



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

PROPOSTA DE UMA ARQUITETURA COMPUTACIONAL INTELIGENTE PARA A GESTÃO DOS INDICADORES UTILIZADOS NO PLANEJAMENTO E NO MONITORAMENTO DO SISTEMA DE TRANSPORTE URBANO

TÚLIO PHILIPPE FERREIRA E VIEIRA

Orientador: Gray Farias Moita
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadora: Magali Rezende Gouvêa Meireles
Pontifícia Universidade Católica de Minas Gerais

Coorientador: Renato Guimarães Ribeiro
Centro Federal de Educação Tecnológica de Minas Gerais

BELO HORIZONTE
NOVEMBRO DE 2024

TÚLIO PHILIPPE FERREIRA E VIEIRA

**PROPOSTA DE UMA ARQUITETURA COMPUTACIONAL
INTELIGENTE PARA A GESTÃO DOS INDICADORES
UTILIZADOS NO PLANEJAMENTO E NO
MONITORAMENTO DO SISTEMA DE TRANSPORTE
URBANO**

Tese apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Sistemas Inteligentes

Orientador: Gray Farias Moita
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadora: Magali Rezende Gouvêa Meireles
Pontifícia Universidade Católica de Minas Gerais

Coorientador: Renato Guimarães Ribeiro
Centro Federal de Educação Tecnológica de Minas Gerais

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL
BELO HORIZONTE
NOVEMBRO DE 2024

V658p

Vieira, Túlio Philipe Ferreira e

Proposta de uma arquitetura computacional inteligente para a gestão dos indicadores utilizados no planejamento e no monitoramento do sistema de transporte urbano / Túlio Philipe Ferreira e Vieira. – 2023.
128 f.

Tese de doutorado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.

Orientador: Gray Farias Moita.

Coorientadora: Magali Rezende Gouvêa Meireles.

Coorientador: Renato Guimarães Ribeiro.

Tese (doutorado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Transporte urbano – Teses. 2. Aprendizado do computador – Teses.
3. Filtro de Kalman – Teses. 4. Séries temporais – Teses. 5. Cidades inteligentes – Teses. 6. Sistemas de transporte inteligentes – Teses. I. Moita, Gray Farias.
II. Meireles, Magali Rezende Gouvêa. III. Ribeiro, Renato Guimarães.
IV. Centro Federal de Educação Tecnológica de Minas Gerais. V. Título.

CDD 004.0151



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

***“PROPOSTA DE UMA ARQUITETURA COMPUTACIONAL
INTELIGENTE PARA A GESTÃO DOS INDICADORES UTILIZADOS NO
PLANEJAMENTO E NO MONITORAMENTO DO SISTEMA DE
TRANSPORTE URBANO”.***

Tese de Doutorado apresentada por **Túlio Philipe Ferreira e Vieira**, em 29 de agosto de 2023, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Gray Farias Moita

Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Renato Guimarães Ribeiro (Coorientador)

Centro Federal de Educação Tecnológica de Minas Gerais

Profª. Drª. Magali Rezende Gouvêa Meireles (Coorientadora)

Pontifícia Universidade Católica de Minas Gerais

Prof. Dr. Marcos Paulino Roriz Junior

Universidade Federal de Goiás

Prof. Dr. Guilherme de Castro Leiva

Centro Federal de Educação Tecnológica de Minas Gerais

Profª. Drª. Elisângela Martins de Sá

Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Alisson Marques da Silva

Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida a impressão.

Prof. Dr. Thiago de Souza Rodrigues

Coordenador do Programa de Pós-Graduação Stricto Sensu em

CÓPIA DE FOLHA DE ASSINATURAS Nº 16/2023 - DEC (11.56.07)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 21/12/2023 08:34)

ALISSON MARQUES DA SILVA
PROFESSOR ENS BASICO TECN TECNOLOGICO
CTINFDV (11.50.29)
Matrícula: ###529#8

(Assinado digitalmente em 21/12/2023 08:26)

ELISANGELA MARTINS DE SA
PROFESSOR ENS BASICO TECN TECNOLOGICO
DCSA (11.56.04)
Matrícula: ###280#9

(Assinado digitalmente em 20/12/2023 16:22)

GRAY FARIAS MOITA
PROFESSOR DO MAGISTERIO SUPERIOR
DEC (11.56.07)
Matrícula: ###431#6

(Assinado digitalmente em 04/01/2024 15:51)

GUILHERME DE CASTRO LEIVA
PROFESSOR ENS BASICO TECN TECNOLOGICO
CETR (11.51.07)
Matrícula: ###154#6

(Assinado digitalmente em 29/12/2023 18:16)

RENATO GUIMARAES RIBEIRO
PROFESSOR ENS BASICO TECN TECNOLOGICO
DET (11.55.11)
Matrícula: ###233#1

(Assinado digitalmente em 21/12/2023 09:26)

MARCOS PAULINO RORIZ JUNIOR
ASSINANTE EXTERNO
CPF: ###.###.081-##

(Assinado digitalmente em 20/12/2023 18:57)

MAGALI REZENDE GOUVÊA MEIRELES
ASSINANTE EXTERNO
CPF: ###.###.046-##

Visualize o documento original em <https://sig.cefetmg.br/documentos/> informando seu número: 16, ano: 2023, tipo: **CÓPIA DE FOLHA DE ASSINATURAS**, data de emissão: 20/12/2023 e o código de verificação: 2e3b885f5f

Dedico este trabalho e toda a minha formação como pós-graduado ao meu professor, orientador e amigo Paulo Eduardo Maciel de Almeida, com quem pude aprender e conhecer o real significado da palavra Mestre. Sentirei muita saudade, velhinho.

Resumo

De acordo com dados da Organização das Nações Unidas, mais de 3,8 bilhões de pessoas vivem em áreas urbanas, atualmente, e espera-se que esse número cresça significativamente até 2050. Com o aumento da urbanização, especialmente em países em desenvolvimento, a mobilidade urbana torna-se um desafio crítico. O sistema de transporte, essencial para a movimentação das pessoas nas cidades, influencia diretamente a qualidade de vida dos cidadãos, pois é por meio dele que se garante o acesso a serviços essenciais como saúde, educação, cultura e lazer. Além disso, a análise da mobilidade pode revelar disparidades sociais, econômicas e culturais dentro de uma cidade. Portanto, para garantir um desenvolvimento urbano equilibrado e sustentável, é imperativo focar no planejamento e na gestão eficaz da mobilidade urbana, buscando soluções inovadoras e sustentáveis para o transporte e a circulação nas cidades. A fim de contribuir para a construção de um sistema eficiente de transporte, este trabalho apresenta o desenvolvimento de um *framework* composto por *hardware* e *software*, de baixo custo, que possa ser utilizado para geração e monitoramento de indicadores do sistema de transporte. O *framework* proposto utiliza dados de diferentes fontes como *probe request*, censo e características regionais e emprega algoritmos de *machine learning* como Filtro de Kalman e Redes Neurais Artificiais para gerar a matriz de origem-destino e a outros indicadores do sistema de transporte. Os resultados indicam que a estratégia utilizada consegue reduzir os erros de previsão de indicadores, como a taxa de carregamento, para menos de 5%.

Palavras-chave: Transporte público. Aprendizado de máquina. Filtro de Kalman. Séries temporais. Matriz origem-destino. Cidades inteligentes. Transporte inteligente.

Abstract

According to data from the United Nations, over 3.8 billion people currently live in urban areas, and this number is expected to grow significantly by 2050. With the rise of urbanization, especially in developing countries, urban mobility becomes a critical challenge. The transportation system, vital for people's movement within cities, directly impacts the quality of life of its residents, as it ensures access to essential services such as health, education, culture, and leisure. Furthermore, mobility analysis can reveal social, economic, and cultural disparities within a city. Therefore, to ensure balanced and sustainable urban development, it is imperative to focus on the planning and effective management of urban mobility, seeking innovative and sustainable solutions for transportation and circulation in cities. To contribute to the construction of an efficient transportation system, this study presents the development of a low-cost framework, comprised of both hardware and software, that can be used for generating and monitoring transportation system indicators. The proposed framework utilizes data from various sources such as probe requests, census, and regional characteristics and employs machine learning algorithms like the Kalman Filter and Artificial Neural Networks to generate the origin-destination matrix and other transportation system indicators. The results indicate that the employed strategy can reduce prediction errors for indicators, such as the loading rate, to less than 5%.

Keywords: Public transportation. Machine learning. Kalman filter. Time series. Origin-destination matrix. Smart cities. Intelligent transportation.

Lista de Figuras

Figura 1 – Urbanização brasileira por região em 2015	2
Figura 2 – Crescimento da população e variação de outros indicadores de mobilidade urbana em Belo Horizonte, Brasil, entre os anos de 2002 e 2012	5
Figura 3 – Comparação entre a diferença do tempo de viagens realizadas por meio do transporte coletivo e individual, na cidade de Belo Horizonte, entre os anos de 2002 e 2012.	6
Figura 4 – Ciclo vicioso da falta de planejamento urbano	9
Figura 5 – Tempo médio das viagens realizadas por classe social	9
Figura 6 – Percentual de transporte realizado por classes, dividido entre coletivo e individual	10
Figura 7 – Modelo conceitual das Cidades Inteligentes	28
Figura 8 – Exemplo de uma matriz O-D	32
Figura 9 – Exemplo de uma matriz O-D para a cidade de Maceió/AL	32
Figura 10 – Componentes para a solução de uma tarefa utilizando AM	36
Figura 11 – Representação de um Perceptron Multicamadas (PMC)	38
Figura 12 – Representação do processo de treinamento de uma RNA	39
Figura 13 – Exemplos de RNA e principais aplicações	41
Figura 14 – Representação da Equação 2 como um DAG	42
Figura 15 – Representação gráfica da Equação 3	43
Figura 16 – Representação gráfica da Equação 4	44
Figura 17 – Estrutura para as células LSTM	46
Figura 18 – Estrutura para as células GRU	47
Figura 19 – Representação de uma topologia matricial	48
Figura 20 – Representação da operação de convolução	49
Figura 21 – Exemplo de convolução	50
Figura 22 – Camada de convolução	51
Figura 23 – Representação gráfica do processo de <i>pooling</i>	51
Figura 24 – Representação gráfica de uma CNN	52
Figura 25 – Representação gráfica da SOM	53
Figura 26 – Topologias para ligação dos nós em uma SOM	54
Figura 27 – Representação de um processo utilizando o filtro de Kalman	57
Figura 28 – Transformação linear entre os espaços da variável de estados $x(t)$ e das medidas da variável $z(t)$	60
Figura 29 – Representação gráfica do algoritmo do filtro de Kalman	62
Figura 30 – <i>Hardware</i> utilizado para a construção do sistema	64
Figura 31 – Esquema dos componentes principais do sistema proposto	66

Figura 32 – Rota para a linha 1404A	70
Figura 33 – Mapa geral do transporte público em Saquarema-RJ	72
Figura 34 – Linha 1404A	79
Figura 35 – Linha 1404B	79
Figura 36 – Linha 1404C	79
Figura 37 – Linha 1404A	80
Figura 38 – Linha 1404B	80
Figura 39 – Linha 1404C	80
Figura 40 – Valores e <i>boxplot</i> para as taxas de carregamento	82
Figura 41 – Boxplot para RNR LSTM, RNR GRU e CNN	84
Figura 42 – Previsão da taxa de carregamento utilizando RNR LSTM, RNR GRU e CNN	85
Figura 43 – <i>Autocorrelation Function</i> (ACF) para o valor médio da taxa de carregamento	86
Figura 44 – Previsão utilizando ARIMA(0,2,1)	86
Figura 45 – Trajetos e pontos de retorno do transporte público em Saquarema-RJ	87
Figura 46 – Itinerário Bacaxá-Jaconé	88
Figura 47 – Itinerário Saquarema-Palmital (ida)	88
Figura 48 – Itinerário Bacaxá-Vilatur (ida e volta)	89
Figura 49 – Itinerário Saquarema-Rio Mole	90
Figura 50 – Demanda de passageiros em setembro/2022 por faixa horária (Plano de mobilidade urbana)	90
Figura 51 – Demanda de passageiros em setembro/2022 por faixa horária (Sistema proposto)	91
Figura 52 – Carregamento dos pontos de embarque	91
Figura 53 – Comparação e localização dos principais PED encontrados pelo sistema	92
Figura 54 – Comparação entre as origens e os destinos das viagens geradas pelo sistema proposto e pelo Plano de Mobilidade Urbana de Saquarema	94
Figura 55 – Marcadores dos <i>clusters</i> em um mapa de calor	98
Figura 56 – <i>Framework</i> proposto	107

Lista de Tabelas

Tabela 1 – Evolução do tempo gasto na ida ao trabalho em algumas regiões do Brasil entre os anos de 1992 e 2012	4
Tabela 2 – Modelo de arquivo contendo a marcação temporal, latitude e longitude dos MAC coletados	65
Tabela 3 – Modelo de arquivo contendo os MAC coletados e a marcação temporal	66
Tabela 4 – Exemplo de dados consolidados após a etapa adicional	73
Tabela 5 – Exemplo de dados utilizados para a construções dos modelos de análise de séries temporais	75
Tabela 6 – Análise estatística para RNR com LSTM	81
Tabela 7 – Análise estatística para RNR com GRU	83
Tabela 8 – Análise estatística para CNN	83
Tabela 9 – Análise estatística para RNR LSTM vs RNR GRU vs CNN	84
Tabela 10 – Análise estatística para o modelo ARIMA(0,2,1)	84
Tabela 11 – Dados das linhas e datas utilizadas para coleta e avaliação do sistema proposto	87
Tabela 12 – Conjunto de variáveis utilizadas pela rede SOM	98

Lista de Abreviaturas e Siglas

AFC	<i>Automated Fare Collection</i>
AM	Aprendizado de Máquina
ARIMA	<i>Autoregressive Integrated Moving Average</i>
BRT	<i>Bus Rapid Transit</i>
CDR	<i>Call Detail Record</i>
CNN	<i>Convolutional Neural Network</i>
GPS	<i>Global Positioning System</i>
GRU	<i>Gated Recurrent Units</i>
IA	Inteligência Artificial
ICT	<i>Information and Communications Technology</i>
IoT	Internet das Coisas
ITS	<i>Intelligent Transport System</i>
KPI	<i>Key Performance Indicator</i>
LSTM	<i>Long short-term memory</i>
MAC	<i>Medium Access Control</i>
Matriz O-D	Matriz Origem-Destino
MG	Minas Gerais
MUS	Mobilize - Mobilidade Urbana Saudável
NTU	Empresas de Transportes Urbanos
ONU	Organização das Nações Unidas
PCA	<i>Principal Component Analysis</i>
PED	Pontos de Embarque e Desembarque
PeMS	<i>Caltrans Performance Measurement System</i>

PNAD	Pesquisa Nacional por Amostra de Domicílios
RFID	<i>Radio-frequency Identification</i>
RMSE	<i>Root-Mean-Square Error</i>
RNR	Rede Neural Recorrente
RJ	Rio de Janeiro
RMSPE	<i>Root Mean Square Percent Errors</i>
RNA	Redes Neurais Artificiais
RNR	Redes Neurais Recorrentes
Rpi	Raspberry Pi
SAE	<i>Stacked Autoencoders</i>
SMS	<i>Short Message Service</i>
SO	Sistema Operacional
SOM	<i>Self-organizing Map</i>
UFRGS	Universidade Federal do Rio Grande do Sul
UFSC	Universidade Federal de Santa Catarina
UnB	Universidade de Brasília
VAR	<i>Vector Autoregressive</i>

Sumário

1 – Introdução	1
1.1 Justificativa	3
1.2 Motivação	8
1.3 Objetivos	10
1.4 Organização do Trabalho	11
2 – Trabalhos Relacionados	12
2.1 Planejamento do sistema de transporte por meio de pesquisas domiciliares	13
2.2 Planejamento e operação do sistema de transporte por meio de dados de cartões magnéticos de cobrança	16
2.3 Planejamento e operação do sistema de transporte por meio de dados de redes de telefonia	19
2.4 Planejamento e operação do sistema de transporte por meio de dados de diferentes fontes	22
3 – Fundamentação Teórica	27
3.1 Cidades Inteligentes	27
3.2 Matriz O-D	31
3.3 Inteligência Artificial (IA)	34
3.3.1 Aprendizado de Máquina (AM)	35
3.3.2 Redes Neurais Artificiais (RNA)	37
3.3.3 Redes Neurais Recorrentes (RNR)	41
3.3.4 <i>Gated Recurrent Unit</i> (GRU)	47
3.3.5 Rede Neural Convolucional	47
3.3.6 Self Organizing Map (SOM)	50
3.4 Filtro de Kalman	55
4 – Metodologia	63
4.1 Coleta e Armazenamento dos Dados	63
4.1.1 Equipamentos Utilizados	63
4.1.2 Processo de Coleta de Dados	67
4.1.3 Armazenamento dos Dados	67
4.2 Modelagem da Taxa de Carregamento	68
4.2.1 Coleta de Dados Experimentais	69
4.2.2 Validação Manual da Taxa de Carregamento	70
4.3 Modelagem da Matriz Origem-Destino (O-D)	71

4.3.1	Criação da Matriz O-D	71
4.3.2	Validação da Matriz O-D	73
4.4	Técnicas e Algoritmos Utilizados	74
4.4.1	Modelagem e Análise do Sistema Proposto	74
4.4.2	Análise da Dinâmica das Viagens	76
5	– Apresentação e Análise dos Resultados	77
5.1	Estimativa da Taxa de Carregamento	77
5.2	Modelagem da Taxa de Carregamento	81
5.3	Análise dos Dados do Experimento de Saquarema-RJ	85
5.4	Estimativa da Matriz Origem-Destino	93
5.5	Análise da Dinâmica das Viagens	96
6	– Proposta de <i>Framework</i> para o Planejamento e o Gerenciamento do Sistema de Transporte	106
6.1	Estrutura do <i>Framework</i> Proposto	106
6.2	Características do <i>Framework</i> Proposto	108
7	– Discussões Finais	111
7.1	Contribuições	112
7.1.1	Contribuições Teóricas	112
7.1.2	Contribuições Práticas	114
7.1.3	Contribuições Sociais e Ambientais	115
7.2	Propostas de Continuidade	117
	Referências	119

Capítulo 1

Introdução

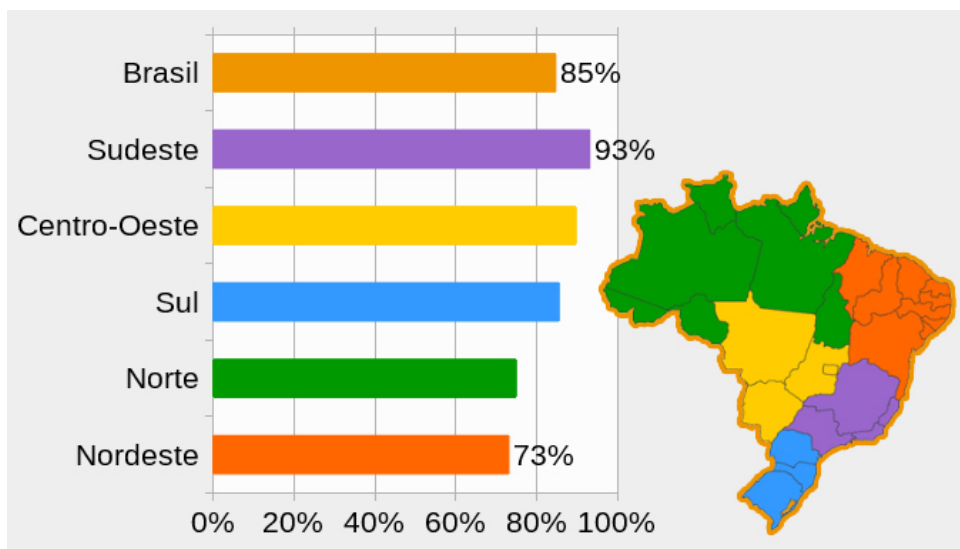
A mobilidade urbana representa um elemento essencial na configuração e evolução das cidades contemporâneas (CAGNEY et al., 2020). Mobilidade urbana refere-se à dinâmica e organização dos deslocamentos de pessoas e bens no espaço urbano, envolvendo os diferentes modos de transporte e suas interações (PINHEIRO; FRISCHTAK, 2015). No século XX, a popularização dos automóveis e outros meios de transporte motorizados revolucionou a dinâmica das áreas urbanas, trazendo consigo uma série de implicações para o planejamento e a sustentabilidade das cidades (SCHIPPER; EMANUEL; OLDENZIEL, 2020). Embora tais veículos tenham proporcionado uma maior conectividade e conveniência, eles também introduziram desafios substanciais, particularmente no que diz respeito à segurança e à infraestrutura urbana (PINHEIRO; FRISCHTAK, 2015).

No contexto dos países em desenvolvimento, esses desafios são ainda mais pronunciados. A urbanização rápida e, em muitos casos, não planejada, combinada com infraestruturas de transporte frequentemente inadequadas, tem levado a complicações em termos de mobilidade e acessibilidade (PINHEIRO; FRISCHTAK, 2015). Dados da Organização das Nações Unidas (ONU) prevê que, em 2050, mais de 68% da população global viverá em áreas urbanas. Isso representa mais de 6,5 bilhões de pessoas nos centros urbanos. No Brasil, dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2015 já indicavam que 84,7% da população residia em áreas urbanas. A Figura 1 ilustra o percentual de urbanização por região no país.

Em resposta a esses desafios, surgiu o conceito de "mobilidade sustentável". A mobilidade sustentável, derivada do conceito de "desenvolvimento sustentável", visa garantir que os sistemas de transporte atendam às demandas econômicas, sociais e ambientais da sociedade, minimizando impactos negativos (GALLO; MARINELLI, 2020). Esse conceito vai além da mera redução da poluição, abordando também os impactos sociais e econômicos. Para alcançar a verdadeira mobilidade sustentável, é imperativo adotar estratégias que não se restrinjam apenas a sistemas de transporte menos poluentes, mas que também

considerem uma abordagem holística que englobe todas as facetas da sustentabilidade.

Figura 1 – Urbanização brasileira por região em 2015



Fonte: Adaptado de [Transporte \(2017b\)](#)

[Araújo et al. \(2011\)](#) definem o sistema de transporte urbano como o deslocamento de pessoas e produtos dentro do espaço das cidades. Pesquisa realizada pelo projeto Mobilize - Mobilidade Urbana Saudável (MUS), que contou com a participação de membros de universidades brasileiras, como a Universidade de Brasília (UnB), Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Rio Grande do Sul (UFRGS) e da Oxford Brookes University do Reino Unido, buscou compreender o impacto que a mobilidade gera na saúde e no bem-estar em comunidades de baixa renda no Brasil e no Reino Unido ([JONES, 2019](#)). Esse estudo ressalta que a qualidade de vida e o bem-estar social podem estar, diretamente, associados ao maior acesso à mobilidade urbana. Os pesquisadores desse estudo também apontam que o planejamento e uma política pública integrada são fatores determinantes para maior acesso à mobilidade urbana e, consequentemente, para o melhor bem-estar da população.

[Ferraz e Torres \(2001\)](#) indicam a mobilidade urbana como um elemento balizador do desenvolvimento das cidades. Portanto, como a maioria da população mora nas cidades, é de suma importância que o planejamento, a gestão e a operação do sistema de transporte seja realizada de maneira integrada. Por meio dessa integração e de uma visão holística da mobilidade urbana, é possível melhorar a qualidade de vida das pessoas, promover um maior desenvolvimento econômico e social das cidades e proporcionar um crescimento sustentável.

Assim, quando não existe um planejamento adequado que garanta a mobilidade dentro dos centros urbanos, toda a população passa a ser afetada, principalmente, a porção de

baixa renda da sociedade. Isso ocorre pois, por meio do sistema de transporte, acontece o acesso a serviços de saúde, lazer, trabalho e educação. A ocupação desordenada e não planejada gera um fenômeno conhecido como o espraiamento das cidades (MACEDO, 2012). Nas cidades espraiadas, as atividades econômicas estão pouco dispersas em todo o território, o que ocasiona maior necessidade de deslocamento da população por toda a zona urbana. Portanto, o sistema de transporte urbano é ainda mais sobrecarregado e, como a população de baixa renda, normalmente, vive nas periferias das cidades e necessita do transporte para se deslocar ao local de trabalho, essa porção da população passa a ser ainda mais prejudicada. Nesse sentido, assegurar a mobilidade urbana é um dos elementos essenciais para melhorar a qualidade de vida dos cidadãos e garantir um desenvolvimento social, econômico e ambiental sustentável para as cidades.

Este trabalho apresenta um *framework* composto por *software* e *hardware* que auxilia na coleta de dados, na construção de indicadores como a taxa de carregamento dos ônibus e na tomada de decisão. Assim, ele poderá contribuir para que o planejamento e a operação do sistema de transporte público urbano possa ocorrer de maneira adequada. Desse modo, empregar esse *framework* no auxílio à tomada de decisão para o sistema de transporte público poderá promover impactos positivos na população urbana, melhorando a qualidade de vida das pessoas e contribuindo para o desenvolvimento da mobilidade urbana sustentável.

1.1 Justificativa

O planejamento de todo o sistema de transporte das cidades é fundamental para a melhoria da qualidade de vida de todos os residentes nas áreas urbanas. Campos (2006) apresenta a mobilidade urbana como um fator determinante para a garantia do desenvolvimento sustentável das cidades em um contexto sócioeconômico e ambiental. Nesse sentido, Campos (2006) demonstra que é necessária a adoção de uma 'oferta inteligente de transporte' por meio do planejamento inteligente da mobilidade que privilegie o transporte coletivo de passageiros. Entretanto, o que é observado, principalmente nos países em desenvolvimento, é a utilização cada vez menor do transporte coletivo. A Tabela 1 ilustra essa situação, mostrando a evolução do tempo de deslocamento de casa ao trabalho entre 1992 e 2012 nas principais regiões do Brasil. Há um aumento notável no tempo médio gasto, especialmente nas regiões metropolitanas. Em Salvador, por exemplo, o número de pessoas que levam mais de uma hora no trajeto cresceu 11,8 pontos percentuais. Estes dados ressaltam a necessidade urgente de ações concretas para melhorar o planejamento e o gerenciamento do sistema de transporte urbano.

Ferraz e Torres (2001) elencam algumas características que tornam o transporte coletivo de passageiros cada vez menos atrativo aos cidadãos. Dentre as características apresentadas,

Tabela 1 – Evolução do tempo gasto na ida ao trabalho em algumas regiões do Brasil entre os anos de 1992 e 2012

	Tempo médio gasto de casa ao trabalho (minutos)			Pessoas que levam mais de 1 hora até o trabalho (%)		
	1992	2012	Variação (%)	1992	2012	Variação (p.p)
RM Belém	24,3	33,0	36,0	3,3	10,5	7,1
RM Curitiba	30,4	32,2	6,0	8,5	11,1	2,6
RM Rio de Janeiro	43,9	47,4	7,8	23,2	27,1	3,9
RM Salvador	31,6	40,0	26,8	6,9	18,7	11,8
RM São Paulo	38,4	46,0	19,7	16,9	23,5	6,6
RM Belo Horizonte	32,6	36,8	12,9	9,6	15,8	6,2
Distrito Federal	33,0	35,2	6,5	10,3	11,0	0,7
Áreas não metropolitanas	22,5	23,7	5,1	3,6	4,6	1,1
Regiões metropolitanas	36,7	41,1	12,2	14,8	19,3	4,4
Brasil	28,4	30,4	6,8	8,3	10,2	2,0

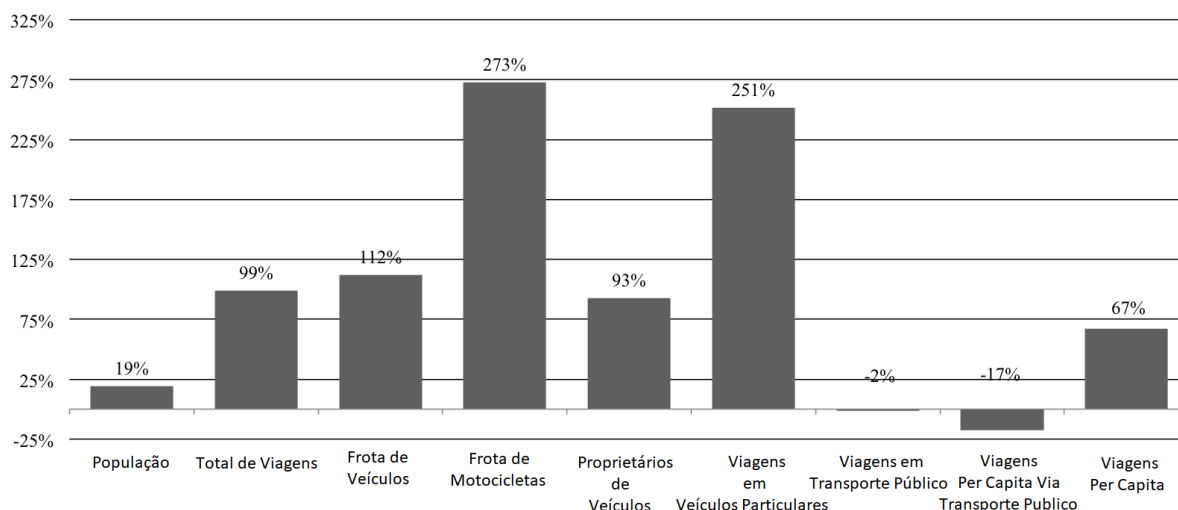
Fonte: Adaptado de [Pinheiro e Frischtak \(2015\)](#)

a baixa qualidade do serviço prestado e o acesso a veículos de transporte individual são vistos como fatores determinantes para o aumento de viagens por meio do transporte individual. Resultados da pesquisa da Associação Nacional das Empresas de Transportes Urbanos (NTU), conduzida em algumas capitais do Brasil, no ano de 2017, mostra a redução do número de passageiros do transporte coletivo. Essa redução do número de usuários do transporte público pode ser explicada pela baixa qualidade do sistema, pelo valor da tarifa e pelo aumento do tempo de viagem no transporte coletivo quando comparado ao transporte individual.

A [Figura 2](#) mostra dados da pesquisa de origem-destino realizada na cidade de Belo Horizonte, Minas Gerais, entre 2002 a 2012. Por meio das pesquisas de origem-destino, é possível identificar o comportamento da mobilidade urbana de toda uma região. Os dados apresentados na [Figura 2](#) evidenciam que, por um lado, ocorreu um aumento da frota de veículos de transporte individual como motocicletas, automóveis e, por outro lado, ressalta que ocorreu uma redução dos indicadores do número de viagens por meio do transporte público. Esse aumento do número de veículos de transporte individual em detrimento do transporte coletivo, agrava ainda mais o trânsito nas grandes cidades e reflete o aumento ainda maior das diferenças entre os tempos de viagens do transporte coletivo e do individual. Quanto maior essa diferença, mais os passageiros tendem a optar pelo transporte individual em detrimento do coletivo, prejudicando ainda mais a qualidade de vida nas cidades.

Fatores como o aumento dos congestionamentos, maior poluição atmosférica, necessidade de investimento em grandes infraestruturas, maior número de acidentes e o consumo

Figura 2 – Crescimento da população e variação de outros indicadores de mobilidade urbana em Belo Horizonte, Brasil, entre os anos de 2002 e 2012



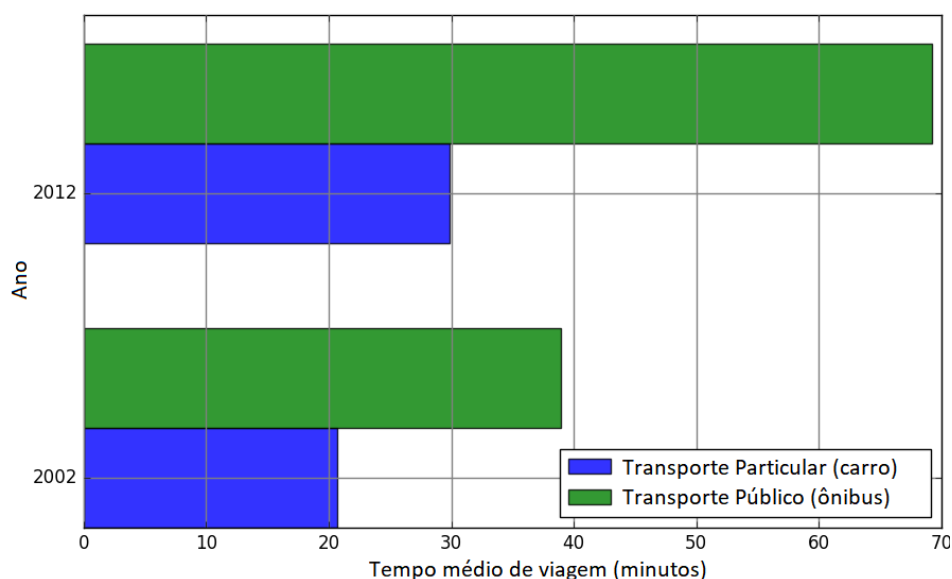
Fonte: Adaptado de [Ribeiro \(2015\)](#)

desordenado de combustíveis não renováveis passam a ser mais evidentes em cidades com uma menor utilização do transporte coletivo. A [Figura 3](#) mostra que, em 2002, o tempo médio das viagens no transporte público era 1,88 vezes maior que o transporte individual. Já em 2012, essa diferença passou para 2,38 na cidade de Belo Horizonte. Assim, é necessário que seja adotado um planejamento eficiente do transporte de passageiros a fim de torná-lo cada vez mais atrativo para que os cidadãos possam optar por esse meio de transporte e contribuir com a maior sustentabilidade das cidades. Desse modo, garantir o acesso a um sistema de transporte de qualidade que possibilite a mobilidade de pessoas e produtos nos centros urbanos é fundamental para o desenvolvimento sustentável das cidades.

Com o crescimento e o desenvolvimento cada vez maior dos centros urbanos, surge o conceito da construção de cidades sustentáveis. Nas cidades que planejam e executam projetos que visam promover a sustentabilidade, as necessidades (sociais, políticas, econômicas e ambientais) dos residentes são atendidas sem comprometer a possibilidade das gerações futuras. [Wey \(2019\)](#) elenca vários fatores que são determinantes para a construção das cidades sustentáveis. Dentre esses fatores, evidencia que o planejamento do sistema de transporte possui papel essencial para a preservação ambiental e para a qualidade de vida dos cidadãos residentes nos centros urbanos.

O desenvolvimento e utilização cada vez maior de sistemas computacionais, sensores e atuadores contribui para que a construção de bases de dados mais confiáveis e que possam ser produzidas de maneira automática ([WELCH; WIDITA, 2019](#)). Esses métodos automáticos de coleta reduzem os possíveis erros gerados pelos processos manuais,

Figura 3 – Comparação entre a diferença do tempo de viagens realizadas por meio do transporte coletivo e individual, na cidade de Belo Horizonte, entre os anos de 2002 e 2012.



Fonte: Adaptado de [Vieira et al. \(2020\)](#)

apoiam a construção do *Big Data* e auxiliam no desenvolvimento de modelos de análise de dados e tomada de decisão mais robustos. Esses modelos mais eficientes possibilitam a avaliação mais precisa de cenários e, conseqüentemente, propiciam um planejamento mais adequado de todo o sistema de transporte público.

[Zhao e Zhu \(2021\)](#) definem o *Big Data* dos sistemas de transporte como a massa de dados gerada pelos sistemas de transporte a fim de serem analisados e processados por ferramentas específicas do *Big Data*. Além disso, após analisados, esses dados são capazes de produzir programas de transporte mais eficazes, convenientes e com maior qualidade. Assim, esses dados podem ser empregados para melhorar o planejamento e a operação do sistema de transporte. Por meio de um planejamento e operação eficientes do transporte coletivo de passageiros, é possível aprimorar a governança do transporte urbano, trazer comodidade para a viagem das pessoas, melhorar a eficiência do transporte, economizar recursos energéticos e reduzir as emissões de poluição.

[Welch e Widita \(2019\)](#) apresentam um estudo sobre os principais tópicos e aplicações do *Big Data* para o planejamento e operação dos sistemas de transporte público. Nesse estudo, os autores destacam que, mesmo existindo uma crescente utilização das fontes de dados constituintes do *Big Data*, a análise desses dados ainda é pouco explorada de maneira consistente. Vários trabalhos como [Neilson et al. \(2019\)](#) e [Ceder \(2021\)](#) apresentam a forma que diferentes autores abordam a importância da construção e análise do *Big Data* no

planejamento, na gestão e na operação do sistema de transporte, seja utilizando *software*, *hardware* ou coletando dados de outras fontes como redes sociais. Essas fontes de dados são utilizadas para medir o nível e a *performance* do serviço ofertado, o comportamento dos usuários, a demanda, o gerenciamento e a segurança do transporte público.

Como pode ser visto no trabalho de [Ceder \(2021\)](#), várias tecnologias como o *Big Data*, a inteligência artificial e a coleta inteligente de dados são elementos essenciais para a construção de um sistema eficiente de transporte de passageiros dentro das cidades. Portanto, o desenvolvimento de um arcabouço que utilize essas tecnologias pode contribuir com a construção de cidades sustentáveis e com o bem-estar da população.

Desse modo, este trabalho apresenta contribuições teóricas e práticas para o desenvolvimento das cidades sustentáveis por meio do planejamento e gerenciamento adequados dos sistemas de transporte nos centros urbanos. Devido à importância da mobilidade urbana para o desenvolvimento sustentável, várias são as pesquisas realizadas no sentido de prover algoritmos e ferramentas que auxiliem no planejamento do sistema de transporte urbano ([ZHU et al., 2018](#)). Com o advento de novas tecnologias como a Internet das Coisas (IoT), a integração dos sistemas de comunicação, a geração de um grande número de dados (que possibilitou a construção do *Big Data*) e os avanços no campo da inteligência artificial, possibilitou o surgimento do *Intelligent Transport System* (ITS). ITS é uma denominação atribuída aos sistemas de transporte urbano que utilizam o estado da arte de sistemas de informação e telecomunicação para melhorar a segurança, a eficiência, o conforto, a preservação ambiental e para reduzir os congestionamentos nas cidades ([ANDERSEN; SUTCLIFFE, 2000](#)).

Uma maneira de alcançar a melhoria da mobilidade urbana é empregar as tecnologias desenvolvidas por meio do ITS e do *Big Data* para obter e monitorar os indicadores do sistema de transporte público de passageiros. [Ferraz e Torres \(2001\)](#) elencam alguns indicadores essenciais para um planejamento eficiente do sistema de transporte urbano. Indicadores como a taxa de carregamento e a pontualidade nas viagens são fundamentais para um planejamento do sistema de transporte que garanta a acessibilidade e a mobilidade urbana. Nesse sentido, este trabalho apresenta uma contribuição teórica para um melhor planejamento do transporte público, uma vez que apresenta uma nova metodologia que utiliza a inteligência artificial, os sistemas de informação e um conjunto de *hardware* e *software*, de baixo custo, para a construção automática das matrizes de origem-destino em linhas do transporte coletivo de passageiros e para o monitoramento constante de indicadores como a taxa de carregamento e os tempos de atrasos das linhas de transporte coletivo. Desse modo, é possível obter informações importantes para a construção de um planejamento eficiente da mobilidade nas grandes cidades.

O desenvolvimento de um sistema de baixo custo para o mapeamento do sistema de trans-

porte, como apresentado neste trabalho, representa uma contribuição prática significativa. A adoção do sistema proposto pode auxiliar na melhoria da qualidade do transporte coletivo de passageiros. Isso é possível graças à metodologia empregada, que pode ser aplicada a projetos de transporte público e associada a outras políticas urbanas. Essa abordagem é capaz de reduzir custos de operação, tempos de viagem e congestionamentos, aumentar o conforto dos usuários, diminuir a poluição ambiental, contribuir para uma melhoria da qualidade de vida nos centros urbanos e ajudar na construção das cidades sustentáveis. Portanto, o trabalho desenvolvido tem implicações práticas significativas para a gestão e o planejamento do transporte público urbano.

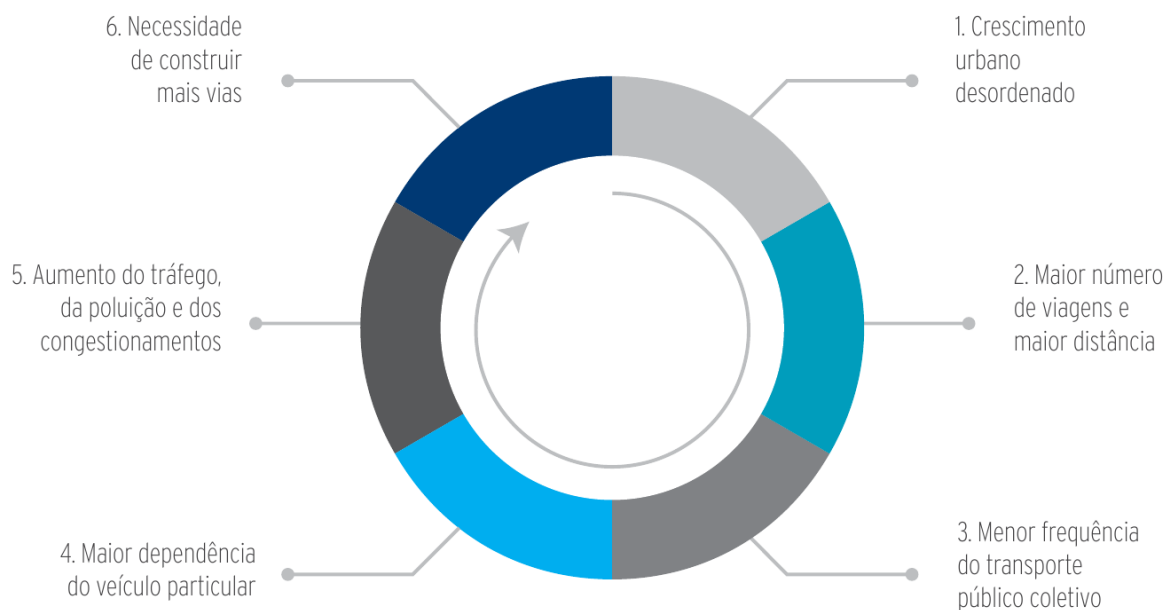
A viabilidade teórica, prática e econômica da metodologia apresentada neste trabalho foi comprovada pelos artigos publicados em congressos e revistas internacionais (([VIEIRA et al., 2020](#)), ([VIEIRA et al., 2019](#)) e ([VIEIRA et al., 2018](#))) e por meio de pesquisas de campo realizadas em conjunto com empresas de planejamento e operação do transporte coletivo urbano. Esses resultados preliminares mostram que a metodologia apresentada pode ser uma grande aliada na construção de um sistema de transporte urbano mais eficiente, justo e sustentável.

1.2 Motivação

A qualidade do sistema de transporte está entre os principais fatores que interferem na qualidade de vida das pessoas nos grandes centros urbanos. No Brasil, principalmente, a adoção de políticas públicas que privilegiem a mobilidade das pessoas nos centros urbanos foi negligenciada por muitos anos. Essa afirmação pode ser confirmada pelo estudo “A Conta do Desmonte – Balanço Geral do Orçamento da União” apresentado pelo Instituto de Estudos Socioeconômicos (INESC) em que os dados mostram que o governo federal reduziu em 70% o orçamento destinado aos transportes coletivos urbanos, entre 2019 e 2021 ([ALVAREZ, 2022](#)). Esse descaso com o sistema de transporte, a falta de planejamento urbano, a rápida urbanização e a preferência pelo transporte individual de passageiros resultam em um ciclo vicioso da falta de planejamento urbano. A [Figura 4](#) mostra um exemplo desse ciclo vicioso que prejudica a acessibilidade e a mobilidade urbana.

Quando é analisado o impacto econômico-social e ambiental que essa falta de planejamento adequado gera para as cidades, é possível perceber que as classes mais vulneráveis são ainda mais prejudicadas. A pesquisa de mobilidade da população urbana, realizada pela Confederação Nacional do Transporte (CNT) em parceria com a Associação Nacional das Empresas de Transportes Urbanos (NTU), no ano de 2017, mostra que as classes D e E são as que gastam mais tempo nos deslocamentos ([Figura 5](#)) e são as classes que mais necessitam do transporte coletivo para os deslocamentos ([Figura 6](#)). Quando esse acesso à mobilidade urbana é cerceado, torna-se ainda mais difícil o acesso a serviços de saúde,

Figura 4 – Ciclo vicioso da falta de planejamento urbano

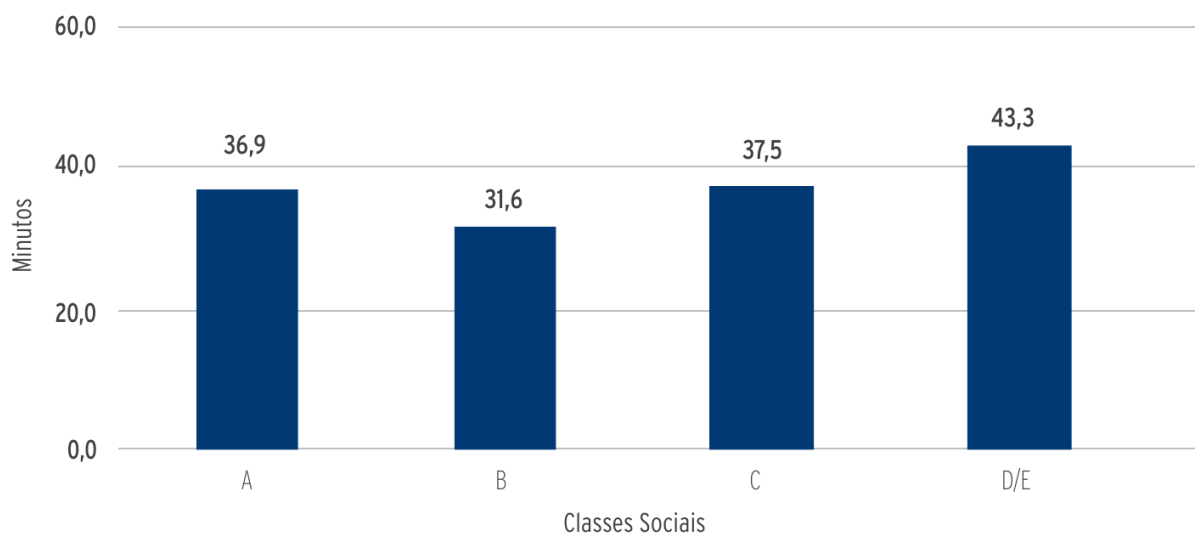


Fonte: Adaptado de [Transporte \(2017a\)](#)

trabalho e lazer. Consequentemente, realizar o planejamento adequado da mobilidade urbana também é uma forma de reduzir as diferenças sociais existentes no país.

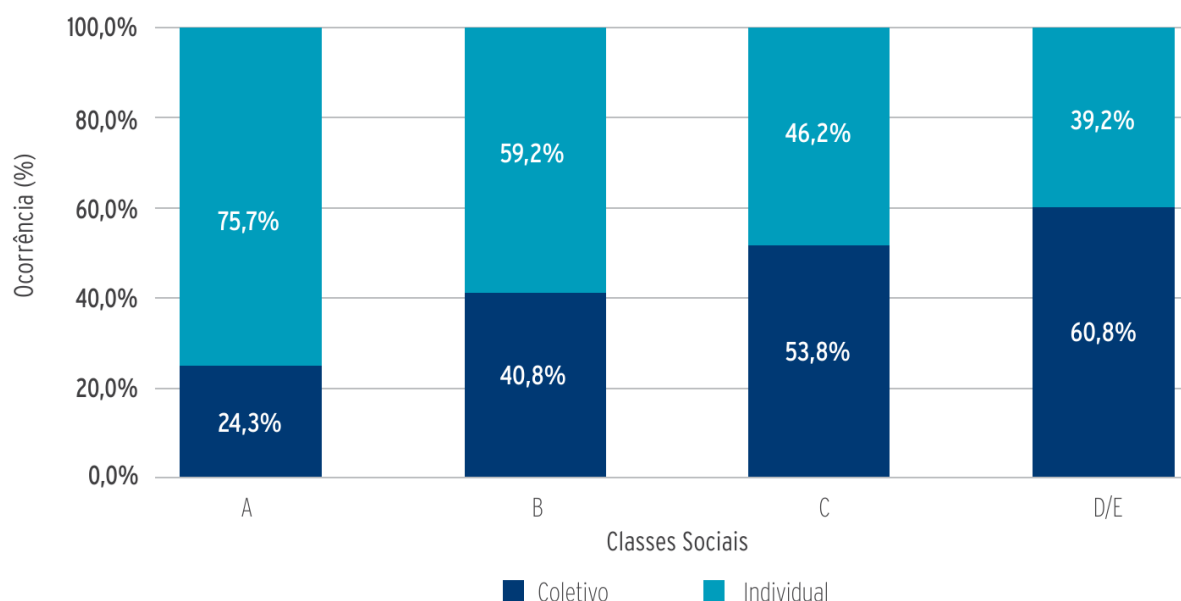
O desenvolvimento cada vez maior das tecnologias de comunicação, o barateamento dos sistemas de *hardware* e *software*, a evolução dos modelos de inteligência artificial e o

Figura 5 – Tempo médio das viagens realizadas por classe social



Fonte: Adaptado de [Transporte \(2017a\)](#)

Figura 6 – Percentual de transporte realizado por classes, dividido entre coletivo e individual



Fonte: Adaptado de [Transporte \(2017a\)](#)

crecente interesse em utilizar essas técnicas para a construção de cidades inteligentes sustentáveis possibilitam a melhoria da qualidade de vida dos cidadãos residentes nos centros urbanos. Portanto, aplicar sistemas computacionais que auxiliem na coleta de dados, no planejamento e no monitoramento do sistema de transporte é fundamental para a construção das cidades sustentáveis e para a diminuição das diferenças sociais enfrentadas, principalmente, pelos países em desenvolvimento.

1.3 Objetivos

Este trabalho tem como objetivo geral propor um *framework*, de baixo custo, que consiga utilizar diferentes fontes de dados para gerar e monitorar indicadores, e, assim, possibilitar um melhor planejamento, gerenciamento e operação do sistema de transporte, especialmente, em países em desenvolvimento.

Para alcançar esse objetivo geral proposto, são traçados os seguintes objetivos específicos:

- Construir um sistema de *hardware* e *software* de baixo custo para a coleta dos *probes request* dos aparelhos celulares dentro de ônibus e estações de transporte coletivo.
- Utilizar técnicas de aprendizado de máquina para o ajuste dos dados coletados.
- Modelar a variação da taxa de carregamento diária para as linhas do transporte coletivo.

- Desenvolver um modelo para a construção da matriz de origem-destino a partir do modelo de variação da taxa de carregamento.
- Coletar, a partir do sistema desenvolvido, outros indicadores para o planejamento do sistema de transporte.
- Comparar a metodologia desenvolvida com outras já existentes na literatura.
- Validar a metodologia proposta com dados coletados em campo.

1.4 Organização do Trabalho

Este trabalho está organizado da seguinte maneira:

- No Capítulo 2, são apresentados os trabalhos relacionados ao tema deste trabalho, ou seja, são discutidos os trabalhos presentes na literatura e que possuem pontos de interseção com a presente pesquisa.
- No Capítulo 3, é apresentado o referencial teórico utilizado para o desenvolvimento desta pesquisa.
- No Capítulo 4, são apresentados os métodos para o desenvolvimento do sistema de coleta, preparação, análise e construção dos modelos de estimação das matrizes de origem-destino.
- No Capítulo 5, são apresentados e discutidos os resultados obtidos.
- No Capítulo 6, é apresentado e discutido o *framework* proposto.
- No Capítulo 7, é apresentada a conclusão do trabalho e as propostas para a continuidade dos estudos.

Capítulo 2

Trabalhos Relacionados

Neste capítulo introdutório da revisão de trabalhos relacionados, o foco é direcionado para o tema central da pesquisa: a construção da matriz de origem-destino (matriz O-D) e a identificação de indicadores como a taxa de carregamento no contexto do planejamento e operação de sistemas de transporte. A análise abrange uma variedade de abordagens e metodologias, fornecendo uma visão abrangente do campo e estabelecendo o contexto para as seções subsequentes.

Os trabalhos relacionados são organizados em quatro grupos distintos, que serão detalhados nas próximas seções deste capítulo. Esses grupos refletem diferentes abordagens para o planejamento e operação do sistema de transporte, incluindo o uso de pesquisas domiciliares, [Seção 2.1](#), dados de cartões magnéticos de cobrança, [Seção 2.2](#), dados de redes de telefonia, [Seção 2.3](#), e a combinação de diferentes fontes de dados, [Seção 2.4](#).

A análise de [Litman \(2007\)](#) serve como um ponto de partida para a discussão, enfatizando a importância de escolher indicadores que reflitam as características econômicas, sociais e ambientais do sistema de transporte. Essa perspectiva abre caminho para a exploração das oportunidades e desafios da utilização do *Big Data*, um tema que é aprofundado por [Neilson et al. \(2019\)](#). Eles examinam a utilização do *Big Data* na construção de um sistema eficiente de transporte, uma análise que ressoa com as categorias de trabalhos relacionados que serão exploradas nas seções subsequentes.

[Welch e Widita \(2019\)](#) complementam essa análise, reforçando a importância dos dados gerados pelo *Big Data* para o planejamento e operação do sistema de transporte. Eles destacam lacunas e desafios que se conectam diretamente às categorias e abordagens discutidas anteriormente, sublinhando a relevância das próximas seções deste capítulo.

Em resumo, este capítulo introdutório estabelece o cenário para uma exploração detalhada das técnicas, oportunidades e desafios na área de planejamento e gerenciamento de sistemas de transporte. A revisão da literatura serve como uma base sólida para a compreensão

das contribuições teóricas e práticas deste estudo, integrando as várias abordagens e preparando o terreno para as análises específicas que serão apresentadas nas seções subsequentes.

2.1 Planejamento do sistema de transporte por meio de pesquisas domiciliares

A utilização de pesquisas domiciliares no planejamento do sistema de transporte é um elemento-chave no desenvolvimento urbano, promovendo melhorias significativas na mobilidade e acessibilidade em diferentes regiões (RODRIGUES; SILVA; BANDEIRA, 2021). Essa metodologia envolve coletar diretamente nos domicílios informações detalhadas sobre os padrões de viagem e as características socioeconômicas dos moradores. A complexidade e o elevado custo associados a essas pesquisas são justificados pela necessidade crucial de obter dados precisos e atualizados, essenciais para a eficácia do planejamento do transporte urbano. Utilizada como uma ferramenta comum no planejamento de transporte, essa metodologia apresenta complexidades e nuances específicas. Nesta seção, serão discutidos os trabalhos de diversos autores que exploram as diferentes metodologias, desafios e oportunidades associadas a essa abordagem.

Richardson, Ampt e Meyburg (1995) apresentam diferentes metodologias para a realização de pesquisas domiciliares no planejamento do sistema de transporte. Além disso, os autores discutem vários métodos para a realização dessas pesquisas e os resultados esperados após a realização do processo de análise dos dados. Entretanto, os autores indicam que os custos para a construção dessas pesquisas podem ser elevados. Desse modo, é necessário que exista um *trade-off* entre a qualidade dos dados, a quantidade de dados e os recursos disponíveis para a realização das pesquisas. Estratégias mais modernas, como a construção e análise de grandes bases de dados gerados por sensores e *smartphones* minimizam a diferença entre a qualidade, quantidade e recursos disponíveis.

O'Flaherty (1997) também apresenta algumas estratégias para o planejamento do sistema de transporte utilizando pesquisas de campo. O autor sugere um conjunto de passos para a construção e para a condução das pesquisas de campo. O'Flaherty (1997) indica que o treinamento dos pesquisadores e a correta seleção dos métodos de pesquisas são fundamentais para assegurar a eficiência do planejamento por meio desses dados coletados. Assim, a necessidade de treinamento e a motivação dos pesquisadores para a condução dessas pesquisas envolvem maiores gastos para a geração de indicadores de planejamento e monitoramento do sistema de transporte e torna essa metodologia mais suscetível a erros quando comparada com modelos automáticos de geração e controle de indicadores.

Stopher e Jones (2003) apresentam alguns requisitos que devem ser observados para a

condução de pesquisas domiciliares que auxiliem na construção das matrizes O-D e outros indicadores do sistema de transporte. Mesmo indicando que as pesquisas de domicílios possam ser utilizadas para o planejamento do sistema, os próprios autores ressaltam que pode existir uma baixa taxa de resposta e que os dados coletados podem apresentar baixa confiabilidade. Nesse sentido, combinar diferentes técnicas para a construção de bases de dados mais confiáveis torna-se necessário para o sucesso do planejamento e da gestão do sistema de transporte.

A utilização de pesquisas domiciliares em conjunto com tecnologias como GPS (do inglês *Global Positioning System*) e RFID (do inglês *Radio-frequency Identification*) é proposta por [Wolf et al. \(2006\)](#). Os autores sugerem a aplicação dessas tecnologias a fim de reduzir os custos das pesquisas e aumentar a acurácia dos dados coletados. Entretanto, esse método proposto não utiliza modelos mais modernos de inteligência artificial que possam melhorar a precisão das pesquisas realizadas e garantir análises em tempo real dos dados coletados.

[Stopher e Greaves \(2007\)](#) mostram algumas mudanças existentes entre a forma de como eram executadas as pesquisas domiciliares nos anos 1970's e nos dias atuais. Esses autores indicam que tecnologias como painéis de pesquisa, GPS e a fusão de diferentes fontes de dados podem ser utilizadas para a construção de pesquisas mais eficientes. Nesse trabalho, os autores concluem que a adoção de fontes de dados mais confiáveis deve ser realizada.

[Stopher, FitzGerald e Xu \(2007\)](#) buscaram melhorar a qualidade e confiabilidade dos dados gerados por pesquisas de campo e análise da origem e destino das viagens, adotando o uso de dados de GPS. Eles aplicaram esse modelo na cidade de Sidney, na Austrália, e alcançaram resultados promissores, reduzindo a sub-notificação das viagens realizadas para apenas 7%. Essa abordagem permitiu que os dados obtidos representassem de maneira mais fiel a realidade das viagens. No entanto, como mostrado por [Kurauchi e Schmöcker \(2017\)](#), os custos para a adoção dessa estratégia tendem a ser mais elevados, uma vez que a maior parte dos custos está associada às pesquisas domiciliares.

Uma estratégia mais moderna e que utiliza os *smartphones* como meio de realizar pesquisas é proposto por [Safi, Mesbah e Ferreira \(2014\)](#). Nesse trabalho, os autores propõem a utilização dos *smartphones* para realização de pesquisas e coleta automática dos dados das viagens. Desse modo, é possível que participantes da pesquisa possam enviar os dados coletados de maneira automática para a construção dos indicadores necessários. Apesar de ser uma estratégia inovadora, fatores como a necessidade de o 'entrevistado/passageiro' manter uma aplicação no *smartphone* pode ser um complicador, pois nem todos os usuários se sentiriam confortáveis deixando uma aplicação de monitoramento no celular.

[Litman \(2022\)](#) também apresenta uma abordagem que utiliza as pesquisas domiciliares para

o planejamento do transporte. Ele destaca algumas mudanças nas abordagens de avaliação da equidade no transporte, com foco especial nas pesquisas domiciliares. [Litman \(2022\)](#) indica que essas pesquisas fornecem dados sobre os gastos com transporte em diferentes grupos de renda e podem ser usadas para entender como os custos de transporte afetam as famílias de baixa renda. Nesse trabalho, o autor conclui que os gastos com transporte tendem a aumentar com o uso do veículo e diminuir com o uso do transporte público. No entanto, [Litman \(2022\)](#) também sugere que uma possível limitação dessas pesquisas é a falta de consideração de custos indiretos, como estacionamento residencial, o que pode levar a uma subestimação do ônus financeiro real. A aplicabilidade dessas pesquisas em países em desenvolvimento e diferentes contextos socioeconômicos também é discutida, exigindo uma análise mais detalhada.

O estudo realizado por [Mishra, Aithal et al. \(2022\)](#) detalha a coleta de dados primários sobre as características atuais das residências e viagens, características de tráfego, acessibilidade e nível de mobilidade dentro da área pesquisada. Nesse trabalho, os autores concluem que as informações sobre as características pessoais, domésticas e de viagem podem ser obtidas por meio de um método de questionário domiciliar. Esses dados são utilizados para conduzir uma pesquisa de Origem e Destino. Os autores destacam que essa pesquisa pode ajudar a visualizar a acessibilidade e o cenário de mobilidade da rede rodoviária e do transporte público. Essa abordagem permite uma análise detalhada das necessidades e características de transporte da população, contribuindo para um planejamento de transporte mais eficaz e equitativo. Destacando o potencial dessa abordagem para a construção de dados qualitativos para a análise do sistema.

As diferentes abordagens utilizando pesquisas domiciliares no planejamento do sistema de transporte revelam-se como uma ferramenta essencial, mas complexa. A integração de tecnologias modernas, como GPS e *smartphones*, com métodos tradicionais de pesquisa, oferece oportunidades para aprimorar a precisão e eficiência dessas pesquisas. No entanto, os desafios associados aos custos, confiabilidade dos dados e aplicabilidade em diferentes contextos exigem uma consideração cuidadosa. A combinação de diferentes técnicas e a adaptação às necessidades específicas da região podem levar a uma representação mais fiel da realidade das viagens e contribuir significativamente para o planejamento e gestão eficazes do sistema de transporte. A evolução contínua dessas metodologias, considerando as limitações e potenciais inovações, permanece fundamental para a sustentabilidade e equidade no desenvolvimento do transporte urbano.

2.2 Planejamento e operação do sistema de transporte por meio de dados de cartões magnéticos de cobrança

Com o desenvolvimento dos sistemas de cobrança automática (do inglês *Automated Fare Collection* (AFC)), uma nova possibilidade para obter dados e melhorar o planejamento e o monitoramento do sistema de transporte passa a ser utilizada. Desse modo, vários são os trabalhos que utilizam dados do AFC para obter indicadores que possibilitem a melhoria do sistema de transporte urbano.

Segundo [Kurauchi e Schmöcker \(2017\)](#), a utilização dos cartões magnéticos de cobrança (*smart cards*) no sistema de transporte possibilita maior conveniência para os passageiros, facilita o trabalho dos operadores do transporte, reduz a utilização de dinheiro nos ônibus e, além disso, permite a integração dos sistemas de pagamento com os sistemas de gerenciamento do transporte.

[Lathia et al. \(2013\)](#) utilizam dados de cartões para encontrar os padrões individuais de comportamento dos usuários do transporte coletivo da cidade de Londres, Inglaterra. Diferente de algumas estratégias existentes, onde são analisados os comportamentos de grupos de indivíduos, esse trabalho realiza uma análise sobre o comportamento individual dos passageiros baseado na utilização dos cartões magnéticos. Nesse trabalho, são empregados dendogramas e outros algoritmos para identificar as diferenças de comportamento entre cada usuário do sistema. Assim, os autores propõem uma estratégia capaz de melhorar a acurácia das previsões de tempo de viagem e compreender as características que diferentes granularidades nas análises podem trazer.

[Ma e Wang \(2014\)](#) empregam modelos de aprendizado de máquina, como as árvores de decisão bayesiana, sobre dados de cartões magnéticos de cobrança em conjunto com valores de posicionamento obtidos pelos sinais do GPS para realizar a construção de matrizes O-D. Os autores utilizam essa estratégia para propor um *framework* capaz de obter indicadores como velocidade da rede, confiabilidade do tempo de viagem, número de passageiros no nível de parada e variação de intervalo entre viagens. Esse *framework* desenvolvido, como relatado pelos autores, necessita de grandes investimentos, pois demanda um alto poder computacional. Além disso, é necessário realizar mais estudos em situações em que os ônibus não são tarifados segundo a distância percorrida.

[Zhao et al. \(2014\)](#) também empregam dados de cartões magnéticos de cobrança. Entretanto, aplicam o algoritmo *K-means* para encontrar características similares e possíveis anomalias nesses dados. A partir dessa estratégia, os autores buscam identificar padrões de comportamento temporais e espaciais dos passageiros no metrô da cidade de Shenzhen, China. Os resultados mostram que existem grupos de passageiros com características

similares e que existe uma tendência desses passageiros em manter esse comportamento. Nesse estudo, não é mostrado como essa estratégia pode auxiliar no planejamento e no monitoramento do sistema de transporte.

Costa et al. (2015) comparam 3 diferentes métodos (árvore de decisão, Naive Bayes e Top-K) para a classificação de passageiros em quatro diferentes grupos. Foram selecionadas características temporais e espaciais das viagens. Esses dados foram obtidos por meio de cartões de cobrança da cidade de Porto, Portugal. Os resultados alcançados mostram que, dependendo da situação em análise, cada um desses algoritmos possui uma melhor aplicabilidade. Em cenários mais específicos, em que existem diferenças entre a quantidade de dados históricos disponíveis, o dia da semana escolhido para análise e quais são os padrões das viagens, pode ser mais interessante selecionar um ou mais desses algoritmos. Assim, não é definida qual é a melhor estratégia para a classificação de passageiros, o que gera uma maior complexidade para o planejamento do sistema de transporte.

No trabalho de Sun et al. (2015), os dados de pagamentos por meio de cartões magnéticos, no metrô de Singapura, são utilizados para inferir as rotas dos passageiros. Modelos de inferência bayesiana e o algoritmo de Metropolis são empregados nesse trabalho. Os resultados mostram que o *framework* desenvolvido pode ser empregado para avaliar o fluxo de passageiros no metrô. Entretanto, algumas premissas como as rotas possíveis dos passageiros serem independentes e normalmente distribuídas não foi, empiricamente, comprovada. Desse modo, em aplicações reais, pode existir uma inferência equivocada sobre o fluxo dos passageiros.

No trabalho de Toqué et al. (2016), os autores indicam que as matrizes O-D possuem um papel fundamental para o desenvolvimento de um eficiente sistema de transporte urbano, pois é possível, por meio dessas matrizes, decidir sobre a correta alocação de recursos. Os autores aplicam as Redes Neurais Recorrentes (RNR) do tipo *Long short-term memory* (LSTM) sobre os dados de cartões de transporte no metrô de Rennes, na França, para a construção da matriz O-D de maneira dinâmica. Os resultados mostram que a utilização das redes LSTM em conjunto com outras fontes de dados pode ser empregada para a construção de matrizes O-D dinâmicas e a precisão alcançada mostra-se superior à utilização de modelos tradicionais como *Vector Autoregressive* (VAR) ou a utilização do contagem média diária de passageiros. Quando apenas dados de cartões foram utilizados, a acurácia do modelo empregando LSTM não foi muito alta. A partir da adoção de dados externos, o desempenho foi melhorado. Com a abordagem relatada, é possível construir matrizes O-D de maneira dinâmica por meio de dados dos cartões de metrô em conjunto com dados externos.

Ding et al. (2017) propõem um modelo para análise do fluxo de passageiros por meio da utilização dos cartões de ônibus. Para isso, é empregado o *Hadoop MapReduce*. O

framework Hadoop é utilizado, uma vez que possibilita trabalhar com grandes volumes de dados e reduzir a latência das análises. Este artigo demonstra que a utilização de *frameworks* distribuídos na análise do comportamento do fluxo de passageiros auxilia na construção de sistemas mais eficientes.

Kurauchi e Schmöcker (2017) apresentam um conjunto de diferentes estratégias para, a partir dos dados coletados pelos cartões magnéticos de cobrança, obter alguns *Key Performance Indicator* (KPI) que possibilitem o planejamento e o gerenciamento do sistema de transporte. Indicadores como a quantidade de vendas de passagens em uma viagem, veículos por quilômetro, passageiro por quilômetro, aderência ao itinerário, variação da demanda durante o dia e taxa de carregamento dos veículos são construídos pelos dados desses cartões. Além desses KPI, são apresentados alguns modelos para a construção da matriz O-D. Nesse trabalho, os autores também apresentam estratégias para a identificação dos motivos de viagens dos passageiros e para medir a qualidade do sistema. Entretanto, esse trabalho realiza apenas análises para o transporte público de passageiros e os métodos empregados necessitam que várias premissas sejam verificadas. Premissas como o local de destino da última viagem ser idêntico ao local de origem do passageiro e que o passageiro deve pegar o próximo ônibus assim que deixar o local de transferência nem sempre são verdadeiras, pois essas características estão relacionadas a questões sociais e culturais (JAMAL et al., 2019).

Como apresentados em alguns trabalhos existentes na literatura (KURAUCHI; SCHMÖCKER, 2017; LI et al., 2018), para a adoção dos dados obtidos por meio do AFC, algumas considerações devem ser realizadas:

- **A taxa de passageiros que utilizam os cartões.** Por mais que existam cidades que possuem uma taxa de utilização de cartões superior a 80%, podem existir grandes diferenças entre o comportamento dos passageiros que utilizam e não os cartões.
- **A seleção de usuário.** Usuários de cartão, normalmente, são usuários com características próprias como os estudantes, trabalhadores, aposentados etc. Desse modo, a análise pode ser enviesada.
- **A utilização do cartão por outros usuários.** Usuários do sistema de transporte que recebem cartões com passagens e não utilizam o transporte público podem comercializar esses bilhetes eletrônicos, que passam a ser utilizados por outros usuários.
- **Sistemas em que os usuários apenas precisam utilizar o cartão no momento do embarque em uma estação.** Assim, podem desembarcar em outras estações ou até em locais externos às estações. Logo, para esse tipo de transporte, algumas das premissas adotadas não podem ser verificadas.

A adoção de dados de cartões magnéticos de cobrança no planejamento e operação do sistema de transporte tem emergido como uma ferramenta valiosa para a análise e melhoria do transporte urbano. Essa abordagem permite uma compreensão mais profunda do comportamento dos passageiros e facilita a construção de indicadores de desempenho. No entanto, a implementação desses métodos requer considerações cuidadosas, como a taxa de utilização dos cartões e o potencial enviesamento na seleção de usuários. Apesar desses desafios, a integração desses dados com tecnologias modernas podem gerar avanços significativos na eficiência e gerenciamento do transporte público.

2.3 Planejamento e operação do sistema de transporte por meio de dados de redes de telefonia

Os Registros Detalhados de Chamadas (do inglês *Call Detail Records* (CDR)) são registros de interações entre os usuários e os sistemas de telecomunicação. Quando qualquer interação entre o usuário e a rede de telecomunicação ocorrer, são gravados os dados dessa interação. Dados como a data e a hora de início e fim da interação, antenas participantes da rota de transmissão e tipo de interação são coletados e armazenados. As interações podem ocorrer, por exemplo, por meio de chamadas, envio de *Short Message Service* (SMS) e conexões com a internet. Portanto, os CDR correspondem a uma importante fonte de dados para o planejamento dos sistemas de transporte.

Järv, Ahas e Witlox (2014) utilizam os dados de CDR para realizar uma análise sobre a variação de comportamento dos usuários do sistema de transporte. Esse trabalho utiliza dados coletados em um período de 12 meses para identificar a variação de comportamento das viagens dos indivíduos durante o ano. Foram selecionados dados da Estônia em uma região específica denominada de Tallinn. Os resultados indicam que existe uma pequena variação de deslocamentos para atividades entre meses consecutivos e, dentro de um mês, ocorre uma maior variação das atividades realizadas. Desse modo, é possível identificar quais são os motivos da viagens dos indivíduos e como ocorre esse deslocamento espacial. Apesar de não enfatizado no trabalho, esses dados podem auxiliar no planejamento e no monitoramento do sistema de transporte, pois é possível identificar as características das regiões e oferecer diferentes modos de transporte que atendam as particularidades da região.

No trabalho de Berlingerio et al. (2014), os autores desenvolveram um sistema chamado de AllAboard que utiliza os dados de CDR para extrair a origem e destino dos usuários, inferir uma matriz O-D e sugerir novas rotas que otimizem o tempo de viagem a partir do modelo de matriz O-D construído. Aplicando o sistema para a cidade de Abidjan, Costa do Marfim, os autores indicam que é possível, após a sugestão de novas rotas, reduzir em até 10% o

tempo de viagem dos passageiros. Contudo, mesmo apresentando esses resultados, por ser um sistema comercial, em países em desenvolvimento, a utilização desses sistema pode ser comprometida devido aos custos de aquisição, treinamento de pessoal e manutenção.

[Iqbal et al. \(2014\)](#) também utilizam dados de CDR em conjunto com os dados de laços magnéticos para a contagem de veículos a fim de desenvolver uma metodologia para a construção de matrizes O-D. Esses dados foram extraídos da cidade de Dhaka, Bangladesh. Os resultados, quando comparados com simulações realizadas por meio do *software* MITSIMLab, indicam um *Root Mean Square Percent Errors* (RMSPE) próximo a 14%. Desse modo, os resultados são promissores, mas uma série de simplificações como a função objetivo utilizada para identificar o fator de escala (necessário para encontrar a variação real dos deslocamentos), pode levar a matriz O-D a não representar a realidade dos deslocamentos urbanos.

Os autores [Alexander et al. \(2015\)](#) utilizam dados de CDR da cidade de Boston, nos Estados Unidos, para desenvolver matrizes O-D diárias. Utilizando algoritmos de clusterização hierárquicos, os autores empregam a frequência de viagens, dia da semana da viagem e os horários das viagens para inferir o local de residência, trabalho e atividades de lazer dos indivíduos. Assim, é possível construir as matrizes O-D para a região analisada. Os resultados mostram que, a partir de algumas premissas, pode ser desenvolvida uma metodologia para obter matrizes O-D utilizando os CDR. Como não foram identificados os tipos de veículos utilizados pelos indivíduos, definir uma política de transporte baseada nessas matrizes torna-se mais complexa, necessitando, assim de outros dados para a realização de um planejamento eficiente.

[Jiang, Ferreira e Gonzalez \(2017\)](#) empregam os CDR, dados coletados por meio de censo e de limites geográficos, a fim de construir um *framework* capaz de identificar qual é o propósito da viagem dos usuários do sistema de transporte em Singapura. O *framework* proposto apresenta resultados promissores, mas a identificação do modo de transporte não foi evidenciada nesse estudo. Desse modo, é necessário que outras estratégias sejam utilizadas em conjunto para assegurar um planejamento eficiente do sistema de transporte.

[Lwin, Sekimoto et al. \(2020\)](#) utilizam os dados extraídos a partir do CDR para a identificação do modo de transporte dos usuários. Esses dados foram extraídos e analisados por meio de uma perspectiva dos países em desenvolvimento, uma vez que foram retirados a partir de dados das companhias telefônicas da cidade de Yangon, em Mianmar. Ao realizar a análise dos dados, os autores utilizam as diferentes velocidades de deslocamento do usuário a fim de inferir qual é o meio de transporte utilizado. Os resultados, quando comparados a entrevistas realizadas com os usuários, mostram que é possível aplicar esse método para identificar o meio de transporte utilizado. Entretanto, nesse trabalho, os autores não demonstram como podem ser gerados indicadores para o planejamento do sistema mediante

o resultado obtido. Além disso, empregar essa metodologia para a definição dos modos de transporte durante viagens curtas pode enviesar as análises (GOLDING, 2018).

Barboza et al. (2021) exploram a utilização de CDR para identificar padrões de mobilidade humana na área metropolitana do Rio de Janeiro. Os autores integram os CDRs com dados demográficos e outras fontes para modelar o transporte. No entanto, o estudo também aponta limitações, como a falta de informações precisas sobre a localização exata dos indivíduos e a possibilidade de viés devido à propriedade e uso de telefones móveis. Apesar dessas restrições, eles concluem que os CDRs representam uma opção robusta e de baixo custo para o planejamento e monitoramento urbano, mas enfatizam a necessidade de considerar essas limitações na interpretação dos resultados. Desso modo, integrar os CDR com outras fontes de dados pode auxiliar na interpretação dos resultados.

O trabalho realizado por Zhao, Koutsopoulos e Zhao (2022) utiliza CDR para analisar padrões de mobilidade humana e informações de transporte urbano. Os autores destacam a capacidade dos CDRs em estimar origem-destino e detectar atividades humanas, mas também apontam limitações. A principal restrição é a dispersão dos dados de CDR, já que o uso do telefone é esporádico, resultando em períodos sem registros, o que pode levar a estimativas de viagem propensas a erros. Eles concluem que uma abordagem supervisionada é necessária para superar essas limitações.

Tsumura et al. (2022) exploram o uso de CDRs no planejamento de transporte, especialmente em países em desenvolvimento. Os autores enfatizam que os CDRs oferecem uma abordagem econômica para estimar a distribuição temporal das pessoas, contrastando com os métodos convencionais. No entanto, eles também discutem limitações e restrições práticas, incluindo possíveis impactos das medidas de proteção de dados, o que pode afetar a eficácia dos CDRs como uma ferramenta de planejamento de transporte.

A utilização de CDR no planejamento e operação do sistema de transporte tem se mostrado uma abordagem promissora, oferecendo *insights* valiosos sobre o comportamento dos usuários e permitindo a otimização de rotas e tempos de viagem. A integração dos CDRs com outros dados, como demográficos e de localização, tem possibilitado a construção de modelos mais precisos e robustos. No entanto, a aplicação desses registros também apresenta limitações, como a falta de informações exatas sobre a localização dos indivíduos, a possibilidade de viés, e desafios relacionados à proteção de dados e custos em alguns contextos. Essas restrições enfatizam a necessidade de uma abordagem cuidadosa e integrada, combinando CDRs com outras fontes de dados e considerando o contexto específico da região, para assegurar um planejamento eficiente e responsável do sistema de transporte.

2.4 Planejamento e operação do sistema de transporte por meio de dados de diferentes fontes

O planejamento e operação eficientes do sistema de transporte são cruciais para a mobilidade urbana e o desenvolvimento sustentável das cidades. Com o avanço tecnológico e a crescente disponibilidade de dados, a integração de diferentes fontes de informações, como *Big Data*, CDR, GPS, e sensores inteligentes, tem se tornado uma abordagem promissora. Essa integração permite uma análise mais profunda e precisa do comportamento dos usuários, fluxo de tráfego, e padrões de mobilidade. No entanto, a aplicação dessas fontes de dados também traz desafios e complexidades, como a necessidade de garantir a qualidade, segurança, e privacidade dos dados, além de considerar o contexto específico de cada região. Esta seção explora as oportunidades e restrições na utilização de diferentes fontes de dados para o planejamento e operação do sistema de transporte, destacando as inovações, métodos, e casos de sucesso na área.

No trabalho realizado por [Lv et al. \(2014\)](#), é destacada a aplicação de técnicas de inteligência artificial, especificamente o algoritmo *Stacked Autoencoders* (SAE), para aprimorar o planejamento do sistema de transporte urbano. Utilizando dados do *Caltrans Performance Measurement System* (PeMS), os autores aplicam o SAE para prever o fluxo de veículos, e os resultados são avaliados através de métricas como o *Root-Mean-Square Error* (RMSE). Essa abordagem demonstrou ser superior aos modelos paramétricos tradicionais. No entanto, o estudo se concentra apenas nesse indicador específico de planejamento e monitoramento, deixando de explorar outros possíveis indicadores. Portanto, apesar dos resultados promissores, o trabalho sugere a necessidade de desenvolver mais indicadores para implementar o planejamento do transporte de forma mais abrangente e efetiva.

[Oort e Cats \(2015\)](#) apresentam dois casos de sucesso na utilização do *Big Data* em conjunto com a inteligência artificial para o planejamento e análise do sistema de transporte. Esses dois estudos de casos estão localizados, respectivamente, na Holanda e na Suécia. Na Holanda, como existe um sistema de transporte público completamente integrado por meio dos cartões de cobrança, é possível realizar análise do comportamento dos passageiros e realizar previsões de demanda. Na Suécia, também são utilizados dados de cartões de cobrança e AVL para identificar o comportamento dos usuários e encontrar indicadores de qualidade do sistema de transporte. Esses dois casos de sucesso mostram o potencial que diferentes fontes de dados possuem para o planejamento do sistema de transporte. Entretanto, esses casos estão concentrados em países que possuem características específicas (desenvolvidos com pequena população). Para a realização de um planejamento eficiente em países em desenvolvimento, por exemplo, é necessário que estratégias mais abrangentes sejam utilizadas.

No trabalho de [Toole et al. \(2015\)](#), é apresentado um sistema que utiliza dados de CDR, localização, censo e pesquisas realizadas com os passageiros do transporte para propor a construção da matriz O-D. Além disso, neste trabalho, é apresentado um sistema de *software* capaz de exibir a representação dessa matriz e dar suporte à decisão. O desempenho desse sistema foi avaliado utilizando dados de 5 diferentes cidades (Boston e San Francisco, nos EUA, Lisboa e Porto, em Portugal, e Rio de Janeiro, no Brasil). Os resultados mostram a aplicabilidade do sistema para a construção da matriz O-D. Entretanto, mesmo com a utilização do *Big Data*, como são empregados dados do CDR, problemas como a maior dificuldade em traçar rotas de menor distância continuam sendo evidenciados.

No trabalho de [Herrera-Quintero et al. \(2016\)](#), é apresentada a implementação de um sensor inteligente capaz de detectar sinais de bluetooth dos *smartphones* para a construção das matrizes O-D em veículos do Bus Rapid Transit (BRT) na cidade de Bogotá, Colômbia. Os testes realizados mostram que o sistema construído pode ser empregado como uma das fontes de dados para o planejamento do transporte público. Como a construção da arquitetura e os testes foram restritos ao BRT, estender esse trabalho para as diferentes linhas de ônibus que não pertencem ao sistema de BRT, faz com que essa estratégia necessite de outros diferentes sistemas para apoiar o planejamento e o monitoramento do transporte urbano.

[Shukla, Balachandran e Sumitha \(2016\)](#) apresentam um *framework* composto por 4 camadas (entrada, armazenamento, análise e comunicação) a fim de possibilitar a utilização de dados provindos de sensores e de sistemas de controle de tráfego para a geração de conhecimento. Esse *framework* reduz as dificuldades de implementação de estratégias que utilizem o *Big Data*. A qualidade dos dados, a conectividade e a velocidade de análise são algumas das características empregadas por esse *framework*. Apesar de auxiliar na construção, integração e análise de dados do transporte, a aplicabilidade desse *framework* não foi testada pelos autores. Portanto, para se comprovar essa eficácia, é necessário que exista a implementação desse *framework* em aplicações reais.

[Rathore et al. \(2016\)](#) propõem a construção de cidades inteligentes mediante a utilização do *Big Data* em uma estrutura de 4 camadas capaz de integrar redes veiculares, casas inteligentes, estacionamentos e segurança. Essa arquitetura é similar à proposta por [Jan et al. \(2019\)](#), mas não é específica para o sistema de transporte. Portanto, existe uma maior complexidade para a implementação e análise do sistema. Mesmo os resultados mostrando que essa arquitetura pode ser empregada sobre diferentes base de dados, para aplicações de ITS, existe a necessidade de adaptações da estrutura proposta.

No estudo conduzido por [Torre-Bastida et al. \(2018\)](#), os autores exploram os desafios significativos associados à utilização efetiva do *Big Data* nos sistemas de transporte. Esses desafios incluem a complexidade da análise de dados em tempo real, questões de

segurança e privacidade, e a necessidade de trabalhar de maneira eficiente com diversas fontes de dados, cada uma com características específicas. Essas complexidades tornam evidente que o uso do *Big Data* no planejamento do sistema de transporte urbano não é uma tarefa simples. Requer o desenvolvimento de *frameworks* especializados que sejam capazes de lidar com esses desafios, garantindo que as informações sejam utilizadas de maneira responsável e eficaz para melhorar o planejamento e a operação do transporte.

Zin et al. (2018) utilizam dados do CDR para estimar a matriz O-D. Além desses dados, foram utilizados dados de sensores para estimar o comportamento das viagens. Para ajustar os problemas gerados pela utilização do CDR como estimador dos movimentos da população, é utilizado um fator de correção para estender a análise na construção da matriz O-D. Os resultados obtidos demonstram que essa estratégia pode de ser utilizada como forma de construção da matriz O-D. Entretanto, a disponibilidade desses dados exige parcerias com as empresas de telefonia móvel, o que, muitas vezes, não é fácil conseguir.

Welch e Widita (2019) indicam a importância da utilização do *Big Data* para a tomada de decisão nos sistemas de transporte urbano. Nesse trabalho, os autores indicam que, apesar da relevância do *Big Data* em diferentes pesquisas, existem poucos trabalhos que utilizam técnicas estatísticas e ferramentas mais avançadas que foram concebidas para lidar, especificamente, com o *Big Data*. Por exemplo, apenas 10% dos trabalhos empregam técnicas de aprendizado de máquina para geração de conhecimento e um número menor de trabalhos utiliza *frameworks* como o Hadoop e Spark para o tratamento e o processamento de um grande volume de dados. Desse modo, empregar técnicas modernas capazes de trabalhar com os desafios impostos pelo *Big Data* para o planejamento e para o gerenciamento do sistema de transporte é essencial para a melhoria do sistema de transporte nas cidades.

Babar e Arif (2019) propõem o desenvolvimento de uma arquitetura composta de 3 fases (*Big Data Service Managemen*, *Big Data Processing and Analysis* e *Big Data Organization and Management*), capaz de realizar o gerenciamento em tempo real do sistema de transporte por meio dos dados coletados por diferentes fontes. A arquitetura utiliza o *framework Apache Spark* para o processamento em tempo real por meio do *streaming* de dados. Essa arquitetura é utilizada para a construção de diferentes gráficos como a velocidade média dos veículos e o tempo de chegada. Os resultados indicam que essa arquitetura pode ser empregada para o gerenciamento do sistema de transporte. Esse trabalho apresenta uma arquitetura de *software* que pode ser utilizada para o gerenciamento em tempo real do sistema de transporte. Entretanto, não são apresentados indicadores ou novos métodos para a planejamento do transporte urbano.

Jan et al. (2019) também utilizam Apache Spark e Hadoop para o desenvolvimento de um sistema capaz de realizar o processamento de dados em tempo real. Nesse trabalho, os

autores propõem a construção de um esquema de 4 camadas (coleta de dados, comunicação, processamento e aplicação) para a análise de dados coletados de diferentes fontes. O sistema proposto foi testado utilizando dados de estacionamento e da contagem de veículos mediante laços magnéticos. Os resultados obtidos mostram que essa arquitetura em 4 camadas pode ser utilizada para a análise de dados do sistema de transporte. Entretanto, nesse artigo não é realizada a construção ou a análise de indicadores do sistema de transporte. Desse modo, essa arquitetura funciona apenas como suporte à coleta e à análise de dados do sistema de transporte.

[Kaffash, Nguyen e Zhu \(2020\)](#) apresentam uma revisão da literatura sobre a utilização de diferentes algoritmos e estratégias para análise do *Big Data* no campo do ITS. Esse trabalho mostra a relevância do tema para o planejamento e para o monitoramento do sistema de transporte. Por meio da revisão de mais de 580 trabalhos, é evidenciado que a utilização das técnicas de inteligência artificial sobre *Big Data* representa uma tendência para os estudos no campo do transporte inteligente.

A utilização de diferentes fontes de dados como redes sociais, *smartphones*, GPS e sensores inteligentes para a construção e para a análise do *Big Data* passam a ser cada vez mais relevantes para o planejamento e monitoramento do ITS. Entretanto, caso algumas características não sejam observadas durante o processo de coleta e análise de dados, podem ser introduzidos vieses. [Griffin et al. \(2020\)](#) apresentam um estudo que divide esses possíveis vieses em 4 diferentes categorias e indica ações que devem ser tomadas para mitigar a influência desses vieses no planejamento e no monitoramento do sistema de transporte. Desse modo, é possível utilizar a amplitude dos dados do *Big Data* sem comprometer a qualidade das análises.

[Iliashenko, Iliashenko e Lukyanchenko \(2021\)](#) exploram as complexidades e desafios do planejamento de transporte nas cidades em crescimento. Eles indicam que tecnologias como *Big Data*, Aprendizado de Máquina e Internet das Coisas (IoT) podem ser integradas para resolver problemas como congestionamentos de tráfego e poluição do ar. A pesquisa inclui exemplos de otimização de rotas de transporte em Moscou e coleta de informações sobre fluxos de tráfego em Kazan, utilizando essas tecnologias. No entanto, os autores também destacam limitações significativas, como a necessidade de infraestrutura tecnológica robusta, desafios na integração de diferentes fontes de dados, preocupações com a privacidade e segurança dos dados, e a complexidade na interpretação e aplicação dos *insights* derivados dessas tecnologias. Nesse trabalho, eles concluem que, apesar do potencial promissor, a adoção de *Big Data* e aprendizado de máquina no planejamento de transporte requer uma abordagem cuidadosa e integrada para superar essas limitações e garantir um planejamento eficiente e responsável.

O trabalho apresentado por [Lemondé, Arsenio e Henriques \(2021\)](#) fornece *insights* sobre

as dinâmicas complexas do tráfego urbano em Lisboa, destacando a integração de várias fontes de dados, como análises de *Big Data* e Aprendizado de Máquina. Os autores enfatizam o uso de fontes de dados históricas e prospectivas, incluindo condições climáticas, eventos públicos e interdições de estradas, para aprimorar a modelagem e previsão do tráfego. Eles também exploram a aplicação de camadas de redes neurais recorrentes, como componentes de memória de longo e curto prazo, para melhorar as previsões de tráfego, incorporando tanto o contexto histórico quanto o prospectivo. No entanto, o estudo também aponta vários desafios, como a natureza estocástica do tráfego, o tamanho massivo dos dados produzidos e a presença de mudanças emergentes no tráfego urbano que podem facilmente tornar estudos estáticos irrelevantes. Esses desafios exigem requisitos rigorosos de escalabilidade e a capacidade de aprender com os fluxos de dados de tráfego para detectar padrões emergentes, refletindo as mudanças contínuas na mobilidade.

[Pettit et al. \(2022\)](#) destacam a transformação na abordagem do planejamento e operação do sistema de transporte urbano, impulsionada pelo uso de *Big Data*. Eles indicam que tecnologias como análise de dados de cartões inteligentes de transporte, *smartphones* e mídias sociais podem ser utilizadas para entender os padrões de mobilidade e construir modelos mais eficientes. No entanto, esse trabalho também aponta limitações, como a necessidade de uma compreensão adequada das forças e limitações desses dados, e enfatiza o papel essencial da coleta de dados propositada. Nesse trabalho, os autores concluem que, embora o *Big Data* ofereça promessas significativas para o planejamento futuro, a adoção de fontes de dados mais confiáveis e complementares deve ser realizada para assegurar uma tomada de decisão eficaz e responsável no sistema de transporte urbano.

A utilização de diferentes fontes de dados, incluindo *Big Data* e técnicas de inteligência artificial, no planejamento e operação do sistema de transporte urbano, tem emergido como uma estratégia inovadora, fornecendo análises profundas e permitindo a melhoria da eficiência e qualidade do transporte. A combinação dessas tecnologias com dados específicos, como CDR, GPS, e sensores inteligentes, tem facilitado a construção de modelos mais precisos e a previsão do fluxo de tráfego. No entanto, essa abordagem também traz desafios e limitações, como a necessidade de mitigar possíveis vieses, garantir a segurança e privacidade dos dados, e considerar os custos e características específicas de cada região. Essas complexidades indicam a importância de uma implementação cuidadosa e adaptada, integrando diversas fontes de dados e tecnologias, para garantir um planejamento e monitoramento eficientes e responsáveis do sistema de transporte.

Capítulo 3

Fundamentação Teórica

Neste capítulo, serão apresentados os principais conceitos que formam a base teórica desta tese, incluindo temas como Cidades Inteligentes, Matriz Origem-Destino, Inteligência Artificial e Filtro de Kalman. A discussão dessas ideias é essencial para contextualizar o estudo e conectar os trabalhos anteriores na área com a abordagem adotada nesta pesquisa. A exploração desses tópicos ajudará a fornecer uma compreensão clara do que será discutido nas próximas seções.

3.1 Cidades Inteligentes

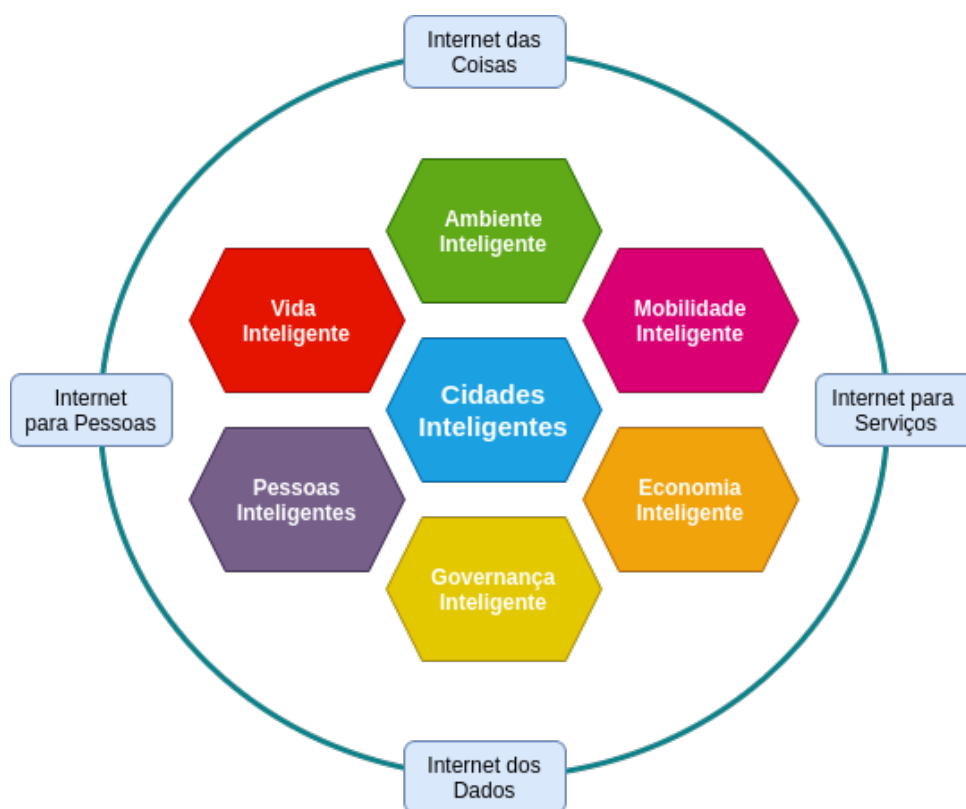
O conceito de Cidades Inteligentes, ou *Smart Cities*, tem evoluído significativamente desde sua introdução na década de 1990, quando estava ligado principalmente à construção de infraestrutura urbana utilizando tecnologias de informação e comunicação (ICT - do inglês *Information and Communications Technology*) ([ALAWADHI et al., 2012](#)). Hoje, essa ideia se expandiu para abranger uma visão mais holística e integrada da vida urbana, focando na interconexão, sustentabilidade, conforto e segurança. Essa evolução reflete a crescente necessidade de soluções inovadoras para enfrentar os desafios complexos das cidades modernas. No contexto deste trabalho, o conceito de Cidades Inteligentes é particularmente relevante, pois o objetivo é propor um *framework* de baixo custo que utilize diferentes fontes de dados para melhorar o planejamento, gerenciamento e operação do sistema de transporte, especialmente em países em desenvolvimento. A abordagem das Cidades Inteligentes oferece um caminho promissor para alcançar esses objetivos, integrando tecnologia e inovação para criar soluções eficazes e sustentáveis.

Com o passar do tempo e o surgimento de novas tecnologias, várias outras definições surgiram, buscando traduzir a importância dos diferentes modelos tecnológicos na construção de cidades sustentáveis. [Albino, Berardi e Dangelico \(2015\)](#) apresentam várias diferentes visões sobre o conceito das Cidades Inteligentes. Dentre os diferentes conceitos apresentados, uma definição reduzida, mas que transmite o objetivo principal para a construção das

Cidades Inteligentes, é apresentada por [Lazaroiu e Roscia \(2012\)](#). Esses autores definem as Cidades Inteligentes como uma comunidade tecnológica, interconectada, sustentável, confortável, atraente e segura. Desse modo, além das tecnologias adotadas, para que um centro urbano seja “inteligente”, é necessário que os diferentes sistemas que constituem uma cidade estejam conectados e gerem valor para toda a comunidade.

Sendo assim, as Cidades Inteligentes podem ser vistas como um ambiente urbano inteligente, que utiliza as tecnologias de informação e comunicação embarcada para a construção de cidades interativas. Essa interatividade permite a integração do mundo virtual com o real, facilitando a solução de problemas urbanos e um melhor gerenciamento dos recursos e espaços públicos. Autores como [Albino, Berardi e Dangelico \(2015\)](#) questionam a definição de que as Cidades Inteligentes são apenas tecnológicas. Para [Albino, Berardi e Dangelico \(2015\)](#) e outros autores, as Cidades Inteligentes devem ser vistas como espaços da pluralidade e sustentabilidade e que possibilitem o bem estar geral da população. Assim, as Cidades Inteligentes podem ser vistas como um conjunto de sistemas integrados ([KHATOUN; ZEADALLY, 2016](#)). A [Figura 7](#) apresenta o modelo de construção das Cidades Inteligentes.

Figura 7 – Modelo conceitual das Cidades Inteligentes



Fonte: Adaptado de [Khatoun e Zeadally \(2016\)](#)

Dentre os 6 sistemas descritos na [Figura 7](#), a Mobilidade Inteligente se apresenta como fundamental para o sucesso das Cidades Inteligentes. A maneira como ocorre o deslocamento

de pessoas e mercadorias dentro das cidades está, diretamente, relacionada à qualidade de vida dos cidadãos, à economia e ao meio ambiente. Desse modo, utilizar o desenvolvimento tecnológico para a construção de uma mobilidade cada vez mais inteligente é essencial para a implantação e disseminação do conceito das Cidades Inteligentes.

O ITS, que representa a união de diversas tecnologias e aplicações, tem como objetivo compreender e otimizar a mobilidade urbana. Essa otimização tem como finalidade promover segurança e conforto aos motoristas e passageiros, além de assegurar a sustentabilidade do transporte nas cidades. O ITS utiliza sensores presentes na infraestrutura urbana e nos veículos de transporte a fim de proporcionar um melhor gerenciamento e monitoramento dos sistemas de transporte das cidades. Com um melhor planejamento e gerenciamento, é possível reduzir congestionamentos, diminuir o consumo de combustíveis fósseis, mitigar emissões de CO_2 e reduzir perdas monetárias.

Por meio dos dados coletados por esses sistemas presentes no ITS é possível, por exemplo, realizar uma análise mais profunda de todo o sistema de transporte a fim de identificar os períodos e as regiões da cidade mais frequentadas e os motivos para que essas regiões serem as mais frequentadas. [Kurauchi e Schmöcker \(2017\)](#) elencam a importância da coleta de dados em três diferentes frentes para o planejamento adequado dos sistemas de transporte.

- **Estudos estratégicos:** planejamento de longo prazo como previsão de demanda, *marketing* e classificação de usuários.
- **Estudos táticos:** determinação de padrões de comportamento dos usuários e adequação de linhas.
- **Estudos operacionais:** coleta de indicadores de *performance* e qualidade do serviço.

Para a realização de todos esses estudos, é necessário que alguns indicadores sejam determinados. São esses indicadores que norteiam as decisões estratégicas, táticas e operacionais na construção de uma mobilidade urbana cada vez mais inteligente. Vários são os trabalhos que mostram a utilização de diferentes indicadores para o planejamento do sistema de transporte. [Ferraz e Torres \(2001\)](#) mostram que os indicadores são importantes ferramentas para avaliar o desempenho do transporte público urbano, identificar pontos críticos e propor soluções para melhorar a qualidade do serviço oferecido aos usuários. Alguns indicadores apresentados por [Ferraz e Torres \(2001\)](#) são:

- **Índice de Passageiros Transportados por Quilômetro (IPTK):** mede a eficiência do transporte público em transportar passageiros por quilômetro de operação.

- **Índice de Passageiros Transportados por Hora de Pico (IPTH):** mede a eficiência do transporte público em transportar passageiros durante as horas de pico, quando a demanda é maior.
- **Taxa de ocupação:** mede a proporção de assentos ocupados em relação ao total de assentos disponíveis em um veículo.
- **Índice de Pontualidade:** mede a eficiência do transporte público em cumprir os horários programados de chegada e partida.
- **Índice de Confiabilidade:** mede a capacidade do transporte público em manter a regularidade e consistência nos horários e frequências dos serviços.
- **Índice de Segurança:** mede a segurança dos passageiros durante a utilização do transporte público.
- **Índice de Satisfação dos Usuários:** mede a satisfação dos usuários em relação ao transporte público, levando em consideração diversos aspectos, como qualidade de serviço, acessibilidade, conforto, segurança e informações.

Esses indicadores são ferramentas fundamentais para avaliar o desempenho do transporte público urbano. Eles não apenas medem a eficiência, regularidade e segurança do sistema, mas também ajudam a identificar pontos críticos que podem necessitar de intervenção. Por exemplo, um baixo IPTK pode indicar uma necessidade de otimização de rotas, enquanto uma Taxa de Ocupação baixa pode sinalizar um desperdício de recurso. Os Índices de Pontualidade e Confiabilidade são vitais para garantir que o serviço seja confiável e consistente, impactando diretamente na satisfação do usuário. A análise desses índices permite que os gestores do transporte público proponham soluções específicas para melhorar a qualidade do serviço oferecido aos usuários, seja por meio de ajustes nas rotas, horários, segurança ou outros aspectos que contribuam para uma experiência de transporte mais eficiente e satisfatória.

Nesse contexto, é fundamental compreender as características e as possibilidades geradas pelas Cidades Inteligentes, bem como reconhecer a importância dos indicadores no planejamento e no gerenciamento do sistema de transporte urbano. Essa compreensão alinha-se ao objetivo deste trabalho, que é propor um *framework* de baixo custo, capaz de utilizar diferentes fontes de dados para gerar e monitorar indicadores. A análise e aplicação desses índices têm o potencial de facilitar um planejamento, gerenciamento e operação mais eficazes do sistema de transporte. Isso contribui diretamente para a realização dos objetivos da tese e para a melhoria da qualidade do serviço oferecido aos usuários, com especial relevância em países em desenvolvimento.

3.2 Matriz O-D

A matriz O-D é uma ferramenta essencial para a gestão de sistemas de transporte público (PETERSON, 2007). Ela permite entender como as pessoas se deslocam dentro de uma determinada área geográfica, por meio da identificação das origens e destinos dos deslocamentos realizados pelos usuários do sistema, bem como da quantidade de viagens realizadas entre cada par de pontos.

A análise da matriz origem-destino é importante para identificar as demandas dos usuários do sistema de transporte público e planejar as rotas e horários de ônibus, trens, metrô e outras formas de transporte coletivo. Outra importante característica da matriz O-D é a possibilidade de identificar as áreas de maior concentração de demanda, bem como a frequência das viagens realizadas entre os diferentes pontos da cidade.

Existem diversas maneiras de coletar as informações que compõem a matriz origem-destino, como pesquisas por questionários aplicados aos usuários do transporte público, registros eletrônicos de passagens, imagens de câmeras de segurança nos pontos de ônibus e outras tecnologias de monitoramento de tráfego (GUERRA, 2011).

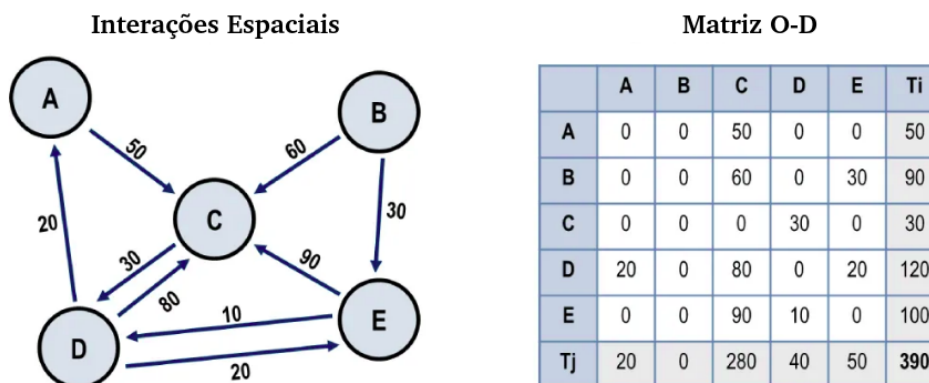
A matriz origem-destino pode ser representada em diversas formas, como mapas, gráficos e tabelas. Em sua forma mais comum, ela é apresentada em uma tabela que contém as origens e destinos dos deslocamentos realizados pelos usuários do sistema, bem como o número de viagens realizadas entre cada par de pontos (ZARGARI; MEMARNEJAD; MIRZAHOSSEIN, 2021).

Utilizando, por exemplo, a representação matricial, cada célula dessa matriz deve conter um número que representa o fluxo (quantidade de viagens) de um mesmo local de origem e de destino (EKOWICAKSONO; BUKHARI; AMAN, 2016). As linhas dessa matriz contêm a região de origem dessa viagem e as colunas contêm o destino dessa viagem. A Figura 8 apresenta um exemplo de uma matriz O-D hipotética obtida por meio de um grafo direcional. Na Figura 9 tem-se a representação gráfica de uma matriz O-D para a cidade de Maceió, Alagoas (GUERRA, 2011).

A partir da matriz O-D, é possível conhecer toda a dinâmica dos movimentos populacionais dentro do espaço urbano. De posse dessa movimentação, os gestores dos sistemas de transporte têm a capacidade de identificar quais são as áreas de maior atração das viagens, em quais horários ocorre essa movimentação, qual deve ser a quantidade de meios de transporte motorizados a ser ofertados para a população e, assim, otimizar os gastos e oferecer um sistema de melhor qualidade para a população.

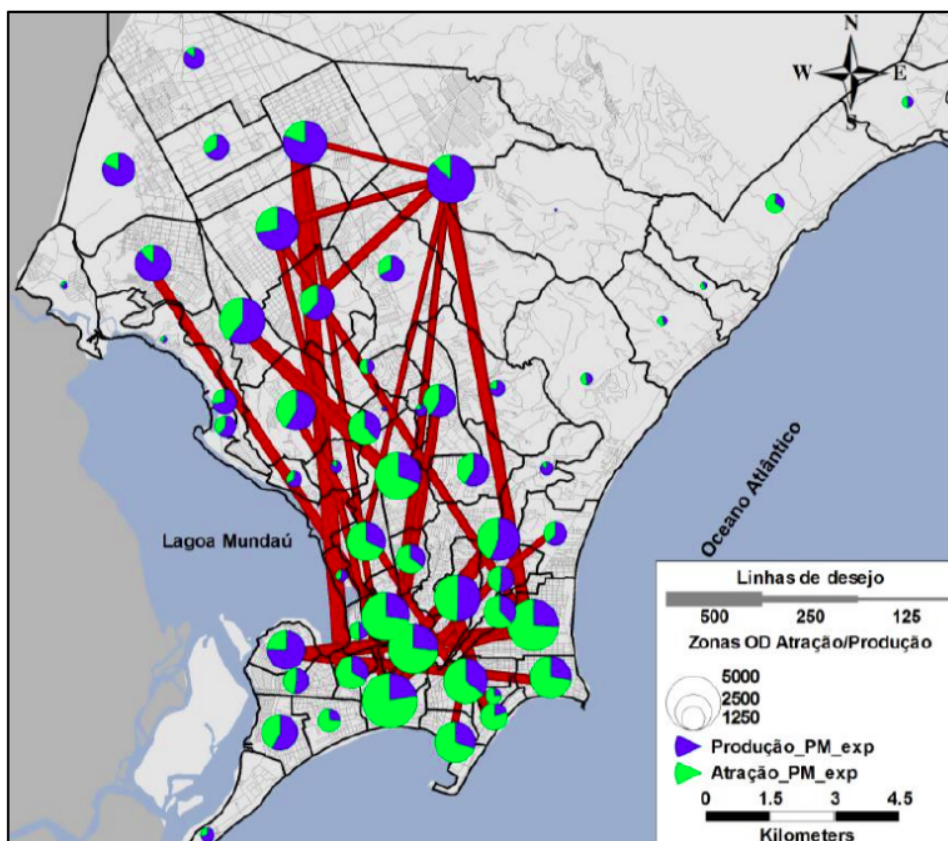
A construção da matriz O-D é feita por meio da coleta de dados sobre a origem e o destino das viagens realizadas pelos usuários do sistema. Existem diferentes métodos para estimar

Figura 8 – Exemplo de uma matriz O-D



Fonte: Adaptado de Winder (2016)

Figura 9 – Exemplo de uma matriz O-D para a cidade de Maceió/AL



Fonte: Adaptado de Guerra (2011)

a matriz O-D em um sistema de transporte público. Alguns desses métodos são (TIMMS, 2001):

- **Pesquisa de Origem e Destino (O-D):** Este é o método mais comum de coleta de dados para a estimação da matriz O-D. Neste método, é realizada uma pesquisa

por amostragem entre os usuários do sistema de transporte público para determinar as suas origens e destinos (ZARGARI; MEMARNEJAD; MIRZAHOSSEIN, 2021). A amostra é selecionada a partir de um universo representativo de usuários do sistema, e os dados coletados são extrapolados para a população total. A pesquisa de O-D pode ser realizada por meio de entrevistas pessoais e questionários, por exemplo.

- **Registro Eletrônico de Passagens (REP):** Este método utiliza a informação coletada pelos sistemas de bilhetagem eletrônica para estimar a matriz O-D. Esses sistemas registram a hora, a data, o local de embarque e o local de desembarque dos usuários. Com base nesses dados, é possível obter informações sobre as origens e destinos das viagens, que podem ser utilizadas para construir a matriz O-D (GUERRA, 2011).
- **Sensoriamento Remoto:** Este método utiliza imagens de satélite ou aéreas para estimar a matriz O-D. Com base nas informações sobre a distribuição de pessoas e atividades em uma determinada área, é possível estimar a demanda de transporte público entre diferentes regiões. Este método é particularmente útil em áreas de difícil acesso ou onde a coleta de dados por outros meios é inviável (PROSPERI et al., 2021).
- **Modelos de Equilíbrio de Transporte:** Estes modelos são usados para estimar a matriz O-D, utilizando informações sobre as características do sistema de transporte, a rede viária, a distribuição da população e as atividades econômicas. Os modelos de equilíbrio de transporte são geralmente usados para planejar e avaliar políticas públicas de transporte, e podem ser usados para estimar a demanda de transporte público em diferentes regiões (YANG; IIDA; SASAKI, 1994).
- **Modelo de Gravidade:** Este método utiliza a teoria da gravidade para estimar a matriz O-D. O modelo de gravidade pressupõe que a demanda de transporte público é inversamente proporcional à distância entre as origens e destinos, e proporcional à população e à atividade econômica das regiões em questão. Este método é útil para estimar a demanda de transporte público entre regiões próximas (EKOWICAKSONO; BUKHARI; AMAN, 2016).

Portanto, a matriz O-D desempenha um papel crucial no planejamento do sistema de transporte, fornecendo uma análise detalhada dos padrões de deslocamento dentro de uma área geográfica. Essa análise permite uma compreensão profunda das demandas dos usuários do sistema de transporte público, sendo fundamental para o planejamento eficaz e a otimização dos serviços de transporte. A utilização da matriz O-D como uma das fontes de dados na construção do *framework* proposto, que busca ser de baixo custo e capaz de gerar e monitorar indicadores, representa um passo importante na direção de construir um sistema de transporte mais eficiente.

3.3 Inteligência Artificial (IA)

O termo Inteligência Artificial (IA) foi proposto por John McCarthy, em 1956, durante a conferência de *Dartmouth* (ANDRESEN, 2002). Esse é considerado o marco para o surgimento da IA. John McCarthy utilizou esse termo para descrever a habilidade das máquinas que desempenham tarefas que necessitam de uma destreza inerente aos seres humanos, ou seja, a capacidade das máquinas agirem utilizando algoritmos e sistemas capazes de executar tarefas que, normalmente, requerem inteligência humana. Desse modo, a IA pode ser uma aliada fundamental no gerenciamento e planejamento do sistema de transporte, oferecendo soluções dinâmicas e adaptativas para enfrentar desafios complexos. Utilizando a análise de grandes volumes de dados, tais como padrões de tráfego, horários de pico, condições climáticas e preferências dos usuários, a IA é capaz de otimizar rotas, prever demandas e ajustar horários de maneira eficiente. Essa abordagem, orientada por dados, facilita uma tomada de decisão mais informada e ágil, o que contribui para a criação de um sistema de transporte mais responsivo, econômico e focado no usuário. Assim, alinha-se com as necessidades em constante evolução das cidades modernas, especialmente no contexto de países em desenvolvimento.

Na literatura, existem várias diferentes definições sobre o que é a IA. Vários autores como Russell e Norvig (2022) e Campesato (2020) abordam a definição de inteligência artificial como uma análise em separado das palavras inteligência e artificial. Russell e Norvig (2022) afirmam que a inteligência pode ser vista como a combinação de 4 (quatro) diferentes elementos: **agir como um ser humano, pensar como um ser humano, pensamento racional e agir de maneira racional.**

Segundo Russell e Norvig (2022), **agir como um ser humano** significa que a resolução de tarefas deve ser realizada por meio de um comportamento similar ao comportamento dos seres humanos. Tarefas como a comunicação por meio da linguagem humana, a habilidade de responder questões complexas e gerar novas conclusões, a capacidade de se adaptar a novos cenários, a aptidão em aprender padrões e extrapolar esse aprendizado, além de interagir com o meio ao redor (visão computacional e manipulação de objetos), devem ser realizadas pelas máquinas de tal modo que os resultados obtidos sejam similares aos alcançados pelos seres humanos. Um exemplo de teste para avaliar o comportamento de uma máquina é conhecido como o Teste de Turing (FRENCH, 2000).

Pensar como os seres humanos significa que as máquinas devem ser capazes de construir os seus próprios pensamentos utilizando um dos 3 (três) tipos de pensamento humano: utilizando os próprios atos (introspecção), observando os atos realizados por outras pessoas ou utilizando a imaginação. Assim, o comportamento mental de uma máquina será similar ao comportamento de um ser humano. Agir como um ser humano significa que a máquina tem a capacidade de resolver uma tarefa como um ser humano e o pensar como um ser

humano possibilita à máquina pensar e modelar a solução de um problema como o ser humano.

O **pensamento racional** requer que as máquinas possuam a capacidade de realizar conclusões lógicas. Por exemplo, uma máquina deve ser capaz de inferir a sentença “Todos os homens são mortais. Sócrates é homem. Logo, Sócrates é mortal”. Em um cenário de incertezas, as máquinas devem escolher conclusões com a maior probabilidade de acontecer.

Agir de maneira racional refere-se à capacidade da máquina de gerar ações que visem alcançar os melhores resultados possíveis dentro do contexto específico em que está inserida. Por exemplo, em um jogo eletrônico de xadrez, o agente (máquina) desenvolvido deve tomar decisões estratégicas com o objetivo de vencer o jogo. Essa abordagem racional garante que as decisões do agente sejam direcionadas e alinhadas com o resultado esperado, demonstrando uma atuação lógica e orientada a objetivos.

Analisando essas visões sobre IA é possível defini-la como a área da ciência responsável por criar sistemas capazes de desenvolver tarefas que normalmente exigem inteligência humana, como o aprendizado, o raciocínio, a percepção e a tomada de decisão.

3.3.1 Aprendizado de Máquina (AM)

O Aprendizado de Máquina (AM) é uma subárea da IA ([CAMPESATO, 2020](#)). Essa subárea é a responsável por fornecer algoritmos computacionais com a capacidade de melhorar o desempenho na realização de uma tarefa por meio da experiência. Os algoritmos de AM adquirem experiência à medida que absorvem o conhecimento da tarefa a ser desenvolvida. Esse conhecimento é adquirido durante o processo de treinamento. É durante esse processo que o algoritmo utiliza os dados fornecidos para aprender as características do problema e consegue gerar melhores resultados.

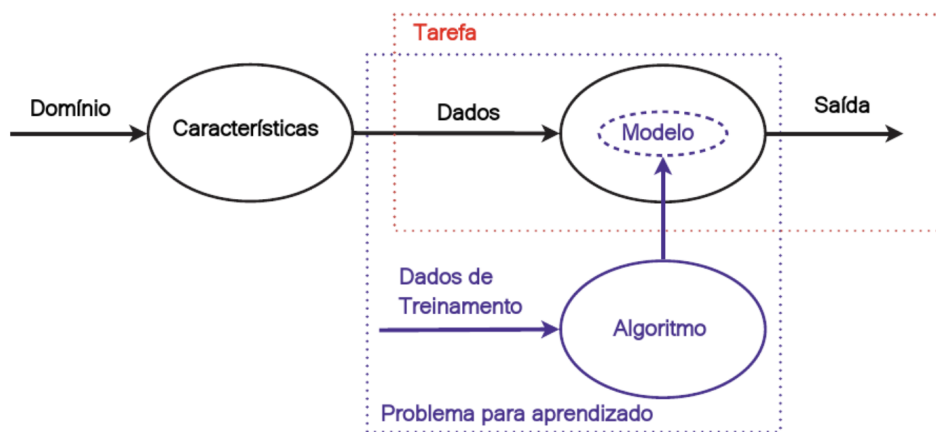
Tarefas como realizar a previsão de demanda semanal por ônibus em uma grande cidade ou prever qual é o objetivo das viagens dos passageiros, representam tarefas típicas desempenhadas por algoritmos de aprendizado de máquina. Nesses 2 (dois) exemplos tem-se que os algoritmos de AM utilizam dados com características específicas de cada uma dessas tarefas a fim de aprender a dinâmica do sistema e, posteriormente, serem capazes de generalizar esse aprendizado para situações que não estejam explícitas nos dados utilizados durante a etapa de aprendizado. A cada nova iteração realizada durante o treinamento, o algoritmo tende a melhorar os resultados obtidos.

[Flach \(2012\)](#) indica que a principal diferença entre os algoritmos de AM e os algoritmos tradicionais como os de ordenação por meio do *Merge Sort* ou do *Bubble Sort* está,

justamente, nessa capacidade dos algoritmos de AM melhorarem o desempenho por meio da experiência adquirida.

Nesse sentido, os algoritmos de AM podem ser vistos como os responsáveis por resolver diferentes tarefas necessárias em aplicações de IA. A [Figura 10](#) apresenta os componentes necessários para que uma tarefa seja executada utilizando AM.

Figura 10 – Componentes para a solução de uma tarefa utilizando AM



Fonte: Adaptado de [Flach \(2012\)](#)

Analisando a [Figura 10](#), é possível perceber que, para o funcionamento correto de um algoritmo, são necessários alguns componentes. Esses componentes são a **tarefa** a ser resolvida, o **modelo** e o **algoritmo** de aprendizagem.

No exemplo de realizar a previsão da demanda horária de passageiros em uma grande cidade, a **tarefa** pode ser vista como sendo a realização dessa previsão dentro de um horizonte de tempo e em uma área específica. Um exemplo pode ser a previsão horária da demanda de ônibus para a próxima semana no bairro Castelo na cidade de Belo Horizonte, no estado Minas Gerais. Desse modo, é necessário construir um **modelo** que seja capaz de, para cada hora durante os próximos 7 (dias), prever a variação horária da quantidade de passageiros em cada uma das linhas que possuem origem ou destino no bairro Castelo.

Nesse contexto, os **dados** a serem utilizados devem representar essa variação horária do número de passageiros em cada uma das linhas. Para esses dados, podem ser empregadas as séries temporais históricas de utilização das linhas. Dados como a bilhetagem eletrônica ou de pesquisas de origem-destino podem ser utilizadas para construir a base de **treinamento** dos algoritmos. O algoritmo deve ser capaz de receber todas essas variações de demanda e conseguir encontrar a dinâmica da movimentação desses passageiros residentes no bairro Castelo. Desse modo, a tarefa de realizar a previsão de demanda é alcançada por meio de um **modelo**, que é obtido mediante o treinamento de um **algoritmo**

que “aprendeu” a dinâmica do movimento da população que utiliza o transporte público. É por meio do **treinamento** que o algoritmo aprende a encontrar o perfil de movimentação da população e, após esse treinamento, é que o algoritmo será capaz de gerar o **modelo** responsável por realizar a previsão do fluxo de passageiros do transporte público na região do bairro Castelo.

Essa capacidade de resolver tarefas complexas utilizando dados de diferentes fontes e, por meio deles, conseguir gerar previsões corretas e auxiliar na identificação de oportunidades é fundamental para o desenvolvimento da IA.

Devido a essas características, o AM pode desempenhar um papel vital no planejamento e gerenciamento do sistema de transporte, oferecendo *insights* e soluções personalizadas. Ao analisar e interpretar grandes volumes de dados, como padrões de tráfego, horários de pico, condições climáticas e comportamento dos usuários, os algoritmos de aprendizado de máquina podem identificar tendências e prever necessidades futuras. Isso permite a otimização de rotas, a previsão de demandas, e o ajuste de horários de forma proativa e precisa. Além disso, o aprendizado contínuo desses sistemas possibilita a adaptação às mudanças nas condições e necessidades urbanas, contribuindo para um sistema de transporte mais eficiente, econômico e centrado no usuário. Essa abordagem baseada em dados e IA transforma o planejamento e gerenciamento do transporte em um processo mais ágil e responsivo, alinhado com as complexidades e dinâmicas das cidades modernas.

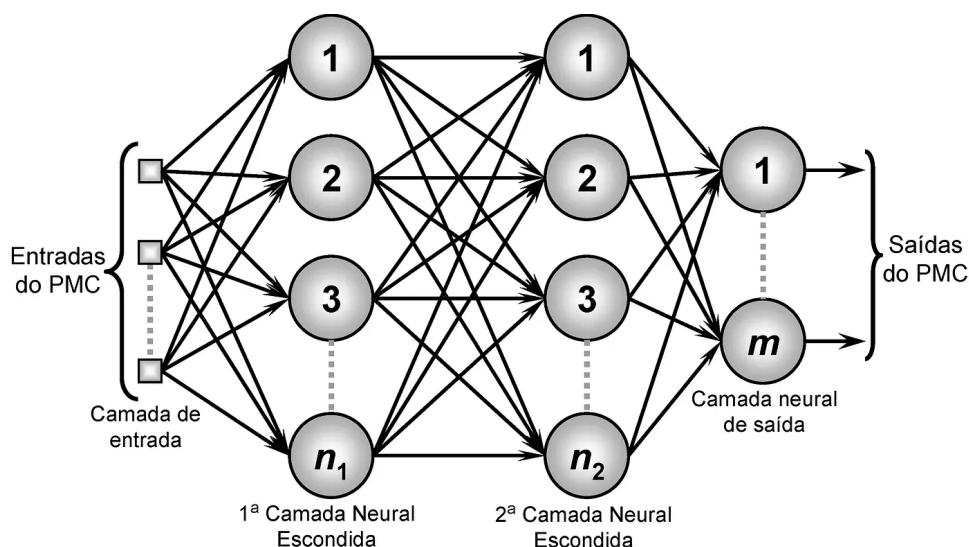
3.3.2 Redes Neurais Artificiais (RNA)

O desenvolvimento de algoritmos computacionais que possuem como inspiração a observação de fenômenos da natureza é uma estratégia bastante empregada tanto no meio acadêmico quanto na resolução de problemas reais ([ARANA-DANIEL; LOPEZ-FRANCO; ALANIS, 2018](#)). As Redes Neuras Artificiais (RNA) são um exemplo desses algoritmos. Portanto, as RNA podem ser vistas como sendo um conjunto de algoritmos de AM que possuem a característica de serem inspirados no funcionamento do sistema nervoso central humano.

As RNA correspondem a um conjunto de unidades computacionais com características similares ao neurônio humano. Essas unidades computacionais, quando dispostas em camadas conectadas, possibilitam a construção de estruturas computacionais complexas que constituem as bases de vários algoritmos de AM. A [Figura 11](#) apresenta um exemplo de uma RNA conhecida como Perceptron Multicamadas (PMC - do inglês *Multilayer Perceptron*). O perceptron pode ser visto como a RNA mais simples existente.

Quando os dados são apresentados à camada de entrada de uma PMC, como mostrado na [Figura 11](#), esses dados são multiplicados por valores predefinidos, conhecidos como

Figura 11 – Representação de um Perceptron Multicamadas (PMC)



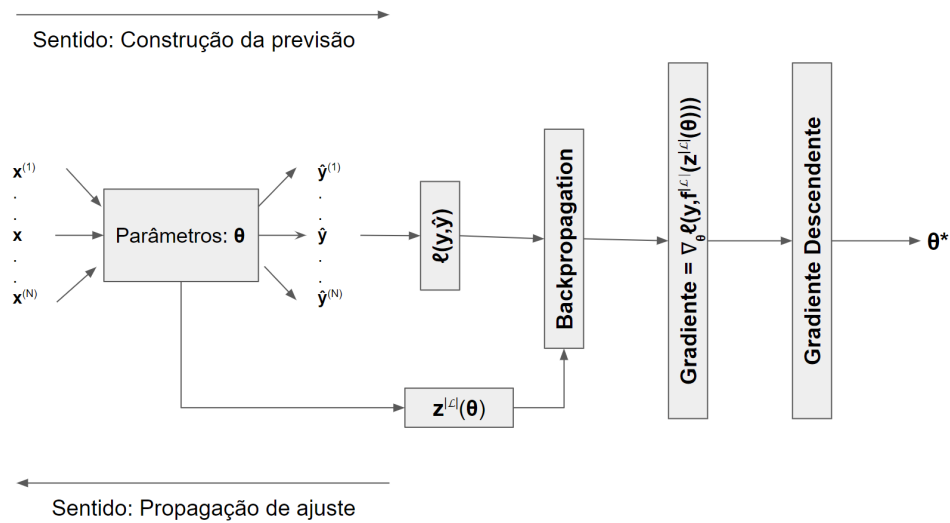
Fonte: Adaptado de [Moreira \(2018\)](#)

peso. O resultado dessa operação é transmitido para os neurônios presentes nas próximas camadas. Como pode ser visto, as setas indicam como essa operação é propagada dentro de uma RNA do tipo PMC. Além dos pesos, há também o *bias*, que é uma constante adicionada à saída do neurônio para deslocar a função de ativação, permitindo uma melhor adaptação do modelo aos dados. Desse modo, cada neurônio realiza a soma dos resultados obtidos pelas camadas anteriores, ajusta os pesos e *bias* utilizando as funções de ativação e propaga os novos resultados para os elementos subsequentes. É esse processo iterativo de ajuste dos pesos e *bias* que garante o aprendizado da RNA. Essas iterações ajustam esses valores no sentido de minimizar os erros de previsão. Esse processo de aprendizado ocorre durante a chamada “*learning phase*” (fase de aprendizado).

Uma das estratégias mais utilizadas para realizar essa fase de aprendizado é conhecida como *backpropagation*. A [Figura 12](#) mostra em mais detalhes os elementos que compõem o processo de treinamento de uma RNA. O processo de treinamento consiste em empregar o *backpropagation* para que a equação matemática contendo os pesos gere um resultado de previsão utilizando os dados e os valores propagados pelos neurônios (sentido de previsão). Essa é a primeira etapa do processo iterativo para o treinamento de uma RNA.

Na segunda etapa (sentido de ajuste,) o resultado desejado (y) é comparado com resultado da previsão ($\hat{y}$). Essa diferença gera um erro de previsão. Esse erro de previsão é gerado por meio da função perda (do inglês *loss function*). Os ajustes dos pesos e *bias* presentes nessa equação ocorrem no sentido de encontrar os novos pesos que minimizam esse erro de previsão. Normalmente, o processo de otimização empregado para encontrar esses parâmetros que minimizam a função perda durante as iterações de treinamento é o gradiente descendente. [Damadi, Moharrer e Cham \(2023\)](#) apresentam um exemplo de implementação

Figura 12 – Representação do processo de treinamento de uma RNA



Fonte: Adaptado de Damadi, Moharrer e Cham (2023)

do *backpropagation* (algoritmo 1). O cálculo do gradiente descendente está representado no algoritmo 1. Esse cálculo indica o sentido de decaimento da função perda e, portanto, mostra quais devem ser os próximos pesos e bias que precisam ser utilizados a fim de minimizar os erros de previsão do algoritmo. Desse modo, em cada iteração no processo o algoritmo passa a aprender melhor as características dos dados utilizados para treinamento, ou seja, melhora o desempenho das previsões realizadas.

Algorithm 1 Algoritmo *backpropagation*

Data: Dado uma RNA com L -Camadas, uma função perda l e um conjunto de dados (x,y) .

Seja $a^{[0]} := x$

Calcule $\nabla_{\mathbf{z}^{[L]}} l(\mathbf{y}, f(\mathbf{z}^{[L]}))$

for $l=1, \dots, L$ **do**

if $l \neq L$ **then**

$$\mathbf{J}_{\mathbf{W}^{[l]}, \mathbf{b}^{[l]}}(\mathbf{z}^{[l]}) = (\mathbf{W}^{[L]})^T \mathbf{J}_{\mathbf{W}^{[L-1]}, \mathbf{b}^{[L-1]}}(f^{[L-1]}(\mathbf{z}^{[L-1]})) \dots \mathbf{J}_{\mathbf{W}^{[l]}, \mathbf{b}^{[l]}}(f^{[l]}(\mathbf{z}^{[l]})) \begin{bmatrix} (\mathbf{a}^{[l]})^T & 0 & 0 & \\ 0 & \ddots & 0 & I \\ 0 & 0 & (\mathbf{a}^{[l]})^T & \end{bmatrix}$$

if $l = L$ **then**

$$\mathbf{J}_{\mathbf{W}^{[L]}, \mathbf{b}^{[L]}}(\mathbf{z}^{[L]}) = \begin{bmatrix} (\mathbf{a}^{[L]})^T & 0 & 0 & \\ 0 & \ddots & 0 & I \\ 0 & 0 & (\mathbf{a}^{[L]})^T & \end{bmatrix}$$

end

 Construa:

$$\mathbf{J}_{\theta \mathbf{z}^{[L]}}(\theta) = [\mathbf{J}_{\mathbf{W}^{[L]}, \mathbf{b}^{[L]}}(\mathbf{z}^{[L]}(\theta)) \dots \mathbf{J}_{\mathbf{W}^{[1]}, \mathbf{b}^{[1]}}(\mathbf{z}^{[1]}(\theta))]$$

 Calcule o gradiente:

$$\nabla_{\theta} l(\mathbf{y}, \hat{\mathbf{y}}(\theta)) = (\mathbf{J}_{\theta \mathbf{z}^{[L]}}(\theta))^T \nabla_{\mathbf{z}^{[L]}} l(\mathbf{y}, f(\mathbf{z}^{[L]}))$$

end

end

onde:

- L é o número total de camadas na RNA.
- l é a função de perda utilizada.
- (x, y) são os dados de entrada e saída, respectivamente.
- $a^{[0]}$ é a ativação da camada de entrada, igual a x .
- $\nabla_{\mathbf{z}^{[L]}} l(\mathbf{y}, f(\mathbf{z}^{[L]}))$ é o gradiente da função de perda em relação à saída da última camada.
- $\mathbf{W}^{[l]}$ e $\mathbf{b}^{[l]}$ são os pesos e bias da camada l , respectivamente.
- $\mathbf{J}_{\mathbf{W}^{[l]}, \mathbf{b}^{[l]}}$ é a matriz Jacobiana para a camada l .
- $\mathbf{a}^{[l]}$ é a ativação da camada l .
- $\mathbf{z}^{[l]}$ é a entrada ponderada da camada l antes da função de ativação.
- $f^{[l]}$ é a função de ativação da camada l .
- $\mathbf{J}_{\theta \mathbf{z}^{[L]}}(\theta)$ é a matriz Jacobiana total para a última camada.
- $\nabla_{\theta} l(\mathbf{y}, \hat{\mathbf{y}}(\theta))$ é o gradiente da função de perda em relação a todos os parâmetros.
- $\hat{\mathbf{y}}(\theta)$ é a saída prevista da RNA.
- θ representa todos os parâmetros da RNA.

Esse processo de treinamento ocorre em uma variedade de RNA. Cada uma dessas RNA possui características específicas que as tornam adequadas para a resolução dos mais variados problemas e em diferentes áreas do conhecimento (TKÁČ; VERNER, 2016).

Áreas como negócio (TKÁČ; VERNER, 2016), engenharia (ALANIS; ARANA-DANIEL; LOPEZ-FRANCO, 2019), *marketing* (FERASSO; ALNOOR, 2022), visão computacional (ANSARI, 2020), cuidado com a saúde (do inglês *healthcare*) (SHAHID; RAPPON; BERTA, 2019), educação (KASNECI et al., 2023) e transporte (SEGUNDO; SILVA; FARINES, 2016) utilizam as RNA como auxílio à tomada de decisão, previsão e análise de cenários, identificação de gargalos produtivos e oportunidades de negócio, redução de custos e automação de processos. A Figura 13 apresenta alguns tipos populares de RNA e algumas das principais áreas de aplicação.

Portanto, no contexto do transporte público de passageiros, as RNA podem ser treinadas para prever demandas de tráfego, identificar horários de pico, otimizar rotas e até mesmo detectar anomalias ou falhas em tempo real. Por exemplo, ao analisar dados históricos de fluxo de passageiros, uma RNA pode prever com precisão quando e onde ocorrerão congestionamentos, permitindo que os gestores de transporte reajam proativamente, redistribuindo recursos ou ajustando horários. Além disso, as RNA se adaptam continuamente à medida que recebem novos dados, tornando os sistemas de transporte mais resilientes e adaptativos às mudanças urbanas e às necessidades dos usuários. Assim, a implementação de RNA no setor de transporte podem contribuir com o planejamento e gerenciamento da mobilidade, tornando-a mais eficiente, segura e centrada no usuário.

3.3.3 Redes Neurais Recorrentes (RNR)

Redes Neurais Recorrentes (RNR) pertencem a uma família de algoritmos de AM que realiza o processamento de dados sequenciais (GOODFELLOW; BENGIO; COURVILLE, 2016). Essas redes são responsáveis por receber um conjunto de valores sequenciais $(x(1), x(2), \dots, x(n))$ e realizar o processamento desses valores propagando características

Figura 13 – Exemplos de RNA e principais aplicações

01	Rede Neural Recorrente (RNR)	• Análise de Séries Temporais; • Tradução de Textos; • Reconhecimento de Discurso.
02	Rede Neural Convolutacional (RNC)	• Visão Computacional ; • Análise de Séries Temporais; • Sistemas de Recomendação.
03	Self Organizing Map (SOM)	• Clusterização e Segmentação de Clientes; • Classificação de Imagens; • Processamento de Sinais .

Fonte: O autor (2023)

de valores passados para o processamento de valores atuais. Essas características são propagadas por meio da preservação dos estados dessa rede neural.

Devido a essa característica, as RNR podem lembrar informações anteriores na sequência, o que é essencial para entender padrões temporais, como flutuações no tráfego durante diferentes horas do dia ou dias da semana. Essa capacidade de capturar dependências temporais permite que as RNR façam previsões mais precisas sobre a demanda futura de transporte, otimizem rotas com base em padrões históricos e ajustem a alocação de recursos em tempo real. Por exemplo, uma RNR pode ser treinada para prever o volume de passageiros em uma linha de ônibus durante a hora de pico, permitindo que os operadores ajustem a frequência dos ônibus de acordo com a necessidade prevista. Além disso, as RNR podem ser usadas para detectar anomalias no sistema, como atrasos inesperados, e responder rapidamente a essas situações. Assim, as RNR oferecem uma abordagem dinâmica e adaptativa para o gerenciamento do transporte, contribuindo para um sistema de transporte mais eficiente e responsivo.

Um sistema dinâmico recorrente pode ser representado pela [Equação 1](#).

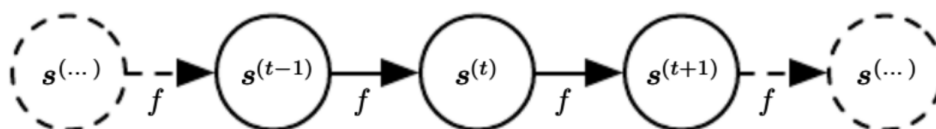
$$s^t = f(s^{(t-1)}; \theta) \quad (1)$$

Na [Equação 1](#) tem-se que s^t é conhecido como o estado do sistema. A recorrência desse sistema pode ser visualizada pelo fato do valor (t-1) ser utilizado para encontrar o próximo valor temporal t. θ representa um conjunto de parâmetros do modelo ou sistema. Esses parâmetros são elementos cruciais que definem o comportamento do sistema ou modelo descrito pela função f . Desse modo, a [Equação 2](#) representa o estado s para o tempo ($t=2$).

$$s^2 = f(f(s^{(1)}; \theta); \theta) \quad (2)$$

A [Equação 2](#) pode ser expressa como um gráfico acíclico dirigido (do inglês *Directed Acyclic Graph* - DAG). A representação da DAG para a [Equação 2](#) pode ser vista na [Figura 14](#).

Figura 14 – Representação da [Equação 2](#) como um DAG

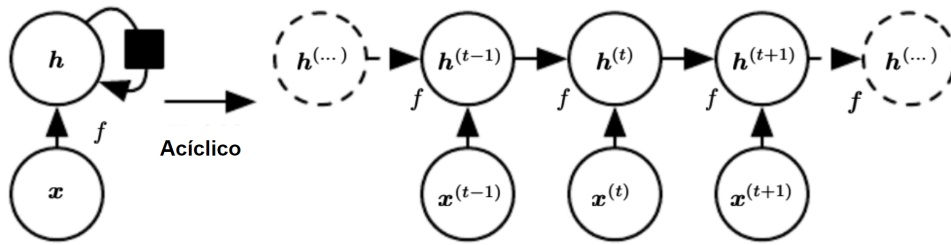


Fonte: [Goodfellow, Bengio e Courville \(2016\)](#)

Quando existe alguma outra característica externa que influencia o sistema expresso pela [Equação 2](#), essa equação pode ser escrita como mostrado pela [Equação 3](#). Na [Equação 3](#) tem-se que o estado é expresso por $\mathbf{h}$ e a variável $\mathbf{x}$ representa uma entrada sequencial externa de valores $(x^{(t)}, x^{(t-1)}, \dots, x^{(2)}, x^{(1)})$. A [Figura 15](#) mostra a representação dessa estrutura. Nessa figura, cada nova informação ($\mathbf{x}$) é incorporada ao estado $\mathbf{h}$. Desse modo, durante o processo de treinamento, são selecionadas as principais características dos estados $\mathbf{h}$ que devem ser propagados para os estados futuros.

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; \theta) \quad (3)$$

Figura 15 – Representação gráfica da [Equação 3](#)



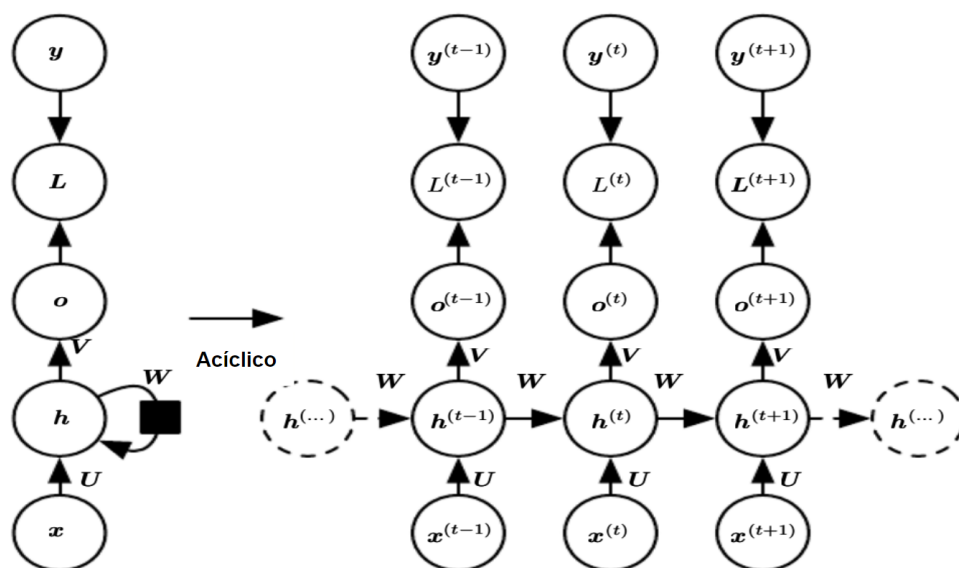
Fonte: [Goodfellow, Bengio e Courville \(2016\)](#)

A [Equação 4](#) mostra a representação de uma RNR. Nessa equação tem-se que $\mathbf{U}$, $\mathbf{W}$ e $\mathbf{V}$ representam, respectivamente, a matriz de pesos para a conexão das entradas com a camada escondida, a matriz de pesos para a conexão entre as camadas escondidas e a matriz de pesos para a conexão das camadas escondidas com as saídas. A [Figura 16](#) mostra a representação gráfica da [Equação 4](#).

$$\begin{aligned} \mathbf{a}^{(t)} &= \mathbf{b} + \mathbf{W}\mathbf{h}^{(t-1)} + \mathbf{U}\mathbf{x}^{(t)} \\ \mathbf{h}^{(t)} &= \tanh(\mathbf{a}^{(t)}) \\ \mathbf{o}^{(t)} &= \mathbf{c} + \mathbf{V}\mathbf{h}^{(t)} \\ \hat{\mathbf{y}}^{(t)} &= \text{softmax}(\mathbf{o}^{(t)}) \end{aligned} \quad (4)$$

Pela [Figura 16](#) e pela [Equação 4](#), é possível identificar que a função perda para o processo de treinamento de uma RNR que recebe uma sequência de valores de entrada $\mathbf{x}$ e realiza o mapeamento para uma sequência de valores de saída $\mathbf{o}$ é representada por L . Essa função é utilizada para medir o quanto a previsão dos valores $\mathbf{o}$ está distante dos valores alvo $\mathbf{y}$. Esses valores alvo devem ser fornecidos para que a RNR, durante o processo de treinamento, seja capaz de ajustar os pesos que conectam cada uma das camadas dessa rede. A função L computa $\hat{\mathbf{y}} = \text{softmax}(\mathbf{o})$ e compara esse valor com o alvo $\mathbf{y}$. O cálculo realizado pela função L pode ser realizado como apresentado na [Equação 5](#).

Figura 16 – Representação gráfica da Equação 4



Fonte: Goodfellow, Bengio e Courville (2016)

$$L(\{x^{(1)}, \dots, x^{(\tau)}\}, \{y^{(1)}, \dots, y^{(\tau)}\}) = - \sum \log \mathbf{p}_{model}(y^{(t)} | \{x^{(1)}, \dots, x^{(t)}\}) \quad (5)$$

Durante o processo de treinamento, assim como ocorre em outras RNA, a ideia é minimizar as diferenças entre os valores previstos pela rede ($\mathbf{o}$) e os valores alvo ($\mathbf{y}$). Essa minimização ocorre, normalmente, utilizando o gradiente descendente (GOODFELLOW; BENGIO; COURVILLE, 2016). Para as RNR, o cálculo do gradiente é efetuado para cada nó $\mathbf{N}$, conforme ilustrado na Figura 16. Esse cálculo resulta no gradiente $\Delta_{\mathbf{N}}L$, que indica o sentido de crescimento da função. Para reduzir a diferença entre os valores reais e os calculados, normalmente se escolhe o sentido inverso de atualização em relação ao encontrado pelo gradiente. Assim, o gradiente $\Delta_{\mathbf{o}(t)}L_i$ representa as previsões ($\mathbf{o}$) para cada momento ($\mathbf{t}$) no processo de treinamento.

$$(\Delta_{\mathbf{o}^{(t)}} L)_i = \frac{\partial L}{\partial o_i^{(t)}} = \frac{\partial L}{\partial L^{(t)}} \frac{\partial L^{(t)}}{\partial o_i^{(t)}} = y_i^{\hat{(t)}} - \mathbf{1}_{i, y^{(t)}} \quad (6)$$

Utilizando a [Equação 6](#) como um processo iterativo de treinamento para cada um dos estados $\mathbf{h}$ e utilizando as matrizes $\mathbf{U}$, $\mathbf{W}$ e $\mathbf{V}$ como representação dos pesos, encontramos a [Equação 7](#).

$$\Delta_{h(t)}L = \mathbf{W}^T(\Delta_{h(t+1)}L)diag(1 - (\mathbf{h}^{(t+1)})^2) + \mathbf{V}^T(\Delta_{\mathbf{o}(t)}L) \quad (7)$$

Nessa equação tem-se que $diag(1 - (\mathbf{h}^{(t+1)})^2)$ representa a matriz diagonal que contém os valores para $1 - (\mathbf{h}^{(t+1)})^2$. Após calcular o gradiente para cada um dos nós $\mathbf{N}$, é possível encontrar os valores de atualização dos pesos presentes em cada uma das matrizes $\mathbf{U}$, $\mathbf{W}$ e $\mathbf{V}$ e dos parâmetros (bias) $\mathbf{c}$ e $\mathbf{b}$. A Equação 8, a Equação 9, a Equação 10, a Equação 11 e a Equação 12 apresentam os modelos matemáticos de atualização de cada um dos elementos presentes na RNR.

$$\Delta_c L = \sum_t \left(\frac{\partial o^{(t)}}{\partial \mathbf{c}} \right)^T \Delta_{o^{(t)}} L = \sum_t \Delta_{o^{(t)}} L \quad (8)$$

$$\Delta_b L = \sum_t \left(\frac{\partial \mathbf{h}^{(t)}}{\partial \mathbf{b}^{(t)}} \right)^T \Delta_{h^{(t)}} L = \sum_t diag(1 - (\mathbf{h}^{(t)})^2) \Delta_{h^{(t)}} L \quad (9)$$

$$\Delta_v L = \sum_t \sum_i \left(\frac{\partial L}{\partial o_i^{(t)}} \right) \Delta_v o_i^{(t)} = \sum_t (\Delta_{o^{(t)}} L) \mathbf{h}^{(t)T} \quad (10)$$

$$\Delta_w L = \sum_t \sum_i \left(\frac{\partial L}{\partial h_i^{(t)}} \right) \Delta_w h_i^{(t)} = \sum_t diag(1 - (\mathbf{h}^{(t)})^2) (\Delta_{h^{(t)}} L) \mathbf{h}^{(t-1)T} \quad (11)$$

$$\Delta_u L = \sum_t \sum_i \left(\frac{\partial L}{\partial h_i^{(t)}} \right) \Delta_u h_i^{(t)} = \sum_t diag(1 - (\mathbf{h}^{(t)})^2) (\Delta_{h^{(t)}} L) \mathbf{x}^{(t)T} \quad (12)$$

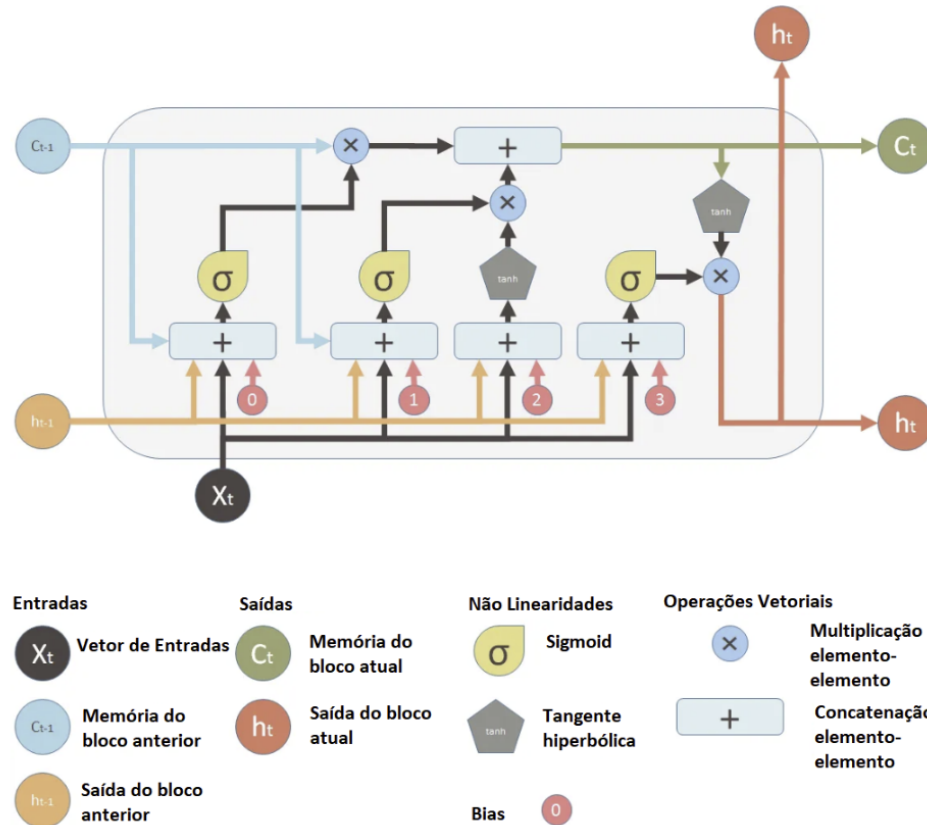
Quando existe uma dependência de longo prazo entre as variáveis de entrada e essas dependências são propagadas por meio dos estados $\mathbf{h}$, o gradiente tende a “desaparecer”. Esse efeito é conhecido como *Vanishing-gradient problem*. Ele ocorre, pois, pela Equação 12 e pela Figura 16, é possível perceber que a cada iteração de treinamento, existe o produto da matriz $\mathbf{W}$ pelos estados $\mathbf{h}^{(t)}$, para o cálculo do gradiente. Desse modo, essa propagação pode gerar um aumento significativo do gradiente (*Exploding Gradient*) ou uma redução (tendendo a zero) *Vanishing Gradient*.

Hochreiter e Schmidhuber (1997) propuseram uma solução para o *Vanishing Gradient Problem*. Essa solução ficou conhecida como *Long Short-Term Memory* (LSTM). Para o caso em que a matriz $\mathbf{W} > 1$, gera-se o problema do *Exploding Gradient*, quando $\mathbf{W} < 1$, identifica-se o problema do *Vanishing Gradient*. De uma maneira simplificada, a grande ideia para a solução desses problemas é manter $\mathbf{W} = 1$. Desse modo, é possível manter o ajuste adequado dos parâmetros do modelo durante o processo de treinamento.

A ideia implementada por Hochreiter e Schmidhuber (1997) consiste em adicionar células LSTM à estrutura da RNR. Essas células possuem um *loop* interno. Cada uma das células

possui as mesmas entradas e saídas de uma RNR convencional, mas são adicionados um conjunto de parâmetros que permite às células determinar quais devem ser as informações transferidas dentro da rede. A Figura 17 apresenta a estrutura dessas células com o *self-loop* associado. As operações vetoriais e as não linearidades controlam as informações que devem ser repassadas para as próximas células.

Figura 17 – Estrutura para as células LSTM



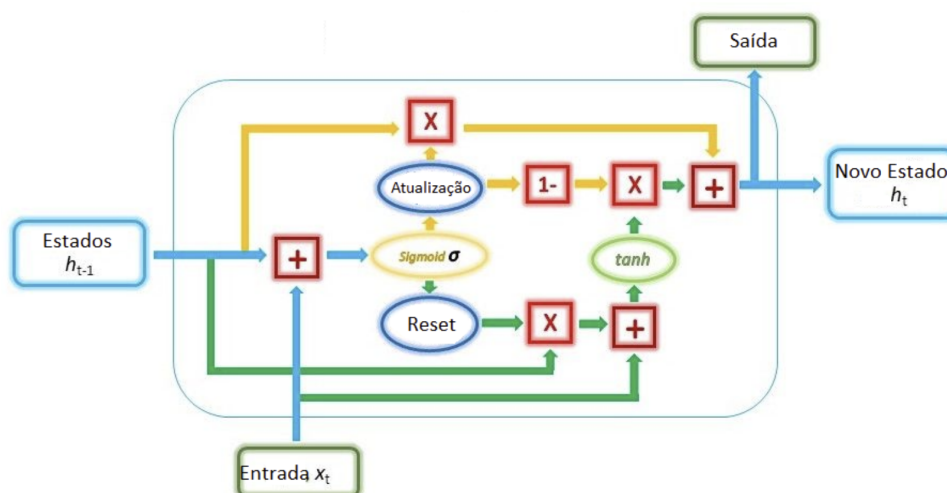
Fonte: Adaptado de Yan (2016)

Os estados $h_{(t-1)}$ e entradas $x_{(t-1)}$ são obtidas das células anteriores. Além disso, tem-se os valores $c_{(t-1)}$ que representam a memória da célula. Os valores $h_{(t-1)}$ e $x_{(t-1)}$ são combinados e passam pela função de ativação *sigmoid*. Essas funções (*sigmoid*) funcionam como válvulas e regulam a passagem dos sinais a fim de decidir quais características devem ser transmitidas para o próximo passo. Como pode ser visto, esses valores combinados são também enviados para a função *tangente hiperbólica*. É essa função que decide quais valores devem ser transmitidos para a memória dessa célula. Por exemplo, se a válvula da primeira *sigmoid* estiver aberta e a válvula da *tangente hiperbólica* estiver fechada, tem-se que a memória não será alterada e, assim, será transmitida integralmente para a próxima iteração. Desse modo, utilizando as células LSTM é possível regular como deve ocorrer a transferência de valores entre as células e ajustar os valores do gradiente sem que ocorra o problema do *Vanishing Gradient*.

3.3.4 Gated Recurrent Unit (GRU)

A GRU representa uma estrutura bastante semelhante à encontrada nas LSTM. A GRU foi apresentada em 2014 por [Cho et al. \(2014\)](#). A principal diferença entre essas duas arquiteturas está no fato de a GRU reduzir o número de estruturas necessárias para *resetar* a memória transmitida ou realizar a atualização dessa memória. A [Figura 18](#) apresenta a estrutura interna de uma célula GRU.

Figura 18 – Estrutura para as células GRU



Fonte: Adaptado de [Academy \(2022\)](#)

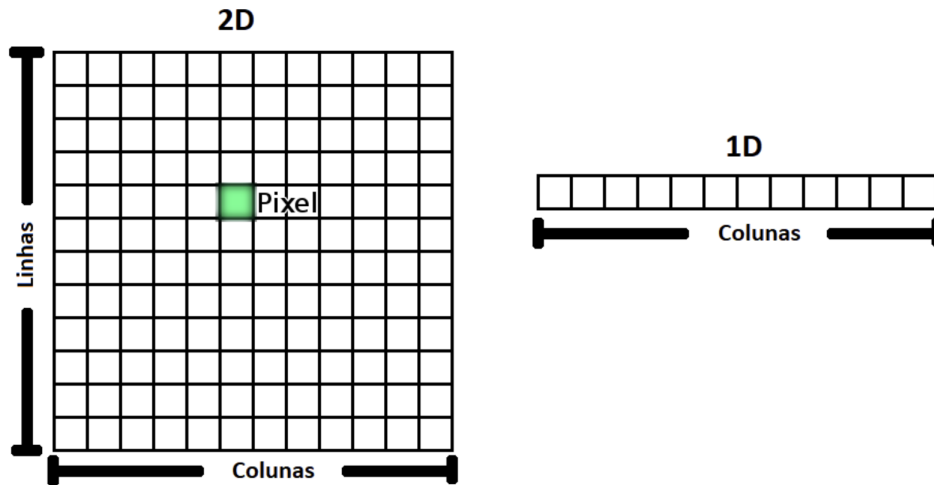
Como pode ser visto pela [Figura 18](#), existe uma estrutura responsável por decidir quais das características devem ser transmitidas (*reset* ou *atualização*) para a memória da célula. Esse estado de memória é utilizado para manter as características mais relevantes e excluir aquelas menos importantes em um longo prazo. Essas características obtidas durante o processo de treinamento da rede são responsáveis pelo aprendizado da rede.

3.3.5 Rede Neural Convolucional

As RNC são redes neurais artificiais especialmente desenvolvidas para tratar problemas que possuem uma estrutura de dados com uma topologia matricial ([GOODFELLOW; BENGIO; COURVILLE, 2016](#)). Uma representação dessa topologia pode ser encontrada na [Figura 19](#).

A representação em 2 (duas) dimensões pode ser encontrada no processamento de imagens, já a representação em 1 (uma) dimensão é observada em séries temporais em que os dados são coletados em períodos regulares. Além de sua aplicação mais comum em imagens, essas redes podem ser adaptadas para entender padrões temporais em dados sequenciais, como variações diárias e semanais no tráfego, congestionamentos recorrentes em certas rotas e flutuações na demanda durante eventos especiais ou mudanças sazonais. Por exemplo, a análise de séries temporais através de CNN pode ajudar a prever o aumento

Figura 19 – Representação de uma topologia matricial



Fonte: O autor (2023)

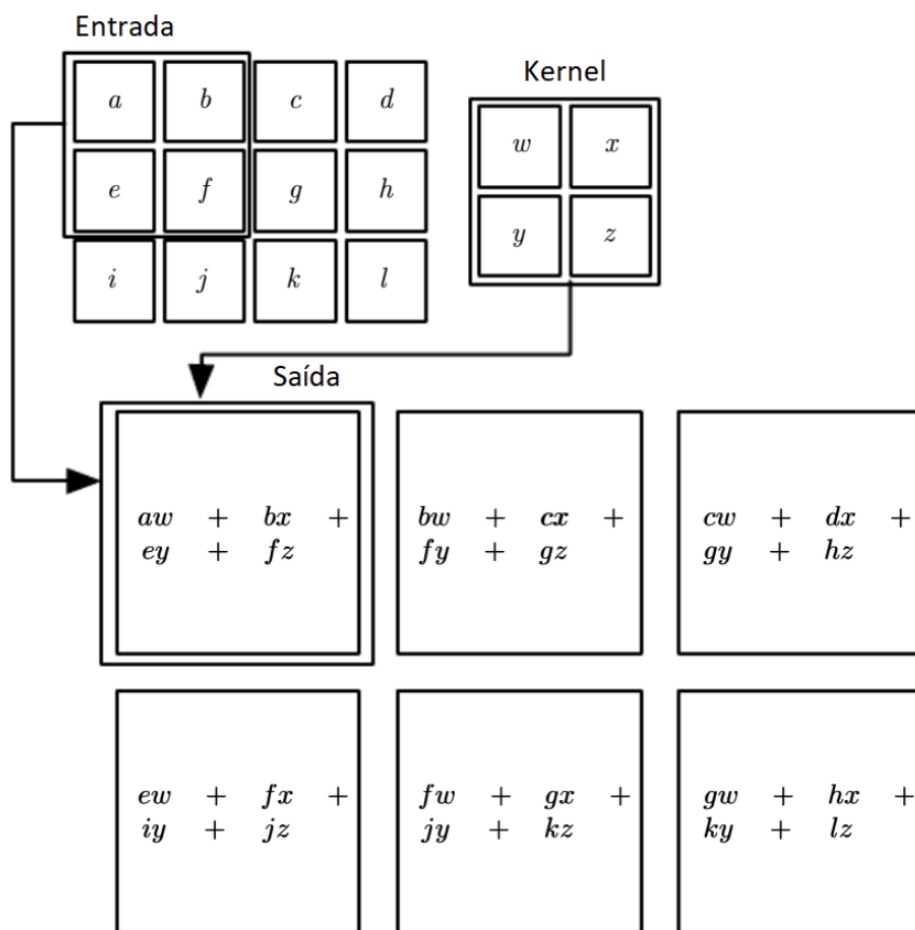
da demanda de transporte público durante grandes eventos esportivos ou festivais na cidade, permitindo um planejamento antecipado de rotas adicionais e horários ajustados. Além disso, essa abordagem pode revelar tendências de longo prazo, como o crescimento gradual da demanda em certas áreas devido ao desenvolvimento urbano, contribuindo para decisões estratégicas de longo prazo. Utilizando CNN para analisar essas e outras séries temporais, os gestores de transporte podem criar um sistema mais responsivo e eficiente, capaz de se adaptar às necessidades em constante mudança da população urbana.

As CNN possuem uma característica marcante, e que dá origem ao nome, que é a camada de convolução. Essa camada implementa a operação matemática linear conhecida como convolução. A operação convolucional entre duas funções, x e w , é expressa pela [Equação 13](#). Para essa equação tem-se que a função x é conhecida como entrada e a função w é dita como sendo o *kernel*. A saída s é nomeada como função de mapeamento. O símbolo $*$ denota a operação de convolução entre x e w , uma operação matemática que produz uma terceira função descrevendo a influência mútua dessas duas funções.

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t - a) \quad (13)$$

A [Figura 20](#) apresenta a representação gráfica dessa operação convolucional sobre um conjunto de dados com característica de 2 (duas) dimensões. Por meio dessa figura, é possível perceber que essa operação ocorre deslocando os elementos do *kernel* por toda a extensão dos valores de entrada. Essa operação gera os resultados também expressos nessa figura.

Figura 20 – Representação da operação de convolução

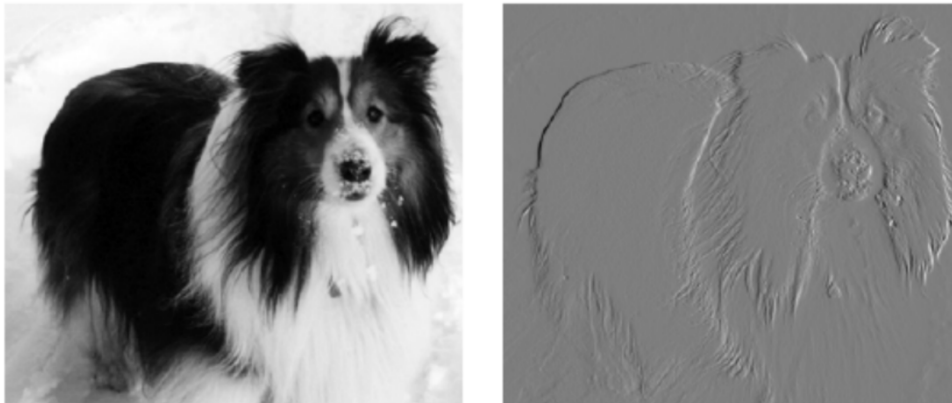


Fonte: Adaptado de [Goodfellow, Bengio e Courville \(2016\)](#)

A motivação principal para a utilização da operação de convolução no aprendizado de máquina está no fato dessa operação contribuir para o **tratamento de interações esparsas**, facilitar o **compartilhamento de parâmetros**, contribuir com **representações equivalentes** dos dados e possibilitar o tratamento de dados de entrada com diferentes tamanhos ([GOODFELLOW; BENGIO; COURVILLE, 2016](#)). As **interações esparsas** ocorrem, por exemplo, no processamento de imagens. Essa característica advém do fato das imagens, como valores de entrada, possuírem um conjunto grande de *pixels* e nem todos esses *pixels* possuem relevância para o aprendizado da rede. Desse modo, ao aplicar a operação de convolução é possível gerar um conjunto de *pixels* que seja relevante para o aprendizado da rede. O **compartilhamento de parâmetros** significa que o mesmo parâmetro será utilizado por mais de uma função do modelo. Em uma rede convolucional, como mostrado na [Figura 20](#), os valores do *kernel* são utilizados em operações sobre todo o conjunto de valores de entrada. Assim, existe a redução do número necessário de parâmetros, o que reduz a quantidade de memória necessária para o armazenamento dessa rede. A eficiência computacional gerada pelas **interações esparsas** e pelo **compartilhamento de parâmetros** pode ser visualizado

na [Figura 21](#). Nessa figura, é possível perceber o mapa de características gerado após o processo de convolução. É esse mapa que destaca certos aspectos da imagem original e preserva a relação espacial entre os *pixels*, o que é crucial para entender a forma e a estrutura de objetos na imagem. As **representações equivalentes** indicam que uma função pode ser equivalente a outra, ou seja, $f(g(x)) = g(f(x))$. Portanto, caso ocorra uma mudança nos dados de entrada, os valores de saída serão modificados da mesma forma.

Figura 21 – Exemplo de convolução



Fonte: Adaptado de [Goodfellow, Bengio e Courville \(2016\)](#)

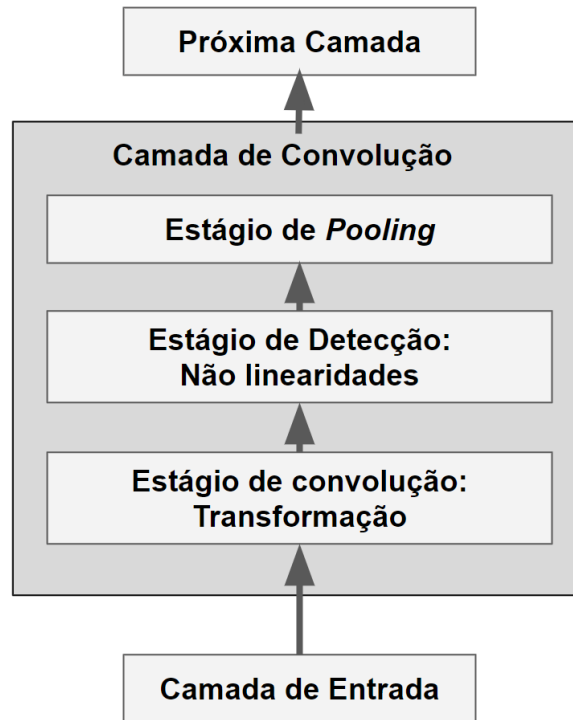
A representação da camada de convolução presente em uma arquitetura CNN pode ser vista na [Figura 22](#). A camada de *pooling* reduz a amostra da saída das camadas convolucionais. Essa redução ocorre por meio da modificação do filtro. Esse processo de redução pode ocorrer, por exemplo, utilizando o *max-pooling* em que é escolhido o maior valor obtido após a convolução ou escolhendo o valor médio *average-pooling*. A [Figura 23](#) apresenta uma possível representação desse processo de *pooling*. Assim, utilizando essa camada, é possível reduzir o tamanho do conjunto dos dados gerados após o processo de convolução.

A [Figura 24](#) apresenta as camadas de uma CNN. Nessa figura, além das camadas apresentadas anteriormente, temos a presença da camada de *Flatten* e a camada *Densa*. A camada *Flatten* é empregada para converter todas as matrizes bidimensionais resultantes do processo de *pooling* em um único vetor linear contínuo. Essa matriz linear resultante é utilizada como entrada para a camada *Densa*. Essa camada possui a característica de que todos os neurônios estão totalmente conectados à saída e será responsável por classificar a imagem, por exemplo.

3.3.6 Self Organizing Map (SOM)

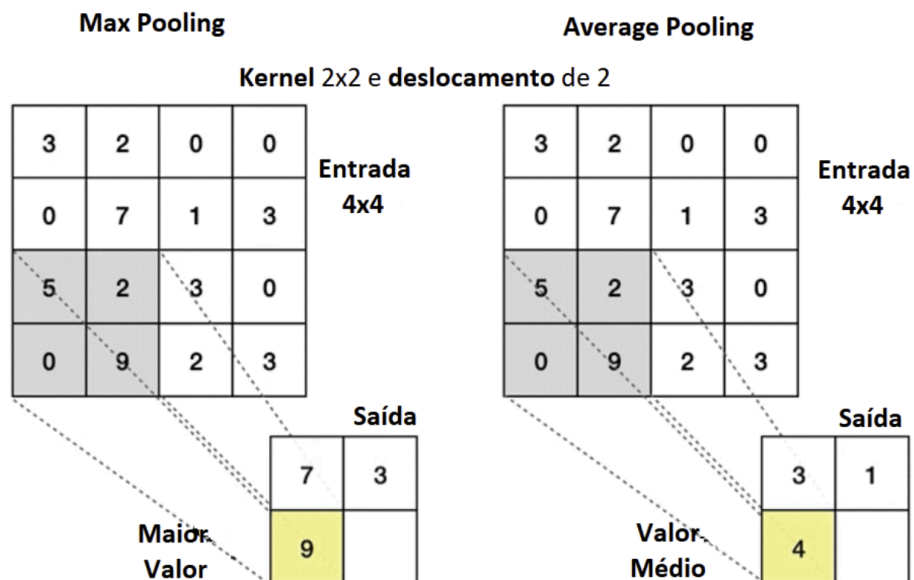
As redes SOM, introduzidas em 1988 por [Kohonen \(1990\)](#), representam uma rede neural artificial não supervisionada. A SOM é um tipo de RNA com apenas 2 (duas) camadas (entrada e saída). Essa rede apresenta características similares ao algoritmo *K-means*

Figura 22 – Camada de convolução



Fonte: Adaptado de [Goodfellow, Bengio e Courville \(2016\)](#)

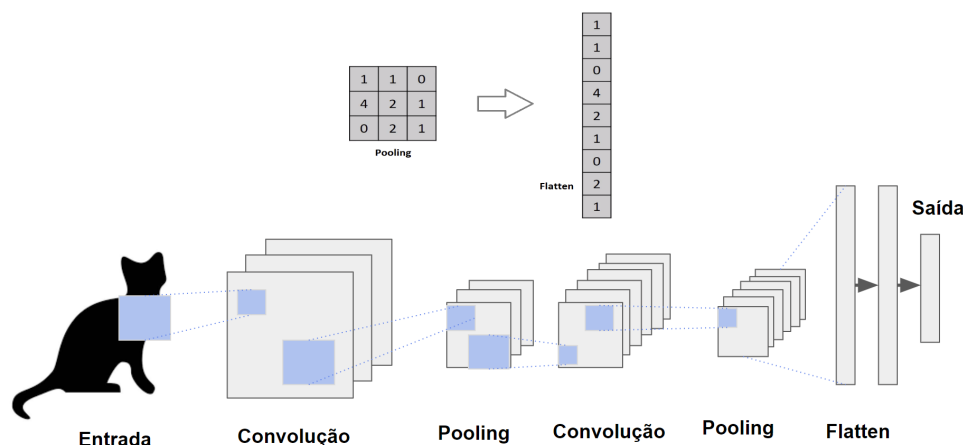
Figura 23 – Representação gráfica do processo de *pooling*



Fonte: Adaptado de [Goodfellow, Bengio e Courville \(2016\)](#)

([SELIM; ISMAIL, 1984](#)), uma vez que emprega a análise de similaridade entre um conjunto de dados. É por meio dessa análise de similaridade que os dados são agrupados. Esse agrupamento ocorre utilizando o processo conhecido como clusterização. Ao final do

Figura 24 – Representação gráfica de uma CNN



Fonte: O autor (2023)

processo de clusterização, é possível identificar conjuntos de dados em que as instâncias pertencentes a cada um desses *clusters* possuem características similares.

Diferente de algoritmos como o *K-means*, a SOM não necessita receber como hiperparâmetro o número de *cluster* a ser encontrado. A SOM considera, além dos dados presentes em um *cluster*, uma estrutura de vizinhança entre os agrupamentos. Dessa forma, os dados próximos pertencem ao mesmo *cluster* (como ocorre no *k-means*) ou a *clusters* vizinhos. Essa propriedade de identificação dos dados além dos *clusters* contribui para a visualização e análise de dados multidimensionais. Assim, a SOM tende a ter um desempenho melhor do que outros algoritmos de clusterização, pois os *clusters* podem ser exibidos de acordo com a estrutura da vizinhança e não apenas em relação aos dados do mesmo *cluster* (COTTRELL et al., 2018).

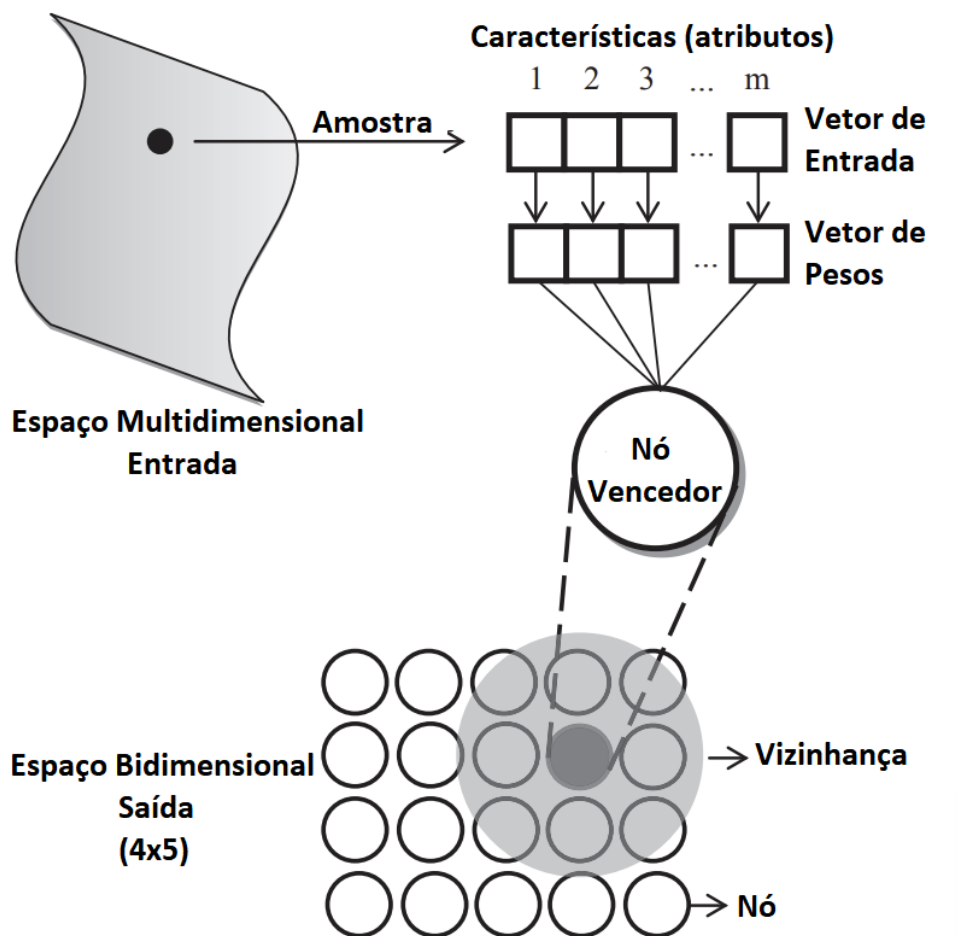
A ideia central da SOM é transformar um conjunto de dados com várias características (espaço de alta dimensão) em um espaço de baixa dimensão (normalmente duas dimensões). Essa transformação espacial é obtida preservando a topologia inicial do conjunto de dados. Como em outros algoritmos, como o *Principal Component Analysis* (PCA) (KURITA, 2019), a SOM pode ser empregada para a redução de dimensionalidade. Entretanto, a SOM possui algumas características construtivas que a faz ser bastante empregada em uma variedade de problemas. Isso se deve ao fato de não ser necessário assumir a independência das variáveis, ser capaz de resolver problemas complexos, e conseguir trabalhar em conjuntos de dados com um número limitado de instâncias e em conjuntos de dados de diferentes dimensionalidades (ASAN; ERCAN, 2012).

Utilizando a capacidade da SOM de mapear padrões complexos e multidimensionais em uma representação bidimensional, os gestores de transporte podem visualizar e entender as relações intrincadas entre diferentes variáveis, como tráfego, demanda de passageiros,

condições climáticas e horários de pico. Por exemplo, a SOM pode ser usada para identificar *clusters* de rotas com características semelhantes, permitindo a otimização de horários e a alocação de recursos de acordo com as necessidades específicas de cada *cluster*. Essa abordagem orientada por dados contribui para um sistema de transporte capaz de se adaptar às mudanças dinâmicas da paisagem urbana e às necessidades dos usuários, resultando em um serviço mais econômico e centrado no usuário.

A [Figura 25](#) apresenta uma representação visual do algoritmo implementado por uma SOM. Nessa figura, tem-se que uma amostra de um conjunto multidimensional de dados é utilizada como entrada desse algoritmo. O comprimento dessa amostra indica a quantidade de características existentes no conjunto de dados. Essa amostra é utilizada como vetor de entradas ($x = (x_1, x_2, \dots, x_m)$) para o algoritmo. A entrada é conectada à saída por meio do vetor de pesos. É a atualização desse vetor de pesos que possibilita o aprendizado da rede.

Figura 25 – Representação gráfica da SOM



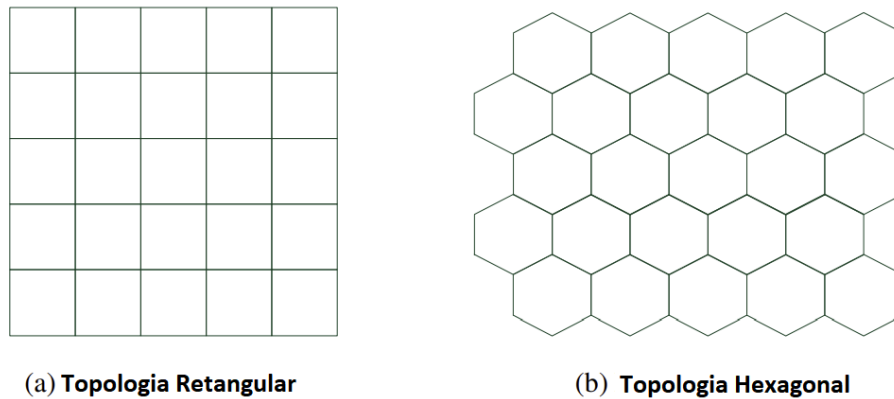
Fonte: O autor (2023)

Na [Figura 25](#), também é possível encontrar a camada de saída (*output layer* ou *Kohonen layer*). O número de neurônios presentes nessa camada indica a quantidade máxima de

clusters possível. É a topologia dessa camada de saída que indica como os nós (neurônios) se relacionam com os demais nós vizinhos.

Existem diferentes tipos de topologia. A [Figura 26](#) apresenta as topologias retangular e hexagonal. Na topologia retangular, cada um dos nós do centro está diretamente ligado a outros 4 (quatro) nós (vizinhança). Já na topologia hexagonal, cada nó do centro está conectado a outros 6 (seis) nós.

Figura 26 – Topologias para ligação dos nós em uma SOM



Fonte: O autor (2023)

Uma possível estrutura para a construção do algoritmo SOM é apresentado por [Asan e Ercan \(2012\)](#):

- i Defina a dimensão e o comprimento da saída. Escolha quais devem ser os parâmetros iniciais do algoritmo $w_i(0)$, $\alpha(0)$, σ que representam, respectivamente, o vetor inicial dos pesos, a taxa de aprendizado inicial e a função que descreve a variação da região de ativação dos neurônios ao redor do nó vencedor. O vetor inicial dos pesos $w_i(0)$ é uma estimativa inicial para os pesos da rede. A taxa de aprendizado inicial $\alpha(0)$ determina a magnitude do ajuste dos pesos a cada iteração. A função σ descreve como a região de ativação dos neurônios muda ao longo do tempo. Um exemplo dessa função é a *cutgauss* $h_{ci}(t) = \exp\left(-\frac{d_{ci}^2}{2\sigma^2(t)}\right) 1\left(\sigma(t) - d_{ci}\right)$, onde d_{ci} é a distância entre o neurônio vencedor e o neurônio i .
- ii Selecione uma amostra dos dados que será utilizada como vetor de entrada $x(t)$, onde t é o instante de tempo atual.
- iii Encontre a similaridade entre o vetor de pesos (w_i) e a amostra de entrada ($x(t)$). Essa similaridade pode ser calculada, por exemplo, utilizando a distância euclidiana $d_i(t) = |x(t) - w_i(t)| = \sqrt{\sum_j 1^m (x_{tj} - w_{tji})^2}$, onde m é o número de dimensões da entrada. Selecione o neurônio (nó) de maior similaridade com o vetor de entrada:

$c(t) = \operatorname{argmin}_i \left\{ |x(t) - w_i(t)| \right\}$, onde $c(t)$ é o índice do neurônio vencedor no instante t .

- iv Realize o ajuste dos pesos do nó vencedor e da vizinhança definida pela topologia da rede utilizando a equação $w_i(t+1) = w_i + \alpha(t) [x(t) - w_i(t)]$, onde $\alpha(t)$ é a taxa de aprendizado no instante t .
- v Repita o processo até que o critério de parada seja alcançado. O critério de parada pode ser, por exemplo, definido pelo número máximo de interações possíveis ou pelo percentual de alterações nos valores dos pesos após uma nova atualização.

3.4 Filtro de Kalman

Em 1960, R.E. Kalman, publicou o famoso trabalho que descreve um algoritmo que fornece estimativas de algumas variáveis desconhecidas, dadas as medições observadas ao longo do tempo (KALMAN, 1960). O filtro de Kalman, como ficou conhecido, pode ser visto como um algoritmo que utiliza os dados ruidosos coletados por sensores, uma estimativa de estado anterior e um modelo dinâmico a fim de produzir uma estimativa de estado atual mais próximo da realidade (MILLIDGE et al., 2021).

O Filtro de Kalman opera de forma iterativa, combinando um modelo matemático do sistema com observações obtidas de sensores para produzir uma estimativa otimizada do estado verdadeiro do sistema (VIEIRA et al., 2020). O algoritmo do Filtro de Kalman é estruturado em duas fases principais: a *Fase de Previsão* e a *Fase de Atualização*. Na *Fase de Previsão*, representada pelas equações Equação 14 e Equação 15, o filtro utiliza o modelo do sistema para prever o próximo estado e sua incerteza. Especificamente, a Equação 14 descreve a evolução do estado do sistema, enquanto a Equação 15 quantifica a incerteza associada a essa previsão. Neste contexto, o símbolo A é de particular importância, pois representa a matriz de transição de estados que modela a dinâmica do sistema. As equações 14 a 19 descrevem o funcionamento do Filtro de Kalman.

$$x_k = Ax_{k-1} + Bu_k + w_k \quad (14)$$

Para a Equação 14 - Predição do Estado:

- x_k : Representa o estado estimado do sistema no instante k .
- A : Matriz de transição de estado.
- x_{k-1} : Estado estimado no instante anterior.

- B : Matriz que relaciona a entrada de controle ao estado.
- u_k : Entrada de controle.
- w_k : Ruído do processo.

$$P_k = AP_{k-1}A^T + Q_k \quad (15)$$

Para a [Equação 15](#) - Predição da Covariância do Erro:

- P_k : Matriz de covariância do erro de estimativa.
- Q_k : Matriz de covariância do ruído do processo.

$$K_k = \frac{P_k H^T}{H P_k H^T + R} \quad (16)$$

Para a [Equação 16](#) - Ganho de Kalman:

- K_k : Ganho de Kalman.
- H : Matriz que relaciona o estado ao valor medido.
- R : Matriz de covariância do ruído de medição.

$$x_k = x_k + K_k[y_k - Hx_k] \quad (17)$$

Para a [Equação 17](#) - Atualização do Estado:

- y_k : Valor medido.

$$y_k = Cy_{km} + Z_k \quad (18)$$

Para a [Equação 18](#) - Relação de Medição:

- C : Matriz que relaciona o estado ao valor medido.
- y_{km} : Valor medido anterior.

- Z_k : Ruído associado à medição.

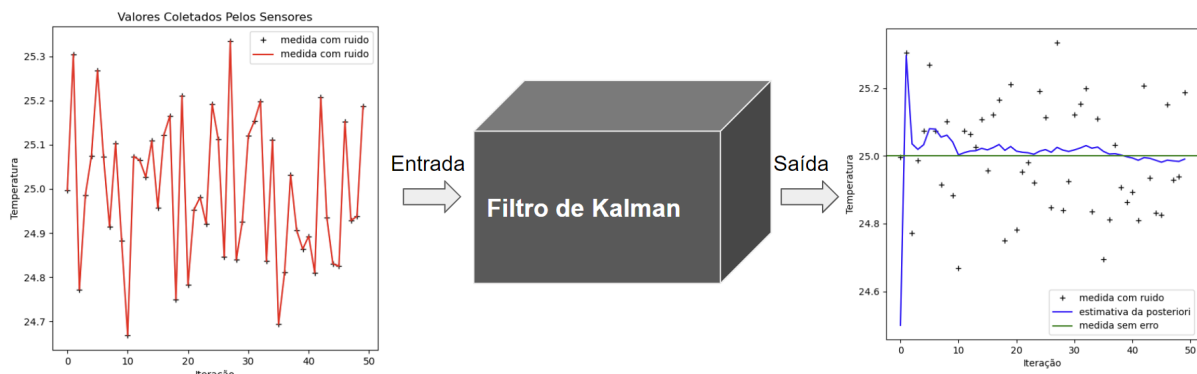
$$P_k = (\mathbf{I} - K_k \mathbf{H}) P_k \quad (19)$$

Para a [Equação 19](#) - Atualização da Covariância do Erro:

- $\mathbf{I}$: Matriz identidade.

A ideia central do filtro de Kalman pode ser vista na [Figura 27](#). Por meio da [Figura 27](#), é possível perceber que o filtro de Kalman, visto como uma caixa preta, possui como entrada dados ruidosos e que, após aplicar a filtragem, os dados de saída são menos ruidosos e mais precisos. Desse modo, o filtro de Kalman estima parâmetros do sistema que foram e não foram observados durante a definição do sistema dinâmico linear. Assim, apesar do nome “filtro”, esse processo matemático pode ser visto como um estimador de estados.

Figura 27 – Representação de um processo utilizando o filtro de Kalman



Fonte: O autor (2023)

Como o filtro de Kalman estima parâmetros não observados e não medidos pelos sensores, além da velocidade em encontrar uma estimativa para o valor real de uma variável, é possível empregá-lo para a construção de sistemas que operam em tempo real ([LI et al., 2015](#)), na área da robótica para a movimentação mais precisa dos equipamentos ([PATNAIK; KONAR; MANDAL, 2002](#)), em veículos autônomos ([COSTANZI et al., 2018](#)) e ([KASPER; SCHMIDT, 2008](#)), no diagnóstico e no tratamento de pacientes ([FOUSSIER et al., 2014](#)) e ([SCHMITT et al., 1995](#)), na precisão do tempo ([CASSOLA; BURLANDO, 2012](#)) dentre várias outras aplicações que empregam sistemas dinâmicos e necessitam de um resultados mais precisos.

O filtro de Kalman é uma ferramenta estatística que estima o estado de um sistema dinâmico a partir de uma série de medições incompletas e ruidosas. No contexto do planejamento e

operação do sistema de transporte urbano de passageiros por meio de ônibus, por exemplo, entender o funcionamento do filtro de Kalman é de grande importância para compreender a dinâmica dos passageiros. Isso ocorre porque o tráfego urbano é altamente dinâmico e sujeito a inúmeras variáveis imprevisíveis, como condições climáticas, acidentes e flutuações na demanda de passageiros. O filtro de Kalman pode processar dados em tempo real de várias fontes, como GPS, sensores de tráfego e relatórios de passageiros, para fornecer estimativas precisas e atualizadas sobre a localização dos ônibus, o fluxo de tráfego e os padrões de demanda. Essas informações são relevantes para a tomada de decisões informadas sobre o roteamento, agendamento e alocação de recursos, permitindo uma operação mais eficiente e confiável do sistema de transporte. Ao adaptar-se continuamente às mudanças nas condições de tráfego, o filtro de Kalman ajuda a minimizar atrasos, melhorar a utilização da frota e aumentar a satisfação dos passageiros, tornando-se uma ferramenta importante para os gestores de transporte urbano moderno.

Para a construção do filtro de Kalman, deve ser considerado que um sistema dinâmico que é caracterizado por um conjunto de n variáveis de estado expressa pelo vetor $x(t)$, de dimensão $n \times 1$. Desse modo, a evolução do sistema de estados (de t para $t + 1$) pode ser descrita de forma recorrente. Seja $F(t)$ a matriz, $n \times n$, de transição de estados, ou seja, a matriz que promove a evolução do sistema dinâmico durante a variação do tempo t . Considerando o sistema como linear, essa evolução pode ser obtida por meio do produto entre o vetor $x(t)$ e a matriz $F(t)$. Além disso, considerando que esse sistema dinâmico pode sofrer a influência de outras variáveis e utilizando o vetor $u(t)$, com dimensões $m \times 1$, para representar essas variáveis e a influência dessas variáveis para o estado $x(t + 1)$ como uma matriz $G(t)$, é possível representar a transição de estados como presente na [Equação 20](#):

$$x(t + 1) = F(t)x(t) + G(t)u(t) \quad (20)$$

Como a caracterização do sistema dinâmico pode não representar fielmente a transição de estados de um sistema real, pode-se utilizar medidas do ambiente externo para ajustar a modelagem presente na [Equação 20](#). Considerando $z(t)$ como um vetor $m \times 1$ que representa as medidas externas coletadas no tempo t , pode ser expressa a variação dessas medidas por meio da [Equação 21](#):

$$z(t) = H(t)x(t) \quad (21)$$

Nesse processo de obter o valor do estado $x(t + 1)$, o filtro de Kalman utiliza duas fases principais. A cada previsão de estado encontrada após a aplicação da [Equação 21](#), existe uma incerteza associada, pois variáveis externas podem interferir nos resultados obtidos.

Essas incertezas podem ser modeladas como uma matriz de covariância. Para as incertezas presentes em $x(t)$, associa-se uma matriz $P(t)$ e para as incertezas presentes na vetor $u(t)$, associa-se uma matriz de covariância $Q(t)$. Adotando essas definições para as matrizes de covariância $P(t)$ e $Q(t)$, é possível associar a [Equação 22](#) como a atualização da matriz $P(t + 1)$.

$$P(t + 1) = F(t)P(t)F(t)^T + Q(t) \quad (22)$$

A [Equação 20](#) e a [Equação 22](#) são conhecidas como as equações da fase de previsão do filtro de Kalman. Assim, essas equações são utilizadas para realizar a previsão de estados, ou seja, a [Equação 20](#) apresenta o valor inicial obtido pelo modelo desenvolvido e a [Equação 22](#) apresenta a covariância associada a essa estimativa inicial de estados.

Pela [Equação 20](#) e pela [Equação 22](#), é possível realizar a previsão de estados apenas utilizando o modelo “analítico” do sistema. Como esse modelo pode apresentar incertezas, a adição de outras fontes de informação podem auxiliar no melhor ajuste desse modelo. A fim de prover outra fonte de informação para a previsão de estados do sistema e melhorar o ajuste presente na [Equação 20](#) e na [Equação 22](#), são coletados, a cada iteração t , dados externos. Esses dados podem ser relacionados ao sistema dinâmico por meio da [Equação 21](#).

Assim como ocorre na previsão realizada pelo modelo analítico presente na [Equação 21](#) e na [Equação 22](#), os dados coletados por sensores podem apresentar incertezas, sejam elas devido ao processo de medição empregado ou ao próprio equipamento utilizado. Essas incertezas, da mesma maneira como apresentado na [Equação 20](#), podem ser representadas por uma matriz de covariância $R(t)$.

Analisando as equações para a modelagem analítica da variação de estados ([Equação 20](#) e [Equação 22](#)) e a construção empírica dessa variação de estados ([Equação 21](#)) é possível identificar que existem dois modelos distintos para a construção das transições de estados (analítico e empírico). Desse modo, a fim de definir uma estimativa mais precisa dos estados do sistema dinâmico, é possível utilizar os conceitos da Estimativa Bayesiana para realizar a junção desses dois modelos.

O Teorema de Bayes poder ser enunciado por meio da [Equação 23](#).

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (23)$$

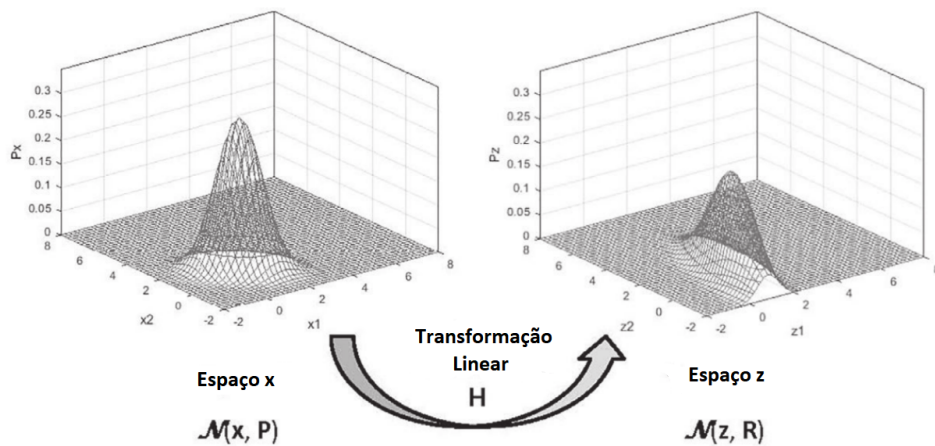
onde:

- $P(A|B)$ é a probabilidade posterior de A dado B .
- $P(B|A)$ é a probabilidade de B dado A .
- $P(A)$ é a probabilidade a priori de A .
- $P(B)$ é a probabilidade total de B .

Na [Equação 23](#), a porção referente *a priori* pode ser expressa, para o filtro de Kalman, como sendo a [Equação 20](#) e a [Equação 22](#), ou seja, a modelagem analítica do sistema dinâmico. As medidas $z(t)$ e a matriz de covariância $R(t)$ podem ser vistas como a função de verosimilhança (*likelihood*) correspondente ao modelo empírico. A marginalização ($P(B)$), nesse contexto, pode ser vista apenas como uma porção necessária para a normalização das estimativas de estados.

Analisando a [Equação 21](#) é possível perceber que a matriz $H(t)$ representa o mapeamento do espaço das variáveis $x(t)$ para o espaço das variáveis medidas $z(t)$. Desse modo, adotando uma distribuição normal para as variáveis de estado $x(t)$ e $z(t)$, tem-se uma transformação linear que pode ser representada pela [Figura 28](#).

Figura 28 – Transformação linear entre os espaços da variável de estados $x(t)$ e das medidas da variável $z(t)$



Fonte: Adaptado de [Barreto et al. \(2020\)](#)

Para o Teorema de Bayes apresentado na [Equação 23](#), que leva ao ajuste da previsão dos estados, considere que a distribuição *a priori* pode ser expressa pela [Equação 24](#) e a função de verosimilhança pode ser representada pela [Equação 25](#)

$$N(Hx_B, HP_BH^T) \quad (24)$$

$$N(z, R) \quad (25)$$

Desse modo, o produto entre as [Equação 24](#) e [Equação 25](#) resulta na [Equação 26](#), em que $\mu_p = \mu_0$ e a covariância é $\Sigma_p = \Sigma_0$ e K é a matriz $K = \Sigma_0$:

$$N(\mu_p, \Sigma_p) \quad (26)$$

Assim, a posteriori será a distribuição com o vetor de médias presente na [Equação 27](#) e com a matriz de covariância presente na [Equação 28](#) em que o K é apresentado na [Equação 29](#).

$$Hx_B + K(z - Hx_B) \quad (27)$$

$$HP_BH^T - KHP_BH^T \quad (28)$$

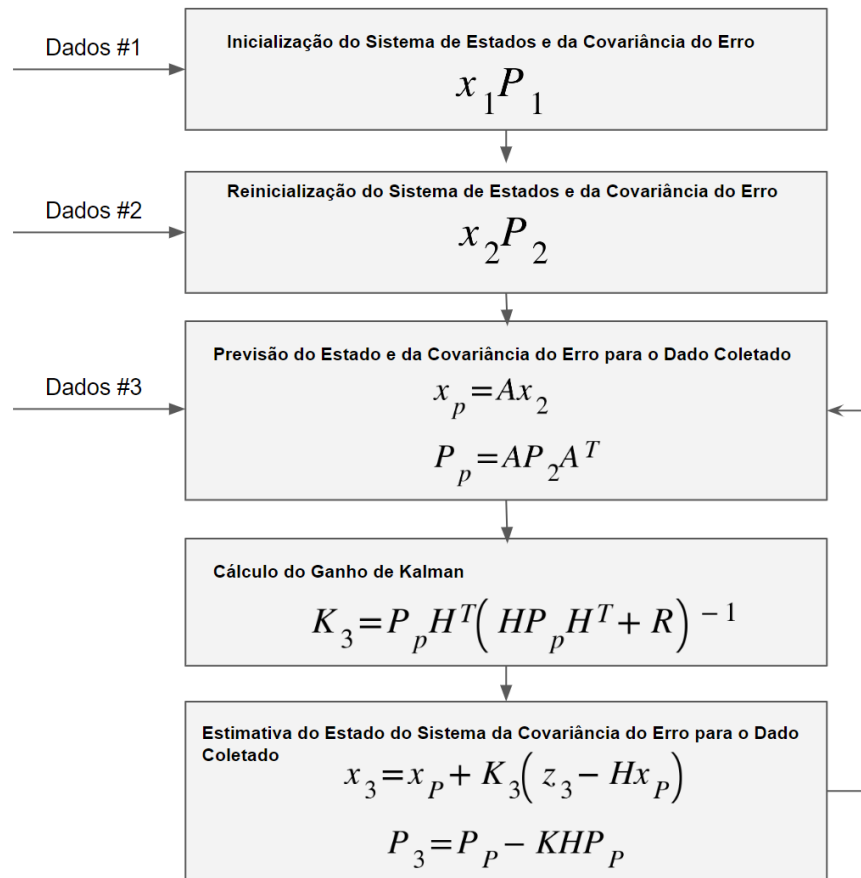
$$K = HP_BH^T(HP_BH^T + R)^{-1} \quad (29)$$

Aplicando todas as transformações necessárias, utilizando a [Equação 20](#) a [Equação 29](#), é possível construir o filtro de Kalman utilizando 5 (cinco) passos principais. Esses passos podem ser visualizados na [Figura 29](#).

Para a [Figura 29](#) temos que cada uma das variáveis utilizadas possuem as seguintes definições:

- x : vetor (n x 1) que define a variável de estado;
- P : matriz (n x n) que define a covariância de estados;
- z : vetor (m x 1) que define os dados coletados;
- A : matriz (n x n) que define a transição de estados;
- H : matriz (m x n) que define a transição de estados dos dados coletados;
- R : matriz (m x m) que define a covariância dos dados coletados;
- Q : matriz (n x n) que define o ruído nos dados coletados;
- K : matriz (n x m) que define o ganho de Kalman.

Figura 29 – Representação gráfica do algoritmo do filtro de Kalman



Fonte: Adaptado de Barreto et al. (2020)

Utilizando o algoritmo presente na Figura 29, é possível obter uma estimação mais precisa para o vetor de estados $x(t)$ utilizando duas fontes distintas de informação: (1) o modelo analítico e (2) o modelo empírico, $z(t)$, obtido por meio das medidas externas. Desse modo, o filtro de Kalman pode ser visto como um estimador ótimo, desde que todas as distribuições envolvidas (modelos e erros na medição) sejam Gaussianas (BARRETO et al., 2020).

Portanto, o filtro de Kalman emerge como uma ferramenta essencial para a análise e otimização do sistema de transporte urbano. Sua capacidade de processar informações em tempo real e ajustar-se dinamicamente a variáveis imprevisíveis torna-o uma solução robusta para enfrentar os desafios inerentes ao tráfego urbano. A precisão e eficiência deste filtro, ao lidar com incertezas e fornecer estimativas confiáveis, são importantes para a tomada de decisões informadas no planejamento e gerenciamento do transporte. Conectando-se diretamente aos objetivos desta tese, o filtro de Kalman não só reforça a necessidade de ferramentas avançadas no cenário atual de transporte, mas também destaca o potencial de inovações tecnológicas para transformar e melhorar a mobilidade urbana, especialmente em contextos desafiadores como o de países em desenvolvimento.

Capítulo 4

Metodologia

Este capítulo é dedicado à apresentação da metodologia empregada nesta tese, estabelecendo o caminho para a realização dos objetivos propostos. Inicialmente, a [Seção 4.1](#) abordará a coleta e armazenamento dos dados, detalhando os equipamentos, processos e estratégias utilizados. A [Seção 4.2](#) focará na modelagem da taxa de carregamento, explicando como os dados experimentais são coletados e validados. A modelagem da Matriz O-D será explorada na [Seção 4.3](#), incluindo a construção e validação da matriz. Por fim, a [Seção 4.4](#) discutirá as técnicas e algoritmos utilizados, fornecendo uma visão abrangente das abordagens computacionais adotadas. Juntas, essas seções fornecem uma compreensão completa da metodologia, permitindo a replicação e avaliação crítica do trabalho.

4.1 Coleta e Armazenamento dos Dados

Esta seção se dedica a detalhar o processo de coleta e armazenamento de dados, componentes cruciais para garantir a qualidade e confiabilidade das informações utilizadas neste trabalho. Serão inicialmente apresentados os equipamentos utilizados, oferecendo uma visão clara das ferramentas e tecnologias empregadas. Em seguida, o processo de coleta de dados será descrito, elucidando as etapas e metodologias adotadas para a obtenção das informações. Por fim, a seção abordará as estratégias e sistemas utilizados para o armazenamento dos dados, garantindo sua integridade e acessibilidade para análises futuras.

4.1.1 Equipamentos Utilizados

Uma importante característica dessa proposta metodológica reside no desenvolvimento de um equipamento de baixo custo capaz de coletar, gerar indicadores e analisar dados relativos ao sistema de transporte coletivo urbano. Este equipamento é concebido para coletar de forma passiva os endereços de *Medium Access Control* (MAC) dos dispositivos

móveis. A coleta ocorre por meio de *probe requests* emitidos pelos smartphones. Os *probe requests* são pacotes de dados enviados por um dispositivo de rede sem fio para procurar redes *Wireless Fidelity* (Wi-Fi) disponíveis ao seu redor. Esses pacotes são enviados periodicamente pelos dispositivos para identificar redes próximas e estabelecer uma conexão. As informações contidas nos *probe requests* podem incluir o nome da rede *Service Set Identifier* (SSID), endereço MAC do dispositivo e outros dados relevantes para identificação (FREUDIGER, 2015).

Na estrutura do equipamento, cada endereço MAC coletado é considerado uma entidade única, seja ela um passageiro embarcado, um pedestre, veículos, redes Wi-Fi etc. A parte de *hardware* do equipamento é composta por um Raspberry Pi (RPi) modelo B, alimentado por uma bateria externa. O *software*, por sua vez, é baseado no Sistema Operacional (SO) Kali Linux, versão 2022.1. Este equipamento é ilustrado na Figura 30.

Figura 30 – *Hardware* utilizado para a construção do sistema



Fonte: o autor (2021)

Os principais diferenciais deste equipamento incluem a facilidade de utilização, o baixo custo de construção, a sua natureza compacta, eficiência energética e a conectividade. A conectividade é importante, pois permite que o equipamento seja facilmente integrado às

operações dos sistemas de transporte existentes sem necessidade de alterações ou interferências. Essa integração facilita a comunicação e o compartilhamento de dados, tornando a implementação, manutenção e treinamento pessoal mais viáveis. Isso é particularmente útil para a coleta de dados em sistemas de transporte coletivo de passageiros, onde a conectividade pode melhorar a eficiência e a resposta às mudanças nas condições de tráfego.

Para complementar a funcionalidade do RPi, um *smartphone* com Sistema Operacional Android é usado. Um aplicativo desenvolvido para este *smartphone* recebe os dados MAC coletados pelo RPi via *Bluetooth*. No *smartphone*, esses dados são enriquecidos com informações de localização obtidas por GPS e *timestamps*. Esses dados aprimorados são então armazenados em um arquivo de texto.

Este sistema de *hardware* e *software* foi inicialmente projetado para coletar dados para calcular a taxa de carregamento (a variação do número de passageiros durante uma viagem em um ônibus de transporte coletivo). Para gerar essa taxa, foi desenvolvida uma arquitetura de dois estágios, ilustrada na [Figura 31](#).

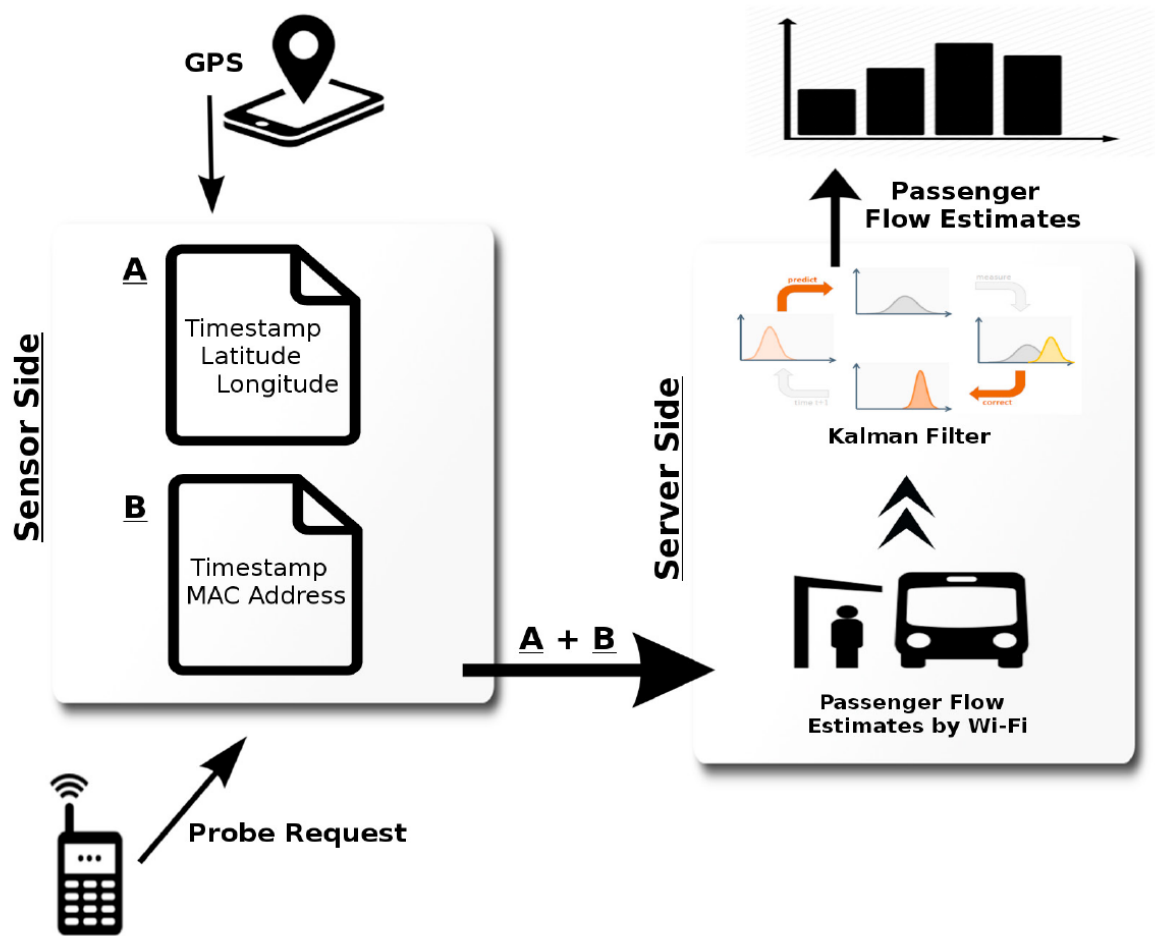
O primeiro estágio corresponde ao sensoriamento (*sensor side*). Aqui, os *probe requests* são capturados pelo RPi e enviados ao *smartphone*, onde são anexados *timestamps*. A partir desses dados, o *smartphone* gera dois arquivos. Um desses arquivos contém o *timestamp*, latitude e longitude, e o outro contém todos os endereços MAC coletados durante uma viagem. Estes arquivos são então enviados para o segundo estágio. A [Tabela 2](#) mostra um exemplo de dados de latitude e longitude coletados e a [Tabela 3](#) apresenta um exemplo de dados MAC coletados.

O segundo estágio é o servidor de processamento (*server side*). Neste estágio, um terceiro arquivo contendo a localização dos pontos de embarque e desembarque para a linha de ônibus em análise é adicionado aos dois arquivos recebidos. Os dados desse terceiro arquivo são incorporados ao sistema por meio do *software* desenvolvido. Com base nestes três arquivos, é aplicado um algoritmo para inferir os momentos de embarque e desembarque dos passageiros. Nesta etapa, considera-se que cada passageiro possui um único

Tabela 2 – Modelo de arquivo contendo a marcação temporal, latitude e longitude dos MAC coletados

Unix Timestamp	Latitude	Longitude
1531765160	-19.89871416	-44.03903528
1531765161	-19.89866144	-44.0390124
1531765162	-19.89861313	-44.03899247
1531765163	-19.89855613	-44.03897109
1531765164	-19.8985008	-44.03894966

Figura 31 – Esquema dos componentes principais do sistema proposto



Fonte: Adaptado de [Vieira et al. \(2020\)](#)

smartphone. Portanto, cada entrada e saída de um MAC no sistema pode ser interpretada como um passageiro.

Tabela 3 – Modelo de arquivo contendo os MAC coletados e a marcação temporal

Unix Timestamp	Endereço MAC
1531765161	44:33:4c:64:85:8d
1531765163	50:92:b9:85:89:1a
1531765164	44:33:4c:63:9e:25
1531765164	44:33:4c:63:9e:25
1531765164	44:34:4c:63:9e:25
1531765166	38:9a:f6:7f:f8:3b

4.1.2 Processo de Coleta de Dados

O processo de coleta dos dados para a construção do sistema proposto ocorre em duas etapas principais: a detecção de dispositivos móveis e a captação de *probe requests*.

A detecção de dispositivos móveis é realizada por meio do protocolo Wi-Fi. Em geral, os dispositivos móveis emitem sinais regulares, conhecidos como *probe requests*, em busca de redes disponíveis para conexão. Esses *probe requests* contêm informações únicas para cada dispositivo, incluindo o endereço MAC, um identificador singular atribuído a cada equipamento habilitado para rede, como um *smartphone*. O endereço MAC permanece constante, independentemente da rede à qual o dispositivo está conectado (FREUDIGER, 2015).

O equipamento, equipado com um Raspberry Pi model B, que já possui um adaptador Wi-Fi, opera como um *sniffer* de Wi-Fi, uma ferramenta que captura os sinais de tráfego Wi-Fi ao redor e coleta *probe requests* emitidos pelos dispositivos móveis nas proximidades. Isso ocorre de forma passiva, ou seja, o equipamento apenas escuta e coleta os sinais de Wi-Fi sem interferir ou interagir com eles. Isso torna o processo de coleta de dados não intrusivo e compatível com as regulamentações de privacidade, pois as informações coletadas não são suficientes para identificar ou rastrear indivíduos (NOBRE et al., 2019).

No momento da coleta, o equipamento registra o endereço MAC e a hora da detecção do *probe request*, conhecida como *timestamp*. Este *timestamp* é essencial para determinar quando e por quanto tempo um determinado dispositivo móvel esteve dentro do alcance do equipamento.

A coleta de dados é contínua durante o funcionamento do sistema de transporte, fornecendo uma fonte constante de dados atualizados sobre a utilização do sistema. Esses dados são então usados para gerar indicadores úteis para o planejamento e a gestão do sistema de transporte. Com o tempo, essa coleta contínua de dados permite a identificação de padrões e tendências, o que pode auxiliar na identificação de problemas e oportunidades para melhorias no sistema de transporte coletivo de passageiros.

4.1.3 Armazenamento dos Dados

Para maximizar a eficiência e a simplicidade do sistema proposto, o armazenamento dos dados coletados é realizado em arquivos de texto (.txt). Esta decisão decorre de algumas vantagens inerentes a este formato.

Primeiramente, os arquivos de texto são extremamente leves e eficientes em termos de espaço de armazenamento. Isso é de suma importância para o nosso sistema de baixo custo, uma vez que permite a coleta e o armazenamento de grandes volumes de dados

sem sobrecarregar os recursos de *hardware* ou de armazenamento.

Em segundo lugar, os arquivos de texto são compatíveis com uma ampla variedade de linguagens e ferramentas de programação, tornando-os altamente versáteis para análise de dados. Podem ser lidos e manipulados facilmente por bibliotecas de análise de dados em linguagens como Python e R, além de poderem ser importados para plataformas de análise de dado, como o Excel ou o Tableau.

Além disso, os arquivos de texto oferecem a vantagem de serem facilmente transformáveis em outros formatos, como *Comma-Separated Values* (CSV) ou *Java Object Notation* (JSON), dependendo das necessidades específicas da análise. Isso proporciona uma grande flexibilidade ao sistema proposto, permitindo a adaptação a diferentes metodologias de análise de dados e requisitos de processamento (YUSOF; MAN, 2018).

Outro característica importante é que os arquivos de texto são facilmente legíveis tanto por humanos quanto por máquinas. Isso facilita o rastreamento de problemas, a depuração e a manutenção do sistema, uma vez que os dados coletados podem ser facilmente examinados e compreendidos.

No entanto, uma desvantagem na utilização desse formato para o armazenamento dos dados é que, ao contrário de formatos de dados mais estruturados, como os bancos de dados *Structured Query Language* (SQL) ou *Not Only SQL* (NoSQL), os arquivos de texto não possuem mecanismos de consulta integrados (LI; MANOHARAN, 2013). Isso pode dificultar a extração de informações específicas dos dados armazenados, especialmente quando se trabalha com grandes volumes de dados. Dessa maneira, os dados armazenados em arquivos podem ser consolidados utilizando bancos relacionais, a fim de contornar essa possível desvantagem.

Portanto, ao optar por armazenar os dados coletados em arquivos de texto, maximiza-se a eficiência, a simplicidade, a versatilidade e a transparência do sistema proposto, facilitando a realização de análises robustas e significativas para o planejamento e monitoramento do sistema de transporte coletivo urbano.

4.2 Modelagem da Taxa de Carregamento

Nesta seção, o foco é a sobre a modelagem da taxa de carregamento, um elemento-chave para entender a dinâmica e eficiência do sistema de transporte. Será discutido o processo de coleta de dados experimentais, que serve como base para a construção e validação do modelo proposto. Além disso, a validação manual da taxa de carregamento será explorada, demonstrando os métodos e critérios adotados para assegurar a precisão e relevância dos resultados obtidos.

4.2.1 Coleta de Dados Experimentais

Para verificar a aplicabilidade do sistema e inferir a taxa de carregamento a partir dos *probe requests*, foi realizado um experimento coletando dados entre os dias 16/07/2019 e 20/07/2019 para três diferentes linhas de ônibus (1404A, 1404B e 1404C) na cidade de Belo Horizonte, Minas Gerais, Brasil. Essas linhas foram escolhidas devido à rota que percorrem, atravessando a região central de Belo Horizonte, uma área de alta atração de passageiros.

A efetiva aplicação dessa estratégia levou em consideração algumas características cruciais:

- Identificação precisa dos Pontos de Embarque e Desembarque (PEDs);
- Diferenciação dos MACs (passageiros) embarcados dos outros MACs coletados (redes Wi-Fi, veículos, passageiros não embarcados etc);
- Encontro dos momentos de embarque e desembarque dos passageiros;
- Tratamento de casos onde os passageiros não possuem celular ou não estão com o Wi-Fi ligado.

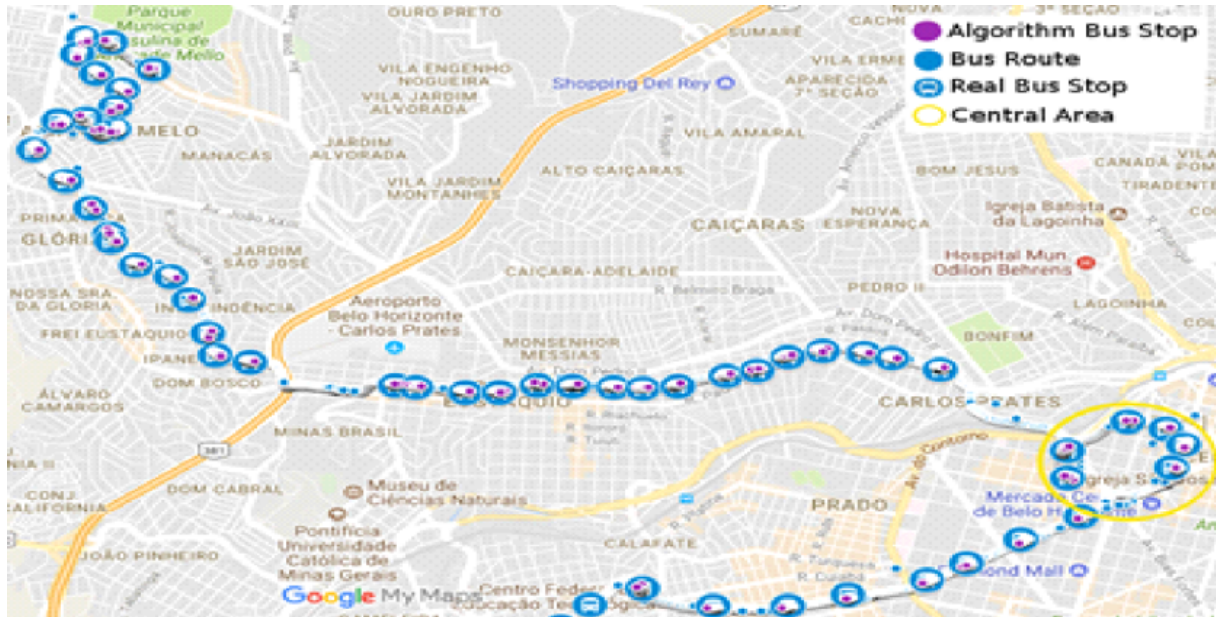
Para tratar esses desafios, foram adotadas algumas estratégias. Para cada PED, a localização geográfica (latitude e longitude) foi obtida. Dado que cada PED representa um ponto de localização e há um erro inerente ao processo de geolocalização, cada PED foi representado como uma região e não apenas um ponto. Isso permite determinar quando um ônibus entrou ou saiu de um PED.

Os PED definidos como uma região foram encontrados por meio da utilização da distância de Haversine. A distância de Haversine, ou grande círculo, corresponde a distância angular entre dois pontos em uma superfície esférica ([CHOPDE; NICHAT, 2013](#)). Desse modo, a partir desse modelo, é possível encontrar uma região provável de um PED levando em consideração a latitude e longitude do ponto de parada do ônibus.

A [Figura 32](#) ilustra um dos trajetos realizados pela linha 1404A. No mapa, o círculo amarelo representa os PEDs existentes na região central de Belo Horizonte. Os pontos em roxo indicam as paradas identificadas pelo sistema, enquanto as marcações em azul representam os PEDs reais fornecidos pela companhia de transporte.

Para identificar os MACs que correspondem aos passageiros verdadeiramente embarcados, foram aplicados alguns filtros. O primeiro filtro é responsável por identificar o primeiro e o último momento em que cada um dos MACs é coletado. Assim, são selecionados apenas os MACs que apresentam um intervalo de tempo entre a primeira e a última incidência superior a 10min. Este intervalo é considerado o tempo mínimo para uma viagem válida

Figura 32 – Rota para a linha 1404A



Fonte: Adaptado de [Vieira et al. \(2020\)](#)

de um passageiro. Além disso, MACs coletados menos de 10 vezes foram excluídos do sistema, uma vez que tendem a não representar uma amostra significativa para análise.

4.2.2 Validação Manual da Taxa de Carregamento

A fim de validar a efetividade da estratégia empregada, foi realizada uma contagem manual do fluxo de passageiros que embarcaram e desembarcaram no veículo em cada um dos PEDs estudados durante a viagem.

A validação manual do carregamento proporciona uma comparação direta e efetiva entre os dados coletados pelo sistema e a realidade observada no contexto do transporte público coletivo. Esta etapa é fundamental para identificar eventuais discrepâncias, ajustar parâmetros e otimizar a precisão do modelo proposto.

Nesta contagem, cada entrada e saída de passageiro foi registrada pelo pesquisador, formando um conjunto de dados que reflete a variação do número de passageiros ao longo do trajeto. Este conjunto de dados torna-se a referência, a base real, contra a qual o fluxo de passageiros indicado pelo sistema é comparado.

Por exemplo, ao observar uma parada específica, o pesquisador anota quantos passageiros entraram e saíram do veículo. Posteriormente, essas informações são comparadas com os dados fornecidos pelo sistema. Se o sistema indicar que dez passageiros embarcaram, mas a contagem manual indicar que apenas oito entraram, isso apontará para uma possível

ineficiência que deve ser investigada e resolvida.

A semelhança entre esses números – a taxa de carregamento indicada pelo sistema e a contagem manual realizada – é uma forte indicativa da eficácia da estratégia proposta. Divergências, por sua vez, podem levar a importantes *insights* para aprimorar a técnica e tornar a inferência da taxa de carregamento cada vez mais precisa e confiável.

4.3 Modelagem da Matriz Origem-Destino (O-D)

Esta seção apresenta a modelagem da matriz O-D, um componente essencial no planejamento e gerenciamento do sistema de transporte. Esta seção é dividida em duas partes principais. A [Subseção 4.3.1](#) aborda a Criação da Matriz O-D, detalhando os métodos e processos envolvidos na construção dessa matriz, que representa os padrões de deslocamento dentro de uma área geográfica. A [Subseção 4.3.2](#), por sua vez, se concentra na Validação da Matriz O-D, explicando as técnicas utilizadas para assegurar que a matriz seja precisa e confiável. Juntas, essas subseções fornecem uma visão abrangente da importância da matriz O-D e de como ela pode ser efetivamente modelada e validada.

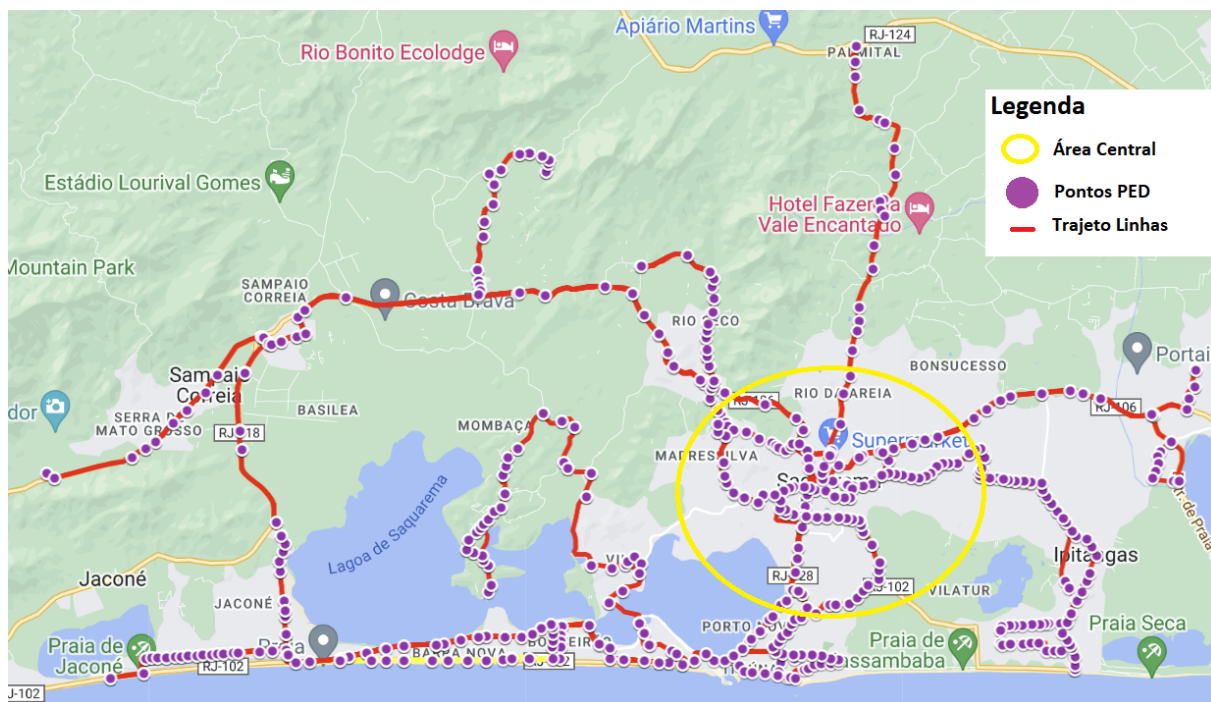
4.3.1 Criação da Matriz O-D

O segundo experimento corresponde à criação da matriz O-D. Para esse experimento foram coletados dados de múltiplas linhas de transporte coletivo em Saquarema, no Rio de Janeiro. Em um intervalo de tempo que se estende de 14/09/2022 a 27/09/2022, foram instalados oito aparelhos em diferentes veículos, cobrindo, assim, 11 (onze) rotas únicas de transporte coletivo. A [Figura 33](#) ilustra os trajetos e os (PEDs) existentes na cidade. A escolha da cidade de Saquarema-RJ como local de experimentação está relacionada ao desenvolvimento do Plano de Mobilidade Urbana na cidade, que estava em andamento no momento da realização desta pesquisa.

Localizada no Rio de Janeiro, Saquarema é uma cidade com uma população de 89 mil habitantes, conforme dados do [Instituto Brasileiro de Geografia e Estatística \(2022\)](#). Economicamente, a cidade se destaca, com a indústria contribuindo com 64,9% do seu Produto Interno Bruto (PIB) e os serviços com 29,9%. Este desempenho resulta em um PIB per capita que supera as médias estadual e regional. Em termos demográficos, Bacaxá e Porto Da Roça são os bairros mais populosos, enquanto Itaúna e Rio Seco se destacam pelo alto percentual de residentes com ensino superior. Estas características, juntamente com um elevado potencial de consumo e uma forte concentração de empregos no setor de comércio e serviços, refletem o perfil da mobilidade urbana da população e as dinâmicas econômicas locais de Saquarema.

Nesse experimento, assim como no anterior, foram priorizadas as linhas de transporte

Figura 33 – Mapa geral do transporte público em Saquarema-RJ



Fonte: O autor (2023)

que percorriam o centro da cidade, dada a expressiva quantidade de passageiros que circulam por essa área. Essa decisão possibilitou um exame mais detalhado do sistema nas regiões de maior interesse para o planejamento do transporte público. Considerando que os aparelhos foram instalados em veículos específicos, que poderiam servir a rotas diferentes em dias diferentes, foi necessário empregar um algoritmo para identificar em qual linha o veículo estava operando, com base nos PED e trajetos dispostos na Figura 33. Isso permitiu a aquisição do código identificador dos PEDs, possibilitando a verificação da mudança de linha pelo passageiro durante sua viagem.

O mesmo *hardware* utilizado no experimento anterior também foi utilizado para coleta de dados na construção da matriz O-D. Os códigos MAC obtidos passaram pelos mesmos filtros usados anteriormente. Porém algumas alterações foram necessárias. Devido à instalação dos aparelhos em diversos ônibus e linhas de transporte, foi preciso adicionar mais duas etapas ao processo de criação da matriz O-D.

A primeira etapa adicional corresponde à fase de filtragem de dados, onde foi preciso inserir filtros capazes de detectar as mudanças de linha feitas pelos passageiros. Essa necessidade se deve ao fato de que o mesmo endereço MAC pode ser registrado em linhas diferentes. Dessa maneira, o tempo de viagem e as coordenadas (latitude e longitude) são empregados para determinar quando um passageiro muda de ônibus durante o percurso.

A segunda etapa adicional se relaciona à consolidação dos dados. Nesse momento, é preciso reunir os diversos endereços MAC que representam os passageiros embarcados para identificar os pontos de embarque e desembarque, bem como os horários de cada viagem realizada. A [Tabela 4](#) apresenta um exemplo da matriz Origem-Destino (O-D) inicial gerada com a inclusão dessas duas etapas adicionais.

O produto final desse segundo experimento é uma matriz contendo informações essenciais como: código MAC do passageiro, número do veículo onde o código MAC foi capturado, data e turno de embarque, latitude e longitude do ponto de embarque, código identificador do PED de embarque, data e turno de desembarque, latitude e longitude do ponto de desembarque, código identificador do PED de desembarque, distância total da viagem em metros e o tempo total da viagem. Estas informações são de vital importância para a formação da matriz O-D, permitindo uma análise minuciosa do fluxo de passageiros e das alterações de linha que ocorrem durante a viagem.

4.3.2 Validação da Matriz O-D

Para a validação da matriz O-D construída, foi necessário recorrer aos dados históricos. Esses dados, apesar de não refletirem os mesmos períodos e trajetos investigados durante o experimento, serviram como uma base relevante para comparar e validar os resultados obtidos pelo sistema proposto.

Ainda que não houvesse uma correspondência direta em termos de período e trajeto, os dados históricos contribuíram para um parâmetro comparativo. Esta comparação foi viabilizada pela semelhança na natureza dos dados, permitindo, assim, uma avaliação consistente da precisão e eficiência do método proposto.

No processo de validação, cada variável da matriz O-D gerada foi confrontada com os respectivos dados históricos. Isso incluiu aspectos como o código MAC do passageiro, o número do veículo, os dados de embarque e desembarque e a distância total da viagem.

Logo, mesmo sem uma correspondência direta, a validação da matriz O-D por meio da aplicação de dados históricos desempenhou um papel fundamental na metodologia adotada, permitindo uma avaliação detalhada da acurácia do sistema proposto e fornecendo *insights* valiosos para seu contínuo aprimoramento.

Tabela 4 – Exemplo de dados consolidados após a etapa adicional

MAC	DATA EMB	LAT EMB	LONG EMB	...	DATA DES	LAT DES	LONG DES
00:12:36:c1:f0:8a	1663240431	-22,8917	-42,4335	...	1663243266	-22.8959	-42.4693
a8:e3:ee:ba:ad:21	1663240439	-22,8917	-42,4335	...	1663243277	-22.8955	-42.4693
c8:3a:35:32:e1:49	1663240444	-22,8918	-42,4333	...	1663243746	-22.8930	-42.4699

4.4 Técnicas e Algoritmos Utilizados

Esta seção aborda as técnicas e algoritmos empregados nesta pesquisa, dividindo-se em duas partes principais. A primeira, [Subseção 4.4.1](#), foca na Modelagem e Análise do Sistema Proposto, detalhando os métodos e ferramentas utilizados para construir e avaliar o sistema proposto. Isso inclui a seleção de algoritmos, a estruturação do modelo e a avaliação de sua eficácia e eficiência. A segunda parte, [Subseção 4.4.2](#), concentra-se na Análise da Dinâmica das Viagens, explorando as técnicas aplicadas para entender e interpretar o comportamento e os padrões de viagem dos passageiros. Essa análise é importante para otimizar o planejamento e a operação do sistema de transporte, permitindo ajustes e melhorias com base em dados e *insights* concretos.

4.4.1 Modelagem e Análise do Sistema Proposto

As características únicas da coleta de dados de transporte público, incluindo o fato de que nem todos os passageiros possuem *smartphones* ou mantêm o Wi-Fi ligado durante a viagem, demandam um tratamento especial dos dados coletados. Para garantir a viabilidade dos dados como indicador confiável da taxa de carregamento e da matriz O-D, foram empregadas técnicas de filtragem e modelagem.

Um importante componente da estratégia proposta nesta tese é o uso do Filtro de Kalman no servidor de processamento de dados. Este filtro, conhecido por suas propriedades de estimativa ótima em sistemas com ruído gaussiano ([AUGER et al., 2013](#)), foi aplicado para corrigir e melhorar a estimativa do número de passageiros embarcados durante a viagem. Esse método possibilita a definição de um modelo matemático para a variação dos passageiros embarcados e ajusta esse modelo de acordo com os dados coletados.

Inicialmente, um modelo auto-regressivo foi empregado para construir a equação representativa da variação do fluxo de passageiros. A [Equação 30](#) é expressa da seguinte forma:

$$x_k = \alpha x_{k-1} + \text{norm}(2, 3) \quad (30)$$

Nessa [Equação 30](#), x_k é a quantidade de passageiros dentro do ônibus no PED k e α é determinado pela diferença entre os dados dos MAC e os dados históricos. A função $\text{norm}(2, 3)$ é utilizada para atualizar a quantidade de passageiros após cada PED, representando uma distribuição normal com média 2 e desvio padrão 3.

Os dados sendo coletados periodicamente (a cada 1 PED) demandaram uma investigação sobre a aplicabilidade de diferentes modelos de análise de séries temporais. Modelos paramétricos (ARIMA) e não paramétricos (RNR com LSTM e RNR com GRU), além

das CNN foram considerados para otimizar a representação do modelo de variação dos passageiros em uma viagem.

Nos modelos de redes neurais recorrentes, foram experimentadas configurações com um, dois, três e cinco atrasos entre as observações. As RNRs LSTM e GRU foram construídas com duas camadas ocultas de 64 e 32 neurônios, com uma taxa de abandono de 20% e função de perda MSE. Para as CNN, foi usada a convolução de uma dimensão com janela de tamanho 3, 64 filtros e *average pooling*. A [Tabela 5](#) mostra um exemplo dos dados utilizados para a construção dos modelos utilizando séries temporais.

Além disso, um modelo ARIMA (0,2,1) - com defasagem igual a 0, grau de diferenciação igual a 2 e média móvel igual a 1 - foi utilizado como estratégia paramétrica para comparar com as técnicas de RNA.

Para avaliar o desempenho de cada algoritmo, foram utilizadas métricas de erro como o Erro Médio Absoluto (MAE), a Raiz do Erro Quadrático Médio (RMSE) e o Coeficiente de Determinação (R^2). A utilização dessas métricas permite verificar quão bem cada modelo se ajustava aos dados, a magnitude dos erros produzidos e o quão explicativo era cada modelo no contexto da variação do número de passageiros em uma viagem. Adicionalmente, os tempos de treinamento e predição também foram considerados na análise de desempenho, a fim de avaliar a eficiência dos algoritmos em termos de velocidade de processamento.

Cada uma dessas abordagens foi implementada e avaliada com base na eficácia de representar o comportamento de embarque e desembarque dos passageiros ao longo dos trajetos. A análise dessas técnicas e a comparação entre seus desempenhos constituem a base para a melhoria contínua do sistema.

Tabela 5 – Exemplo de dados utilizados para a construções dos modelos de análise de séries temporais

Porta	1		2		3		Total		Carregamento
Hora	P. in	P. out	P. in	P. out	P. in	P. out	P. in	P. out	
5:59:45	33	0	0	0	0	0	33	0	33
6:00:06	5	0	0	0	0	0	5	0	38
6:00:52	5	0	0	0	0	0	5	0	43
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
7:42:19	0	0	0	5	0	0	0	5	11
7:42:58	0	0	0	3	0	2	0	5	6
7:44:33	0	1	0	5	0	3	0	6	0

4.4.2 Análise da Dinâmica das Viagens

Além das técnicas de modelagem descritas anteriormente, foi também realizado um esforço para entender a dinâmica dos passageiros no transporte público de uma maneira mais ampla. Isso foi feito por meio de uma análise utilizando técnicas de agrupamento (*clustering*), que permitem segmentar os dados de acordo com as similaridades existentes entre eles.

Para essa análise de clusterização, características como o tempo de viagem, a localização dos pontos de embarque e desembarque, o número de viagens realizadas em diferentes períodos do dia (manhã, tarde e noite) e ao longo da semana foram consideradas. Adicionalmente, foi possível incorporar informações sobre a transferência de passageiros entre diferentes linhas, oferecendo uma representação mais precisa e abrangente da dinâmica dos passageiros.

O algoritmo de mapeamento auto-organizável, SOM, foi escolhido para executar essa análise de clusterização. A rede SOM é uma técnica de aprendizado não supervisionado que é capaz de transformar um espaço de entrada multivariável complexo em um espaço de saída de menor dimensão, enquanto preserva as relações topológicas dos dados de entrada. Assim, a análise com a SOM possibilita identificar como grupos de passageiros se deslocam e utilizam o sistema de transporte público, proporcionando *insights* valiosos para a gestão e o planejamento do transporte coletivo de passageiros.

A fim de avaliar a efetividade da aplicação da SOM para compreender o comportamento dos passageiros do sistema de transporte público, foram utilizadas algumas métricas de desempenho. Essas métricas correspondem à medida de Silhueta e os erros de quantização e topográfico (KOHONEN; KOHONEN, 2001). A medida de Silhueta oferece um ponto de vista sobre a distância média entre cada objeto e os demais dentro do mesmo grupo, além da distância média entre cada objeto e os que estão situados em outros grupos. Um valor de Silhueta mais próximo de 1 indica uma categorização eficiente, o que, no contexto do transporte público, pode ajudar a entender a homogeneidade do comportamento dos passageiros dentro de um determinado grupo (SHAHAPURE; NICHOLAS, 2020). O erro de quantização mostra o erro médio entre cada entrada de dado e o neurônio correspondente na camada de saída. Isso pode ajudar a avaliar a precisão com que o comportamento dos passageiros está sendo mapeado pela SOM. O erro topográfico, por sua vez, é a proporção de todas as entradas de dados para as quais as duas primeiras *Best Matching Units* (BMU) não são adjacentes, indicando possivelmente uma qualidade insatisfatória do mapa. Este é um indicativo de quão bem a SOM está mantendo a similaridade comportamental dos passageiros quando agrupados (KOHONEN, 2012).

Capítulo 5

Apresentação e Análise dos Resultados

Neste capítulo, são apresentados os resultados obtidos após a realização dos experimentos apresentados no [Capítulo 4](#). Além disso, é realizada uma análise da utilização do sistema na criação de indicadores do transporte urbano. São discutidas também algumas limitações e apresentadas algumas estratégias a serem desenvolvidas para a construção do *framework* proposto.

5.1 Estimativa da Taxa de Carregamento

Como apresentado no [Capítulo 4](#), os testes do sistema para a estimativa da taxa de carregamento foram realizados em 3 diferentes linhas urbanas do transporte coletivo de Belo Horizonte. A [Figura 34](#), a [Figura 35](#) e a [Figura 36](#) mostram a comparação da taxa de carregamento (número de passageiros embarcados) encontrada por meio do sistema proposto e os valores reais (coletados pelo pesquisador) durante uma viagem em cada uma das linhas.

Os dados presentes na [Figura 34](#), na [Figura 35](#) e na [Figura 36](#) (barras em verde) correspondem às estimativas realizadas apenas pelo algoritmo presente no sistema (lado do servidor), antes da aplicação do filtro de Kalman. Como já era esperado, os valores encontrados pelo sistema são, normalmente, inferiores aos valores reais. Isso ocorre pois nem todos os passageiros possuem *smartphone*, ou possuem o aparelho, mas estão com a função de conexão à rede Wi-Fi desligada. Desse modo, não é possível rastrear todos os passageiros apenas com a utilização dos endereços MAC. Analisando as figuras, é possível perceber que essa diferença é mais evidenciada nos PED finais da linha 1404C.

Analisando a [Figura 35](#), é possível perceber também que existe um descompasso entre os valores reais e os estimados pelo sistema. Assim, entre os PED 11 a 20 ([Figura 35](#)), a quantidade de passageiros embarcados estimada pelo sistema é superior à encontrada pela contagem manual. Caso seja realizado um deslocamento dos dados para a direita (barras

em verde), é possível reduzir essas diferenças, uma vez que as curvas (estimadas e real) possuem formas similares. Isso aconteceu porque a contagem manual de passageiros não utilizou a localização dos pontos de ônibus (latitude e longitude). Portanto, não foi possível comparar, exatamente, os PED obtidos pelo modelo proposto com a taxa de carregamento real. Assim, foi utilizada a sequência de dados obtidos pelo pesquisador para comparar com a abordagem proposta.

A partir dos gráficos presentes na [Figura 34](#), na [Figura 35](#) e na [Figura 36](#), nota-se que a taxa de carregamento em cada uma das linhas pesquisadas apresenta um comportamento distinto. Para realizar um planejamento adequado do sistema de transporte coletivo, é importante entender o comportamento de cada uma das linhas, já que elas possuem características particulares. Portanto, empregar estratégias como a proposta neste trabalho, pode auxiliar no entendimento dos padrões de movimentação dos passageiros que utilizam os ônibus urbanos.

Analizando a [Figura 34](#), a [Figura 35](#) e a [Figura 36](#), é possível identificar que os erros entre os valores da taxa de carregamento coletados manualmente (real) e os encontrados pelo sistema variam entre 0% e 60%. A fim de minimizar essa diferença, é utilizado o filtro de Kalman para reduzir as incertezas no processo de medição. A [Figura 37](#), a [Figura 38](#) e a [Figura 39](#) apresentam os resultados obtidos pelo sistema após a correção realizada pelo filtro de Kalman.

Como pode ser visto por meio dos gráficos presentes na [Figura 37](#), na [Figura 38](#) e na [Figura 39](#), existe uma aproximação entre os valores reais (curva vermelha) e os previstos pelo sistema (curva verde). A maior diferença entre esses valores passou a ser de 17% (PED 20 da linha 1404C).

Essa importante melhoria ocorreu devido à adoção do modelo de estado presente na [Equação 30](#) e as correções geradas pelo Filtro de Kalman ([Equação 14](#) a [Equação 19](#)). Como o modelo presente na [Equação 14](#) incorpora uma parcela do erro gerado pela impossibilidade de medir todos os MAC dos passageiros, o modelo apresenta uma variação relevante para a previsão do fluxo de passageiros em cada um dos PED. Esta parte é implicitamente usada pelo filtro de Kalman para introduzir incertezas estatísticas ao quadro geral de previsão do fluxo de passageiros e, assim, retratar com mais fidelidade as características de variação da taxa de carregamento durante uma viagem.

Para cada um dos gráficos presentes na [Figura 34](#), na [Figura 35](#) e na [Figura 36](#) (linhas 1404A, 1404B e 1404C), foi calculado o erro médio quadrático entre os valores real e o obtido pelo sistema sem o ajuste do filtro. Os valores encontrados foram, respectivamente, 64,93; 95,41 e 345,1. Para os dados presentes na [Figura 37](#), na [Figura 38](#) e na [Figura 39](#), também foi calculado o erro médio quadrático. Os valores encontrados após o modelo ajustado pelo

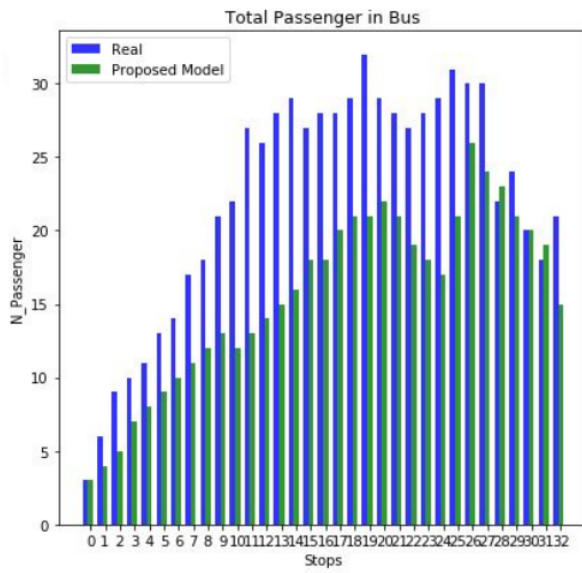


Figura 34 – Linha 1404A

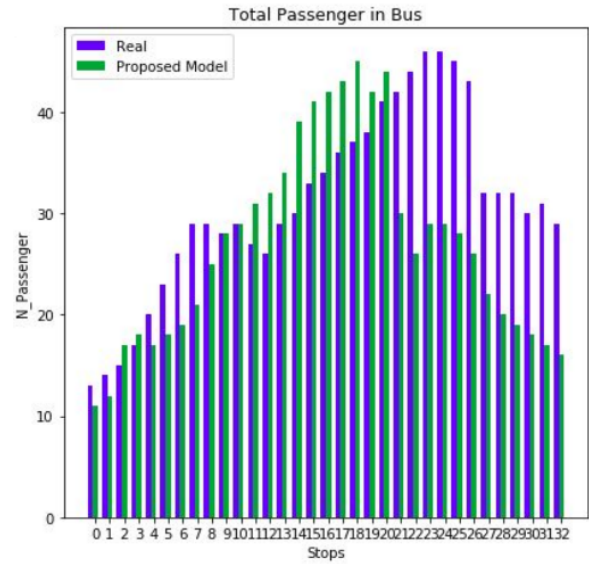


Figura 35 – Linha 1404B

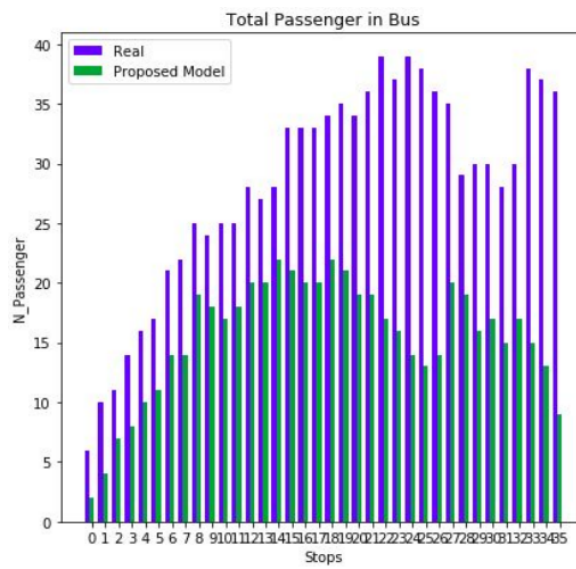


Figura 36 – Linha 1404C

filtro de Kalman, foram, respectivamente, 11,16; 41,32 e 119,4. Como mostrado, mesmo após a correção realizada pelo filtro de Kalman, a linha 1404C ainda apresenta o maior erro de estimativas para a taxa de carregamento. Além disso, esses dados mostram que as correções realizadas pelo filtro de Kalman contribuem para uma melhoria significativa nas estimativas da taxa de carregamento realizada pelo sistema proposto.

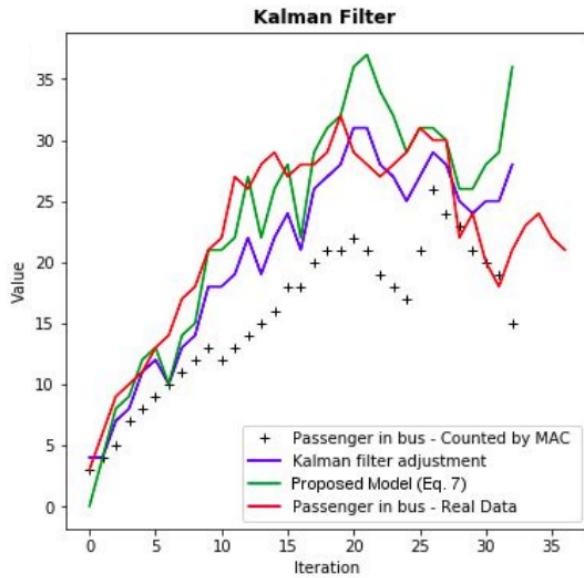


Figura 37 – Linha 1404A

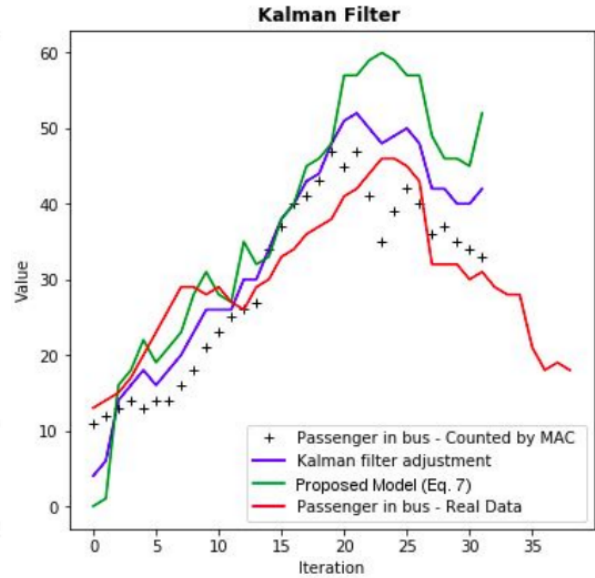


Figura 38 – Linha 1404B

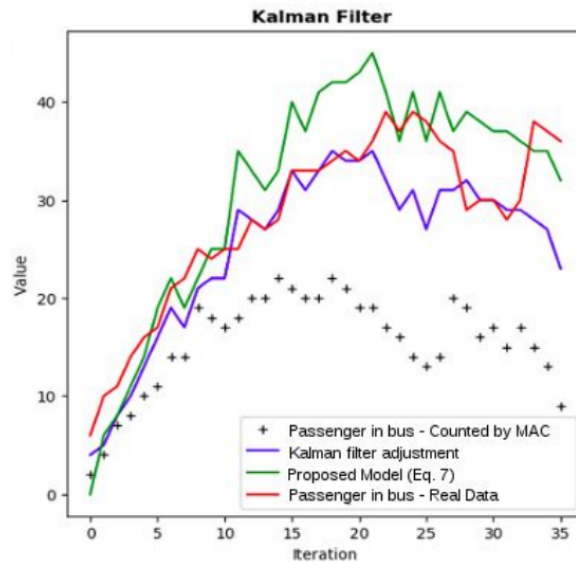


Figura 39 – Linha 1404C

5.2 Modelagem da Taxa de Carregamento

Os dados presentes na [Tabela 5](#), discutida no [Capítulo 4](#), representam os valores de passageiros que entraram e saíram por cada uma das portas do veículo durante uma viagem. A coleta desses dados foi realizada por meio da contagem manual da entrada e saída de passageiros nos PED estabelecidos pelo rota do ônibus. Desse modo, esses dados podem ser vistos como o valor real da variação da taxa de carregamento do veículo durante uma viagem.

Como a taxa de carregamento de uma mesma linha pode variar durante todo o dia (horários de pico da manhã, pico da tarde e entre picos), para as análises, foi utilizado o valor médio da taxa de carregamento entre esses três períodos distintos (pico manhã, tarde e entre picos). A [Figura 40-a](#) mostra os dados dos três horários coletados e o valor médio da taxa de carregamento empregado.

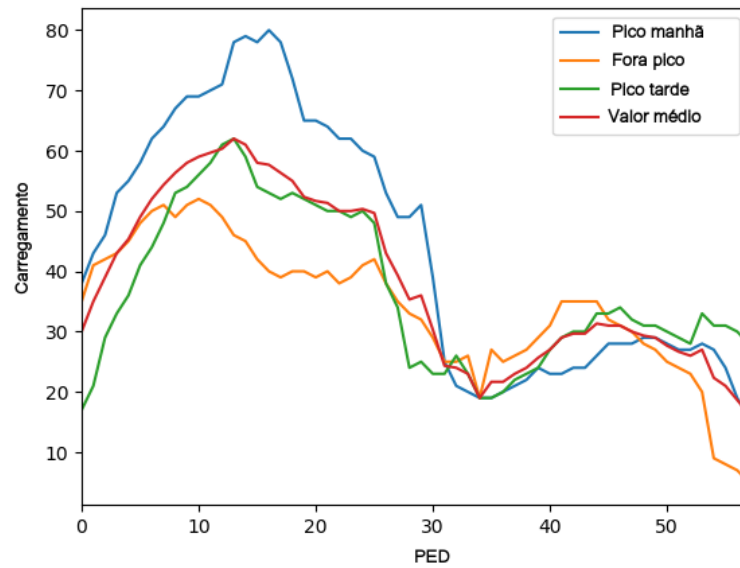
Pelos *boxplot* presentes na [Figura 40-b](#), é possível perceber que não existem *outliers* para os dados coletados durante as três viagens. Assim, é calculado o valor médio dessas viagens sem a necessidade de remover valores muito dispersos. Por meio dos dados presentes na [Figura 40-b](#), também é possível concluir que, durante a viagem no período do pico da manhã, existe uma maior dispersão dos dados quando comparados com os demais horários coletados.

A [Tabela 6](#), que contém os resultados da análise estatística após 100 execuções de cada uma das configurações (1, 2, 3 e 5 delays) para a RNR com o LSTM, fornece informações cruciais sobre a normalidade dos dados e as diferenças de desempenho entre as configurações. Os testes de normalidade (Shapiro-Wilk) ([RAZALI; WAH et al., 2011](#)) revelam valores de *p-valor* muito próximos de zero, indicando que os dados não seguem uma distribuição normal. Além disso, o teste não paramétrico (Nemenyi test) ([BRESLOW, 1970](#)) mostra diferenças significativas nos *p-valores* entre as configurações, comprovando estatisticamente que existem diferenças na *performance* entre as configurações de RNR com LSTM utilizando os diferentes delays (1, 2, 3 e 5) para prever a taxa de carregamento durante uma viagem. Esses resultados, portanto, reforçam a importância da escolha adequada da configuração para alcançar a previsão mais precisa da taxa de carregamento.

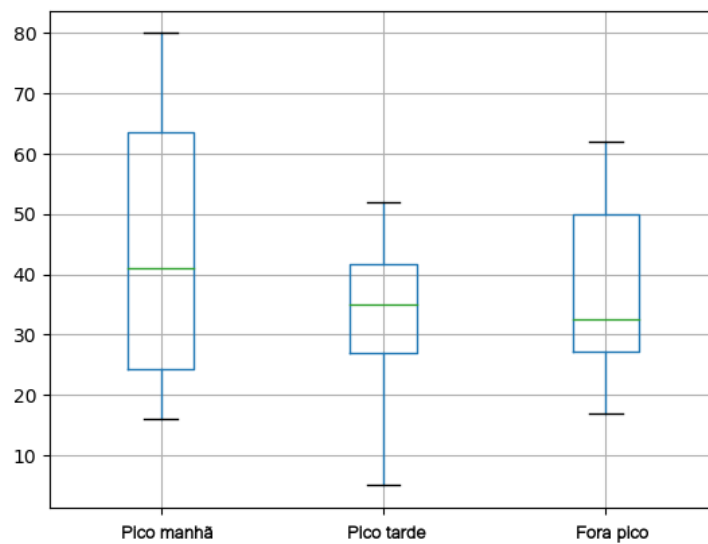
Tabela 6 – Análise estatística para RNR com LSTM

		p_valor: Nemenyi test			
	p_valor: Shapiro-Wilk	LSTM_1	LSTM_2	LSTM_3	LSTM_5
LSTM_1	2.3e-15	-1	0.001	0.001	0.001
LSTM_2	0.791		-1	0.885	0.578
LSTM_3	0.87			-1	0.900
LSTM_5	0.71				-1

Figura 40 – Valores e *boxplot* para as taxas de carregamento



a) Taxa de carregamento (picos, fora pico e médio)



b) Boxplot para as taxas de carregamento

Tabela 7 – Análise estatística para RNR com GRU

		p_valor: Nemenyi test			
	p_valor: Shapiro-Wilk	GRU_1	GRU_2	GRU_3	GRU_5
GRU_1	3.8e-9	-1	0.452	0.452	0.004
GRU_2	0.820		-1	0.900	0.221
GRU_3	0.949			-1	0.221
GRU_5	0.828				-1

Tabela 8 – Análise estatística para CNN

		p_valor: Nemenyi test			
	p_valos: Shapiro-Wilk	CNN_1	CNN_2	CNN_3	CNN_5
CNN_1	4.6e-12	-1	0.001	0.900	0.900
CNN_2	0.857		-1	0.001	0.001
CNN_3	0.655			-1	0.900
CNN_5	0.92				-1

A [Tabela 7](#) apresenta os resultados estatísticos obtidos para os testes utilizando a RNR com GRU. Como pode ser visto ([Tabela 7](#)), após as 100 execuções, apenas para a configuração com 1 delay não foi verificada a normalidade dos dados. Os resultados obtidos por meio da CNN com a janela de convolução com 1 dimensão e comprimento 3 são apresentados na [Tabela 8](#). Assim como ocorreu no caso anterior, apenas não foi identificada a normalidade com a utilização da janela de comprimento 1.

A [Tabela 9](#) e a [Figura 41](#) apresentam a comparação entre cada um dos métodos não paramétricos utilizando o delay de 1 unidade temporal. A [Figura 41](#) mostra o boxplot comparativo entre o erro alcançado após cada uma das 100 execuções (RNR LSTM, GRU e CNN) utilizando a configuração com 1 delay. Como pode ser visto, existe uma menor dispersão dos dados com a utilização da Rede Neural Recorrente com o LSTM.

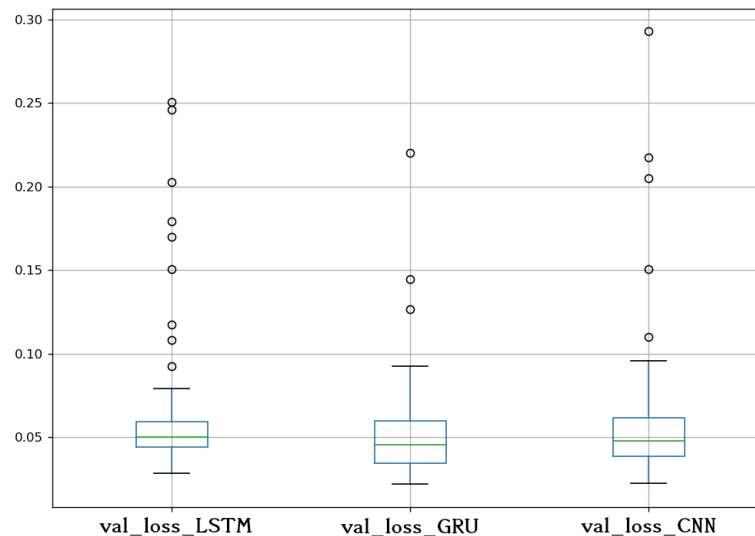
O coeficiente de determinação, frequentemente denotado como R^2 , é uma medida estatística que representa a proporção da variação na variável dependente que é previsível a partir das variáveis independentes. No contexto da identificação da taxa de carregamento, o R^2 pode ser utilizado para avaliar o quão bem o modelo se ajusta aos dados observados. Um valor de R^2 próximo de 1 indica que uma grande proporção da variabilidade na taxa de carregamento é explicada pelo modelo, enquanto um valor próximo de 0 sugere que o modelo não explica bem a variabilidade. Assim, o coeficiente de determinação é utilizado para avaliar a eficácia do modelo (DRAPER; SMITH, 1998). Na [Tabela 9](#) também está presente o R^2 para as previsões por meio de cada um dos algoritmos.

A [Figura 42](#) apresenta a previsão da taxa de carregamento encontrada para cada um dos algoritmos não paramétricos utilizando a melhor configuração alcançada (menor erro na previsão). Analisando a [Figura 42](#) e a [Tabela 9](#), é possível perceber que os resultados

Tabela 9 – Análise estatística para RNR LSTM vs RNR GRU vs CNN

	p_value: Shapiro-Wilk	R2 score	p_value: Nemenyi test		
			LSTM_1	GRU_1	CNN_1
LSTM_1	7.3e-16	0.827	-1	0.235	0.860
GRU_1	1.9e-10	0.893		-1	0.885
CNN_1	2.9e-15	0.883			-1

Figura 41 – Boxplot para RNR LSTM, RNR GRU e CNN



são bastante similares, apesar de existir uma menor variação entre os resultados obtidos empregando a RNR com LSTM.

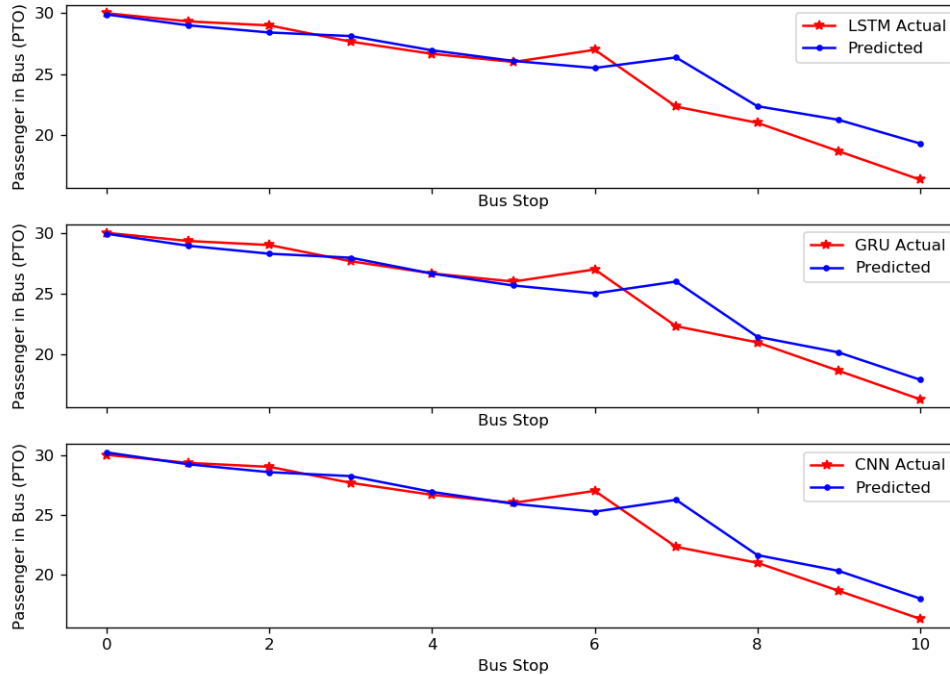
A fim de comparar as estratégias de AM com os algoritmos paramétricos, é utilizado o modelo ARIMA(0,2,1). Essa configuração foi escolhida, pois representa o modelo com os melhores resultados alcançados (Tabela 10 e Figura 43). A previsão da taxa de carregamento utilizando o ARIMA(0,2,1) é mostrada na Figura 44.

Tabela 10 – Análise estatística para o modelo ARIMA(0,2,1)

	Performance		Coef	Std err	P_valor
AIC	224.447	Const.	-0.1152	0.086	0.185
BIC	230.523	MA	-0.6337	0.091	0.000

Comparando os resultados de previsão por meio dos diferentes algoritmos (RNR com LSTM e GRU, CNN e ARIMA(0,2,1)) é possível perceber que ambos possuíam resultados satisfatórios (R^2 próximo a 0,9). Entretanto, como a modelagem paramétrica permite gerar um modelo matemático (Tabela 10) que demonstra a variação temporal da taxa de carregamento, esse modelo (ARIMA(0,2,1)) deve ser utilizado em conjunto com o filtro de Kalman

Figura 42 – Previsão da taxa de carregamento utilizando RNR LSTM, RNR GRU e CNN



na previsão da taxa de carregamento.

Portanto, os resultados da previsão utilizando diferentes algoritmos (RNR com LSTM e GRU, CNN e ARIMA(0,2,1)) apresentaram desempenhos satisfatórios, com valores de R2 próximos a 0,9. Embora todos os métodos tenham mostrado bons resultados, a modelagem paramétrica, especificamente o ARIMA(0,2,1), possibilitou a criação de um modelo matemático que demonstra a variação temporal da taxa de carregamento. Esse modelo pode ser utilizado em conjunto com o filtro de Kalman para aprimorar a precisão das previsões da taxa de carregamento, tornando-o uma abordagem promissora para a previsão de demanda no sistema de transporte coletivo.

5.3 Análise dos Dados do Experimento de Saquarema-RJ

Segundo os dados apresentados por [Saquarema \(2023\)](#), o sistema de transporte público coletivo de Saquarema é composto por 20 linhas principais, com rotas e pontos de embarque e desembarque bem definidos. As linhas de ônibus podem ser categorizadas como radiais, que ligam bairros periféricos aos centros da cidade, ou circulares, que circulam por um determinado trajeto com um ou dois pontos de controle. A [Figura 45](#) apresenta o trajeto das 20 linhas do sistema de transporte público coletivo da cidade de Saquarema.

Figura 43 – *Autocorrelation Function* (ACF) para o valor médio da taxa de carregamento

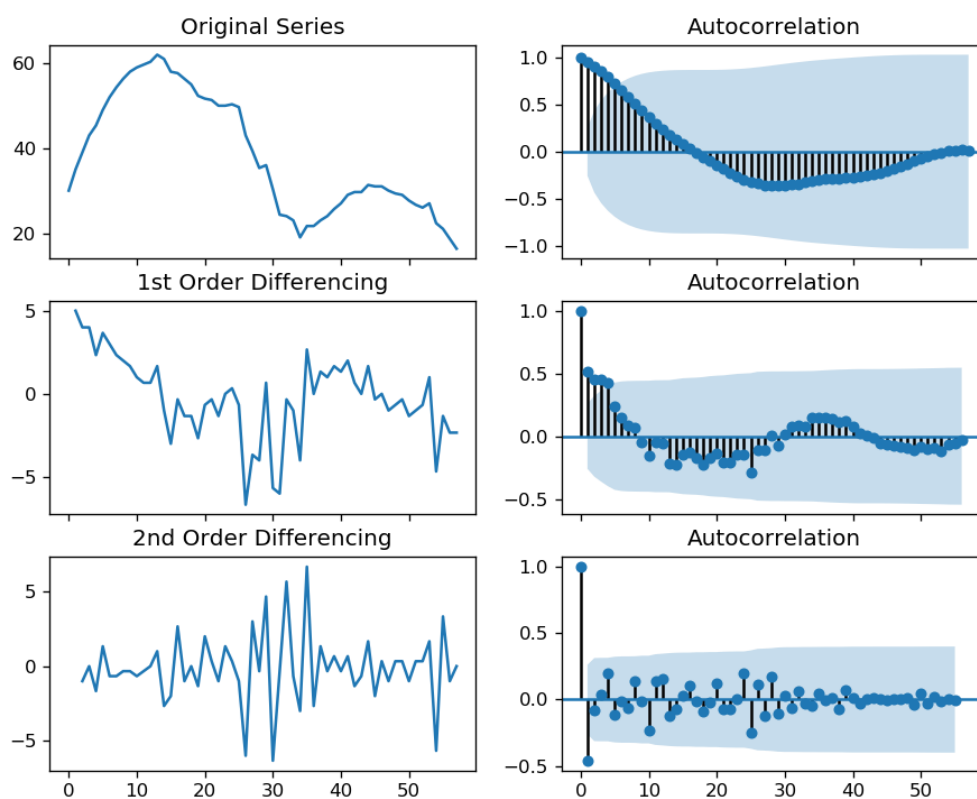
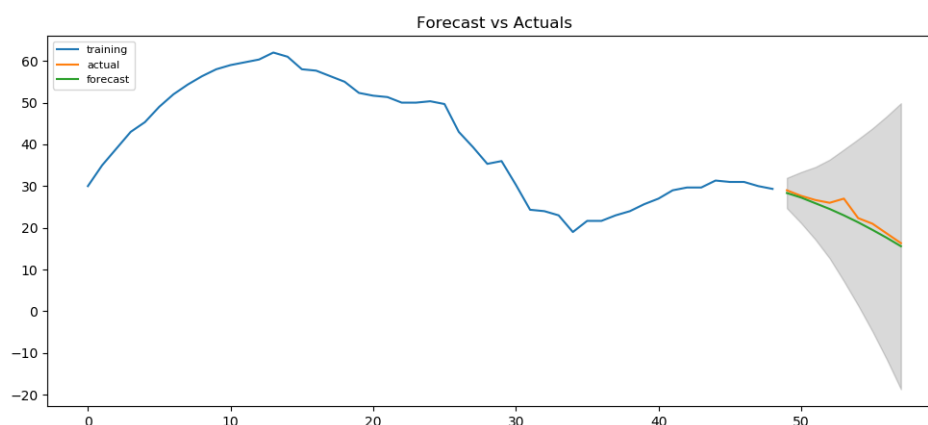
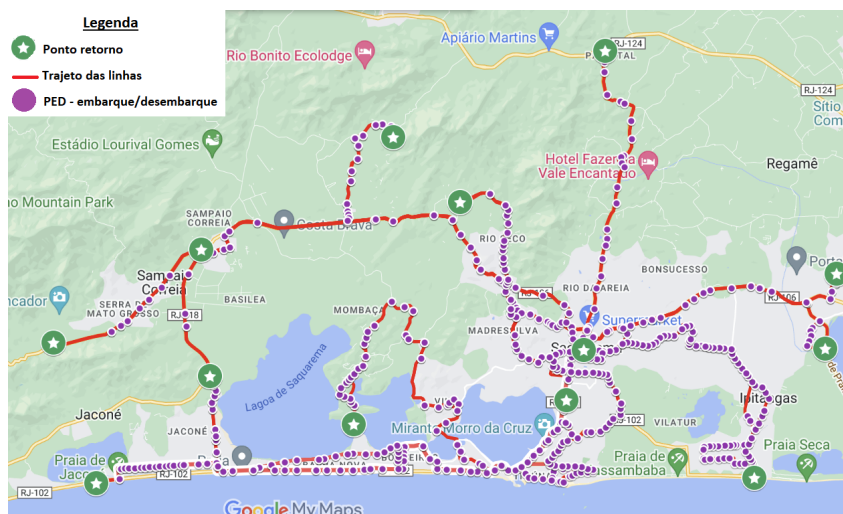


Figura 44 – Previsão utilizando ARIMA(0,2,1)



A cobertura da rede de transporte público é ampla, mas existem áreas urbanas densamente povoadas que estão distantes da rede, com uma distância de caminhada superior a 500 metros. Essas características do sistema de transporte público de Saquarema podem ser

Figura 45 – Trajetos e pontos de retorno do transporte público em Saquarema-RJ



Fonte: O autor (2023)

comparadas com os dados gerados pelo sistema proposto para avaliar a eficiência do serviço e identificar áreas de melhoria. Das 20 (vinte) linhas principais, 11 (onze) foram utilizadas para a construção da matriz OD. As linhas presentes no estudo e os dias em que esse estudo foi realizado estão presentes na [Tabela 11](#). Como pode ser visto pela [Tabela 11](#), não foi possível coletar os dados em todos os dias da semana para cada uma das linhas monitoradas. Isso se deve ao fato de, no momento da realização do experimento, a quantidade de equipamentos necessária para cobrir todos os ônibus do transporte coletivo de Siquemema-RJ não estar disponível.

Para a geração dos itinerários que operam essas 11 (onze) linhas nos dias informados na [Tabela 11](#), foram utilizados os próprios GPS presentes no sistema desenvolvido. A

Tabela 11 – Dados das linhas e datas utilizadas para coleta e avaliação do sistema proposto

Linha	Datas Coleta	Extensão da Linha (km)
Bacaxá-Ipitangas	21/09/2022	23,32
Bacaxá-Jaconé	22/09/2022	44,83
Bacaxá-Sampaio	23/09/2022	27,77
Bacaxá-Squarema (via Itaúna)	22/09/2022	21,02
Bacaxá-Vilatur	23/09/2022	26,88
Squarema-Bacaxá (via Guarani)	22/09/2022	25,75
Squarema - Rio Mole	23/09/2022	50,25
Squarema-Palmital	19/09/2022 - 21/09/2022 22/09/2022 - 23/09/2022	38,97
Squarema-Sampaio	19/09/2022 - 21/09/2022 22/09/2022	51,86
Squarema-Sampaio-Serra	20/09/2022	63,19
Bacaxá - Jaconé (Trevo Barra Nova)	19/09/2022 - 20/09/2022	30,80

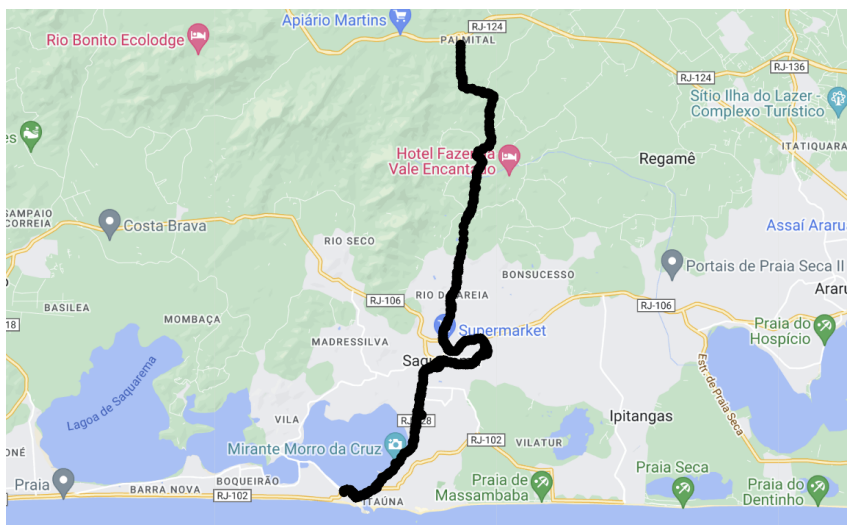
Figura 46 – Itinerário Bacaxá-Jaconé



Fonte: O autor (2023)

Figura 46, a Figura 47, a Figura 48 e a Figura 49 apresentam amostras dos dados gerados pelos equipamentos instalados para os itinerários Bacaxá-Jaconé, Saquarema-Palmital, Bacaxá-Vilatur e Saquarema-Rio Mole, respectivamente.

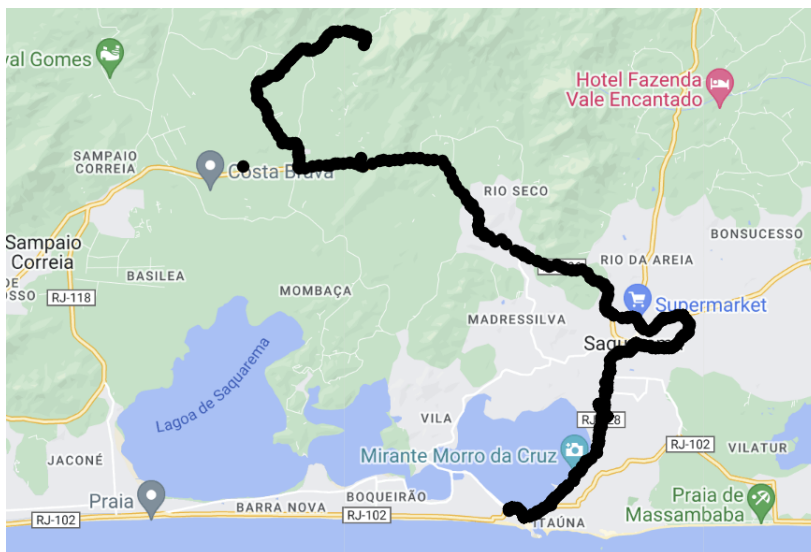
Figura 47 – Itinerário Saquarema-Palmital (ida)



Fonte: O autor (2023)

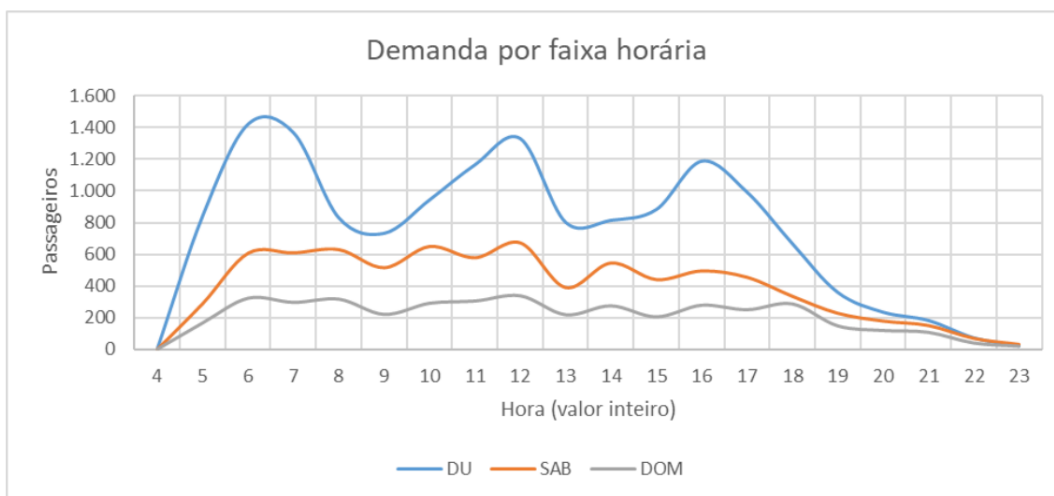
A fim de verificar a eficiência do sistema proposto, alguns dados como a variação de demanda foram comparados com os dados apresentados em Saquarema (2023). A Figura 50 apresenta a variação dessa demanda de passageiros do transporte público para o mês de setembro/2022, encontrada pelo Plano de Mobilidade Urbana Saquarema (2023). Por meio da Figura 50, é possível perceber que existem 3 períodos de pico de demanda. O primeiro pico ocorre no período da manhã entre às 06:00h e 07:59h. O segundo pico ocorre no período do almoço, entre 11:00h e 12:59h, e o terceiro pico, entre 16:00h e 17:00h. Com o intuito de comparar a variação desse demanda com o sistema desenvolvido, é apresentada

Figura 49 – Itinerário Saquarema-Rio Mole



Fonte: O autor (2023)

Figura 50 – Demanda de passageiros em setembro/2022 por faixa horária (Plano de mobilidade urbana)



Fonte: Adaptado de [Saquarema \(2023\)](#)

transporte em Saquarema-RJ. Na [Figura 51](#), embora o número de passageiros obtidos seja inferior ao apresentado na [Figura 50](#) devido à não realização da pesquisa em todas as linhas, a grande proximidade entre as duas curvas sugere que o sistema foi capaz de capturar as variações do fluxo horário de passageiros, mesmo com uma pequena amostra da demanda total.

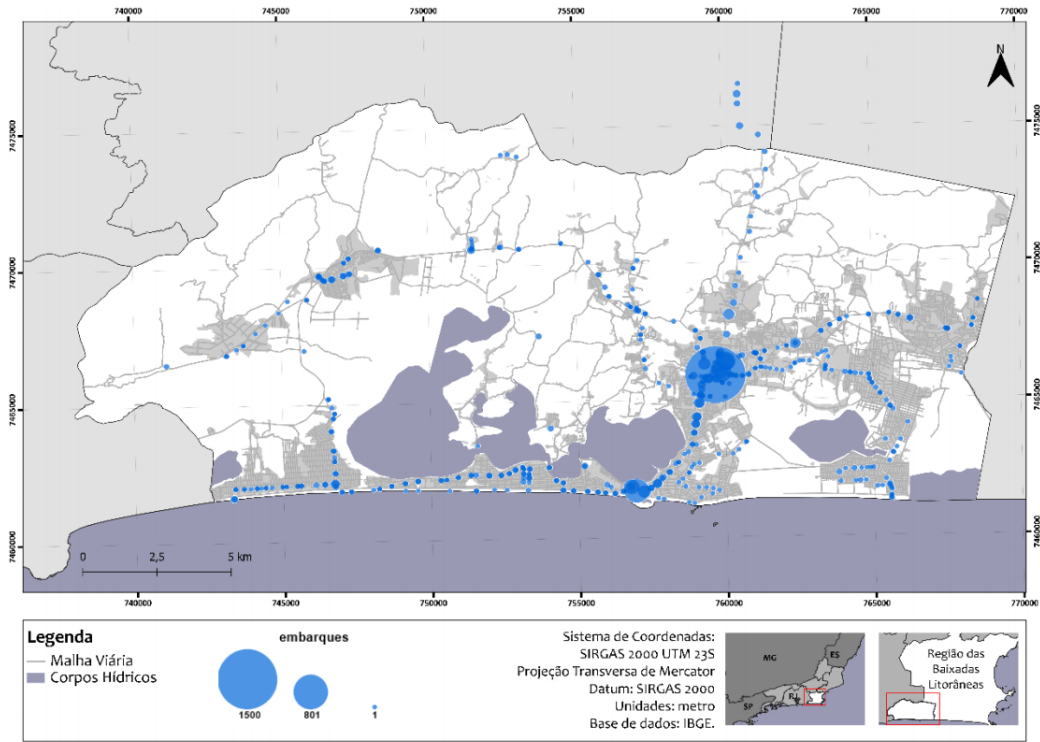
A **Figura 52** revela uma concentração significativa de pontos de embarque em duas principais regiões do município, representadas por dois grandes círculos azuis, identificada por

Figura 51 – Demanda de passageiros em setembro/2022 por faixa horária (Sistema proposto)



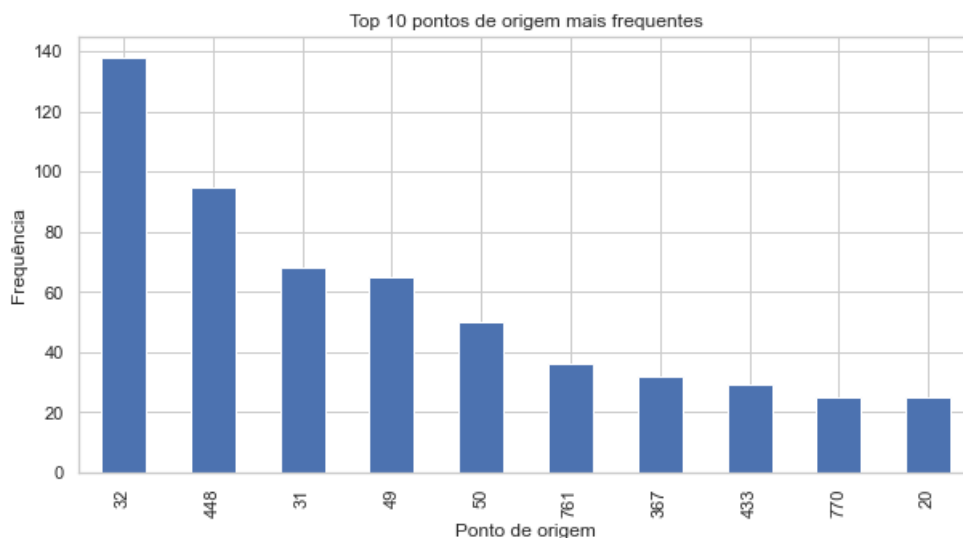
Fonte: O autor (2023)

Figura 52 – Carregamento dos pontos de embarque

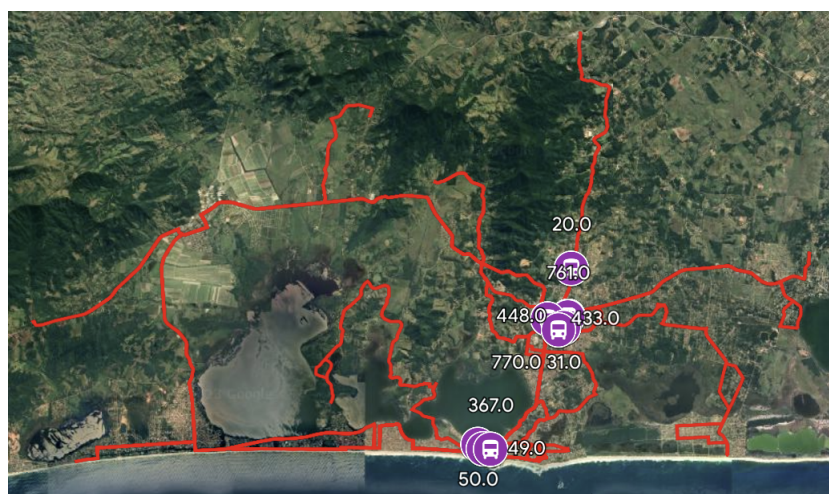


Fonte: Adaptado de Saquarema (2023)

Figura 53 – Comparação e localização dos principais PED encontrados pelo sistema



(a) Principais PED encontrados pelo sistema



(b) Localização dos principais PED encontrados pelo sistema

meio do estudo de mobilidade urbana. Essas concentrações coincidem com os principais pontos de embarque identificados pelo sistema proposto, conforme mostrado na [Figura 53](#), onde estão presentes os principais pontos encontrados pelo sistema. A correspondência exata desses locais, evidenciada pelas figuras, reforça a capacidade da ferramenta de detectar as distribuições desses pontos de embarque de forma precisa, sem a necessidade de pesquisas manuais. Esses resultados atestam a qualidade e relevância do sistema, evidenciando sua aplicabilidade como uma valiosa ferramenta para o planejamento e gerenciamento eficiente do sistema de transporte coletivo em Saquarema-RJ.

5.4 Estimativa da Matriz Origem-Destino

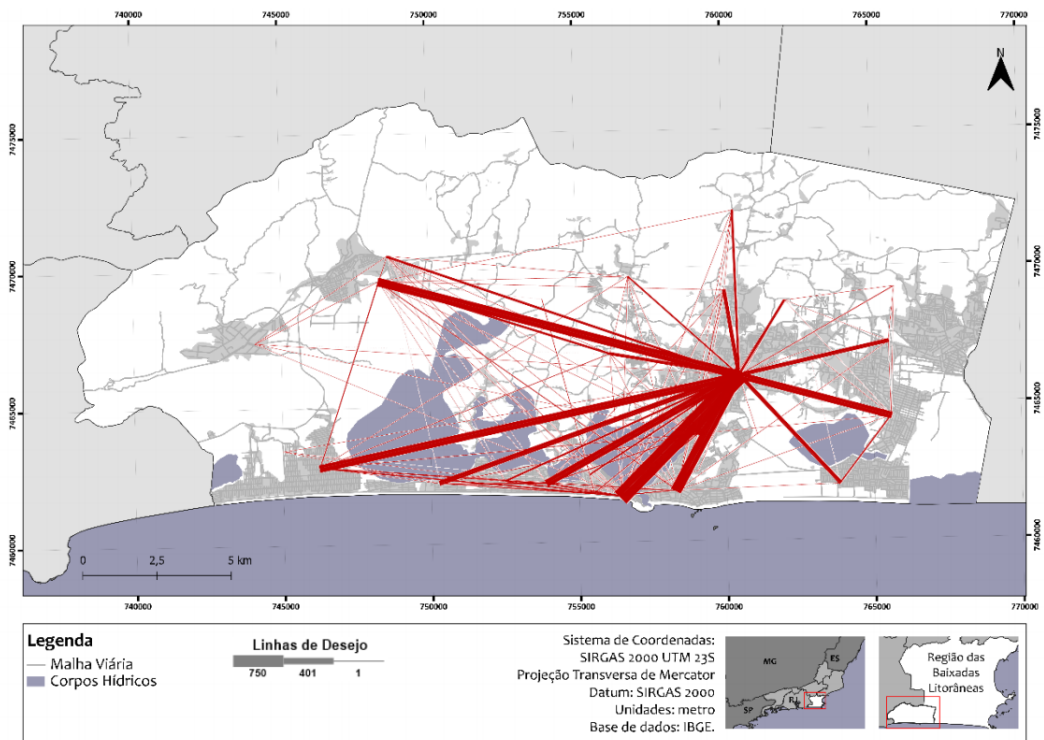
Conforme apresentado no [Capítulo 4](#), para estimar a matriz O-D, os equipamentos foram instalados em alguns ônibus do transporte coletivo de passageiros na cidade de Saquarema, no estado do Rio de Janeiro, Brasil. Para avaliar a eficácia dessa abordagem, a matriz O-D gerada pelo sistema proposto foi comparada com os dados da pesquisa de mobilidade urbana realizada na mesma cidade e disponível no *site* oficial do município de Saquarema-RJ ([SAQUAREMA, 2023](#)). Esta pesquisa fornece uma visão abrangente do sistema de transporte público da cidade, incluindo mapas de rotas e dados sobre demanda e oferta de passageiros. Ao comparar a matriz O-D gerada pelo sistema proposto com os dados da pesquisa, foi possível avaliar a precisão da abordagem sugerida e identificar possíveis discrepâncias ou semelhanças entre os dois conjuntos de dados. Esta comparação permitiu não apenas validar os resultados, mas também identificar áreas de melhoria potencial no sistema de transporte público de Saquarema-RJ.

A construção da matriz O-D é um processo que envolve a coleta e análise de um amplo espectro de dados. No sistema de *hardware* e *software* proposto, esta matriz é gerada por meio do rastreamento de diversos parâmetros de cada passageiro em seu trajeto no transporte público. Os parâmetros incluem identificação do passageiro (representado por meio do MAC), o veículo utilizado, datas e horas de embarque e desembarque, coordenadas geográficas de embarque e desembarque, identificação dos pontos de embarque e desembarque, a distância percorrida na viagem e o tempo de viagem. Por meio da coleta dessas informações, o sistema é capaz de mapear a origem e o destino de cada passageiro, bem como a rota e o tempo de viagem, proporcionando uma visão sobre a dinâmica dos movimentos dos passageiros. Esses dados são, então, processados e consolidados em um gráfico que representa a matriz O-D, fornecendo um panorama sobre os padrões de fluxo dos passageiros no sistema de transporte coletivo.

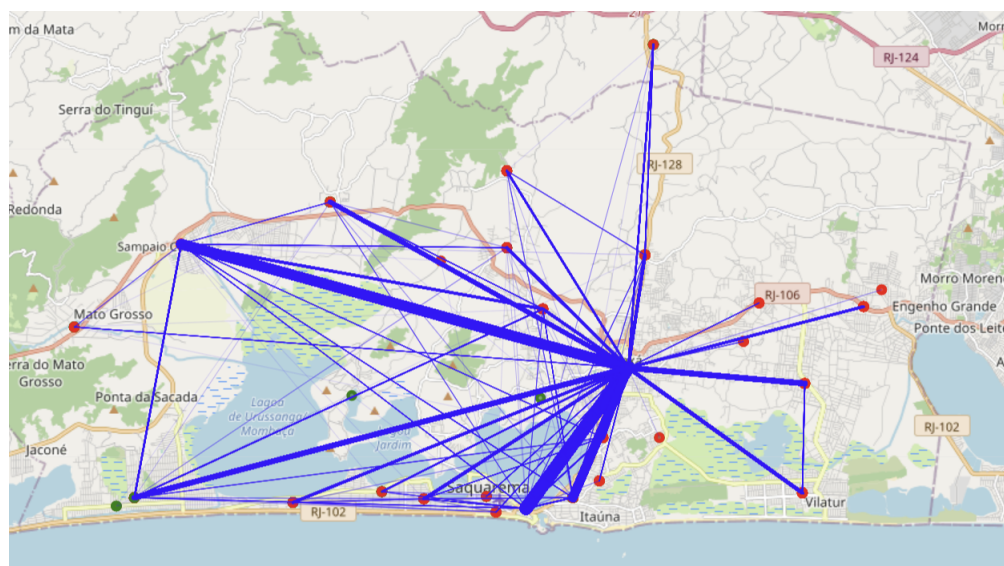
Para avaliar a precisão e a eficácia do sistema proposto, foi realizada uma análise comparativa entre a matriz O-D gerada pelo sistema proposto e a apresentada no Plano de Mobilidade Urbana de Saquarema ([SAQUAREMA, 2023](#)). O gráfico presente no Plano de Mobilidade Urbana foi construído por meio da pesquisa Sobe-Desce e com abrangência em quase todas as linhas do sistema de transporte coletivo da cidade de Saquarema-RJ. Desse modo, fornece um parâmetro confiável para comparação e validação do sistema proposto como alternativa para a construção da matriz O-D. A [Figura 54](#) apresenta a representação das 2 (duas) matrizes geradas, respectivamente, pelo Plano de Mobilidade Urbana de Saquarema e pelo sistema de *hardware* e *software* proposto.

Ao analisar as características dos dois gráficos presentes na [Figura 54](#), é possível perceber que eles apresentam semelhanças consideráveis, principalmente em relação à forma e aos principais trajetos, indicando que o sistema desenvolvido conseguiu efetivamente capturar a

Figura 54 – Comparação entre as origens e os destinos das viagens geradas pelo sistema proposto e pelo Plano de Mobilidade Urbana de Saquarema



(a) Linhas de trajeto do transporte coletivo de Saquarema-RJ presentes no Plano de Mobilidade Urbana



(b) Linhas de trajeto do transporte coletivo de Saquarema-RJ encontradas pelo sistema

dinâmica dos movimentos dos passageiros.

Essa dinâmica é evidenciada pela espessura das linhas: as mais espessas indicam um maior fluxo de passageiros, enquanto as mais finas sugerem um tráfego menor. Como pode ser visto na [Figura 54](#), a região de maior atração encontra-se na região de Bacaxá. A região de Bacaxá se destaca como o centro comercial da cidade, desempenhando um papel crucial na dinâmica do transporte público local. Como essa região possui robusta infraestrutura comercial, gera uma alta demanda por transporte. Esta demanda é atendida por várias linhas de ônibus que circulam pela região, tornando Bacaxá um importante nó na rede de transporte público de Saquarema-RJ. A presença de um fluxo constante e significativo de ônibus para e a partir de Bacaxá enfatiza sua importância como um centro de atividade. Esse fato é evidenciado na [Figura 54](#) que apresenta Bacaxá como o principal eixo da cidade.

Entretanto, é importante notar que, enquanto o gráfico da matriz O-D gerado pelo sistema proposto nesta tese apresenta várias características em comum com o estudo de mobilidade da cidade de Saquarema-RJ, nem todos os trajetos estão representados. Isso pode ser atribuído ao fato de que o sistema foi inicialmente implementado em uma amostra de 11 (onze) linhas de transporte, e não em todo o sistema. Essa observação ilustra uma oportunidade para expansão futura do sistema, ampliando sua implementação para abranger um espectro maior de linhas e, conseqüentemente, proporcionar uma representação mais completa dos trajetos de transporte coletivo.

A eficácia da abordagem proposta é evidenciada pelas semelhanças entre a matriz O-D gerada pelo sistema desenvolvido e a pesquisa de Mobilidade Urbana de Saquarema, presentes na [Figura 54](#). Apesar da implementação restrita em apenas 11 (onze) linhas de transporte, o sistema capturou a dinâmica dos movimentos dos passageiros, atestando seu potencial como ferramenta de coleta de dados eficiente e de baixo custo na construção de matrizes O-D.

Para mitigar as limitações observadas e potencializar a utilidade do sistema, algumas medidas são propostas. Inicialmente, a expansão da implementação do sistema para mais linhas de transporte seria vantajosa, ampliando a representação da dinâmica do transporte coletivo. Apesar de um custo inicial, tal medida sustenta a vantagem de ser economicamente viável em comparação com as alternativas existentes. Adicionalmente, a formação de parcerias com operadoras de transporte coletivo poderia ser promovida, uma vez que o sistema pode oferecer uma ferramenta valiosa para um planejamento e gerenciamento mais eficazes do transporte coletivo.

Os resultados apresentados destacam o potencial do sistema proposto como uma ferramenta eficiente e de baixo custo para a geração e análise de indicadores fundamentais

para o planejamento e para o gerenciamento de sistemas de transporte. A semelhança considerável entre a matriz O-D gerada pelo sistema e a presente no Plano de Mobilidade Urbana de Saquarema demonstra a habilidade do sistema em capturar a dinâmica dos movimentos dos passageiros, mesmo quando implementado em uma seleção limitada de linhas de transporte.

Ainda que algumas limitações tenham sido observadas, como a representação incompleta de trajetos devido à implantação em apenas 11 (onze) linhas de transporte, elas ilustram uma oportunidade para futura expansão do sistema. Além disso, a possibilidade de formação de parcerias com operadoras de transporte coletivo enfatiza o potencial de aplicação prática do sistema, o que reforça ainda mais a relevância e a contribuição do trabalho em questão para o campo de planejamento e gerenciamento de transporte coletivo.

5.5 Análise da Dinâmica das Viagens

Compreender a estrutura social e as características sociais e demográficas de uma cidade, é fundamental para entender as movimentações presentes no sistema de transporte coletivo. Trabalhos, como o de [Litman \(2015\)](#), destacam que a eficiência do transporte público está diretamente relacionada à densidade populacional e à distribuição de empregos e moradias. Cidades com alta densidade populacional e distribuição equilibrada tendem a ter sistemas de transporte público mais eficientes.

A utilização de algoritmos de clusterização, como a rede SOM, é essencial para analisar esse comportamento dos passageiros em sistemas de transporte coletivo. Essa abordagem permite identificar padrões de mobilidade, segmentar os passageiros com base em suas preferências de viagem e auxiliar na otimização do sistema. Utilizando algoritmos como a rede SOM e o K-Means, por exemplo, decisões estratégicas podem ser tomadas para aprimorar a eficiência e a oferta de transporte, atendendo melhor às necessidades da população.

Essa combinação de informações proporcionada pelo uso de algoritmos de clusterização e pelo conhecimento da estrutura social da cidade pode ser de grande valia no planejamento do sistema de transporte coletivo. Essas ferramentas podem fornecer *insights* importantes para a tomada de decisões e a implementação de medidas que tornem o transporte público mais eficiente, sustentável e adaptado às necessidades dos usuários, promovendo, assim, um sistema de transporte coletivo mais inclusivo e amigável ao meio ambiente. Desse modo, foi conduzido um experimento com o objetivo de analisar os padrões de deslocamento dos passageiros em um sistema de transporte público urbano.

A fim de realizar esse experimento, os dados apresentados na [Seção 5.3](#) e consolidados na [Seção 5.4](#), foram utilizados como entrada para a rede SOM. Essa consolidação dos dados

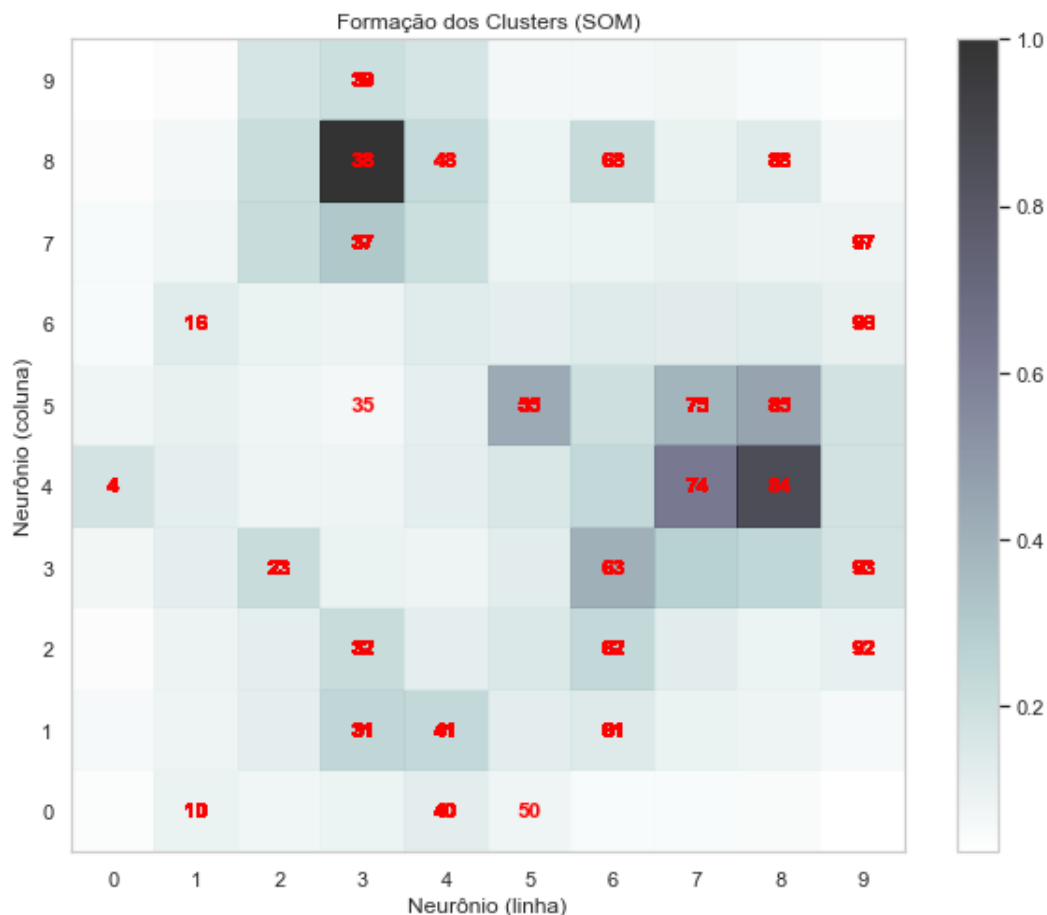
tem como objetivo fornecer informações relevantes para identificar características comuns a cada grupo de passageiros.

Em cada viagem, é associado um identificador único do passageiro (MAC), o que possibilita o rastreamento dos deslocamentos individuais ao longo do tempo. Informações como data e turno de embarque e desembarque oferecem uma visão abrangente dos padrões de viagens em diferentes momentos do dia. Além disso, as coordenadas de latitude e longitude para os pontos de embarque e desembarque permitem a análise espacial dos deslocamentos. As variáveis de distância percorrida e tempo de viagem revelam detalhes sobre a extensão e eficiência das viagens, e as consolidações e linhas específicas agrupam as viagens de acordo com rotas ou direções específicas. Assim, por meio desses dados, é possível entender o comportamento geral dos grupos de passageiros do sistema de transporte coletivo.

A fim de agrupar os passageiros com comportamentos semelhantes, foi utilizada a rede SOM. Essa técnica de análise permitiu mapear os dados multidimensionais em um *grid* de neurônios, possibilitando a identificação de padrões espaciais nas viagens. A SOM tornou possível visualizar regiões com concentração de viagens e identificar possíveis *clusters* geográficos. Diferentemente do K-Means, que agrupa os passageiros com base em suas características de viagem, a rede SOM se concentra na organização espacial dos dados, oferecendo uma abordagem complementar à análise dos padrões de mobilidade. Com essas informações, foi possível compreender melhor o comportamento dos passageiros em sistemas de transporte coletivo, contribuindo para a otimização do sistema e para o planejamento de ações estratégicas que atendam às necessidades dos usuários. A [Figura 55](#) apresenta o resultado da formação dos *clusters*. Esse resultado é visualizado em um mapa de calor, onde cada neurônio é representado por uma célula. O mapa de calor mostra a distância média entre os dados de entrada e os vetores dos neurônios associados. Os marcadores em vermelho nos neurônios indicam os *clusters* aos quais as entradas pertencem.

A rede SOM foi configurada com 10 linhas e 10 colunas, formando um *grid* de 100 neurônios. Cada neurônio possui pesos associados que representam um vetor no espaço de entrada, correspondente às características normalizadas dos dados. Durante o treinamento, a SOM ajusta esses pesos para formar regiões no espaço que refletem padrões e características nos dados de entrada. A configuração em formato de *grid* permite uma organização espacial ordenada dos neurônios, facilitando a visualização e a interpretação dos resultados. Ao mapear os dados para os neurônios vencedores mais próximos, a SOM fornece uma representação compacta e visualmente informativa dos padrões presentes nos dados, auxiliando na identificação de agrupamentos e estruturas espaciais relevantes nos comportamentos dos passageiros. A [Tabela 12](#) apresenta algumas das variáveis utilizadas no processo de

Figura 55 – Marcadores dos *clusters* em um mapa de calor



Fonte: O autor (2023)

clusterização por meio da rede SOM.

Tabela 12 – Conjunto de variáveis utilizadas pela rede SOM

mac	veiculo	data_embarque
turno_embarque	lat_embarque	long_embarque
id_ped_emb	data_desembarque	turno_desembarque
lat_desembarque	long_desembarque	id_ped_desemb
distancia_viagem_metros	tempo_viagem_segundos	consolidado
nome_diretorio	linhas	zona_emb
zona_desemb		

Após o treinamento da rede SOM, observou-se a formação de um total de 28 *clusters* distintos, cada um com características e comportamentos próprios relacionados às viagens de passageiros. É a análise das características desses *clusters* que pode auxiliar na identificação e a geração de *insights* importantes para a tomada de decisões e a implementação de medidas que tornem o transporte público mais eficiente, sustentável e adaptado.

Esses *clusters* encontrados apresentam algumas características que podem ser utilizadas

para entender o comportamento dos usuários, por exemplo:

Cluster 1 (n = 73):

- Predominância de viagens realizadas no turno da manhã (M) e no turno tarde (T).
- Duração média de viagem em torno de 1.800 segundos.
- Distância média percorrida de aproximadamente 7.500 metros.
- Concentração de locais de embarque em áreas residenciais e educacionais.

O *Cluster 1* representa um grupo de passageiros que realiza viagens tanto no turno da manhã quanto no turno da tarde, possivelmente relacionadas a atividades diárias, incluindo trabalho, estudo e outras tarefas cotidianas. A duração média e distância percorrida indicam trajetos de média extensão, sugerindo deslocamentos entre áreas residenciais e educacionais da cidade durante o período da manhã e tarde. A concentração dos locais de embarque em áreas residenciais e educacionais sugere que esse *cluster* pode ser composto por estudantes e trabalhadores que se deslocam para suas instituições de ensino e trabalho durante o dia.

Cluster 2 (n = 60):

- Predominância de viagens realizadas no turno da manhã (M) e no turno da tarde (T).
- Duração média de viagem em torno de 1.600 segundos.
- Distância média percorrida de aproximadamente 6.800 metros.
- Locais de embarque distribuídos em áreas residenciais, comerciais e educacionais.

O *Cluster 2* representa um grupo de passageiros que realiza viagens tanto no turno da manhã quanto no turno da tarde, indicando uma mobilidade diversificada para diferentes atividades durante o dia. A duração média e distância percorrida apontam para trajetos de média extensão, sugerindo deslocamentos entre diferentes áreas da cidade, incluindo regiões residenciais, comerciais e educacionais. A distribuição dos locais de embarque em áreas residenciais, comerciais e educacionais reflete a diversidade de atividades que podem ser realizadas pelos membros desse *cluster* ao longo do dia, abrangendo deslocamentos para trabalho, estudo e outras atividades cotidianas.

Cluster 6 (n = 85):

- Predominância de viagens realizadas no turno da noite (N).
- Duração média de viagem em torno de 700 segundos.
- Distância média percorrida de aproximadamente 2.000 metros.
- Localidades de embarque distribuídas em diversas áreas urbanas.

O *Cluster 6* representa um grupo de viagens noturnas com trajetos mais curtos e diversificados em relação aos outros *clusters*. Os passageiros que pertencem a este *cluster* parecem estar mais concentrados em deslocamentos noturnos, provavelmente relacionados a atividades de entretenimento e lazer. A curta duração e distância média das viagens indicam que esse grupo pode ser composto por passageiros que buscam opções de transporte mais próximas de seus locais de residência ou lazer durante a noite.

Cluster 8 (n = 85):

- Predominância de viagens realizadas no turno da tarde (T) e no turno da noite (N).
- Duração média de viagem em torno de 1.400 segundos.
- Distância média percorrida de aproximadamente 6.200 metros.
- Locais de embarque distribuídos em áreas residenciais e comerciais.

O *Cluster 8* representa um grupo de passageiros que realiza viagens tanto no turno da tarde quanto noite, sugerindo uma maior diversidade de atividades que podem incluir lazer, compras ou eventos noturnos. A duração média e distância percorrida indicam trajetos de média extensão, possivelmente relacionados a deslocamentos entre áreas residenciais e comerciais da cidade durante o período da tarde e à noite. A distribuição dos locais de embarque em áreas residenciais e comerciais sugere uma mobilidade diversificada, onde os membros desse *cluster* podem se deslocar para diferentes atividades em diferentes regiões da cidade.

Cluster 13 (n = 42):

- Predominância de viagens realizadas no turno da tarde (T) e no turno da noite (N).
- Duração média de viagem em torno de 2.000 segundos.
- Distância média percorrida de aproximadamente 7.700 metros.
- Concentração de locais de embarque em áreas residenciais e comerciais.

O **Cluster 13** se destaca por representar viagens realizadas principalmente durante o período da tarde e noite, indicando uma mobilidade diversificada para atividades diurnas e noturnas. A duração média e distância percorrida sugerem trajetos de média extensão, possivelmente relacionados a deslocamentos entre áreas residenciais e comerciais da cidade. A concentração dos locais de embarque em áreas residenciais e comerciais sugere uma mobilidade diversificada, onde os membros desse *cluster* podem se deslocar para diferentes atividades em diferentes regiões da cidade ao longo do dia e à noite.

Cluster 15 (n = 119):

- Predominância de viagens realizadas no turno da tarde (T).
- Duração média de viagem em torno de 1.900 segundos.
- Distância média percorrida de aproximadamente 8.000 metros.
- Concentração de locais de embarque em áreas residenciais e comerciais.

O *Cluster 15* se destaca por representar viagens realizadas principalmente durante o período da tarde, apontando possivelmente para um perfil de passageiros que se deslocam para atividades diárias, como trabalho ou estudo. A longa duração média e distância percorrida sugerem que esse grupo de passageiros realiza trajetos mais extensos, o que pode estar relacionado a deslocamentos entre diferentes áreas da cidade, como regiões residenciais e centros comerciais.

Cluster 18 (n = 48):

- Predominância de viagens realizadas no turno da manhã (M) e no turno da noite (N).
- Duração média de viagem em torno de 1.700 segundos.
- Distância média percorrida de aproximadamente 6.400 metros.
- Locais de embarque distribuídos em áreas residenciais, comerciais e industriais.

O *Cluster 18* representa um grupo de passageiros que realiza viagens tanto no turno da manhã quanto no turno da noite, indicando uma mobilidade diversificada para atividades diárias e noturnas. A duração média e distância percorrida sugerem trajetos de média extensão, possivelmente relacionados a deslocamentos entre áreas residenciais, comerciais e industriais da cidade. A distribuição dos locais de embarque em áreas residenciais, comerciais e industriais indica uma mobilidade diversificada, onde os membros desse

cluster podem se deslocar para diferentes atividades em diferentes regiões da cidade durante o dia e à noite.

Cluster 19 (n = 38):

- Predominância de viagens realizadas no turno da manhã (M).
- Duração média de viagem em torno de 1.300 segundos.
- Distância média percorrida de aproximadamente 5.100 metros.
- Concentração de locais de embarque em áreas residenciais e educacionais.

O *Cluster 19* representa um grupo de passageiros que realiza viagens principalmente no turno da manhã, sugerindo uma associação com deslocamentos de trabalho ou estudo durante o período matinal. A duração média e distância percorrida indicam trajetos de curta extensão, possivelmente relacionados a deslocamentos entre áreas residenciais e educacionais da cidade. A concentração dos locais de embarque em áreas residenciais e educacionais sugere que os membros desse *cluster* podem ser estudantes e trabalhadores que utilizam o transporte público para deslocamentos diários entre suas residências e suas instituições de ensino ou trabalho.

Cluster 20 (n = 67):

- Predominância de viagens realizadas no turno da manhã (M).
- Duração média de viagem em torno de 1.100 segundos.
- Distância média percorrida de aproximadamente 4.200 metros.
- Locais de embarque concentrados em áreas residenciais.

O *Cluster 20* representa um grupo de passageiros que realiza viagens principalmente no turno da manhã, sugerindo uma possível associação com deslocamentos de trabalho ou estudo no período matinal. A duração média e distância percorrida indicam trajetos de curta extensão, possivelmente relacionados a deslocamentos entre áreas residenciais da cidade. A concentração dos locais de embarque em áreas residenciais sugere que os membros desse *cluster* são moradores locais que utilizam o transporte público para deslocamentos curtos dentro de suas regiões de residência.

Cluster 21 (n = 95):

- Predominância de viagens realizadas no turno da manhã (M).
- Duração média de viagem em torno de 1.100 segundos.
- Distância média percorrida de aproximadamente 4.500 metros.
- Locais de embarque concentrados em áreas comerciais e industriais.

O *Cluster 21* representa um grupo de viagens matinais que provavelmente está relacionado a deslocamentos de trabalho ou estudo. A duração média e distância percorrida indicam que esse grupo de passageiros realiza trajetos de média extensão, principalmente entre áreas comerciais e industriais da cidade. Isso sugere que os membros desse *cluster* podem ser trabalhadores ou estudantes que realizam viagens diárias para seus locais de atividades profissionais ou educacionais.

Cluster 23 (n = 26):

- Predominância de viagens realizadas no turno da manhã (M).
- Duração média de viagem em torno de 1.400 segundos.
- Distância média percorrida de aproximadamente 5.600 metros.
- Concentração de locais de embarque em áreas residenciais e comerciais.

O *Cluster 23* se destaca por representar viagens realizadas principalmente no turno da manhã, apontando para um possível perfil de passageiros que se deslocam para atividades diárias, como trabalho ou estudo, durante o período matinal. A duração média e distância percorrida sugerem trajetos de curta a média extensão, possivelmente relacionados a deslocamentos entre áreas residenciais e comerciais da cidade. A concentração dos locais de embarque em áreas residenciais e comerciais sugere uma mobilidade diversificada, onde os membros desse *cluster* podem se deslocar para diferentes atividades em diferentes regiões da cidade durante a manhã.

Cluster 26 (n = 19):

- Predominância de viagens realizadas no turno da tarde (T).
- Duração média de viagem em torno de 1.800 segundos.
- Distância média percorrida de aproximadamente 6.900 metros.
- Concentração de locais de embarque em áreas comerciais e industriais.

O *Cluster 26* representa um grupo de passageiros que realiza viagens principalmente no turno da tarde, apontando para um possível perfil de deslocamentos para atividades diárias e comerciais durante esse período. A duração média e distância percorrida indicam trajetos de média a longa extensão, possivelmente relacionados a deslocamentos entre áreas comerciais e industriais da cidade durante a tarde. A concentração dos locais de embarque em áreas comerciais e industriais sugere que os membros desse *cluster* podem ser trabalhadores ou pessoas envolvidas em atividades comerciais, utilizando o transporte público para suas movimentações diárias.

Cluster 28 (n = 83):

- Predominância de viagens realizadas no período da manhã (M).
- Duração média de viagem em torno de 2.000 segundos.
- Distância média percorrida de aproximadamente 7.500 metros.
- Concentração de locais de embarque em áreas residenciais.

O *Cluster 28* também é caracterizado por viagens realizadas no período da manhã, sugerindo um padrão similar ao *Cluster 21*. No entanto, as viagens neste *cluster* apresentam uma duração média e distância percorrida ligeiramente maiores. Isso pode indicar que os membros desse grupo se deslocam para diferentes áreas da cidade durante a manhã, possivelmente relacionado a atividades profissionais, estudo ou outras responsabilidades diárias.

Na análise dos resultados obtidos pela rede SOM, duas métricas foram utilizadas para avaliar a qualidade dos *clusters* formados. A primeira métrica é a Silhueta, que varia de -1 a 1 e indica o grau de separação e coesão dos *clusters*. Os resultados obtidos apresentaram uma pontuação média de Silhueta de 0.39, sugerindo uma razoável separação dos grupos de viagens dos passageiros. Esse valor indica a eficácia do algoritmo na identificação de padrões distintos de deslocamento. A segunda métrica utilizada é o Erro de Quantização, que representa a média das distâncias entre os dados de entrada e os vetores dos neurônios associados aos *clusters*. Os resultados mostraram um Erro de Quantização médio de 0.19, indicando uma boa capacidade da rede SOM em representar os dados de entrada em um espaço de menor dimensionalidade. Essas métricas comprovam a adequação da SOM como uma ferramenta eficiente para agrupar os comportamentos dos passageiros e facilitar a análise das características das viagens no sistema de transporte coletivo.

O uso da rede SOM na análise do comportamento de passageiros em sistemas de transporte coletivo apresenta algumas limitações que devem ser consideradas. A escolha inadequada

do tamanho da grade da SOM pode afetar a representação dos padrões de viagem, e parâmetros mal ajustados podem impactar a qualidade dos resultados. Além disso, em conjuntos de dados de alta dimensionalidade, a interpretação dos *clusters* pode ser complexa, e o pré-processamento dos dados é crucial para evitar conclusões incorretas. O método também pode ser computacionalmente custoso para grandes conjuntos de dados. No entanto, com o devido tratamento e interpretação criteriosa dos *clusters*, a rede SOM pode ser uma ferramenta valiosa para o planejamento e gerenciamento eficiente do transporte público, especialmente em países em desenvolvimento onde a mobilidade urbana é uma questão crítica.

Desse modo, os resultados podem ser utilizados para fornecer *insights* valiosos para a gestão e operação do sistema de transporte público urbano. A análise dos padrões de deslocamento dos passageiros pode ser útil para o planejamento de rotas e horários, visando melhorar a eficiência e qualidade dos serviços oferecidos. A aplicação das técnicas de agrupamento e de visualização dos dados demonstrou-se eficaz na identificação de comportamentos de deslocamento e na revelação de informações importantes para aprimorar o transporte público e proporcionar uma experiência mais satisfatória aos usuários.

Capítulo 6

Proposta de *Framework* para o Planejamento e o Gerenciamento do Sistema de Transporte

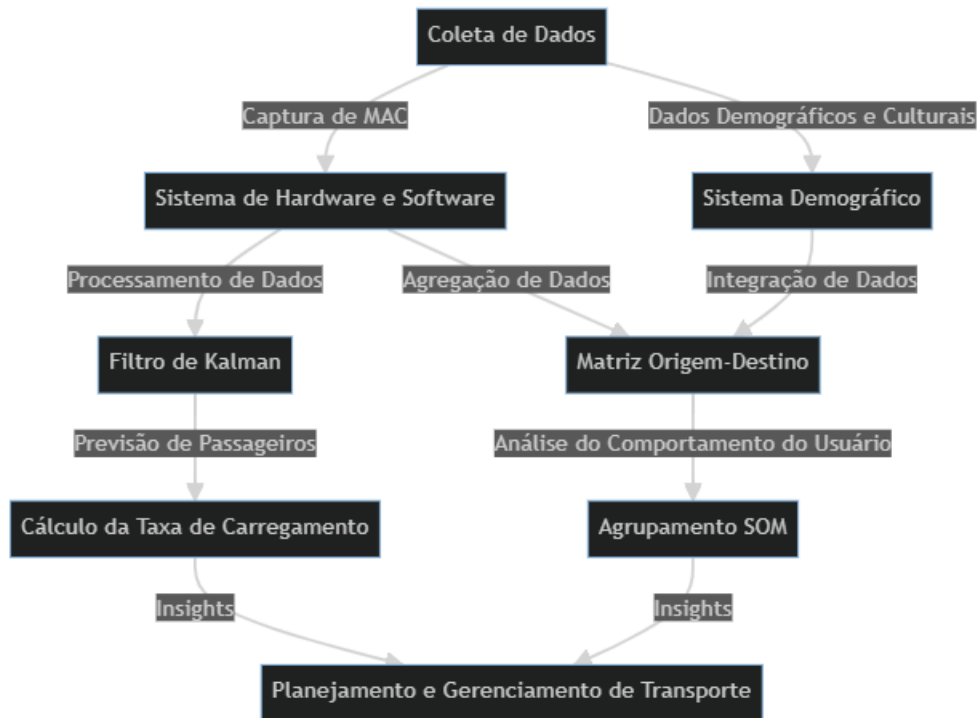
Este capítulo apresenta a proposta de um *framework* para o planejamento e o gerenciamento do sistema de transporte, uma contribuição central desta pesquisa. A [Seção 6.1](#) descreve a estrutura do *Framework*, fornecendo uma visão completa de sua composição, funcionamento e interações entre seus componentes. A [Seção 6.2](#), por sua vez, detalha as características distintivas do *framework* proposto, explicando os elementos que o tornam uma solução eficaz e adaptável para os desafios contemporâneos do transporte.

6.1 Estrutura do *Framework* Proposto

A gestão eficiente do transporte público representa um desafio significativo, particularmente em países em desenvolvimento onde a infraestrutura e os recursos podem ser escassos. Para enfrentar esses desafios, propõe-se um *framework* inovador e de baixo custo. Ele é especificamente projetado para melhorar o planejamento e o gerenciamento de sistemas de transporte público de passageiros. Este *framework* baseia-se em um sistema integrado de coleta e de análise de dados, utilizando várias fontes, para gerar *insights* valiosos que podem informar e aprimorar a tomada de decisões no contexto do transporte público, tornando-o uma solução eficiente e acessível para otimizar a mobilidade urbana em cenários de recursos limitados.

O diagrama presente na [Figura 56](#) apresenta uma visão geral do *framework* proposto, ilustrando as várias etapas do processo, desde a coleta de dados até a compreensão precisa e profunda do sistema público de transporte de passageiros. Cada componente do diagrama representa uma etapa crucial no processo, e, juntos, eles formam um sistema

Figura 56 – *Framework* proposto



Fonte: O autor (2023)

integrado que é tanto eficiente quanto de baixo custo.

Em países em desenvolvimento, os sistemas de transporte público de passageiros enfrentam desafios únicos, como infraestrutura inadequada, falta de financiamento e alta demanda (VASCONCELLOS, 2014). Além disso, a coleta e análise de dados podem ser particularmente desafiadoras devido à falta de tecnologia e recursos (CERVERO, 2001).

O *framework* começa com a **Coleta de Dados**, que inclui a captura de endereços MAC de dispositivos móveis dos passageiros, além de dados demográficos e culturais. Esses dados são coletados por meio de um **Sistema de Hardware e Software** capaz de rastrear os usuários e capturar os endereços MAC de seus dispositivos móveis.

Os dados coletados são, então, processados por meio de um **Filtro de Kalman** para prever o número de passageiros e calcular a taxa de carregamento dos ônibus. Esses dados processados fornecem *insights* cruciais para o planejamento e gerenciamento do transporte. Este tipo de previsão pode ser extremamente útil em ambientes onde a demanda por transporte público é alta e os recursos são limitados (CERVERO, 2001).

Paralelamente, os dados coletados são agregados para construir uma **Matriz Origem-Destino**. Esta matriz é alimentada pelos dados processados pelo **Sistema Demográfico**, que integra os dados demográficos e culturais coletados. A compreensão das origens

e destinos dos passageiros é crucial para o planejamento eficaz do transporte público (VASCONCELLOS, 2014).

A **Matriz Origem-Destino** é então utilizada para analisar o comportamento do usuário por meio de um processo de **Agrupamento SOM**. O resultado deste agrupamento gera um outro conjunto de *insights* que alimentam o **Planejamento e o Gerenciamento de Transporte**.

Em resumo, este *framework* permite a coleta e a análise de várias fontes de dados para gerar *insights* que podem melhorar significativamente o planejamento e o gerenciamento de sistemas de transporte público de passageiros em países em desenvolvimento. Além disso, a sua implementação de baixo custo o torna uma solução viável para esses ambientes desafiadores.

6.2 Características do *Framework* Proposto

O transporte público é um pilar essencial da mobilidade urbana em países em desenvolvimento, onde a demanda é elevada e os recursos podem ser limitados. No [Capítulo 5](#), foram apresentados os principais resultados obtidos por meio do sistema de baixo custo, capaz de utilizar diversas fontes de dados para gerar e analisar indicadores. Esses resultados são sintetizados por meio do *framework* e demonstram o potencial em informar e aprimorar os processos de tomada de decisão no contexto do transporte público. Desse modo, os objetivos específicos como a construção de um sistema de *hardware* e *software* de baixo custo, a identificação de técnicas de aprendizado de máquina para ajuste de dados e a modelagem da variação da taxa de carregamento diário para linhas de transporte público, são alcançados. A avaliação e validação dos resultados, incluindo a construção de uma matriz O-D e a previsão da taxa de ocupação dos ônibus, reforçam a eficácia do *framework* proposto.

O *framework* proposto nesta pesquisa aborda diretamente vários dos desafios e lacunas identificados na literatura sobre o planejamento do sistema de transporte por meio de pesquisas domiciliares, em contraste com os métodos tradicionais, que podem ser caros e suscetíveis a erros, principalmente quando são utilizados apenas dados de pesquisas domiciliares, como destacado por [Richardson, Ampt e Meyburg \(1995\)](#) e por [O'Flaherty \(1997\)](#). O *framework* proposto utiliza uma abordagem integrada que combina várias fontes de dados, incluindo sensores e *smartphones*. Isso minimiza o *trade-off* entre qualidade, quantidade e recursos disponíveis. Além disso, ao incorporar técnicas modernas de IA, o *framework* vai além das propostas de [Wolf et al. \(2006\)](#), oferecendo análises integradas e melhorando a precisão das pesquisas realizadas.

Além disso, esse *framework* também se alinha com as inovações sugeridas por autores

como [Stopher e Greaves \(2007\)](#) e como [Safi, Mesbah e Ferreira \(2014\)](#), que enfatizam a importância de fontes de dados mais confiáveis e a utilização de *smartphones* para coleta de dados. No entanto, o *framework* proposto supera as limitações associadas ao uso exclusivo de *smartphones*, como a resistência dos usuários em manter uma aplicação de monitoramento, conforme mencionado por [Safi, Mesbah e Ferreira \(2014\)](#). Ao combinar diferentes técnicas e tecnologias, o *framework* oferece uma solução mais robusta e versátil, capaz de fornecer *insights* valiosos para o planejamento e gestão eficazes do sistema de transporte, sem os altos custos associados às pesquisas domiciliares tradicionais, como evidenciado por [Kurauchi e Schmöcker \(2017\)](#) e por [Stopher, FitzGerald e Xu \(2007\)](#). Essa abordagem integrada e inovadora posiciona o *framework* como uma ferramenta valiosa para autoridades de transporte e formuladores de políticas em países em desenvolvimento, contribuindo para a melhoria da mobilidade e acessibilidade urbanas.

O *framework* proposto oferece uma abordagem integrada e adaptável que endereça os desafios e complexidades identificados na literatura. Ao suportar a combinação de diferentes fontes de dados, como *Big Data*, CDR, GPS e sensores inteligentes, ele permite uma análise mais profunda e precisa do comportamento dos usuários e padrões de mobilidade. Essas características são essenciais para o planejamento adequado do sistema de transporte, conforme destacado por [Iliashenko, Iliashenko e Lukyanchenko \(2021\)](#) e por [Lemondé, Arsenio e Henriques \(2021\)](#). Além disso, o *framework* considera as nuances específicas de cada região, abordando as limitações observadas em estudos de casos concentrados em países desenvolvidos, como mencionado por [Oort e Cats \(2015\)](#). Ao incorporar técnicas de AM e utilizar a qualidade, segurança e privacidade dos dados como premissa, o *framework* responde aos desafios e restrições na utilização de diferentes fontes de dados para o planejamento e operação do sistema de transporte, fornecendo uma solução robusta e eficaz.

O *framework* proposto oferece uma abordagem mais integrada para o planejamento e gestão do sistema de transporte, alinhando-se com as tendências emergentes e superando desafios identificados na literatura existente. Ele considera diferentes fontes de dados, permitindo uma análise mais completa e precisa.

Primeiramente, o *framework* leva em conta as premissas específicas associadas ao uso de cartões, como a taxa de passageiros que os utilizam e a possibilidade de utilização do cartão por outros usuários, conforme destacado por [Kurauchi e Schmöcker \(2017\)](#) e por [Li et al. \(2018\)](#). Isso evita os potenciais vieses e limitações que podem surgir em sistemas que dependem exclusivamente de dados de AFC.

Além disso, essa abordagem integrada garante uma representação mais precisa e confiável do comportamento dos passageiros e do desempenho do sistema de transporte. Isso torna o *framework* uma ferramenta poderosa, capaz de se adaptar a diferentes técnicas e fontes

de dados.

Finalmente, o *framework* também aborda desafios específicos, como os problemas de trabalhar com *Big Data*, apontados por [Torre-Bastida et al. \(2018\)](#), e a necessidade de técnicas modernas, enfatizada por [Welch e Widita \(2019\)](#). Essa capacidade de integrar e adaptar-se a diferentes abordagens reforça sua relevância e eficácia no contexto atual do sistema de transporte.

Embora o *framework* proposto represente uma abordagem promissora e inovadora para o planejamento e gestão do sistema de transporte, existem aspectos que merecem atenção. A complexidade inerente à integração de múltiplas fontes de dados e à consideração das características regionais pode exigir uma calibração e harmonização dos diferentes elementos que compõem esse *framework*. A aplicação de técnicas modernas de AM pode apresentar desafios que necessitam de recursos e conhecimentos especializados. Além disso, a escalabilidade e a adaptabilidade do *framework* a diferentes cenários podem ser vistas como áreas que requerem estudo e desenvolvimento contínuos. A limitação de não ter sido possível testar o *framework* em diferentes cenários também acrescenta uma camada de complexidade, tornando essencial uma avaliação futura para garantir sua eficácia e robustez em uma maior variedade de contextos de transporte.

Portanto, o *framework* proposto representa um avanço no planejamento e gestão do sistema de transporte, especialmente em países em desenvolvimento. Sua capacidade de integrar múltiplas fontes de dados e considerar características regionais o diferencia de abordagens tradicionais, proporcionando uma análise mais precisa do comportamento dos usuários e padrões de mobilidade. No entanto, como qualquer inovação, ele apresenta desafios inerentes que necessitam de atenção e refinamento contínuo. A necessidade de calibração, a aplicação de técnicas avançadas e a adaptabilidade em diferentes cenários são áreas que requerem estudos adicionais. Ainda assim, a proposta e os resultados indicam um caminho promissor para melhorar a mobilidade urbana, alinhando-se com as tendências emergentes e atendendo às demandas específicas de regiões em desenvolvimento.

Capítulo 7

Discussões Finais

A questão da mobilidade urbana ganha cada vez mais relevância no atual contexto de crescimento acelerado e transformação das cidades. O desafio de criar sistemas de transporte eficientes e sustentáveis - essenciais para a qualidade de vida dos residentes urbanos - torna-se uma necessidade premente para as metrópoles em desenvolvimento. Este trabalho discute essa problemática, oferecendo contribuições teóricas e práticas para o planejamento e gerenciamento adequados dos sistemas de transporte nos centros urbanos.

Nesse contexto, o sistema de transporte urbano emerge como um pilar fundamental para o desenvolvimento sustentável das cidades, pois é por meio de um sistema de mobilidade urbana de qualidade que todos os cidadãos podem ter acesso a atividades comerciais, industriais, educacionais e recreativas. Para alcançar esse desenvolvimento sustentável, é fundamental a administração adequada do trânsito urbano.

Durante o desenvolvimento deste trabalho, tornou-se evidente que o transporte coletivo, se adequadamente implementado, tem o potencial de promover uma mobilidade mais sustentável, economizando recursos energéticos, reduzindo a poluição atmosférica e melhorando a qualidade de vida da população. O *framework* desenvolvido neste trabalho é um passo significativo nessa direção, oferecendo uma abordagem inovadora e integrada para o planejamento e gestão do transporte coletivo. Desse modo, contribui para que o transporte coletivo de passageiros possa promover uma mobilidade mais sustentável e responsiva às necessidades da comunidade.

O *framework* proposto, também, evidenciou o uso de tecnologias emergentes, como *Big Data* e IA, contribui para a criação de um sistema de transporte eficiente. O uso dessas tecnologias permite a coleta e análise de dados mais precisos e confiáveis, proporcionando uma avaliação mais detalhada das necessidades de transporte e permitindo um planejamento mais eficaz.

A metodologia apresentada neste trabalho, que combina essas tecnologias com *hardware*

e *software* de baixo custo para a construção eficiente das matrizes de origem-destino em linhas do transporte coletivo, oferece uma solução viável e econômica para a questão do planejamento do transporte urbano, especialmente em países em desenvolvimento. Assim, a implementação desta metodologia pode melhorar a qualidade do transporte coletivo, reduzindo custos operacionais e tempos de viagem.

Adicionalmente, esta pesquisa pode ser utilizada como ponto de partida para que os gestores das cidades considerem a adoção de soluções tecnológicas no planejamento e no gerenciamento do transporte coletivo, já que empregar essas tecnologias, é essencial para a construção de cidades mais adequadas às necessidades dos cidadãos. Nesse sentido, em um contexto onde a qualidade de vida urbana está cada vez mais vinculada ao acesso a sistemas de transporte eficientes, a adoção de ferramentas como as apresentadas nessa tese pode ser um passo importante para a construção de cidades verdadeiramente sustentáveis.

Ao proporcionar uma análise da situação atual do transporte urbano em cidades como Belo Horizonte-MG e Saquarema-RJ e sugerir um *framework* capaz de agregar diferentes fontes de dados e desenvolver um sistema de baixo custo, este trabalho reafirma a importância da mobilidade urbana para a qualidade de vida e o desenvolvimento sustentável das cidades.

7.1 Contribuições

Esta seção é organizada para apresentar as principais contribuições e descobertas da pesquisa. As contribuições teóricas são exploradas, incluindo aspectos como a utilização de *Big Data* e a comparação com métodos tradicionais. Em seguida, são abordadas as contribuições práticas, com ênfase na implementação realizada em Saquarema-RJ e Belo Horizonte-MG. Por fim, a seção também discute as contribuições sociais e ambientais, como a melhoria da qualidade do transporte público e a preservação do meio ambiente.

7.1.1 Contribuições Teóricas

Esta tese apresenta o desenvolvimento e a implementação de um *framework* de baixo custo para aprimorar o planejamento e o gerenciamento de sistemas de transporte público, com foco em países em desenvolvimento. Por meio da utilização de diversas fontes de dados e técnicas de AM, o estudo propõe uma nova abordagem para a coleta e análise de indicadores relevantes, visando a otimização dos serviços de transporte coletivo. Após um estudo sistemático da literatura e uma comparação com os resultados obtidos por métodos tradicionais, como a pesquisa Sobe-Desce manual, as contribuições teóricas e práticas desta tese para a administração do transporte público são destacadas. Os resultados obtidos demonstram a eficácia deste *framework*.

Ao avaliar os resultados alcançados e compará-los com os trabalhos e considerações apresentadas no [Capítulo 2](#), torna-se evidente que o presente estudo gerou contribuições teóricas significativas na área de coordenação e operação do sistema público de transporte de passageiros. Essas contribuições ilustram o potencial do *framework* proposto para otimizar os sistemas de transporte público, particularmente em cenários de países em desenvolvimento.

Em primeiro lugar, observa-se que vários trabalhos citados no [Capítulo 2](#) ressaltam a importância da seleção de indicadores adequados e o potencial do *Big Data* para melhorar os sistemas de transporte inteligentes. O *framework* sugerido se alinha a essas ideias, apresentando um sistema eficiente para a geração e análise de indicadores essenciais. Mais especificamente, ao desenvolver e implementar o sistema em uma seleção limitada de linhas de transporte, foi possível demonstrar a capacidade de capturar dinâmicas de movimento de passageiros, corroborando e ampliando as descobertas presentes na literatura sobre o tema.

Ademais, a necessidade de equilibrar a qualidade e a quantidade de dados com os recursos disponíveis, foi igualmente abordada nesta pesquisa. Isso ocorre, pois foi explorada a IA como um meio de melhorar a precisão das pesquisas conduzidas, permitindo a análise eficiente dos dados coletados, contribuindo assim para o preenchimento de lacunas identificadas na literatura.

Os resultados alcançados também se relacionam com estudos que utilizaram modelos de AM para a construção de matrizes O-D. Embora esses estudos tenham identificado o investimento significativo requerido como uma barreira, a pesquisa presente nesta tese propõe um sistema de baixo custo que mantém a eficácia e a eficiência na geração de matrizes O-D, demonstrando que é possível contornar alguns dos desafios discutidos na literatura.

A implementação bem-sucedida do sistema na cidade de Saquarema-RJ tem implicações significativas para o campo de transporte público, especialmente em relação a desafios previamente identificados. Trabalhos anteriores ressaltaram a complexidade da classificação de passageiros, a necessidade de estratégias mais amplas nos países em desenvolvimento, e as dificuldades na rastreabilidade de rotas mais curtas. O *framework* proposto nesta tese responde diretamente a essas questões, oferecendo uma solução inovadora. Ele permite categorizar de maneira eficaz os usuários do transporte coletivo e rastreá-los em trajetos mais curtos, demonstrando sua aplicabilidade em contextos semelhantes.

Adicionalmente, a estrutura de baixo custo deste *framework* contribui para a sua viabilidade em cenários de recursos limitados. Este é um ponto crucial para países em desenvolvimento, onde a implementação de tecnologias caras pode ser proibitiva. Portanto, este *framework*

oferece uma solução eficiente e acessível para melhorar a gestão do transporte público nestes países.

Como apresentado no decorrer dessa tese, é possível perceber a relevância do uso de técnicas de IA no campo do ITS. Desse modo, o *framework*, ao utilizar a aprendizagem não supervisionada por meio do algoritmo SOM, está alinhado com essa tendência, demonstrando a aplicabilidade prática de técnicas de IA no planejamento e monitoramento do sistema de transporte.

Nesse mesmo sentido, o *framework* apresentado também aborda as preocupações levantadas sobre os possíveis vieses introduzidos durante o processo de coleta e análise de dados. Ao utilizar uma abordagem baseada em AM, o *framework* minimiza a introdução de vieses, proporcionando a qualidade das análises.

Assim, os resultados não apenas complementam os estudos existentes na área, como também oferecem novas perspectivas e abordagens para superar alguns dos desafios identificados na literatura atual. Tais contribuições destacam a relevância do trabalho para o avanço do campo de planejamento e gerenciamento de transporte coletivo.

7.1.2 Contribuições Práticas

Com base nos experimentos realizados em campo e nos resultados obtidos, fica evidente que o sistema desenvolvido e o *framework* proposto têm potencial para contribuir de maneira prática para a melhoria da qualidade do transporte público. Por meio da coleta e análise de dados, o sistema revelou ter a capacidade de fornecer *insights* valiosos que podem fomentar decisões sobre a alocação de recursos, planejamento de rotas e horários, e estratégias de gerenciamento de demanda.

Inicialmente, o sistema proposto demonstrou sua capacidade de estimar a taxa de carregamento em três diferentes linhas urbanas de transporte coletivo em Belo Horizonte. Apesar dos valores estimados serem geralmente inferiores aos reais, devido a fatores como a não utilização de *smartphones* ou a desativação da função Wi-Fi por alguns passageiros, a aplicação do filtro de Kalman resultou em uma melhoria significativa nas estimativas. O erro médio quadrático de estimativa para a taxa de carregamento, após a correção realizada pelo filtro de Kalman, foi de 11,16; 41,32 e 119,4 para as linhas testadas.

Em segundo lugar, a análise dos dados da taxa de carregamento permitiu identificar padrões e tendências, como picos de demanda. Esta informação é de grande utilidade para informar decisões sobre a alocação de recursos e o planejamento de rotas e horários.

Adicionalmente, a aplicação da rede SOM na análise do comportamento dos passageiros resultou em uma pontuação média de silhueta de 0,39. Este valor sugere uma razoável

separação dos grupos de viagens dos passageiros, indicando a eficácia do algoritmo na identificação de padrões distintos de deslocamento. Como complemento ao resultado anterior, o erro de quantização médio foi de 0,19, indicando uma boa capacidade da SOM em representar os dados de entrada em um espaço de menor dimensionalidade.

Além disso, a validação desse *framework* na cidade de Saquarema-RJ, realizada por meio da comparação entre as origens e destinos das viagens geradas pelo sistema proposto e pelo Plano de Mobilidade Urbana da cidade mostrou uma semelhança considerável, especialmente em relação à forma e aos principais trajetos. Isso sugere que o sistema desenvolvido conseguiu efetivamente capturar a dinâmica dos movimentos dos passageiros. Sugerindo, assim, que esse *framework* pode contribuir com o melhor planejamento de rotas e horários dos ônibus, pois ajuda a identificar os principais fluxos de passageiros na cidade, o que permite que os planejadores de transporte otimizem as rotas e horários dos ônibus e atender melhor à demanda. Ademais, o sistema proposto contribui para a melhor alocação de recursos, já que ao entender onde e quando a demanda é mais alta, os operadores de transporte público podem alocar seus recursos de maneira mais eficiente - por exemplo, colocando veículos maiores ou mais veículos em rotas de alta demanda. Também podem ser identificadas contribuições para a preservação do meio ambiente, uma vez que ao otimizar o uso do transporte público por meio da matriz O-D, pode-se reduzir a dependência de veículos particulares, contribuindo para a redução da poluição e do congestionamento urbano.

7.1.3 Contribuições Sociais e Ambientais

Com a aplicação deste *framework*, a qualidade dos serviços prestados à população pode ser aprimorada. Isso ocorre, pois os operadores de transporte público têm a oportunidade de melhorar a experiência do cliente, pois é possível gerar a redução dos tempos de viagem, utilizando um planejamento mais eficiente das rotas e horários, e a diminuição da superlotação, por meio de uma alocação mais eficaz dos recursos de transporte. Além disso, ao fornecer dados como a taxa de carregamento e a matriz O-D, o *framework* pode auxiliar na identificação de padrões e tendências de uso, o que pode informar decisões estratégicas e políticas de transporte. Portanto, a implementação deste *framework* proposto pode levar a um sistema de transporte público mais eficiente, eficaz e centrado no usuário.

Além das melhorias diretas no transporte público, o *framework* proposto também traz benefícios sociais e ambientais significativos. Ao otimizar o uso do transporte público, há uma potencial redução na dependência de veículos particulares, o que pode contribuir para a diminuição da emissão de poluentes e do congestionamento urbano. Isso não apenas melhora a qualidade do ar nas cidades, mas também promove um ambiente urbano mais saudável e habitável. Adicionalmente, ao facilitar o acesso ao transporte público, o sistema

pode promover a inclusão social, garantindo que mais pessoas, independentemente de sua localização ou condição socioeconômica, tenham acesso a oportunidades de trabalho, educação e lazer.

As contribuições ambientais do *framework* vão além da redução de emissões e congestionamento. Ao promover um transporte público mais eficiente e acessível, incentiva-se o uso de modos de transporte mais sustentáveis, o que pode levar a uma mudança cultural em relação à mobilidade urbana. Essa mudança pode fomentar uma maior conscientização sobre a importância de escolhas de transporte responsáveis e seu impacto no meio ambiente. No âmbito social, a otimização do transporte público pode facilitar o acesso a áreas urbanas mais remotas, promovendo a integração e coesão social. Isso pode resultar em uma distribuição mais equitativa de oportunidades e serviços, reduzindo disparidades sociais e contribuindo para uma sociedade mais justa e igualitária. Essas contribuições ambientais e sociais reforçam a relevância do *framework* não apenas como uma ferramenta técnica, mas como um instrumento para promover o desenvolvimento sustentável e a justiça social nas cidades.

Os resultados alcançados demonstram o potencial do sistema proposto para melhorar o planejamento e a operação do transporte público urbano. Ao fornecer dados e análises detalhadas, o sistema pode auxiliar os operadores de transporte a otimizar seus serviços, melhorar a eficiência e aprimorar a experiência do usuário. Nesse sentido, a aplicação das técnicas de agrupamento e visualização dos dados pode ser eficaz na identificação de comportamentos de deslocamento e na revelação de informações importantes para aprimorar o transporte público e proporcionar uma experiência mais satisfatória aos usuários.

Portanto, a implementação do *framework* proposto representa uma ferramenta importante no campo do transporte público, com implicações tanto em níveis técnicos quanto sociais e ambientais. Ao otimizar o planejamento e a operação do transporte público, melhora-se a eficiência e a experiência do usuário, ao mesmo tempo em que se abrem caminhos para uma mobilidade urbana mais sustentável e inclusiva. A capacidade de reduzir os tempos de viagem, diminuir a superlotação e fornecer informações relevantes para decisões estratégicas e políticas de transporte reflete uma abordagem centrada no usuário. Ademais, ao promover o uso de modos de transporte mais sustentáveis e facilitar o acesso a áreas urbanas mais remotas, o *framework* contribui para uma sociedade mais justa e igualitária e para a preservação do meio ambiente. Essa confluência de benefícios técnicos, sociais e ambientais destaca a relevância do trabalho realizado, posicionando-o como um instrumento relevante para o avanço do campo de planejamento e gerenciamento de transporte coletivo, especialmente em contextos de países em desenvolvimento.

7.2 Propostas de Continuidade

Levando em conta os desafios e as oportunidades de aprimoramento identificados ao longo deste estudo, é possível delinear algumas propostas para a continuidade da pesquisa. Estas propostas visam aprimorar o trabalho já realizado, expandir o escopo da pesquisa e explorar novas possibilidades para o desenvolvimento e aplicação do *framework* proposto.

Primeiramente, uma área de foco para trabalhos futuros poderia ser a melhoria na estimativa da taxa de carregamento. Embora o filtro de Kalman tenha contribuído para uma melhoria significativa nas estimativas, a linha 1404C ainda apresentou o maior erro de estimativas para a taxa de carregamento. Portanto, a exploração de outras técnicas de filtragem ou aprimoramento do modelo atual poderia ser uma área de pesquisa valiosa. A investigação de abordagens alternativas para a estimativa da taxa de carregamento, como a aplicação de outros algoritmos de aprendizado de máquina ou a integração de mais fontes de dados como, CDR ou bilhetagem eletrônica, poderia ajudar a melhorar a precisão das estimativas e a eficácia do sistema.

Em segundo lugar, a expansão do sistema para incluir mais linhas de transporte e cidades poderia ser considerada. Embora o sistema tenha sido testado em três linhas de transporte em Belo Horizonte e na cidade de Saquarema, a inclusão de mais linhas e cidades poderia ajudar a validar ainda mais o sistema e aprimorar sua aplicabilidade. Isso permitiria uma avaliação mais robusta do desempenho do sistema em diferentes contextos e poderia fornecer *insights* adicionais sobre como o sistema pode ser adaptado para atender a diferentes necessidades e condições de transporte.

Terceiro, a análise de autocorrelação para o valor médio da taxa de carregamento poderia ser aprofundada. Isso poderia fornecer *insights* adicionais sobre os padrões de uso do transporte público e poderia ser usado para melhorar o planejamento e a alocação de recursos. A aplicação de técnicas de análise de séries temporais mais avançadas poderia ajudar a identificar padrões e tendências subjacentes nos dados da taxa de carregamento, o que poderia informar estratégias de gerenciamento de demanda e o planejamento de serviços.

Por fim, a comparação entre as origens e destinos das viagens geradas pelo sistema e os planos de mobilidade urbana existentes poderia ser expandida. Isso poderia, também, ajudar a validar ainda mais o sistema e fornecer *insights* valiosos para o planejamento e gerenciamento do transporte público. Além disso, a análise poderia ser estendida para incluir a comparação de outros indicadores, como tempos de viagem e padrões de uso. Isso poderia fornecer uma avaliação mais abrangente da eficácia do sistema e poderia ajudar a identificar áreas onde o sistema poderia ser aprimorado para melhor atender às necessidades dos usuários e operadores de transporte.

Ademais, a implementação de um módulo de *feedback* do usuário poderia ser uma adição valiosa ao sistema. Isso permitiria que os usuários fornecessem *feedback* direto sobre sua experiência com o serviço de transporte, o que poderia ser usado para informar melhorias no sistema e no serviço de transporte. Além disso, a integração de dados de *feedback* do usuário poderia ajudar a melhorar a precisão das estimativas da taxa de carregamento e fornecer uma avaliação mais precisa da experiência do usuário.

Outra proposta para a continuidade do trabalho seria a realização de uma análise mais aprofundada dos dados coletados pelo sistema. Isso poderia incluir a aplicação de técnicas de aprendizado de máquina mais avançadas para identificar padrões e tendências nos dados, ou a realização de estudos de caso mais detalhados para explorar como o sistema pode ser usado para informar o planejamento e o gerenciamento do transporte público em diferentes contextos.

Por fim, a exploração de parcerias com operadoras de transporte público e autoridades de transporte poderia ser uma área de foco para trabalhos futuros. Isso poderia permitir a implementação do sistema em uma escala maior e fornecer uma oportunidade para avaliar seu desempenho e impacto em um ambiente operacional real. Além disso, essas parcerias poderiam fornecer uma plataforma para a troca de conhecimentos e experiências, o que poderia ajudar a informar o desenvolvimento contínuo do sistema e aprimorar sua relevância e aplicabilidade para a indústria de transporte público.

Referências

ACADEMY, D. S. 2022. Disponível em: <<https://www.deeplearningbook.com.br/>>. Citado na página 47.

ALANIS, A. Y.; ARANA-DANIEL, N.; LOPEZ-FRANCO, C. **Artificial neural networks for engineering applications**. [S.l.]: Academic Press, 2019. Citado na página 41.

ALAWADHI, S. et al. Building understanding of smart city initiatives. In: SPRINGER. **International conference on electronic government**. [S.l.], 2012. p. 40–53. Citado na página 27.

ALBINO, V.; BERARDI, U.; DANGELICO, R. M. Smart cities: Definitions, dimensions, performance, and initiatives. **Journal of urban technology**, Taylor & Francis, v. 22, n. 1, p. 3–21, 2015. Citado 2 vezes nas páginas 27 e 28.

ALEXANDER, L. et al. Origin–destination trips by purpose and time of day inferred from mobile phone data. **Transportation research part c: emerging technologies**, Elsevier, v. 58, p. 240–250, 2015. Citado na página 20.

ALVAREZ, S. **A conta do desmonte - balanço do orçamento geral da união 2021**. 2022. Disponível em: <<https://www.inesc.org.br/acontadodesmonte/>>. Citado na página 8.

ANDERSEN, J.; SUTCLIFFE, S. Intelligent transport systems (its)-an overview. **IFAC Proceedings Volumes**, Elsevier, v. 33, n. 18, p. 99–106, 2000. Citado na página 7.

ANDRESEN, S. L. John mccarthy: father of ai. **IEEE Intelligent Systems**, IEEE, v. 17, n. 5, p. 84–85, 2002. Citado na página 34.

ANSARI, S. **Building Computer Vision Applications Using Artificial Neural Networks**. [S.l.]: Springer, 2020. Citado na página 41.

ARANA-DANIEL, N.; LOPEZ-FRANCO, C.; ALANIS, A. Y. **Bio-inspired algorithms for engineering**. [S.l.]: Butterworth-Heinemann, 2018. Citado na página 37.

ARAÚJO, M. R. M. d. et al. Transporte público coletivo: discutindo acessibilidade, mobilidade e qualidade de vida. **Psicologia & Sociedade**, SciELO Brasil, v. 23, n. 3, p. 574–582, 2011. Citado na página 2.

ASAN, U.; ERCAN, S. An introduction to self-organizing maps. In: **Computational Intelligence Systems in Industrial Engineering: with Recent Theory and Applications**. [S.l.]: Springer, 2012. p. 295–315. Citado 2 vezes nas páginas 52 e 54.

AUGER, F. et al. Industrial applications of the kalman filter: A review. **IEEE Transactions on Industrial Electronics**, IEEE, v. 60, n. 12, p. 5458–5471, 2013. Citado na página 74.

BABAR, M.; ARIF, F. Real-time data processing scheme using big data analytics in internet of things based smart transportation environment. **Journal of Ambient Intelligence and Humanized Computing**, Springer, v. 10, n. 10, p. 4167–4177, 2019. Citado na página 24.

BARBOZA, M. H. C. et al. Identifying human mobility patterns in the rio de janeiro metropolitan area using call detail records. **Transportation Research Record**, v. 2675, n. 4, p. 213–221, 2021. Disponível em: <<https://doi.org/10.1177/0361198120977655>>. Citado na página 21.

BARRETO, A. et al. **Intuitive Understanding of Kalman Filtering with MATLAB®**. [S.l.]: CRC Press, 2020. Citado 2 vezes nas páginas 60 e 62.

BERLINGERIO, M. et al. **AllAboard: a system for exploring urban mobility and optimizing public transport using cellphone data**. [S.l.], 2014. Citado na página 19.

BRESLOW, N. A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship. **Biometrika**, Oxford University Press, v. 57, n. 3, p. 579–594, 1970. Citado na página 81.

CAGNEY, K. A. et al. Urban mobility and activity space. **Annual Review of Sociology**, Annual Reviews, v. 46, p. 623–648, 2020. Citado na página 1.

CAMPESATO, O. **Artificial intelligence, machine learning, and deep learning**. [S.l.]: Mercury Learning and Information, 2020. Citado 2 vezes nas páginas 34 e 35.

CAMPOS, V. B. G. Uma visão da mobilidade urbana sustentável. **Revista dos Transportes Públicos**, v. 2, n. 99-106, p. 4, 2006. Citado na página 3.

CASSOLA, F.; BURLANDO, M. Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output. **Applied energy**, Elsevier, v. 99, p. 154–166, 2012. Citado na página 57.

CEDER, A. Urban mobility and public transport: future perspectives and review. **International Journal of Urban Sciences**, Taylor & Francis, v. 25, n. 4, p. 455–479, 2021. Citado 2 vezes nas páginas 6 e 7.

CERVERO, R. Efficient urbanisation: economic performance and the shape of the metropolis. **Urban Studies**, Sage Publications Sage UK: London, England, v. 38, n. 10, p. 1651–1671, 2001. Citado na página 107.

CHO, K. et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. **arXiv preprint arXiv:1406.1078**, 2014. Citado na página 47.

CHOPDE, N. R.; NICHAT, M. Landmark based shortest path detection by using a* and haversine formula. **International Journal of Innovative Research in Computer and Communication Engineering**, v. 1, n. 2, p. 298–302, 2013. Citado na página 69.

COSTA, V. et al. Prediction of journey destination in urban public transport. In: SPRINGER. **Portuguese Conference on Artificial Intelligence**. [S.l.], 2015. p. 169–180. Citado na página 17.

COSTANZI, R. et al. Ukf-based navigation system for auvs: Online experimental validation. **IEEE Journal of Oceanic Engineering**, IEEE, v. 44, n. 3, p. 633–641, 2018. Citado na página 57.

COTTRELL, M. et al. Self-organizing maps, theory and applications. **Revista de Investigaçao Operacional**, v. 39, n. 1, p. 1–22, 2018. Citado na página 52.

DAMADI, S.; MOHARRER, G.; CHAM, M. The backpropagation algorithm for a math student. **arXiv preprint arXiv:2301.09977**, 2023. Citado 2 vezes nas páginas 38 e 39.

DING, W. et al. A passenger flow analysis method through ride behaviors on massive smart card data. In: SPRINGER. **International Conference on Collaborative Computing: Networking, Applications and Worksharing**. [S.l.], 2017. p. 374–382. Citado na página 17.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. [S.l.]: John Wiley & Sons, 1998. v. 326. Citado na página 83.

EKOWICAKSONO, I.; BUKHARI, F.; AMAN, A. Estimating origin-destination matrix of bogor city using gravity model. In: **IOP Conference Series: Earth and Environmental Science**. [S.l.: s.n.], 2016. v. 31, n. 1, p. 6–11. Citado 2 vezes nas páginas 31 e 33.

FERASSO, M.; ALNOOR, A. Artificial neural network and structural equation modeling in the future. In: **Artificial Neural Networks and Structural Equation Modeling: Marketing and Consumer Research Applications**. [S.l.]: Springer, 2022. p. 327–341. Citado na página 41.

FERRAZ, A. C. P.; TORRES, I. G. E. **Transporte público urbano**. [S.l.]: RiMa Editora, 2001. Citado 4 vezes nas páginas 2, 3, 7 e 29.

FLACH, P. **Machine learning: the art and science of algorithms that make sense of data**. [S.l.]: Cambridge university press, 2012. Citado 2 vezes nas páginas 35 e 36.

FOUSSIER, J. et al. An adaptive kalman filter approach for cardiorespiratory signal extraction and fusion of non-contacting sensors. **BMC medical informatics and decision making**, BioMed Central, v. 14, n. 1, p. 1–15, 2014. Citado na página 57.

FRENCH, R. M. The turing test: the first 50 years. **Trends in cognitive sciences**, Elsevier, v. 4, n. 3, p. 115–122, 2000. Citado na página 34.

FREUDIGER, J. How talkative is your mobile device? an experimental study of wi-fi probe requests. In: **Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks**. [S.l.: s.n.], 2015. p. 1–6. Citado 2 vezes nas páginas 64 e 67.

GALLO, M.; MARINELLI, M. Sustainable mobility: A review of possible actions and policies. **Sustainability**, MDPI, v. 12, n. 18, p. 7499, 2020. Citado na página 1.

GOLDING, J. Best practices and methodology for od-matrix creation from cdr-data. **NLAB, University of Nottingham**, 2018. Citado na página 21.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. (Adaptive Computation and Machine Learning series). ISBN 9780262035613. Disponível em: <<https://books.google.com.br/books?id=Np9SDQAAQBAJ>>. Citado 8 vezes nas páginas 41, 42, 43, 44, 47, 49, 50 e 51.

GRIFFIN, G. P. et al. Mitigating bias in big data for transportation. **Journal of Big Data Analytics in Transportation**, Springer, v. 2, n. 1, p. 49–59, 2020. Citado na página 25.

GUERRA, A. L. Determinação de matriz origem/destino utilizando dados do sistema de bilhetagem eletrônica. Universidade Federal de Minas Gerais, 2011. Citado 3 vezes nas páginas 31, 32 e 33.

HERRERA-QUINTERO, L. F. et al. Smart its sensor for the transportation planning using the iot and bigdata approaches to produce its cloud services. In: **2016 8th Euro American Conference on Telematics and Information Systems (EATIS)**. [S.l.: s.n.], 2016. p. 1–7. Citado na página 23.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 45.

ILIASHENKO, O.; ILIASHENKO, V.; LUKYANCHENKO, E. Big data in transport modelling and planning. **Transportation Research Procedia**, v. 54, p. 900–908, 2021. ISSN 2352-1465. International Scientific Siberian Transport Forum - TransSiberia 2020. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352146521003240>>. Citado 2 vezes nas páginas 25 e 109.

Instituto Brasileiro de Geografia e Estatística. **Panorama Censo 2022 Saquarema (RJ)**. Rio de Janeiro, 2022. Disponível em: <<https://cidades.ibge.gov.br/brasil/rj/saquarema/panorama>>. Acesso em: 26 out. 2023. Citado na página 71.

IQBAL, M. S. et al. Development of origin–destination matrices using mobile phone call data. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 40, p. 63–74, 2014. Citado na página 20.

JAMAL, E. et al. Investigating factors affecting the mode choices of commuters in kuwait city & surrounding urban areas: Strategies for a higher quality and more sustainable public transport system. **Urban Development Issues**, Sciendo, v. 63, n. 1, p. 65–76, 2019. Citado na página 18.

JAN, B. et al. Designing a smart transportation system: An internet of things and big data approach. **IEEE Wireless Communications**, IEEE, v. 26, n. 4, p. 73–79, 2019. Citado 2 vezes nas páginas 23 e 24.

JÄRV, O.; AHAS, R.; WITLOX, F. Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 38, p. 122–135, 2014. Citado na página 19.

JIANG, S.; FERREIRA, J.; GONZALEZ, M. C. Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. **IEEE Transactions on Big Data**, IEEE, v. 3, n. 2, p. 208–219, 2017. Citado na página 20.

JONES, T. 2019. Disponível em: <<http://www.mobilize.org.br/estudos/389/brasil-e-reino-unido--mobilidade-urbana-saudavel.html>>. Citado na página 2.

KAFFASH, S.; NGUYEN, A. T.; ZHU, J. Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis. **International Journal of Production Economics**, Elsevier, v. 231, p. 107868, 2020. Citado na página 25.

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Transactions of the ASME–Journal of Basic Engineering**, v. 82, n. Series D, p. 35–45, 1960. Citado na página 55.

KASNECI, E. et al. Chatgpt for good? on opportunities and challenges of large language models for education. **Learning and Individual Differences**, Elsevier, v. 103, p. 102274, 2023. Citado na página 41.

- KASPER, R.; SCHMIDT, S. Sensor-data-fusion for an autonomous vehicle using a kalman-filter. In: IEEE. **2008 6th International Symposium on Intelligent Systems and Informatics**. [S.l.], 2008. p. 1–5. Citado na página 57.
- KHATOUN, R.; ZEADALLY, S. Smart cities: concepts, architectures, research opportunities. **Communications of the ACM**, ACM New York, NY, USA, v. 59, n. 8, p. 46–57, 2016. Citado na página 28.
- KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, IEEE, v. 78, n. 9, p. 1464–1480, 1990. Citado na página 50.
- KOHONEN, T. **Self-organizing maps**. [S.l.]: Springer Science & Business Media, 2012. v. 30. Citado na página 76.
- KOHONEN, T.; KOHONEN, T. Learning vector quantization. **Self-organizing maps**, Springer, p. 245–261, 2001. Citado na página 76.
- KURAUCHI, F.; SCHMÖCKER, J.-D. **Public transport planning with smart card data**. [S.l.]: CRC Press, 2017. Citado 5 vezes nas páginas 14, 16, 18, 29 e 109.
- KURITA, T. Principal component analysis (pca). **Computer Vision: A Reference Guide**, Springer, p. 1–4, 2019. Citado na página 52.
- LATHIA, N. et al. Individuals among commuters: Building personalised transport information services from fare collection systems. **Pervasive and Mobile Computing**, Elsevier, v. 9, n. 5, p. 643–664, 2013. Citado na página 16.
- LAZAROIU, G. C.; ROSCIA, M. Definition methodology for the smart cities model. **Energy**, Elsevier, v. 47, n. 1, p. 326–332, 2012. Citado na página 28.
- LEMONDE, C.; ARSENIO, E.; HENRIQUES, R. Integrative analysis of multimodal traffic data: addressing open challenges using big data analytics in the city of lisbon. **European Transport Research Review**, v. 13, n. 1, p. 64, Dec 2021. ISSN 1866-8887. Disponível em: <<https://doi.org/10.1186/s12544-021-00520-3>>. Citado 2 vezes nas páginas 25 e 109.
- LI, Q. et al. Kalman filter and its application. In: IEEE. **2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)**. [S.l.], 2015. p. 74–77. Citado na página 57.
- LI, T. et al. Smart card data mining of public transport destination: A literature review. **Information**, Multidisciplinary Digital Publishing Institute, v. 9, n. 1, p. 18, 2018. Citado 2 vezes nas páginas 18 e 109.
- LI, Y.; MANOHARAN, S. A performance comparison of sql and nosql databases. In: IEEE. **2013 IEEE Pacific Rim conference on communications, computers and signal processing (PACRIM)**. [S.l.], 2013. p. 15–19. Citado na página 68.
- LITMAN, T. Developing indicators for comprehensive and sustainable transport planning. **Transportation Research Record**, SAGE Publications Sage CA: Los Angeles, CA, v. 2017, n. 1, p. 10–15, 2007. Citado na página 12.
- LITMAN, T. **Evaluating public transit benefits and costs**. [S.l.]: Victoria transport policy institute Victoria, BC, Canada, 2015. Citado na página 96.

LITMAN, T. M. Evaluating transportation equity: Guidance for incorporating distributional impacts in transport planning. **Institute of Transportation Engineers. ITE Journal**, Institute of Transportation Engineers, v. 92, n. 4, p. 43–49, 2022. Citado 2 vezes nas páginas 14 e 15.

LV, Y. et al. Traffic flow prediction with big data: a deep learning approach. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 16, n. 2, p. 865–873, 2014. Citado na página 22.

LWIN, K. K.; SEKIMOTO, Y. et al. Identification of various transport modes and rail transit behaviors from mobile cdr data: A case of yangon city. **Asian Transport Studies**, Elsevier, v. 6, p. 100025, 2020. Citado na página 20.

MA, X.; WANG, Y. Development of a data-driven platform for transit performance measures using smart card and gps data. **Journal of Transportation Engineering**, American Society of Civil Engineers, v. 140, n. 12, p. 04014063, 2014. Citado na página 16.

MACEDO, J. Cidades espraiadas: porque o modelo americano deve ser evitado no brasil. In: **Desafios Urbanos: Para a Sustentabilidade Ambiental nas Cidades Brasileiras**. [S.l.]: ProUrb/UFRJ, 2012. p. 34–49. Citado na página 3.

MILLIDGE, B. et al. Neural kalman filtering. **arXiv preprint arXiv:2102.10021**, 2021. Citado na página 55.

MISHRA, K.; AITHAL, P. et al. Planning assessment of transport system: A case from nepal. **International Journal of Applied Engineering and Management Letters (IJAEML)**, v. 6, n. 1, p. 280–300, 2022. Citado na página 15.

MOREIRA, S. **Rede Neural perceptron Multicamadas**. Ensina.AI, 2018. Disponível em: <<https://medium.com/ensina-ai/rede-neural-perceptron-multicamadas-f9de8471f1a9>>. Citado na página 38.

NEILSON, A. et al. Systematic review of the literature on big data in the transportation domain: Concepts and applications. **Big Data Research**, Elsevier, v. 17, p. 35–44, 2019. Citado 2 vezes nas páginas 6 e 12.

NOBRE, J. et al. Segurança da informação para internet das coisas (iot): uma abordagem sobre a lei geral de proteção de dados (lgpd). **Revista Eletrônica de Iniciação Científica em Computação**, v. 17, n. 4, 2019. Citado na página 67.

O'FLAHERTY, C. A. **Transport planning and traffic engineering**. [S.l.]: CRC Press, 1997. Citado 2 vezes nas páginas 13 e 108.

OORT, N. V.; CATS, O. Improving public transport decision making, planning and operations by using big data: Cases from sweden and the netherlands. In: **2015 IEEE 18th International Conference on Intelligent Transportation Systems**. [S.l.: s.n.], 2015. p. 19–24. Citado 2 vezes nas páginas 22 e 109.

PATNAIK, S.; KONAR, A.; MANDAL, A. K. Building 3-d visual perception of a mobile robot employing extended kalman filter. **Journal of Intelligent and Robotic Systems**, Springer, v. 34, n. 1, p. 99–120, 2002. Citado na página 57.

PETERSON, A. **The origin-destination matrix estimation problem: analysis and computations**. Tese (Doutorado) — Institutionen för teknik och naturvetenskap, 2007. Citado na página 31.

PETTIT, C. J. et al. Big data: The engine to future cities—a reflective case study in urban transport. **Sustainability**, v. 14, n. 3, 2022. ISSN 2071-1050. Disponível em: <<https://www.mdpi.com/2071-1050/14/3/1727>>. Citado na página 26.

PINHEIRO, A. C.; FRISCHTAK, C. **Mobilidade urbana: desafios e perspectivas para as cidades brasileiras**. [S.l.]: Elsevier, 2015. Citado 2 vezes nas páginas 1 e 4.

PROSPERI, T. et al. Traffic zones discretization and origin-destination matrix estimation by means of transport demand and satellite data fusion. In: IEEE. **2021 IEEE International Intelligent Transportation Systems Conference (ITSC)**. [S.l.], 2021. p. 3217–3222. Citado na página 33.

RATHORE, M. M. et al. Urban planning and building smart cities based on the internet of things using big data analytics. **Computer Networks**, Elsevier, v. 101, p. 63–80, 2016. Citado na página 23.

RAZALI, N. M.; WAH, Y. B. et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. **Journal of statistical modeling and analytics**, v. 2, n. 1, p. 21–33, 2011. Citado na página 81.

RIBEIRO, R. G. **Estudo dos Deslocamentos Urbanos da Classe Média Brasileira na Região Metropolitana de Belo Horizonte**. Tese (Doutorado) — Doctorate Thesis, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2015. Citado na página 5.

RICHARDSON, A. J.; AMPT, E. S.; MEYBURG, A. H. **Survey methods for transport planning**. [S.l.]: Eucalyptus Press Melbourne, 1995. Citado 2 vezes nas páginas 13 e 108.

RODRIGUES, S.; SILVA, O. da; BANDEIRA, R. Pesquisa domiciliar de origem-destino: Proposta de metodologia de aquisição de dados com baixo custo. In: **9 Congresso Luso-Brasileiro para o Planejamento Urbano, Regional, Integrado e Sustentável (PLURIS 2021) Pequenas cidades, grandes desafios, muitas oportunidades**. [S.l.: s.n.], 2021. Citado na página 13.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach, 4th, Global ed.** [S.l.]: Prentice Hall, 2022. Citado na página 34.

SAFI, H.; MESBAH, M.; FERREIRA, L. Smartphone-assisted travel surveys: a smart way for transport planning. In: **Sydney, Australia: CAITR (Conference of Australian Institutes of Transport Research)**. [S.l.: s.n.], 2014. Citado 2 vezes nas páginas 14 e 109.

SAQUAREMA, P. M. **Plano de Mobilidade urbana - prefeitura de saquarema**. Prefeitura Municipal de Saquarema, 2023. Disponível em: <<https://saquarema.rj.gov.br/plano-de-mobilidade-urbana/>>. Citado 5 vezes nas páginas 85, 88, 90, 91 e 93.

SCHIPPER, F.; EMANUEL, M.; OLDENZIEL, R. Sustainable urban mobility in the present, past, and future. **Technology and culture**, Johns Hopkins University Press, v. 61, n. 1, p. 307–317, 2020. Citado na página 1.

SCHMITT, T. et al. Restoration of nuclear medicine images using a kalman filtering approach. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Image and Video Processing III**. [S.l.], 1995. v. 2421, p. 64–69. Citado na página 57.

SEGUNDO, F. R.; SILVA, E. S. e; FARINES, J.-M. A dtn routing strategy based on neural networks for urban bus transportation system. **Journal of Network and Computer Applications**, Elsevier, v. 64, p. 216–228, 2016. Citado na página 41.

SELIM, S. Z.; ISMAIL, M. A. K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, n. 1, p. 81–87, 1984. Citado na página 51.

SHAHAPURE, K. R.; NICHOLAS, C. Cluster quality analysis using silhouette score. In: IEEE. **2020 IEEE 7th international conference on data science and advanced analytics (DSAA)**. [S.l.], 2020. p. 747–748. Citado na página 76.

SHAHID, N.; RAPPON, T.; BERTA, W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. **PloS one**, Public Library of Science San Francisco, CA USA, v. 14, n. 2, p. e0212356, 2019. Citado na página 41.

SHUKLA, S.; BALACHANDRAN, K.; SUMITHA, V. A framework for smart transportation using big data. In: IEEE. **2016 International Conference on ICT in Business Industry & Government (ICTBIG)**. [S.l.], 2016. p. 1–3. Citado na página 23.

STOPHER, P.; FITZGERALD, C.; XU, M. Assessing the accuracy of the sydney household travel survey with gps. **Transportation**, Springer, v. 34, n. 6, p. 723–741, 2007. Citado 2 vezes nas páginas 14 e 109.

STOPHER, P. R.; GREAVES, S. P. Household travel surveys: Where are we going? **Transportation Research Part A: Policy and Practice**, Elsevier, v. 41, n. 5, p. 367–381, 2007. Citado 2 vezes nas páginas 14 e 109.

STOPHER, P. R.; JONES, P. **Transport survey quality and innovation**. [S.l.]: pergamon New York, NY, 2003. Citado na página 13.

SUN, L. et al. An integrated bayesian approach for passenger flow assignment in metro networks. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 52, p. 116–131, 2015. Citado na página 17.

TIMMS, P. A review of methods for estimating origin-destination trip matrices using link counts. Institute of Transport Studies, University of Leeds, 2001. Citado na página 32.

TKÁČ, M.; VERNER, R. Artificial neural networks in business: Two decades of research. **Applied Soft Computing**, Elsevier, v. 38, p. 788–804, 2016. Citado na página 41.

TOOLE, J. L. et al. The path most traveled: Travel demand estimation using big data resources. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 58, p. 162–177, 2015. Citado na página 23.

TOQUÉ, F. et al. Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks. In: IEEE. **2016 IEEE 19th international conference on intelligent transportation systems (ITSC)**. [S.l.], 2016. p. 1071–1076. Citado na página 17.

TORRE-BASTIDA, A. I. et al. Big data for transportation and mobility: recent advances, trends and challenges. **IET Intelligent Transport Systems**, IET, v. 12, n. 8, p. 742–755, 2018. Citado 2 vezes nas páginas 23 e 110.

TRANSPORTE, C.-C. do. Ntu associação nacional das empresas de transporte urbanos. **Pesquisa mobilidade da população urbana**. Brasília, 2017. Citado 2 vezes nas páginas 9 e 10.

TRANSPORTE, C. N. do. **Pesquisa Mobilidade da População Urbana 2017**. [S.l.], 2017. Citado na página 2.

TSUMURA, Y. et al. Examining potentials and practical constraints of mobile phone data for improving transport planning in developing countries. **Asian Transport Studies**, v. 8, p. 100043, 2022. ISSN 2185-5560. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2185556021000110>>. Citado na página 21.

VASCONCELLOS, E. A. **Urban Transport Environment and Equity: The case for developing countries**. [S.l.]: Routledge, 2014. Citado 2 vezes nas páginas 107 e 108.

VIEIRA, T. et al. Improving the management of public transport through modeling and forecasting passenger occupancy rate. **International Conference on Time Series**, 2019. Citado na página 8.

VIEIRA, T. et al. Public transport occupancy estimation using wlan probing and mathematical modeling. **Transportation Research Procedia**, Elsevier, v. 48, p. 3299–3309, 2020. Citado 5 vezes nas páginas 6, 8, 55, 66 e 70.

VIEIRA, T. P. et al. Use of computational intelligence for scheduling of pumps in water distribution systems: a comparison between optimization algorithms. In: IEEE. **2018 IEEE Congress on Evolutionary Computation (CEC)**. [S.l.], 2018. p. 1–8. Citado na página 8.

WELCH, T. F.; WIDITA, A. Big data in public transportation: a review of sources and methods. **Transport reviews**, Taylor & Francis, v. 39, n. 6, p. 795–818, 2019. Citado 5 vezes nas páginas 5, 6, 12, 24 e 110.

WEY, W.-M. Constructing urban dynamic transportation planning strategies for improving quality of life and urban sustainability under emerging growth management principles. **Sustainable Cities and Society**, Elsevier, v. 44, p. 275–290, 2019. Citado na página 5.

WINDER, I. **Computational Ants: Agent Based Visualization Technique with CDR OD Matrix**. Interactive Simulation, 2016. Disponível em: <<https://ira.mit.edu/blog/agent-based-visualization>>. Citado na página 32.

WOLF, J. et al. Applications of new technologies in travel surveys. **Travel Survey Methods- Quality and Future Directions**, Elsevier Oxford, p. 531–544, 2006. Citado 2 vezes nas páginas 14 e 108.

YAN, S. **Understanding LSTM and its diagrams**. 2016. <<https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714>>. (Accessed on 04/09/2023). Citado na página 46.

YANG, H.; IIDA, Y.; SASAKI, T. The equilibrium-based origin-destination matrix estimation problem. **Transportation Research Part B: Methodological**, Elsevier, v. 28, n. 1, p. 23–33, 1994. Citado na página 33.

YUSOF, K.; MAN, M. Efficiency of flat file database approach in data storage and data extraction for big data. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 9, n. 2, p. 460–473, 2018. Citado na página 68.

ZARGARI, S. A.; MEMARNEJAD, A.; MIRZAHOSSEIN, H. Hourly origin–destination matrix estimation using intelligent transportation systems data and deep learning. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 21, n. 21, p. 7080, 2021. Citado 2 vezes nas páginas 31 e 33.

ZHAO, G.; ZHU, G. **Big Data Transportation Systems**. World Scientific, 2021. ISBN 9789811235993. Disponível em: <<https://books.google.com.br/books?id=JGsvzgEACAAJ>>. Citado na página 6.

ZHAO, J. et al. Understanding temporal and spatial travel patterns of individual passengers by mining smart card data. In: IEEE. **17th International IEEE Conference on Intelligent Transportation Systems (ITSC)**. [S.l.], 2014. p. 2991–2997. Citado na página 16.

ZHAO, Z.; KOUTSOPOULOS, H. N.; ZHAO, J. Identifying hidden visits from sparse call detail record data. **Transactions in Urban Data, Science, and Technology**, v. 1, n. 3-4, p. 121–141, 2022. Disponível em: <<https://doi.org/10.1177/27541231221124164>>. Citado na página 21.

ZHU, L. et al. Big data analytics in intelligent transportation systems: A survey. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 20, n. 1, p. 383–398, 2018. Citado na página 7.

ZIN, T. A. et al. Estimation of originating-destination trips in yangon by using big data source. **Journal of Disaster Research**, Fuji Technology Press Ltd., v. 13, n. 1, p. 6–13, 2018. Citado na página 24.