CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA
DE MINAS GERAIS - Campus Leopoldina
Programa de Pós-Graduação em Automação e Sistemas

Natália da Mata Campos

Técnicas Inteligentes na Previsão de Demanda em Sistemas de Compartilhamento de Bicicletas

Leopoldina

Janeiro, 2025

Natália da Mata Campos

Técnicas Inteligentes na Previsão de Demanda em Sistemas de Compartilhamento de Bicicletas

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Automação e Sistemas do Centro Federal de Educação Tecnológica de Minas Gerais, Campus Leopoldina, como parte dos requisitos para obtenção do título de Mestre em Automação e Sistemas.

Orientador: Prof. Dr. Alisson Marques da Silva

Leopoldina

Janeiro, 2025

Campos, Natália da Mata.
C198t Técnicas inteligentes na previsão de demanda em sistemas de compartilhamento de bicicletas. / Natália da Mata Campos. – 2025.
62 f. : il.

Dissertação apresentada ao Programa de Pós-Graduação em Automação e Sistemas (PPGAS), 2025.
Orientador: Alisson Marques da Silva.
Bibliografia.
Dissertação (mestrado) – Centro Federal de Educação Tecnológica de Minas Gerais. *Campus* Leopoldina.

1. Sistemas inteligentes de controle. 2. Previsão. 3. Bicicleta. 4. Regressão. 5. Lógica fuzzy. I. Silva, Alisson Marques da. II. Título.

CDU: 004.89:681.51

Elaboração da ficha catalográfica pela bibliotecária Luzia Adriana Damasceno, CRB 6º 2305 / Cefet/MG c*ampus* Leopoldina.

MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM AUTOMAÇÃO E SISTEMAS - LP

ATA DE DEFESA DE DISSERTAÇÃO Nº 6 / 2025 - PPGAS (11.52.20)

Nº do Protocolo: 23062.007735/2025-00

Leopoldina-MG, 11 de fevereiro de 2025.

DISSERTAÇÃO DE MESTRADO

TÉCNICAS INTELIGENTES NA PREVISÃO DE DEMANDA EM SISTEMAS DE COMPARTILHAMENTO DE BICICLETAS

AUTORA: NATÁLIA DA MATA CAMPOS

ORIENTADOR: Professor Doutor Alisson Marques da Silva

A Banca Examinadora composta pelos membros abaixo aprovou em 30 de janeiro de 2025 esta Dissertação:

Professor Doutor Alisson Marques da Silva (ORIENTADOR)

Centro Federal de Educação Tecnológica de Minas Gerais – CEFET/MG

Professor Doutor Gustavo Campos Menezes (MEMBRO INTERNO)

Centro Federal de Educação Tecnológica de Minas Gerais – CEFET/MG

Professor Doutor José Geraldo Ribeiro Júnior (MEMBRO INTERNO)

Centro Federal de Educação Tecnológica de Minas Gerais – CEFET/MG

Professor Doutor Thiago Magela Rodrigues Dias (MEMBRO EXTERNO)

Centro Federal de Educação Tecnológica de Minas Gerais – CEFET/MG

(Assinado digitalmente em 12/02/2025 09:26)
ALISSON MARQUES DA SILVA
PROFESSOR ENS BASICO TECN TECNOLOGICO
DECOMDV (11.60.11)
Matrícula: 3552958

(Assinado digitalmente em 11/02/2025 15:27)
GUSTAVO CAMPOS MENEZES
PROFESSOR ENS BASICO TECN TECNOLOGICO
DCCN (11.58)
Matrícula: 1550421

(Assinado digitalmente em 11/02/2025 15:20)
JOSE GERALDO RIBEIRO JUNIOR
PROFESSOR ENS BASICO TECN TECNOLOGICO
DCLP (11.61)
Matrícula: 1322715

(Assinado digitalmente em 11/02/2025 15:57)
THIAGO MAGELA RODRIGUES DIAS
PROFESSOR ENS BASICO TECN TECNOLOGICO
DECOMDV (11.60.11)
Matrícula: 2646669

Visualize o documento original em **https://sig.cefetmg.br/public/documentos/index.jsp** informando seu número: **6**, ano: **2025**, tipo: **ATA DE DEFESA DE DISSERTAÇÃO**, data de emissão: **11/02/2025** e o código de verificação: **bb3bb74079**

Agradecimentos

Agradeço primeiramente a Deus por me conceder forças, saúde e determinação para concluir essa etapa tão importante em minha vida.

Agradeço também a minha família, que sempre acreditou e me apoiou em meus projetos de vida e esteve ao meu lado me oferecendo amor e suporte durante toda esta jornada.

Ao meu orientador, Alisson, agradeço imensamente por ter aceitado a minha orientação, por ter me acompanhando atentamente, pela paciência e orientação em cada etapa deste trabalho. E por me guiar de forma dedicada em cada dificuldade encontrada na pesquisa. Suas contribuições foram fundamentais para a conclusão dessa etapa.

Ao Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), agradeço por sempre fazer parte da minha história profissional. À Coordenação do Programa de Pós-Graduação em Automação e Sistemas, em especial à Gláucia pelo apoio e o carinho por todos os alunos.

Agradeço aos meus colegas do Programa de Mestrado que tornaram essa jornada mais leve, em especial, ao Reinaldo, que foi meu parceiro e amigo em todas as aulas, projetos e atividades durante estes anos.

Resumo

Nos últimos anos, os sistemas de bicicletas compartilhadas evoluíram significativamente, desde iniciativas com bicicletas gratuitas para uso comum, passando por estações onde o aluguel era gerenciado presencialmente por atendentes, até os sistemas tecnologicamente avançados que predominam nas cidades atualmente. Apesar dessas inovações, o objetivo principal dos sistemas de compartilhamento de bicicletas permanece inalterado: facilitar que os usuários retirem uma bicicleta em um local desejado e a devolvam em outro, tornando esse meio de transporte mais prático e acessível. Com o crescimento da demanda, um problema recorrente se torna evidente: o desequilíbrio entre a procura dos usuários e a oferta de bicicletas. A indisponibilidade do serviço em determinadas estações leva os usuários a buscarem alternativas de transporte, o que pode comprometer o sucesso e a expansão desses sistemas. Por outro lado, investimentos em infraestrutura e novas bicicletas podem não ser financeiramente vantajosos se a operação não atender à demanda de forma eficiente. Além disso, a má redistribuição de bicicletas e a ineficiência operacional podem aumentar as emissões de carbono, já que os usuários recorrem a meios de transporte motorizados. Esse cenário destaca a importância de desenvolver técnicas que minimizem a discrepância entre a procura e a disponibilidade de bicicletas. Nesse contexto, o presente trabalho propõe uma nova abordagem para prever a demanda de usuários por bicicletas utilizando dados históricos de deslocamentos e informações meteorológicas. Para selecionar os atributos mais relevantes, é empregada a técnica de Eliminação Recursiva de Atributos (RFE). A previsão da demanda é realizada por modelos tradicionais, como Rede Neural Artificial, *Random Forest*, *Support Vector Machine*, Árvore de Decisão e Regressão Linear, e também por modelos evolutivos, como *ALMMo*, *eNFN*, *eFCE* e *eFLS*. Os experimentos computacionais realizados permitiram comparar o desempenho das abordagens propostas. Os resultados obtidos indicam que a abordagem apresentada é uma alternativa viável e eficaz para prever a demanda de bicicletas em sistemas de compartilhamento.

Palavras-chaves: Previsão; Demanda; Bicicletas; Métodos de Regressão; Sistemas Fuzzy Evolutivos.

Abstract

In recent years, bike-sharing systems have evolved significantly from distributing free bikes for public use to stations where rentals were managed by an attendant to the advanced technology systems that are popular in today's cities. Despite these changes, the main objective of bike-sharing systems has remained: to make it easier for users to pick up a bike at a desired location and return it to another, making this means of transportation more practical and accessible. As this demand increases, a recurring problem becomes clearer: the relationship between user demand and the supply of bikes. The lack of service availability at stations causes users to opt for alternative transportation options, thus hindering the growth of the bike-sharing system. However, investments in infrastructure and bicycles may only generate financial returns if demand is adequately met. Otherwise, inefficiencies in system operation and poor redistribution of bicycles may result in higher carbon emissions, as there will be an increase in the use of alternative motorized transportation. This scenario suggests the need to develop techniques that reduce the gap between user demand and the number of bicycles available to meet this demand. In this context, this paper presents a new approach to forecasting user demand for bicycles. The proposed approach uses historical data on bicycle trips and weather information. The RFE method is used to select the most relevant attributes. Five intelligent regression techniques are evaluated to make the prediction: Artificial Neural Network, Random Forest, Support Vector Machine, Decision Tree, and Linear Regression. Four evolving fuzzy systems were also used to predict ALMMo, eNFN, eFCE, and eFLS. Computational experiments were conducted to compare the performance of the proposed approach. The results suggest that the proposed approach is a viable alternative to forecast bicycle demand.

Keywords: Forecasting; Demand; Bicycles; Regression Methods; Evolving Fuzzy Systems.

Lista de ilustrações

Figura 3.1 – Etapas da abordagem proposta. 20
Figura 3.2 – Etapas da seleção de variáveis. 26
Figura 3.3 – Processo de previsão de demanda. 27
Figura 4.1 – Distribuição das estações de compartilhamento de bicicletas na cidade de Washington D.C., EUA. 29
Figura 4.2 – Gráficos de resultados dos métodos *offline*. 33
Figura 4.3 – Gráficos de resultados dos métodos *online*. 35

Lista de tabelas

Tabela 3.1 – Registro dos deslocamentos de bicicletas. 21
Tabela 3.2 – Registro de início do deslocamento e nome da estação. 21
Tabela 3.3 – Intervalos de tempo. 23
Tabela 3.4 – Identificação das variáveis. 25
Tabela 3.5 – Sugestão de sistemas inteligentes para previsão de demanda. 27
Tabela 4.1 – Comparação de diferentes métodos e métricas. 32
Tabela 4.2 – Comparação de diferentes métodos e métricas. 34
Tabela A.1 – Resumos dos trabalhos da literatura. 44
Tabela B.1 – Resumos dos trabalhos da literatura. 45
Tabela C.1 – Resumos dos trabalhos da literatura. 46
Tabela D.1 – Resumos dos trabalhos da literatura. 47

Lista de abreviaturas e siglas

ACO	*Ant Colony Optimization Algorithm*
ALMMo	*Autonomous Learning Multi-Model System*
ANN	*Artificial Neural Network*
ARIMA	*Autoregressive Integrated Moving Average*
BMA	*Bayesian Model Averaging*
CART	*Classification And Regression Tree*
CNN	*Convolutional Neural Networks*
CNN-LSTM	*Convolutional Neural Network - Long Short-Term Memory*
conv-LSTM	*Convolutional Long Short-Term Memory*
CP	*Constraint Programming*
DBSCAN	*Density-Based Spatial Clustering of Applications with Noise*
DC	*District of Columbia*
DCNN	*Dual-Channel Convolutional Neural Network*
DHI	Demanda Histórica
DHD	Demanda Histórica dos Dias Anteriores
MDHI	Média da Demanda Histórica
MDHD	Média da Demanda Histórica dos Diárias Anteriores
DT	*Decision Tree*
EDA	*Empirical Data Analytics*
eFCE	*evolving Fuzzy with Multivariable Gaussian Participatory Learning and Recursive Maximum Correntropy*
eFLS	*evolving Fuzzy with Multivariable Gaussian Participatory Learning and Weighted Recursive Least Squares*
eMG	*evolving Multivariable Gaussian Fuzzy System*

eNFN	*evolving Neo-Fuzzy Neuron*
EUA	*United States of America*
FDS	*Fully Dependent Stations*
GACNN	*Graph Attention Convolutional Neural Network*
GBM	*Gradient Booster Machine*
GPS	*Global Positioning System*
GPR	*Gaussian Process Regression*
GRU	*Gated Recurrent Unit*
Hg	Mercúrio
IIS	*Individual and Independent Stations*
IPDS	*Individual and Partially Dependent Stations*
LSTM	*Long Short-Term Memory*
LR	*Linear Regression*
LSR	*Least Squares Regression*
LSM	*Least Square Method*
MAE	*Mean Absolut Error*
MAPE	*Mean Absolute Percentage Error*
MIP	*Mixed Integer Programming*
MLP	*Multi-Layer Perceptron*
MLR	*Multivariate Linear Regression*
MSE	*Mean Square Error*
mph	Milhas por hora
NCEI	*National Centers for Environmental Information*
NFN	*Neo-Fuzzy-Neuron*
NRMSE	*Normalized Root Mean Square Error*
OOB	*Out-Of-Bag*

PLSR	*Partial-Least-Squares Regression*
POI	Pontos de Interesse
PU	*Pick-up*
RBF	*Radial Basis Function*
RF	*Random Forest*
RFE	*Recursive Feature Elimination*
RFWLS	*Recursive Fuzzily Weighted Least Squares*
RMSE	*Root Mean Square Error*
RMSLE	*Root Mean Squared Logarithmic Error*
RNA	*Rede Neural Artificial*
RNN	*Recurrent neural networks*
R^2	*Coefficient of Determination*
SARIMA	*Seasonal AutoRegressive Integrated Moving Average*
SMVP	*Stacking Model For Variation Prediction*
SVM	*Support Vector Machine*
SVR	*Support Vector Regression*
UDF	*User Dissatisfaction Function*
VP-RNN	*Variational Predictive Recurrent Neural Network*
WRLS	*Weighted Recursive Least Squares*

Sumário

1 INTRODUÇÃO 1
1.1 Motivação e Relevância 1
1.2 Caracterização do Problema 5
1.3 Objetivos 6
1.3.1 Objetivo Geral 6
1.3.2 Objetivos Específicos 6
1.4 Organização do Trabalho 6

2 FUNDAMENTOS E REVISÃO 8
2.1 Fundamentação Teórica 8
2.1.1 Sistemas Tradicionais (*offline)* 8
2.1.2 Sistemas Evolutivos *(online)* 11
2.2 Revisão da Literatura 13

3 ABORDAGEM PROPOSTA 20
3.1 Construção da Base de Dados 20
3.2 Criação de Variáveis de Entrada 24
3.3 Seleção de Variáveis de Entrada 26
3.4 Previsão de Demanda 26

4 EXPERIMENTOS E RESULTADOS 28
4.1 Metodologia 28
4.2 Resultados dos Modelos Offline 30
4.3 Resultados dos Modelos Online 32
4.4 Análise Comparativa entre os Modelos Offline e Online 35

5 CONSIDERAÇÕES FINAIS 37

REFERÊNCIAS 38

APÊNDICES 43

APÊNDICE A – TABELA COM OS RESUMOS DOS TRABALHOS DA LITERATURA 44

APÊNDICE B – TABELA COM OS RESUMOS DOS TRABALHOS DA LITERATURA 45

APÊNDICE C – TABELA COM OS RESUMOS DOS TRABALHOS DA LITERATURA 46

APÊNDICE D – TABELA COM OS RESUMOS DOS TRABALHOS DA LITERATURA 47

1 Introdução

Este capítulo introduz o problema em questão, destacando os desafios e descobertas encontrados ao longo da pesquisa. A Seção 1.1 contextualiza o problema dentro de um panorama mais amplo e explica as razões que motivaram este estudo. A Seção 1.2 apresenta a caracterização do problema. Em seguida, a Seção 1.3 especifica os objetivos gerais e específicos do trabalho. Por fim, a Seção 1.4 descreve a organização do trabalho.

1.1 Motivação e Relevância

Os sistemas de transporte urbano enfrentam grandes desafios devido ao aumento do número de veículos particulares e das taxas de motorização (relação entre o número de veículos particulares e o número de habitantes) (OSKARBSKI; BIRR; ŻARSKI, 2021). Segundo dados da IOMVH (2024), o total de veículos em uso globalmente cresceu de 1,28 bilhão de unidades em 2015 para 1,59 bilhão em 2020, representando um aumento significativo de aproximadamente 4% ao ano.

Muitas cidades ao redor do mundo enfrentam a expansão urbana, o crescimento acelerado da motorização, sistemas de transporte público inadequados, uma alta proporção de viagens realizadas por automóveis, níveis elevados de poluição e infraestrutura insuficiente para pedestres e ciclistas. Esse crescimento insustentável das atividades de transporte exerce pressão sobre os ecossistemas e os recursos do planeta. As emissões de gases de efeito estufa associadas à produção de energia são uma das principais causas das mudanças climáticas (OSKARBSKI; BIRR; ŻARSKI, 2021).

O aumento do número de veículos motorizados e dos sistemas de transporte baseados em veículos particulares levanta questões críticas relacionadas à sustentabilidade. À medida que a população global, cada vez mais urbanizada, continua a crescer, a demanda por mobilidade ultrapassará a capacidade dos sistemas fortemente dependentes de veículos particulares. Esse cenário, aliado ao aumento das emissões de gases de efeito estufa, também apresenta desafios significativos para o cumprimento das metas climáticas globais (WHO, 2024). Em 2023, as emissões totais de CO2 aumentaram 1,1%. Longe de apresentar a redução acelerada necessária para alcançar as metas climáticas globais estabelecidas no Acordo de Paris, as emissões atingiram um novo recorde de 37,4 Gt nesse mesmo ano (IEA, 2024).

Os governos precisam liderar estratégias baseadas em dados sólidos, apoiadas por legislações e financiamentos robustos, que envolvam todos os setores relevantes. As

empresas devem colocar a segurança e a sustentabilidade no centro de suas cadeias de valor. A academia e a sociedade civil têm o papel de gerar evidências e responsabilizar os tomadores de decisão, enquanto os jovens podem exigir ações e participar ativamente na sua implementação. Essa abordagem holística à mobilidade traz benefícios que vão além das questões de transporte. Por exemplo, ao incentivar a caminhada e o ciclismo, é possível reduzir o sedentarismo, diminuir a poluição e combater as mudanças climáticas (WHO, 2024).

Os benefícios para a saúde associados ao compartilhamento de bicicletas apresentam duas vertentes principais: a melhoria da qualidade do ar, resultante da redução do congestionamento no tráfego devido à substituição modal que dispensa o uso de automóveis particulares, e o aumento da atividade física. A exposição ao material particulado proveniente das emissões de automóveis está associada a sérios problemas de saúde respiratória, sendo que as concentrações desse material são mais elevadas nas proximidades de vias de trânsito intenso (ITDP, 2018).

Muitas cidades incluem metas de redução de quilômetros percorridos por automóveis (*Vehicle Kilometers Traveled - VKT*) em suas políticas de sustentabilidade municipal e o compartilhamento de bicicletas pode ser uma intervenção significativa para alcançar tais metas. Um estudo da *Bikeplus*, realizado com usuários no Reino Unido em 2017, revelou que 23% dos usuários optaram por bicicletas compartilhadas em vez de automóveis para completar seus trajetos (ITDP, 2018).

Segundo o relatório da WHO (2024), mortes e ferimentos no trânsito continuam sendo um dos maiores desafios globais em saúde e desenvolvimento. Em 2019, os acidentes de trânsito foram a principal causa de morte de crianças e jovens entre 5 e 29 anos, e a 12^a principal causa de óbitos considerando todas as idades. Dois terços dessas mortes ocorreram entre pessoas em idade produtiva (18 a 59 anos), gerando impactos profundos na saúde, além de prejuízos sociais e econômicos para a sociedade como um todo.

Entre os tipos de usuários afetados, motociclistas e condutores de veículos motorizados de duas ou três rodas representam 30% das mortes, enquanto ocupantes de veículos de quatro rodas respondem por 25%. Pedestres são responsáveis por 21% das mortes, e ciclistas por 5%. Os ocupantes de veículos que transportam mais de 10 pessoas, veículos pesados de carga e outros usuários compõem os 19% restantes das fatalidades, conforme relatado em WHO (2024).

A segurança das bicicletas compartilhadas pode ser amplamente avaliada por meio da proporção de fatalidades em relação ao número total de viagens realizadas, comparada à mesma relação para viagens em automóveis. Dados mais detalhados devem incluir o número total de ciclistas mortos ou gravemente feridos *(Killed or Seriously Injured - KSI)* nos anos anteriores ao lançamento do sistema de bicicletas, comparando esses números

após a sua implantação. Também é relevante analisar esses indicadores dentro e fora da área de serviço das bicicletas compartilhadas, particularmente em sistemas baseados em estações (ITDP, 2018).

A transição para o ciclismo contribui para mitigar os problemas mencionados anteriormente. O ciclismo influencia a demanda por veículos motorizados e pelo transporte público, e deve ser considerado em estudos de viabilidade para investimentos em infraestrutura e em medidas organizacionais voltadas a todos os modos de transporte, bem como no processo de tomada de decisão (OSKARBSKI; BIRR; ŻARSKI, 2021).

Os sistemas de transporte urbano desempenham um papel crucial na promoção da habitabilidade e sustentabilidade das cidades. Esse papel se torna ainda mais relevante em grandes cidades com altos níveis de desenvolvimento econômico. Frente aos crescentes desafios impostos pela rápida urbanização e pelo aumento populacional, cidades em todo o mundo estão buscando alternativas de transporte mais eficientes e sustentáveis, especialmente devido às preocupações cada vez maiores com a poluição no ambiente urbano (TIAN; WEN; SUN, 2023).

Um estudo realizado em 2017 demonstrou que, em viagens curtas de até 3 km durante os horários de pico de congestionamento, como nas horas de rush nos dias úteis, o uso de bicicletas compartilhadas em Nova York é uma alternativa mais rápida do que pegar um táxi. De maneira similar, a análise dos trajetos de bicicleta no bairro de *Bayview*, em São Francisco, mostrou que viagens curtas feitas em bicicletas elétricas são mais rápidas e menos onerosas do que aquelas realizadas com outros meios de transporte disponíveis (ITDP, 2018).

Em comparação com outros meios de transporte, as viagens de bicicleta tendem a ser mais curtas em termos de tempo de deslocamento e mais baratas. Dessa forma, o levantamento feito entre usuários de bicicletas, para estimar a economia de tempo e dinheiro – apresentada como um percentual da renda individual – em comparação com outros modos de transporte, pode ser uma maneira eficaz de avaliar o benefício econômico de um sistema de bicicletas compartilhadas. Um estudo realizado em 2017 demonstrou que, em viagens curtas de até 3 km durante os horários de pico de congestionamento, como nas horas de rush nos dias úteis, o uso de bicicletas compartilhadas em Nova York é uma alternativa mais rápida do que pegar um táxi. De maneira similar, a análise dos trajetos de bicicleta no bairro de *Bayview*, em São Francisco, mostrou que viagens curtas feitas em bicicletas elétricas são mais rápidas e menos onerosas do que aquelas realizadas com outros meios de transporte disponíveis (ITDP, 2018).

Como parte da mobilidade urbana sustentável, o compartilhamento de bicicletas está crescendo em cidades ao redor do mundo. Pesquisas empíricas têm se concentrado em destacar os requisitos e os impactos benéficos das mudanças relacionadas no transporte

urbano. Um planejamento cuidadoso é necessário para tornar os sistemas de micromobilidade compartilhada mais atraentes em comparação com alternativas menos ecológicas, como o uso de carros (RENNIE et al., 2023).

Para sistemas operados por empresas privadas, seja sob licença ou outra estrutura regulatória, a implementação pode começar assim que os requisitos definidos forem aprovados pelos órgãos governamentais responsáveis e a equipe estiver pronta para iniciar o processamento dos pedidos de licença. Por exemplo, os requisitos de licença em *Seattle* entraram em vigor em 30 de junho de 2017, e as duas primeiras operadoras aprovadas iniciaram suas operações duas semanas depois – menos de três meses após o término das operações de um sistema baseado em estações na cidade (ITDP, 2018).

Em sistemas operados por uma agência pública ou por meio de uma parceria público-privada, o cronograma de implementação depende da assinatura de contratos, da aquisição e instalação de equipamentos e do desenvolvimento ou aquisição de software. O sistema *Vélib'* e o *Ecobici* levaram seis meses para sua completa implementação. Já o sistema de bicicletas compartilhadas de *Nova York* demorou dois anos para ser totalmente implementado, em parte devido a problemas contratuais entre a operadora do sistema e a subcontratada responsável pelo desenvolvimento do software (ITDP, 2018).

O sistema deve oferecer formas para que os usuários se registrem, efetuem pagamentos, enviem reclamações e alertas sobre equipamentos defeituosos, além de disponibilizar pontos de venda para a compra de assinaturas e uma linha direta para esclarecer dúvidas. Após o lançamento do sistema *Relay* em Atlanta, embaixadores voluntários promoveram mesas-redondas em eventos comunitários para incentivar a adesão de novos usuários e ajudá-los a utilizar o aplicativo do sistema (ITDP, 2018).

Medir o acesso às bicicletas compartilhadas é fundamental para a construção de um sistema equitativo, funcionando como um meio de transporte confiável e acessível para todos os moradores e visitantes da cidade. Uma medida abrangente de acessibilidade é a porcentagem da população da cidade que vive a 500 metros de uma estação de bicicletas, em comparação com a área total da cidade. Caso camadas adequadas do sistema de informação geográfica (GIS) – como dados sobre a população por setor censitário ou áreas geográficas semelhantes – estejam disponíveis para uma cidade e as estações sejam georreferenciadas, essa análise pode ser realizada com facilidade (ITDP, 2018).

O planejamento de um futuro incerto exige flexibilidade e sensibilidade aos eventos imprevistos, além da disponibilização de mecanismos que permitam medir o impacto de mudanças nas políticas e procedimentos existentes. O cenário das bicicletas compartilhadas está em constante evolução. Atualmente, as cidades têm a oportunidade de explorar essas novas tecnologias para viabilizar redes de transporte mais sustentáveis (ITDP, 2018).

Planejar operações de compartilhamento de bicicletas envolve, por exemplo, de-

terminar a melhor distribuição de estações na área de serviço e como distribuir as bicicletas entre as estações. Ao adotar uma boa estratégia no planejamento, os procedimentos de otimização que definem os níveis de estoque por estação dependem da demanda prevista (RENNIE et al., 2023).

1.2 Caracterização do Problema

Uma das maneiras mais adequadas de realizar o planejamento de operações de compartilhamento de bicicletas é por meio da previsão de demanda. Ao antecipar os padrões de uso, é possível melhorar a alocação de recursos, reduzir custos operacionais e aumentar a satisfação dos usuários. A previsão de demanda abrange dados que variam ao longo do tempo, apresentando sazonalidade, tendência ou variações de curto prazo. Esse processo pode ser tratado como um problema de previsão de séries temporais, em que os dados históricos de uso do sistema de bicicletas são analisados para identificar padrões que podem ser usados para prever a demanda futura.

A modelagem de séries temporais e a análise de regressão são duas abordagens comuns para a previsão da demanda por bicicletas. Modelos de séries temporais são usados para modelar e prever dados que mudam ao longo do tempo, como a demanda por bicicletas em uma base horária ou diária. Esses modelos conseguem capturar padrões como tendências, sazonalidade e ciclicidade nos dados, usando essas informações para fazer previsões (SUBRAMANIAN et al., 2023). Por outro lado, a análise de regressão é utilizada para modelar a relação entre uma ou mais variáveis preditoras e uma variável de resposta. No contexto da previsão de demanda por bicicletas, a análise de regressão pode modelar a relação entre a demanda e fatores como clima, hora do dia, dia da semana e feriados (SUBRAMANIAN et al., 2023).

Tanto a modelagem de séries temporais quanto a análise de regressão têm pontos fortes e limitações, e a escolha da abordagem depende das especificidades do problema em questão. Os modelos de séries temporais são especialmente úteis quando o foco está na previsão da demanda futura com base em padrões históricos, enquanto a análise de regressão é mais adequada quando o objetivo é entender os fatores que influenciam a demanda por bicicletas e como esses fatores podem ser usados para aprimorar a precisão das previsões (SUBRAMANIAN et al., 2023).

Além das dependências temporais, a seleção cuidadosa dos recursos espaciais também é essencial para prever a demanda por compartilhamento de bicicletas. A escolha do melhor modelo de previsão depende do conjunto de dados específico e do horizonte de previsão. A engenharia de recursos relevantes, considerando tanto aspectos temporais quanto espaciais, pode aumentar a capacidade dos modelos de capturar padrões subjacentes (MA

et al., 2022).

Portanto, a combinação de modelos de séries temporais e análise de regressão, aliados a técnicas inteligentes, é crucial para enfrentar desafios específicos e garantir a disponibilidade adequada de bicicletas em diferentes cenários urbanos.

1.3 Objetivos

1.3.1 Objetivo Geral

Este trabalho tem como objetivo geral o desenvolvimento de uma abordagem baseada em técnicas inteligentes para prever a demanda de bicicletas.

1.3.2 Objetivos Específicos

- Desenvolver e implementar uma abordagem baseada em técnicas inteligentes para previsão da demanda de bicicletas.
- Realizar previsões de demanda em ambientes *offline* e *online*.
- Conduzir experimentos computacionais utilizando bases de dados reais sobre deslocamentos de bicicletas para avaliar o desempenho da metodologia proposta.
- Comparar os resultados obtidos com a metodologia proposta aos resultados reportados na literatura.

1.4 Organização do Trabalho

Para facilitar o entendimento deste trabalho, o texto está organizado em 5 capítulos da seguinte forma:

- Este capítulo introdutório apresentou a motivação, a caracterização do problema e os objetivos desta pesquisa.
- O Capítulo 2 oferece o embasamento teórico necessário para a compreensão do trabalho, revisando a literatura e destacando as principais abordagens relacionadas à demanda de bicicletas em sistemas de compartilhamento.
- O Capítulo 3 descreve a metodologia proposta, detalhando cada uma de suas etapas.

- O Capítulo 4 apresenta os experimentos computacionais realizados para avaliar e comparar o desempenho da metodologia proposta.
- Por fim, o Capítulo 5 discute as considerações finais e propõe direções para futuras pesquisas.

2 Fundamentos e Revisão

Este capítulo relata os fundamentos e a revisão da literatura. A Seção 2.1 apresenta as revisões das pesquisas de outros autores referente ao tema abordado na pesquisa. A Seção 2.2 evidencia o processo de leitura, de análise para a pesquisa científica, além de contextualizar o problema.

2.1 Fundamentação Teórica

Esta seção apresenta uma breve revisão dos sistemas inteligentes utilizados neste trabalho. Inicialmente, na Seção 2.1.1, revisita-se os sistemas inteligentes tracionais que possuem métodos de treinamento *offline*. Em seguida, na Seção 2.1.2, os sistemas evolutivos com aprendizado *online* e incremental são detalhados.

2.1.1 Sistemas Tradicionais (*offline)*

Os sistemas inteligentes tradicionais[1] utilizam um aprendizado em *batch/offline*. Esses sistemas requerem um conjunto de dados completo para realizar o processo de treinamento, que geralmente envolve múltiplas passagens pelos dados para ajustar os parâmetros do modelo, conhecidas como épocas de treinamento. Após o ajuste ideal dos parâmetros, eles são fixados, e o sistema é colocado em produção ou avaliado (HAYKIN, 2017). Diversos sistemas inteligentes adotam essa abordagem. A seguir, apresenta-se uma breve descrição dos modelos com essa abordagem que são utilizados neste trabalho.

SVM - Support Vector Machine: As *SVMs* são classificadores que utilizam separadores lineares com margem máxima. Essa margem máxima é a largura da área delimitada por linhas tracejadas - o dobro da distância do separador até o ponto do exemplo mais próximo. Por meio de métodos de *kernel*, as *SVMs* transformam implicitamente os dados de entrada em um espaço de alta dimensionalidade, permitindo a separação linear mesmo para conjuntos de dados originalmente não separáveis (RUSSEL, 2022). O kernel linear é o mais simples de todos os kernels. Tecnicamente, os dados não são projetados em dimensões superiores quando este kernel é usado, então é apenas o produto interno de x e y com um termo constante opcional c. O benefício do *kernel linear* é que ele é simples e tem apenas o termo constante c como

[1] Nomenclatura utilizada em alguns trabalhos para diferenciar sistemas com aprendizado em *batch/offline* daqueles com aprendizado incremental e contínuo.

parâmetro. A função *kernel* polinomial é direcional, ou seja, a saída depende da direção dos dois vetores no espaço de baixa dimensão. A magnitude da saída também depende da magnitude do vetor. A *RBF*, do inglês *radial basis function* é uma das funções *kernel* mais populares. Ela adiciona uma proeminência ao redor de cada ponto de dados (PRAJAPATI; PATLE, 2010). O hiperplano separador é definido pelos exemplos mais próximos da margem, conhecidos como vetores de suporte, reduzindo a complexidade do modelo ao evitar a dependência de todos os pontos do conjunto de dados. Além disso, as *SVMs* combinam as vantagens de modelos não-paramétricos e paramétricos: têm flexibilidade para representar funções complexas, mas são resistentes ao sobreajuste *(overfitting)* (RUSSEL, 2022).

DT - Decision Tree: O algoritmo *DT* é uma representação de uma função que mapeia um vetor de valores de atributos para um único valor de saída - uma "decisão". A árvore toma sua decisão realizando uma sequência de testes, começando na raiz e seguindo o ramo apropriado até alcançar uma folha. Cada nó interno da árvore corresponde a um teste do valor de um dos atributos de entrada, os ramos que partem do nó são rotulados com os possíveis valores do atributo, e os nós folha especificam o valor a ser retornado pela função. Uma técnica chamada poda de árvore de decisão combate o ajuste excessivo. Portanto, funciona eliminando nós que não são claramente relevantes. Desta forma, começa-se com uma árvore completa. Em seguida, olha-se para um nó de teste que tem apenas nós folha como descendentes. Se o teste parecer irrelevante - detectando apenas ruído nos dados - então elimina-se o teste, substituindo-o por um nó folha. Repete-se esse processo, considerando cada teste apenas com descendentes folha, até que cada um tenha sido podado ou aceito no estado atual (RUSSEL, 2022).

RNA - Rede Neural Artificial: é composta por neurônios organizados em camadas, onde cada neurônio realiza cálculos baseados em suas entradas, aplicando funções de ativação que determinam a resposta do neurônio. Durante o processo de treinamento, uma rede neural capta muitos valores de entrada e muitos valores de saída correspondentes. Ao processar um vetor de entrada, a rede neural realiza vários cálculos intermediários antes de produzir a saída. Pode-se pensar nos valores computados em cada camada da rede como uma representação diferente para a entrada. Cada camada transforma a representação produzida pela camada anterior para produzir uma nova representação. A composição de todas essas transformações é bem-sucedida - se o processo transcorrer sem falhas - em transformar a entrada na saída desejada. De fato, uma hipótese para o porquê de o aprendizado profundo funcionar corretamente é que a complexa transformação de ponta a ponta que mapeia da entrada para saída - diga-se, de uma imagem de entrada para a categoria de saída - é decomposta pelas muitas camadas na composição de muitas transformações relativamente simples,

cada uma das quais é bastante fácil de aprender por um processo de atualização local (RUSSEL, 2022).

LR - Regressão Linear A regressão é uma técnica usada para duas teorias. Primeiro, as análises de regressão são geralmente usadas para previsão e predição, nas quais sua aplicação têm grandes sobreposições com a área de aprendizado de máquina. Segundo, a análise de regressão pode ser empregada para investigar possíveis relações causais entre variáveis independentes e dependentes. É importante ressaltar que as regressões sozinhas mostram apenas relações entre uma variável dependente e uma coleção de conjuntos de dados fixos de diferentes variáveis. A *LR* é um modelo de caso no qual há apenas uma variável independente. A *LR* define a dependência da variável permitindo distinguir a influência de variáveis independentes da interação de variáveis dependentes. A *MLR*, do inglês *multivariate linear regression* é uma técnica estatística para prever o resultado de uma variável de resposta, usando uma série de variáveis explicativas. O objetivo da *MLR* é modelar a relação linear entre as variáveis independentes x e a variável dependente y que serão analisadas. A regressão polinomial é um tipo de análise de regressão na modelagem polinomial de grau n da relação entre variáveis independentes e dependentes. A regressão polinomial é um caso especial de *MLR* em que a equação polinomial de dados se mistura com a interação curvilínea das variáveis dependentes e independentes. O *LSM*, do inglês *least square method* é usado para encontrar a linha ou curva de melhor ajuste para um conjunto de dados, reduzindo a quantidade de quadrados dos deslocamentos (parte residual) dos pontos da curva. A motivação por trás da abordagem dos mínimos quadrados tem sido encontrar as estimativas dos parâmetros usando os mínimos quadrados, que é a linha mais próxima de todos os pontos. A solução do *LSM* na regressão linear é obtida resolvendo um sistema de equações lineares que garante o melhor ajuste aos dados (MAULUD, 2020).

RF - Random Forest: O modelo *RF* é um conjunto de árvores de decisão onde cada árvore é construída a partir de uma versão *bootstrapped* (amostragem aleatória de um conjunto de dados com substituição) do conjunto de dados de treinamento, ou seja, utiliza o método *ensemble*. Cada árvore é cultivada por meio do princípio de partição repetitiva, onde, a partir do nó raiz, o mesmo procedimento de divisão de nó é aplicado repetidamente até que certas regras de parada sejam atendidas. Seu poder na previsão vem da agregação de muitos aprendizes mais fracos (árvores de decisão). O desempenho é bom se as correlações entre árvores na floresta forem baixas. Para uma árvore de decisão binária como *CART*, do inglês *classification and regression tree*, o processo de divisão de nó consiste em selecionar uma variável de divisão e determinar a regra de divisão. O princípio para a divisão de nó é minimizar a impureza dos valores de resposta em cada nó, que geralmente é medido

pelo índice de *Gini*, se a variável de resposta for categórica ou pela variância, se for quantitativa. O crescimento de cada árvore de decisão é interrompido quando os nós a serem divididos pertencem à mesma classe ou possuem o mesmo valor de resposta, ou ainda quando outras regras de parada pré-determinadas são atendidas. Os nós na camada final de uma árvore são chamados de folhas e são usados para previsão de novas observações. Para fazer previsões com *RF*, uma observação passa por cada árvore de decisão na floresta. A previsão final para a observação da *RF* é feita por votação majoritária ou média, com base nos resultados de todas as árvores de decisão na floresta. Como o algoritmo *RF* usa amostras *bootstrap* para aumentar cada árvore de decisão, algumas observações são deixadas de fora na construção de uma determinada árvore. Ao tratar essas amostras *OOB*, do inglês *out-of-bag* como observações que precisam ser previstas, ele pode, portanto, fornecer uma estimativa do erro de previsão da floresta construída. Além disso, a chamada medida de importância variável pode ser obtida para cada preditor, que mede sua relevância para a predição (BAI et al., 2022).

2.1.2 Sistemas Evolutivos *(online)*

Sistemas Evolutivos são caracterizados pelo aprendizado incremental (*online*), realizado em uma única passagem pelos dados. Em outras palavras, o aprendizado ocorre de forma contínua, processando as amostras individualmente, que podem ser descartadas após o processamento. Além disso, esses sistemas possuem uma estrutura flexível que se adapta dinamicamente ao fluxo de dados de entrada, podendo se expandir ou se contrair conforme necessário (RODRIGUES; SILVA; LEMOS, 2022). Um ponto importante é que, diferentemente dos sistemas com aprendizado *offline*, os sistemas evolutivos não requerem retreinamento completo, mesmo em situações de mudanças no contexto dos dados (AFFONSO; MOITA; SILVA, 2024).

ALMMo - Autonomous Learning Multi-Model System: O modelo *ALMMo* é não-paramétrico e orientado a dados para dados em fluxo contínuo dentro do *framework EDA*, do inglês *Empirical Data Analytics*. Ele pode ser interpretado como regras do tipo *IF-THEN* de *AnYa*, ou como uma *RNN* com uma estrutura simples de múltiplas camadas que se assemelha a uma *RBF*. Sua estrutura é construída com base em nuvens de dados não-paramétricas, livres de restrições externas. Todos os meta-parâmetros são extraídos diretamente dos dados observados empiricamente, sem necessidade de conhecimento prévio específico do usuário ou do problema e podem ser atualizados de forma recursiva. Assim, o sistema também é eficiente em termos de memória e computação. Sua estrutura é capaz de evoluir *online* para acompanhar possíveis mudanças ou derivações nos padrões de dados no caso de da-

dos em fluxo contínuo. A qualidade das nuvens de dados existentes é monitorada *online* para manter o sistema otimizado (ANGELOV; GU; PRíNCIPE, 2018).

eNFN - evolving Neo-Fuzzy Neuron: O modelo *eNFN* atualiza seus pesos utilizando a busca por gradiente com taxa de aprendizagem ótima. No entanto, em contraste com o NFN, a abordagem evolutiva atualiza a estrutura do modelo sempre que novos dados são inseridos. Além disso, os valores modais das funções de pertinência, o número de funções de pertinência e o número de neurônios podem ser modificados simultaneamente durante a atualização dos pesos. Para o aprendizado do *eNFN*, deve-se escolher os valores iniciais dos parâmetros das funções de pertinência. Este passo é realizado apenas uma vez, utilizando os limites inferior e superior dos domínios das variáveis de entrada. Calcular os graus de pertinência da entrada *xt*, encontrar as funções de pertinência mais ativas e atualizar seus valores modais. Verificar se a função de pertinência mais ativa representa bem a vizinhança de *xt*. Decidir se uma nova função de pertinência deve ser criada e inserida para refinar a vizinhança de *xt*. Encontrar a função de pertinência inativa mais antiga. Decidir se essa função de pertinência deve ser excluída. E por fim, uma função de pertinência é excluída se seu tempo de inatividade, ou seja, o número de etapas durante as quais ela esteve inativa, for maior que o seu limite (SILVA et al., 2014).

eFCE - evolving Fuzzy with Multivariable Gaussian Participatory Learning and Recursive Maximum Correntropy: Neste método, busca-se reduzir o erro de saída ao maximizar a densidade de probabilidade do erro na origem. As técnicas de ajuste de parâmetros geralmente se fundamentam no critério *MSE*. Por outro lado, a atualização dos parâmetros na máxima correntropia inclui um fator de escala, representado por uma função exponencial do erro. Essa função tem como finalidade atenuar o efeito de possíveis discrepâncias ou níveis anômalos de ruído (RODRIGUES; SILVA; LEMOS, 2021a). A correntropia representa uma medida não linear de similaridade entre as saídas desejadas e reais em um espaço de recursos do kernel (BAO et al., 2018).

eFLS - evolving Fuzzy with Multivariable Gaussian Participatory Learning and Weighted Recursive Least Squares: A adaptação dos parâmetros do consequente do eFLS é realizada por meio do algoritmo *WRLS*, do inglês *weighted recursive least squares*. O *WRLS* visa aprimorar a eficiência na estimativa dos parâmetros, atribuindo a cada amostra o nível apropriado de influência sobre o resultado da estimativa. Vale destacar que, devido à função de otimização do tipo parábola, o *WRLS* é capaz de alcançar a otimalidade global em um único ciclo de iteração devido à parábola da função objetivo (RODRIGUES; SILVA; LEMOS, 2021a).

2.2 Revisão da Literatura

A revisão da literatura tem como foco a aplicação de técnicas computacionais para previsão de demanda em sistemas de compartilhamento de bicicletas. São abordados trabalhos relevantes ao tema, destacando os métodos utilizados, as contribuições e as análises realizadas ao longo do tempo. Para facilitar a compreensão, os artigos são apresentados em ordem cronológica, enfatizando o desenvolvimento das abordagens e os resultados obtidos.

Em Gast et al. (2015), foi utilizado o conjunto de dados do sistema Vélib, de Paris, França, que inclui medições da disponibilidade de bicicletas coletadas a cada minuto, além de variáveis externas, como dados meteorológicos. As variáveis relevantes foram identificadas por meio da análise de processos de *Poisson*. O estudo teve como objetivo principal desenvolver um modelo preditivo para a disponibilidade de bicicletas, validá-lo com dados reais e propor novas métricas para avaliar a precisão de previsões probabilísticas. Utilizando um modelo teórico de filas não homogêneo no tempo, as previsões probabilísticas estimaram a distribuição de probabilidade do número de bicicletas disponíveis em uma estação. Essa abordagem foi útil tanto para usuários planejarem suas viagens quanto para operadores gerenciarem o sistema. A análise foi realizada em intervalos de tempo de 5, 15, 30 e 60 minutos, com o *RMSE*, do inglês *root mean-square error*, sendo utilizado como métrica de avaliação. O trabalho destaca a importância de considerar não apenas a quantidade esperada de bicicletas, mas também a distribuição de probabilidade, oferecendo uma visão mais realista e abrangente das dinâmicas do sistema.

No estudo de Salaken et al. (2015), sistemas *fuzzy* do tipo 1 e tipo 2 foram aplicados para prever a demanda por bicicletas compartilhadas. Dados do sistema (Capital Bikeshare, 2023), em *Washington*, *D.C.*, *EUA*, referentes aos anos de 2011 e 2012, foram utilizados, incluindo variáveis como data, estação, condições climáticas, temperatura, umidade e velocidade do vento. O método *Wang-Mendel* foi empregado para gerar uma base de regras *fuzzy* a partir de dados históricos, utilizando funções de pertinência Gaussianas. O estudo demonstrou que mecanismos de inferência *fuzzy* capturam de forma eficaz a diversidade da demanda, apresentando desempenho superior às redes neurais tradicionais em precisão, mesmo sem otimização dos parâmetros. As métricas *RMSLE*, do inglês *root mean squared logarithmic error*, e *MAPE* do inglês *mean absolute percentage error*, foram utilizadas para avaliar as previsões, destacando a capacidade do modelo em lidar com a natureza complexa da demanda.

Na pesquisa de Zhang et al. (2017), foi proposta uma metodologia dinâmica para reposicionamento de bicicletas que integra, em um único modelo, a previsão de estoque, a chegada de usuários, o reposicionamento de bicicletas e as rotas dos veículos. O objetivo foi otimizar a operação dos sistemas de compartilhamento em *Washington, D.C.*, *EUA*, e

Paris, França. O modelo utiliza *MIP*, do inglês *mixed integer programming* e inclui dados históricos, condições climáticas e intervalos de 3 horas para análise. Processos de *Poisson* e cadeias de *Markov* foram empregados para selecionar as variáveis mais relevantes. As previsões e a qualificação da demanda foram realizadas por meio de um algoritmo heurístico.

Por sua vez, Schuijbroek, Hampshire e van Hoeve (2017) propuseram uma abordagem focada no reequilíbrio de inventário e no roteamento de veículos em sistemas de compartilhamento de bicicletas. O objetivo foi garantir níveis adequados de bicicletas e docas disponíveis para os usuários, minimizando os custos operacionais associados ao reequilíbrio. Modelos baseados em *MIP* e Programação por Restrições foram utilizados, com o desenvolvimento de uma heurística para resolver o problema. Os dados foram obtidos dos sistemas *Hubway* (Boston, MA) e *Capital Bikeshare (Washington, D.C.)*, sendo o reequilíbrio realizado durante a noite, quando a atividade dos usuários é reduzida. Variáveis como níveis de estoque, previsões de chegada de usuários e rotas de reposicionamento foram modeladas utilizando processos de *Poisson* e cadeias de *Markov*. A abordagem assegura que os inventários atendam aos requisitos de serviço, otimizando tanto as retiradas quanto as devoluções de bicicletas.

No estudo de Ruffieux et al. (2017), os dois principais algoritmos comparados foram o RF, do inglês *Random Forest* e a *CNN*, do inglês *convolutional neural networks*, aplicados para prever a disponibilidade de bicicletas e vagas em estações de sistemas de bicicletas compartilhadas. Os dados foram coletados a cada 5 minutos, e as previsões foram realizadas para intervalos de 5, 10 e 15 minutos. As informações incluíam dados da rede de bicicletas *Citybik2*, dados meteorológicos do serviço *OpenWeatherMap* (como temperatura, umidade, vento e condições gerais do tempo), além de informações sobre feriados. O estudo envolveu seis cidades europeias: *Namur* (Bélgica), *Essen* (Alemanha), *Glasgow* (Reino Unido), *Budapeste* (Hungria), *Viena* (Áustria) e *Nice* (França). Três modelos de dependência entre as estações foram testados: *IIS*, do inglês *individual and independent stations*, *IPDS*, do inglês *individual and partially dependent stations*, e *FDS*, do inglês *fully dependent stations*. O objetivo principal foi melhorar a gestão das redes de bicicletas compartilhadas por meio da coleta de dados e desenvolvimento de algoritmos preditivos, comparando a precisão dos modelos utilizando o *RMSE* e sua versão normalizada *NRMSE*, do inglês *normalized root mean square error*.

Na pesquisa de Duc-Nghiem et al. (2018), foi utilizado um modelo de seleção de variáveis baseado na *BMA*, do inglês, *bayesian model averaging*, complementado com estratégias de modelagem de regressão. A abordagem considerou variáveis independentes como características da infraestrutura, condições de tráfego e características dos ciclistas. Os dados foram coletados em tempo real na cidade de Saitama, Japão, a partir de vídeos gravados nas ruas entre 8h00 e 12h00. As variáveis incluíam a largura das faixas de

tráfego, a presença de veículos parados, o tipo de bicicleta, a existência de pontos de ônibus, o ciclismo em grupo, barreiras entre a calçada e a faixa de rodagem, e a presença de veículos parando. Os resultados indicaram que variáveis como a existência de pontos de ônibus, a largura da calçada e o tipo de bicicleta influenciam significativamente a decisão dos ciclistas ao escolherem ciclovias. O objetivo do estudo foi desenvolver um modelo para prever a escolha de ciclovias com base em variáveis de infraestrutura. O modelo demonstrou boa capacidade preditiva e aplicabilidade prática, sendo validado internamente e externamente.

No estudo de Xu, Ji e Liu (2018), um modelo de previsão de demanda para bicicletas compartilhadas sem estação foi desenvolvido utilizando aprendizado profundo. O *LSTM*, do inglês *long short-term memory* foi empregado para prever a produção e as viagens de bicicletas em intervalos de 10, 15, 20 e 30 minutos. Os dados foram coletados na área central de *Nanjing*, China, ao longo de duas semanas, incluindo informações sobre qualidade do ar, condições meteorológicas (temperatura, ponto de orvalho, umidade, velocidade e pressão do vento) e condições climáticas gerais (como tempo claro, nublado e chuva). A importância das variáveis foi avaliada por meio da Análise de Importância de Variáveis, considerando diferentes conjuntos de dados, como viagens, meteorologia e qualidade do ar. O objetivo do estudo foi desenvolver um modelo preditivo para o sistema de bicicletas sem estação, utilizando aprendizado profundo e grandes volumes de dados. A precisão do modelo foi avaliada por meio de validação cruzada, utilizando métricas como o *RMSE* e *NRMSE*.

No estudo de Ruffieux, Mugellini e Khaled (2018), foi proposto um sistema para otimizar as operações de reequilíbrio de bicicletas em redes de compartilhamento, utilizando algoritmos de aprendizado de máquina como *RF* e *MLP*, do inglês *multi-layer perceptron*. O sistema prevê a disponibilidade de bicicletas em estações específicas para diferentes séries temporais, com intervalos de 15, 30, 60, 120, 360 e 720 minutos. Os dados históricos e as condições climáticas foram coletados a cada 5 minutos, incluindo hora, dia, mês, tipo de feriado, temperatura, umidade, vento e condições meteorológicas, por meio de redes gerenciadas pela *Intermobility SA*. O principal objetivo foi melhorar a gestão das operações de campo nos sistemas de compartilhamento de bicicletas, garantindo uma distribuição equilibrada das bicicletas para atender à demanda dos usuários. O estudo foi realizado nas cidades de *Bienne*, *Vevey* e *Geneva*, na Suíça.

Wang e Kim (2018) utilizou técnicas de aprendizado de máquina, como *RNNs*, do inglês *recurrent neural networks* com *LSTM* e *GRU*, do inglês *gated recurrent unit* , além de *RF*, para prever a disponibilidade de bicicletas em estações do sistema de compartilhamento em *Suzhou*, China. Os dados foram coletados a cada 1, 5 e 10 minutos, refletindo o tempo que os usuários normalmente levam para acessar uma estação de bicicletas. Os dados incluíam o número de bicicletas disponíveis em cada estação, ID da estação e *times-*

tamp, coletados a cada minuto. O objetivo principal foi melhorar a gestão e redistribuição das bicicletas, prevendo a disponibilidade em *Suzhou*. As métricas utilizadas para avaliação foram o *MSE*, do inglês *mean square error*, *MAE*, do inglês *mean absolute error* e *MAPE*. O estudo concluiu que todos os modelos são eficazes para previsões de curto prazo, com o *RF* mostrando vantagem em termos de tempo de treinamento, enquanto o *LSTM* se destacou nas previsões de longo prazo.

Na pesquisa de Zhou et al. (2018), os autores propuseram um modelo para prever a demanda em estações de bicicletas compartilhadas, utilizando cadeias de *Markov*, modelos de regressão, redes neurais e *SVM*. O estudo foi conduzido com dados coletados diariamente ao longo de 2013, que foram usados para prever a demanda em 2014. Os dados foram obtidos do sistema de compartilhamento de bicicletas de *Zhongshan* e incluíam variáveis como estações do ano, feriados, finais de semana, temperatura, precipitação, condições meteorológicas e eventos especiais. O objetivo principal foi desenvolver um modelo para prever a demanda de bicicletas em nível de estação, levando em conta os fatores de influência mais significativos e as restrições entre as estações. O estudo foi realizado em *Zhongshan*, China.

No estudo de Lozano et al. (2018), foi proposto um sistema que oferece ferramentas de previsão para sistemas de compartilhamento de bicicletas, utilizando diversos algoritmos de regressão, como *RF*, *Gradient Boosting* e *Extra Trees*. Os dados foram analisados em intervalos diários, das 7:00 às 22:00, e incluíam informações sobre as viagens (origem e destino), dados climáticos (temperatura, velocidade do vento, pressão e chuva). O estudo foi realizado no sistema de compartilhamento de bicicletas *SalenBici*, em *Salamanca*, Espanha. A seleção de variáveis foi realizada utilizando *GridSearchCV*, que testa várias combinações de parâmetros para identificar os melhores modelos de previsão. A métrica de avaliação utilizada foi o *RMSLE*. O sistema proposto inclui um agente *WebAPI* que oferece acesso aos dados históricos e serviços de previsão de demanda por meio de uma interface web, permitindo aos operadores do sistema visualizar dados históricos e solicitar previsões para auxiliar na realocação de bicicletas.

O estudo de Lin et al. (2018) apresentou o modelo *SMVP*, do inglês *stacking model for variation prediction*, que integra múltiplos modelos treinados pelo *XGBoost*, com diferentes combinações de características, para melhorar o desempenho preditivo. Além dos fatores tradicionais, como temporais, espaciais, históricos e meteorológicos, o algoritmo *K-Medoids* foi utilizado para agrupar as estações com base nesses fatores. Os dados históricos de aluguel e devolução de bicicletas foram extraídos das plataformas *Citi Bike System* (*Nova York*, *EUA*) e *Hangzhou Public Transport Ltda* (*Hangzhou*, China). O estudo utilizou dados diários e temporais para prever o fluxo de bicicletas, com intervalos de 2 horas.

No estudo de Yi et al. (2018), foi utilizada uma abordagem de aprendizado pro-

fundo, especificamente a *conv-LSTM*, do inglês *convolutional long short-term memory*, para prever a distribuição espaço-temporal de curto prazo em um sistema de compartilhamento de bicicletas sem estação na cidade de *Chengdu*, China. Usando dados coletados do aplicativo *Mobike*, o estudo considerou um intervalo de previsão de 1 hora, analisando a quantidade de bicicletas em uma área, a uniformidade da distribuição, a distribuição de uso e a hora do dia. O modelo *conv-LSTM* se mostrou mais eficaz do que a *LSTM* na captura das características espaço-temporais, com avaliação baseada no *MSE* e no *MAE*.

Boufidis et al. (2020) propõe uma ferramenta de previsão de demanda e visualização para operadores de sistemas de compartilhamento de bicicletas, com foco no sistema *ThessBike,* em *Thessaloniki*, Grécia. Utilizando algoritmos de regressão como *RF*, *Gradient Boosting*, *XGBoost* e Redes Neurais, o estudo considerou variáveis históricas e climáticas (temperatura média, velocidade do vento, cobertura do céu e precipitação) e aplicou técnicas de engenharia de recursos para seleção de variáveis temporais e sazonais. A análise foi realizada em intervalos de 1, 2 e 3 horas, e a eficácia dos modelos foi avaliada por meio do *MAE*, *RMSE* e R^2, em inglês *coefficient of determination*. A visualização dos resultados em um ambiente interativo visa apoiar a tomada de decisão dos operadores.

No trabalho de He e Shin (2020), uma *GACNN*, do inglês *graph attention convolutional neural network* foi empregada para prever o fluxo de bicicletas em sistemas de compartilhamento, utilizando dados históricos de viagens, correlações entre estações e fatores externos (como condições meteorológicas e eventos). A previsão foi feita em intervalos de 15 minutos, com dados provenientes dos sistemas *Citi Bike* em Nova York, *Divvy* em Chicago e *Metro Bike* em *Los Angeles*. A seleção de variáveis utilizou a Correlação de *Pearson*, e a eficácia do modelo foi avaliada com base no *RMSE*.

Cantelmo, Kucharski e Antoniou (2020) desenvolveu um modelo de previsão de demanda baseado em um vetor de movimentos, utilizando dados do *Citi Bike* em *Nova York*, com intervalos de 3 horas. Variáveis como o centro de gravidade para origens e destinos, vetores de movimentos específicos para horários de pico, além de dados contextuais (temperatura e precipitação), foram incorporadas ao modelo. A seleção de variáveis foi realizada por meio de *clustering* para reduzir a complexidade dos dados. O desempenho do modelo foi avaliado com a métrica *RMSE*.

O estudo de Hua et al. (2020) aplicou *RF* para prever a demanda dinâmica em sistemas de compartilhamento de bicicletas, utilizando dados fornecidos pela *Beijing Mobike Technology* e analisando intervalos de 15, 30, 45 e 60 minutos. A seleção de variáveis foi realizada por meio de análise de correlação de *Pearson*. A metodologia focou na criação de modelos capazes de prever dinamicamente o uso e a distribuição de bicicletas, com base em dados de viagem coletados em *Nanjing*, China.

Em Ma et al. (2021), foi investigado o impacto do tamanho da grade na previsão

de lacunas de entrada e saída em sistemas de bicicletas compartilhadas sem estação fixa. Utilizando algoritmos de aprendizado de máquina como *LR*, do inglês *linear regression*, *SVR*, do inglês *support vector regression*, *RF* e *GBM*, do inglês *gradient boosting machine*, a performance dos algoritmos foi comparada para diferentes tamanhos de grade, com dados fornecidos pelo Laboratório de Inovação Conjunta de *Big Data* de *Shangai*. A previsão foi realizada em intervalos de 1 hora, utilizando uma técnica de janela deslizante espaço-temporal para a seleção de variáveis.

Jiménez-Bravo et al. (2021) propõem o uso de dados dos usuários para melhorar a previsão de demanda em sistemas de compartilhamento de bicicletas, aplicando técnicas de aprendizado de máquina para comparar modelos baseados em dados das estações e dos usuários no sistema *Salenbici BBS* em Salamanca, Espanha. As previsões foram feitas em três intervalos de tempo ao longo do dia, com a seleção de variáveis realizada por meio de *GridSearch* de hiperparâmetros.

No estudo de Qiao et al. (2021), uma abordagem com *DCNN*, do inglês *dual-channel convolutional neural network* foi utilizada para prever a demanda por bicicletas compartilhadas. Utilizando dados do sistema de compartilhamento de bicicletas da *Bay Area*, o estudo construiu uma matriz de características dinâmica com base em variáveis selecionadas pelo coeficiente de correlação de *Pearson*. O estudo foi realizado em *Chengdu*, China.

Ashqar Mohammed Elhenawy, Almannaa e House (2022) desenvolveram modelos para prever a disponibilidade de bicicletas no sistema de compartilhamento da *Bay Area*, utilizando algoritmos de regressão como *RF*, *LSR*, do inglês *least squares regression* e *PLSR*, do inglês *partial-least-squares regression*. A previsão foi realizada em intervalos de 15 minutos, com dados de viagens e condições meteorológicas. O objetivo foi fornecer detalhes para auxiliar na tomada de decisões pelos operadores do sistema.

Em Dastjerdi e Morency (2022), foi investigada a previsão de demanda de bicicletas compartilhadas durante a pandemia de COVID-19, utilizando modelos *LSTM* e *CNN-LSTM*, do inglês *Convolutional Neural Network - Long Short-Term Memory* com dados históricos do sistema *BIXI* em *Montreal*. O estudo focou na previsão de curto prazo com intervalos de 15 minutos, incorporando dados climáticos e temporais. A avaliação dos modelos foi feita com *MAE* e *RMSE*.

O estudo de Gammelli et al. (2022) explorou o uso de modelos *VP-RNN*, do inglês *variational predictive recurrent neural network* para otimizar a gestão de inventário em sistemas de compartilhamento de bicicletas, com previsões feitas em intervalos de 5, 15, 30 e 60 minutos, utilizando dados históricos e condições climáticas do sistema *Citi Bike* de *Nova York*. As variáveis foram identificadas por meio de processos de *Poisson*.

Lee e Kim (2022) investigaram a previsão de demanda em sistemas de comparti-

lhamento de bicicletas em *Seul*, Coreia do Sul, utilizando *CNN* e *Conv-LSTM*. O estudo considerou dados históricos e condições climáticas, com seleção de variáveis por autocorrelação espacial. As métricas utilizadas incluíram *MAE* e *RMSE*.

No estudo de Lim, Chung e Lee (2022), foi desenvolvida uma abordagem baseada em *RNN* com *LSTM* (modelo *DeepAR*) para prever a demanda futura de bicicletas em *Seul*, Coreia do Sul. O modelo foi treinado com dados históricos de demanda e condições climáticas, focando em previsões diárias e semanais. A avaliação incluiu *RMSE* e *MAPE*.

Por fim, Nannapaneni (2023) propuseram um modelo de previsão de demanda utilizando redes Bayesianas quânticas, aplicadas ao sistema *Citi Bike* em *Nova York*. Utilizando *LSTM* e *GPR* do inglês *gaussian process regression*, o modelo foi avaliado com *MAE* e *RMSE*, visando melhorar a gestão dos sistemas de compartilhamento de bicicletas.

Subramanian et al. (2023) realizaram uma comparação de diversos algoritmos de aprendizado de máquina e séries temporais para prever a demanda de bicicletas no sistema *Capital Bikeshare* em Washington, D.C., EUA. Utilizando *GRU* do inglês *gated recurrent unit*, *ARIMA*, do inglês *autoregressive integrated moving average*, *SARIMA* do inglês *seasonal autoregressive integrated moving average*, *LSTM* e *RF*, o estudo avaliou a performance dos modelos com *MSE*, *RMSE* e *MAE*, com a seleção de variáveis baseada em análise de correlação.

As Tabelas A.1, B.1, C.1 e D.1, que se encontram nos Apêndices A, B, C e D apresentam um resumo sucinto dos trabalhos discutidos nesta seção. No total, foram analisados 30 estudos relevantes, organizados de forma a proporcionar uma visão objetiva das principais contribuições à pesquisa acadêmica. Os estudos analisados foram selecionados a partir de bases de dados científicas reconhecidas, incluindo, mas não se limitando a, *IEEE Xplore, ScienceDirect, Springer International Publishing e IET Intelligent Transport Systems*. Essas tabelas foram estruturadas para sintetizar as informações mais importantes de cada trabalho, facilitando a identificação das características chave e permitindo uma comparação das diferentes abordagens adotadas.

3 Abordagem Proposta

Este capítulo apresenta a abordagem proposta para a previsão da demanda de bicicletas. A metodologia adotada é composta por quatro etapas, conforme ilustrado na Figura 3.1. A primeira etapa, detalhada na Seção 3.1, envolve a definição, extração e construção da base de dados, composta, neste trabalho, por dados históricos de deslocamentos e informações meteorológicas. Na segunda etapa, descrita na Seção 3.2, realiza-se a criação das variáveis de entrada. Em seguida, um método estatístico é aplicado para identificar e ranquear as variáveis mais relevantes, conforme destacado na Seção 3.3. Por fim, na quarta etapa, apresentada na Seção 3.4, a previsão da demanda é realizada utilizando uma técnica inteligente.

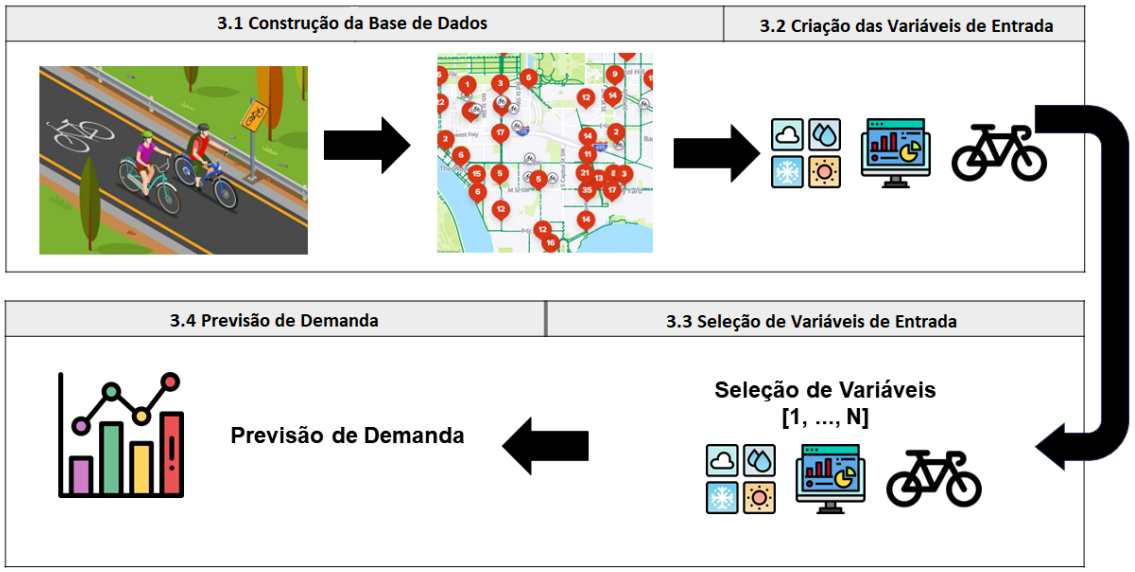


Figura 3.1 – Etapas da abordagem proposta.

3.1 Construção da Base de Dados

Os dados de deslocamentos de bicicletas são normalmente coletados por sistemas de monitoramento equipados com dispositivos *GPS* instalados nas bicicletas. Esses sistemas garantem eficiência e segurança ao permitir que as empresas acompanhem em tempo real a localização dos veículos.

As informações coletadas incluem identificação do passeio, tipo de bicicleta (elétrica ou convencional), duração do trajeto, data e horário de início e término, nomes e *IDs*

das estações de partida e chegada, coordenadas geográficas (latitude e longitude), tipo de usuário (membro ou casual), plano de assinatura e método de pagamento.

A Tabela 3.1[1] apresenta um exemplo de dados registrados pelo sistema de bicicletas compartilhadas de Washington D.C., EUA, contendo campos como *ride-id, rideable-type, started-at, ended-at, start-station-name, start-station-id, end-station-name, end-station-id, start-lat, start-lng, end-lat, end-lng, e member-casual*.

Tabela 3.1 – Registro dos deslocamentos de bicicletas.

rideable_type	**...**	**start_station_name**	**end_lng**	**member_casual**
docked_bike	...	Anacostia Roller Skating Pavillion	-7.69793	casual
classic_bike	...	Anacostia Roller Skating Pavillion	-7.697093	member
classic_bike	...	N Roosevelt St Roosevelt Blvd	-77.153953	casual
docked_bike	...	15th St Constitution Ave NW	-770.446	casual
classic_bike	...	Crystal Dr 15th St S	-77.052808	member
classic_bike	...	Crystal Dr 15th St S	-77.052.808	member
classic_bike	...	Good Hope Rd MLK Ave SE	-7.699523	member

Após a coleta, os dados são processados e filtrados para incluir apenas as informações relevantes: data e horário de início (*Pick-Up*), local de partida, data e horário de término (*Drop-Off*) e local de chegada. A Tabela 3.2 exemplifica registros filtrados.

Tabela 3.2 – Registro de início do deslocamento e nome da estação.

started_at	**start_station_name**
01/07/2023 00:00	17th St & Rhode Island Ave NW
01/07/2023 00:00	17th St & Rhode Island Ave NW
01/07/2023 00:00	17th St & Rhode Island Ave NW
01/07/2023 00:00	Pennsylvania & Minnesota Ave SE
01/07/2023 00:00	13th & O St NW/ Logan Circle
01/07/2023 00:00	15th & P St NW
01/07/2023 00:00	Calvert St & Woodley Pl NW
01/07/2023 00:01	National Harbor Carousel
01/07/2023 00:01	15th & Euclid St NW
01/07/2023 00:01	7th St & Florida Ave NW
01/07/2023 00:02	North Capitol St & F St NW
01/07/2023 00:02	New Jersey Ave & R St NW
01/07/2023 00:02	22nd & I St NW / Foggy Bottom

Para a divisão da área de interesse, existem várias abordagens que podem ser aplicadas para garantir uma separação adequada dos dados, facilitando a análise e a

[1]Na tabela algumas informações foram omitidas devido à limitação de espaço e para melhor visualização.

previsão de demanda por bicicletas. Cada método possui características e benefícios específicos, permitindo a escolha da técnica mais apropriada com base nas necessidades e peculiaridades da região analisada.

Uma das abordagens mais simples consiste em dividir a área de estudo em uma grade regular composta por regiões quadradas ou retangulares. Embora essa técnica facilite a análise espacial, ela pode não refletir adequadamente as variações na densidade de demanda, especialmente em áreas onde as estações estão muito próximas umas das outras. Outra possibilidade é agrupar a área de interesse com base nos prefixos dos nomes das ruas. Por exemplo, ruas que começam com Travessa podem ser agrupadas em uma região, enquanto ruas com os prefixos Avenida ou Rua formam outras regiões. Essa técnica aproveita a estrutura existente nos nomes das vias e pode ser especialmente útil em áreas urbanas, onde os prefixos frequentemente indicam diferentes tipos de zonas, como comerciais, residenciais ou de tráfego intenso. Uma outra abordagem interessante para a divisão da área de interesse é a utilização de pontos de interesse *(POIs)*, como teatros, shoppings, parques, universidades, hospitais, estações de transporte público, entre outros. Esses pontos podem ser usados como referências para identificar regiões com padrões específicos de uso, permitindo uma segmentação mais contextualizada e adaptada às necessidades locais.

Técnicas de agrupamento, como *K-means* ou *DBSCAN* do inglês, *density-based spatial clustering of applications with noise* são usadas para identificar zonas com base na densidade de pontos de dados (localizações de empréstimos e devoluções de bicicletas). Esta abordagem pode identificar áreas de alta demanda, enquanto outras têm atividades mais dispersas e ainda, detectar *outliers* que podem indicar pontos de alta demanda esporádica, como eventos ou locais turísticos ou, descobrir que existem algumas zonas principais onde a maioria das atividades ocorre. A abordagem de análise de *Kernel Density* pode estimar a densidade de eventos de empréstimo e devolução de bicicletas em diferentes pontos da área de estudo, criando uma superfície de densidade contínua que pode ser usada para identificar zonas de alta e baixa demanda.

A divisão da área com base na rede de ruas e ciclovias pode criar regiões que refletem melhor a acessibilidade e conectividade das vias. Esta abordagem fornece *insights* detalhados sobre a utilização das infraestruturas de transporte. Por exemplo, é possível identificar que as avenidas principais têm alta demanda durante os horários de pico, enquanto as ciclovias em áreas residenciais mantêm uma demanda constante ao longo do dia. Além disso, observa-se que certas interseções próximas a estações de transporte público apresentam um aumento significativo no uso de bicicletas.

Após a aplicação dessas técnicas de divisão de regiões, a base de dados original é segmentada em N bases de dados menores, cada uma correspondente a uma das regiões definidas. Isso permite analisar a demanda de bicicletas em diferentes partes da área de

estudo. Com as bases de dados divididas, cada amostra desse conjunto de dados corresponde a uma locação de bicicletas. Ao final do processo, obtém-se um conjunto de dados *Pick-Up* e um conjunto de dados *Drop-off* para cada área de interesse.

O próximo passo consiste em agrupar os dados por intervalos de tempo para realizar a previsão de demanda. Esse agrupamento temporal permite capturar variações na demanda de bicicletas ao longo do dia, da semana ou de períodos específicos, como finais de semana e feriados. Os intervalos podem variar de minutos a horas, dependendo do nível de previsão desejado. Por exemplo, intervalos de 15 minutos, 30 minutos, 1 hora ou até mesmo períodos maiores, como manhã, tarde e noite. Os dados de cada região podem ser agrupados conforme os intervalos de tempo definidos, com cada registro de deslocamento sendo associado a um intervalo específico com base no campo *started-at*. A análise dos dados agrupados pode identificar padrões temporais na demanda de bicicletas, como horários de pico (manhãs e finais de tarde em dias úteis) e variações sazonais ou diárias.

A demanda D por bicicletas em uma área específica p é calculada pela soma dos deslocamentos de partida ou chegada ocorridos dentro de um intervalo de tempo t previamente estabelecido. Um intervalo de tempo pode ser definido como uma divisão do tempo contínuo em períodos sequenciais e uniformes. Por exemplo, considerando um período de 2 horas, um dia de 24 horas resultaria em 12 intervalos de tempo. Para ilustrar essa abordagem, a Tabela 3.3 apresenta um exemplo com os intervalos de tempo, a data de início do deslocamento e o nome da estação.

Tabela 3.3 – Intervalos de tempo.

started_at	**start_station_name**	**Intervalo de tempo**
2021-08-01	1st St & Potomac Ave SE	0h-2h
2021-08-01	1st St & Potomac Ave SE	2h-4h
2021-08-01	1st St & Potomac Ave SE	4h-6h
2021-08-01	1st St & Potomac Ave SE	6h-8h
2021-08-01	1st St & Potomac Ave SE	8h-10h
2021-08-01	1st St & Potomac Ave SE	10h-12h
2021-08-01	1st St & Potomac Ave SE	12h-14h
2021-08-01	1st St & Potomac Ave SE	14h-16h
2021-08-01	1st St & Potomac Ave SE	16h-18h
2021-08-01	1st St & Potomac Ave SE	18h-20h
2021-08-01	1st St & Potomac Ave SE	20h-22h
2021-08-01	1st St & Potomac Ave SE	22h-24h

Ao final dessa etapa, obtém-se uma base de dados consolidada, contendo todas as informações por região e agrupadas por um determinado período. Dessa forma, dispõem-se de N conjuntos de dados de *Pick-Up* e N de *Drop-Off*, correspondentes a N áreas

definidas, em que cada amostra de cada conjunto de dados corresponde a um intervalo de tempo.

3.2 Criação de Variáveis de Entrada

A previsão de demanda pode ser caracterizada como uma previsão de série temporal, ou seja, um conjunto de observações ordenadas cronologicamente. Para criar as variáveis de entrada desta série temporal são utilizadas até 8 defasagens nas variáveis originais, isto é, para prever a demanda um passo à frente $t+1$, serão utilizados os valores defasados na série de $t-1$ até $t-7$. As variáveis construídas com base nas informações do sistema de compartilhamento de bicicletas são: demanda histórica *(DHI)*, demanda histórica dos dias anteriores *(DHD)*, média da demanda histórica *(MDHI)*, média da demanda histórica dos diárias anteriores *(MDHD)*. Além das variáveis construídas com base em informação dos sistemas de compartilhamento serão construídas variáveis utilizando dados provenientes de estações meteorológicas. Ao final, o conjunto de dados será composto por 23 variáveis de entrada, sendo 18 derivadas das informações dos deslocamentos e 5 provenientes de dados meteorológicos. A lista a seguir apresenta detalhes sobre cada uma dessas variáveis:

- Demanda histórica (DHI): demanda de bicicletas dos 8 últimos intervalos de tempo

$$DHI_t, DHI_{t-1}, \ldots, DHI_{t-6}, DHI_{t-7}.$$

- Demanda histórica dos dias anteriores (DHD): demanda de 8 dias anteriores para o mesmo intervalo de tempo

$$DHD_{ta}, DHD_{ta-1}, \ldots, DHD_{ta-6}, DHD_{ta-7}.$$

- Média da demanda histórica ($MDHI$): média da demanda de bicicletas dos 8 últimos intervalos de tempo

$$MDHI = \frac{1}{8}\sum_{i=0}^{7} DHI_{t-i}. \tag{3.1}$$

- Média da demanda histórica dos dias anteriores ($MDHD$): média da demanda de 8 dias anteriores para o mesmo intervalo de tempo

$$MDHD = \frac{1}{8}\sum_{i=0}^{7} DHd_{ta-i}. \tag{3.2}$$

em que t indexa intervalos de tempo e ta indexa os dias.

A inclusão de dados meteorológicos é fundamental, pois as condições climáticas pode ter um impacto significativo no comportamento dos usuários de bicicletas. Diferentes aspectos do clima, como temperatura, precipitação, umidade e velocidade do vento, podem influenciar a decisão das pessoas de utilizarem bicicletas como meio de transporte. Por exemplo, em dias de chuva ou temperaturas extremas (muito frio ou muito calor), a demanda por bicicletas tende a diminuir, pois as condições não são favoráveis para o ciclismo. Por outro lado, em dias com clima ameno e ensolarado, a demanda pode aumentar devido às condições ideais para o uso de bicicletas. A análise dos dados meteorológicos permite identificar esses padrões e ajustar os modelos de previsão de demanda para refletir a influência do clima. A lista detalhada a seguir apresenta as variáveis meteorológicas:

- Temperatura: medida em *Fahrenheit*;
- Umidade relativa do ar: medida em porcentagem;
- Precipitação: medida em polegadas;
- Velocidade do vento: medida em milhas por hora (mph);
- Pressão atmosférica do ar: medida em polegadas de Mercúrio (Hg).

Ao final dessa etapa, obtém-se um conjunto de dados para cada uma das p áreas de interesse, contendo as amostras agrupadas por demandas históricas. Cada amostra incluirá 23 variáveis de entrada e uma variável de saída, que representa a quantidade de bicicletas. A Tabela 3.4 apresenta as 23 variáveis de entrada, com a variável de saída destacada em negrito.

Tabela 3.4 – Identificação das variáveis.

Variáveis de entrada/saída		
1. dhi_1	9. dhd_1	17. media_dhi
2. dhi_2	10. dhd_2	18. media_dhd
3. dhi_3	11. dhd_3	19. Temperatura
4. dhi_4	12. dhd_4	20. Precipitação
5. dhi_5	13. dhd_5	21. Umidade
6. dhi_6	14. dhd_6	22. Pressão
7. dhi_7	15. dhd_7	23. Velocidade_Vento
8. dhi_8	16. dhd_8	**24.Quantidade_de_Bicicletas**

3.3 Seleção de Variáveis de Entrada

Os conjuntos de dados de cada uma das p áreas de interesse são compostos por 23 variáveis de entrada e uma variável de saída (vide Tabela 3.4). Vale destacar que, para cada uma das p áreas, há um conjunto de dados distinto para *Pick-Up* e outro para *Drop-Off*. As 23 variáveis de entrada são, então, submetidas a um processo de ranqueamento. Existem diversos métodos que podem ser utilizados para essa seleção, como a *RFE*, do inglês *Recursive Feature Elimination*, o teste de *Kruskal-Wallis* e a Correlação de *Pearson*.

Neste trabalho, será empregado o método *RFE*. Trata-se de uma técnica que cria um *ranking* das variáveis ao avaliar sua importância de forma iterativa. O *RFE* começa treinando um modelo inicial com todas as variáveis disponíveis. Em seguida, a variável menos importante é removida e o modelo é retreinado com as variáveis restantes. Esse processo de remoção e retreinamento continua até que todas as variáveis tenham sido eliminadas, uma por uma. Ao final, o *RFE* gera um *ranking* das variáveis baseado em sua importância. Essa definição pode ser feita com base em critérios como desempenho do modelo ou análise de variância.

Após a seleção de características, novos conjuntos são formados para cada um com N variáveis de entrada, sendo $N \leq 23$. A Figura 3.2 ilustra as fases desse processo. Ao final, obtém-se um conjunto de dados com todas as informações por região, agrupadas com as N variáveis mais relevantes.

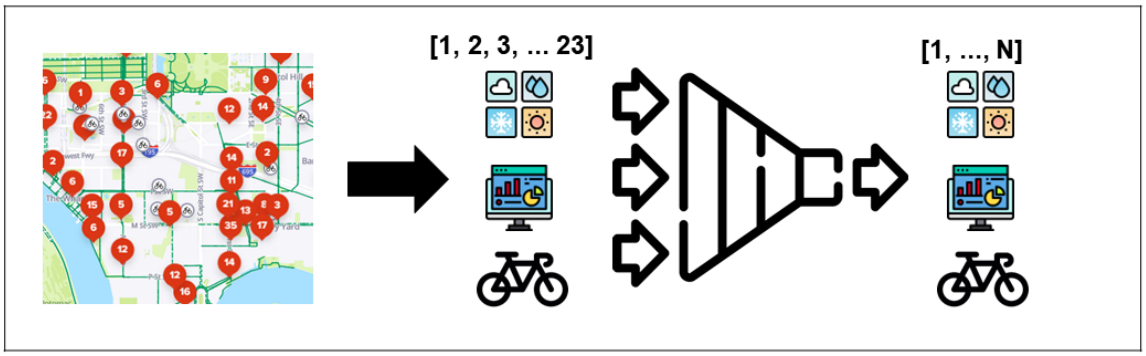


Figura 3.2 – Etapas da seleção de variáveis.

3.4 Previsão de Demanda

Na abordagem proposta, a previsão da demanda é realizada por um modelo de regressão (vide Tabela 3.5), o qual estima a previsão para o próximo intervalo de tempo. A previsão é gerada individualmente para cada uma das regiões de *Pick-Up* e/ou *Drop-Off*, utilizando como entrada o conjunto de variáveis selecionadas na fase anterior. Assim, caso

as variáveis mais significativas para um determinado conjunto de dados sejam $MDHP$, DHI_2, DHD_8, DHD_2, $pressao$, $umidade$, o modelo para este conjunto de dados será definido por:

$$\hat{y}_t = f(MDHP,\ DHI_2,\ DHD_8,\ DHD_2,\ pressao,\ umidade) \tag{3.3}$$

A previsão de demanda de bicicletas pode ser realizada utilizando qualquer algoritmo projetado para tarefas de previsão. A Tabela 3.5 apresenta alguns dos modelos que podem ser aplicados a essa tarefa, organizados em dois grupos: sistemas tradicionais com treinamento *offline* e sistemas *fuzzy* evolutivos com processamento *online*. A Figura 3.3 ilustra como a previsão de demanda por regressão é realizada.

Tabela 3.5 – Sugestão de sistemas inteligentes para previsão de demanda.

Sistemas tradicionais (*offline*)	**Sistemas evolutivos (*online*)**
Rede Neural (SHAH et al., 2023)	ALMMo (ANGELOV; GU; PRÍNCIPE, 2017)
Random Forest (SHEYKHMOUSA et al., 2020)	eNFN (SILVA et al., 2014)
SVM (BANSAL et al., 2022)	eFCE (RODRIGUES; SILVA; LEMOS, 2021a)
Árvore de Decisão (BANSAL et al., 2022)	eFLS (RODRIGUES; SILVA; LEMOS, 2021b)
Regressão Linear (SHAH et al., 2023)	

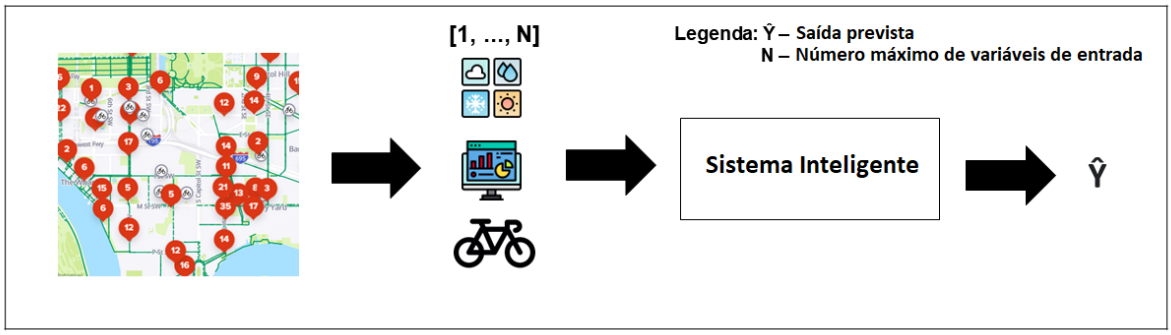


Figura 3.3 – Processo de previsão de demanda.

4 Experimentos e Resultados

Este capítulo descreve os experimentos computacionais conduzidos para analisar o desempenho da abordagem proposta. Nesta pesquisa, a abordagem proposta foi implementada em cinco técnicas tradicionais com treinamento *offline* e quatro técnicas evolutivas com treinamento *online*. A Tabela 3.5, apresentada no Capítulo 3, apresentou os modelos e suas respectivas referências. Inicialmente, na Seção 4.1, é apresentada a metodologia empregada nos experimentos. Em seguida, a Seção 4.2 ilustra os resultados obtidos com os modelos tradicionais. A Seção 4.3 detalha os resultados experimentais alcançados pelos modelos evolutivos. Por fim, a Seção 4.4 traz uma análise comparativa entre o desempenho dos modelos com aprendizado *offline* e aqueles com aprendizado *online*.

4.1 Metodologia

A metodologia utilizada nos experimentos computacionais pode ser sumarizada nas seguintes etapas:

Construção da base de dados: Este estudo utilizou bases de dados contendo registros de deslocamentos de bicicletas, fornecidos pelo serviço de compartilhamento de bicicletas de Washington (Capital Bikeshare, 2023), além de dados meteorológicos disponibilizados por National Centers for Environmental Information (2023). Os deslocamentos foram analisados em intervalos de 2 horas, totalizando 12 intervalos diários. As bases de dados foram combinadas associando os registros de deslocamentos as condições climáticas correspondentes a cada viagem. As áreas de interesse foram definidas a partir do prefixo dos nomes das ruas, resultando em 50 áreas distintas. A Figura 4.1 ilustra a distribuição das estações de bicicletas na cidade de Washington D.C., EUA.

O objetivo dos experimentos foi prever a demanda de bicicletas (*Pick-Up*) em cada área de interesse. Os dados utilizados cobrem o período de 1º de agosto de 2021 a 31 de julho de 2023. Conforme descrito na Seção 3.2, cada base de dados contém 23 variáveis de entrada e uma variável de saída. Para os experimentos, os dados foram normalizados no intervalo de 0 a 1.

Seleção de variáveis de entrada: Foram utilizados conjuntos de dados contendo as 5, 10 e 20 variáveis mais relevantes, selecionadas pelo método *RFE*. Além disso, foi consi-

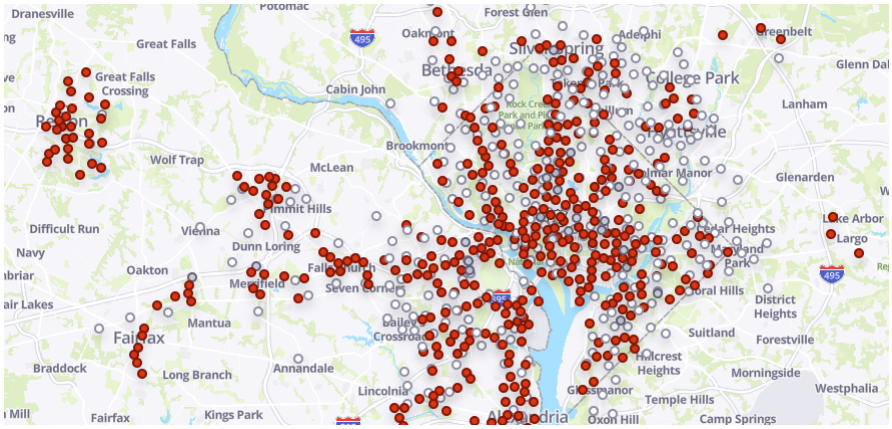


Figura 4.1 – Distribuição das estações de compartilhamento de bicicletas na cidade de Washington D.C., EUA.

derado um conjunto completo com todas as 23 variáveis. Essa abordagem permitiu avaliar o desempenho dos preditores em diferentes configurações de entrada.

Medidas de desempenho: O desempenho dos preditores foi avaliado utilizando três métricas: o Erro Quadrático Médio Raiz (RMSE - *Root Mean Square Error*), o Coeficiente de Determinação (R^2) e o Índice de Erro Não Dimensional (NDEI - *Non-Dimensional Error Index*), conforme definido nas Equações (4.1) a (4.3).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{4.1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y})^2}, \tag{4.2}$$

$$\text{NDEI} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\sigma}, \tag{4.3}$$

onde y_i são os valores observados, $\hat{y}_i$ são os valores previstos, n é o número total de observações e σ representa o desvio padrão dos valores observados.

Preditores Tradicionais (*off-line*): Foram empregados os seguintes preditores tradicionais: SVM, Árvore de Decisão, *Random Forest*, Redes Neurais e Regressão Linear, listados na Tabela 3.5. Como esses preditores possuem comportamento estocástico, os experimentos foram repetidos 21 vezes e os resultados apresentados correspondem à média das execuções na base de teste.

Para este estudo, implementou-se uma rede neural do tipo *Multi-Layer Perceptron (MLP)* utilizando o *MLPRegressor* da biblioteca *scikit-learn* do *Python* e executados no Jupyter versão 3.6.3 (Projeto Jupyter, 2023), para prever a demanda de bicicletas. Inicialmente, os dados foram coletados a partir de múltiplos arquivos .xlsx contidos

em um diretório específico. Cada arquivo foi lido e convertido em um *DataFrame* utilizando a biblioteca pandas, onde a coluna Quantidade de Bicicletas foi definida como a variável alvo, e as demais colunas foram utilizadas como variáveis preditoras.

Os dados foram então divididos em conjuntos de treino e teste, com 70% dos dados destinados ao treinamento e 30% aos testes, utilizando a função *train-test-split*. O modelo MLP foi instanciado através do *MLPRegressor* e treinado com os dados de treino. Após o treinamento, o modelo realizou previsões sobre o conjunto de teste.

Para os outros modelos de regressão, incluindo a árvore de decisão *(DecisionTreeRegressor)*, a floresta aleatória *(RandomForestRegressor)*, a regressão linear (*(LinearRegression)* e a regressão por máquinas de vetores de suporte *(SVR)*, seguimos as mesmas etapas de processamento para todos os modelos. Todos esses modelos foram implementados utilizando a biblioteca *scikit-learn*.

Preditores Evolutivos (*on-line*): Os preditores evolutivos empregados nos experimentos incluem os sistemas fuzzy ALMMo (ANGELOV; GU; PRÍNCIPE, 2017), eNFN (SILVA et al., 2014), eFCE (RODRIGUES; SILVA; LEMOS, 2021b) e eFLS (RODRIGUES; SILVA; LEMOS, 2021a). Esses sistemas foram implementados em Matlab e disponibilizados por seus respectivos autores.

A construção da base de dados seguiu a mesma metodologia utilizada para os modelos *offline*. No entanto, diferentemente dos modelos tradicionais, que são treinados em um conjunto de dados fixo, os modelos *online* foram atualizados continuamente à medida que novos dados eram recebidos.

Os dados foram processados e inseridos continuamente nos modelos, que se ajustaram dinamicamente por meio da inserção ou atualização de regras e neurônios conforme novas informações eram disponibilizadas.

Por serem determinísticos, cada experimento foi executado uma única vez. Os hiperparâmetros utilizados foram os recomendados na literatura, sem ajustes finos, conforme listado a seguir:

- **ALMMo:** $forgettingfactor = 0{,}1$, $densitythreshold = 0{,}8$, $\omega = 10$.
- **eNFN:** $\eta = 10$, $\beta = 0{,}01$, $\omega = 100$.
- **eFCE:** $\lambda = 0{,}05$, $w = 20$, $\Sigma_{\text{init}} = 1$, $N_{\max} = 2$, $\rho = 0{,}10$.
- **eFLS:** $\lambda = 0{,}05$, $w = 20$, $\Sigma_{\text{init}} = 1$, $N_{\max} = 2$, $\rho = 0{,}10$.

4.2 Resultados dos Modelos *Offline*

Nesta seção, são apresentados os resultados obtidos pelos modelos treinados de forma *offline*, ou seja, com os dados previamente divididos em conjuntos de treinamento e

teste. As métricas foram calculadas para avaliar o desempenho dos modelos nas previsões realizadas em intervalos de 2 horas para as demandas de *Pick-Up* (PU). A análise comparativa entre os modelos e o impacto do número de variáveis selecionadas é detalhada a seguir.

A Tabela 4.1 apresenta o desempenho dos modelos de previsão utilizando as métricas *RMSE*, *NDEI* e R^2. Dentre os modelos analisados, o *SVM* obteve o melhor desempenho ao utilizar 20 e 23 variáveis, alcançando um *RMSE* de 0,0683 e um *NDEI* de 0,5456. A análise do R^2 sugere que o uso de mais de 20 variáveis não proporciona ganhos significativos no desempenho, embora ajude a manter resultados consistentes. Assim, 20 variáveis são suficientes, considerando o custo computacional.

A Figura 4.2a ilustra as previsões do modelo *SVM* com 23 variáveis nos últimos 8 dias do período analisado (24/07/2023 a 31/07/2023). Observa-se uma boa correlação entre os valores desejados e previstos, com previsões mais suavizadas nos picos de demanda, apesar de desvios pontuais. De modo geral, o desempenho do *SVM* melhora progressivamente à medida que o número de variáveis aumenta de 5 para 23.

O modelo de Árvore de Decisão também apresentou melhorias à medida que o número de variáveis aumentou. Com 23 variáveis, alcançou um *RMSE* de 0,1165, *NDEI* de 0,7418 e R^2 de 0,4512. O desempenho estabiliza a partir de 10 variáveis, indicando que este número oferece um bom equilíbrio entre custo computacional e precisão. No entanto, com 23 variáveis, o modelo atinge sua melhor performance, capturando picos de demanda e minimizando desvios, conforme ilustrado na Figura 4.2b. A análise do R^2 demonstra que, com 5 variáveis, o modelo sofre com alta variabilidade e erros, enquanto o uso de 23 variáveis proporciona previsões mais consistentes.

O *Random Forest* mostrou um desempenho crescente com o aumento do número de variáveis, estabilizando com 20 e 23 variáveis. O R^2 apresentou um dos maiores aumentos entre os modelos analisados, indicando que cada variável adicional contribuiu significativamente para melhorar o desempenho. Com 23 variáveis, o modelo alcançou um *NDEI* de 0,5119 e demonstrou uma forte capacidade de aprendizado de padrões complexos. A Figura 4.2c destaca a eficácia do *Random Forest* em capturar padrões com maior precisão, reduzindo a variância e os erros de previsão.

O modelo de Rede Neural apresentou uma diferença de desempenho mínima entre 20 e 23 variáveis, indicando que 20 variáveis já são suficientes para capturar as informações relevantes. O R^2 aumentou de 0,6902 para 0,7136, enquanto o *NDEI* diminuiu de 0,5537 para 0,5206. Embora o modelo apresente maior precisão com 23 variáveis, desvios ainda ocorrem em alguns picos de demanda, como mostrado na Figura 4.2d.

Por fim, o modelo de Regressão Linear apresentou um aumento modesto no desempenho com o acréscimo de variáveis. O *RMSE* reduziu de 0,0879 para 0,0845, enquanto

o R^2 aumentou de 0,6902 para 0,7136. O *NDEI* apresentou uma diminuição de 0,5562 para 0,5363, indicando que o modelo é menos sensível ao aumento no número de variáveis. A Figura 4.2e mostra que, com 23 variáveis, as previsões apresentam menor variabilidade e maior correlação com os valores desejados.

Tabela 4.1 – Comparação de diferentes métodos e métricas.

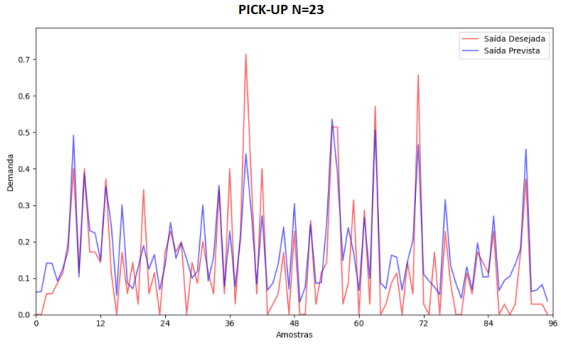
Modelo	Métrica	5	10	20	23
SVM	RMSE	0.0978	0.0885	0.0683	**0.0683**
	NDEI	0.6211	0.5631	0.5456	0.5498
	R^2	0.6149	0.6816	0.6995	0.7005
Random Forest	RMSE	0.0922	0.0832	0.0816	0.0813
	NDEI	0.5847	0.5266	0.5158	0.5119
	R^2	0.6605	0.7208	0.7310	0.7333
Rede Neural	RMSE	0.0866	0.0840	0.0830	0.0824
	NDEI	0.5537	0.5286	0.5231	0.5206
	R^2	0.6902	0.7100	0.7118	0.7136
Regressão Linear	RMSE	0.0879	0.0852	0.0847	0.0845
	NDEI	0.5562	0.5387	0.5393	0.5363
	R^2	0.6902	0.7100	0.7118	0.7136
Árvore de Decisão	RMSE	0.1244	0.1183	0.1165	0.1167
	NDEI	0.7842	0.7569	0.7433	0.7418
	R^2	0.3888	0.4428	0.4423	0.4512

4.3 Resultados dos Modelos *Online*

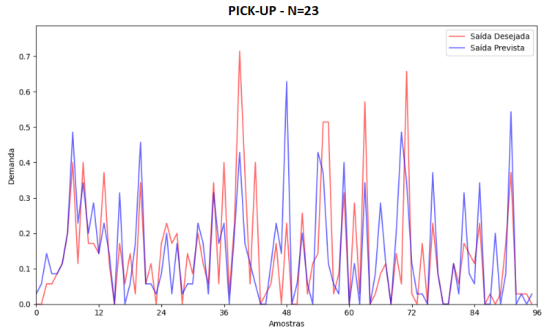
Nesta seção são apresentados os resultados dos modelos preditivos treinados e avaliados de forma *online*, ou seja, com aprendizado incremental à medida que os dados são recebidos. Essa abordagem é essencial em cenários dinâmicos, onde os dados podem variar ao longo do tempo. A análise das métricas realizadas permite identificar o modelo que melhor se adapta a diferentes conjuntos de variáveis.

A Tabela 4.2 exibe o desempenho de quatro Sistemas *Fuzzy* Evolutivos (ALMMO, eNFN, eFCE e eFLS) em termos de *RMSE*, *NDEI* e R^2, considerando quatro diferentes configurações de variáveis: 5, 10, 20 e 23. Essa análise busca esclarecer como cada modelo responde às particularidades de cada conjunto de variáveis e identificar o método mais adequado para cenários que demandam alta precisão preditiva.

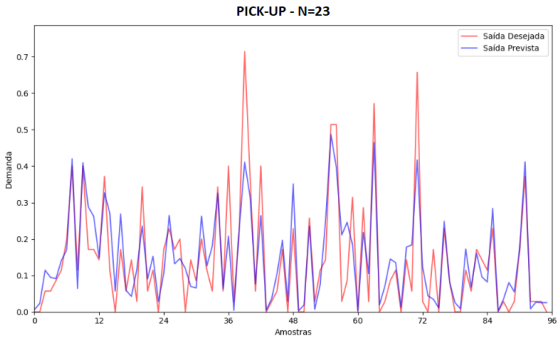
O ALMMO apresentou bom desempenho com 20 variáveis (N=20), onde o *RMSE* diminuiu de 0.0875 (N=5) para 0.0836 (N=20 e N=23), com pouca diferença entre essas últimas configurações. Isso indica que N=20 é a configuração ideal, pois combina bom



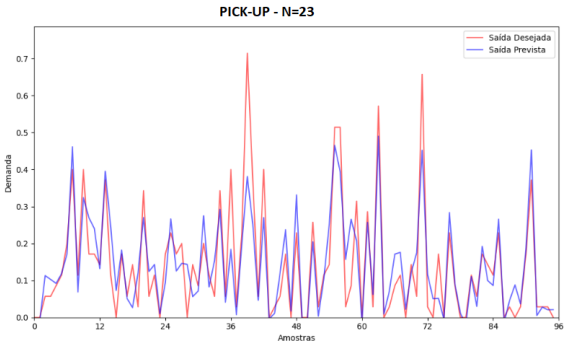
(a) SVM.



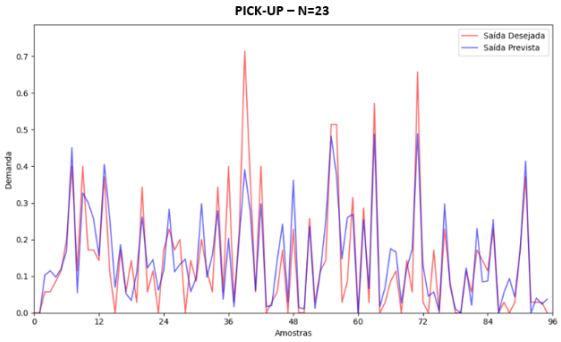
(b) Árvore de Decisão.



(c) *Random Forest*.



(d) Rede Neural.



(e) Regressão Linear.

Figura 4.2 – Gráficos de resultados dos métodos *offline*.

desempenho com menor custo computacional. O valor de R^2 cresceu de 0.6915 (N=5) para 0.7184 (N=20), mantendo-se estável com N=23. O *NDEI* também apresentou uma redução significativa, de 0.5554 (N=5) para 0.5307 (N=20), com pouca variação em N=23. O gráfico 4.3a mostra que, embora eficiente na previsão geral da demanda, o modelo tem dificuldade em prever os picos. Em resumo, o ALMMO com N=20 oferece o melhor equilíbrio entre desempenho e custo computacional.

O modelo eNFN apresentou o maior *RMSE* entre os modelos analisados, começando em 0.1137 (N=5) e reduzindo-se para 0.1078 (N=23), indicando uma melhoria com o aumento do número de variáveis. Apesar disso, os valores de *RMSE* permanecem altos em comparação com outros modelos. O R^2 aumentou de 0.4793 (N=5) para 0.5317

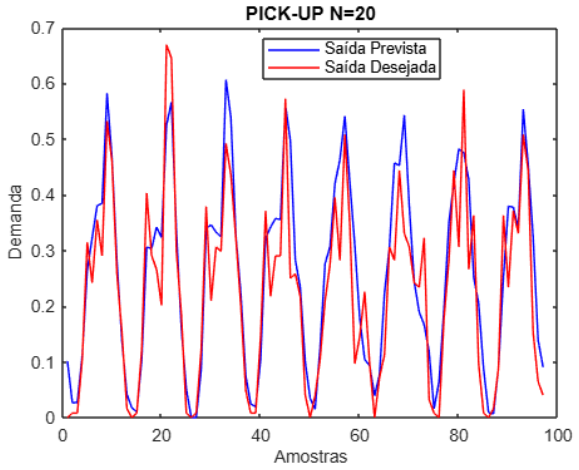
(N=23), mostrando um ganho incremental, mas ainda limitado. O *NDEI* reduziu-se de 0.7216 (N=5) para 0.6843 (N=23), com uma melhoria modesta. O gráfico 4.3b evidencia que o desempenho do eNFN não melhora significativamente com o aumento do número de variáveis.

O modelo eFCE apresentou uma evolução consistente no *RMSE*, que diminuiu de 0.0902 (N=5) para 0.0882 (N=10). Contudo, com N=20 e N=23, o *RMSE* estabilizou-se em 0.0889, indicando que o modelo atinge seu desempenho ideal com N=10. Em relação ao R^2, o modelo manteve-se estável, com valores próximos a 0.6804 desde N=5 até N=23. O *NDEI* apresentou uma redução inicial de 0.5724 (N=5) para 0.5598 (N=10), mas aumentou ligeiramente para 0.5639 (N=20) e 0.5653 (N=23). O gráfico 4.3c, que apresenta o desempenho do eFCE na previsão de demanda com N=20, mostra que o modelo acompanha bem as variações da curva de demanda, destacando-se por sua consistência.

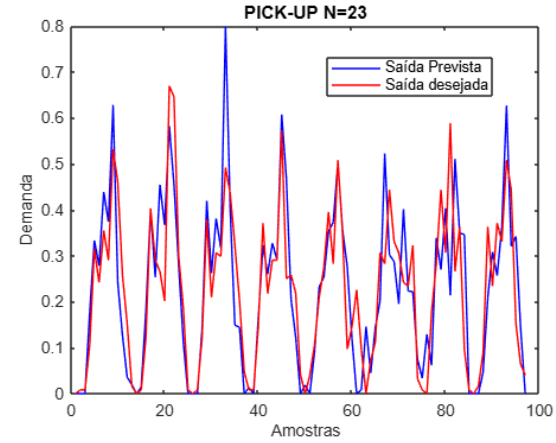
O modelo eFLS seguiu um padrão similar ao eFCE. O *RMSE* diminuiu com o aumento do número de variáveis, alcançando seu valor mínimo de 0.0879 (N=10). Contudo, a partir de N=20, o *RMSE* apresentou um leve aumento. O R^2 acompanhou a tendência de melhoria, com valores próximos a 0.7200 até N=10, estabilizando-se a partir daí. O gráfico 4.3d evidencia que o eFLS, com N=10, ajusta-se bem às variações da demanda, embora enfrente limitações na previsão de picos. Essas variações podem ser atribuídas à influência de variáveis exógenas ajustadas à série temporal. Em resumo, o eFLS com N=10 demonstra bom desempenho, mas sua capacidade de prever picos de demanda ainda pode ser aprimorada.

Tabela 4.2 – Comparação de diferentes métodos e métricas.

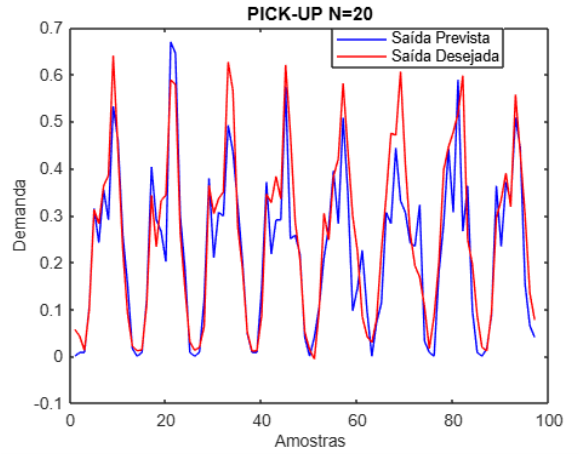
Modelo	Métrica	5	10	20	23
ALMMO	RMSE	0.0875	0.0846	0.0836	**0.0836**
	NDEI	0.5554	0.5370	0.5307	0.5306
	R^2	0.6915	0.7116	0.7184	0.7184
eFLS	RMSE	0.0898	0.0879	0.0886	0.0892
	NDEI	0.5698	0.5578	0.5627	0.5661
	R^2	0.6804	0.6805	0.6805	0.6805
eFCE	RMSE	0.0902	0.0882	0.0888	0.0890
	NDEI	0.5724	0.5597	0.5639	0.5652
	R^2	0.6804	0.6805	0.6805	0.6805
eNFN	RMSE	0.1137	0.1086	0.1079	0.1078
	NDEI	0.7216	0.6893	0.6850	0.6843
	R^2	0.4792	0.5247	0.5307	0.5317



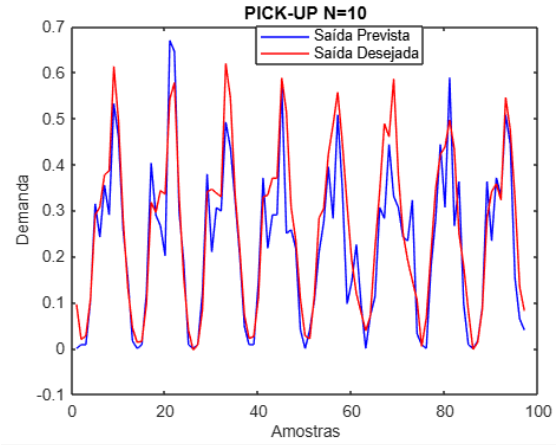
(a) ALMMO.



(b) eNFN.



(c) eFCE.



(d) eFLS.

Figura 4.3 – Gráficos de resultados dos métodos *online*.

4.4 Análise Comparativa entre os Modelos *Offline* e *Online*

Uma análise comparativa entre os modelos *offline* e *online* demonstrou diferenças em termos de desempenho e melhor coleta dos dados. Os modelos *offline*, como *SVM*, *Random Forest*, Redes Neurais, Árvore de Decisão e Regressão Linear são treinados utilizando um conjunto estável de dados históricos. Permite-se que esses métodos coletem determinados comportamentos com uma análise abrangente. Por outro lado, os modelos online, como *ALMMO, eNFN, eFCE* e *eFLS*, ajustam-se dinamicamente durante a execução, sendo mais seguros para situações onde há mudanças frequentes nos valores da demanda.

Em termos de métricas de desempenho, os modelos offline apresentaram resultados mais consistentes em relação ao *RMSE, R^2 e NDEI*, destacando-se o *SVM* e a Rede Neural, que obtiveram os melhores valores com conjuntos de 20 ou 23 variáveis. A Regressão Linear mostrou-se menor desempenho com o aumento do número de variáveis. Entre

os modelos online, o *ALMMO* se destacou por apresentar os menores valores de *RMSE e NDEI*, além do maior R^2. Já os modelos *eFCE e eFLS* alcançaram bom desempenho com um número de N = 10. O *eNFN*, por sua vez, apresentou os maiores valores de RMSE e os menores de R^2 em todas as variáveis.

Em relação ao impacto do número de variáveis, tanto os modelos *offline* quanto *online* mostraram bons resultados no desempenho ao aumentar a quantidade de variáveis. Nos modelos *offline*, o desempenho melhorou até N = 20, com melhores resultados ao incluir mais variáveis, como distribuídas no *SVM* e na Rede Neural. Nos modelos *online*, o *ALMMo* atingiu seu melhor desempenho com N = 20, enquanto os modelos *eFCE e eFLS* alcançaram bons resultados com N = 10.

Conclui-se que a escolha entre os modelos *offline* e *online* é baseada em várias características do problema e nas demandas específicas de cada previsão. Os resultados alcançados nesta pesquisa demonstram que os modelos *online*, especialmente o *ALMMO*, apresentam desempenho superior, destacando-se ao coletar melhor os dados com as mudanças na demanda.

5 Considerações Finais

Este estudo apresenta uma nova metodologia para prever a demanda de usuários de bicicletas utilizando técnicas inteligentes baseadas em aprendizado *offline* e *online*. A abordagem utiliza dados históricos de deslocamentos de bicicletas, combinados com informações meteorológicas, como variáveis de entrada. As variáveis mais significativas foram selecionadas por meio do método *RFE* (Eliminação Recursiva de Atributos), fornecendo uma base otimizada para os modelos preditivos. Para avaliar a eficácia da metodologia proposta, foram conduzidos experimentos computacionais com os modelos de aprendizado *offline* (Rede Neural, *Random Forest*, *SVM*, Árvore de Decisão e Regressão Linear) e os modelos de aprendizado *online* (*ALMMo, eNFN, eFCE e eFLS*). A demanda foi prevista para eventos de *Pick-Ups* em intervalos de 2 horas na cidade de *Washington*, D.C., EUA. Os resultados obtidos foram comparados com uma abordagem de referência e entre os próprios modelos ao longo de um período predeterminado. Entre os modelos de aprendizado *offline*, a *SVM* destacou-se com o melhor desempenho. Já no aprendizado *online*, o *ALMMo* apresentou os resultados mais consistentes. A metodologia proposta revelou-se eficaz para a previsão de demanda de bicicletas, evidenciando: (i) capacidade de processamento de dados em cenários *offline* e *online*; (ii) desempenho competitivo em comparação com abordagens da literatura; e (iii) consistência nas previsões realizadas ao longo de períodos prolongados.

Referências

AFFONSO, E.; MOITA, G.; SILVA, A. efc-evolving fuzzy classifier with incremental clustering algorithm based on samples mean value. *Big Data and Cognitive Computing*, v. 8, p. 183, 12 2024.

ANGELOV, P.; GU, X.; PRÍNCIPE, J. Autonomous learning multi-model systems from data streams. *IEEE Transactions on Fuzzy Systems*, n. 99, p. 1, 2017. ISSN 1063-6706.

ANGELOV, P. P.; GU, X.; PRíNCIPE, J. C. Autonomous learning multimodel systems from data streams. *IEEE Transactions on Fuzzy Systems*, v. 26, n. 4, p. 2213–2224, 2018.

ASHQAR MOHAMMED ELHENAWY, H. A. R. H. I.; ALMANNAA, M.; HOUSE, L. Network and station-level bike-sharing system prediction: a san francisco bay area case study. *Journal of Intelligent Transportation Systems*, Taylor & Francis, v. 26, n. 5, p. 602–612, 2022. Disponível em: <https://doi.org/10.1080/15472450.2021.1948412>.

BAI, J. et al. Multinomial random forest. *Pattern Recognition*, v. 122, p. 108331, 2022. ISSN 0031-3203. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320321005112>.

BANSAL, M. et al. A comparative analysis of k-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decision Analytics Journal*, v. 3, p. 100071, 2022. ISSN 2772-6622. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2772662222000261>.

BAO, R.-J. et al. Correntropy-based evolving fuzzy neural system. *IEEE Transactions on Fuzzy Systems*, v. 26, n. 3, p. 1324–1338, 2018.

BOUFIDIS, N. et al. Development of a station-level demand prediction and visualization tool to support bike-sharing systems' operators. *Transportation Research Procedia*, v. 47, p. 51–58, 01 2020.

CANTELMO, G.; KUCHARSKI, R.; ANTONIOU, C. Low-dimensional model for bike-sharing demand forecasting that explicitly accounts for weather data. *Transportation Research Record*, v. 2674, n. 8, p. 132–144, 2020. Disponível em: <https://doi.org/10.1177/0361198120932160>.

Capital Bikeshare. *Capital Bikeshare: Metro DC's bikeshare service*. 2023. Disponível em: <https://capitalbikeshare.com/>.

DASTJERDI, A. M.; MORENCY, C. Bike-sharing demand prediction at community level under covid-19 using deep learning. *Sensors*, v. 22, n. 3, 2022. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/22/3/1060>.

DUC-NGHIEM, N. et al. Modeling cyclists' facility choice and its application in bike lane usage forecasting. *IATSS Research*, v. 42, n. 2, p. 86–95, 2018. ISSN 0386-1112. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0386111216300516>.

FAN, Y. et al. Distributed forecasting and ant colony optimization for the bike-sharing rebalancing problem with unserved demands. *PLOS ONE*, v. 14, p. e0226204, 12 2019.

GAMMELLI, D. et al. Predictive and prescriptive performance of bike-sharing demand forecasts for inventory management. *Transportation Research Part C: Emerging Technologies*, v. 138, p. 103571, 2022. ISSN 0968-090X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0968090X22000183>.

GAST, N. et al. Probabilistic forecasts of bike-sharing systems for journey planning. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2015. (CIKM '15), p. 703–712. ISBN 9781450337946. Disponível em: <https://doi.org/10.1145/2806416.2806569>.

HAYKIN, S. *Redes Neurais: Princípios e Práticas*. 3ª edição. ed. [S.l.]: Bookman, 2017. Acesso em: 14 dez. 2024.

HE, S.; SHIN, K. G. Towards fine-grained flow forecasting: A graph attention approach for bike sharing systems. In: *Proceedings of The Web Conference 2020*. New York, NY, USA: Association for Computing Machinery, 2020. (WWW '20), p. 88–98. ISBN 9781450370233. Disponível em: <https://doi.org/10.1145/3366423.3380097>.

HUA, M. et al. Forecasting usage and bike distribution of dockless bike-sharing using journey data. *IET Intelligent Transport Systems*, v. 14, 12 2020.

IEA. *International Energy Agency*. 2024. <https://www.iea.org/>. 2024-09-09.

IOMVH. *International Organization of Motor Vehicle Manufacturers - OICA: Production Statistics*. 2024. <https://www.oica.net/>. 2024-09-09.

ITDP. *Instituto de Políticas de Transporte e Desenvolvimento - Guia de Sistemas de Bicicletas Compartilhadas*. 2018. <https://itdpbrasil.org/guia-de-sistemas-de-bicicletas-compartilhadas/>. 2024-09-09.

JIMéNEZ-BRAVO, D. M. et al. Can user data improve bike sharing systems demand forecasting? In: *2021 Telecoms Conference (ConfTELE)*. [S.l.: s.n.], 2021. p. 1–6.

LEE, J.; KIM, J. Evaluation of spatial and temporal performance of deep learning models for travel demand forecasting: Application to bike-sharing demand forecasting. *Journal of Advanced Transportation*, v. 2022, n. 1, p. 5934670, 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2022/5934670>.

LI, Y.; SHUAI, B. Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms. *Multimedia Tools and Applications*, v. 79, p. 1–12, 02 2020.

LIM, H.; CHUNG, K.; LEE, S. Probabilistic forecasting for demand of a bike-sharing service using a deep-learning approach. *Sustainability*, v. 14, n. 23, 2022. ISSN 2071-1050. Disponível em: <https://www.mdpi.com/2071-1050/14/23/15889>.

LIN, F. et al. A stacking model for variation prediction of public bicycle traffic flow. *Intelligent Data Analysis*, v. 22, p. 911–933, 06 2018.

LOZANO, A. et al. Multi-agent system for demand prediction and trip visualization in bike sharing systems. *Applied Sciences*, v. 8, n. 1, 2018. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/8/1/67>.

MA, X. et al. Short-term prediction of bike-sharing demand using multi-source data: A spatial-temporal graph attentional lstm approach. *Applied Sciences*, v. 12, n. 3, 2022. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/12/3/1161>.

MA, Y. et al. Investigating the impact of spatial-temporal grid size on the microscopic forecasting of the inflow and outflow gap in a free-floating bike-sharing system. *Journal of Transport Geography*, v. 96, p. 103208, 2021. ISSN 0966-6923. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0966692321002611>.

MAULUD, A. M. A. D. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, v. 1, n. 2, p. 140–147, Dec. 2020. Disponível em: <https://jastt.org/index.php/jasttpath/article/view/57>.

NANNAPANENI, R. S. Forecasting bike sharing demand using quantum bayesian network. *Expert Systems with Applications*, v. 221, p. 119749, 2023. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417423002506>.

National Centers for Environmental Information. *National Centers for Environmental Information (NCEI)*. 2023. Disponível em: <https://www.ncei.noaa.gov/>.

OSKARBSKI, J.; BIRR, K.; ŻARSKI, K. Bicycle traffic model for sustainable urban mobility planning. *Energies*, v. 14, n. 18, 2021. ISSN 1996-1073. Disponível em: <https://www.mdpi.com/1996-1073/14/18/5970>.

PRAJAPATI, G. L.; PATLE, A. On performing classification using svm with radial basis and polynomial kernel functions. In: *2010 3rd International Conference on Emerging Trends in Engineering and Technology*. [S.l.: s.n.], 2010. p. 512–515.

Projeto Jupyter. *Projeto Jupyter*. 2023. 2023-08-01. Disponível em: <https://jupyter.org/>.

QIAO, S. et al. A dynamic convolutional neural network based shared-bike demand forecasting model. *ACM Trans. Intell. Syst. Technol.*, Association for Computing Machinery, New York, NY, USA, v. 12, n. 6, nov 2021. ISSN 2157-6904. Disponível em: <https://doi.org/10.1145/3447988>.

RENNIE, N. et al. *Analysing and visualising bike-sharing demand with outliers*. 2023. Disponível em: <https://arxiv.org/abs/2204.06112>.

RODRIGUES, F. P. S.; SILVA, A. M.; LEMOS, A. P. Evolving fuzzy system with multivariable gaussian participatory learning and recursive maximum correntropy - efce. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. [S.l.: s.n.], 2021. p. 1–6.

RODRIGUES, F. P. S.; SILVA, A. M.; LEMOS, A. P. Evolving fuzzy system with multivariable gaussian participatory learning and recursive maximum correntropy - efce. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. [S.l.: s.n.], 2021. p. 1–6.

RODRIGUES, F. P. S.; SILVA, A. M.; LEMOS, A. P. Evolving fuzzy predictor with multivariable gaussian participatory learning and multi-innovations recursive weighted least squares: efmi. *Evolving Systems*, v. 13, n. 5, p. 667–686, Oct 2022. ISSN 1868-6486.

RUFFIEUX, S.; MUGELLINI, E.; KHALED, O. A. Bike usage forecasting for optimal rebalancing operations in bike-sharing systems. In: . [S.l.: s.n.], 2018. p. 854–858.

RUFFIEUX, S. et al. Real-time usage forecasting for bike-sharing systems: A study on random forest and convolutional neural network applicability. In: *2017 Intelligent Systems Conference (IntelliSys)*. [S.l.: s.n.], 2017. p. 622–631.

RUSSEL, P. N. S. *Artificial Intelligence, A Modern Approach*. 4ª edição. ed. São Paulo: GEN LTC, 2022. 1080 p. ISBN 978-8595158870.

SALAKEN, S. M. et al. *Forecasting Bike Sharing Demand Using Fuzzy Inference Mechanism*. Cham: Springer International Publishing, 2015. 567–574 p.

SCHUIJBROEK, J.; HAMPSHIRE, R.; van Hoeve, W.-J. Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research*, v. 257, n. 3, p. 992–1004, 2017. ISSN 0377-2217. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221716306658>.

SHAH, A. et al. A comprehensive study on skin cancer detection using artificial neural network (ann) and convolutional neural network (cnn). *Clinical eHealth*, v. 6, p. 76–84, 2023. ISSN 2588-9141. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2588914123000205>.

SHEYKHMOUSA, M. et al. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 13, p. 6308–6325, 2020.

SILVA, A. et al. A fast learning algorithm for evolving neo-fuzzy neuron. *Applied Soft Computing*, v. 14, p. 194–209, 2014. ISSN 1568-4946. Evolving Soft Computing Techniques and Applications.

SUBRAMANIAN, M. et al. Enhancing sustainable transportation: Ai-driven bike demand forecasting in smart cities. *Sustainability*, v. 15, n. 18, 2023. ISSN 2071-1050. Disponível em: <https://www.mdpi.com/2071-1050/15/18/13840>.

TIAN, D.; WEN, Z.; SUN, Y. Analyzing the spatial interaction characteristics of urban area shared bicycle systems: A case study of beijing's central area. *Buildings*, v. 13, n. 10, 2023. ISSN 2075-5309. Disponível em: <https://www.mdpi.com/2075-5309/13/10/2646>.

WANG, B.; KIM, I. Short-term prediction for bike-sharing service using machine learning. *Transportation Research Procedia*, v. 34, p. 171–178, 2018. ISSN 2352-1465. International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)Emerging Transport Technologies for Next Generation Mobility. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352146518303181>.

WHO. *World Health Organization - Global Status Report on Road Safety 2023*. 2024. <https://www.who.int/>. 2024-09-09.

XU, C.; JI, J.; LIU, P. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies*, v. 95, p. 47–60, 2018. ISSN 0968-090X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0968090X18306764>.

YANG, H. et al. Short-term demand forecasting for bike sharing system based on machine learning. In: *2019 5th International Conference on Transportation Information and Safety (ICTIS)*. [S.l.: s.n.], 2019. p. 1295–1300.

YI, A. et al. Retracted article: A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Computing and Applications*, v. 31, p. 1–13, 04 2018.

ZHANG, D. et al. A time-space network flow approach to dynamic repositioning in bicycle sharing systems. *Transportation Research Part B: Methodological*, v. 103, p. 188–207, 2017. ISSN 0191-2615. Green Urban Transportation. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0191261516302697>.

ZHOU, Y. et al. A markov chain based demand prediction model for stations in bike sharing systems. *Mathematical Problems in Engineering*, v. 2018, n. 1, p. 8028714, 2018. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2018/8028714>.

Apêndices

APÊNDICE A – Tabela com os resumos dos trabalhos da literatura

Autor	Técnica	Intervalo	Dados Históricos e Condições Climáticas	Conjuntos de Dados	Seleção de Variáveis	Objetivos	Cidade
Salaken et al. (2015)	Mecanismo de Fuzzy	Diário e Horário	Data, estação, ano, mês, dia da semana, hora, fim de semana, feriado público, condição climática, temperatura, sensação térmica, umidade e velocidade do vento	Capital Bikeshare, Washington DC, EUA	Regras de Wang-Mendel	Prever a demanda de compartilhamento de bicicletas com boa precisão em qualquer ponto do tempo com o conhecimento de informações do calendário juntamente com variáveis climáticas como temperatura, velocidade do vento e umidade	Washington DC, EUA
Gast et al. (2015)	Modelo Tóerico de Filas	5, 15, 30, 60min	Disponibilidade de bicicletas e dados metereológicos	Velib BSS	Processos de Poisson	Previsões probabilísticas obtidas a partir de um modelo teórico de filas não homogêneo no tempo de um sistema de compartilhamento de bicicletas	Paris, França
Zhang et al. (2017)	MIP	3h	Nível de estoque, previsão de chegada de usuários, reposicionamento de bicicletas e rota de veículos	Capital Bikeshare DC e Vélib	Processo de Poisson e cadeia de Markov	Propor uma metodologia dinâmica de reposicionamento de bicicletas que considera a previsão do nível de estoque, a previsão de chegada de usuários, o reposicionamento de bicicletas e a rota de veículos de maneira unificada	Washington, DC, EUA e Paris, França
Ruffieux et al. (2017)	Random Forest e CNN	5, 10, 15 min	Temperatura, umidade, vento e condições gerais do tempo	Citybik2	Estações Individuais e Independentes (IIS), Estações Individuais e Parcialmente Dependentes (IPDS), Estações Totalmente Dependentes (FDS)	Facilitar a coleta e uso de dados de sistemas de compartilhamento de bicicletas para pesquisa, principalmente para desenvolver e comparar algoritmos de previsão de uso de bicicletas	Namur (Bélgica), Essen (Alemanha), Glasgow (Reino Unido), Budapeste (Hungria), Viena (Áustria) e Nice (França)
Schuijbroek, Hampshire e van Hoeve (2017)	MIP e CP	6h às 9h	Dados históricos de demanda, incluindo níveis de estoque nas estações, previsões de chegada de usuários e rotas dos veículos de reposicionamento	Hubway e Capital Bikeshare	Processos de Poisson e cadeias de Markov	Unificar restrições de nível de serviço de limite duplo (que adicionam flexibilidade de inventário) e roteamento de veículos para reequilíbrio estático em sistemas de compartilhamento de bicicletas	Washington e Boston, EUA
Duc-Nghiem et al. (2018)	Regression Modeling Strategy	Início 8h00 e o horário de término variando de 10h00 às 12h00.	Largura das faixas de tráfego, a existência de veículos parados em tempo real, o tipo de bicicleta, existência de ponto de ônibus e ciclismo em grupo	Processamento de vídeo gravado em tempo real nas ruas	Bayesian Model Averaging (BMA)	Prever a escolha de instalações de ciclistas entre instalações na rua (meio-fio, faixa de tráfego e ciclovia (BL)) e instalações fora da rua (calçadas)	Saitama, Japão

Tabela A.1 – Resumos dos trabalhos da literatura.

APÊNDICE B – Tabela com os resumos dos trabalhos da literatura

Autor	Técnica	Intervalo	Dados Históricos e Condições Climáticas	Conjuntos de Dados	Seleção de Variáveis	Objetivos	Cidade
Ruffieux, Mugellini e Khaled (2018)	Random forest e Multilayer Perceptron (MLP)	15, 30, 60, 120, 360, 720 min	Hora, dia, mês e tipo de feriado, temperatura, umidade, vento na corrente, tempo e condições metereológicas	Intermobility (IMY)	GridSearch	Otimizar as operações diárias de reequilíbrio que precisam ser realizadas pelos operadores de sistemas de compartilhamento de bicicletas usando algoritmos de aprendizado de máquina.	Bienne, Vevey, Genebra Suíça
Zhou et al. (2018)	Redes Neurais e SVM	Diário	Estações do ano, feriados, finais de semana, temperatura, precipitação, condições meteorológicas e eventos especiais	Sistema de compartilhamento de bicicletas da cidade de Zhongshan	Cadeia de Markov	Propõe um método de predição baseado no modelo de cadeia de Markov. O modelo proposto é avaliado com base em dados de campo coletados do sistema de compartilhamento de bicicletas	Zhongshan, China
Xu, Ji e Liu (2018)	LSTM NNs	10, 15, 20, 30 minutos	Temperatura, ponto de orvalho, umidade, velocidade e pressão do vento, bem como as condições climáticas gerais, incluindo tempo claro, tempo nublado, chuva fraca, chuva mediana e chuva forte	Crawler multi-processamento	Análise de Importância de Variáveis	Desenvolver um modelo dinâmico de previsão de demanda para compartilhamento de bicicletas sem estações utilizando a abordagem de aprendizagem profunda.	Nanjing, China
Wang e Kim (2018)	RNNs (LSTM e GRU) e Random Forest	1, 5, 10 min	Número de bicicletas disponíveis em cada estação, ID da estação e timestamp	Suzhou Youon Public Bicycle Systems	Não foi mencionado no artigo	Previsão de curto prazo para o uso da estação de acoplamento no caso de Suzhou, na China	Suzhou, China
Yi et al. (2018)	CNN, LSTM, conv-LSTM	1h	Número de bicicletas na área, uniformidade de distribuição, distribuição de uso e hora do dia	Mobike	Arquitetura do modelo conv-LSTM	Previsão de distribuição de curto prazo sob compartilhamento de bicicletas sem estação sistema	Chengdu, China
Lin et al. (2018)	SMVP e XGBoost	2h	Dados diários e temporais	Citi Bike System e Hangzhou Public Transport Ltd	Algoritmo K-Medoids	Propor um modelo de empilhamento para previsão de variação do fluxo de tráfego de bicicletas públicas com base em conjuntos de dados do mundo real	Hangzhou China e Nova York EUA
Lozano et al. (2018)	RandomForest, Dummy, Extra-Trees, Gradiant-Boosting	7h00–22h00	Temperatura, pressão, velocidade do vento e chuva	Salenbici BBS	GridSearch	Propõe um sistema multiagente que fornece ferramentas de visualização e previsão para sistemas de compartilhamento de bicicletas	Salamanca, Espanha
Fan et al. (2019)	Random Forest e AGO	7h00-8h00	Temperatura, ponto de orvalho, umidade, pressão do ar, velocidade do vento e chuva	Hangzhou Public Transport Ltd	Coeficientes de autocorrelação	Destaca o problema de reequilíbrio do compartilhamento de bicicletas (BRP), que consiste em dois aspectos: determinar as demandas desejadas em cada estação e projetar rotas para redistribuir as bicicletas entre as estações	Hangzhou, China

Tabela B.1 – Resumos dos trabalhos da literatura.

APÊNDICE C – Tabela com os resumos dos trabalhos da literatura

Autor	Técnica	Intervalo	Dados Históricos e Condições Climáticas	Conjuntos de Dados	Seleção de Variáveis	Objetivos	Cidade
Yang et al. (2019)	Regressão linear, Árvore de decisão e Árvore de elevação de gradiente	1h	Hora, temperatura, velocidade do vento, visibilidade, condições climáticas, fins de semana, feriados	Citi Bike	Análise de Correlação	Propor um método de previsão de demanda de curto prazo para o valor total do empréstimo das estações	Nova York, EUA
Hua et al. (2020)	Random forest	15, 30, 45, 60 min	|D da viagem, ID do usuário, ID da bicicleta, longitude de origem, latitude de origem, hora de partida, longitude de destino, latitude de destino e hora de chegada.	Beijing Mobike Technology Co. Ltd	Análise de correlação de Pearson	Previsão de distribuição de curto prazo sob compartilhamento de bicicletas sem estação.	Nanjing, China
He e Shin (2020)	GACNN	15 min	Nublado, ensolarado, nebuloso, nebuloso, enevoado, direção do vento, umidade, pressão, dia da semana	Citi Bike, NYC, Metro Bike, LA-Chicago	Correlação de Pearson	Provê projetos espaço-temporais abrangentes, levando em conta correlação e proximidade temporal das estações e seus fluxos de bicicletas	Nova York, Chicago e Los Angeles, EUA
Cantelmo, Kucharski e Antoniou (2020)	Vetores de Movimentos (VM)	3h	Temperatura e precipitação	Citi Bike Nova York	Clustering Espacial e Temporal	Prever o padrão de demanda das bicicletas da cidade de Nova York com uma abordagem de baixa dimensão, utilizando agrupamento de dados em três níveis	Nova York, EUA
Boufidis et al. (2020)	Random Forests, Gradient Boosting, XGBoost e Redes Neurais	1h, 2h e 3h	Temperatura média, velocidade do vento, cobertura do céu e precipitação líquida	Thessbike	Engenharia de Recursos	Permitir decisões precisas e baseadas em dados para otimizar a operação dos sistemas e pode fornecer informações muito úteis para o desenvolvimento de estratégias que incentivem os usuários a devolver/alugar bicicletas para estações específicas de acordo com demanda esperada	Thessaloniki, Grécia
Li e Shuai (2020)	LSTM, CNN e CLTFP	1h	Distância de viagem, densidade de viagem e clima	Mobike e OFO	Análise das características espaciais, temporais e meteorológicas	Prever a distância percorrida e a distribuição de bicicletas compartilhadas sob diferentes condições de tempo e espaço	Shangai, China
Qiao et al. (2021)	DCNN	Não definido	Temperatura, umidade, cobertura de nuvens, velocidade do vento, visibilidade	Bay Area	Coeficiente de correlação de Pearson	Previsão de demanda de bicicletas compartilhadas baseado em redes neurais convolucionais dinâmicas com a necessidade de prever a demanda de bicicletas compartilhadas	Chengdu, China
Jiménez-Bravo et al. (2021)	RandomForest, Dummy, Extra-Trees, Gradiant-Boosting	7h às 12h - 12h às 17h - 17h às 23h	Temperatura, pressão, velocidade do vento e chuva	Salenbici BBS	GridSearch	Propõe-se explorar os padrões de utilização dos usuários cadastrados nesses sistemas para construir modelos capazes de prever a demanda futura	Salamanca, Espanha

Tabela C.1 – Resumos dos trabalhos da literatura.

APÊNDICE D – Tabela com os resumos dos trabalhos da literatura

Autor	Técnica	Intervalo	Dados Históricos e Condições Climáticas	Conjuntos de Dados	Seleção de Variáveis	Objetivos	Cidade
Ma et al. (2021)	LR, SVR, RF, GBM	1h	Períodos do dia, dias úteis, fins de semana, estações do ano e condições climáticas, para prever as lacunas de entrada e saída	Laboratório Conjunto de Inovação de Big Data de Xanga	Sliding window espaço-temporal	Prever a lacuna entre o comportamento de travamento e destravamento em uma grade específica, que é definida como a lacuna de entrada e saída	Shangai, China
Lim, Chung e Lee (2022)	RNN-LSTM DeepAR	Diário	Estado do tempo, temperatura, umidade, velocidade do vento e precipitação	Seoul Metropolitan City	Combinação com o modelo DeepAR	Desenvolver previsões probabilísticas de séries temporais para a demanda de serviços de compartilhamento de bicicletas	Seul, Coreia
Gammelli et al. (2022)	VP-RNN	5, 15, 30, 60 min	Temperatura, probabilidade de chuva, dia da semana, hora do dia	New York Citi Bike	Processos de Poisson	Previsão para otimizar a gestão de inventário em sistemas de compartilhamento de bicicletas.	Nova York, EUA
Lee e Kim (2022)	CNN e Conv-LSTM	1h	EEstado do tempo, temperatura, umidade, velocidade do vento e precipitação	Open Data Plaza e National Climate Data Center	Autocorrelação Espacial	Explorar as dependências espaço-temporais das demandas de viagens usando uma abordagem de aprendizagem profunda	Seul, Coreia
Dastjerdi e Morency (2022)	LSTM e CNN-LSTM	15min	Temperatura, precipitação, velocidade do vento e umidade relativa, além de variáveis temporais como hora do dia e tipo de dia	BIXI de Montreal	Engenharia de características	Previsão de curto prazo, com 15 minutos de antecedência, da procura de bicicletas partilhadas em Montreal, utilizando uma abordagem de aprendizagem profunda	Montreal, Canadá
Ashqar Mohammed Elhenawy, Almannaa e House (2022)	Random Forest e Least-Squares Boosting (LSR) , Partial Least-Squares Regression (PLSR)	15min	Contagem de bicicletas disponíveis em cada estação, condições meteorológicas,estação de origem, estação de destino e horário de cada viagem	Bay Area	Combinação de métodos de análise de importância como RF e LSBoost	Modelar a contagem de bicicletas no BSS da área da baía de São Francisco no nível da estação usando algoritmos de aprendizado de máquina	São Francisco, EUA
Nannapaneni (2023)	LSTM e GPR	2h	Temperatura mínima, média e máxima, velocidade do vento, pressão e chuva, dia da semana e do fim de semana, duração das viagens e localizações geográficas iniciais e finais das estações de bicicletas	New York Citi Bike	Engenharia de características	Desenvolver um modelo de previsão de conjunto com memória de longo e curto prazo (LSTM) e regressão de processo gaussiano (GPR) como modelos individuais cujos resultados são agregados usando um Quantum Rede Bayesiana	Nova York, EUA
Subramanian et al. (2023)	GRU, ARIMA, SARIMA, LSTM, RF	1h	Mês, hora, dia da semana, ano, estação, feriado, dia útil, clima e temperatura	Capital Bikeshare	Análise de Correlação	Comparar o desempenho de várias séries temporais e algoritmos de aprendizado de máquina para prever a demanda de bicicletas usando um registro histórico de dois anos do sistema Capital Bikeshare	Washington D.C., EUA

Tabela D.1 – Resumos dos trabalhos da literatura.