

APRENDIZADO DE MÁQUINAS AUTOMATIZADO PARA PREDIÇÃO DE UMIDADE, TEMPERATURA E ILUMINAÇÃO EM AMBIENTES INTERNOS

AUTOMATED MACHINE LEARNING FOR HUMIDITY, TEMPERATURE AND LIGHTING PREDICTION IN INDOOR ENVIRONMENTS

Daniel Campos Lisboa
danielcamposlisboa@gmail.com

Natália Nakamura Barros
nataliabarros@facens.br

Johannes Von Lochter
johannes.locheter@facens.br

Centro Universitário Facens – Sorocaba, SP, Brasil

Submetido em: 11 mar. 2022. Aceito em: 22 jun. 2022

RESUMO

Diversos fatores ambientais afetam a saúde e bem-estar do ser humano, podendo ser prejudiciais quando fora dos valores ideais. Por meio da utilização de dispositivos relacionados à Internet das Coisas (*Internet of Things* - IoT), é possível medir esses fatores, possibilitando uma melhor análise da qualidade do ambiente interno. Com o Aprendizado de Máquinas (*Machine Learning* - ML), é possível analisar, de forma rápida e precisa, grandes quantidades de dados, permitindo criar modelos que podem prever valores futuros com base em valores passados registrados em um banco de dados. Este estudo busca utilizar o Aprendizado de Máquinas automatizado (*Automated Machine Learning* - *Auto ML*) para encontrar os modelos mais eficazes para a predição de umidade, temperatura e iluminação em ambientes internos. Os dados utilizados no experimento foram coletados de dois dispositivos sensíveis localizados em ambientes internos distintos de uma mesma instituição de ensino no Brasil. Foi feita uma comparação entre modelos preditivos baseados em hipóteses de aprendizagem diferentes, objetivando encontrar o modelo que se destacasse nos quesitos “acurácia” e “velocidade”. Os resultados mostraram que os modelos que utilizaram o algoritmo *Voting Ensemble*, juntamente com os algoritmos *XGBoostRegressor* e *LightGBM*, foram os mais eficazes, atingindo acurácia semelhante aos modelos que utilizaram o *Stack Ensemble*, porém em um tempo consideravelmente menor. Este estudo contribuirá para futuros pesquisadores que precisem analisar grandes quantidades de dados envolvendo alguma das variáveis do estudo, assim como auxiliar no processo de encontrar um bom modelo de ML para predição de outras variáveis através da utilização do *Auto ML*.

Palavras-chave: Aprendizado de máquinas. ML Automatizado. Qualidade do Ambiente Interno. Predição.

ABSTRACT

Some environmental factors affect human's health and well-being and can be harmful when its values are outside the ideal range. By using Internet of Things (IoT) devices, it is possible to measure indicators of these factors, enabling a better analysis of indoor environmental quality. With Machine Learning (ML), it is possible to analyze big data in a fast and accurate way, making it possible to create models that can predict future values based on past values recorded in a database. This study seeks to use Automated Machine Learning (Auto ML) to get the most effective machine learning predictive model for humidity, temperature and lighting in indoor environments. The data used for the experiment were collected from two Smart Citizen Kits,

located at two different indoor environments of an educational institution in Brazil. A comparison was made between different learning hypotheses and predictive models in order to find the most effective one, considering accuracy and speed. The results showed that the predictive models that used the Voting Ensemble algorithm along with XGBoostRegressor and LightGBM were the most effective ones, achieving an accuracy similar to the models that used Stack Ensemble, but in a considerably shorter time. This study will contribute to future researchers who need to analyze big data involving some of the variables analyzed in this study, as well as to help in the process of finding a good machine learning predictive model for other variables by using Auto ML.

Keywords: Machine learning. Auto ML. Indoor Environmental Quality. Prediction.

1. INTRODUÇÃO

Diversos fatores ambientais influenciam na qualidade de ambientes internos, impactando o nível de conforto e a saúde de quem frequenta o local (FRONTCZAK; WARGOCKI, 2011). Dentre alguns desses fatores estão a umidade, a temperatura e a iluminação do ambiente. Desse modo, a análise desses fatores se torna essencial para que se possa medir a qualidade de um projeto arquitetônico e garantir o bem-estar dos usuários de determinada edificação. Fatores como temperatura e iluminação são normalmente estimados a partir de simulações em softwares, como o EnergyPlus, DOE-2, eQUEST, IES, ECOTECT, entre outros. Embora essas simulações possam ser bastante efetivas na fase de projeto de novas edificações, elas perdem em acurácia na medida em a edificação já esteja em funcionamento, uma vez que surgem novos fatores difíceis de se simular à priori, tais como ocupação, propriedades térmicas dos materiais, sistemas de climatização e iluminação artificial, entre outros. Assim, o Aprendizado de Máquinas (*Machine Learning* – ML) se demonstra como um método mais eficaz para a predição de dados ambientais em edificações que já possuam o registro de dados ambientais passados (DEB, 2017).

Para auxiliar no registro desses dados foi desenvolvido o Smart Citizen Kit, uma tecnologia ligada à Internet das coisas (*Internet of Things* - IoT) criada em 2012 pelo Fab Lab Barcelona, localizado no *Institute for Advanced Architecture of Catalonia* (IAAC), o qual consiste de um aparelho sensível que capta ininterruptamente, no ambiente, informações de 8 fatores diferentes: luz, ruído, temperatura do ar, umidade relativa, pressão atmosférica, material particulado atmosférico (PM1, PM2.5 e PM10), concentração estimada de dióxido de carbono (CO₂) e compostos orgânicos voláteis (TVOC). O dispositivo tem como objetivo principal promover melhorias para a sociedade, a partir de observações e análises dos dados coletados e, por ser um hardware *open source* e de fácil compreensão, possibilita um maior acesso a dados de conforto e bem-estar das pessoas (IAAC, 2020).

Juntamente com a IoT, a Inteligência Artificial (IA) vem ganhando grande importância nos últimos anos pela sua eficiência no processamento e entendimento de dados (GILCHRIST, 2016). Adjacente a esses conceitos está o Machine Learning (ML), considerado como um subcampo da IA e responsável pelo aprendizado autônomo das máquinas em relação a grandes volumes de dados (GERON, 2017). Esse subcampo se sobressai pela sua capacidade de produzir decisões e resultados confiáveis a partir do estudo e construção de algoritmos, fazendo também previsões sobre dados.

Os algoritmos de ML operam construindo modelos a partir de *inputs* preliminares, com o objetivo de fazer previsões ou decisões guiadas pelas

amostras de dados, ao invés de simplesmente seguir instruções programadas. A fim de se obter bons modelos de ML para determinado conjunto de dados, algumas ações devem ser feitas, tais como o pré-processamento dos dados, a escolha de um algoritmo adequado e a otimização dos hiper-parâmetros dos algoritmos (FEURER et al, 2018). Esse pode ser um processo exaustivo que demanda diversos testes até se encontrar o modelo ideal. Uma forma de facilitar esse processo é por meio do Aprendizado de Máquinas Automatizada (*Automated Machine Learning* - Auto ML), um processo que automatiza estas tarefas, melhorando a produtividade dos cientistas de dados. Além disso, os métodos automatizados podem, até mesmo, produzir modelos superiores aos produzidos por seres humanos (DROZDAL et al, 2020).

Dentre os estudos já realizados, Pasupuleti et al (2020) utilizaram os algoritmos de ML: *Linear Regression*, *Decision Tree* e *Random Forest* para predição do índice de qualidade do ar e verificaram que o *Random Forest* obteve a melhor acurácia. Zhou et al (2019) utilizaram os algoritmos de ML: *Random Forest*, *XGBoost* e *K-NN* para estimar a temperatura do ar e verificaram que o *XGBoost* obteve a melhor acurácia. Os fatores ambientais analisados por Pasupuleti et al (2020) são diferentes dos analisados nesta pesquisa. Zhou et (2019) estimou a temperatura do ar, porém a estimou isoladamente. Em ambos os estudos, os autores utilizaram algoritmos de ML pré-estabelecidos para os testes, comparando apenas um número limitado de modelos e parâmetros.

O objetivo do presente artigo é encontrar o modelo de ML mais eficaz, considerando acurácia e tempo de processamento, para fins de predição de umidade, temperatura e iluminação em ambientes internos de edificações institucionais de ensino, por meio do auxílio do Auto ML.

2. MÉTODO

O método classifica-se, segundo sua finalidade, em desenvolvimento experimental, se consistindo em um trabalho sistemático com vistas à melhoria de sistemas e serviços. Classifica-se, ainda, segundo seus objetivos, em pesquisa exploratória, de modo a proporcionar maior familiaridade com o problema (GIL, 2002).

O delineamento da pesquisa seguirá as seguintes etapas, conforme demonstrado na Figura 1.

Figura 1 – Etapas do método utilizado



Fonte: elaborado pelos autores

2.1. Coleta de dados

A coleta dos dados ocorreu a partir de dois Smart Citizen Kits localizados em dois ambientes internos distintos de uma mesma instituição de ensino. O primeiro foi posicionado no L1, em uma mesa de trabalho, enquanto o segundo, no L2, em uma prateleira próxima à janela, conforme demonstrado na Figura 2. A atividade predominante do L1 é de trabalho, no qual possui computadores, mesas, cadeiras e ar-condicionado split. Já o L2 é um FabLab, que possui diversos equipamentos voltados à fabricação digital, como cortadora a laser, impressora 3D, Router CNC, entre outros. Estima-se uma média de 5 a 6 pessoas trabalhando concomitantemente em cada ambiente. Salienta-se que o período de análise corresponde ao período da pandemia da Covid-19 e, desse modo, os ambientes não estiveram em pleno uso durante todo o tempo.

Figura 2 – Localização dos Smart Citizen Kits



Fonte: Google Earth

O Smart Citizen Kit localizado no L1 obteve 453.757 linhas de dados entre novembro de 2019 e maio de 2021, não registrando dados entre junho e setembro de 2020 devido ao desligamento do aparelho. O Smart Citizen Kit localizado no L2 obteve 439.311 linhas de dados entre setembro de 2020 e agosto de 2021, não registrando dados entre dezembro de 2020 e janeiro de 2021 devido ao desligamento do aparelho. O desligamento dos aparelhos ocorreu devido à ausência de pessoas no local durante estes períodos, motivada pela pandemia da Covid-19.

2.2. Dados de entrada

Os dados obtidos pelos aparelhos foram exportados no formato csv e salvos em formato tabular com separadores. Em seguida, foi utilizada a biblioteca Pandas para importação do arquivo contendo os dados para a plataforma Google Colab, com uso da linguagem de programação Python, conforme demonstrado na Figura 3.

Figura 3 – Dados importados no Google Colab

	TEMPO	BATERIA	PRESSAO	UMIDADE	TEMPERATURA	ECO2	TVOC	LUZ	RUIDO	PM1	PM25	PM10
0	2019-11-25 14:09:00 UTC	100.0	95.00	57.30	23.40	NaN	NaN	693.0	56.58	5.0	8.0	8.0
1	2019-11-25 14:10:03 UTC	100.0	95.00	57.68	23.66	NaN	NaN	705.0	56.93	5.0	7.0	7.0
2	2019-11-25 14:11:05 UTC	100.0	95.00	56.74	23.79	NaN	NaN	694.0	59.29	6.0	9.0	10.0
3	2019-11-25 14:12:07 UTC	100.0	95.00	55.69	23.73	NaN	NaN	683.0	56.42	3.0	6.0	7.0
4	2019-11-25 14:13:09 UTC	100.0	95.00	55.78	23.48	2212.0	1165.0	707.0	56.24	6.0	10.0	10.0
...
453752	2021-05-02 14:18:58 UTC	99.0	95.63	56.09	24.62	671.0	41.0	68.0	49.88	12.0	18.0	20.0
453753	2021-05-02 14:19:58 UTC	99.0	95.63	56.07	24.62	674.0	41.0	68.0	48.09	15.0	23.0	24.0
453754	2021-05-02 14:20:58 UTC	99.0	95.63	56.09	24.62	671.0	41.0	68.0	48.03	13.0	17.0	17.0
453755	2020-03-31 13:50:44 UTC	NaN	NaN	NaN	NaN	466.0	NaN	NaN	NaN	NaN	NaN	NaN
453756	2020-03-31 13:51:44 UTC	NaN	NaN	NaN	NaN	463.0	NaN	NaN	NaN	NaN	NaN	16.0

453757 rows × 12 columns

Fonte: elaborado pelos autores

2.3. Limpeza e transformação

A limpeza e transformação dos dados visa a melhorar a qualidade dos mesmos, por meio da remoção dos valores nulos e *outliers*. Para auxiliar na análise dos mesmos, foram extraídas as informações de mês e hora da coluna Tempo, a qual indica a data em que o dado foi registrado.

Após uma verificação prévia da porcentagem que os valores nulos representavam em relação ao conjunto completo, os mesmos foram removidos. A Tabela 1 mostra a quantidade de linhas de dados contendo valores nulos em comparação com a quantidade total de linhas de dados. As linhas contendo dados nulos foram removidas.

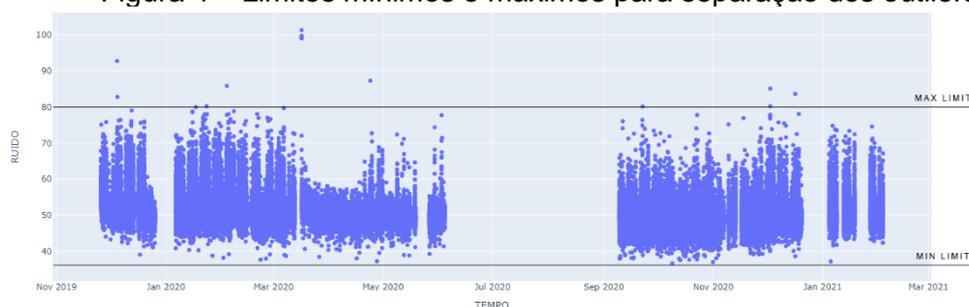
Tabela 1 – Quantidade de linhas contendo valores nulos

Conjunto de dados	Total de linhas	Linhas contendo dados nulos
L1	453.757	10.963 (2,41%)
L2	439.311	3.472 (0,79%)

Fonte: elaborado pelos autores

Utilizando a biblioteca Plotly Express, foi feita uma análise de padrões dos dados por meio de gráficos, como exemplificado na Figura 4. Com base nos padrões identificados, foram definidos valores mínimos e máximos para cada variável de modo a facilitar a identificação e remoção de valores *outliers*. Os valores mínimos e máximos definidos e a quantidade de outliers identificados se encontram na Tabela 2 e na Tabela 3.

Figura 4 – Limites mínimos e máximos para separação dos outliers



Fonte: elaborado pelos autores

Tabela 2 – Quantidade de outliers em L1

Variável	Valor mínimo	Valor máximo	Outliers
Umidade (%)	35	80	7
Temperatura (°C)	18	35	0
eCO2 (ppm)	400	2000	81
Luz (Lux)	0	1100	28
Ruído (dBA)	40	75	67
PM10 (ug/m ³)	0	100	1

Fonte: elaborado pelos autores

Tabela 3 – Quantidade de outliers em L2

Variável	Valor mínimo	Valor máximo	Outliers
Umidade (%)	25	80	58
Temperatura (°C)	12	43	8
eCO2 (ppm)	400	4500	10
Luz (Lux)	0	5000	120
Ruído (dBA)	35	80	239
PM10 (ug/m ³)	0	500	67

Fonte: elaborado pelos autores

2.4. Análise e exploração

A Figura 5 mostra uma matriz de correlação entre as variáveis de L1, feita utilizando as bibliotecas Seaborn e Matplotlib. A partir da análise da matriz, verificou-se que as variáveis Umidade e Temperatura tiveram pouca correlação com as demais variáveis, devido ao pouco contato com o ambiente externo que L1 possui e, ainda, por ser um ambiente de baixa circulação, com climatização artificial. A variável Luz, que representa a iluminação, obteve uma correlação considerável com as variáveis Hora e Ruído, em função de o uso da iluminação artificial ocorrer apenas nos horários de ocupação do ambiente.

Figura 5 – Matriz de correlações em L1



Fonte: elaborado pelos autores

A Figura 6 mostra a matriz de correlação entre as variáveis de L2. Como este ambiente sofre maior influência do ambiente externo, uma vez que o Smart Citizen está posicionado na janela que recebe incidência solar direta e de a porta do ambiente ficar frequentemente aberta, foram identificadas maiores correlações entre as variáveis. A variável Hora demonstrou grande correlação com as variáveis Umidade e Temperatura, diferente do percebido em L1, onde

havia o uso de climatização artificial. Além da correlação com Hora, foi identificada uma correlação negativa considerável entre Umidade e Temperatura, significando que, à medida que a temperatura sobe, o nível da umidade desce. A variável Luz, porém, teve baixa correlação direta com a variável Hora, tendo maior correlação com as variáveis Ruído, Umidade e Temperatura que, estas sim, estando correlacionadas com a variável Hora.

Figura 6 – Matriz de correlações em L2



Fonte: elaborado pelos autores

Em ambos os casos, em L1 e L2, as variáveis PM1, PM2.5 e PM10, que indicam a quantidade de partículas inaláveis no ambiente, tiveram alta correlação entre si, próximo a 1, já que se diferem apenas pelo diâmetro das mesmas. O mesmo aconteceu entre as variáveis eCO2 e TVOC, as quais indicam a quantidade de dióxido de carbono e compostos orgânicos voláteis, respectivamente. Para evitar variáveis redundantes, PM1, PM2.5 e TVOC foram descartadas para a próxima etapa. As demais variáveis foram mantidas para auxiliar no treino do modelo de Machine Learning, sendo estas: Umidade, Temperatura, eCO2, Luz, Ruído, PM10, Mês e Hora.

2.5. Criação dos modelos

Após a etapa de pré-processamento, os dados foram exportados para um arquivo CSV e importados para o Microsoft Azure Auto ML. Foi criada uma execução de Machine Learning automatizada para cada variável-alvo de L1 e L2, com duração de treino de 6 horas por execução cada.

2.6. Interpretação dos resultados

O Azure AutoML gerou uma lista dos algoritmos de ML que obtiveram a melhor acurácia, ou seja, cujo coeficiente de determinação (R^2) fosse mais próximo a 1. A próxima seção apresentará os algoritmos que apresentaram os valores mais elevados de R^2 , para cada variável-alvo de cada conjunto de dados.

3. RESULTADOS

A comparação dos resultados entre os modelos de ML utilizou o coeficiente de determinação, R^2 , o qual avalia a dispersão dos pontos de dados em uma linha de regressão. Valores maiores de R^2 representam diferenças

menores entre os dados reais e os dados previstos.

$R^2 = \text{Variação explicada pelo modelo} / \text{Variação total}$

Os modelos que atingiram maior acurácia para a predição de todas as variáveis foram o *Stack Ensemble* e o *Voting Ensemble*, tendo uma diferença média de 0,0016 de R^2 . Por este motivo, embora os modelos que utilizaram o *Stack Ensemble* tenham atingido uma acurácia maior, os modelos considerados mais eficazes foram os que utilizaram o *Voting Ensemble*, uma vez que estes atingiram acurácia semelhante, porém em um tempo consideravelmente menor, como visto na Tabela 4 e na Tabela 5.

Stack Ensemble é um algoritmo de ML que aprende como combinar melhor a predição de diversos outros algoritmos. Seu principal benefício é o de aproveitar as capacidades de diferentes algoritmos para criar um modelo que supere todos aqueles pertencentes ao *ensemble* (BROWNIEE, 2020). *Voting Ensemble* é um algoritmo de ML que combina predições de diversos outros algoritmos com o objetivo de aumentar o desempenho do modelo, podendo ser usado para regressão ou classificação. No caso da regressão, a predição final é resultado da média da predição dos algoritmos pertencentes ao *ensemble* (BROWNIEE, 2020).

Tabela 4 – R^2 e tempo de processamento dos modelos para as variáveis de L1

Algoritmo	Temperatura		Umidade		Luz	
	R^2	Tempo	R^2	Tempo	R^2	Tempo
StackEnsemble	0,994	24m 32s	0,989	24m 05s	0,997	23m 32s
VotingEnsemble	0,994	01m 32s	0,988	02m 02s	0,996	01m 33s
LightGBM	0,938	00m 18s	0,882	07m 25s	0,960	00m 18s
XGBoostRegressor	0,994	07m 13s	0,988	01m 18s	0,997	06m 56s
ExtremeRandomTrees	0,796	01m 04s	0,652	01m 17s	0,836	01m 17s
DecisionTree	0,578	01m 14s	0,359	01m 18s	0,701	01m 36s

Fonte: elaborado pelos autores

Tabela 5 – R^2 e tempo de processamento dos modelos para as variáveis de L2

Algoritmo	Temperatura		Umidade		Luz	
	R^2	Tempo	R^2	Tempo	R^2	Tempo
StackEnsemble	0,997	22m 44s	0,997	22m 52s	0,808	03m 11s
VotingEnsemble	0,996	01m 21s	0,996	01m 21s	0,802	01m 20s
LightGBM	0,960	00m 18s	0,951	00m 30s	0,678	00m 18s
XGBoostRegressor	0,997	06m 41s	0,996	06m 44s	0,802	00m 55s
ExtremeRandomTrees	0,836	02m 00s	0,792	01m 24s	0,137	01m 52s
DecisionTree	0,701	01m 07s	0,837	01m 57s	0,089	01m 24s

Fonte: elaborado pelos autores

Os algoritmos utilizados nos modelos *Voting Ensemble* foram o *XGBoostRegressor* e o *LightGBM*, os quais atingiram as terceira e quarta maiores acurácias, respectivamente. O *XGBoostRegressor* consiste em uma implementação do algoritmo *Gradient Tree Boosting*, sendo um método de aprendizagem supervisionado que tenta, de forma consistente, prever determinada variável alvo por meio da combinação de predições de modelos de aprendizagem mais simples (AMAZON, 2021). Já o *LightGBM* consiste em um framework de *Gradient Boosting*, o qual utiliza algoritmos de aprendizagem baseados em *Trees* (LIGHT GBM, 2021).

Em ambos os algoritmos, o Auto ML efetuou variações nos parâmetros utilizados, de modo a gerar modelos distintos que melhor se adequassem a cada variável alvo, como visto no Quadro 1.

Quadro 1 – Modelos utilizados nos Voting Ensemble

Variável alvo	Modelos utilizados
Temperatura L1	XGBR v1, XGBR v2, XGBR v3
Umidade L1	XGBR v1, XGBR v2, XGBR v3
Luz L1	XGBR v1, XGBR v2, XGBR v3
Temperatura L2	XGBR v1, XGBR v2, XGBR v3, LGBM v1, LGBM v2
Umidade L2	XGBR v1, XGBR v2, XGBR v3, XGBR v4, LGBM v2
Luz L2	XGBR v1

Fonte: elaborado pelos autores

Os parâmetros utilizados nos algoritmos XGBoostRegressor e LightGBM podem ser vistos nos Quadro 2 e Quadro 3, respectivamente.

Quadro 2 – Parâmetros dos modelos XGBoostRegressor

Parâmetro	XGBR v1	XGBR v2	XGBR v3	XGBR v4
Transformação de dados	Standard Scaler	SparseNormalizer	Standard Scaler	StandardScaler
booster	gbtree	gbtree	gbtree	gbtree
eta	0.5	0.5	0.001	0.001
gamma	0	0.1	0	0.01
max_depth	9	10	9	0
max_leaves	0	255	63	127
n_estimators	100	600	100	100
objective	reg:linear	reg:linear	reg:linear	reg:linear
reg_alpha	0	0.7291	0	1.4583
reg_lambda	2.3958	1.3541	0.625	0.7
subsample	0.9	0.5	0.7	
tree_method	auto	auto	auto	hist
colsample_bytree				0.7
max_bin				1023
grow_policy				lossguide

Fonte: elaborado pelos autores

Quadro 3 – Parâmetros dos modelos LightGBM

Parâmetro	LGBM v1	LGBM v2
Transformação de dados	SparseNormalizer	MaxAbsScaler
reg_alpha	0.075	
reg_lambda	0.975	
subsample	0.95	
boosting_type	Gbdt	
colsample_bytree	0.7	

learning_rate	0.1157	
max_bin	255	
max_depth	8	
min_data_in _leaf	0.0027	20
min_split_gain	0.5789	
n_estimators	200	
num_leaves	31	
subsample_freq	7	

Fonte: elaborado pelos autores

A apresentação dos parâmetros utilizados pelo Auto ML permitirá que o algoritmo possa ser executado e validado em outros testes, bem como permitirá a sua eventual implementação em softwares de simulação, de modo a fornecer maior precisão à tomada de decisão do projetista.

4. CONCLUSÃO

O uso de aprendizado de máquinas pode fornecer análises ambientais de ambientes internos com maior precisão, auxiliando na melhoria da qualidade no projeto arquitetônico e fornecendo maior bem-estar para os usuários da edificação. Da análise das variáveis ambientais, temperatura do ar, umidade e luz, realizada neste estudo, infere-se que os modelos que utilizaram o algoritmo *StackEnsemble* com o *XGBoostRegressor* e *LightGBM* atingiram maior acurácia porém, os modelos que utilizaram o algoritmo *VotingEnsemble* com o *XGBoostRegressor* e *LightGBM* se mostraram mais eficazes, visto que atingiram acurácia similar porém com menos tempo de processamento.

Este trabalho demonstrou que o algoritmo *VotingEnsemble* com o *XGBoostRegressor* e o *LightGBM* possui maior potencial para predição de umidade, temperatura e iluminação em ambientes internos de instituições de ensino. Além disso, entende-se que o presente estudo contribuirá para futuros pesquisadores que precisem analisar *big data* contendo alguma das variáveis do estudo, bem como servir de auxílio no processo de busca por um bom modelo de *Machine Learning* para predição de outras variáveis ambientais.

As limitações do estudo são: coleta de dados limitada a apenas um *Smart Citizen Kit* por ambiente analisado, assim como a ausência de dados em ambos os dispositivos durante alguns meses causado pelo desligamento dos *Smart Citizen Kits*. Vale ressaltar ainda que o período de coleta dos dados ocorreu majoritariamente durante a pandemia da Covid-19 e, portanto, poucas pessoas frequentaram os espaços durante este período.

Pesquisas futuras poderão utilizar o *VotingEnsemble* com *XGBoostRegressor* e *LightGBM* em edificações com características, usos e localização distintas.

5. AGRADECIMENTOS

Agradecemos o apoio financeiro concedido pela OMITIR durante o período da pesquisa.

REFERÊNCIAS

Amazon SageMaker: developer guide. **Amazon**, 2021. Disponível em: docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html. Acesso em: 24 nov. 2021.

BROWNIEE, J. How to Develop Voting Ensembles With Python. **Machine learning mastery**, 2020. Disponível em: machinelearningmastery.com/voting-ensembles-with-python. Acesso em: 15 out. 2021.

BROWNIEE, J. Stacking Ensembles Machine Learning With Python. **Machine learning mastery**, 2020. Disponível em: machinelearningmastery.com/stacking-ensemble-machine-learning-with-python. Acesso em: 15 out. 2021.

DEB, C. et al. A review on time series forecasting techniques for Building energy consumption. **Renewable and Sustainable Energy Reviews**, 74, p.902-924, fev. 2017.

DROZDAL, J. *et al.* Trust in AutoML: Exploring Information Needs for Establishing Trust in Automated Machine Learning Systems. In: 25th International Conference on Intelligent User Interfaces (IUI), 2020, Cagliari Italia. **Anais [...]**. Nova York: Association for Computing, 2020.

FEURER, M. et al. Practical Automated Machine Learning for the AutoML Challenge 2018. **Informatik**, 2018. Disponível em: ml.informatik.uni-freiburg.de/wp-content/uploads/papers/18-AUTOML-AutoChallenge.pdf. Acesso em: 24 fev. 2022.

FRONTCZAK, M; WARGOCKI, P. Literature survey on how different factors influence human comfort in indoor environments. **Building and Environment**, 46, p. 922-937, 2011.

GERON, A. **Hands-On Machine Learning with Scikit-Learn and TensorFlow**. Estados Unidos: Nicole Tache, 2017. 718 p.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2009. 176 p.

GILCHRIST, A. **Industry 4.0: The Industrial Internet of Things**. Tailândia: Apress, 2016. 250 p.

INSTITUTE FOR ADVANCED ARCHITECTURE OF CATALONIA (IAAC). **IAAC**, 2020. Disponível em: iaac.net. Acesso em: 01 fev. 2021.

LightGBM Documentation. **LightGBM**, 2021. Disponível em: lightgbm.readthedocs.io/en/latest/#:~:text=LightGBM%20is%20a%20gradient%20boosting,efficient%20with%20the%20following%20advantages%3A&text=Support%20of%20parallel%2C%20distributed%2C%20and,of%20handling%20large%2Dscale%20data. Acesso em: 24 nov. 2021.

PASUPULETI, V. et al. Air Quality Prediction Of Data Log By Machine Learning. In: International Conference on Advanced Computing and Communication Systems (ICACCS), 6, 2020, Coimbatore, India. **Anais [...]**. Nova Jersey: IEEE, 2020, p. 1395-1399.

ZHOU, K. et al. Estimation of Air Temperature from FY-4A AGRI Data: A Comparison of Different Machine Learning Algorithm. In: International Conference on Meteorology Observations (ICMO), 2019, Chengdu, China. **Anais [...]**. Nova Jersey: IEEE, 2019, p. 1-4.