



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS QUIXADÁ
CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

RODRIGO BEZERRA MONTEIRO

**COMPARAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA
PREDIÇÃO DA DISPONIBILIDADE DE BICICLETAS NO PROJETO BICICLETAR
FORTALEZA**

QUIXADÁ
2018

RODRIGO BEZERRA MONTEIRO

COMPARAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA
DISPONIBILIDADE DE BICICLETAS NO PROJETO BICICLETAR FORTALEZA

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Sistemas de Informação
do Campus Quixadá da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Sistemas de Informação.
Área de Concentração: Computação

Orientador: Prof. Me. Regis Pires Magalhães

QUIXADÁ

2018

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

M779c Monteiro, Rodrigo Bezerra.

Comparação de técnicas de aprendizado de máquina para predição da disponibilidade de bicicletas no projeto Bicicletar Fortaleza / Rodrigo Bezerra Monteiro. – 2018.
37 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Sistemas de Informação, Quixadá, 2018.

Orientação: Prof. Me. Regis Pires Magalhães.

1. Aprendizagem do computador. 2. Advinhação. 3. Sistemas de bicicletas públicas. 4. Bicicletar. 5. Controle preditivo. I. Título.

CDD 005

RODRIGO BEZERRA MONTEIRO

COMPARAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA
DISPONIBILIDADE DE BICICLETAS NO PROJETO BICICLETAR FORTALEZA

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Sistemas de Informação
do Campus Quixadá da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Sistemas de Informação.
Área de Concentração: Computação

Aprovado em: ___ / ___ / ____

BANCA EXAMINADORA

Prof. Me. Regis Pires Magalhães (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Me. Dr. Tércio Jorge da Silva
Universidade Federal do Ceará - UFC

Prof. Dra. Ticiania Linhares Coelho da Silva
Universidade Federal do Ceará - UFC

À Deus e minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, a esperança para seguir. Pai, sua presença e apoio significou segurança e certeza de que não estou sozinho nessa caminhada.

AGRADECIMENTOS

Agradeço à minha mãe, Francisca, e ao meu pai, Francisco, pela boa educação que sempre me foi dada, pelas motivações para que eu não desistisse dos meus objetivos e por sempre estarem ao meu lado.

Agradeço ao Prof. Me. Regis Pires Magalhães, por ter me apoiado, pela paciência, e pela excelente orientação que permitiu a conclusão deste trabalho.

Agradeço aos professores Tércio Jorge da Silva e Ticiania Linhares Coelho da Silva, pela disponibilidade em participar da banca desse trabalho e pelas excelentes colaborações e sugestões.

Agradeço ao meu irmão Roberto Bezerra por ter feito o esforço de viajar para Quixadá para me matricular, mesmo quando eu estava em uma classificação inacreditável para ingressar na universidade.

Agradeço a minha irmã Jacqueline por sempre estar me apoiando e incentivando em toda minha trajetória.

Agradeço aos professores Regis Pires, Neto Feitosa, Carlos Diego, Enyo José, Francisco Helder, Germana Ferreira, Rubens Fernandes e Marcos Dantas pelas aulas e pelos conselhos.

Agradeço aos meus amigos, Anthony, Iarlen, Jonas, Teotônio, Chiquim, Alines Mendes, Jecivando, Marcel, Kairão, Igor(gago), George, Daniel, Allan, Messias pelos conselhos, ajudas, incentivos, brincadeiras, saídas e pelos melhores momentos da minha vida acadêmica.

Agradeço aos meus amigos de Fortaleza, Lucileudo, Leilto, Junior Marques, Vitor, Leleo, Onezio, Lucas Saraiva, Pablo, Solon, Roger, Jairo e João Batista que de alguma forma contribuíram para a realização desse sonho.

Agradeço ao meu primo João Victor por fazer parte de toda essa trajetória, dividindo praticamente as mesmas angústias e conquistas.

Agradecer aos servidores técnicos-administrativos do campus Quixadá em especial Roberta Dutra, Gerlyson, Gilmário, Glauco, Rayson, RYanne, Venício, Zarathon por todo apoio e ajuda fornecida.

A todos que direta ou indiretamente fizeram parte da minha formação.

“Tudo o que um sonho precisa para ser realizado
é alguém que acredite que ele possa ser reali-
zado.”

(Roberto Shinyashiki)

RESUMO

O projeto Bicicletar, consiste num programa de bicicletas compartilhadas situado na cidade de Fortaleza no Ceará. O projeto hoje conta com um sistema *web* e *mobile*, tendo como principal funcionalidade o fornecimento de informações das estações de bicicletas, como o número de bicicletas livres, localização, estado e quantidade de vagas. Contudo, através das funcionalidades dos aplicativos os usuários não conseguem obter informações antecipadas sobre a disponibilidade de bicicletas em determinadas estações e horários, gerando atrasos e incômodos para os usuários do programa. Utilizando aprendizagem de máquina, este trabalho propõe a comparação de técnicas de aprendizagem de máquina para predições de bicicletas livres em estações do projeto Bicicletar Fortaleza.

Palavras-chave: Predição. Bicicletar. Bicicletas compartilhadas. Aprendizagem de Máquina. Modelos preditivos.

ABSTRACT

The Bicicletar project consists of a shared bike program located in the city of Fortaleza, Ceará. The project nowadays has a web system and a mobile system, which has as main functionality the provision of information of the bike stations, such as the free bikes numbers, location, status and amount of vacancy. However, through the main functionalities of the applications, the users can't get information in advance about the availability of bikes at specific stations and times, causing delays and inconvenience to users of the program. This article proposes a comparison of machine learning techniques to predict free bikes on stations of the Bicicletar project.

Keywords: Prediction. Bicicletar. Shared bikes. Machine Learning. Predictive models.

LISTA DE ILUSTRAÇÕES

Figura 1 – A Hierarquia do aprendizado	16
Figura 2 – Crescimento de programas de bicicletas compartilhadas	18
Figura 3 – Mapa estações Bicicletar	20
Figura 4 – Estado das estações	20
Figura 5 – Agrupamento de estações individuais em clusters	23
Figura 6 – Passos para execução do trabalho	25
Figura 7 – Bicicletas disponíveis na estação Luíza Távora	29
Figura 8 – Gráfico desempenho dos algoritmos	33

LISTA DE TABELAS

Tabela 1 – Amostra arquivo CSV contendo dados de estações do biciletar	28
Tabela 2 – Dados gerais da estação Luíza Távora	29
Tabela 3 – <i>Features</i> em slots de 1 hora	30
Tabela 4 – Resultados das métricas sem <i>StandardScaler</i> na estação Luíza Távora . . .	32
Tabela 5 – Resultados das métricas com <i>StandardScaler</i> na estação Luíza Távora . . .	32
Tabela 6 – Resultados	33
Tabela 7 – Media ponderada da avaliação do resultado das predições	33

LISTA DE SÍMBOLOS

<i>RMSE</i>	<i>Root Mean Square Error</i> (Erro médio quadrático)
<i>MAE</i>	<i>Mean Absolute Error</i> (Erro Médio Absoluto)
<i>API</i>	<i>Application Programming Interface</i> (Interface de programação de aplicações)
<i>JSON</i>	<i>JavaScript Object Notation</i>
<i>AWS</i>	<i>Amazon Web Service</i>
<i>CSV</i>	<i>Comma-Separated Values</i>
<i>LR</i>	<i>Linear Regressor</i> (Regressão Linear)
<i>SGDR</i>	<i>Stochastic Gradient Descent Regressor</i>
<i>LSVR</i>	<i>Linear Support Vector Regressor</i>
<i>SVR</i>	<i>Support Vector Regressor</i>
<i>RFR</i>	<i>Random Forest Regressor</i>
<i>GBR</i>	<i>Gradient Boosting Regressor</i>
<i>DTR</i>	<i>Decision Tree Regressor</i>
<i>KNN</i>	<i>KNeighbors Regressor</i>

SUMÁRIO

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Aprendizagem de Máquina	15
2.1.1	<i>Algoritmos de aprendizagem de Máquina</i>	16
2.1.2	<i>Métricas para avaliação</i>	17
2.2	Bicicletas Compartilhadas	18
2.2.1	<i>Bicicletar</i>	19
3	TRABALHOS RELACIONADOS	22
4	PROCEDIMENTOS METODOLÓGICOS	25
4.1	Coleta dos dados	25
4.2	Limpeza dos dados	25
4.3	Análise	26
4.4	Definição de <i>features</i> para uso nos modelos preditivos	26
4.5	Criação dos modelos preditivos	26
4.6	Validação e comparação dos modelos preditivos	26
5	RESULTADOS E DISCUSSÕES	27
5.1	Coleta dos dados	27
5.2	Limpeza dos dados	28
5.3	Análise	28
5.4	Definição de <i>features</i> para uso	29
5.5	Criação dos modelos preditivos para cada estação	31
5.6	Validação e comparação dos modelos preditivos	31
6	CONCLUSÕES E TRABALHOS FUTUROS	34
	REFERÊNCIAS	35
	APÊNDICE A	37

1 INTRODUÇÃO

O número de programas de bicicletas compartilhadas e de pessoas que os utilizam vem crescendo pelo mundo nos últimos anos. Esse crescimento decorre de políticas públicas de incentivo ao uso de bicicletas como meio alternativo de transporte em grandes centros urbanos, melhorando a qualidade de vida com a prática de atividade física, diminuição do uso de automóveis e o número de emissão de CO₂ (Gás Carbônico).

Com a facilidade e viabilidade para adesão aos programas por parte da população, o uso das bicicletas se torna, além de um meio para se locomover no dia a dia, uma opção de lazer para pessoas de todas as idades. Isso se torna mais atrativo por conta das estações de bicicletas, que são pontos de locação e devolução de bicicletas, situados em diversos pontos da cidade.

Contudo, o uso do sistema é muitas vezes impossibilitado por falta de bicicletas nas estações próximas ou de preferências dos usuários, causando transtornos como atrasos em compromissos ou impossibilitando diversões de famílias que se programaram para pedalar. Isso se deve a falta de funcionalidades nos sistemas, que antecipem informações para usuários sobre a disponibilidade de bicicletas em determinadas estações e horários.

Hoje, existem alternativas computacionais que proporcionam reconhecimento de padrões, e Aprendizagem de Máquina é um deles. Segundo (SHWARTZ-SHALEV; BEN-DAVID, 2014), o termo refere-se à detecção automatizada de significativos padrões de dados, ou seja, são algoritmos que geram modelos preditivos ou descritivo com base no conjunto de dados fornecidos. Atualmente, percebe-se que técnicas de Aprendizagem de Máquina estão sendo utilizadas em campos como diagnósticos médicos, filtragem de *spam*, reconhecimento de fala, dentre outros.

Partindo do potencial mostrado pela Aprendizagem de Máquina, este trabalho consiste na avaliação de algoritmos de Aprendizagem de Máquina que possam mostrar melhores previsões sobre bicicletas livres no projeto Bicicletar em Fortaleza-CE. As previsões são sobre determinadas estações de bicicletas e horários, ou seja, a partir da hora dada como entrada o resultado será a previsão da média do número de bicicletas disponíveis de uma hora depois da fornecida como entrada, com a finalidade de proporcionar uma melhor programação para os usuários do programa, otimizando assim o tempo dos mesmos.

(LOPES, 2017), realizou um trabalho sobre o Bicicletar onde criou uma ferramenta para análise de dados do programa, com base em perguntas elaboradas direcionadas para os usuários e administradores do Bicicletar. O trabalho gerou contribuições importantes, mostrando

dados de estações e horários onde havia uma maior movimentação de locação e devolução de bicicletas. Contudo, o foco do trabalho foi gerar estatísticas que até então não existiam.

Para este trabalho, foi escolhido Aprendizagem de Máquina como instrumento, com ênfase nos algoritmos de regressão como *Linear Regression*, *SGD Regressor*, *SVR*, *SVR Linear*, *Random Forest Regressor*, *Gradient Boosting Regressor*, *Decision Tree Regressor* e *KNeighbors Regressor*, pois mostram um potencial para aquisição de informações futuras com base em dados iniciais, como a média do número de bicicletas disponíveis dos últimos meses. Isso é possível, por conta do uso e combinação de diversos campos de trabalho como estatística, programação, banco de dados, entre outros, além de proporcionar otimização de sistemas e auxiliar na tomadas de decisões. Diante disso, Aprendizagem de Máquina fornece o suporte necessário para as contribuições deste trabalho, que consiste em gerar resultados comparativos de algoritmos que melhor predizem a média do número de bicicletas disponíveis no Bicicletar.

Os demais capítulos que compõem este trabalho são apresentadas a seguir. No Capítulo 2, são apresentados os fundamentos teóricos para uma melhor compreensão deste trabalho; o Capítulo 3, apresenta os trabalhos relacionados; no Capítulo 4 são apresentados os procedimentos metodológicos que descrevem os passos para se alcançar a solução. Em seguida, o Capítulo 5 demonstra os resultados e discussões deste trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Na fundamentação teórica deste trabalho, são abordados os principais conceitos utilizados em seu desenvolvimento, como aprendizagem de máquina, bicicletas compartilhadas e métricas de avaliação.

2.1 Aprendizagem de Máquina

Aprendizagem de máquina é um campo crescente que ganhou destaque nos últimos anos, por sua capacidade de fornecer respostas com base em experiências passadas. Por conta disso, essa área vem sendo utilizada em campos que necessitam de otimização como áreas médicas, reconhecimento de fala, reconhecimento de imagens, dentre outros.

Segundo (SHWARTZ-SHALEV; BEN-DAVID, 2014), o termo Aprendizagem de Máquina refere-se à detecção automatizada de significativos padrões de dados. Ou seja, é um método de análise de dados que de forma inteligente desenvolve modelos analíticos, através do uso de algoritmos que aprendem interativamente a partir de entradas de dados, permitindo que computadores encontrem resultados preditivos com base no que aprendeu.

Em geral, Aprendizagem de Máquina pode ser classificada em dois tipos: supervisionadas e não supervisionadas. A supervisionada engloba a construção de um modelo estatístico para prever ou estimar um resultado baseado em uma ou mais entradas de dados classificados (MONARD MARIA CAROLINA; BARANAUSKAS, 2003). Para exemplificar, imagine o caso onde se deseja classificar pessoas saudáveis e não saudáveis, onde já existe uma amostra que associa cada pessoa saudável a uma série de variáveis. Então o algoritmo de aprendizagem usaria explicitamente essa informação para no futuro separar uma pessoa saudável de uma não saudável.

De acordo com (BATISTA *et al.*, 2003), técnicas não supervisionadas fazem uso de um conjunto de exemplos (conjunto de dados), com o objetivo de construir um modelo que procura regularidades nos exemplos, formando agrupamentos ou clusters com características similares.

Na Figura 1, é mostrada a hierarquia do aprendizado indutivo, base da aprendizagem de máquina. O foco deste trabalho é no aprendizado supervisionado com ênfase nos algoritmos de regressão, que tem o foco na busca por funções lineares, sendo a variável a prever um atributo numérico (contínuo) (GALVÃO; MARIN, 2009).

Figura 1 – A Hierarquia do aprendizado



Fonte: Elaborado pelo autor

2.1.1 Algoritmos de aprendizagem de Máquina

Existe uma infinidade de algoritmos de aprendizagem de máquina e cada um tem seus pontos positivos e negativos para diferentes funcionalidades. Nesta Seção são abordados os algoritmos utilizados neste trabalho. A escolha dos algoritmos ocorreu com base na popularidade e na eficiência dos mesmos para resolução do problema proposto.

Segundo (PESTANA; GAGEIRO, 2003), Regressão Linear constitui uma equação matemática linear que descreve o relacionamento entre duas variáveis. A Formula 2.1, expressa a equação linear, onde a e b são valores que se determinam com base nos dados amostrais, ou seja, a é uma constante e b é o coeficiente angular. A variável y é o que se deve prever e x é o valor preditor.

$$y = a + bx \quad (2.1)$$

SGD (*Stochastic Gradient Descent*) é um método de predição, onde o parâmetro a prever é atualizado a cada iteração, como na maioria dos algoritmos de otimização, esperando obter uma aproximação do método do gradiente descendente (AGUIAR, 2016). De acordo com (PONTI; COSTA, 2017), SGD *Regressor* tem a formula descrita na Equação 2.2

$$W_{t+1} = W_t - \eta \sum_{j=1}^B \nabla \mathcal{L}(W; x_j^B) \quad (2.2)$$

SVR (*Support Vector Regressor*) baseia-se na metodologia do aprendizado supervisionado, onde tem como objetivo encontrar uma função que apresente uma margem de erro caracterizada por um intervalo. Já o SVR Linear, descende do SVR, tendo mais flexibilidade na escolha de penalidades e funções de perda e deve ser dimensionado melhor para um grande número de amostras, que é o caso deste trabalho.

Segundo (SATO *et al.*, 2013), *Decision Tree Regressor*, são árvores de decisão que apresentam como entrada um vetor de atributos e uma resposta como valor de saída. A árvore funciona com divisões de conjuntos de dados em subconjuntos de forma recursiva. A separação dos dados acontece até que cada subconjunto esteja homogêneo.

Random Forest Regressor é formado por uma coleção de árvores de regressão, onde cada árvore é construída a partir de uma reamostra aleatória do conjunto de treinamento original (BASTOS *et al.*, 2013).

Gradient Boosting Regressor Constrói um modelo aditivo de maneira progressiva, ou seja, em cada estágio uma árvore de regressão é ajustada.

KNeighbors Regressor, conhecido como k-vizinhos mais próximos, trata-se de um método utilizado como classificador ou regressor em reconhecimentos de padrões. Segundo (NETO *et al.*, 2016), *KNeighbors Regressor* utiliza métricas de distância ou similaridade para encontrar K pontos mais próximos no espaço das amostras de treinamento.

Os algoritmos apresentados nessa seção, serão utilizados na construção dos modelos preditivos para prever a média de bicicletas disponíveis no programa Bicicletar Fortaleza.

2.1.2 Métricas para avaliação

Em Aprendizagem de Máquina, para avaliar a qualidade do modelo em relação a um conjunto de dados, utiliza-se medidas quantitativas de erros, que consiste na diferença entre os valores reais e os preditos (LEGATES; MCCABE, 1999).

Perante o exposto, neste trabalho são utilizadas duas métricas com nível significativo de aceitação e utilização na comunidade científica para obter quantificação dos erros: RMSE e MAE, com as seguintes definições:

- *Root Mean Squared Error* (RMSE): é a medida da magnitude média dos erros estimados. Tem valor sempre positivo e quanto mais próximo de zero, maior a

qualidade dos valores medidos ou estimados, calculado pela Equação 2.3.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - O_i)^2} \quad (2.3)$$

- *Mean Absolute Error* (MAE): erro médio absoluto, definido pela Equação 2.4 é a medida média do erro absoluto.

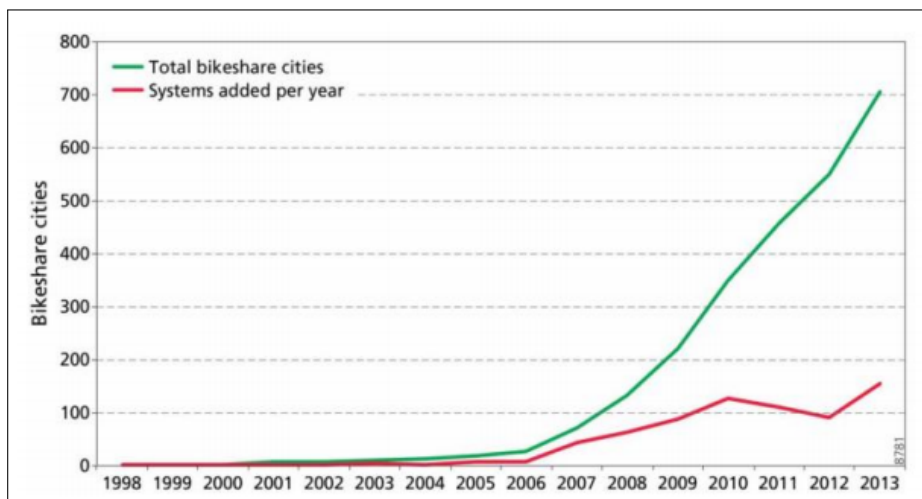
$$MAE = \frac{1}{n} \sum_{i=1}^n |E_i - O_i| \quad (2.4)$$

Onde, E_i e O_i são os valores estimados e observados (medidos), respectivamente, e n é o número de observações (ALVES; VECCHIA, 2011)

2.2 Bicicletas Compartilhadas

Segundo (SHAHEEN *et al.*, 2011), o primeiro programa de bicicleta compartilhadas lançado no mundo foi na Europa em 1965, o que desde então teve um crescimento com inúmeros programas surgindo pelo mundo. Atualmente, estima-se que existam mais de 500 programas de compartilhamento globalmente (WANG, 2016). A Figura 2 mostra uma visão do crescimento entre os anos de 1998 e 2013, onde a linha verde mostra o número total de cidades com programas de bicicletas compartilhadas e a linha vermelha mostra a quantidade de sistemas criados por ano.

Figura 2 – Crescimento de programas de bicicletas compartilhadas



Fonte: wang

Alguns autores categorizam os programas em quatro gerações, desde a sua criação, levando em conta os sucessos e fracassos ao longo do tempo de existência, como (PARKES *et al.*, 2013). A primeira é conhecida como *White Bikes*, que consistiu na inserção de bicicletas nas ruas de Amsterdam para uso público de forma gratuita. Contudo, o programa foi categorizado como de 1º geração por conta dos roubos e danos das bicicletas (FISHMAN, 2016). Já na segunda geração, *Coin Deposit Systems*, foi imposto um depósito de moedas para fazer uso das bicicletas, porém também deixou a desejar no quesito segurança. A terceira geração, *IT-Based Systems*, foi caracterizada pelo uso de estações de engate para locação e devolução de bicicletas, com método de pagamento através de cartões de crédito e um sistema de rastreamento das mesmas. Por fim, a quarta geração chamada *Demand Responsive, Multi-Modal Systems* onde busca a integração com sistemas de transportes maiores como trens, metrô e ônibus (SHAHEEN *et al.*, 2011).

Uma das definições mais aceitas de bicicletas compartilhadas é "Permitir que qualquer pessoa retire uma bicicleta em um local e a devolva ao sistema em outro local viabilizando assim o transporte ponto-a-ponto por tração humana" (GAUTHIER COLIN HUGHES, 2014). De uma forma mais direta, a provisão de bicicletas que podem ser retiradas e devolvidas em estações dedicadas e autosuficientes por tempos curtos, geralmente 30 minutos (FISHMAN, 2016).

Neste trabalho são utilizados os dados do programa Bicicletar¹ Fortaleza, que se encaixa na definição de programas de bicicletas compartilhadas, bem como representa uma possível contribuição para a melhoria do programa.

2.2.1 Bicicletar

Situado na cidade de Fortaleza, o Bicicletar é um projeto de bicicletas compartilhadas gerido pela empresa Sertell² com apoio da Unimed³ Fortaleza. O projeto pertence a prefeitura de Fortaleza e foi lançado para melhorar a mobilidade da cidade. Hoje o projeto conta com 80 estações para adultos e 4 estações para crianças (Mini Bicicletar), onde a maioria das estações contém 12 vagas para disposições das bicicletas.

Para utilização das bicicletas, as pessoas precisam se cadastrar, obter passes e liberar as bicicletas por meio do aplicativo do programa ou sem cadastro, telefonando para a central do Bicicletar e adquirindo um passe diário. Para que a locação seja possível, cada estação conta

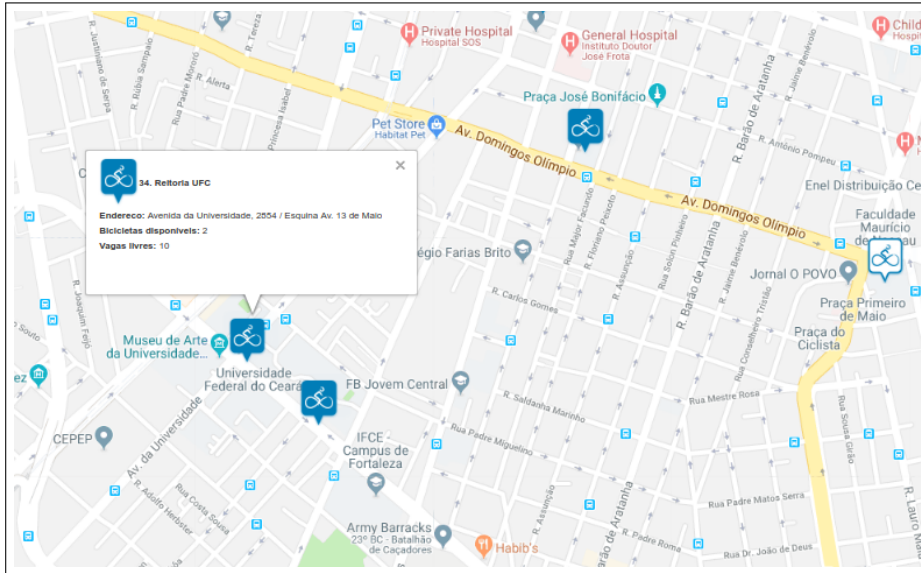
¹ <http://www.bicicletar.com.br>

² <http://www.serttel.com.br/>

³ <http://www.unimedfortaleza.com.br/>

com conexão *wireless*, meio pelo qual é feito a comunicação e troca de dados para retirada ou devolução de bicicletas (BICICLETAR, 2018).

Figura 3 – Mapa estações Bicletar



Fonte: (BICICLETAR, 2018)

Figura 4 – Estado das estações



Fonte: (BICICLETAR, 2018)

Como pode-se observar na Figura 3, além de informações sobre o uso do sistema, no site do programa, é disponibilizado para os usuários um mapa contendo informações em tempo real de todas as estações como, bicicletas disponíveis, número de vagas e estado da estação.

O estado da estação pode assumir diferentes estados como em operação, implemen-

tação, manutenção, todas vagas ocupadas, nenhuma bicicleta disponível e *offline*, como ilustrado na Figura 4.

3 TRABALHOS RELACIONADOS

Nesta seção, são apresentados trabalhos que fazem previsões de tráfego em sistema compartilhamento de bicicletas que servem de base para a construção deste trabalho. Um quadro comparativo é apresentado ao final da seção.

O trabalho realizado por (LI *et al.*, 2015), intitulado *Traffic Prediction in a Bike-Sharing System*, tem como objetivo tratar o desequilíbrio causado pelas locações e devoluções de bicicletas em diferentes estações e em variados períodos, sendo necessário haver uma redistribuição de bicicletas entre as estações do sistema de compartilhamento de bicicletas. Então foi implementado um algoritmo nos sistemas Nova York e Washington DC, onde foram agrupadas estações formando *clusters*, como mostrado na Figura 5, onde os pontos demarcados por círculos denotam estações de um mesmo *cluster*, ou seja, C1, C2 e C3 são *clusters* e S1 e S2 indicam estações dentro do *cluster* C1. Com isso, o número total de bicicletas alugadas em uma cidade é predita através da técnica *Gradient Boosting Regression*. Em seguida, um modelo de inferência baseado em multi-similaridade é proposto para prever a proporção de aluguel e devolução entre *clusters* e a transição *inter-cluster*, com base no número de bicicletas alugadas e devolvidas. O trabalho aqui proposto utiliza técnicas de aprendizagem de máquina para obter resultados semelhantes, porém a avaliação das técnicas são o principal alvo de pesquisa do mesmo.

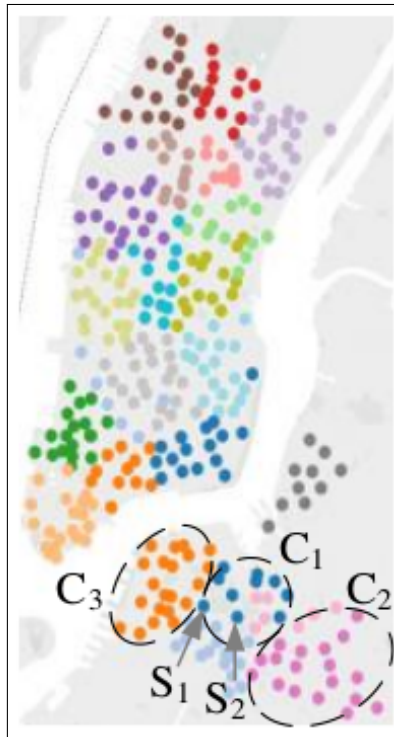
O trabalho de (WANG, 2016) baseia-se na plataforma Kaggle¹, a maior comunidade do mundo voltada para modelagem e competições de análise preditiva, onde pesquisadores e empresas podem disponibilizar seus dados para que analistas do mundo todo possam competir mostrando as melhores soluções. Em cada desafio da plataforma é mantido um *ranking* onde são classificados os melhores resultados da competição. Demanda de bicicletas compartilhadas² é uma competição no Kaggle, que fornece dados do Capital *Bikesharing* em Washington DC para combinar padrões de uso histórico com dados meteorológicos com a finalidade de prever a demanda de aluguel de bicicletas.

Em (WANG, 2016), centra-se no *City Bike*, um dos maiores projetos de partilha de bicicletas do mundo, usando como base o problema da plataforma Kaggle chamado *Bike Sharing Demand*, que fornece dados do Capital *Bikeshare* na cidade Washington, Estados Unidos. O trabalho utiliza dados do *City Bike*, dados meteorológicos, dados de férias e integra-os a fim de fornecer uma solução de visão geral para o problema de reequilíbrio de bicicletas, para ajudar

¹ <https://www.kaggle.com>

² <https://www.kaggle.com/c/bike-sharing-demand>

Figura 5 – Agrupamento de estações individuais em clusters



Fonte: (LI *et al.*, 2015)

a gerenciar melhor o programa *City Bike*. A semelhança entre o trabalho de (WANG, 2016) e este trabalho de conclusão de curso é que ambos tentam prever bicicletas disponíveis utilizando aprendizagem de máquina, contudo diferem pelo contexto de países e programas distintos, e a não utilização de dados de férias.

(LOPES, 2017), realizou um trabalho sobre o Bicicletar, onde criou uma ferramenta para análise de dados do programa, com base em perguntas elaboradas direcionadas para os usuários e administradores do Bicicletar. O trabalho gerou contribuições importantes, mostrando dados de estações e horários onde haviam uma maior movimentação de locação e devolução de bicicletas. Contudo, o foco do trabalho foi gerar estatísticas que até então não haviam.

O Quadro 1 apresenta um resumo comparativo entre os trabalhos relacionados e esta obra.

Quadro 1 – Trabalhos Relacionados

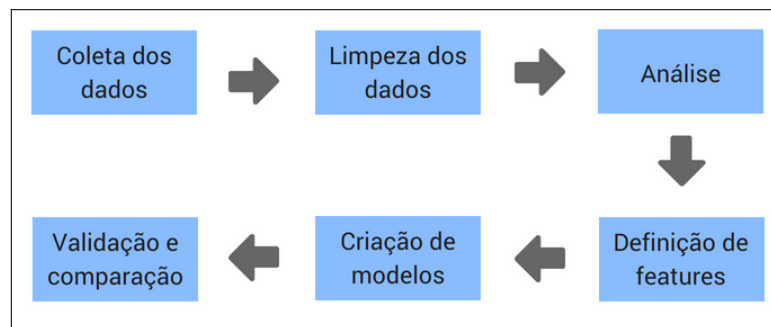
Trabalhos	Métodos Preditivos	Programas de Compartilhamento	Métricas de Avaliação
Li et al. (2015)	GBR	City Bike Capital Bikeshare	RMSLE ER
Wang (2016)	Random Forest, Árvore de Decisão, Rede Neural	City Bike	RMSLE
Lopes (2017)	Não utilizou	Bicicletar Fortaleza	Perguntas Elaboradas pelo autor
Este Trabalho	LR, SGDR, LSVR, SVR, RFR, GBR, DTR, KNN	Bicicletar Fortaleza	RMSE MAE

Fonte: Elaborado pelo autor

4 PROCEDIMENTOS METODOLÓGICOS

Nos procedimentos metodológicos, são abordados os passos para o desenvolvimento deste trabalho.

Figura 6 – Passos para execução do trabalho



Fonte: Elaborado pelo autor

4.1 Coleta dos dados

Obter dados do programa Bicicletar através de uma API (*Application Programming Interface*) pública, disponível na plataforma GitHub¹, que utiliza um *script* escrito na linguagem Python. O *script* recebe como parâmetro o endereço ou uma palavra chave de um endereço, e retorna como resultado os dados da estação buscada. Existe também uma função para buscar os dados de todas as estações em uma única consulta. Neste trabalho utiliza-se esta última função, pois pretendemos comparar técnicas de predição para todas as estações.

4.2 Limpeza dos dados

Devido a possíveis falhas na coleta, é comum que os dados não estejam preparados para que os métodos de predição sejam aplicados diretamente. Dependendo da qualidade desses dados, algumas ações podem ser necessárias. Este processo de limpeza dos dados geralmente envolve filtrar, combinar e preencher valores vazios.

Esta é uma parte crucial no processo, pois a qualidade dos dados vai determinar a eficiência dos algoritmos de predição. Por conta dessa importância, esse passo se faz presente neste trabalho.

¹ <https://github.com/paoloo/DadosBicicletar>

4.3 Análise

Nesta etapa é realizada uma análise dos dados para verificar se realmente estão preparados para os próximos processos. Além disso, uma análise mais aprofundada dos dados se faz presente nesse passo para o melhor entendimento das informações e um melhor planejamento das técnicas de predições que serão utilizadas.

4.4 Definição de *features* para uso nos modelos preditivos

Nem sempre o uso de muitas *features* (características) é sinônimo de predições precisas. Isso se deve ao grande número de informações desnecessárias e complexas ao real foco da solução do problema, imposto por informações sem relevância. Para evitar estes excessos, neste trabalho é realizado um levantamento e definição das *features* com maior importância para a predição de bicicletas disponíveis. A escolha das *features* ocorre a partir dos dados coletados.

4.5 Criação dos modelos preditivos

A criação dos modelos preditivos é um dos passos da aprendizagem de máquina e tem uma relevante importância para todo processo. Esse passo tem como finalidade a criação de um modelo a partir dos dados de treinamento usando uma determinada técnica de predição. Os modelos são estruturas de aprendizado resultante do treinamento dos algoritmos de aprendizagem de máquina, e são utilizados para prever informações futuras.

Os modelos são obtidos nessa fase a partir dos algoritmos citados na Seção 2.1.1 e servem para aplicar no conjunto de testes, dados que não utilizam o seu rótulo de resposta, com o objetivo de obter uma predição e estimar o desempenho obtido com esse algoritmo.

4.6 Validação e comparação dos modelos preditivos

A partir de um modelo construído, os dados são validados em um conjunto de testes usando como base o modelo fornecido. O resultado deste processo é a predição de bicicletas disponíveis. A partir disso é realizado um processo de medição usando como métricas RMSE e MAE para avaliar os resultados obtidos com uma determinada técnica de predição utilizada. Quanto menores os valores das métricas do RMSE e MAE, mais a predição se aproxima do valor real(rótulo).

5 RESULTADOS E DISCUSSÕES

Neste Capítulo, os resultados obtidos são apresentados e discutidos. Na Seção 5.1, são apresentados os dados coletados. Na Seção 5.2, são apresentados os resultados da limpeza dos dados. Na Seção 5.3, são apresentados os resultados da análise. Na Seção 5.4, são apresentados o resultado das escolhas das *features*. Na Seção 5.5, é apresentado o resultado da criação dos modelos preditivos. Na Seção 5.6, são apresentados os resultados da validação e comparação dos modelos preditivos.

Todos os experimentos reportados aqui foram executados em um computador portátil com um processador Intel Core™ i7, 6GB de memória RAM e sistema operacional Ubuntu 16.04 LTS.

5.1 Coleta dos dados

Este procedimento foi viabilizado através de uma API¹ pública disponibilizada na plataforma GitHub. Trata-se de um extrator de dados, que acessa o site do Bicicletar e fornece as informações tanto diretamente no terminal, como através de API, retornando um GeoJSON² das estações, em tempo real. Um exemplo do GeoJSON gerado pode ser visto no Apêndice A. As informações são tratadas de forma a propiciar tanto uma visão completa de todas as estações quanto apenas informações resultantes de uma busca de estações por local.

Os dados foram coletados com auxílio de um servidor na plataforma Microsoft Azure³ e outro na *Amazon Web Service*⁴(AWS), onde foi colocado em execução um *script* para coleta dos dados. Os dados foram coletados a cada minuto e no período de 26/02/2018 a 21/03/2018.

O armazenamento das informações foi realizado em pastas nomeadas com a data do dia da coleta. Cada pasta possui vários arquivos no formato CSV⁵, onde cada um representa uma estação diferente. Nos arquivos estão contidos 12 *features* como demonstra a Tabela 1.

¹ *Application Programming Interface*

² <http://geojson.org/geojson-spec.html>

³ <https://azure.microsoft.com>

⁴ <https://aws.amazon.com/pt/>

⁵ *Comma-Separated Values*

Tabela 1 – Amostra arquivo CSV contendo dados de estações do bicicletar

Data/ Hora	Id Estação	Nome Estação	Lat	Log	Endereço da estação	Status online	Status operacional	Status interno	Bikes Disp 1	Bikes Disp 2	Vagas
2018-02-26 23:47:03	5	Frei Mansueto	-3.735 9524	-38.48 8797	Rua Dom Luís, 1400B	A	EO	Est_ Vazia -1	0	0	12
2018-02-26 23:48:04	5	Frei Mansueto	-3.735 9524	-38.48 8797	Rua Dom Luís, 1400B	A	EO	Est_ Vazia -1	0	0	12
2018-02-26 23:49:04	5	Frei Mansueto	-3.735 9524	-38.48 8797	Rua Dom Luís, 1400B	A	EO	Est_ Vazia -1	0	0	12
2018-02-26 23:50:05	5	Frei Mansueto	-3.735 9524	-38.48 8797	Rua Dom Luís, 1400B	A	EO	Est_ Vazia -1	0	0	12
2018-02-26 23:51:06	5	Frei Mansueto	-3.735 9524	-38.48 8797	Rua Dom Luís, 1400B	A	EO	Est_ Vazia -1	0	0	12
...

Fonte: Elaborado pelo autor

5.2 Limpeza dos dados

Na limpeza dos dados fez-se necessária a seleção das informações mais importantes como hora/data, quantidade de bicicletas disponíveis e quantidade de vagas. Toda essa informação foi agrupada por estação e todo seu período de coleta de dados foi posteriormente colocado em um novo arquivo CSV para cada estação. Durante essa separação foi observado que em determinados momentos houve indisponibilidade no sistema de coleta, causado por uma possível queda no sistema do Bicicletar. Essa falha foi tratada excluindo as linhas onde o dado estava inconsistente. Houve ocasiões onde determinadas estações sumiram e reapareceram no sistema de coleta, dando a entender que possivelmente foi desativada para manutenção e reativada em seguida.

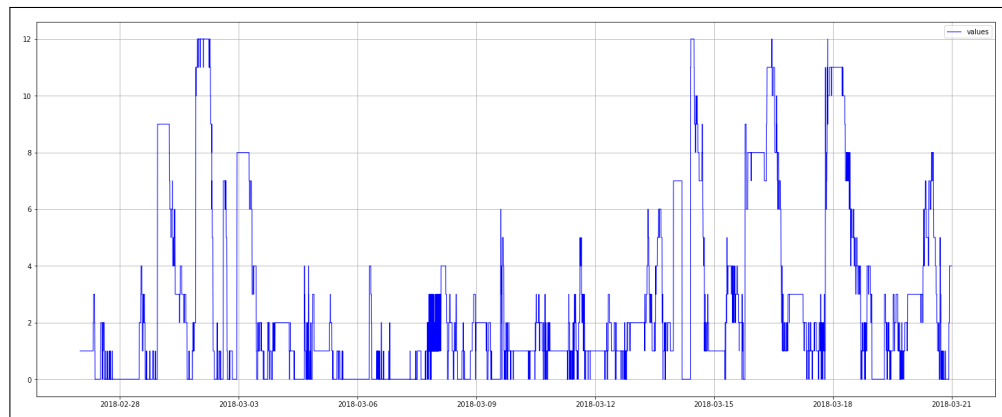
5.3 Análise

Inicialmente, foram carregados os dados dos arquivos csv para análise, obtidos após a limpeza dos dados. Com isso pode-se observar de uma forma mais clara a consistência dos dados e analisar os comportamentos dos mesmos.

A Figura 7 e a Tabela 2 mostram os resultados coletados a partir da estação 1 localizada na praça Luíza Távora. Na Figura 7, pode-se perceber a variação do número de bicicletas disponíveis bem como a grande atividade de locação e devolução de bicicletas no período de 26/02/2018 a 21/03/2018.

Na Tabela 2, tem-se um conjunto de dados como contador, média, desvio padrão,

Figura 7 – Bicicletas disponíveis na estação Luíza Távora



Fonte: Elaborado pelo autor

Tabela 2 – Dados gerais da estação Luíza Távora

	Bikes Disponíveis 1	Vagas
count	31.445	31.445
mean	2,492924	9,507076
std	3,028118	3,028118
min	0	0
max	12	12

Fonte: Elaborado pelo autor

mínimo, máximo de bicicletas disponíveis e vagas nas estações. O atributo *Bikes Disponíveis 1*, aponta a quantidade de bicicletas disponíveis em cada estação sendo esse atributo utilizado neste trabalho. Em contrapartida o atributo *Vagas* é um complemento de *Bikes Disponíveis 1* apontando a quantidade de vagas disponíveis para depositar bicicletas. O mesmo levantamento estatístico foi realizado para as demais estações, o que ajudou a entender melhor os dados.

5.4 Definição de *features* para uso

Após a análise dos dados, adotou-se uma estratégia de agrupamento dos registros (linhas do arquivo csv) em *slots* de tempo, para obter e salvar as *features* em cada *slot*. Foram formados *slots* de 1 hora, já que a proposta deste trabalho é a predição de bicicletas disponíveis para 1 hora depois da hora dada como entrada.

O agrupamento dos *slots* foi realizado percorrendo os arquivos de cada estação, onde cada *slot* reunia informações de bicicletas disponíveis como:

- Quantidade mínima e máxima;
- Média de bicicletas disponíveis;
- Desvio padrão;

- Dia da semana (0 a 6 iniciando da segunda-feira);
- Feriado:
 - 0 não é feriado.
 - 1 é feriado.
- Número do *slot*:
 - 1 hora de 0 a 23.

Tabela 3 – *Features* em slots de 1 hora

	min	max	media	desvio padrão	dia semana	feriado	slot	label
0	1	1	1.000000	0.000000	0	0	23	1.000000
1	1	1	1.000000	0.000000	1	0	0	1.000000
2	1	1	1.000000	0.000000	1	0	1	1.000000
3	1	1	1.000000	0.000000	1	0	2	1.000000
4	1	1	1.000000	0.000000	1	0	3	1.000000
5	1	1	1.000000	0.000000	1	0	4	1.000000
6	1	1	1.000000	0.000000	1	0	5	1.000000
7	1	1	1.000000	0.000000	1	0	6	1.450000
8	1	3	1.450000	0.693421	1	0	7	2.084746
9	0	3	2.084746	1.225273	1	0	8	0.000000
10	0	0	0.000000	0.000000	1	0	9	0.000000
11	0	0	0.000000	0.000000	1	0	10	0.000000
12	0	0	0.000000	0.000000	1	0	11	1.483333
13	0	2	1.483333	0.806053	1	0	12	1.169492
14	0	2	1.169492	0.667719	1	0	13	0.779661
15	0	2	0.779661	0.975423	1	0	14	0.016949
16	0	1	0.016949	0.129081	1	0	15	0.101695
17	0	1	0.101695	0.302247	1	0	16	0.283333
18	0	1	0.283333	0.450617	1	0	17	0.135593
19	0	1	0.135593	0.342356	1	0	18	0.033898
20	0	1	0.033898	0.180967	1	0	19	0.000000
21	0	0	0.000000	0.000000	1	0	20	0.000000
22	0	0	0.000000	0.000000	1	0	21	0.000000
23	0	0	0.000000	0.000000	1	0	22	0.000000

Fonte: Elaborado pelo autor

A *feature* de média de bicicletas disponíveis, principal alvo desse trabalho, foi calculada de acordo com a soma de bicicletas disponíveis em cada *slot* de hora e dividida pela o número de coletas naquela hora.

Em seguida, foi percorrido cada *slot* criando um campo de *label* e inserindo o valor da média de bicicletas disponíveis do próximo *slot*, já que o objetivo deste trabalho é a predição média de bicicletas. Esse passo se fez presente, pois esse campo foi usado como resposta no treino dos algoritmos. Na Tabela 3, podemos ver o resultado da formação das *features* em *slots* de 1 hora.

5.5 Criação dos modelos preditivos para cada estação

Com as *features* e com seus respectivos *labels* definidos, começa um dos passos mais importante desse trabalho, a criação dos modelos preditivos. Com a ajuda da biblioteca *scikit-learn*⁶ em conjunto com a ferramenta jupyter, os dados foram separados em porcentagens, sendo 75% para treino e 25% para teste com finalidade de efetuar as predições. Vale relembrar que as porcentagens definidas são para cada estação.

Logo após a separação iniciou-se a criação e treinamento dos modelos com os seguintes algoritmos de aprendizagem de máquina fornecidos pela biblioteca *scikit-learn*:

- *Linear Regression*;
- *SGD Regressor*;
- *Linear SVR*;
- *SVR*;
- *Random Forest Regressor*;
- *Gradient Boosting Regressor*;
- *Decision Tree Regressor*;
- *KNeighbors Regressor*.

Com isso, os dados de treino foram colocados para treinar em cada algoritmo citado acima para obter os modelos. Além disso, foi realizado o escalonamento dos dados usando a técnica de estandardização com o algoritmo *StandardScaler*, a fim de se ajustar os dados para que possam ter entradas sob a mesma escala para alguns algoritmos. O processo de estandardização padroniza os recursos de forma que a sua distribuição fique ajustada, (FIGUEIREDO; FIGUEIREDO, 2018).

5.6 Validação e comparação dos modelos preditivos

Neste passo, utilizou-se os modelos obtidos sem e com *StandardScaler* na Seção 5.5, na porção de dados escolhidas para teste, ou seja, os dados sem o campo de *label* como resposta. Logo depois, foi realizado a medição do resultados obtidos com cada modelo, utilizando as métricas RMSE e MAE.

A partir da medição obtida, foram geradas tabelas para exibição dos resultados obtidos de cada estação, como pode-se ver nas Tabelas 4 e 5.

⁶ <http://scikit-learn.org/stable/index.html>

Tabela 4 – Resultados das métricas sem *StandardScaler* na estação Luíza Távora

	rmse	mae
DTR	1.72281	0.936314
GBR	1.1874	0.737916
KNN	1.3542	0.788198
LR	1.18098	0.729775
LSVR	1.37942	0.826286
RFR	1.31364	0.790098
SGDR	1.26523	0.725296
SVR	2.05438	1.13168

Fonte: Elaborado pelo autor

Tabela 5 – Resultados das métricas com *StandardScaler* na estação Luíza Távora

	rmse_std	mae_std
DTR	1.52962	0.908354
GBR	1.15095	0.729281
KNN	1.36003	0.850357
LR	1.18098	0.729775
LSVR	1.34176	0.724223
RFR	1.31426	0.790086
SGDR	1.29846	0.770556
SVR	1.44783	0.811062

Fonte: Elaborado pelo autor

Avaliando e comparando os resultados de forma geral, no caso da estação Luíza Távora, pode-se perceber que os modelos que tiveram um melhor desempenho foram nos casos em que foi utilizado o *StandardScaler*. Além disso, pela Tabela 4 os algoritmos que tiveram melhores avaliações pela métrica RMSE foi o GBR e por MAE foi SGDR. Já a Tabela 5, nos mostra que o GBR também foi o melhor para a métrica RMSE e o LSVR foi o melhor para o MAE, o que nos permite dizer que nem sempre o modelo que melhor se aplica no caso usando *StandardScaler* se aplica ao caso onde não se faz uso do *StandardScaler*.

(LEGATES; MCCABE, 1999) dizem que para uma melhor interpretação dos valores RMSE e MAE, seja calculados a média e desvio padrão do *label*, que neste caso é a média de bicicletas disponíveis. A Tabela 6 apresenta esses valores. Além disso, (LEGATES; MCCABE, 1999) diz que um bom indicativo de que um modelo se ajustou bem aos dados, é quando suas taxas de erro forem menores que o desvio padrão dos valores.

A Tabela 7 nos mostra uma média ponderada dos resultados obtidos em todas estações e se comparados a desvio padrão e média da Tabela 6 os algoritmos ficaram com resultados abaixo, mostrando um bom desempenho.

A Figura 8 permite ver graficamente os melhores e piores algoritmos para a predi-

Tabela 6 – Resultados

Desvio Padrão	Media
2.140613	1.348163

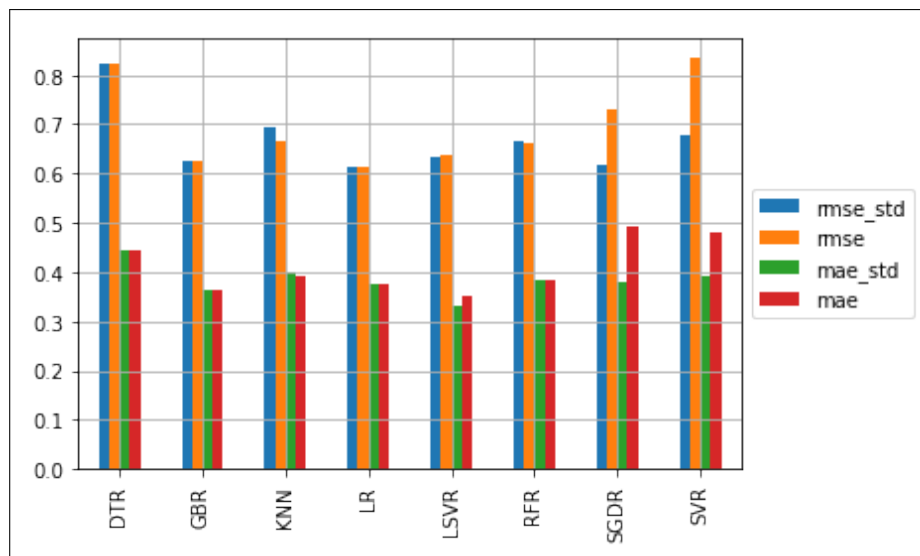
Fonte: Elaborado pelo autor

Tabela 7 – Media ponderada da avaliação do resultado das predições

	rmse_std	rmse	mae_std	mae
DTR	0.821334	0.823603	0.443152	0.442429
GBR	0.624853	0.625083	0.362060	0.362054
KNN	0.693530	0.666835	0.401219	0.392222
LR	0.611229	0.611229	0.375771	0.375771
LSVR	0.633817	0.638156	0.329740	0.349544
RFR	0.663638	0.662553	0.383213	0.382388
SGDR	0.617009	0.728285	0.379348	0.492513
SVR	0.676735	0.834076	0.392855	0.479793

Fonte: Elaborado pelo autor

Figura 8 – Gráfico desempenho dos algoritmos



Fonte: Elaborado pelo autor

ção de bicicletas pela perspectiva das métricas RMSE, MAE, RMSE (com *StandardScaler*) e MAE(com*StandardScaler*). O algoritmo LSVR foi o melhor avaliado, no geral, seguido pelo GBR pela métrica MAE com uso do *StandardScaler*, já o pior resultado foi do SVR na métrica RMSE.

6 CONCLUSÕES E TRABALHOS FUTUROS

Com o número de programas de bicicletas compartilhadas e de pessoas que os utilizam crescendo pelo mundo no últimos anos, a bicicleta se tornou um dos meios de locomoção mais utilizados em grande centros urbanos. Com grandes atrativos aos usuários, como facilidade de locação, rápido deslocamento em trechos congestionados e um meio de transporte não poluente, a locação de bicicletas tem grande potencial para estudos.

Neste trabalho, foram realizados estudos para comparação de técnicas de aprendizagem de máquina para a predição de bicicletas disponíveis no projeto Bicicletar Fortaleza. Para obter tais resultados, foi colocado um coletor de dados em execução para obter os dados do site do programa. Logo depois, foram realizados diversos tratamentos nos dados para que fossem utilizados na criação dos modelos preditivos e por fim comparar os melhores algoritmos para prever a disponibilidade de bicicletas.

Com o trabalho concluído, alcançam-se os objetivos propostos e adquirem-se informações valiosas para outros estudos e pesquisas na área de locação, não somente de bicicletas, mas de carros compartilhados. Apesar dos bons resultados obtidos, ainda faz-se necessário mais estudos para otimização das predições.

Contudo, fica como trabalho futuro um aprofundamento nos parâmetros fornecidos por cada algoritmo de predição, para obter resultados ainda melhores. Outro trabalho de grande valor, seria a criação de uma aplicação que forneça predições para os usuários, já que o site oficial disponibiliza apenas informações em tempo real. Este trabalho também pode se estender para o compartilhamento de outros meios de transportes. As predições alcançadas neste trabalho foram para predições adiante a 1 hora. Outros trabalhos poderiam explorar outros horizontes de tempo para predições.

REFERÊNCIAS

- AGUIAR, E. M. d. **Aplicação do Word2vec e do Gradiente descendente dstocástico em tradução automática**. Tese (Doutorado) — Faculdade Getúlio Vargas, Rio de Janeiro, 2016.
- ALVES, E. D. L.; VECCHIA, F. A. S. Análise de diferentes métodos de interpolação para a precipitação pluvial no estado de goiás. **Acta Scientiarum. Human and Social Sciences**, Universidade Estadual de Maringá, v. 33, n. 2, 2011.
- BASTOS, D. G.; NASCIMENTO, P. S.; LAURETTO, M. S. Proposta e análise de desempenho de dois métodos de seleção de características para random forests. **IX Simpósio Brasileiro de Sistemas de Informaç ao**, p. 49–60, 2013.
- BATISTA, G. E. d. A. P. *et al.* **Pré-processamento de dados em aprendizado de máquina supervisionado**. Tese (Doutorado) — Universidade de São Paulo, 2003.
- BICICLETAR. **Bicicletar - Bicycles compartilhadas de Fortaleza (Ng)**. 2018. Disponível em: <<http://www.bicicletar.com.br>>. Acesso em: 11 jun. 2018.
- FIGUEIREDO, A.; FIGUEIREDO, F. **Métodos Fatoriais de Análise de Dados e Big Data**. [s.l: s.n]. 2018.
- FISHMAN, E. Bikeshare: A review of recent literature. **Transport Reviews**, Taylor & Francis, v. 36, n. 1, p. 92–113, 2016.
- GALVÃO, N. D.; MARIN, H. de F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, Escola Paulista de Enfermagem, v. 22, n. 5, 2009.
- GAUTHIER COLIN HUGHES, C. K. S. L. C. L. S. L. J. M. C. P. C. R. B. S. e. X. T. A. **Guia de Compartilhamento de Bicycles Compartilhadas**. ITPD, 2014.
- LEGATES, D. R.; MCCABE, G. J. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. **Water resources research**, Wiley Online Library, v. 35, n. 1, p. 233–241, 1999.
- LI, Y.; ZHENG, Y.; ZHANG, H. *et al.* Traffic prediction in a bike-sharing system. In: ACM. **Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems**. [S.l.], 2015. p. 33.
- LOPES, F. C. **UMA FERRAMENTA PARA ANÁLISE DE DADOS DO PROGRAMA DE BICYCLES COMPARTILHADAS BICYCLETAR**. [s.n]. Quixadá, 2017.
- MONARD MARIA CAROLINA; BARANAUSKAS, J. A. **Conceitos sobre aprendizado de máquina: Sistemas inteligentes-fundamentos e aplicações**. v. 1, n. 1, 2003.
- NETO, B.; SOUSA, L. de *et al.* **Sistemas wearable baseados em métodos de visão computacional para auxiliar pessoas com deficiência visual**. [S.l: s.n], 2016.
- PARKES, S. D.; MARSDEN, G.; SHAHEEN, S. A.; COHEN, A. P. Understanding the diffusion of public bikesharing systems: evidence from europe and north america. **Journal of Transport Geography**, Elsevier, v. 31, p. 94–103, 2013.
- PESTANA, M. H.; GAGEIRO, J. N. **Análise de dados para ciências sociais: a complementaridade do spss**. Sílabo Lisboa, 2003.

PONTI, M. A.; COSTA, G. B. P. da. **Como funciona o Deep Learning**. [S.l: s.n], 2017.

SATO, L. Y.; SHIMABUKURO, Y. E.; KUPLICH, T. M.; GOMES, V. C. F. Análise comparativa de algoritmos de árvore de decisão do sistema weka para classificação do uso e cobertura da terra. **XVI Simpósio Brasileiro de Sensoriamento Remoto**, p. 2353–2360, 2013.

SHAHEEN, S.; ZHANG, H.; MARTIN, E.; GUZMAN, S. China's hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing. **Transportation Research Record: Journal of the Transportation Research Board**, Transportation Research Board of the National Academies, n. 2247, p. 33–41, 2011.

SHWARTZ-SHALEV, S.; BEN-DAVID, S. **Understanding Machine Learning: From theory to algorithms**. NY, USA: Cambridge University Press, 2014.

WANG, W. **Forecasting Bike Rental Demand Using New York Citi Bike Data**. Dublin Institute of Technology, 2016.

APÊNDICE A – EXEMPLO GEOJSON

Neste apêndice é apresentado um exemplo de retorno obtido com as consultas através do *script* utilizado na coleta dos dados no site do Bicicletar. O formato retornado é um GeoJSON, contendo dados como, quantidade de bicicletas disponíveis, vagas disponíveis, coordenadas das estações, *status*, nome da estação e endereço.

Código-fonte 1 – Exemplo GeoJSON

```
1 {
2   "type": "FeatureCollection",
3   "features": [{
4     "geometry": {
5       "type": "Point",
6       "coordinates": [-38.547983, -3.732094]
7     },
8     "type": "Feature",
9     "properties": {
10      "qtd_bikes_disp_1": "8",
11      "statusInterno": "Est_Normal 1",
12      "status_operacional": "E0",
13      "qtd_vagas_total": "4",
14      "nome": "Parque Arax&#225;",
15      "endereco": "Avenida Bezerra de Menezes, 334 /
16        Esquina Rua Ribeiro da Silva",
17      "qtd_bikes_disp_2": "8",
18      "estacao": "Parque Arax&#225;",
19      "id": 37,
20      "status_online": "A"
21    }
22  ]
}
```