

Detecção de Deriva de Conceito e Anomalias no Consumo de Energia Elétrica com Árvores de Decisão

Rodrigo Mattos Scavassin

*Programa de Pós-Graduação em Computação Aplicada
Universidade Presbiteriana Mackenzie
São Paulo, Brasil*

Leandro Augusto da Silva

*Programa de Pós-Graduação em Computação Aplicada
Universidade Presbiteriana Mackenzie
São Paulo, Brasil*

Resumo — As Cidades Inteligentes demandam sistemas avançados de monitoramento e eficiência energética, especialmente no consumo de energia elétrica. Nesse contexto, a adaptação de algoritmos de Inteligência Artificial para análise contínua de dados (*streaming*), em que decisões rápidas são necessárias, pode evitar desperdícios e falhas operacionais. Este artigo propõe uma abordagem baseada em ciência de dados para monitoramento inteligente do consumo de energia elétrica, com foco na detecção de deriva de modelo usado para previsão de consumo. O caso em que o estudo foi aplicado se trata de uma base histórica de dez anos de consumo de uma Universidade. Avaliou-se na proposta o uso de Árvore de Decisão para previsão de consumo, com o desempenho medido pelo RMSE. A detecção de anomalias e deriva foi realizada a partir do erro preditivo com o método ADWIN (*Adaptive Windowing*). O modelo foi retreinado automaticamente com os dados do último ano apenas quando múltiplas detecções de deriva ocorriam em curto intervalo de tempo, evitando reações excessivas a variações normais. Os resultados demonstraram que a aplicação de retreinamento após detecção de deriva melhorou em 56% em relação ao modelo sem retreinamento enquanto a estratégia de desconsiderar anomalias apresentou uma melhoria de 50% sendo que o número de retreinamentos foi 88% menor. Também foi possível verificar que não houve melhora visível ao segmentar os dados por mês e por dia da semana.

Palavras-chave — aprendizagem de máquina, consumo de energia, deriva de modelo, detectores de deriva.

I. INTRODUÇÃO

A energia elétrica é um dos pilares fundamentais para o funcionamento das sociedades modernas. Presente em praticamente todas as atividades humanas, desde o uso doméstico até os processos industriais mais complexos, seu consumo está diretamente relacionado ao desenvolvimento econômico e ao bem-estar social [1].

No contexto empresarial, o consumo de energia elétrica representa não apenas um insumo essencial, mas também um dos principais componentes do custo operacional. Empresas dos mais diversos setores dependem de um fornecimento estável e eficiente de energia para manter sua produtividade, competitividade e sustentabilidade. Em muitos casos, variações no consumo podem impactar diretamente a margem de lucro e a estratégia de crescimento do negócio [2].

Do ponto de vista macroeconômico, o consumo de energia pode ser percebido como indicador de atividade econômica. Em

períodos de expansão, é comum observar um aumento na demanda energética; já em momentos de recessão, essa demanda tende a cair. Além disso, políticas energéticas eficazes podem contribuir significativamente para a redução de custos e melhoria da eficiência [1].

Com o avanço das tecnologias de medição, armazenamento e análise de dados, tornou-se possível monitorar o consumo de energia com maior granularidade e precisão. Isso abre oportunidades para o desenvolvimento de modelos inteligentes que auxiliem na detecção de padrões de uso, anomalias e mudanças de comportamento, contribuindo para a tomada de decisão mais estratégica por parte das empresas.

Nesse cenário, ferramentas baseadas em ciência de dados tornam-se cada vez mais relevantes para otimizar o uso de energia, prever variações de consumo e identificar desvios significativos, tanto operacionais quanto comportamentais. Essas abordagens não apenas apoiam a gestão mais eficiente dos recursos, como também proporcionam vantagem competitiva às organizações que as adotam [3].

Apesar dos avanços tecnológicos na medição e monitoramento do consumo de energia elétrica, muitas organizações ainda enfrentam desafios significativos para interpretar os dados de maneira eficiente e extrair informações que realmente apoiem a tomada de decisão. Uma das principais dificuldades está na identificação de mudanças relevantes no padrão de consumo, especialmente quando essas mudanças ocorrem de forma sutil ou gradual [4].

Essas alterações podem se manifestar de duas formas principais: anomalias e deriva de conceito. As anomalias são eventos isolados, caracterizados por desvios bruscos e inesperados em relação ao comportamento habitual da série temporal de consumo [5]. Elas podem indicar falhas operacionais, desperdícios, uso indevido de equipamentos ou até erros nos sistemas de medição. Embora pontuais, essas ocorrências podem representar desperdício de recursos e aumento de custos se não forem rapidamente detectadas e corrigidas.

A deriva de conceito (*concept drift*) refere-se a mudanças graduais, mas significativas, na distribuição estatística dos dados ao longo do tempo. Ela pode ser classificada em diferentes tipos: súbita (mudança abrupta), incremental (mudança que acontece aos poucos), recorrente (padrões antigos que reaparecem) e gradual (transição contínua entre dois conceitos) [6]. Esse tipo de alteração pode ser decorrente

de transformações estruturais no funcionamento da empresa, mudanças no perfil de consumo, sazonalidades, ou adoção de novas tecnologias. O maior desafio nesse caso é que os modelos preditivos utilizados tendem a perder acurácia à medida que o comportamento dos dados se altera, sem que essa perda de desempenho seja prontamente identificada [7].

A ausência de mecanismos eficazes para monitorar e distinguir entre essas mudanças compromete a confiabilidade dos modelos analíticos, dificulta a gestão eficiente do consumo energético e pode levar a decisões baseadas em informações desatualizadas ou distorcidas. Assim, torna-se fundamental o desenvolvimento de abordagens que permitam detectar automaticamente anomalias e indícios de deriva nos dados.

O presente trabalho tem como objetivo desenvolver um método analítico baseado em ciência de dados para monitorar o consumo de energia elétrica de uma Universidade, com foco na detecção automática de anomalias e deriva de modelo. A proposta visa apoiar a tomada de decisão gerencial, contribuindo para a otimização do uso energético e a redução de custos operacionais, por meio de análise de séries temporais e aprendizado de máquina.

O artigo foi estruturado da seguinte forma: a Seção 2 traz uma explicação dos principais conceitos utilizados, a Seção 3 descreve a metodologia de como os dados foram processados e modelados, Seção 4 apresenta os resultados obtidos e a Seção 5 resume as conclusões e propõe trabalhos futuros.

II. FUNDAMENTAÇÃO TEÓRICA

O crescimento do volume de dados gerados por sistemas de medição inteligente impulsionou o uso de técnicas de ciência de dados na gestão do consumo de energia elétrica. A literatura mostra que a análise de séries temporais é uma das abordagens mais utilizadas para modelar o comportamento de consumo ao longo do tempo, especialmente pela capacidade de capturar tendências, sazonalidades e padrões repetitivos. A aplicação de modelos preditivos permite estimar o consumo futuro com boa precisão, sendo os métodos baseados em árvores de decisão, ensemble e redes neurais amplamente empregados por sua flexibilidade e robustez [8].

Foram considerados para a previsão de consumo modelos como Árvore de decisão, Random Forest, Gradient Boosting e Redes Neurais. A árvore de decisão foi escolhida, pois apresenta uma estrutura com menos parâmetros e de fácil retreinamento, fator crítico em ambientes onde o modelo precisa ser atualizado com frequência. Além disso, em relação às Redes Neurais, a Árvore de Decisão apresenta maior transparência na tomada de decisão [9], facilitando a explicação dos resultados para gestores e usuários não técnicos, o que é desejável no atual contexto. Embora modelos mais sofisticados possam alcançar maior acurácia em alguns cenários, a Árvore de Decisão representa um bom equilíbrio entre desempenho, interpretabilidade e agilidade na adaptação a novos padrões de dados.

Segundo [10], a árvore de decisão é um algoritmo de aprendizado de máquina usado para tarefas de classificação e regressão, por sua simplicidade e facilidade de interpretação.

Ela funciona como um fluxograma: em cada nó de decisão, é feita uma pergunta baseada em alguma variável dos dados e as respostas conduzem para novos nós ou folhas, que contêm o valor de saída final (número previsto). O algoritmo constrói a árvore buscando dividir os dados de maneira que as partições fiquem o mais “puras” possível, ou seja, com dados semelhantes entre si, visando minimizar a diferença entre os valores previstos e esperados. A construção da árvore é recursiva, dividindo os dados até atingir um critério de parada (como profundidade máxima, número mínimo de amostras ou pureza).

Entretanto, modelos preditivos tradicionais assumem que os dados seguem uma distribuição estável, o que nem sempre é verdade em ambientes reais. Alterações operacionais, mudanças de perfil de uso e fatores externos, como clima e eventos econômicos, podem causar anomalias ou levar à deriva, fenômeno no qual a distribuição dos dados muda com o tempo, tornando os modelos anteriores obsoletos [11]. Por isso, a detecção automática de mudanças no comportamento do consumo tornou-se um tema de grande relevância, tanto na literatura quanto na prática empresarial.

Para a detecção de anomalias, determinados estudos utilizam abordagens baseadas em teste estatístico e erro preditivo, onde grandes desvios entre o valor observado e o estimado sinalizam comportamento atípico [5]. A anomalia é detectada através da avaliação do quão bem um novo ponto de dados se ajusta ao modelo de regressão que descreve o comportamento esperado, examinando a diferença entre os valores observados e os previstos pelo modelo. Grandes diferenças podem indicar anomalias.

Diversos métodos têm sido propostos para detecção de deriva em fluxos de dados, entre eles o Page-Hinkley Test (PH), o Hoeffding Drift Detection Method (HDDM) e o Adaptive Windowing (ADWIN). O PH é um teste estatístico baseado na detecção de mudanças na média de uma variável monitorada, sendo particularmente eficaz para mudanças abruptas e de baixa complexidade computacional [12]. Por outro lado, o HDDM busca detectar desvios significativos entre distribuições de erro em janelas fixas, sendo adequado para detectar tanto deriva abrupta quanto gradual [13]. Já o ADWIN mantém uma janela de tamanho variável e realiza testes estatísticos para detectar alterações significativas entre duas sub-janelas, ajustando-se dinamicamente ao ritmo de mudança dos dados, o que o torna eficaz mesmo em ambientes não estacionários [13].

Foi adotado o ADWIN por sua flexibilidade e robustez em ambientes não estacionários. Diferentemente do Page-Hinkley, que utiliza uma abordagem acumulativa com parâmetros fixos e é mais sensível a mudanças abruptas [12], o ADWIN é capaz de lidar tanto com derivas súbitas quanto graduais por meio do ajuste dinâmico do tamanho da janela de observação. Esse mecanismo adaptativo permite que o ADWIN mantenha o equilíbrio entre sensibilidade a mudanças e estabilidade em relação a ruídos. Em comparação ao HDDM, que requer a definição prévia de janelas fixas, o ADWIN elimina a necessidade de configuração manual do tamanho da janela, reduzindo o risco de erros de detecção causados por parametrizações inadequadas.

O critério para detecção do ADWIN foi o Erro Quadrático Médio (Root Mean Squared Error, ou RMSE). Esta é uma das métricas mais utilizadas para avaliar a precisão de um modelo de regressão. Ele mede a raiz quadrada da média dos quadrados dos erros entre os valores previstos pelo modelo e os valores reais. O RMSE indica a magnitude média dos erros do modelo, com um peso maior para erros maiores devido ao termo ao quadrado [14]. Um valor de RMSE mais baixo sugere que o modelo tem um desempenho melhor, indicando que as previsões estão mais próximas dos valores observados.

No contexto da energia elétrica, estudos recentes exploram o uso de técnicas de previsão e detecção de anomalias para melhorar a eficiência energética, identificar falhas operacionais e fornecer alertas em tempo real para gestores [15]. Reforçando também que, além de prever o consumo, tais modelos podem ser integrados a sistemas de monitoramento inteligente que identificam desvios e ajudam a diagnosticar comportamentos anormais, promovendo maior controle sobre custos e eficiência no uso de energia.

A etapa de pré-processamento das séries temporais desempenha um papel fundamental na melhoria do desempenho de modelos preditivos. (Coimbra et al., 2023) [16] realizou-se uma análise comparativa entre diferentes métodos de remoção de tendência e dessazonalização aplicados a séries históricas de consumo, avaliando seu impacto na acurácia de modelos. Os resultados indicam que a combinação de diferenciação para tratar a tendência e normalização para remoção da sazonalidade produziu os menores erros de previsão. Os autores também demonstraram que o uso isolado de padronizações simples, sem tratar a estrutura temporal dos dados, é insuficiente para garantir boa performance preditiva. Esses achados reforçam a importância de aplicar técnicas adequadas de transformação de séries temporais antes do treinamento dos modelos, especialmente em domínios onde os padrões de consumo são fortemente influenciados por variações sazonais e tendências de longo prazo.

O presente trabalho integra conceitos de modelagem de aprendizagem de máquina, análise de séries temporais, detecção de anomalias e detecção de deriva para propor uma solução prática de monitoramento inteligente do consumo de energia elétrica.

III. METODOLOGIA

A primeira etapa consiste no pré-processamento da base de dados, que contém registros mensais de consumo de energia elétrica ao longo de dez anos. Os registros foram levantados através das contas de consumo da Universidade pela concessionária de energia. Para viabilizar a aplicação de técnicas baseadas em séries temporais de alta granularidade, os dados mensais foram extrapolados para obter estimativas horárias de consumo. Essa extrapolação foi realizada com base em perfis típicos de carga, fazendo a diferenciação entre os dias da semana e variações sazonais utilizando uma distribuição proporcional. A distribuição diária e semanal é apresentada nas figuras 1 e 2 e a distribuição geral dos dados que foi simulada é apresentada na figura 3.

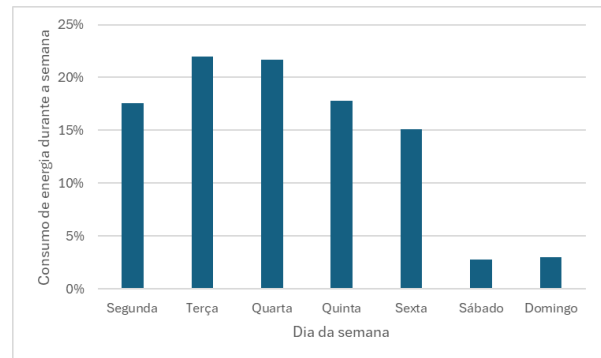


Figura 1. Distribuição do consumo de energia ao longo da semana.

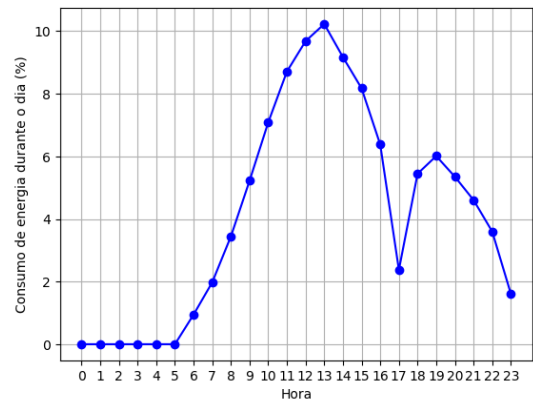


Figura 2. Distribuição do consumo de energia ao longo do dia.

Com a série temporal simulada construída, foi realizada a análise exploratória dos dados, com o objetivo de identificar padrões, tendências, sazonalidades e possíveis outliers. Métricas estatísticas, histogramas, boxplots e mapas de calor foram utilizados para compreender o comportamento do consumo ao longo do tempo e embasar as decisões de modelagem.

Na etapa seguinte, foi aplicado o modelo de árvore de decisão para a previsão de consumo energético, com o intuito de estimar valores esperados e detectar desvios. A cada instante, um vetor de entrada composto pelos 24 valores anteriores de consumo é utilizado para prever o valor da próxima hora. A performance preditiva dos modelos foi avaliada com base na taxa de erro RMSE (Root Mean Squared Error).

Foi implementada uma abordagem para identificação de deriva de modelo, que consiste em mudanças persistentes no padrão estatístico dos dados ao longo do tempo. Para isso, foi utilizado o detector de deriva ADWIN. Além da detecção de deriva, foi aplicada uma heurística na qual se a detecção de deriva ocorrer de forma isolada, deve ser considerada uma anomalia e o modelo não precisa ser retreinado. Para este trabalho foi considerada uma detecção isolada o evento no qual não ocorra mais de uma detecção no período de dez dias. O retreinamento foi feito apenas com os dados dos últimos 360 dias.

A biblioteca River [17] (especializada em aprendizado de máquina incremental) foi utilizada para essa etapa, dada sua compatibilidade com fluxos de dados contínuos e sensibilidade a mudanças no conceito.

Foi avaliado se o desempenho dos modelos de previsão melhora ao segmentar os dados por características temporais, como mês ou dia da semana. Modelos específicos foram treinados para cada segmento, permitindo verificar se a

granularidade temporal influencia a sensibilidade à detecção de desvios. A figura 4 apresenta o diagrama que resume o método aplicado:

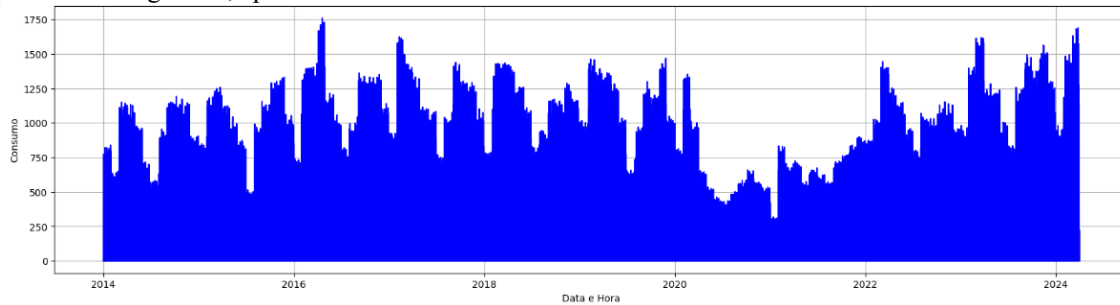


Figura 3. Distribuição do consumo ao longo de dez anos

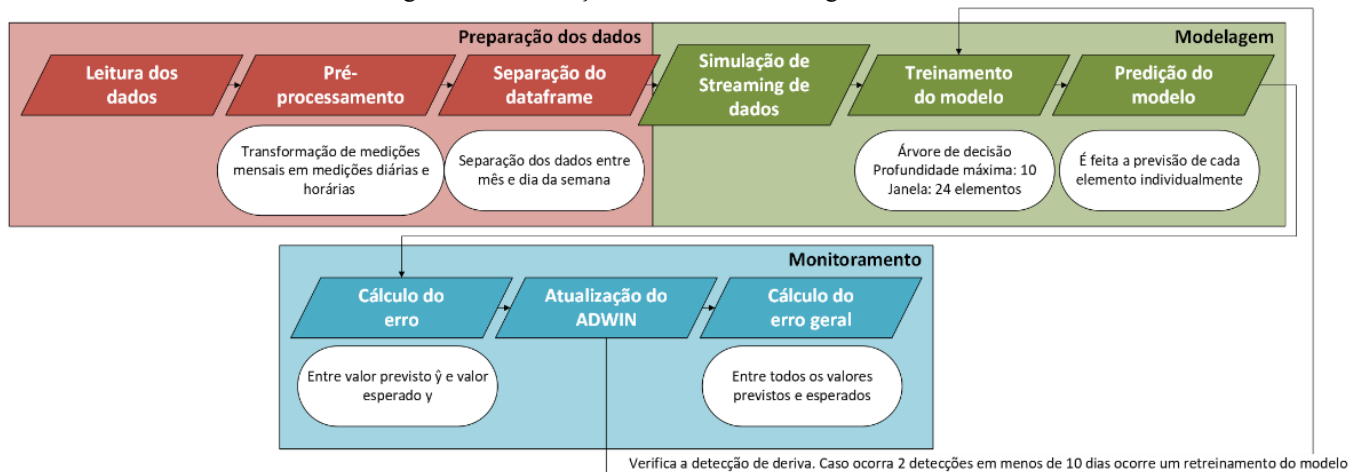


Figura 4. Fluxograma do método aplicado

IV. RESULTADOS

A avaliação dos modelos de previsão de consumo foi conduzida com o objetivo de analisar o impacto da detecção de deriva de modelo e do tratamento de anomalias no desempenho preditivo. Para isso, foram testados três cenários distintos: (tabela I) sem retreinamento, (tabela II) com retreinamento ao detectar deriva por ADWIN, e (tabela III) com retreinamento apenas quando mais de duas detecções de deriva ocorriam em um intervalo inferior a dez dias (tratando as demais como anomalias). Em cada cenário, compararam-se três abordagens: utilizando todo o conjunto de dados, com separação por mês, e com separação por dia da semana. Os modelos foram avaliados com base na taxa de erro RMSE, além do número de derivações e anomalias detectadas. As figuras 5, 6 e 7 apresentam o gráfico com o conjunto de dados completo, com os dados de um único dia da semana e com os dados de um único mês respectivamente. A figura 8 resume as tabelas, apresentando a média do RMSE no caso do cenário em que os dados estão separados por dia da semana e mês.

TABELA I: Métricas de desempenho para o cenário sem retreinamento do modelo

DADOS UTILIZADOS	RMSE	DRIFT
TODO O CONJUNTO DE DADOS	71,7451	253
SEPARADO POR MÊS	88,1940	217
SEPARADO POR DIA DA SEMANA	61,1483	282

MÊS	RMSE	DRIFT
JANEIRO	56,4062	15
FEVEREIRO	150,2560	15
MARÇO	117,7499	19
ABRIL	102,3305	17
MAIO	82,1628	33
JUNHO	96,8343	18
JULHO	61,4100	16
AGOSTO	67,5433	14
SETEMBRO	81,0188	20
OUTUBRO	73,3145	12
NOVEMBRO	98,3068	20
DEZEMBRO	70,9945	18
DIA DA SEMANA	RMSE	DRIFT
SEGUNDA-FEIRA	77,5000	44
TERÇA-FEIRA	101,4854	43
QUARTA-FEIRA	88,3323	39
QUINTA-FEIRA	72,4147	43
SEXTA-FEIRA	61,3685	36
SÁBADO	11,8792	41
DOMINGO	15,0580	36

TABELA II: Métricas de desempenho para o cenário com retreinamento após detecção de deriva

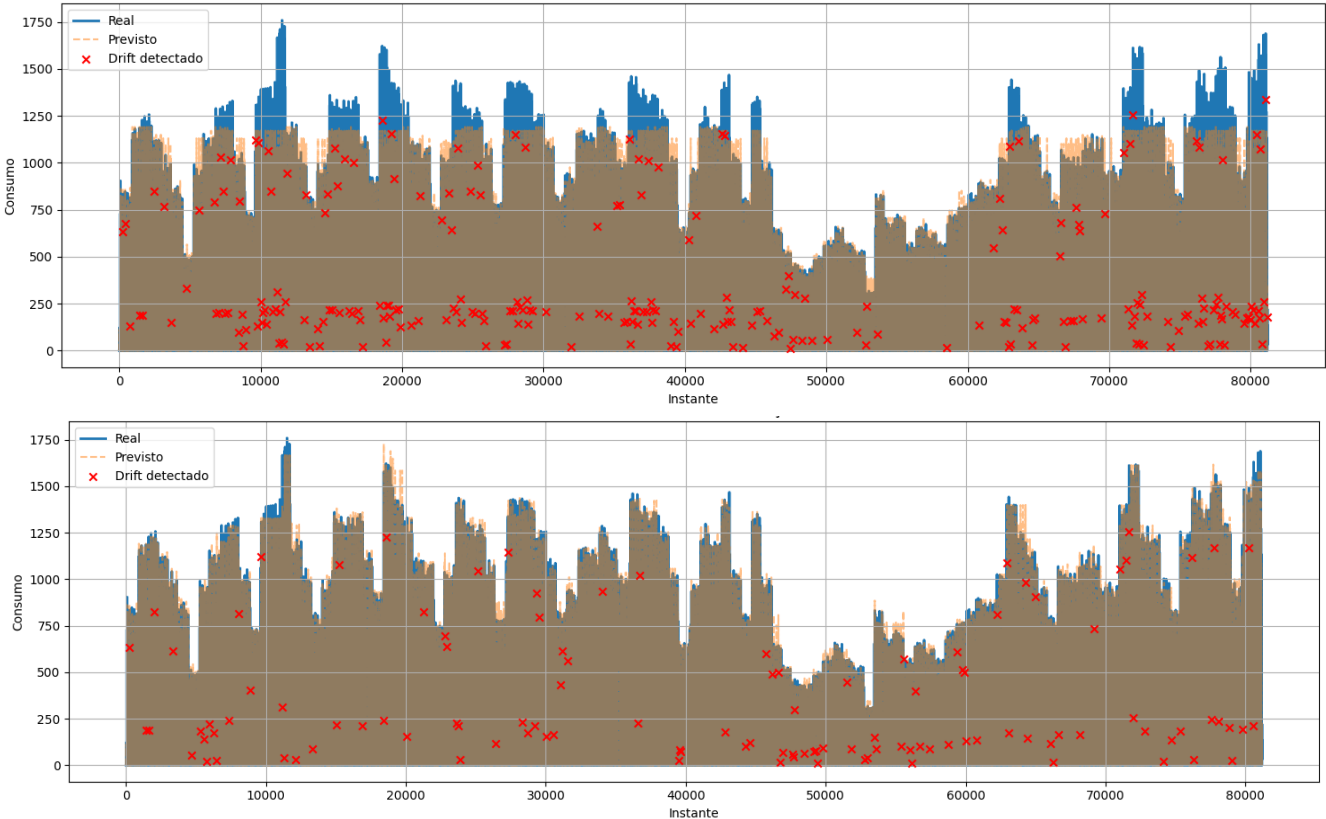
DADOS UTILIZADOS	RMSE	DRIFT
TODO O CONJUNTO DE DADOS	31,3496	115
SEPARADO POR MÊS	70,0163	82

SEPARADO POR DIA DA SEMANA	46,6376	242
MÊS	RMSE	DRIFT
JANEIRO	49,7422	6
FEVEREIRO	80,2041	5
MARÇO	114,6504	5
ABRIL	89,8758	6
MAIO	64,3668	7
JUNHO	61,9063	7
JULHO	40,8002	8
AGOSTO	63,0334	8
SETEMBRO	62,7652	11
OUTUBRO	72,7172	5
NOVEMBRO	82,7849	7
DEZEMBRO	57,3493	7
DIA DA SEMANA	RMSE	DRIFT
SEGUNDA-FEIRA	58,2512	32
TERÇA-FEIRA	70,8234	34
QUARTA-FEIRA	70,5713	27
QUINTA-FEIRA	58,6209	37
SEXTA-FEIRA	49,9860	37
SÁBADO	8,6417	39
DOMINGO	9,5690	36

TABELA III: Métricas de desempenho para o cenário com treinamento após detecção de deriva desconsiderando anomalias

DADOS UTILIZADOS	RMSE	DRIFT	ANOMALIAS
------------------	------	-------	-----------

TODO O CONJUNTO DE DADOS	35,5088	13	174
SEPARADO POR MÊS	91,4141	1	131
SEPARADO POR DIA DA SEMANA	60,8566	16	211
MÊS	RMSE	DRIFT	ANOMALIAS
JANEIRO	56,4062	0	9
FEVEREIRO	150,2560	0	11
MARÇO	117,7499	0	14
ABRIL	102,3305	0	8
MAIO	82,1628	0	14
JUNHO	96,8343	0	11
JULHO	61,4100	0	8
AGOSTO	106,1846	1	15
SETEMBRO	81,0188	0	11
OUTUBRO	73,3145	0	7
NOVEMBRO	98,3068	0	7
DEZEMBRO	70,9945	0	16
DIA DA SEMANA	RMSE	DRIFT	ANOMALIAS
SEGUNDA-FEIRA	77,9188	8	31
TERÇA-FEIRA	78,0327	1	30
QUARTA-FEIRA	86,1988	1	26
QUINTA-FEIRA	83,6868	2	31
SEXTA-FEIRA	76,6988	2	35
SÁBADO	11,8792	0	32
DOMINGO	11,5807	2	26



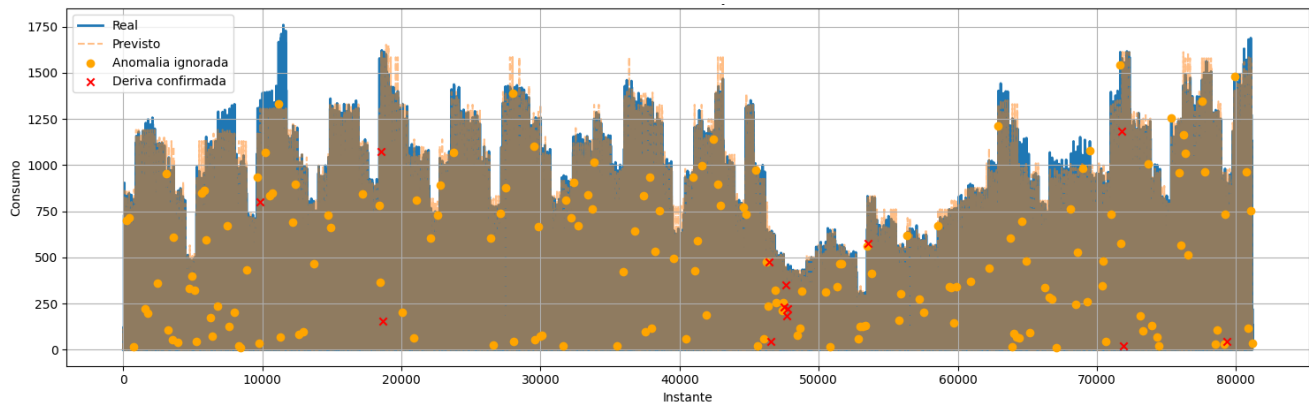
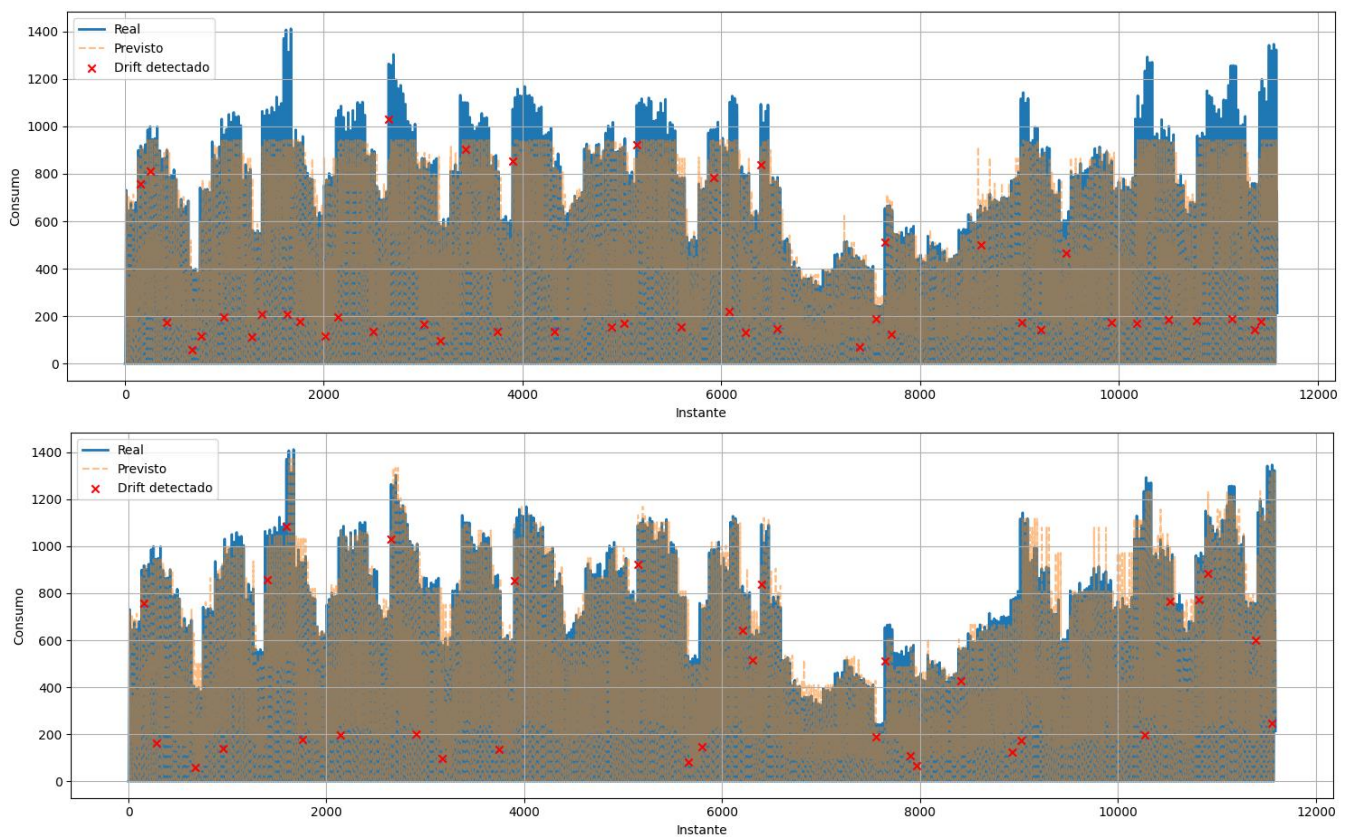


Figura 5. Detecção de deriva em todo o conjunto de dados sem retreinamento (cima) com retreinamento na detecção de deriva (meio) e desconsiderando anomalias (baixo)

A Figura 5 apresenta a comparação entre três estratégias distintas de tratamento de deriva em séries temporais de consumo de energia elétrica, considerando o conjunto de dados completo. No primeiro cenário (gráfico superior), o modelo demonstrou elevada divergência entre os valores previstos e os reais, bem como em uma quantidade significativa de derivas detectadas. No segundo cenário (gráfico central), observou-se

uma melhora na aderência das previsões ao consumo real e uma redução no número de detecções, evidenciando que o retreinamento contribuiu para a adaptação do modelo às mudanças. Por fim, o terceiro cenário (gráfico inferior) resultou em menor quantidade de derivas confirmadas permitindo que o modelo fosse retreinado apenas diante de mudanças significativas.



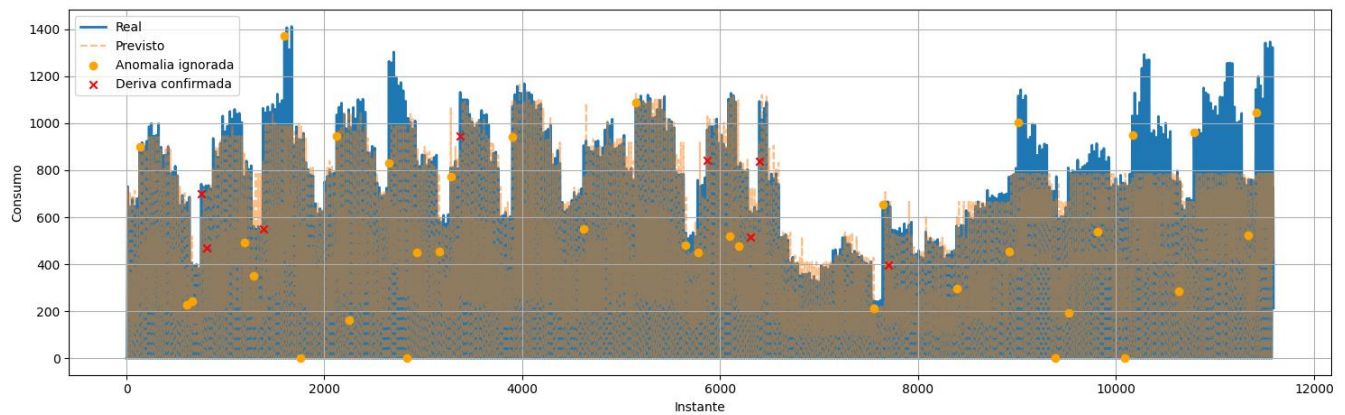
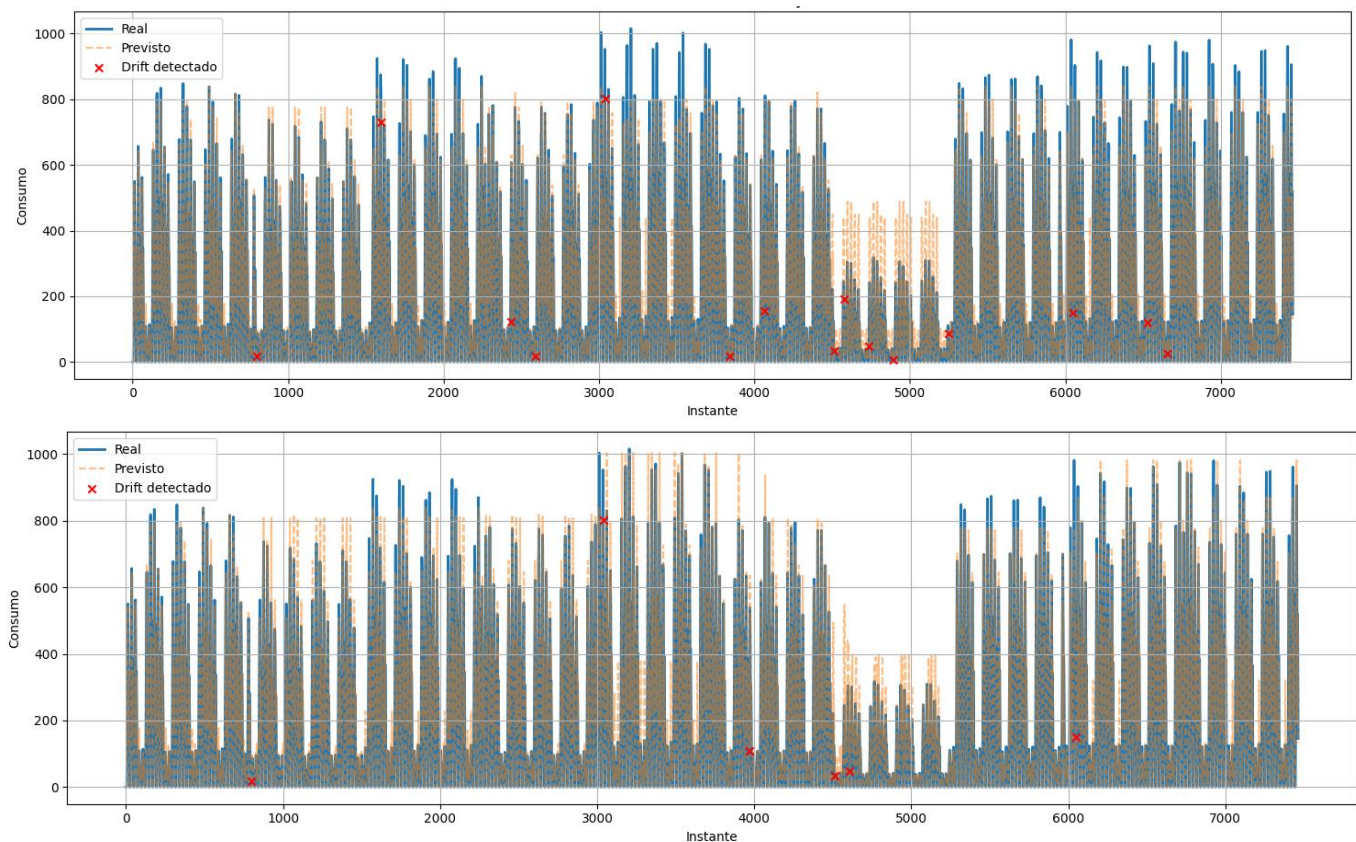


Figura 6. Detecção de deriva com dados apenas da segunda-feira sem retreinamento (cima) com retreinamento na detecção de deriva (meio) e desconsiderando anomalias (baixo)

A Figura 6 ilustra os resultados da detecção de deriva considerando exclusivamente os dados das segundas-feiras, o que permite avaliar o desempenho do modelo em um subconjunto com padrão semanal recorrente. No primeiro cenário, observa-se que a ausência de retreinamento levou a uma degradação da acurácia, além de frequentes detecções de deriva. Isso evidencia a limitação do modelo em se manter adaptado mesmo sob um regime de dados mais homogêneo. No

segundo cenário, a qualidade das previsões melhora, com menor incidência de deriva. Já o terceiro cenário apresenta um desempenho ainda mais estável: o modelo retreina apenas quando ocorrem mudanças estruturais reais no padrão de consumo, evitando atualizações desnecessárias causadas por ruídos pontuais. Entretanto, falhou em identificar uma necessidade de retreinar o modelo em algumas instâncias.



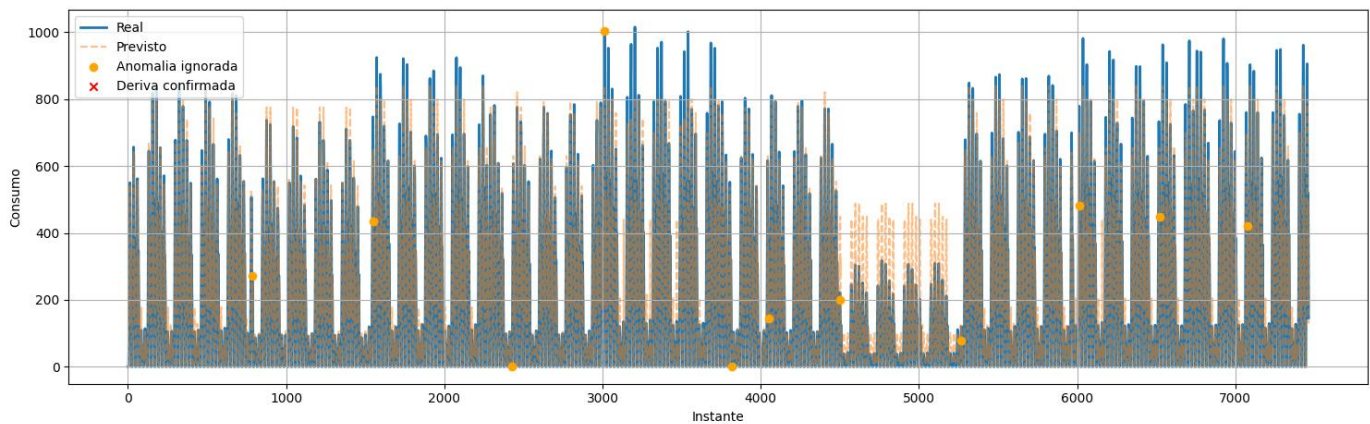


Figura 7. Detecção de deriva com dados apenas de janeiro sem retreinamento (cima) com retreinamento na detecção de deriva (meio) e desconsiderando anomalias (baixo)

A Figura 7 apresenta a detecção de deriva aplicada ao subconjunto de dados referente apenas ao mês de janeiro, caracterizado por um comportamento mais regular e com padrão de consumo bem definido. No primeiro cenário, é possível observar uma degradação da acurácia preditiva, evidenciada pelo descolamento entre os valores previstos e os reais, sobretudo no intervalo entre os instantes 4000 e 5000. No segundo cenário, observa-se uma melhora nas previsões, com menor erro em relação aos valores reais e maior estabilidade ao longo do tempo, o que mostra a eficácia do retreinamento para restaurar a acurácia após alterações nos padrões. Por fim, no terceiro cenário, não foi detectada nenhuma deriva, sendo todos os pontos de alerta entendidos apenas como anomalias, levando a previsão se comportar exatamente igual em relação ao primeiro cenário.

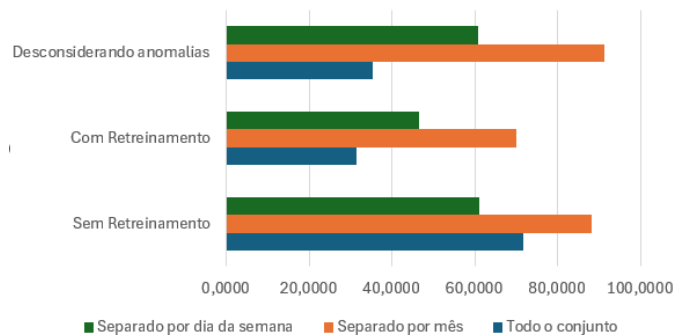


Figura 8. Comparação do RMSE entre os três cenários (quanto menor o valor, melhor)

A análise dos resultados indica que a aplicação de retreinamento após detecção de deriva melhora significativamente o desempenho do modelo.

Nos exemplos apresentados o RMSE diminuiu em aproximadamente 56% para o conjunto de dados completo, 20% para o conjunto de dados separado por mês e 23% para o conjunto de dados separado por dias da semana, ao custo de fazer vários retreinamentos durante o processo. Já com a desconsideração de anomalias, o RMSE diminuiu em aproximadamente 50% em relação ao conjunto de dados completo sem retreinamento. Apesar de a queda do erro ter sido menor neste cenário, o número de retreinamentos também caiu

88%, demonstrando a eficácia da filtragem de anomalias, que permite reduzir o número de retreinamentos desnecessários sem comprometer a acurácia do modelo. Entretanto vale notar que os resultados não foram positivos em relação aos modelos feitos para os conjuntos de dados separados por mês e por dia da semana.

Outra conclusão que pode ser feita é que separar conjunto de dados em conjuntos menores não foi uma estratégia que trouxe resultados positivos neste cenário. Considerando o retreinamento a cada nova detecção de deriva, o modelo treinado com o conjunto de dados completo apresentou RMSE 55% menor em relação à média do modelo treinado com o conjunto de dados separado por mês e 32% menor em relação à média do modelo treinado com o conjunto de dados separado por dia da semana.

V. CONCLUSÃO E TRABALHOS FUTUROS

A crescente complexidade dos padrões de consumo de energia elétrica impõe desafios significativos à previsão precisa e à detecção precoce de comportamentos anômalos. Este trabalho propôs uma abordagem baseada em aprendizado de máquina e monitoramento adaptativo para identificar anomalias e deriva de modelo em séries temporais de consumo energético. O retreinamento de modelos conforme sua deriva é detectada, combinada com técnicas de pré-processamento, mostrou-se promissora para capturar variações relevantes nos dados. Além disso, a utilização de métodos de detecção de anomalias permitiu implementar um sistema de atualização dinâmica dos modelos, reduzindo retreinamentos desnecessários, melhorando a performance geral. Espera-se que este trabalho contribua para uma gestão energética mais eficiente e baseada em dados. Futuros trabalhos podem explorar outras técnicas de detecção de deriva, diferentes modelos e a combinação com dados externos (como temperatura e atividade econômica), aumentando o número de atributos e a complexidade do conjunto de dados. Também é necessário estudar novas estratégias para se adequar a sazonalidade e tendência dos dados.

Referências

- [1] Barbosa, P. A. M., & Mattos, L. B. de. (2021). Relação entre consumo de energia elétrica e crescimento econômico no Brasil: uma análise a nível setorial. *Anais do Encontro Nacional de Economia*
- [2] MACHADO, Cícero Ferreira. Relacionamento entre produtividade e eficiência energética, em uma empresa gráfica. 2017. 91 f. Dissertação (Mestrado em Engenharia de Produção), Universidade Federal da Paraíba, João Pessoa, 2017.
- [3] Hasan Yesilyurt, Yesim Dokuz, Ahmet Sakir Dokuz, (2024). Data-driven energy consumption prediction of a university office building using machine learning algorithms. <https://doi.org/10.1016/j.energy.2024.133242>
- [4] Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., & Amira, A. (2020). Artificial Intelligence based Anomaly Detection of Energy Consumption in Buildings: A Review, Current Trends and New Perspectives <https://doi.org/10.48550/arXiv.2010.04560>
- [5] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 15. <https://doi.org/10.1145/1541880.1541882>
- [6] Arora, S., Rani, R., & Saxena, N. (2024). A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques. *WIREs Data Mining and Knowledge Discovery*, 14(4), e1536. <https://doi.org/10.1002/widm.1536>
- [7] Supriya Agrahari, Anil Kumar Singh. (2022). Concept Drift Detection in Data Stream Mining: A literature review. <https://doi.org/10.1016/j.jksuci.2021.11.006>
- [8] Zhang, Y., Wang, J., Wang, X., & Zhao, Y. (2018). Short-term electric load forecasting based on multivariate LSTM. *Energy*, 166, 893–903. <https://doi.org/10.1016/j.energy.2018.10.148>
- [9] Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: From efficient prediction to responsible AI. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.1124553>
- [10] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," in *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838
- [11] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 44. <https://doi.org/10.1145/2523813>
- [12] Page, E.S. (1954) Continuous Inspection Schemes. *Biometrika*, 41, 100–115. [dx.doi.org/10.1093/biomet/41.1-2.100](https://doi.org/10.1093/biomet/41.1-2.100)
- [13] Bifet, A., & Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining* (pp. 443–448). <https://doi.org/10.1137/1.9781611972771.42>
- [14] Hodson, T. O.: Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not (2022), *Geosci. Model Dev.*, 15, 5481–5487, <https://doi.org/10.5194/gmd-15-5481-2022>
- [15] Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2018). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74, 902–924. <https://doi.org/10.1016/j.rser.2017.02.085>
- [16] Coimbra, T., Boccato, L., Siqueira, H. V., & Attux, R. (2023). Comparative Analysis of Deseasonalization and Detrending Methods in Energy Consumption Forecasting. In *Anais do Congresso Brasileiro de Inteligência Computacional (CBIC 2023)*. <https://doi.org/10.21528/CBIC2023-105>
- [17] Montiel, J., Read, J., Bifet, A., & Abdessalem, T. (2021). River: machine learning for streaming data in Python. *Journal of Machine Learning Research*, 22(123), 1–8.