2006). Segmentação e Modelagem Comportamental de Usuários dos Serviços de Transporte Urbano Brasileiros. 97f. Tese (Doutorado em Engenharia de Transportes), Escola Politécnica da Universidade de São Paulo.
- Pappas, I.O.; Kourouthanassis, P.E.; Giannakos, M.N.; Chrissikopoulos, V. (2015). Explaining online shopping behavior with fsQCA: The role of cognitive and affective perceptions. *Journal of Business Research*, 69, 794-803. DOI: 10.1016/j.jbusres.2015.07.010
- Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of machine Learning research*, 12, 2825-2830. Available at <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Prebreza, A.; Shala, B. (2021). The Trust in Online Shopping during COVID-19: Case Study from Kosovo. *Open Access Library Journal*, 8, 1-15. DOI: 10.4236/oalib.1107288.
- Rabitz, H. (1989). Systems analysis at the molecular scale. *Science*, 246, 221–226. DOI: 10.1126/science.246.4927.221
- Ramachandran, P.; Zoph, B.; Le, Q.V. (2017). *Searching for activation functions*. Cornell University, USA. arXiv: 1710.05941
- Razavi, S.; Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of ‘global’ sensitivity in earth and environmental systems models. *Water Resource Research*, 51(5), 3070-3092. DOI: 10.1002/2014WR016527

- Regan, A.C.; Garrido, R.A. (2002). *Modelling freight demand and shipper behaviour: state of the art, future directions*. Travel Behaviour Research: The Leading Edge. Pergamon, Amsterdam. 185-215.
- Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511812651
- Robertson, T. S.; Zielinski, J.; Ward, S. (1984). *Consumer behavior*. Glenview: Scott Foresman.
- Rodrigue, J.P. (1997). Parallel modelling and neural networks: An overview for transportation/land use systems. *Transportation Research Part C*, 5, 259-271. DOI:10.1016/S0968-090X(97)00014-4
- Schalkoff, R.J. (1997). *Artificial Neural Networks*. McGraw-Hill, New York.
- Simpson, P.K. (1990). *Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations*. Elmsford, NY: Pergamon Press, Inc.
- Sippel, R. (1997). An Experiment on the Pure Theory of Consumer's Behaviour. *Economic Journal*, 107(444), 1431-44. DOI: 10.1111/j.1468-0297.1997.tb00056.x.
- Streiner, D. L. (2003). Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80, 217-222. DOI: 10.1207/S15327752JPA8003\_01
- Swinyard, W.R.; Smith, S.M, (2003). Why people (don't) shop online: A lifestyle study of the Internet consumer. *Journal Psychology & Marketing*, 20(7), 567-597. DOI: 10.1002/mar.10087
- Tolles, J.; Meurer; W. J. (2016). Logistic Regression Relating Patient Characteristics to Outcomes. *JAMA*, 316(5), 533-4. DOI:10.1001/jama.2016.7653.
- van Cranenburgh, S.; Alwosheel A. (2019). An artificial neural network based approach to investigate travellers' decision rules. *Transportation Research: Part C*, 98 152-166. DOI: 10.1016/j.trc.2018.11.014
- Vasić, N.; Kilibarda, M.; Kaurin, T. (2019). The Influence of Online Shopping Determinants on Customer Satisfaction in the Serbian Market. *Journal of theoretical and applied electronic commerce research*, 14(2), 70-89. DOI: 10.4067/S0718-18762019000200107
- Visa, S.; Ramsay, B.; Ralescu, A.; van der Knaap, E. (2011). Confusion Matrix-based Feature Selection. *Proceedings of the Twenty second Midwest Artificial Intelligence and Cognitive Science Conference*, Cincinnati, Ohio, 710, 120-127
- Wang, G.; Liu, Y.; Xu, J. (2020). Short-term failure mechanism triggered by hydraulic fracturing. *Energy Science and Engineering*, 8, 592-601. DOI: 10.1002/ese3.535
- Warner, B., Misra, M. (1996). Understanding neural networks as statistical tools. *The American Statistician*, 50(4), 284-293. DOI: 10.2307/2684922

- Wild, C.J., Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-248. DOI: 10.1111/j.1751-5823.1999.tb00442.x
- Xiao, Z., Wang, J.J., and Liu, Q. (2018). The impacts of final delivery solutions on e-shopping usage behaviour: The case of Shenzhen, China. *International Journal of Retail & Distribution Management*, 46(1), 2-20. DOI:10.1108/IJRDM-03-2016-0036
- Yuen, K.; Wang, X.; Ng, L.; Wong, Y. (2018). An investigation of customers' intention to use self-collection services for last-mile delivery. *Transport Policy*, 66, 1-8. DOI: 10.1016/j.tranpol.2018.03.001
- Zeng, M.; Cao, H.; Chen, M.; Li, Y. (2018). User behavior modeling, recommendations, and purchase prediction during shopping festivals. *Electronic Markets*, 29, 263-274. DOI: 10.1007/s12525-018-0311-8
- Zhang, Y.; Wallace, B.C. (2016). A sensitivity analysis of (and Practitioners' Guide to) convolutional neural networks for sentence classification; Proceedings of the Eighth International Joint Conference on Natural Language Processing, 1, 253-263. arXiv:1510.03820
- Zhao, Q.; Hastie, T. (2019). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39, 272-281. DOI: 10.1080/07350015.2019.1624293
- Zhao, X.; Yanb, X.; Hentenrycka, P.V. (2019). *Modeling Heterogeneity in Mode-Switching Behavior Under a Mobility-on-Demand Transit System: An Interpretable Machine Learning Approach*. ArXiv ID: 1902.02904

## 8 APÊNDICE

### 8.1 Questionário aplicado

#### Quais os motivos para você comprar um produto pela Internet? Quais os motivos para pedir um delivery?

Esta pesquisa tem por objetivo coletar informações para entender o comportamento do consumidor do comércio eletrônico no momento de uma compra pela Internet e no processo de recebimento do produto.

Neste estudo, devem ser consideradas compras realizadas exclusivamente pela Internet e que geram uma entrega física. Desta forma, produtos comprados para download não são objeto de análise desta pesquisa, como mídias virtuais (exemplo: filmes, livros, músicas e jogos).

O tempo médio para responder esta pesquisa é de 10 minutos.

A pesquisa faz parte das pesquisas desenvolvidas pelo Grupo de Pesquisa em Transporte e Logística Urbana, da Universidade Federal de Minas Gerais. As informações serão tratadas como confidenciais e não serão divulgadas e/ou utilizadas para outros fins que não acadêmicos. Mais informações podem ser obtidas através do email [leise@etg.ufmg.br](mailto:leise@etg.ufmg.br)

**\*Obrigatório**

#### 1. 1.1 Você faz compras pela

**Internet?** \* Marcar apenas uma

oval.

☐

Sim

Não

*Ir para a pergunta 17.*

#### A compra pela Internet e a entrega dos produtos

Esta seção tem por objetivo coletar informações para entender o processo de compra pela Internet e como você recebe o produto em casa.

#### 2. 2.1 Por onde você mais acessa a Internet para fazer a

**compra?** \* Marcar apenas uma oval.

☐

Smartphone

☐

Tablet

Computador

3. **2.2 Em relação aos produtos abaixo, quais você costuma comprar pela Internet? \*** *Marque todas que se aplicam.*

- ☐ Moda e acessórios
- ☐ Alimentos, bebida e suplementos
- ☐ Medicamentos
- ☐ Livros, CD, Papelaria, Gráfica
- ☐ Eletrônicos (smartphone, computador, tablets)
- ☐ Eletrodomésticos
- ☐ Cosméticos e perfumaria
- ☐ Acessórios automotivos
- ☐ Outro:

4. **2.3 Qual o número de entregas que você recebe em casa de produtos adquiridos pela Internet? \***

*Marcar apenas uma oval.*

- ☐ Uma entrega ou menos por mês
- ☐ De 2 a 4 entregas por mês
- ☐ De 5 a 8 entregas por mês
- ☐ Acima de 9 entregas por mês

5. **2.4 O valor do frete é um fator determinante para você comprar pela Internet. \*** *Marcar apenas uma oval.*

- ☐ Discordo totalmente
- ☐ Discordo
- ☐ Neutro
- ☐ Concordo
- ☐ Concordo totalmente

- 2.6 O valor que você está disposto a pagar pelo frete é influenciado por quais dos fatores descritos abaixo:

*Marque todas que se aplicam.*

- ☐ Tamanho do produto
- ☐ Peso do produto
- ☐ Valor da compra
- ☐ Prazo de entrega
- ☐ Distância do local de origem do produto
- ☐ Outro:

7. 2.5 Para uma compra acima de R\$300, qual valor você aceita pagar pelo frete? (apenas número) \*

8. 2.6 A possibilidade de escolher o dia e o horário de entrega influencia na decisão de realizar uma compra pela Internet. \*

*Marcar apenas uma oval.*

- ☐ Discordo totalmente
- ☐ Discordo
- ☐ Neutro
- ☐ Concordo
- ☐ Concordo totalmente

9. 2.7 Você estaria disposto a pagar mais pelo frete para receber um produto em um prazo menor. \*

*Marcar apenas uma oval.*

- ☐ Discordo totalmente
- ☐ Discordo
- ☐ Neutro
- ☐ Concordo

Concordo totalmente ☐

**2.8 Você estaria disposto a pagar mais pelo frete para escolher dia e horário da entrega. \***

*Marcar apenas uma oval.*

- ☐ Discordo totalmente
- ☐ Discordo
- ☐ Neutro
- ☐ Concordo
- ☐ Concordo totalmente

**11. 2.9 Qual é o período preferencial para recebimento das mercadorias compradas pela Internet?**

*Marque todas que se aplicam.*

- ☐ Indiferente
- ☐ Segunda a sexta-feira, 8h as 18h
- ☐ Segunda a sexta-feira, 18h as 20h
- ☐ Segunda a sexta-feira, após as 20h
- ☐ Sábado
- ☐ Domingo ou feriado

**Fatores que Influenciam na Realização de uma Compra pela Internet**

Esta seção, queremos entender o grau de importância para alguns fatores no momento da compra e da entrega que influenciam na tua decisão de comprar determinado produto por computador, smartphones ou tablets.

Caso esteja respondendo este questionário em um smartphone, é sugerido girar a tela para visualizar todas as colunas.

**3.1. Qual é o grau de importância dos seguintes fatores na hora de comprar pela Internet? \***  
*Marcar apenas um oval por linha.*

|                                                                      | Irrelevante           | Pouco importante      | Neutro (nem muito nem pouco importante) | Importante            | Muito importante      |
|----------------------------------------------------------------------|-----------------------|-----------------------|-----------------------------------------|-----------------------|-----------------------|
| Comodidade de comprar sem sair de casa                               | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Privacidade                                                          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Segurança na hora da compra                                          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Preço do produto                                                     | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Oportunidade de Compra (aproveitar ofertas relâmpagos ou benefícios) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Condições de pagamento maior que da loja física                      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Desconto para pagamento com boleto                                   | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Prazo de entrega                                                     | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Valor do Frete                                                       | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Frete grátis                                                         | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Horário da entrega                                                   | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |
| Ponto de retirada                                                    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>                   | <input type="radio"/> | <input type="radio"/> |

## Caracterização do entrevistado

As informações abaixo solicitadas serão tratadas como confidenciais e não serão divulgadas e/ou utilizadas para outros fins que não acadêmicos.

### 29. 8.1 Faixa etária \*

*Marcar apenas uma oval.*

- ☐ 15 a 24 anos
- ☐ 25 a 34 anos
- ☐ 35 a 49
- ☐ 50 a 69 anos
- ☐ Mais de 70 anos

### 30. 8.2 Renda familiar \*

*Marcar apenas uma oval.*

- ☐ Até um salário mínimo (R\$954,00)
- ☐ De 2 a 4 salários mínimos (R\$955,00 a R\$3.816,00)
- ☐ De 4 a 10 salários mínimos (R\$3.817,00 a R\$9.540,00)
- ☐ De 10 a 20 salários mínimos (R\$9.541,00 a R\$19.080,00)
- ☐ Mais de 20 salários mínimos (R\$19.081,00)

### 31. 8.3 Sexo \*

*Marcar apenas uma oval.*

- ☐ Feminino
- ☐ Masculino

**8.4 Em qual Estado você mora? \****Marcar apenas uma oval.*

- ☐ Acre
- ☐ Alagoas
- ☐ Amapá
- ☐ Amazonas
- ☐ Bahia
- ☐ Ceará
- ☐ Distrito Federal
- ☐ Espírito Santo
- ☐ Goiás
- ☐ Maranhão
- ☐ Mato Grosso
- ☐ Mato Grosso do Sul
- ☐ Minas Gerais
- ☐ Pará
- ☐ Paraíba
- ☐ Paraná
- ☐ Pernambuco
- ☐ Piauí
- ☐ Rio de Janeiro
- ☐ Rio Grande do Norte
- ☐ Rio Grande do Sul
- ☐ Rondônia
- ☐ Roraima
- ☐ Santa Catarina
- ☐ São Paulo
- ☐ Sergipe
- ☐ Tocantins

**33. 8.5 Em qual cidade você mora? (Ex.: Belo Horizonte) \*****8.6 Qual o CEP da residência? (Ex. 11111000)**

\*

---

## 8.2 Análise gráfica: Perfil dos respondentes

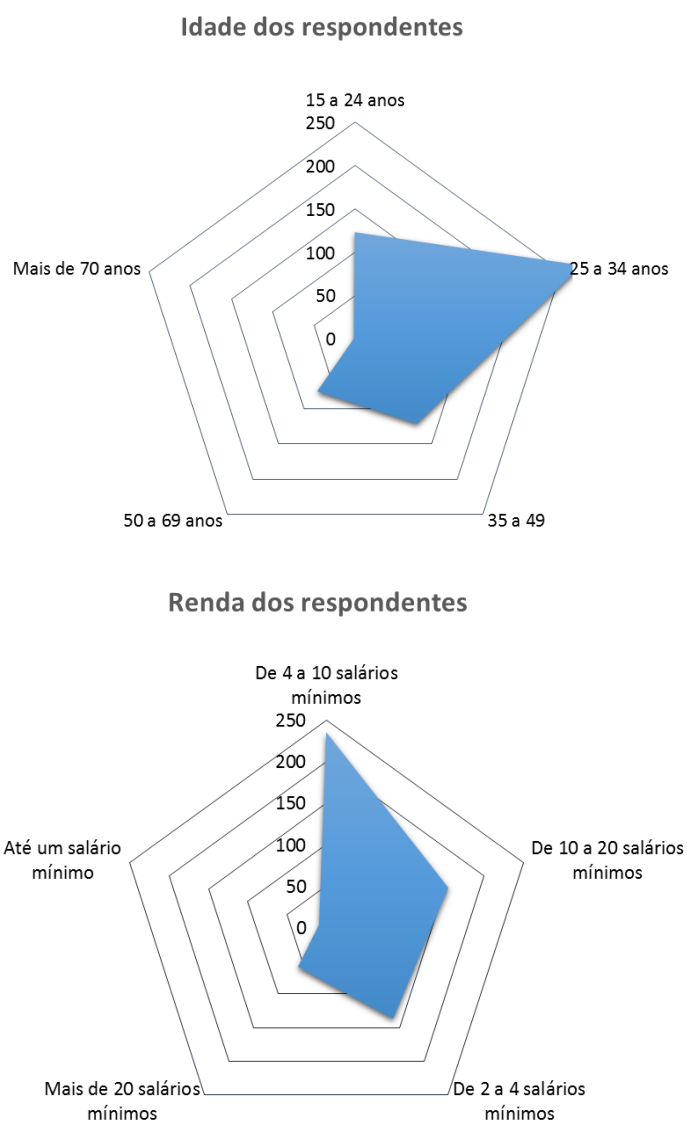


Figura 8.1 Perfil socioeconômico dos respondentes

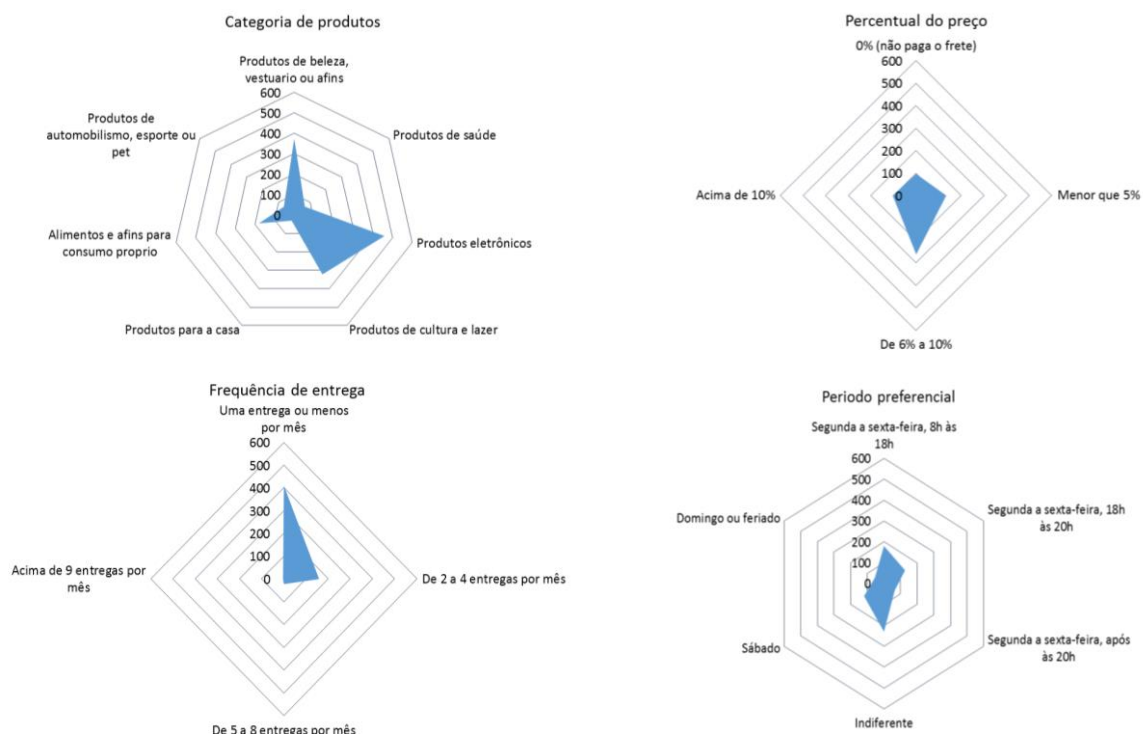


Figura 8.2 Perfil dos respondentes quanto aos hábitos de compras

### 8.3 Distribuição de frequência das variáveis dos modelos

A figura 8.3 apresenta histogramas para as variáveis de V01 a V18, descrito na seção 4.3. Através desses, para cada variável, é possível visualizar a representação gráfica da distribuição de frequências (eixo das ordenadas) das classes de resposta (eixo das abcissas). Ademais, a análise da frequência é completada por curvas de estimativa de densidade Kernel, representadas em cada gráfico.

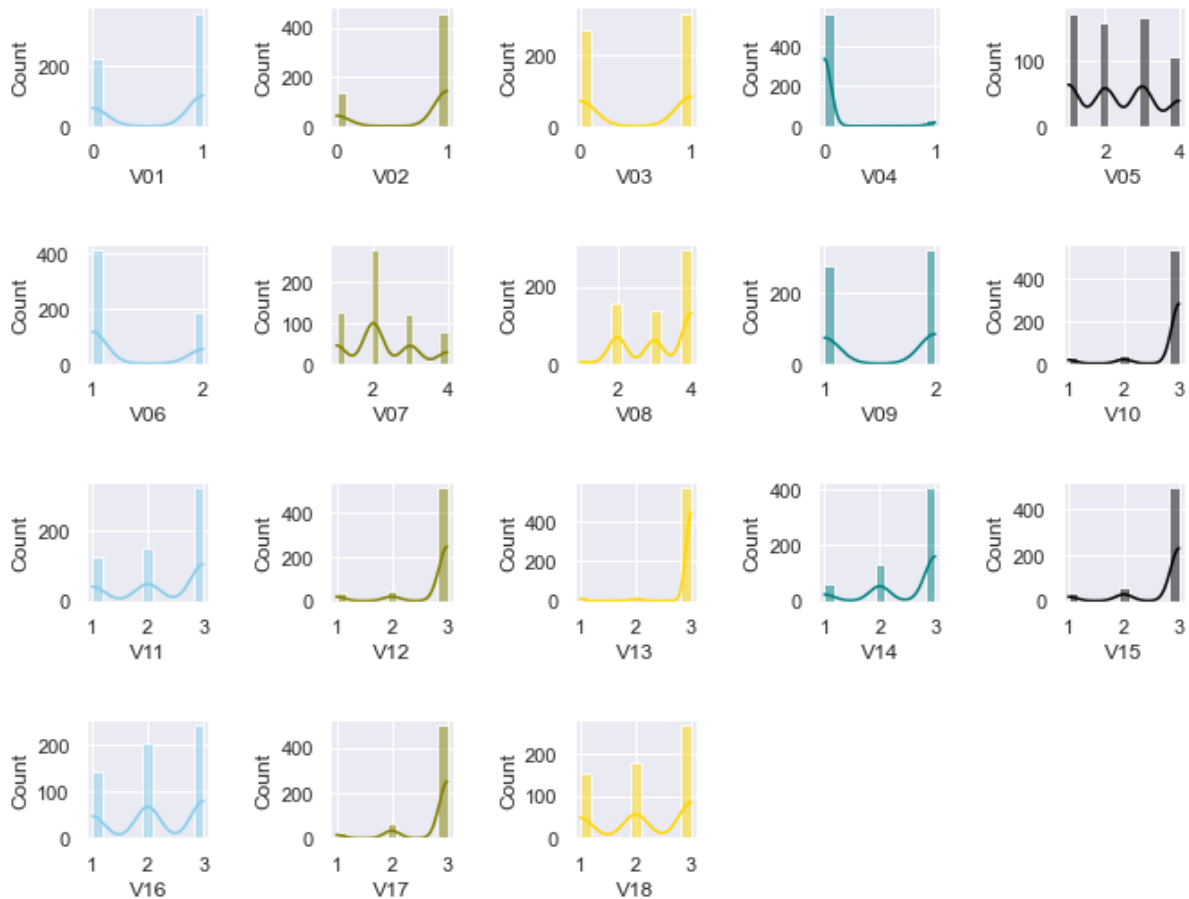


Figura 8.3 Histogramas e Curvas de Kernel – Variáveis V01 a V18

#### 8.4 Métricas de qualidade dos modelos estimados por RNA

Tal como descrito na seção 4.4.1, a análise completa das métricas de qualidade aferidas para os modelos estimados por RNA são apresentadas, por categoria de resposta, nas tabelas a seguir.

Tabela 8.1 Métricas de qualidade RNA – Modelo 1

| Categorias da variável dependente | Precisão | Revocação | <i>F1-score</i> | Frequência (n = 90) | Matrix de confusão |
|-----------------------------------|----------|-----------|-----------------|---------------------|--------------------|
| (0) Resposta negativa             | 0,00     | 0,00      | 0,00            | 7                   | [0, 7],            |
| (1) Resposta positiva             | 0,92     | 0,93      | 0,92            | 83                  | [6, 77]            |
|                                   |          |           |                 | <b>Acurácia</b>     | <b>86%</b>         |

Tabela 8.2 Métricas de qualidade RNA – Modelo 2

| Categorias da variável dependente | Precisão | Revocação | <i>F1-score</i> | Frequência (n = 90) | Matrix de confusão |
|-----------------------------------|----------|-----------|-----------------|---------------------|--------------------|
| (0) Resposta negativa             | 0,27     | 0,32      | 0,29            | 19                  | [6, 13],           |
| (1) Resposta positiva             | 0,81     | 0,77      | 0,79            | 71                  | [16, 55]           |
|                                   |          |           |                 | <b>Acurácia</b>     | <b>68%</b>         |

Tabela 8.3 Métricas de qualidade RNA – Modelo 3

| Categorias da variável dependente | Precisão | Revocação | <i>F1-score</i> | Frequência (n = 90) | Matrix de confusão |
|-----------------------------------|----------|-----------|-----------------|---------------------|--------------------|
| (0) Resposta negativa             | 0,29     | 0,13      | 0,18            | 15                  | [2, 13],           |
| (1) Resposta positiva             | 0,84     | 0,93      | 0,89            | 75                  | [5, 70]            |
|                                   |          |           |                 | <b>Acurácia</b>     | <b>80%</b>         |

Tabela 8.4 Métricas de qualidade RNA – Modelo 4

| Categorias da variável dependente | Precisão | Revocação | <i>F1-score</i> | Frequência (n = 90) | Matrix de confusão |
|-----------------------------------|----------|-----------|-----------------|---------------------|--------------------|
| (0) Resposta negativa             | 0,50     | 0,10      | 0,17            | 10                  | [1, 9],            |
| (1) Resposta positiva             | 0,90     | 0,99      | 0,94            | 80                  | [1, 79]            |
|                                   |          |           |                 | <b>Acurácia</b>     | <b>89%</b>         |

Tabela 8.5 Métricas de qualidade RNA – Modelo 5

| Categorias da variável dependente | Precisão | Revocação | <i>F1-score</i> | Frequência (n = 90) | Matrix de confusão |
|-----------------------------------|----------|-----------|-----------------|---------------------|--------------------|
| (0) Resposta negativa             | 0,33     | 0,30      | 0,31            | 27                  | [8, 19],           |
| (1) Resposta positiva             | 0,71     | 0,75      | 0,73            | 63                  | [16, 47]           |
|                                   |          |           |                 | <b>Acurácia</b>     | <b>61%</b>         |