

Metodologia para Identificação de Perdas Não Técnicas em Unidades Consumidoras localizadas em Áreas Rurais com Cultivo de Arroz Irrigado

Henrique S. Eichkoff*. Daniel P. Bernardon*. Natália B. Souza*. Pedro Marcolin*.
Luciana M. Kopp**. Juliano S. Andrade***. Lucas M. Chiara***

*Universidade Federal de Santa Maria (UFSM)
Santa Maria, Brasil (e-mail: henriquekoff@gmail.com; dpbernardon@ufsm.br; rhyo_gabaglia@hotmail.com;
pedro_macolin@hotmail.com)

**Universidade Federal de Pelotas (UFPEL)
Pelotas, Brasil (e-mail: lucianakopp@gmail.com)

*** Companhia Paulista Força e Luz (CPFL Energia)
Campinas, Brasil, (e-mail: julianoandrade@cpfl.com.br; lucaschiara@cpfl.com.br)

Abstract: This work presents a proposal for the identification of non-technical losses in rural consumers directed to irrigated rice cultivation. The methodology employs in the structure of its algorithm, the k-Means and Random Forest methods, for the clustering and classification of data, using as inputs, monthly records of billed energy from a group of irrigating consumers in the county of Uruguaiana/RS. In addition, the algorithm proposed by this study performs the estimate of electricity consumption of rural consumers based on information on crop planting areas and variables of irrigation systems.

Resumo: Este trabalho apresenta uma proposta para a identificação de perdas não técnicas em consumidores rurais direcionados para o cultivo do arroz irrigado. A metodologia emprega na estrutura de seu algoritmo, os métodos de *k-Means* e *Random Forest*, para o agrupamento e classificação de dados, utilizando como entradas, registros mensais de energia faturada de um conjunto de consumidores irrigantes do município de Uruguaiana/RS. Além disso, o algoritmo proposto por esse estudo, realiza a estimativa de consumo de energia elétrica das unidades consumidoras rurais a partir de informações sobre as áreas de plantio de arroz e variáveis dos sistemas de irrigação.

Keywords: Irrigation Systems; Rural Consumers; k-Means; Non-Technical Losses; Random Forest.

Palavras-chaves: Consumidores Rurais; *k-Means*; Perdas Não Técnicas; *Random Forest*; Sistemas de Irrigação.

1. INTRODUÇÃO

As Perdas Não Técnicas correspondem à diferença entre as perdas globais e as perdas técnicas de um sistema elétrico, e ocorrem quando uma determinada quantidade de energia elétrica não é contabilizada pelos sistemas de medições nos processos de faturamento, sendo essencialmente um problema de receita das empresas de distribuição de energia elétrica (Barros et al. 2021). As perdas não técnicas podem ser causadas por (Medeiros et al. 2021): falhas (problemas técnicos em medidores ou na infraestrutura de medição); furto (conexão clandestina à rede de distribuição); fraude (adulteração e alterações irregulares em medidores) e ataques cibernéticos (invasões na rede de telecomunicações da concessionária para alterar os registros de medições).

A presença de perdas não técnicas é uma situação recorrente em sistemas de distribuição localizados em áreas rurais, sendo um problema extremamente significativo para as concessionárias de energia elétrica. O processo de identificação das mesmas neste âmbito é complexo, devido à presença de unidades consumidoras em locais remotos ou de difícil acesso e também pelas grandes extensões de redes presentes nos alimentadores rurais, fatores que acabam

dificultando a investigação de irregularidades por equipes de peritos.

O procedimento mais elementar para identificar perdas não técnicas é a partir do método por inspeção local. Em alguns países, as inspeções manuais são programadas durante um período específico para detectar atividades ilícitas de furto de energia elétrica. No entanto, vistoriar presencialmente cada unidade consumidora em um alimentador, não é visto como uma solução realista para as empresas do setor de energia elétrica (Ghori et al. 2020).

No âmbito rural, alguns fatores dificultam a prática da fiscalização local, como os elevados custos com equipes de peritos, tempo necessário para investigar extensas áreas e as próprias adversidades no processo simples de inspeção, pois em muitos casos, as irregularidades não são perceptíveis a uma simples inspeção visual (Evaldt 2018). Desse modo, o método de inspeção local pode apresentar respostas imprecisas para o problema de identificação de perdas não técnicas em alimentadores rurais, se não for realizado o mapeamento correto de todas as variáveis que envolvem esse processo.

Diante de algumas adversidades apresentadas, metodologias

utilizando técnicas de inteligência artificial orientada a dados vem sendo desenvolvidas com o intuito de auxiliar as distribuidoras de energia elétrica no processo de identificação de perdas não técnicas, restringindo as regiões de busca e indicando os consumidores suspeitos de irregularidades. Em Evaldt (2018) foi proposto um estudo de identificação de perdas não técnicas em consumidores rurais em lavouras de arroz irrigado, empregando um Sistema Neural constituídos de Redes Neurais Artificiais (RNAs) que utilizam como entradas, as correlações dos padrões de consumo de energia elétrica, características das áreas irrigadas e condições climáticas da região durante a época de irrigação.

Além dos RNAs, outras técnicas orientadas a dados podem sem utilizadas no problema de identificação de perdas não técnicas em alimentadores rurais. A aplicação de metodologias empregando métodos de agrupamento e classificação de dados, podem apresentar respostas eficientes e simples para esse problema, utilizando como dados de entrada, informações de consumo de energia elétrica de um conjunto de unidades consumidoras de uma determinada região. Um exemplo de aplicação seria através dos algoritmos de *Clustering (k-Means, Fuzzy cMeans e Density Based Spatial Clustering of Applications with Noise (DBSCAN))*, *Decision Trees*, *Random Forest*, *Nearest neighbor (k-NN)* e *Support Vector Machine (SVM)* (Messinis et al. 2018).

O estudo do problema de perdas não técnicas em alimentadores localizados em áreas rurais é fundamental para as distribuidoras, pois o consumidor rural direciona o uso da energia elétrica em grande parte para a produção agroindustrial, apresentando um elevado consumo sazonal, sendo assim, extremamente relevantes para os alimentadores rurais. Na Região da Fronteira Oeste do Rio Grande do Sul, a energia elétrica é empregada para o funcionamento das estações de bombeamento utilizadas na distribuição de água nas lavouras de arroz irrigado. Esses sistemas são compostos por bombas hidráulicas, motores de indução trifásicos (com potência ativa nominal variando de 100 a 300 cv), sistema de transmissão mecânica (polias ou correias), tubulações (recalque e sucção), entre outros componentes (Kopp et al. 2016). Dessa forma, a atividade de irrigação representa uma parcela muito significativa no consumo da instalação rural, pois as estações de bombeamento funcionam de maneira contínua em praticamente todo o período em que compreende a safra do arroz irrigado (Pfitscher et al. 2012).

Nesse contexto, esse trabalho tem por objetivo apresentar uma metodologia de identificação de perdas não técnicas em unidades consumidoras localizadas em áreas rurais com cultivo de arroz irrigado, relacionando o consumo de energia elétrica com as características das lavouras e dos sistemas de irrigação. Para isso, serão empregadas técnicas de agrupamento e classificação de dados em um conjunto de consumidores irrigantes reais do município de Uruguaiana, no estado do Rio Grande do Sul.

Para a etapa de agrupamento de dados, será empregado o algoritmo de *k-Means*, que é definido como o cálculo de centroides das amostras de um conjunto, utilizando o número de clusters como entrada e separando os elementos em grupos para minimizar algum critério (Passos Júnior et al. 2016).

Para a etapa de classificação, será utilizado o método de *Random Forest*, que é um algoritmo de classificação e regressão de dados desenvolvido por Breiman (2001), induzido aleatoriamente através de um conjunto de árvores, empregando um determinado número de atributos em cada nó, onde a árvore vota na classe mais popular para definir uma entrada. Esse algoritmo é um complemento direto do método de *Decision Tree* (em português, Árvores de Decisão).

Além disso, como parte da avaliação de possibilidade de ocorrência de perdas não técnicas para a presente metodologia, será realizado o cálculo da estimativa de consumo de energia elétrica para os consumidores rurais selecionados, a partir de variáveis presentes na cultura do arroz irrigado e nos sistemas de irrigação.

Este trabalho está estruturado conforme descrição a seguir. A Seção 2 descreve a metodologia proposta para identificação de perdas não técnicas em unidades consumidoras irrigantes, caracterizando todas os procedimentos e etapas presentes na mesma. A Seção 3 apresenta os estudos de casos e resultados obtidos com a metodologia proposta e a Seção 4, as considerações finais sobre o trabalho.

2. METODOLOGIA PROPOSTA

A metodologia proposta nesse trabalho tem por objetivo identificar possíveis ocorrências de perdas não técnicas em unidades consumidoras rurais, em que o uso da energia elétrica é direcionado para a atividade de irrigação nas lavouras de arroz. A metodologia é estruturada a partir de uma arquitetura composta por métodos matemáticos computacionais, cuja funcionalidade é agrupar e classificar os dados característicos para um determinado conjunto de consumidores, e também por informações relevantes associadas a cultura e as estações de bombeamento de arroz irrigado que venham a influenciar no consumo de energia elétrica da instalação. A Fig. 1 apresenta um fluxograma contendo a arquitetura estrutural da metodologia proposta nesse trabalho e seus respectivos estágios.

A metodologia proposta foi desenvolvida a partir de um algoritmo escrito em linguagem de programação *Python 3.7*, elaborado na plataforma computacional *The Scientific Python Development Environment (Spyder®)*, sendo este, uma ferramenta gratuita de análises em ambiente de desenvolvimento integrado, que possibilita o estudo de recursos de edição de scripts, testes, *debugging* (em português, depuração de dados) e visualização gráfica. A escolha por esse software é consequência de o mesmo tratar-se de uma plataforma com acesso livre e apresentar integração com as principais bibliotecas da linguagem *Python*.

O estágio de aprendizado compreende as etapas de pré-processamento do banco de dados (*dataset*) de unidades consumidoras, e do agrupamento das amostras presentes deste conjunto. A etapa de pré-processamento tem por objetivo identificar unidades consumidoras que apresentem dados de cadastro ausentes ou incorretos, constatar correlações de atributos, extrair as informações relevantes

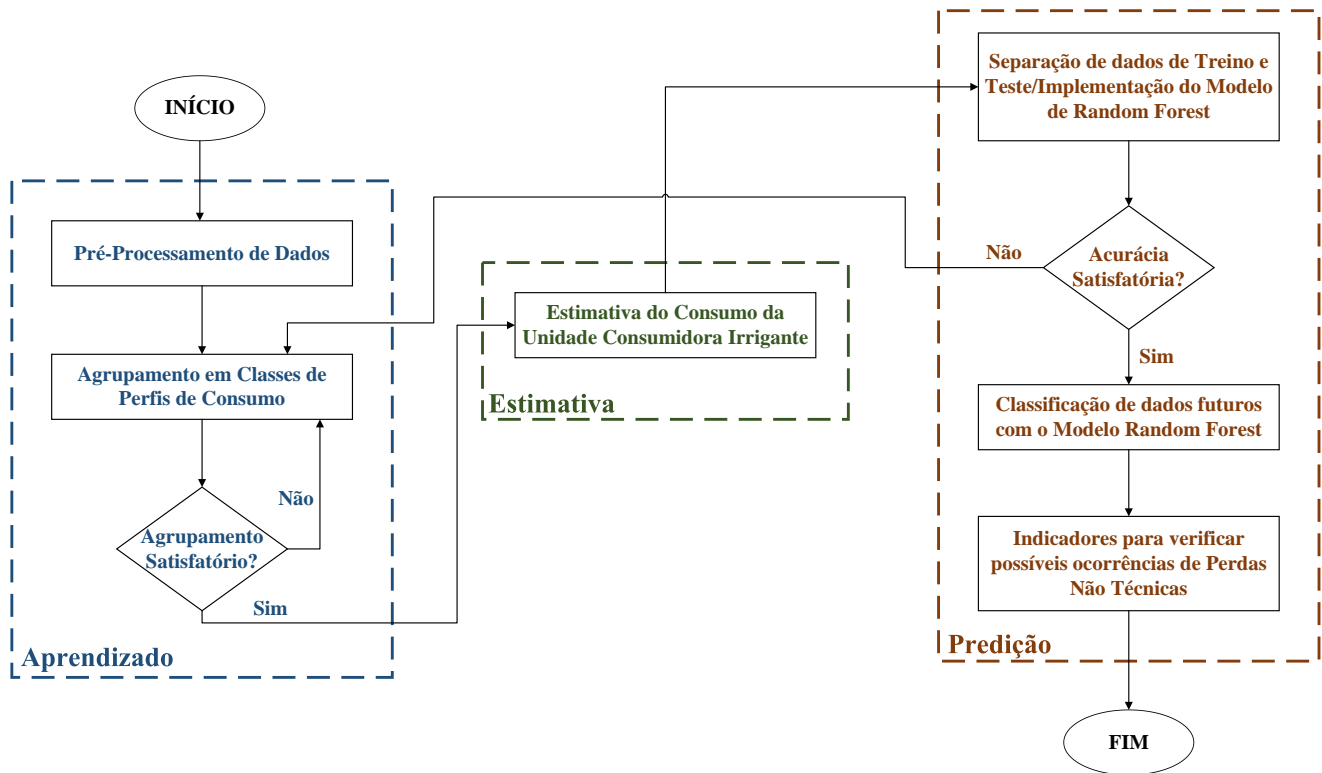


Fig. 1 Fluxograma da metodologia proposta.

para o estudo e aprimorar a compreensão dos elementos presentes no *dataset*. Já a etapa de agrupamento de dados, tem o intuito de criar grupos ou classes de perfil de consumo nas amostras remanescentes da etapa anterior. O algoritmo de *K-Means* foi o método adotado para a realização do agrupamento de dados.

O estágio de estimativa é responsável por realizar a previsão de consumo de energia elétrica das unidades consumidoras irrigantes extraídas nas amostras resultantes pelas etapas anteriores. O mesmo, tem como objetivo comparar e estabelecer um perfil de consumo esperado para cada consumidor, através de diversos atributos presentes nas lavouras de arroz irrigado que podem ser relacionados com o uso da energia elétrica para a atividade de irrigação, tais como: área de plantio e dados específicos das estações de bombeamento da lavoura.

O estágio de predição é composto pelas etapas de separação das amostras de treinos e testes do conjunto de dados obtidos pelas etapas iniciais, implementação de um modelo de classificação utilizando o algoritmo de *Random Forest*, e a verificação da acurácia do modelo. Para apurar a precisão do presente modelo desenvolvido, será empregada a métrica de avaliação realizada pela tabela de confusão de dados, considerando um valor percentual de acurácia pré-definido. Por fim, são propostos indicadores para averiguar possíveis ocorrências de perdas não técnicas a partir dos perfis de consumo presumidos pelo modelo de previsão e por divergências entre o consumo previsto e faturado para o conjunto de unidades consumidoras correspondente.

Nas subseções a seguir, serão detalhadas e brevemente contextualizadas as principais etapas da metodologia.

2.1 Pré-Processamento de Dados

O conjunto de dados empregado como objeto de estudo da metodologia proposta por esse trabalho, é disponibilizado por uma distribuidora de energia elétrica, e contém o histórico de consumo mensal referentes aos anos de 2017 a 2020, de unidades consumidoras rurais classificadas como Irrigantes em Período Reservado de acordo com o sistema setorial da concessionária.

Inicialmente, o conjunto de dados era constituído por 1085 unidades consumidoras rurais e 112 atributos associadas a esses clientes. Para o interesse desse estudo, foram selecionadas as seguintes informações: código de instalação, data de referência (no formato de 4 dígitos para “ano” e 2 dígitos para “mês”) e energia mensal faturada do consumidor. As informações referentes ao banco de dados foram processadas em três etapas a fim de selecionar as amostras a serem utilizadas como objeto de estudo na metodologia desse trabalho.

Na primeira etapa do pré-processamento foram selecionadas no banco de dados disponibilizado pela distribuidora, as unidades consumidoras irrigantes localizados no município de Uruguaiana, na Região da Fronteira Oeste do estado do Rio Grande do Sul (Região piloto do presente estudo). Essa filtragem da dados, resultou em 530 unidades consumidoras rurais.

Na segunda etapa de pré-processamento foram inseridos no banco de dados resultante da primeira etapa, duas novas colunas contendo o desvio padrão e a média do consumo de energia elétrica para cada unidade consumidora irrigante desse conjunto.

Na terceira e última etapa de pré-processamento foram eliminados do conjunto de dados, clientes com dados faltantes, nulos e com erros de cadastro. Desse modo, esses processos resultaram em 378 unidades consumidoras irrigantes do município de Uruguaiana.

2.2 Agrupamento de Dados em Classes de Perfil de Consumo

O agrupamento é uma técnica de mineração de dados não supervisionada que desassocia dados estruturados e não estruturados em grupos significativos denominados de *clusters*. (Umar et al. 2019). Para o processo de alocação das amostras em classes de perfil de consumo de energia elétrica, foram utilizados como dados de entrada, a média (m) e desvio padrão (σ) dos consumos de energia elétrica de cada unidade consumidora irrigante, definidas pela etapa de pré-processamento. O número de *clusters* (grupos de perfis de consumo) foi definido como 4 para a parametrização do algoritmo *k-Means*, sendo uma escolha particular, visando estimular a mudança de classes para a quantidade de amostras resultantes pela etapa inicial.

Para a construção das classes de perfil de consumo, priorizou-se os limites máximos e mínimos da média de consumo de energia elétrica durante o ano de 2019. Dessa forma, o algoritmo determina os valores máximos do atributo m de cada perfil de consumo de todas as unidades consumidoras.

Os dados apresentados na Tabela 1, apresentam a quantidade de unidades consumidores por classe de perfil de consumo e os limites máximos e mínimos do valor da média de consumo de energia elétrica para cada grupo. As unidades serão referidas como “UC” para apresentação dos dados.

Tabela 1. Número de amostras por classes de perfil de consumo.

Classe	UCs	Min/Max	m (kWh)	σ (kWh)
Perfil_1	117	Min	0	0
		Max	8622,50	13753,60
Perfil_2	115	Min	8688,60	1979,90
		Max	19254,00	28515,90
Perfil_3	77	Min	19617,70	5457,03
		Max	38118,30	70955,20
Perfil_4	69	Min	38305,00	4707,48
		Max	410992,00	860181,00

As amostras de consumidores por classe de perfil de consumo ficaram desbalanceadas devido ao algoritmo considerar como prioridade para delimitação dos *clusters*, a média de consumo de energia elétrica, evitando que um elemento pertencente a um determinado grupo seja alocado em outra classe de consumo. O desbalanceamento pode ocasionar alguns problemas de predições, pois o modelo tende a classificar corretamente as classes com maior número de amostras (Perfil_1 e Perfil_2) e erroneamente os grupos com números inferiores de clientes (Perfil_3 e Perfil_4).

2.3 Estimativa de Consumo de Energia Elétrica das Unidades Consumidoras Irrigantes

O estágio de estimativa de consumo de energia elétrica de unidades consumidoras irrigantes é composto por diversos atributos relacionados a cultura do arroz irrigado e aos sistemas de irrigação utilizados na lavoura. A Fig. 2 apresenta o diagrama sequencial de estimativa de consumo de energia elétrica em unidades consumidoras irrigantes de lavouras de arroz.

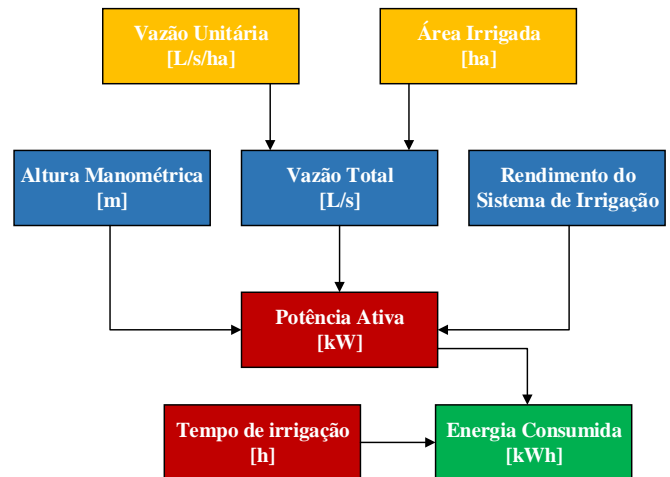


Fig. 2 Rotina de estimativa de consumo de energia elétrica para unidades consumidoras irrigantes.

A seguir, serão brevemente contextualizados os atributos associados a etapa de estimativa de consumo de energia elétrica em unidades consumidoras irrigantes.

- **Área Irrigada:** Representa a parcela da lavoura cultivada com a cultura do arroz e que necessita de irrigação. Esse atributo possui representatividade significativa no consumo de energia elétrica do sistema de irrigação instalado da unidade consumidora rural;

- **Vazão Unitária:** Aplicada para manutenção de água para cada hectare da lavoura de arroz. A mesma, é originada através de uma lâmina de água estabelecida em toda a área de plantio;

- **Altura Manométrica Total:** Representa a energia que a bomba hidráulica deverá transmitir a água para transportar uma determina vazão entre as tubulações de sucção e recalque. A altura manométrica total de uma estação de bombeamento é o resultado do montante entre a altura geométrica (desnível no terreno) e as perdas de cargas totais na tubulação;

- **Vazão Total:** Relação entre a vazão unitária de manutenção de água e área total de plantio da lavoura de arroz irrigado;

- **Rendimento do Sistema de Irrigação:** Definido como o produto dos rendimentos da bomba, motor e transmissão.

O cálculo da estimativa de consumo de energia elétrica em unidades consumidoras irrigantes, é iniciado a partir da

previsão da potência instalada do sistema de irrigação apresentada em Kopp et al. (2016), conforme descrito em (1).

$$P = \frac{Q \cdot AMT}{\eta_g} \cdot 0,0098$$

$$P = \frac{q \cdot A \cdot AMT}{\eta_g} \cdot 0,0098 \quad (1)$$

Onde P é a potência ativa do sistema de irrigação (kW); Q é a vazão total da lavoura de arroz (L/s); q é a vazão unitária para manutenção de irrigação (L/s/ha); A é a área irrigada (ha); AMT é a altura manométrica total da estação de bombeamento (m) e η_g é o rendimento global do sistema de irrigação.

A partir disso, pode-se estimar o consumo mensal de energia elétrica da unidade consumidora irrigante:

$$E = P \cdot t \quad (2)$$

Onde E é a consumo mensal estimado de energia elétrica da unidade consumidora irrigante (kWh); P é a potência ativa estimada para o sistema de irrigação (kW) e t é o tempo de funcionamento mensal da estação de bombeamento.

Algumas variáveis são fundamentais no processo de irrigação de lavoura de arroz e essenciais para produtividade da cultura e para o dimensionamento das estações de bombeamento, são elas: área irrigada, volume hídrico por hectare, altura da lâmina, vazão e tempo de irrigação.

2.4 Implementação do Modelo de Random Forest para Classificação de Dados

No contexto do *machine learning* (em português, aprendizado de máquina), a classificação de dados é realizada por meio de um algoritmo, onde o principal objetivo é prever o rótulo de novas entradas com base nos exemplos de entrada que já foram rotulados (Ferreira 2018). A classificação de dados é um processo de duas etapas, consistindo em uma etapa de aprendizado (onde os dados de treinamento são analisados por um algoritmo de classificação) e uma etapa de classificação (onde os dados de testes são utilizados para estimar a precisão das regras de classificação) (Han et al. 2014).

A implementação do modelo de classificação é iniciada com a separação das amostras de treino e teste do conjunto de consumidores irrigantes, definidos pela etapa inicial de pré-processamento. Os dados de treino são definidos como os elementos que serão apresentados ao algoritmo de *Random Forest* para a criação do modelo. Já os dados de teste, são aqueles demonstrados ao sistema de predição após a sua origem, reproduzindo previsões reais e avaliando o desempenho do modelo para uma determinada situação.

Para o presente modelo, os dados de treino foram divididos em 80% de todo o *dataset* (302 unidades consumidoras irrigantes), enquanto que as amostras de testes representam 20% da base de dados original (76 unidades consumidoras irrigantes). É importante salientar que, esses dados foram

divididos de acordo com as classes de perfis de consumo estabelecidas pela etapa de agrupamento. A Tabela 2 apresenta a distribuição das amostras de treino e teste para cada classe de perfil de consumo.

Tabela 2. Amostras de Treino e Teste para cada classe.

Classe	Amostras de Treino	Amostras de Teste
Perfil_1	87	30
Perfil_2	93	22
Perfil_3	66	11
Perfil_4	56	13

Após o processo de separação dos dados de treino e teste, é verificada a acurácia do modelo preditivo a partir de uma métrica de avaliação, apresentada na subseção a seguir.

2.5 Métrica de Avaliação para o Modelo Preditivo

Na aprendizagem de máquina, é necessário utilizar algumas métricas apropriadas para avaliar os resultados obtidos com um algoritmo classificador. Uma maneira de avaliar o desempenho de um modelo de classificação de dados é através de uma matriz de confusão, também conhecida como tabela de contingência. A mesma, é aplicada para tal finalidade, fornecendo não apenas a contagem de erros e acertos, mas também as variáveis necessárias para o cálculo de outras medidas (Beckmann et al. 2015).

O principal pontuador de uma matriz de confusão é a acurácia, que indica quantas amostras de dados foram classificadas corretamente, independente da classe. Por exemplo, se um conjunto de dados possui 100 amostras e 90 delas foram classificadas acertadamente, o modelo apresentou uma acurácia de 90%.

A matriz de confusão resultante para os dados de teste classificados pelo Modelo *Random Forest* é apresentada na Tabela 3.

Tabela 3. Matriz de confusão para o modelo preditivo.

Original		Classificadas pelo Modelo Random Forest				Acurácia
		Perfil_1	Perfil_2	Perfil_3	Perfil_4	
Perfil_1	30	29	1	0	0	0,97
Perfil_2	22	2	14	5	1	0,64
Perfil_3	11	0	1	6	4	0,55
Perfil_4	13	1	3	2	7	0,54

A matriz de confusão representada pela Tabela 3 expressa que, 29 unidades consumidoras irrigantes pertencentes ao grupo “Perfil_1” foram classificadas corretamente (elementos presentes na diagonal principal da matriz). Já para amostras da classe “Perfil_2”, 14 clientes foram classificados adequadamente e 8 amostras, erroneamente. Para o grupo “Perfil_3”, apenas 6 consumidores foram classificados corretamente. Para as amostras que compõe a classe “Perfil_4”, somente 7 unidades consumidoras foram categorizadas apropriadamente. Como esperado, os grupos “Perfil_3” e “Perfil_4” apresentaram as piores pontuações de precisão. A acurácia geral do Modelo *Random Forest* para o

conjunto de dados de teste atingiu o percentual de 74%, enquanto que para as amostras de treino foi de 100%.

2.5 Indicadores de ocorrências de Perdas Não Técnicas

Para averiguar a possibilidade de ocorrência de perdas não técnicas em unidades consumidoras, são propostos indicadores para avaliar a classe de perfil de consumo e o consumo mensal de energia elétrica estimado da instalação rural, utilizando como dados de entrada, registros mensais de energia faturada referentes ao ano de 2020 para o consumidor selecionado.

A Tabela 4 apresenta os indicadores e as situações propostas para avaliar os consumidores irrigantes que compõem a base de estudos de caso.

Tabela 4. Indicadores propostos para ocorrências de perdas não técnicas.

Indicador		Situação Proposta
Classe de Perfil de Consumo (Comparação entre 2019 e 2020)	Sem alteração de Classe	Consumidor regular
	Mudança de Classe (Classe Superior)	
	Mudança de Classe (Classe Inferior)	Possível prática de perdas não técnicas
Estimativa de Consumo Mensal (2020)	Real \geq 60% Estimado	Consumidor regular
	Real $<$ 60% Estimado	Possível prática de perdas não técnicas

O indicador de classe de perfil de consumo aponta para irregularidades se houver a mudança para um grupo inferior. A comparação é realizada entre a classe original (definida pela Etapa de Agrupamento com dados de 2019) e classe prevista (definida pela Etapa de Classificação com dados de 2020).

O segundo indicador realiza a comparação entre o consumo de energia elétrica mensal registrado da unidade consumidora irrigante no ano de 2020 com o valor estimado durante o mesmo intervalo. A possibilidade de ocorrências perdas não técnicas é indicada no caso em que o consumo real for inferior a 60% do consumo estimado. O presente indicador será aplicado apenas aos meses de Janeiro, Fevereiro, Novembro e Dezembro, pois os mesmos apresentam registros de consumos de energia elétrica mais significativos devido a atividade de irrigação intensa na lavoura de arroz.

3. ESTUDOS DE CASOS E RESULTADOS

Os estudos de casos serão realizados em duas unidades consumidoras irrigantes localizadas no município de Uruguai/RS na Região da Fronteira Oeste do Rio Grande do Sul, selecionadas no conjunto de dados (*dataset*) definidos pela Etapa de Pré-Processamento. Os consumidores rurais foram selecionados de maneira manual, com base no conhecimento das áreas de plantio de arroz irrigado associadas. Esse parâmetro é essencial para o modelo

preditivo, pois o mesmo é fundamental para o cálculo da estimativa de consumo de energia elétrica dos sistemas de irrigação das lavouras.

Nas subseções a seguir, serão apresentadas as características dos consumidores rurais selecionados e os resultados do modelo preditivo aplicado a estes estudos de caso.

3.1 Dados Característicos das Unidades Consumidoras Irrigantes

A Tabela 5 apresenta os dados característicos das unidades consumidoras irrigantes selecionadas para estudos de casos como parte da aplicação da metodologia proposta. Os consumidores serão identificados pelas nomenclaturas UC-1 e UC-2.

Tabela 5. Dados característicos das unidades consumidoras irrigantes.

Atributo	UC-1	UC-2
Classe de Perfil de Consumo	Perfil_2	Perfil_2
A (ha)	716,18	116,52
AMT (m)	10	10
q (L/s/ha)	1,5	1,5
η_g	0,65	0,65

As classes de perfil de consumo foram definidas pela Etapa de Agrupamento de Dados. O atributo da área irrigada foi determinado a partir das informações de intenção de plantio para cada unidade consumidora irrigante associada. Já os valores atribuídos as variáveis dos sistemas de irrigação foram referenciadas nos estudos realizados por Kopp et al. (2015) e Evaldt (2018).

A Tabela 6 apresenta as estimativas de potência ativa instalada e o consumo mensal de energia elétrica previsto para as unidades consumidoras irrigantes analisadas.

Tabela 6. Potência ativa e consumo mensal estimados para os consumidores rurais.

Atributo	UC-1	UC-2
P (kW)	162,11	26,35
t (h)	450	450
E (kWh)	72949,20	11858,66

A potência ativa estimada e consumo mensal previsto foram calculadas conforme (1) e (2), apresentados na Subseção 2.3. Para o tempo de irrigação, é atribuído um valor médio mensal de bombeamento, considerando um funcionamento contínuo em grande parte do período, e alguns dias de desligamentos devido a variáveis meteorológicas e quantidade de água acumulada nas lavouras.

Para os meses de Janeiro, Fevereiro, Novembro e Dezembro, será considerado um percentual de 100% do consumo de energia elétrica previsto, devido às incertezas quanto as variações nas áreas de cultivo e técnicas de manejo na irrigação. Para os demais meses, não serão atribuídos registros de energia faturada.

3.1 Resultados

3.1.1. Unidade Consumidora UC-1

A Fig. 3 descreve o relatório final de simulação para o indicador de classe de perfil de consumo realizada no *software* Spyder®.

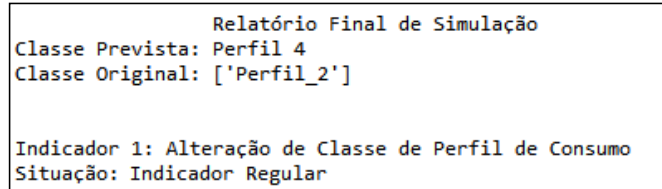


Fig. 3 Relatório do indicador de classe de consumo para a UC-1.

A Fig. 4 apresenta um comparativo na curva de carga mensal entre os consumos reais e previstos, e os respectivos limites de tolerância para o indicativo de ocorrência de perdas não técnicas para o consumidor irrigante UC-1. Já a Fig. 5, ilustra o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no *software* Spyder®.

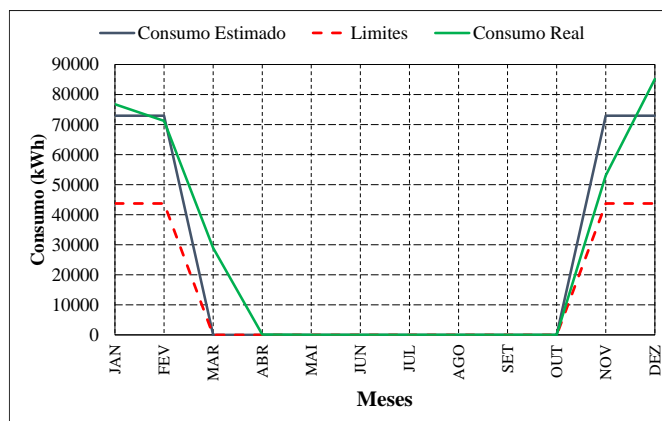


Fig. 4 Curvas de consumo para a UC-1.

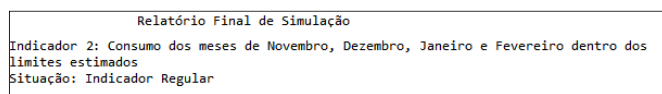


Fig. 5 Relatório do indicador de estimativa de consumo para a UC-1.

De acordo com o relatório apresentado pela Fig. 3, a classe prevista para o ano de 2020 da unidade consumidora irrigante UC-1, classificadas pelo modelo preditivo, difere do grupo de consumo definido pela Etapa de Agrupamento. A classe de perfil de consumo original pertencente a este consumidor é o grupo “Perfil_2”, enquanto que a classe prevista determinada pelo modelo é o cluster “Perfil_4”. No entanto, como ocorreu a mudança para uma classe superior de perfil de consumo, não é assinalada a possibilidade de ocorrência de perdas não técnicas pela unidade consumidora UC-1, para avaliação desse indicador.

Os resultados apresentados pelas Fig. 4 e Fig. 5, evidenciam que os consumos registrados para a respectiva unidade

consumidora nos meses de Janeiro, Fevereiro, Novembro e Dezembro, não infringiram os limites mínimos mensais estabelecidos para o indicativo de perdas não técnicas. Desse modo, o presente indicador é estabelecido como regular.

O presente consumidor apresentou regularidades em ambos indicadores, sendo descartado pelo modelo preditivo como um possível praticante de perdas não técnicas.

3.1.1. Unidade Consumidora UC-2

A Fig. 6 apresenta o relatório final de simulação para o indicador de classe de perfil de consumo realizada no *software* Spyder®.

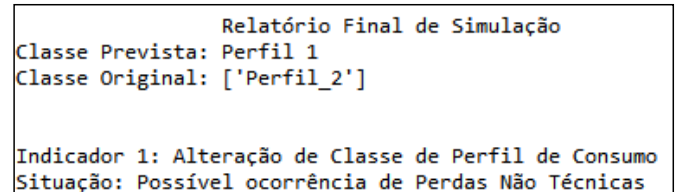


Fig. 6 Relatório do indicador de classe de consumo para a UC-2.

A Fig. 7 apresenta um comparativo na curva de carga mensal entre os consumos reais e previstos, e os respectivos limites de tolerância para o indicativo de ocorrência de perdas não técnicas para o consumidor irrigante UC-2. Já a Fig. 8, apresenta o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no *software* Spyder®.

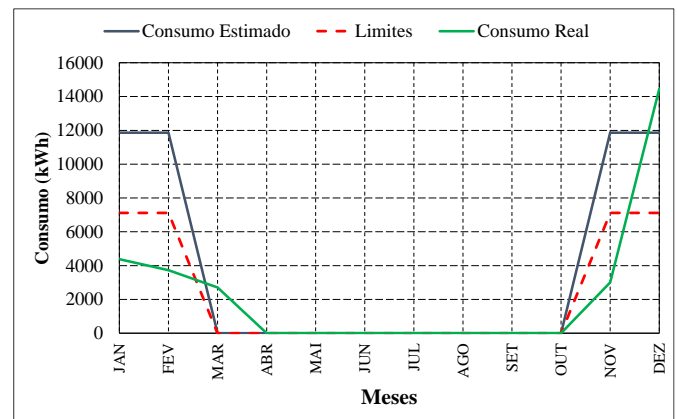


Fig. 7 Curvas de consumo para a UC-2.

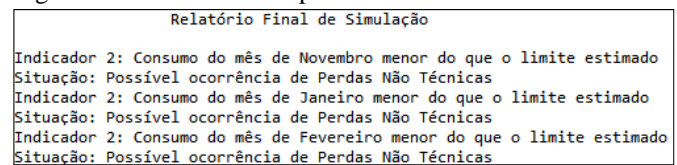


Fig. 8 Relatório do indicador de estimativa de consumo para a UC-2.

De acordo com o relatório demonstrado na Fig. 6, a classe prevista para o ano de 2020 da unidade consumidora irrigante UC-2, classificada pelo modelo preditivo, diverge do grupo de consumo definido pela Etapa de Agrupamento. A classe de perfil de consumo original pertencente a este cliente irrigante é o grupo “Perfil_2”, enquanto que a classe prevista

determinada pelo modelo preditivo é a classe “Perfil_1”. Desse modo, devido a mudança para uma classe inferior de perfil de consumo, o consumidor rural UC-2 é indicado nesse avaliador como uma possibilidade de práticas de perdas não técnicas.

Os resultados apresentados nas Fig. 7 e Fig. 8, demonstram que os consumos mensais registrados para a unidade consumidora UC-2 nos meses de Janeiro, Fevereiro e Novembro, foram inferiores aos limites mínimos propostos para o indicativo de perdas não técnicas. Dessa forma, o consumidor rural avaliado é apontando para esse indicador, como um possível atuator de práticas de perdas não técnicas.

O presente consumidor apresentou irregularidades em ambos indicadores, sendo apontado pelo modelo preditivo como um possível suspeito de práticas de perdas não técnicas.

6. CONCLUSÕES

O presente trabalho apresentou uma metodologia para identificar possíveis ocorrências de perdas não técnicas em unidades consumidoras rurais, utilizando um algoritmo que engloba métodos de inteligência artificial para agrupamento e classificação de dados. Além disso, como parte da metodologia proposta, foram realizadas estimativas de consumo mensal de energia elétrica dos sistemas de irrigação dos consumidores rurais, utilizando dados previamente pesquisados na literatura.

Os resultados dos estudos de casos mostraram-se satisfatórios, sendo apresentadas diferentes situações para as unidades consumidoras irrigantes avaliadas. Os indicadores propostos identificaram irregularidades no consumo de energia elétrica para o segundo consumidor avaliado. As estimativas mensais de consumo de energia elétrica, realizadas através das características dos sistemas de irrigação e das áreas de plantio associadas, foram convenientes ou permaneceram dentro de um padrão aceitável na maioria dos casos.

Para trabalhos futuros, pretende-se considerar o impacto das variáveis meteorológicas no consumo mensal de energia elétrica dos sistemas de irrigação, aprimorar os processos de agrupamento e classificação de dados, expandir os testes de validação para outras regiões produtoras de arroz irrigado e avaliar esse estudo em um maior intervalo de safras.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES/PROEX) – Código de Financiamento 001. Os autores gostariam de agradecer o apoio técnico e financeiro da CPFL Energia ao projeto “Sistema de Detecção de Perdas Não Técnicas em Áreas de Irrigação empregando Técnicas de Inteligência Artificial” (desenvolvido no âmbito do Programa de PD da ANEEL PD 00063-3065/2020). Este estudo também foi realizado com apoio do Instituto Nacional de Ciência e Tecnologia em Geração Distribuída (INCT-GD) da Universidade Federal de Santa Maria (UFSM) (processo

CNPq 465640/2014-1, processo CAPES 23038.000776/2017-54 e FAPERGS 17/2551-0000517-1).

REFERÊNCIAS

- Barros, R., Costa, E.G., e Araújo, J. (2021). Maximizing the Financial Return of Non-Technical Loss Management in Power Distribution Systems. *IEEE Transactions on Power Systems*, pp 1-6.
- Beckmann, M., Ebecken, N.F.F., Lima, B.S.L.P. (2015). A KNN Undersampling Approach for Data Balancing. *Journal of Intelligent Learning Systems and Applications*, vol. 7 (n. 4), pp. 104-116.
- Breimann, L. (2001). Random forests. *Machine Learning*, vol. 45, pp. 5-32.
- Evaldt, M. C. (2018). Sistema Neural Artificial para Identificação de Perdas Não Técnicas em Consumidores Rurais. Dissertação (Mestrado em Engenharia Elétrica). Universidade Federal de Santa Maria, Santa Maria, Brasil.
- Ferreira, L.D. (2018). Técnicas de aprendizado de máquina aplicadas à classificação de estudantes a partir de estilos de aprendizagem. Dissertação (Mestrado em Ciência da Computação), Universidade de São Paulo, São Paulo, Brasil.
- Ghori, K.M., Ayaz, A.R., Awais, M., Imran, M., Ullah, A., e Szathmary, L. (2020). Impact of Feature Selection on Non-technical Loss Detection. Em *CDMA 2020*, Riad, Arábia Saudita.
- Han, J. Kamber, M., Pei, J. (2014). *Data mining: Data mining concepts and techniques*. Morgan Kaufmann, Waltham, Estados Unidos.
- Kopp, L.M., Peiter, M.X, Robaina, A. D., e Toescher, C. F. (2016). Caracterização de Estações de Bombeamento em Lavouras de Arroz do Rio Grande do Sul. *Journal of the Brazilian Association of Agricultural Engineering*, vol. 36 (n. 2), pp 342-351.
- Medeiros, M.H., Sanz-Bobi, M.A., Domingo, J.M., e Picchi, D. (2021). Network Oriented Approaches Using Smart Metering Data for Non-Technical Losses Detection. Em *2021 IEEE Madrid PowerTech*, Madri, Espanha.
- Messinis, G.M., e Hatziargyriou, N.D. (2018). Review of non-technical loss detection methods. *Electric Power Systems Research*, vol. 158, pp 1-6.
- Passos Júnior, L.A., Ramos, C.C.O., Rodrigues, D., Pereira D.R., Souza, A.N., Costa, K.A.P., e Papa, J.P. (2016). Unsupervised non-technical losses identification through optimum-path forest. *Electric Power Systems Research*, vol. 140, pp. 413-423.
- Pfitscher, L.L., Bernardon, D.P., Kopp, L.M., Heckler, M. V. T., Behrens, J. Montani, P. B., e Thomé, B. (2012). Automatic control of irrigation systems aiming at high energy efficiency in rice crops. Em *ICCDCS 2012*, Playa del Carmén, México.
- Umar, H.A., Prasad, R., e Fonkam, M. (2019). Assessing severity of non-technical losses in power using clustering Algorithms. Em *ICECCO 2019*, Abuja, Nigéria.