

APLICAÇÃO DA IOT EM ANÁLISE PREDITIVA DE FALHAS

Edilberto Strauss

Ph.D. em Ciência e Tecnologia, ITLab/POLI da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil
strauss@poli.ufrj.br

Ana Gabriella Amorim Abreu Pereira

D. Sc. em Ciência em Engenharia de Produção, ITLab/POLI da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil
aga.amorim@gmail.com

Manoel Villas Bôas Júnior

Mestre em Computação Aplicada, ITLab/POLI da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil
mvbjunior@poli.ufrj.br

Maurício Pereira de Oliveira

Engenheiro mecânico pela Universidade Severino Sombra (USS), Vassouras, RJ, Brasil
mauriciocomrad@gmail.com

Manuel Oliveira Lemos Alexandre

Mestre em Engenharia de Transportes, COPPE - Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil
manuel.alexandre@pep.ufrj.br

RESUMO

Desde a publicação do Plano Nacional de Internet das Coisas no Brasil em 2019, a tecnologia tem sido objeto de interesse pelas administrações públicas, principalmente voltada para a melhoria dos seus processos internos, aproveitando suas inúmeras aplicações e potenciais. Este artigo apresenta o resultado de uma Prova de Conceito que testa a hipótese de se utilizar os dados de monitoramento energético de uma solução de IoT instalada experimentalmente em diversas agências de um banco público, capaz de subsidiar a tomada de decisão através de uma análise preditiva de falhas, com o objetivo de identificar tempestivamente o momento para acionamento da manutenção dos equipamentos de ar condicionado, antes da ocorrência da falha ou defeito, utilizando técnicas de exploração dos dados, desenvolvimento e validação de um modelo de previsão, bem como propõe uma arquitetura de referência e um pipeline de dados a fim de subsidiar uma futura aplicação em escala.

Palavras-chave: IoT, Prova de Conceito, Ar-condicionado, Pipeline de dados.

APPLICATION OF IOT IN PREDICTIVE FAILURE ANALYSIS

ABSTRACT

Since the publication of the National Plan for the Internet of Things in Brazil in 2019, the technology has been the object of interest by public administrations, mainly focused on improving their internal processes, taking advantage of its numerous applications and potentials. This article presents the result of a Proof of Concept that tests the hypothesis of using the energy monitoring data of an IoT solution installed experimentally in several branches of a public bank, capable of subsidizing decision making through a predictive analysis of failures, in order to identify in a timely manner the moment to trigger the maintenance of air conditioning equipment, before the occurrence of the fault or defect, using data exploration techniques, development and validation of a prediction model, as well as proposing a reference architecture and a data pipeline in order to subsidize a future application at scale.

Keywords: IoT, Proof of Concept, Air Conditioning, Data Pipeline.

1 INTRODUÇÃO

Internet das Coisas ou “Internet of Things” (IoT), é definido pelo PNIC (2019), como sendo a infraestrutura que integra a prestação de serviços de valor adicionado com capacidades de conexão física ou virtual de coisas com dispositivos baseados em tecnologias da informação e comunicação existentes e nas suas evoluções, com interoperabilidade, sendo coisas, objetos no mundo físico ou no mundo digital, capazes de serem identificados e integrados pelas redes de comunicação, e dispositivos, os equipamentos ou subconjuntos de equipamentos com capacidade mandatória de comunicação e capacidade opcional de sensoriamento, de atuação, de coleta, de armazenamento e de processamento de dados.

O uso da IoT em conjunto com a aplicação de análises preditivas tem potencial de trazer grandes benefícios para as empresas. Levantamento realizado pelo TCU (2020), mostra que o impacto da IoT na produtividade dos processos fabris pode gerar uma economia de 10% a 20% no uso de recursos energéticos, de 10% a 25% na eficiência de mão de obra e uma redução nos custos de manutenção de equipamentos de 10% a 40% para esses ambientes até o ano de 2025. Somente com as inovações oportunizadas aos sistemas fabris de manutenção preditiva prevê-se um impacto econômico mundial de cerca de U\$ 1,6 trilhão até 2025, portanto, ainda com um enorme potencial a ser explorado pelas organizações.

Com este propósito, um banco público publicou em 2020 um edital para contratação de um “piloto” de solução de IoT, especialmente direcionada a gestão energética em 300 agências bancárias, correspondentes à aproximadamente 6% do seu portfólio, com o objetivo testar o potencial da solução para identificar oportunidades de melhoria na eficiência energética,

alcançando uma economia média mensal de 12% no consumo de energia, apenas com a redução de desperdícios, conseguida pela operação remota dos equipamentos de ar condicionado, sem utilizar análises avançadas dos dados.

Esse artigo tem por objetivo, apresentar o resultado de uma Prova de Conceito que testa a hipótese de se utilizar os dados de monitoramento energético das 300 agências bancárias contempladas no “piloto”, cedidos pelo banco público, para subsidiar à tomada de decisão sob a ótica da manutenção, através de uma análise preditiva de falhas nos equipamentos de ar condicionado, com o objetivo de identificar tempestivamente o momento para acionamento da mantenedora, antes da ocorrência da falha ou defeito, utilizando aprendizado de máquina, bem como, propõe uma arquitetura de referência e um pipeline de dados a fim de contribuir para o aprimoramento e desenvolvimento de novas soluções.

2 DESENVOLVIMENTO

2.1 REVISÃO DA LITERATURA

Apresentação das metodologias, principais conceitos e abordagens utilizadas no desenvolvimento do estudo.

2.1.1 PROVA DE CONCEITO

Prova de Conceito ou “Prove of Concept” (PoC) é uma etapa crucial no desenvolvimento e implementação de soluções de engenharia de dados. Ela envolve a criação de um protótipo funcional para validar a viabilidade técnica e operacional de uma solução hipotética antes de sua implementação em larga escala (Databricks, 2023).

Uma PoC realizada de forma adequada, possibilita tomar decisões mais informadas e obter uma compreensão mais clara dos potenciais benefícios e desafios de uma solução proposta antes de sua implementação em produção.

2.1.2 MINERAÇÃO DE DADOS

“Data mining” ou mineração de dados, é um processo de descoberta de padrões, relações e informações significativas em conjuntos de dados grandes e complexos. É uma disciplina que

combina técnicas de estatística, inteligência artificial, aprendizado de máquina e banco de dados para extrair conhecimento útil a partir dos dados (Shmueli, et al. 2019).

O Quadro 1 apresenta as principais etapas envolvidas no processo de mineração dos dados.

Quadro 1 – Etapas do processo de data mining

Etapa	Processo
Definição do problema	Definição clara do problema a ser resolvido, descrição dos fatores, desafios e metas a serem alcançados.
Estabelecimento dos objetivos	Estabelecer claramente os objetivos do modelo de dados, quais questões deverão ser respondidas e quais informações serão relevantes para o melhor entendimento do contexto.
Preparação dos dados	Seleção e pré-processamento dos dados relevantes para a análise. Inclui limpeza, transformação e integração dos dados, garantindo sua qualidade e consistência.
Exploração dos dados	Aplicação de técnicas estatísticas e de visualização para identificar padrões, tendências e relações nos dados. Envolve a utilização de gráficos, tabelas, medidas de tendência central, variabilidade, correlação, entre outros.
Modelagem dos dados	Construção de modelos preditivos ou descritivos com base nos padrões encontrados na etapa anterior. Aplicação de algoritmos de aprendizado de máquina para construir esses modelos, utilizados para fazer previsões ou para entender melhor os dados.
Validação	Avaliação quanto à sua precisão e eficácia na extração de informações relevantes. Os resultados obtidos são interpretados para fornecer insights e conhecimentos úteis para subsidiar tomadas de decisão

Fonte: Resultados originais da pesquisa

2.1.3 ANÁLISES PREDITIVAS

As análises preditivas são uma área da ciência de dados que visa utilizar algoritmos e técnicas estatísticas para fazer previsões ou estimativas sobre eventos futuros ou comportamentos. Essa abordagem se baseia no uso de dados históricos e padrões identificados para extrapolar informações e realizar predições confiáveis (Fávero, 2017).

Uma das técnicas utilizadas é a árvores de Decisão, que são estruturas em formato de árvore que ajudam a tomar decisões ou prever eventos com base em uma série de condições e variáveis. Elas são usadas em uma variedade de aplicações e são particularmente úteis quando se deseja entender o raciocínio por trás das previsões.

2.2 SITUAÇÃO ATUAL

A solução de IoT compreende o fornecimento e instalação de medidores, sensores, infraestrutura interna de comunicação de dados, com fio e sem fio, gerenciadores e modem celular GPRS de comunicação com o servidor de dados da contratada, bem como o fornecimento de plataforma web para acesso pelo contratante, contendo gráficos, relatórios, alarmes e a possibilidade de download dos dados de consumo em formato .CSV.

Foram instalados medidores de grandezas elétricas nos quadros de energia de alimentação principal das agências denominados Quadros Gerais de Baixa Tensão (QGBT) e nos quadros de alimentação dos sistemas de ar condicionado denominados Quadro de Força de Ar Condicionado (QFAC).

Além disso, foram instalados sensores de temperatura em pontos estratégicos das agências, para monitoramento de Temperatura de Bulbo Seco (TBS) nos dutos de insuflamento, nas áreas internas climatizadas, no retorno do ar para as salas de máquinas e do ar exterior.

O escopo não prevê o tratamento analítico dos dados monitorados para uso pela contratante, bem como não foram previstos requisitos para os dados disponibilizados para “download”.

2.3 ESTUDO DE CASO

Os dados baixados da plataforma da contratada foram submetidos ao processo de mineração de dados, com o objetivo de identificar anomalias, padrões e correlações, com potencial de prever resultados de falhas ou defeitos nos equipamentos de ar condicionado.

2.3.1 DEFINIÇÃO DOS PROBLEMAS E OBJETIVOS

Dada a natureza da solução de IoT implementada, direcionada a gestão energética das agências, os dados gerados se achem a grandezas elétricas medidas nos quadros de alimentação, bem como das temperaturas relacionadas ao conforto térmico nos ambientes climatizados, não sendo fornecidos outros dados relativos ao funcionamento interno dos componentes dos equipamentos de ar condicionado.

Outro desafio encontrado, foi a diversidade de tamanhos de equipamentos de ar condicionado encontrados nas agências, assim como as medições das grandezas elétricas serem

realizadas nos quadros de alimentação, havendo a possibilidade da ocorrência de dois ou mais equipamentos alimentados pelo mesmo quadro.

Assim, foram estabelecidos os seguintes objetivos:

- a) descobrir padrões nos dados de monitoramento da IoT que possam expressar o funcionamento dos equipamentos de ar condicionado;
- b) criar indicadores que permitam avaliar o desempenho dos equipamentos;
- c) identificar padrões de comportamento anômalos dentro (rotular os dados);
- d) executar análise preditiva nos dados rotulados;
- e) analisar o nível de confiança dos resultados para validação da hipótese.

2.3.2 PREPARAÇÃO DOS DADOS

O estudo foi desenvolvido usando serviços de computação em nuvem para facilitar a coleta de informações necessárias para definir um pipeline de dados para apoiar futuras implementações em escala, incluindo coleta, processamento, armazenamento e análise de dados.

2.3.2.1 EXTRAÇÃO DOS DADOS

Para o desenvolvimento deste estudo, foram utilizados os dados de monitoramento da solução de IoT baixados diretamente da plataforma WEB disponibilizada pela contratada.

A plataforma possui alguns recursos de filtros de dados, possibilitando a seleção de agências, UF, TBS dos ambientes e de “setpoint” das máquinas, QGBT, QFAC e períodos, porém, com poucos recursos de filtros simultâneos.

Grandezas elétricas disponibilizadas: consumo de energia em “kWh” e “kVArh”, tensão, corrente elétrica, fator de potência e frequência.

Os “downloads” são disponibilizados em formato .CSV, limitados a 20 MB por operação. Também só é possível selecionar um período mensal por vez.

Intervalo mínimo para comunicação com o servidor de dados para transferência dos registros, conforme premissa do contrato, a cada 15 minutos, permitido atraso na troca de dados por até 4 horas, em caso de anormalidade.

A solução de IoT monitora sete pontos de medição de temperatura diferentes em cada agência. Estes dados precisam ser baixados individualmente.

O estudo se baseou em dados diários de grandezas elétricas medidos nos QFAC e de temperaturas entre abril de 2022 a maio de 2023, com o objetivo de se obter pelo menos um ciclo completo de observações, prevendo-se efeitos da sazonalidade, como por influência do clima ou de operações do contratante.

Quadro 2 – Identificação dos pontos de medição de temperatura

SET POINT	“Setpoint” de temperatura dos equipamentos de ar condicionado
TEMP. CAIXAS	Temperatura do ar no guichê dos caixas
TEMP. INSUFLAMENTO	Temperatura do ar que sai do equipamento de ar condicionado
TEMP. RETORNO	Temperatura do ar que retorna para a casa de máquinas
TEMP. SALA ATENDIMENTO	Temperatura do ar na sala de atendimento ao público
TEMP. SALA ONLINE	Temperatura do ar na sala onde ficam os racks de TI
TMA	Temperatura do ar externo

Fonte: Resultados originais da pesquisa

2.3.2.2 INTERPRETAÇÃO DOS DADOS

As tabelas baixadas da plataforma são constituídos por séries temporais dos dados elétricos medidos nos quadros de alimentação dos equipamentos de ar condicionado e pelas respectivas temperaturas medidas nos ambientes climatizados, medidos 24 horas por dia, inclusive finais de semana e feriados, com intervalos de 15 minutos entre as medições. Já as medições de consumo são disponibilizadas com intervalos de 1 hora.

A Figura 1 apresenta a tabela de dados baixada da plataforma do contratado em formato .CSV, contendo medições de TBS das agências, em (°C), em ordem cronológica.

Figura 1. Dados de temperatura medidos na sala de atendimento

Ponto de Medição	Data	Valor	Unidade
001-0584 - BB/VILA MARIA (SP) - Temp. Ambiente - Sala de Atendimento	07/06/2023 00:16	23,8	°C
001-0584 - BB/VILA MARIA (SP) - Temp. Ambiente - Sala de Atendimento	07/06/2023 00:31	23,8	°C
001-0584 - BB/VILA MARIA (SP) - Temp. Ambiente - Sala de Atendimento	07/06/2023 00:46	23,8	°C
001-0584 - BB/VILA MARIA (SP) - Temp. Ambiente - Sala de Atendimento	07/06/2023 01:01	23,8	°C
001-0584 - BB/VILA MARIA (SP) - Temp. Ambiente - Sala de Atendimento	07/06/2023 01:17	23,8	°C

Fonte: Resultados originais da pesquisa

2.3.2.3 TRANSFORMAÇÃO DOS DADOS

Para seleção da amostra foi adotado o método de amostragem por conveniência mencionada por Tashakkori (2003), conhecido como um método rápido e fácil para representar adequadamente uma população, baseada na disponibilidade e conveniência dos elementos. Esta etapa foi realizada com a ajuda da equipe de engenharia da contratante da solução de IoT, que indicou uma agência de referência por possuir poucos equipamentos de ar condicionado, possibilitando a obtenção de dados mais individualizados.

2.3.2.4 INTEGRAÇÃO DE DADOS

As tabelas baixadas da plataforma foram combinadas, obtendo-se duas grandes tabelas unificadas e relacionadas, a primeira contendo somente os dados elétricos e a segunda os dados de medição de temperatura. Estas tabelas foram tratadas separadamente, em que foram realizados processos de limpeza e padronização dos dados.

Posteriormente, as duas tabelas foram unificadas para formar um único conjunto de dados compatibilizados, com o objetivo de facilitar a análise das variáveis do processo. A integração dos dados em duas etapas foi escolhida para facilitar o processo de equalização dos intervalos de medição entre os dados elétricos e de temperaturas.

2.3.2.5 LIMPEZA DE DADOS

Foram identificados e corrigidos erros, preenchidos valores ausentes, removidas duplicatas e tratadas inconsistências nos dados. Os campos em branco, identificados em consulta a plataforma como falha na comunicação ou interrupção de energia, foram preenchidos utilizando-se a média calculada a partir dos valores imediatamente anterior e posterior. Os demais, sem possibilidade de identificação da causa pela plataforma, ou com longos intervalos sem medições registradas, não foram alterados, e mantidos na tabela.

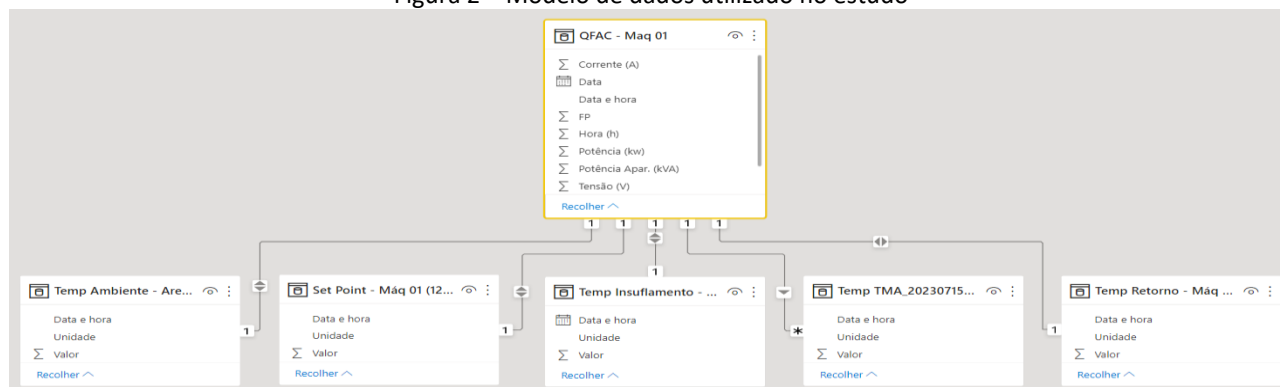
2.3.2.6 PADRONIZAÇÃO DE DADOS

Foram identificadas e corrigidas inconsistências e falta de padronizado, incluindo equalização de casas decimais e arredondamentos. Por fim, foram removidos os outliers, como medições de temperatura acima de 50°C ou negativas, consideradas como provável defeito nos sensores em consulta a equipe técnica de engenharia da contratante.

Os registros horários foram equalizados para que na tabela resultante constassem de forma sincronizada e na mesma escala temporal. Para tanto, os dados foram arredondados para que aparecessem como “hora cheia” e selecionados os primeiros registros das observações repetidas.

A Figura 2 apresenta o modelo de dados resultantes das integrações, constando o relacionamento entre a tabela de dados elétricos medidos nos QFAC e das temperaturas.

Figura 2 – Modelo de dados utilizado no estudo



Fonte: Resultados originais da pesquisa

2.3.3 CARREGAMENTO DOS DADOS

O processo de tratamento resultou numa tabela contendo um único conjunto de dados estruturados e estratificados, possibilitando análises visuais e interpretações das variáveis do processo.

O conjunto de dados resultante deste processo representa as condições potencialmente encontradas caso a análise preditiva de falhas fosse implementada em escala, incluindo estrutura, formatação e inconsistências, visando a automação do processo.

A Figura 3 apresenta o leiaute final obtido com o tratamento e preparação dos dados baixados na plataforma da contratada com o objetivo de subsidiar o processo de data mining, resultando em uma única tabela contendo datas e horas equalizadas.

Figura 3 – Dados elétricos e de temperatura consolidados

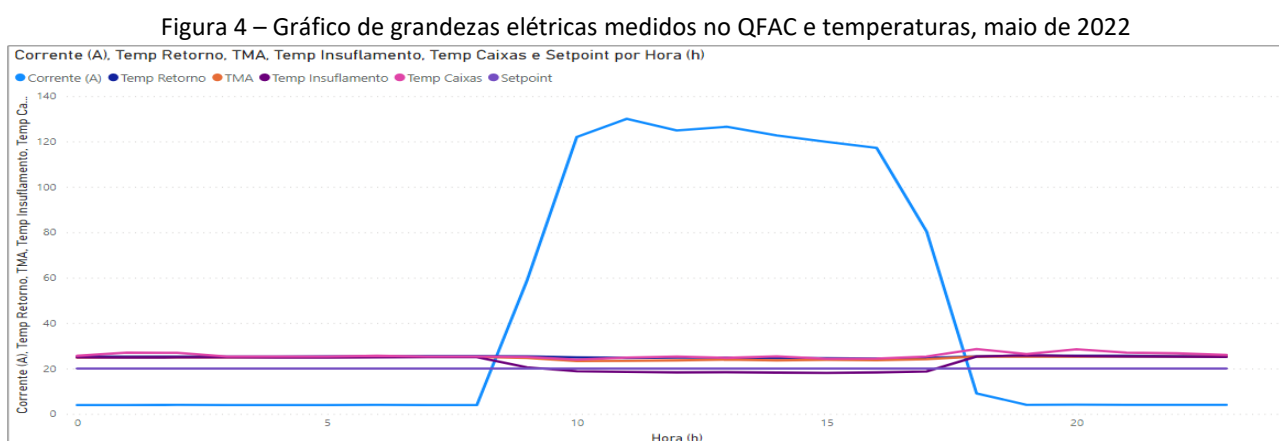
Data e hora	Potência (kW)	Potência aparente (kVA)	Tensão (V)	Corrente (A)	Temp Insuflamento (°C)	TMA (°C)	Temp Retorno (°C)	Temp Caixas (°C)	Setpoint (°C)
02/05/2023 09:00:00	25,82	19,44	127,75	67,36	25,80	26,10	26,10	26,10	20,00
02/05/2023 10:00:00	59,23	40,06	127,68	154,64	19,30	20,70	25,90	22,50	20,00
02/05/2023 11:00:00	73,37	47,38	127,65	191,59	18,20	20,30	24,80	23,60	20,00
02/05/2023 12:00:00	71,58	46,63	127,65	186,91	18,00	19,60	24,50	23,00	20,00
02/05/2023 13:00:00	71,03	46,46	127,65	185,46	17,80	20,40	24,20	22,00	20,00

Fonte: Resultados originais da pesquisa

2.3.4 EXPLORAÇÃO DOS DADOS

Nesta etapa, foram aplicadas técnicas estatísticas e de visualização para identificar padrões, tendências e relações entre as variáveis do processo. Para tanto, foram utilizadas tabelas, gráficos, medidas de tendência central, variabilidade, correlação, entre outros.

A Figura 4 apresenta um gráfico contendo, simultaneamente, medições das grandezas elétricas do QFAC e de temperatura durante o período de 24 horas, onde se pode observar a relação entre estas grandezas e as horas do dia.



Fonte: Resultados originais da pesquisa

Os resultados da análise mostraram que a corrente exibiu a maior variabilidade entre os dados elétricos, respondendo com maior sensibilidade ao funcionamento do equipamento de ar condicionado. Por outro lado, a tensão não mudou significativamente e não foi afetada no processo.

Com o objetivo de confirmar o padrão observado na análise visual, foi aplicada análise descritiva, calculando-se medidas de tendência central em diversos períodos. A análise considerou apenas dias úteis e o horário de funcionamento dos equipamentos, com o objetivo de se obter as curvas características das variáveis do processo. Também foram dispensadas da amostra os dias cujas variáveis contivessem valor zero, por considerar como possível indicação de problema local, como falha na comunicação ou defeito nos sensores.

A Figura 5 apresenta os resultados obtidos na análise descritiva, em que foram calculadas medidas de tendência central e de dispersão para corrente elétrica, temperatura do ambiente climatizado, temperatura de insuflamento de ar e temperatura do ar externo de diversos dias seguidos. Estes dados foram selecionados por apresentarem maior variabilidade diária.

Figura 5 – Medidas de tendência central e de dispersão de diversos dias, maio de 2023

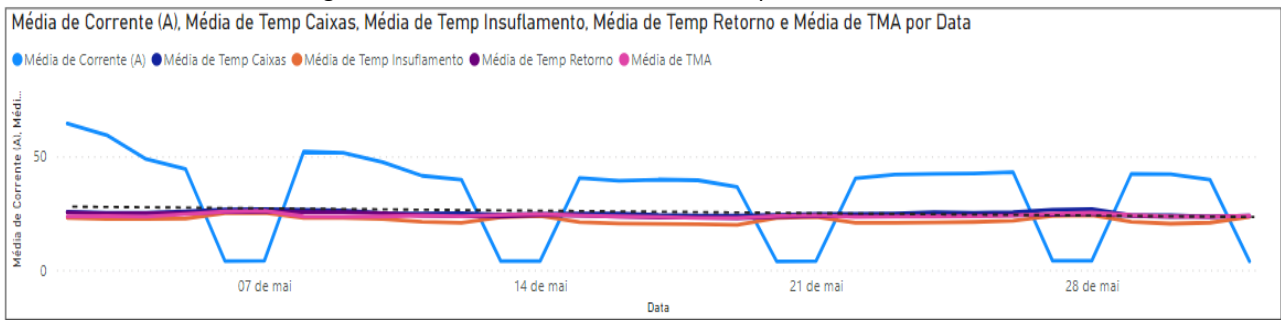
Data	Corr. Méd.	Corr. Des.	Corr. Mín.	Corr. Máx.	Corr. Var.	Temp. Amb. Méd.	Temp. Amb. Des.	Temp. Amb. Mín.	Temp. Amb. Máx.	Temp. Amb. Var.	Temp. Insuf. Méd.	Temp. Insuf. Des.	Temp. Insuf. Mín.	Temp. Insuf. Máx.	Temp. Insuf. Var.	TMA Méd.	TMA Des.	TMA Mín.	TMA Máx.	TMA Var.
terça-feira, 2 de maio de 2023	158,29	41,08	67,36	191,59	1687,81	24,09	1,66	22,00	26,40	2,76	18,93	2,47	17,70	25,80	6,12	20,64	1,99	19,00	26,10	3,97
quarta-feira, 3 de maio de 2023	151,42	38,83	77,17	189,48	1507,60	23,61	1,73	21,50	26,10	3,00	18,36	2,43	17,00	25,10	5,88	21,54	1,91	20,00	25,90	3,64
quinta-feira, 4 de maio de 2023	123,18	30,47	57,37	149,38	928,30	23,61	1,74	21,50	27,00	3,04	18,53	2,26	16,90	24,80	5,13	21,34	2,00	19,40	25,30	4,00
sexta-feira, 5 de maio de 2023	111,34	23,22	58,58	130,00	538,97	24,80	0,49	23,90	25,40	0,24	18,64	0,69	18,10	20,50	0,48	23,79	0,40	23,30	24,70	0,16
segunda-feira, 8 de maio de 2023	132,01	26,93	69,01	154,00	725,01	24,36	1,50	22,90	26,60	2,26	18,90	2,33	17,60	25,40	5,44	20,41	2,04	19,10	26,10	4,15
terça-feira, 9 de maio de 2023	129,98	25,64	69,73	154,48	657,25	24,31	2,32	21,70	28,20	5,37	18,79	0,62	18,30	20,40	0,39	20,42	0,92	19,60	22,90	0,85

Fonte: Resultados originais da pesquisa

Pela análise dos resultados apresentados, observa-se correlações entre as medidas calculadas das variáveis do processo, em que se pode inferir que o comportamento do equipamento em responder as variações diárias da carga térmica se mantiveram semelhantes nos períodos observados.

A Figura 6 apresenta as médias diárias das variáveis do processo calculadas para um período de 30 dias e a respectiva linha de tendência representada pela linha tracejada, onde se pode observar a manutenção destes valores.

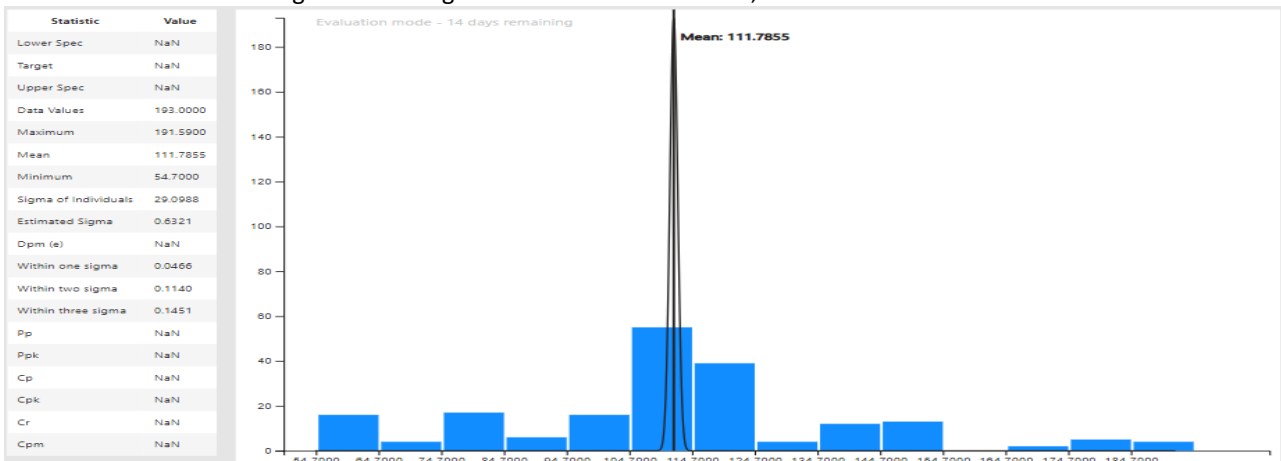
Figura 6 – Médias diárias das variáveis do processo, maio 2022



Fonte: Resultados originais da pesquisa

Também se pode observar uma maior concentração de valores de corrente elétrica próximas ao valor médio, bem como uma distribuição dos demais valores em torno deste, como mostrado na Figura 7.

Figura 7 – Histograma de correntes elétricas, abril 2022 a maio 2023

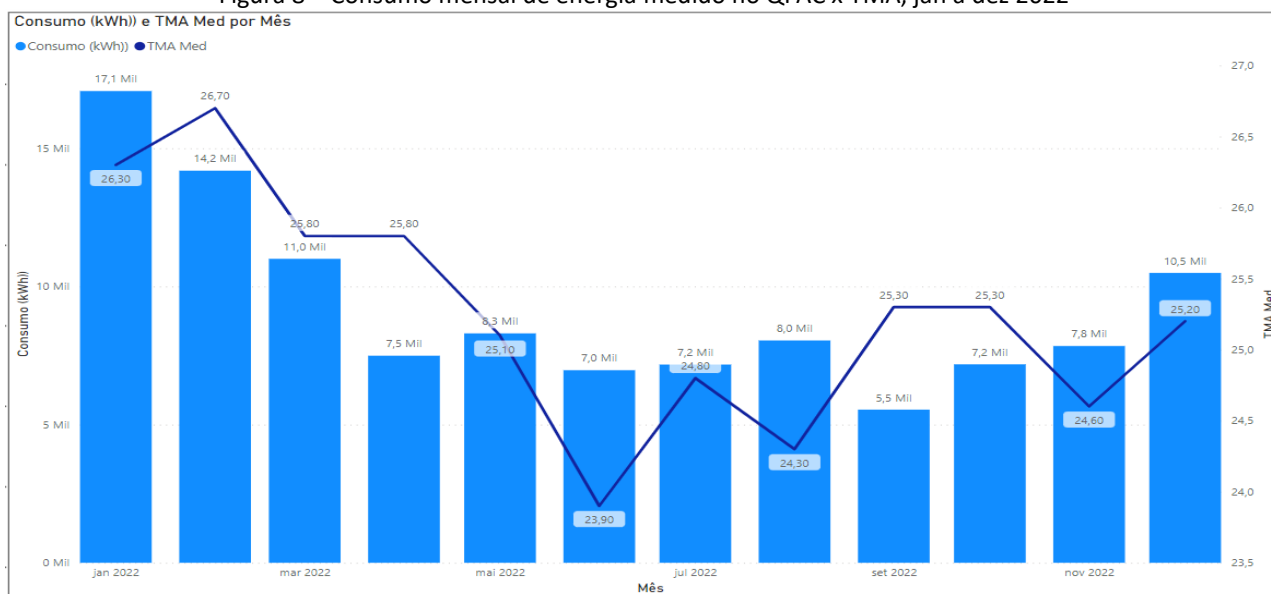


Fonte: Resultados originais da pesquisa

Também foi possível observar sazonalidade por influência do clima, cujas medidas de tendência central retornam ao valor inicial ao final de um ciclo, correspondente a um ano de observação.

A Figura 8 apresenta o comparativo entre os valores de consumo energia mensal medidos no QFAC, correspondente ao período de um ano, onde se pode observar uma sazonalidade por influência da temperatura externa e uma tendência de retorno ao valor inicial.

Figura 8 – Consumo mensal de energia medido no QFAC x TMA, jan a dez 2022



Fonte: Resultados originais da pesquisa

O processo de exploração de dados permitiu identificar uma relação entre os dados elétricos e as respectivas temperaturas operacionais medidas no equipamento de ar condicionado durante o horário de funcionamento da agência. Por esta hipótese, o comportamento observado nas variáveis do processo representaria o padrão de funcionamento desses equipamentos para exercer sua função, que é manter a temperatura de conforto nos ambientes.

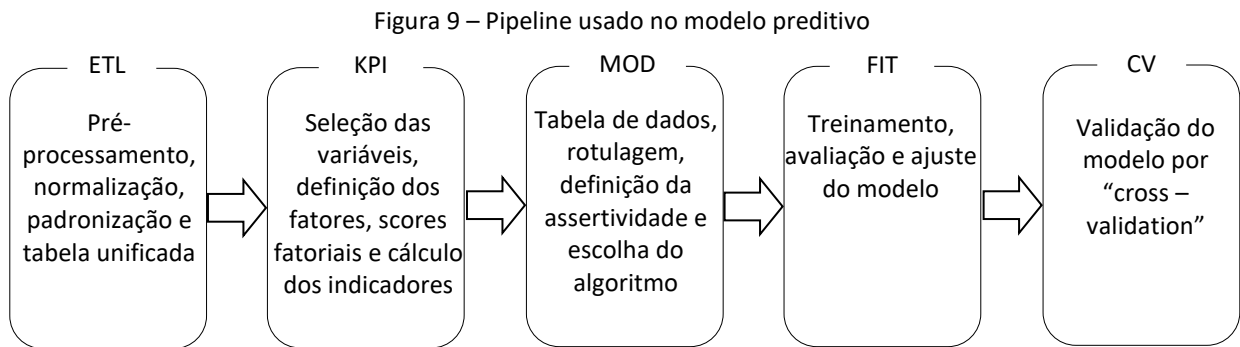
A análise também confirmou comportamentos semelhantes para outros períodos de observação, sendo percebidas mesmas tendências centrais e variações influenciadas por fatores externos, como temperatura externa TMA e do ambiente interno, sendo a primeira atribuída ao tempo meteorológico e ao clima (sazonalidade) e a segunda à quantidade de clientes. Por esta hipótese, supôs-se que os dados históricos utilizados na análise exploratória representariam o funcionamento normal do equipamento.

2.3.5 CONSTRUÇÃO DO MODELO

Com base nos resultados obtidos na exploração dos dados, foi proposto modelo de previsão capaz de representar estes comportamentos por meio de indicadores de desempenho calculados a partir das variáveis do processo, resultante do tratamento e unificação das massas de dados baixadas da solução de IoT.

Para tanto, foi utilizado algoritmo de classificação para classificar os indicadores diários entre resultados bons, ou seja, próprias do funcionamento normal do equipamento de ar condicionado, ou ruins, ou seja, fora dos padrões e, portanto, com risco de indisponibilidade futura.

A Figura 9 apresenta o pipeline utilizado no modelo de previsão, que inclui a preparação dos dados, a definição dos indicadores do processo, a criação da tabela de treinamento, o processo de treinamento e a validação do modelo.



Fonte: Resultados originais da pesquisa

2.3.5.1 PREPARAÇÃO DOS DADOS

Nesta etapa, foram selecionados os dados elétricos e as respectivas temperaturas operacionais de um equipamento de ar condicionado selecionado pelo método de amostragem por conveniência, considerando como principais requisitos a completude e a constâncias dos registros de monitoramento, para que se obtivessem dados históricos de um ciclo anual completo.

Estes dados foram tratados e normalizados, sendo selecionados apenas os registros efetuados em dias úteis e durante o horário de funcionamento da agência, obtendo-se os dados representativos típicos do processo de funcionamento do equipamento de ar condicionado.

A Figura 10 apresenta o resultado do tratamento dos dados correspondente a um dia de funcionamento do equipamento de ar condicionado.

Figura 10 – Tabela resultante do tratamento dos dados, 2022 a 2023

	Data e hora	Potência (kW)	Potência aparente (kVA)	Tensão (V)	Corrente (A)	Temp Insuflamento (°C)	TMA (°C)	Temp Retorno (°C)	Temp Caixas (°C)
0	2023-05-05 08:00:00	15	2	12701	393	2500	251	255	253
1	2023-05-05 09:00:00	2247	197	12788	5858	205	247	254	251
2	2023-05-05 10:00:00	4677	3552	12776	12202	188	233	249	239
3	2023-05-05 11:00:00	4982	3718	12775	13000	185	234	247	248
4	2023-05-05 12:00:00	4786	3623	12776	12487	183	236	248	253

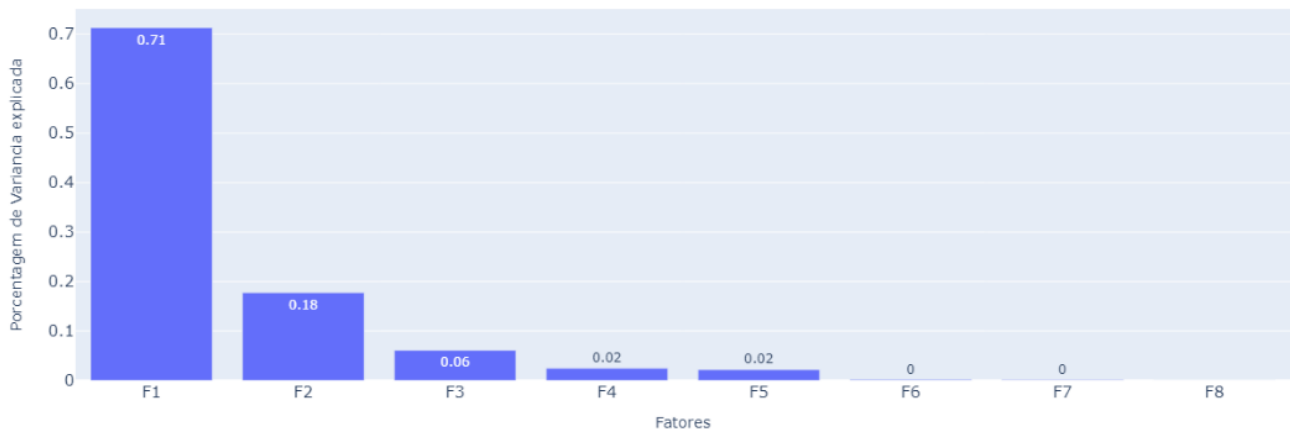
Fonte: Resultados originais da pesquisa

2.3.5.2 DEFINIÇÃO DOS INDICADORES

Para definição dos indicadores, foram selecionadas as variáveis de maior representatividade do processo, consideradas aquelas com maior variabilidade, utilizando-se análise PCA, obtendo-se os escores fatoriais calculados a partir dos componentes principais, a partir dos quais foram calculados os indicadores de desempenho, correspondentes às medidas de tendência central e de dispersão diárias.

A Figura 11 apresenta o resultado da análise PCA em que se pode observar que os dois primeiros componentes principais respondem por aproximadamente 90% das variáveis do processo.

Figura 11 – Scree Plot resultado da análise PCA



Fonte: Resultados originais da pesquisa

A Tabela 1 apresenta o resumo dos resultados obtidos com a análise PCA, onde a primeira componente principal representa 5,7 de um total de 8 variáveis do processo.

Tabela 1 – Resumo da análise PCA

Index	Fator	Autovalor	Variância explicada	Variância acumulada
0	F1	5.703687e+00	7.129609e-01	0.712961
1	F2	1.418913e+00	1.773642e-01	0.890325
2	F3	4.870269e-01	6.087836e-02	0.951203
3	F4	1.960934e-01	2.451168e-02	0.975715
4	F5	1.740017e-01	2.175021e-02	0.997465
5	F6	1.318170e-02	1.647712e-03	0.999113
6	F7	7.095475e-03	8.869344e-04	1.000000
7	F8	3.234374e-09	4.042968e-10	1.000000

Fonte: Resultados originais da pesquisa

2.3.5.3 MODELO DE TREINAMENTO

Os indicadores diários calculados na etapa anterior foram rotulados com base nos resultados obtidos na análise descritiva, que por sua vez foram classificados como “0” os resultados ruins e “1” os resultados bons. Como referência para definição dos resultados ruins, adotou-se os dois dias que antecederam os chamados de manutenção corretiva urgente, conforme informações fornecidas pelo contratante da solução de IoT, obtidos de sua plataforma de gestão de manutenção, num total de 3% da amostra durante o período de um ano.

A Figura 12 representa a tabela de dados para treinamento proposto contendo os indicadores de desempenho do processo e os respectivos rótulos.

Figura 12 – Tabela de dados de treinamento

Data	Média de F1	Mínimo de F1	Máximo de F1	Desvio padrão de F1	Variação de F1	Média de F2	Mínimo de F2	Máximo de F2	Desvio padrão de F2	Variação de F2	Rótulo
02/05/2023	-4E+015	-4E+016	3E+016	1E+016	2E+032	-3E+015	-2E+016	2E+016	1E+016	2E+032	1
03/05/2023	1E+015	-9E+015	2E+016	9E+015	9E+031	-3E+015	-5E+016	2E+016	2E+016	3E+032	1
04/05/2023	-2E+015	-2E+016	2E+016	1E+016	1E+032	-5E+015	-4E+016	7E+015	1E+016	1E+032	1
05/05/2023	-6E+015	-3E+016	7E+015	9E+015	7E+031	5E+014	-1E+016	2E+016	8E+015	7E+031	0
08/05/2023	4E+014	-1E+016	2E+016	9E+015	9E+031	-2E+015	-2E+016	1E+016	1E+016	1E+032	1

Fonte: Resultados do trabalho

2.3.5.4 DEFINIÇÃO DOS ALGORITMOS E MÉTRICAS

Para treinamento do modelo, foi proposto o algoritmo de Árvore de Decisão que realiza previsões aprendendo regras de decisão a partir dos dados de treinamento (Fávero, 2017). Considerou-se como o modelo mais indicado tendo em vista a tabela de treinamento ser formada por dez indicadores calculados a partir das variáveis do processo, que possuem variações sutis entre os dias bons e ruins, devendo o modelo ser capaz de identificar estas variações.

Por se tratar de uma PoC para confirmar a possibilidade de uma análise preditiva de falhas em equipamentos de ar condicionado, utilizando apenas dados básicos de monitoramento de energia e temperaturas, e, considerando que geralmente é adotada uma assertividades de 80% a 90% nos métodos de manutenção preditiva em que se utiliza medição direta dos parâmetros operacionais dos dispositivos, como vibração, pressão, etc (Serkan, A.; Horay, A., 2021), estimou-se como razoável uma assertividade mínima de 60% para confirmação do potencial da hipótese apresentada, a partir da qual se poderia sugerir um estudo mais aprofundado para uma futura aplicação em escala.

Também foi utilizado o Classificador Dummy para criar uma linha de base para comparar o desempenho com o modelo de árvore de decisão, numa proporção entre dados de treino e de teste de 75/25 a partir da referência anterior.

Quadro 3 – Algoritmos de previsão selecionados

Classificador Dummy (Dummy Classifier)	Este algoritmo faz previsões com base na distribuição das classes nos dados de treinamento, é rápido e não requer ajuste de hiperparâmetros ou treinamento complexo.
Árvore de decisão (Decision Tree Classifier)	O algoritmo de árvore de decisão é uma técnica de aprendizado de máquina supervisionado que pode ser usada para problemas de classificação e regressão. O método consiste em dividir o conjunto de dados de treinamento em subconjuntos menores, com base nos atributos das instâncias, de forma que cada subconjunto contenha instâncias de uma única classe (no caso da classificação) ou possua valores de saída semelhantes (no caso da regressão).

2.3.5.5 TREINAMENTO DO MODELO

Os dados foram submetidos ao processo de treinamento, aplicando-se técnicas de Machine Learning Supervisionado, com uso dos algoritmos selecionados, com o objetivo de verificar sua capacidade de prever situações de possível falha ou defeito nos equipamentos de ar condicionado.

Primeiramente foi calculada a acurácia baseada no classificador *Dummy Classifier*, com utilização de linguagem Python e biblioteca *sklearn.dummy*, onde foram obtidos os seguintes resultados apresentados na Tabela 2.

Tabela 2 – Resultados do classificador Dummy Classifier

Classificador	Amostra	Dados de treino	Dados de teste	Acurácia
Dummy Classifier	253	190	63	83,33%

Fonte: Resultados originais da pesquisa

Posteriormente foi aplicado o método de classificação *Decision Tree Classifier*, com utilização de linguagem Python e biblioteca *sklearn.tree*, neste caso adotado “seed” randômico e profundidade 5, indicada como a mínima suficiente na literatura disponível do site da (Scikit-Learn, 2023), onde foram obtidos os seguintes resultados apresentados na Tabela 3.

Tabela 3 – Resultados do classificador Decision Tree Classifier

Classificador	Amostra	SEED	Depth	Acurácia
Decision Tree Classifier	253	Random	5	86,36%

Fonte: Resultados originais da pesquisa

2.3.5.6 VALIDAÇÃO DO MODELO E ANÁLISE DOS RESULTADOS

Para validação do modelo foi utilizado o método de validação cruzada ou “cross validate”, com utilização de linguagem Python e biblioteca *sklearn.model_selection*, as funções *cross_validate*, *KFold* e *StratifiedKFold*, aplicados os recursos de *shuffle* e *stratify* para reduzir efeitos de aleatoriedade, desbalanceamento e “overfitting”.

Tabela 4 – Resultados obtidos no processo de cross validation

Método	Amostra	Splits	Depth	Acurácia média
Cross validation	253	6	5	81,82%

Fonte: Resultados originais da pesquisa

Pela análise dos resultados obtidos na análise preditiva foi observada uma acurácia média de 82% no modelo proposto, portanto, acima do estabelecido no estudo para confirmar o potencial da hipótese de análise preditiva de falhas usando apenas os dados de gestão energética da solução de IoT.

3 CONSIDERAÇÕES FINAIS

Assim, visando subsidiar futura tomada de decisão sobre a possibilidade de uma aplicação em escala, o estudo confirma este potencial, devendo para tanto ser complementado por outros estudos mais aprofundados, abrangendo uma massa maior de dados e outras tipologias de equipamentos.

O estudo também permitiu identificar melhorias necessárias na solução de IoT para viabilizar uma futura aplicação de análise preditiva em escala, como: integração dos dados de monitoramento

da plataforma da contratada com a aplicação da contratante; disponibilização dos dados já tratados e estruturados para aplicação da análise preditiva; individualização do monitoramento dos equipamentos de ar condicionado, para que sejam obtidos dados de grandezas elétricas individualizados; identificação individual dos equipamentos de ar condicionado, para que a solução possa fornecer “insights” valiosos, como indicar equipamentos mais eficientes energeticamente ou mais confiáveis sob o ponto de vista da manutenção; integração entre a aplicação que executa a análise preditiva e a ferramenta de gestão de manutenção, possibilitando abertura de chamados de manutenção preditiva de forma automática, dada a grande quantidade de equipamentos.

REFERÊNCIAS

DATABRICKS. **The Big Book of Data Engineering**. 2023. E-book. Disponível em: <https://www.databricks.com/sites/default/files/2023-06/big-book-of-data-engineering-2nd-edition-final.pdf>. Acesso em: 14 jun. 2023.

Fávero, L.P.; Belfiore, P. **Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®**. GEN LTC, 2017. E-book.

PNIC. BRASIL. Decreto nº 9.854, de 25 de junho de 2019. **Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas**.

Serkan, A.; Koray, A. **Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time, Expert Systems with Applications**. Elsevier, v. 173, 2021. E-book.

Shmueli, G.; Sarstedt, M.; Hair, J.F.; Cheah, J.-H.; Ting, H.; Vaithilingam, S; Ringle, C.M. **Predictive model assessment in PLS-SEM: guidelines for using PLSpredict**. European Journal of Marketing, 2019. ISSN: 0309-0566.

Tashakkori, A.; Teddlie, C. **Handbook of Mixed Methods in Social and Behavioral Research**. International Journal of Epidemiology, Volume 33, Número 6, 2004. E-book.

TCU. BRASIL. Tribunal de Contas da União. Relatório de Levantamento TC 028.109/2020-1. **Políticas públicas e programas do Governo Federal relacionados à internet das coisas (IoT)**. Obtenção de subsídios para a implementação do Plano Nacional de Internet das Coisas no Brasil, assim como para o desenvolvimento de futuros trabalhos do TCU relacionados ao tema. Encaminhamento do acórdão para conhecimento de órgãos do poder executivo e legislativo. 2020.