UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS DE RUSSAS

CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE

ANA CIBELE RODRIGUES LIMA

ESTUDO DA INCERTEZA DE JANELAS DE ATENDIMENTO EM PROBLEMAS DE ROTEIRIZAÇÃO DE VEÍCULOS

RUSSAS

2022

ANA CIBELE RODRIGUES LIMA

ESTUDO DA INCERTEZA DE JANELAS DE ATENDIMENTO EM PROBLEMAS DE ROTEIRIZAÇÃO DE VEÍCULOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Software do Campus de Russas da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Software.

Orientadora: Prof. Dra. Rosineide Fernando da Paz

Coorientador: Prof. Dr. Alexandre Matos Arruda

RUSSAS

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

L696e Lima, Ana Cibele Rodrigues.
Estudo da incerteza de janelas de atendimento em problemas de roteirização de veículos / Ana Cibele Rodrigues Lima. – 2022.
49 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Russas, Curso de Engenharia de Software, Russas, 2022.
Orientação: Profa. Dra. Rosineide Fernando da Paz.
Coorientação: Prof. Dr. Alexandre Matos Arruda.

1. Logística. 2. Janelas de Atendimento. 3. Roteirização de Veículos. 4. Otimização Combinatória. I. Título.

CDD 005.1

ANA CIBELE RODRIGUES LIMA

ESTUDO DA INCERTEZA DE JANELAS DE ATENDIMENTO EM PROBLEMAS DE ROTEIRIZAÇÃO DE VEÍCULOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Software do Campus de Russas da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Software.

Aprovada em:

BANCA EXAMINADORA

Prof. Dra. Rosineide Fernando da Paz (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Alexandre Matos Arruda (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Anderson Feitoza Leitão Maia
Universidade Federal do Ceará (UFC)

Edson Rocha Patricio
GreenMile (GM)

AGRADECIMENTOS

O desenvolvimento deste trabalho de conclusão de curso contou com a ajuda de diversas pessoas dentre as quais agradeço.

Agradeço aos meus pais Iolanda Rodrigues e Luiz Lima, por todo amor e esforço que vocês fizeram para que eu obtivesse essa conquista. A vocês, todo o meu amor e a minha gratidão.

As minhas irmãs Vanessa Rodrigues e Ana Lívia Rodrigues por estarem sempre presentes, me ajudarem em momentos difíceis, me ouvirem repetidamente ensaiando para apresentações ou estudando e me incentivarem todos os dias a ser melhor, melhor irmã, melhor amiga, melhor estudante, melhor profissional. Eu amo vocês.

Ao Professor Pablo Soares, por me ajudar todos os dias no inicio dessa trajetória a aprender a programar, mesmo papel e caneta sendo o único recurso. Obrigada por enxergar potencial onde eu mesma não conseguia e ter me dado forças para continuar quando pensei em desistir.

Ao Professor Filipe Maciel, por ter me dado a oportunidade de ingressar no Laboratório de Tecnologias Inovadoras e ser sua orientanda. Grata por todas as horas de discussão, conselhos acadêmicos e profissionais. Os aprendizados adquiridos foram imensos e me tornaram melhor em todos os âmbitos.

A minha amiga Ana Lara, por estar presente durante toda a graduação e ser minha duplinha. Obrigada por me ouvir, me abraçar e dividir comigo muitos momentos, sejam de alegria ou tristeza. Você foi essencial na minha trajetória e agradeço muito por ter você comigo.

Ao meu amigo Lucas Talam, por todas as conversas e apoio durante essa trajetória. Você sempre me deu força para continuar e nunca me deixou desanimar. Obrigada por me fazer enxergar meu potencial e proporcionar tantos momentos felizes.

A todos os meus amigos que a Universidade Federal do Ceará me proporcionou conhecer, em especial ao Marlo Henrique, Cleiton Monteiro, Artur Sampaio, Mateus Franco, Susana Moreira e Leonardo David que tornaram meus dias melhores e me ajudaram nas dificuldades. Obrigada por tudo, foi muito mais fácil tendo vocês ao lado.

Ao João Lucas, que me acompanhou nesses últimos anos de trajetória. Obrigada por todo incentivo, carinho e colo em todos os momentos.

Aos meus amigos de infância Eduardo Rubens e Sarah Oliveira, por estarem sempre torcendo por mim, mesmo depois da distância física que a graduação causou. Obrigada por toda

amizade, carinho e compreensão, amo vocês.

Por fim, a todos que contribuíram direta ou indiretamente para a minha graduação. Muito obrigada.

"O primeiro pecado da humanidade foi a fé; a primeira virtude foi a dúvida."

(Carl Sagan)

RESUMO

O Problema de Roteirização de Veículos com Janela de Atendimento é uma extensão do Problema de Roteirização de Veículos adicionando janelas de atendimento. Embora amplamente estudado na literatura, os métodos propostos para resolução do Problema de Roteirização de Veículos com Janela de Atendimento (PRVJA) assumem que as janelas fornecidas estão corretas. A partir disso, em decorrência de janelas de atendimento inválidas, entregas mal sucedidas podem ocorrer com maior frequência. O presente trabalho propõe estratégias para otimizar a chance de uma entrega ser bem sucedida. Para esse propósito, são utilizados dados históricos de tentativas de entregas, os quais foram fornecidos pela empresa GreenMile. A esses dados são aplicados métodos estatísticos para dados em intervalos limitados e métodos de regressão para resposta binária. O horário da tentativa de entrega é modelado na abordagem para dados em intervalo limitado. Outra abordagem adotada é o uso do modelo de regressão logística. Nessa abordagem, o objetivo é utilizar um conjunto de variáveis explicativas na obtenção da probabilidade de se obter sucesso, ou fracasso, numa tentativa de entrega. Como resultado, obteve-se uma estratégia para tomada de decisão sobre o horário para realização de uma entrega, considerando as características do estabelecimento atendido e os dados históricos das entregas realizadas.

Palavras-chave: Logística. Janelas de Atendimento. Roteirização de Veículos. Otimização Combinatória.

ABSTRACT

The Vehicle Routing Problem with Time Window is a Vehicle Routing Problem extension by adding time windows. Although extensively studied in the literature, the methods proposed for solving the Vehicle Routing Problem with Time Window assume that the given time windows are correct. Then due to invalid time windows, unsuccessful deliveries may occur more frequently. The present work proposes strategies to optimize the chance of successful delivery. For this purpose, historical data of delivery attempts are used, provided by GreenMile company. Statistical methods for data in limited intervals and regression methods for binary response apply to these data. Delivery attempt hour is modeled in the approach for data in limited intervals. Another approach adopted is the use of the logistic regression model. In this approach, the goal is to use a set of explanatory variables to obtain the probability of achieving success, or failure, in a delivery attempt. As a result, it was getting a decision-making strategy on the optimal time interval to carry out delivery, considering the characteristics of the establishment served and the historical data of the deliveries made.

Keywords: Logistic. Time Windows. Vehicle Routing. Combinatorial Optimization.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

BD	Banco de Dados
BT	Busca Tabu
CD	Centro de Distribuição
CRISP-DM	Processo Padrão de Indústria Cruzada para Exploração de Dados
PCV	Problema do Caixeiro Viajante
PCVJA	Problema do Caixeiro Viajante com Janelas de Atendimento
PRV	Problema de Roteirização de Veículos
PRVJA	Problema de Roteirização de Veículos com Janela de Atendimento
SQL	Structured Query Language
VA	Variável Aleatória

SUMÁRIO

1 INTRODUÇÃO

Introduzido por Dantzig e Ramser (1959) o Problema de Roteirização de Veículos (PRV) pertence a família de problemas de Otimização Combinatória, com alta complexidade computacional, aplicados à distribuição de mercadorias. Essa teoria possui diversas vertentes que incorporam complexidades associadas a problemas do mundo real. Dentre elas: PRV com janelas de atendimento (GENDREAU; TARANTILIS, 2010), problemas de entrega e coleta (BERBEGLIA *et al.*, 2007) e problema de roteirização com múltiplos depósitos (MONTOYA-TORRES *et al.*, 2015).

De acordo com Eksioglu *et al.* (2009), a literatura sobre o PRV têm crescido exponencialmente a uma taxa de 6% ao ano, demonstrando sua relevância no contexto logístico atual. Segundo Medeiros (2020), o *Ecommerce* teve em 2020 uma expansão de 40% no número de usuários, favorecendo as empresas transportadoras, visto que o elemento essencial no setor de logística é o sistema de transporte, que é requerido durante todo o processo de produção, desde a fabricação, até a entrega da mercadoria ao consumidor final, e, ainda, o potencial retorno da mercadoria.

O Problema de Roteirização de Veículos com Janela de Atendimento é um problema, dentre vários, derivado do PRV. Em problemas com janela de atendimento, um conjunto de veículos é necessário para atender um grupo de solicitações geograficamente dispersas. Cada requisição leva uma determinada quantidade de mercadorias que deve ser transportada de uma origem e destino especificados dentro das janelas de atendimento definidas. Cada janela representa o intervalo de tempo que deve ser respeitado para uma entrega ser bem sucedida.

Para o PRVJA, são propostos algoritmos exatos ou heurísticos, que têm como objetivo projetar rotas ótimas de menor custo de um Centro de Distribuição (CD) para um conjunto de clientes com demanda conhecida, de forma que cada cliente seja visitado apenas uma vez por exatamente um veículo em um determinado intervalo de tempo, sem violar as restrições de capacidade. Dentre os trabalhos recentemente desenvolvidos nessa área podem ser citados Pan *et al.* (2021), Sun *et al.* (2020), Sun *et al.* (2019).

Todos os trabalhos citados aqui supõem a existência de uma janela de atendimento em que o serviço de entrega ou coleta possa ser iniciado. Ou seja, os algoritmos propostos, sejam heurísticos ou exatos, pressupõem que as janelas não sejam contestáveis, ignorando, até mesmo, erro humano durante o cadastro. No entanto, em problemas práticos, nem sempre o cadastro dessas janelas é realizado pontualmente, ou é realizado sem contato com o cliente, ou

são cadastrados de forma ampla podendo constar janelas que tomam as 24 horas do dia. Quando isso ocorre, estabelecimentos podem ser visitados em horários inapropriados, o que permite aos usuários do serviço de entrega recusar a mesma ou entrar num estado de insatisfação com o serviço.

Se a informação sobre a janela de atendimento não está presente de forma confiável, tentativas de entregas ou coletas podem resultar em insucesso da operação. Assim, uma maneira de otimizar as rotas é reduzir a quantidade de vezes que uma tentativa de entrega, ou coleta, não é bem sucedida. Motivados pelas questões apresentadas, este trabalho busca desenvolver estratégias para reduzir o problema da ocorrência de entregas mal sucedidas devido a janelas de atendimento. Para tal, uma possível solução foi construída a partir da aplicação de métodos estatísticos, incluindo modelos probabilísticos para dados em intervalos limitados, como descritos em Paz *et al.* (2019), Gupta e Nadarajah (2004), Pearson (1895) e modelos de regressão para resposta binária aos dados fornecidos pela empresa (GREENMILE, n.d). Desta forma, este trabalho tem como principal função a criação de um modelo que reduz a probabilidade de ocorrência de uma entrega mal sucedida, buscando reduzir os prejuízos atrelados a este evento. Os resultados poderão ser utilizados para melhorar a performance do algoritmo de roteirização com janelas de atendimento utilizado pela empresa fazendo com que o mesmo encontre uma rota ótima com redução da probabilidade de entregas mal sucedidas.

Este trabalho esta dividido em sete capítulos. O primeiro introduz a problemática, juntamente com a importância do problema abordado na monografia. Os objetivos esperados no desdobramento da pesquisa são retratados no Capítulo 2. O Capítulo 3 apresenta a definição dos conceitos essenciais para a compreensão e resolução do problema. O Capítulo 4 descreve os trabalhos relacionados a este trabalho. A metodologia adotada para o desenvolvimento desta pesquisa é apresentada no Capítulo 5. No Capítulo 6 são apresentados os resultados obtidos. Por fim, o Capítulo 7 apresenta as conclusões e trabalhos futuros.

2 OBJETIVOS

2.1 Objetivo geral

Desenvolver estratégias para inclusão de informações mais precisas sobre janelas de atendimento para algoritmos de roteirização com o propósito de otimizar a chance de uma entrega ser bem sucedida.

2.2 Objetivos específicos

- Estruturar os dados fornecidos pela empresa, a fim de possibilitar o uso de ferramentas de análise de dados (matemáticas e computacionais);
- Realizar uma exploração nos dados de entrega fornecidos pela empresa com o objetivo de verificar se existem entregas sendo feitas fora da janela de atendimento cadastrada dos clientes;
- Realizar uma exploração nos dados de entrega da empresa a fim de investigar a frequência de ocorrência de tentativas de entrega que deixaram de ser realizadas devido ao cliente não estar aceitando entregas;
- Modelar os dados de horários das tentativas de entregas por meio de modelos de regressão para resposta limitada;
- Modelar os dados de tentativa de entregas usando modelos de regressão para resposta binária, levando em conta um conjunto de variáveis explicativas.
- Construir um indicador de horários propícios a entregas, com base na probabilidade de insucesso de entrega, a partir dos resultados obtidos nos objetivos anteriores.

3 FUNDAMENTAÇÃO TEÓRICA

Para a compreensão do problema abordado na pesquisa e como será modelada a solução para tal, é necessária a assimilação de alguns conceitos básicos sobre PRV que serão apresentados nas seções 3.1 e 3.2 respectivamente.

3.1 O Problema Clássico

O PRV é um problema clássico da área de Otimização Combinatória derivado do Problema do Caixeiro Viajante (PCV) com adição de restrição de capacidade dos veículos. Nesse problema consideram-se uma quantidade determinada de estabelecimentos espacialmente distribuídos com uma demanda de mercadorias associada. As mercadorias são entregues a partir de um CD por uma frota homogênea de veículos. Os veículos iniciam no CD e devem retornar ao mesmo. Durante o percurso, as mercadorias são entregues para um subconjunto de estabelecimentos, satisfazendo as necessidades de demanda de cada um. A rota de cada veículo deve obedecer a algumas restrições como: a quantidade de mercadoria entregue não deve exceder a capacidade do veículo e o tempo limite para realização de uma rota não deve ser ultrapassado. O objetivo do PRV é minimizar o custo total de transporte no atendimento aos clientes, isto é, minimizar custos fixos, custos operacionais e o número de veículos envolvidos no transporte.

Matematicamente, o VRP pode ser formulado como sendo um grafo $G = (V,A)$, não direcionado, onde $V = \{v_0, v_1, \cdots, v_n\}$ é o conjunto dos vértices, representando as cidades ou consumidores e $A = \{(v_i, v_j) : v_i, v_j \in V, i < j\}$ é o conjunto de arestas que ligam dois vértices. O vértice v_0 representa o depósito, sendo este a base de uma frota de veículos idênticos de capacidade Q. Cada consumidor v tem uma demanda não negativa q_i por um determinado produto. À cada aresta (v_i, v_j) está associada uma distância não negativa c_{ij}, que representa a distância entre os consumidores.

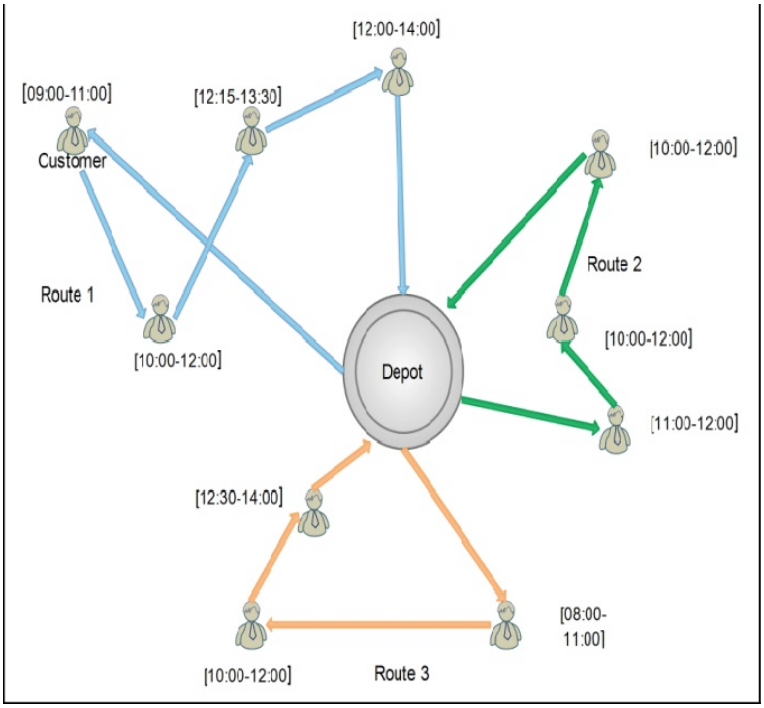
O conjunto de restrições que devem ser satisfeitas no PRV são:

- cada rota inicie e termine no CD;
- cada estabelecimento é visitado exatamente uma vez por exatamente um veículo;
- a demanda total de cada rota não pode exceder a capacidade do veículo Q.

3.2 Problema de Roteirização de Veículos com Janela de Atendimento

O PRVJA (FISHER *et al.*, 1997) é uma importante generalização do PRV que pode ser utilizado para modelar muitos problemas do mundo real e consiste na geração de um conjunto de rotas de custo mínimo, iniciadas e finalizadas no CD, para uma frota de veículos que atende um conjunto de estabelecimentos com demandas conhecidas. Os estabelecimentos devem ser atendidos exatamente por um veículo, de modo que a capacidade dos veículos não sejam excedidas. O atendimento a um estabelecimento deve ser iniciado dentro da janela de atendimento definida pela primeira e última hora em que o cliente permitir o início do atendimento. Algumas das aplicações do PRVJA a problemas do mundo real incluem entregas bancárias, entregas postais, coleta de lixo industrial, roteamento de ônibus escolar e serviços de patrulha de segurança. A Figura 1 mostra um exemplo de PRVJA com as arestas em azul representando o caminho percorrido pelo veículo para atender todos os 6 estabelecimentos geograficamente distribuídos e com restrições de janela de atendimento.

Figura 1 – Ilustração do PRVJA



Fonte: AGGARWAL *et al.*

4 TRABALHOS RELACIONADOS

Neste capítulo, será apresentada uma revisão de literatura nesse tema, que engloba trabalhos que abordam o PRVJA no que diz respeito as janelas de atendimento, sejam elas através de métodos heurísticos ou algoritmos exatos. Também é disponibilizado uma tabela para nível de comparação entre os trabalhos abordados neste trabalho, afim de destacar o diferencial da pesquisa.

Zhang *et al.* (2019) realizaram estudos sobre o Problema do Caixeiro Viajante com Janelas de Atendimento (PCVJA) em que as janelas de atendimento são incertas e propuseram um critério de decisão, chamado de *essential riskiness index*, baseado no *riskiness index* de Aumann e Serrano (2007). O objetivo é formular e resolver o problema do PCVJA de forma mais eficaz. Os autores também propuseram um método de decomposição de Benders para resolver as versões estocásticas e distribucionalmente robustas do PCVJA.

Yang *et al.* (2020) produziram um método de recuperação de distúrbios em rotas, com base na ideia de gerenciamento de distúrbios, que identifica quando houver mudanças na janela de atendimento durante a execução da rota planejada. Um método de distribuição baseado na meta-heurística Busca Tabu (BT), que emprega métodos de pesquisa local para otimização matemática, foi também proposto para ajustar uma rota ótima que garanta um desvio mínimo de custo generalizado. O método de distribuição proposto provou ser mais rápido e eficaz quando testado em um caso prático real e em um problema padrão de teste.

Žunić *et al.* (2020) propuseram um algoritmo para solucionar os PRV do mundo real com custo mínimo, atendendo todas as restrições consideradas realísticas como, por exemplo, restrição de janela de atendimento e, para configurar e ajustar os parâmetros do algoritmo, um modelo de predição que utiliza dados históricos. Para validar os resultados, os autores testaram sua solução em duas fases. O primeiro teste foi realizado utilizando dados padronizados, em que o algoritmo proposto apresentou resultados elevados e satisfatórios. O segundo teste foi realizado utilizando dados do mundo real, em que o algoritmo, na maioria das vezes, conseguiu atender a todas as restrições e, apesar do custo mínimo, teve um tempo de execução satisfatório dada a aplicação no mundo real.

A seguir, são apresentados na Tabela 1 os aspectos que diferenciam os trabalhos relacionados a este. Os aspectos marcados com X indicam sua existência nos trabalhos e o X destacado de vermelho simboliza o diferencial desta monografia em relação aos trabalhos relacionados.

Tabela 1 – Comparação da monografia com os trabalhos relacionados

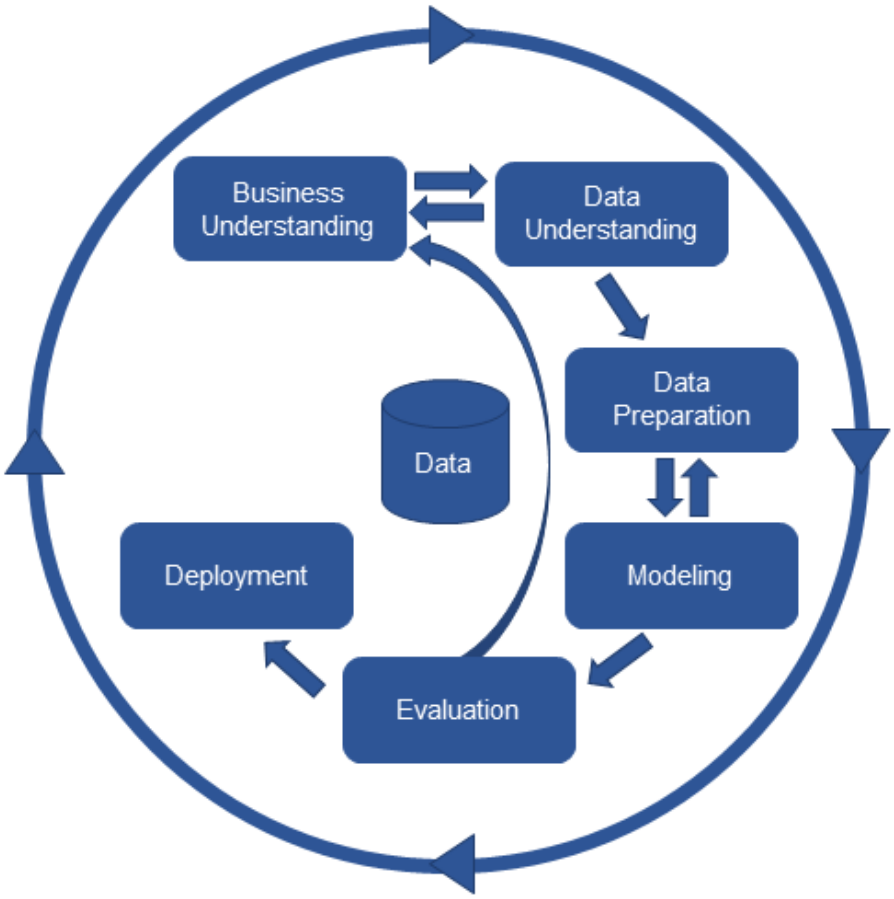
Lista de atividades	**Zhang et al. (2019)**	**Yang et al. (2020)**	**Zunic et al. (2020)**	**TCC Cibele**
Estudo do efeito da incerteza de janelas de atendimento.			X	X
Estudo do efeito da incerteza do tempo de serviço.	X			
Desenvolver ferramentas para otimização de rotas.	X	X	X	
Utilização de dados históricos.			X	X
Indicar probabilidade de insucesso de entrega.				X

Fonte: elaborado pela autora (2021).

5 METODOLOGIA

Este capítulo descreve os procedimentos metodológicos usados para a realização da monografia. A metodologia deste trabalho foi baseada no modelo Processo Padrão de Indústria Cruzada para Exploração de Dados (CRISP-DM), devido a sua ampla literatura disponível (RIVO *et al.*, 2012). A fase de *Deployment*, que corresponde a implantação do modelo, não será abordada nesta monografia. A empresa que forneceu seus dados para este trabalho identificou a necessidade de um estudo de implantação, dada a complexidade do seu processo de negócios. A Figura 2 ilustra as fases do processo.

Figura 2 – Esquema ilustrativo das fases do CRISP-DM



Fonte: (GONZALEZ, 2019)

5.1 Entendimento do negócio (*Business Understanding*)

A primeira fase consiste em pensar cuidadosamente sobre o cenário de uso da solução e o problema a ser resolvido. Nesta etapa, é analisada a viabilidade e estruturada possíveis soluções para o problema.

À medida que avançamos, voltamos e percebemos que, muitas vezes, o cenário de uso deve ser ajustado para refletir melhor a necessidade real de negócios. Através de reuniões, são apresentadas as regras de negócio da empresa que devem ser respeitadas, o problema enfrentado pelo cliente e a importância do projeto, para tornar clara a expectativa sobre o produto final.

5.2 Entendimento dos dados (*Data Understanding*)

Nesta fase é realizada uma análise exploratória dos dados disponíveis, buscando entender as informações que os dados podem fornecer. Nessa fase, as informações consideradas importantes são extraídas e analisadas.

Essa fase constitui no acesso, seleção e aquisição dos dados. Para acesso ao Banco de Dados (BD) foi utilizado o DataGrip[1], que possibilitou estruturar e realizar consultas a base de dados. Na fase de seleção, foi estudado quais informações poderiam ser úteis para atingir o objetivo proposto nesta monografia. Por fim, na fase de aquisição, é utilizado um microserviço, uma função essencial de uma aplicação utilizada de maneira independente. O microserviço é responsável por conectar a linguagem de programação Python (ROSSUM; JR, 1995) com o BD. Por meio de recursos da biblioteca Pandas (TEAM, 2020), a consulta em *Structured Query Language (SQL)* foi executada. Os dados resultantes são armazenados temporariamente, em tempo de execução, em uma variável em formato de *Dataframe*. Na Tabela 2 há uma descrição de cada variável selecionada nessa fase e que será tratada na etapa de preparação dos dados.

Os dados extraídos são do período de 01/01/2020 a 01/07/2021. Dessa forma, toda análise é desenvolvida utilizando os dados registrados nesse intervalo de tempo. Além disso, na consulta, são aplicados filtros que excluem estabelecimentos que não possuem janela de atendimento cadastrada, indicando possivelmente não terem restrições de janela de atendimento.

5.3 Preparação dos dados (*Data Preparation*)

Nesta fase os dados obtidos na fase de Entendimento dos dados são preparados para a etapa de Modelagem dos dados. Os dados são manipulados e convertidos para que rendam melhores resultados.

[1] https://www.jetbrains.com/datagrip/

Tabela 2 – Descrição das variáveis extraídas

Variável	Descrição	Tipo
routeStart	Horário de inicio da rota	datetime
origCDLat	Latitude do Centro de Distribuição	float
origCDLong	Longitude do Centro de Distribuição	float
storeId	Identificador do estabelecimento	int
arrival	Horário de chegada no estabelecimento	datetime
departure	Horário de saída no estabelecimento	datetime
stopLat	Latitude do estabelecimento	float
stopLong	Longitude do estabelecimento	float
twOpen	Horário de abertura da janela de atendimento do estabelecimento	datetime
twClose	Horário de fechamento da janela de atendimento do estabelecimento	datetime
serviceTimeType	Tipo de serviço do estabelecimento	categorical
undeliveryReason	Motivo de não ter realizado a entrega estabelecimento	categorical

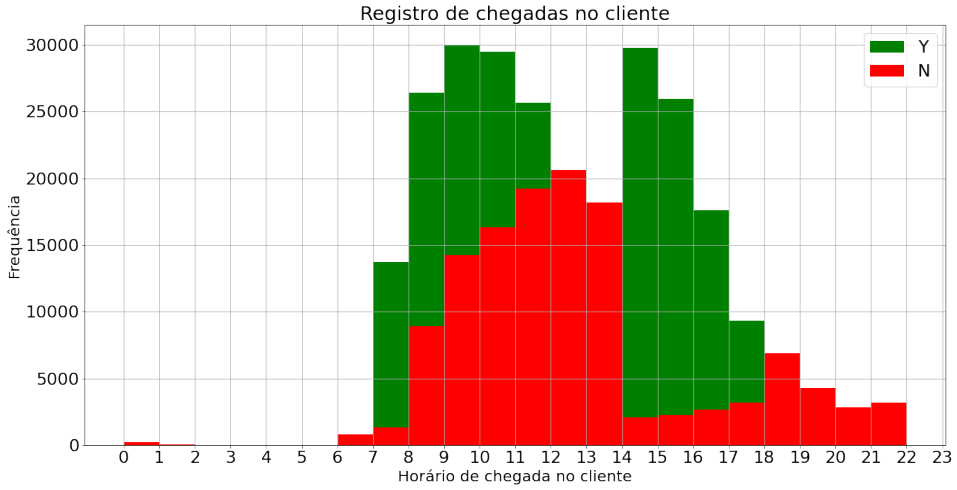
Fonte: elaborado pela autora (2021).

5.3.1 Conformidade de janela de atendimento

A conformidade de janela de atendimento se refere a verificação do horário de entrega de acordo com a janela de atendimento cadastrada. Neste estudo, é dito que uma entrega tem conformidade de janela de atendimento se seu horário de realização estiver dentro da janela de atendimento cadastrada do estabelecimento.

Para compreender e analisar o volume de entregas sem conformidade de janela de atendimento foi fixado um cliente da empresa. Cada cliente da empresa possui um conjunto de estabelecimentos. Dessa maneira, é estudada a frequência de entregas feitas fora da janela de atendimento dos estabelecimentos que possuem essa informação. Dessa forma, utilizando apenas os dados de entregas realizadas e as janelas cadastradas é possível visualizar que é preciso que haja melhoria em relação ao período em que o estabelecimento está aceitando entregas. Pois, uma grande parte das entregas são realizadas fora do intervalo em que o sistema considera o estabelecimento aberto. A Figura 3 mostra horários de início de entregas para os estabelecimentos atendidos pelo cliente selecionado. Todas as entregas foram realizadas, no entanto valores em vermelho representam as entregas realizadas em horários fora da janela de atendimento cadastrada, enquanto a cor verde representa as entregas realizadas dentro da janela de atendimento. Esse gráfico fornece evidências de que existem janelas que poderiam ser menos restritas, ou maiores, do que o que está atualmente cadastrado no sistema da empresa.

Figura 3 – Horários das entregas para um determinado cliente, nos diversos estabelecimentos atendidos.



Fonte: elaborado pela autora (2021).

5.3.2 Tempo de serviço

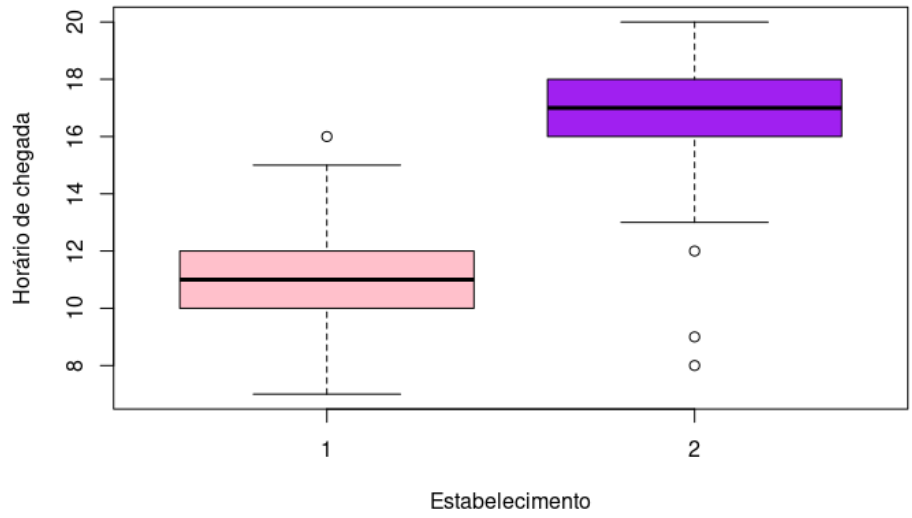
Os motoristas que utilizam a aplicação de roteirização podem, muitas vezes, não seguirem o processo esperado e registrar chegada e/ou saída no estabelecimento em horários que não representam a realidade. Para controle desse fato, existe uma variável de controle, que verifica se a geolocalização do motorista quando esse registrou a chegada e/ou saída do estabelecimento e a geolocalização do estabelecimento são iguais ou estão em um raio de proximidade definido pela empresa. Contudo, somente essa variável, por vezes, não é suficiente.

Para que os dados utilizados nesta monografia tenham maior nível de confiabilidade, é retirado do conjunto de dados todos os registros de chegada no estabelecimento que não tiveram, no mínimo, 1 minuto de tempo de serviço. O tempo de serviço representa o tempo que o motorista gastou no atendimento. Esse tempo é calculado através de uma simples subtração entre o horário de saída do estabelecimento e o horário de chegada. Tempos de serviço inferiores a 1 minuto representam uma informação não confiável devido ao fluxo que o motorista realiza ao chegar no estabelecimento. Mesmo se considerado uma entrega de peso mínimo, é necessário que o destinatário da entrega, no mínimo, assine um documento. Qualquer fluxo de entrega, desde o mais simples possível, é humanamente impossível de ser realizado em um intervalo de tempo inferior a 1 minuto.

5.3.3 Horário de chegada

O horário de chegada no estabelecimento é a variável com maior importância nesta monografia. Por meio dela é observado, além do intervalo de horários que o estabelecimento está aceitando entregas, quais horários tem a maior proporção de entregas bem sucedidas. A Figura 4 apresenta um *boxplot* referente aos horários de entrega em dois estabelecimentos, onde podemos observar a distribuição dos horários de chegada e suas medianas.

Figura 4 – Boxplot dos horários de entregas em dois estabelecimentos



Fonte: elaborado pela autora (2021).

Para que seja possível o ajuste de um modelo para dados no intervalo unitário, os horários de entrega são transformados, por meio da divisão dos valores originais por 23, em uma variável contínua limitada ao intervalo $(0, 1)$ ao contrário de usar os valores no intervalo $(0, 23)$.

A distribuição beta, descrita na seção 5.4.1.1, é utilizada para modelar dados no intervalo unitário. Dessa forma, testes de hipóteses são realizados para obter indícios de adequabilidade do modelo aos dados trabalhados.

5.4 Modelagem (*Modeling*)

Motivados pela análise descritiva e pelo problema em questão, são propostos os modelos descritos nas subseções 5.4.1 e 5.4.2. Para ajuste destes modelos a linguagem de programação R (R Core Team, 2021) em conjunto com os pacotes *lme4* (BATES *et al.*, 2015),

merTools (KNOWLES; FREDERICK, 2020) e *caret* (KUHN, 2008) foi aplicada. O método de inferência empregado foi o frequentista (FISHER, 1958).

5.4.1 Modelo de Regressão para resposta limitada

Os modelos de regressão para resposta limitada são aqueles em que a variável resposta assume valores no intervalo $(0,1)$. Essa abordagem é utilizada neste trabalho devido a sua popularidade, embora existam muitos modelos para esse tipo de dados.

5.4.1.1 Distribuição beta

Uma Variável Aleatória (VA) X cujos valores possíveis estão no intervalo unitário tem distribuição beta se sua f.d.p. pode ser escrita como:

$$Be(x|\nu_1,\nu_2) = \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} x^{\nu_1-1}(1-x)^{\nu_2-1}, \tag{5.1}$$

com $0 < x < 1$ e zero, caso contrário. Nessa densidade ν_1 e ν_2 $(\nu_1 > 0, \nu_2 > 0)$ são os parâmetros do modelo e $\frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} = \int_0^1 x^{\nu_1-1}(1-x)^{\nu_2-1}dx$ é a função beta.

A esperança e a variância da v.a. X com distribuição beta com parâmetros ν_1 e ν_2 são dadas, respectivamente, por

$$E[X] = \frac{\nu_1}{\nu_1+\nu_2} \quad \text{e} \quad Var[X] = \frac{\nu_1\nu_2}{(\nu_1+\nu_2)^2(\nu_1+\nu_2+1)}.$$

A fim de permitir a inclusão de variáveis explicativas na modelagem da média da distribuição beta, pode ser assumida as seguintes relações: $\nu_1 = \mu\phi$ e $\nu_2 = (1-\mu)\phi$, em que $\mu = E[Y]$ e $\phi = \frac{\mu}{\nu_1}$, parâmetro que controla a dispersão da distribuição. Com isso pode ser obtida uma nova parametrização da distribuição beta dada por:

$$Be(x|\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} x^{(\mu\phi)-1}(1-x)^{((1-\mu)\phi)-1}, \qquad 0 < x < 1.$$

A partir disso, a variância da distribuição beta pode ser escrita da seguinte forma: $Var(X) = \frac{\nu_1\nu_2}{(\nu_1-\nu_2)^2(\nu_1+\nu_2+1)} = \frac{\mu\phi(1-\mu)}{\phi^2(\phi+1)} = \frac{\mu(1-\mu)\phi}{\phi+1}$. Mais detalhes podem ser vistos em Ferrari e Cribari-Neto (2004).

5.4.1.2 Função de verossimilhança

Uma vez observada uma amostra aleatória $x = (x_1, \cdots, x_n)$, de uma variável aleatória X com distribuição beta, a função de verossimilhança pode ser escrita como:

$$L(\mu, \phi) \quad = \quad \prod_{i=1}^{n} Be(x_i|\mu, \phi).$$

O método de verossimilhança é utilizada para estimar os parâmetros do modelo através do método da máxima verossimilhança. O estimador de máxima verossimilhança dos parâmetros μ, ϕ são os valores mais verossímeis que estes parâmetros podem assumir dado o observado na amostra, ou seja, os valores que maximizam a probabilidade de ocorrer a amostra observada.

5.4.2 Modelo de Regressão para resposta binária

Os modelos de regressão binários são aqueles em que a variável resposta pode assumir somente dois valores denotados geralmente por 1 para a ocorrência do evento de interesse ("sucesso") e 0 para a ocorrência do evento complementar ("fracasso"). Ou seja:

$$Y = \begin{cases} 1, & \text{se ocorre sucesso,} \\ 0, & \text{se corre fracasso.} \end{cases}$$

Esse tipo de variável é assumida seguir uma distribuição de Bernoulli. Uma variável aleatória Y segue uma distribuição de Bernoulli com probabilidade de sucesso p se sua função de probabilidade é dada por:

$$p(y) = \quad p^y(1-p)^{1-y}, \quad \text{se} \;\; y = 0, 1$$

e zero caso contrário. Usa-se a notação: $Y \sim Bernoulli(p)$ para afirmar que a variável aleatória Y segue a distribuição de Bernoulli com parâmetro p.

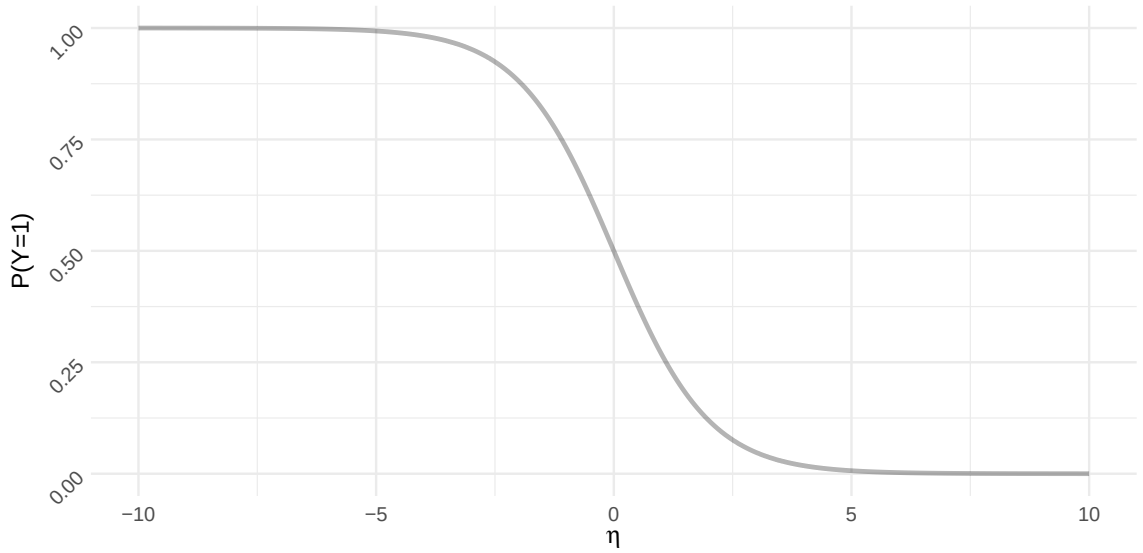
A variável resposta está comumente associada a outras variáveis, que podem ser contínuas, discretas ou categóricas. Consideramos que a probabilidade de sucesso possa ser explicada por estas outras variáveis, denominadas variáveis explicativas ou co-variáveis.

5.4.2.1 Regressão logística

O modelo de regressão logístico é empregado quando a variável resposta é binária. O modelo recebe esse nome devido a função que liga os preditores (que é um número real) a probabilidade de ocorrência do sucesso (que é um número entre 0 e 1), de uma variável binária y, ser a função logística. A função logística com valor máximo 1 é dada por:

$$p(\eta) = P(y=1) \quad = \quad \frac{e^{\eta}}{1+e^{\eta}} = \frac{1}{1+e^{-\eta}}, \ \ \eta \in \mathbb{R}. \tag{5.2}$$

Figura 5 – Curva da função logística descrevendo a probabilidade da resposta y assumir o valor 1 para um número real.



Fonte: elaborada pela autora (2021).

A Figura 5 mostra a curva da função logística com valor máximo igual a 1. A inversa da função logística é denominada função logito e é dada por:

$$\eta(p) \quad = \quad \log(\frac{p}{1-p}), \ \ 0 \leq p \leq 1. \tag{5.3}$$

A fim de definir o modelo de regressão logístico, considere $y_1, \cdots, y_n$ sendo n observações das variáveis $Y_1, \cdots, Y_n$, respectivamente e com $n > 1$, tal que $Y_i \in \{0,1\}$ para $i = 1, \cdots, n$. Assim define-se:

$$\begin{aligned} Y_i &\sim Bernoulli(p_i), \text{ com} \\ logit(p_i) &= \eta_i, \end{aligned} \tag{5.4}$$

em que $Bernoulli(p)$ representa a distribuição de Bernoulli com probabilidade de sucesso; $logit(.)$ é a função logito;

$$\begin{aligned}
\eta_i &= x_i^T\beta \text{ é dito ser o preditor linear do modelo;} \\
x_i &= (x_{i1}, x_{i2}, ..., x_{ip})^T \text{ é um vetor de variáveis independentes fixadas;} \\
\beta &= (\beta_0, \beta_1, ..., \beta_{p-1})^T \text{ são os coeficientes da regressão e} \\
p_i &= \frac{e^{\eta_i}}{1+e^{\eta_i}} \text{ é a probabilidade de } Y_i \text{ assumir o valor 1 (probabilidade de sucesso).}
\end{aligned}$$

5.4.2.2 Função de Verossimilhança da Regressão Logística

Após serem observados os dados, pode-se escrever a função de verossimilhança como sendo o produtório da função de probabilidade aplicada em cada observação, ou seja:

$$\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} \\
&= \prod_{i=1}^{n} \left(\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)^{y_i} \left(\frac{1}{1+e^{\eta_i}}\right)^{1-y_i} \\
&= \frac{\exp\{\sum_{i=1}^{n}\eta_i y_i\}}{\prod_{i=1}^{n}(1+\exp\{\eta_i\})}.
\end{aligned}$$

O modelo foi ajustado a partir da função *glm* disponível no pacote *lme4*, que realiza o ajuste de um modelo linear generalizado utilizando o método da máxima verossimilhança.

5.4.2.3 Modelo de Regressão Logístico Misto

Muitas vezes, em problemas práticos, a população de interesse possui subgrupos que podem ser identificados. Como, por exemplo, uma transportadora atender diversos estabelecimentos periodicamente. Nesse sentido, as entregas realizadas para um dado estabelecimento possuem um efeito comum, que corresponde a amostras de atendimento a um mesmo estabelecimento em diferentes intervalos de tempo. Assim, podem ser identificados subgrupos de entregas feitas a um mesmo estabelecimento. Nesse caso, assumir independência entre todas as entregas controladas pela empresa de transporte não seria razoável. Deve-se, no entanto, considerar a dependência existente entre as entregas a um mesmo estabelecimento e uma possível independência entre os estabelecimentos. Se o interesse é investigar se entregas são ou não realizadas, tem-se uma resposta dicotômica, sugerindo que um modelo legístico pode ser apropriado. Para levar em consideração os subgrupos, pode ser pensado um modelo de regressão logístico misto. Para definir o modelo de regressão logístico misto, considere a variável Y_{ij} uma resposta dicotômica ($Y_{ij} = 1$ para o sucesso e $Y_{ij} = 0$ para o fracasso), em que cada $i = 1, \cdots, m$ indexa o i-ésimo sub-

grupo e cada $j = 1, \cdots, n_i$ indexa a j-ésima observação dentro do i-ésimo subgrupo da população de interesse. Assim $Y_{i1}, \cdots, Y_{in_i}$ estão sujeitos ao mesmo efeito. Como $i = 1, \cdots, m$, existem m efeitos a serem considerados. Seja o modelo:

$$\begin{aligned} Y_{ij}|\cdots &\sim Bernoulli(p_{ij}), \text{ com} \\ logit(p_{ij}) &= \eta_{ij}, \end{aligned} \tag{5.5}$$

em que $|\cdots$ representa todos os parâmetros do modelo e todas as variáveis, explicativas e latentes;

$\eta_{ij} = \mathbf{x}_{ij}^T\beta + z_{ij}^T\mathbf{b}_i$ é o preditor linear;
$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, ..., x_{ijp})^T$ é um vetor de variáveis explicativas associadas aos efeitos fixos (globais);
$\beta = (\beta_0, \beta_1, ..., \beta_{p-1})^T$ são os coeficientes associados aos efeitos fixos da regressão;
$\mathbf{z}_{ij} = (z_{ij1}, z_{ij2}, ..., z_{ijq})^T$ é um vetor de variáveis explicativas associadas aos efeitos aleatórios (locais);
$\mathbf{b}_i = (b_0, b_1, ..., b_{q-1})^T$ são os coeficientes associados aos efeitos aleatórios (variáveis latentes) e
$p_{ij} = \frac{e^{\eta_{ij}}}{1+e^{\eta_{ij}}}$ é a probabilidade de Y_{ij} assumir o valor 1 (probabilidade de sucesso).

O modelo foi ajustado a partir da função *glmer* presente no pacote *lme4*, que realiza o ajuste de um modelo misto linear generalizado a partir da incorporação dos parâmetros de efeitos fixos e aleatórios em um preditor linear utilizando a máxima verossimilhança.

6 RESULTADOS

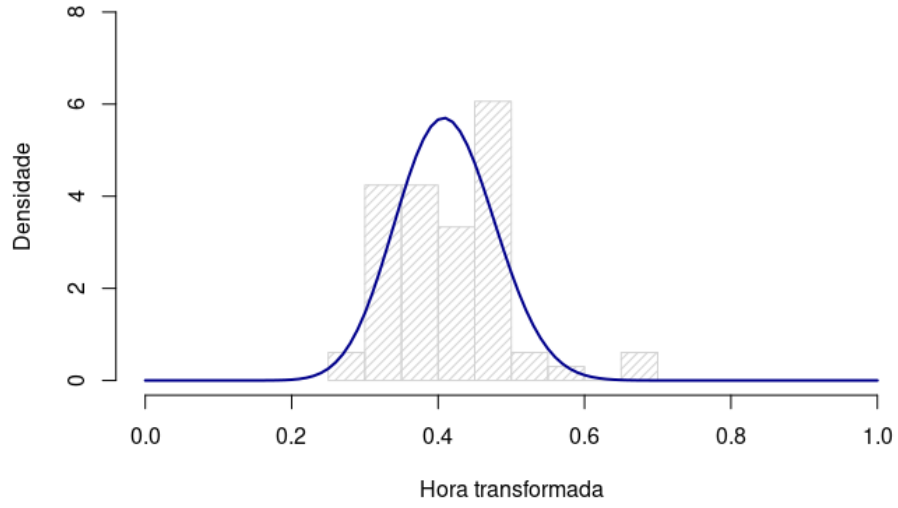
Este capítulo expõe os resultados dos modelos utilizados nesta monografia. Na Seção 6.1 é apresentado o resultado obtido através do modelo de regressão para resposta limitada: regressão beta, que tem como objetivo estimar um intervalo de tempo que maximize a probabilidade de sucesso de uma entrega. Enquanto a Seção 6.2 apresenta os resultados obtidos através dos modelos de regressão para resposta binária: regressão logística e regressão logística mista com o intuito de prever a ocorrência de uma entrega mal-sucedida.

6.1 Análise do modelo de resposta limitada

O algoritmo de regressão para resposta limitada proposto utiliza apenas os registros de sucesso de entrega para modelagem. Para obter um intervalo de tempo que maximize a probabilidade de sucesso de uma entrega é assumindo que as entregas seguem uma distribuição beta e seus parâmetros são estimados conforme descrito na seção 5.4.1.

A Figura 6 apresenta a densidade estimada para um estabelecimento por meio da estimativa dos parâmetros da distribuição das entregas em um estabelecimento. A partir disso, é realizado o cálculo do intervalo de confiança, que fornece o intervalo de tempo, com os respectivos horários de abertura e fechamento da janela de atendimento.

Figura 6 – Densidade estimada para um estabelecimento



Fonte: elaborado pela autora (2021).

Em conhecimento dos parâmetros estimados da distribuição, é obtido a janela de

atendimento que maximiza a probabilidade de sucesso de entrega.

Para verificar se a janela de atendimento estimada está dentro dos intervalos que são disponibilizados pela empresa, é realizada uma comparação entre as mesmas. Essa verificação consiste em avaliar se a janela estimada está contida em algumas das janelas fornecidas. Em decorrência de obter-se uma janela não adequada as ofertadas pela empresa, elas são ajustadas novamente, de acordo com um critério de proximidade.

A janela previamente estabelecida mais próxima é definida como aquela que menos realiza alterações na janela estimada. Dessa forma, esse novo ajuste comprime ou expande a janela estimada para fins de adequação. Como resultado, é obtido uma janela de atendimento apropriada dado os critérios estabelecidos pela empresa.

6.2 Análise dos modelos de resposta binária

Para prever a ocorrência de entregas mal-sucedidas devido a inconsistências na janela de atendimento cadastrada, são utilizados os métodos descritos na Seção 5.4.2.

6.2.1 Conformidade das entregas

Como apresentado na seção 5.3.1, podemos medir se uma entrega foi ou não realizada respeitando os horários cadastrados no sistema. A Tabela 3 apresenta a quantidade de entregas conformes e não conformes para entregas realizadas e não realizadas a partir da base utilizada para esta monografia.

Tabela 3 – Tabela de não conformidade das entregas

	Não conforme	**Conforme**	**Total**
Não entregue	1047	784	1831
Entregue	127401	242959	370360
Total	128448	243743	372191

Fonte: Elaborado pela autora (2022).

A partir dessa tabela, temos que, das entregas realizadas com sucesso, 34.4% corresponde a entregas fora da janela de atendimento do estabelecimento e 64.4% a entregas dentro da janela de atendimento cadastrada. Das entregas não realizadas 57.2% reflete a entregas sem conformidade de janela de atendimento e 42.8% refere-se com conformidade de janela de atendimento. Assim, podemos observar que dentre as entregas não realizadas, existe uma

maior incidência de não conformidade, evidenciando que tentativas de entrega fora da janela de atendimento cadastrada resultam, muitas vezes, em insucesso de entrega.

É possível observar que a não conformidade está agravando a ocorrência de entregas mal sucedidas. Temos, analisando a tabela 3, que a taxa de não entregue e não conforme é de 0.82% enquanto que a taxa de não entregue e conforme é de apenas 0.32%. Dessa forma, é identificado a necessidade de que as janelas de atendimento cadastradas sejam respeitadas para diminuir a incidência de entregas mal sucedidas.

6.2.2 Correlação entre as entregas

A base de dados utilizada para desenvolvimento desta monografia apresenta uma quantidade n de atendimentos registrados para cada estabelecimento, onde $n >= 1$. Esses atendimentos em um mesmo estabelecimento são replicações de um mesmo indivíduo (estabelecimento). Essas replicações possuem correlação, sendo necessário inserir o efeito dos estabelecimentos sobre as entregas na modelagem. Para isso, a população é dividida em duas amostras.

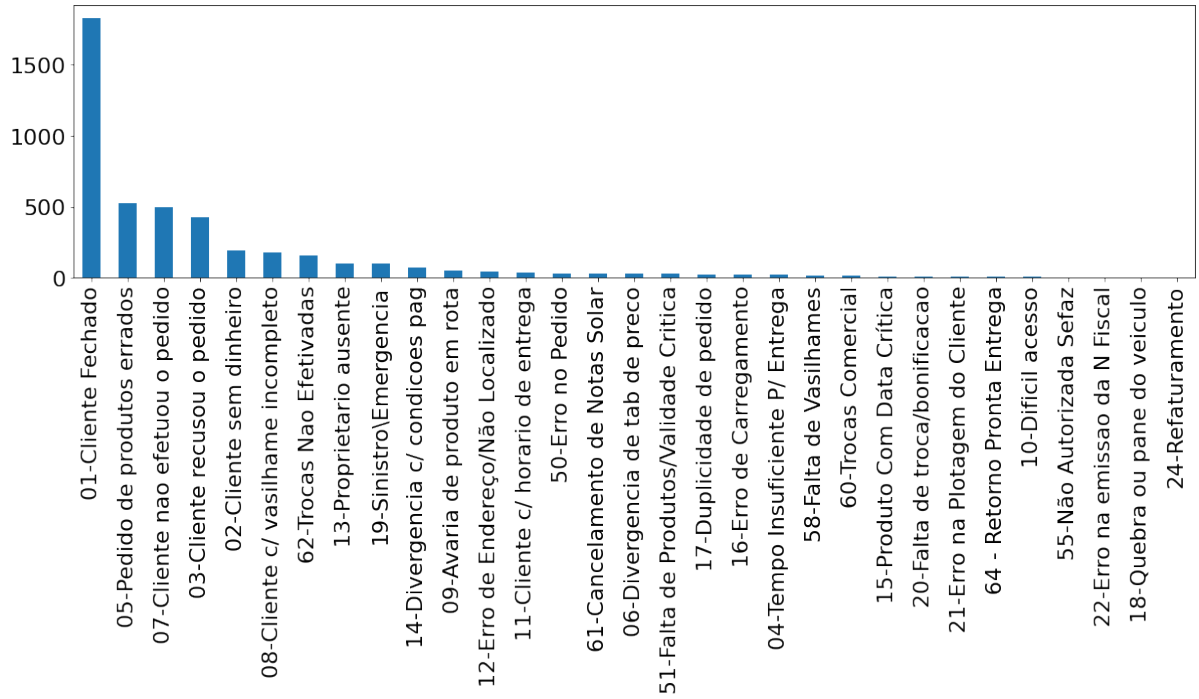
A primeira amostra extraída da população corresponde aos dados utilizados para o modelo de regressão logístico, descrito na seção 5.4.2.1 e contêm apenas os estabelecimentos que possuem menos de 4 registros de entregas, ou seja, estabelecimentos com pouco volume de entregas ou novos estabelecimentos registrados no intervalo de tempo de estudo. Na segunda amostra estão contidos todos os estabelecimentos que possuem pelo menos 4 atendimentos registrados no período e, estes, são modelados utilizando o modelo de regressão logístico misto descrito na seção 5.4.2.4.

6.2.3 Variável resposta

Na definição da variável resposta foi analisado os motivos de insucesso de entrega que sugerissem erro na atual janela de atendimento utilizada pelo algoritmo de roteirização.

Na Figura 7 é possível observar o motivo da entrega ter sido mal-sucedida e a respectiva frequência de ocorrências. Dentre os motivos listados "Cliente Fechado" possui maior frequência. Além disso, esse motivo está diretamente relacionado ao alvo deste estudo, visto que queremos minimizar a probabilidade de insucesso de entregas por razão do cliente não se encontrar no estabelecimento ou estar fora do horário em que ele aceita entregas no momento da tentativa de entrega.

Figura 7 – Motivos de insucesso na entrega



Fonte: elaborado pela autora (2021).

Como descrito na Seção 5.4.2 a variável resposta é definida como:

$$Y = \begin{cases} 1, & \text{se a entrega não é realizada,} \\ 0, & \text{caso contrário,} \end{cases}$$

em que $Y =$ "tentativa de entrega" representa a resposta para o problema abordado nesta monografia.

6.2.4 Variáveis explicativas

Para investigar o problema apresentado, foi selecionado, a partir de discussões sobre os dados disponibilizados, um conjunto de variáveis que podem ser identificadas como explicativas durante a fase de modelagem. A Tabela 4 apresenta esse conjunto de variáveis com suas respectivas descrições.

Dentre as variáveis apresentadas, com exceção de *twOpen*, *twClose* e *StoreId*, todas foram obtidas na fase de Engenharia de Características. A variável *normHour* é um horário

Tabela 4 – Variáveis investigadas como explicativas.

Variável	Descrição	Tipo
twOpen	Horário inicial da janela cadastrada.	datetime
twClose	Horário final da janela cadastrada.	datetime
normHour	Horário de chegada no estabelecimento padronizado.	float
Distance	Distância do centro de distribuição até o estabelecimento.	float
timeUntilService	Tempo de duração (em minutos) do percurso do CD ao estabelecimento.	int
dayName	Dia da semana que foi realizada a tentativa de entrega.	categorical
serviceTime	Duração (em minutos) do tempo de atendimento no estabelecimento.	int
conformity	Verificação do horário de chegada de acordo com janela de atendimento cadastrada.	bool
storeId	Identificador único do estabelecimento	int

Fonte: Elaborado pela autora (2022).

estimado pelo algoritmo de roteirização e, dessa forma, conseguimos obter essa informação para execução do modelo antes que a rota seja iniciada.

6.2.5 Modelo de Regressão Logístico

O evento definido como de interesse é dado pela ocorrência de uma entrega mal sucedida. Com o objetivo de prever a ocorrência deste evento são ajustados três modelos de Regressão Logística com diferentes amostras de dados. As amostras são divididas da seguinte forma:

- Apenas entregas sinaladas como conformes;
- Apenas entregas sinaladas como não conformes;
- Todas as entregas.

O propósito de analisar os resultados com diferentes amostras segundo a conformidade de janela de atendimento, deve-se a hipótese de que a não conformidade da janela de atendimento é uma das causas que levam ao insucesso de uma entrega.

A Tabela 5 apresenta a quantidade de entregas utilizada para modelagem. Temos, na primeira coluna, a quantidade de entregas que foram realizadas dentro da janela de atendimento do estabelecimento e então, na segunda coluna, a quantidade de entregas que foram realizadas sem respeitar a restrição de horário de entrega cadastrada.

Tabela 5 – Quantidade de entregas: Modelo de Regressão Logístico

	Conformes	**Não conformes**	**Total**
Não entregues	9219	2622	11841
Entregues	86	79	165

6.2.5.1 Análise da significância dos coeficientes de regressão

Para analisar o conjunto de dados com melhor performance em prever a ocorrência de uma entrega mal sucedida, são apresentados os resultados dos três distintos modelos de Regressão Logística. As Tabelas 6, 8 e 10 apresentam os coeficientes de regressão estimados para cada modelo desenvolvido.

Os coeficientes de regressão, apresentados na forma logarítmica, auxiliam na interpretabilidade do modelo. Existe maior chance de ocorrência de uma entrega mal sucedida conforme eleva-se o valor da variável preditora, caso o coeficiente seja maior do que zero. Caso contrário, onde o coeficiente é menor do que zero, temos uma menor chance de ocorrência do evento de interesse seguindo a elevação da variável preditora.

O modelo desenvolvido utilizando apenas as entregas realizadas dentro da janela de atendimento cadastrada possui resultados apresentados nas Tabelas 6 e 7. Temos, para este modelo, uma maior incidência de entregas mal sucedidas conforme o crescimento da variável *normHour*, ou seja, quanto mais tarde for o horário da tentativa de entrega maior será a probabilidade de que a entrega seja mal sucedida. Obtemos então, a partir da interpretação dos coeficientes de regressão, que as tentativas de entrega realizadas próximas ou durante o final do horário comercial padrão e aos sábados são mais suscetíveis a serem mal sucedidas. Além disso, janelas de atendimento com maiores intervalos de tempo tem menor probabilidade de ocorrência de insucesso de entrega.

A significância estatística das variáveis é apresentada pela coluna $Pr(>|z|)$. Os resultados apresentados na Tabela 6 mostram que a variável *twClose* não possui significância estatística, dado que possui valor superior ao p-valor, sendo $p < 0,05$. Para as demais variáveis, obtemos valores inferiores ao p-valor, mostrando que possuem importância para o modelo de regressão proposto.

A Tabela 7 expõe a matriz de confusão, ou também denominada tabela de confusão, obtida a partir da execução do modelo apenas com dados conformes na base de teste. O objetivo dessa matriz é exibir a distribuição das classificações em termos de suas classes atuais e de suas classes previstas. Na diagonal principal (os quadrantes (1, 1) e (2, 2)) podemos observar os

Tabela 6 – Coeficientes de regressão: amostra conforme

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-9,115527	1,7464235	-5,219540	0,0000002
normHour	10,590145	1,7658294	5,997264	0,0000000
twOpen	3,596527	1,0175938	3,534344	0,0004088
twClose	-4,126700	2,4955746	-1,653607	0,0982074
dayName_Saturday	0,725026	0,2973607	2,438204	0,0147604

Fonte: Elaborado pela autora (2022).

acertos do modelo e na diagonal secundária (os quadrantes (2, 1) e (1, 2)) vemos os erros do modelo.

A partir dos resultados apresentados na Tabela 7, é possível observar que o modelo classificou como bem sucedida um total de 2726 entregas e, dentre essas, houveram 19 entregas que, na realidade, eram entregas mal sucedidas. A partir disso, obtemos que 99% do total de entregas classificadas como bem sucedidas de fato são tal. Além disso, temos que, das entregas bem sucedidas tiveram um total de 103 entregas que foram classificadas como mal sucedidas, ou seja, do total de 2810 entregas bem sucedidas 96% foi classificada como tal. Por meio dessas visualizações sobre o comportamento do modelo em relação a entregas bem sucedidas, é possível constatar que ele, de maneira geral, é capaz de identificar uma entrega bem sucedida a partir das variáveis explicativas, dado que apenas uma pequena percentagem de entregas bem sucedidas é erroneamente classificada como mal sucedida. As entregas mal sucedidas erroneamente classificadas como bem sucedidas possuem menor impacto no modelo, dado que são entregas que sem a utilização do modelo seriam, de todo modo, mal sucedidas.

Para a classificação de entregas mal sucedidas, ainda na Tabela 7, temos que do total de entregas classificadas como mal sucedidas, apenas $0,06\%$ são de fato mal sucedidas, visto que existem 103 entregas bem sucedidas classificadas como mal sucedidas. No entanto, do total de entregas mal sucedidas temos 27% das entregas corretamente classificadas como mal sucedidas. A partir dos valores apresentados e discutidos, para este modelo, os resultados indicam que apenas aproximadamente 4% do total de entregas bem sucedidas é classificada como mal sucedida e 27% das entregas mal sucedidas são classificadas como tal.

Tabela 7 – Matriz de confusão: amostra conforme

	Entregue	**Não entregue**
Pred. Entregue	2707	19
Pred. Não Entregue	103	7

Fonte: Elaborado pela autora (2022).

A interpretação dos coeficientes de regressão é a mesma apresentada utilizando apenas a amostra com dados conformes, conforme pode ser observado nas Tabelas 8 e 10. No entanto, temos diferentes resultados quanto a significância das variáveis e resultados apresentados na matriz de confusão.

Para o experimento modelado utilizando apenas as entregas que foram realizadas fora da janela de atendimento cadastrada, o resultado do modelo é apresentado na Tabela 8. Temos, conforme exposto, as variáveis *twOpen* e *twClose* sem significância estatística. A matriz de confusão obtida a partir desse modelo classifica corretamente 14 de 23 entregas mal sucedidas, enquanto que classifica 703 de 829 entregas bem sucedidas corretamente, conforme pode ser observado na Tabela 9.

Tabela 8 – Coeficientes de regressão: amostra não conforme

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-7,2483109	1,3567399	-5,3424468	0,0000001
normHour	6,0127359	1,1356237	5,2946551	0,0000001
twOpen	1,0545494	2,1252922	0,4961903	0,6197601
twClose	-0,9979870	2,0354744	-0,4902970	0,6239238
dayName_Saturday	0,9786569	0,3162595	3,0944744	0,0019716

Fonte: Elaborado pela autora (2022).

Tabela 9 – Matriz de confusão: amostra não conforme

	Entregue	**Não entregue**
Pred. Entregue	703	9
Pred. Não Entregue	126	14

A população foi utilizada, por fim, neste último experimento e o modelo resultante é apresentado na Tabela 10. Nota-se que há significância estatística na utilização de todas as variáveis para o modelo, mostrando que são variáveis com importância para o modelo de regressão desenvolvido. A matriz de confusão deste modelo é apresentada na Tabela 11.

Tabela 10 – Coeficientes de regressão: irrestrito

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-7,949893	1,0060492	-7,902092	0,0000000
normHour	7,799502	0,8181244	9,533394	0,0000000
twOpen	4,972259	0,8856706	5,614117	0,0000000
twClose	-4,518578	1,1706947	-3,859741	0,0001135
dayName_Saturday	1,063581	0,2189775	4,857033	0,0000012

Fonte: Elaborado pela autora (2022).

Tabela 11 – Matriz de confusão: irrestrito

	Entregue	**Não entregue**
Pred. Entregue	3399	31
Pred. Não Entregue	239	18

As frequências de classificação para cada classe dos modelos desenvolvidos apresentadas nas Tabelas 7, 8 e 11, são utilizadas para calcular as métricas de validação de modelo apresentadas na Tabela 12, onde podemos identificar a qualidade dos modelos desenvolvidos.

De acordo com Castro e Ferrari (2017), em problemas de classificação binária, predições podem assumir quatro possíveis classes:

- **Verdadeiro Positivo (VP):** classificada como positiva e, de fato, positiva;
- **Verdadeiro Negativo (VN):** classificada como negativa e, de fato, negativa;
- **Falso Positivo (FP):** classificada como positiva, mas, na realidade, negativa;
- **Falso Negativo (FN):** classificada como negativa, mas, na realidade, positiva.

A sensibilidade, também conhecida como *recall*, representa a proporção de entregas bem sucedidas identificadas corretamente, ou seja, avalia a capacidade do modelo de detectar com sucesso resultados classificados como positivos. Ela pode ser obtida através da equação:

$$sensibilidade = \frac{VP}{VP+FN} \tag{6.1}$$

A especificidade avalia a capacidade do classificador de detectar resultados negativos. Pode ser calculada a partir da equação:

$$especificidade = \frac{VN}{VN+FP} \tag{6.2}$$

O valor preditivo positivo representa a proporção de verdadeiros positivos entre todos as entregas classificadas como bem sucedidas. Pode ser obtido através da equação:

$$VPP = \frac{VP}{VP+FP} \tag{6.3}$$

O valor preditivo negativo consiste na proporção de verdadeiros negativos entre todas entregas classificadas como mal sucedidas. É obtido a partir da equação:

$$VPN = \frac{VN}{FN+VN} \tag{6.4}$$

Em decorrência do desbalanceamento dos dados, a métrica acurácia balanceada é utilizada. Essa métrica não é influenciada pelo desbalanceamento das classes, dado que sua equação ocorre em função da taxa de verdadeiros positivos e verdadeiros negativos, como demonstrado na equação (6.5).

$$Acurcia\,Balanceada = \frac{1}{2}\left(\frac{VP}{VP+FN} + \frac{VN}{VN+FP}\right) \tag{6.5}$$

Tabela 12 – Performance dos modelos.

	Conforme	**Não conforme**	**Soma**
Sensibilidade	0,96	0,85	0,93
Especificidade	0,27	0,67	0,37
Valor Pred. Pos.	0,99	0,99	0,99
Valor Pred. Neg.	0,06	0,10	0,07
Acurácia balanceada	0,62	0,73	0,65

A partir dos resultados apresentados na Tabela 12 é possível concluir que o modelo com melhor performance é aquele que utiliza apenas as entregas que foram realizadas fora da janela de atendimento cadastrada. Neste modelo, temos 10% de entregas mal sucedidas sendo classificadas corretamente sem a ocorrência de um grande volume de entregas bem sucedidas sendo classificada incorretamente, dada a especificidade calculada de $0,67\%$ e a observação, a partir da matriz de confusão, de apenas 9 entregas bem sucedidas serem classificadas como mal sucedidas.

6.2.6 Modelo de Regressão Logístico Misto

Utilizando o Modelo de Regressão Logístico Misto foram desenvolvidos três experimentos. Serão analisados, assim como na Seção 6.2.5 :

- Apenas entregas sinaladas como conformes;
- Apenas entregas sinaladas como não conformes;
- Todas as entregas.

Para tratar o desbalanceamento dos dados, onde uma entrega mal sucedida ocorre em, aproximadamente, uma vez a cada 215 ocorrências de entregas bem sucedidas e necessidade de minimizar o esforço computacional, foi colhida uma amostra aleatória que corresponde a 15% da população. A Tabela 13 apresenta a quantidade total de entregas realizadas e não realizadas utilizadas para modelagem.

Tabela 13 – Quantidade de entregas: Modelo de Regressão Logístico Misto

	Conformes	**Não conformes**	**Irrestrito**
Não entregues	15011	20987	51943
Entregues	698	968	1666

Fonte: Elaborado pela autora (2022).

6.2.6.1 Análise da significância dos coeficientes de regressão

O Modelo de Regressão Logístico Misto, como apresentado na Seção 5.4.2.3, possui dois vetores de variáveis, diferentemente do modelo de Regressão Logístico já apresentado. Temos, neste modelo misto, um vetor para as variáveis associadas aos efeitos fixos globais e um outro vetor associado aos efeitos aleatórios locais.

O vetor de variáveis **x** compreende as variáveis que influenciam as entregas entre o estabelecimento, enquanto que o vetor de variáveis **z** possui as variáveis que possuem efeito em cada estabelecimento. Foi definido para **x**, o conjunto de variáveis apresentado na Tabela 4 com exceção de *storeId*. Para o conjunto de variáveis **z**, foram definidas as variáveis *twOpen*, *twClose* e *storeId*.

As Tabelas 14, 15 e 16 apresentam os coeficientes de regressão estimados para os modelos desenvolvidos utilizando as três amostras selecionadas. Observando essas tabelas é possível identificar o diferente comportamento das variáveis de acordo com a amostra utilizada para modelagem.

O modelo desenvolvido utilizando o conjunto de entregas realizadas fora da janela de atendimento e demonstrado na Tabela 15 possui coeficientes de regressão positivo para as variáveis *dayName_Tuesday* e *dayName_Thursday* sinalizando uma maior probabilidade de entrega mal sucedidas nestes dias da semana. Por outro lado, na Tabela 14 e 16 temos as mesmas variáveis com coeficientes negativos, apontando para uma interpretação inversa.

A significância estatística foi identificada nas variáveis *twOpen*, *twClose*, *dayName_Monday* e *dayName_Saturday* nos modelos utilizando apenas as entregas conformes e utilizando apenas as não conformes. Em contrapartida, para o modelo desenvolvido com a população, todas as variáveis utilizadas possuem valores de significância de aproximadamente 10^{-7} fornecendo evidências de possuírem importância para o modelo desenvolvido.

As frequências de classificação para cada classe dos modelos desenvolvidos são apresentadas nas Tabelas 17, 18 e 19. A partir dessas frequências são obtidas as métricas apresentadas na Tabela 20, onde podemos identificar a qualidade dos modelos.

O resultado das métricas apresentadas na Tabela 20 demonstra que, assim como na

Tabela 14 – Coeficientes de regressão modelo logístico misto: amostra conforme

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
normHour	2,3225221	1,5598437	1,488945	0,365017
twOpen	14,5707374	1,8199501	8,006119	0,0000000
twClose	-16,7879784	1,3906898	-12,071692	0,0000000
dayName_Friday	-0,1549499	0,1890096	-0,819799	0,4123307
dayName_Monday	0,5683815	0,2452074	2,317962	0,0204514
dayName_Saturday	0,6760766	0,1930920	3,501319	0,0004630
dayName_Thursday	-0,2471634	0,1939337	-1,274473	0,2024957
dayName_Tuesday	-0,2171779	0,2029680	-1,070010	0,2846147

Fonte: Elaborado pela autora (2022).

Tabela 15 – Coeficientes de regressão modelo logístico misto: amostra não conforme

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
normHour	-0,2413279	0,4084770	-0,5907992	0,5546550
twOpen	6,6759937	2,1222633	3,1456953	0,0016569
twClose	-10,8907096	1,8112971	-6,0126577	0,0000000
dayName_Friday	-0,2258397	0,1842086	-1,2260003	0,2201986
dayName_Monday	0,8301142	0,2102511	3,9482033	0,0000787
dayName_Saturday	0,5309711	0,1630836	3,2558221	0,0011306
dayName_Thursday	0,0222237	0,1790155	0,1241441	0,9012012
dayName_Tuesday	0,1354738	0,1760324	0,7695960	0,4415396

Fonte: Elaborado pela autora (2022).

Tabela 16 – Coeficientes de regressão modelo logístico misto: irrestrito

	Estimate	**Std. Error**	**z value**	**Pr(|z|)**
normHour	5,021903	0,3937227	12,754924	0
twOpen	8,540907	1,1288720	7,565878	0
twClose	-7,868901	1,1962010	-6,578243	0
dayname_Friday	-5,851790	0,4750391	-12,318543	0
dayname_Monday	-4,924890	0,4721184	-10,431471	0
dayname_Saturday	-5,126264	0,4646378	-11,032817	0
dayname_Thursday	-5,708432	0,4705095	-12,132447	0
dayname_Tuesday	-5,656693	0,4712280	-12,004153	0
dayname_Wednesday	-5,987831	0,4736597	-12,641632	0

Fonte: Elaborado pela autora (2022).

Tabela 17 – Matriz de confusão modelo logístico misto: amostra conforme

	Entregue	**Não entregue**
Pred. Entregue	4202	104
Pred. Não Entregue	577	169

Tabela 18 – Matriz de confusão modelo logístico misto: amostra não conforme

	Entregue	**Não entregue**
Pred. Entregue	6288	187
Pred. Não Entregue	431	144

Tabela 19 – Matriz de confusão modelo logístico misto: irrestrito

	Entregue	**Não entregue**
Pred. Entregue	15565	325
Pred. Não Entregue	780	239

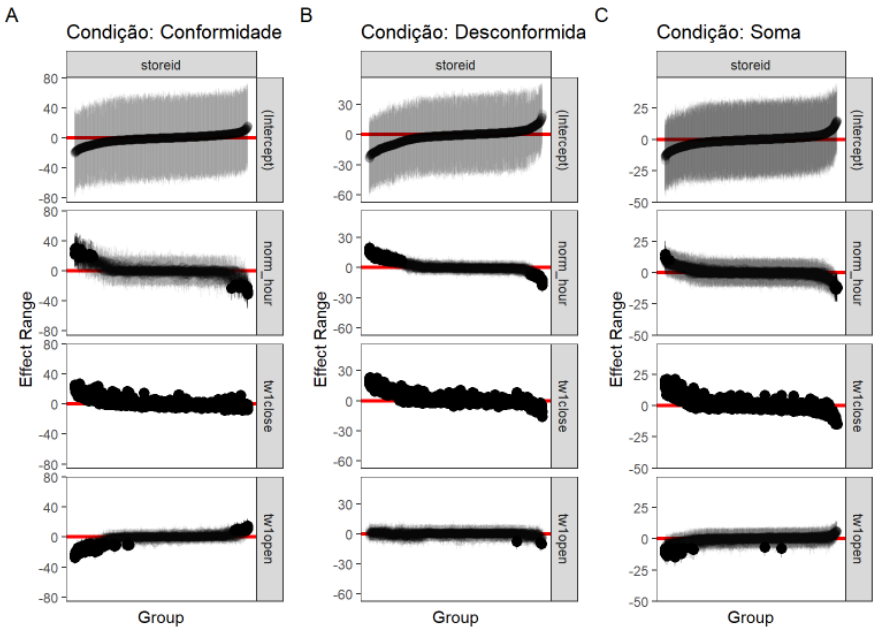
Seção 6.2.5, o modelo que obteve melhores resultados foi o que utiliza apenas as amostras não conformes. No entanto, o modelo de regressão logístico misto constata melhores resultados. Este evento, compreende que, deve-se ao fato de uma não entrega poder ocorrer de forma diferente em estabelecimentos distintos, destacando a importância do conjunto de variáveis que modelam o efeito em cada estabelecimento.

Tabela 20 – Performance dos modelos: modelo logístico misto

	Conforme	**Não conforme**	**Irrestrito**
Sensibilidade	0,88	0,94	0,95
Especificidade	0,62	0,44	0,42
Valor Pred. Pos.	0,98	0,97	0,98
Valor Pred. Neg.	0,23	0,25	0,23
Acurácia balanceada	0,75	0,69	0,69

Os efeitos aleatórios, isto é, os coeficientes associados as variáveis em $\mathbf{z}$, das variáveis para os três experimentos desenvolvidos pode ser estudado através da Figura 8. Para os três experimentos apresentados, existem coeficientes que podem não ser considerados 0, visto que, pode ser observado alguns intervalos de confiança, representados no gráfico pelos segmentos da reta vertical, não contendo o valor 0. A partir disso, podemos constatar a evidência da existência da influência dos estabelecimentos.

Figura 8 – Efeito dos fatores para os estabelecimentos



Fonte: Elaborado pela autora (2022).

7 CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo são descritas as considerações e lições aprendidas a respeito da metodologia desenvolvida, para estudo das janelas de atendimento, assim como resultados obtidos.

7.1 Considerações gerais

Neste trabalho foi proposta uma metodologia para minimizar a probabilidade de insucesso de uma entrega quando associada a problemas na janela de atendimento do estabelecimento. Para desenvolver a proposta foram analisadas diversas técnicas e trabalhos relacionados, visando minimizar os desafios encontrados neste tipo de problema do mundo real, aplicado em empresas de logística.

Foi apresentado duas soluções para o problema. A primeira solução consiste em, a partir dos dados históricos de entregas bem sucedidas, obter um intervalo de tempo que maximize a probabilidade de sucesso de uma entrega. Para essa solução não há maneira de validação dos parâmetros estimados e o intervalo estimado pode ser restritivo, podendo ocasionar a necessidade de alocar mais recursos para concluir a rota.

A segunda solução tem como objetivo utilizar os dados históricos para prever o acontecimento de uma entrega mal sucedida devido a janela de atendimento. Para o modelo de regressão logístico, o experimento com melhor resultado corresponde a amostra com entregas realizadas apenas fora da janela de atendimento cadastrada, obtendo $sensibilidade = 0,85$, $especificidade = 0,67$ e $VPN = 0,10$. Para o modelo de regressão logístico misto, o melhor resultado também foi obtido utilizando apenas as entregas não conformes com $sensibilidade = 0,94$, $especificidade = 0,44$ e $VPN = 0,25$. No entanto, apesar de obter melhores resultados, esse modelo não é útil na prática, uma vez que não é considerado pelo algoritmo que a entrega irá ser realizada fora da janela de atendimento.

Os resultados obtidos demonstram que entregas realizadas fora da janela de atendimento cadastrada possuem maior probabilidade de ocorrência de uma entrega mal sucedida, tornando essa uma evidência de que um dos problemas a serem resolvidos pela empresa de transporte, é a garantia de que a entrega seja realizada dentro da janela de atendimento do estabelecimento a ser atendido.

Por fim, concluímos que é necessário modelar os efeitos de cada estabelecimento e

possuir um histórico de entregas realizado dentro da janela de atendimento cadastrada, para que seja possível os algoritmos identificarem padrões no acontecimento de entregas mal sucedidas levando em consideração as janelas de atendimento.

7.2 Trabalhos futuros

Como sugestão para trabalhos futuros a serem pesquisados:

- Utilização de técnicas estatísticas para balanceamento dos dados;
- Utilização de modelos não lineares;
- Incorporação da incerteza nos parâmetros do modelo, como por exemplo, o tempo para atendimento do estabelecimento, utilizando modelos de programação estocástica ou robusta.

REFERÊNCIAS

AGGARWAL, D.; CHAHAR, V.; GIRDHAR, A. Lagrangian relaxation for the vehicle routing problem with time windows. In: . [S.l.: s.n.], 2017. p. 1601–1606.

AUMANN, R.; SERRANO, R. **An economic index of riskiness**. [S.l.], 2007. Disponível em: <https://EconPapers.repec.org/RePEc:imd:wpaper:wp2007-08>.

BATES, D.; MÄCHLER, M.; BOLKER, B.; WALKER, S. Fitting linear mixed-effects models using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 1–48, 2015.

BERBEGLIA, G.; CORDEAU, J.-F.; GRIBKOVSKAIA, I.; LAPORTE, G. Static pickup and delivery problems: a classification scheme and survey. **TOP**, v. 15, n. 1, p. 1–31, jul 2007. ISSN 1863-8279. Disponível em: <https://doi.org/10.1007/s11750-007-0009-0>.

CASTRO, L. de; FERRARI, D. **Introdução a mineração de dados**. Saraiva Educação S.A., 2017. ISBN 9788547200992. Disponível em: <https://books.google.com.br/books?id=SSlrDwAAQBAJ>.

DANTZIG, G. B.; RAMSER, J. H. The truck dispatching problem. **Management Science**, v. 6, n. 1, p. 80–91, 1959. Disponível em: <https://doi.org/10.1287/mnsc.6.1.80>.

EKSIOGLU, B.; VURAL, A. V.; REISMAN, A. The vehicle routing problem: A taxonomic review. **Computers and Industrial Engineering**, v. 57, n. 4, p. 1472–1483, 2009. ISSN 03608352. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0360835209001405>.

FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, Taylor & Francis, v. 31, n. 7, p. 799–815, 2004.

FISHER, M. L.; JöRNSTEN, K. O.; MADSEN, O. B. G. Vehicle routing with time windows: Two optimization algorithms. **Operations Research**, v. 45, n. 3, p. 488–492, 1997. Disponível em: <https://doi.org/10.1287/opre.45.3.488>.

FISHER, R. **Statistical Methods for Research Workers**. Hafner, 1958. (Biological monographs and manuals). Disponível em: <https://books.google.com.br/books?id=qqNpAAAAMAAJ>.

GENDREAU, M.; TARANTILIS, C. Solving large-scale vehicle routing problems with time windows: The state-of-the-art. In: . [S.l.: s.n.], 2010.

GONZALEZ, M. **CRISP-DM na prática**. 2019. Disponível em: <https://medium.com/matgonz/crisp-dm-na-pr%C3%A1tica-65be0ee92ada>.

GREENMILE. <https://greenmile.com/>, acessado em 29-06-2021. Disponível em: <https://greenmile.com/>.

GUPTA, A. K.; NADARAJAH, S. **Handbook of beta distribution and its applications**. [S.l.]: CRC press, 2004.

KNOWLES, J. E.; FREDERICK, C. **merTools: Tools for Analyzing Mixed Effect Regression Models**. [S.l.], 2020. R package version 0.5.2. Disponível em: <https://CRAN.R-project.org/package=merTools>.

KUHN, M. Building predictive models in r using the caret package. **Journal of Statistical Software, Articles**, v. 28, n. 5, p. 1–26, 2008. ISSN 1548-7660. Disponível em: <https://www.jstatsoft.org/v028/i05>.

MEDEIROS, M. A. **Webshoppers 42: Ecommerce tem a maior alta em 20 anos**. 2020. Disponível em: <https://ecommercedesucesso.com.br/ecommerce-bate-recorde>.

MONTOYA-TORRES, J. R.; López Franco, J.; Nieto Isaza, S.; Felizzola Jiménez, H.; HERAZO-PADILLA, N. A literature review on the vehicle routing problem with multiple depots. **Computers Industrial Engineering**, v. 79, p. 115–129, 2015. ISSN 0360-8352. Disponível em: <https://www.sciencedirect.com/science/article/pii/S036083521400360X>.

PAN, B.; ZHANG, Z.; LIM, A. Multi-trip time-dependent vehicle routing problem with time windows. **European Journal of Operational Research**, Elsevier, v. 291, n. 1, p. 218–231, 2021.

PAZ, R. F. da; BALAKRISHNAN, N.; BAZÁN, J. L. *et al.* L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 33, n. 3, p. 455–479, 2019.

PEARSON, K. Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. **Philosophical Transactions of the Royal Society of London. A**, The Royal Society, v. 186+, p. 343–414, 1895.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>.

RIVO, E.; AGUADO, J. De la fuente; VáZQUEZ, ; GARCíA-FONTáN, E.; CAñIZARES, M.-; GIL, P. Cross-industry standard process for data mining is applicable to the lung cancer surgery domain, improving decision making as well as knowledge and quality management. **Clinical translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico**, v. 14, p. 73–9, 01 2012.

ROSSUM, G. V.; JR, F. L. D. **Python reference manual**. [S.l.]: Centrum voor Wiskunde en Informatica Amsterdam, 1995.

SUN, B.; YANG, Y.; SHI, J.; ZHENG, L. Dynamic pick-up and delivery optimization with multiple dynamic events in real-world environment. **IEEE Access**, IEEE, v. 7, p. 146209–146220, 2019.

SUN, P.; VEELENTURF, L. P.; HEWITT, M.; WOENSEL, T. V. Adaptive large neighborhood search for the time-dependent profitable pickup and delivery problem with time windows. **Transportation Research Part E: Logistics and Transportation Review**, Elsevier, v. 138, p. 101942, 2020.

TEAM, T. pandas development. **pandas-dev/pandas: Pandas**. Zenodo, 2020. Disponível em: <https://doi.org/10.5281/zenodo.3509134>.

YANG, H.; ZHAO, L.; YE, D.; MA, J. Disturbance management for vehicle routing with time window changes. **Operational Research**, v. 20, n. 2, p. 1093–1112, jun 2020. ISSN 1866-1505. Disponível em: <https://doi.org/10.1007/s12351-017-0363-0>.

ZHANG, Y.; BALDACCI, R.; SIM, M.; TANG, J. Routing optimization with time windows under uncertainty. **Mathematical Programming**, v. 175, n. 1/2, p. 263 – 305, 2019. ISSN 00255610. Disponível em: <http://search-ebscohost-com.ez11.periodicos.capes.gov.br/login.aspx?direct=true&db=aph&AN=136067538&lang=pt-br&site=ehost-live>.

ŽUNIĆ, E.; , D.; BUZA, E. An Adaptive Data-Driven Approach to Solve Real-World Vehicle Routing Problems in Logistics. **Complexity**, Hindawi, v. 2020, p. 7386701, jan 2020. ISSN 1076-2787. Disponível em: <https://doi.org/10.1155/2020/7386701>.