



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

# **SISTEMAS *Fuzzy* EVOLUTIVOS NA PREVISÃO E NA CLASSIFICAÇÃO DA DEMANDA DE PASSAGEIROS DE TÁXI**

**LUÍS FERNANDO LINHARES**

Orientador: Prof. Alisson Marques da Silva  
CEFET-MG

BELO HORIZONTE  
MARÇO DE 2022

**LUÍS FERNANDO LINHARES**

**SISTEMAS *Fuzzy* EVOLUTIVOS NA PREVISÃO E NA  
CLASSIFICAÇÃO DA DEMANDA DE PASSAGEIROS DE  
TÁXI**

Dissertação de Mestrado 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 Mestre em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Sistemas Inteligentes

Orientador: Prof. Alisson Marques da Silva  
CEFET-MG

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL  
BELO HORIZONTE  
MARÇO DE 2022

L755s Linhares, Luís Fernando  
Sistemas Fuzzy evolutivos na previsão e na classificação da demanda de passageiros de táxi / Luís Fernando Linhares. – 2022.  
88 f.

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.

Orientador: Alisson Marques da Silva.

Dissertação (mestrado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Sistemas difusos – Teses. 2. Computação evolutiva – Teses. 3. Teoria da previsão – Teses. 4. Táxis – Teses. 5. Transporte urbano – Modelos matemáticos – Teses. I. Silva, Alisson Marques da. II. Centro Federal de Educação Tecnológica de Minas Gerais. III. Título.

CDD 006.3



SERVIÇO PÚBLICO FEDERAL  
MINISTÉRIO DA EDUCAÇÃO  
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS  
COORDENAÇÃO DO CURSO DE MESTRADO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

**“SISTEMAS FUZZY EVOLUTIVOS NA PREVISÃO E NA  
CLASSIFICAÇÃO DA DEMANDA DE PASSAGEIROS DE TÁXI”**

Dissertação de Mestrado apresentada por **Luís Fernando Linhares**, em 04 de abril de 2022,  
ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-  
MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Alisson Marques da Silva  
Centro Federal de Educação Tecnológica de Minas Gerais

Prof.<sup>a</sup>. Dr.<sup>a</sup>. Magali Rezende Gouvêa Meireles  
Pontifícia Universidade Católica de Minas Gerais

Prof. Dr. Eduardo Habib Bechelane Maia  
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Gustavo Campos Menezes  
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida a impressão,

Prof.<sup>a</sup>. Dr.<sup>a</sup>. Elizabeth Fialho Wanner  
Presidenta do Colegiado do Programa de Pós-Graduação em  
Modelagem Matemática e Computacional

Dedico este trabalho aos meus pais, por todo o amor, carinho e dedicação em minha criação. Por permitir que eu tivesse acesso a uma educação de qualidade, e ensinar que o conhecimento é um dos bens mais preciosos que se pode ter.

# Agradecimentos

Em primeiro lugar, agradeço a Deus pela vida e por permitir que eu pudesse concluir mais esta etapa em minha caminhada. Uma das poucas certezas desta vida é que, tendo fé Nele, nunca estarei desamparado.

Agradeço aos meus queridos pais, Marilene e Geraldo, por toda a dedicação e amor que dispenderam em minha criação, pelos meus valores morais, e não obstante, por priorizar seus recursos em minha formação, permitindo que eu tivesse acesso à uma boa educação.

Ao meu irmão, Fabiano, por estar sempre ao meu lado, pelas palavras de conforto em momentos difíceis e pelas valiosas discussões sobre este trabalho.

À minha querida esposa Sara, pelo, amor, paciência, compreensão, carinho e companheirismo. Saber que você está ao meu lado me faz ter forças para continuar a perseguir objetivos maiores.

Ao meu orientador, Alisson, por todo o cuidado e zelo na orientação deste trabalho, pela paciência, pelas valiosas discussões e contribuições, e não menos importante, por acreditar em mim, quando muitas vezes eu mesmo não conseguia fazê-lo.

Ao professor Paulo Almeida (*in memoriam*), pelas valiosas contribuições nas etapas iniciais do projeto.

Aos colegas do CEFET-MG, por todo o apoio e pelas palavras de incentivo.

Aos colegas do MMC, por compartilhar momentos de alegria e tensão no decorrer do mestrado.

Aos amigos Paulo e Cristiano, pelas inúmeras discussões e valiosas contribuições para este trabalho.

A todos que de alguma forma me apoiaram e estiveram ao meu lado, permitindo que eu finalizasse mais esta etapa da minha vida. A vocês, minha gratidão!

*“Se quiser triunfar na vida, faça da perseverança a sua melhor amiga; da experiência, o seu conselheiro; da prudência, o seu irmão mais velho; e da esperança, o seu anjo da guarda.” (Joseph Addison)*

# Resumo

Utilizar transportes privados tem se tornado uma opção viável e cada vez mais popular como alternativa para transportes urbanos. Com este crescimento, um problema antigo e recorrente torna-se mais latente, que é a relação entre demandas de passageiros e oferta de táxis. Enquanto taxistas gastam considerável tempo em busca de novos passageiros, em contrapartida, têm-se passageiros que não conseguem encontrar taxistas. Com este desequilíbrio, os taxistas acabam gastando mais com combustível e desgastando o veículo devido à procura excessiva por passageiros e passageiros acabam não conseguindo táxis, ocasionando, assim, uma descrença na disponibilidade do serviço prestado. Este problema sugere a criação e a utilização de técnicas que possibilitem diminuir a lacuna entre a procura de passageiros por táxis e o contingente efetivo de veículos necessários para suprir essa demanda. Neste contexto, este trabalho apresenta uma nova abordagem para realizar a previsão e a classificação da demanda de passageiros de táxi. A abordagem proposta utiliza dados históricos de corridas de táxis e dados meteorológicos. O método de *Kruskal-Wallis* é empregado para identificar as variáveis mais relevantes e um sistema *fuzzy* evolutivo para realizar a previsão/classificação da demanda. Neste trabalho, são avaliados três sistemas evolutivos: *Autonomous Learning Multi-Model (ALMMo)*; *evolving Gaussian Multivariable Fuzzy System (eMG)* e; *evolving Neo-Fuzzy Neuron (eNFN)*. Experimentos computacionais foram realizados para comparar o desempenho da abordagem proposta. Os resultados experimentais indicam que a abordagem proposta obteve resultados comparáveis ou superiores ao estado da arte. O desempenho obtido sugere, portanto, que a abordagem proposta é promissora como alternativa para a previsão e classificação da demanda de passageiros.

**Palavras-chave:** Previsão, Classificação, Demanda de Táxi, Sistemas Evolutivos, Sistemas *Fuzzy*.

# Abstract

Private transport has become a viable and increasingly popular option as an alternative to urban transportation. With this growth, an old and recurring problem becomes more latent: the relationship between passenger demands and taxi supply. While taxi drivers spend considerable time looking for new passengers, on the other hand, some passengers cannot find taxi drivers. With this imbalance, taxi drivers spend more on fuel and wear out the vehicle due to excessive demand by passengers. As a result, passengers end up not getting taxis, thus causing disbelief in the service availability. This problem suggests the creation and use of techniques that make it possible to reduce the gap between the demand for taxi passengers and the effective contingent of vehicles needed to meet this demand. In this context, this work presents a new approach to forecasting and classifying taxi passenger demand. The proposed approach uses historical data from taxi rides and meteorological data. The Kruskal-Wallis method is used to identify the most relevant variables and an evolving fuzzy system to perform demand forecasting/classification. In this work, three evolving systems are evaluated: Autonomous Learning Multi-Model (ALMMo); evolving Multivariable Gaussian Fuzzy System (eMG), and; evolving Neo-Fuzzy Neuron (eNFN). Computational experiments were performed to evaluate and compare the performance of the proposed approach. The results showed that the proposed approach obtained results comparable or superior to those of state-of-the-art. The performance obtained by the proposed approach suggest, therefore, that the proposed approach is promising as an alternative for forecasting and classification of passenger demand.

**Keywords:** Prediction, Classification, Taxi Demand, Evolving Systems, Fuzzy Systems.

# Lista de Figuras

Figura 1 – Exemplos de aplicações em ITS. . . . .	7
Figura 2 – Demandas de <i>Pick-Up</i> e <i>Drop-Off</i> em determinada região e horário. . . . .	10
Figura 3 – Exemplos de divisões de zonas: (A) divisão retangular na cidade de Mianyang (China); (B) divisão baseada em diagramas de Voronoi e; (C) divisão em zonas irregulares na cidade de Nova Iorque (Estados Unidos da América). . . . .	13
Figura 4 – Exemplo de mapa de calor de solicitações de táxi em Nova Iorque. . . . .	14
Figura 5 – Nuvens de dados e seus pontos focais. . . . .	17
Figura 6 – Representação de três <i>clusters</i> no eMG. . . . .	20
Figura 7 – Funções de pertinência triangulares e complementares. . . . .	22
Figura 8 – Etapas da abordagem proposta para a previsão da demanda. . . . .	31
Figura 9 – Representação das coordenadas de corridas dentro e fora dos limites de uma determinada zona. . . . .	32
Figura 10 – Etapas da seleção de características. . . . .	35
Figura 11 – Processo para previsão da demanda. . . . .	36
Figura 12 – Etapas da abordagem proposta para a classificação da demanda. . . . .	37
Figura 13 – Exemplo de histograma para Chengdu. . . . .	38
Figura 14 – Exemplo de mapa de calor para 63 zonas da cidade de Nova Iorque. . . . .	39
Figura 15 – Divisão das 25 zonas em Chengdu. . . . .	42
Figura 16 – Previsões do ALMMo (com 5 variáveis de entrada) em intervalos de 15 minutos, para a zona 15, no período de 24/11/2016 até 30/11/2016. . . . .	43
Figura 17 – Previsões do ALMMo (com 5 variáveis de entrada) em intervalos de 30 minutos, para a zona 15, no período de 24/11/2016 até 30/11/2016. . . . .	44
Figura 18 – Histogramas com as faixas de valores em Chengdu. . . . .	45
Figura 19 – Mapas de calor nas 25 zonas em Chengdu - dia 30/11/2016 no intervalo de 07:00 às 07:15: (A) mapa obtido pelo ALMMo $\{N = 5\}$ para <i>Pick-Up</i> ; (B) mapa desejado para <i>Pick-Up</i> ; (C) mapa obtido pelo ALMMo $\{N = 5\}$ para <i>Drop-Off</i> e; (D) mapa desejado para <i>Drop-Off</i> . . . . .	47
Figura 20 – Mapas de calor nas 25 zonas em Chengdu - dia 30/11/2016 no intervalo de 07:00 até 07:30: (A) mapa obtido pelo ALMMo $\{N = 5\}$ para <i>Pick-Up</i> ; (B) mapa desejado para <i>Pick-Up</i> ; (C) mapa obtido pelo ALMMo $\{N = 5\}$ para <i>Drop-Off</i> e; (D) mapa desejado para <i>Drop-Off</i> . . . . .	48
Figura 21 – Divisão das 86 zonas em Nova Iorque. . . . .	49
Figura 22 – Previsões do ALMMo (com 20 variáveis de entrada) em intervalos de 15 minutos, para a Times Square (zona 230), no período de 01/07/2018 até 31/10/2018. . . . .	50

Figura 23 – Previsões do ALMMo (com 20 variáveis de entrada) em intervalos de 30 minutos, para a Times Square (zona 230), no período de 01/07/2018 até 31/10/2018. . . . .	51
Figura 24 – Histogramas com as faixas de valores em Nova Iorque com 86 zonas. . . . .	52
Figura 25 – Mapa de calor nas 86 zonas em Nova Iorque - dia 31/10/2018 no intervalo de 12:00 às 12:15: (A) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Pick-Up</i> ; (B) mapa desejado para <i>Pick-Up</i> ; (C) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Drop-Off</i> e; (D) mapa desejado para <i>Drop-Off</i> . . . . .	53
Figura 26 – Mapas de calor nas 86 zonas em Nova Iorque - dia 31/10/2018 no intervalo de 12:00 às 12:30: (A) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Pick-Up</i> ; (B) mapa desejado para <i>Pick-Up</i> ; (C) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Drop-Off</i> e; (D) mapa desejado para <i>Drop-Off</i> . . . . .	55
Figura 27 – Divisão das 63 zonas em Nova Iorque. . . . .	56
Figura 28 – Previsões do ALMMo (com 20 variáveis de entrada), em intervalos de 15 minutos, para Yorkville West (zona 263), no período de 10/12/2019 até 30/06/2020. . . . .	57
Figura 29 – Previsões do ALMMo (com 20 variáveis de entrada), em intervalos de 30 minutos para Yorkville West (zona 263), no período de 10/12/2019 até 30/06/2020. . . . .	58
Figura 30 – Histogramas com as faixas de valores em Nova Iorque com 63 zonas. . . . .	59
Figura 31 – Mapas de calor nas 63 zonas em Nova Iorque - dia 01/01/2020 no intervalo de 18:00 às 18:15: (A) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Pick-Up</i> ; (B) mapa desejado para <i>Pick-Up</i> ; (C) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Drop-Off</i> e; (D) mapa desejado para <i>Drop-Off</i> . . . . .	60
Figura 32 – Mapas de calor nas 63 zonas em Nova Iorque - dia 01/01/2020 no intervalo de 18:00 às 18:30: (A) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Pick-Up</i> ; (B) mapa desejado para <i>Pick-Up</i> ; (C) mapa obtido pelo ALMMo $\{N = 20\}$ para <i>Drop-Off</i> e; (D) mapa desejado para <i>Drop-Off</i> . . . . .	61
Figura 33 – Mapa de calor das variáveis de entrada para <i>Pick-Up</i> em Chengdu com intervalos de 15 minutos. . . . .	62
Figura 34 – Mapa de calor das variáveis de entrada para <i>Drop-Off</i> em Chengdu com intervalos de 15 minutos. . . . .	63
Figura 35 – Mapa de calor das variáveis de entrada para <i>Pick-Up</i> em Chengdu com intervalos de 30 minutos. . . . .	63
Figura 36 – Mapa de calor das variáveis de entrada para <i>Drop-Off</i> em Chengdu com intervalos de 30 minutos. . . . .	64
Figura 37 – Mapa de calor das variáveis de entrada para <i>Pick-Up</i> em Nova Iorque (86 zonas) com intervalos de 15 minutos. . . . .	65

Figura 38 – Mapa de calor das variáveis de entrada para <i>Drop-Off</i> em Nova Iorque (86 zonas) com intervalos de 15 minutos. . . . .	65
Figura 39 – Mapa de calor das variáveis de entrada para <i>Pick-Up</i> em Nova Iorque (86 zonas) com intervalos de 30 minutos. . . . .	66
Figura 40 – Mapa de calor das variáveis de entrada para <i>Drop-Off</i> em Nova Iorque (86 zonas) com intervalos de 30 minutos. . . . .	66
Figura 41 – Mapa de calor das variáveis de entrada para <i>Pick-Up</i> em Nova Iorque (63 zonas) com intervalos de 15 minutos. . . . .	67
Figura 42 – Mapa de calor das variáveis de entrada para <i>Drop-Off</i> em Nova Iorque (63 zonas) com intervalos de 15 minutos. . . . .	67
Figura 43 – Mapa de calor das variáveis de entrada para <i>Pick-Up</i> em Nova Iorque (63 zonas) com intervalos de 30 minutos. . . . .	68
Figura 44 – Mapa de calor das variáveis de entrada para <i>Drop-Off</i> em Nova Iorque (63 zonas) com intervalos de 30 minutos. . . . .	68

# Lista de Tabelas

Tabela 1 – Representação dos dados em uma base de corridas. . . . .	31
Tabela 2 – Desempenho na previsão de demanda em intervalos de 15 minutos para Chengdu. . . . .	43
Tabela 3 – Desempenho na previsão de demanda em intervalos de 30 minutos para Chengdu. . . . .	44
Tabela 4 – Definição das demandas em faixas de valores para Chengdu. . . . .	46
Tabela 5 – Acurácia em intervalos de 15 minutos para Chengdu. . . . .	46
Tabela 6 – Acurácia em intervalos de 30 minutos para Chengdu. . . . .	46
Tabela 7 – Desempenho na previsão da demanda em intervalos de 15 minutos para Nova lorque com 86 zonas. . . . .	50
Tabela 8 – Desempenho na previsão da demanda em intervalos de 30 minutos para Nova lorque com 86 zonas. . . . .	51
Tabela 9 – Definição das demandas em faixas de valores para Nova lorque com 86 zonas. . . . .	52
Tabela 10 – Acurácia em intervalos de 15 minutos para Nova lorque com 86 zonas. . . . .	53
Tabela 11 – Acurácia em intervalos de 30 minutos para Nova lorque com 86 zonas. . . . .	54
Tabela 12 – Desempenho na previsão da demanda em intervalos de 15 minutos para Nova lorque com 63 zonas. . . . .	56
Tabela 13 – Desempenho na previsão da demanda em intervalo de 30 minutos para Nova lorque com 63 zonas. . . . .	57
Tabela 14 – Definição das demandas em faixas de valores para Nova lorque com 63 zonas. . . . .	58
Tabela 15 – Acurácia em intervalos de 15 minutos para Nova lorque com 63 zonas. . . . .	59
Tabela 16 – Acurácia em intervalos de 30 minutos para Nova lorque com 63 zonas. . . . .	60
Tabela 17 – Índices e condições do tempo. . . . .	80
Tabela 18 – Desempenho na previsão de demanda em intervalos de 15 minutos para Chengdu. . . . .	82
Tabela 19 – Desempenho na previsão de demanda em intervalos de 30 minutos para Chengdu. . . . .	83
Tabela 20 – Acurácia em intervalos de 15 minutos para Chengdu. . . . .	83
Tabela 21 – Acurácia em intervalos de 30 minutos para Chengdu. . . . .	84
Tabela 22 – Desempenho na previsão da demanda em intervalos de 15 minutos para Nova lorque com 86 zonas. . . . .	84
Tabela 23 – Desempenho na previsão da demanda em intervalos de 30 minutos para Nova lorque com 86 zonas. . . . .	85
Tabela 24 – Acurácia em intervalos de 15 minutos para Nova lorque com 86 zonas . . . . .	85

Tabela 25 – Acurácia em intervalos de 30 minutos para Nova Iorque com 86 zonas .	86
Tabela 26 – Desempenho em intervalos 15 minutos para Nova Iorque com 63 zonas.	86
Tabela 27 – Desempenho na previsão da demanda em intervalos de 30 minutos para Nova Iorque com 63 zonas. . . . .	87
Tabela 28 – Acurácia em intervalos de 15 minutos para Nova Iorque com 63 zonas. .	87
Tabela 29 – Acurácia em intervalos de 30 minutos para Nova Iorque com 63 zonas. .	88

# Lista de Quadros

Quadro 1 – Resumos dos trabalhos da literatura. . . . .	28
Quadro 1 – Resumos dos trabalhos da literatura (continuação). . . . .	29

# Lista de Abreviaturas e Siglas

ALMMo	<i>Autonomous Learning Multi-Model System</i>
ANN	<i>Artificial Neural Networks</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
AUF	<i>Augmented Dickey-Fuller</i>
CNN	<i>Convolutional Neural Networks</i>
CACRNN	<i>Context-Aware Attention-Based Convolutional Recurrent Neural Network</i>
CRNN	<i>Convolutional Recurrent Neural Network</i>
Conv-LSTM	<i>Convolutional Long Short-Term Memory</i>
DBSCAN	<i>Density-Based Spatial Clustering of Applications with Noise</i>
DTW	<i>Dynamic Time Warping</i>
EDA	<i>Empirical Data Analytics</i>
EFS	<i>Evolving Fuzzy Systems</i>
eMG	<i>evolving Gaussian Multivariable Fuzzy System</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
eNFN	<i>evolving Neo-Fuzzy Neuron</i>
GCN	<i>Graph Convolutional Networks</i>
GPS	<i>Global Positioning System</i>
GRU	<i>Gated Recurrent Unit</i>
ITS	<i>Intelligent Transportation Systems</i>
MLP	<i>Multilayer Perceptron</i>
MLRNN	<i>Multi-Level Recurrent Neural Networks</i>
MT	<i>Multi-Task</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>

LSTM	<i>Long Short-Term Memory</i>
ResNet	<i>Residual Neural Networks</i>
RMSE	<i>Root Mean Squared Error</i>
RNN	<i>Recurrent Neural Networks</i>
ST	<i>Spatio-Temporal</i>
ST-Ann	<i>Spatio-Temporal Attention</i>
SVM	<i>Support Vector Machine</i>
USTN	<i>Unified Spatial-Temporal Network</i>
VAR	<i>Vector Auto Regression</i>
XGBoost	<i>eXtreme Gradient Boosting</i>

# Lista de Símbolos

$a$	Índice de alerta
$argmin$	Método que busca a nuvem de dados mais próxima
$ativa$	Tempo em que uma função de pertinência foi ativada
$b$	Valor modal (centro) de uma função de pertinência triangular
$\beta$	Taxa de aprendizagem do eNFN
$b^*$	Índice da função de pertinência mais ativa no eNFN
$b^-$	Índice da função de pertinência menos ativa no eNFN
$c$	Centro de um grupo/nuvem de dados
$cc$	Classes corretamente classificadas
$ct$	Condição do tempo
$D$	Demanda de táxi
$\bar{D}$	Densidade discreta unimodal
$dhd$	Demanda histórica dos dias
$dhi$	Demanda histórica dos intervalos
$ds$	Dia da semana
$E$	Número de variáveis de entrada
$e$	Erro residual
$F$	Quantidade de passos à frente
$fs$	Final de semana
$G$	Número de nuvens de dados/grupos
$H$	Funções de pertinência gaussianas
$hd$	Horário do dia
$i$	Índice dos grupos/nuvens de dados

<i>idade</i>	Tempo de inatividade de uma função de pertinência
<i>j</i>	Índice do tempo de ativação de uma função de pertinência
<i>K</i>	Número máximo de zonas
<i>L</i>	Determinada região do globo terrestre
<i>l</i>	Instante de tempo onde uma regra/nuvem e dados foi ativada
<i>M</i>	Distância entre grupos
<i>mdhi</i>	Média da demanda histórica dos intervalos
<i>mdhd</i>	Média da demanda histórica dos dias
<i>N</i>	Número máximo de variáveis definidas na seleção de características
<i>P</i>	Número máximo de observações passadas
<i>p</i>	Medida de compatibilidade
<i>pr</i>	Precipitação
<i>S</i>	Número total de observações
<i>T<sub>a</sub></i>	Limiar do índice de alerta
<i>T<sub>p</sub></i>	Limiar da medida de compatibilidade
<i>t</i>	Instante de tempo / intervalo de tempo
<i>tr</i>	Intervalo de tempo relativo aos dias passados
<i>te</i>	Temperatura
<i>ur</i>	Umidade relativa do ar
<i>vv</i>	Velocidade do vento
<i>w</i>	Tamanho da janela utilizado para computar o índice de alerta no eMG
<i>x</i>	Vetor de entradas
$\bar{X}$	Média escalar de determinada regra/nuvem de dados
<i>y</i>	Valor real
$\hat{y}$	Valor previsto
<i>z</i>	Índice de determinada zona

$\alpha$	Nível de significância
$\delta$	Nível de ativação de uma regra/nuvem de dados
$\eta$	Medida de utilidade de uma regra/nuvem de dados
$\eta_0$	Limiar de tolerância para exclusão de regras/nuvem de dados
$\lambda$	Taxa de aprendizado do eMG
$\mu$	Média global de uma regra/nuvem de dados
$\mu_{b_i^*}$	Valor médio do erro global correspondente a função de pertinência mais ativa $b_i^*$
$\hat{\mu}$	Média do erro global
$\omega$	Limiar para exclusão de funções de pertinência por tempo de inatividade no eNFN / valor para iniciar a matriz de covariância no ALMMo
$\hat{\sigma}^2$	Variância do erro global da rede
$\Sigma_{init}$	Matriz de dispersão inicial
$\Xi$	Nuvem de dados

# Sumário

<b>1 – Introdução</b>	<b>1</b>
1.1 Motivação e Relevância	1
1.2 Objetivos	3
1.2.1 Objetivo Geral	3
1.2.2 Objetivos Específicos	3
1.3 Publicações	3
1.4 Principais Contribuições	4
1.5 Organização do Trabalho	4
<b>2 – Fundamentação Teórica</b>	<b>6</b>
2.1 Sistemas de Transporte Inteligente	6
2.1.1 Gestão de Preços	8
2.1.2 Oferta e Demanda de Passageiros	9
2.1.3 Previsão e Classificação da Demanda de Passageiros	9
2.2 Técnicas de Divisão de Regiões	11
2.3 Mapas de Calor	12
2.4 Aprendizagem <i>Offline</i> e <i>Online</i>	14
2.5 Sistemas <i>Fuzzy</i> Evolutivos	15
2.5.1 ALMMo - Autonomous Learning Multi-Model System	16
2.5.2 eMG - evolving Gaussian Multivariable Fuzzy System	19
2.5.3 eNFN - evolving Neo-Fuzzy Neuron	22
<b>3 – Trabalhos Relacionados</b>	<b>24</b>
3.1 Previsão com Análise Temporal	24
3.2 Previsão com Inclusão de Variáveis Exógenas	25
3.3 Combinações de Técnicas Espaciais e Temporais para a Previsão	26
3.4 Resumo dos Trabalhos	27
<b>4 – Abordagem Proposta</b>	<b>30</b>
4.1 Previsão da Demanda	30
4.1.1 Extração da Base de Corridas	30
4.1.2 Caracterização das Variáveis de Entrada	32
4.1.3 Seleção de Variáveis	35
4.1.4 Previsão da Demanda	35
4.2 Classificação da Demanda	36
4.2.1 Definição das Classes	36

4.2.2	Classificação da Demanda . . . . .	37
4.2.3	Geração dos Mapas de Calor . . . . .	38
<b>5</b>	<b>– Experimentos Computacionais . . . . .</b>	<b>40</b>
5.1	Previsão e Classificação da Demanda em Chengdu (China) . . . . .	42
5.1.1	Previsão da Demanda em Chengdu . . . . .	43
5.1.2	Classificação da Demanda em Chengdu . . . . .	45
5.2	Previsão e Classificação da Demanda em Nova Iorque (Estados Unidos da América) com 86 zonas . . . . .	48
5.2.1	Previsão da Demanda em Nova Iorque com 86 zonas . . . . .	49
5.2.2	Classificação da Demanda em Nova Iorque com 86 zonas. . . . .	51
5.3	Previsão e Classificação da Demanda em Nova Iorque (Estados Unidos da América) com 63 zonas . . . . .	54
5.3.1	Previsão da Demanda em Nova Iorque com 63 zonas . . . . .	55
5.3.2	Classificação da Demanda em Nova Iorque com 63 zonas . . . . .	57
5.4	Análise das Variáveis de Entrada . . . . .	61
5.4.1	Análise das Variáveis de Entrada em Chengdu . . . . .	62
5.4.2	Análise das Variáveis de Entrada em Nova Iorque com 86 zonas . . . . .	65
5.4.3	Análise das Variáveis de Entrada em Nova Iorque com 63 zonas . . . . .	67
<b>6</b>	<b>– Considerações Finais . . . . .</b>	<b>69</b>
6.1	Perspectivas de Continuidade . . . . .	70
	<b>Referências . . . . .</b>	<b>71</b>
	<b>Apêndices . . . . .</b>	<b>79</b>
	<b>APÊNDICE A – Lista das Condições do Tempo . . . . .</b>	<b>80</b>
	<b>APÊNDICE B – Índices das 63 e 86 Zonas em Nova Iorque . . . . .</b>	<b>81</b>
	<b>APÊNDICE C – Listagem Completa dos Resultados e as Respectivas Variáveis de Entrada . . . . .</b>	<b>82</b>

# Capítulo 1

## Introdução

Este capítulo apresenta uma descrição introdutória do problema abordado e ilustra os desafios e descobertas que permeiam esta pesquisa. A Seção 1.1 contextualiza o problema tratado por este trabalho e os motivos que levaram ao seu desenvolvimento. Posteriormente, na Seção 1.2, são detalhados os objetivos gerais e os específicos. Na Seção 1.3, são descritas as publicações geradas a partir deste trabalho e as principais contribuições são descritas na Seção 1.4. Por último, a Seção 1.5 mostra um breve detalhamento dos capítulos subsequentes.

### 1.1 Motivação e Relevância

Os problemas relacionados à mobilidade urbana nos grandes centros têm aumentado constantemente. A popularização de veículos particulares em circulação, somada à ineficiência e precariedade de transportes públicos coletivos (ônibus, metrô, etc.) têm contribuído para o aumento do tráfego urbano, impactando negativamente a vida de milhões de pessoas (CUNHA et al., 2017). Um dos desafios principais para os órgãos de transportes destas cidades é como gerenciar estratégias para mitigar estes problemas (YAO et al., 2018).

As corridas de táxi são alternativas convidativas em face aos tradicionais meios de transporte coletivo, por oferecerem serviços eficientes e convenientes aos passageiros, permitindo uma experiência personalizada e de fácil acesso (JIANG et al., 2019). Contudo, um problema pertinente e constante para este sistema de transporte ainda não foi solucionado: a oferta e a demanda de passageiros. Passageiros podem ficar sem atendimento devido a uma grande demanda em determinada região e, em contrapartida, há locais onde a oferta de veículos é grande, mas poucos clientes solicitarão corridas (BAI et al., 2019). Sabendo-se com antecedência qual a demanda de determinada região, empresas de táxi podem organizar de forma inteligente sua frota, aumentando ou diminuindo a oferta de acordo com a necessidade. Com isso, diminui-se o tempo de espera dos passageiros, a ociosidade dos táxis e, conseqüentemente, os taxistas podem atender mais clientes e otimizar os seus

lucros (WEI et al., 2016). Além disso, o desgaste dos veículos e o tempo gasto em trânsito também diminui, gerando redução dos custos com combustível e manutenção (XU et al., 2017). Algumas pesquisas revelam que, se a oferta de táxis conseguisse suprir a demanda desejada, a ineficiência do mercado de transportes privados cairia em mais de 60% (ZHAN; QIAN; UKKUSURI, 2016).

Para endereçar esta questão, diversas técnicas de aprendizado de máquina têm sido utilizadas para resolver ou minimizar o problema da oferta e demanda de passageiros. *Auto-regressive Integrated Moving Average* (ARIMA) (MOREIRA-MATIAS et al., 2012; MOREIRA-MATIAS et al., 2013), *Convolutional Neural Networks* (CNN) (ZHANG; LIU; ZHENG, 2019; ZHANG et al., 2020), *Long Short-Term Memory* (LSTM) (KE et al., 2017; XU et al., 2018) e *Support Vector Machine* (SVM) (JIANG et al., 2019; YAO et al., 2019) são exemplos de algumas das principais técnicas utilizadas. Contudo, estas técnicas, em sua grande maioria, apresentam determinadas limitações inerentes à sua natureza, a saber:

- necessitam de revisões periódicas dos modelos, treinando-os e adequando-os novamente à realidade dos dados (ANGELOV; ZHOU, 2006);
- não lidam com situações nas quais há mudanças nas condições ambientais, típicas de configurações *online* com dados não estacionários (LUGHOFER; ANGELOV, 2011);
- raramente conseguem lidar com sistemas complexos onde há múltiplos modos de operação (LUGHOFER, 2015).

Estas limitações geram lacunas metodológicas que dificultam uma previsão acurada das demandas de táxi. Isso se deve, principalmente, ao fato de a demanda estar relacionada simultaneamente a diversos fatores exógenos com alta volatilidade. Temperatura, horário do dia, eventos importantes na região e tráfego são alguns exemplos desses fatores (LIU et al., 2019). Além disso, dada a sua natureza, extrair conhecimento de conjuntos de dados que possuem estas características é uma tarefa complexa e pouco intuitiva, favorecendo a incerteza na interpretação de tais informações (MACIEJEWSKI; BISCHOFF; NAGEL, 2016). Dado este contexto, é desejável que as técnicas utilizadas para previsão da demanda de táxi possuam aprendizado contínuo e incremental, adaptando-se aos fatores externos presentes no trânsito. Adicionalmente, estes sistemas devem, preferencialmente, suprir as restrições inerentes às técnicas tradicionais já citadas.

Os sistemas *fuzzy* evolutivos (do inglês, *Evolving Fuzzy Systems* - EFS) foram propostos para o processamento incremental de grandes fluxos de dados, nos quais as amostras são apresentadas uma única vez ao modelo e depois são descartadas (LUGHOFER, 2015). Estes sistemas têm como principal característica a evolução de sua estrutura (neste caso, novos neurônios, grupos e/ou regras *fuzzy*) e de seus parâmetros *online*, eventualmente em

tempo real, à medida que novas informações são recebidas de um fluxo contínuo de dados (ANGELOV; GU; PRINCIPE, 2017; SOUZA; LUGHOFER, 2021). Além disso, destaca-se a característica dos EFS em tratar a incerteza das informações, e estas são inerentes ao problema em questão. Os EFS têm sido utilizados com sucesso em diversas aplicações, como por exemplo em previsões no mercado financeiro (ANGELOV; GU; PRINCIPE, 2017; AMARAL et al., 2019; ANGELOV; GU, 2019), no tratamento de dados ausentes (GARCIA; LEITE; ŠKRJANC, 2019; JUNIOR; SILVA, 2021), na mobilidade urbana e gerenciamento de tráfego (TANG et al., 2017; ANDONOVSKI et al., 2020), na previsão e monitoramento de dados climáticos (SOARES et al., 2018a; SOARES et al., 2018b), no reconhecimento de padrões (SOUZA; LUGHOFER; GUIMARAES, 2021; SOUZA; LUGHOFER, 2021) e em diagnósticos na área de saúde (SOARES; ANGELOV; GU, 2020), para citar somente alguns.

Neste contexto, o objetivo deste trabalho é apresentar uma nova abordagem para previsão e classificação da demanda de passageiros de táxi. Na abordagem proposta, a previsão/classificação é realizada por um sistema *fuzzy* evolutivo utilizando como entrada informações históricas de corridas de táxis e dados climáticos.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

O objetivo geral deste trabalho é propor uma abordagem utilizando sistemas *fuzzy* evolutivos para a previsão e classificação da demanda de passageiros de táxi.

### 1.2.2 Objetivos Específicos

- Propor e implementar uma abordagem utilizando sistemas *fuzzy* evolutivos para a previsão da demanda de passageiros de táxi.
- Propor e implementar uma abordagem utilizando sistemas *fuzzy* evolutivos para a classificação da demanda de passageiros de táxi.
- Realizar experimentos computacionais em bases de dados contendo corridas de táxi reais para avaliar o desempenho da abordagem proposta.
- Comparar os resultados obtidos pela abordagem proposta em face aos resultados da literatura.

## 1.3 Publicações

Esta seção apresenta as publicações realizadas no decorrer desta pesquisa.

1. Linhares, L.F; Silva, A.M: **Previsão de Demanda de Passageiros de Táxi por Sistemas *Fuzzy* Evolutivos**. Em: VI Congresso Brasileiro de Sistemas Fuzzy (CBSF 2021) (Publicado).
2. Linhares, L.F; Silva, A.M; Meireles, M.R.G: **An Incremental Approach to Forecasting and Classification of Taxi Demand Based on Evolving Fuzzy Systems**. (Artigo em desenvolvimento, a ser submetido em periódico).

## 1.4 Principais Contribuições

Resumidamente, as principais contribuições deste trabalho para o setor de transportes, bem para o meio acadêmico, são:

- Descrição de uma nova abordagem para a previsão da demanda de passageiros de táxi utilizando sistemas *fuzzy* evolutivos, dados históricos de corridas e informações meteorológicas.
- Proposição de uma nova abordagem para a classificação da demanda de táxis. A abordagem proposta utiliza a demanda prevista pelo modelo evolutivo para realizar a classificação em quatro classes que são determinadas de acordo com a demanda: Muito Baixa; Baixa; Média e; Alta.
- As abordagens propostas foram desenvolvidas para configurações *online* com condições ambientais voláteis e dados não estacionários. Nestes ambientes, os dados chegam como um fluxo, têm tamanho ilimitado e estão sujeitos a variações e mudanças de conceito. Estas são características típicas do problema em questão.
- Representação da demanda de táxis em mapas de calor. Essa forma de visualização é mais intuitiva, permitindo uma rápida visualização da demanda em diversas zonas da cidade e facilitando uma tomada de decisão.

## 1.5 Organização do Trabalho

Para melhor entendimento deste trabalho, o texto está dividido em 6 capítulos como se segue:

- Este capítulo inicial demonstrou as motivações e os objetivos que permeiam esta pesquisa.
- O Capítulo 2 apresenta todo o arcabouço teórico necessário para o entendimento do trabalho. Discorre-se sobre os Sistemas de Transporte Inteligentes, que incluem

a oferta e demanda a de passageiros. É feita também uma breve explicação sobre as técnicas de divisão de regiões, mapas de calor e aprendizagem *online/offline*. Por último, é realizada uma sucinta e importante fundamentação sobre sistemas *fuzzy* evolutivos.

- O Capítulo 3 traz uma revisão da literatura apresentando as principais abordagens relacionadas à demanda de passageiros de táxis.
- O Capítulo 4 introduz a abordagem proposta e apresenta, detalhadamente, cada uma de suas etapas.
- O Capítulo 5 discorre sobre os experimentos computacionais realizados para avaliar e comparar o desempenho da abordagem proposta.
- O Capítulo 6 resume as contribuições dessa dissertação e sugere as propostas de continuidade para esta pesquisa.

# Capítulo 2

## Fundamentação Teórica

Neste capítulo, são apresentados os principais conceitos relacionados aos temas que compõem este trabalho. A Seção 2.1 discorre sobre os Sistemas de Transporte Inteligente e suas principais aplicações, dando ênfase à Oferta e Demanda de Passageiros e à Previsão e Classificação da Demanda. A Seção 2.2 trata das técnicas disponíveis para divisão de regiões. Uma breve explicação sobre mapas de calor é apresentada na Seção 2.3. A Seção 2.4 trata dos conceitos de aprendizagem *offline* e *online*. Por fim, a Seção 2.5 apresenta os conceitos de sistemas *fuzzy* evolutivos e detalha os algoritmos que serão utilizados neste trabalho.

### 2.1 Sistemas de Transporte Inteligente

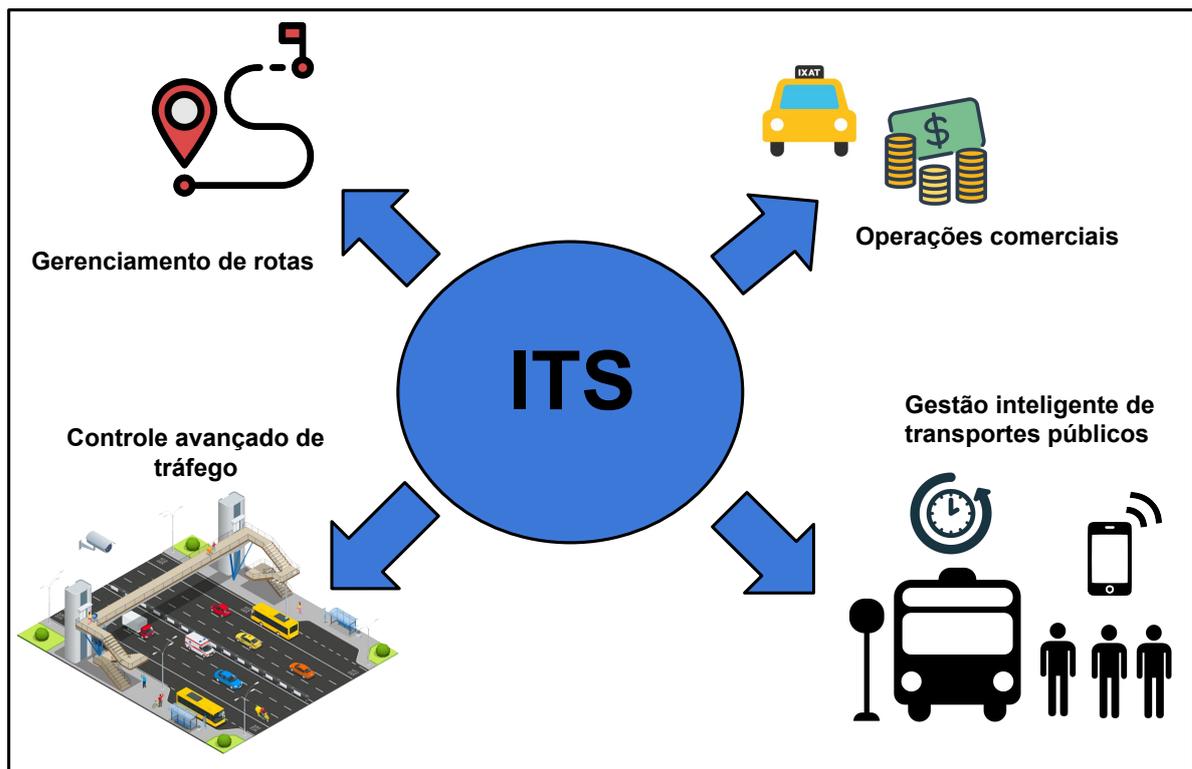
Os Sistemas de Transporte Inteligente (*Intelligent Transportation Systems* - ITS) são sistemas integrados que agrupam uma gama de tecnologias em comunicação, controle e sensores veiculares para resolver e gerenciar problemas que permeiam o tráfego urbano (SINGH; GUPTA, 2015). Estes problemas são causados principalmente pelo crescimento desenfreado da quantidade de veículos que trafegam em grandes megalópoles, colaborando com a geração de diversos problemas de mobilidade urbana (CUNHA et al., 2017). Congestionamentos, aumento de acidentes, poluição sonora e do ar são alguns dos problemas causados pelo excesso de veículos (DIMITRAKOPOULOS; DEMESTICHAS, 2010). Os ITS não são propostos apenas para gerenciar as condições de tráfego urbano, mas também têm como objetivo tornar o setor de transportes mais seguro, sustentável e eficiente (CUNHA et al., 2017). Para Singh e Gupta (2015), as aplicações em ITS oferecem os seguintes benefícios para a sociedade:

- melhor mobilidade para a população, incluindo maior acesso aos transportes para idosos, deficientes e pessoas que vivem em regiões remotas;
- diminuição nas mortes e acidentes no trânsito;

- melhor gerenciamento do tráfego;
- menores incertezas no trajeto de viagens, possibilitando melhor planejamento por parte dos usuários.

As tecnologias da computação, robótica, automação e comunicações têm sido usadas em pesquisas e desenvolvimentos de diversos ITS. Gerenciamento inteligente de tráfego, prevenção de acidentes, operações comerciais de transporte, automação veicular e gerenciamento de transportes públicos são exemplos de aplicações abordadas (QU; WANG; YANG, 2010). O crescimento desses sistemas fomentou a criação de congressos e revistas específicas para discutir as soluções de ITS e suas diversas aplicações, como o *Transportation Research Part C: Emerging Technologies*<sup>1</sup>, *IEEE Transactions on Intelligent Transportation Systems*<sup>2</sup>, *International IEEE Conference on Intelligent Transportation Systems*<sup>3</sup>, *Springer Geography*<sup>4</sup>, para citar somente alguns. A Figura 1 ilustra alguns dos campos de pesquisa mais abordados em ITS.

Figura 1 – Exemplos de aplicações em ITS.



Nos trabalhos de Sussman (2005) e Qu, Wang e Yang (2010) foram identificados vários nichos de aplicações para ITS. Para efeito de objetividade, neste trabalho serão abordados

<sup>1</sup> <https://www.journals.elsevier.com/transportation-research-part-c-emerging-technologies>

<sup>2</sup> <https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6979>

<sup>3</sup> <https://www.ieee.org/conferences/index.html>

<sup>4</sup> <https://www.springer.com/series/10180>

apenas os Sistemas de Gerenciamento Avançado de Tráfego e as Operações Comerciais de Transporte.

Os Sistemas de Gerenciamento Avançado de Tráfego são aqueles que auxiliam motoristas e unidades regionais de trânsito a entender e prever problemas inerentes ao tráfego em tempo real. Estes sistemas preveem congestionamentos e fornecem alternativas de roteamento para veículos em áreas regionais para melhorar a eficiência da rodovia, adaptando-se ao dinamismo do ambiente em que foi inserido (SUSSMAN, 2005). Estes sistemas agregam diversos dados relacionados ao tráfego e, a partir de fontes heterogêneas, utilizam diversas técnicas para selecionar e fundir esses dados, objetivando a geração de informações consistentes. Por fim, utilizam essas informações para conceber aplicações e serviços que auxiliarão na detecção, controle e redução de congestionamentos (CUNHA et al., 2017).

Nas Operações Comerciais de Transporte, empresas privadas e táxis começam a adotar soluções de ITS para incrementar a qualidade dos serviços prestados e gerenciar melhor a força de trabalho das suas frotas. Este ramo apresenta-se como uma tecnologia emergente e promissora, visto que os sistemas criados para este fim proporcionam vantagens diretas e lucrativas (SUSSMAN, 2005). Neste sentido, duas vertentes das Operações Comerciais de Transporte serão detalhadas: a Gestão de Preços e a Oferta e Demanda de Passageiros.

### 2.1.1 Gestão de Preços

Os estudos da Gestão de Preços visam identificar e mitigar problemas críticos na precificação em serviços oferecidos por transportes privados, como estratégias de precificação indevida, viagens seletivas mais lucrativas, dentre outros. Sendo assim, abordagens são propostas como auxílio à mitigação destas situações (SUSSMAN, 2005).

As viagens seletivas são corridas de táxi de longa duração escolhidas propositalmente por serem mais lucrativas. O excesso da oferta de táxis é visto em aeroportos, rodoviárias e outros pontos onde a demanda por táxis é grande. Taxistas acordam preços fechados com clientes, eliminando a necessidade de taxímetros e incentivando a concorrência desleal com taxistas legalizados, ou seja, que utilizam meios legais (taxímetros, aplicativos) para precificar suas corridas (YANG et al., 2010).

São estudadas também estratégias para minimizar os impactos financeiros providos do cancelamento de corridas. Estes cancelamentos (sem custo para o cliente) acontecem, seja porque houve a desistência da corrida, ou o cliente encontrou outro táxi no último instante, causando vacância do táxi solicitado e uma potencial perda de clientes que seriam atendidos por ele (HE et al., 2018).

## 2.1.2 Oferta e Demanda de Passageiros

A oferta e a demanda de passageiros remete ao seguinte problema: reduzir uma lacuna existente entre a oferta de veículos disponíveis (táxis) e a demanda de passageiros em uma determinada região e horário (MOREIRA-MATIAS et al., 2016). O custo com combustível e a busca por clientes é uma consequência afetada por esta lacuna, isto é, enquanto taxistas gastam tempo procurando por passageiros, têm-se passageiros que não encontram táxis disponíveis, aumentando consequentemente a insatisfação dos clientes pelos serviços prestados e o gasto excessivo de combustível nos veículos de táxi (MOREIRA-MATIAS et al., 2013).

Uma vez que se sabe de antemão qual a demanda de sua clientela, empresas de táxi conseguem organizar de forma logística sua frota, aumentando ou diminuindo sua oferta, diminuindo o tempo de espera dos passageiros, reduzindo o tempo ocioso dos taxistas sem clientes e atendendo mais passageiros e, consequentemente, obtendo maiores lucros (WEI et al., 2016). Além disso, o custo com combustível, o desgaste dos veículos e o tempo em trânsito também diminui, gerando redução dos custos de uma frota de táxis (XU et al., 2017).

Segundo Moreira-Matias et al. (2013), para conseguir quantificar a demanda de passageiros, é possível extrair conhecimento de diversas fontes de dados disponíveis. Utilizar dados de tráfego urbano é a forma principal de agregar informações da demanda. Taxímetros e aplicativos próprios contidos em *smartphones* contêm, em seu núcleo, a tecnologia de rastreamento conhecida como *Global Positioning System* (GPS). O GPS é um sistema que busca informações de satélites posicionados estrategicamente para identificar a localização aproximada em tempo real de veículos, pessoas, objetos, etc (HEGARTY, 2017). Com esta tecnologia, é possível saber com certa confiabilidade as seguintes informações: local de embarque, velocidade média do veículo, trajetória e tempo percorridos, local de desembarque, dentre outros. Diversas empresas utilizam-se destas tecnologias para controlar sua frota e ter registros consistentes. Com isso, por meio de técnicas computacionais, é possível transformar estes dados em informação útil, que servirá de apoio para entender o fluxo da demanda em determinada região.

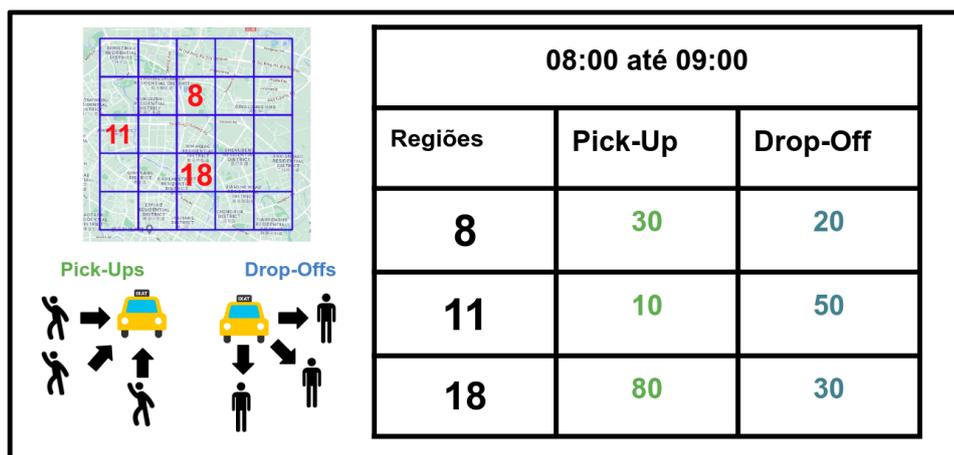
## 2.1.3 Previsão e Classificação da Demanda de Passageiros

A previsão da demanda de passageiros é considerada como um problema de séries temporais (ZHANG et al., 2020; TONG et al., 2017; ZHANG; LIU; ZHENG, 2019). Uma série temporal pode ser descrita como um conjunto de observações organizadas sequencialmente ao longo do tempo. Supondo que se tenha um conjunto de variáveis observadas ( $x_t, x_{t-1}, x_{t-2}, \dots, x_{t-P}$ ) de  $P$  observações passadas até o instante  $t$ , deseja-se assim prever o próximo valor ( $x_{t+1}$ ) (CHATFIELD, 2000). Para entender como a previsão da demanda é

realizada, é necessário primeiro apresentar alguns conceitos preliminares:

- **Zona de Táxi:** uma zona de táxi pode ser definida como a partição de uma região  $L$  em  $K$  regiões menores, definidas como zonas. Cada zona é definida por um índice  $z$ , sendo  $1 \leq z \leq K$ ;
- **Intervalo de Tempo:** é o tempo contínuo particionado em intervalos sequenciais idênticos. Considere o período observado de 30 dias e um intervalo de 60 minutos. Este período representaria então 720 (24 x 30) observações;
- **Corrida de Táxi:** uma corrida de táxi representa um evento, onde o motorista busca o passageiro em determinada região e em determinado horário (*Pick-Up*) e o deixa em seu destino também em determinada região e horário (*Drop-Off*) (ZHANG et al., 2020);
- **Demanda de Passageiros:** trabalhos na literatura, como os de Ke et al. (2017), Bai et al. (2019) e Zhang et al. (2020) classificam as demandas de passageiros como o somatório de *Pick-Up* e *Drop-Off* em determinado intervalo de tempo  $t$ , em uma determinada zona  $z$ . A Figura 2 mostra um exemplo da quantidade da demanda de (*Pick-Up* e *Drop-Off*) por táxis nas zonas 8, 11 e 18, ocorridas entre 08:00 e 09:00 h da manhã.

Figura 2 – Demandas de *Pick-Up* e *Drop-Off* em determinada região e horário.



Introduzidos os conceitos supracitados, dado um número  $P$  de observações de demandas passadas ( $D_t, D_{t-1}, D_{t-2}, D_{t-3}, \dots, D_{t-P}$ ), objetiva-se prever a demanda de passageiros  $F$  passos à frente ( $D_{t+F}$ ). A previsão da demanda pode ser definida para períodos curtos (*short-term*) e para períodos longos (*long-term*). Basicamente, o que diferencia uma demanda de curtos para longos períodos é a quantidade de passos à frente em que é realizada a previsão. Demandas de longos períodos aumentam a janela de passos (*steps*),

prevendo  $F$  demandas à frente, onde  $F > 1$ . Já nas de curto período, as demandas à frente são definidas como  $F = 1$ , ou seja, um passo à frente (ZHANG; LIU; ZHENG, 2019).

A classificação é uma das tarefas mais comuns presentes em aprendizado de máquina. Esta consiste em reconhecer e categorizar padrões, dadas determinadas características que estão pré-determinadas em diferentes classes (NOVAKOVIĆ et al., 2017). Os problemas de classificação podem ser divididos em duas categorias: classificação binária e classificação multi-classe (SEN; HAJRA; GHOSH, 2020). Nos problemas de classificação binária, só é possível que se obtenha duas classes. Já nos problemas multi-classe, é possível que se tenha três ou mais classes.

No contexto da demanda de passageiros, uma das formas para se classificar a demanda é definir uma faixa de valores na qual o valor discreto da demanda se encaixa. Cada faixa representaria uma classe, e estas seriam definidas por nomenclaturas, tornando-se, assim, um problema multi-classes (ZANDER, 2017). Como exemplo, tome as faixas  $[0 - 10]$ ,  $[10 - 50]$ ,  $[50 - 100]$  e  $[ > 100]$ , representadas respectivamente pelas classes "Muito Baixa", "Baixa", "Média" e "Alta". Tome também como exemplo uma previsão que retornou o valor discreto 45. Este, portanto, seria representado pela classe "Baixa", pois  $10 < 45 < 50$ .

Uma vez definido como trabalhos da literatura geralmente endereçam a previsão e a classificação da demanda de passageiros, esforços têm sido feitos para a criação de sistemas e técnicas que consigam realizar esta tarefa, isto é, saber previamente qual o quantitativo de passageiros que provavelmente virão a solicitar um transporte privado, indo ao encontro dos ITS. Além de ser tema deste trabalho, realizar a previsão e a classificação da demanda também foi alvo de estudos em diversos trabalhos, sendo o Capítulo 3 dedicado a sumarizar os principais estudos relacionados a este tema.

## 2.2 Técnicas de Divisão de Regiões

A divisão de uma região no mapa feita de forma eficiente pode melhorar a acurácia de modelos preditores, justamente por moldar formatos específicos de cada região baseados em estudos empíricos regionais. Alguns trabalhos da literatura, como os de Wang et al. (2019), Zhang, Liu e Zheng (2019), Zhang et al. (2020) realizam a divisão da seguinte forma: escolha de uma determinada região  $L$ , delimitando suas latitudes e longitudes máximas e dividindo-a em  $K$  polígonos semelhantes, sendo retângulos ou hexágonos as formas comumente utilizadas. Cada limiar do polígono é definido por um conjunto de latitudes e longitudes. Portanto, uma zona  $z$  de táxi pode ser descrita como  $L_z$ , onde  $1 \leq z \leq K$ . É possível encontrar diversas ferramentas para auxiliar no processo de delimitação e divisão

de zonas, como o *QGIS*<sup>5</sup>, *ArcGIS*<sup>6</sup> e a biblioteca *GeoPandas*<sup>7</sup>.

Outras formas de divisão incluem identificar centros de alta densidade de demandas por meio de algoritmos de agrupamento, como o *K-means*. Com isso, determinam-se essas densidades com técnicas de delimitação, como os diagramas de Voronoi (DAVIS; RAINA; JAGANNATHAN, 2018). O trabalho de Moreira-Matias et al. (2016) propõe uma matriz de origem e destino evolutiva, dividindo a cidade de Porto, Portugal em regiões moldadas pela densidade das solicitações de táxi. O algoritmo que gera estas regiões trabalha de forma evolutiva, moldando a forma de cada região da matriz conforme a sazonalidade da cidade.

Adicionalmente, a seleção de locais turísticos (Estádios, Museus, etc.) (VANICHRUJEE et al., 2018; RODRIGUES; MARKOU; PEREIRA, 2019) também é utilizada como formas de delimitação das regiões. Para isso, é determinado um raio de distância predefinido partindo da latitude/longitude central do local. Há também a divisão espacial por bairros ou formatos irregulares pré-definidos, por exemplo, pela prefeitura da cidade (ZHANG et al., 2020). A Figura 3 ilustra três técnicas de divisão das zonas, a saber: (A) divisão por polígonos; (B) divisão por diagramas de Voronoi e; (C) divisão irregular na cidade de Nova Iorque (Estados Unidos da América).

## 2.3 Mapas de Calor

Os mapas de calor são representações gráficas bidimensionais de dados, onde os valores de uma determinada variável são representados por cores. Sua forma de apresentação minimiza a quantidade de aprendizado necessário para compreendê-lo. Em um mapa de calor, é fácil identificar que a proporção de determinada cor é relativa ao nível da variável representada. Os mapas de calor mostram as informações diretamente sobre o estímulo, pois os dados podem não estar mais perto dos elementos a que pertencem, necessitando de pouco esforço mental para interpretá-los (BOJKO, 2009).

Segundo Bojko (2009), é necessário se atentar a algumas diretrizes antes de utilizar os mapas de calor em um trabalho:

- utilize mapas de calor apenas se eles forem relevantes para a pesquisa;
- mapas de calor são utilizados para visualização ao invés de análise de dados;
- conheça as limitações de cada mapa de calor para prever interpretações equivocadas nos resultados;

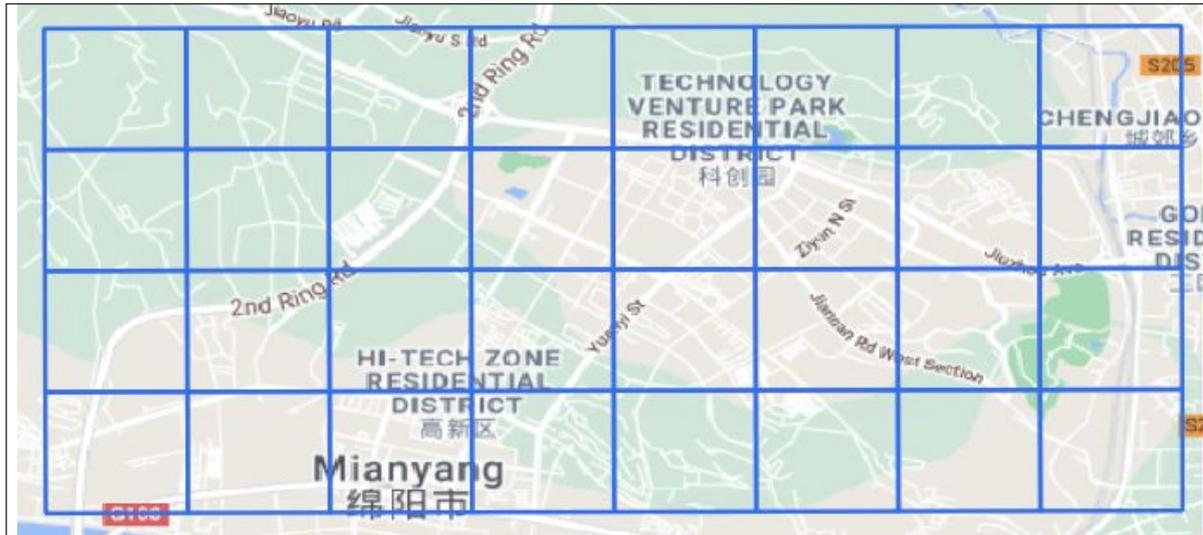
---

<sup>5</sup> <https://qgis.org/>

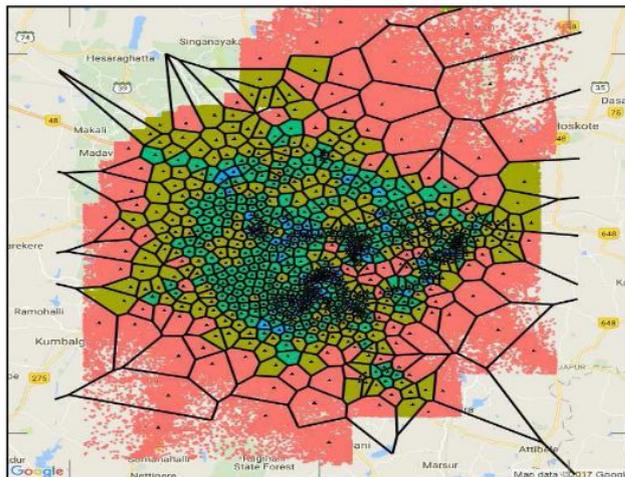
<sup>6</sup> <https://www.arcgis.com>

<sup>7</sup> <https://geopandas.org>

Figura 3 – Exemplos de divisões de zonas: (A) divisão retangular na cidade de Mianyang (China); (B) divisão baseada em diagramas de Voronoi e; (C) divisão em zonas irregulares na cidade de Nova Iorque (Estados Unidos da América).



(A)



(B)



(C)

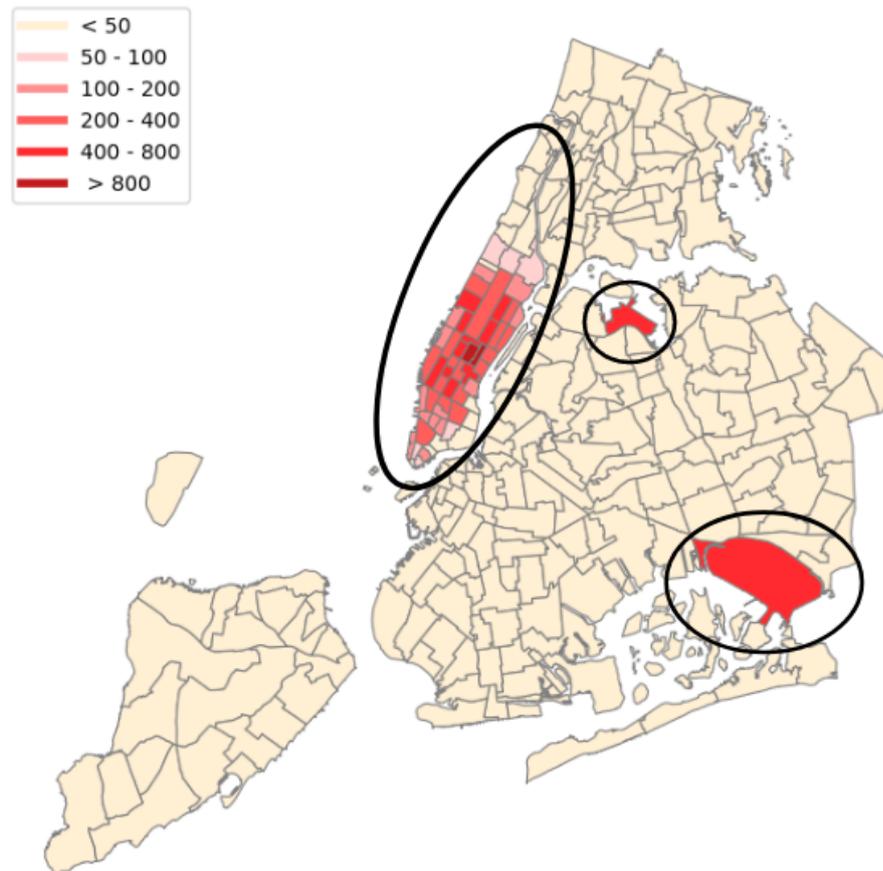
Fonte: Figura (B) adaptada de [Davis, Raina e Jagannathan \(2018\)](#).

- defina os limítrofes da variável para a escala de cores.

Os mapas de calor coropletos são mapas utilizados geralmente para representar informações em uma determinada região geográfica. A principal diferença entre os mapas coropletos em relação aos convencionais é o limite de representação das cores na imagem. Enquanto os mapas de calor tradicionais representam as cores sem um limítrofe, os mapas coropletos apresentam as informações em cores limitadas por um formato pré-definido, seja ele irregular ou não ([BABICKI et al., 2016](#)). Entre suas aplicações, pode-se citar: fluxo de tráfego; rotas e densidade de veículos; densidade populacional; alastramento de doenças; para citar somente alguns ([CRAMPTON, 2004](#)). A Figura 4 ilustra o exemplo de um mapa

de calor coropleto, representando a quantidade de solicitações de táxis de 263 regiões pré-definidas dos bairros de Nova Iorque (Estados Unidos da América). Os círculos em preto mostram as regiões com maior densidade de solicitações.

Figura 4 – Exemplo de mapa de calor de solicitações de táxi em Nova Iorque.



## 2.4 Aprendizagem *Offline* e *Online*

A aprendizagem *offline* (também conhecida como *batch*) representa o treinamento de modelos tradicionais de aprendizado de máquina feito por amostras acumuladas em um período de tempo, isto é, seu treinamento é feito em lote. Conforme [Ben-David, Kushilevitz e Mansour \(1997\)](#), as principais características desse tipo de aprendizado incluem:

- necessidade de possuir todo o conjunto de dados disponível previamente;
- o treinamento dos modelos é feito utilizando várias iterações nos dados para que se encontre o melhor conjunto de pesos e/ou parâmetros;
- sistemas uma vez treinados não são mais atualizados.

Para este tipo de aprendizagem, o conjunto de dados, geralmente, é dividido em treinamento e teste. Os dados de treinamento são executados em lote nos modelos para atualização dos

parâmetros. Já os dados de teste são executados com o modelo já treinado previamente para avaliação de sua performance (CHATFIELD, 2000). Em algumas abordagens, a divisão dos conjuntos de dados é feita por treinamento, validação e teste, sendo a validação um período necessário para avaliar se o treinamento do modelo foi satisfatório.

O conceito de aprendizagem *online* trata do processamento de dados por um fluxo contínuo. Sendo assim, os modelos que utilizam desta aprendizagem processam as amostras incrementalmente, à medida que são apresentadas. Além disso, o treinamento destes modelos é constante, ou seja, não há interrupção na aprendizagem enquanto há disponibilidade de novas amostras de dados (SAYED-MOUCHAWEH; LUGHOFER, 2012). Segundo Leite, Škrjanc e Gomide (2020), o processamento de fluxos contínuos de dados traz algumas características inerentes à sua natureza, a saber:

- as amostras chegam continuamente e os fluxos de dados podem ser infinitos;
- os sistemas não têm controle sobre como ou quando os dados chegarão;
- após o processamento, uma amostra deve ser idealmente descartada para evitar problemas de escalabilidade.

Esta forma de aprendizagem também é vantajosa em termos computacionais, pois utiliza poucos recursos de memória, processador e disco (SAYED-MOUCHAWEH; LUGHOFER, 2012).

## 2.5 Sistemas *Fuzzy* Evolutivos

O aumento na disponibilidade de grandes quantidades de informação têm motivado o desenvolvimento de modelos baseados em algoritmos incrementais capazes de processar grandes fluxos de dados de modo contínuo (*online*) (LEMOS; CAMINHAS; GOMIDE, 2011). Estas informações, geralmente, são caracterizadas por serem dinâmicas e não estacionárias (SILVA, 2014). Nesse sentido, modelos tradicionais que utilizam aprendizado *offline* podem não ser uma alternativa viável para esta realidade, visto que necessitam de constantes revisões toda vez que a realidade dos dados apresenta mudanças substanciais (SAYED-MOUCHAWEH; LUGHOFER, 2012).

A necessidade por modelos computacionais com processamento incremental fez surgir um novo campo de pesquisa, voltado a desenvolver metodologias com mecanismos capazes de treinar e atualizar continuamente os modelos, de forma *online*, para que estes se adaptem às mudanças encontradas em um fluxo de dados (SILVA, 2014). Os sistemas *fuzzy* com essas habilidades são conhecidos como sistemas *fuzzy* evolutivos. De acordo com Angelov

(2010), os sistemas inteligentes evolutivos podem ser vistos como uma combinação entre sistemas com estrutura expansível (neste caso específico, as estruturas de sistemas *fuzzy*) e métodos de aprendizado de máquina *online*. Segundo Inácio (2014), sendo os sistemas *fuzzy* evolutivos um tipo de sistema inteligente evolutivo, estes compartilham das seguintes características:

- sua estrutura não é fixa, ou seja, ela se expande e retrai naturalmente conforme o sistema evolui;
- seus parâmetros são adaptados continuamente, à medida que o sistema evolui;
- seu funcionamento é ininterrupto, permitindo o aprendizado de modo *online* e, se necessário, em tempo real.

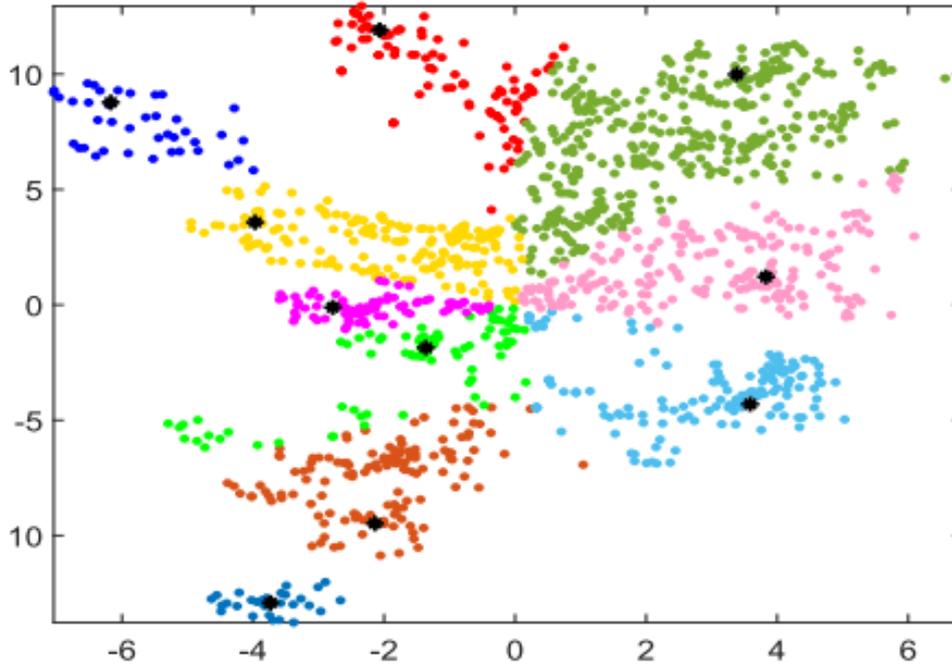
Os sistemas *fuzzy* evolutivos, por possuir um aprendizado incremental, demonstram rápido processamento dos dados, uma vez que as amostras são processadas e logo após são descartadas (LUGHOFER, 2015). Esta característica permite a execução de massivos conjuntos de dados, sem acarretar elevados custos computacionais, o que não ocorreria em modelos de processamento *offline* (KASABOV; FILEV, 2006). Sendo assim, sistemas *fuzzy* evolutivos apresentam-se como uma alternativa viável para o processamento ininterrupto de dados e aprendizagem contínua.

Uma vez percorridos os conceitos de sistemas *fuzzy* evolutivos, as seções subsequentes tratam dos três sistemas utilizados como ferramentas neste trabalho: ALMMo (ANGELOV; GU; PRINCIPE, 2017), eMG (LEMOS; CAMINHAS; GOMIDE, 2011) e eNFN (SILVA et al., 2014).

### 2.5.1 ALMMo - Autonomous Learning Multi-Model System

O ALMMo, proposto em Angelov, Gu e Principe (2017), é um sistema *fuzzy* evolutivo multi-modelo com regras *fuzzy* do tipo AnYa (ANGELOV; YAGER, 2012). O antecedente das regras AnYa é baseado no conceito de nuvens de dados (*data clouds*). O princípio das nuvens de dados é semelhante ao dos algoritmos de agrupamento, ou seja, cada nuvem representa um conjunto de amostras com características semelhantes. O formato de cada nuvem é moldado pelos dados que a constituem, e seus centros são representados por pontos focais. Cada nova amostra é classificada em uma nuvem de dados e esta classificação é realizada tendo como auxílio um método de análise empírica (*Empirical Data Analytics - EDA*) (ANGELOV et al., 2016). O EDA atribui, para cada amostra, um nível de densidade (*discrete density*), baseado na distância desta amostra perante às outras, classificando-as mais próximas ou distantes aos pontos focais de cada nuvem. A Figura 5 mostra a formação das nuvens de dados originada de uma base de dados real.

Figura 5 – Nuvens de dados e seus pontos focais.



Fonte: Angelov, Gu e Principe (2017).

As regras do ALMMo podem ser descritas por:

$$SE \left( j^* = \underset{i=1,2,\dots,G}{\operatorname{argmin}} \left( \|x_t - c_t^i\| \right) \right) \text{ ENTÃO } (\Xi_{j^*} \leftarrow x_t),$$

sendo  $x_t = [x_1, x_2, \dots, x_E]^T$  uma amostra de dados no espaço Euclidiano definido como  $\mathbf{R}^M$ ,  $\operatorname{argmin}$  o método que busca a nuvem de dados mais próxima,  $\Xi_i$  a  $i$ -ésima nuvem de dados e  $c_t^i$  seu ponto focal correspondente. O número de nuvens de dados é representado por  $G$  no universo observado  $\mathbf{R}^M$ , que também representa o número de regras.

O ALMMo se baseia na evolução das suas nuvens de dados, criando ou excluindo-as à medida que novas amostras são inseridas no modelo. Com a chegada de novas amostras, a média escalar dos produtos  $\bar{X}_t$  e a média global  $\mu_t$  no instante  $t$  são calculadas recursivamente por:

$$\bar{X}_t = \frac{t-1}{t} \bar{X}_{t-1} + \frac{1}{t} \|x_t\|^2 \quad (1)$$

e

$$\mu_t = \frac{t-1}{t}\mu_{t-1} + \frac{1}{t}x_t. \quad (2)$$

Em seguida, a densidade unimodal  $\bar{D}$  é calculada entre a amostra atual  $x_t$  e todos os pontos focais identificados da  $i$ -ésima classe por:

$$\bar{D}_t(x_t) = \frac{1}{1 + \frac{\|x_t - \mu_t\|^2}{\bar{X}_t - \|\mu_t\|^2}}. \quad (3)$$

A cada amostra  $x_t$  a estrutura do ALMMo é avaliada e poderá criar uma regra, unir duas regras, excluir uma regra ou atualizar os parâmetros de uma regra. Assim, a cada amostra recebida, verifica-se, pela Condição 1, se uma nova regra deve ser criada.

$$\begin{aligned} \text{Condição 1: } SE \quad & (\bar{D}_{t+1}(x_{t+1}) > \max_{i=1,2,\dots,G_t} (\bar{D}_{t+1}(c_t^i))) \\ \text{OU} \quad & (\bar{D}_{t+1}(x_{t+1}) < \min_{i=1,2,\dots,G_t} (\bar{D}_{t+1}(c_t^i))). \end{aligned}$$

Caso a Condição 1 não seja aceita, a amostra é atribuída a nuvem de dados mais próxima que tem seus parâmetros atualizados. Caso contrário, se a Condição 1 for satisfeita, cria-se uma regra com o ponto focal em  $x_t$  e testa-se a Condição 2. Esta condição verifica se a nuvem criada se sobrepõe a alguma das nuvens existentes.

$$\text{Condição 2: } SE \quad (\bar{D}_{t+1,i} \geq (\frac{1}{1+n^2})).$$

Caso a Condição 2 seja satisfeita, a nuvem sobreposta será substituída pela recém-criada e a nova nuvem herdará os parâmetros do consequente da nuvem sobreposta. Caso contrário, ou seja, a Condição 2 não é satisfeita, os parâmetros do consequente da nova nuvem são inicializados.

Em seguida, a qualidade da base de regras é avaliada por uma medida de utilidade que visa excluir as regras pouco representativas. A medida de utilidade é calculada com base na soma acumulada da contribuição da regra para o cálculo da saída desde o momento de sua criação até a amostra atual. Em outras palavras, é a medida de importância de uma regra *fuzzy* em relação às outras regras ( $i = 1, 2, \dots, G_{t+1}$ ). Seu cálculo é obtido por:

$$\eta_{t+1}^i = \frac{1}{t+1-ativada^i} \sum_{l=ativada^i}^{t+1} \delta_l^i, \quad (4)$$

em que  $ativada^i$  representa o tempo em que a regra/nuvem de dados foi ativada e  $\delta_l^i$  é o nível de ativação da  $i$ -ésima regra/nuvem de dados no instante  $l$ . O nível de ativação  $\delta_l^i$  é calculado para cada regra ( $i = 1, 2, \dots, G_l$ ) em:

$$\delta_l^i = \frac{\overline{D}_l^i(x_l)}{\sum_{j=1}^{G_l} \overline{D}_l^i(x_l)}. \quad (5)$$

Seja  $\eta_0$  uma constante que define um valor de tolerância, as regras com baixa representatividade serão excluídas conforme a Condição 3. Caso a Condição 3 seja atendida, a  $j$ -ésima regra/nuvem de dados será excluída juntamente com os seus parâmetros do consequente.

$$\text{Condição 3: } SE \quad (\eta_t^j + 1 < \eta_0).$$

Além da exclusão das regras, o ALMMo implementa um método para exclusão de entradas inter-correlacionadas visando melhorar o tempo de processamento, reduzir o uso de memória e melhorar o desempenho geral do algoritmo. O método proposto é baseado na soma acumulada normalizada dos valores dos parâmetros da entrada.

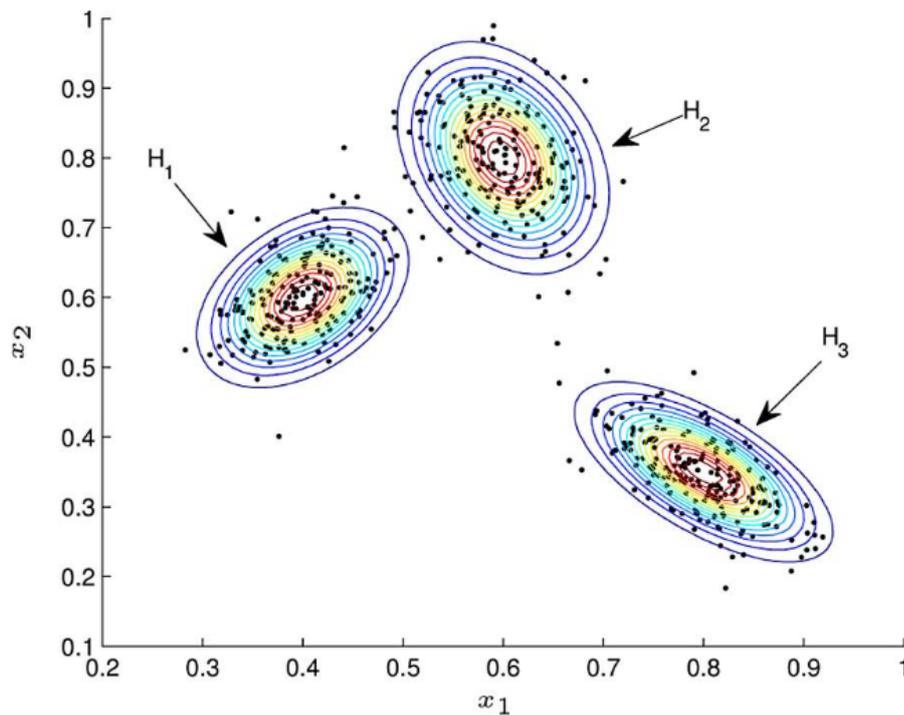
### 2.5.2 eMG - evolving Gaussian Multivariable Fuzzy System

O eMG (LEMOS; CAMINHAS; GOMIDE, 2011) é um modelo *fuzzy* evolutivo que utiliza regras de Takagi-Sugeno de primeira ordem e funções de pertinência Gaussiana multivariável. A base de regras é construída sob um algoritmo de agrupamento evolutivo com aprendizado participativo. A característica principal da aprendizagem participativa é utilizar o que já foi aprendido para avaliar o impacto da inclusão de uma nova amostra no modelo. Resumindo, a relevância de uma nova amostra é avaliada perante o que já foi consolidado como conhecimento (YAGER, 1990). O antecedente das regras é representado por funções de pertinência gaussianas multivariáveis  $H$ , definidas por:

$$H(x) = \exp \left[ -\frac{1}{2}(x - c)\Sigma^{-1}(x - c)^T \right], \quad (6)$$

onde  $x$  é um vetor de entrada,  $\Sigma$  uma matriz simétrica positiva-definitiva,  $c$  é um vetor com valores modais (centros) e representa um elemento presente em  $H(x)$  e  $\Sigma$  representa o nível de dispersão de  $H(x)$ . Tanto  $c$  quanto  $\Sigma$  são parâmetros que fazem parte da função de pertinência, sendo associados ao centro e dispersão da função, respectivamente (LEMOS; CAMINHAS; GOMIDE, 2011). A Figura 6 ilustra a representação de três grupos, definidos respectivamente como  $H_1$ ,  $H_2$  e  $H_3$ .

Figura 6 – Representação de três *clusters* no eMG.



Fonte: Lemos, Caminhas e Gomide (2011).

A estrutura do eMG evolui pela atualização, criação ou união de grupos/regras e a saída é obtida pela média ponderada das contribuições de cada regra. As regras são representadas pelos seus respectivos grupos e, a cada nova amostra apresentada ao eMG, sua estrutura é atualizada por uma medida de compatibilidade computada para cada grupo. A medida de compatibilidade  $p_t^i \in [0, 1]$  é calculada pela distância entre entrada atual  $x_t$  e os centros dos grupos existentes  $c_t^i$ , ou seja,  $p_t^i$  é obtida por:

$$p_t^i = \exp \left[ -\frac{1}{2} M(x_t, c_t^i) \right], \quad (7)$$

na qual a distância  $M$  é calculada por:

$$M(x_t, c_t^i) = (x_t - c_t^i) \left( \sum_i^t \right)^{-1} (x_t - c_t^i)^T, \quad (8)$$

em que  $x_t$  representa a amostra atual e  $c_t^i$  o centro do  $i$ -ésimo grupo. Define-se um limiar para a medida de compatibilidade  $T_p$  que é obtido por:

$$T_p = \exp \left[ -\frac{1}{2} \chi_{m,\alpha}^2 \right], \quad (9)$$

em que  $\chi_{m,\alpha}^2$  é uma distribuição *Qui – Quadrado* com  $m$  graus de liberdade e  $\alpha$  o intervalo de confiança unilateral. Após o cálculo da medida de compatibilidade entre a nova amostra e todos os grupos, encontra-se o índice do grupo com maior grau de compatibilidade  $i^*$ .

Um mecanismo de alerta é utilizado para identificar quando a estrutura dos grupos atuais não representa adequadamente o conhecimento atual do sistema e este necessita de revisão (LEMOS; CAMINHAS; GOMIDE, 2011). A cada nova amostra  $x_t$  apresentada ao modelo, o índice de alerta  $a_t^i \in [0, 1]$  é calculado para todos os grupos e estimado por meio da probabilidade cumulativa de  $V_t$ , isto é:

$$a_t^i = \bar{p}(V_t < \bar{z}), \quad (10)$$

em que  $\bar{p}(V_t = \bar{z})$  é uma distribuição binomial. Define-se também um limiar do índice de alerta  $T_a$  que é utilizado na criação de grupos. O limiar  $T_a$  é calculado por:

$$T_a = 1 - \frac{\alpha}{\omega}, \quad (11)$$

sendo  $\alpha$  um parâmetro que define o nível de significância e  $\omega$  a janela de observações para o cálculo do índice de alerta.

Um novo grupo é criado quando a medida de compatibilidade  $p_t^i$  é menor que o limiar de compatibilidade  $T_p$  para todos os grupos e o índice de alerta para o grupo com o maior grau de compatibilidade  $a_t^{i^*}$  é maior do que seu respectivo limiar  $T_a$ . Em outras palavras, um novo grupo será criado se

$$p_t^i < T_p, \forall i = 1, \dots, c_t^i \text{ and } a_t^{i^*} > T_a \text{ for } i^* = \max_i(p_t^i).$$

Caso contrário, se

$$p_t^i > T_p, \forall i = 1, \dots, c_t^i \text{ and } a_t^{i^*} < T_a \text{ for } i^* = \max_i(p_t^i),$$

então, a amostra é atribuída ao grupo mais compatível, que tem seu centro atualizado usando:

$$c_{t+1}^{i*} = c_t^{i*} + \lambda(p_t^{i*})^{1-a_t^{i*}} (x_t - c_t^{i*}), \quad (12)$$

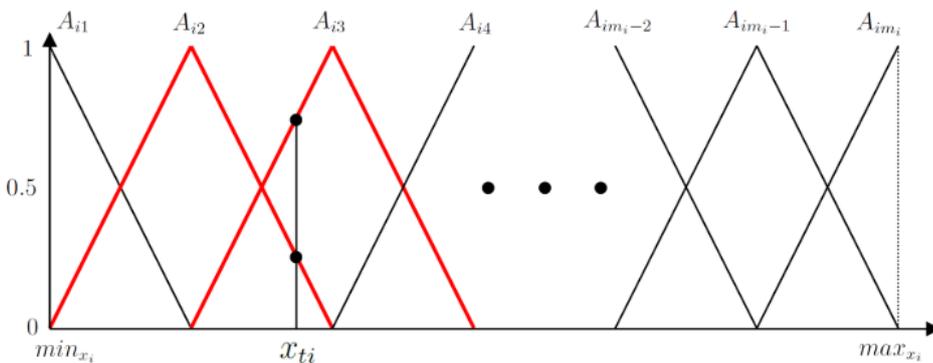
em que  $\lambda \in [0, 1]$  é a taxa de aprendizado. Os parâmetros do consequente do eMG são atualizados por algoritmo recursivo de mínimos quadrados ponderados.

### 2.5.3 eNFN - evolving Neo-Fuzzy Neuron

O eNFN (SILVA et al., 2014) é uma abordagem evolutiva para a rede *Neo-Fuzzy-Neuron* (NFN) proposta por Yamakawa (1992). O modelo possui uma estrutura composta por  $n$  modelos de Takagi-Sugeno de ordem zero, sendo um para cada variável de entrada.

O domínio das variáveis de entrada é particionado por funções de pertinência triangulares e complementares. Sendo assim, independente do número de funções de pertinência, para cada amostra, apenas duas funções de pertinência são ativadas para cada variável de entrada, como pode ser visualizado na Figura 7, na qual as funções de pertinência  $A_{i2}$  e  $A_{i3}$  são ativadas pela amostra  $x_{ti}$ .

Figura 7 – Funções de pertinência triangulares e complementares.



Fonte: Silva (2014).

O eNFN inicia seu aprendizado com duas funções de pertinência para cada variável de entrada e novas funções podem ser adicionadas com base na relação entre o erro médio global do modelo e o erro médio local da função de pertinência mais ativa. O valor médio  $\hat{\mu}_t$  e a variância do erro global  $\hat{\sigma}_t^2$  são recursivamente calculados para a entrada  $x_t$  no tempo  $t$  por:

$$\hat{\mu}_t = \hat{\mu}_{t-1} - \beta (\hat{\mu}_{t-1} - e_t) \quad (13)$$

e

$$\hat{\sigma}_t^2 = (1 - \beta) (\hat{\sigma}_{t-1}^2 + \beta (\hat{\mu}_t - e_t)^2), \quad (14)$$

em que o erro no tempo  $t$  é definido por  $e_t = y_t - \hat{y}_t$  e  $\beta$  é a taxa de aprendizagem. O valor médio do erro local  $\mu_{b_{ii}^*}$  que corresponde à função de pertinência mais ativa  $b_i^*$  é calculado recursivamente para a entrada  $x_{ti}$  por:

$$\mu_{b_{ii}^*} = \mu_{b_{ii-1}^*} - \beta (\mu_{b_{ii-1}^*} - e_t). \quad (15)$$

Um limitador  $\tau$  é utilizado para permitir o controle sobre o número de regras, evitando modelos complexos e *overfitting*. No eNFN, este limitador é comparado com a menor distância (*dist*) permitida entre o valor modal da função a ser criada e as funções adjacentes. Uma nova função de pertinência será criada e inserida no modelo se

$$\mu_{b_{ii}^*} > \hat{\mu}_t + \hat{\sigma}_t^2 \text{ e } dist > \tau.$$

Em contrapartida, a exclusão de funções de pertinência segue o conceito de idade, isto é, uma função será excluída do modelo se ficar muito tempo inativa (LUGHOFER; ANGELOV, 2011). A idade de uma determinada função de pertinência é calculada por:

$$idade_j = t - ativada_j, \quad (16)$$

em que  $t$  é o intervalo de tempo atual e  $ativada_j$  representa o intervalo onde ocorreu a ativação da  $j$ -ésima função de pertinência. Considere  $b_i^-$  o índice da função de pertinência menos ativa no tempo  $t$  que será excluída se

$$b_i^- > \omega,$$

em que  $\omega$  é um parâmetro que indica o limite de tempo para exclusão de uma função.

# Capítulo 3

## Trabalhos Relacionados

Este capítulo apresenta um resumo dos principais trabalhos da literatura, relacionados ao contexto da oferta e demanda de passageiros. De modo geral, serão apresentadas as principais características já abordadas nestes trabalhos, que tangem: Previsão com Análise Temporal (Seção 3.1); Previsão com Inclusão de Variáveis Exógenas (Seção 3.2); e Combinações de Técnicas Espaciais e Temporais para a Previsão (Seção 3.3). Por último, na Seção 3.4, é apresentado um compêndio dos 30 principais trabalhos pesquisados nesse estudo e seus principais objetivos.

### 3.1 Previsão com Análise Temporal

Como descrito em seções pregressas, conseguir prever potenciais clientes que necessitam de transportes privados com determinado nível de confiança é um dos objetivos de estudos em ITS. Para endereçar esse problema, trabalhos na literatura concentram seus esforços para entender como modelar possíveis soluções. Técnicas de mineração de dados são utilizadas para criar conhecimento a partir de bancos de dados diversos. Resultados iniciais indicaram que a quantificação de passageiros era mais bem distribuída por intervalos de horário, e que estratificar uma cidade ou região em microrregiões (zonas) ajudou a entender as peculiaridades em cada uma (LI et al., 2011). A influência do trânsito em determinados horários do dia, finais de semana e determinados dias da semana foram identificados como fatores influentes como estratégias de taxistas para conseguir passageiros (ZHANG et al., 2015).

Com o avanço das pesquisas, os conceitos de agregação da solicitação de passageiros por intervalo de tempo são consolidados e novas abordagens surgem para realizar previsão da demanda. Em Moreira-Matias et al. (2012), é criado um modelo estatístico híbrido para prever a demanda de passageiros para os próximos 30 minutos em 63 estandes na cidade de Porto (Portugal). Adicionalmente, o modelo de Moreira-Matias et al. (2012) é modificado em Moreira-Matias et al. (2013) para simular a previsão da demanda em tempo real. Zander

(2017) prevê a demanda de táxis na cidade de Estocolmo (Suécia), utilizando Redes Neurais Artificiais (*Artificial Neural Networks*). Em [Jiang et al. \(2019\)](#), são criados modelos baseados em *Support Vector Machine* (SVM) para prever demandas nos próximos 5, 10 e 15 minutos na China.

Intuitivamente, presume-se que a previsão da demanda visa descobrir o quantitativo efetivo de solicitações de corridas (*Pick-Up*) em determinado horário em tempos futuros. Porém, foi descoberto que entender o padrão dos destinos de chegadas (*Drop-Off*) também agrega conhecimento. Uma vez que se sabe quantos táxis chegaram uma determinada região, é possível saber de antemão o efetivo de veículos que estarão disponíveis, possibilitando que as empresas de táxis possam despachá-los para locais onde há necessidade de serviços ([YANG; GONZALES, 2017](#)). Entende-se, então, que tanto a previsão de *Pick-Up* quanto a de *Drop-Off* faz-se relevante.

## 3.2 Previsão com Inclusão de Variáveis Exógenas

Em alguns trabalhos, é possível encontrar autores pesquisando como fatores externos afetam o resultado da demanda. A temperatura, condição do tempo, humidade do ar e horário do dia são exemplos destes fatores ([KE et al., 2017](#)). No trabalho de [Liao et al. \(2018\)](#), variáveis de entrada como velocidade do vento, quantidade de neve e temperatura são utilizadas para prever a demanda em Nova Iorque (Estados Unidos da América), testando o desempenho de dois modelos: *Fusion Convolutional Long Short-Term Memory Network* (FCL-Net) e *Spatio-Temporal Residual Network* (ST-ResNet). Já em [Xu et al. \(2018\)](#), são testadas diferentes combinações de características (dados meteorológicos, *Pick-Up*, *Drop-Off*, dia da semana) para prever a demanda no intervalo de 60 minutos. Um modelo apelidado de *Multi-Level Recurrent Neural Networks* (MLRNN) baseado em *Long Short-Term Memory* (LSTM) é tema do trabalho de [Zhang et al. \(2021\)](#). O modelo recebe, como entrada, informações meteorológicas e dados históricos para prever demandas nos próximos 30 minutos em Nova Iorque.

Outros autores analisam o impacto de pontos de grande interesse (estádios, hospitais, etc.) e eventos (shows, apresentações, jogos) no tráfego da região. O estudo de [Markou, Rodrigues e Pereira \(2018\)](#) reúne informações de eventos realizados em duas áreas populares de Nova Iorque (Terminal 5 e *Barclays Center*) e inclui estes eventos como variáveis de entrada em seus modelos, realizando a predição da demanda nestas duas áreas. Ainda neste contexto, outras abordagens exploram a previsão de táxis em regiões de grande interesse, como aeroportos, hospitais, locais turísticos, metrô, etc ([VANICHRUJEE et al., 2018](#)). [Liu et al. \(2020a\)](#) criam modelo apelidado de *Context-Aware Attention-Based Convolutional Recurrent Neural Network* (CACRNN), combinando pontos de interesse (*Points Of Interest* - POI) com dados meteorológicos, para prever demandas nos próximos 15 minutos em

Chengdu (China) e Nova Iorque. Nos estudos de [Zhu et al. \(2021\)](#), um *framework* é proposto combinando Multilayer Perceptron (MLP) e Long Short-Term Memory (LSTM) para capturar as dependências temporais das corridas. Em concomitância, camadas convolucionais são utilizadas para interpretar dados de eventos e convertê-los em informação útil para prever a demanda em Nova Iorque, em dois centros onde ocorrem grandes eventos na cidade (*Barclays Center* e Terminal 5). Em [Liu et al. \(2020b\)](#), é desenvolvido um modelo com a combinação das técnicas *Random Forest* e *Ridge Regression* para prever a demanda de passageiros em pontos de grande interesse em Xian (China) para o intervalo de 1 hora.

Na literatura, é possível encontrar trabalhos que reúnem informações de clientes por meio de aplicativos móveis. Estes aplicativos informam por triangulação GPS a sua localização. Estes dados são usados como entrada para aumentar a acurácia dos modelos. Em [Ishiguro, Kawasaki e Fukazawa \(2018\)](#), é conduzido um experimento utilizando dados de aplicativos móveis de clientes, para prever a demanda para os próximos 30 minutos na cidade de Tóquio (Japão). Os taxistas que participaram do experimento conseguiram um lucro médio de 3,9% maior em relação aos que não participaram.

### 3.3 Combinações de Técnicas Espaciais e Temporais para a Previsão

Algumas pesquisas sugerem a criação de modelos que unificam informações temporais e espaciais, determinados como *Spatiotemporal models*. A análise de mapas (espacial) combinadas com dados históricos de corridas (temporal) cria modelos híbridos mais robustos, capazes de prever a demanda de passageiros com mais acurácia. Alguns modelos, como os de [Yao et al. \(2018\)](#) e [Liu et al. \(2019\)](#), agregam a demanda de táxi em imagens convertidas em mapas de calor e as processa com *Convolutional Neural Networks* (CNN). A saída da CNN serve de entrada a um segundo modelo que prevê a demanda de passageiros. O trabalho de [Liao et al. \(2018\)](#) também apresenta uma proposta semelhante, agregando análise de mapas, dados meteorológicos e informações históricas em um *framework* para testar vários modelos na literatura com o intuito de prever a demanda em Nova Iorque. O trabalho de [Lin et al. \(2021\)](#) utiliza dados climáticos, médias móveis e dados históricos de corridas para alimentar um modelo que utiliza três técnicas de *deep learning*: *Convolutional Neural Networks* (CNN), *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU). Neste estudo, são previstas demandas em pontos de interesse (local de atração turística, hospitais, etc.) na cidade de Kaohsiung (Taiwan). Outros trabalhos compartilham abordagens semelhantes as supracitadas, como os de [Faghieh et al. \(2019\)](#) e [Kong et al. \(2020\)](#).

A relação entre as regiões também faz parte de estudos na literatura. Regiões mais próximas

tendem a compartilhar padrões de demanda semelhantes. Zhang, Liu e Zheng (2019) propõem um seletor de características nomeado *SpatioTemporal Dynamic Time Warping - ST-DTW*, baseado na distância euclidiana entre as zonas. Este método é comparado a outros seletores de características disponíveis na literatura. Wang et al. (2019) apresentam uma proposta agregando pesos a cada região, isto é, regiões mais próximas entre si ganham um peso maior e, conseqüentemente, são classificadas diferentemente pelo modelo. Diferente de outros trabalhos, a previsão da demanda nesta abordagem é feita utilizando uma tupla de origem-destino, isto é, a quantidade de corridas que saíram de determinada região X e tiveram seu destino em determinada região Y. Por outro lado, o estudo de Bai et al. (2019) mede similaridades entre as zonas por meio da análise de preferências dos usuários. Sendo assim, locais com temáticas semelhantes tendem a possuir padrões de demanda semelhantes. Em Luo et al. (2021), é utilizado um seletor de variáveis baseado em um método estatístico conhecido como *Augmented Dickey-Fuller (ADF)*. Após a seleção, o modelo *Multi-Task Deep Learning (MTDL)* é empregado para a previsão da demanda nos próximos 10 minutos em Nova Iorque.

Uma combinação entre modelos de classificação e regressão pode ser visto no trabalho de Zhang et al. (2020), que apresenta duas abordagens: uma baseada em classificação e outra em previsão de séries temporais. Os autores entendem que a demanda de passageiros pode ser transformada em um problema de classificação, isto é, cada demanda é classificada por um rótulo (a demanda de 00:00 até 02:00 é classificada como 1, a de 02:00 até 04:00 como 2, etc). Este classificador serve como variável de entrada a um modelo de regressão que prevê demandas de táxis em intervalos de 30 minutos, na cidade de Nova Iorque.

Dados de empresas concorrentes podem auxiliar a incrementar a acurácia dos modelos. Isso pode ser visto no estudo de Zhao et al. (2020), em que propuseram um *framework* de nome *Unified Spatial-Temporal Network (USTN)*, utilizando mapas processados em camadas convolucionais, juntamente com dados históricos da Uber<sup>1</sup> e informações dos táxis amarelos de Nova Iorque para prever a demanda em intervalos de 1 hora.

### 3.4 Resumo dos Trabalhos

De forma complementar, esta seção apresenta um resumo de 30 dos principais estudos analisados neste trabalho, relacionados ao contexto de previsão da demanda de passageiros. O quadro 1 apresenta, de forma cronológica, os autores, o ano de publicação, as tecnologias utilizadas, os principais objetivos e as cidades onde a previsão foi realizada.

---

<sup>1</sup> Empresa de corridas privadas (<https://www.uber.com>).

Quadro 1 – Resumos dos trabalhos da literatura.

<b>Autor</b>	<b>Modelos e Técnicas</b>	<b>Objetivos</b>	<b>Cidades de Estudo</b>
Ke et al. (2017)	<i>framework</i> próprio - FCL-net	Previsão da demanda em intervalos de 1 hora	Hangzhou - China
Tong et al. (2017)	Modelo próprio de regressão linear (LinUOTD)	Previsão da demanda em intervalo de 1 hora	Pequim e Hangzhou - China
Yang e Gonzales (2017)	Regressão Poisson	Criação de um modelo para identificar padrões de oferta e demanda	Nova Iorque - Estados Unidos da América
Davis, Raina e Jagannathan (2018)	Tesselação Voronoi, Geohash e K-Means	Criação de duas matrizes para otimizar a previsão da demanda	Nova Iorque - Estados Unidos da América
Xu et al. (2018)	<i>Long Short-Term Memory</i> - LSTM	Previsão da demanda em intervalos de 10, 20 e 60 minutos	Nova Iorque - Estados Unidos da América
Markou, Rodrigues e Pereira (2018)	Processos Gaussianos	Previsão da demanda em intervalos de 1 hora	Nova Iorque - Estados Unidos da América
Liao et al. (2018)	<i>Residual Neural Networks</i> - ResNet	Previsão da demanda em intervalos de 1 hora	Nova Iorque - Estados Unidos da América
Vanichrujee et al. (2018)	Modelo combinado com GRU LSTM e XGBoost	Previsão da demanda em intervalos de 1 hora	Bancoque - Tailândia
Yao et al. (2018)	<i>Long Short-Term Memory</i> - LSTM	Previsão da demanda em intervalos de 30 minutos	Guangzhou - China
Ishiguro, Kawasaki e Fukazawa (2018)	<i>Convolutional Neural Networks</i> - CNN	Previsão da demanda em intervalos de 30 minutos	Tóquio - Japão
Jiang et al. (2019)	<i>Support Vector Machine</i> - SVM	Previsão da demanda em intervalos de 10 minutos	Cidade não informada - China.
Bai et al. (2019)	<i>Convolutional Recurrent Neural Networks</i> - CRNN	Previsão da demanda em intervalos de 1 hora	Shenyang - China
Wang et al. (2019)	<i>Graph Convolutional Networks</i> - GCN	Previsão da demanda por matrizes de origem e destino em intervalos de 1 hora	Chengdu e Pequim - China
Zhou et al. (2019)	ST-Ann - Encoders e Decoders	Previsão da demanda em intervalos de 30 minutos e 60 minutos	Pequim - China e Nova Iorque - Estados Unidos da América
Yu et al. (2019)	<i>Density-Based Spatial Clustering of Applications with Noise</i> - DBSCAN	Previsão da demanda em intervalos de 1 hora	Pequim - China

Quadro 1 – Resumos dos trabalhos da literatura (continuação).

<b>Autor</b>	<b>Modelos e Técnicas</b>	<b>Objetivos</b>	<b>Cidades de Estudo</b>
Liu et al. (2019)	<i>Convolutional Long Short-Term Memory</i> Conv-LSTM	Previsão da demanda em intervalos de 30 minutos	Nova Iorque - Estados Unidos da América
Faghni et al. (2019)	<i>Vector Auto Regression - VAR</i> <i>Least Absolute Shrinkage and Selection Operator - LASSO</i>	Previsão da demanda em intervalos de 15 minutos	Nova Iorque - Estados Unidos da América
Rossi et al. (2019)	<i>Long Short-Term Memory - LSTM</i>	Previsão da trajetória de possíveis clientes de táxi	São Francisco e Nova Iorque - Estados Unidos da América Porto - Portugal
Zhang, Liu e Zheng (2019)	<i>Recurrent Neural Networks - RNN</i> e <i>Dynamic Time Warping - DTW</i>	Previsão da demanda em 15 minutos	Chengdu - China e Nova Iorque - Estados Unidos da América
Kong et al. (2020)	framework próprio (TBIZFlow)	Previsão do tráfego de passageiros em 30 minutos	Xangai, China
Xie et al. (2020)	Nenhuma	Revisão bibliográfica sobre previsão de tráfego e demanda	Nenhuma
Zhang et al. (2020)	<i>Long Short-Term Memory - LSTM</i>	Previsão da demanda em 30 minutos	Nova Iorque - Estados Unidos da América
George e Santra (2020)	<i>Long Short-Term Memory - LSTM</i> <i>Fuzzy Neural Networks</i>	Previsão de velocidade média do tráfego em rodovias, em intervalos de 5 minutos	São Diego - Estados Unidos da América e Pequim - China
Zhao et al. (2020)	<i>Unified Spatial-Temporal Network (USTN)</i>	Previsão da demanda em intervalos de 1 hora	Nova Iorque - Estados Unidos da América
Liu et al. (2020a)	<i>Context-Aware Attention-Based Convolutional Recurrent Neural Network - CACRNN</i>	Previsão da demanda em intervalos de 15 minutos	Chengdu - China e Nova Iorque - Estados Unidos da América
Liu et al. (2020b)	<i>Random Forest</i> e <i>Ridge Regression</i>	Previsão de corridas em intervalos de 1 hora	Xian - China
Luo et al. (2021)	<i>Multi-Task Deep Learning</i>	Previsão da demanda em intervalos de 10 minutos	Nova Iorque - Estados Unidos da América
Zhang et al. (2021)	<i>Multi-Level Recurrent Neural Networks (MLRNN)</i>	Previsão da demanda em intervalos de 30 minutos	Nova Iorque - Estados Unidos da América
Zhu et al. (2021)	<i>Multilayer Perceptron - MLP</i> e <i>Long Short-Term Memory LSTM</i>	Previsão da demanda em intervalos de 30 minutos	Nova Iorque - Estados Unidos da América
Lin et al. (2021)	<i>Convolutional Neural Networks - CNN</i> e <i>Long Short-Term Memory - LSTM</i>	Previsão da demanda em intervalos de 30 minutos	Kaohsiung - Taiwan

# Capítulo 4

## Abordagem Proposta

Este capítulo descreve a abordagem proposta para realizar a previsão e a classificação da demanda de passageiros de táxi. Na Seção 4.1, são detalhadas as etapas necessárias para realizar a previsão da demanda. A Seção 4.2 descreve a metodologia utilizada na classificação da demanda.

### 4.1 Previsão da Demanda

A abordagem proposta para previsão da demanda de táxis é ilustrada na Figura 8. Como pode ser visto, a abordagem é composta por 4 etapas. A primeira etapa é a extração da base de corridas que é detalhada na Subseção 4.1.1. Na etapa seguinte, descrita na Subseção 4.1.2, tem-se a caracterização das variáveis de entrada e a construção dos conjuntos de dados. A Subseção 4.1.3 detalha a terceira etapa que é a seleção de variáveis. Por fim, a quarta e última etapa é a previsão da demanda ilustrada na Subseção 4.1.4.

#### 4.1.1 Extração da Base de Corridas

As bases de corridas são conjuntos de dados que contêm informações das corridas de táxi realizadas em determinada cidade ou região, disponibilizadas por empresas que atuam no ramo. Geralmente, estas informações são extraídas dos sistemas de monitoramento instalados nos carros (dispositivos GPS e aplicativos móveis próprios) e apresentadas no seguinte formato: data e horário de entrada, data e horário de chegada, local onde ocorreu a entrada e local onde ocorreu a chegada. Define-se como entrada (do inglês, *Pick-Up*), o instante de início da corrida e a chegada (do inglês, *Drop-Off*) seu instante de término. Adicionalmente, essas bases podem fornecer informações complementares como a trajetória da corrida, o valor pago, o tipo de pagamento (cartão, dinheiro, etc.), número de passageiros, tipo do veículo, dentre outros. A Tabela 1 apresenta um exemplo de dados comumente encontrados nessas bases.

Figura 8 – Etapas da abordagem proposta para a previsão da demanda.

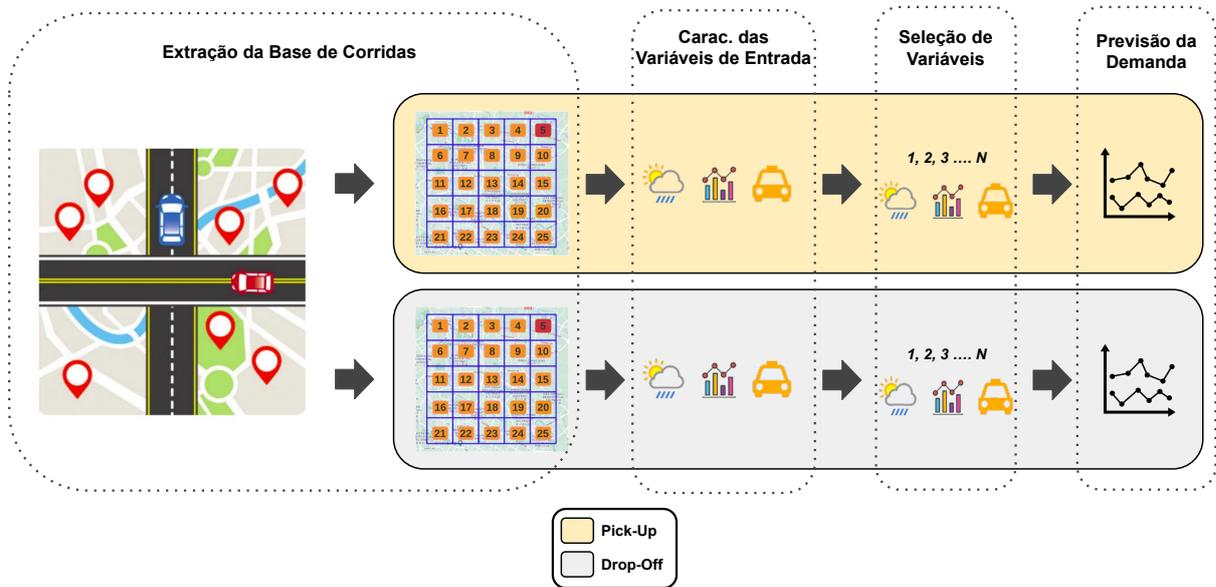


Tabela 1 – Representação dos dados em uma base de corridas.

<i>Pick-Up</i>	<i>Pick-Up Latitude</i>	<i>Pick-Up Longitude</i>	<i>Drop-Off</i>	<i>Drop-off Latitude</i>	<i>Drop-Off Longitude</i>
20/02/2021 - 14:05	43,15220	15,25452	20/02/2021 - 14:30	43,17525	15,30225
21/02/2021 - 15:40	43,10252	15,50025	21/02/2021 - 16:00	43,12345	15,60021
21/02/2021 - 18:26	43,80025	16,12450	21/02/2021 - 19:00	43,60014	16,40042
22/02/2021 - 09:20	44,10548	17,66521	21/02/2021 - 11:15	42,60151	17,98452
21/02/2021 - 13:42	43,80025	16,12500	21/02/2021 - 14:00	43,60014	16,40042

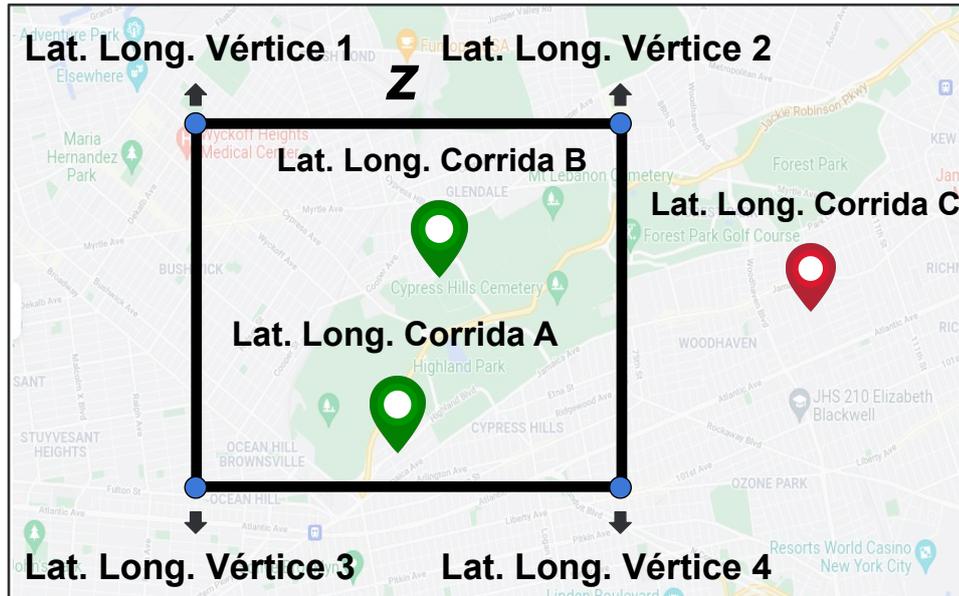
Após a obtenção da base de corridas, é realizada a extração das informações que serão utilizadas nas próximas etapas. Assim, são selecionadas as seguintes informações: data e horário de *Pick-Up*, data e horário de *Drop-Off*, local de ocorrência de *Pick-Up* e local de ocorrência de *Drop-Off*.

Em seguida, a região de interesse é dividida em microrregiões, denominadas zonas. Esta divisão tem como intuito enquadrar a região de estudos em limites pré-definidos e agrupar as localizações das corridas. Na literatura, é possível encontrar diversas técnicas para a divisão em zonas. Este trabalho focou em algumas destas discorridas na Seção 2.2 para realizar a divisão.

Concluída a divisão da região de interesse em zonas, as localizações (longitude e latitude de *Pick-Up* e *Drop-Off*) são convertidas para o seu respectivo número de zona. Considerando uma divisão em zonas por polígonos, um par de coordenadas será pertencente a uma zona  $z$ , se sua latitude e longitude estiverem dentro dos limítrofes das coordenadas de todos os vértices do polígono referente a aquela zona. A Figura 9 ilustra um exemplo de como uma coordenada pode pertencer a uma determinada zona  $z$ . As corridas A e B pertencem a  $z$ , pois suas latitudes e longitudes estão dentro dos limítrofes dos vértices de 1 a 4 da referida zona. A corrida C não pertence a  $z$ , pois sua latitude e longitude estão fora dos limítrofes

da zona. Destaca-se que a conversão das latitudes e longitudes em zonas também pode ser realizada com o auxílio das ferramentas já citadas neste trabalho para a criação dos polígonos.

Figura 9 – Representação das coordenadas de corridas dentro e fora dos limites de uma determinada zona.



Considerando um cenário em que é realizada a previsão da demanda para *Pick-Up* e *Drop-Off*, é criado, para cada uma das  $K$  zonas, um conjunto de dados de *Pick-Up* e um *Drop-Off*. Tendo como exemplo um conjunto de dados de *Pick-Up* da zona  $z$ , este conjunto é composto pela data e horário de *Pick-Up* de todas as corridas realizadas dentro dos limites da referida zona.

Uma vez determinada a qual zona cada corrida pertence e construído o conjunto de dados de cada zona, a próxima etapa é o agrupamento das corridas para caracterizar as demandas de táxi. Para isso, inicialmente, o tempo contínuo (dias e horários) é particionado em períodos sequenciais idênticos, definidos como intervalos de tempo. Em seguida, todas as corridas realizadas em um determinado intervalo de tempo são agrupadas para caracterizar a demanda de táxi para aquele intervalo. Em outras palavras, a demanda de táxi  $D$  em uma zona  $z$  é obtida pelo número de corridas que ocorreram naquela zona em um determinado intervalo de tempo. Assim, ao concluir esta etapa, tem-se para cada zona um conjunto de dados contendo os intervalos de tempo e as suas respectivas demandas de táxi.

#### 4.1.2 Caracterização das Variáveis de Entrada

Uma série temporal é caracterizada por um conjunto de observações geradas sequencialmente ao longo do tempo (BOX; JENKINS, 1976). Portanto, dada as suas características, a previsão da demanda de táxi pode ser definida como uma série temporal e esta pode

ser descrita como um problema não linear dinâmico e representada (LÜTKEPOHL, 2005; RODRIGUES; SILVA; LEMOS, 2022) por:

$$\hat{y}_t = f(y_{t-1}, \dots, y_{t-n_y}, \bar{a}_1, \dots, \bar{a}_{n_a}, \bar{h}_t, \dots, \bar{h}_{t-n_h}), \quad (17)$$

em que a previsão no passo  $t$  é realizada usando  $n_y$  defasagens para modelar a série, sendo  $t$  o passo atual,  $y$  os valores defasados da série,  $\bar{a}$  as variáveis auxiliares,  $\bar{h}$  as variáveis exógenas,  $n_a$  o número de variáveis auxiliares,  $n_h$  o número de variáveis exógenas e  $f$  uma função não linear adequada para modelar a série.

Neste trabalho, os valores passados a serem utilizados para modelar a série são caracterizados por:

- **Demanda Histórica dos Intervalos ( $dhi$ ):** demanda de táxi dos 8 últimos intervalos de tempo<sup>1</sup>. A demanda histórica dos intervalos da zona  $z$  é representada por:

$$dhi_{z,t} = dhi_{z,t-1}, dhi_{z,t-2}, \dots, dhi_{z,t-8}. \quad (18)$$

- **Demanda Histórica dos Dias ( $dhd$ ):** demanda de 8 dias anteriores<sup>2</sup>. A demanda histórica dos dias na zona  $z$  pode ser descrita como:

$$dhd_{z,tr} = dhd_{z,tr-1}, dhd_{z,tr-2}, \dots, dhd_{z,tr-8}, \quad (19)$$

para o mesmo intervalo relativo de tempo, onde  $tr$  representa o intervalo de tempo relativo do dia anterior da demanda a ser prevista.

- **Média da Demanda Histórica dos Intervalos<sup>3</sup> ( $mdhi$ ):** a média simples das 8 últimas defasagens da demanda histórica ( $dhi$ ). A  $mdhi_z$  é obtida para a zona  $z$  por:

$$mdhi_{z,t} = 1/8 \sum_{i=1}^8 dhi_{z,t-i}. \quad (20)$$

- **Média da Demanda Histórica dos Dias ( $mdhd$ ):** média simples das 8 últimas defasagens da demanda histórica dos dias anteriores ( $dhd_z$ ). Seu cálculo é representado da seguinte forma:

<sup>1</sup> A janela de tempo dos dados históricos foi definida para 8 intervalos passados como em Zhang et al. (2020), Ke et al. (2017), Yao et al. (2018).

<sup>2</sup> O número histórico escolhido foi 8, por englobar também os todos os dias da semana anterior, permitindo aos modelos o aprendizado das tendências específicas de cada dia.

<sup>3</sup> A maioria dos trabalhos identifica a média móvel de dias anteriores como média histórica, portanto, este trabalho seguirá a mesma nomenclatura.

$$mdhd_{z,tr} = 1/8 \sum_{i=1}^8 dhd_{z,tr-i}. \quad (21)$$

Assim, são geradas 18 variáveis extraídas de valores históricos de corridas (defasagens da série). Ademais, utilizando informações extraídas das corridas de táxi são obtidas as seguintes 3 variáveis auxiliares:

- **Dia da Semana** ( $ds$ ): Domingo (1), Segunda-Feira (2), ..., Sábado (7).
- **Final de Semana** ( $fs$ ): dia de semana (0) ou fim de semana (1).
- **Horário do Dia** ( $hd$ ): divide-se em três períodos: horário de descanso (0), horário normal (1) e horário de pico (2). O horário de descanso vai de 20:00 até 06:00. O horário de pico engloba os seguintes períodos: [06:00 - 09:00, 12:00 - 15:00, 17:00 - 20:00]. Os horários restantes do dia são determinados como horários normais<sup>4</sup>.

As variáveis exógenas utilizadas são obtidas de informações meteorológicas, conforme sugerido em [Ke et al. \(2017\)](#), [Tong et al. \(2017\)](#). Os dados meteorológicos utilizados foram extraídos da plataforma *Wheater Underground*<sup>5</sup>. Nesta plataforma, os dados meteorológicos são disponibilizados a cada hora, sendo inseridos no conjunto de dados no intervalo anterior à demanda a ser prevista. A lista a seguir detalha as 5 variáveis exógenas utilizadas neste trabalho:

- **Temperatura** ( $te$ ): medida em Fahrenheit.
- **Umidade Relativa do Ar** ( $ur$ ): medida em porcentagem.
- **Precipitação** ( $pr$ ): medida em milímetros.
- **Velocidade do Vento** ( $vv$ ): medida em milhas por hora (mph).
- **Condição do Tempo** ( $ct$ ): descrição textual que foi representada por índices de 1 (*Fog*) até 50 (*Windy*) (vide Apêndice A).

Totalizam-se então, para cada conjunto de dados, 26 variáveis de entrada, sendo 21 extraídas de informações contidas na base de corridas e 5 variáveis exógenas obtidas de informações meteorológicas. Assim, ao final desta etapa, tem-se para cada zona um conjunto de dados contendo os índices dos intervalos, as 26 variáveis e suas respectivas demandas de táxi.

<sup>4</sup> A divisão do horário foi baseada em resultados obtidos por observações empíricas do tráfego local das regiões, disponibilizadas por prefeituras municipais.

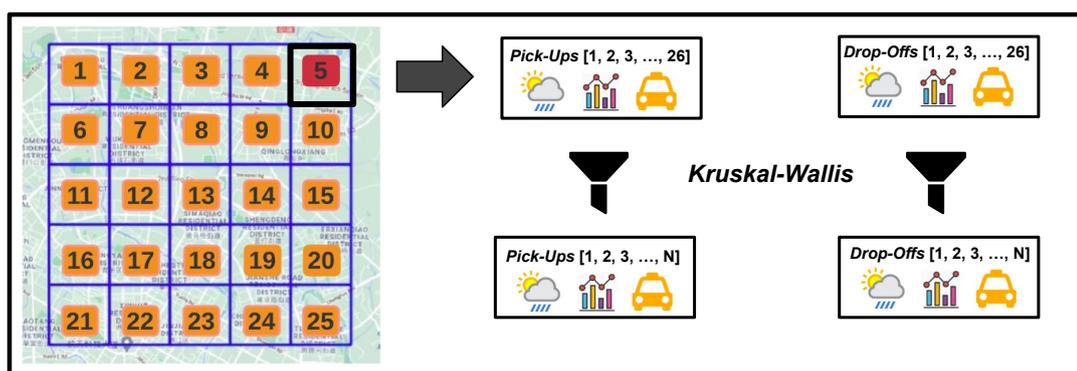
<sup>5</sup> <https://www.wunderground.com/>

### 4.1.3 Seleção de Variáveis

Concluída a criação dos conjuntos de dados com as variáveis de entrada e suas respectivas saídas desejadas (demandas de táxi) é realizada uma etapa de seleção de variáveis. Essa etapa visa identificar e selecionar, dentre as 26 variáveis de entrada, as que são relevantes para realizar a previsão da demanda. Neste trabalho, a seleção é realizada pelo método estatístico de *Kruskal-Wallis* (KRUSKAL; WALLIS, 1952) que ordena as variáveis de acordo com seu grau de relevância. A seleção de variáveis se faz necessária, pois um número elevado de dimensões apresentadas aos modelos acarreta dois principais problemas: (i) maior tempo de execução dos modelos e; (ii) maldição da dimensionalidade, termo utilizado para explicar o problema causado pelo aumento exponencial no volume de dados associado à inclusão de dimensões extras no espaço Euclidiano (BELLMAN, 1966). Em termos práticos, uma vez que o número de variáveis ultrapassa determinado limiar, os modelos provavelmente perderão desempenho, afetando conseqüentemente seus resultados.

Uma vez ordenadas, são selecionadas as  $N$  variáveis mais relevantes que comporão os novos conjuntos de dados. O método de *Kruskal-Wallis* não indica explicitamente qual valor de  $N$  é o ideal para que se obtenha o melhor desempenho. Sendo assim, este valor deve ser encontrado empiricamente ou por algum outro método. Ao término da etapa de seleção de variáveis tem-se para cada zona um conjunto de dados com os índices dos intervalos, as  $N$  variáveis mais relevantes e suas respectivas demandas de táxi. A Figura 10 ilustra as etapas desta seleção.

Figura 10 – Etapas da seleção de características.



### 4.1.4 Previsão da Demanda

Na abordagem proposta, a previsão da demanda é realizada por um sistema *fuzzy* evolutivo que realiza a previsão 1 passo à frente, isto é, para o próximo intervalo de tempo. A previsão é obtida individualmente para cada uma das zonas de *Pick-Up* e/ou *Drop-Off* utilizando como entrada o conjunto de variáveis selecionado na fase anterior.

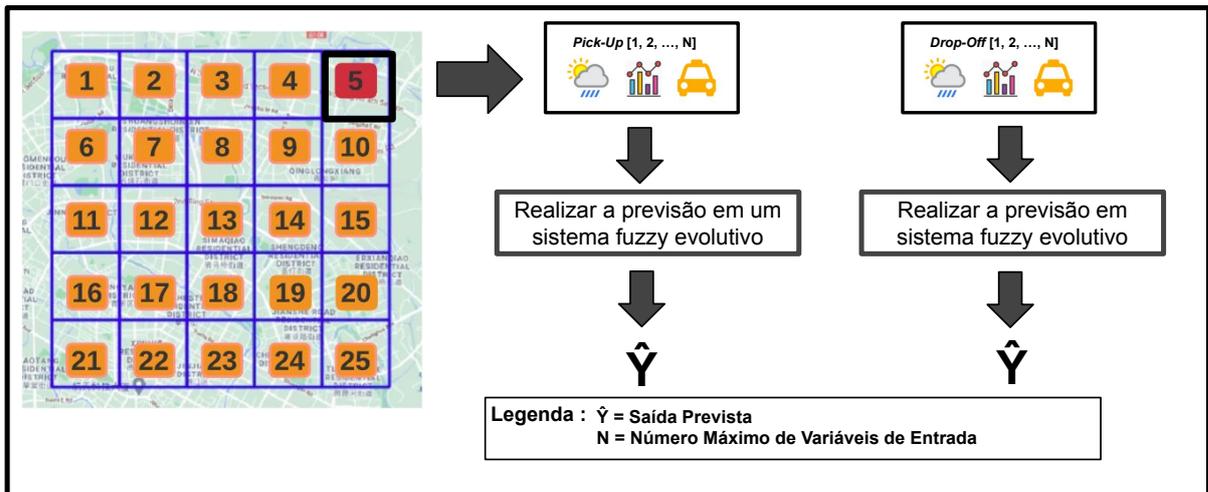
Portanto, caso as variáveis mais relevantes para um determinado conjunto de dados sejam

$dhi_t - 1, dhi_{t-2}, dhd_{tr-1}, dhd_{tr-5}, fs$  e  $hd$ , o modelo para este conjunto de dados será definido por:

$$\hat{y}_t = f(dhi_t - 1, dhi_{t-2}, dhd_{tr-1}, dhd_{tr-5}, fs, hd). \quad (22)$$

A Figura 11 descreve como a previsão da demanda é realizada.

Figura 11 – Processo para previsão da demanda.



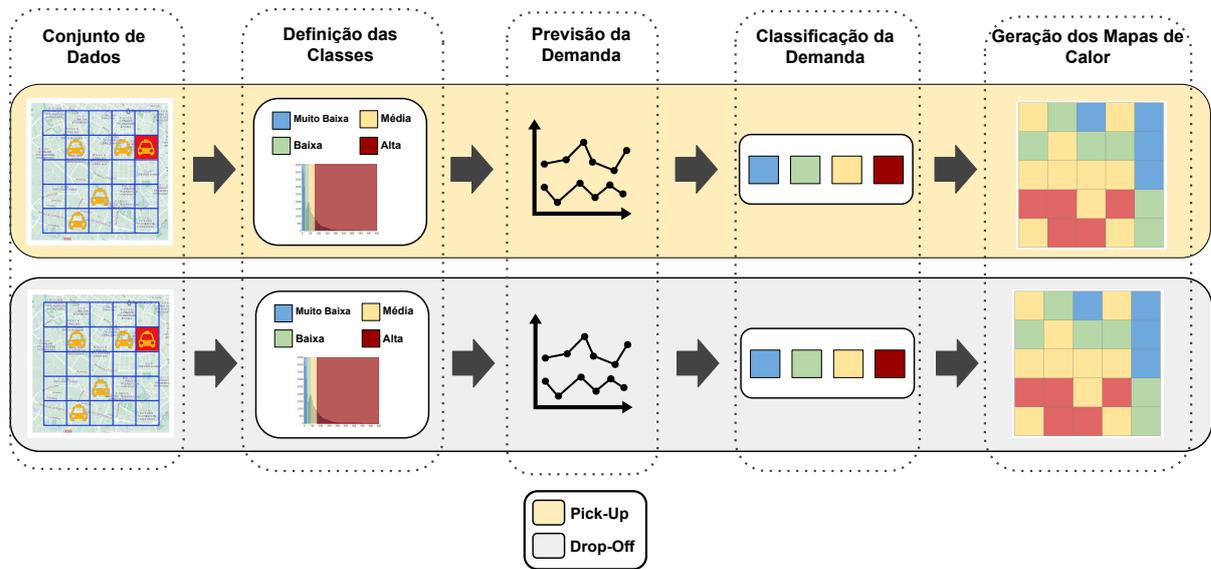
## 4.2 Classificação da Demanda

A abordagem proposta para classificação da demanda é resumida pela Figura 12, que ilustra suas principais etapas. As etapas de construção do conjunto de dados e previsão da demanda são, *mutatis mutandis*, como descritas na Seção 4.1 e, por conseguinte, não serão descritas novamente. A Seção 4.2.1 trata da definição das classes. Neste trabalho a classificação da demanda de táxi é realizada transformando a tarefa de previsão da demanda em um problema de classificação, conforme sugerido em [Salman e Kecman \(2012\)](#). Assim, a classificação é realizada utilizando as saídas da etapa de previsão. A etapa de classificação da demanda é detalhada na Seção 4.2.2. Por fim, a Seção 4.2.3 ilustra o processo de geração dos mapas de calor.

### 4.2.1 Definição das Classes

Inicialmente é definido o número de classes. Optou-se por trabalhar com 4 classes da demanda, a saber: (i) Muito Baixa; (ii) Baixa; (iii) Média e; (iv) Alta. Em seguida, é necessário especificar o domínio de cada classe, isto é, a faixa de valores que determina seus limites inferior e superior. É desejável que estes sejam definidos, para cada classe, mantendo o

Figura 12 – Etapas da abordagem proposta para a classificação da demanda.



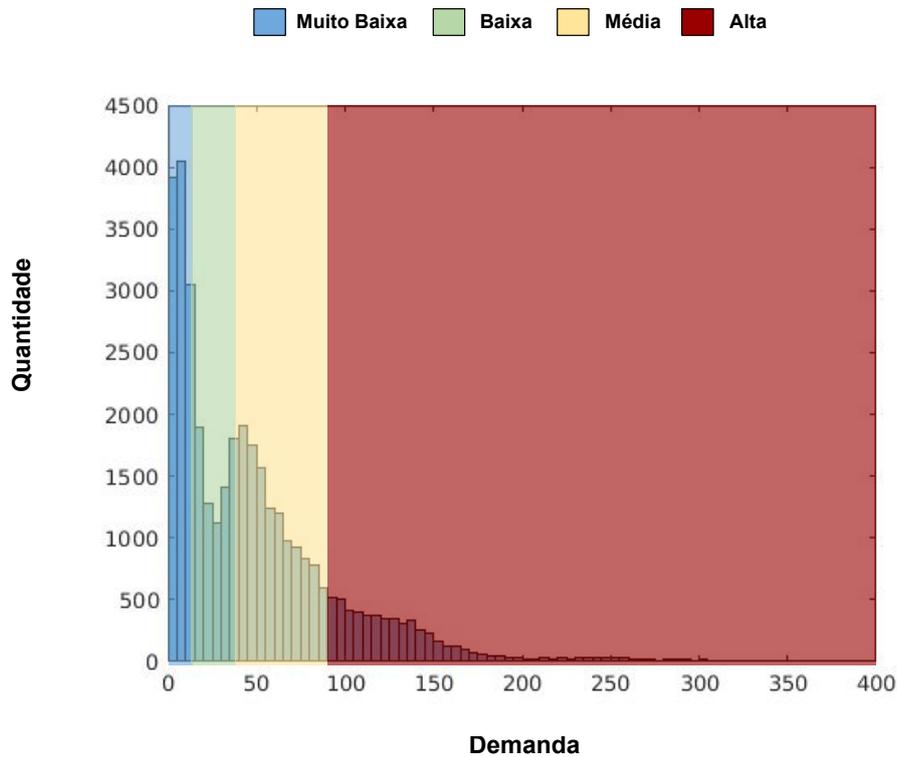
balanceamento entre o número de corridas entre as classes. Assim, evita-se os problemas causados por classes desbalanceadas (TYAGI; MITTAL, 2020). A identificação dos limites das classes é realizada com base nas demandas de todas as zonas. Neste trabalho, os limítrofes de todas as classes são sempre valores múltiplos de 5.

Neste trabalho, a identificação dos limites é realizada através de uma análise auxiliada por um histograma, composto pelas demandas de todas as zonas de um conjunto de dados das corridas. Através do histograma, é possível visualizar a distribuição das demandas por quantidade, isto é, sua frequência de aparição, facilitando a estratificação de cada limite. A Figura 13 ilustra um exemplo de divisão em classes baseado no histograma das 25 zonas em Chengdu (China) para o intervalo de 15 minutos. No eixo  $x$ , é descrita a demanda registrada e, no eixo  $y$ , a quantidade de vezes na qual determinada demanda ocorreu.

#### 4.2.2 Classificação da Demanda

Conforme descrito na seção anterior, a classificação é realizada por meio da transformação de uma tarefa de previsão em uma tarefa de classificação. Assim, na classificação da demanda são utilizadas as saídas obtidas na etapa de previsão (vide Seção 4.1.4). Os valores previstos pelo modelo evolutivo são interpretados e convertidos em uma das classes. Em outras palavras, verifica-se no domínio de qual classe o valor previsto pelo modelo evolutivo ( $\hat{y}_t$ ) se enquadra. A conversão é realizada conforme:

Figura 13 – Exemplo de histograma para Chengdu.



$$Classe = \begin{cases} Muito Baixa, & \text{Se } 0 \leq \hat{y}_t \leq LS_{Muito Baixa}, \\ Baixa, & \text{Se } LI_{Baixa} \leq \hat{y}_t \leq LS_{Baixa}, \\ Média, & \text{Se } LI_{Média} \leq \hat{y}_t \leq LS_{Média}, \\ Alta, & \text{Se } LI_{Alta} \leq \hat{y}_t \leq LS_{Alta}, \end{cases} \quad (23)$$

em que  $LI$  e  $LS$  são respectivamente os limites inferior e o superior da classe.

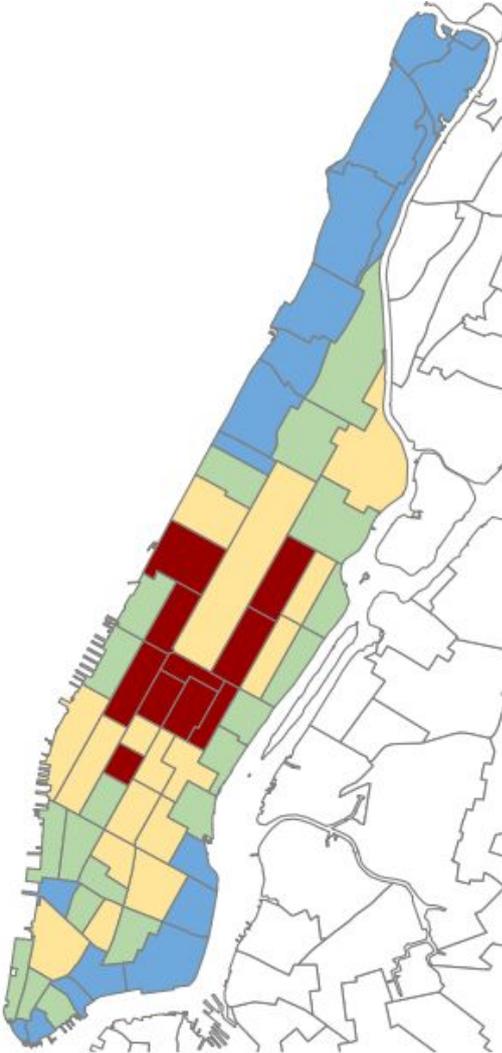
### 4.2.3 Geração dos Mapas de Calor

Neste trabalho, o mapa de calor representará graficamente a demanda de táxi em uma cidade ou região. Por meio do mapa de calor é possível, de forma simples, intuitiva e fácil, visualizar e interpretar a demanda de táxi simultaneamente em todas as zonas de uma cidade ou região.

No mapa, cada zona é representada por cores, sendo as com demandas mais altas identificadas por cores mais quentes e as com menores demandas, por cores mais frias. Neste trabalho, o mapa de calor é gerado considerando as 4 classes definidas na Seção 4.2.1. A Figura 14 ilustra um mapa gerado para 63 zonas da cidade de Nova Iorque (Estados Unidos da América).

Figura 14 – Exemplo de mapa de calor para 63 zonas da cidade de Nova Iorque.

 **Muito Baixa**    **Baixa**    **Média**    **Alta**



# Capítulo 5

## Experimentos Computacionais

Esta seção apresenta experimentos computacionais realizados para avaliar e comparar o desempenho da abordagem proposta em tarefas de previsão e classificação da demanda de *Pick-Up* e *Drop-Off*. A Seção 5.1 ilustra a previsão e a classificação da demanda para a cidade de Chengdu (China). Os experimentos com a previsão e classificação de demanda para a cidade de Nova Iorque (Estados Unidos da América) com 86 zonas são ilustrados na Seção 5.2 e com 63 zonas na Seção 5.3. Por último, a Seção 5.4 detalha a análise realizada nas variáveis de entrada selecionadas pelo método de *Kruskal-Wallis*.

Os experimentos foram conduzidos utilizando a seguinte metodologia:

- **Intervalos de Tempo:** foram definidos intervalos de tempo de 15 e 30 minutos, conforme [Moreira-Matias et al. \(2012\)](#), [Moreira-Matias et al. \(2013\)](#), [Zhang, Liu e Zheng \(2019\)](#), [Zhang et al. \(2020\)](#).
- **Tipo de Previsão:** os experimentos contemplam a previsão de *Pick-Up* e de *Drop-Off*. Assim, para cada zona  $z$ , são criados 4 conjuntos de dados: (i) previsão de *Pick-Up* com 15 minutos; (ii) previsão de *Drop-Off* com 15 minutos; (iii) previsão de *Pick-Up* com 30 minutos e; (iv) previsão de *Drop-Off* com 30 minutos.
- **Seleção de Variáveis:** para os experimentos foram utilizados conjuntos de dados contendo as 5, 10 e 20 variáveis mais relevantes indicadas pelo método de *Kruskal-Wallis*, ou seja,  $N = \{5, 10, 20\}$ . Além de um conjunto contendo todas as 26 variáveis. Para efeitos de objetividade, nos resultados descritos nas tabelas [2](#), [3](#), [5](#), [6](#), [7](#), [8](#), [10](#), [11](#), [12](#), [13](#), [15](#) e [16](#) são apresentados apenas o melhor resultado obtido por cada modelo, juntamente com o respectivo número de variáveis utilizado. Adicionalmente, o Apêndice [C](#) ilustra as tabelas contendo o desempenho dos modelos para cada conjunto de variáveis.

- **Modelos Fuzzy Evolutivos:** para realizar os experimentos foram utilizados os algoritmos ALMMo (ANGELOV; GU; PRINCIPE, 2017), eMG (LE MOS; CAMINHAS; GOMIDE, 2011) e eNFN (SILVA et al., 2014). Os três modelos são determinísticos, isto é, seus resultados não mudam a cada nova iteração. Portanto, cada algoritmo foi executado apenas uma vez. O processamento das amostras foi realizado de forma *online*, isto é, os modelos evoluem sua estrutura e parâmetros para todas as amostras. Destaca-se que não foi realizado nenhum procedimento para ajuste fino nos hiperparâmetros dos modelos e, visando manter a abordagem generalista, optou-se por utilizar o mesmo conjunto de parâmetros em todos os experimentos. Os valores dos hiperparâmetros para cada modelo evolutivo seguiram os limites métricos definidos pelos seus respectivos autores, sendo listados como se segue:
  - **ALMMo:** *forgettingfactor* = 0,1; *densitythreshold* = 0,8;  $\omega = 10$ .
  - **eMG:**  $\alpha = 0,01$ ;  $\lambda = 0,05$ ;  $w = 40$ ,  $\Sigma = 10^{-1}$ .
  - **eNFN:**  $\eta = 10$ ;  $\beta = 0,01$ ;  $\omega = 100$ .
- **Experimentos de Previsão:** nestes experimentos, o desempenho da abordagem proposta foi avaliado e comparado com abordagens alternativas por meio do RMSE (*Root Mean Square Error*), conforme Zhang, Liu e Zheng (2019), Zhang et al. (2020), que pode ser obtido por:

$$RMSE = \sqrt{\frac{1}{S} \sum_{t=1}^S |\hat{y}_t - y_t|^2}, \quad (24)$$

em que  $\hat{y}_t$  é demanda prevista,  $y_t$  demanda real e  $S$  o número de amostras utilizadas para avaliar o desempenho. Os resultados são avaliados pela média dos erros das zonas.

- **Experimentos de Classificação:** a primeira etapa para os experimentos de classificação é a definição das classes. Conforme discutido na Seção 4.2.1, foram definidas quatro classes (Muito Baixa, Baixa, Média e Alta). Para avaliar o desempenho na classificação será empregada a Acurácia, definida por:

$$Acurácia = \frac{cc}{S} * 100, \quad (25)$$

sendo  $cc$  o número de amostras classificadas corretamente e  $S$  o número total de amostras. Além da Acurácia, o resultado da classificação será representado em um mapa de calor.

## 5.1 Previsão e Classificação da Demanda em Chengdu (China)

Nesta seção, são apresentados os experimentos computacionais realizados no conjunto de dados derivado da base de corridas realizadas em Chengdu, China. A base de dados para estes experimentos é provida pela Didi Chuxing<sup>1</sup>. Os arquivos disponibilizados contêm corridas no intervalo de 01/11/2016 até 30/11/2016.

O local de estudos está localizado na 2ª via do anel rodoviário da cidade, com aproximadamente 65 km<sup>2</sup>. Os limites da região vão de 104,043 até 104,129 graus Leste em longitude, e 30,653 até 30,726 Norte em latitude. A área em questão foi particionada em 25 zonas retangulares com aproximadamente 1,6 km de lado, ilustradas na Figura 15. O conjunto de dados foi dividido em 80% (01/11/2016 até 23/11/2016) para a seleção de variáveis pelo Kruskal-Wallis e definição das classes de valores. Os 20% restantes (24/11/2016 até 30/11/2016) são reservados para avaliação e comparação do desempenho. Os conjuntos de dados foram normalizadas entre [-1 e 1]. Para efeitos de comparação de resultados, as configurações citadas anteriormente foram as mesmas utilizadas por Zhang, Liu e Zheng (2019).

Figura 15 – Divisão das 25 zonas em Chengdu.



Em seu trabalho, Zhang, Liu e Zheng (2019) avaliaram o desempenho da previsão de demanda comparando 11 modelos com aprendizado *offline*, os quais foram: SVM, XGBoost, STL-LSTM, STL-ConvLSTM, STL-TCNN, MTL-LSTM, MTL-ConvLSTM, FLC-Net, FTCNN-Net, MTL-TCNN e MTL-TCNN (ST-DTW). O melhor resultado foi obtido pelo MTL-TCNN (ST-DTW) que será usado como *benchmark*.

<sup>1</sup> Disponível em: <https://gaia.didichuxing.com/>

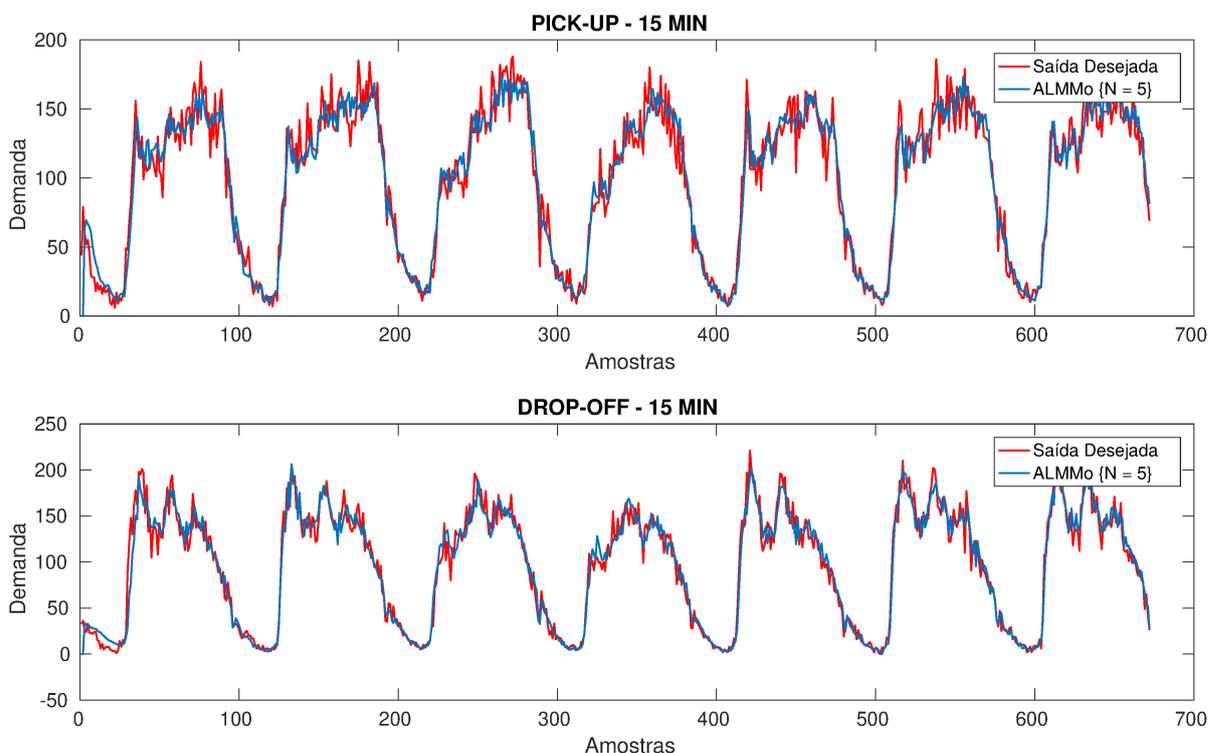
### 5.1.1 Previsão da Demanda em Chengdu

A Tabela 2 ilustra o desempenho dos modelos na previsão de *Pick-Up* e *Drop-Off* com intervalos de 15 minutos. A Tabela ilustra os melhores resultados obtidos por cada um dos modelos e sugere que o melhor desempenho foi obtido pelo ALMMo (com 5 variáveis de entrada) seguido pelo MTL-TCNN (ST-DTW) (ZHANG; LIU; ZHENG, 2019)<sup>2</sup>, eMG (com 5 variáveis de entrada) e eNFN (com 10 variáveis de entrada). A Figura 16 apresenta a previsão de *Pick-Up* e *Drop-Off* do ALMMo (com 5 variáveis de entrada) em intervalos de 15 minutos, na zona 15, no período selecionado para validação dos modelos.

Tabela 2 – Desempenho na previsão de demanda em intervalos de 15 minutos para Chengdu.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 5}	8,720	8,669
MTL-TCNN (ST-DTW) (ZHANG; LIU; ZHENG, 2019) <sup>2</sup>	8,935	—
eMG - {N = 5}	9,673	9,948
eNFN - {N = 10}	10,210	10,001

Figura 16 – Previsões do ALMMo (com 5 variáveis de entrada) em intervalos de 15 minutos, para a zona 15, no período de 24/11/2016 até 30/11/2016.



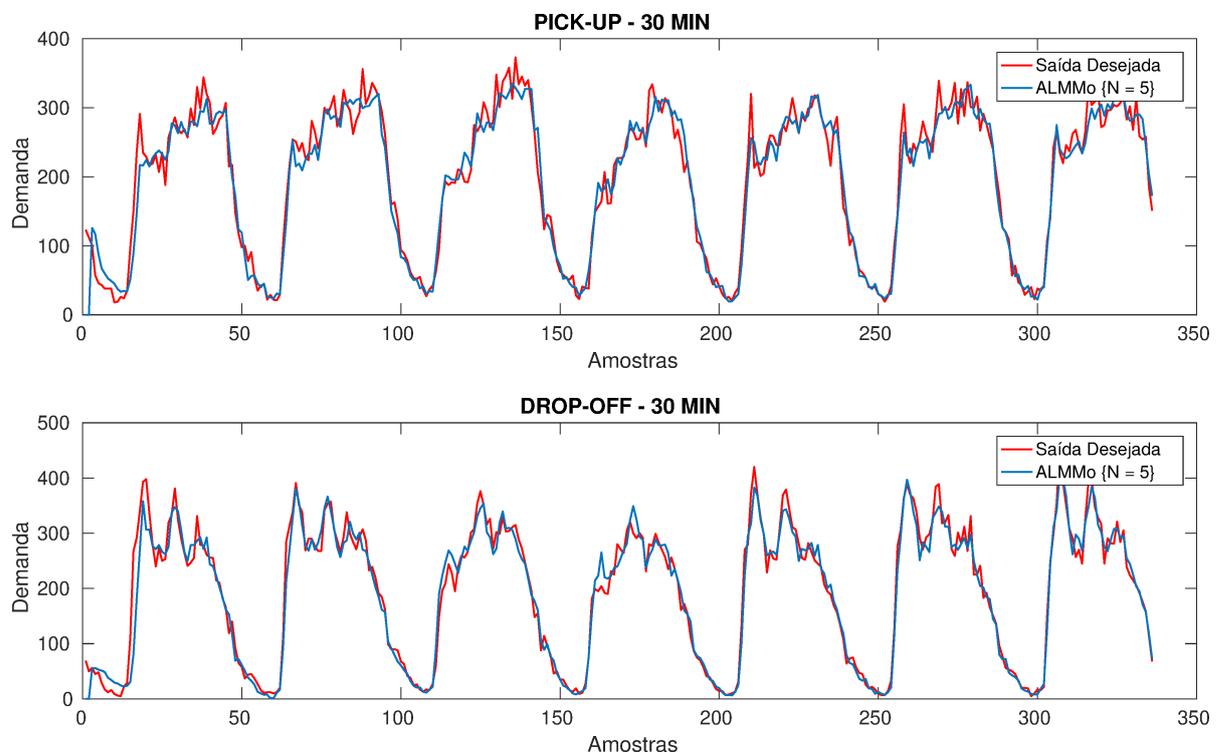
<sup>2</sup> Em Zhang et al. (ZHANG; LIU; ZHENG, 2019) não houve experimentos que realizassem a previsão de demanda para *Drop-Off*.

A Tabela 3 mostra os melhores resultados obtidos por cada um dos modelos na previsão de *Pick-Up* e *Drop-Off* para o intervalo de 30 minutos. Assim como na previsão para o intervalo de 15 minutos, nos experimentos de previsão de demanda para o intervalo de 30 minutos, o melhor desempenho foi obtido pelo ALMMo (com 5 variáveis de entrada), seguido pelo eMG (com 5 variáveis de entrada) e pelo eNFN (com 10 variáveis de entrada). A Figura 17 ilustra a previsão de *Pick-Up* e *Drop-Off* do ALMMo (com 5 variáveis de entrada) em intervalo de 30 minutos, para a zona 15, no período de 24/11/2016 até 30/11/2016.

Tabela 3 – Desempenho na previsão de demanda em intervalos de 30 minutos para Chengdu.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 5}	<b>15,172</b>	<b>15,640</b>
eMG {N = 5}	17,339	17,508
eNFN {N = 10}	17,490	17,514
MTL-TCNN (ST-DTW)(ZHANG; LIU; ZHENG, 2019) <sup>3</sup>	—	—

Figura 17 – Previsões do ALMMo (com 5 variáveis de entrada) em intervalos de 30 minutos, para a zona 15, no período de 24/11/2016 até 30/11/2016.

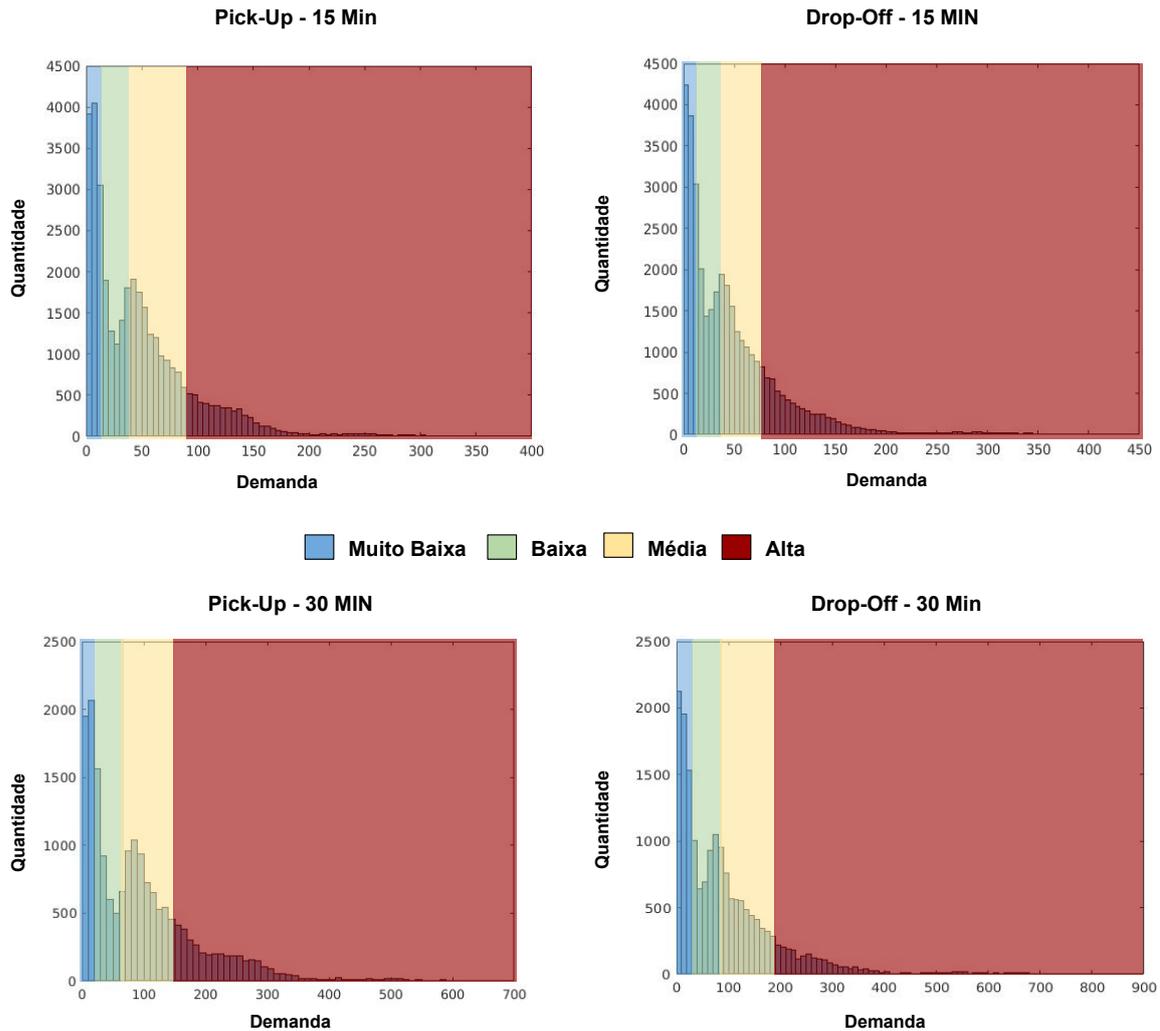


<sup>3</sup> Em Zhang, Liu e Zheng (2019) não houve experimentos que realizassem a previsão de demanda com intervalo de 30 minutos.

### 5.1.2 Classificação da Demanda em Chengdu

A Tabela 4 ilustra, para *Pick-Up* e *Drop-Off*, a faixa de valores de cada classe para os intervalos de 15 e 30 minutos, na qual  $D$  representa a demanda. Além da faixa de valores, a tabela destaca (entre parênteses) o número de amostras em cada classe. As informações apresentadas na Tabela 4 são ilustradas graficamente na Figura 18 por meio de histogramas.

Figura 18 – Histogramas com as faixas de valores em Chengdu.



O desempenho dos modelos na classificação da demanda por faixa de valores em intervalos de 15 minutos para a cidade de Chengdu é ilustrada na Tabela 5 pela média das acurácias, bem como pelo desvio padrão. O melhor desempenho de *Pick-Up* e *Drop-Off* foi alcançado pelo ALMMo (com 5 variáveis de entrada), seguido pelo eMG (com 5 variáveis de entrada) e pelo eNFN (com 10 variáveis de entrada).

A Figura 19 ilustra os mapas de calor para as 25 zonas de Chengdu. A Figura 19 (A) apresenta o mapa gerado com base na classificação do ALMMo para *Pick-Up* e a Figura 19 (B) o mapa desejado. Por outro lado, a Figura 19 (C) retrata o mapa para *Drop-Off* obtido com base nos resultados do ALMMo e a Figura 19 (D) o desejado. Como pode ser

Tabela 4 – Definição das demandas em faixas de valores para Chengdu.

Tipo	Intervalo	Faixas e Quantidade			
		Muito Baixa	Baixa	Média	Alta
Pick-Up	15	$D \leq 10$ (8607)	$10 < D \leq 35$ (8452)	$35 < D \leq 70$ (10283)	$D > 70$ (8658)
Pick-Up	30	$D \leq 20$ (4196)	$20 < D \leq 70$ (4164)	$70 < D \leq 140$ (5320)	$D > 140$ (4320)
Drop-Off	15	$D \leq 10$ (8788)	$10 < D \leq 35$ (9432)	$35 < D \leq 70$ (9542)	$D > 70$ (8238)
Drop-Off	30	$D \leq 20$ (4275)	$20 < D \leq 70$ (4713)	$70 < D \leq 140$ (4869)	$D > 140$ (4143)

Tabela 5 – Acurácia em intervalos de 15 minutos para Chengdu.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	$\pm$ Desv. Pad.	Acurácia (%)	$\pm$ Desv. Pad.
ALMMo {N = 5}	<b>83,845</b>	<b>4,779</b>	<b>82,738</b>	<b>5,605</b>
eMG {N = 5}	83,065	5,558	81,435	7,121
eNFN {N = 10}	81,571	5,206	80,268	6,790

visto, há poucas diferenças visuais entre os mapas previstos e desejados, mostrando assim um bom desempenho do modelo evolutivo em questão. Os destaques em preto ilustram alguns exemplos de zonas onde a classe prevista não corresponde com a classe desejada. Estes destaques exemplificam que, mesmo o modelo não tendo efetuado corretamente a classificação, a diferença deste erro é de apenas uma classe adjacente.

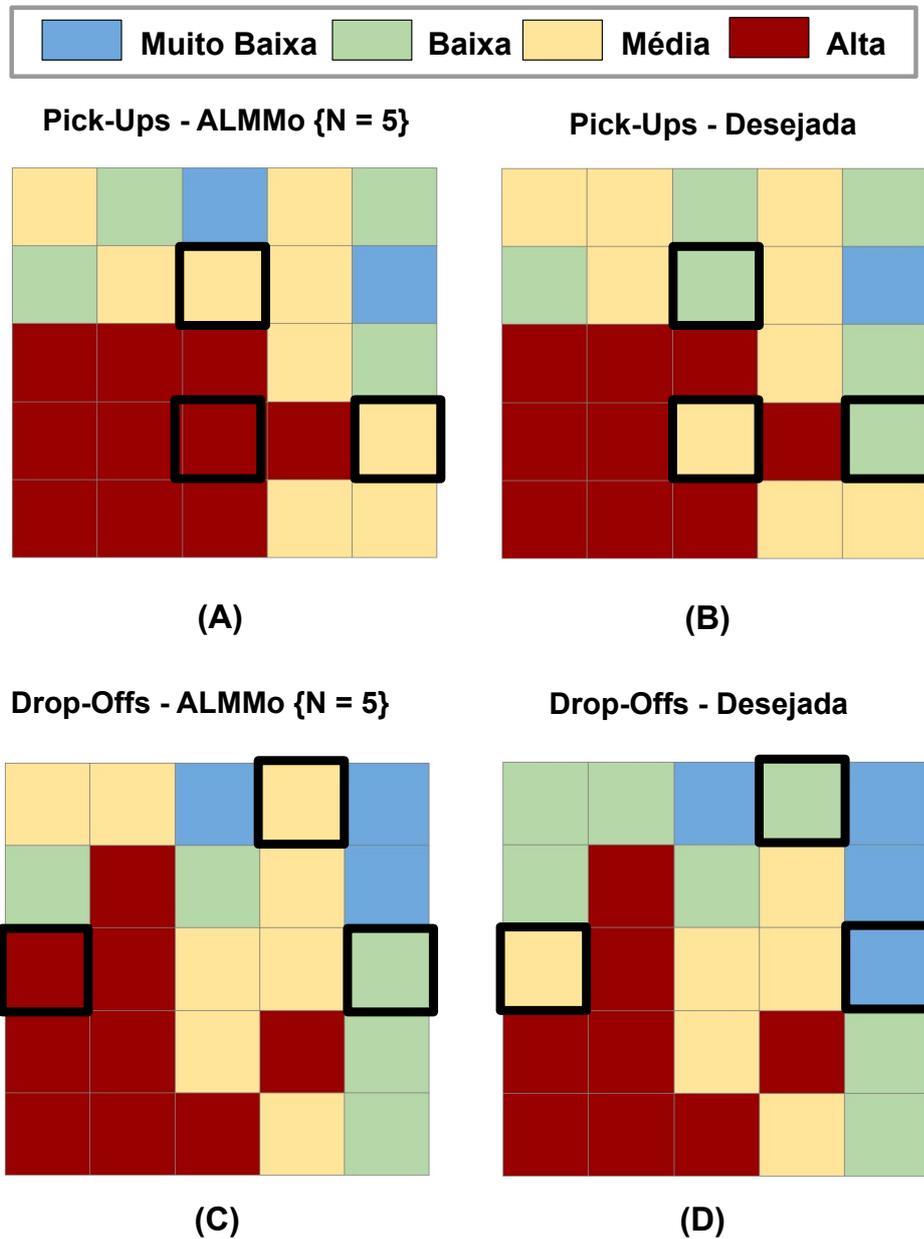
A Tabela 6 representa as médias das acurácias das zonas, bem como seu desvio padrão para Chengdu, em intervalos de 30 minutos. Assim como nos resultados de 15 minutos (Tabela 5), o melhor desempenho de *Pick-Up* e *Drop-Off* foi obtido pelo ALMMo (com 5 variáveis de entrada). Na sequência, os melhores desempenhos foram alcançados, respectivamente, pelo eMG (com 5 variáveis de entrada) e eNFN (com 10 variáveis de entrada).

Tabela 6 – Acurácia em intervalos de 30 minutos para Chengdu.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	$\pm$ Desv. Pad.	Acurácia (%)	$\pm$ Desv. Pad.
ALMMo {N = 5}	<b>86,952</b>	<b>4,118</b>	<b>85,286</b>	<b>5,670</b>
eMG {N = 5}	86,274	4,533	84,964	6,117
eNFN {N = 10}	84,869	4,803	83,143	6,068

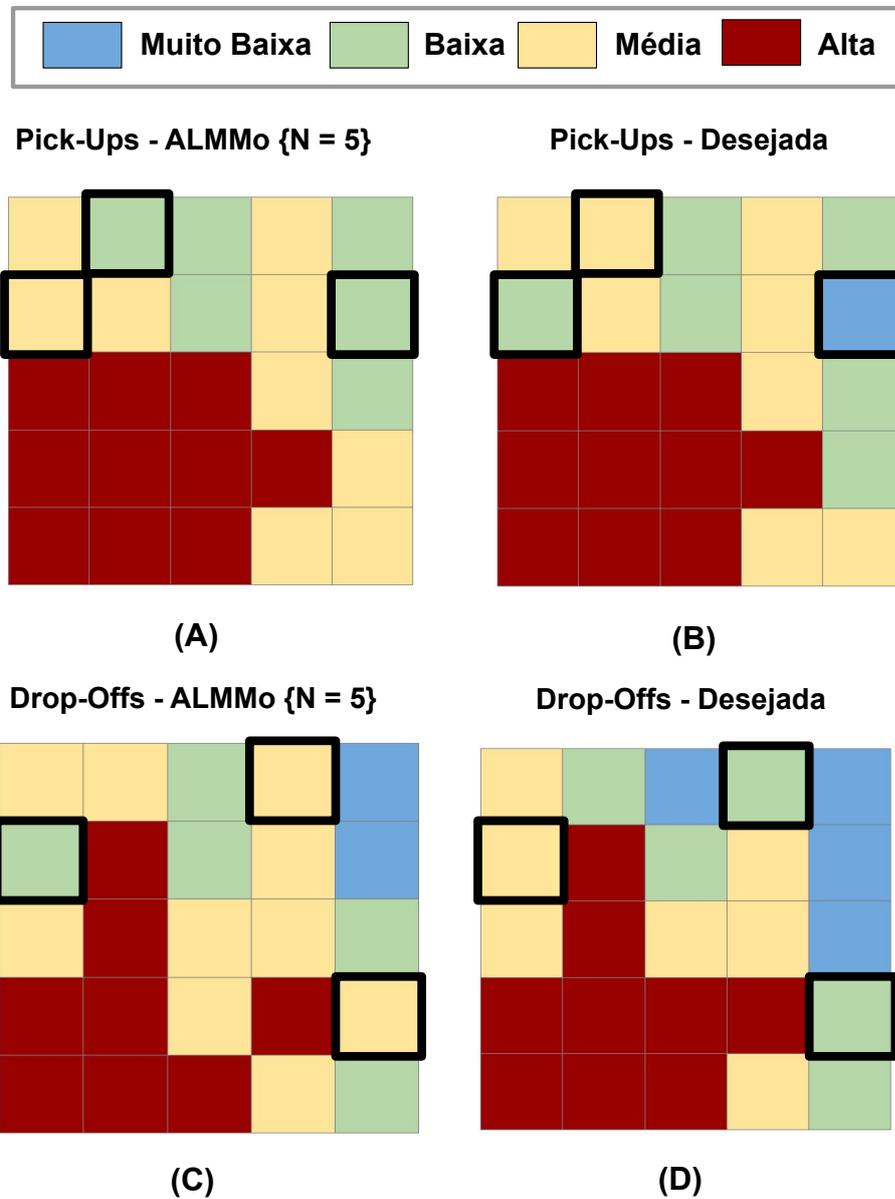
Os mapas de calor gerados para as 25 zonas de Chengdu, considerando o intervalo das 07:00 às 07:30 no dia 30/11/2016 são apresentados na Figura 20. O mapa composto

Figura 19 – Mapas de calor nas 25 zonas em Chengdu - dia 30/11/2016 no intervalo de 07:00 às 07:15: (A) mapa obtido pelo ALMMo  $\{N = 5\}$  para *Pick-Up*; (B) mapa desejado para *Pick-Up*; (C) mapa obtido pelo ALMMo  $\{N = 5\}$  para *Drop-Off*; e (D) mapa desejado para *Drop-Off*.



utilizando os resultados do ALMMo para *Pick-Up* é apresentado na Figura 20 (A) e o desejado na Figura 20 (B). Em contrapartida, o mapa para *Drop-Off* gerado com base nas previsões do ALMMo é ilustrado na Figura 20 (C) e o desejado para *Drop-Off* na Figura 20 (D). Assim como nos mapas da Figura 19, pode-se perceber que os mapas previstos e os desejados são similares. Novamente, os destaques em preto apresentam exemplos de zonas onde há divergência entre a classificação prevista e a desejada. A diferença nas classificações também é de apenas uma classe adjacente.

Figura 20 – Mapas de calor nas 25 zonas em Chengdu - dia 30/11/2016 no intervalo de 07:00 até 07:30: (A) mapa obtido pelo ALMMo  $\{N = 5\}$  para *Pick-Up*; (B) mapa desejado para *Pick-Up*; (C) mapa obtido pelo ALMMo  $\{N = 5\}$  para *Drop-Off*; (D) mapa desejado para *Drop-Off*.



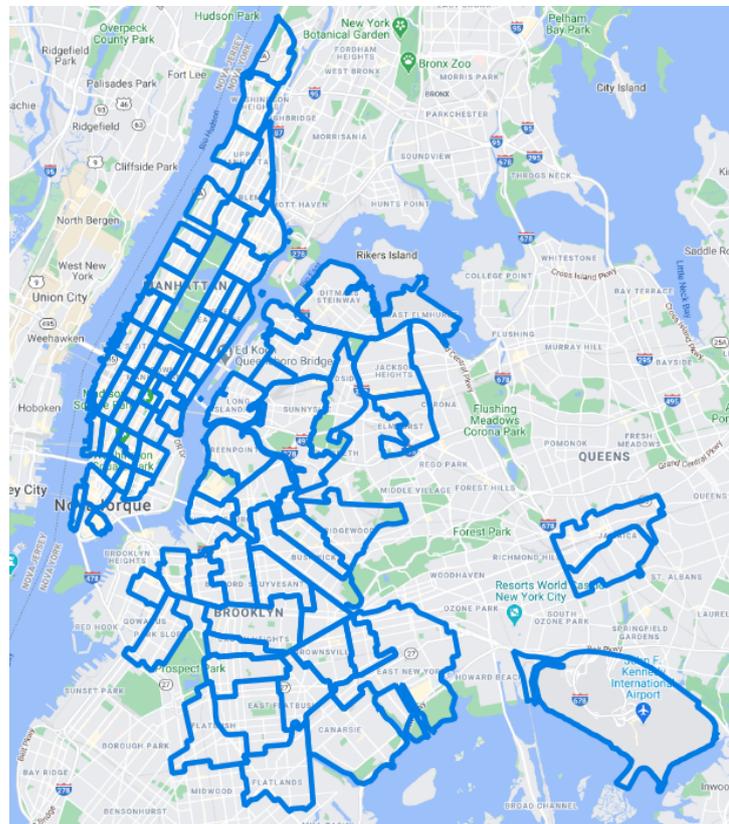
## 5.2 Previsão e Classificação da Demanda em Nova Iorque (Estados Unidos da América) com 86 zonas

Nesta seção, são apresentados os experimentos computacionais realizados no conjunto de dados derivado da base de corridas da cidade de Nova Iorque, nos Estados Unidos da América. A base provém da *NYC Taxi and Limousine Commission*<sup>4</sup>, principal frota de táxis que atua na cidade e possui registros diários das corridas de três tipos de táxis: *yellow*,

<sup>4</sup> Disponível em: <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

*green* e FHV (*For Hire Vehicle*). Nesta base de corridas, a cidade de Nova Iorque está dividida em 263 zonas (microrregiões) já predefinidas pela prefeitura da cidade e os registros disponíveis compreendem o período de 01/01/2009 até os dias atuais. Os experimentos foram realizados em um subconjunto de 86 zonas (vide Apêndice B), representadas na Figura 21, no período de 01/01/2018 até 31/10/2018. Os conjuntos de dados foram divididos em 60% (01/01/2018 até 30/06/2018) para a seleção de variáveis pelo *Kruskal-Wallis* e definição das classes de valores. Os 40% (01/07/2018 até 31/10/2018) restantes foram utilizados para a validação dos modelos. Os conjuntos de dados foram normalizados entre 0 e 1. Para efeitos de comparação entre resultados, as configurações descritas anteriormente foram, *mutatis mutandis*, as mesmas utilizadas por Zhang et al. (2020).

Figura 21 – Divisão das 86 zonas em Nova Iorque.



Em seu trabalho, Zhang et al. (2020) avaliaram o desempenho da previsão de demanda comparando 5 modelos com aprendizado *offline*, a saber: ARIMA, MLP, sLSTM, mLSTM e pmLSTM. O melhor resultado foi obtido pelo pmLSTM que será usado como *benchmark*.

### 5.2.1 Previsão da Demanda em Nova Iorque com 86 zonas

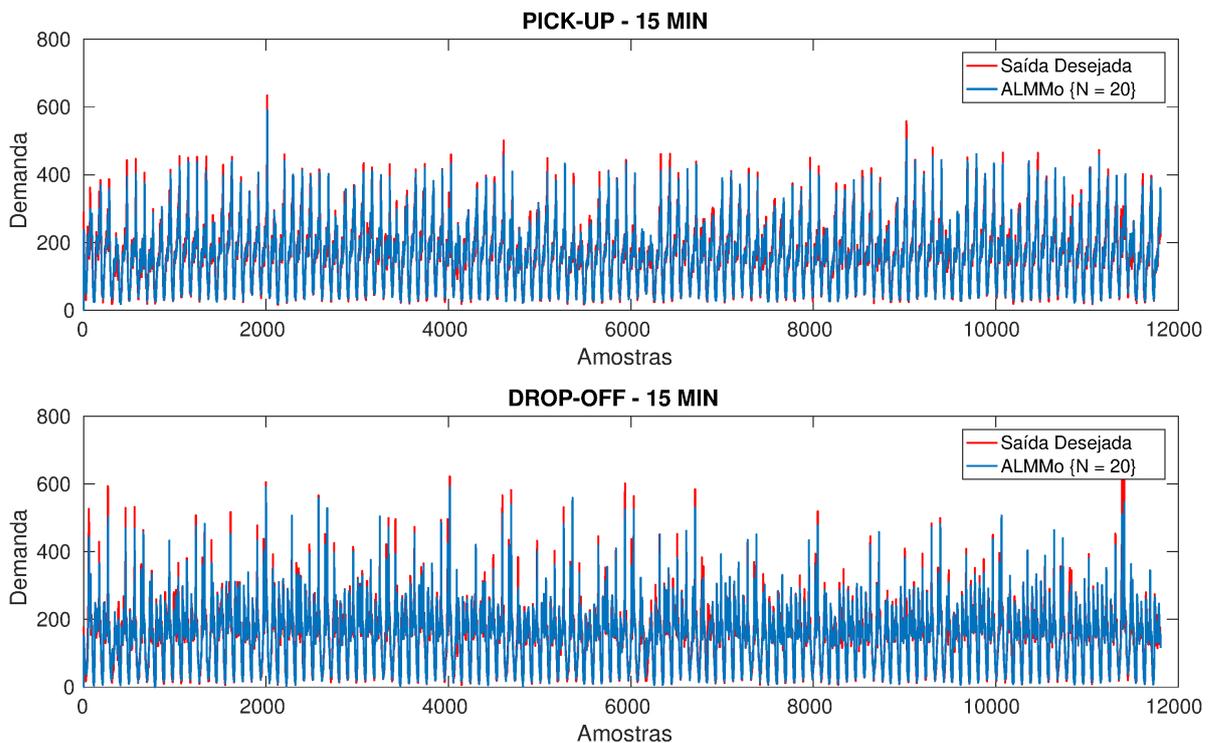
A Tabela 7 demonstra o desempenho dos preditores para o intervalo de 15 minutos. A tabela apresenta os melhores resultados obtidos por cada um dos modelos, sendo o melhor desempenho obtido pelo ALMMo (com 20 variáveis de entrada) seguido pelo eMG (com 10 variáveis de entrada) e pelo eNFN (com 5 variáveis de entrada). A Figura 22 ilustra a

previsão pelo ALMMo (com 20 variáveis de entrada) para o intervalo de 15 minutos, no período de 01/07/2018 até 31/10/2018 na Times Square (zona 230).

Tabela 7 – Desempenho na previsão da demanda em intervalos de 15 minutos para Nova Iorque com 86 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	<b>13,329</b>	<b>12,460</b>
eMG {N = 10}	14,793	15,466
eNFN {N = 5}	15,464	14,580
pmlLSTM (ZHANG et al., 2020) <sup>5</sup>	—	—

Figura 22 – Previsões do ALMMo (com 20 variáveis de entrada) em intervalos de 15 minutos, para a Times Square (zona 230), no período de 01/07/2018 até 31/10/2018.



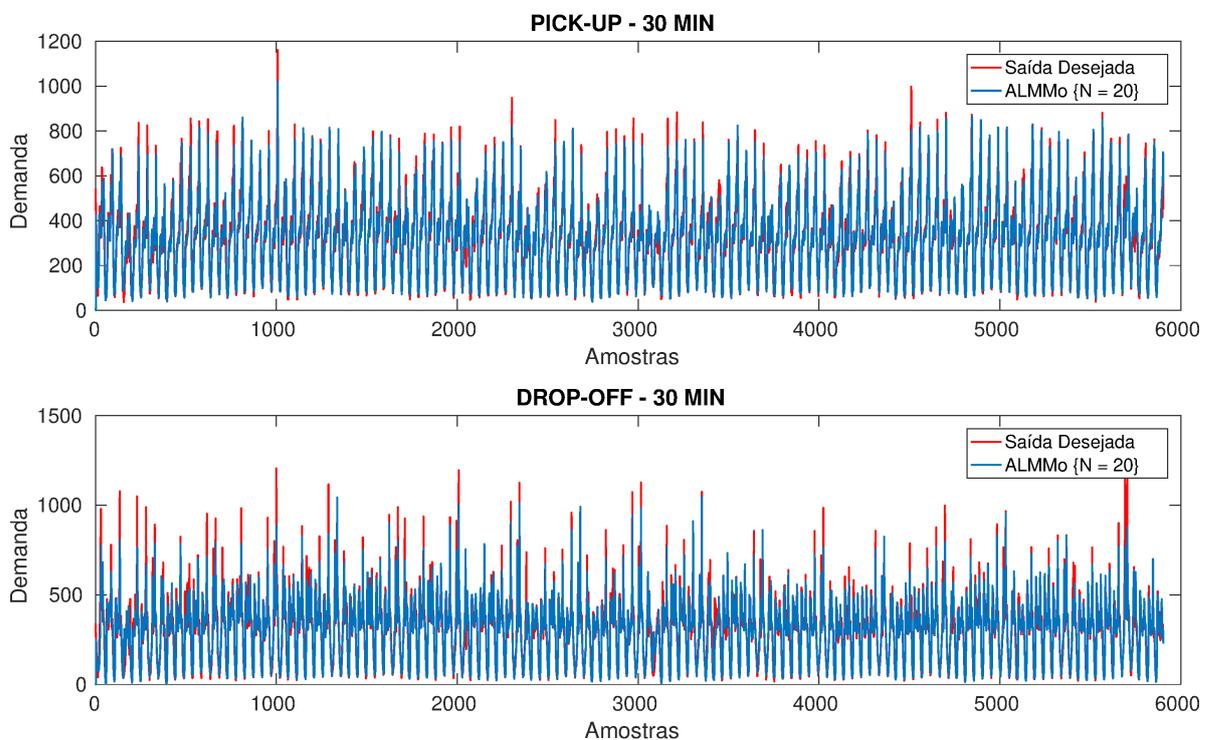
A Tabela 8 descreve os melhores resultados dos preditores para o intervalo de 30 minutos. O ALMMo (com 20 variáveis) obteve o menor erro para *Pick-Up*. Para *Drop-Off*, os melhores resultados foram obtidos pelo pmlLSTM. O ALMMo (com 20 variáveis) mostrou-se competitivo, apresentando uma diferença de apenas 3% em relação aos resultados de *Drop-Off* obtidos pelo pmlLSTM. Na sequência, segue-se o eMG (com 10 variáveis) e o eNFN (com 5 variáveis). Na Figura 23 é representada a previsão pelo ALMMo (com 20 variáveis de entrada), em intervalos de 30 minutos, de 01/07/2018 até 31/10/2018 na Times Square (zona 230).

<sup>5</sup> Em Zhang et al. (2020) não houve experimentos que realizassem a previsão de demanda com intervalo de 15 minutos.

Tabela 8 – Desempenho na previsão da demanda em intervalos de 30 minutos para Nova lorque com 86 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	<b>25,072</b>	22,957
pmlSTM (ZHANG et al., 2020)	26,311	<b>22,411</b>
eMG {N = 10}	28,512	26,473
eNFN {N = 5}	29,017	25,358

Figura 23 – Previsões do ALMMo (com 20 variáveis de entrada) em intervalos de 30 minutos, para a Times Square (zona 230), no período de 01/07/2018 até 31/10/2018.



### 5.2.2 Classificação da Demanda em Nova lorque com 86 zonas.

A Tabela 9 ilustra, para *Pick-Up* e *Drop-Off*, a faixa de valores de cada classe para os intervalos de 15 e 30 minutos em Nova lorque (86 zonas), na qual  $D$  representa a demanda e o número de amostras em cada classe é representado em parênteses. Os histogramas da Figura 24 representam graficamente os dados descritos na Tabela 9.

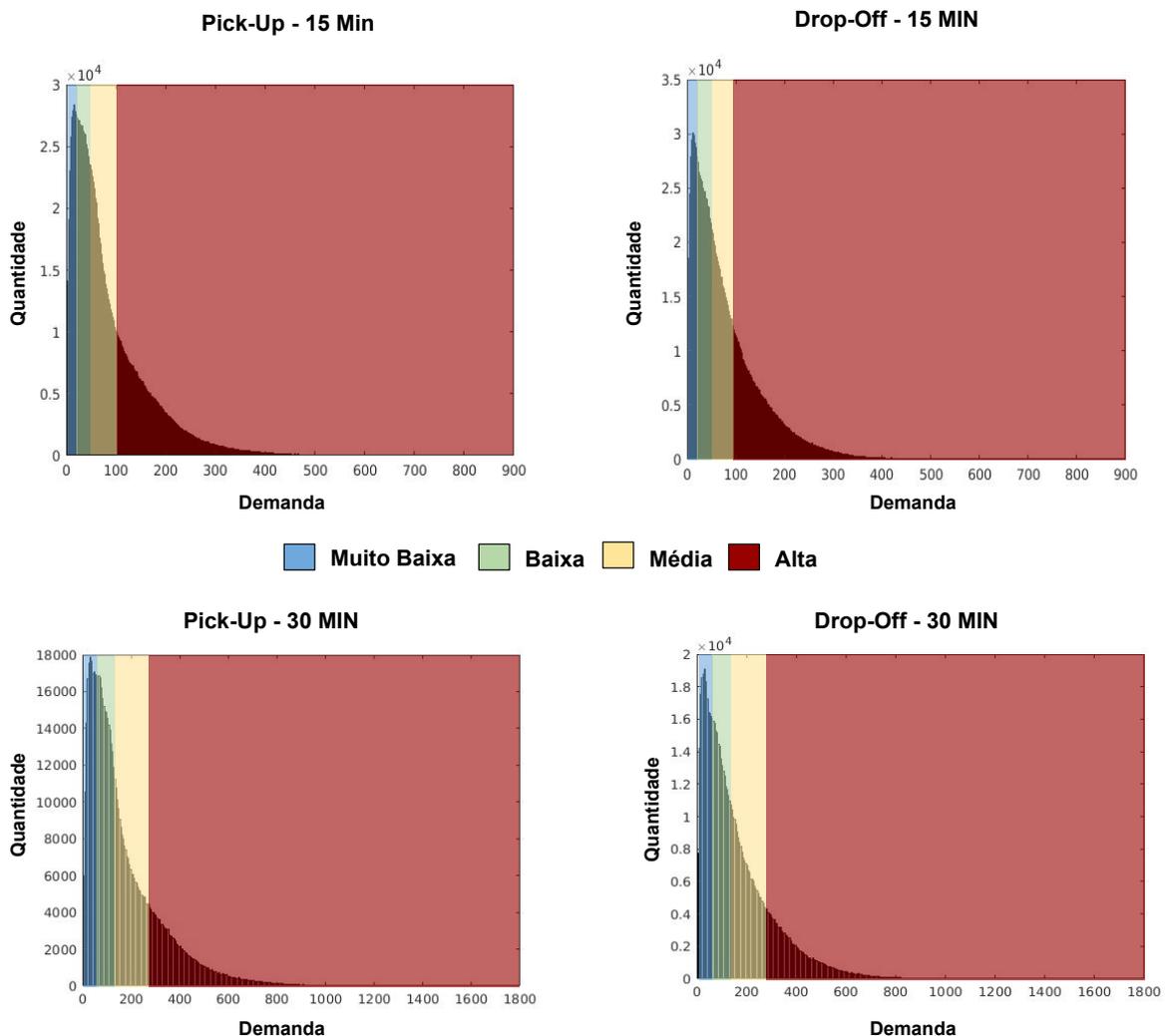
A Tabela 10 demonstra as melhores acurácias e desvio padrão dos modelos evolutivos para as 86 zonas de Nova lorque, em intervalos de 15 minutos. Os resultados sugerem que a melhor acurácia foi obtida pelo ALMMo (com 20 variáveis de entrada), seguido pelo eMG (com 10 variáveis de entrada) e eNFN (com 5 variáveis de entrada).

A Figura 25 apresenta os mapas de calor previstos e os desejados para *Pick-Up* e *Drop-*

Tabela 9 – Definição das demandas em faixas de valores para Nova lorque com 86 zonas.

Tipo	Intervalo	Faixas e Quantidade			
		Muito Baixa	Baixa	Média	Alta
Pick-Up	15	$D \leq 25$ (287283)	$25 < D \leq 50$ (318128)	$50 < D \leq 100$ (389045)	$D > 100$ (433832)
Pick-Up	30	$D \leq 60$ (173844)	$60 < D \leq 120$ (182183)	$120 < D \leq 240$ (185552)	$D > 240$ (172565)
Drop-Off	15	$D \leq 25$ (313466)	$25 < D \leq 50$ (300727)	$50 < D \leq 100$ (391842)	$D > 100$ (422253)
Drop-Off	30	$D \leq 60$ (185395)	$60 < D \leq 120$ (169884)	$120 < D \leq 240$ (196772)	$D > 240$ (162093)

Figura 24 – Histogramas com as faixas de valores em Nova lorque com 86 zonas.



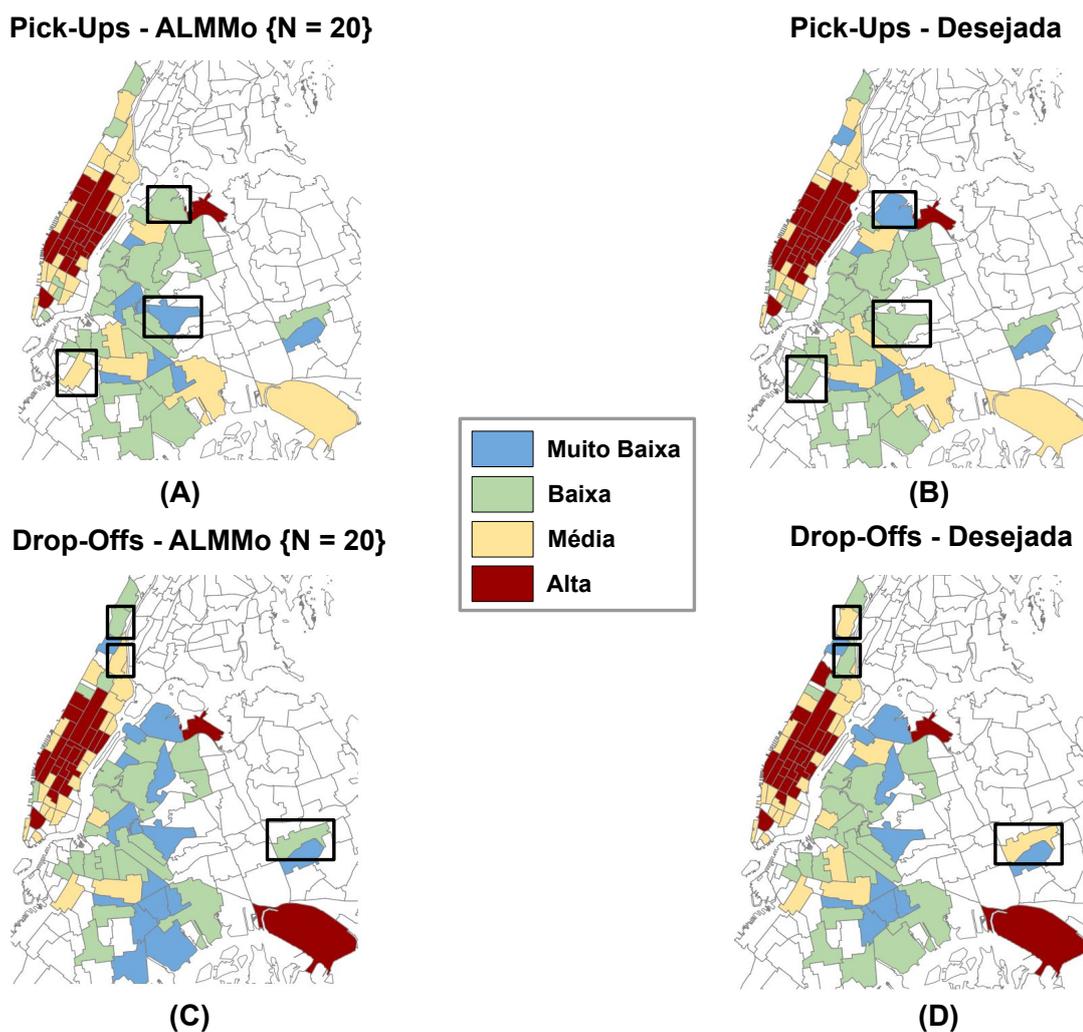
Off para as 86 zonas de Nova lorque abrangendo o intervalo das 12:00 às 12:15 no dia 31/10/2018. A Figura 25 (A) mostra o mapa baseado na classificação realizada pelo ALMMo para *Pick-Up* e a Figura 25 (B) apresenta o mapa desejado para *Pick-Up*. Considerando *Drop-Off*, o mapa gerado a partir dos resultados do ALMMo é ilustrado na Figura 25 (C)

Tabela 10 – Acurácia em intervalos de 15 minutos para Nova lorque com 86 zonas.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 20}	<b>84,796</b>	<b>4,334</b>	<b>84,865</b>	<b>4,593</b>
eMG {N = 10}	83,368	4,426	82,790	5,137
eNFN {N = 5}	81,950	5,101	81,824	5,373

e o desejado na Figura 25 (D). Percebe-se a semelhança entre os mapas previstos e os desejados, o que indica o bom desempenho dos modelos. Além disso, nota-se que a diferença nos mapas, quando ocorre, é de apenas uma classe.

Figura 25 – Mapa de calor nas 86 zonas em Nova lorque - dia 31/10/2018 no intervalo de 12:00 às 12:15: (A) mapa obtido pelo ALMMo {N = 20} para *Pick-Up*; (B) mapa desejado para *Pick-Up*; (C) mapa obtido pelo ALMMo {N = 20} para *Drop-Off* e; (D) mapa desejado para *Drop-Off*.



Os melhores desempenhos de acurácia e desvio padrão dos modelos evolutivos para as 86 zonas de Nova lorque, em intervalos de 30 minutos é ilustrada na Tabela 11. O ALMMo

(com 20 variáveis de entrada) foi o modelo que obteve os melhores resultados, seguidos pelo eMG (com 10 variáveis de entrada) e eNFN (com 5 variáveis de entrada).

Tabela 11 – Acurácia em intervalos de 30 minutos para Nova Iorque com 86 zonas.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 20}	<b>86,529</b>	<b>3,538</b>	<b>87,256</b>	<b>3,329</b>
eMG {N = 10}	85,203	3,732	85,843	3,530
eNFN {N = 5}	83,579	4,666	85,084	3,865

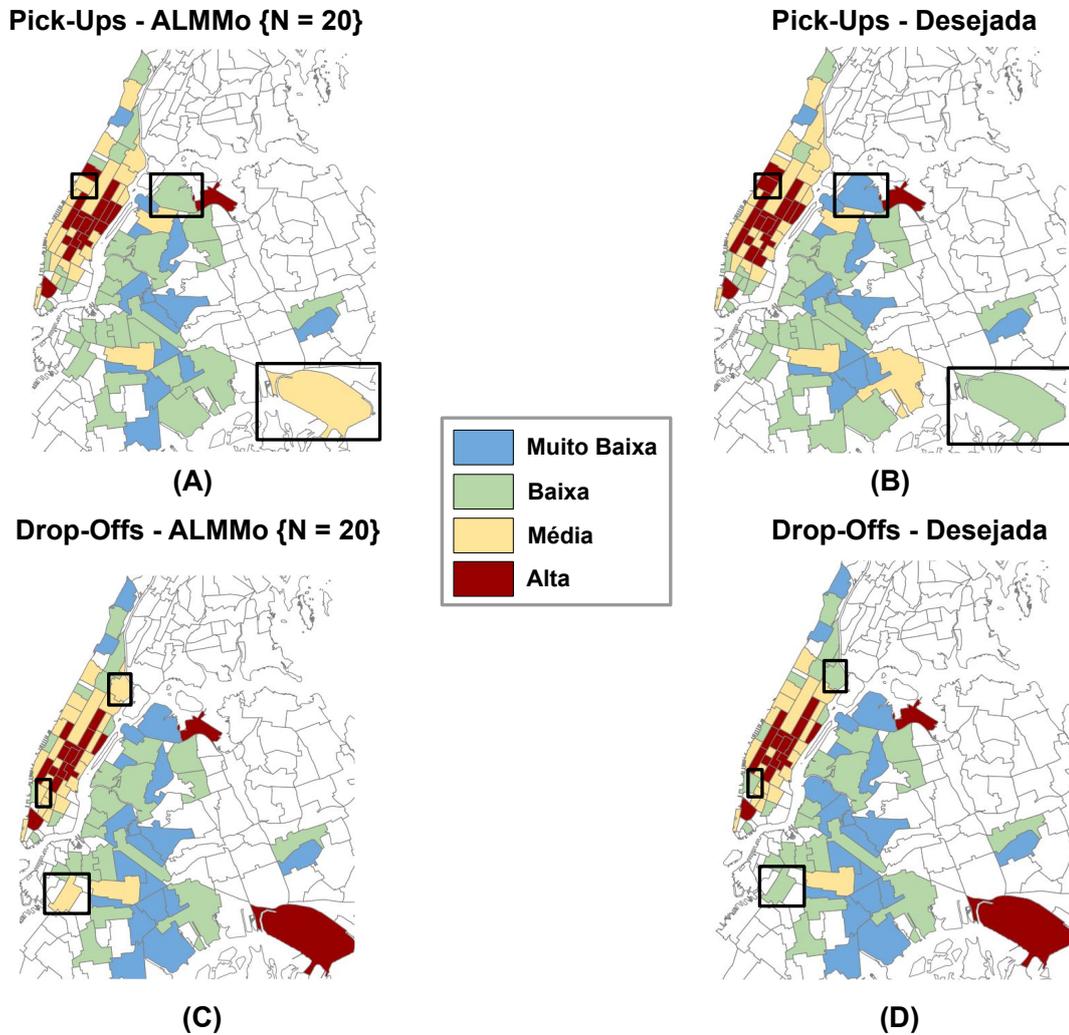
A Figura 26 ilustra os mapas de calor para *Pick-Up* e *Drop-Off* nas 86 zonas de Nova Iorque para o intervalo das 12:00 às 12:30 no dia 31/10/2018. Considerando os mapas para *Pick-Up*, a Figura 26 (A) mostra o mapa provindo da execução do ALMMo e a Figura 26 (B) apresenta o mapa desejado. Por outro lado, a Figura 26 (C) destaca o mapa para *Drop-Off* obtido com base nos resultados do ALMMo e a Figura 26 (D), o mapa desejado para *Drop-Off*. Não obstante, notam-se poucas divergências entre o que é previsto e desejado. Pode-se perceber também que há poucas diferenças entre os padrões das cores dos mapas de calor previstos e os desejados. As zonas demarcadas em preto mostram alguns casos em que houve erro na classificação, no qual a diferença do erro demonstrada é de apenas uma classe adjacente.

### 5.3 Previsão e Classificação da Demanda em Nova Iorque (Estados Unidos da América) com 63 zonas

Esta seção também apresenta experimentos computacionais baseados no conjunto de dados descrito na Seção 5.2. Porém, será utilizado um subconjunto de 63 zonas (vide o Apêndice B) (ZHANG; LIU; ZHENG, 2019) pertencentes ao distrito de Manhattan, Nova Iorque, ilustrados na Figura 27. Os experimentos foram realizados no período de 01/07/2017 até 30/06/2020. Os primeiros 20% (01/07/2017 até 31/12/2017) foram reservados para a seleção de variáveis pelo *Kruskal-Wallis* e a classificação da faixa de valores. Os 80% (01/01/2018 até 30/06/2020) das amostras foram utilizados para avaliação dos modelos. Este período engloba também o momento da pandemia de COVID-19<sup>6</sup>, que afetou negativamente as demandas por serviços de corridas privadas. Os conjuntos de dados foram normalizados pelo intervalo [0, 1].

<sup>6</sup> <https://www.cdc.gov/mmwr/volumes/69/wr/mm6946a2.htm>

Figura 26 – Mapas de calor nas 86 zonas em Nova Iorque - dia 31/10/2018 no intervalo de 12:00 às 12:30: (A) mapa obtido pelo ALMMo  $\{N = 20\}$  para *Pick-Up*; (B) mapa desejado para *Pick-Up*; (C) mapa obtido pelo ALMMo  $\{N = 20\}$  para *Drop-Off* e; (D) mapa desejado para *Drop-Off*.



### 5.3.1 Previsão da Demanda em Nova Iorque com 63 zonas

A Tabela 12 apresenta o melhor desempenho dos modelos em Nova Iorque, com 63 zonas, para o intervalo de 15 minutos, em *Pick-Up* e *Drop-Off*. O melhor resultado foi apresentado pelo ALMMo (com 20 variáveis de entrada), seguido pelo eMG (com 10 variáveis de entrada) e eNFN (com 5 variáveis de entrada). A Figura 28 ilustra as previsões de *Pick-Up* e *Drop-Off* do ALMMo (com 20 variáveis de entrada), de 10/12/2019 até 30/06/2020 em Yorkville West (zona 263), em intervalos de 15 minutos. Pelo gráfico, nota-se uma diminuição drástica nas demandas no instante 9000. Essa diminuição se deve ao decreto que instaurou o *lockdown* devido à pandemia de COVID-19. Mesmo com a queda da demanda e a mudança na tendência das linhas, o modelo conseguiu acompanhar adequadamente a nova dinâmica dos dados.

Figura 27 – Divisão das 63 zonas em Nova Iorque.

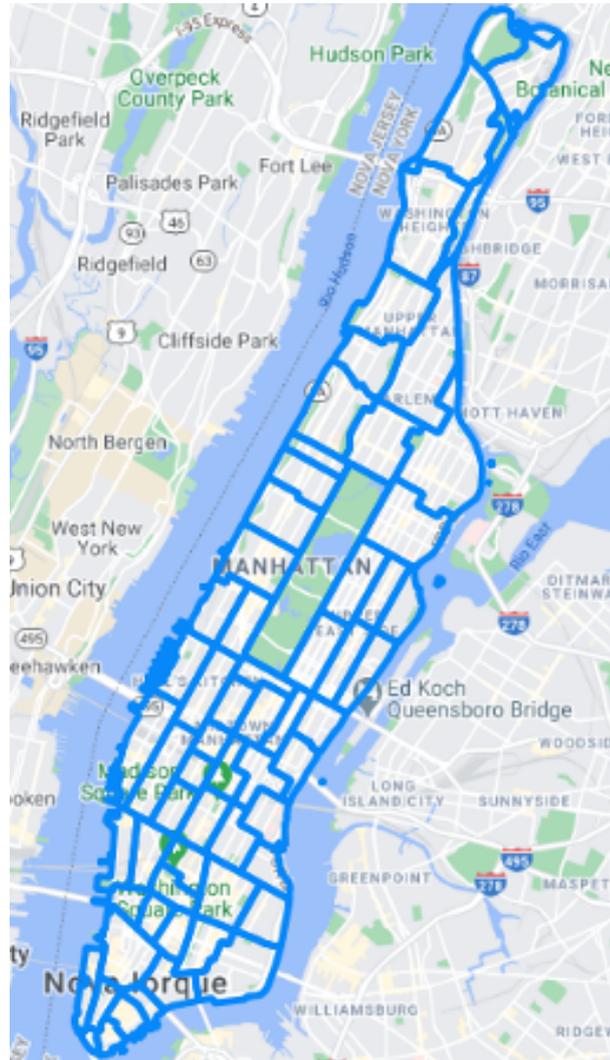
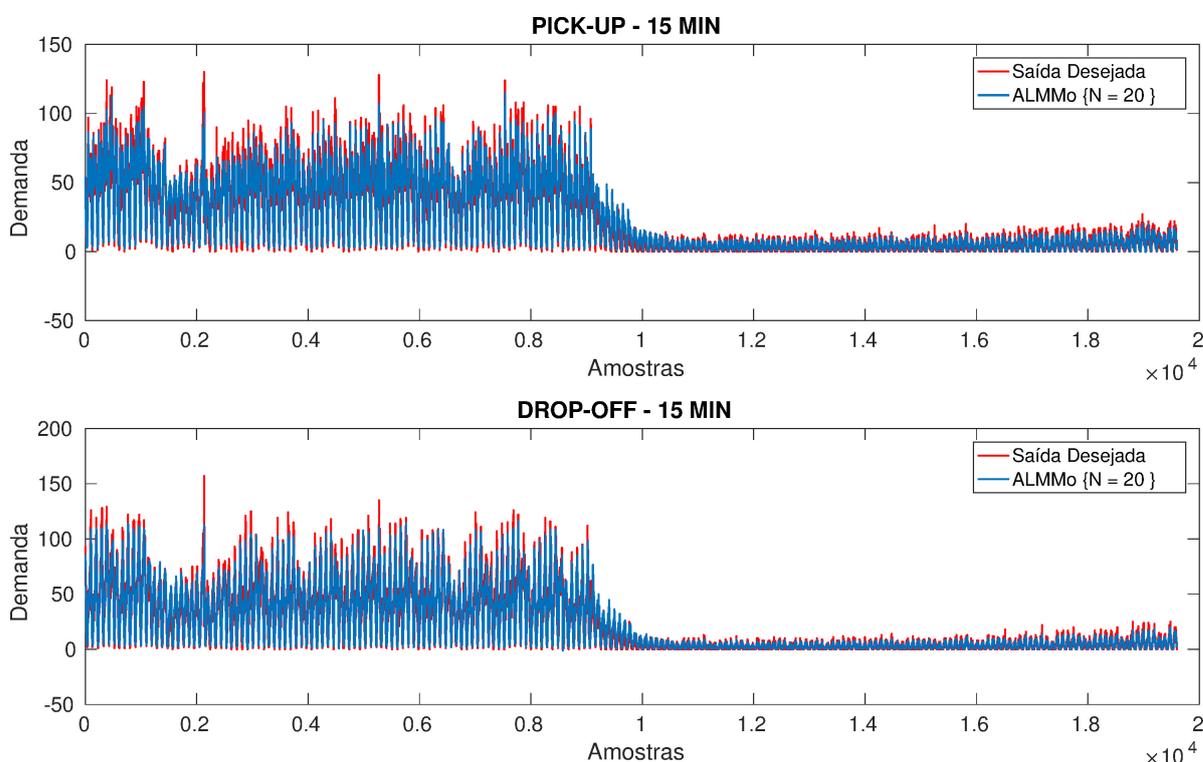


Tabela 12 – Desempenho na previsão da demanda em intervalos de 15 minutos para Nova Iorque com 63 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	7,705	7,145
eMG {N = 10}	8,041	7,518
eNFN {N = 5}	8,897	8,334

A Tabela 13 ilustra os melhores resultados dos modelos para Nova Iorque, com 63 zonas, em intervalos de 30 minutos, para *Pick-Up* e *Drop-Off*. O ALMMo (com 20 variáveis de entrada) foi o modelo que apresentou os melhores resultados. O eMG (com 10 variáveis de entrada) e o eNFN (com 5 variáveis de entrada) descrevem os melhores resultados na sequência. A Figura 29 descreve as previsões de *Pick-Up* e *Drop-Off* do ALMMo (com 20 variáveis de entrada), de 10/12/2019 até 30/06/2020 em Yorkville West (zona 263), em

Figura 28 – Previsões do ALMMo (com 20 variáveis de entrada), em intervalos de 15 minutos, para Yorkville West (zona 263), no período de 10/12/2019 até 30/06/2020.



intervalos de 30 minutos. Assim como visto na Figura 28, há uma mudança drástica no padrão de tendência das demandas. Novamente, o modelo conseguiu manter a previsão adequada a nova dinâmica dos dados, indicando boa adaptabilidade e acurácia em ambos os intervalos de tempo.

Tabela 13 – Desempenho na previsão da demanda em intervalo de 30 minutos para Nova lorque com 63 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	<b>14,087</b>	<b>12,709</b>
eMG {N = 10}	14,536	13,177
eNFN {N = 5}	16,105	14,415

### 5.3.2 Classificação da Demanda em Nova lorque com 63 zonas

Na Tabela 14, são descritas as faixas e os respectivos valores contidos em cada uma das 4 classes para o conjunto de dados de Nova lorque com 63 zonas, onde os valores em parênteses representam a quantidade em cada classe e  $D$ , a demanda prevista. As faixas de valores são representadas graficamente por meio dos histogramas da Figura 30.

Figura 29 – Previsões do ALMMo (com 20 variáveis de entrada), em intervalos de 30 minutos para Yorkville West (zona 263), no período de 10/12/2019 até 30/06/2020.

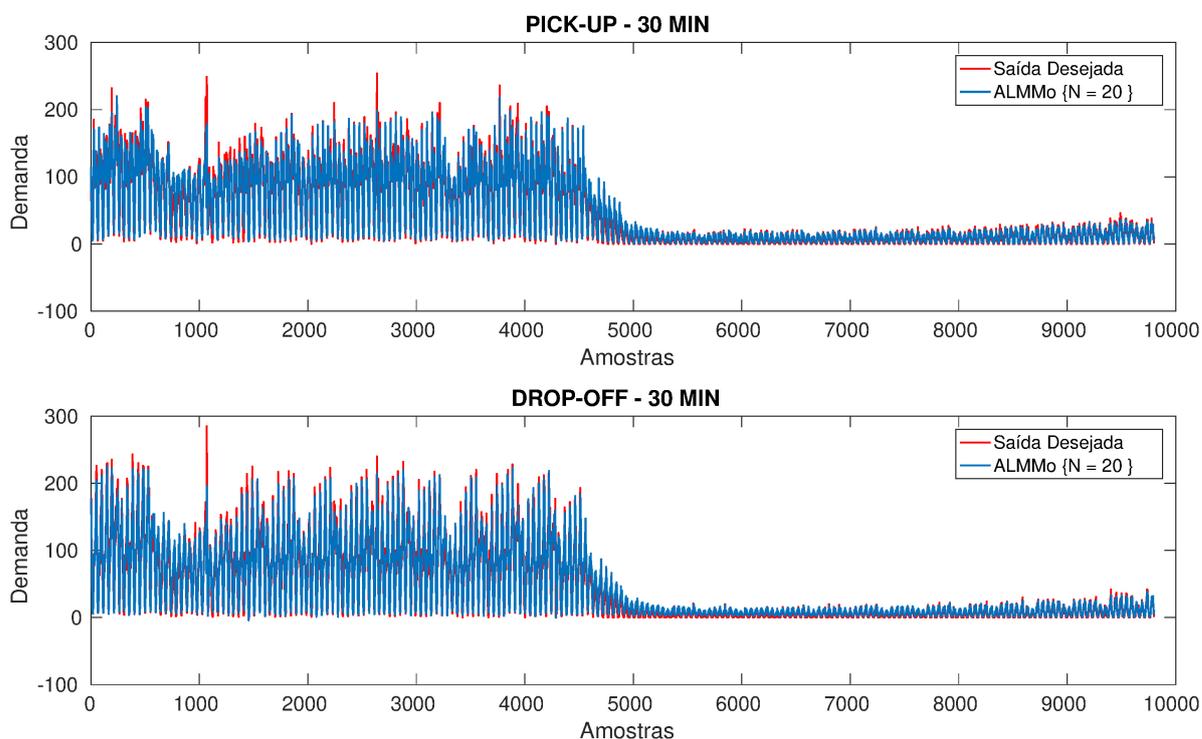


Tabela 14 – Definição das demandas em faixas de valores para Nova lorque com 63 zonas.

Tipo	Intervalo	Faixas e Quantidade			
		Muito Baixa	Baixa	Média	Alta
Pick-Up	15	$D \leq 10$ (315197)	$10 < D \leq 40$ (309900)	$40 < D \leq 80$ (224854)	$D > 80$ (214497)
Pick-Up	30	$D \leq 20$ (154801)	$20 < D \leq 60$ (114231)	$60 < D \leq 120$ (108218)	$D > 120$ (154974)
Drop-Off	15	$D \leq 10$ (301136)	$10 < D \leq 40$ (341414)	$40 < D \leq 80$ (231584)	$D > 80$ (190314)
Drop-Off	30	$D \leq 20$ (146694)	$20 < D \leq 60$ (129418)	$60 < D \leq 120$ (111408)	$D > 120$ (144704)

A acurácia e o desvio padrão para *Pick-Up* e *Drop-Off* obtidos por cada modelo, para o intervalo de 15 minutos em Nova lorque com 63 zonas, são descritos na Tabela 15. O melhor desempenho foi obtido pelo ALMMo (com 20 variáveis). Logo após, segue-se o eMG (com 20 variáveis de entrada) e o eNFN (com 5 variáveis de entrada).

A Figura 31 apresenta os mapas de calor para as 63 zonas de Nova lorque no dia 01/01/2020 para o intervalo das 18:00 às 18:15. A Figura 31 (A) ilustra o mapa obtido com base nos resultados do ALMMo para *Pick-Up* e a Figura 31 (B) o mapa desejado para *Pick-Up*. Em contrapartida, a Figura 31 (C) retrata o mapa para *Drop-Off* construído com base nos resultados do ALMMo e a Figura 31 (D) o desejado para *Drop-Off*. Os destaques em preto

Figura 30 – Histogramas com as faixas de valores em Nova Iorque com 63 zonas.

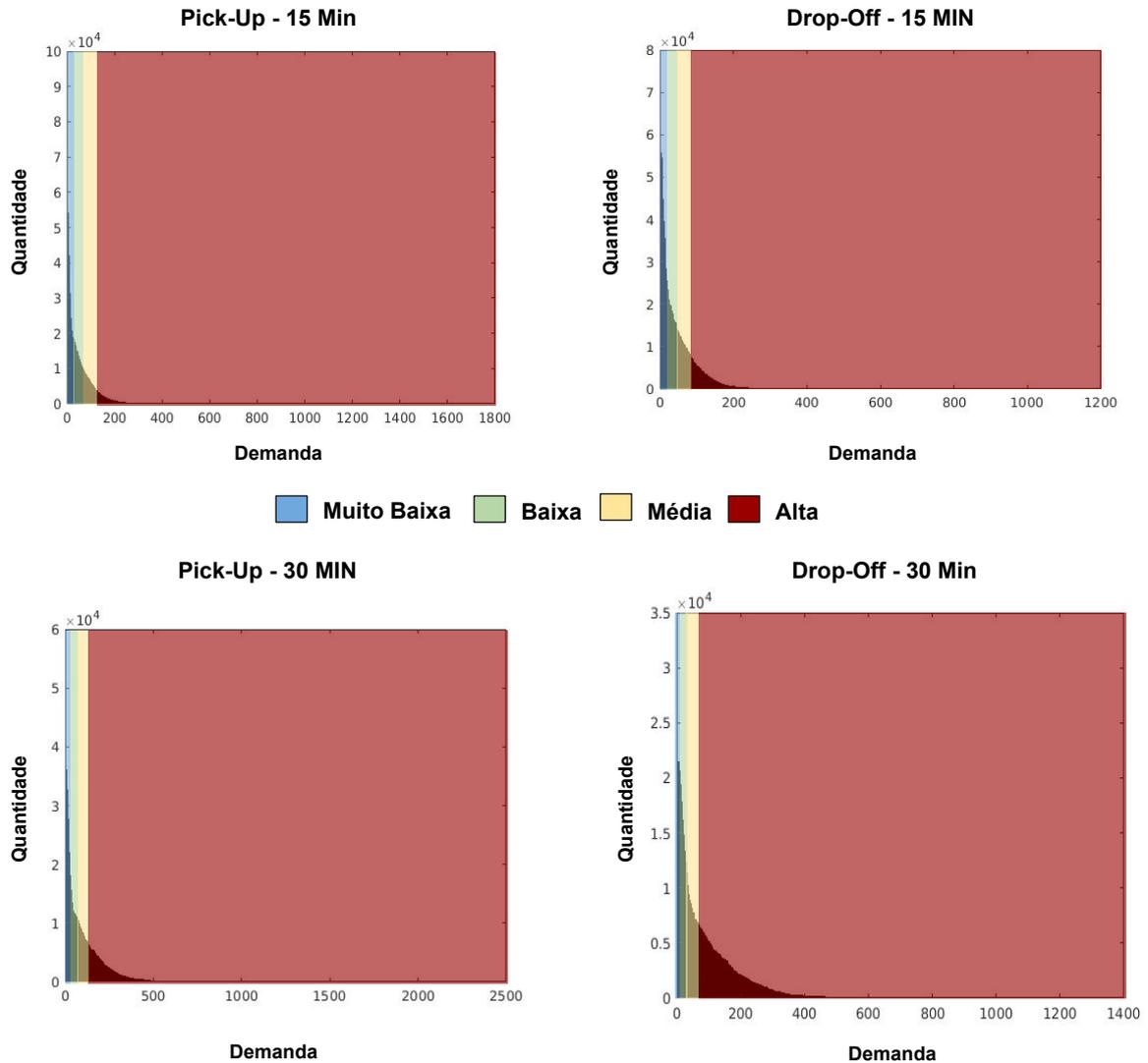


Tabela 15 – Acurácia em intervalos de 15 minutos para Nova Iorque com 63 zonas.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	$\pm$ Desv. Pad.	Acurácia (%)	$\pm$ Desv. Pad.
ALMMo {N = 20}	87,154	5,318	86,524	4,696
eMG {N = 20}	86,537	5,502	85,899	4,926
eNFN {N = 5}	85,171	5,870	84,430	5,236

mostram algumas zonas onde houve divergências na classificação. Como pode ser visto, os poucos erros ocorridos na classificação se mantêm em classes adjacentes.

Para o intervalo de 30 minutos em Nova Iorque com 63 zonas, os resultados são apresentados na Tabela 16. O ALMMo (com 20 variáveis de entrada) foi o modelo com a melhor acurácia, seguido pelo eMG (com 10 variáveis de entrada) e eNFN (com 5 variáveis de entrada).

Figura 31 – Mapas de calor nas 63 zonas em Nova lorque - dia 01/01/2020 no intervalo de 18:00 às 18:15: (A) mapa obtido pelo ALMMo  $\{N = 20\}$  para *Pick-Up*; (B) mapa desejado para *Pick-Up*; (C) mapa obtido pelo ALMMo  $\{N = 20\}$  para *Drop-Off* e; (D) mapa desejado para *Drop-Off*.

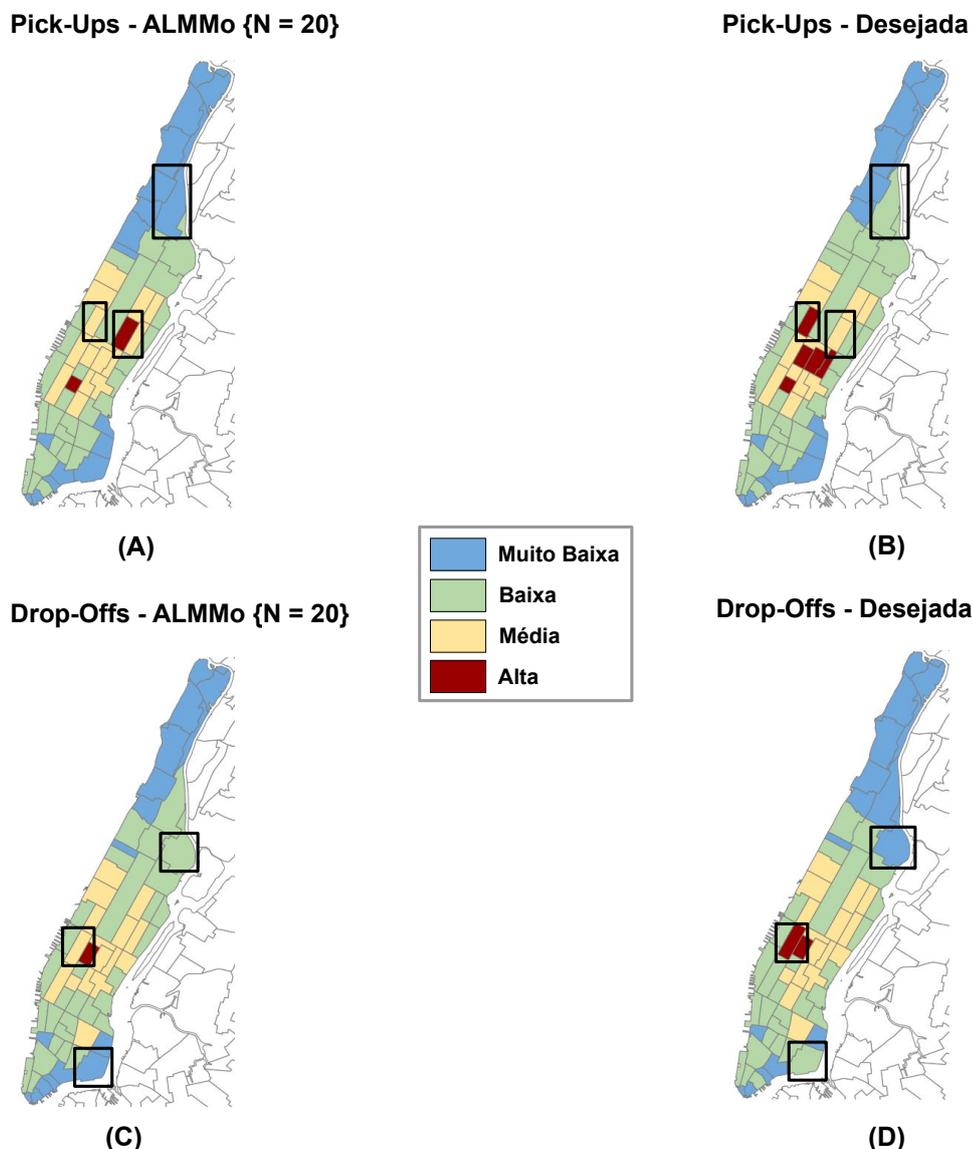


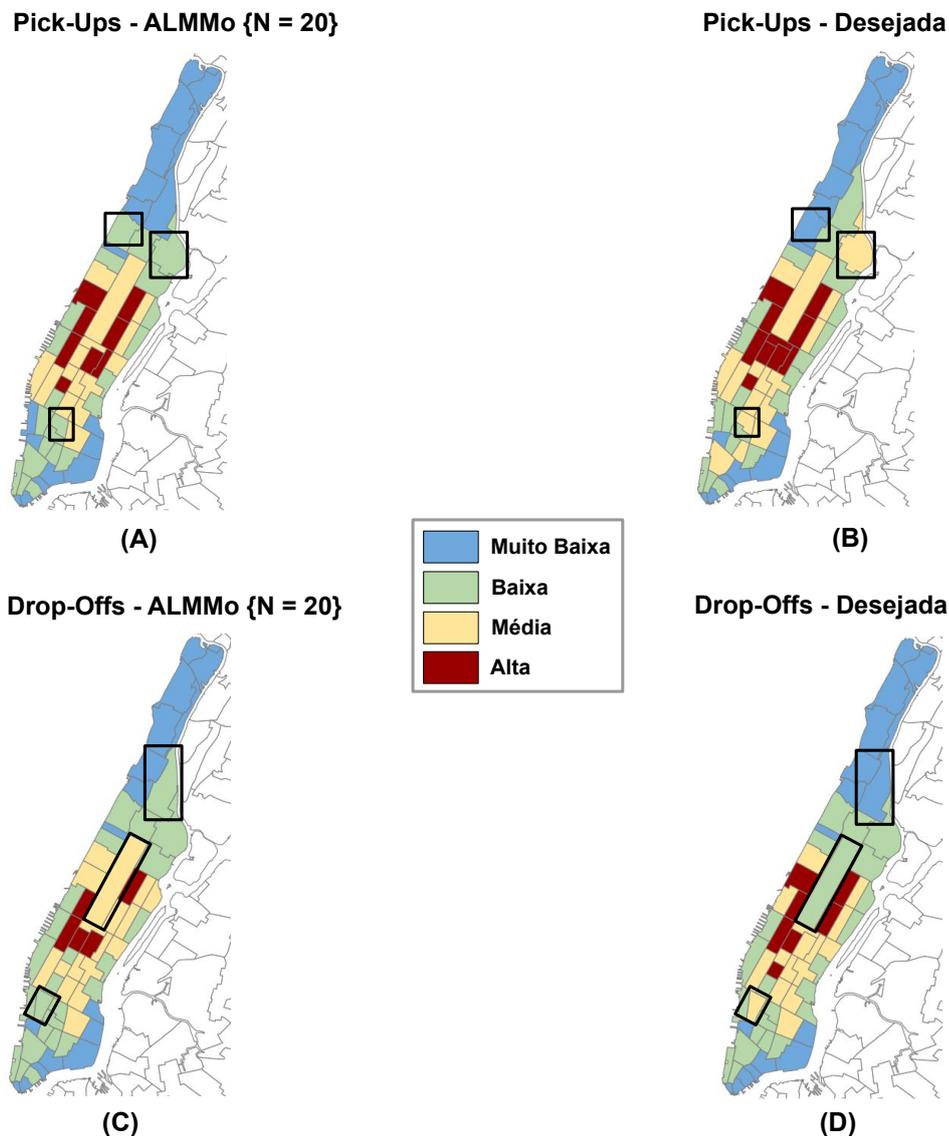
Tabela 16 – Acurácia em intervalos de 30 minutos para Nova lorque com 63 zonas.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo $\{N = 20\}$	88,060	5,123	87,449	4,467
eMG $\{N = 10\}$	87,806	5,111	87,065	4,515
eNFN $\{N = 5\}$	86,252	5,723	85,520	5,164

Os mapas de calor gerados para as 63 zonas de Nova lorque, abrangendo o intervalo das 18:00 às 18:30 no dia 01/01/2020 são apresentados na Figura 32. O mapa construído utilizando os resultados do ALMMO para *Pick-Up* é apresentado na Figura 32 (A) e o

desejado na Figura 32 (B). Por outro lado, o mapa para *Drop-Off* gerado com base nas previsões do ALMMo é ilustrado na Figura 32 (C) e o desejado para *Drop-Off* na Figura 32 (D). Assim como os mapas de calor anteriores, as marcações em preto exemplificam as zonas com divergências entre as classificações. Novamente, como nos mapas supracitados, o erro da classificação é entre classes adjacentes.

Figura 32 – Mapas de calor nas 63 zonas em Nova Iorque - dia 01/01/2020 no intervalo de 18:00 às 18:30: (A) mapa obtido pelo ALMMo  $\{N = 20\}$  para *Pick-Up*; (B) mapa desejado para *Pick-Up*; (C) mapa obtido pelo ALMMo  $\{N = 20\}$  para *Drop-Off* e; (D) mapa desejado para *Drop-Off*.



## 5.4 Análise das Variáveis de Entrada

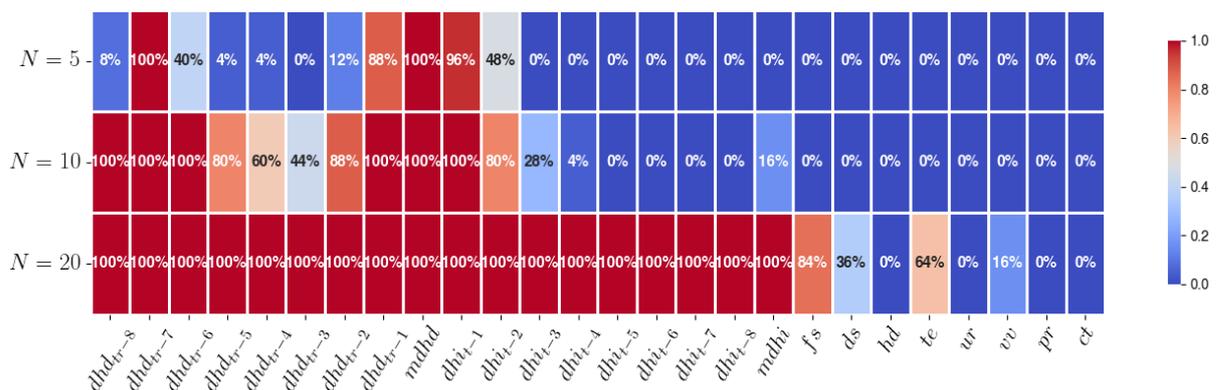
Esta seção apresenta uma análise detalhada das variáveis de entrada identificadas pelo método de *Kruskal-Wallis*. Para auxiliar a visualização e apresentação das informações, serão usados mapas de calor, identificando a porcentagem de seleção de cada variável.

As cores mais quentes indicam porcentagens maiores e as mais frias, menores. Para facilitar o entendimento de como o cálculo da porcentagem é realizado, considere o seguinte exemplo: tome um estudo realizado em 25 zonas para *Pick-Up* de 15 minutos, e a seleção considerando as  $N = 5$  variáveis mais relevantes. Têm-se, portanto, 25 conjuntos de dados. Supondo que a variável  $dhd_{tr-2}$  foi identificada pelo *Kruskal-Wallis* entre as  $N = 5$  variáveis mais relevantes em 24 das 25 zonas, sua porcentagem será de 96% ( $24/25 * 100$ ) e sua representação se dará por uma cor mais quente. A Subseção 5.4.1 detalha as análises das variáveis de entrada na região de Chengdu. Na Subseção 5.4.2, são apresentados os mapas de calor relativos a seleção de variáveis nas 86 zonas de Nova Iorque. Finalmente, é detalhada na Subseção 5.4.3 a seleção de variáveis realizada no subconjunto de 63 zonas em Nova Iorque.

### 5.4.1 Análise das Variáveis de Entrada em Chengdu

A Figura 33 ilustra o mapa de calor das variáveis selecionadas para *Pick-Up* em intervalos de 15 minutos. Nota-se que as demandas dos dias anteriores apresentaram maior percentual de seleção, tanto para seleções de 5, 10 e 20 variáveis. No geral, percebe-se que os dados históricos (demandas de dias anteriores e demanda histórica) são as variáveis mais utilizadas, principalmente quando  $N = \{10, 20\}$ . As variáveis exógenas e auxiliares não sofreram seleção quando  $N = \{5, 10\}$ , mostrando pouca utilidade destas como possíveis candidatas aos modelos. Além disso, quando  $N = 20$ , apenas o final de semana ( $fs$ ) e temperatura ( $te$ ) demonstraram percentual de seleção acima de 50%.

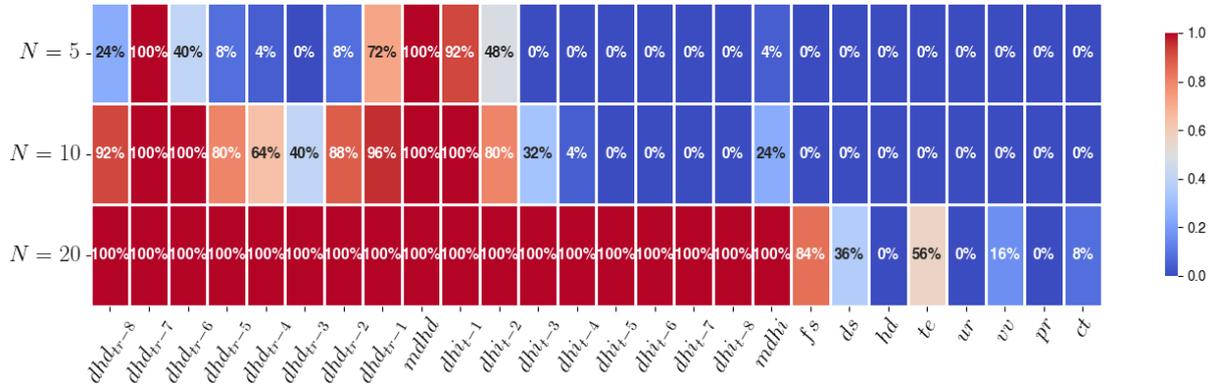
Figura 33 – Mapa de calor das variáveis de entrada para *Pick-Up* em Chengdu com intervalos de 15 minutos.



A Figura 34 apresenta o mapa de calor das variáveis selecionadas para *Drop-Off* em intervalos de 15 minutos. Assim como a Figura 33, percebe-se que os dados históricos são as variáveis com maior percentual de seleção. Em relação as demandas históricas dos intervalos ( $dhi$ ), nota-se que o maior percentual de seleção é visualizado nos dados históricos com menor defasagem ( $t - 1, t - 2$ ). A média da demanda histórica dos dias ( $mdhd$ ), a demanda histórica dos dias ( $dhd$ ) na defasagem  $tr - 7$  e a demanda histórica

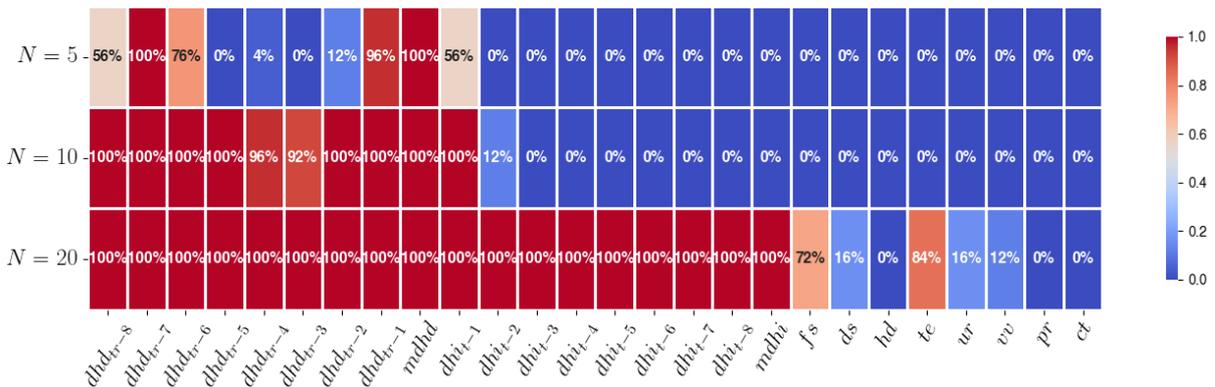
dos intervalos ( $dhi$ ) na defasagem  $t - 1$ , foram as variáveis com percentagem maior do que 90% nas três configurações apresentadas. Mais uma vez, apenas o final de semana ( $fs$ ) e a temperatura ( $te$ ) obtiveram percentual maior do que 50% quando  $N = 20$ .

Figura 34 – Mapa de calor das variáveis de entrada para *Drop-Off* em Chengdu com intervalos de 15 minutos.



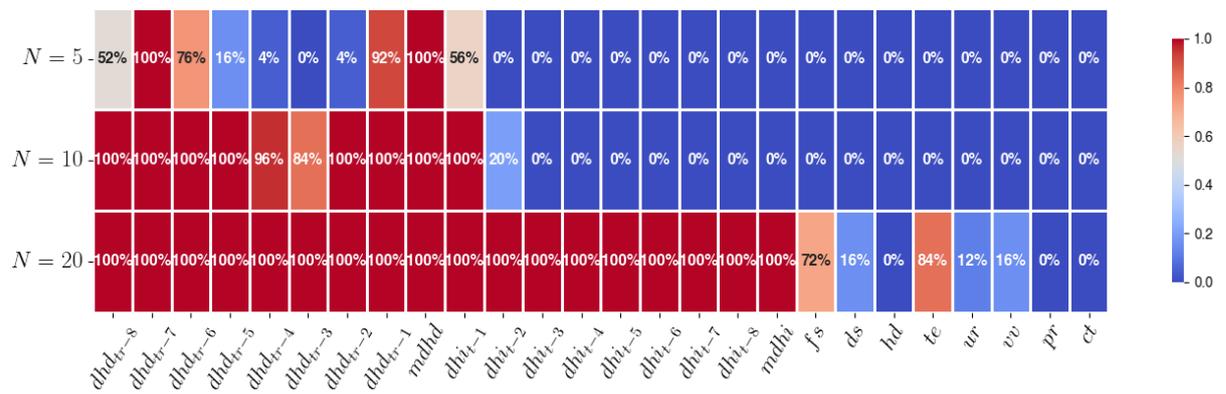
O mapa de calor para *Pick-Up* em intervalos de 30 minutos é descrito na Figura 35. Pode-se observar um padrão semelhante em relação ao mapa de calor para 15 minutos, isto é, a demanda histórica dos dias e a demanda histórica dos intervalos demonstram maior percentual de seleção. Novamente, percebe-se que as variáveis exógenas e auxiliares só possuem percentual de seleção quando  $N = 20$ . Além disso, destas 8 variáveis, apenas o final de semana ( $fs$ ) e temperatura ( $te$ ) se destacaram como relevantes.

Figura 35 – Mapa de calor das variáveis de entrada para *Pick-Up* em Chengdu com intervalos de 30 minutos.



A Figura 36 mostra o mapa de calor para *Drop-Off* em intervalos de 30 minutos. Assim como nos mapas de 15 e 30 minutos, para *Pick-Up* e *Drop-Off*, as variáveis de demanda histórica mostraram alto percentual de seleção. O final de semana ( $fs$ ) e a temperatura ( $te$ ) foram as únicas candidatas com maior representatividade dentre as variáveis exógenas e auxiliares. Este padrão se repete nos mapas anteriores, sugerindo experimentos futuros sem a utilização, por exemplo, do dia da semana ( $ds$ ), horário do dia ( $hd$ ), umidade relativa ( $ur$ ), velocidade do vento ( $vv$ ), precipitação ( $pr$ ) e condição do tempo ( $ct$ ).

Figura 36 – Mapa de calor das variáveis de entrada para *Drop-Off* em Chengdu com intervalos de 30 minutos.



### 5.4.2 Análise das Variáveis de Entrada em Nova Iorque com 86 zonas

As figuras 37 e 38 demonstram os mapas de calor em intervalos de 15 minutos, para *Pick-Up* e *Drop-Off*, respectivamente. Quando  $N = 5$ , todas as demandas dos dias anteriores demonstraram percentuais de seleção, indicando grande relevância como possíveis candidatas aos modelos. De  $N = 5$  para  $N = 10$ , nota-se o incremento da variável  $dhi$  nos tempos  $t - 1$  até  $t - 5$ . Quando  $N = 20$ , todas as variáveis de demanda histórica mostraram-se relevantes. Com relação as variáveis exógenas e auxiliares, o final de semana ( $fs$ ), dia da semana ( $ds$ ) e temperatura ( $te$ ) mostraram percentuais relevantes.

Figura 37 – Mapa de calor das variáveis de entrada para *Pick-Up* em Nova Iorque (86 zonas) com intervalos de 15 minutos.

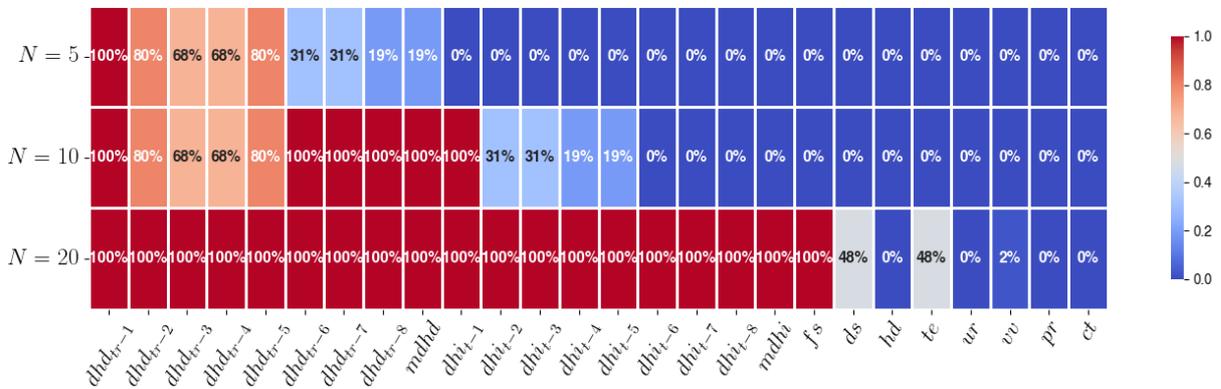
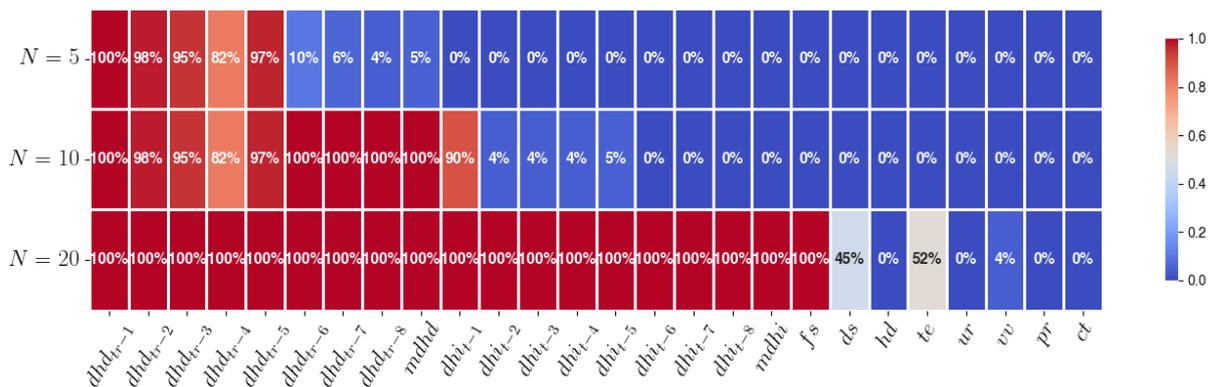


Figura 38 – Mapa de calor das variáveis de entrada para *Drop-Off* em Nova Iorque (86 zonas) com intervalos de 15 minutos.



As figuras 39 e 40 representam os mapas de calor em intervalos de 30 minutos, para *Pick-Up* e *Drop-Off*, respectivamente. Comparado aos mapas de 15 minutos, percebe-se que em  $N = \{5, 10\}$ , além das demandas dos dias anteriores, há uma maior seleção das demandas históricas dos intervalos ( $dhi$ ). Em  $N = 20$ , têm-se apenas o final de semana ( $fs$ ), dia da semana ( $ds$ ) e temperatura ( $te$ ) com algum percentual de seleção.

Figura 39 – Mapa de calor das variáveis de entrada para *Pick-Up* em Nova lorque (86 zonas) com intervalos de 30 minutos.

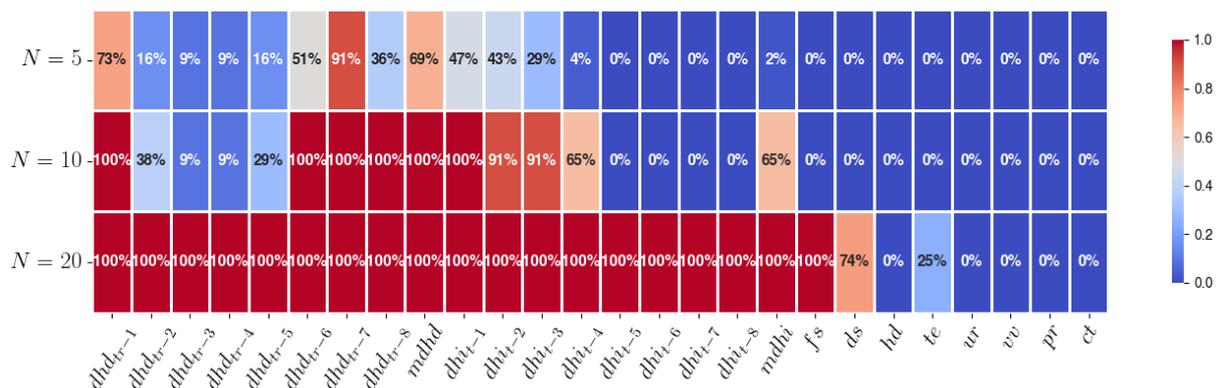
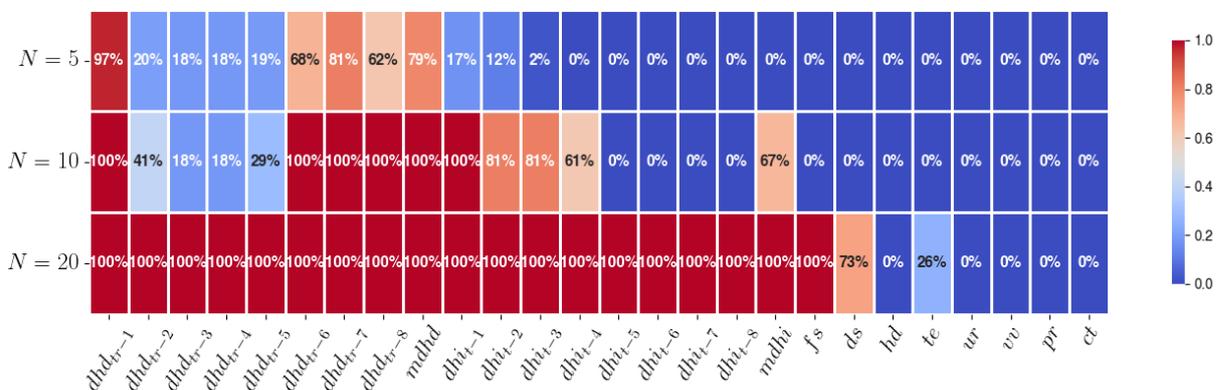


Figura 40 – Mapa de calor das variáveis de entrada para *Drop-Off* em Nova lorque (86 zonas) com intervalos de 30 minutos.



### 5.4.3 Análise das Variáveis de Entrada em Nova Iorque com 63 zonas

Os mapas de calor em intervalos de 15 minutos, para *Pick-Up* e *Drop-Off* são apresentados nas figuras 41 e 42. Quando  $N = \{5, 10\}$ , nota-se que há mais de 70% das demandas históricas com algum percentual de seleção. Ainda nesta configuração, não há percentual para variáveis exógenas e auxiliares. Na configuração  $N = 20$ , além das demandas históricas como principais selecionadas, percebe-se que há percentual de seleção em 7 das 8 variáveis exógenas/auxiliares, sendo o final de semana (*fs*) a variável com maior indicação (> 80 %).

Figura 41 – Mapa de calor das variáveis de entrada para *Pick-Up* em Nova Iorque (63 zonas) com intervalos de 15 minutos.

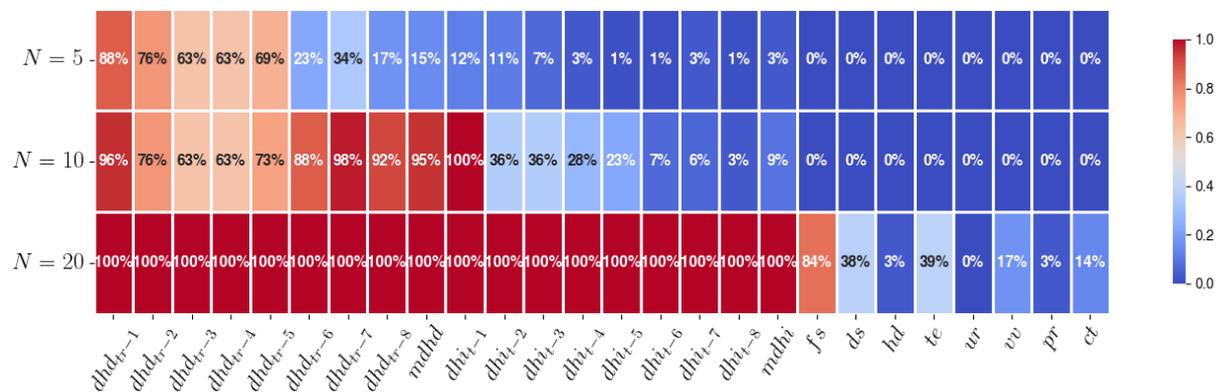
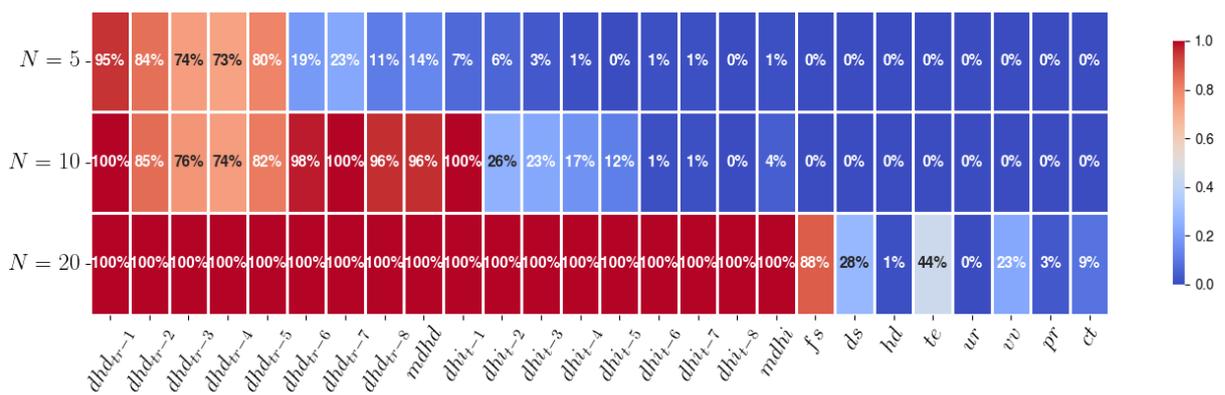


Figura 42 – Mapa de calor das variáveis de entrada para *Drop-Off* em Nova Iorque (63 zonas) com intervalos de 15 minutos.



Em intervalos de 30 minutos, os mapas de calor para *Pick-Up* e *Drop-Off* são ilustrados nas figuras 43 e 44. Para  $N = 5$ , têm-se 3 variáveis com percentual de seleção acima de 50 %: demanda histórica dos dias em  $tr - 1$ ,  $tr - 7$  e média da demanda histórica dos dias (*mdhd*). Em  $N = 10$ , as variáveis com percentual acima de 50 % aumenta para 10. Destaca-se que entre a demanda histórica dos intervalos, as mais relevantes são as que representam uma defasagem temporal menor ( $t - 1, t - 2, t - 3$ ), além da média da demanda histórica dos intervalos (*mdhi*). Quando a configuração é de  $N = 20$ , nota-se que todas as variáveis exógenas/auxiliares apresentaram algum percentual de seleção, porém, apenas uma delas

( $f_s$ ) indicou alto percentual. Sendo assim, conclui-se que há pouca relevância na utilização destas variáveis, indicando uma possível exclusão das mesmas em experimentos futuros.

Figura 43 – Mapa de calor das variáveis de entrada para *Pick-Up* em Nova Iorque (63 zonas) com intervalos de 30 minutos.

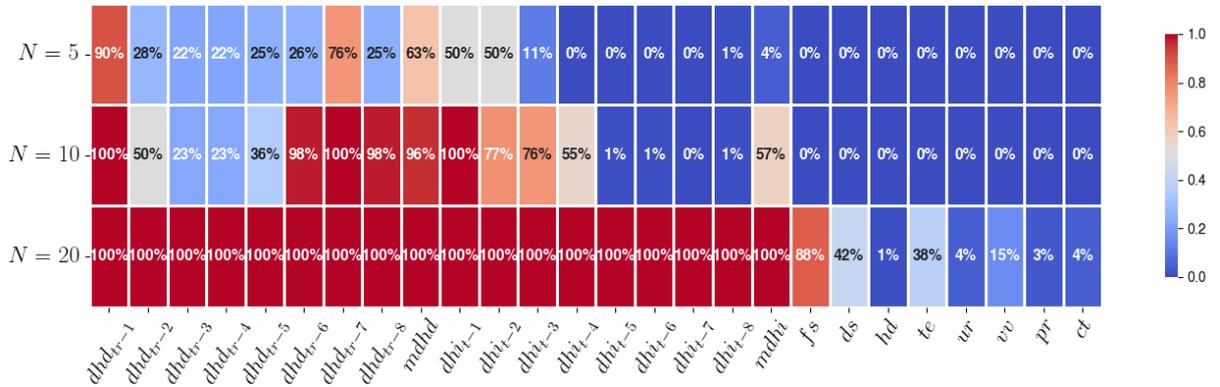
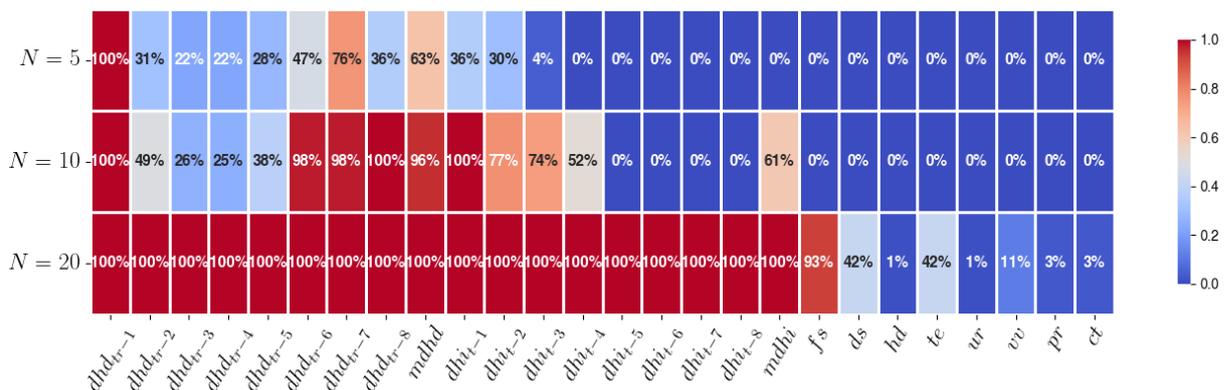


Figura 44 – Mapa de calor das variáveis de entrada para *Drop-Off* em Nova Iorque (63 zonas) com intervalos de 30 minutos.



# Capítulo 6

## Considerações Finais

Este trabalho apresentou uma nova abordagem para a previsão e classificação da demanda de passageiros de táxi utilizando sistemas *fuzzy* evolutivos. A abordagem proposta inclui dados históricos de corridas de táxis e informações meteorológicas como variáveis de entrada. As melhores variáveis são ranqueadas pelo método estatístico de *Kruskal-Wallis* e servem como entrada para um *fuzzy* evolutivo, que realiza a previsão da demanda de táxi para o próximo intervalo de tempo. A classificação da demanda utiliza as saídas da previsão para realizar a classificação considerando quatro classes de demanda.

Para avaliar o desempenho da abordagem proposta, experimentos computacionais foram conduzidos considerando o ALMMo, eMG e eNFN, três sistemas *fuzzy* evolutivos já consolidados para avaliar séries temporais. Os experimentos foram realizados para *Pick-Up* e *Drop-Off* em intervalos de 15 e 30 minutos. Nos experimentos com a abordagem para a previsão da demanda, os resultados de duas abordagens estado da arte são comparados aos resultados previstos pelos modelos evolutivos. Além disso, os modelos são avaliados entre si perante a um extenso período. O ALMMo apresentou os melhores resultados, seja em comparação com os outros modelos evolutivos, seja em comparação com o estado da arte. Os resultados obtidos pelo eMG e eNFN mostraram-se efetivos, principalmente, ao efetuar a previsão por um longo período de tempo. Nos experimentos com a abordagem para a classificação da demanda, a comparação foi realizada somente entre os modelos evolutivos. O ALMMo foi o modelo que apresentou a maior acurácia em todos os experimentos realizados. Todos os algoritmos apresentaram acurácias acima de 80%, mostrando boa efetividade da abordagem para a classificação.

Ao realizar uma análise detalhada das variáveis de entrada, foi possível identificar maiores índices de seleção pelas demandas históricas (*dhd* e *dhi*), juntamente com suas respectivas médias. As variáveis exógenas e auxiliares foram as menos selecionadas pelo método de *Kruskal-Wallis*, indicando pouca relevância destas como candidatas aos modelos. Para experimentos futuros, uma possível sugestão seria descartar as variáveis exógenas e auxili-

ares como possíveis entradas aos modelos, submetendo apenas as demandas históricas ao método de *Kruskal-Wallis*.

Os resultados computacionais e as comparações realizadas sugerem que as abordagens propostas para a previsão e classificação da demanda de passageiros são promissoras para a resolução do problema da oferta e demanda de passageiros. Adicionalmente, são descritos alguns fatores que reforçam a aplicabilidade das abordagens propostas: (i) apresentação do processamento dos dados *online*, diferenciando-se da maioria dos trabalhos com a mesma temática na literatura ; (ii) obtenção de resultados melhores ou semelhantes perante as abordagens do estado da arte e; (iii) consistência nas previsões em extensos períodos de tempo. Além disso, destaca-se a proposta de utilização dos mapas de calor como uma forma de visualização mais intuitiva dos dados, facilitando assim a tomada de decisão mais assertiva.

## 6.1 Perspectivas de Continuidade

Baseando-se nos experimentos realizados e nas conclusões extraídas, sugere-se as seguintes perspectivas de trabalhos futuros:

- Explorar outras técnicas de seleção de características como alternativas ao *Kruskal-Wallis*. Implementar métodos adicionais como complemento à seleção de características.
- Expandir os experimentos utilizando mais modelos evolutivos e bases de corridas de outras cidades.
- Considerar utilizar conjuntos *fuzzy* para classificar as demandas, em alternativa aos histogramas.
- Incorporar, à abordagem, técnicas que analisam e decodificam imagens, com o intuito de capturar informações espaciais das regiões estudadas.
- Estudar novas técnicas de divisão de região e incorporá-las a abordagem proposta.

# Referências

- AMARAL, V. L. et al. New Fuzzy Approaches to Cryptocurrencies Investment Recommendation Systems. **Advances in Intelligent Systems and Computing**, v. 1000, p. 135–147, 2019. ISSN 21945365. Citado na página 3.
- ANDONOVSKI, G. et al. Detection of driver maneuvers using evolving fuzzy cloud-based system. In: **2020 IEEE Symposium Series on Computational Intelligence (SSCI)**. Canberra, ACT, Australia: IEEE, 2020. p. 700–706. Citado na página 3.
- ANGELOV, P. Evolving Takagi-Sugeno Fuzzy Systems from Streaming Data (eTS+). **Evolving Intelligent Systems: Methodology and Applications**, p. 21–50, 2010. Citado na página 16.
- ANGELOV, P. et al. Empirical data analysis: A new tool for data analytics. In: **2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**. Budapest, Hungary: IEEE, 2016. p. 000052–000059. Citado na página 16.
- ANGELOV, P.; YAGER, R. A new type of simplified fuzzy rule-based system. **International Journal of General Systems**, Taylor & Francis, v. 41, n. 2, p. 163–185, 2012. Citado na página 16.
- ANGELOV, P.; ZHOU, X. Evolving fuzzy systems from data streams in real-time. In: **IEEE. 2006 International symposium on evolving fuzzy systems**. Ambleside, UK: IEEE, 2006. p. 29–35. Citado na página 2.
- ANGELOV, P. P.; GU, X. Applications of autonomous learning multi-model systems. In: **Empirical Approach to Machine Learning**. Studies in Computational Intelligence, vol 800: Springer, Cham, 2019. p. 277–293. Citado na página 3.
- ANGELOV, P. P.; GU, X.; PRINCIPE, J. C. Autonomous Learning Multimodel Systems from Data Streams. **IEEE Transactions on Fuzzy Systems**, v. 26, n. 4, p. 2213–2224, 2017. ISSN 10636706. Citado 4 vezes nas páginas 3, 16, 17 e 41.
- BABICKI, S. et al. Heatmapper: web-enabled heat mapping for all. **Nucleic acids research**, Oxford University Press, v. 44, n. W1, p. W147–W153, 2016. Citado na página 13.
- BAI, L. et al. Passenger demand forecasting with multi-task convolutional recurrent neural networks. In: **Pacific-Asia Conference on Knowledge Discovery and Data Mining**. Macau, China: Springer, 2019. p. 29–42. Citado 4 vezes nas páginas 1, 10, 27 e 28.
- BELLMAN, R. Dynamic programming. **Science**, American Association for the Advancement of Science, v. 153, n. 3731, p. 34–37, 1966. Citado na página 35.
- BEN-DAVID, S.; KUSHILEVITZ, E.; MANSOUR, Y. Online Learning versus Offline Learning. **Machine Learning**, v. 29, n. 1, p. 45–63, 1997. ISSN 08856125. Citado na página 14.
- BOJKO, A. A. Informative or misleading? heatmaps deconstructed. In: JACKO, J. A. (Ed.). **Human-Computer Interaction. New Trends**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 30–39. ISBN 978-3-642-02574-7. Citado na página 12.

BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. Hoboken, Nova Jersey, EUA: Wiley, 1976. Citado na página 32.

CHATFIELD, C. **Time-series forecasting**. 1. ed. CRC: CRC press, 2000. Citado 2 vezes nas páginas 9 e 15.

CRAMPTON, J. Gis and geographic governance: reconstructing the choropleth map. **Cartographica: The International Journal for Geographic Information and Geovisualization**, University of Toronto Press, v. 39, n. 1, p. 41–53, 2004. Citado na página 13.

CUNHA, F. D. da et al. Sistemas de Transporte Inteligentes: Conceitos, Aplicações e Desafios. In: **XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC) - 2017**. Belém, Pará: SBC, 2017. p. 1–47. Citado 3 vezes nas páginas 1, 6 e 8.

DAVIS, N.; RAINA, G.; JAGANNATHAN, K. Taxi Demand Forecasting: A HEDGE-Based Tessellation Strategy for Improved Accuracy. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 19, n. 11, p. 3686–3697, 2018. ISSN 15249050. Citado 3 vezes nas páginas 12, 13 e 28.

DIMITRAKOPOULOS, G.; DEMESTICHAS, P. Intelligent transportation systems: Systems based on cognitive networking principles and management functionality. **IEEE Vehicular Technology Magazine**, v. 5, n. 1, p. 77–84, 2010. ISSN 15566072. Citado na página 6.

FAGHIH, S. S. et al. Predicting Short-Term Uber Demand in New York City Using Spatio-temporal Modeling. **Journal of Computing in Civil Engineering**, v. 33, n. 3, p. 05019002, 2019. ISSN 0887-3801. Citado 2 vezes nas páginas 26 e 29.

GARCIA, C.; LEITE, D.; ŠKRJANC, I. Incremental missing-data imputation for evolving fuzzy granular prediction. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 28, n. 10, p. 2348–2362, 2019. Citado na página 3.

GEORGE, S.; SANTRA, A. K. An improved long short-term memory networks with Takagi-Sugeno fuzzy for traffic speed prediction considering abnormal traffic situation. **Computational Intelligence**, v. 36, n. 3, p. 964–993, 2020. ISSN 14678640. Citado na página 29.

HE, F. et al. Pricing and penalty/compensation strategies of a taxi-hailing platform. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 86, n. June 2017, p. 263–279, 2018. ISSN 0968090X. Citado na página 8.

HEGARTY, C. J. The Global Positioning System (GPS). **Springer Handbooks**, v. 50, n. 3, p. 197–218, 2017. ISSN 25228706. Citado na página 9.

INÁCIO, M. J. **Diagnóstico de Falhas baseado em Sistema Inteligente Evolutivo**. 105 p. Tese (Doutorado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, 2014. Citado na página 16.

ISHIGURO, S.; KAWASAKI, S.; FUKAZAWA, Y. Taxi demand forecast using real-time population generated from cellular networks. In: **Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers**. Singapore: Association for Computing Machinery, New York, 2018. p. 1024–1032. Citado 2 vezes nas páginas 26 e 28.

JIANG, S. et al. Short-term demand prediction method for online car-hailing services based on a least squares support vector machine. **IEEE Access**, IEEE, v. 7, n. c, p. 11882–11891, 2019. ISSN 21693536. Citado 4 vezes nas páginas [1](#), [2](#), [25](#) e [28](#).

JUNIOR, G. A. da S.; SILVA, A. M. da. A simple and efficient incremental missing data imputation method for evolving neo-fuzzy network. **Evolving Systems**, Springer, p. 1–20, 2021. Citado na página [3](#).

KASABOV, N.; FILEV, D. Evolving intelligent systems: Methods, learning, applications. In: **2006 International Symposium on Evolving Fuzzy Systems**. Ambleside, UK: IEEE, 2006. p. 8–18. Citado na página [16](#).

KE, J. et al. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 85, n. June, p. 591–608, 2017. ISSN 0968090X. Citado 6 vezes nas páginas [2](#), [10](#), [25](#), [28](#), [33](#) e [34](#).

KONG, X. et al. TBI2Flow: Travel behavioral inertia based long-term taxi passenger flow prediction. **World Wide Web**, World Wide Web, v. 23, n. 2, p. 1381–1405, 2020. ISSN 15731413. Citado 2 vezes nas páginas [26](#) e [29](#).

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American statistical Association**, Taylor & Francis Group, v. 47, n. 260, p. 583–621, 1952. Citado na página [35](#).

LEITE, D.; ŠKRJANC, I.; GOMIDE, F. An overview on evolving systems and learning from stream data. **Evolving Systems**, Springer Berlin Heidelberg, p. 1–18, 2020. Citado na página [15](#).

LEMONS, A.; CAMINHAS, W.; GOMIDE, F. Multivariable Gaussian Evolving Fuzzy Modeling System. **IEEE Transactions on Fuzzy Systems**, v. 19, n. 1, p. 91–104, 2011. ISSN 10636706. Citado 6 vezes nas páginas [15](#), [16](#), [19](#), [20](#), [21](#) e [41](#).

LI, B. et al. Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset. In: **2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)**. Seattle, WA, USA: IEEE, 2011. p. 63–68. Citado na página [24](#).

LIAO, S. et al. Large-scale short-term urban taxi demand forecasting using deep learning. In: IEEE. **2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC)**. Jeju, Korea (South): IEEE, 2018. p. 428–433. Citado 3 vezes nas páginas [25](#), [26](#) e [28](#).

LIN, C.-Y. et al. Predictions of taxi demand based on neural network algorithms. **International Journal of Intelligent Transportation Systems Research**, Springer, v. 19, n. 3, p. 477–495, 2021. Citado 2 vezes nas páginas [26](#) e [29](#).

LIU, L. et al. Contextualized spatial–temporal network for taxi origin-destination demand prediction. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 20, n. 10, p. 3875–3887, 2019. Citado 3 vezes nas páginas [2](#), [26](#) e [29](#).

LIU, T. et al. Predicting taxi demands via an attention-based convolutional recurrent neural network. **Knowledge-Based Systems**, Elsevier, v. 206, p. 106294, 2020. Citado 2 vezes nas páginas [25](#) e [29](#).

LIU, Z. et al. Taxi demand prediction based on a combination forecasting model in hotspots. **Journal of Advanced Transportation**, Hindawi, v. 2020, 2020. Citado 2 vezes nas páginas 26 e 29.

LUGHOFER, E. Evolving fuzzy systems: Fundamentals, reliability, interpretability, useability and applications. In: **2015 7th International Joint Conference on Computational Intelligence (IJCCI)**. Lisbon, Portugal: IEEE, 2015. v. 1, p. 11–11. Citado 2 vezes nas páginas 2 e 16.

LUGHOFER, E.; ANGELOV, P. Handling drifts and shifts in on-line data streams with evolving fuzzy systems. **Applied Soft Computing Journal**, Elsevier B.V., v. 11, n. 2, p. 2057–2068, 2011. ISSN 15684946. Citado 2 vezes nas páginas 2 e 23.

LUO, H. et al. A multi-task deep learning model for short-term taxi demand forecasting considering spatiotemporal dependences. **Journal of Traffic and Transportation Engineering (English Edition)**, Elsevier, v. 8, n. 1, p. 83–94, 2021. Citado 2 vezes nas páginas 27 e 29.

LÜTKEPOHL, H. **New Introduction to Multiple Time Series Analysis**. Springer: Springer Science Business Media, 2005. Citado na página 33.

MACIEJEWSKI, M.; BISCHOFF, J.; NAGEL, K. An Assignment-Based Approach to Efficient Real-Time City-Scale Taxi Dispatching. **IEEE Intelligent Systems**, v. 31, n. 1, p. 68–77, 2016. ISSN 15411672. Citado na página 2.

MARKOU, I.; RODRIGUES, F.; PEREIRA, F. C. Real-time taxi demand prediction using data from the web. In: **2018 21st International Conference on Intelligent Transportation Systems (ITSC)**. Maui, HI, USA: IEEE, 2018. p. 1664–1671. Citado 2 vezes nas páginas 25 e 28.

MOREIRA-MATIAS, L. et al. A predictive model for the passenger demand on a taxi network. In: IEEE. **2012 15th International IEEE Conference on Intelligent Transportation Systems**. Anchorage, AK, USA: IEEE, 2012. p. 1014–1019. Citado 3 vezes nas páginas 2, 24 e 40.

MOREIRA-MATIAS, L. et al. Predicting taxi-passenger demand using streaming data. **IEEE Transactions on Intelligent Transportation Systems**, v. 14, n. 3, p. 1393–1402, 2013. ISSN 15249050. Citado 4 vezes nas páginas 2, 9, 24 e 40.

MOREIRA-MATIAS, L. et al. Time-evolving O-D matrix estimation using high-speed GPS data streams. **Expert Systems with Applications**, Elsevier Ltd, v. 44, p. 275–288, 2016. ISSN 09574174. Citado 2 vezes nas páginas 9 e 12.

NOVAKOVIĆ, J. D. et al. Evaluation of classification models in machine learning. **Theory and Applications of Mathematics & Computer Science**, v. 7, n. 1, p. 39–46, 2017. Citado na página 11.

QU, F.; WANG, F. Y.; YANG, L. Intelligent transportation spaces: Vehicles, traffic, communications, and beyond. **IEEE Communications Magazine**, v. 48, n. 11, p. 136–142, 2010. ISSN 01636804. Citado na página 7.

RODRIGUES, F.; MARKOU, I.; PEREIRA, F. C. Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach. **Information Fusion**, Elsevier B.V., v. 49, p. 120–129, 2019. ISSN 15662535. Citado na página 12.

RODRIGUES, F. P. S.; SILVA, A. M.; LEMOS, A. P. An evolving fuzzy predictor with multivariable gaussian participatory learning and multi-innovations recursive weighted least squares - efmi. **Evolving Systems**, Springer, p. 1–17, 2022. Citado na página 33.

ROSSI, A. et al. Modelling Taxi Drivers' Behaviour for the Next Destination Prediction. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 21, n. 7, p. 2980–2989, 2019. ISSN 15580016. Citado na página 29.

SALMAN, R.; KECMAN, V. Regression as classification. In: **2012 Proceedings of IEEE Southeastcon**. Orlando, Florida, USA: Institute of Electrical and Electronics Engineers (IEEE), 2012. Citado na página 36.

SAYED-MOUCHAWEH, M.; LUGHOFER, E. **Learning in Non-Stationary Environments**. 1. ed. New York: Springer-Verlag New York, 2012. 440 p. ISBN 9781441980199. Citado na página 15.

SEN, P. C.; HAJRA, M.; GHOSH, M. Supervised classification algorithms in machine learning: A survey and review. In: **Emerging technology in modelling and graphics**. Singapore: Springer, 2020. p. 99–111. Citado na página 11.

SILVA, A. M. et al. A fast learning algorithm for evolving neo-fuzzy neuron. **Applied Soft Computing**, Elsevier, v. 14, p. 194–209, 2014. Citado 3 vezes nas páginas 16, 22 e 41.

SILVA, A. M. da. **Sistemas Neuro-Fuzzy Evolutivos: Novos Algoritmos de Aprendizado e Aplicações**. 145 p. Tese (Doutorado) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, 2014. Citado 2 vezes nas páginas 15 e 22.

SINGH, B.; GUPTA, A. Recent trends in intelligent transportation systems: a review. **Journal of Transport Literature**, v. 9, n. 2, p. 30–34, 2015. ISSN 2238-1031. Citado na página 6.

SOARES, E.; ANGELOV, P.; GU, X. Autonomous learning multiple-model zero-order classifier for heart sound classification. **Applied Soft Computing**, Elsevier, v. 94, p. 106449, 2020. Citado na página 3.

SOARES, E. et al. Ensemble of evolving data clouds and fuzzy models for weather time series prediction. **Applied Soft Computing**, Elsevier, v. 64, p. 445–453, 2018. Citado na página 3.

SOARES, E. A. et al. Incremental gaussian granular fuzzy modeling applied to hurricane track forecasting. In: IEEE. **2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)**. Rio de Janeiro, Brazil: IEE, 2018. p. 1–8. Citado na página 3.

SOUZA, P. V. d. C.; LUGHOFER, E.; GUIMARAES, A. J. Evolving fuzzy neural network based on uni-nullneuron to identify auction fraud. In: **19th World Congress of the International Fuzzy Systems Association (IFSA), 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and 11th International Summer School on Aggregation Operators (AGOP)**. Bratislava, Slovakia: Atlantis Press, 2021. p. 314–321. Citado na página 3.

SOUZA, P. V. de C.; LUGHOFER, E. An evolving neuro-fuzzy system based on uni-nullneurons with advanced interpretability capabilities. **Neurocomputing**, v. 451, p. 231–251, 2021. ISSN 0925-2312. Citado na página 3.

SUSSMAN, J. S. **Perspectives on intelligent transportation systems (ITS)**. Springer: Springer Science & Business Media, 2005. Citado 2 vezes nas páginas 7 e 8.

TANG, J. et al. An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 18, n. 9, p. 2340–2350, 2017. Citado na página 3.

TONG, Y. et al. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. In: **Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining**. Halifax NS Canada: Association for Computing Machinery, New York, 2017. p. 1653–1662. Citado 3 vezes nas páginas 9, 28 e 34.

TYAGI, S.; MITTAL, S. Sampling approaches for imbalanced data classification problem in machine learning. In: **Proceedings of ICRIC 2019**. Springer, Cham: Springer, 2020. p. 209–221. Citado na página 37.

VANICHRUJEE, U. et al. Taxi demand prediction using ensemble model based on rnns and xgboost. In: **2018 International Conference on Embedded Systems and Intelligent Technology International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES)**. Khon Kaen, Thailand: IEEE, 2018. p. 1–6. Citado 3 vezes nas páginas 12, 25 e 28.

WANG, Y. et al. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling. In: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. Anchorage AK USA: Association for Computing Machinery, New York, 2019. p. 1227–1235. Citado 3 vezes nas páginas 11, 27 e 28.

WEI, H. et al. Zest: a hybrid model on predicting passenger demand for chauffeured car service. In: **Proceedings of the 25th ACM International on Conference on Information and Knowledge Management**. Indianapolis Indiana USA: Association for Computing Machinery, New York, 2016. p. 2203–2208. Citado 2 vezes nas páginas 2 e 9.

XIE, P. et al. Urban flow prediction from spatiotemporal data using machine learning: A survey. **Information Fusion**, Elsevier B.V., v. 59, p. 1–12, 2020. ISSN 15662535. Citado na página 29.

XU, J. et al. A sequence learning model with recurrent neural networks for taxi demand prediction. In: **2017 IEEE 42nd Conference on Local Computer Networks (LCN)**. Singapore: IEEE, 2017. p. 261–268. Citado 2 vezes nas páginas 2 e 9.

XU, J. et al. Real-Time prediction of taxi demand using recurrent neural networks. **IEEE Transactions on Intelligent Transportation Systems**, v. 19, n. 8, p. 2572–2581, 2018. ISSN 15249050. Citado 3 vezes nas páginas 2, 25 e 28.

YAGER, R. A model of participatory learning. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 20, n. 5, p. 1229–1234, 1990. Citado na página 19.

YAMAKAWA, T. A neo fuzzy neuron and its applications to system identification and prediction of the system behavior. In: **Proc. of the 2nd Int. Conf. on Fuzzy Logic & Neural Networks**. Lizuka, Japan: Fuzzy Logic Systems Institute, 1992. v. 1, p. 477–483. Citado na página 22.

YANG, C.; GONZALES, E. J. Modeling taxi demand and supply in New York city using large-scale taxi GPS data. **Springer Geography**, p. 405–425, 2017. ISSN 21943168. Citado 2 vezes nas páginas [25](#) e [28](#).

YANG, H. et al. Nonlinear pricing of taxi services. **Transportation Research Part A: Policy and Practice**, Elsevier Ltd, v. 44, n. 5, p. 337–348, 2010. ISSN 09658564. Citado na página [8](#).

YAO, H. et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In: **Proceedings of the AAAI conference on artificial intelligence**. Honolulu, Hawaii, USA: AAAI Press, 2019. v. 33, n. 01, p. 5668–5675. Citado na página [2](#).

YAO, H. et al. Deep multi-view spatial-temporal network for taxi demand prediction. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. New Orleans, Louisiana, USA: AAAI Press, 2018. v. 32, n. 1. Citado 4 vezes nas páginas [1](#), [26](#), [28](#) e [33](#).

YU, H. et al. Taxi-Based Mobility Demand Formulation and Prediction Using Conditional Generative Adversarial Network-Driven Learning Approaches. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 20, n. 10, p. 3888–3899, 2019. ISSN 15580016. Citado na página [28](#).

ZANDER, G. **Predicting taxi passenger demand using artificial neural networks**. 51 p. Dissertação (Mestrado) — KTH, School of Computer Science and Communication (CSC), 2017. Citado 2 vezes nas páginas [11](#) e [25](#).

ZHAN, X.; QIAN, X.; UKKUSURI, S. V. A graph-based approach to measuring the efficiency of an urban taxi service system. **IEEE Transactions on Intelligent Transportation Systems**, v. 17, n. 9, p. 2479–2489, 2016. Citado na página [2](#).

ZHANG, C. et al. Mlrnn: Taxi demand prediction based on multi-level deep learning and regional heterogeneity analysis. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, 2021. Citado 2 vezes nas páginas [25](#) e [29](#).

ZHANG, C. et al. Taxi Demand Prediction Using Parallel Multi-Task Learning Model. **IEEE Transactions on Intelligent Transportation Systems**, p. 1–10, 2020. ISSN 1524-9050. Citado 13 vezes nas páginas [2](#), [9](#), [10](#), [11](#), [12](#), [27](#), [29](#), [33](#), [40](#), [41](#), [49](#), [50](#) e [51](#).

ZHANG, D. et al. Understanding taxi service strategies from taxi GPS traces. **IEEE Transactions on Intelligent Transportation Systems**, v. 16, n. 1, p. 123–135, 2015. ISSN 15249050. Citado na página [24](#).

ZHANG, K.; LIU, Z.; ZHENG, L. Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 21, n. 4, p. 1480–1490, 2019. Citado 11 vezes nas páginas [2](#), [9](#), [11](#), [27](#), [29](#), [40](#), [41](#), [42](#), [43](#), [44](#) e [54](#).

ZHAO, J. et al. Unifying uber and taxi data via deep models for taxi passenger demand prediction. **Personal and Ubiquitous Computing**, Springer, p. 1–13, 2020. Citado 2 vezes nas páginas [27](#) e [29](#).

ZHOU, Y. et al. St-attn: Spatial-temporal attention mechanism for multi-step citywide crowd flow prediction. In: **2019 International Conference on Data Mining Workshops (ICDMW)**. Beijing, China: IEEE, 2019. p. 609–614. Citado na página [28](#).

ZHU, K. et al. A novel hybrid deep learning model for taxi demand forecasting based on decomposition of time series and fusion of text data. **Journal of Intelligent & Fuzzy Systems**, v. 41, p. 3355–3371, 2021. Citado 2 vezes nas páginas [26](#) e [29](#).

# Apêndices

# APÊNDICE A – Lista das Condições do Tempo

Tabela 17 – Índices e condições do tempo.

Índice	Descrição	Índice	Descrição
1	<i>Fog</i>	26	<i>Light Freezing Rain</i>
2	<i>Light Rain</i>	27	<i>Light Rain / Windy</i>
3	<i>Fair</i>	28	<i>Light Rain with Thunder</i>
4	<i>Partly Cloudy</i>	29	<i>Light Sleet</i>
5	<i>Haze / Cloudy</i>	30	<i>Light Snow</i>
6	<i>Haze</i>	31	<i>Light Snow and Sleet</i>
7	<i>Mostly Cloudy</i>	32	<i>Light Snow and Sleet / Windy</i>
8	<i>Blowing Snow / Windy</i>	33	<i>Light Snow / Windy</i>
9	<i>Cloudy</i>	34	<i>Mostly Cloudy / Windy</i>
10	<i>Cloudy / Windy</i>	35	<i>Partly Cloudy / Windy</i>
11	<i>Drizzle and Fog</i>	36	<i>Rain</i>
12	<i>Drizzle and Fog / Windy</i>	37	<i>Rain / Windy</i>
13	<i>Drizzle / Windy</i>	38	<i>Small Hail</i>
14	<i>Fair / Windy</i>	39	<i>Snow</i>
15	<i>Fog / Windy</i>	40	<i>Snow and Sleet</i>
16	<i>Haze / Windy</i>	41	<i>Snow and Sleet / Windy</i>
17	<i>Heavy Rain</i>	42	<i>Snow / Windy</i>
18	<i>Heavy Rain / Windy</i>	43	<i>Squalls / Windy</i>
19	<i>Heavy Snow</i>	44	<i>Thunder</i>
20	<i>Heavy Snow / Windy</i>	45	<i>Thunder in the Vicinity</i>
21	<i>Heavy T-Storm</i>	46	<i>Thunder / Windy</i>
22	<i>Heavy T-Storm / Windy</i>	47	<i>T-Storm</i>
23	<i>Light Drizzle</i>	48	<i>T-Storm / Windy</i>
24	<i>Light Drizzle / Windy</i>	49	<i>Wintry Mix</i>
25	<i>Light Freezing Drizzle</i>	50	<i>Wintry Mix / Windy</i>

# **APÊNDICE B – Índices das 63 e 86 Zonas em Nova Iorque**

Índices das 63 zonas utilizadas nos experimentos da Seção 5.3:

4, 12, 13, 24, 41, 42, 43, 45, 48, 50, 68, 74, 75, 79, 87, 88, 90, 100, 107, 113, 114, 116, 120, 125, 127, 128, 137, 140, 141, 142, 143, 144, 148, 151, 152, 158, 161, 162, 163, 164, 166, 170, 186, 209, 211, 224, 229, 230, 231, 232, 233, 234, 236, 237, 238, 239, 243, 244, 246, 249, 261, 262, 263.

Índices das 86 zonas utilizadas nos experimentos da Seção 5.2:

7, 13, 17, 25, 35, 36, 37, 39, 41, 42, 43, 48, 49, 50, 61, 62, 68, 72, 74, 75, 76, 77, 79, 80, 82, 87, 89, 90, 91, 97, 100, 107, 112, 113, 114, 116, 129, 130, 132, 137, 138, 140, 141, 142, 143, 144, 145, 146, 148, 151, 158, 161, 162, 163, 164, 166, 170, 177, 179, 181, 186, 188, 198, 211, 215, 223, 225, 226, 229, 230, 231, 233, 234, 236, 237, 238, 239, 243, 244, 246, 249, 255, 256, 260, 262, 263.

# APÊNDICE C – Listagem Completa dos Resultados e as Respectivas Variáveis de Entrada

Tabela 18 – Desempenho na previsão de demanda em intervalos de 15 minutos para Chengdu.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 5}	<b>8,720</b>	<b>8,669</b>
ALMMo {N = 10}	8,755	8,672
ALMMo {N = 20}	9,051	8,961
ALMMo {N = 26}	9,081	9,029
eMG - {N = 5}	9,673	9,948
eNFN - {N = 10}	10,210	10,001
eNFN - {N = 5}	10,424	10,265
eNFN - {N = 20}	10,581	10,355
eNFN - {N = 26}	10,619	10,458
eMG - {N = 10}	10,745	11,921
eMG - {N = 20}	14,600	14,549
eMG - {N = 26}	18,841	18,295

Tabela 19 – Desempenho na previsão de demanda em intervalos de 30 minutos para Chengdu.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 5}	<b>15,172</b>	<b>15,640</b>
ALMMo {N = 10}	15,223	15,602
ALMMo {N = 20}	16,202	16,580
ALMMo {N = 26}	16,438	16,833
eMG {N = 5}	17,339	17,508
eNFN {N = 10}	17,490	17,514
eNFN {N = 5}	17,759	17,668
eNFN {N = 20}	18,614	18,635
eNFN {N = 26}	18,750	18,947
eMG {N = 10}	19,441	20,365
eMG {N = 20}	41,057	46,872
eMG {N = 26}	50,234	54,522

Tabela 20 – Acurácia em intervalos de 15 minutos para Chengdu.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 5}	<b>83,845</b>	<b>4,779</b>	<b>82,738</b>	<b>5,605</b>
ALMMo {N = 10}	83,778	4,435	82,613	5,794
ALMMo {N = 20}	83,744	4,546	81,905	5,788
ALMMo {N = 26}	83,661	4,608	81,988	5,653
eMG {N = 5}	83,065	5,558	81,435	7,121
eNFN {N = 10}	81,571	5,206	80,268	6,790
eMG {N = 10}	81,327	5,665	79,101	6,970
eNFN {N = 5}	81,030	5,334	79,554	6,632
eNFN {N = 26}	80,625	5,442	79,375	6,395
eNFN {N = 20}	80,565	5,620	79,833	6,490
eMG {N = 20}	75,720	6,528	74,607	7,437
eMG {N = 26}	71,375	7,692	69,423	10,415

Tabela 21 – Acurácia em intervalos de 30 minutos para Chengdu.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 5}	<b>86,952</b>	<b>4,118</b>	<b>85,286</b>	<b>5,670</b>
ALMMo {N = 10}	86,762	4,212	85,000	5,440
eMG {N = 5}	86,274	4,533	84,964	6,117
ALMMo {N = 20}	85,964	4,208	84,476	5,119
ALMMo {N = 26}	85,762	4,475	83,881	5,336
eNFN {N = 10}	84,869	4,803	83,143	6,068
eNFN {N = 5}	84,536	5,042	83,024	6,068
eMG {N = 10}	83,917	5,272	81,810	7,482
eNFN {N = 26}	82,512	5,357	80,405	5,693
eNFN {N = 20}	82,357	5,427	81,048	5,696
eMG {N = 20}	71,167	7,060	67,762	8,215
eMG {N = 26}	65,357	10,147	60,131	10,183

Tabela 22 – Desempenho na previsão da demanda em intervalos de 15 minutos para Nova Iorque com 86 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	<b>13,329</b>	<b>12,460</b>
ALMMo {N = 26}	13,330	12,432
ALMMo {N = 10}	13,546	12,781
eMG {N = 10}	14,793	15,466
eNFN {N = 5}	15,464	14,580
eNFN {N = 26}	16,475	15,147
eNFN {N = 20}	16,775	15,375
eNFN {N = 10}	16,909	15,237
eMG {N = 20}	20,068	17,273
ALMMo {N = 5}	20,966	21,998
eMG {N = 26}	21,019	19,242
eMG {N = 5}	23,436	25,350

Tabela 23 – Desempenho na previsão da demanda em intervalos de 30 minutos para Nova lorque com 86 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	<b>25,072</b>	<b>22,957</b>
ALMMo {N = 26}	25,116	23,028
ALMMo {N = 10}	25,281	23,308
eMG {N = 10}	28,512	26,473
eNFN {N = 5}	29,017	25,358
eNFN {N = 10}	29,386	26,190
eNFN {N = 26}	29,396	26,634
eNFN {N = 20}	29,449	26,496
ALMMo {N = 5}	30,464	31,440
eMG {N = 5}	33,333	34,814
eMG {N = 20}	37,259	37,073
eMG {N = 26}	39,551	38,263

Tabela 24 – Acurácia em intervalos de 15 minutos para Nova lorque com 86 zonas

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 20}	<b>84,796</b>	<b>4,334</b>	<b>84,865</b>	<b>4,593</b>
ALMMo {N = 26}	84,788	4,351	84,808	4,568
ALMMo {N = 10}	84,477	4,400	84,314	4,706
eMG {N = 10}	83,368	4,426	82,790	5,137
eNFN {N = 5}	81,950	5,101	81,824	5,373
eMG {N = 20}	81,002	5,638	81,504	5,537
eNFN {N = 26}	80,632	5,610	80,913	5,839
eNFN {N = 20}	80,312	5,703	80,617	6,019
eNFN {N = 10}	80,309	5,785	80,869	5,872
eMG {N = 26}	79,929	5,589	79,784	5,257
ALMMo {N = 5}	77,572	5,654	75,399	6,373
eMG {N = 5}	76,006	6,287	73,330	7,348

Tabela 25 – Acurácia em intervalos de 30 minutos para Nova Iorque com 86 zonas

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 20}	<b>86,529</b>	<b>3,538</b>	<b>87,256</b>	<b>3,329</b>
ALMMo {N = 26}	86,438	3,533	87,210	3,229
ALMMo {N = 10}	86,299	3,525	86,996	3,245
eMG {N = 10}	85,203	3,732	85,843	3,530
eNFN {N = 5}	83,579	4,666	85,084	3,865
eNFN {N = 26}	83,442	3,944	84,509	3,725
eNFN {N = 20}	83,348	3,926	84,541	3,756
eNFN {N = 10}	83,254	4,169	84,478	4,034
ALMMo {N = 5}	84,322	4,074	83,741	4,477
eMG {N = 5}	83,618	4,274	82,544	4,721
eMG {N = 20}	82,397	4,801	82,491	6,146
eMG {N = 26}	80,628	4,937	81,290	5,241

Tabela 26 – Desempenho em intervalos 15 minutos para Nova Iorque com 63 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	<b>7,705</b>	<b>7,143</b>
ALMMo {N = 26}	7,705	7,144
ALMMo {N = 10}	7,893	7,364
eMG {N = 10}	8,041	7,518
eMG {N = 20}	8,163	7,620
eMG {N = 26}	8,268	7,667
eNFN {N = 5}	8,897	8,334
eNFN {N = 10}	9,296	8,501
eNFN {N = 26}	9,419	8,570
eNFN {N = 20}	9,451	8,604
ALMMo {N = 5}	11,752	11,231
eMG {N = 5}	12,474	12,025

Tabela 27 – Desempenho na previsão da demanda em intervalos de 30 minutos para Nova lorque com 63 zonas.

Modelo	Pick-Up	Drop-Off
	RMSE	RMSE
ALMMo {N = 20}	<b>14,087</b>	<b>12,709</b>
ALMMo {N = 26}	14,090	12,709
ALMMo {N = 10}	14,212	12,834
eMG {N = 10}	14,536	13,177
eMG {N = 20}	14,982	13,408
eMG {N = 26}	15,393	13,922
eNFN {N = 5}	16,105	14,415
eNFN {N = 10}	16,373	14,705
eNFN {N = 20}	16,414	14,712
eNFN {N = 26}	16,494	14,732
ALMMo {N = 5}	18,012	16,732
eMG {N = 5}	18,798	17,408

Tabela 28 – Acurácia em intervalos de 15 minutos para Nova lorque com 63 zonas.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 20}	<b>87,154</b>	<b>5,318</b>	<b>86,524</b>	<b>4,696</b>
ALMMo {N = 26}	87,146	5,325	86,523	4,696
ALMMo {N = 10}	86,792	5,453	86,074	4,837
eMG {N = 20}	86,537	5,502	85,899	4,926
eMG {N = 10}	86,542	5,502	85,774	4,894
eMG {N = 26}	86,472	5,606	85,853	4,945
eNFN {N = 5}	85,171	5,870	84,430	5,236
eNFN {N = 10}	84,972	6,026	84,636	5,220
eNFN {N = 20}	84,715	6,167	84,515	5,291
eNFN {N = 26}	84,562	6,195	84,304	5,308
ALMMo {N = 5}	81,038	8,541	79,457	8,031
eMG {N = 5}	80,568	8,689	78,824	8,189

Tabela 29 – Acurácia em intervalos de 30 minutos para Nova Iorque com 63 zonas.

Modelo	Pick-Up		Drop-Off	
	Acurácia (%)	± Desv. Pad.	Acurácia (%)	± Desv. Pad.
ALMMo {N = 20}	<b>88,060</b>	<b>5,123</b>	<b>87,449</b>	<b>4,467</b>
ALMMo {N = 26}	88,054	5,120	87,421	4,478
ALMMo {N = 10}	87,949	5,110	87,253	4,483
eMG {N = 10}	87,806	5,111	87,065	4,515
eMG {N = 20}	87,476	5,281	86,960	4,603
eMG {N = 26}	87,276	5,403	86,650	4,868
eNFN {N = 5}	86,252	5,723	85,520	5,164
eNFN {N = 20}	85,957	5,805	85,380	5,083
eNFN {N = 26}	85,861	5,771	85,369	5,014
eNFN {N = 10}	85,820	5,907	85,177	5,195
ALMMo {N = 5}	85,549	6,244	84,334	5,951
eMG {N = 5}	85,315	6,402	84,029	6,094