Avaliação da Qualidade Ambiental de um Ambiente de Smart Home Utilizando Internet of Things e Inteligência Artificial

Laciene Alves Melo¹, Felipe Reis¹, Leonardo Nunes Gonçalves¹, Haelton Carvalho¹, Jasmine Priscyla Leite de Araujo¹, Fabrício Brito¹, Fabrício Farias¹

¹Universidade Federal do Pará (UFPA) – Cametá – PA – Brazil

lacienealvesmelo@gmail.com, felipe.vitow@gmail.com,
leo.widgeon16@gmail.com, haeltoncarvalho@gmail.com,
jasmine.araujo2018@gmail.com, fjbbrito@gmail.com, fabriciosf@ufpa.br

Abstract. This article aims to provide a solution based on artificial intelligence (AI) for smart home classification by considering the monitoring of environmental parameters. In order to colect the data, an Internet of Things (IoT) network was proposed for real-time monitoring of a bedroom and living room of a certain residence. After the data acquisition, they were presented to the steps of selection, pre-process, transformation and, finally, to the classification through AI algorithm based on decision tree. Results indicate that is possible to use IoT collected date to for monitoring the environment quality through AI.

Resumo. Este artigo tem como propósito construir uma solução baseada em inteligência artificial (IA) para classificação de um ambiente de smart home, considerando o monitoramento de parâmetros ambientais. Para isso, foi proposto a implantação de uma rede de Internet of Things (IoT) para monitoramento em tempo real de um quarto e sala de estar de determinada residência. Após a coleta dos dados, estes passaram pelas etapas de seleção, pré-processamento, transformação e apresentação ao algoritmo de IA baseado em árvore de decisão. Após os experimentos, os resultados indicam que é possível usar dados coletados de uma rede IoT para monitorar a qualidade de determinado ambiente por meio da aplicação de IA.

1. Introdução

A Internet of Things (IoT) é uma área de pesquisa que tem por finalidade interconectar dispositivos ou objetos utilizados no dia a dia, através da Internet, e assim influenciando na qualidade de vida e decisões de uma sociedade (Zhamanov, 2017). Atualmente, a IoT já se popularizou mundialmente, sendo uma realidade em diversos campos de aplicação, por exemplo, em cenários de *smart* home existem diferentes problemas a serem resolvidos que quando solucionados podem melhorar o dia a dia da população.

Em relação as soluções já apresentadas na literatura, os autores Elçi, Yalçin e Ünaldi (2021) apresentaram uma proposta de *smart home* visando solucionar o problema da acessibilidade para deficientes físicos, idosos e doentes. O sistema incluiu o uso de sensores de ventilação, iluminação, gás e controle de entrada e saída. Para amenizar certas restrições para seus dependentes, abrangeu-se componentes como comando de voz, permitindo o autogerenciamento de todos esses recursos através da

rede IoT. Os resultados obtidos no trabalho indicaram o sucesso da aplicação no que diz respeito a melhoria da qualidade de vida dos sujeitos beneficiados pelo projeto.

Já Dash e Choudekar (2021) desenvolveram um sistema autômato residencial, que inclui vigilância inteligente, medidor de energia elétrica e uma rede IoT configurada para controlar os dispositivos. Os resultados indicaram que é possível utilizar redes IoT como solução para o monitoramento e segurança de um ambiente residencial.

Por fim, Wang, Fu e Du (2021) realizaram um estudo de caso para identificar anomalias em *smart home*, para isso propuseram o uso de árvore de decisão para controlar as operações da rede IoT e seus respectivos dispositivos inteligentes. Neste trabalho os dados foram coletados em ecossistemas reais, o que demonstrou a eficácia da abordagem proposta. Além disso, os autores provaram a eficácia do uso de Inteligência Artificial (IA) como parte da solução em cenários IoT.

Embora existam diversas soluções já presentes ao cenário de *smart home*, ainda existem lacunas a serem preenchidas com novas soluções. Dentre as lacunas, destaca-se o desafio de monitoramento do controle da qualidade ambiental presente em cenários *smart home*. Como solução, surgem alternativas da produção de projetos que permitam a coleta, processamento e tomada de decisões inteligentes a partir do grande volume de dados produzidos por sensores instalados em tais ambientes.

De acordo com Ferreira e Araújo (2018), a infraestrutura tecnológica precisa ser capaz de atender as necessidades oriundas do contexto em que está inserida, por exemplo, para o monitoramento ambiental há como solução a implantação de uma arquitetura IoT composta por diferentes sensores ambientais, como os de temperatura e umidade, conectados aos nós do tipo *endnode*, sendo estes nós conectados ao *gateway* para transmissão dos dados coletados em tempo real.

Além disso, também é necessário fazer a devida implantação de uma infraestrutura em nuvem para recebimento e armazenamento dos dados coletados pelos sensores distribuídos na residência, sendo comumente adotado em projetos brasileiros o *middleware open source* da Dojot para gerenciamento do tráfego e armazenamento das informações na nuvem. Por fim, se faz necessário planejar uma solução baseada em IA com o intuito de recuperar e analisar os dados armazenados para posterior extração de padrões úteis aos usuários.

Em relação a aplicação de IA, isso pode ser feito na forma de aprendizagem supervisionada, neste caso a solução requer a utilização de algoritmos que tenham uma ampla taxa de predição, dentre eles, destaca-se o algoritmo árvore de decisão [Rajeshkanna e Arunesh, 2020], sendo a árvore de decisão um tipo de solução que facilita a análise de dados em modelos não-lineares. Segundo Jalal e Ezzedine (2020) a árvore de decisão é um algoritmo de aprendizagem supervisionada e eficaz na identificação de anomalias, devido a sua simplicidade, rápida adaptação e seu poder de classificação, que por sua vez, permite a criação de um modelo que prevê a saída de uma variável baseado nas variáveis de entrada.

Diante deste contexto, este trabalho visa apresentar um protótipo de solução de *smart home* composto por uma rede IoT e arquitetura para coleta, armazenamento e processamento de dados, com o intuito de monitorar os parâmetros ambientais presentes em um cenário residencial. Além disso, este trabalho também tem como objetivo aplicar os dados coletados a um algoritmo de IA baseado em árvore de decisão com o intuito de

classificar o ambiente como agradável ou desagradável, isto é, em relação a qualidade do ambiente para a permanência de pessoas no recinto em discussão. A partir deste trabalho, espera-se determinar se é possível classificar a qualidade de um ambiente de *smart home* a partir do uso de dados reais coletados pela rede IoT e aplicação de IA.

Além desta seção introdutória, a Seção 2 apresenta os métodos e procedimentos adotados. A seção 3 apresenta os resultados desta pesquisa. Por fim, as considerações finais são apresentadas na Seção 4.

2. Métodos e Procedimentos

Esta seção apresenta a implantação de uma rede IoT para coleta de dados em um cenário *smart home*. Deste modo, serão apresentadas a arquitetura do sistema, a forma de aquisição de dados, as métricas de análise, e por fim, as etapas de pré-processamento de dados e classificação utilizando o algoritmo de IA.

2.1. Arquitetura do Sistema

Para este estudo foram utilizados dois endnodes e um gateway. Os endnodes foram instalados em ambiente residencial contendo os sensores de temperatura, umidade, ultrassônico, chama, luminosidade e som. Após a instalação, foram coletados dados em tempo real, os quais foram enviados por um payload transmitido via protocolo LoRa e estabelecendo a comunicação entre os endnodes e gateways [Pinto et al., 2020]. Ao chegar no gateway, o payload foi interpretado e organizado para um segundo envio a partir do gateway até os servidores instalados na nuvem, sendo neste caso utilizado o protocolo Message Queuing Telemetry Transport (MQTT) [Carvalho, 2021].

Em relação aos servidores, os dados foram direcionados para um instalado na Amazon e outro instalado na Universidade Federal do Pará (UFPA), isto é, garantindo a redundância dos dados em caso de eventuais falhas. Ao chegar nos servidores, os dados foram lidos a partir da porta TCP 1883 e publicados no ecossistema Dojot [Silva, 2021]. A Dojot, por sua vez, é responsável em permitir o gerenciamento das informações e consulta dos dados guardados, assim, garantindo que possam ser recuperados posteriormente para a aplicação dos algoritmos de IA. A Figura 1 ilustra a arquitetura geral da rede IoT implantada para sensoriamento.

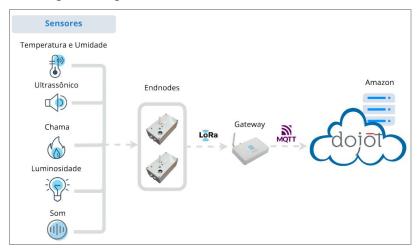


Figura 1. Arquitetura de execução do projeto de Smart Home implantado.

2.2. Manipulação da Base de Dados

2.2.1. Aquisição de Dados

Para a coleta dos dados utilizados neste trabalho, os equipamentos foram instalados em ambiente residencial, sendo um *endnode* instalado na sala de estar e outro no quarto. O período de monitoramento para coletas foi determinado por 12 horas, tendo início às 18h e encerrando às 6h do próximo dia. Após a coleta e armazenamento dos dados na Dojot, estes foram apresentados para a etapa de processamento de dados.

2.2.2. Processamento de Dados

Durante a etapa de processamento de dados foi observada a presença dos seguintes atributos: dispositivo, dia, chama, som, device_id, ts, metadata, distancia, luminosidade, rssi, temperatura e umidade. Nesta etapa, foi adotado o uso de uma técnica de IA, do tipo de classificação não supervisionada, com objetivo de agrupar os dados coletados pelos sensores em duas ou três possíveis classes de saída. Durante a definição das classes, foram levadas em consideração as possibilidades de ambiente agradável e ambiente desagradável para o agrupamento em duas classes, e de ambiente agradável, ambiente desagradável, e ambiente nem agradável e nem desagradável para o agrupamento em três classes. Para validação e seleção da melhor opção de agrupamento, o especialista do ambiente foi consultado para validar, sendo selecionada como saída, o agrupamento que retornou duas classes, isto é 1 – ambiente agradável ou 2 – ambiente desagradável.

Para a realização do agrupamento, foi aplicada o algoritmo de IA denominado Kohonen [Kohonen, 1990], que por sua vez, é responsável em organizar os registros com características similares. Para cada agrupamento, o algoritmo foi executado por mil épocas. Ao final dessa etapa, os dados foram apresentados para a etapa de exploração de dados.

2.2.3. Exploração de Dados

Na exploração de dados, a base de dados foi exportada para um formato inteligível para a linguagem de programação Python. Em seguida, foi utilizado o método de exploração de dados, a fim de se separar as informações úteis das irrelevantes para o processo de análise das condições ambientais da residência.

A Figura 2 ilustra a base de dados após a etapa de exploração de dados, onde em vermelho são apresentados os atributos considerados irrelevantes e em azul são apresentados os atributos relevantes para serem utilizados pelo algoritmo de IA para classificação.

2.2.4. Eliminação Manual de Atributos

Após a identificação dos dados mais relevantes, foi realizada a aplicação do método de Eliminação Manual de Atributos, onde foram eliminados o conjunto de atributos que não apresentavam contribuição para a estimativa e treinamento das classes de saída.

[dispositi	vo Di	ia	chama	Som	device_id	t	metadata	distancia	luminosidade	rssi	temperatura	umidade	Classe com 3 Saidas	Classe com 2 Saidas
0		1	1	0.0	NaN	fbf5d3	2021-07-30T08:47:58,503000	Z {}	10.0	950	-69	29.0	78.0	3	2
1		1	1	0.0	NaN	fbf5d3	2021-07-30T08:47:40,995000	Z {}	10.0	950	-49	29.0	78.0	2	1
2		1	1	0.0	NaN	fbf5d3	2021-07-30T08:47:34,536000	Z {}	10.0	950	-69	29.0	78.0	3	2
3		1	1	0.0	NaN	fbf5d3	2021-07-30T08:47:22,981000	Z {}	9.0	950	-68	29.0	78.0	3	2
4		1	1	0.0	NaN	fbf5d3	2021-07-30T08:47:11,207000	Z {}	10.0	950	-68	29.0	78.0	3	2
5652		2	1	NaN	572.0	e23155	2021-07-29T22:45:05,956000	Z {}	NaN	205	-51	28.2	84.6	2	2
5653		2	1	NaN	535.0	e23155	2021-07-29T22:44:59,577000	Z {}	NaN	205	-50	28.2	84.6	2	2
5654		2	1	NaN	542.0	e23155	2021-07-29T22:44:47,547000	Z {}	NaN	206	-51	28.2	84.3	2	2
5655		2	1	NaN	527.0	e23155	2021-07-29T22:44:36,067000	Z {}	NaN	204	-50	28.2	84.4	2	2
5656		2	1	NaN	569.0	e23155	2021-07-29T22:44:24,007000	Z {}	NaN	204	-51	28.2	84.4	2	2
5657 ro	ws × 14 colu	mns													

Figura 2. Exploração de dados coletados para o experimento.

A Figura 3 ilustra o processo de eliminação e verificação da nova base de dados, onde a Figura 3 (a) apresenta a função *drop*, que por sua vez, recebe como parâmetros um *array* com o nome das colunas que serão eliminadas da base de dados, e o *axis* para indicação da eliminação da linha que contém o nome de cada atributo. Já a Figura 3 (b) exibe os atributos selecionados como mais relevantes para análise, assim como, as respectivas classes de saída associadas a cada registro. Por fim, a Figura 3 (c) ilustra o resultado obtido a partir da função *sum*, que por sua vez, apresenta os atributos selecionados e suas respectivas quantidades de registros vazios.

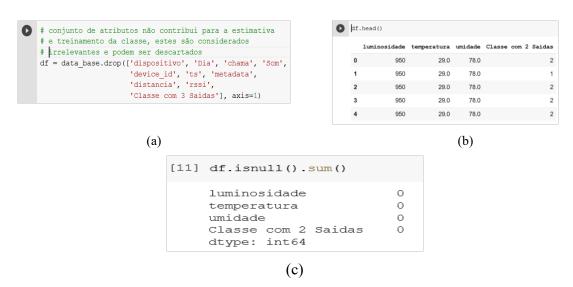


Figura 3. Conjunto de atributos eliminados. (a) Execução da função *drop*. (b) Dados após a Avaliação. (c) Contagem de Registros Vazios.

Após as etapas realizadas, foram selecionados como atributos de entrada os dados dos sensores de *luminosidade*, *temperatura* e *umidade*. Em seguida, foi aplicada a função *pairplot* do Python para a determinação da correlação entre os atributos selecionados e classes de saída. A Figura 4, ilustra a correlação entre os atributos selecionados em função das classes de saída. A partir deste resultado, é possível observar que a tomada de decisão em relação aos critérios de níveis ambientais pode ser determinada a partir dos atributos selecionados nas etapas de exploração e eliminação de dados. Além disso, é possível observar a separação entre as classes de saída e atributos de entrada, isto é, a impureza dos dados.

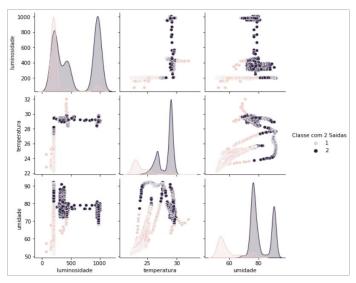


Figura 4. Correlação dos atributos selecionados em relação as classes de saída.

2.2.5. IA para Classificação dos Dados

O processo de classificação utilizando IA, é definido pela classificação do item de uma classe, portanto, a aprendizagem pode ser usada para mapeamentos entre classes discretas definidas previamente. Para a etapa de classificação, foi utilizado o algoritmo de IA de Árvore de decisão.

A Árvore de decisão é um método aplicado em aprendizagem de máquina, no qual, é possível treinar o modelo, utilizando as classes esperadas, considerando que estejam parcialmente ou totalmente classificadas. Para este caso considerou-se o cálculo de entropia para quantificar o ganho de cada atributo em relação a entropia geral do sistema, assim, computando a posição de cada atributo em relação a topologia da árvore de decisão resultante. A Equação (1) apresente o cálculo da entropia do sistema:

$$E(S) = \sum_{i=1}^{C} -P_i \log_2 P_i \tag{1}$$

onde S representa o conjunto de elementos presentes na base de dados; i=1 ... C diz respeito aos registros que serão apresentados para a aprendizagem da árvore; P_i representa a quantidade de elementos de uma classe cuja decisão é i, dividido pelo total elementos da base de dados.

2.2.6. Treinamento e Teste utilizando o Algoritmo de Árvore de Decisão

Para a obtenção dos melhores resultados, utilizou-se a função *train_test_split* para separar as amostras. A função recebeu como parâmetros os valores de *x*, *y* e *teste_size*, sendo o último parâmetro determinado com valor igual a 0.20.

O parâmetro *x* representa os atributos de luminosidade, temperatura, umidade, já *y* representa as classes de saída, por fim *test_size* indica a divisão da base de dados, que para este experimento foi determinada em 20% (1.132) dos registros para teste, e 80% (4.525) dos registros para treinamento do algoritmo, conforme ilustrado na Figura 5.

```
# dividindo os dados em treino e teste
# obs.: 20% para teste e 80% para treino
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20)
```

Figura 5. Determinação dos parâmetros para separação das amostras de treino e teste.

Por fim, foi utilizada a função *DecisionTreeClassifier* para determinação da melhor topologia da Árvore de decisão, tendo a função a definição dos seguintes parâmetros: *max_depth* para determinação da profundidade máxima de 5 nós, e a métrica de *criterion* com argumento "entropy" para determinação do método de cálculo do algoritmo. Para ilustrar, a Figura 6 apresenta a configuração dos parâmetros via a função.

```
[119] # instância o classificador de árvore
   model = tree.DecisionTreeClassifier(max_depth=5, criterion='entropy')
```

Figura 6. Parâmetros de definição da árvore.

3. Resultados

Durante o processamento das amostras destinadas para teste, o algoritmo de árvore de decisão rotulou os atributos para esse procedimento, sendo luminosidade definida como X[0], temperatura como X[1] e umidade como X[2]. É válido reforçar que na aprendizagem supervisionada as classes de saída foram definidas como 1 – Ambiente Agradável e 2 – Ambiente Desagradável. Por meio dos atributos, a árvore foi criada e retornou X[2] como nó raiz e os demais como nós filhos. A Figura 7 ilustra o resultado final da árvore de decisão obtida.

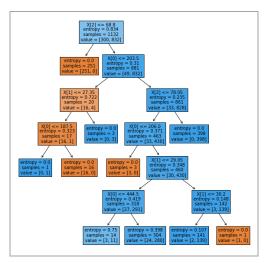


Figura 7. Árvore de decisão de variável categórica.

O resultado obtido indica o sucesso da aplicação de IA com o algoritmo de árvore de decisão para a determinação do nível de qualidade do ambiente residencial, o que pode ser útil para a tomada de decisão do usuário ou de equipamentos instalados na residência.

Em termos de desempenho do algoritmo, a Figura 8 (a) ilustra os resultados obtidos na etapa de teste, enquanto que a Figura 8 (b) ilustra os resultados obtidos na etapa de treinamento. Para ambos os casos, foi obtida uma precisão superior a 95%, sendo um resultado satisfatório para uma tomada de decisão em relação a qualidade do ambiente. Detalhadamente, na etapa de testes, isto é para os 1132 registros, houve uma precisão total de 96%, sendo uma precisão de 99% de precisão em relação a classe 1 (agradável), e 95% para a classe 2 (desagradável). Já na etapa de treinamento, isto é para os 4525 registros, foi obtida uma precisão total de 97%, sendo 99% e 97% para, respectivamente, classe 1 e classe 2.

	Precisão	Amostra
Agradável	99%	295
Desagradável	95%	837
Total	96%	1132

(a)

	Precisão	Amostra
Agradável	99%	1206
Desagradável	97%	3319
Total	98%	4525

(b)

Figura 8. Comparação de Amostras. (a) Relatório de Teste. (b) Relatório de Treino.

Logo após, para fins de reforço dos resultados de precisão obtidos no modelo de teste, utilizou-se a métrica da matriz de confusão para ilustrar o desempenho da árvore de decisão enquanto classificador, vale ressaltar que a matriz gerada é binária, tendo duas saídas possíveis para ambiente agradável ou desagradável, neste caso é possível analisar a classe esperada em função da classe prevista.

Partindo desse princípio, foram obtidos quatro resultados ao final da classificação, conforme ilustrado na Figura 9, onde a métrica verdadeiro positivo (TP) ocorre quando a classe esperada é igual a classe de saída, neste caso, obteve-se a previsão correta para ambiente agradável em 233 amostras. Para a métrica falso positivo (FP) a classe esperada era para ambiente desagradável, entretanto a classe prevista classificou como agradável, deste modo, o algoritmo de IA errou a previsão de 34 amostras. Em relação a métrica de falso negativo (FN) a classe esperada era para ambiente agradável, porém a classe prevista classificou como desagradável, errando a previsão de 1 amostra para ambiente agradável. Por fim, na métrica verdadeiro negativo (TN), a classe de saída esperada era para ambiente desagradável e a classe de previsão também agradável, no total foram classificados corretamente 864 amostras como clima desagradável.



Figura 9. Matriz de confusão da árvore de decisão resultante.

Portanto, averiguou-se mais uma vez a eficácia da IA em conjunto com árvore decisão na previsão de condições ambientais em residência, a partir dos elevados números de acertos alcançados em TP e TN.

4. Conclusão

Este trabalho apresentou uma solução baseada em IA, com a aplicação do algoritmo de árvore de decisão para prever as condições ambientais de uma base de dados, coletada a partir de uma rede IoT com monitoramento em tempo real de um cenário de *smart home*.

Seguindo as conjecturas de pré-processamento de dados para aplicação de IA, foi possível verificar o passo a passo para a manipulação de uma base de dados, destacando as seguintes etapas: aquisição de dados, processamento de dados, exploração de dados, eliminação manual de atributos, e aplicação de IA para classificação dos dados. Todas essas etapas foram imprescindíveis para garantir a estabilidade e integridade do conjunto amostral, tendo em vista, o momento da classificação dos dados pelo algoritmo de árvore de decisão.

Como resultados, a árvore de decisão foi construída a partir da definição do atributo umidade como nó raiz, isto é, utilizando a métrica para o cálculo de entropia. Com o algoritmo pronto, dividiu-se a base de dados entre treinamento e teste, possibilitando a comparação do nível de precisão entre os dois conjuntos de amostra.

Ao final verificou-se que a Árvore de decisão foi um método adequado para o problema proposto, predizendo as saídas com uma taxa acima de 95%. Deste modo, é possível concluir que é possível determinar se o ambiente está agradável ou desagradável a partir do uso de dados coletados em uma rede IoT e aplicando uma solução baseada em IA, especificamente baseada no algoritmo de árvore de decisão.

Para trabalhos futuros pretende-se realizar teste mais aprofundados para melhoria dos resultados obtidos pela árvore de decisão, considerando, por exemplo, o aumento da base de dados, expansão da quantidade de testes para assegurar a eficácia do algoritmo, assim como testes individuas para cada atributo. Por fim, espera-se instalar os sensores para realizar coletas em outros ambientes de *smart home*.

Referências

Carvalho, H., Sá, J. e Farias, F. (2021). Implantação de uma Arquitetura de Software para Monitoramento de Dados Ambientais em um Cenário de Smart Campus. In

- Anais do XII Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais, p. 167–170. SBC.
- Dash, S. e Choudekar, P. (2021). Home Automation using Smart Devices and IoT. In *IEEE Latin America Transactions. 9th International Conference on Reliability, Infocom Technologies and Optimization*, p. 1–5. IEEE.
- Elçi, E., Yalçin, N. e Ünaldi, S. (2021). Internet of Things based Smart Home System Design for Individuals With Disabilities. In *IEEE Latin America Transactions*. 6th International Conference on Computer Science and Engineering (UBMK), p. 619–622. IEEE.
- Ferreira, F. H. C., e Araújo, R. M. (2018). Campus Inteligentes: Conceitos, aplicações, tecnologias e desafios. *RelaTe-DIA*, *11*(1). Recuperado de http://www.seer.unirio.br/monografiasppgi/article/view/7147
- Kohonen, T. (1999). The self-organizing map. In *Proceedings of the IEEE*, p 1464–1990. IEEE.
- Silva, E. do N. da, (2021). Uma Solução De Software Back End Para Gestão E Consulta De Dados Em Bancos De Dados Relacional E Em Tempo Real: Um Caso De Estudo No Contexto Da Mobilidade Urbana. In *Trabalho de Conclusão de Curso (Graduação) Curso de Sistemas de Informação, Campus Universitário de Cametá, Universidade Federal do Pará, Cametá*, p. 51.
- Smith, A. and Jones, B. (1999). On the complexity of computing. In *Advances in Computer Science*, pages 555–566. Publishing Press.
- Pinto, K. S. et al. (2020) "Um Estudo de Caso de Internet das Coisas Utilizando o Ecossistema Dojot: Da Configuração a Utilização", Conferência Nacional em Comunicações, Redes e Segurança da Informação, X Edição, p. 95.
- Rajeshkanna, A.; Arunesh, K. (2020). ID3 Decision Tree Classification: An Algorithmic Perspective based on Error rate. In *IEEE Latin America Transactions*. 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), p. 787–790. IEEE.
- Wang, M., Fu, C., Du, X. (2021). Decision-Tree Based Root Cause Localization for Anomalies in Smart IoT Systems. In *IEEE Latin America Transactions*. *ICC 2021 IEEE International Conference on Communications*, p. 1–5. IEEE.
- Zhamanov, A. et al. (2021). IoT Smart Campus Review and Implementation of IoT Applications into Education Process of University. In 13th International Conference on Electronics, Computer and Computation (ICECCO), p. 1–4. IEEE.