

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Henrique Silveira Eichkoff

**METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO
TÉCNICAS EM UNIDADES CONSUMIDORAS LOCALIZADAS EM
ÁREAS RURAIS COM CULTIVO DE ARROZ IRRIGADO**

Santa Maria, RS, Brasil

2021

Henrique Silveira Eichkoff

**METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM
UNIDADES CONSUMIDORAS LOCALIZADAS EM ÁREAS RURAIS COM
CULTIVO DE ARROZ IRRIGADO**

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica, Área de Concentração em Processamento de Energia Elétrica, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Mestre em Engenharia Elétrica**

Orientador: Prof. Dr. Daniel Pinheiro Bernardon

Santa Maria, RS, Brasil

2021

Eichkoff, Henrique Silveira

Metodologia para Identificação de Perdas Não Técnicas em Unidades Consumidoras localizadas em Áreas Rurais com Cultivo de Arroz Irrigado / Henrique Silveira Eichkoff.- 2021.

174 p.; 30 cm

Orientador: Daniel Pinheiro Bernardon

Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, RS, 2021

1. Perdas Não Técnicas 2. Unidades Consumidoras Rurais 3. Sistemas de Irrigação 4. Estimativa de Consumo de Energia Elétrica 5. Agrupamento e Classificação de Dados I. Pinheiro Bernardon, Daniel II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

©2021

Todos os direitos autorais reservados a Henrique Silveira Eichkoff. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

Endereço: Avenida Roraima, 1000, Prédio 10, Bairro Camobi, Santa Maria, RS, 97105-900

E-mail: henriquekoff@gmail.com

Henrique Silveira Eichkoff

**METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM
UNIDADES CONSUMIDORAS LOCALIZADAS EM ÁREAS RURAIS COM
CULTIVO DE ARROZ IRRIGADO**

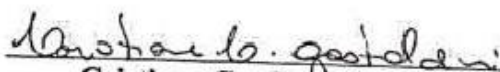
Dissertação apresentada ao Curso de
Mestrado do Programa de Pós-Graduação
em Engenharia Elétrica, Área de
Concentração em Processamento de
Energia Elétrica, da Universidade Federal
de Santa Maria (UFSM, RS), como
requisito parcial para a obtenção do grau
de:

Mestre em Engenharia Elétrica

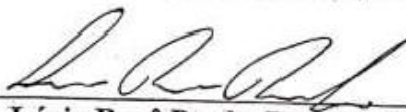
Aprovado em 10 de Dezembro de 2021:



Daniel Pinheiro Bernardon, Dr. Eng. (UFSM)
(Presidente/Orientador)



Cristiane Caudure Gastaldini, Dr^a. Eng. (UFSM)



Lúcio Renê Prade, Dr. Eng. (UNISINOS)

Santa Maria, RS, Brasil.

2021

Dedico este trabalho as pessoas mais importantes da minha vida: meus pais, Roberto e Marilei, e minha irmã Vanessa.

AGRADECIMENTOS

Aos meus pais Roberto e Marilei, e minha irmã Vanessa, pelo o amor, carinho, compreensão, ensinamentos, entusiasmo, incentivo, paciência e por todo o apoio destinado a mim, durante o desenvolvimento desse trabalho.

Ao Professor Dr. Daniel Pinheiro Bernardon, pela orientação, confiança, incentivo, motivação e diversas oportunidades de aprendizagem proporcionadas.

Aos docentes que compõem a banca examinadora, Professores Drs. Cristiane Cauduro Gastaldini e Lúcio Renê Prade, pelas avaliações e contribuições realizadas a esta dissertação.

A Companhia Paulista Força e Luz Energia (CPFL Energia) pelo suporte técnico e incentivo financeiro ao Projeto “Sistemas de Detecção de Perdas Não Técnicas em Áreas de Irrigação empregando Técnicas de Inteligências Artificial” (desenvolvido no âmbito do programa de P&D da ANEEL PD-00063-3065/2020) que orientou essa dissertação. O presente trabalho também foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível (CAPES) – Código Financeiro 001.

Aos engenheiros eletricitistas do Grupo CPFL Energia, Juliano Andrade Silva e Lucas Melo Chiara, por todo auxílio e suporte técnico destinados ao Projeto de Pesquisa e Desenvolvimento (P&D) que orientou essa dissertação.

A Professora Dr^a Luciana Marini Köpp da Universidade Federal de Pelotas (UFPFL), pelas contribuições valiosas proporcionadas a este estudo, expressando minha enorme gratidão pela oportunidade de trabalhar em conjunto no desenvolvimento do P&D que originou essa dissertação.

Aos amigos e colegas Natália Souza e Pedro Marcolin, pelo o apoio, amizade, aprendizagem e convívio, os quais tive o imenso prazer de trabalhar em conjunto durante a realização do P&D que orientou esse estudo.

Aos amigos e colegas do Centro de Excelência em Energia e Sistemas de Potência (CEESP) pelo bom convívio e amizade.

Aos professores e técnicos administrativos do Programa de Pós-Graduação em Engenharia Elétrica (PPGEE) da Universidade Federal de Santa Maria (UFSM).

A minha amiga, professora e orientadora durante o período da graduação em Engenharia Elétrica na Universidade Federal do Pampa (UNIPAMPA), Ana Paula Carboni de Mello, que sempre me incentivou a seguir os estudos na pós-graduação e a quem serei sempre extremamente grato por todas as oportunidades proporcionadas.

Aos amigos e engenheiros eletricitistas, Alex Itczak, Arian Fagundes, Dimas Schuetz, Guilherme Pinheiro, João Alberto Machiavelli, Lucas Tarone, Luis Pedro Gonçalves, Marcel Stalter, Micael Oliveira, Paulo Ricardo Marques e Pietro Cassol, pela grande amizade e o incentivo de sempre.

Aos amigos e colegas, Daniel Lemes, Daniela Bueno, Matheus Veiga e Rodolfo Londero, pelas jornadas de estudos e trabalhos desenvolvidos nas disciplinas durante o período do mestrado.

A todos que conviveram comigo e de maneiras diferentes, me incentivaram e contribuíram para a realização deste trabalho.

“Tente uma, duas, três vezes e se possível tente a quarta, a quinta e quantas vezes for necessário. Só não desista nas primeiras tentativas, a persistência é amiga da conquista. Se você quer chegar aonde a maioria não chega, faça o que a maioria não faz.”

Bill Gates

RESUMO

METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM UNIDADES CONSUMIDORAS LOCALIZADAS EM ÁREAS RURAIS COM CULTIVO DE ARROZ IRRIGADO

AUTOR: Henrique Silveira Eichkoff
ORIENTADOR: Daniel Pinheiro Bernardon

As perdas não técnicas são um dos maiores problemas enfrentados pelas empresas do setor energético, uma vez que, afetam diretamente as receitas financeiras das concessionárias, a segurança nos sistemas elétricos e a qualidade no fornecimento de energia elétrica para os consumidores. As mesmas, estão associadas a erros de leituras e medições, ligações clandestinas a rede secundária de distribuição e fraudes em medidores. Em alimentadores rurais, o problema das perdas não técnicas torna-se mais complexo, pois as grandes extensões das redes de distribuição e a presença de unidades consumidoras alocadas em regiões remotas ou de difícil acesso, dificultam a fiscalização pelas equipes de inspeções. Além disso, os consumidores rurais são extremamente relevantes para as distribuidoras, pois representam uma parcela significativa do consumo de energia elétrica, devido aos sistemas de irrigação instalados com bombas e motores de elevada potência ativa e utilizados de maneira praticamente contínua durante o período da safra. Este trabalho apresenta uma proposta para a identificação de perdas não técnicas em unidades consumidoras localizadas em áreas rurais com cultivo de arroz irrigado. A metodologia desenvolvida nesse estudo, é baseada na correlação dos padrões de consumos de energia elétrica e das características das áreas de plantio e dos sistemas de irrigação. A metodologia emprega na estrutura de seu algoritmo, os métodos de *k-Means* e *Random Forest*, para o desenvolvimento das etapas de agrupamento e classificação de dados, respectivamente, sendo utilizados como entradas, registros mensais de energia faturada de um conjunto de consumidores rurais. Além disso, o algoritmo tem a capacidade de estimar o consumo de energia elétrica, a partir de informações referenciais de algumas variáveis que compõem os sistemas de irrigação e das áreas de plantio cultivadas para uma determinada safra. A validação da metodologia proposta foi realizada em um conjunto de sete unidades consumidoras rurais localizadas na Região da Fronteira Oeste do estado do Rio Grande do Sul. Para verificar possíveis irregularidades, foram avaliados indicadores referentes a classe e a comparação entre consumos reais e estimados para cada consumidor. Os resultados demonstraram a operação adequada do modelo, indicando cinco casos suspeitos de ocorrências de perdas não técnicas.

Palavras-chave: Agrupamento. Classificação. Estimativa de Consumo. *k-Means*. Perdas Não Técnicas. *Random Forest*. Sistemas de Irrigação. Unidades Consumidoras Rurais.

ABSTRACT

METHODOLOGY FOR IDENTIFYING NON-TECHNICAL LOSSES IN CONSUMERS LOCATED IN RURAL AREAS WITH IRRIGATED RICE CULTIVATION

AUTHOR: Henrique Silveira Eichkoff

ADVISOR: Daniel Pinheiro Bernardon

Non-technical losses are one of the biggest problems faced by companies in the energy sector, since they directly affect the financial revenues of concessionaires, the security of electrical systems and the quality of electricity supply to consumers. They are associated with errors in readings and measurements, clandestine connections to the secondary distribution network and fraud in meters. In rural feeders, the problem of non-technical losses becomes more complex, as the large extensions of the distribution networks and the presence of consumers located in remote or difficult-to-access regions make it difficult for inspection teams to investigation. Furthermore, rural consumers are extremely relevant for concessionaires, as they represent a significant portion of electricity consumption, due to irrigation systems installed with pumps and motors with high active power and used practically continuously during the crop period. This work presents a proposal for the identification of non-technical losses in consumers located in rural areas with irrigated rice cultivation. The methodology developed in this study based on the correlation of electricity consumption patterns and the characteristics of planting areas and irrigation systems. The methodology employs in the structure of its algorithm, the k-Means and Random Forest methods, for the development of the data clustering and classification stages, respectively, being used as inputs, monthly records of billed energy from a set of rural consumers areas. Furthermore, the algorithm has the capacity to estimate the consumption of electricity, based on referential information of some variables that make up the irrigation systems and the cultivated planted areas for a certain crop. The validation of the proposed methodology was accomplished in a group of seven rural consumers located in the West Frontier Region of the state of Rio Grande do Sul. In order to verify for possible irregularities, indicators referring to the class and the comparison between actual and estimated consumptions for each consumer were evaluated. The results demonstrated the proper operation of the model, indicating five suspected cases of occurrences of non-technical losses.

Keywords: Classification. Clustering. Consumption Estimate. Irrigation Systems. k-Means. Non-Technical Losses. Random Forest. Rural Consumers.

LISTA DE FIGURAS

Figura 2.1 – Cultivo de arroz de sequeiro sem irrigação.	47
Figura 2.2 – Cultivo de arroz irrigado por inundação.	49
Figura 2.3 – Cultivo de arroz irrigado por pivô central.	49
Figura 2.4 – Esquema de um sistema de irrigação por inundação.	50
Figura 2.5 – Representação das perdas de água por percolação (P), fluxo lateral (FL), evaporação (E) e transpiração (T) em uma lavoura de arroz.	53
Figura 2.6 – Mapa do Rio Grande do Sul com suas respectivas regiões orizícolas.	54
Figura 2.7 – Captação da água em uma fonte hídrica pela tubulação de sucção da bomba hidráulica – Início do levante em lavouras de arroz.	61
Figura 2.8 – Desague no canal de irrigação – Término do levante em lavouras de arroz.	61
Figura 2.9 – Conjunto Bomba-Motor de um sistema de irrigação.	63
Figura 2.10 – Acoplamento por polia e correias (a) eixo horizontal e (b) eixo vertical.	64
Figura 2.11 – Perdas sobre a Energia Injetada no período entre 2008 e 2019.	70
Figura 2.12 – Composição da Tarifa de Energia Elétrica no Brasil.	72
Figura 3.1 – Fluxograma da metodologia proposta.	77
Figura 3.2 – Exemplo de um processo de <i>Clustering</i>	83
Figura 3.3 – Exemplo do algoritmo <i>k-Means</i>	85
Figura 3.4 – Representação gráfica do resultado de agrupamento de unidades consumidoras para diferentes classes de perfil de consumo.	86
Figura 3.5 – Rotina de estimativa de consumo de energia elétrica para unidades consumidoras irrigantes.	88
Figura 3.6 – Identificação de uma área de plantio de arroz irrigado através do processamento de imagens de satélite.	89
Figura 3.7 – Diagrama contendo diferentes alturas de lâmina de água na lavoura arroz irrigado.	91
Figura 3.8 – Diagrama ilustrando a altura manométrica em uma estação de bombeamento. ..	95
Figura 3.9 – Curva característica de uma bomba centrífuga.	96
Figura 3.10 – Processo de Classificação de dados.	101
Figura 3.11 – Exemplo de arquitetura do método de <i>Random Forest</i>	104
Figura 3.12 – Distribuição das amostras de treino e teste para cada classe de perfil de consumo.	106
Figura 4.1 – Área Irrigada para a Unidade Consumidora UC-1.	115
Figura 4.2 – Área Irrigada para a Unidade Consumidora UC-2.	116
Figura 4.3 – Área Irrigada para a Unidade Consumidora UC-3.	116
Figura 4.4 – Área Irrigada para a Unidade Consumidora UC-4.	117

Figura 4.5 – Área Irrigada para a Unidade Consumidora UC-5.	117
Figura 4.6 – Área Irrigada para a Unidade Consumidora UC-6.	118
Figura 4.7 – Área Irrigada para a Unidade Consumidora UC-7.	118
Figura 4.8 – Curva característica da estimativa anual de consumo.	120
Figura 4.9 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-1 no IDE Spyder®.	122
Figura 4.10 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-1.	122
Figura 4.11 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-1 no IDE Spyder®.	123
Figura 4.12 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-2 no IDE Spyder®.	124
Figura 4.13 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-2.	125
Figura 4.14 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-2 no IDE Spyder®.	125
Figura 4.15 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-3 no IDE Spyder®.	127
Figura 4.16 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-3.	128
Figura 4.17 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-3 no IDE Spyder®.	128
Figura 4.18 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-4 no IDE Spyder®.	130
Figura 4.19 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-4.	131
Figura 4.20 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-4 no IDE Spyder®.	131
Figura 4.21 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-5 no IDE Spyder®.	132
Figura 4.22 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-5.	133
Figura 4.23 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-5 no IDE Spyder®.	134
Figura 4.24 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-6 no IDE Spyder®.	135
Figura 4.25 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-6.	136
Figura 4.26 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-6 no IDE Spyder®.	136

Figura 4.27 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-7 no IDE Spyder®.....	138
Figura 4.28 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-7.....	139
Figura 4.29 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-7 no IDE Spyder ®.....	139
Figura A.1 – <i>Script</i> para inicialização das Bibliotecas <i>Python</i> do algoritmo proposto para a metodologia.	157
Figura A.2 – Código para a leitura do conjunto de dados das unidades consumidoras irrigantes.....	158
Figura A.3 – Código para a etapa de agrupamento de dados.	159
Figura A.4 – Código para a etapa de classificação de dados.....	160
Figura A.5 – Código da etapa de estimativa de consumo para o estudo de caso da UC-1. ...	161
Figura A.6 – Código para saída do modelo aplicado ao estudo de caso da UC-1.....	162
Figura A.7 – Código da etapa de estimativa de consumo para o estudo de caso da UC-2. ...	163
Figura A.8 – Código para saída do modelo aplicado ao estudo de caso da UC-2.....	164
Figura A.9 – Código da etapa de estimativa de consumo para o estudo de caso da UC-3. ...	165
Figura A.10 – Código para saída do modelo aplicado ao estudo de caso da UC-3.....	166
Figura A.11 – Código da etapa de estimativa de consumo para o estudo de caso da UC-4...	167
Figura A.12 – Código para saída do modelo aplicado ao estudo de caso da UC-4.....	168
Figura A.13 – Código da etapa de estimativa de consumo para o estudo de caso da UC-5...	169
Figura A.14 – Código para saída do modelo aplicado ao estudo de caso da UC-5.....	170
Figura A.15 – Código da etapa de estimativa de consumo para o estudo de caso da UC-6...	171
Figura A.16 – Código para saída do modelo aplicado ao estudo de caso da UC-6.....	172
Figura A.17 – Código da etapa de estimativa de consumo para o estudo de caso da UC-7...	173
Figura A.18 – Código para saída do modelo aplicado ao estudo de caso da UC-7.....	174

LISTA DE TABELAS

Tabela 2.1 – Intenção de plantio e área semeada por região orizícola no Rio Grande do Sul.	58
Tabela 2.2 – Intenção de plantio e área semeada nos principais municípios produtores de arroz do Rio Grande do Sul.	58
Tabela 2.3 – Dados gerais de levante para o arroz irrigado.	60
Tabela 3.1 – Atributos selecionados do banco de dados.	80
Tabela 3.2 – Número de unidades consumidoras por perfil de consumo e valores de máximo e mínimo.	87
Tabela 3.3 – Valores típicos para os atributos de lâmina de água e vazão unitária.	92
Tabela 3.4 – Valor típico para o atributo de altura manométrica total.	95
Tabela 3.5 – Valor típico para o atributo de rendimento global do sistema de irrigação.	97
Tabela 3.6 – Valor típico para o atributo de tempo de bombeamento do sistema de irrigação.	98
Tabela 3.7 – Exemplo de uma matriz de confusão.	108
Tabela 3.8 – Matriz de Confusão e Pontuadores do Modelo <i>Random Forest</i> .	110
Tabela 4.1 – Indicadores propostos pelo Modelo Preditivo para identificação de perdas comerciais.	112
Tabela 4.2 – Classes de perfil de consumo correspondente a cada unidade consumidora irrigante.	113
Tabela 4.3 – Dados mensais de consumo de energia elétrica referentes ao ano de 2020.	114
Tabela 4.4 – Variáveis do sistema de irrigação para todas as unidades consumidoras.	114
Tabela 4.5 – Dados de área irrigada para cada unidade consumidora associada.	119
Tabela 4.6 – Potência Ativa e Consumo de Energia Elétrica mensais estimados para cada unidade consumidora irrigante.	120
Tabela 4.7 – Resumo geral dos indicadores avaliados para a UC-1.	123
Tabela 4.8 – Resumo geral dos indicadores avaliados para a UC-2.	126
Tabela 4.9 – Resumo geral dos indicadores avaliados para a UC-3.	129
Tabela 4.10 – Resumo geral dos indicadores avaliados para a UC-4.	132
Tabela 4.11 – Resumo geral dos indicadores avaliados para a UC-5.	134
Tabela 4.12 – Resumo geral dos indicadores avaliados para a UC-6.	137
Tabela 4.13 – Resumo geral dos indicadores avaliados para a UC-7.	140

LISTA DE ABREVIATURAS E SIGLAS

A*	Algoritmo A-Star
AavNNet	Model Averaged Neural Netowrks
ANA	Agência Nacional de Águas
ANEEL	Agência Nacional de Energia Elétrica
ANOVA	Analysis of Variance
BN	Bayesian Network
CONAB	Companhia Nacional de Abastecimento
CSS	Charged System Search
DBSCAN	Density Based Spatial Clustering of Applications with Noise
FS	Fator de Serviço
GAM	Generalized Additive Model
GIS	Geographic Information System
GMM	Gaussian Mixture Model
IA	Inteligência Artificial
IFS	Incremental Feature Selection
IRGA	Instituto Rio-Grandense do Arroz
k-NN	Nearest neighbor
LDA	Linear Discriminant Analysis
LOF	Local Outlier Factor
LSTM	Long Short-Term Memory
MEC	Maximum Entropy Classifier
MGD	Gaussian Distribution
MLP	Multi-Layer Perceptrons
NB	Naive Bayes Classifier
Nnet	Model Neural Network
OpenDSS®	Open Distribution System Simulator
OPF	Optimum Path Forrest
P-OPF	Probabilistic Optimum-Path Forest
PRORET	Procedimentos de Regulação Tarifária
PSM	Power Summation Method
PSO	Particle Swarm Optimization
QDA	Quadratic Discriminant Analysis

REHs	Resoluções Homologatórias
RNA	Redes Neurais Artificiais
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
SOM	Self Organizing Map
Spyder®	The Scientific Python Development Environment
SVM	Support Vector Machine
TE	Tarifa de Energia
TUSD	Tarifa de Uso do Sistema de Distribuição
UC	Unidade Consumidora
WLS	Weighted Least Squares

LISTA DE SÍMBOLOS

A	Área da lavoura a ser irrigada
AMT	Altura manométrica total
C	Coefficiente de rugosidade de Hazzen-Willians
D	Diâmetro da tubulação
E	Energia consumida pelo sistema de irrigação
FN	Falso Negativo
FP	Falso Positivo
h_f	Perda de carga na tubulação
h_{fR}	Perda de carga na tubulação de recalque
h_{fS}	Perda de carga na tubulação de sucção
h_{fT}	Perda de carga total na estação de bombeamento
h_g	Altura geométrica
h_r	Altura de recalque
h_s	Altura de sucção
k	Número de classes definidas
L	Comprimento da tubulação
$L_{virtual}$	Comprimento virtual equivalente às peças especiais presentes na instalação
m	Média do consumo de energia elétrica da unidade consumidora
n	Número de peças especiais presentes na estação de bombeamento
N	Número de diâmetros equivalente a peça especial
P	Potência ativa do sistema de irrigação
Q	Vazão total
q	Vazão unitária para manutenção de irrigação na lavoura
Q_{md}	Vazão medida
t	Tempo de funcionamento mensal da estação de bombeamento
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
η_b	Rendimento da bomba
η_g	Rendimento global do sistema de irrigação
η_m	Rendimento do motor
η_s	Velocidade específica da bomba

η_t	Rendimento da transmissão
σ	Desvio Padrão do consumo de energia elétrica da unidade consumidora

SUMÁRIO

1. INTRODUÇÃO.....	29
1.1 MOTIVAÇÃO E JUSTIFICATIVA.....	32
1.2 OBJETIVOS GERAIS E ESPECÍFICOS.....	33
1.3 ESTADO-DA-ARTE.....	33
1.4 CONTRIBUIÇÕES DO TRABALHO.....	41
1.5 DELIMITAÇÕES DO TRABALHO.....	42
1.6 ORGANIZAÇÃO DOS CAPÍTULOS.....	43
2. PERDAS NÃO TÉCNICAS EM CONSUMIDORES IRRIGANTES DE LAVOURAS DE ARROZ.....	45
2.1 LAVOURAS DE ARROZ.....	45
2.1.1 Cultivo de Arroz Irrigado.....	46
2.1.2 Impacto das Variáveis Meteorológicas nas Lavouras de Arroz Irrigado.....	51
2.1.3 Perdas de Água nas Lavouras de Arroz.....	52
2.1.4 Regiões Orizícolas do Rio Grande do Sul.....	53
2.1.5 Plantio por regiões orizícolas e municípios do Rio Grande do Sul.....	57
2.2 SISTEMAS DE IRRIGAÇÃO.....	59
2.2.1 Processo de Irrigação nas Lavouras de Arroz.....	60
2.2.2 Características das Estações de Bombeamento.....	62
2.3 PERDAS NÃO TÉCNICAS.....	66
2.3.1 Perdas Não Técnicas no Brasil.....	70
2.3.2 Impacto das Perdas Não Técnicas na Estrutura Tarifária do Brasil.....	71
2.3.3 Tratamento Regulatório de Perdas Não Técnicas no Brasil.....	73
2.3.4 Perdas Não Técnicas em Alimentadores Rurais.....	74
3. METODOLOGIA PROPOSTA.....	77
3.1 CONJUNTO DE UNIDADES CONSUMIDORAS RURAIS E PRÉ-PROCESSAMENTO DE DADOS.....	79
3.2 AGRUPAMENTO DE DADOS.....	81
3.2.1 Método <i>k-Means</i>.....	83
3.2.2 Alocação de Unidades Consumidoras Irrigantes em Classes de Consumo de Energia Elétrica.....	84
3.3 ESTIMATIVA DO CONSUMO DE ENERGIA ELÉTRICA DE UNIDADES CONSUMIDORAS IRRIGANTES.....	87
3.3.1 Área Irrigada.....	88
3.3.2 Lâmina de água e Vazão.....	89
3.3.3 Altura Manométrica Total.....	92

3.3.4	Rendimento Global do Sistema de Irrigação	95
3.3.5	Tempo de funcionamento do Sistema de Irrigação	97
3.3.6	Estimativa do Consumo de Energia Elétrica do Sistema de Irrigação.....	98
3.4	CLASSIFICAÇÃO DE DADOS.....	100
3.4.1	Método de Random Forest.....	102
3.4.2	Modelo Preditivo para Detecção de Perdas Não Técnicas aplicando o Algoritmo de <i>Random Forest</i>.....	105
3.5	MÉTRICAS DE AVALIAÇÃO.....	107
3.5.1	Métrica de Avaliação do Modelo Preditivo de Identificação de Perdas Não Técnicas em Unidades Consumidoras Irrigantes	110
4.	ESTUDOS DE CASOS	111
4.1	DADOS CARACTERÍSTICOS DAS UNIDADES CONSUMIDORAS IRRIGANTES	112
4.2	RESULTADOS E DISCUSSÕES.....	121
4.2.1	Unidade Consumidora Irrigante UC-1.....	121
4.2.2	Unidade Consumidora Irrigante UC-2.....	124
4.2.3	Unidade Consumidora Irrigante UC-3.....	126
4.2.4	Unidade Consumidora Irrigante UC-4.....	129
4.2.5	Unidade Consumidora Irrigante UC-5.....	132
4.2.6	Unidade Consumidora Irrigante UC-6.....	135
4.2.7	Unidade Consumidora Irrigante UC-7.....	137
5.	CONCLUSÃO	141
5.1	TRABALHOS FUTUROS	144
5.2	PUBLICAÇÕES.....	144
	REFERÊNCIAS.....	147
	APÊNDICE A – CÓDIGO DO MODELO PROPOSTO PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM UNIDADES CONSUMIDORAS IRRIGANTES ...	157

1. INTRODUÇÃO

As perdas de energia elétrica representam uma enorme preocupação e necessitam de um monitoramento preciso por parte do setor elétrico, uma vez que proporcionam um elevado custo econômico para as concessionárias e seus consumidores, devido ao consumo de energia não faturado. Um sistema elétrico que apresente grandes quantidades de perdas, é considerado ineficiente, pois a energia elétrica produzida não é devidamente aproveitada e os investimentos em expansões e melhorias na capacidade de produção não são completamente desfrutados (EVALDT, 2014).

As perdas de energia elétrica são definidas como a diferença entre a energia requerida pelo sistema elétrico e a energia total comercializada e entregue a unidades consumidoras a esse mesmo sistema (OLIVEIRA, 2009). Nesse contexto, as mesmas referem-se à energia elétrica produzida que passa pelas linhas de transmissão (Rede Básica) ou redes de distribuição, mas que não chega a ser comercializada, seja por motivos técnicos ou comerciais. Essas perdas são classificadas em dois grupos, são eles: perdas técnicas e perdas não técnicas ou comerciais. As perdas técnicas estão relacionadas à transformação de energia elétrica em energia térmica nos condutores (denominado de Efeito Joule), perdas nos núcleos dos transformadores, perdas dielétricas, entre outras (ANEEL, 2015). Já as perdas não técnicas, originam-se através de ações ilícitas por consumidores e de processos comerciais das concessionárias de energia elétrica.

As perdas não técnicas, também denominadas de perdas comerciais, objetivo desse estudo, correspondem à diferença entre as perdas globais ou totais de energia elétrica e as perdas técnicas, e estão associadas em grande parte, à gestão das concessionárias e às características socioeconômicas das áreas de concessão (ANEEL, 2019). Estas perdas são causadas principalmente por ações de furtos de energia elétrica (conexões clandestinas na rede secundária de distribuição), falhas ou fraudes nos medidores (equipamento defeituoso ou ação intencional de violação para registrar menor consumo de energia) e irregularidades no faturamento da concessionária (estimativa equivocada do consumo e instalações sem medidores) (AGÜERO, 2012; AHMAD, 2017; BUZAU et al., 2019).

Dessa forma, devido aos fatores que estão associados a natureza de perdas não técnicas, é improvável encontrar esse tipo de perdas nos sistemas de geração e transmissão, isto é, sua presença é mais específica nos sistemas de distribuição, pois estes estão mais

susceptíveis a ações ilegais intencionais pelos usuários (MADRIGAL; RICO; UZCATEGUI, 2017).

De acordo com Zhang, Gao e Sun (2018), a presença de perdas não técnicas nos sistemas de potência em larga escala, pode impactar em alguns aspectos como: (i) sobrecarga nos sistemas de distribuição, afetando o desempenho ou mesmo danificando os equipamentos presentes na rede elétrica; (ii) interrupções constantes no fornecimento de energia elétricas das unidades consumidoras irregulares; (iii) quedas de energia e *blackouts* durante os períodos de pico, devido a carga adicional imprevisível causada pelas perdas comerciais; (iv) prejuízos econômicos as empresas de energia, impondo tarifas adicionais excessivas a todos os consumidores para recuperar o déficit causado por essas perdas, prejudicando consumidores regulares; (v) riscos de choques elétricos na pessoa que pratica a atividade de furto de energia e (vi) operação inadequada das redes de distribuição, com alta possibilidade de ocorrência de incêndios em condições climáticas extremas, colocando a comunidade em risco.

A identificação de perdas não técnicas em um sistema elétrico, ainda é considerado um processo avaliativo complexo para as concessionárias de energia elétrica. Segundo Madrigal, Rico e Uzcategui (2017), o cálculo de perdas não técnicas em redes de distribuição ainda é um enorme desafio, devido a fatores como: (i) o número de circuitos envolvidos aumenta consideravelmente em relação ao número de elementos que devem ser considerados na avaliação de perdas técnicas que devem ser considerados na avaliação de perdas técnicas nas redes elétricas; (ii) as taxas de amostras ou de registro nas redes de distribuição são determinados por processos de faturamento com poucas possibilidades de determinar com precisão o comportamento horário da demanda e (iii) detectar a ocorrência de um evento ilícito em meio a uma dinâmica de assentamentos irregulares importantes.

O procedimento mais comum para detectar perdas não técnicas é a inspeção local. Em alguns países, as inspeções manuais são programadas durante um período específico para detectar atividades ilícitas de furto de energia. Entretanto, uma desvantagem óbvia de realizar-se a inspeção no local, é que não é uma solução realista vistoriar fisicamente todas as unidades consumidoras. Além disso, existem diversos fatores que dificultam a prática da fiscalização local, como os elevados custos com equipes de peritos, tempo necessário para inspecionar grandes áreas e a própria dificuldade de investigação, pois em muitos casos, as irregularidades não são perceptíveis com uma simples inspeção visual. Dessa forma, essa prática acaba apresentando uma resposta com baixa eficácia para o problema de detecção de perdas não técnicas (EVALDT, 2018; GHORI et al., 2020).

Diante dessas adversidades, diversas metodologias para detecção de perdas não

técnicas são propostas com o intuito de auxiliar as equipes de inspeção, restringindo as regiões de busca e indicando os consumidores suspeitos de irregularidades (EVALDT, 2018). As principais metodologias empregadas atualmente na identificação de perdas não técnicas são desenvolvidas a partir da diferença entre perdas globais e técnicas e, da aplicabilidade de métodos estatísticos e técnicas de Inteligência Artificial (IA).

Metodologias utilizando modelos matemáticos computacionais, estão cada vez mais presentes em análises nos sistemas de potência. As aplicações dessas metodologias, consistem em apresentar respostas mais precisas e eficazes em diversas áreas de estudo de um sistema elétrico. Para o problema de perdas não técnicas, as mesmas possibilitam identificar situações irregulares, como fraudes ou erros de medições, através da mineração de dados, classificação dos perfis de consumidores e monitoramento de medidores eletrônicos de unidades consumidoras. Dessa forma, seu uso é imprescindível para as concessionárias de energia elétrica, pois possibilita a redução de custos com o equipes de peritos, limitando as regiões de busca e indicando as unidades consumidoras suspeitas de irregularidades, além de proporcionar uma maior autonomia e controle sobre suas áreas de concessão.

Os métodos de detecção de perdas não técnicas são categorizados como orientados a dados, orientados a rede e híbridos (HAMMERSCHMITT et al., 2020; MESSINIS; HATZIARGYRIOU, 2018a). Atualmente, metodologias orientadas a dados vem sendo aplicadas em maior escala para o estudo de identificação de perdas não técnicas. Segundo Messinis e Hatziargyriou (2018a), os principais algoritmos utilizados são: *Support Vector Machine* (SVM), Redes Neurais Artificiais (RNA), *Optimum Path Forrest* (OPF), *Decision Trees*, *Random Forest*, *Nearest neighbor* (k-NN), *Naive Bayes Classifier* (NB), *Generalized Additive Model* (GAM), *Self Organizing Map* (SOM), Lógica Fuzzy, Modelos de Regressão Linear e técnicas de agrupamento de dados, conhecido como *Clustering* (*k-Means*, *Fuzzy c-Means* e *Density Based Spatial Clustering of Applications with Noise* (DBSCAN)).

As ocorrências de perdas não técnicas são situações comuns encontradas nas redes de distribuição localizadas nas áreas rurais. Isso se deve, a alguns fatores como: o elevado consumo de energia elétrica dos sistemas de irrigação, acarretando altos custos financeiros a esses consumidores, e a presença de unidades consumidoras alocadas em regiões remotas ou de difícil acessibilidade, o que acaba dificultando a prestação de serviço da distribuidora (medições, manutenção e inspeções para averiguar irregularidades) e incentivando a atividade de fraude de energia, com a adulteração de medidores para registrar um consumo inferior ao efetivamente utilizado. Além disso, as grandes extensões dos alimentadores rurais é um fator que contribui para a dificuldade de inspeção. Dessa forma, as metodologias empregando

modelos matemáticos computacionais podem ser uma ferramenta imprescindível para as distribuidoras no combate as perdas não técnicas nessas regiões.

O consumo de energia elétrica em áreas é direcionado em grande parte para a produção agroindustrial. No estado do Rio Grande do Sul, destacam-se o cultivo e a produtividade do arroz irrigado, sendo este, o maior produtor do Brasil. A energia elétrica é empregada para o funcionamento dos sistemas de bombeamento de água. Os mesmos, são compostos por bombas hidráulicas, motores de acionamentos, tubulações, e peças especiais, e, sua funcionalidade é distribuir a água captada de uma fonte disponível (rios, lagos, riachos, arroios ou barragens) para canais de distribuição de água em todas da parcelas da lavoura de arroz. A atividade de irrigação representa uma parcela muito significativa no consumo de energia, pois as estações de bombeamento normalmente estão localizadas no final de uma rede alimentadora rural, concentrada e de uso praticamente contínuo durante o período da safra (PFITSCHER et al., 2012).

1.1 MOTIVAÇÃO E JUSTIFICATIVA

As perdas não técnicas representam uma preocupação expressiva para as concessionárias, pois afetam diretamente a qualidade do fornecimento de energia elétrica, e os seus prejuízos financeiros são extremamente significativos para o setor comercial das empresas. Além disso, a identificação das perdas não técnicas é um processo complexo, principalmente em instalações alocadas em regiões rurais, normalmente, distantes dos centros urbanos e de difíceis acesso.

Dessa forma, tornar o processo de identificação e inspeção de perdas não técnicas em alimentadores rurais mais simples e preciso, é uma alternativa essencial para as empresas de energia, pois algumas adversidades podem ser amenizadas, tais como, despesas de deslocamento com equipes para inspecionar extensas áreas e o tempo para investigação de instalações irregulares nessas regiões.

A partir do contexto apresentado, essa dissertação retrata o problema de identificação de perdas não técnicas nos alimentadores localizados em áreas rurais. Dessa forma, esse estudo envolve conjuntos de unidades consumidoras rurais que utilizam sistemas de bombeamento de água para a atividade de irrigação na lavoura de arroz. Esses sistemas, também são conhecidos como levantes. O propósito da realização desse estudo, está relacionado com a escassez de trabalhos sobre perdas não técnicas para esse tipo de

consumidores. Além disso, os sistemas de irrigação normalmente apresentam um elevado consumo sazonal, principalmente, no período corresponde a safra do arroz irrigado, sendo extremamente relevantes nos alimentadores das distribuidoras.

1.2 OBJETIVOS GERAIS E ESPECÍFICOS

O objetivo geral desse trabalho consiste em:

- Desenvolver uma metodologia para detectar possíveis ocorrências de perdas não técnicas em unidades consumidoras alocadas em áreas rurais com cultivo de arroz irrigado, relacionando o consumo de energia elétrica com as características das lavouras e dos sistemas de irrigação, empregando técnicas de agrupamento e classificação de dados.

Com o intuito de atender o objetivo geral, essa dissertação possui os seguintes objetivos específicos:

- Analisar e compreender a relação do consumo de energia elétrica dos clientes irrigantes com as características das lavouras de arroz e dos sistemas de irrigação;
- Realizar o agrupamento (*clustering*) de dados de consumo de energia elétrica em um conjunto de unidades consumidoras, empregando o método *k-Means*;
- Classificar unidades consumidoras com base em seu perfil de consumo utilizando o algoritmo classificador de *Random Forest*;
- Estimar o consumo de energia elétrica de unidades consumidoras irrigantes;
- Aplicar esse estudo em um conjunto de unidades consumidoras irrigantes reais do município de Uruguaiana/RS.

1.3 ESTADO-DA-ARTE

Essa seção apresenta alguns trabalhos relevantes e recentes na literatura, relacionados ao tema de detecção de perdas não técnicas. Nesse contexto, serão apresentadas algumas metodologias elaboradas a partir de análises estatísticas, modelos matemáticos de inteligência artificial e aplicações de métodos supervisionados e não-supervisionados para orientação de dados.

Em Ghori et al. (2020) é proposta uma metodologia para identificação das características mais relevantes em unidades consumidoras que venham a auxiliar no processo

de detecção das perdas não técnicas. Estes conjuntos de dados são filtrados e classificados devido a sua respectiva importância, a partir de um algoritmo de seleção desenvolvido pelos autores, denominado de *Incremental Feature Selection* (IFS) (em tradução livre, Seleção de Recurso Incremental). Os métodos classificadores *CatBoost*, *Decision Trees* e *k-NN*, são utilizados para identificar possíveis ocorrências de perdas não técnicas em consumidores de uma empresa de energia elétrica do Paquistão, a partir do conjunto de dados característicos definidos pelo algoritmo IFS.

O trabalho desenvolvido por Raggi et al. (2020) apresenta uma técnica de análise de dados para detecção e localização de perdas não técnicas ocasionadas por conexões ilegais a redes primárias e secundária de distribuição, utilizando informações extraídas de medidores inteligentes instalados junto a unidades consumidoras. A técnica de análise de dados empregada nesse estudo, é fundamentada em detectar dados incorretos em um determinado conjunto de medições. Essa técnica é conhecida como *Bad Data Analysis* (em tradução livre, Análise de Dados Ruins). A formulação de *Weighted Least Squares* (WLS) (em tradução livre, Mínimos Quadrados Ponderados) foi utilizada para representar um circuito equivalente monofásico na estrutura matemática da metodologia. Para a validação do modelo proposto nesse trabalho, estudos de casos foram realizados em um alimentador real de distribuição.

Em Sharma e Majumdar (2020) é apresentado um modelo dinâmico não supervisionado de detecção de perdas não técnicas em tempo real. O modelo de aprendizagem de máquina desenvolvido pelos os autores, é denominado de *Recursive Transform Learning* (em tradução livre, Aprendizagem por Transformação Recursiva), e é responsável por estimar o consumo da unidade consumidora baseado nos registros anteriores de energia elétrica faturada, previstos a partir do uso de técnicas de análises de séries temporais. Dessa forma, para detectar a ocorrência de perdas não técnicas em um conjunto de consumidores, o modelo realiza a diferença entre o consumo previsto e o real, baseado em um limite calculado sobre a incerteza.

Um modelo de detecção de perdas não técnicas baseado nos métodos de *Decision Tree*, *Random Forest* e *Gradient Boosting*, é desenvolvido em Tehrani, Moghaddam e Asadi (2020). Esses algoritmos classificadores são implementados e executados em dados de consumo de energia elétrica extraídos de um pequeno conjunto de unidades consumidoras de uma rede inteligente. Dessa forma, o modelo proposto pelos autores identifica perdas comerciais causadas por furtos de energia. A metodologia ainda emprega o algoritmo *k-Means* para o agrupamento de dados e seleção de clientes com perfis de consumo semelhantes.

Em Buzau et al. (2019) são desenvolvidas Redes Neurais Híbridas com o objetivo de detectar perdas não técnicas em medidores inteligentes. A arquitetura proposta pelos autores, consiste em analisar perfis de consumos diários, demanda contratada e informações geográficas de determinados conjuntos de unidades consumidoras de uma distribuidora de energia elétrica da Espanha. O perfil de carga diário é analisado por uma técnica de memória de longo prazo, conhecida como *Long Short-Term Memory* (LSTM), enquanto que os demais dados são tratados através de uma Rede Neural de múltiplas camadas, intitulada de *Multi-Layer Perceptrons* (MLP).

Em Luya e Pedrasa (2019) é proposto um método de detecção de perdas não técnicas em redes secundárias de distribuição utilizando Regressão Linear. Nesse trabalho, os autores consideraram como efeito das perdas comerciais, apenas a atividade de furto de energia. Dessa forma, ações como conexões clandestinas de carga ao sistema de distribuição e desvios *bypass* em medidores de energia elétrica, foram modelados na plataforma *Open Distribution System Simulator* (OpenDSS®), com o intuito de simular os efeitos das perdas não técnicas na rede de baixa tensão. A partir das medições extraídas nas simulações do IDE de análises de sistemas elétricos, o algoritmo de Regressão Linear é empregado para identificar os clientes irregulares e estimar a quantidade de energia elétrica furtada da rede de distribuição.

Em Wang et al. (2019) é analisada a resposta de diferentes tipos de algoritmos de clusterização na detecção de perdas não técnicas em sistemas elétricos de média tensão. Os algoritmos de agrupamento de dados *k-Means*, DBSCAN, *Spectral Clustering* e *Gaussian Mixture Model* (GMM), são empregados pelos os autores para detectar perdas não técnicas ocasionadas por furtos de energia elétrica e fraudes de medidores em um conjunto de unidades consumidores de uma concessionária de energia. O algoritmo GMM apresentou as melhores respostas para o estudo de caso proposto pelos autores nesse trabalho.

O trabalho proposto por Toma et al. (2019) apresenta uma metodologia para detectar perdas não técnicas relacionadas a atividade de furtos de energia elétrica em Redes Elétricas Inteligentes (*Smart Grids*) utilizando o algoritmo de SVM. Com isso, os autores buscam eliminar ou reduzir os índices de perdas não técnicas em um conjunto de unidades consumidoras. O objetivo da aplicabilidade do algoritmo de SVM nesse estudo, é para analisar e treinar os dados registrados nos medidores inteligentes desses consumidores, e dessa forma, classifica-los em regulares ou irregulares com base em seu histórico e perfil de consumo de energia elétrica.

Outra abordagem utilizando informações extraídas de medidores inteligentes é apresentada em Persson e Lindskog (2019). Nesse trabalho, os autores propõem um método

para a detecção e localização de perdas não técnicas em sistemas de distribuição a partir de aferições nos valores de tensão em medidores inteligentes. Para a modelagem das perdas não técnicas, foram consideradas variações estatísticas (análise de regressão linear) no comportamento das cargas no alimentador. A identificação de possíveis ocorrências de perdas não técnicas é determinada a partir de um valor percentual definido de perdas por energia transferida pela carga. Por fim, para a localização das perdas não técnicas, os autores propuseram três estratégias de detecção baseadas nas medições de tensão e energia das cargas, foram elas: erro de tensão no barramento, correlação entre tensão e energia e correlação entre estimativa e medição de tensão.

Em Fernandes et al. (2019) é proposta uma metodologia de detecção de perdas não técnicas em sistemas de distribuição utilizando o algoritmo classificador *Probabilistic Optimum-Path Forest* (P-OPF). Esse estudo é empregado em um conjunto de dados privados de uma concessionária de energia elétrica brasileira, para avaliação de possíveis unidades consumidoras suspeitas de irregularidades com base na resposta demonstrada pelo algoritmo. A abordagem proposta pelos autores, foi comparada com as respostas apresentadas pelos demais métodos de inteligência artificial aplicados ao processo de identificação de perdas não técnicas, tais como: SVM, *Logistic Regression*, NB e *Particle Swarm Optimization* (PSO).

Em Araujo, Almeida e Mello (2019) é apresentado um comparativo entre os algoritmos de Redes Neurais, *Model Neural Network* (Nnet) e *Model Averaged Neural Networks* (AavNNet), aplicados à detecção de perdas não técnicas em unidades consumidoras. Para esse trabalho, as perdas não técnicas estão relacionadas apenas para as atividades ilícitas de furto e fraude de energia elétrica. As análises com os modelos de redes neurais são realizadas a partir dos históricos de consumos de energia elétrica e outros dados característicos de um conjunto de unidades consumidoras. Dessa forma, os algoritmos buscam classificar o consumidor como suspeito ou regular. Os autores enfatizam que a principal contribuição desse trabalho, é determinar uma composição eficiente de algoritmos de previsão capazes de prover os melhores resultados na identificação de unidades consumidoras irregulares.

Em Ghori et al. (2019) é realizada uma análise de desempenho de diferentes tipos de classificadores de aprendizado de máquina para detecção de perdas não técnicas. Esses estudos, são realizados em um conjunto de unidades consumidoras de uma empresa de fornecimento de energia elétrica localizada no Paquistão. Os autores aferiram as perdas não técnicas a partir de um modelo de métrica de avaliação, utilizando os seguintes tipos de

classificadores de aprendizado de máquinas: NB, *Linear Discriminant Analysis* (LDA), *Quadratic Discriminant Analysis* (QDA), *Logistic Regression*, *Stochastic Gradient Descent* (SGD), SVM, *Decision Trees*, k-NN, Redes Neurais e *Ensemble Methods* (AdaBoost, CatBoost, LightGBM e XGBoost). Os dois últimos classificadores citados apresentaram as respostas mais precisas para a detecção de perdas não técnicas para esse conjunto de consumidores.

O trabalho proposto por Guerrero et al. (2018) apresenta uma metodologia para melhorar a precisão do processo de detecção de perdas não técnicas em uma concessionária de energia elétrica da Espanha. A arquitetura desenvolvida pelos os autores é estruturada em dois módulos coordenados. A primeira etapa, consiste em filtrar as unidades consumidoras utilizando as técnicas de *Text Mining* (em tradução livre, Mineração de texto) e Redes Neurais Artificiais, a partir de informações de inspeções anteriores realizadas pela empresa de energia. O segundo módulo é a etapa de mineração de dados, onde analisou-se o histórico de consumo dos clientes, utilizando um modelo fundamentado nos métodos de *Decision Tree* e *Self-Organizing Map* (em tradução livre, Mapa Auto-Organizável). Esse último módulo, resultou em uma lista de unidades consumidoras selecionadas para inspeção.

Em Leite e Mantovani (2018) é desenvolvido uma metodologia de detecção e localização de perdas não técnicas causadas por diferentes tipos de ataques cibernéticos em sistemas modernos de distribuição. Segundo os autores, essas atividades ilícitas estão associadas ao acesso não autorizado a banco de dados das concessionárias e as práticas de adulteração digital de medidores inteligentes. O algoritmo *A-Star* (A^*) é empregado para detectar os pontos de consumo sujeitos a prática de irregularidades no registro de sua energia faturada. Por fim, os autores utilizaram uma ferramenta computacional com base em informações extraídas de sistemas de informações geográficas (*Geographic Information System* (GIS)), para determinar as coordenadas dos pontos de consumos irregulares identificados pelo algoritmo anterior.

Um estudo utilizando técnicas de mineração de dados aplicadas ao processo de identificação de perdas não técnicas em sistemas de distribuição é desenvolvido em Monteiro e Maciel (2018). Essas técnicas tem por objetivo, executar a tarefa de agrupamento de dados. Dessa forma, os algoritmos de clusterização, *k-Means* e *k-Medoids* são utilizados para determinar os grupos de unidades consumidoras suspeitas de irregularidades, a partir de atributos extraídos de cada consumidor em um conjunto de informações de uma concessionária de energia elétrica.

Em Messinis e Hatzigargyriou (2018b) é apresentado uma metodologia de identificação

de perdas não técnicas utilizando métodos de mineração de dados não supervisionados. Para esse trabalho, os autores elaboraram um modelo de fraudes de energia elétrica para a representação das perdas comerciais. O modelo em questão, compreendeu a quantidade de energia fraudada pela unidade consumidora e o dia do ano em que a ação ilegal foi iniciada. As técnicas de classificação de dados empregadas na metodologia são: *Gaussian Distribution* (MGD), *Local Outlier Factor* (LOF), *k-Means*, Lógica Fuzzy, DBSCAN e SOM. Métricas de avaliação são calculadas para cada conjunto classificador a fim de avaliar a melhor precisão. Os testes foram realizados em um banco de dados de um sistema de distribuição real.

Em Evaldt (2018) é apresentada uma proposta de identificação de perdas não técnicas em alimentadores rurais contendo estações de bombeamento para irrigação de lavouras de arroz. A metodologia desenvolvida pelo autor, emprega um Sistema Neural Artificial constituídos de RNAs que utilizam como entradas, correlações dos padrões de consumo de energia elétrica, características da área irrigada e condições climáticas do período de irrigação. Para os estudos de casos, foram considerados um conjunto de unidades consumidoras irrigantes a partir de uma base de dados reais de safras de lavouras de arroz irrigado do estado do Rio Grande do Sul, para um período de cinco anos. Dessa forma, é indicado pelos resultados das análises, o risco percentual de perdas não técnicas para cada consumidor avaliado.

O trabalho proposto por Madrigal, Rico e Uzcategui (2017) apresenta uma metodologia para estimar e identificar perdas não técnicas em redes de distribuição causadas por roubo e fraude de energia elétrica. O método desenvolvido pelos autores, é fundamentado em medições de variáveis estatisticamente selecionadas por amostragem aleatória simples em diversos setores de um sistema elétrico, tais como: roubos, fraudes, tipo e condições da rede elétrica, condições de medidores, número de residências com acesso à rede, tipo de população, níveis socioeconômicos, vandalismo, entre outros. Além disso, um modelo matemático de furto e fraude de energia elétrica foi desenvolvido através de uma análise de regressão linear múltipla, com o intuito de identificar as variáveis mais influentes na ocorrência dessas ações ilícitas em determinadas regiões do alimentador.

Em Terciyanlı, Emre e Caliskan (2017) é apresentado uma modelagem computacional utilizando *Score Based* aplicado a detecção de perdas não técnicas em redes de distribuição. A metodologia proposta pelos autores é composta por três etapas. No primeiro estágio, é atribuído uma pontuação a cada medidor, considerando a área em que a unidade consumidora se encontra. Na segunda etapa, é realizada um agrupamento de dados utilizando *Fuzzy C-means* para encontrar e classificar consumidores com perfis de consumo semelhantes,

atribuindo uma pontuação a cada conjunto encontrado. No último cenário, são efetuadas estimativas de consumo nas instalações com base na demanda fornecida pela distribuidora. Essas estimativas são comparadas com o consumo real registrado e a diferença entre ambos corresponde a uma pontuação para este cenário. Dessa forma, uma pontuação final é constituída a partir dos resultados obtidos pelos três cenários, e utilizada pelos autores para apontar possíveis consumidores irregulares.

O trabalho desenvolvido por Chatterjee et al. (2017) apresenta um sistema de detecção de perdas não técnicas baseado nos dados cadastrais de unidades consumidoras, como o consumo de energia elétrica no período de meia hora e seus respectivas localizações geográficas. Redes Neurais Recorrentes (*Recurrent Neural Network* (RNN)) são utilizadas para detectar irregularidades no consumo de energia na área de concessão analisada. O algoritmo desenvolvido nesse estudo, ainda empregou unidades de memória de longo prazo (LSTM) para processar os dados sequenciais de consumo de energia elétrica.

Em Bezerra et al. (2016) é apresentada uma metodologia para a detecção e estimativa de perdas não técnicas em alimentadores de distribuição, a partir das tarifas de energia dos consumidores e medições das potências ativas e reativas na subestação. A metodologia proposta pelo autor, utiliza o *Power Summation Method* (PSM) (em tradução livre, Método do Somatório das Potências) para a cálculo do fluxo de potência do sistema elétrico em questão, onde o mesmo é alterado para incluir as medições de injeções de potência ativa e reativa disponíveis na saída do alimentador na subestação referente.

Uma metodologia de detecção de perdas não técnicas utilizando o algoritmo de OPF de forma não supervisionada é apresentada em Passos Júnior et al. (2016). Essa técnica é empregada para identificar consumidores irregulares de perfis comerciais e industriais, extraídos de um conjunto de dados de uma empresa brasileira de energia elétrica. Esse trabalho apresenta algumas contribuições como o uso de técnicas de Distribuição Gaussiana para modelar o problema das perdas não técnicas e a aplicação do algoritmo OPF para estimar os parâmetros dessas técnicas. Por fim, a abordagem proposta pelos autores foi comparada com algumas técnicas não supervisionadas aplicadas ao problema de detecção de perdas comerciais, como o *k-Means*, *GMM*, *Affinity propagation* e *Birch*, onde a mesma apresentou resultados mais eficientes.

Em Huang, Lo e Lu (2013) é apresentada uma abordagem utilizando estimativas de estado e análise de variância para detectar perdas não técnicas em um alimentador. A estimativa de estado é empregada pelos autores, para definir os critérios de ocorrências de irregularidades no uso da energia elétrica, como fraudes e adulterações em medidores de

energia elétrica, bem como localizar as mesmas em um conjunto de unidades consumidoras. Após isso, com o auxílio da técnica de análise de variância (*Analysis of Variance* (ANOVA)), foram estabelecidas uma relação de prováveis clientes irregulares para a distribuidora, com base nas informações obtidas pela etapa anterior. Os testes foram efetuados em um conjunto de dados de um alimentador de distribuição de uma companhia de energia elétrica de Taiwan.

Uma metodologia de treinamento de Redes Neurais do tipo MLP utilizando uma técnica meta-heurística aplicada a detecção de perdas não técnicas, é desenvolvida em Pereira et al. (2013). O método heurístico empregado nesse estudo, é o algoritmo de otimização denominado *Charged System Search* (CSS) (em tradução livre, Busca de Sistemas Carregados), que é baseado no princípio físico do campo elétrico, observando as interações entre partículas eletricamente carregadas. A abordagem proposta pelos autores, tem por objetivo identificar a ocorrência de perdas não técnicas em conjuntos de unidades consumidoras de uma concessionária de energia elétrica no Brasil e comparar os resultados obtidos pelo método com outras técnicas de otimização. Esse trabalho destacou-se por apresentar uma método inovador de detecção perdas não técnicas.

Em Silva (2013) é desenvolvido um modelo para detecção de perdas não técnicas utilizando Lógica *Fuzzy*, denominada nesse trabalho por lógica nebulosa, como principal agente preditivo. O principal objetivo desse estudo, é melhorar os critérios de seleção de unidades consumidoras para fins de inspeção, e assim, reduzir as perdas não técnicas em um sistema elétrico. Dessa forma, foi elaborado um algoritmo com as principais variáveis pertinentes ao problema, tais como: grupo de tensão, clientes ativos, consumidores residenciais, unidades consumidoras em áreas de risco. A partir disso, é construído um sistema baseado em regras, através de um controlador nebuloso aplicando o Método Mandani para determinar as instalações mais favoráveis à inspeção pela distribuidora.

Em Ribeiro et al. (2012) é proposta uma metodologia de otimização de investimento aplicada as ações que busquem identificar e reduzir as perdas não técnicas em um sistema de distribuição. A abordagem utilizada pelos os autores, consiste em um modelo estatístico utilizando dados históricos de consumo que são processados em um algoritmo baseado em árvores de decisão (*Decision Tree*) para maximizar a função objetivo desse estudo. Além disso, esse trabalho proporcionou a elaboração de uma ferramenta computacional para uma distribuidora de energia elétrica no Brasil, a fim de observar e compreender esse estudo em seus alimentadores.

A partir da contextualização dos trabalhos que compõem o estado-da-arte desta dissertação, observa-se que o problema da identificação de perdas não técnicas vem sendo

explorado nos diversos segmentos dos sistemas de potência, desde subestações a medidores de redes elétricas inteligentes. A grande parte dos autores empregam dados reais de unidades consumidoras para validação dos estudos de casos de suas metodologias propostas. No entanto, nota-se que ainda há uma baixa quantidade de estudos na literatura sobre esse problema nos alimentador rurais, sendo abordado com tema principal apenas em Evaldt (2018).

Os métodos *k-Means* e *Random Forest*, empregados na metodologia proposta nessa dissertação, também foram aplicados de forma individual ou conjunta, em problemas de identificação de perdas não técnicas nos trabalhos desenvolvidos por Tehrani, Moghaddam, Asadi (2020); Wang et al. (2019); Monteiro e Maciel (2018); Messinis e Hatziargyriou (2018b) e Passos Júnior et al. (2016). Além disso, a estimativa do consumo, como indicativo de possíveis ocorrências de perdas não técnicas, são apresentadas como parte do desenvolvimento das metodologias elaboradas por Tehrani, Moghaddam e Asadi (2020) e Terciyarli, Emre, Caliskan (2017).

1.4 CONTRIBUIÇÕES DO TRABALHO

Diferente da grande parte das metodologias presentes na literatura, que abordam o processo de identificação de perdas não técnicas em equipamentos elétricos, blocos de carga localizados em centros urbanos e redes de distribuição de média tensão, este trabalho emprega o estudo da detecção das perdas não técnicas em unidades consumidoras localizadas em áreas rurais, especificamente em lavouras de arroz que utilizem a energia elétrica para a atividade de irrigação, onde há uma baixa demanda de estudos relacionados ao tema específico.

A aplicação de técnicas de agrupamento e classificação de dados vem sendo bastante utilizadas em metodologias de identificação de perdas não técnicas em unidades consumidoras. Dessa forma, trabalhando com informações relativas ao histórico de consumo e características dos sistemas de irrigação e do clima, torna-se possível empregar essas técnicas para desenvolver uma ferramenta com excelente eficácia, englobando todas as características possíveis dos consumidores.

Desse modo, esse trabalho pretende contribuir para a literatura com os seguintes pontos:

- Desenvolver uma metodologia de identificação de possíveis ocorrências de perdas não técnicas em unidades consumidoras irrigantes, aplicando dados característicos disponíveis (consumo de energia elétrica e dados dos sistemas de irrigação);
- Definir classes de perfis de consumo em um conjunto de unidades consumidoras através de técnicas de agrupamento de dados;
- Elaborar indicadores para averiguar possíveis ocorrências de perdas não técnicas em unidades consumidoras irrigantes, verificando eventuais irregularidades nos seus respectivos registros mensais de consumo de energia elétrica;
- Estabelecer uma rotina de cálculo de estimativa de consumo das unidades consumidoras irrigantes.

1.5 DELIMITAÇÕES DO TRABALHO

Este trabalho tem como propósito principal, a identificação de possíveis ocorrências de perdas não técnicas em unidades consumidoras localizadas em áreas rurais com cultivo de arroz irrigado, a partir de correlações dos registros mensais de consumo de energia elétrica e dados característicos das lavouras. Como parte do desenvolvimento da metodologia proposta para essa dissertação, serão realizadas estimativas mensais de consumo de energia elétrica em consumidores irrigantes, como um indicativo para avaliar possíveis irregularidades nessas instalações. Para isso, serão utilizados como parâmetros de entrada, a altura manométrica total, vazão, rendimento do sistema de irrigação tempo de funcionamento das estações de bombeamento e as áreas de plantio da lavoura, sendo a última variável citada, fundamental para o resultado final da previsão de consumo. No entanto, algumas adversidades em relação as áreas cultivadas foram encontradas durante o desenvolvimento desse trabalho, tais como incertezas, imprecisões e em muitos casos, ausências de informações para associar a alguma consumidor rural próximo. Dessa forma, serão utilizados como estudos de casos para validação da metodologia proposta, unidades consumidoras irrigantes em que as áreas de plantio de suas respectivas lavouras de arroz são conhecidas de forma completa ou parcial.

Além disso, para essas análises, não foram considerados a influência das variáveis meteorológicas no consumo de energia elétrica do consumidor rural, podendo restringir essa proposta de formulação em análises mais complexas. Os conceitos dos aspectos meteorológicos no cenário da cultura do arroz irrigado, serão brevemente apresentados no capítulo de revisão bibliográfica (Capítulo 2).

1.6 ORGANIZAÇÃO DOS CAPÍTULOS

Esta dissertação foi dividida em cinco capítulos, conforme apresentada a seguir:

- Capítulo 1: No presente capítulo introdutório, é realizada contextualização do tema abordado, definição dos objetivos do trabalho e apresentação do estado-da-arte de estudos recentes presentes na literatura sobre a identificação de perdas não técnicas, bem como as contribuições e delimitações da dissertação;
- Capítulo 2: Apresenta a revisão bibliográfica, dividida em três partes principais: caracterização das lavouras de arroz irrigado, dos sistemas de irrigação por inundação e das perdas não técnicas;
- Capítulo 3: Nesse capítulo, é apresentada a metodologia desenvolvida para identificação de perdas não técnicas em unidades consumidoras irrigantes;
- Capítulo 4: Apresenta os estudos de casos para validação da metodologia proposta, retratando os dados característicos das unidades consumidoras irrigantes avaliadas pelo modelo preditivo, bem como os resultados e discussões;
- Capítulo 5: Nesse capítulo, serão apresentadas as considerações finais sobre a metodologia proposta e os resultados obtidos, além das sugestões de continuidade para trabalhos futuros.

2. PERDAS NÃO TÉCNICAS EM CONSUMIDORES IRRIGANTES DE LAVOURAS DE ARROZ

Nesse capítulo será apresentada uma contextualização geral sobre os principais temas abordados nesse trabalho. Inicialmente, serão demonstradas as principais características das lavouras de arroz irrigado, tais como: aspectos socioeconômicos, peculiaridades do sistema de cultivo por irrigação, métodos de irrigação, influências das variáveis meteorológicas, perdas de águas nas lavouras e particularidades das regiões orizícolas do estado do Rio Grande do Sul. Após isso, serão descritos o processo de irrigação na lavoura de arroz e os principais equipamentos que compõem uma estação de bombeamento de água. Por fim, será salientado um tópico sobre perdas não técnicas em sistemas elétricos, relatando alguns pontos como: principais causas associadas a esse problema, presença dessas perdas nos alimentadores rurais, histórico, impactos nas tarifas de energia e aspectos regulatórios no Brasil.

2.1 LAVOURAS DE ARROZ

O arroz é o segundo cereal mais cultivado no mundo, ocupando uma área aproximada de 161 milhões de hectares e produção de cerca de 756,5 milhões de toneladas, representando 29% do total de grãos utilizados na alimentação humana (SOSBAI, 2018). O arroz é considerado um alimento básico e essencial para mais da metade da população mundial. O continente asiático destaca-se no cultivo desse grão, responsável por aproximadamente 90% da produção mundial (SOUZA, 2015). Os países com as maiores produções em ordem decrescente são: China, Índia, Indonésia, Bangladesh, Vietnã, Tailândia, Myanmar, Filipinas e Brasil.

O Brasil participa com 78% da produção do Mercosul (na média de 2009/10 até 2017/18), seguido pelo Uruguai (8,0%), Argentina (8,0%) e por último, o Paraguai (6,0%) (SOSBAI, 2018). Desse modo, o cultivo de arroz apresenta grande representatividade na agricultura brasileira. A produção para a safra 2020/21 foi de 11,1 milhões de toneladas, redução de 0,9% em relação à safra anterior (2019/20). A produção nacional de arroz tem sua maior concentração na Região Sul, responsável por mais de 80% da oferta, com grande

destaque para os estados de Santa Catarina e principalmente, o Rio Grande do Sul (CONAB, 2021a).

De acordo com o 7º Levantamento da Safra 2020/21 no Boletim de Grãos da Companhia Nacional de Abastecimento (CONAB), o Rio Grande do Sul durante a última safra (2019/20) apresentou a maior produção total de arroz no Brasil, cerca de 56,07%. O segundo maior produtor foi Santa Catarina, também na região sul, com 8,86% da produção nacional, seguido pelos estados de Tocantins (7,50%), Mato Grosso (7,26%) e Maranhão (5,64%) (CONAB, 2021b).

Dessa forma, o arroz é uma cultura que está presente em todas as regiões brasileiras, sendo comercializada e consumida por todas as classes sociais. Esse cereal apresenta enorme importância socioeconômica, pois é responsável por suprir a população brasileiras com um importante aporte de calorias e proteínas na sua nutrição.

2.1.1 Cultivo de Arroz Irrigado

Segundo Uberti (2017), o cultivo do arroz basicamente divide-se em dois grupos, são eles: cultivo de sequeiro, também denominado de “cultivo em terras altas”, e cultivo de arroz irrigado.

O cultivo de sequeiro é tipicamente comum nas região Centro-Oeste, onde o ecossistema do cerrado é predominante. O plantio é realizado de forma que o período de maior exigência hídrica da planta coincida com a época mais chuvosa da região (UBERTI, 2017). Por se tratar de solos com baixa capacidade de retenção de água, os produtores destas regiões podem aderir a irrigação suplementar por aspersão, a fim de compensar possíveis irregularidades na distribuição das chuvas. De acordo com Marcolan et al. (2008), o cultivo de arroz em terras alta, tende a desaparecer devido aos altos impactos ambientais ocasionados após a derrubada de vegetação nativa, como a mudança climática, a severa perda de biodiversidade e o elevado custo financeiro para conversão de sistemas naturais. Os estados com as maiores produção são: Mato Grosso, Maranhão, Roraima, Piauí, Pará e Tocantins. A Figura 2.1 ilustra uma lavoura de arroz com cultivo de sequeiro sem irrigação.

O arroz irrigado, objeto de estudo desse trabalho, é responsável por cerca de 75% da produção nacional, com produtividade superior a 7.000 kg/ha (UBERTI, 2017). Na Safra 2020/21, a área total semeada de arroz irrigado no Brasil ultrapassou os 1.300 mil hectares. Esse sistema de cultivo está presente em todas as regiões do país, destacando-se a Região Sul,

Figura 2.1 – Cultivo de arroz de sequeiro sem irrigação.



Fonte: Embrapa (2020).

que é responsável atualmente, por aproximadamente 85% da produção total de arroz irrigado no país. Dentre os estados que compõe essa região, o Rio Grande do Sul é o maior produtor, com 84,93% de área semeada (946,70 mil hectares), enquanto que Santa Catarina e Paraná apresentam parcelas bem inferiores, cerca de 13,42% (149,60 mil hectares) e 1,65% (18,40 mil hectares), respectivamente (CONAB, 2021b).

Segundo Evaldt (2018), os produtores de arroz do Rio Grande do Sul, plantaram cerca de 1,1 milhões de hectares no ano de 2016. Isso equivale a uma área 22 vezes maior que a cidade de Porto Alegre, capital do estado, e corresponde a 3,1% do território do Rio Grande do Sul. Além disso, a irrigação corresponde a 5,28% do custo de produção de arroz irrigado. Estima-se que durante o ano de 2016, o custo total foi próximo de R\$ 6.715,00 por hectare plantado.

O arroz irrigado tipicamente é cultivado em ecossistemas de várzea. Os mesmos, são encontrados nas planícies de rios, lagoas e lagunas, formando-se em diversas condições de deficiência de drenagem (hidroformismo). Essa característica permite manter o solo saturado em períodos com maiores níveis de precipitação pluviométrica. A drenagem deficiente está relacionada com a ocorrência de horizontes argilosos, que, por disporem de uma condutividade hidráulica muito baixa, dificultam a percolação da água no terreno. Estas

características, geralmente desfavoráveis as práticas de outros tipos de cultivo, tornam-se apropriadas para a cultura do arroz irrigado, possibilitando a manutenção de uma lâmina de água sobre o superfície do solo Além disso, os solos de várzea ocupam extensas áreas com relevo variando de plano a suavemente ondulado, apresentam baixas altitudes (0-200 m) (DUTRA et al., 2017; LORENSI, 2011).

No Rio Grande do Sul, são encontrados cerca de 5,4 milhões de hectares de várzeas e em Santa Catarina, aproximadamente 684 mil hectares. No Paraná, estima-se que existem aproximadamente 400 mil hectares, totalizando uma área de cerca de 6,5 milhões de hectares de várzeas na Região Sul do Brasil (ALONÇO et al., 2004).

De acordo com Uberti (2017), o arroz irrigado é considerado uma cultura temporária de verão. Seu período de safra tem início no mês de Setembro, realizando a preparação do solo. Já o plantio ocorre no período entres os meses de Outubro a Dezembro e a colheita entre o final de Janeiro e o início de Abril, variando de acordo com os cultivares plantados e da região de plantio.

O arroz irrigado necessita de irrigação suplementar, sendo o principal método empregado, o de irrigação por inundação. No entanto, o método de irrigação por aspersão através de um pivô central, vem sendo praticado recentemente em algumas lavouras de arroz na Região da Fronteira Oeste do Rio Grande do Sul. Estima-se que com o uso da aspersão via pivô central, pode-se obter uma economia de mais de 50% no uso da água em comparação ao sistema tradicional de irrigação por inundação (EMBRAPA, 2020a).

As Figuras 2.2 e 2.3 apresentam lavouras de arroz irrigado pelo método de inundação e aspersão por pivô central, respectivamente.

A irrigação suplementar dispõem de demanda hídrica suficiente para saturar o solo, formar a lâmina de água, compensar a evapotranspiração, e repor as perdas por percolação vertical e as perdas laterais nos canais de irrigação. Estima-se, que é necessário um volume de água entre 8.000 e 14.000 m³/ha para irrigação de uma safra, proveniente de levantes e precipitações somados (UBERTI, 2017).

As principais vantagens do método de irrigação por inundação são: reduções nos custos relativos a mão de obra; controle de ervas daninhas e máximo aproveitamento da água provenientes das chuvas. Por outro lado, a irrigação por inundação apresenta algumas desvantagens, tais como: maiores custos de água, necessidade de sistematização da lavoura para terrenos que apresentem declividades inferiores a 2%, restrição para culturas resistentes a submersão e perda de área de plantio para construção das taipas e canais de distribuição (CASTRO, 2003).

Figura 2.2 – Cultivo de arroz irrigado por inundação.



Fonte: Uberti (2017).

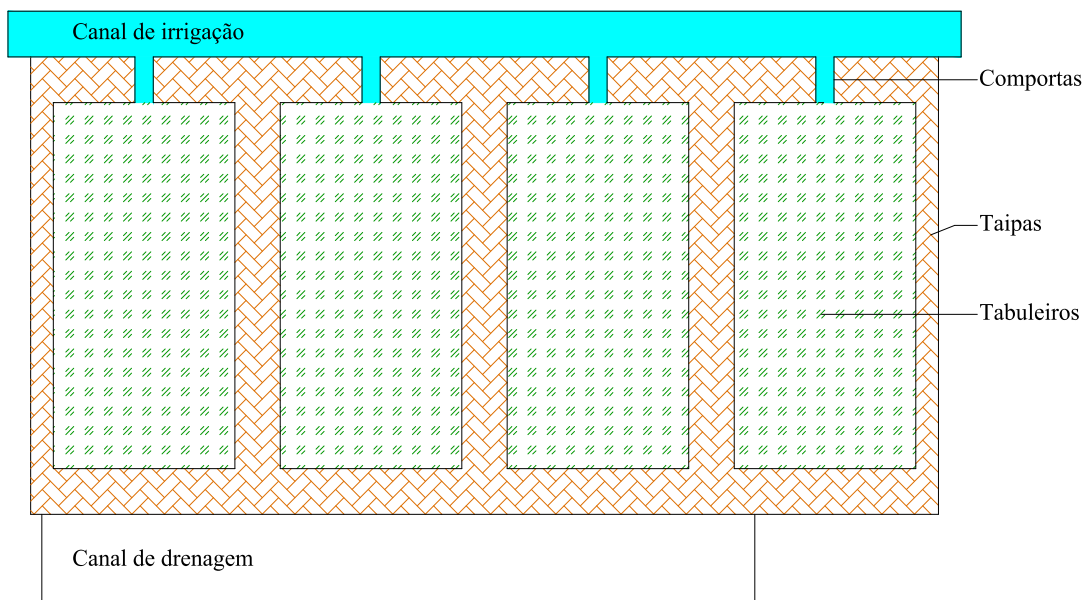
Figura 2.3 – Cultivo de arroz irrigado por pivô central.



Fonte: Scivittaro e Parfitt (2011).

Segundo Castro (2003), para utilizar o método de irrigação por inundação é necessário a aplicação de uma lâmina de água em toda a lavoura de arroz. Para isso, são construídos quadros de irrigação, que são delimitados por taipas, além de canais de irrigação e drenagem. Para o caso em que o reservatório de água estiver localizado em uma cota inferior aos tabuleiros de arroz, é necessário uma estação de bombeamento para transportar a água do reservatório até os canais de irrigação. Dessa forma, os principais componentes do método de irrigação são: reservatório de água, estações de bombeamento, tabuleiros, taipas, canais de irrigação e drenagem, sifões ou comportas para a entrada de água na lavoura. A Figura 2.4 apresenta um esquema de um sistema de irrigação por inundação.

Figura 2.4 – Esquema de um sistema de irrigação por inundação.



Fonte: Adaptado de Castro (2003).

De acordo com Evaldt (2018), para a produção de arroz irrigado são empregadas três modalidades de irrigação, são elas: contínua, estática e de solo saturado. A irrigação por inundação contínua, caracteriza-se por utilizar um enorme volume de água e com baixa eficiência energética, pois o sistema de irrigação permanece ligado em praticamente todo o período de desenvolvimento da cultura até alguns dias antes da colheita. Já a irrigação por inundação estática, a lâmina de água é preservada entre 5 e 10 cm, todavia com comunicação entre os diversos quadros da lavoura. Por fim, a irrigação de solo saturado, o mesmo é mantido encharcado, com uma lâmina mínima de água. O método de irrigação por inundação contínua é predominante em todas as regiões orizícolas do Rio Grande do Sul.

2.1.2 Impacto das Variáveis Meteorológicas nas Lavouras de Arroz Irrigado

As variáveis meteorológicas são um conjunto de dados contendo informações extremamente relevantes a serem consideradas no regime de irrigação em uma lavoura de arroz. Esses elementos estão diretamente relacionados com o processo de evapotranspiração e a necessidade hídrica nas áreas de plantio da respectiva cultura (UBERTI, 2017). As principais variáveis meteorológicas utilizadas são: precipitação pluviométrica, evaporação, velocidade do vento, radiação solar, temperatura e umidade relativa do ar.

A precipitação pluviométrica colabora para a reposição e o acúmulo de água nas lavouras de arroz, enquanto que outras variáveis meteorológicas como a evaporação, radiação solar e velocidade do vento, contribuem para acelerar o processo de evapotranspiração nas áreas de plantio de arroz irrigado.

A precipitação pluviométrica contribui de maneira favorável para a reposição das perdas hídricas por percolação e por evapotranspiração nas lavouras de arroz. Esta variável meteorológica, é uma importante forma de irrigação natural da agricultura. Quanto maior a precipitação no período da safra do arroz irrigado, menor será a necessidade de irrigação complementar na lavoura, e conseqüentemente, menor será o consumo de energia elétrica do sistema de irrigação (UBERTI, 2017).

A evaporação representa a transferência de água de uma superfície úmida para a atmosfera e contribui para o aumento da necessidade de reposição de água nas áreas de plantio. Quanto maior a evaporação medida em uma lavoura de arroz, maior será a necessidade de reposição hídrica e, por consequência, maior será o consumo de energia elétrica do sistema irrigante. A evaporação está diretamente relacionada a outras variáveis meteorológicas como radiação solar, pressão atmosférica, temperatura e umidade relativa do ar (UBERTI, 2017).

O vento é responsável por acelerar o processo de evapotranspiração das lavouras de arroz irrigado, além de contribuir para as perdas laterais, devido ao arraste superficial da água para drenos. Desta forma, uma alta velocidade de vento impacta em um acréscimo na necessidade de irrigação da área de plantio, implicando em uma elevação no consumo de energia elétrica pelo sistema de irrigação (UBERTI, 2017).

A radiação solar é uma variável meteorológica essencial no desenvolvimento e na produtividade das sementes de arroz, principalmente durante as fases fenológicas reprodutiva e de maturação. Por outro lado, elevados índices de radiação solar colaboram com a reposição hídrica nas lavouras de arroz, uma vez que auxiliam de forma muito significativa no processo

de evaporação da água. Dessa forma, essa variável implica em um acréscimo na necessidade de irrigação na área de plantio, e conseqüentemente em um aumento de consumo de energia elétrica do sistema irrigante.

A radiação solar absorvida pela atmosfera e o calor emitido pelo solo elevam a temperatura do ar. O calor sensível do ar circulante transfere energia para a cultura e em seguida, exerce algum controle sobre a taxa de evapotranspiração no solo. Em dias ensolarados e com altas temperaturas, as perdas de água por evapotranspiração são mais elevadas em relação a dias nublados que apresentem temperaturas mais baixas (ALLEN et al., 2006).

Enquanto a radiação solar combinada com a alta temperatura do ar contribuir significativamente para a evaporação da água, a diferença entre a pressão do vapor de água na superfície de evapotranspiração e o ar circulante é um fator determinante para excisão do vapor. Áreas de plantio bem irrigadas em regiões quentes e secas devido à alta disponibilidade de energia e a capacidade de extrair vapor da atmosfera. Por outro lado, em regiões tropicais úmidas, apesar da elevada radiação solar, a alta umidade relativa do ar reduzirá a demanda de evapotranspiração. Nesse último caso, devido ao ar já estar próximo da saturação, o mesmo pode absorver uma quantidade menor de água e, portanto, apresentar uma taxa de evapotranspiração menor em comparação as regiões áridas (ALLEN et al., 2006).

2.1.3 Perdas de Água nas Lavouras de Arroz

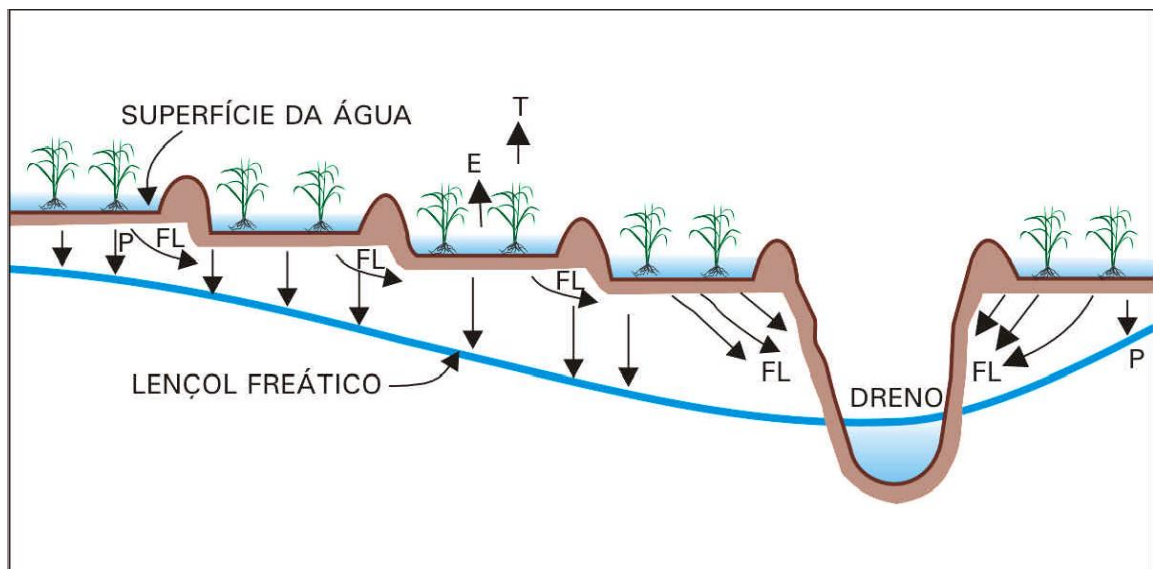
As perdas de água nos tabuleiros de arroz são fomentadas por três maneiras, são elas: evapotranspiração, percolação e fluxos laterais. As mesmas, contribuem significativamente para a baixa eficiência do uso da água nas lavouras de arroz e por consequência, em reduções na produtividade desses grãos.

A evapotranspiração é o movimento ascendente da água do solo em direção a atmosfera, a partir da transpiração das plantas e da evaporação da superfície solo-água. No início da cultura, a evapotranspiração é na sua maior parte, concebida pela evaporação da superfície da água, mas, à medida que a fase fenológica da cultura evolui e a planta começa a sombrear a água, a evaporação decresce e a transpiração da planta aumenta (STONE, 2005).

A percolação é a infiltração vertical hídrica no sentido do lençol freático. Já o fluxo lateral é a perda de água para os drenos, o que ocorre nas últimas parcelas da lavoura, que são adjacentes com áreas não plantadas, canais de drenagem, rios ou riachos. Este fluxo lateral

será pequeno, quanto menor for desnível entre a parcela de arroz e o dreno. Os fluxos laterais que sucedem de uma parcela para outra de nível mais baixo não são considerados perdas do sistema de irrigação (UBERTI, 2017). As drenagens por percolação e fluxo lateral podem ser observadas na Figura 2.5.

Figura 2.5 – Representação das perdas de água por percolação (P), fluxo lateral (FL), evaporação (E) e transpiração (T) em uma lavoura de arroz.



Fonte: Stone (2005).

Os valores típicos de percolação e fluxo lateral em lavouras de arroz variam de 1 a 5 mm dia^{-1} em solos muito argilosos e de 25 a 30 mm dia^{-1} em arenosos. Para a evapotranspiração, os valores típicos para as lavouras de arroz são de 4 a 5 mm dia^{-1} , para a estação chuvosa, e 6 a 7 mm dia^{-1} , na estação seca (TABBAL et al., 2002; UBERTI, 2017).

2.1.4 Regiões Orizícolas do Rio Grande do Sul

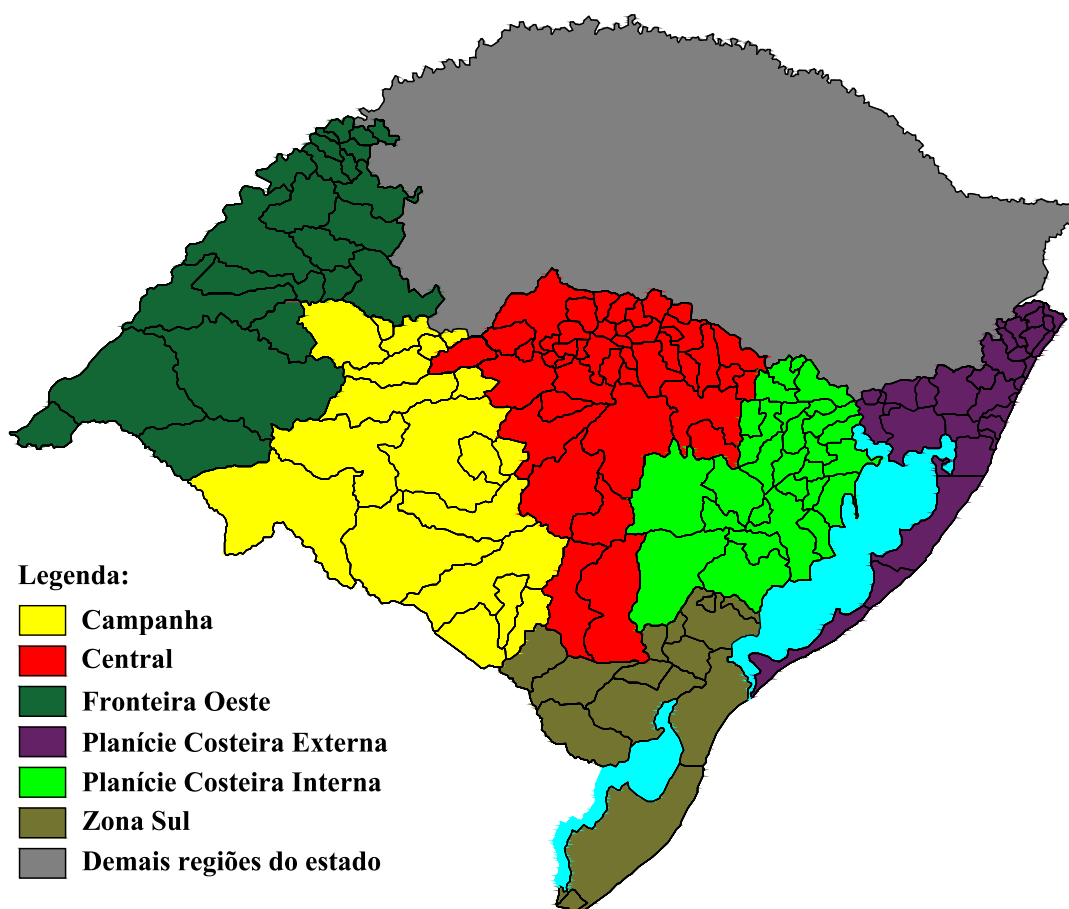
O Rio Grande do Sul é o estado que concentra a maiores áreas de lavouras irrigadas no Brasil, com mais de 900 mil hectares cultivados por safra. A orizicultura gaúcha caracteriza-se pela predominância do cultivo irrigado extensivo, dentro de um sistema empresarial, efetivado principalmente por grandes e médios produtores, onde é tipicamente conduzido em rotação com pastagens (EMBRAPA, 2020b). As regiões produtoras de arroz irrigado no estado do Rio Grande do Sul, são definidas a partir das divisões políticas administrativas

estabelecidas pelo o Instituto Rio-Grandense do Arroz (IRGA), entidade responsável por incentivar, coordenar e superintender a defesa da produção, da indústria e do comércio do arroz produzido em terras gaúchas e promover o desenvolvimento sustentável do setor orizícola do Rio Grande do Sul (IRGA, 2020a).

Desse modo, o arroz irrigado é cultivado nas seguintes regiões (IRGA, 2020a): Fronteira Oeste, Depressão Central, Campanha, Sul, Costa Externa da Laguna dos Patos e Costa Interna da Laguna dos Patos. Essas regiões apresentam diferenças, principalmente, quanto à topografia, clima, solos, disponibilidade de água para irrigação, tamanho da lavoura, variações em termos de área cultivada, produção e rendimento médio (KLERING, 2012).

A Figura 2.6 ilustra o mapa do Rio Grande do Sul com as respectivas divisões regionais produtoras de arroz irrigado.

Figura 2.6 – Mapa do Rio Grande do Sul com suas respectivas regiões orizícolas.



Fonte: Elaborado pelo o autor.

Região da Fronteira Oeste

A Fronteira Oeste é a região que apresenta o maior potencial produtivo do estado, e consequentemente a maior intenção de plantio de arroz irrigado na Safra 2020/21, cerca de 29,89% (IRGA, 2020c). Estima-se que na última safra (2019/20), a produtividade média na Fronteira Oeste atingiu 9.091 kg/ha e foram colhidos 2.590 toneladas de grãos (IRGA, 2020b). Os municípios de Uruguaiana, Itaqui e Alegrete são as principais produtoras dessa região.

A região da Fronteira Oeste apresenta como características climáticas, elevados índices de radiação solar durante o período do verão, atributo essencial para a cultura do arroz, e que colabora para que a região apresente um ótimo valor de produtividade média por safra. Além disso, a grande variação de amplitude térmica entre as estações do inverno e verão, é uma característica peculiar no clima da Fronteira Oeste (KUHN, 2017).

As lavouras localizadas nessa região são de médio e grande porte, e as barragens são as principais fontes de água para a irrigação (ALONÇO et al., 2004). A Fronteira Oeste possui extensas áreas de várzea, com ocorrência principalmente de Planossolos, localizados nas encostas dos rios Uruguai, Butuí, Quaraí, Ibicuí e Ibirapuitã. Esses solos apresentam drenagem ruim, permanecendo saturados por um longo período. Mesmo apresentando um manejo dificultado, a região apresenta excelente aptidão para a cultura do arroz (KUHN, 2017).

Região da Zona Sul

A Zona Sul é a região que apresentou a segunda maior intenção de plantio de arroz irrigado na Safra 2020/21, aproximadamente 16,60% (IRGA, 2020c). Estima-se que na última safra (2019/20), a produtividade média na Zona Sul atingiu 8.788 kg/ha e foram colhidos 1.318 toneladas de grãos (IRGA, 2020b). O município de Santa Vitória do Palmar é o principal produtor dessa região.

A região da Zona Sul possui verões quentes e chuvas distribuídas ao longo do ano, mesmo que a ausência de níveis de precipitação pluvial seja bastante comum durante o período do verão, devido as altas temperaturas e o baixo regime hídrico anual. O clima é bem definido em termos de temperatura durante todo o ano (TRAPP, 2018). Além disso, a região apresenta grande disponibilidade de água para a irrigação, já que está localizada entre a Laguna dos Patos e as Lagoas Mirim e Mangueira. Entretanto, a água normalmente tem de percorrer grandes distâncias, encarecendo o custo de irrigação (ALONÇO et al., 2004).

Os solos de várzea dessa região caracterizam-se pelo hidromorfismo, apresentando alternância textural abrupta, acidez e baixa fertilidade, sendo a última, a principal limitação no solo dessa região. Em geral são solos planos e aptos para o cultivo do arroz (TRAPP, 2018). Nessa região, a semeadura e a colheita são realizadas em períodos curtos e bem definidos, refletindo num investimento expressivo em máquinas e equipamentos (ALONÇO et al., 2004).

Região da Campanha

A Campanha é a região que apresentou a terceira maior intenção de plantio irrigado na Safra 2020/21, cerca de 14,55% (IRGA, 2020c). Estima-se que na última safra (2019/20), a produtividade média na Campanha atingiu 8.364 kg/ha e foram colhidos 1.150 toneladas de grãos (IRGA, 2020b). O município de Dom Pedrito e São Gabriel são os principais produtores dessa região.

Essa região apresenta todas as estações bem definidas, com verões quentes e invernos relativamente frios, inclusive com ocorrências de geadas nos meses de Julho e Agosto. A Região da Campanha possui uma grande variedade de solos e relevos ondulados, abrangendo áreas com altos declives e coxilhas, de topografias planas e baixas caracterizando as chamadas “várzeas”(CARNEIRO, 2017). Nessa região, os solos de várzea são encontrados juntos aos rios Santa Maria, Ibicuí da Armada, Negro e Jaguarão. Além dos rios, barragens também são utilizadas como fontes para irrigação nas lavouras de arroz localizadas nessa região (ALONÇO et al., 2004; SCIVITTARO; PARFITT, 2011).

Região da Planície Costeira Interna

A Planície Costeira Interna é a região que apresentou a quarta maior intenção de plantio de arroz irrigado na Safra 2020/21, aproximadamente 14,50% (IRGA, 2020c). Estima-se que na última safra (2019/20), a produtividade média na Zona Sul atingiu 7.970 kg/ha e foram colhidos 1.039 toneladas de grãos (IRGA, 2020b). O município de Camaquã é o principal produtor da região.

Essa região apresenta solos encharcados e de fácil manejo de irrigação, com boa absorção e retenção de água, poucos profundos, resistentes a compactação e fertilidade média. São solos considerados aptos para o cultivo do arroz irrigado (FERREIRA NETO, 2011). Os solos de várzea são encontrados próximo a Laguna dos Patos. No ponto de vista das condições meteorológicas, essa região apresenta verões quentes e invernos longos.

Região Central

A Região Central, também denominada de Depressão Central, apresentou a segunda menor intenção de plantio de arroz irrigado na Safra 2020/21, cerca de 13,50% (IRGA, 2020c). Estima-se que na última safra (2019/20), a produtividade média na Zona Sul atingiu 7.689 kg/ha e foram colhidos 982 toneladas de grãos (IRGA, 2020b). O município de Cachoeira do Sul é o principal produtor da região.

Na Região Central, as lavouras de arroz são de pequeno e médio porte, com uso intensivo dos solos, localizados principalmente nas várzea dos rios, sendo estes as fontes de irrigação. Os solos de várzea dessa região são encontrados próximos aos rios Jacuí, Vacaraí, Pardo, Taquari, Caí e Sinos. Além disso, possuem médias e baixas fertilidades (ALONÇO et al., 2004; SCIVITTARO; PARFITT, 2011). A Região Central apresenta verões com altas temperaturas, sendo uma das regiões quentes do estado.

Região da Planície Costeira Externa

A Região da Planície Costeira Externa apresentou a menor intenção de plantio de arroz irrigado na Safra 2020/21, aproximadamente 11,05% (IRGA, 2020c). Estima-se que na última safra (2019/20), a produtividade média na Zona Sul atingiu 7.398 kg/ha e foram colhidos 757 toneladas de grãos (IRGA, 2020b). O municípios de Mostradas e Viamão são os principais produtores dessa região.

No ponto de vista das características meteorológicas, o mês de Setembro é o que apresenta o maior volume de chuvas enquanto, que os meses de Abril e Maio possuem os menores índices de precipitação pluviométrica. Essa região possui um relevo de variações, com várias regiões de várzea (NUNES, 2011). Os solos de várzea nessa região, são encontrados próximos a Laguna dos Patos (SCIVITTARO; PARFITT, 2011).

2.1.5 Plantio por regiões orizícolas e municípios do Rio Grande do Sul

A Tabela 2.1 apresenta um levantamento da intenção de plantio e área semeada durante a última safra nas regiões orizícolas do Rio Grande do Sul. Já na Tabela 2.2, serão listados os principais municípios produtores do estado. Os dados foram disponibilizados pelo IRGA através do Boletim de Evolução de Semeadura da Safra 2020/21. Essas informações foram atualizados pela última vez em 22 de Dezembro de 2020.

Tabela 2.1 – Intenção de plantio e área semeada por região orizícola no Rio Grande do Sul.

Região	Intenção de plantio (ha)	Área semeada (ha)	Área semeada (%)
Fronteira Oeste	289.737	269.938	93,17%
Zona Sul	160.284	163.071	101,70%
Campanha	141.540	127.727	90,24%
Planície Costeira Interna	140.487	138.628	98,68%
Central	130.202	128.890	99%
Planície Costeira Externa	106.989	107.138	100,14%

Fonte: Adaptado de IRGA (2020c).

Tabela 2.2 – Intenção de plantio e área semeada nos principais municípios produtores de arroz do Rio Grande do Sul.

Município	Região	Intenção de plantio (ha)	Área semeada (ha)	Área semeada (%)
Uruguaiana	Fronteira Oeste	78.500	69.551	88,60%
Santa Vitória do Palmar	Zona Sul	68.436	68.618	100,30%
Itaqui	Fronteira Oeste	56.492	58.885	104,20%
Alegrete	Fronteira Oeste	54.837	50.178	91,50%
Dom Pedrito	Campanha	38.300	36.889	96,30%
São Borja	Fronteira Oeste	36.782	37.584	102,20 %
Arrio Grande	Zona Sul	32.528	34.031	101,70%
Camaquã	Planície Costeira Interna	32.302	31.160	96,50%
Mostardas	Planície Costeira Externa	31.620	31.620	100%
São Gabriel	Campanha	27.000	22.350	87,80%
Cachoeira do Sul	Central	26.266	26.051	99,20%
Barra do Quaraí	Fronteira Oeste	25.000	20.300	81,20%
Maçambará	Fronteira Oeste	20.349	18.404	90,40%
Viamão	Planície Costeira Externa	18.698	18.698	100%
Jaguarão	Zona Sul	18.000	18.019	100,10%

Fonte: Adaptado de IRGA (2020c).

Nota-se que pelo dados apresentados nas Tabelas 2.1 e 2.2, a Fronteira Oeste apresenta uma enorme representatividade na produção de arroz irrigado no Rio Grande do Sul, como mencionado. Dentre os quinze municípios listados, cinco estão localizados nessa região, com grande destaque para a cidade de Uruguaiana, sendo o maior produtor desse grão no Brasil. A Zona Sul também representa uma grande parcela no cultivo de arroz irrigado no estado, destacando-se o município de Santa Vitória do Palmar, no extremo sul do Rio Grande do Sul, segundo maior produtor em terras gaúchas.

2.2 SISTEMAS DE IRRIGAÇÃO

Os sistemas de irrigação utilizados nas lavouras de arroz, são constituídos de uma bomba hidráulica, um motor de acionamento, um sistema de transmissão mecânica, tubulações de sucção e de recalque, peças especiais e um sistema de suporte (KÖPP, 2015). Os mesmos, também podem ser denominados de sistemas de levante e estações de bombeamento.

No Rio Grande do Sul, prepondera o sistema de cultivo através de taipas com parcelas em nível e o método de irrigação superficial por inundação contínua, onde a água é colocada no tabuleiro mais elevado e após ter sua lâmina estabelecida, passa ao tabuleiro imediatamente inferior, e assim sucessivamente, até atingir o último tabuleiro (KÖPP, 2015; UBERTI, 2017).

Alguns aspectos são fundamentais no processo de irrigação de uma lavoura de arroz e essenciais para garantir uma excelente produtividade dos grãos. Volume de água por hectare, tempo de irrigação, altura da lâmina de água, são algumas características importantes na irrigação. Dessa forma, a Tabela 2.3 demonstra os dados gerais dos sistemas de irrigação para as lavouras de arroz, apresentados em Evaldt (2018).

A Área Irrigada, representa a área total da lavoura cultivada com arroz e que precisa de irrigação. Áreas irrigadas com até 30 ha, são consideradas como pequenas. Em áreas entre 30 e 150 hectares, as mesmas são classificadas como parcialmente pequenas e médias, entre 150 e 200 ha são áreas puramente médias, e entre 200 e 400 hectares são parcialmente médias e grandes. As áreas superiores de 400 ha são consideradas grandes (UBERTI, 2017).

O dimensionamento dos conjuntos elevatórios é realizado a partir da vazão unitária, que por sua vez é dependente de fatores como lâmina adotada, tempo de bombeamento diário, aspectos físicos do solo e da topografia da área irrigada (KÖPP, 2015). A quantidade de água

Tabela 2.3 – Dados gerais de levante para o arroz irrigado.

Dados de irrigação	Grandezas e Unidades	
Volume por hectare (total)	8.000 m ³ (mín)	14.000 m ³ (máx)
Volume por hectare proveniente do levante	4.200 m ³ (mín)	7.000 m ³ (máx)
Volume por hectare proveniente de precipitação	7.000 m ³ (mín)	9.800 (máx)
Vazão por hectare no início da irrigação	3 L/s/ha (mín)	5 L/s/ha (máx)
Vazão por hectare para manutenção da irrigação	1 L/s/ha (mín)	1,5 L/s/ha (máx)
Tempo de irrigação	80 dias (mín)	100 dias (máx)
Espessura da lâmina de água	2,5 mm (mín)	10 mm (máx)
Altura do levante	< 0 m (gravidade)	50 m (máx)

Fonte: Evaldt (2018).

de reposição está propriamente relacionada com as condições climáticas no período da safra e das características do solo (UBERTI, 2017). Em Köpp (2015), é mencionado que sob o aspecto de engenharia, os fatores determinantes do uso de água são a vazão adotada e tempo de funcionamento das estações de bombeamentos, sendo possível altera-los desde que sejam atendidas as necessidades agrônômicas do processo.

2.2.1 Processo de Irrigação nas Lavouras de Arroz

O processo de irrigação tem início com a captação da água em reservatórios, rios, riachos, barragens, lagos ou arroios, realizada pelos sistemas de bombeamento eletromecânicos, também denominados de bombas de levante. Esses equipamentos são responsáveis por recalcar a água para um reservatório em um nível acima dos tabuleiros do arroz, ou diretamente para os canais de distribuição de água para os tabuleiros, sendo o mais usual (KÖPP, 2015; UBERTI, 2017). As Figuras 2.7 e 2.8 ilustram o início e término do levante em lavouras de arroz, respectivamente.

O controle de entrada de água e de drenagem de taipas são realizados de forma manual pelo o aguador, individuo responsável pela manutenção da lâmina de água contínua em todos os tabuleiros de arroz da lavoura. Entretanto, em muitas situações ocorrem excesso de água em algumas taipas, devido principalmente as grandes extensões das lavouras. Isso representa um elevado desperdício de água e energia elétrica no bombeio, ou ainda um déficit hídrico momentâneo, comprometendo a produtividade da planta. Para as situações em que terreno

Figura 2.7 – Captação da água em uma fonte hídrica pela tubulação de sucção da bomba hidráulica – Início do levante em lavouras de arroz.



Fonte: Elaborado pelo o autor.

Figura 2.8 – Desague no canal de irrigação – Término do levante em lavouras de arroz.



Fonte: Mundstock et al. (2011).

apresente maior declividade, é utilizado um método de cultivo de parcelas em desnível. Nessas circunstâncias, o recalque do sistema de bombeamento alimenta a parcela mais elevada, e as demais recebem a água da parcela anterior por gravidade (UBERTI, 2017).

De acordo com Köpp (2015), estima-se que 79% das estações de bombeamento no Rio Grande do Sul sejam acionados por motores elétricos, 6,7% por motores Diesel e 14,3 % empregam o desnível natural do terreno.

2.2.2 Características das Estações de Bombeamento

Conjunto Bomba-Motor

Como mencionado anteriormente, a bomba hidráulica é um equipamento que possui a função de distribuir a água captada de uma fonte hídrica disponível (rios, lagos, entre outros) para as áreas de plantio através dos dutos de recalque. As bombas são acionadas por motores elétricos e suas dimensões dependem da área de plantio e da quantidade de volume de água exigido para a atividade de irrigação.

As bombas são classificadas de acordo com a sua velocidade específica (η_s) em radial centrifuga lenta ($\eta_s < 80$ RPM), radial centrifuga normal ($80 < \eta_s < 150$ RPM), radial centrifuga rápida ($150 < \eta_s < 365$ RPM), mista ($365 < \eta_s < 547$ RPM), semiaxial ($547 < \eta_s < 1095$ RPM) e axial ($\eta_s > 1095$ RPM) (KÖPP, 2015). Segundo Gomes et al. (2012), a eficiência da bomba está relacionada a velocidade específica e vazão recalçada pela mesma, sendo que os maiores rendimentos são registrados para vazões superiores a 650 L.s^{-1} e velocidades específicas entre 40 e 60 RPM.

Na Região da Fronteira Oeste do Rio Grande do Sul, as bombas hidráulicas mais comuns utilizadas para a irrigação são de característica do tipo turbo com fluxo radial, capazes de transportar grandes vazões a médias alturas podendo apresentar rendimentos superiores a 85%. As perdas de energia se dividem em internas, ocasionadas pelo movimento do fluido, e externas que são oriundas nos mancais e elementos de vedação (GOMES et al., 2012; KÖPP, 2015).

Em relação aos motores que compõem as estações de bombeamento, a maioria são do tipo *standart* e muitos são rebobinados mais de uma vez, devido as constantes quedas de tensão da rede elétrica da distribuidora, que produzem elevadas correntes elétricas na instalação e avaria nos mesmos. A rebobinagem pode reduzir a eficiência do motor de 3 a 4% (KÖPP, 2015).

Uma situação muito comum nas estações de bombeamento em lavouras de arroz, são

motores operando com potência ativa superior a sua potência nominal. Isso se deve, pela variação da altura geométrica ao longo do ciclo de irrigação com consequente aumento da altura manométrica. Desse modo, é importante conhecer o fator de serviço (FS) do motor, que indica a carga permissível que pode ser aplicada continuamente em relação a sua potência nominal. A NBR 7094 estabelece que o FS de um motor pode variar de 1,0 a 1,25, variando de acordo com as características operacionais dos mesmos (KÖPP, 2015).

Os motores elétricos utilizados nesses sistemas de irrigação são do tipo de indução trifásicos, com potências nominais que variam usualmente entre 100 e 300 cv (EVALDT, 2018). As chaves de partidas atualmente utilizadas para o acionamento dos motores do sistema de irrigação são do tipo *soft-starter*. A mesma, efetua a ligação dos motores em rampa suave, possibilitando a eliminação do registro de gaveta, anteriormente utilizado na saída da bomba para produzir inclusão suave da carga através da sua abertura gradual (KÖPP, 2015).

A Figura 2.9 demonstra um conjunto bomba-motor de uma estação de bombeamento utilizada para atividade de irrigação nas lavouras de arroz.

Figura 2.9 – Conjunto Bomba-Motor de um sistema de irrigação.



Apresentado as características das bombas hidráulicas e motores elétricos que compõem as estações de bombeamentos, o próximo tópico tem por objetivo descrever uma breve contextualização de como é realizado a transmissão de energia entre esses equipamentos para contemplar a atividade de irrigação da lavoura de arroz.

Transmissão de Energia

Outro aspecto extremamente relevante para os sistemas de irrigação, é o processo de transmissão de energia entre o motor elétrico e a bomba hidráulica. De acordo com Köpp (2015), o mesmo pode ser realizado por acoplamento direto com luva estática, quando a rotação da bomba é idêntica à rotação do motor de acionamento, ou por acoplamento mecânico com polias e correias, quando a rotação da bomba se diferenciar daquela do motor, o que é mais usual devido a maioria das bombas utilizadas em lavouras de arroz operarem em baixas rotações (inferiores a 1800 RPM). A Figura 2.10 ilustra dois modelos de acoplamento mecânico com polias e correias no conjunto bomba-motor.

Figura 2.10 – Acoplamento por polia e correias (a) eixo horizontal e (b) eixo vertical.



Fonte: Köpp (2015).

O uso de acoplamento por polias e correias proporciona perdas de energia de elétrica e pode ser substituído por alternadores de rotação, chamados de inversores de frequência que modificam a rotação do eixo do motor e por consequência altera as revoluções do eixo da bomba e podem proporcionar bons índices de economia no consumo de energia elétrica.

Entretanto, a baixa qualidade da energia fornecida em áreas rurais ainda é um fator limitante para os inversores de frequência, pois se tratam de equipamentos eletrônicos e necessitam de estabilidade de tensão em todo seu período de funcionamento, dissemelhantemente das chaves de ligação do tipo *soft-starter*, que também são eletrônicas porém acionadas por períodos curtos (KÖPP, 2015).

Regime de Funcionamento

O regime de funcionamento dos sistemas de irrigação é praticamente contínuo durante o período de irrigação na lavoura de arroz. As bombas permanecem ligadas por um intervalo de 21 horas diárias durante um período que varia entre 80 e 100 dias, dependendo do ciclo da cultivar, das características físicas do solo e aspectos meteorológicos da região (UBERTI, 2017).

As estações de bombeamento são desligadas apenas no período de horário de ponta (geralmente das 18 às 21 horas) devido ao enquadramento tarifário. Na região da Fronteira Oeste do Rio Grande do Sul, o horário de ponta para o desligamento dos sistemas de bombeamentos é definido através de um rodízio entre as unidades consumidoras para não sobrecarregar a rede de distribuição com o acionamento de todos os motores simultaneamente (KÖPP, 2015).

A Agência Nacional de Energia Elétrica (ANEEL) ainda incentiva que os consumidores rurais utilizem suas instalações de bombeamento utilizadas nas atividades de irrigação, funcionem em horário especial (Intervalo entre 21:30 e 6:00) com percentual de desconto variando de 70% a 90% dos valores da Tarifa de Energia (TE) e Uso do Sistema de Distribuição (TUSD) da modalidade verde (ANEEL, 2016)

Conforme Uberti (2017), o uso de energia elétrica para irrigação, representa cerca de 3,83% do custo total da safra. Esse percentual é equivalente a R\$ 335 milhões, considerando a área plantada no país. O percentual de despesas com energia elétrica na lavoura, é superior aos gastos com sementes (3,14%). Na região da Fronteira Oeste do Rio Grande do Sul, estima-se que a parcela gasta com energia elétrica ultrapassa 10% do custo total da safra.

Eficiência do uso da água

A eficiência do uso da água no setor agrícola em nível mundial é considerada muita baixa. Estima-se, que apenas cerca de 45% da vazão derivada dos mananciais, são

efetivamente utilizadas pelas culturas. No Rio Grande do Sul, o consumo médio de água por uma lavoura arrozeira é tradicionalmente de $2 \text{ L}\cdot\text{seg}^{-1}\cdot\text{ha}^{-1}$ ($17,3 \text{ mm dia}^{-1}$) com um eficácia de irrigação de 40 a 45% que, em condições convenientes de solo, relevo e manejo de água, pode alcançar os 60% (RIGHES, 2013).

A eficiência do sistema de irrigação para o uso da água, está relacionada a fatores como topografia do terreno, tipo e perfil do solo, características físicas dos horizontes, localização da camada impermeável, fluxo lateral, tamanho dos quadros, planejamento adequado a locação, construção dos drenos e canais de irrigação e cuidados operacionais (RIGHES, 2013). Segundo a Agência Nacional de Águas (ANA), em condições adequadas, o rendimento de aplicação em um sistema de irrigação por inundação pode alcançar um percentual de 60%. No caso de irrigação por aspersão via pivô central, a mesma se torna próxima de 85%, porém com perdas por evaporação e arraste de 8% (ANA, 2019).

Eficiência do uso de energia elétrica

Segundo Loureiro (2009), a eficiência energética em estações de bombeamento pode ser promovida das seguintes maneiras: (i) pelo uso de instalações hidráulicas adequadas; (ii) evitar ou reduzir o consumo de energia elétrica, principalmente no período de horário de ponta; (iii) redução de vazamentos nas tubulações e (iv) melhoria dos rendimentos do conjunto bomba-motor. Dessa forma, torna-se necessário uma análise para identificar as perdas que reduzem a eficiência de um sistema de irrigação.

De acordo com Köpp (2015), o desperdício de energia elétrica em instalações de bombeamento de lavouras de arroz é superior a 50%, e é ocasionado por uma combinação de: consumo hídrico exagerado (em média 20% acima do desejável), baixo rendimento da transmissão mecânica entre motor e bomba, rendimento global baixo do sistema de irrigação (escolhas equivocadas de equipamentos e falta de manutenção em polias) e por altura manométrica muito elevada (erros de dimensionamento).

2.3 PERDAS NÃO TÉCNICAS

Como mencionado no Capítulo 1, as perdas não técnicas, ou perdas comerciais, correspondem à diferença entre as perdas totais ou globais e as perdas técnicas. Essas perdas estão associadas a ineficiências do setor comercial da distribuidora, como problemas de faturamento da energia distribuída, erros de leitura e medição, ou por ações de consumidores,

como furtos e fraudes de energia. Assim, as perdas não técnicas estão relacionadas à gestão da concessionária e às características socioeconômicas das áreas de concessão (ANEEL, 2019). Dessa forma, quaisquer perdas que não possam ser qualificadas como técnicas, são consideradas não técnicas (HUBACK, 2018).

De acordo com Penin (2008) e Huback (2018), as causas mais comuns das perdas não técnicas são:

- **Inexistências de medidores:** A ausência de dispositivos de medição em unidades consumidoras é um fato bastante comum nos países da América Latina. Essas faltas podem ser atribuídas a falhas da concessionária ou a casos previstos em legislação, como bancas de jornal, quiosques, iluminação pública. Para esses casos, o consumo é estimado a partir da carga instalada e fatores de carga e demanda. Essas estimativas podem ocasionar níveis menores ou maiores de perdas não técnicas. A ausência de medidores também pode ser associada a casos de furtos de energia, pois em muitas situações, o consumidor conecta-se a rede elétrica não estando inserido na base comercial da distribuidora;

- **Falha no funcionamento do medidor:** Com o decorrer do tempo, a degradação desse equipamento colabora para o aumento das perdas que não podem ser estimadas por metodologias de perdas técnicas conhecidas. Assim, essas perdas serão colocadas na parcela referente a perdas não técnicas;

- **Erros na leitura dos medidores ou no faturamento das unidades consumidoras:** Esse tipo de situação acontece, quando é realizada uma leitura equivocada no medidor da unidade consumidora, causando uma perda não técnica, que é compensada na leitura do próximo mês. Outra situação típica de estimativa incorreta do consumo de energia, são em unidades consumidoras localizadas em locais remotos ou com difícil acesso (AGÜERO, 2012). Além disso, os medidores podem ser afetados por alguma anomalia que origine um faturamento distorcido. Essas anomalias, podem estar relacionadas aos desgastes sofridos pelo equipamento ao longo de sua vida útil e implicam em registros incorretos no consumo de energia, em muitos casos valores abaixo do consumo real que de fato ocorrer na instalação;

- **Fraude de Energia:** A fraude de energia está relacionado a adulterações no medidor, ou a desvios no ramal de entrada efetuados antes do equipamento que registra o consumo de energia elétrica da instalação. Na maioria dos casos, esse aumento de carga à revelia da distribuidora, em um circuito clandestino, é realizado de modo sofisticado, praticado por ação intencional de violação por parte da unidade consumidora a fim de registrar um menor consumo de energia elétrica na instalação. Outra situação que pode ser

associada a fraude de energia, são os ataques cibernéticos a banco de dados e medidores inteligentes, conforme relatado em Leite e Mantovani (2018). Esses ataques bem sucedidos, devem resultar em perdas não técnicas sempre que os dados de consumo são excluídos ou corrompidos;

• **Furto de energia:** O furto de energia está associado a ligações irregulares realizadas na rede secundária de distribuição, sem o consentimento e autorização da concessionária. A maioria dos furtos da energia elétrica, ocorrem a partir de desvios ou ligações clandestinas, como perfuração de eletroduto ou ganchos de rede para ligação *by-pass*, desvios de energia com instalação de acionamento remoto, fraudes em medidores de forma intencional para reduzir o consumo medido (EVALDT, 2014; REIS, 2005). Os furtos de energia ocorrem principalmente em áreas invadidas ou com baixos índices socioeconômicos, e são popularmente conhecidos como “gatos”.

Diversas são as irregularidades praticadas pelo consumidor que fazem parte de um segmento das perdas comerciais e são caracterizadas como perdas por ação do consumidor. Os tipos de irregularidades mais encontrados são (PENIN, 2008):

- Ligação direta à rede secundária de distribuição;
- Desvio no ramal de entrada (antes do medidor);
- Elemento móvel do medidor bloqueado por meio de perfuração da tampa de vidro;
- Ponteiros do medidor deslocados;
- Ligações do medidor invertidas;
- Terminal de prova aberto;
- Bobina de potencial interrompida;
- Engrenagem do medidor substituída;
- Dentes da engrenagem desgastados;
- Elemento móvel empenado;
- Ponteiro de demanda retrocedido;
- Chave de aferição aberta;
- Lâmina da chave de aferição isolada;
- Condutores de ligação entre a chave de aferição e medidor isolados;
- Condutores de ligação entre a chave de aferição e medidor seccionados;
- Curto-circuito nos secundários dos transformadores de corrente;
- Alimentação do motor de temporização de demanda interrompida;
- Sequência de fases invertidas (reativo);

- Curto-circuito na entrada ou saída do medidor.

O planejamento e a operação do sistema de distribuição podem ser afetados, se no mesmo houver altos índices de perdas não técnicas. Essas perdas em escalas significativas, prejudicam a segurança e qualidade de energia fornecida pela rede elétrica, ocasionando problemas de sobrecargas nas redes e nos transformadores, que conseqüentemente resulta na queima ou redução do tempo de vida útil dos mesmos, acarretando às concessionárias de energia elétrica custos adicionais de manutenção e operação (REIS, 2005).

As perdas não técnicas também afetam os perfis de tensão da rede, refletindo no aumento de quedas de tensão ao longo do alimentador. Dependendo do nível de carregamento do sistema de distribuição, condições operacionais críticas podem ser atingidas, exigindo uma compensação adicional de potência reativa. Dessa forma, esses aspectos salientam a importância de considerar a relevância de perdas não técnicas durante os estudos de planejamento de operação de sistemas de distribuição (BEZERRA et al., 2016).

Além disso, as perdas não técnicas prejudicam o equilíbrio financeiro das empresas de energia, principalmente a ação do furto, que gera problemas na compra e venda de energia elétrica, pois a energia consumida pela unidade consumidora que está praticando o furto, não é considerada no planejamento inicial de compra da energia (REIS, 2005). Assim, as reduções nas receitas ou faturamento das concessionárias ocasionadas principalmente pelas ações de roubo de energia elétrica, são repassadas aos consumidores regulares como forma de compensação através das tarifas (MESSINIS; HATZIARGYRIOU, 2018a).

Segundo ANEEL (2019), os custos relativos as perdas não técnicas no Brasil no ano de 2018, atingiram o valor de R\$ 6,6 bilhões. Nos Estados Unidos, o prejuízo estimado é de aproximadamente US\$ 1,6 bilhões por ano. Já o Canadá, apresenta uma perda anual inferior ao sistema elétrico estadunidense, mais ainda significativa, cerca de US\$ 100 milhões por ano. Em 2004, a Malásia atingiu um recorde no montante de perdas não técnicas, resultando em um custo próximo a US\$ 229 milhões a companhia de energia *Tenaga National Berhad* (DEPURU, 2012). Na Índia, as perdas não técnicas representam mais de 50% da perda total de energia. Essa parcela ocasiona um prejuízo em torno de US\$ 4,5 bilhões anuais, contabilizando todas as empresas de energia elétrica desse país (CHATTERJEE et al., 2017). No Paquistão e em Honduras, os déficits anuais são de US\$ 0,89 bilhões e US\$ 13 milhões, respectivamente (GHORI et al., 2020). A Grécia apresenta uma despesa anual estimada de 150 milhões de Euros (MESSINIS; HATZIARGYRIOU, 2018b).

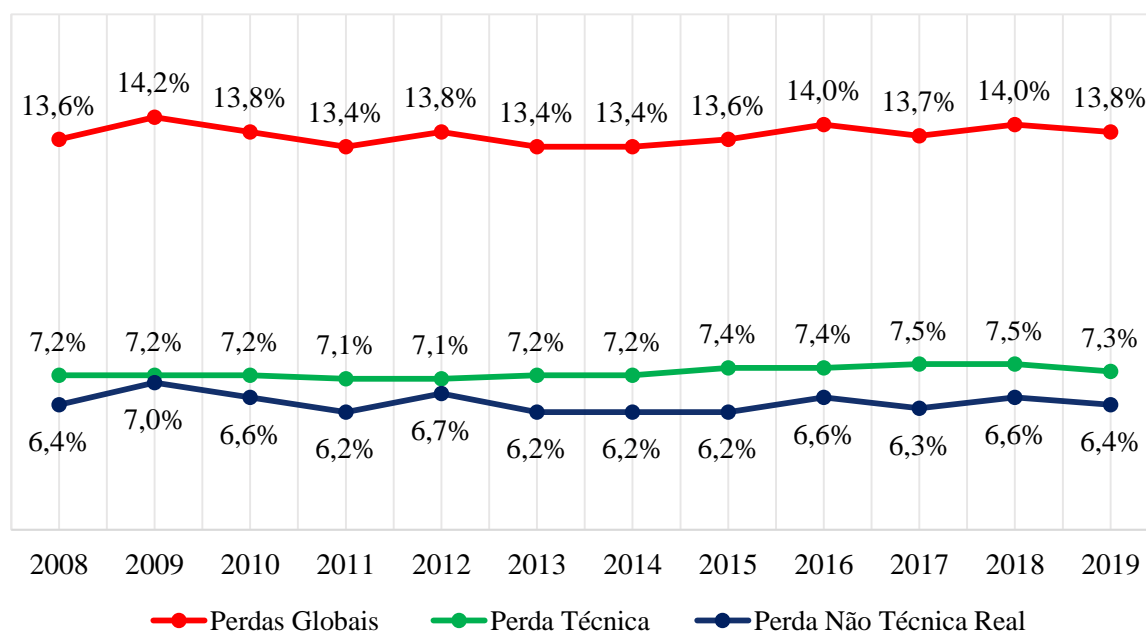
Os menores índices registrados de perdas não técnicas estão em países desenvolvidos ou em emergentes asiáticos, como Alemanha, Áustria, Bélgica, Coreia do Sul, Cingapura,

França, Holanda, Japão e Suíça. Por outro lado, países com baixo grau de desenvolvimento social e econômico, como Albânia, Bangladesh, Myanmar e Nigéria, apresentam elevados níveis de perdas não técnicas e maiores dificuldades no combate ao problema (HUBACK, 2018).

2.3.1 Perdas Não Técnicas no Brasil

Para ilustrar um cenário da presença de perdas não técnicas nos sistemas elétricos de potência do Brasil, a Figura 2.11 apresenta a evolução das perdas globais, técnicas e não técnicas sobre a energia injetada no período entre 2008 e 2019.

Figura 2.11 – Perdas sobre a Energia Injetada no período entre 2008 e 2019.



Fonte: ANEEL (2020).

Nota-se que as perdas sobre a energia injetada no Brasil durante o período de 2008 a 2019, permaneceram praticamente constantes, com pouquíssimas variações entre um ano e outro. Em montantes de energia, as perdas não técnicas corresponderam a 35,90 TWh no Brasil, durante o ano de 2019. Esse valor representa um percentual de 6,44% sobre a energia injetada no país. Se considerada as perdas técnicas, a taxa de energia não utilizada na matriz energética brasileira é de aproximadamente 14%, sendo superior ao consumo de energia elétrica das regiões Norte e Centro-Oeste em 2018 (ANEEL, 2020).

No Brasil, as perdas não técnicas reais são obtidas pela diferença das perdas totais, disponibilizadas pelas distribuidoras, e das perdas técnicas regulatórias, apuradas pela ANEEL. É importante ressaltar que, os valores regulatórios são aqueles que são ratificados na tarifa de energia, enquanto que os valores reais são os que efetivamente acontecem. A diferença de custos entre ambos, é responsabilidade da distribuidora. Além disso, esses valores são definidas pela revisão tarifária periódica de cada distribuidora, a partir dos percentuais regulatórios estabelecidos nas Resoluções Homologatórias (REHs) da ANEEL, que acontecem no intervalo de 3, 4 ou 5 anos. Normalmente, os valores regulatórios das perdas não técnicas são inferiores aos valores praticados pelas distribuidoras, devido a metodologia adotada pela agência regulatória aborda critérios de eficiência, que restringi os repasses das perdas não técnicas reais (ANEEL, 2019).

As concessionárias de grande porte são as responsáveis pela maioria dos montantes de perdas não técnicas no Brasil, devido ao tamanho do mercado e a maior dificuldade no combate a essas perdas. As mesmas, são classificadas como aquelas que possuem um mercado de energia superior a 1 TWh e que atendam mais de 500 mil unidades consumidoras em suas áreas de concessão ou que apresentem mais de 15 mil km de rede elétrica (ANEEL, 2020).

Os níveis de perdas não técnicas estão diretamente associados a características socioeconômicas e de aspectos comportamentais existentes em cada área de concessão, além da política de gestão adotada pelas concessionárias. A homologação das perdas não técnicas são realizadas sobre o mercado de baixa tensão, pois é onde ocorrem a maior incidência desse tipo de perdas (ANEEL, 2019).

2.3.2 Impacto das Perdas Não Técnicas na Estrutura Tarifária do Brasil

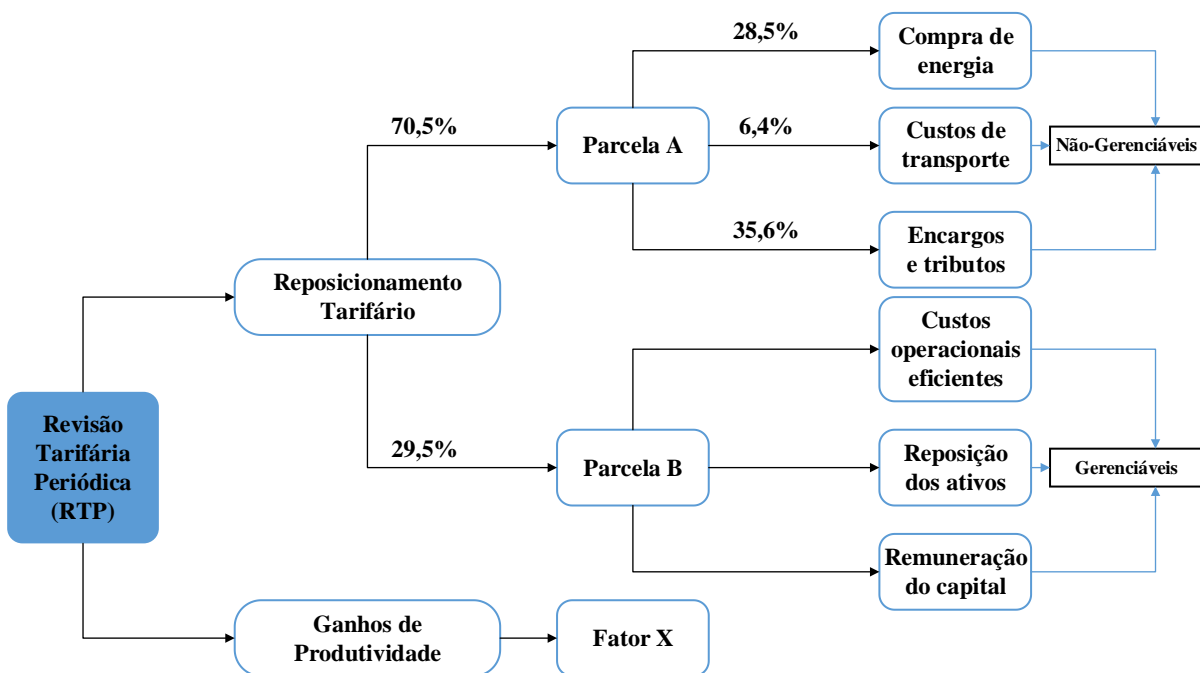
No ano de 2019, as perdas não técnicas no Brasil corresponderam a um custo de aproximadamente R\$ 7,4 bilhões. Entretanto, se considerada as perdas não técnicas regulatórias, calculadas através da metodologia elaborada pela a ANEEL, esse valor ficou em torno de R\$ 5,3 bilhões, representando aos consumidores, cerca de 3% do valor da tarifa de energia elétrica, variando para cada distribuidora (ANEEL, 2020).

De acordo com o decreto n° 4.562 de 2002, que estabelece normas para tarifação e preço de energia elétrica, são consideradas nas tarifas de energia elétrica, as parcelas apropriadas dos custos de transporte e das perdas de energia elétrica, bem como os encargos de conexão e setoriais do segmento de consumo (BRASIL, 2002). Segundo Huback (2018),

os altos níveis de perdas de energia, contribuem para que o Brasil apresente hoje uma das mais altas tarifas de energia elétrica no mundo, implicando assim em elevados índices de inadimplência e furto de energia em diversas regiões do país, principalmente em áreas que ostentem baixas condições socioeconômicas.

A tarifa de energia elétrica no Brasil é calculada a partir da soma da Parcela A e Parcela B da ANEEL, ponderada pelo ganho de produtividade atingido pelas empresas ao longo do ciclo tarifário. As duas principais componentes levam em consideração custos não gerenciáveis (Parcela A) e custos gerenciáveis (Parcela B). A agência regulatória é responsável por estimar as metas eficientes para a Parcela B, enquanto que as distribuidoras são encarregadas pela diferença entre as metas regulatórias e os valores apresentados pelas empresas, não podendo ser repassada à tarifa de energia elétrica (ASEVEDO, 2011). A Figura 2.12 ilustra a composição da tarifa de energia no Brasil.

Figura 2.12 – Composição da Tarifa de Energia Elétrica no Brasil.



Fonte: Adaptado de Asevedo (2011).

Os custos relativos a compra de energia estão em grande parte associados a Parcela A, sendo repassados diretamente para a tarifa. Entretanto, o montante de energia adquirido pelas concessionárias é dimensionado a fim de atender a demanda total da carga, que abrange o mercado efetivamente faturado e o não-faturado. Com isso, os consumidores que são

faturados contraem com o custo total da energia comprada para o seu atendimento e também dos consumidores irregulares. Assim, as perdas não técnicas passaram a ser essenciais para estrutura tarifária, sendo indispensável a definição de um tratamento regulatório para contornar esse problema (ASEVEDO, 2011; HUBACK, 2018).

2.3.3 Tratamento Regulatório de Perdas Não Técnicas no Brasil

A ANEEL como órgão regulador do governo federal, institui definições, regras e metas de redução de perdas não técnicas em seus procedimentos regulatórios. Essa regulação tem por objetivo uniformizar o problema das perdas quanto as diferenças regionais e socioeconômicos para as distribuidoras do Brasil (SILVA, 2013).

A modalidade tarifária é o principal propósito da regulação vigente. O atual regime de regulação empregado no setor de distribuição de energia elétrica é o da Regulação pelo Preço Teto (*Price-Cap*), que se caracteriza por ser uma regulação por incentivos. Neste regime, a ferramenta responsável por incentivar as empresas a buscarem níveis de eficiência mais elevados na gestão do seu negócio está na fixação pelo regulador de parâmetros gerenciáveis durante um período de tempo pré-estabelecido (ANEEL, 2011).

Segundo ANEEL (2011), especificamente para as perdas não técnicas, o incentivo está na fixação do nível de perdas ou da sua trajetória durante um período específico, e posteriormente na avaliação dos resultados a fim de garantir que os ganhos de eficiência sejam repassadas as unidades consumidoras. Dessa forma, as distribuidoras são estimuladas a reprimirem suas perdas a limites estipulados pelo regulador na perspectiva de alcançar ganhos adicionais de receitas ou aproximar as suas perdas de energia elétrica dos limites determinados pelo regulador de forma a reduzir as perdas de receita.

As regras de cálculos das perdas não técnicas estão segmentadas no Submódulo 2.6 dos Procedimentos de Regulação Tarifária (PRORET). O mesmo foi instituído pela ANEEL através da Resolução nº 435 de 2011, e tem como objetivo principal, consolidar a regulamentação acerca dos processos tarifários.

As áreas de concessão de cada distribuidora apresentam características de mercado e socioeconômicas distintas. Assim, a comparação entre as concessionária, somente seria exequível se essas diferenças fossem consideradas. Dessa forma, a ANEEL elaborou um ranking de complexidade socioeconômica, desenvolvido a partir de modelos econométricos, que possibilita comparar o desempenho das perdas não técnicas das distribuidoras, de acordo com o porte e a posição (ANEEL, 2020).

Para a distribuidora cuja perda não técnica será determinada em processo de revisão, esse modelo identifica a existência de uma empresa de referência, denominada de *benchmark*, que comumente está situada em uma área de maior complexidade socioeconômica, mas que pratica perdas menores do que a distribuidora em análise. Com isso, a ANEEL incentiva as distribuidoras a sempre buscarem ações que resultem nas reduções dos seus índices de perdas não técnicas, independentemente do nível regulatório estabelecido, seja para restringir o prejuízo econômico, quando as perdas reais estiverem acima da regulatória, ou para obter ganhos, quando acontecer o oposto. Além disso, eventuais negligências ou ineficiências das concessionárias em combates às perdas não técnicas, não seriam mais repassadas às tarifas de energia, restringindo-se apenas aos níveis regulatórios considerados eficientes (ANEEL, 2020).

2.3.4 Perdas Não Técnicas em Alimentadores Rurais

As redes de distribuição primárias e secundárias que atendem unidades consumidoras presentes em áreas rurais, apresentam características distintas das redes elétricas localizadas em centros urbanos. No âmbito residencial, o uso da energia elétrica é direcionado para a iluminação e eletrodomésticos. Já para o âmbito agroindustrial, a eletricidade é empregada em larga escala para o sistema de bombeamento de água para a irrigação da lavoura, além de equipamentos utilizados para tratamento após a colheita, como compressores de refrigeradores, elevadores de carga, esteiras de transporte de grãos, ventiladores para aeração em silos, entre outros (EVALDT, 2018; UBERTI, 2017).

Segundo Evaldt (2018), Uberti (2017) e Evaldt (2014), as redes de distribuição localizadas em áreas rurais possuem as seguintes particularidades:

- Grandes extensões de redes;
- Baixa densidade populacional por quilômetro de rede;
- Grande dispersão de consumidores, variando entre 0,1 e 5,0 consumidores por quilômetro de rede elétrica;
- Presença de um número expressivo de transformadores de baixa capacidade nominal, a fim de contornar as quedas de tensão em áreas de grande dispersão;
- Curvas de cargas sazonais, impactadas especialmente em regiões com intensa atividade agrícola e agroindustrial, refletindo em um baixo fator de carga dos transformadores. A Eletrobrás considera um fator de carga igual a 0,2 em áreas rurais;

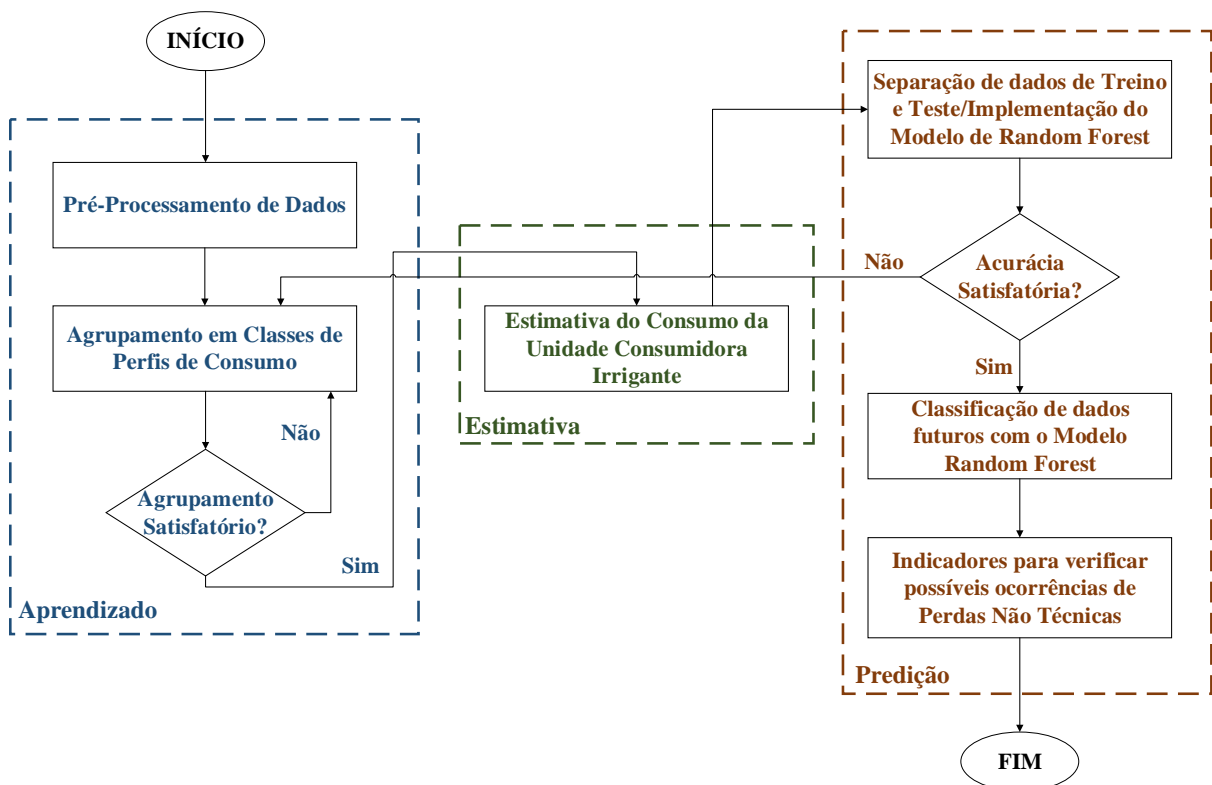
- Inspeção de alguns trechos com acesso dificultado;
- Dificuldade de manutenção dos mesmo padrões de qualidade de energia dos centros urbanos e maior susceptibilidade a falhas;
- Interrupções do fornecimento de energia elétrica devido à sobreposição de vegetação nos cabos (curto-circuito entre fases) e queda de postes (curto-circuito fase-terra) ao longo do tempo, devido as dificuldades de manutenção pela concessionária.

Diante de algumas das adversidades apresentadas, as ocorrências de perdas não técnicas são comuns em alimentadores rurais. Unidades consumidoras com sistemas de irrigação de lavouras de arroz refletem em situações adversas de identificação de perdas comerciais para as distribuidoras, considerando os diferentes processos de irrigação existentes, características climáticas, instalações com difíceis acessos e dificuldades de inspeção (EVALDT, 2018). Além disso, as grandes extensões das redes de distribuição rurais proporcionam elevados custos com equipes de inspeções para as concessionárias de energia elétrica.

3. METODOLOGIA PROPOSTA

A metodologia proposta nesse trabalho tem por objetivo principal, identificar possíveis ocorrências de perdas não técnicas em unidades consumidoras localizadas em áreas rurais, que façam uso da energia elétrica em larga escala para a atividade de irrigação em lavouras de arroz. A mesma, é estruturada a partir de uma arquitetura composta por métodos matemáticos computacionais, cuja funcionalidade é agrupar e classificar 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 possam influenciar no consumo de energia elétrica da instalação. Dessa forma, a Figura 3.1 apresenta um fluxograma exibindo a arquitetura estrutural da metodologia proposta nesse trabalho.

Figura 3.1 – Fluxograma da metodologia proposta.



Fonte: Elaborado pelo o autor.

A metodologia proposta ilustrada pela Figura 3.1, é desenvolvida a partir de um algoritmo escrito em linguagem de programação *Python*, elaborado na plataforma

computacional *The Scientific Python Development Environment* (Spyder®), sendo este, uma ferramenta gratuita de análises em Ambiente Integrado de Desenvolvimento (em inglês, *Integrated Development Environment* (IDE)) que possibilita o estudo de recursos de edição de scripts, testes, *debugging* (em português, depuração de dados) e visualização gráfica. O mesmo foi escolhido por tratar-se de uma multiplataforma com acesso livre e apresentar integração com as principais bibliotecas da linguagem *Python*.

A metodologia para detecção de ocorrências de perdas não técnicas inicia-se a partir da criação de classes de perfil de consumo nas amostras de dados empregadas para esse estudo. Para isso, são utilizados como dados de entrada, os valores de média e desvio padrão de consumo de energia elétrica do ano de 2019 de unidades consumidoras rurais do município de Uruguaiana/RS. Após isso, o algoritmo realiza a estimativa de consumo para cada consumidor irrigante utilizando como entrada, alguns atributos associados ao cultivo da arroz irrigado referentes a Safra 2020/21.

Desse modo, empregando os valores de consumo de energia elétrica faturados pela distribuidora para os anos de 2019 e 2020, o algoritmo verifica se houve a mudança de perfil de consumo para a unidade consumidora comparando a classe alocada durante a etapa de agrupamento (utilizando dados de consumo de 2019 como entrada) e a classe predita pelo modelo preditivo (empregando dados de consumo de 2020 como entrada). A mudança de perfil de consumo serve como um indicador de possível ocorrência de perdas não técnicas. Além disso, outro indicador compara o consumo mensal de energia da instalação com o estimado para o ano de 2020, utilizando um critério percentual pré-estabelecido.

Ainda sobre o fluxograma da metodologia de identificação de perdas não técnicas em unidades consumidoras irrigantes apresentado pela Figura 3.1, o mesmo é segmentado em três estágios de análises e execução de dados (Aprendizado, Estimativa e Predição) subdivididos em diferentes etapas de funcionamento para cada um dos estágios correspondente.

O estágio de aprendizado compreende as etapas de pré-processamento do banco de dados (*dataset*) de unidades consumidoras, o agrupamento das amostras presentes deste conjunto e sua respectiva avaliação para a continuidade de funcionamento do algoritmo, ou seja, se o agrupamento realizado no respectivo conjunto de dados está de acordo com os critérios pré-estabelecidos. 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 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 e implementação de um modelo de classificação utilizando o algoritmo de *Random Forest*, bem como a verificação da acurácia do mesmo e a categorização dos dados futuros para esse modelo classificador. Para a etapa de verificação da precisão do presente modelo desenvolvido, será empregada a métrica de avaliação pela tabela de confusão de dados, considerando um valor percentual de acurácia pré-definido. Por fim, a categorização dos elementos, tem a funcionalidade de averiguar possíveis ocorrências de perdas não técnicas a partir dos perfis de consumo presumidos pelo modelo de previsão e as 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 para detecção de perdas não técnicas em consumidores irrigantes proposta nesse estudo.

3.1 CONJUNTO DE UNIDADES CONSUMIDORAS RURAIS E PRÉ-PROCESSAMENTO DE DADOS

Um banco de dados de cadastro de unidades consumidoras contém diversas informações referentes aos consumidores vinculados a uma determinada concessionária de energia elétrica. Esses elementos são essenciais para os setores técnico e comercial das distribuidoras, permitindo uma maior autonomia e controle sobre as áreas concessão atendidas. Dentre os principais atributos que compõe essa base de dados, destacam-se: identificação do âmbito do consumidor (rural ou urbano), dados mensais de consumo de energia elétrica para cada ano, código de instalação, município da unidade consumidora,

região do cliente, descrição do sistema setorial, latitude ou longitude da instalação (supõe-se: localização do ponto de medição), enumeração do alimentador, status de conexão do consumidor (ativo ou inativo).

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. Além disso, o *dataset* é representado na plataforma computacional em uma tabela de dados de cadastro, onde as linhas representam as unidades consumidoras e as colunas, informações associadas as mesmas (consumo, município, entre outros).

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 faturada do consumidor. Essas informações referente 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, são 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. As colunas dos atributos selecionados no *dataset* são apresentadas pela Tabela 3.1.

Tabela 3.1 – Atributos selecionados do banco de dados.

Atributo	Descrição
COD_INSTALACAO 202001	Código de identificação único para cada unidade consumidora Consumo de Energia: aaaamm
DESVIO_PADRAO MEDIA	Desvio Padrão do consumo da unidade consumidora Média do consumo da unidade consumidora

Fonte: Elaborado pelo o autor.

Na terceira e última etapa de pré-processamento foram eliminados do conjunto de dados, clientes com dados faltantes, nulos e com erros de cadastro, e também amostras que se diferenciam drasticamente de todas as outras, conhecidas como *outliers*, de maneira que apenas pontos dentro dos limites de 25% a 75% dos valores observados de média e desvio padrão sejam aceitados para compor o novo *dataset* a ser utilizado como parte na metodologia de detecção de perdas não técnicas. Dessa forma, esses processos resultaram em 378 unidades consumidoras irrigantes do município de Uruguaiana.

3.2 AGRUPAMENTO DE DADOS

O agrupamento é uma técnica de mineração de dados não supervisionada que desassocia dados estruturados e não estruturados em grupos significativos denominados de *clusters*. Dessa forma, os clusters com semelhanças são desenvolvidos com base nos recursos ou propriedades dos dados (HAN; KAMBER; PEI, 2014; UMAR; PRASAD; FONKAM, 2019). Na literatura, a análise de agrupamento de dados é comumente chamada de *Clustering* (CASSIANO, 2014).

Diferentemente do conceito de classificação, o *Clustering* é um processo mais rudimentar em que nenhuma suposição é realizada em relação aos grupos de dados. Contrariamente dos métodos de classificação, o agrupamento de dados não conta com classes predefinidas e exemplos de treinamento de classes rotuladas. Dessa forma, o mesmo acaba por realizar uma forma de aprendizado não supervisionado. Qualquer método de *Clustering* é definido por um algoritmo específico que determina como será realizada a divisão dos casos nos clusters distintos, e todos os métodos propostos são baseados na ideia de distância ou similaridade entre as observações, definindo a relevância dos objetos a cada cluster, segundo a relação similar que cada elemento tem em relação a dados pertencentes aos outros grupos (CASSIANO, 2014).

O emprego de alguma técnica de *Clustering* pode ser observado em cinco passos, são eles (LACHI; ROCHA, 2005):

- Definição do modo de representatividade dos dados de entrada a serem agrupados;
- Definição de uma medida adequada de aproximação entre os dados;
- Definição da técnica de *Clustering* a ser utilizada para a construção dos clusters;
- Definição de uma abstração dos dados;
- Avaliação do resultado do *Clustering*.

O objetivo central da aplicação de *Clustering*, é agrupar pontos de maneira que as amostras em um cluster específico sejam parecidos entre si e menos semelhantes aos pontos de outros clusters. Assim, todo o processo de agrupamento é realizado com o intuito de maximizar a homogeneidade dentro de cluster e maximizar a heterogeneidade entre clusters. Além disso, agrupar dados similares possibilita retratar de maneira mais eficiente as características peculiares de cada um dos grupos identificados. Desse modo, é possível uma maior compreensão dos conjuntos de dados originais, além de possibilitar o desenvolvimento de esquemas de classificação para novos dados e encontrar correlações importantes entre os atributos de dados que não seriam facilmente visualizadas sem o emprego de tais técnicas. Além disso, o *Clustering* pode ser utilizado como uma etapa de pré-processamento para outros algoritmos, tais como caracterização e classificação, que trabalhariam nos clusters identificados (CASSIANO, 2014).

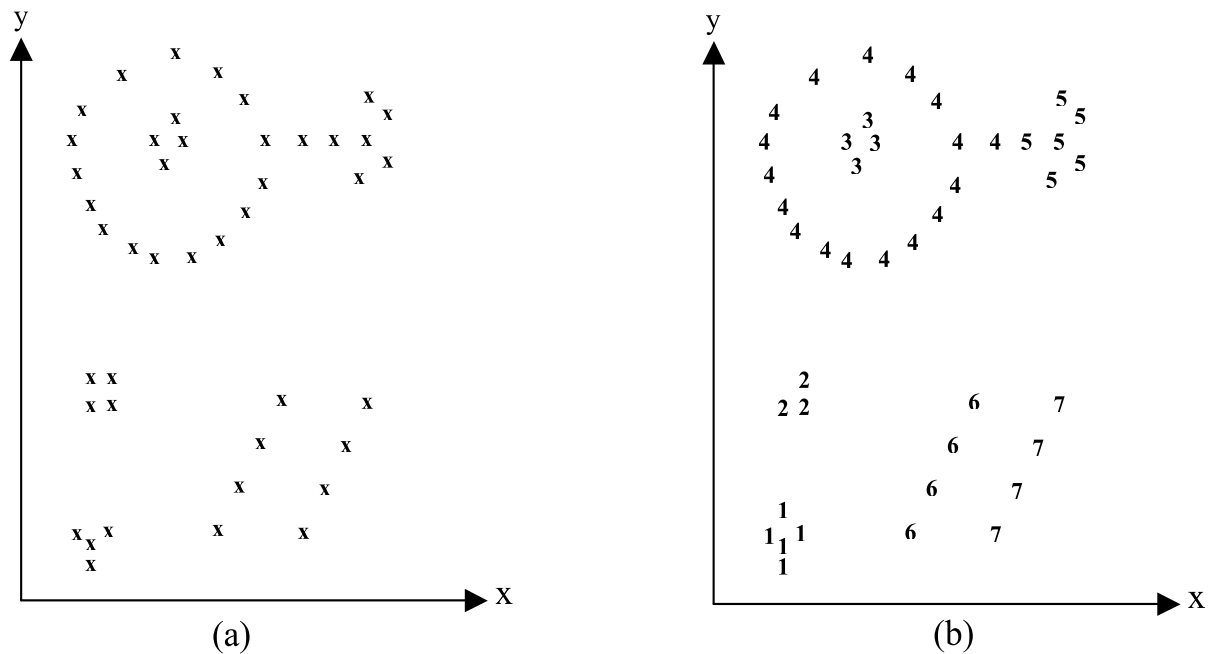
O agrupamento de dados pode ser aplicado também, quando o objetivo do estudo em questão é reduzir o número de objetos para um número de subgrupos característicos, levando as observações a serem consideradas como membros de um grupo e alinhadas de acordo com as características gerais que rotulam distintamente este grupo, ou também, quando o programador deseja formular hipóteses sobre a natureza dos dados ou explorar hipóteses pré-estabelecidas. Se uma determinada estrutura pode ser previamente definida para um certo grupo de objetos, o resultado da análise de clusters pode ser empregado para comparar e validar a estrutura inicial. Outra finalidade do *Clustering* é a identificação de relações entre as observações não identificadas em outras técnicas (CASSIANO, 2014).

Na Figura 3.2 é demonstrado um exemplo ilustrativo de um *Clustering*. Em (a), é apresentado um conjunto de dados de entrada, onde cada elemento é representado pelo símbolo “x”. Em (b), é demonstrado o resultado do agrupamento realizado sobre esse conjunto de dados de entrada, com cada elemento “x” sendo rotulado com um número identificador do conjunto a que pertence ao final do *Clustering*.

Como observado na Figura 3.2, os dados de entrada são apresentados na parte (a), enquanto os 7 clusters construídos são demonstrados na parte (b). Pode-se observar que na parte (b), os dados pertencentes a um mesmo cluster apresentem o mesmo rótulo.

Nas subseção a seguir, será contextualizado o método *k-Means*, utilizado para a etapa de agrupamento de dados na metodologia desenvolvida nesse trabalho, e a sua aplicação e resultados para o processo de composição de classes de perfis de consumo no conjunto de unidades consumidoras irrigantes utilizados para esse estudo.

Figura 3.2 – Exemplo de um processo de *Clustering*.



Fonte: Lachi e Rocha (2005).

3.2.1 Método *k-Means*

O algoritmo *k-Means* executa a análise de agrupamento de dados a partir do cálculo de centroides dos mesmos, ou seja, uma determinada amostra pertence a um grupo (cluster) se a distância euclidiana de seus preditores ao centroide deste mesmo grupo se assentar em um limite de tolerância pré-estabelecido. Dessa forma, o algoritmo agrupa dados de mesma categoria pelas menores distâncias destes aos clusters (MAROTTA, 2019).

Segundo Passos Júnior et al. (2016), o algoritmo requer o número de clusters como entrada e agrupa os dados, separando as amostras em grupos com variâncias semelhantes para minimizar algum critério. Além disso, essa técnica se adapta a um elevado número de amostras e vem sendo amplamente utilizada em diversas áreas de aplicações. Entretanto, uma das maiores adversidades do algoritmo *k-Means* é a sensibilidade a seleção de partida, podendo convergir a um mínimo local do valor da função de critério se a partição inicial não for devidamente selecionada (BELTRAME; FONSECA, 2010; JAIN; MURTY; FLYNN, 1999).

O algoritmo *k-Means* funciona dividindo um conjunto de N amostras em valor k de clusters disjuntos ($C_j, j = 1, 2, \dots, k$), sendo cada grupo descrito pela média μ_j de suas amostras.

As médias são os denominados “centroides” do cluster e não podem ser pontos do conjunto de dados. Assim, esse algoritmo visa escolher centroides que minimizem uma função quadrática de erro (PASSOS JÚNIOR et al., 2016), dada por:

$$\sum_{i=0}^N \min_{\mu_j \in C_j} \left(\|x_i - \mu_j\|^2 \right), \quad \forall j = 1, 2, \dots, k \quad (1)$$

Sendo x_i , uma amostra dos dados de treinamento.

Dessa forma, o algoritmo *k-Means* pode ser descrito pelos seguintes passos de acordo com Beltrame e Fonseca (2010) e Fontana e Naldi (2009) e ilustrado na Figura 3.3.

- **Passo 1:** Atribuem-se valores iniciais para os modelos, seguindo algum tipo de critério. Exemplo: Sorteio aleatório desses valores dentro dos limites de domínio de cada atributo;
- **Passo 2:** Atribui-se cada objeto ao grupo, cujo modelo apresente maior similaridade com o objeto;
- **Passo 3:** Recalcula-se o valor da centroide (modelo) de cada grupo, como sendo a média dos objetos atuais do grupo;
- **Passo 4:** Repetem-se os passos 2 e 3, até que os grupos se consolidem.

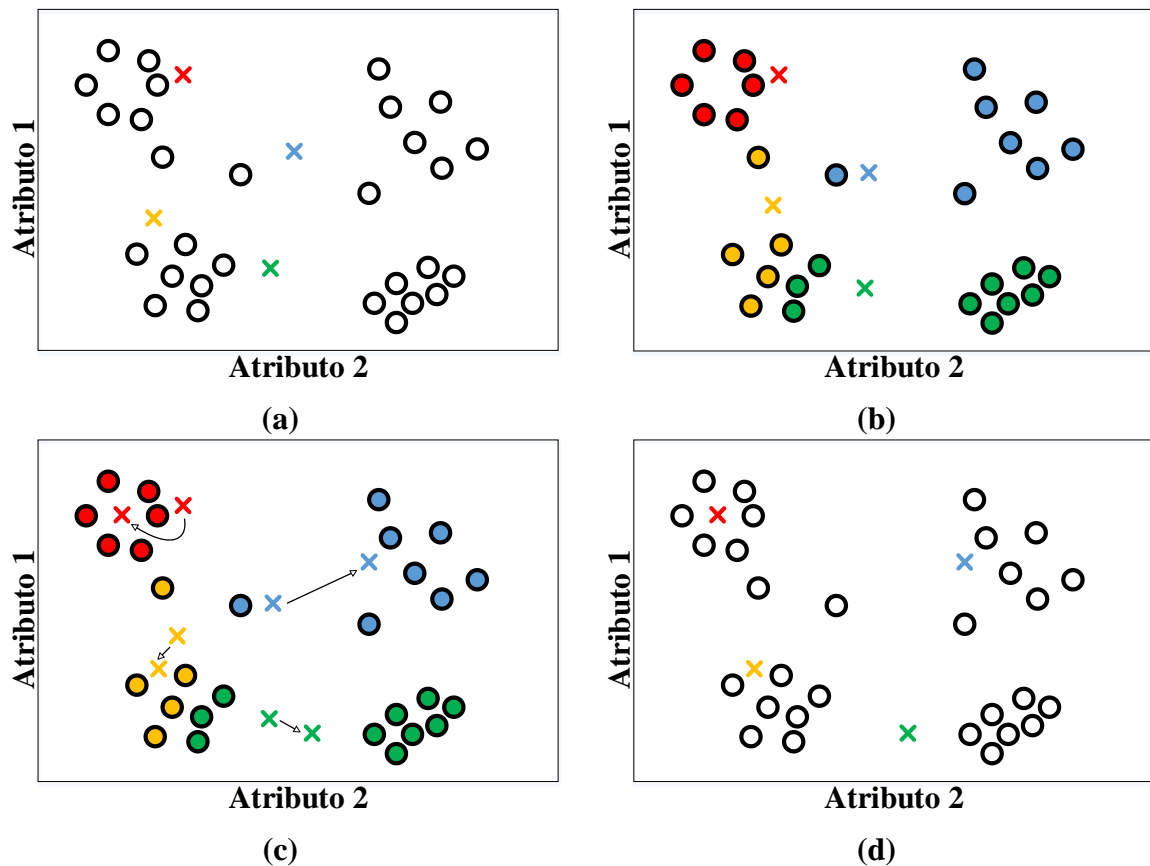
Na Figura 3.3 foi ilustrado um exemplo de atuação do algoritmo *k-Means*, onde os itens: (a) indica o sorteio dos valores iniciais dos centroides (Passo 1); (b) representa a atuação dos grupos em que o modelo apresente a maior similaridade (Passo 2); (c) retrata o novo cálculo dos valores do modelo representando a média dos objetos atuais pertencentes aos grupos (Passo 3) e (d) simboliza a preparação para repetir os Passos 2 e 3 (Passo 4).

3.2.2 Alocação de Unidades Consumidoras Irrigantes em Classes de Consumo de Energia Elétrica

Como mencionado anteriormente, o *k-Means* é um método de aprendizado de máquina não supervisionada utilizado para o agrupamento de dados, que no caso do respectivo estudo, para o processo de criação de classes. O mesmo, tem por objetivo agrupar dados de mesma categoria pelas menores distâncias destes aos grupos, comumente denominados de *clusters*.

Os dados de entrada para a criação dos clusters foram a média (m) e desvio padrão (σ) dos consumos de energia elétrica de cada unidade consumidora irrigante, definidas pela etapa de pré-processamento, apresentada na Seção 3.1. Esse processo resulta como saída, um novo

Figura 3.3 – Exemplo do algoritmo *k-Means*.



Fonte: Adaptado de Beltrame e Fonseca (2010).

atributo ao conjunto de dados fundamento no resultado da etapa de agrupamento de dados, definido como perfil de consumo. O valor de k (número de classes definidas) para o algoritmo *k-Means* foi definido para 4 (quatro classes de consumo), sendo uma escolha particular, de acordo com o número de amostras presentes no *dataset*, visando manter o equilíbrio no número de *clusters* para estimular a mudança de classe de perfil de consumo. Um baixo valor de “ k ”, implicaria em limites muito altos para média de consumo, e o método consideraria a mudança de classe de perfil de consumo, se ocorresse uma alteração muito significativa. Por outro lado, um número alto de *clusters*, tornaria o modelo muito sensível, onde uma pequena alteração na média do consumo, indicaria a mudança de classe para a determinada amostra avaliada.

Para a realização dessa etapa, utilizou-se uma ferramenta da biblioteca da linguagem *Python*, conhecida como *kMeansConstrained*, onde é possível estabelecer um número mínimo e máximo de amostras em cada cluster criado pelo o algoritmo. Desse modo, os valores mínimos foram definidos como o somatório total de amostras dividido pelo número de

clusters (4, valor definido para k) e os valores máximos representando o valor mínimo resultante acrescentado de 1. Além disso, aplicou-se essa ferramenta a fim de mitigar o problema de desbalanceamento entre as classes. Entretanto, a resolução do problema de desbalanceamento resultou em alguns problemas de limites de valores, onde uma unidade consumidora de determina classe, compreendia valores de m dentro dos limites de outra classe, podendo acarretar problemas durante a etapa de predição.

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, e realoca corretamente as amostras que estejam alocadas incorretamente.

A Figura 3.4 ilustra uma representatividade gráfica da etapa de agrupamento relacionando média e desvio padrão de consumo para cada classe de perfil de consumo gerada (Perfil_1, Perfil_2, Perfil_3 e Perfil_4). A distribuição das unidades consumidoras irrigantes por classe de consumo resultante e os limites de máximos e mínimos dos atributos m e σ são apresentados na Tabela 3.2.

Figura 3.4 – Representação gráfica do resultado de agrupamento de unidades consumidoras para diferentes classes de perfil de consumo.

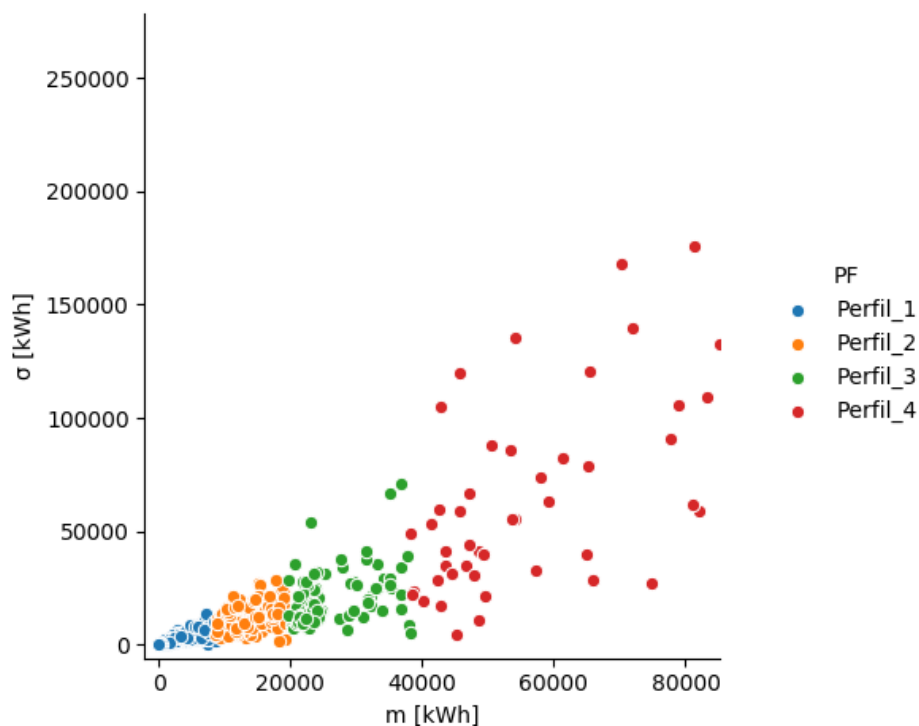


Tabela 3.2 – Número de unidades consumidoras por perfil de consumo e valores de máximo e mínimo.

Perfil de Consumo	Número de Unidades Consumidoras	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

Fonte: Elaborado pelo o autor.

Os dados apresentados na Tabela 3.2, demonstram os limites máximos e mínimos do valor da média de consumo de energia elétrica de cada classe de perfil de consumo. Nota-se, o que o números de amostras de consumidores por classe ficou desbalanceadas, podendo ocasionar problemas de predições futuras, pois o modelo pode tender 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).

3.3 ESTIMATIVA DO CONSUMO DE ENERGIA ELÉTRICA DE UNIDADES CONSUMIDORAS IRRIGANTES

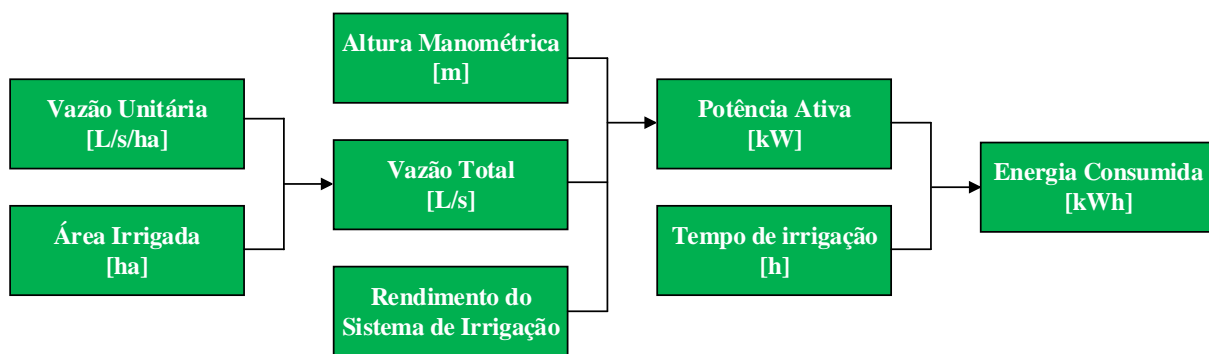
O estágio de estimativa de consumo de energia elétrica de unidades consumidoras irrigantes é composto por diversos atributos que estão relacionados ao uso de eletricidade nas lavouras de arroz irrigado que utilizem o método de irrigação por inundação. Desse modo, essa etapa tem início a partir da relação entre a vazão unitária aplicada para manutenção de irrigação da lavoura e a área de plantio a ser irrigada.

O primeiro elemento é definido através de estudos em literatura, enquanto que o segundo, por informações referentes a intenção de plantio e apresentado por meio de imagens de satélite. A partir dessa relação, é calculada a vazão total da lavoura, considerando toda a área de cultivo, que por sua vez, associado com os atributos de altura manométrica e

rendimento da estação de bombeamento (escolhidos com base em estudos realizados na região piloto), resulta na potência ativa total do sistema de irrigação. Por fim, a consumo de energia elétrica do consumidor pode ser estimado a partir da multiplicação entre a potência ativa do sistema irrigante e tempo de bombeamento da instalação.

A Figura 3.5 apresenta o diagrama sequencial de estimativa de consumo de energia elétrica em unidades consumidoras irrigantes de lavouras de arroz.

Figura 3.5 – Rotina de estimativa de consumo de energia elétrica para unidades consumidoras irrigantes.



Fonte: Elaborado pelo o autor.

Nas subseções a seguir, serão caracterizados os atributos associados a etapa de estimativa de consumo de energia elétrica em unidades consumidoras irrigantes.

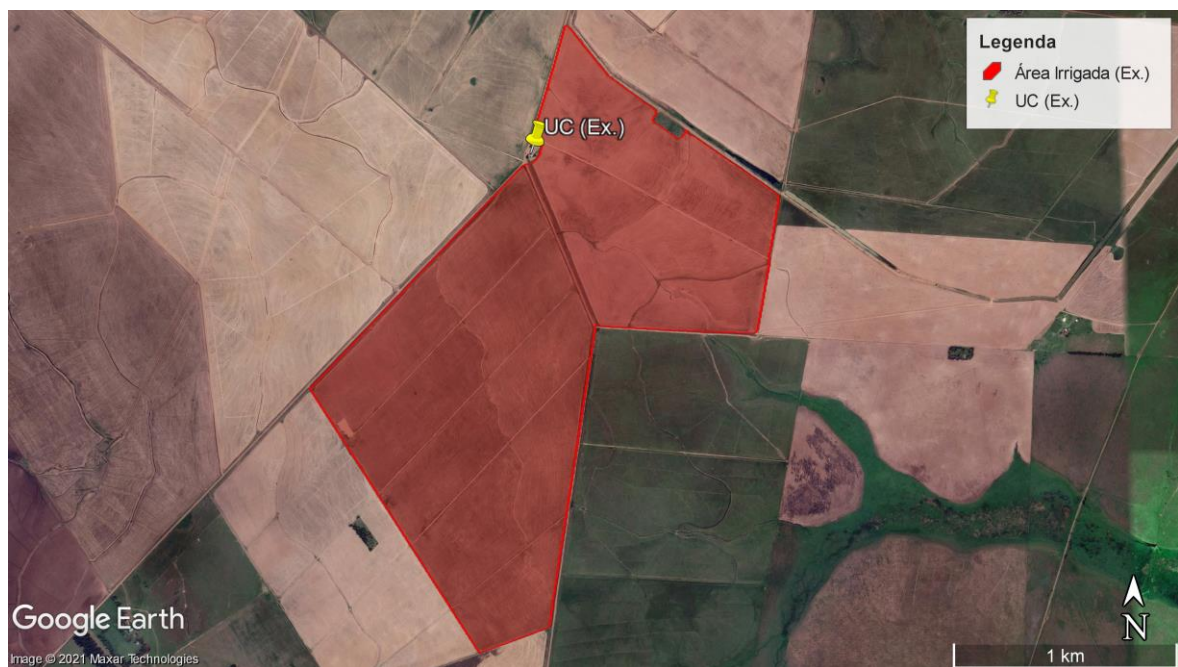
3.3.1 Área Irrigada

Como mencionado na Seção 2.2, a 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.

Esse atributo, é definido com base nas informações referentes a intenção de plantio durante a Safra 2020/21, para cada unidade consumidora irrigante associada a respectiva área. No entanto, esse parâmetro é informado em apenas alguns conjuntos de clientes irrigantes que compõem o *dataset* aplicado a metodologia desse estudo. Dessa forma, serão utilizados para os estudos de casos, apenas os consumidores em que a área de plantio é conhecida de maneira total ou parcial.

Para esse trabalho, serão ilustradas a partir de imagens de satélite, as áreas de plantio e suas respectivas unidades consumidoras irrigantes a serem estudadas, nesse caso, em áreas rurais do município de Uruguaiana na Região da Fronteira Oeste do estado do Rio Grande do Sul. A Figura 3.6 demonstra um exemplo de identificação de uma área de plantio de arroz irrigado através do processamento de uma imagem de satélite, extraída pelo Google Earth Pro®.

Figura 3.6 – Identificação de uma área de plantio de arroz irrigado através do processamento de imagens de satélite.



Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro® (2021).

O exemplo ilustrado na Figura 3.6 representa uma área de lavoura de arroz irrigado, (destacada com preenchimento em vermelho) com a localização da unidade consumidora associada (representada pelo marcador amarelo).

3.3.2 Lâmina de água e Vazão

O manejo da água em arroz irrigado por inundação é essencial para o desempenho da cultura. A água, além de participar de forma ativa no aspecto físico das plantas de arroz, interfere na disponibilidade de nutrientes, na população e espécies de plantas daninhas e na incidência de determinadas pragas e doenças. Além disso, um adequado manejo de água deve

compreender um período de submersão do solo que não afete a produtividade e a qualidade dos grãos (ALONÇO et al., 2004; SOSBAI, 2018).

A escolha do nível de vazão está diretamente relacionada à altura da lâmina de água da lavoura de arroz irrigada. A mesma, representa a técnica praticada pelo produtor em termos de cota média de água acima do nível do solo e a sua altura é uma escolha do produtor. Normalmente, é empregada a técnica de irrigação com lâmina controlada, onde um nível de água é definido por um sistema de automação ou pela observação empírica do aguador, pessoa que responsável pela manutenção da lâmina de água sobre a lavoura (UBERTI, 2017).

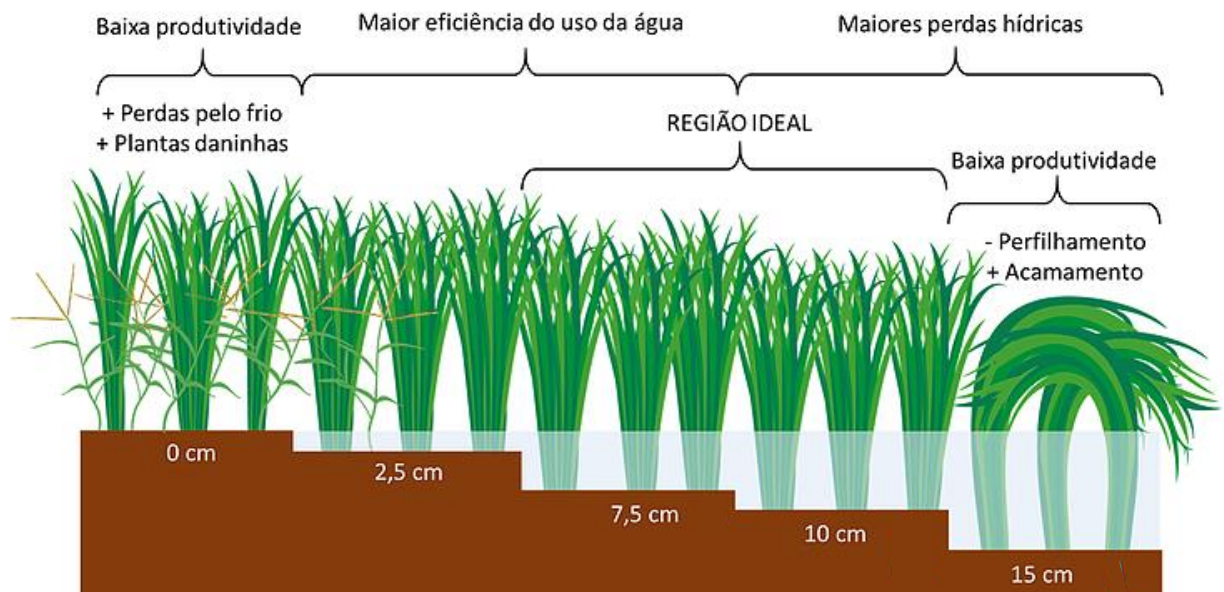
A elevação de nível é um aspecto importante no manejo de água para o cultura do arroz, uma vez que, interfere no volume hídrico utilizado e conseqüentemente, no custo da irrigação. Além disso, a presença da lâmina de água proporciona alguns benefícios como: maiores disponibilidades de nutrientes presentes no solo e o controle físico da plantas daninhas não aquáticas (EMBRAPA, 2020c; SOSBAI, 2018).

Lâminas com alturas de até 2,5 cm proporcionam excelentes índices de rendimento dos grãos, além da uma economia significativa no uso da água em relação as demais alturas utilizadas nessa técnica. Entretanto, essa a mesma requer uma sistematização do solo mais criteriosa e cuidados especiais no controle com plantas daninhas. Já as lâminas de água com altura superiores a 2,5 cm variando até 7,5 cm, apesar do consumo mais elevado de água, são menos exigentes em relação ao nivelamento superficial do solo. Lâminas superiores a 10 cm reduzem o número de perfilhos, e promovem o rápido crescimento das plantas de arroz. No entanto, lâminas com essa altura proporcionam o aumento nas perdas de água por infiltração lateral e percolação, além de elevar a evaporação durante o período noturno.(SOSBAI, 2018). Essas informações podem ser visualizadas na Figura 3.7, que apresenta um diagrama contendo um breve resumo de como as diferentes alturas de lâmina de água podem influenciar na produtividade dos grãos e na eficiência do uso de água nas lavouras de arroz irrigado.

A partir da aplicação da lâmina de água na área de plantio, é possível escolher a vazão unitária a ser aplicada na lavoura de arroz, que por sua vez depende também de outros fatores como: tempo de bombeamento diário, aspectos físicos do solo e topografia da região de cultivo (KÖPP, 2015).

No conceito da física, a vazão de um fluído é definida como um volume que atravessa a seção transversal de um duto por unidade de tempo. A vazão deriva da quantidade de água requerida pela lavoura de arroz e é considerada uma importante variável para o processo de dimensionamento dos sistema irrigantes. Estima-se que durante todo o período da safra, seja utilizado um volume de água em torno de 12.000 m³ por hectare na lavoura de arroz irrigado.

Figura 3.7 – Diagrama contendo diferentes alturas de lâmina de água na lavoura arroz irrigado.



Fonte: Pereira (2019).

No entanto, o mesmo pode variar entre 8.000 m³ e 14.000 m³ por hectare, dependendo das condições meteorológicas durante o período da safra correspondente e da área de plantio a ser irrigada.

Vazões unitárias de 3 L.s⁻¹ a 5 L.s⁻¹ por hectare, são recomendadas no início da etapa de irrigação da área de plantio, enquanto que, para a manutenção da irrigação no restante da safra, são empregadas vazões unitárias de 1,5 L.s⁻¹ a 2 L.s⁻¹ por hectare. O período médio de irrigação da lavoura varia entre 80 e 100 dias.

O armazenamento de água no caso da cultura do arroz irrigado é a saturação da camada do solo em 30 cm, aproximadamente. Assim, o terreno da lavoura deve ser totalmente saturado e ainda formar uma lâmina de água pela área de plantio. Em média, a quantidade de água para saturar o solo é de 45% do volume hídrico da camada de saturação, o que representa um valor aproximado de 135 milímetros. Acrescentando uma lâmina de altura média de 50 mm (5cm) a esse volume, são totalizados 185 milímetros de água em toda a lavoura. Após essa aplicação, a lâmina de água é mantida na área de cultivo, repondo a evapotranspiração diária. A utilização desses valores típicos, implicam em uma vazão unitária de 1,5 L/s/ha.

A Tabela 3.3 apresenta os valores típicos utilizados para esses atributos na metodologia desse trabalho.

Tabela 3.3 – Valores típicos para os atributos de lâmina de água e vazão unitária.

Atributos	Valores Típicos
Lâmina de água	5 cm
Vazão Unitária	1,5 L/s/ha

Fonte: Elaborado pelo o autor.

Os valores apresentados na Tabela 3.3 foram adotados com base em pesquisas bibliográficas realizadas nas literaturas desenvolvidas em Evaldt (2014), Köpp (2015), Uberti (2017) e Evaldt (2018).

3.3.3 Altura Manométrica Total

A altura manométrica total de uma estação de bombeamento envolve alguns aspectos na lavoura de arroz e do próprio sistema de irrigação, como o próprio desnível apresentado no terreno do local de plantio, denominados de altura geométrica, e as perdas nas tubulações desses sistemas, responsáveis por distribuir a água em todas as parcelas da plantação dessa cultura, conhecidas como perdas de carga.

A altura geométrica, também denominada de altura do levante, representa o desnível entre o ponto de captação da fonte hídrica (sucção) e o canal de distribuição de água na lavoura de arroz (recalque). Tipicamente, a altura geométrica das estações de bombeamento podem variar de 0 a 50 metros. Em alguns casos, dependendo da declividade do terreno e do local de instalação do sistema irrigante, o recalque da tubulação está situado abaixo do ponto de captação da água, situação conhecida como irrigação por gravidade (EVALDT, 2018). Quanto maior for o desnível entre a sucção e o recalque da tubulação, maior será a potência instalada do sistema de bombeamento, e conseqüentemente maior será o consumo de energia elétrica na lavoura. A Equação 2 descreve a altura geométrica entre o recalque e a sucção de uma tubulação.

$$h_g = h_r + h_s \quad (2)$$

Onde:

- h_g : Altura geométrica (m);
- h_r : Altura de recalque (m);

- h_s : Altura de sucção (m).

Segundo Köpp (2015), na região da Fronteira Oeste do Rio Grande do Sul, as alturas geométricas das estações bombeamentos variam entre 2,0 e 25,62 metros, o que está relacionado à topografia da região que apresenta relevo suavemente ondulado (3 a 8% de declividade).

Já as perdas de carga, são aquelas decorrentes por peças especiais presentes na tubulação, como curvas, derivações, válvulas, registros, juntas, entre outras. Essa grandeza, pode ser estimada a partir da Equação de Hazzen-Williams, que relaciona a perda de carga e pressão devido a passagem do fluxo de água em tubo. O uso desta equação é muito comum em dimensionamentos hidráulicos e recomendada para tubulações que apresentam peças com diâmetros maiores que 50 milímetros (MELLO; CARVALHO, 1998). A expressão matemática para a estimativa das perdas de cargas nas tubulações desenvolvida por Hazzen-Williams é observada a seguir:

$$h_f = 10,643 \cdot \left(\frac{Q_{md}}{C} \right)^{1,85} \cdot \frac{(L + L_{virtual})}{D^{4,87}} \quad (3)$$

$$L = \sum_{i=1}^{i=n} N \cdot D \quad (4)$$

Onde:

- h_f : Perda de carga na tubulação (m);
- Q_{md} : Vazão medida ($m^3 \cdot s^{-1}$);
- C : Coeficiente de rugosidade de Hazzen-Williams (Valores tabelados de acordo com o material da canalização – Adotado valor “90” para tubulações de aço soldado com mais de 10 anos);
- D : Diâmetro da tubulação (m);
- L : Comprimento da tubulação (m);
- $L_{virtual}$: Comprimento virtual equivalente às peças especiais presentes na instalação (m);
- n : Número de peças especiais presentes na estação de bombeamento;
- N : Número de diâmetros equivalente a peça especial (Valores tabelados de acordo com a peça).

As perdas de cargas são estimadas separadamente para as tubulações de sucção e recalque. Com isso, as perdas de cargas totais de uma estação de bombeamento é dada por:

$$h_{fT} = h_{fS} + h_{fR} \quad (5)$$

Onde:

- h_{fT} : Perda de carga total na estação de bombeamento (m);
- h_{fS} : Perda de carga na tubulação de sucção (m);
- h_{fR} : Perda de carga na tubulação de recalque (m).

A altura manométrica representa a energia que a bomba deverá transmitir a água para transportar uma determina vazão entre as tubulações de sucção e recalque. Dessa forma, a mesma deve vencer o desnível geométrico, as perdas de carga e a diferença de pressões nos reservatórios (SOARES, 2012).

A altura manométrica total de uma estação de bombeamento é o resultado do montante entre a altura geométrica e as perdas de cargas totais na tubulação, conforme pode ser observado pela Equação 6.

$$AMT = h_g + h_{fT} \quad (6)$$

Onde:

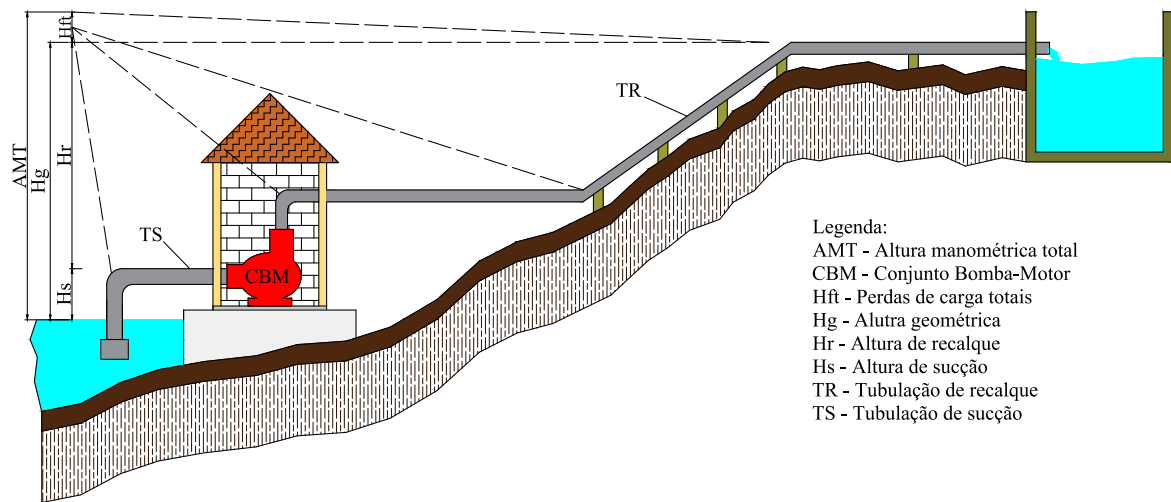
- AMT : Altura manométrica da total da instalação (m);
- h_g : Altura geométrica da instalação (m);
- h_{fT} : Perda de carga total na estação de bombeamento (m).

Desse modo, o diagrama apresentado na Figura 3.8 tem por objeto ilustrar a fim de observação, a relação entre as grandezas das alturas geométricas e perdas de carga nas tubulações de sucção e recalque com a altura manométrica total de uma estação de bombeamento. Nota-se, que a altura manométrica de sistema de bombeamento engloba todas as perdas de cargas e desníveis apresentados nas tubulações de sucção e recalque.

A Tabela 3.4 apresenta o valor típico utilizado para esse atributo na metodologia desse trabalho.

O valor apresentado na Tabela 3.4 foi adotado com base no estudo desenvolvido por Köpp et al. (2016), que caracteriza e avalia as estações de bombeamento de água em lavouras de arroz irrigada na Região da Fronteira Oeste do estado do Rio Grande do Sul, em grande parte no município de Uruguaiana. Segundo os autores, 65% das estações de bombeamento da região da Fronteira Oeste no Rio Grande do Sul possuem alturas manométricas que variam

Figura 3.8 – Diagrama ilustrando a altura manométrica em uma estação de bombeamento.



Fonte: Elaborado pelo o autor.

Tabela 3.4 – Valor típico para o atributo de altura manométrica total.

Atributo	Valor Típico
Altura Manométrica Total	10 m

Fonte: Elaborado pelo o autor.

entre 5 e 15 metros. Esses índices, podem estar relacionados à preferência por áreas com baixas alturas geométricas associadas a medidas de reposicionamento dos sistemas de bombeamento, redução no comprimento das tubulações, adequação dos diâmetros as vazões de projeto e redução de peças especiais. Dessa forma, será atribuído para essa variável, o valor intermediário desse estudo, 10 metros. O mesmo, engloba as grandezas de desvio geométrico e das perdas de energia nas tubulações.

3.3.4 Rendimento Global do Sistema de Irrigação

O rendimento global é o produto dos rendimentos da bomba, motor e transmissão entre ambos. Essa variável, determina a relação entre a energia consumida pelo conjunto bomba-motor e o trabalho realizado na unidade de tempo (KÖPP, 2015). A expressão que define o rendimento global de um sistema de bombeamento é descrita a seguir:

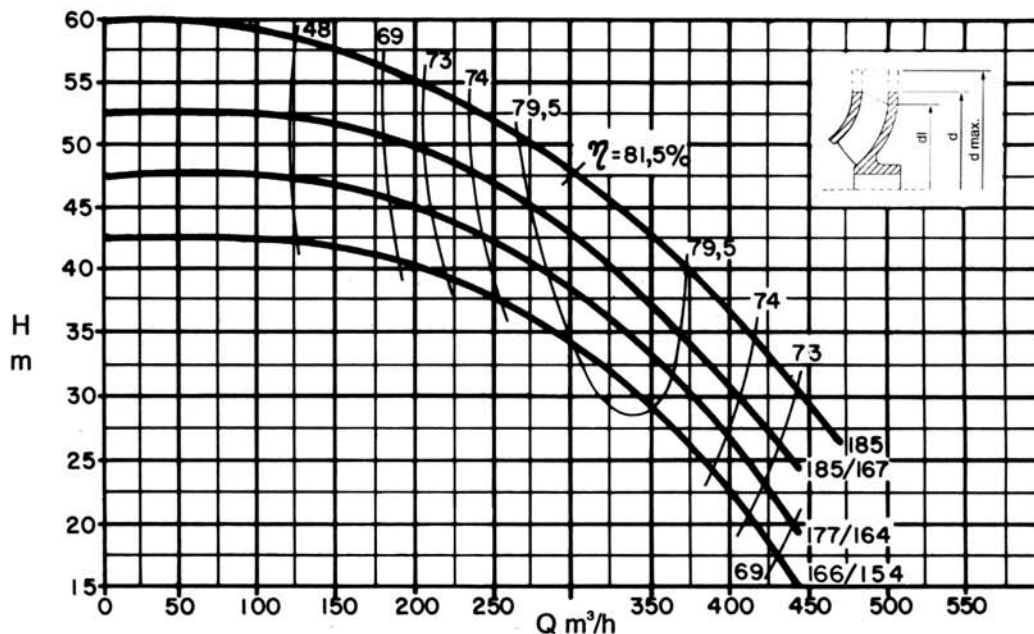
$$\eta_g = (\eta_b \cdot \eta_m \cdot \eta_t) \cdot 100 \quad (7)$$

Onde:

- η_g : Rendimento global do sistema de bombeamento (%);
- η_b : Rendimento da bomba;
- η_m : Rendimento do motor;
- η_t : Rendimento da transmissão.

O rendimento da bomba está relacionado com as condições de conservação do equipamento e à operação em seu ponto de máximo rendimento, a partir de uma curva característica que relaciona vazão e altura manométrica, denominado de diagrama $Q \times H$ (Figura 3.9) (UBERTI, 2017). Para esse trabalho, a nomenclatura referida a altura manométrica será “AMT”. As bombas centrífugas podem apresentar rendimentos em torno de 50% para vazões baixas e velocidades específicas inferiores a 20 rpm, e rendimentos de cerca de 80% para o caso de vazões mais elevadas e velocidades específicas superiores a 100 rpm (KÖPP, 2015).

Figura 3.9 – Curva característica de uma bomba centrífuga.



Fonte: KSB (2013).

Para o caso do rendimento do motor, são levados em consideração seu estado de conservação, modelo (padrão ou alto rendimento), enrolamentos e rendimentos operacionais, determinado pela operação próxima à potência nominal, considerando a condição de sobredimensionamento do motor (UBERTI, 2017). Comercialmente, o rendimento dos

motores elétricos novos, com potência acima de 15 kW (20 cv), é sempre superior a 90%. Para a transmissão por polias e correias, o rendimento fica entre 95% e 99%, dependendo da polia utilizado no processo (KÖPP, 2015).

O rendimento global depende da capacidade de desempenho e das características físicas de cada componente que compõem um sistema de bombeamento. Em Köpp (2015), os rendimentos globais das estações de bombeamento são classificados quanto aos seus respectivos percentuais em: Excelente (η_g superior a 70%); Muito bom (η_g entre 65 e 70%); Bom (η_g entre 50 e 60%); Ruim (η_g entre 40 e 50%) e Muito Ruim (η_g inferior a 40%).

A Tabela 3.5 apresenta o valor típico utilizado para esse atributo na metodologia desse trabalho.

Tabela 3.5 – Valor típico para o atributo de rendimento global do sistema de irrigação.

Atributo	Valor Típico
Rendimento global	65%

Fonte: Elaborado pelo o autor.

De acordo com a Tabela 3.5, foi adotado o percentual mínimo para estações de bombeamento classificadas como “Muito bom” por Köpp (2015), nesse caso, considerando 65% de rendimento global para todos os sistemas de irrigação associados as unidades consumidores rurais.

3.3.5 Tempo de funcionamento do Sistema de Irrigação

Como mencionado na Subseção 2.2.2, o regime de funcionamento de uma estação bombeamento é praticamente contínuo durante a época de irrigação. Os sistemas permanecem ativos por 21 horas diárias, sendo desligados apenas durante o período de horário de ponta, devido ao enquadramento tarifário e permanecendo acionados por 24 horas diárias nos fins de semana e feriados. Desse modo, o tempo usual de funcionamento de um sistema de irrigação em valores mensais considerando todas as condições descritas acima, é de aproximadamente 450 horas, totalizando 1800 horas por ciclo da safra da cultura.

A Tabela 3.6 apresenta o valor típico utilizado para esse atributo na metodologia desse trabalho.

Tabela 3.6 – Valor típico para o atributo de tempo de bombeamento do sistema de irrigação.

Atributo	Valor Típico
Tempo de bombeamento	450 horas

Fonte: Elaborado pelo o autor.

A adoção de valores mensais no tempo de funcionamento de um sistema de irrigação na metodologia desse trabalho, deve-se a estimativa do consumo de energia elétrica do mesmo ser realizada em um período mensal.

3.3.6 Estimativa do Consumo de Energia Elétrica do Sistema de Irrigação

A estimativa do consumo de energia elétrica em uma unidade consumidora irrigante, é iniciada a partir do cálculo de previsão da potência ativa instalada de um sistema de irrigação proposto por Köpp (2015), a partir de um índice de desempenho que avalia o rendimento global de uma estação bombeamento, sendo observada pela expressão matemática descrita a seguir:

$$\eta_g = \frac{Q \cdot AMT}{P \cdot 75} \cdot 100 \quad (8)$$

Deixando a Equação (8) em função de “P” e dividindo-a por 100, tem-se:

$$P = \frac{Q \cdot AMT}{\eta_g \cdot 75} \quad (9)$$

Como a resposta da Equação 9 é expressada em cavalo-vapor (cv) e deseja-se saber a saída em quilowatts (kW), a mesma será multiplicada pelo valor de 0,736 kW/cv, conforme pode ser observada pela Equação 10.

$$P = \frac{Q \cdot AMT}{\eta_g} \cdot 0,0098 \quad (10)$$

Sendo:

$$Q = q \cdot A \quad (11)$$

Substituindo (11) em (10), tem-se:

$$P = \frac{q \cdot A \cdot AMT}{\eta_g} \cdot 0,0098 \quad (12)$$

Onde:

- P : Potência ativa do sistema de irrigação (kW);
- Q : Vazão total ($\text{m}^3 \cdot \text{s}^{-1}$);
- q : Vazão unitária para manutenção de irrigação na lavoura ($\text{m}^3 \cdot \text{s}^{-1} / \text{ha}$);
- A : Área da lavoura a ser irrigada (ha);
- AMT : Altura manométrica total (m);
- η_g : Rendimento global do sistema de bombeamento (valor decimal).

Com isso, pode-se estimar a energia elétrica consumida pelo sistema de irrigação através da expressão descrita a seguir.

$$E = P \cdot t \quad (13)$$

Onde:

- E : Energia consumida pelo sistema de irrigação (kWh);
- P : Potência ativa do sistema de irrigação (kW);
- t : tempo de funcionamento mensal da estação de bombeamento (h).

A presente subseção teve por objetivo apresentar as equações que descrevem a estimativa de consumo mensal de energia elétrica de unidades consumidoras irrigantes, utilizando informações referentes a intenção de plantio e as características dos sistemas de irrigação, aplicadas como parte no desenvolvimento da metodologia proposta para a estudo. Os resultados obtidos por esse estágio são fundamentais no processo de classificação do algoritmo de identificação de possíveis ocorrências de perdas não técnicas em consumidores rurais, apresentadas nas subseções a seguir.

3.4 CLASSIFICAÇÃO DE DADOS

A classificação de dados pode ser compreendida como um processo que proporciona a busca pelo conhecimento através da categorização das respectivas características de um conjunto de dados. Em geral, esse processo corresponde a atribuição de rótulos a determinados objetos não qualificados, a partir da observação de suas similaridades com os outras amostras já classificadas (FERREIRA, 2018).

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. Neste caso, o primeiro passo para a classificação é mapear e modelar o conjunto de dados em que pretende-se aplicar o algoritmo. Desse modo, o banco de dados de entrada é composto por uma coleção de exemplos, onde cada elemento é constituído por um lista de atributos independentes, bem como um atributo classe, sendo esse, o alvo escolhido no processo de classificação (FERREIRA, 2018).

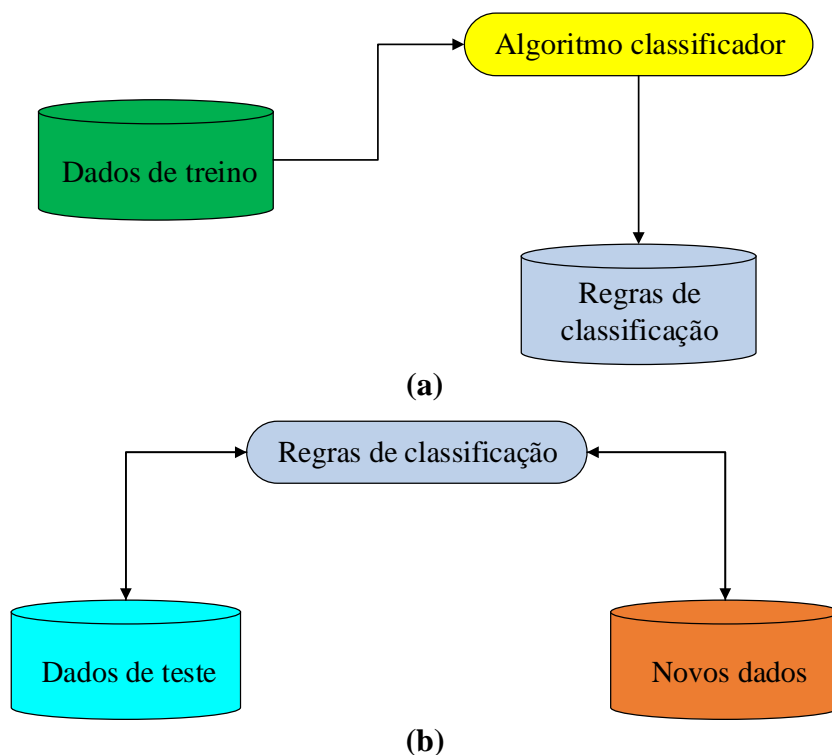
Segundo Han, Kamber e Pei (2014), a classificação de dados é um processo de duas etapas, consistindo em uma etapa de aprendizado (onde um modelo de classificação é desenvolvido) e uma etapa de classificação (onde o modelo é utilizado para prever rótulos de classe para dados fornecidos). O processo de classificação é ilustrado na Figura 3.10.

Na Figura 3.10 foi apresentado o processo de classificação nas suas respectivas etapas, divididas em (a), etapa de aprendizagem, onde os dados de treinamento são analisados por um algoritmo de classificação e (b) etapa de classificação, onde os dados de testes são utilizados para estimar a precisão das regras de classificação. Se a precisão for considerada adequada, as regras podem ser aplicadas à classificação de novas amostras de treinamento de dados (HAN; KAMBER; PEI, 2014).

Na primeira etapa, um modelo classificador é construído descrevendo um conjunto pré-determinado de classes ou conceitos de dados. A partir disso, é desenvolvido um conjunto de treinamento que será utilizado como entrada para o algoritmo de aprendizagem. O principal objetivo do processo de treinamento é alcançar uma função que mapeia as características preditoras do problema. Dessa forma, pretende-se encontrar um modelo que relacione os atributos de um dado exemplo com seu atributo classe (FERREIRA, 2018).

Um conjunto de treinamento é composto por amostras (chamadas de amostras de treinamento) de bancos de dados e seus rótulos de classes associados. Uma amostra (x) é

Figura 3.10 – Processo de Classificação de dados.



Fonte: Adaptado de Han, Kamber e Pei (2014).

representada por um vetor de atributo n -dimensional ($X = x_1, x_2, \dots, x_n$), representando n medições realizadas nas amostras a partir de n atributos do banco de dados (A_1, A_2, \dots, A_n). As amostras são consideradas pertencentes a uma classe predefinida conforme determinado por outro atributo do banco de dados, denominado de atributo de rótulo de classe (HAN; KAMBER; PEI, 2014).

Por fim, é realizada a fase de teste do classificador. Esse estágio, busca avaliar o desempenho do modelo desenvolvido a partir da classificação de novos exemplos ainda não rotuladas. Para isso, é utilizado um conjunto de teste para realizar tal atividade. Idealmente, esse conjunto não deve apresentar exemplos em comum com o conjunto de treinamento (FERREIRA, 2018).

Os algoritmos classificadores são divididos em probabilísticos e lineares. Os primeiros classificadores utilizam modelos de mistura de elementos para classificar os dados. Os principais algoritmos classificadores probabilísticos são: NB, *Bayesian Network* (BN) e *Maximum Entropy Classifier* (MEC). Já o segundos classificadores, tem como objetivo agrupar os itens em grupos que apresentem os mesmos valores característicos. Os principais algoritmos classificadores lineares são: SVM, MLP, *Logistic Regression*, *Rule based*

classifiers, *Decision Tree*, *Random Forest* e Redes Neurais (SARAVANAN; SUJATHA, 2018).

Na subseção a seguir será contextualizado o método de *Random Forest*, utilizado para a etapa de classificação de dados na metodologia desenvolvida nesse trabalho.

3.4.1 Método de Random Forest

O método de *Random Forest* (em português, Florestas Aleatórias) é um algoritmo de classificação e regressão de dados desenvolvido por Breiman (2001), induzido aleatoriamente através de um conjunto de árvores possíveis, utilizando um determinado número de atributos em cada nó. O termo “aleatoriamente” indica que cada árvore apresenta a mesma probabilidade de ser amostrada (OSHIRO, 2013). Esse algoritmo é um desenvolvimento direto e complemento do método de *Decision Tree* (em português, Árvores de Decisão), tornando assim possível reduzir a variabilidade em cada nó e obter estimativas a partir de um modelo linear local (SCHIKOWSKI, 2016).

De acordo com Breiman (2001) e Oshiro (2013), o algoritmo de *Random Forest* é um classificador constituído por um conjunto de árvores de decisão, composto por amostras aleatórias independentes e identicamente distribuídas, e cada árvore vota na classe mais popular para uma determinada entrada.

No ponto de vista das metodologias computacionais, o algoritmo de *Random Forest* é utilizado devido alguns aspectos como (CUTLER; CUTLER; STEVENS, 2011):

- Respostas eficientes para problemas de classificação e regressão;
- Rápido treinamento e prevenção;
- Dependência de apenas um ou dois parâmetros de ajuste;
- Estimativa conjunta de erro;
- Aplicação direta a problemas de alta dimensão;
- Fácil implementação em conjunto com outro método;

Esse método apresenta aprendizado não supervisionado, ou seja, dado as instâncias de treinamento e os rótulos que indicam a classe de cada amostra, o algoritmo aprende a classifica-las sem que seja necessária a intervenção de algum usuário no processo. O *Random Forest* vem sendo aplicado em diversas áreas do aprendizado de máquina, com destaque nas áreas de análises em sistemas de energia elétrica, detecção de ataques cibernéticos,

classificação de texto e reconhecimento facial (GUEDES; GUIMARÃES, 2014; ZHAO; MA, 2017).

Matematicamente, segundo os autores Cutler, Cutler e Stevens (2011), o algoritmo tem por objetivo encontrar uma função de previsão ($f(X)$) para prever uma determinada variável (Y). A mesma é determinada por uma função de perda ($L(Y, f(X))$) e definida para minimizar o valor esperada da perda.

$$E_{XY} (L(Y, f(X))) \quad (14)$$

Onde os elementos subscritos denotam expectativa com relação à distribuição conjunta de X e Y .

Instintivamente, $L(Y, f(X))$ representa a proximidade de $f(X)$ à Y , eliminando os valores de $f(X)$ que estão distantes de Y . As escolhas típicas de L são as perdas de erro quadrático $L(Y, f(X)) = (Y - f(x))^2$ para regressão e a perdas em zero-um para classificação:

$$L(Y, f(X)) = I(Y \neq f(X)) = \begin{cases} 0, & \text{se } Y = f(x) \\ 1, & \text{caso contrário} \end{cases} \quad (15)$$

Minimizando (14) para a perda de erro quadrático, tem-se que:

$$f(x) = E(Y|X = x) \quad (16)$$

A Equação (16) é conhecida como função de regressão de um conjunto de dados. Na situação de classificação, substituindo o conjunto de valores possíveis de Y por ζ e minimizando a Equação (14) para perdas em zero-um, obtém-se:

$$f(x) = \arg \max_{y \in \zeta} P(Y = y|X = x) \quad (17)$$

A Equação (17) também é conhecida como Regra de *Bayes*.

Os conjuntos constroem f em função de uma coleção dos chamados “aprendizes básicos” (descritos por $h_1(x), \dots, h_J(x)$), onde estes são combinados para gerar o “preditor de conjunto” $f(x)$. Na análise de regressão, os “aprendizes básicos” são calculados da seguinte maneira.

$$f(x) = \frac{1}{J} \sum_{j=1}^J h_j(x) \tag{18}$$

Já para a classificação, $f(x)$ é a classe mais frequentemente prevista, denominada de votação.

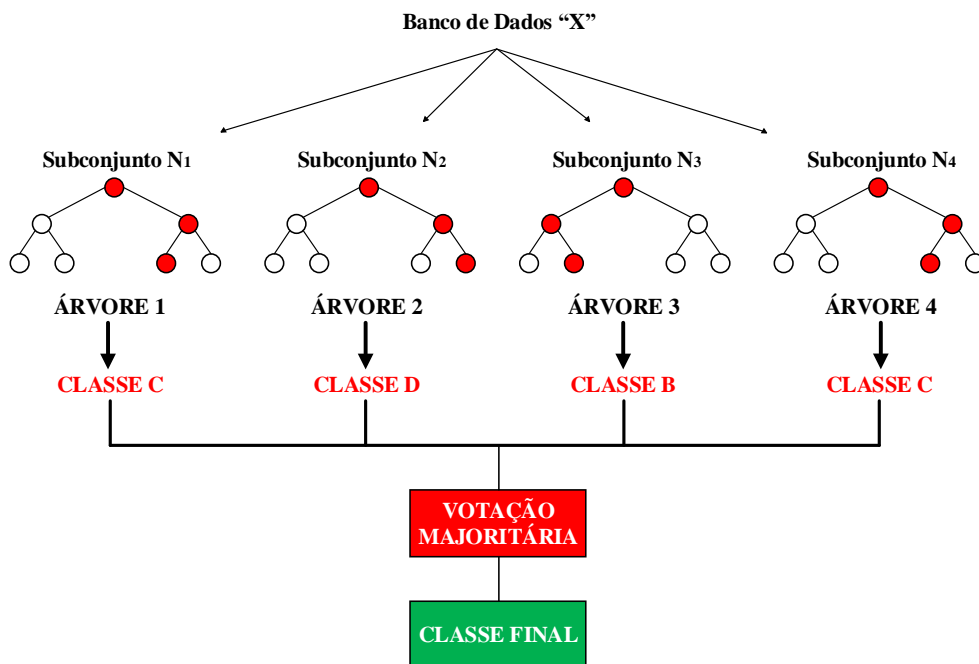
$$f(x) = \arg \max_{y \in \zeta} \sum_{j=1}^J I(y = h_j(x)) \tag{19}$$

Em *Random Forest*, o j -ésimo aprendiz de base é uma árvore de decisão denotado por $h_j(X, \theta_j)$, onde θ_j representa uma coleção de variáveis independentes para $j = 1, \dots, J$.

Em suma, o princípio de funcionamento do *Random Forest* consiste em criar diversas árvores de decisão, onde o atributo de decisão é selecionado de maneira aleatória. Em cada iteração, o método de classificação ignora a parte aleatória selecionada no conjunto de dados de treino, que é utilizada somente para criar uma árvore. Esse processo é repetido diversas vezes, desenvolvendo assim um conjunto de árvores. Para classificar uma instância, ela é passada a todas as árvores do conjunto, que a analisam separadamente. O resultado final do modelo é obtido pela combinação de resultados de cada árvore (SCHIAFFINO, 2020).

A Figura 3.11 ilustra um exemplo do modelo de *Random Forest*.

Figura 3.11 – Exemplo de arquitetura do método de *Random Forest*.



Na Figura 3.11 foi apresentada a estrutura base de um modelo de classificação pelo método de *Random Forest*. Dessa forma, conforme descrito no parágrafo anterior, o funcionamento do algoritmo inicia-se pela seleção de amostras aleatórias de um determinado conjunto de dados. Em seguida, são construídas uma árvore de decisão para cada uma das amostras selecionadas, obtendo-se assim um resultado previsão para as mesmas. Após isso, uma votação majoritária é realizada para todos os resultados previstos. Por fim, é selecionado como resultado final do modelo, a classe mais votada pela etapa anterior.

3.4.2 Modelo Preditivo para Detecção de Perdas Não Técnicas aplicando o Algoritmo de *Random Forest*

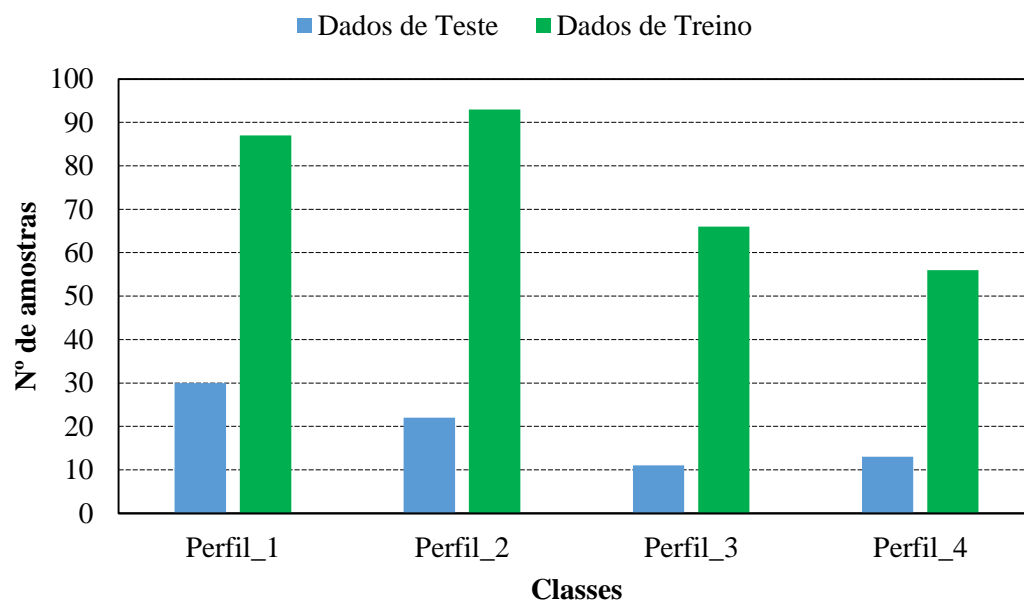
O modelo preditivo desenvolvido na metodologia desse trabalho tem por finalidade, executar o processo de identificação de perdas não técnicas em unidades consumidoras irrigantes alocadas em áreas rurais com cultivo de arroz irrigado, a partir das informações obtidas pelas etapas anteriores (Agrupamento e Estimativa de Consumo). O *Random Forest* é o algoritmo empregado para a predição dos clientes em seus respectivos grupos de Perfil de Consumo associados na etapa de agrupamento e alocação de classes.

A construção do modelo de predição 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). Essas porcentagens foram adotadas visando aumentar o número de unidades consumidoras classificadas em cada um dos treinos. É importante salientar que, esses dados foram divididos de acordo com as classes de perfis de consumo estabelecidas pela etapa de agrupamento, conforme pode ser observado pela Figura 3.12.

Segundo a Figura 3.12, a classe “Perfil_2” apresentou o maior de número de elementos para os dados de treino (93 amostras), seguida pelos grupos “Perfil_1” (87

Figura 3.12 – Distribuição das amostras de treino e teste para cada classe de perfil de consumo.



Fonte: Elaborado pelo o autor.

amostras), “Perfil_3” (66 amostras) e “Perfil_4” (56 amostras). Já para os dados de teste, a classe “Perfil_1” foi a majoritária (30 amostras), seguida pelos grupos “Perfil_2” (22 amostras), “Perfil_4” (13 amostras) e “Perfil_3” (11 amostras). A distribuição dos dados de treino e teste para cada uma classes de perfil de consumo, foi uma definição aleatória do algoritmo de *Random Forest*.

Após o processo de separação dos dados de treino e teste como parte da construção do modelo de predição, é iniciada a implementação do mesmo utilizando como variáveis de entrada, dados mensais de consumo de um conjunto de unidades consumidoras rurais do município de Uruguaiana/RS durante o ano de 2020, com destaque para o meses que compreendem a safra do arroz irrigado, e os grupos de perfil de consumo criados pela etapa de clusterização. Dessa forma, a construção do modelo preditivo é finalizada e a acurácia dos dados das classes de perfil de consumo é avaliada através de uma matriz de confusão. O resultado da métrica avaliativa é apresentado na Subseção 3.5.1.

Realizada as etapas de divisão das amostras de treino e teste e da verificação da acurácia dos dados que constituem o modelo preditivo, parte-se para a saída do mesmo, denominado nessa metodologia de Classificação de dados futuros com o Modelo *Random Forest*. Para a identificação de possíveis ocorrências de perdas não técnicas detectas pelo modelo de predição proposto, serão avaliados os indicadores citados abaixo utilizando

informações obtidas pelas etapas de agrupamento e estimativa de consumo de energia elétrica dos clientes irrigantes.

- **Mudança de Perfil de Consumo:** Esse indicador verifica a possível ocorrência de perdas não técnicas, se houver alteração no grupo de perfil de consumo da unidade consumidora irrigante, realizando a comparação entre a classe alocada durante a etapa de Agrupamento (dados de consumo de 2019) e a classe predita pelo modelo preditivo (dados de consumo de 2020). Desse modo, no caso de mudança de grupo de perfil de consumo para uma classe inferior no determinado cliente analisado, o modelo designara que este consumidor é suspeito de práticas de perdas comerciais;

- **Valor Estimado do Consumo de Energia Elétrica:** Esse indicador compara o consumo de energia elétrica mensal registrado da unidade consumidora irrigante no ano de 2020 com o valor estimado durante o mesmo intervalo, estabelecendo como limite aceitável, a variação de 40%. Desse modo, os valores de consumo que não estejam inseridos dessa faixa percentual são indicados como possíveis ocorrências de perdas não técnicas. 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 durante o período que corresponde a safra do arroz irrigado na Região da Fronteira Oeste do Rio Grande do Sul.

3.5 MÉTRICAS DE AVALIAÇÃO

Na aprendizagem de máquina, é necessário utilizar algumas métricas apropriadas para avaliar os resultados obtidos com um algoritmo classificador. Dessa forma, as respostas apresentadas por essas medidas de desempenho refletem a qualidade do modelo de classificação de dados, ou seja, se o mesmo está atendendo os requisitos necessários para a solução do problema proposto (BECKMANN; EBECKEN; PIRES DE LIMA, 2015; CHEN et al., 2020).

Segundo Chen et al. (2020), um modelo de classificação binária tem como objetivo de decidir em qual classe uma nova amostra de dados pertence dentre duas classes possíveis. Geralmente, duas classes denominadas de positiva (P) e negativa (N) indicam a ocorrência ou não de um determinado evento. Desse modo, a avaliação de um modelo de classificação é realizada a partir da comparação entre as classes previstas pelo modelo e as classes verdadeiras de cada exemplo. Assim, todas as métricas avaliativas tem o intuito de observar o quão distante o modelo está da classificação perfeita.

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; EBECKEN; PIRES DE LIMA, 2015).

A matriz de confusão tem a capacidade de representar duas ou múltiplas classes problemas, sendo empregada para calcular a precisão baseada na classificação correta e incorreta de um determinado conjunto de amostras. No entanto, sua aplicação em pesquisas relacionadas ao conjunto de dados desequilibrados é mais concentrada nas duas classes problemas, também denominadas como problemas binários ou binomiais, ou seja, quando um grupo menos frequente é nomeado como positivo e as demais classes são mescladas e declaradas como negativas (BECKMANN; EBECKEN; PIRES DE LIMA, 2015). Um exemplo de matriz de confusão é ilustrado na Tabela 3.7.

Tabela 3.7 – Exemplo de uma matriz de confusão.

	Previsão Positiva	Previsão Negativa
Classe Positiva	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Classe Negativa	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Adaptado de Beckmann, Ebecken, Pires de Lima (2015).

A matriz apresentada na Tabela 3.7, indica quantos exemplos existem em cada grupo: Falso Negativo (FN), Falso Positivo (FP), Verdadeiro Negativo (VN) e Verdadeiro Positivo (VP). Dessa forma, a matriz de confusão permite visualizar facilmente quantos exemplos foram classificados corretamente e erroneamente em classe, possibilitando compreender se o modelo está favorecendo uma classe em detrimento da outra (CHEN et al., 2020).

Uma série de métricas de desempenho para problemas de classificação calculadas a partir da matriz de confusão podem ser encontradas na literatura. Algumas das métricas mais utilizadas são (BECKMANN; EBECKEN; PIRES DE LIMA, 2015; MESSINIS; HATZIARGYRIOU, 2018b): Acurácia, Precisão, Revocação e F-1 *Score*.

A Acurácia 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% (CHEN et al., 2020; SAMMUT; WEBB, 2010). A Equação (20) descreve a métrica da Acurácia.

$$\text{Acurácia} = \frac{VP + VN}{VP + FN + FP + VN} \quad (20)$$

A Precisão é definida pela razão entre a quantidade de amostras classificadas corretamente como positivas e o total de amostras classificadas como positivas. Em outras palavras, essa métrica de avaliação indica a precisão/exatidão que o modelo de classificação está fora dos previstos positivos (BECKMANN; EBECKEN; PIRES DE LIMA, 2015; CHEN et al., 2020; SAMMUT; WEBB, 2010). A Equação (21) descreve a métrica da Precisão.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (21)$$

A Revocação, também denominada como sensibilidade ou taxa de verdadeiro positivo, é definida pela razão entre a quantidade de amostras classificadas corretamente como positivas e a quantidade de amostras que realmente são positivas (CHEN et al., 2020; SAMMUT; WEBB, 2010). A Equação (22) descreve a métrica da Revocação.

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (22)$$

Por fim, O F-1 *Score* leva em consideração tanto a precisão quanto a revocação. A mesma é definida pela média harmônica entre as duas (CHEN et al., 2020; SAMMUT; WEBB, 2010), conforme pode ser observado na Equação (23).

$$F-1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (23)$$

O F-1 *Score* é um pontuador necessário quando deseja-se buscar um equilíbrio entre a precisão e a revocação. Para um conjunto de dados desbalanceados, as métricas de precisão e revocação são as que possibilitam melhor avaliação de desempenho do modelo de previsão, sumarizados pela pontuação do F-1 *Score* (BECKMANN; EBECKEN; PIRES DE LIMA, 2015).

3.5.1 Métrica de Avaliação do Modelo Preditivo de Identificação de Perdas Não Técnicas em Unidades Consumidoras Irrigantes

Para avaliar se o modelo preditivo de detecção de perdas não técnicas proposta na metodologia desse trabalho apresenta pontuadores aceitáveis a fim de predizer dados futuros (registros mensais de consumo de energia elétrica de 2020), com destaque para a variável de acurácia, foram aplicadas como métricas de avaliação, a matriz de confusão do resultado preditivo para as amostras de teste, que correspondem a 20% do total da base de dados, definidas nas seções anteriores. Dessa forma, o percentual restante de elementos (80%) foram utilizados como dados de treino para composição do modelo, conforme mencionado na Subseção 3.4.2. A matriz de confusão resultante do modelo de predição, bem como os pontuadores para cada grupo de consumo são apresentados na Tabela 3.8.

Tabela 3.8 – Matriz de Confusão e Pontuadores do Modelo *Random Forest*.

Grupo/Dados de Teste	Classificadas pelo Modelo				Precisão	Revocação	F-1 Score	
	<i>Random Forest</i>							
	Perfil_1	Perfil_2	Perfil_3	Perfil_4				
Perfil_1	30	29	1	0	0,91	0,97	0,94	
Perfil_2	22	2	14	5	0,74	0,64	0,68	
Perfil_3	11	0	1	6	0,46	0,55	0,50	
Perfil_4	13	1	3	2	7	0,58	0,54	0,56

Fonte: Elaborado pelo o autor.

A matriz de confusão representada pela Tabela 3.8 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 os 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%, segundo as equações apresentadas na Seção 3.5 e simulações realizadas no IDE Spyder®.

4. ESTUDOS DE CASOS

Os estudos de casos serão realizados em 7 unidades consumidoras irrigantes localizadas no município de Uruguaiana/RS na Região da Fronteira Oeste do Rio Grande do Sul, selecionadas no conjunto de dados (*dataset*) definidos pela Etapa de Pré-Processamento de Dados, descrito na Seção 3.1. 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.

Alguns consumidores rurais podem apresentar imprecisões nas estimativas de consumo de energia elétrica das suas instalações. Em diversas situações, a divergência em larga escala, pode estar relacionada a dois motivos: o primeiro, pelos próprios dados do sistema de irrigação e da área de plantio do consumidor, que em muitos casos, são disponibilizados de forma imprecisa ou incorreta, e o segundo, por fatores meteorológicos que venham a influenciar no consumo de energia elétrica da instalação rural, como por exemplo, elevados índices de precipitação pluviométrica durante a época da safra. É importante destacar que, o impacto das precipitações pluviométricas no consumo mensal de energia elétrica dos sistemas de irrigação, não será considerada para as análises desse estudo.

Como mencionado na Subseção 3.4.2, a possibilidade de ocorrências de perdas não técnicas nas unidades consumidoras analisadas é avaliada pelo modelo preditivo, a partir da classificação de perfil de consumo da amostra que corresponde o cliente irrigante e a comparação entre os valores reais e estimados para os consumidores rurais durante os meses que correspondem aos maiores consumo de energia elétrica durante a época da safra. A Tabela 4.1 apresenta os indicadores e as situações propostas no Modelo Preditivo de *Random Forest* proposto na metodologia para avaliar os consumidores irrigantes que compõem a base de estudos de caso.

É importante ressaltar que, as classes de perfil de consumo atribuídas ao ano de 2019, denominadas para as análises e resultados como “Classe Original”, são estabelecidas pela Etapa de Agrupamento (Clusterização), definido grupos com base na média e desvio padrão do consumo mensal no referente ano para cada unidades consumidoras que o compõem a base de dados. Já para o ano de 2020, os grupos de perfis de consumo pertencente a cada amostra do *dataset*, identificados como “Classe Prevista”, são definidos pelo processo de classificação do modelo preditivo de *Random Forest*.

Tabela 4.1 – Indicadores propostos pelo Modelo Preditivo para identificação de perdas comerciais.

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

Fonte: Elaborado pelo o autor.

Para as análises de estimativas de consumo nas unidades consumidoras rurais, são consideradas que lavouras de arroz apresentem um padrão similar de consumo de energia elétrica, considerando os aspectos que envolvem os sistemas de irrigação, variando apenas a área de plantio a cada cliente associado a essa instalação rural. No entanto, por não haver uma expressão matemática exata que relacione essas variáveis, essas análises podem apresentar alguns resultados imprecisos, pois a energia faturada para esse tipo de consumidor pode ser afetada por uma série de fatores, tais como as variações na cultura, técnicas de manejo de irrigação, contratos comerciais e qualidade de fornecimento de energia elétrica. Desse modo, como proposta para as análises e resultados, considerou-se uma faixa percentual de 40% (Consumo Real menor que 60% do valor do Consumo Previsto) para o indicativo de possíveis ocorrências de perdas não técnicas nas unidades consumidoras irrigantes analisadas.

O principal objetivo do presente capítulo, é apresentar os diferentes cenários e respostas possíveis dos indicadores propostos para o modelo preditivo nas unidades consumidoras irrigantes selecionadas. Nas próximas seções, serão descritos os dados característicos das unidades consumidoras irrigantes utilizadas para a modelagem dos estudos de casos da metodologia desenvolvida nesse trabalho e os resultados das simulações do modelo preditivo para identificação de perdas não técnicas aplicado a estes consumidores.

4.1 DADOS CARACTERÍSTICOS DAS UNIDADES CONSUMIDORAS IRRIGANTES

Como referido no início do presente capítulo, serão empregadas informações referentes a sete consumidores irrigantes para testes de avaliação no modelo de identificação

de possíveis ocorrências de perdas não técnicas. As mesmas, integram um conjunto de duas safras do arroz irrigado (2019/20 e 2020/21) e apresentam lavouras com características semelhantes em relação a técnicas de manejo de irrigação e localizadas no mesmo município, nesse caso, Uruguaiana, Região da Fronteira Oeste do estado do Rio Grande do Sul. Para a presente e seguinte subseções, as unidades consumidoras irrigantes serão identificadas pelas nomenclaturas UC-1, UC-2, UC-3, UC-4, UC-5, UC-6 e UC-7.

A Tabela 4.2 evidencia a definição das classes de perfil de consumo de energia elétrica estabelecidas pela etapa de agrupamento de dados, para cada unidade consumidora irrigante que compõe o conjunto de estudos de caso, com base na média e desvio padrão dos consumos mensais de energia elétrica alusivos ao ano de 2019.

Tabela 4.2 – Classes de perfil de consumo correspondente a cada unidade consumidora irrigante.

Unidade Consumidora Irrigante	m (kWh)	σ (kWh)	Classe de Perfil de Consumo
UC-1	23373	18597,13	Perfil_3
UC-2	17787,33	28515,88	Perfil_2
UC-3	3227	2620,43	Perfil_1
UC-4	20462	7164,91	Perfil_3
UC-5	9063,90	10038,72	Perfil_2
UC-6	8688,60	4463,93	Perfil_2
UC-7	10317,57	16017,85	Perfil_2

Fonte: Elaborado pelo o autor.

As unidades consumidoras irrigantes foram agrupadas em suas respectivas classes, de acordo com os limites estabelecidos pela execução do algoritmo de *k-Means*, apresentados pela Tabela 3.2.

A Tabela 4.3 apresenta os registros mensais de consumo de energia elétrica para as seis unidades consumidoras irrigantes durante o ano de 2020.

A partir dos dados apresentados pela Tabela 4.3, nota-se que o meses de Janeiro, Fevereiro, Novembro e Dezembro apresentam os consumos de energia elétrica mais elevados e expressivos durante todo o ano de 2020, para todas as unidades consumidoras irrigantes analisadas. Esse comportamento característico é bastante comum para esse tipo de consumidor, uma vez que, a atividade de irrigação é realizada em maior escala durante esses intervalos mensais. No entanto, a UC-7 apresentou alguns registros importantes nos meses de

Tabela 4.3 – Dados mensais de consumo de energia elétrica referentes ao ano de 2020.

Dados	Unidades Consumidoras Irrigantes						
Consumo (kWh)	UC-1	UC-2	UC-3	UC-4	UC-5	UC-6	UC-7
JAN	45844	76774	5315	17343	44005	4377	6917
FEV	35538	71258	3339	18286	32601	3731	10379
MAR	0	28928	2453	18693	24	2706	9267
ABR	0	105	0	5576	0	0	299
MAI	0	0	0	0	0	0	3292
JUN	0	13	0	0	0	0	0
JUL	0	13	0	0	0	0	0
AGO	0	8	0	0	0	0	0
SET	0	0	0	0	0	0	0
OUT	0	8	0	0	12	0	0
NOV	21415	53055	44	15162	14832	3007	3587
DEZ	37396	85184	2673	22739	35722	14453	7385

Fonte: Elaborado pelo o autor.

Março e Maio, podendo estar relacionados ao primeiro mês, a operação de irrigação tardia da lavoura de arroz, e ao segundo mês, a alguma outra atividade agropecuária. É importante destacar que, todas unidades consumidoras irrigantes são pertencentes ao Grupo A, ou seja, com tensão de fornecimento igual ou superior a 2,3 kV.

Os dados atribuídos aos sistemas de irrigação dos consumidores rurais foram apresentados e justificados na Seção 3.3. Dessa forma, serão considerados os valores definidos na metodologia para as variáveis das estações de bombeamento em todas unidades consumidoras irrigantes que compõem o conjunto de dados aplicado aos estudos de caso. A caráter de revisão, a Tabela 4.4 demonstra os valores médios aplicados aos seis consumidores para as variáveis de altura monométrica total, vazão unitária e rendimento global do sistema de irrigação.

Tabela 4.4 – Variáveis do sistema de irrigação para todas as unidades consumidoras.

Unidade Consumidora Irrigante	AMT (m)	q (m/s/ha)	η_g
UC-1, UC-2, UC-3, UC-4, UC-5, UC-6 e UC-7	10	1,5	0,65

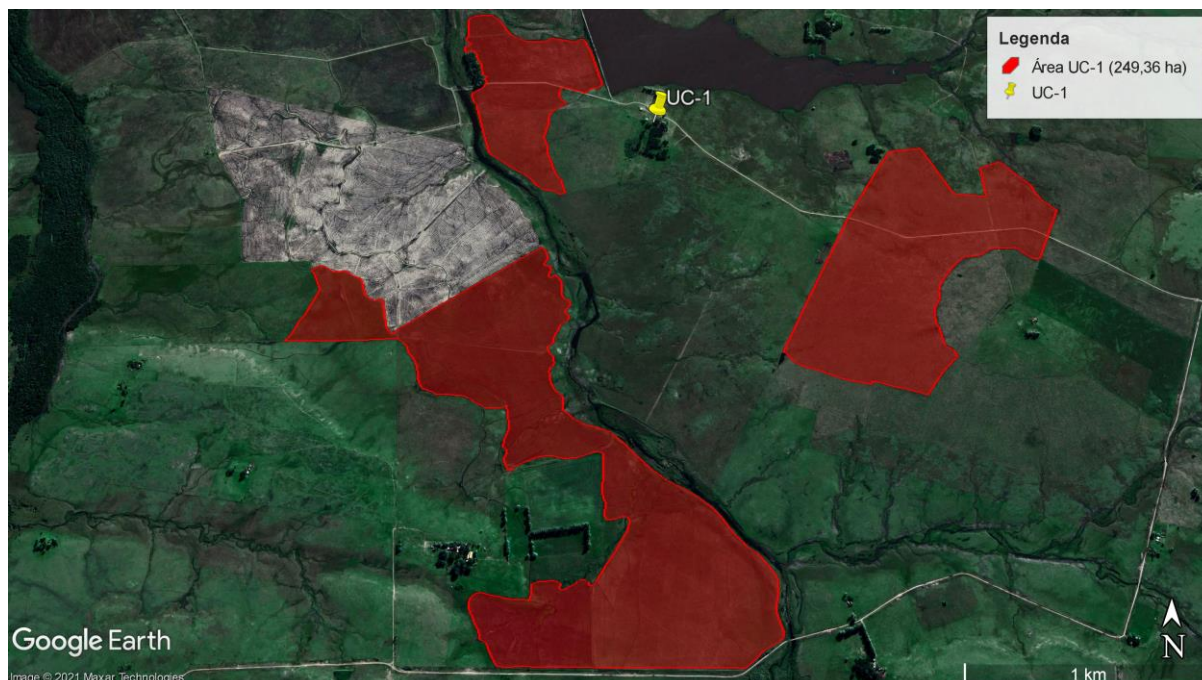
Fonte: Elaborado pelo o autor.

Essas atributos, combinados com a área de plantio necessária para irrigação na lavoura de arroz, são essenciais para estimar a potência ativa instalada nas unidades consumidoras irrigantes, e conseqüentemente, determinar o consumo previsto para essas instalações rurais.

Como reportado na Subseção 3.3.1, a área irrigada é definida a partir das informações de intenção de plantio para cada unidade consumidora irrigante associada. Os dados de plantio são referentes a Safra 2020/21, compreendendo os meses de Novembro e Dezembro de 2020, ano referência nesse estudo para aplicação do modelo preditivo de identificação de possíveis perdas comerciais. Já para os meses de Janeiro e Fevereiro de 2020, serão atributos as áreas de plantio mesma safra, devido a ausência de referências de dados da safra anterior (2019/20).

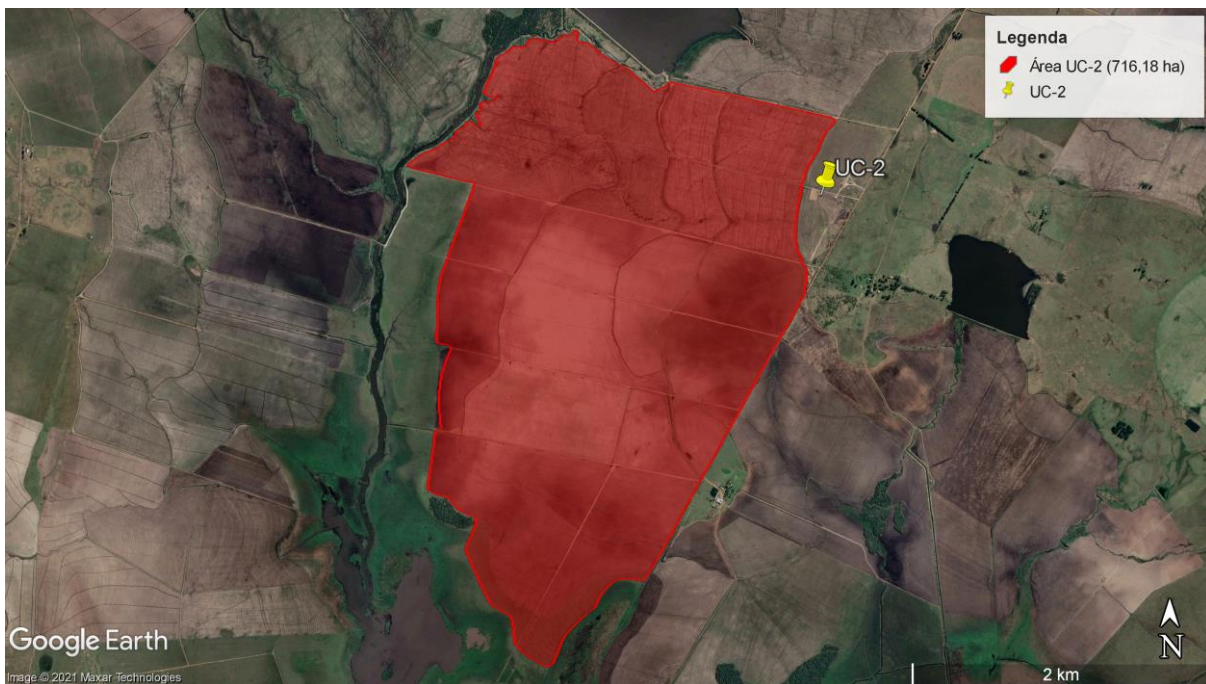
As Figuras 4.1 a 4.7 ilustram imagens de satélite extraídas pelo Google Earth Pro ® para as sete áreas de plantio utilizadas nos estudos de casos.

Figura 4.1 – Área Irrigada para a Unidade Consumidora UC-1.



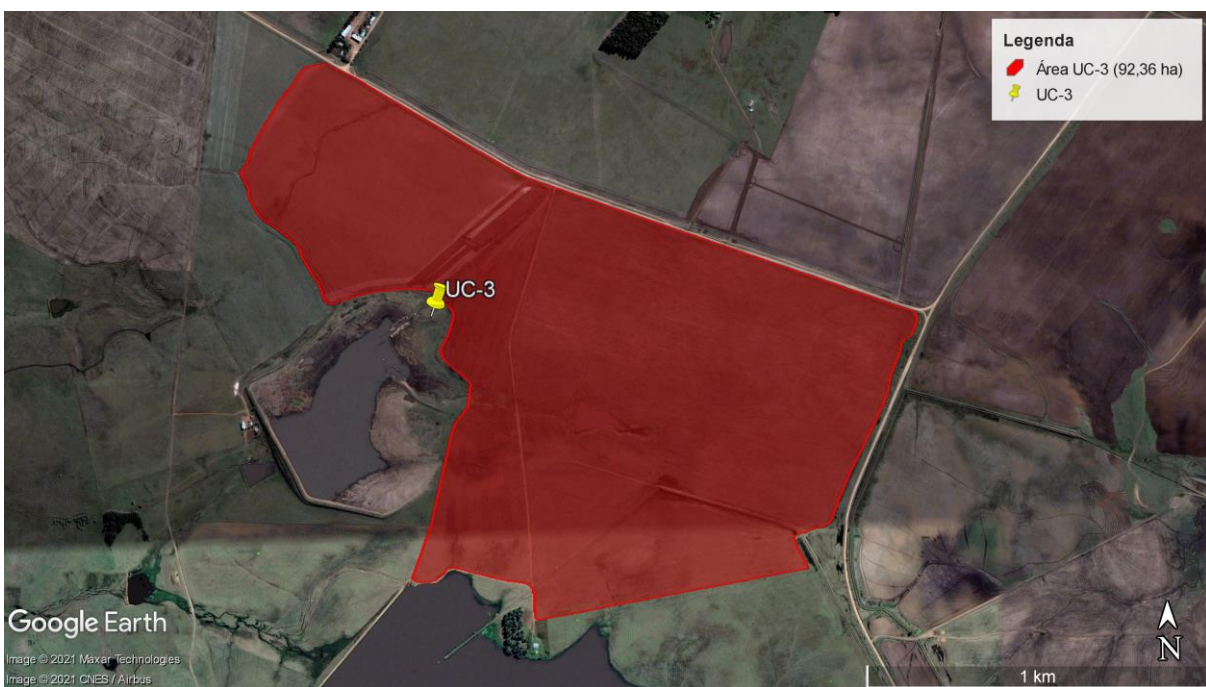
Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro ® (2021).

Figura 4.2 – Área Irrigada para a Unidade Consumidora UC-2.



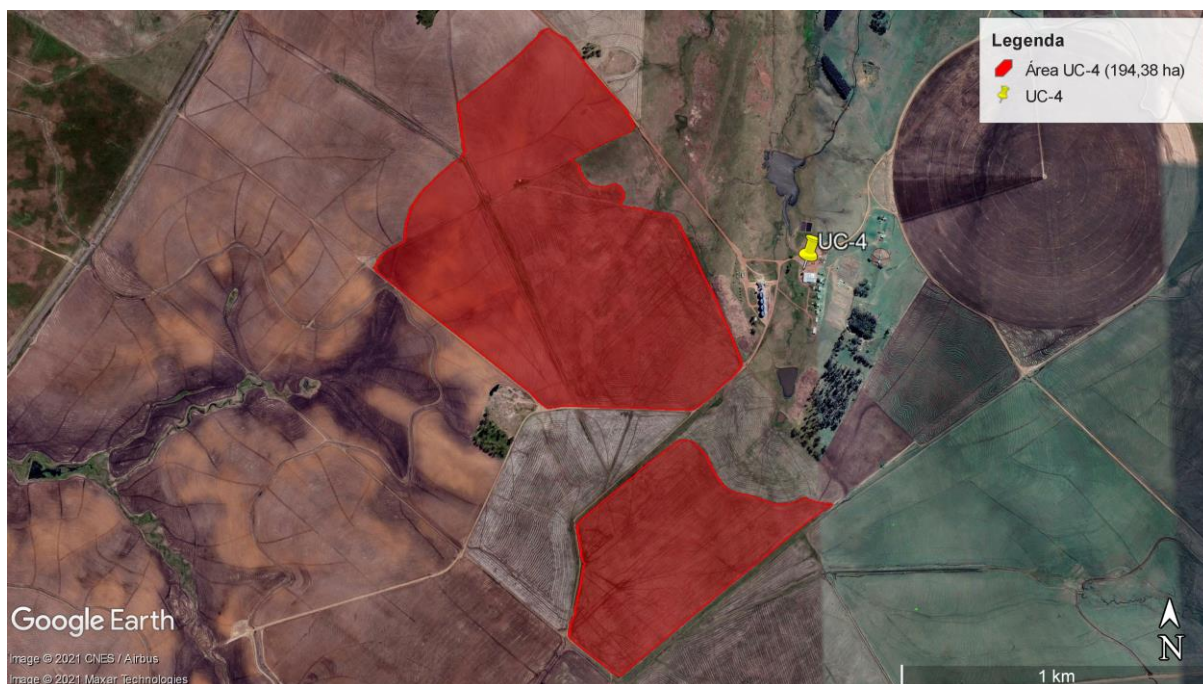
Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro ® (2021).

Figura 4.3 – Área Irrigada para a Unidade Consumidora UC-3.



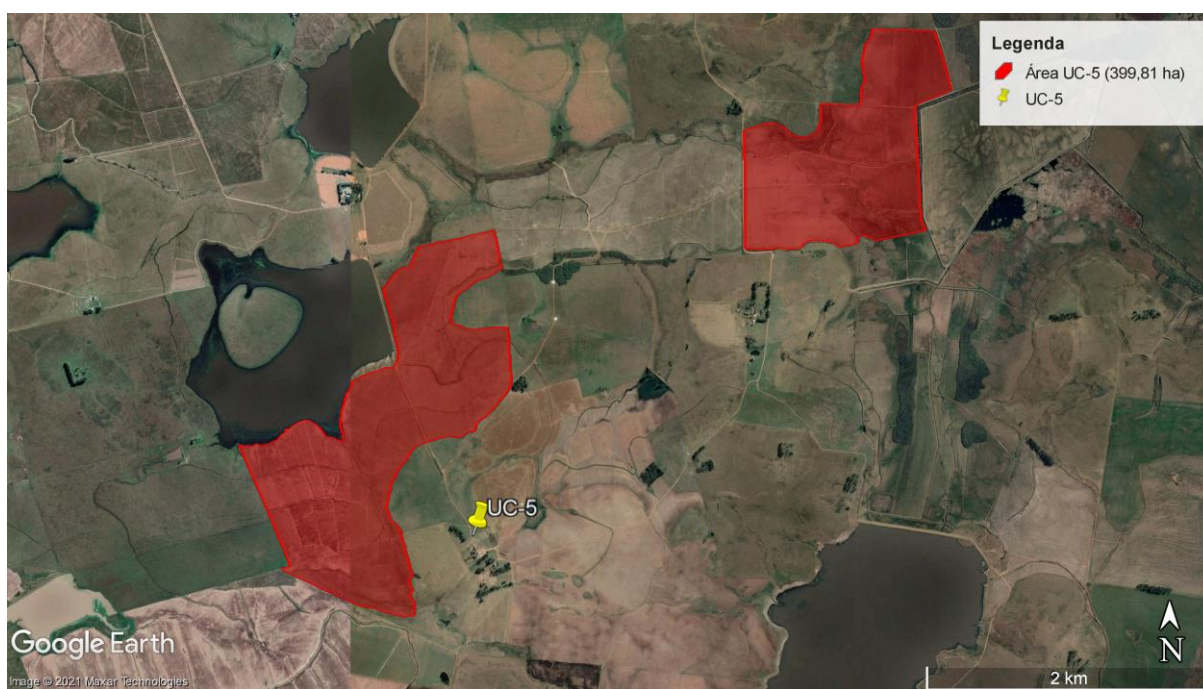
Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro ® (2021).

Figura 4.4 – Área Irrigada para a Unidade Consumidora UC-4.



Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro ® (2021).

Figura 4.5 – Área Irrigada para a Unidade Consumidora UC-5.



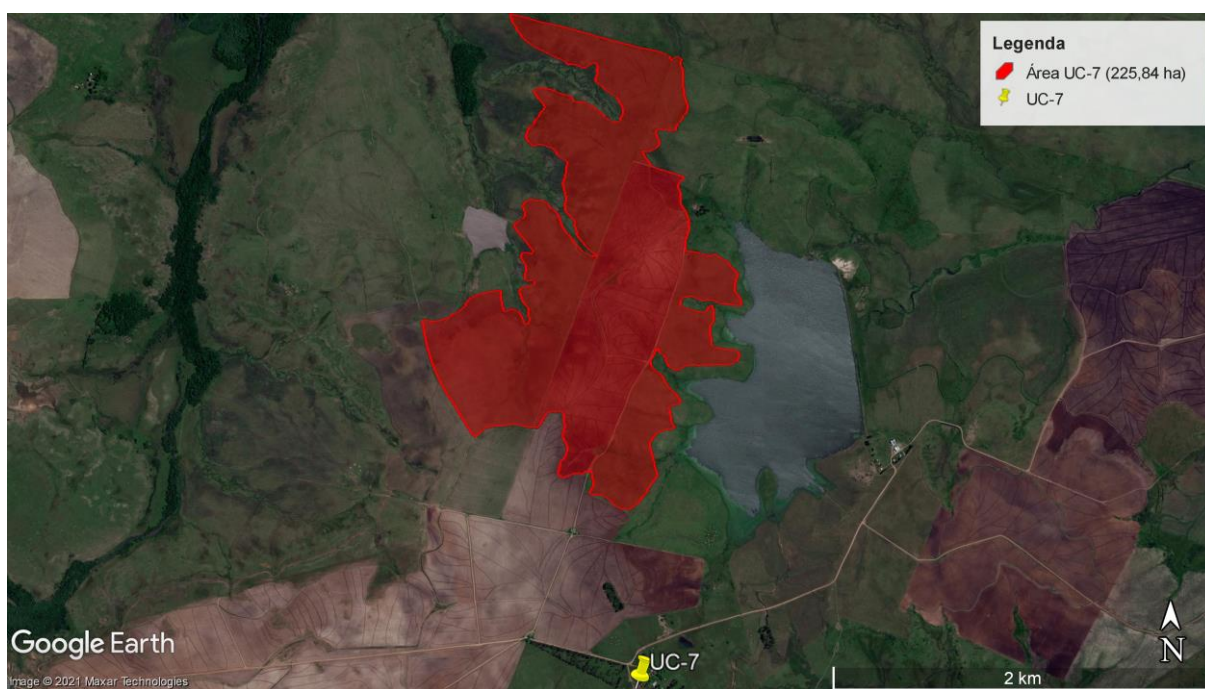
Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro ® (2021).

Figura 4.6 – Área Irrigada para a Unidade Consumidora UC-6.



Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro ® (2021).

Figura 4.7 – Área Irrigada para a Unidade Consumidora UC-7.



Fonte: Elaborado pelo autor com o auxílio do Google Earth Pro ® (2021).

A Tabela 4.5 apresenta os dados gerais de área irrigada das lavouras de arroz para cada unidade consumidora rural associada.

Tabela 4.5 – Dados de área irrigada para cada unidade consumidora associada.

Unidade Consumidora Irrigante	Área Irrigada (ha)
UC-1	249,36
UC-2	716,18
UC-3	92,36
UC-4	194,38
UC-5	399,81
UC-6	116,52
UC-7	225,84

Fonte: Elaborado pelo o autor.

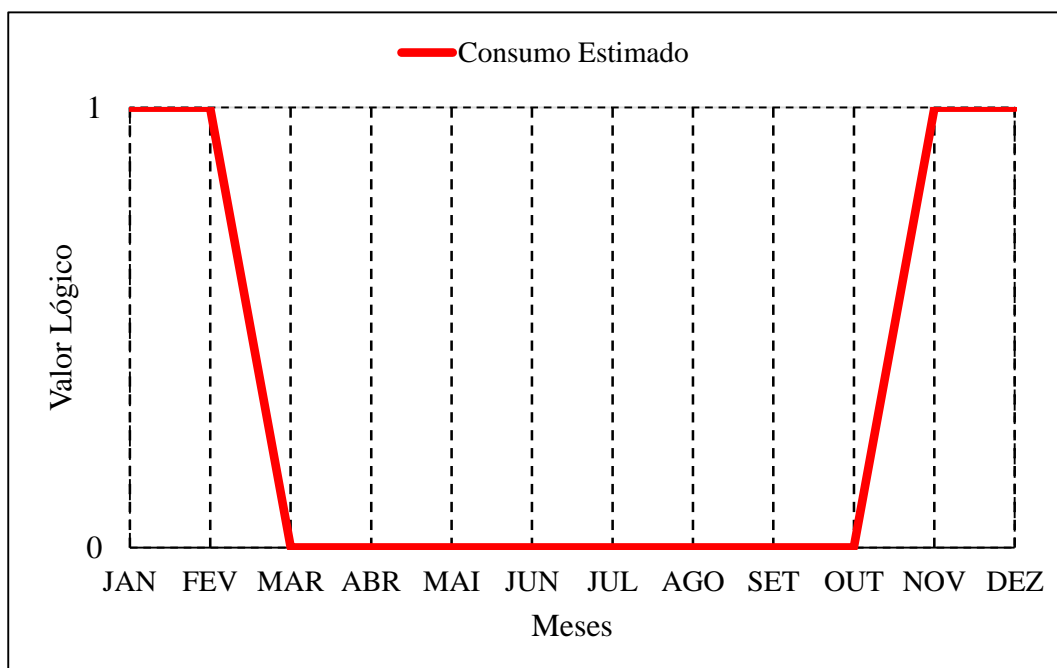
Apresentadas as áreas de plantio associadas as unidades consumidoras irrigantes e as demais variáveis que estruturam os dados dos sistemas de irrigação dos mesmos, são estabelecidas as estimativas da potência ativa instalada e do consumo de energia elétrica das estações de bombeamento para os consumidores rurais analisadas, através das expressões matemáticas apresentadas na Subseção 3.3.6 (Equações 12 e 13).

As estimativas mensais de consumo de energia elétrica das unidades consumidoras irrigantes serão realizadas para os meses de Janeiro, Fevereiro, Novembro e Dezembro, devido a estes, apresentarem aos maiores registros de energia faturada, devido a atividade de irrigação contínua durante esse período. Para os referidos meses, 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 (valor 0). A Figura 4.8 apresenta a curva de carga anual a ser aplicada nas análises de estimativa de consumo de energia elétrica.

Como observado na Figura 4.8, o valor lógico 1 indica o consumo estimado de energia elétrica para o respectivo mês, enquanto que o valor lógico 0 representa os meses em que não serão realizadas as previsões de consumo.

A Tabela 4.6 apresenta as estimativas de potência ativa instalada e consumo de energia elétrica para as unidades consumidoras irrigantes analisadas.

Figura 4.8 – Curva característica da estimativa anual de consumo.



Fonte: Elaborado pelo o autor.

Tabela 4.6 – Potência Ativa e Consumo de Energia Elétrica mensais estimados para cada unidade consumidora irrigante.

Unidade Consumidora Irrigante	Potência Ativa estimada (kW)	Tempo de funcionamento dos Sistemas de Irrigação por mês (h)	Consumo de Energia Elétrica estimado (kWh)
UC-1	56,39	450	25377,17
UC-2	162,11	450	72949,20
UC-3	20,88	450	9399,91
UC-4	43,95	450	19781,90
UC-5	90,42	450	40688,35
UC-6	26,35	450	11858,66
UC-7	51,07	450	22983,56

Fonte: Elaborado pelo o autor.

Apresentado os dados característicos das unidades consumidoras irrigantes empregadas para os testes de validação, a próxima subseção tem por objetivo apresentar os resultados e discussões para a metodologia proposta para identificação de possíveis ocorrência de perdas não técnicas.

4.2 RESULTADOS E DISCUSSÕES

A presente seção, tem por objetivo apresentar os resultados das simulações do modelo preditivo de identificação de perdas não técnicas nas unidades consumidoras irrigantes empregadas como estudos de casos. Como mencionado no Capítulo 3, o algoritmo elaborado para metodologia proposta nesse estudo, foi desenvolvido em linguagem *Python 3.7* e executado em uma plataforma computacional de análises matemáticas para essa programação específica, nesse caso, o IDE Spyder®. A classe de perfil de consumo predita e a faixa de variação aceitável (40%) do consumo estimado em relação ao consumo real, são as variáveis objetivos do algoritmo, que indicará a possibilidade de perdas comerciais pelos consumidores reais analisados.

Nas subseções a seguir, serão apresentados e discutidos os resultados das simulações do modelo de previsão de possíveis perdas não técnicas para cada um dos consumidores rurais, utilizados como objeto para os estudos de caso da presente metodologia desenvolvida. As análises serão realizadas de maneira individual, ou seja, para cada consumidor, avaliando inicialmente, o indicador de classe de perfil de consumo e posteriormente, o parâmetro da estimativa mensal de consumo de energia elétrica.

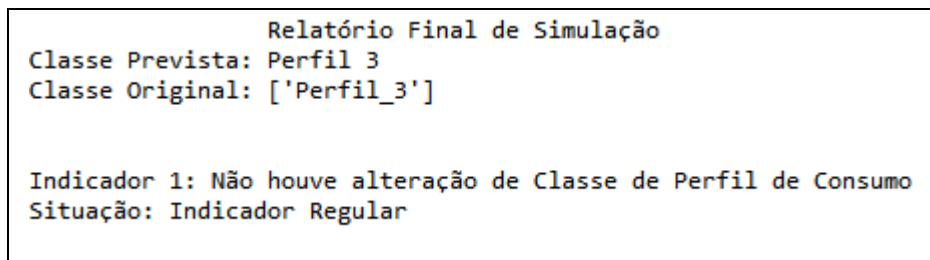
4.2.1 Unidade Consumidora Irrigante UC-1

Nessa subseção, serão apresentados os relatórios finais de apuração dos indicadores propostos pelo modelo preditivo de detecção de possíveis ocorrências de perdas não técnicas para o consumidor irrigante UC-1. O *script* completo do algoritmo para o presente estudo de caso será apresentado no Capítulo de Apêndices, no item A.1.

A Figura 4.9 retrata o relatório final de simulação para o indicador de classe de perfil de consumo realizada no IDE Spyder®.

A partir das informações exibidas na Figura 4.9, nota-se que para a unidade consumidora irrigante UC-1, a classe de perfil de consumo prevista para o ano de 2020 classificada pelo modelo preditivo, é equivalente ao grupo ou cluster definido pela Etapa de Agrupamento realizado a partir dos registros mensais de consumo referentes ao ano de 2019. Assim, tanto a classe prevista quanto a original são pertencentes ao grupo de consumo “Perfil_3”. Dessa modo, o presente indicador é estabelecido como regular, não apontando a

Figura 4.9 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-1 no IDE Spyder®.

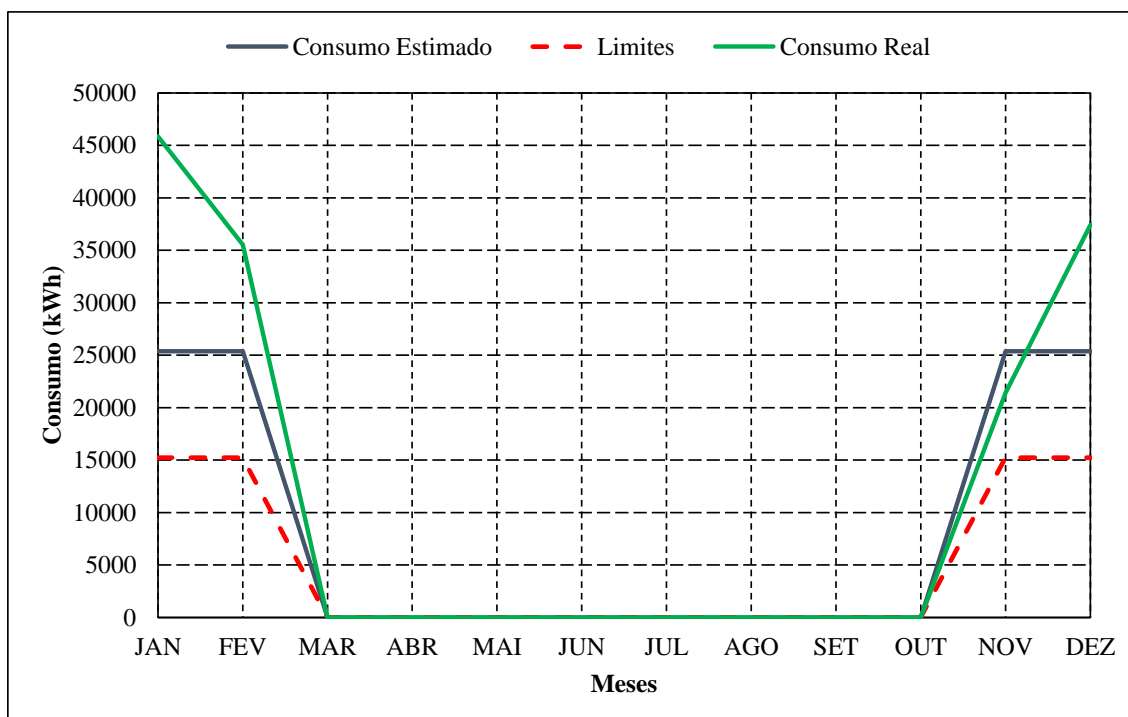


Fonte: Elaborado pelo o autor.

possibilidade de perdas não técnicas para esse consumidor irrigante durante a respectiva avaliação.

A Figura 4.10 apresenta um comparativo na curva de carga mensal entre os consumos mensais reais e previsto e os respectivos limites de tolerância para o indicativo de ocorrência de perdas não técnicas para o consumidor irrigante UC-1. Já a Figura 4.11, ilustra o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no IDE Spyder®.

Figura 4.10 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-1.



Fonte: Elaborado pelo o autor.

Figura 4.11 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-1 no IDE Spyder®.

Relatório Final de Simulação	
Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados	
Situação: Indicador Regular	

Fonte: Elaborado pelo o autor.

Os resultados apresentados pelas Figuras 4.10 e 4.11, 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 comerciais. Dessa modo, o presente indicador é estabelecido como regular, não apontando a possibilidade de perdas não técnicas para esse consumidor irrigante durante a respectiva avaliação.

Para os meses onde as estimativas foram inferiores em uma diferença significativa ao consumo real praticado pela unidade consumidora irrigante, como nos casos de Janeiro, Fevereiro e Dezembro de 2020, a rotatividade de área entre as safras pode ser a justificativa para essa situação, principalmente para os primeiros dois meses citados, uma vez que os dados aplicados nos estudos de caso, não referentes a safra que compreende a esse período, como mencionado e fundamentado na Seção 4.1.

A Tabela 4.7 ilustra um resumo geral dos indicadores avaliados para o cliente UC-1.

Tabela 4.7 – Resumo geral dos indicadores avaliados para a UC-1.

Indicador: Classe de Perfil de Consumo para a UC-1					
Nomenclatura	Classe		Situação		
Original	Perfil_3		Indicador Regular		
Previsto	Perfil_3				
Indicador: Estimativa de consumo de energia elétrica para a UC-1					
Mês	Real (kWh)	Estimado (kWh)	Limite (kWh)	Situação	
JAN	45844	25377,17	15226,30	Indicador Regular	
FEV	35538	25377,17	15226,30		
NOV	21415	25377,17	15226,30		
DEZ	37396	25377,17	15226,30		

Fonte: Elaborado pelo o autor.

Conforme apurado pelos indicadores, a unidade consumidora irrigante UC-1 apresenta regularidade em ambas as situações analisadas. Desse modo, o respectivo consumidor é descartado pelo modelo preditivo como uma situação de possíveis práticas de perdas não técnicas.

4.2.2 Unidade Consumidora Irrigante UC-2

Nessa subseção, serão demonstrados os relatórios finais de apuração dos indicadores propostos pelo modelo preditivo de identificação de possíveis ocorrências de perdas não técnicas para a unidade consumidora rural UC-2. O *script* completo do algoritmo para o presente estudo de caso será apresentado no Capítulo de Apêndices, no item A.2.

A Figura 4.12 ilustra o relatório final de simulação para o indicador de classe de perfil de consumo realizada no IDE Spyder®.

Figura 4.12 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-2 no IDE Spyder®.

```
Relatório Final de Simulação
Classe Prevista: Perfil 4
Classe Original: ['Perfil_2']

Indicador 1: Alteração de Classe de Perfil de Consumo
Situação: Indicador Regular
```

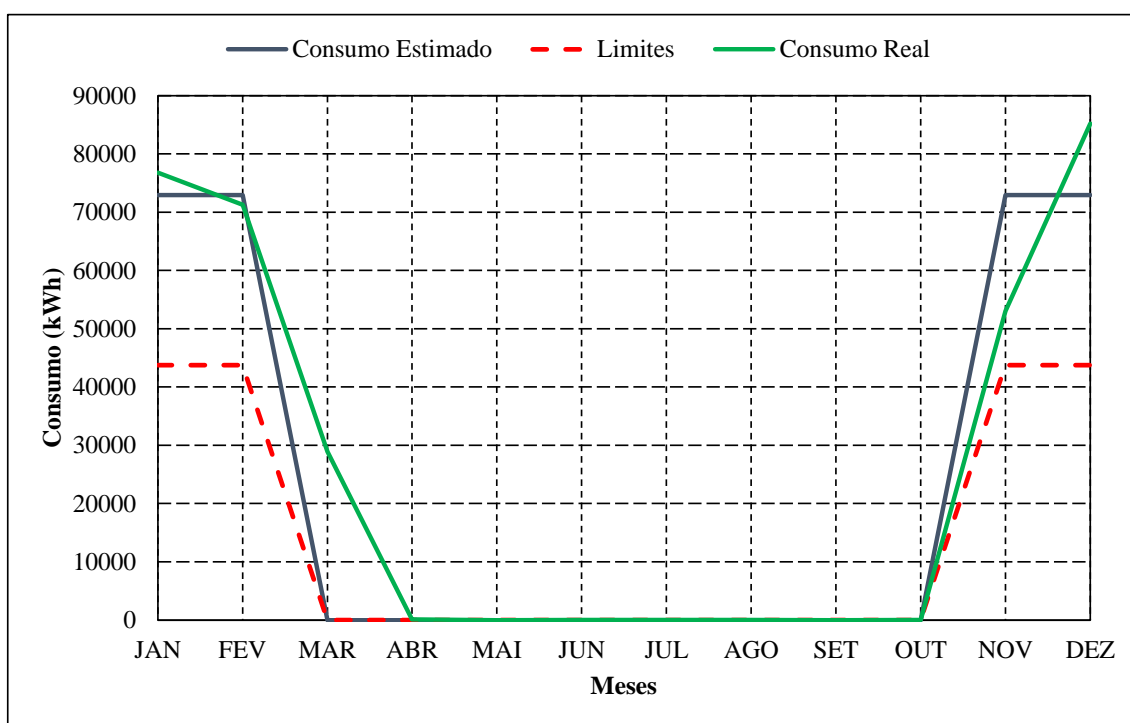
Fonte: Elaborado pelo o autor.

De acordo com o relatório apresentado pela Figura 4.12, a classe prevista para o ano de 2020 da unidade consumidora irrigante UC-2, classificadas pelo modelo preditivo, difere do grupo de consumo definido pela Etapa de Clusterização. 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-2, para avaliação desse indicador.

A Figura 4.13 demonstra um comparativo na curva de carga mensal entre os consumos mensais reais e previsto e os respectivos limites de tolerância para o indicativo de ocorrência de perdas comerciais para o consumidor irrigante UC-2. Já a Figura 4.14, apresenta o relatório

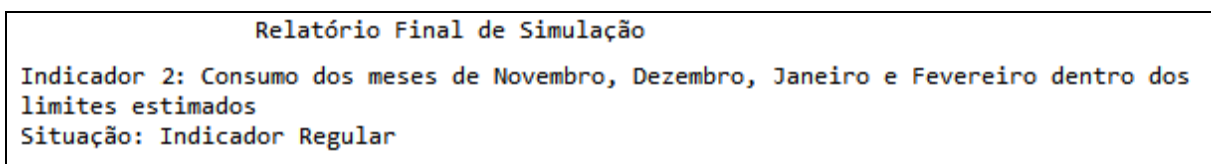
final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no IDE Spyder®.

Figura 4.13 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-2.



Fonte: Elaborado pelo o autor.

Figura 4.14 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-2 no IDE Spyder®.



Fonte: Elaborado pelo o autor.

Os resultados apresentados pelas Figuras 4.13 e 4.14, apontam 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 comerciais. Desse modo, o presente indicador é estabelecido como regular, não indicando a possibilidade de perdas não técnicas para a respectiva avaliação.

Ainda sobre a Figura 4.13, nota-se que os resultados de estimativas de consumo foram satisfatórios para essa unidade consumidora irrigante. Nos meses de Janeiro, Fevereiro e Dezembro, os valores previstos foram muito próximos aos consumo de energia elétrica reais registrados. Para o período de Março, acredita-se que houve uma atividade de irrigação tardia na lavoura de arroz desse consumidor rural. No entanto, esse mês não é avaliado pelo indicador do modelo preditivo, como justificado no início desse capítulo.

A Tabela 4.8 apresenta um resumo geral dos indicadores avaliados para a unidade consumidora irrigante UC-2.

Tabela 4.8 – Resumo geral dos indicadores avaliados para a UC-2.

Indicador: Classe de Perfil de Consumo para a UC-2					
Nomenclatura		Classe	Situação		
Original		Perfil_2	Indicador Regular		
Previsto		Perfil_4			
Indicador: Estimativa de consumo de energia elétrica para a UC-2					
Mês	Real (kWh)	Estimado (kWh)	Limite (kWh)	Situação	
JAN	76774	72949,20	43769,52	Indicador Regular	
FEV	71258	72949,20	43769,52		
NOV	53055	72949,20	43769,52		
DEZ	85184	72949,20	43769,52		

Fonte: Elaborado pelo o autor.

Conforme apurado pelos indicadores, a unidade consumidora irrigante UC-2 apresentou regularidade em ambas as situações analisadas. Dessa forma o respectivo consumidor é descartado pelo modelo preditivo como uma situação de possíveis práticas de perdas não técnicas.

4.2.3 Unidade Consumidora Irrigante UC-3

Nessa subseção, serão apresentados os relatórios finais de apuração dos indicadores propostos pelo modelo preditivo de detecção de possíveis ocorrências de perdas não técnicas para o consumidor irrigante UC-3. O *script* completo do algoritmo para o presente estudo de caso será apresentado no Capítulo de Apêndices, no item A.3.

A Figura 4.15 demonstra o relatório final de simulação para o indicador de classe de perfil de consumo realizada no IDE Spyder®.

Figura 4.15 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-3 no IDE Spyder®.

```
Relatório Final de Simulação
Classe Prevista: Perfil 1
Classe Original: ['Perfil_1']

Indicador 1: Não houve alteração de Classe de Perfil de Consumo
Situação: Indicador Regular
```

Fonte: Elaborado pelo o autor.

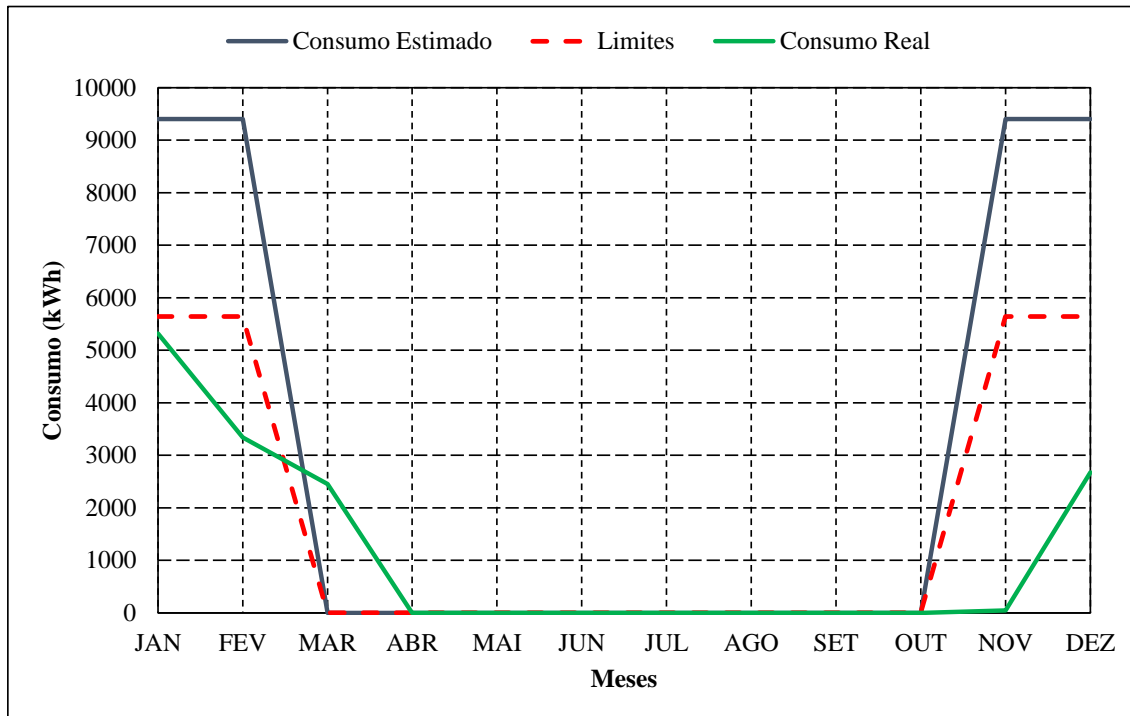
Com base nas informações apresentadas na Figura 4.15, nota-se que para a unidade consumidora irrigante UC-3, a classe de perfil de consumo prevista para o ano de 2020 classificada pelo modelo preditivo, é igual ao cluster definido pela Etapa de Agrupamento. Assim, tanto a classe prevista quanto a original são pertencentes ao grupo de consumo “Perfil_1”. Dessa modo, o presente indicador é instituído como regular, não apontando a possibilidade de perdas não técnicas para esse consumidor irrigante durante a referente avaliação.

A Figura 4.16 apresenta um comparativo na curva de carga mensal entre os consumos mensais reais e previsto e os respectivos limites de tolerância para o indicativo de ocorrência de perdas não técnicas para o consumidor irrigante UC-3. Já a Figura 4.17, demonstra o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no IDE Spyder®.

Os resultados apresentados nas Figuras 4.16 e 4.17, indicam que os consumos registrados para a respectiva unidade consumidora irrigante nos meses de Janeiro, Fevereiro, Novembro e Dezembro, foram inferiores aos limites mínimos mensais estabelecidos para o indicativo de perdas não técnicas. Dessa forma, o consumidor rural avaliado é apontando como um agente suspeito de praticar perdas não técnicas na avaliação do presente indicador.

Observando a Figura 4.16, a situação para o mês de Março é semelhante ao ocorrido nas análises da unidade consumidora irrigante UC-2, ou seja, o consumo significativo de energia elétrica nesse período pode estar relacionado com a atividade de irrigação tardia na lavoura de arroz.

Figura 4.16 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-3.



Fonte: Elaborado pelo o autor.

Figura 4.17 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-3 no IDE Spyder®.

Relatório Final de Simulação

Indicador 2: Consumo do mês de Novembro menor do que o limite estimado
 Situação: Possível ocorrência de Perdas Não Técnicas

Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado
 Situação: Possível ocorrência de Perdas Não Técnicas

Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado
 Situação: Possível ocorrência de Perdas Não Técnicas

Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado
 Situação: Possível ocorrência de Perdas Não Técnicas

Fonte: Elaborado pelo o autor.

A Tabela 4.9 apresenta um resumo geral dos indicadores avaliados para a unidade consumidora irrigante UC-3.

Tabela 4.9 – Resumo geral dos indicadores avaliados para a UC-3.

Indicador: Classe de Perfil de Consumo para a UC-3				
Nomenclatura		Classe	Situação	
Original		Perfil_1	Indicador Regular	
Previsto		Perfil_1		
Indicador: Estimativa de consumo de energia elétrica para a UC-3				
Mês	Real (kWh)	Estimado (kWh)	Limite (kWh)	Situação
JAN	5315	9399,91	5639,95	Possibilidade de ocorrência de Perdas Não Técnicas
FEV	3339	9399,91	5639,95	
NOV	44	9399,91	5639,95	
DEZ	2673	9399,91	5639,95	

Fonte: Elaborado pelo o autor.

Conforme apurado pelos indicadores propostos pelo modelo preditivo, a unidade consumidora irrigante UC-3 apresentou irregularidade no parâmetro de estimativa de consumo de energia elétrica. Dessa forma, o respectivo consumidor é apontado pelo modelo preditivo como uma situação de possíveis práticas de perdas não técnicas.

4.2.4 Unidade Consumidora Irrigante UC-4

Nessa subseção, serão apresentados os relatórios finais de apuração dos indicadores propostos pelo modelo preditivo de detecção de possíveis ocorrências de perdas não técnicas para a unidade consumidora irrigante UC-4. O script completo do algoritmo para o presente estudo de caso será apresentado no Capítulo de Apêndices, no item A.4.

A Figura 4.18 apresenta o relatório final de simulação para o indicador de classe de perfil de consumo realizada no IDE Spyder®.

Segundo o relatório demonstrado na Figura 4.18, a classe prevista para o ano de 2020 da unidade consumidora irrigante UC-5, classificada pelo modelo preditivo, diverge do grupo de consumo definido pela Etapa de Agrupamento de Dados. A classe de perfil de consumo

Figura 4.18 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-4 no IDE Spyder®.

```
Relatório Final de Simulação
Classe Prevista: Perfil 2
Classe Original: ['Perfil_3']

Indicador 1: Alteração de Classe de Perfil de Consumo
Situação: Possível ocorrência de Perdas Não Técnicas
```

Fonte: Elaborado pelo o autor.

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-4 é indicado nesse avaliador como possibilidade de práticas de perdas não técnicas.

A Figura 4.19 apresenta um comparativo na curva de carga mensal entre os consumos mensais reais e previsto e os respectivos limites de tolerância para o indicativo de ocorrência de perdas comerciais para o consumidor irrigante UC-4. Já a Figura 4.20, apresenta o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no IDE Spyder®.

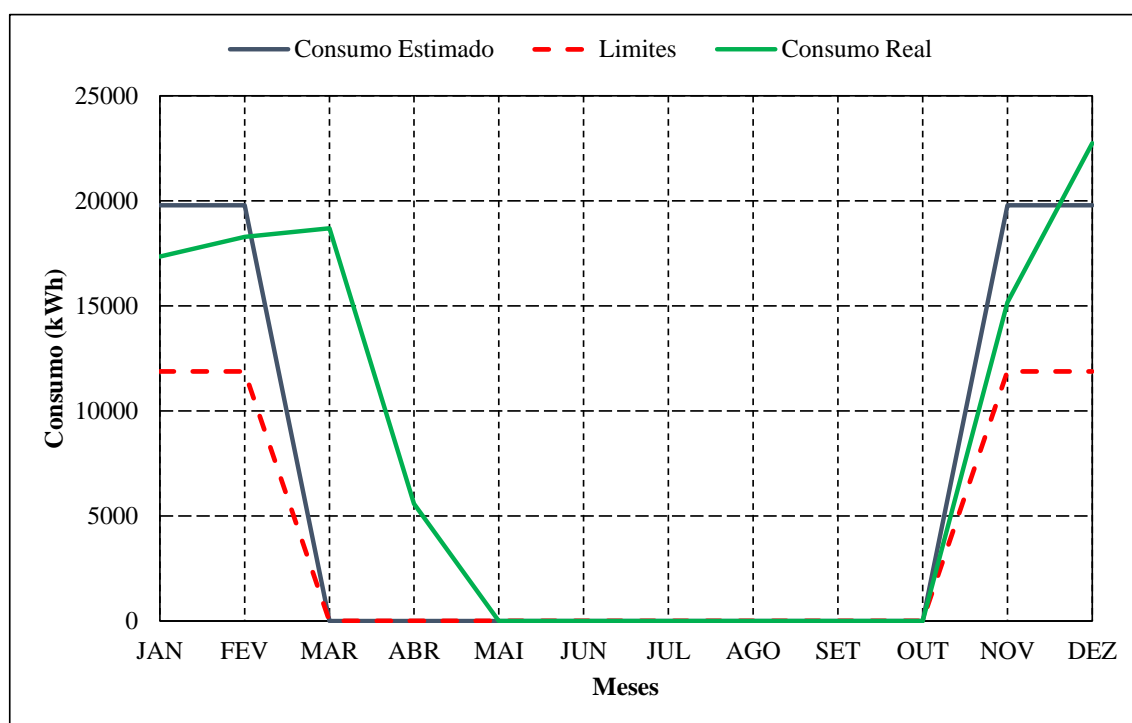
Os resultados apresentados pelas Figuras 4.19 e 4.20, apontam que os consumos registrados para a respectiva unidade consumidora nos meses de Janeiro, Fevereiro, Novembro e Dezembro, não infringiram os limites mínimos mensais estabelecidos para o indicativo de perdas comerciais. Dessa forma, o presente indicador é estabelecido como regular, não indicando a possibilidade de perdas não técnicas para a respectiva avaliação.

Observando a Figura 4.19, nota-se que as estimativas de consumo foram satisfatórias para esse consumidor. Nos meses de Janeiro, Fevereiro e Dezembro, os valores previstos foram muito próximos aos consumo de energia elétrica reais registrados. Para o meses de Março e Abril, acredita-se que houve uma atividade de irrigação tardia na lavoura de arroz dessa unidade consumidora irrigante.

A Tabela 4.10 descreve um resumo geral dos indicadores avaliados para a unidade consumidora irrigante UC-4.

De acordo com as informações apresentadas na Tabela 4.10, a unidade consumidora irrigante UC-4 apresentou irregularidades no indicador de classe de perfil de consumo. Dessa forma, o respectivo consumidor é apontado pelo modelo preditivo como uma situação de possíveis práticas de perdas não técnicas.

Figura 4.19 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-4.



Fonte: Elaborado pelo o autor.

Figura 4.20 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-4 no IDE Spyder®.

Relatório Final de Simulação
Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados
Situação: Indicador Regular

Fonte: Elaborado pelo o autor.

Tabela 4.10 – Resumo geral dos indicadores avaliados para a UC-4.

Indicador: Classe de Perfil de Consumo para a UC-4				
Nomenclatura		Classe	Situação	
Original		Perfil_3	Possibilidade de ocorrência de Perdas Não Técnicas	
Previsto		Perfil_2		
Indicador: Estimativa de consumo de energia elétrica para a UC-4				
Mês	Real (kWh)	Estimado (kWh)	Limite (kWh)	Situação
JAN	17343	19781,90	11869,14	Indicador Regular
FEV	18286	19781,90	11869,14	
NOV	15162	19781,90	11869,14	
DEZ	22739	19781,90	11869,14	

Fonte: Elaborado pelo o autor.

4.2.5 Unidade Consumidora Irrigante UC-5

Nessa subseção, serão demonstrados os relatórios finais de apuração dos indicadores propostos pelo modelo preditivo de identificação de possíveis ocorrências de perdas não técnicas a unidade consumidora rural UC-5. O *script* completo do algoritmo para o presente estudo de caso será apresentado no Capítulo de Apêndices, no item A.5.

A Figura 4.21 descreve o relatório final de simulação para o indicador de classe de perfil de consumo realizada no IDE Spyder®.

Figura 4.21 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-5 no IDE Spyder®.

```

Relatório Final de Simulação
Classe Prevista: Perfil 3
Classe Original: ['Perfil_2']

Indicador 1: Alteração de Classe de Perfil de Consumo
Situação: Indicador Regular

```

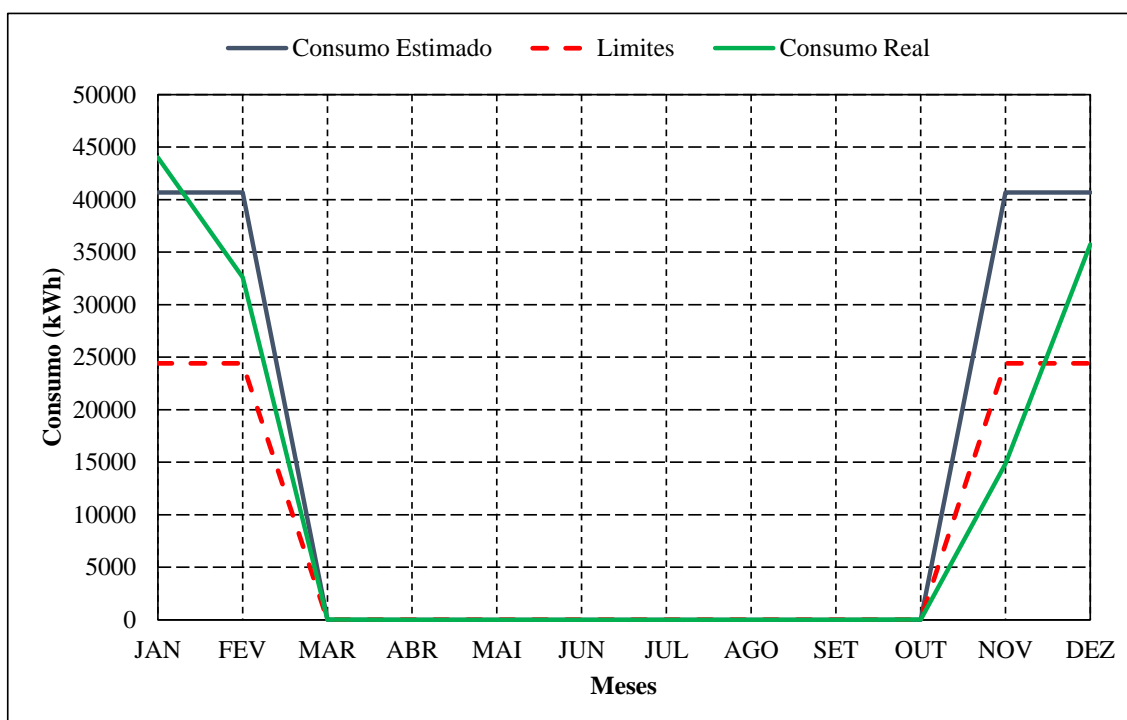
Fonte: Elaborado pelo o autor.

Segundo o relatório apresentado pela Figura 4.21, a classe prevista para o ano de 2020 da unidade consumidora irrigante UC-5, classificada pelo modelo preditivo, diverge do grupo

de consumo definido pela Etapa de Agrupamento de Dados. 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_3”. Entretanto, como ocorreu a mudança para uma classe superior de perfil de consumo, não é assinalada a possibilidade de ocorrência de perdas não técnica pela unidade consumidor irrigante UC-5, para avaliação desse indicador.

A Figura 4.22 ilustra um comparativo na curva de carga mensal entre os consumos mensais reais e previsto e os respectivos limites de tolerância para o indicativo de ocorrência de perdas não técnicas para o consumidor irrigante UC-5. Já a Figura 4.23, descreve o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no IDE Spyder®.

Figura 4.22 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-5.



Fonte: Elaborado pelo o autor.

Figura 4.23 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-5 no IDE Spyder ®.

Relatório Final de Simulação	
Indicador 2: Consumo do mês de Novembro menor do que o limite estimado	
Situação: Possível ocorrência de Perdas Não Técnicas	

Fonte: Elaborado pelo o autor.

Os resultados apresentados nas Figuras 4.22 e 4.23, indicam que o mês de Novembro apresentou um valor de consumo de energia elétrica inferior inferiores aos limites mínimos mensais definidos para o indicativo de ocorrência de perdas não técnicas. Dessa modo, o a unidade consumidora irrigante avaliada é apontada para esse indicador, como possibilidade de práticas de perdas comerciais.

Os resultados das estimativas de consumo foram satisfatórios nos meses de Janeiro, Fevereiro e Dezembro para esse consumidor rural, atingindo patamares próximos aos consumo reais efetivamente praticados.

A Tabela 4.11 apresenta um resumo geral dos indicadores avaliados para a unidade consumidora irrigante UC-5.

Tabela 4.11 – Resumo geral dos indicadores avaliados para a UC-5.

Indicador: Classe de Perfil de Consumo para a UC-5				
Nomenclatura	Classe		Situação	
Original	Perfil_2		Indicador Regular	
Previsto	Perfil_3			
Indicador: Estimativa de consumo de energia elétrica para a UC-5				
Mês	Real (kWh)	Estimado (kWh)	Limite (kWh)	Situação
JAN	44005	40688,35	24413,01	Possibilidade de ocorrência de Perdas Não Técnicas
FEV	32601	40688,35	24413,01	
NOV	14832	40688,35	24413,01	
DEZ	35722	40688,35	24413,01	

Fonte: Elaborado pelo o autor.

De acordo com as informações apresentadas na Tabela 4.11, a unidade consumidora irrigante UC-5 apresentou irregularidades no indicador de estimativa de consumo de energia elétrica. Desse modo, o respectivo consumidor é apontado pelo modelo preditivo como uma situação de possíveis práticas de perdas não técnicas.

4.2.6 Unidade Consumidora Irrigante UC-6

Nessa subseção, serão apresentados os relatórios finais de apuração dos indicadores propostos pelo modelo preditivo de detecção de possíveis ocorrências de perdas não técnicas para o consumidor irrigante UC-6. O *script* completo do algoritmo para o presente estudo de caso será apresentado no Capítulo de Apêndices, no item A.6.

A Figura 4.24 apresenta o relatório final de simulação para o indicador de classe de perfil de consumo realizada no IDE Spyder®.

Figura 4.24 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-6 no IDE Spyder®.

```

                                Relatório Final de Simulação
Classe Prevista: Perfil 1
Classe Original: ['Perfil_2']

Indicador 1: Alteração de Classe de Perfil de Consumo
Situação: Possível ocorrência de Perdas Não Técnicas

```

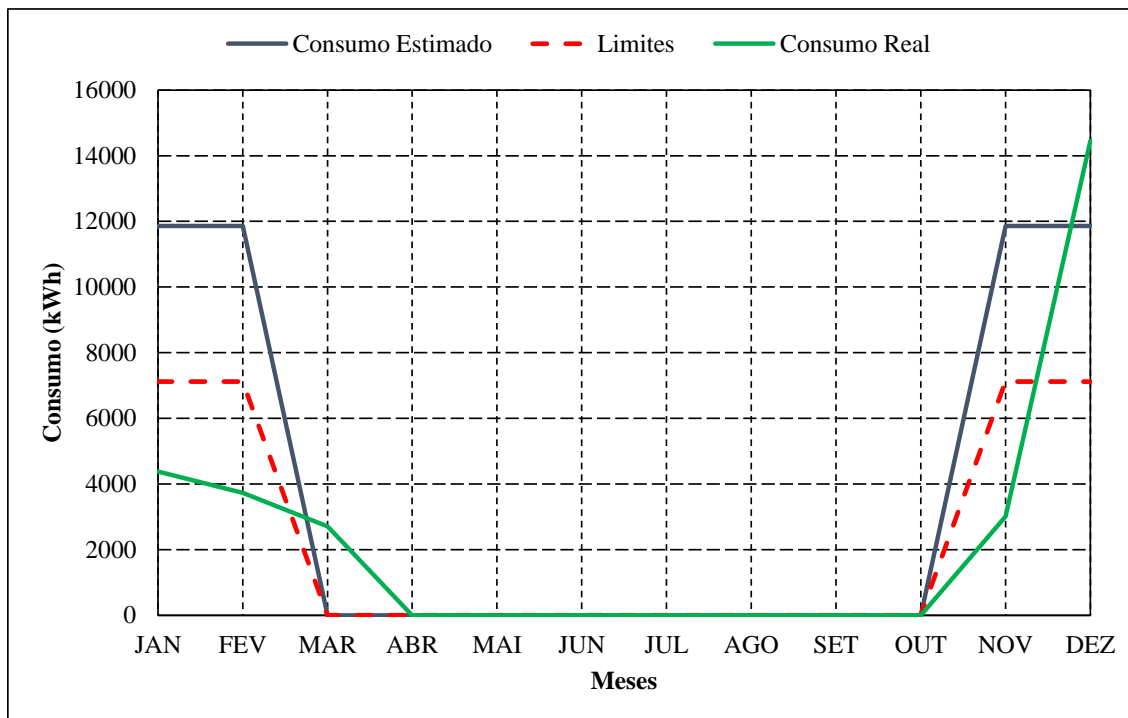
Fonte: Elaborado pelo o autor.

De acordo com o relatório demonstrado na Figura 4.24, a classe prevista para o ano de 2020 da unidade consumidora irrigante UC-6, classificada pelo modelo preditivo, diverge do grupo de consumo definido pela Etapa de Clusterização. 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 de classe de perfil de consumo, o consumidor rural UC-6 é indicado nesse avaliador como possibilidade de práticas de perdas não técnicas.

A Figura 4.25 apresenta um comparativo na curva de carga mensal entre os consumos mensais reais e previsto e os respectivos limites de tolerância para o indicativo de ocorrência

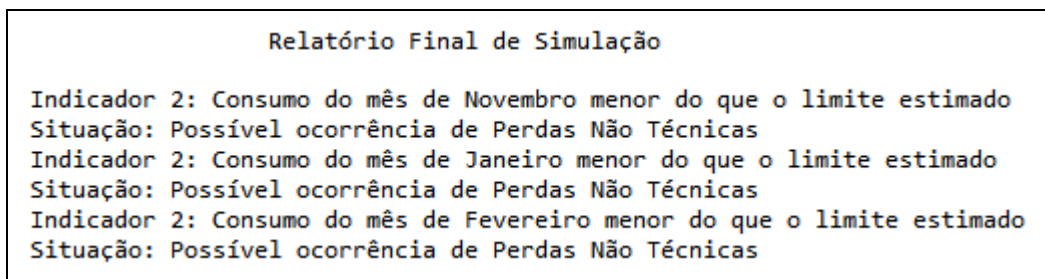
de perdas comerciais para o consumidor irrigante UC-6. Já a Figura 4.26, apresenta o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no IDE Spyder®.

Figura 4.25 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-6.



Fonte: Elaborado pelo o autor.

Figura 4.26 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-6 no IDE Spyder®.



Fonte: Elaborado pelo o autor.

Os resultados apresentados nas Figuras 4.25 e 4.26, demonstram que os consumos mensais registrados para a unidade consumidora UC-6 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 apontado para esse indicador, como um possível atuador de práticas de perdas não técnicas.

Observando a Figura 4.25, a situação para o mês de Março é semelhante ao ocorrido nas análises dos consumidores UC-2, UC-3 e UC-4, ou seja, o consumo de energia elétrica nesse período pode estar relacionado com a atividade de irrigação tardia para a lavoura de arroz associada a esta unidade consumidora.

A Tabela 4.12 apresenta um resumo geral dos indicadores avaliados para a unidade consumidora irrigante UC-6.

Tabela 4.12 – Resumo geral dos indicadores avaliados para a UC-6.

Indicador: Classe de Perfil de Consumo para a UC-6				
Nomenclatura		Classe	Situação	
Original		Perfil_2	Possibilidade de ocorrência de Perdas Não Técnicas	
Previsto		Perfil_1		
Indicador: Estimativa de consumo de energia elétrica para a UC-6				
Mês	Real (kWh)	Estimado (kWh)	Limite (kWh)	Situação
JAN	4377	11858,66	7115,20	Possibilidade de ocorrência de Perdas Não Técnicas
FEV	3731	11858,66	7115,20	
NOV	3007	11858,66	7115,20	
DEZ	14453	11858,66	7115,20	

Fonte: Elaborado pelo o autor.

De acordo com os dados exibidos na Tabela 4.12, a unidade consumidora irrigante UC-6 apresentou irregularidades em ambos os indicadores. Desse modo, a mesma é apontada pelo modelo preditivo como uma situação de possíveis práticas de perdas não técnicas.

4.2.7 Unidade Consumidora Irrigante UC-7

Nessa subseção, serão demonstrados os relatórios finais de apuração dos indicadores propostos pelo modelo preditivo de detecção de possíveis ocorrências de perdas não técnicas para a unidade consumidora rural UC-7. O *script* completo do algoritmo para o presente estudo de caso será apresentado no Capítulo de Apêndices, no item A.7.

A Figura 4.27 descreve o relatório final de simulação para o indicador de classe de perfil de consumo realizada no IDE Spyder®.

Figura 4.27 – Relatório final de simulação do indicador de classe de perfil de consumo para a UC-7 no IDE Spyder®.

```
Relatório Final de Simulação
Classe Prevista: Perfil 1
Classe Original: ['Perfil_2']

Indicador 1: Alteração de Classe de Perfil de Consumo
Situação: Possível ocorrência de Perdas Não Técnicas
```

Fonte: Elaborado pelo o autor.

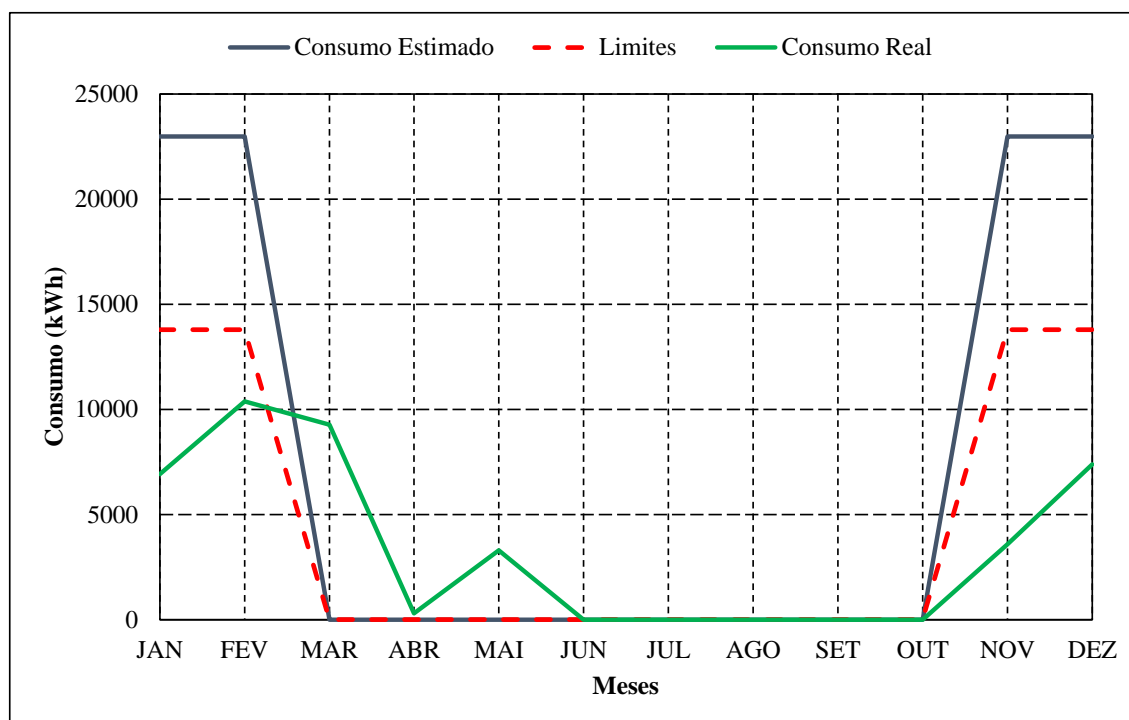
De acordo com o relatório apresentado pela Figura 4.27, a classe prevista para o ano de 2020 da consumidor rural UC-7, classificada pelo modelo preditivo, diverge do grupo de consumo definido pela Etapa de Agrupamento de Dados. A classe de perfil de consumo original pertencente a este cliente irrigante é o grupo “Perfil_2”, enquanto que a classe prevista determinada pelo modelo preditivo é a classe “Perfil_1”. Dessa forma, devido a mudança de classe de perfil de consumo, a unidade consumidora irrigante UC-6 é apontada pelo indicador como possibilidade de práticas de perdas comerciais.

A Figura 4.28 demonstra um comparativo na curva de carga mensal entre os consumos mensais reais e previsto e os respectivos limites de tolerância para o indicativo de ocorrência de perdas não técnicas para o consumidor irrigante UC-7. Já a Figura 4.29, apresenta o relatório final de simulação para o indicador de estimativa de consumo de energia elétrica realizada no IDE Spyder®.

Os resultados apresentados nas Figuras 4.28 e 4.29, indicam que os consumos registrados para a respectivo consumidor rural nos meses de Janeiro, Fevereiro, Novembro e Dezembro, foram inferiores aos limites mínimos mensais definidos para o indicativo de perdas não técnicas. Dessa forma, o consumidor rural avaliado é apontando como um atuador de práticas de perdas comerciais na avaliação do presente indicador.

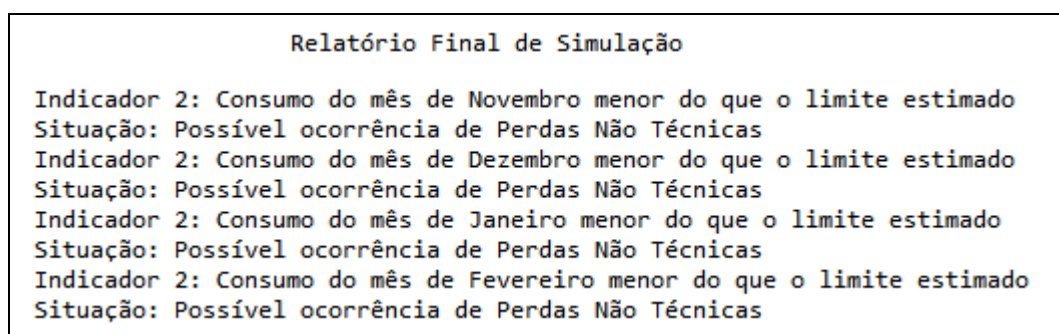
Ressaltando a Figura 4.28, a situação para o mês de Março é semelhante ao ocorrido nas análises dos consumidores UC-2, UC-3, UC-4 e UC-6, ou seja, o considerável consumo de energia elétrica nesse período pode estar relacionado com a atividade de irrigação tardia para as áreas de plantio de arroz desse consumidor. Para o mês de Maio, o consumo de

Figura 4.28 – Curvas de consumos mensais reais e estimados durante o ano de 2020 para a UC-7.



Fonte: Elaborado pelo o autor.

Figura 4.29 – Relatório final de simulação do indicador de estimativa de consumo de energia elétrica para a UC-7 no IDE Spyder ®.



Fonte: Elaborado pelo o autor.

energia elétrica pode estar associada com alguma outra atividade agroindustrial desenvolvida nessa unidade consumidora.

A Tabela 4.13 apresenta um resumo geral dos indicadores avaliados para a unidade consumidora irrigante UC-7.

Tabela 4.13 – Resumo geral dos indicadores avaliados para a UC-7.

Indicador: Classe de Perfil de Consumo para a UC-7				
Nomenclatura		Classe	Situação	
Original		Perfil_2	Possibilidade de ocorrência de Perdas Não Técnicas	
Previsto		Perfil_1		
Indicador: Estimativa de consumo de energia elétrica para a UC-7				
Mês	Real (kWh)	Estimado (kWh)	Limite (kWh)	Situação
JAN	6917	22983,56	13790,14	Possibilidade de ocorrência de Perdas Não Técnicas
FEV	10379	22983,56	13790,14	
NOV	3587	22983,56	13790,14	
DEZ	7385	22983,56	13790,14	

Fonte: Elaborado pelo o autor.

De acordo com dados apresentadas na Tabela 4.13, a unidade consumidora irrigante UC-7 apresentou irregularidades em ambos os indicadores. Desse forma, o respectivo consumidor é apontado pelo modelo preditivo como uma situação de possíveis práticas de perdas não técnicas.

5. CONCLUSÃO

A presente dissertação discutiu o problema das perdas não técnicas em sistemas de distribuição de energia elétrica, com ênfase na aplicação em unidades consumidoras rurais que utilizam sistemas de irrigação para lavouras de arroz localizadas no município de Uruguaiana, na Região da Fronteira Oeste do estado do Rio Grande do Sul. O estudo foi motivado pela relevância desse tipo de consumidor nos alimentadores da região, pelas dificuldades de inspeções e identificações de ações ilícitas em redes de distribuição rurais e pela pouca variedade de trabalhos na literatura que abordam esse problema.

O objetivo geral dessa dissertação foi desenvolver uma metodologia para identificar possíveis ocorrências de perdas não técnicas em unidades consumidoras rurais irrigantes utilizando métodos de agrupamento e classificação de dados, baseada na relação entre o consumo de energia elétrica e as características das lavouras de arroz irrigado e dos sistemas de irrigação.

Dessa forma, com o intuito de atender o objetivo proposto, foi elaborado uma metodologia empregando um algoritmo computacional que relacionou técnicas de inteligência artificial para agrupamento e predição de unidades consumidoras irrigantes presentes em uma base de dados, disponibilizados por uma distribuidora de energia elétrica. O agrupamento de dados foi realizado pelo método *k-Means*, e tinha por finalidade, agrupar as amostras do *dataset* em grupos (*clusters*), com base nas faixas de desvio padrão e média do consumo total de cada consumidor, sendo adotadas quatro classes de perfil de consumo de energia elétrica para todo o conjunto de dados. Já na parte de classificação de dados, foi adotado o método de *Random Forest*, sendo construído um modelo de previsão com a divisão das amostras de treino e teste do conjunto de consumidores irrigantes definidos pela etapa inicial de pré-processamento de dados. Além disso, o algoritmo realizou o cálculo do consumo estimado de energia elétrica das estações de bombeamento das unidades consumidoras rurais, utilizando dados previamente pesquisados na literatura e característicos dos sistemas de irrigação aplicados a lavouras que empregam o método por inundação contínua.

A escolha do método *k-Means* para aplicação na Etapa de Agrupamento de Dados, foi incentivada pela praticidade e facilidade da respectiva técnica na programação em linguagem *Python*, necessitando de poucos parâmetros de entrada no algoritmo. Já a preferência pelo método de *Random Forest* para a Etapa de Classificação de Dados, foi motivado pelas

respostas eficientes apresentadas para problemas de classificação e regressão, pelo rápido treinamento e fácil implementação em conjunto com outras técnicas de IA. O modelo preditivo proposto nesse estudo apresentou acurácia de 74% na classificação dos dados de teste e, 100 % para as amostras de treino, de acordo com os pontuadores estabelecidos como métricas de avaliação. Ambas as classificações, foram consideradas satisfatórias.

Para verificar a possibilidade de ocorrências de perdas não técnicas no conjunto de unidades consumidoras rurais, foram propostos dois indicadores de avaliação a etapa final da metodologia. O primeiro, consistiu na resposta apresentada para o grupo de perfil de consumo classificado pelo modelo preditivo, em comparação com a classe definida durante a Etapa de Agrupamento de Dados, apontando irregularidades, no caso de alteração para um perfil inferior de consumo do cliente irrigante. O segundo indicador, realizou uma confrontação entre os consumos estimados e registrados das unidades consumidores irrigantes para os meses de Janeiro, Fevereiro, Novembro e Dezembro de 2020, indicando a possibilidade de ocorrência de perdas comerciais, quando o consumo real for inferior a 60% do valor mensal estimado para o consumidor rural analisado.

Para a ilustração dos estudos de casos, sete unidades consumidoras irrigantes foram escolhidas do banco de dados, onde o algoritmo desenvolvido para o modelo preditivo foi aplicado individualmente em cada amostra selecionada. Os clientes foram selecionados baseados no conhecimento das áreas de plantio da lavoura de arroz irrigado associada a instalação rural e nos resultados apresentados nas estimativas de consumo de energia elétrica. O principal objetivo da seção de análises e resultados, foi apresentar os diferentes cenários possíveis dos indicadores propostos para o modelo preditivo nos consumidores irrigantes selecionados.

Avaliando os resultados atingidos para os estudos de casos, as unidades consumidoras irrigantes UC-1 e UC-2 não apresentaram irregularidades nos indicadores, sendo apontadas como consumidores regulares pelo modelo preditivo. Os demais consumidores rurais avaliados apresentaram irregularidades em pelo menos um dos indicadores propostos. Desse modo, ambos os clientes foram classificados pelo modelo de previsão como possíveis atuadores de práticas de perdas não técnicas. Para os casos dos consumidores irrigantes UC-3 e UC-5, foram identificadas irregularidades pelo segundo indicador, devido a estes apresentarem registros mensais de energia faturadas inferiores aos limites mínimos estabelecidos pela estimativa de consumo das instalações rurais. Para a unidade consumidora UC-4, a irregularidade foi apontada apenas pelo primeiro indicador, pois este cliente foi classificado pelo modelo preditivo em uma classe inferior ao grupo definido pelo processo de

agrupamento de dados. Os demais consumidores irrigantes, apresentaram irregularidades em ambos os indicadores avaliados.

A proposta para o cálculo da estimativa de consumo de energia elétrica das unidades consumidoras irrigantes proposta para esse trabalho, envolveu alguns atributos relacionados a cultivo da cultura do arroz irrigado, tais como: vazão unitária e total da lavoura; área de plantio; altura manométrica total; rendimento e tempo de funcionamento dos sistemas de irrigação. As estimativas mensais de consumo realizadas através desses parâmetros foram satisfatórias ou permaneceram dentro de uma padrão aceitável na maioria dos casos, sendo as divergências tendendo a valores inferiores aos limites aceitáveis, situações apontadas como possíveis ocorrências de perdas comerciais pelo modelo preditivo. As previsões de consumo de energia elétrica para todas as unidades consumidoras irrigantes, foram realizadas nos meses de Janeiro, Fevereiro, Novembro e Dezembro para o ano de 2020, devido as estes abrangerem os períodos de maiores consumos em virtude do funcionamento praticamente contínuo das estações de bombeamento para a irrigação das lavouras de arroz.

É importante ressaltar que foram utilizados dados característicos da Fronteira Oeste do estado do Rio Grande do Sul, com base em estudos relacionados na literatura e informações sobre o cultivo da cultura e áreas de plantio de arroz irrigado nessa região, e que para análises de lavouras localizadas em outras regiões orizícolas, devem ser realizadas adequações a estes parâmetros. Além disso, alguns cenários não foram considerados na modelagem da estimativa de consumo de energia elétrica das unidades consumidoras irrigantes, como o impacto das variáveis meteorológicas (precipitação pluviométrica, umidade relativa do ar, velocidade do vento e temperatura) e a influência das características do consumo de água da cultura do arroz irrigado (evapotranspiração de referência, evapotranspiração da cultura e balanço hídrico). Os mesmos, apresentam-se como análises propostas para trabalhos futuros.

Destaca-se, por fim, que os resultados gerados pela metodologia proposta nessa dissertação podem futuramente, auxiliar as distribuidoras de energia elétrica no processo de identificação de perdas não técnicas nos seus alimentadores rurais, através do mapeamento das áreas de risco e das unidades consumidoras suspeitas de irregularidades nos registros de consumos mensais de energia elétrica através dos indicadores propostas pelo modelo preditivo. Essas ações, podem proporcionar reduções nos custos com equipes de inspeção para investigar extensas áreas, situação típica no âmbito rural. Nesse sentido, a metodologia aqui desenvolvida pode ser considerada promissora, e com oportunidades de ser aprimorada, considerando as sugestões de trabalhos futuros apresentadas na próxima seção.

5.1 TRABALHOS FUTUROS

Com o objetivo de melhorar o desempenho da metodologia propostas, são sugeridos os seguintes tópicos como propostas para continuidade do trabalho:

- Aprimoramentos no processo de classificação de dados, buscando elevar a precisão do modelo;
- Testes de equilíbrio entre os números de unidades consumidoras irrigantes nas classes de perfil de consumo;
- Delimitação do processo de agrupamento de dados, definido faixas de valores para cada *cluster*;
- Aplicação do presente estudo em um período com mais de 5 anos de histórico de consumo de energia elétrica;
- Considerar o impacto das variáveis meteorológicas no consumo mensal de energia elétrica no conjunto de unidades consumidoras irrigantes, principalmente, a precipitação pluviométrica;
- Ampliar o conjunto de variáveis de entrada da etapa de estimativa de consumo dos sistemas de irrigação, considerando o estágio fenológico da planta durante a época da safra e as características de consumo de água da cultura do arroz irrigado;
- Expandir os testes de validação para outras regiões produtoras de arroz irrigado, buscando dados características das lavouras e das unidades consumidoras irrigantes presentes nessas regiões;
- Aplicação de técnicas de IA para aprimorar o cálculo da estimativa de consumo de energia elétricas das unidades consumidoras irrigantes.
- Considerar os dados históricos de inspeções nas unidades consumidoras irrigantes, como parte do desenvolvimento da metodologia para identificação de possíveis ocorrências de perdas não técnicas.

5.2 PUBLICAÇÕES

Durante a elaboração desta dissertação foram desenvolvidos os seguintes trabalhos científicos que se correlacionavam com o tema base proposto por esse estudo:

- EICHKOFF, H. S.; BERNARDON, D. P.; SOUZA, N. B.; MARCOLIN, P.; MADALUZ, J.; KÖPP, L. M.; CHIARA, L. M.; SILVA, J. A. **Classificação de**

Unidades Consumidoras Irrigantes de Arroz para Análise de Perdas Não Técnicas utilizando o Método de *Random Forest*. 14th Seminar on Power Electronics and Control (SEPOC 2021). **Anais...**Santa Maria, Brasil, 2021.

- MARCOLIN, P.; BERNARDON, D. P.; EICHKOFF, H. S.; SOUZA, N. B.; MADALAZ, J.; KÖPP, L. M.; CHIARA, L. M.; SILVA, J. A. **Metodologia para Detecção de Perdas Não Técnicas de Unidades Consumidoras Irrigantes de Arroz.** XIV Conferência Brasileira sobre Qualidade de Energia Elétrica (CBQEE 2021). **Anais...**Foz do Iguaçu, Brasil, 2021.
- SOUZA, N. B.; BERNARDON, D. P.; EICHKOFF, H. S.; MARCOLIN, P.; MADALAZ, J.; CHIARA, L. M.; SILVA, J. A. **Agrupamento e Classificação de Consumidores de Energia Rural utilizando *Random Forest* e *K-Nearest Neighbors*.** XXIII Congresso Brasileiro de Automática (CBA 2020). **Anais...**Porto Alegre, Brasil, 2020.

REFERÊNCIAS

- AGÜERO, J. R. **Improving the efficiency of power distribution systems through technical and non-technical losses reduction**. 2012 IEEE/PES Transmission and Distribution Conference and Exposition (T&D). **Anais...**Orlando, Estados Unidos: IEEE, 2012.
- AHMAD, T. Non-technical loss analysis and prevention using smart meters. **Renewable and Sustainable Energy Reviews**, v. 72, p. 573–589, 2017.
- ALLEN, R. G. et al. Evapotranspiración del cultivo: Guías para la determinación de los requerimientos de agua de los cultivos. **Estudio Fao Riego y Drenaje 56**, p. 332, 2006.
- ALONÇO, A. DOS S. et al. Sistema de Cultivo de Arroz Irrigado no Brasil. **EMBRAPA - Sistemas de Produção 03**, n. 1, p. 270, 2004.
- ANA. **Manual dos Usos Consuntivos de Água do Brasil**. Brasília, Brasil: Agência Nacional de Águas, 2019. 75 p.
- ANEEL. **Nota Técnica nº 298/2011 - SRE: Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica para o terceiro ciclo de revisão tarifária periódica das concessionárias de distribuição de energia elétrica**. Brasília, 2011. 79 p.
- ANEEL. **Perdas de Energia**. Brasília, 2015. Disponível em: <https://www.aneel.gov.br/metodologia-distribuicao/asset_publisher/e2INtBH4EC4e/content/perdas/654800>. Acesso em: 19 mai. 2020.
- ANEEL. **Tarifas Consumidores**. Brasília, 2016. Disponível em: <<https://www.aneel.gov.br/tarifasconsumidores/>>. Acesso em: 14 set. 2020.
- ANEEL. **Perdas de Energia Elétrica na Distribuição**. Brasília, Brasil: Agência Nacional de Energia Elétrica, Ed. 1, 2019. 21 p.
- ANEEL. **Perdas de Energia Elétrica na Distribuição**. Brasília, Brasil: Agência Nacional de Energia Elétrica, Ed. 1, 2020. 21p.
- ARAUJO, B. S. DE; ALMEIDA, H. L. S. DE; MELLO, F. L. DE. Computational intelligence methods applied to the fraud detection of electric energy consumers. **IEEE Latin America Transactions**, v. 17, n. 1, p. 71–77, 2019.
- ASEVEDO, F. R. DE. **Abordagem Linear Generalizada para Estimar Perdas Não**

Técnicas de Energia Elétrica. 2011. 76 p. Dissertação (Mestrado em Engenharia Elétrica)-Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2011.

BECKMANN, M.; EBECKEN, N. F. F.; PIRES DE LIMA, B. S. L. A KNN Undersampling Approach for Data Balancing. **Journal of Intelligent Learning Systems and Applications**, v. 7, n. 4, p. 104–116, 2015.

BELTRAME, W. A. R.; FONSECA, F. C. S. **Aplicações Práticas dos Algoritmos de Clusterização K-means e Bisecting K-means.** Vitória, 2010. Disponível em: <<https://www.researchgate.net/publication/327121358>>. Acesso em: 13 mai. 2021.

BEZERRA, U. H. et al. **Non-Technical Losses Estimation in Distribution Feeders Using the Energy Consumption Bill and the Load Flow Power Summation Method.** 2016 IEEE International Energy Conference (ENERGYCON 2016). **Anais...**Leuven, Bélgica: IEEE, 2016.

BRASIL, Decreto Nº 4.562, de 31 de Dezembro de 2002. Estabelece normas gerais para celebração, substituição e aditamento dos contratos de fornecimento de energia elétrica; para tarifação e preço de energia elétrica; dispõe sobre compra de energia elétrica das concessionárias de serviço público de distribuição; valores normativos; estabelece a redução do número de submercados; diretrizes para revisão da metodologia de cálculo das Tarifas de Uso do Sistema de Transmissão - TUST e dá outras providências. **Diário Oficial da União**, Poder Executivo, Brasília, DF, 31 dez. 2002. Parágrafo 1.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5–32, 2001.

BUZAU, M. M. et al. Detection of non-technical losses using smart meter data and supervised learning. **IEEE Transactions on Smart Grid**, v. 10, n. 3, p. 2661–2670, 2019.

CARNEIRO, F. N. **Assessoria agrônômica em lavouras de arroz irrigado e de soja no município de Bagé.** 2017. 32 p. Trabalho de Conclusão de Curso (Graduação em Agronomia)-Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2017.

CASSIANO, K. M. **Análise de Séries Temporais usando Análise Espectral Singular (SSA) e Clusterização de suas Componentes baseada em Densidade.** 2014. 172 f. Tese (Doutorado em Engenharia Elétrica)-Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2014.

CASTRO, N. **Apostila de Irrigação (IPH 02 207):** apostila. Porto Alegre: Universidade Federal do Rio Grande do Sul - Instituto de Pesquisas Hidráulicas, 2003. 56 p.

CHATTERJEE, S. et al. **Detection of Non-Technical Losses using Advanced Metering Infrastructure and Deep Recurrent Neural Networks**. 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe). **Anais...**Milão, Itália: IEEE, 2017.

CHEN, D. et al. **Métricas de Avaliação em Machine Learning: Classificação**. 2020. Disponível em: <<https://medium.com/kunumi/metricas-de-avaliacao-em-machine-learning-classificacao-49340dcdb198>>. Acesso em: 25 mai. 2021.

CONAB. **Acompanhamento da Safra Brasileira (Grãos: Safra 2020/21 - 7º Levantamento)**. Brasília, Brasil: Companhia Nacional de Abastecimento, v. 8, n. 7, 2021a. p. 89.

CONAB. **Acompanhamento da Safra Brasileira**. Brasília, 2021b. Disponível em: <<https://www.conab.gov.br/info-agro/safras>>. Acesso em: 9 abr. 2021.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random Forests. **Machine Learning**, v. 45, n. 1, p. 157–176, 2011.

DEPURU, S. S. S. R. **Modeling, Detection, and Prevention of Electricity Theft for Enhanced Performance and Security of Power Grid**. 2012. 141 f. Tese (Doutorado em Filosofia Licenciada em Engenharia)-University of Toledo, Toledo, Estados Unidos, 2012.

DUTRA, A. D. et al. Arroz Irrigado por Aspersão no Rio Grande do Sul. **EMBRAPA - Sistemas de Produção** 24, n. 1, p. 142, 2017.

EMBRAPA. **Métodos de irrigação**. Brasília, 2020a. Disponível em: <<https://www.agencia.cnptia.embrapa.br/gestor/arroz/arvore/CONT000foh49q3602wyiv8065610d5y5f5im.html>>. Acesso em: 10 ago. 2020.

EMBRAPA. **Sistema de cultivo**. Brasília, 2020b. Disponível em: <<https://www.agencia.cnptia.embrapa.br/gestor/arroz/arvore/CONT000g1wcnzza02wx5ok0ha2lipwbeel46.html>>. Acesso em: 10 ago. 2020.

EMBRAPA. **Irrigação e drenagem**. Brasília, 2020c. Disponível em: <<https://www.agencia.cnptia.embrapa.br/gestor/arroz/arvore/CONT000fojvokod02wyiv80bhgp5p9cxb3rg.html>>. Acesso em: 7 dez. 2020.

EVALDT, M. C. **Uma Metodologia para a Identificação de Perdas Não Técnicas em Grandes Consumidores Rurais**. 2014. 118 p. Dissertação (Mestrado em Engenharia Elétrica)-Universidade Federal De Santa Maria, Santa Maria, RS, 2014.

EVALDT, M. C. **Sistema Neural Artificial para Identificação de Perdas Não Técnicas em Consumidores Rurais**. 2018. 150 f. Tese (Doutorado em Engenharia Elétrica)-Universidade Federal de Santa Maria, Santa Maria, RS, 2018.

FERNANDES, S. E. N. et al. A Probabilistic Optimum-Path Forest Classifier for Non-Technical Losses Detection. **IEEE Transactions on Smart Grid**, v. 10, n. 3, p. 3226–3235, 2019.

FERREIRA, L. D. **Técnicas de aprendizado de máquina aplicadas à classificação de estudantes a partir de estilos de aprendizagem**. 2018. 121 p. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional)-Universidade de São Paulo, São Paulo, SP, 2018.

FERREIRA NETO, P. S. **Estudo da cadeia produtiva do arroz em São Lourenço do Sul-RS**. 2011. 34 p. Trabalho de Conclusão de Curso (Graduação em Ciências Econômicas)-Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2011.

FONTANA, A.; NALDI, M. C. **Estudo e Comparação de Métodos para Estimação de Números de Grupos em Problemas de Agrupamento de Dados**. São Paulo, 2009. Disponível em:< <http://repositorio.icmc.usp.br/handle/RIICMC/6697>>. Acesso em 20 mai. 2021.

GHORI, K. M. et al. Performance Analysis of Different Types of Machine Learning Classifiers for Non-Technical Loss Detection. **IEEE Access**, v. 8, p. 16033–16048, 2019.

GHORI, K. M. et al. **Impact of Feature Selection on Non-Technical Loss Detection**. 2020 6th Conference on Data Science and Machine Learning Applications (CDMA 2020). **Anais...Riad**, Arábia Saudita: IEEE, 2020.

GOMES, H. P. et al. **Manual de Sistemas de Bombeamento: Eficiência Energética**. João Pessoa, Brasil: Editora Universitária - UFPB, Ed. 1, 2012. 189 p.

GUEDES, A. R. M.; GUIMARÃES, V. L. **Sistema de Reconhecimento Baseado em Random Forest para Caracteres de Captchas**. Disponível em:<<http://www.decom.ufop.br/menotti/rp142/trab/trab1-dp2-artigo.pdf>>. Acesso em: 20 mai. 2021.

GUERRERO, J. I. et al. Non-Technical Losses Reduction by Improving the Inspections Accuracy in a Power Utility. **IEEE Transactions on Power Systems**, v. 33, n. 2, p. 1209–1218, 2018.

HAMMERSCHMITT, B. K. et al. **Non-Technical Losses Review and Possible Methodology Solutions**. 2020 6th International Conference on Electric Power and Energy Conversion Systems (EPECS 2020). **Anais...Istambul, Turquia: IEEE, 2020**

HAN, J.; KAMBER, M.; PEI, J. **Data mining: Data mining concepts and techniques**. Waltham, Estados Unidos: Morgan Kaufmann, Ed. 2., 2014. 703 p.

HUANG, S. C.; LO, Y. L.; LU, C. N. Non-technical loss detection using state estimation and analysis of variance. **IEEE Transactions on Power Systems**, v. 28, n. 3, p. 2959–2966, 2013.

HUBACK, V. B. DA S. **Medidas ao Combate a Perdas Elétricas Não Técnicas em Áreas com Severas Restrições à Operação de Sistemas de Distribuição de Energia Elétrica**. 2018. 180 p. Dissertação (Mestrado em Planejamento Energético)-Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 2018.

IRGA. **Conheça o Irga**. Porto Alegre, 2020a. Disponível em: <<https://irga.rs.gov.br/quem-somos>>. Acesso em: 19 ago. 2020.

IRGA. **Imagens de satélite confirmam dados do Irga para safra 2019/2020**. Porto Alegre, 2020b. Disponível em: <<https://irga.rs.gov.br/imagens-de-satelite-confirmam-dados-do-irga-para-safra-2019-2020>>. Acesso em: 20 ago. 2020.

IRGA. **Evolução da Semeadura - Safra 2020/21**. Porto Alegre, 2020c. Disponível em: <<https://irga.rs.gov.br/upload/arquivos/202012/28110225-evolucao-da-semeadura-2021.pdf>>. Acesso em: 11 abr. 2021.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**, v. 31, n. 3, 1999.

KLERING, E. V. **Mapeamento da área cultivada e estimativa do rendimento de grãos de arroz irrigado a partir de modelos agrometeorológicos-espectrais para o Rio Grande do Sul**. 2012. 163 f. Tese (Doutorado em Fitotecnia)-Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2012.

KÖPP, L. M. **Índices De Desempenho Para Estações De Bombeamento Em Lavouras De Arroz Irrigado**. 2015. 107 f. Tese (Doutorado em Engenharia Agrícola)-Universidade Federal de Santa Maria, Santa Maria, RS, 2015.

KÖPP, L. M. et al. Caracterização de Estações de Bombeamento em Lavouras de Arroz do Rio Grande do Sul. **Journal of the Brazilian Association of Agricultural Engineering**, v.

36, n. 2, p. 342–351, 2016.

KSB. **Manual de Curvas Características de Bombas centrífugas com corpo espiral dividido radialmente (KSB Meganorm/Megabloc/Megachem/Megachem V)**. Frankenthal, Alemanha: KSB, 2013. 28 p.

KUHN, J. P. B. **Produção de Arroz Irrigado no Município de Uruguaiana-RS**. 2017. 32 p. Trabalho de Conclusão de Curso (Graduação em Agronomia)-Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2017.

LACHI, R. L.; ROCHA, H. V. DA. **Aspectos básicos de clustering: conceitos e técnicos**. Campinas, Brasil: Instituto de Computação - Universidade Estadual de Campinas, 2005. 26 p.

LEITE, J. B.; MANTOVANI, J. R. S. Detecting and locating non-technical losses in modern distribution networks. **IEEE Transactions on Smart Grid**, v. 9, n. 2, p. 1023–1032, 2018.

LORENSI, R. P. **Relações entre atributos do solo e uso de água em arroz irrigado**. 2011. 126 p. Dissertação (Mestrado em Engenharia Agrícola)-Universidade Federal de Santa Maria, Santa Maria, RS, 2011.

LOUREIRO, I. B. **Laboratório Remoto para Ensino de Eficiência Energética em um Sistema de Bombeamento de Água**. 2009. 121 p. Dissertação (Mestrado em Engenharia Elétrica)-Universidade Federal do Pará, Belém, PA, 2009.

LUYA, L. E.; PEDRASA, M. A. **Detecting and Estimating Amount of Energy Theft in the Distribution Network Using Linear Regression**. 2019 9th International Conference on Power and Energy Systems (ICPES 2019). **Anais...Perth**, Austrália: IEEE, 2019.

MADRIGAL, M.; RICO, J. J.; UZCATEGUI, L. Estimation of Non-Technical Energy Losses in Electrical Distribution Systems. **IEEE Latin America Transactions**, v. 15, n. 8, p. 1447–1452, 2017.

MARCOLAN, A. L. et al. Sistema de produção de arroz de terras altas. **EMBRAPA - Sistemas de Produção** 31, n. 4, p. 38, 2008.

MAROTTA, V. **Aprendizado não supervisionado com o kMeans**. Viçosa, 2019. Disponível em: <<https://www.researchgate.net/publication/334574592>>. Acesso em: 13 mai. 2021.

MELLO, C. R. DE; CARVALHO, J. DE A. Análise da Equação de Perda de Carga de Hazen-Williams, associada aos regimes hidráulicos para tubos de PVC e Polietileno de pequeno

diâmetro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 2, n. 3, p. 247–252, 1998.

MESSINIS, G. M.; HATZIARGYRIOU, N. D. Review of non-technical loss detection methods. **Electric Power Systems Research**, v. 158, p. 250–266, 2018a.

MESSINIS, G. M.; HATZIARGYRIOU, N. D. **Unsupervised classification for non-technical loss detection**. 2018 Power Systems Computation Conference (PSCC 2018). **Anais...**Dublin, Irlanda: IEEE, 2018b.

MONTEIRO, M. D.; MACIEL, R. S. **Detection of commercial losses in electric power distribution systems using data mining techniques**. 2018 Simpósio Brasileiro de Sistemas Elétricos (SBSE 2018). **Anais...**Niterói, Brasil: IEEE, 2018.

MUNDSTOCK, C. M. et al. **Manual de boas praticas agrícolas: guia para a sustentabilidade da lavoura de arroz irrigado no Rio Grande do Sul**. Cachoeirinha, Brasil: Instituto Riograndense do Arroz, 2011. 80 p.

NUNES, M. G. **A prática da atividade orizícola pela agricultura familiar no município de Santo Antônio da Patrulha**. 2011. 57 p. Trabalho de Conclusão de Curso (Graduação em Ciências Econômicas)-Universidade Federal do Rio Grande do Sul, Santo Antônio da Patrulha, RS, 2011.

OLIVEIRA, M. E. **Avaliação de metodologias de cálculo de perdas técnicas em sistemas de distribuição de energia elétrica**. 2009. 137 f. Tese (Doutorado em Engenharia Elétrica)-Univerisdade Estadual Paulista, Ilha Solteira, RS, 2009.

OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. 101 p. Dissertação (Mestrado em Bioinformática)-Universidade de São Paulo, Ribeirão Preto, SP, 2013.

PASSOS JÚNIOR, L. A. et al. Unsupervised non-technical losses identification through optimum-path forest. **Electric Power Systems Research**, v. 140, p. 413–423, 2016.

PENIN, C. A. D. S. **Combate , Prevenção e Otimização das Perdas Comerciais de Energia Elétrica**. 2008. 227 f. Tese (Doutorado em Engenharia Elétrica)-Universidade de São Paulo, São Paulo, SP, 2008.

PEREIRA, L. A. M. et al. **Multilayer perceptron neural networks training through charged system search and its Application for non-technical losses detection**. 2013 IEEE

PES Conference on Innovative Smart Grid Technologies (ISGT Latin America 2013). **Anais...**São Paulo, Brasil: IEEE, 2013.

PEREIRA, V. S. **Arroz, porque irrigar: conversa com o especialista**. 2019. Disponível em: <<https://www.austertecnologia.com/single-post/arroz-irrigado>>. Acesso em: 15 dez. 2020.

PERSSON, M.; LINDSKOG, A. **Detection and localization of non-technical losses in distribution systems with future smart meters**. 2019 IEEE Milan PowerTech (PowerTech 2019). **Anais...**Milão, Itália: IEEE, 2019.

PFITSCHER, L. L. et al. **Automatic control of irrigation systems aiming at high energy efficiency in rice crops**. 2012 8th International Caribbean Conference on Devices, Circuits and Systems (ICCDSCS 2012). **Anais...**Playa del Carmen, México: IEEE, 2012.

RAGGI, L. M. R. et al. Non-Technical Loss Identification by Using Data Analytics and Customer Smart Meters. **IEEE Transactions on Power Delivery**, v. 35, n. 6, p. 2700–2710, 2020.

REIS, C. Z. **Eficácia de Solução Tecnológica para Redução de Furtos de Energia Elétrica em Empresas Distribuidoras: Estudo de Caso**. 2005. 81 f. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2005.

RIBEIRO, S. S. et al. **Investment optimization methodology applied to investments on non-technical losses reduction actions**. 2012 IEEE Symposium on Computers and Communications (ISCC). **Anais...**Cappadocia, Turquia: IEEE, 2012.

RIGHES, A. A. **Eficiência em Sistemas com Inundação: O caso do arroz**. Santa Maria, 2013. Disponível em: <<http://cbhpf.upf.br/phocadownload/2seminario/sistemaidundacaoii.pdf>>. Acesso em: 8 dez. 2020.

SAMMUT, C.; WEBB, G. I. **Encyclopedia of Machine Learning**. Springer, Estados Unidos: Springer, Ed. 1, 2010. 1031 p.

SARAVANAN, R.; SUJATHA, P. **Algorithms : A Perspective of Supervised Learning Approaches in Data Classification**. 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018). **Anais...**Madurai, Índia: IEEE, 2018.

SCHIAFFINO, M. C. **Desenvolvimento de um Método para Classificação de Comportamentos de ratos wistar utilizando o Algoritmo de Aprendizado Supervisionado Florestas Aleatórias (Random Forest)**. 2020. 81 p. Dissertação (Mestrado em Neurociências)-Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2020.

SCHIKOWSKI, A. N. A. B. **Estimativa do Volume e da forma do Fuste Utilizando Técnicas de Aprendizado de Máquina**. 2016. 71 p. Dissertação (Mestrado em Engenharia Florestal)-Universidade Federal do Paraná, Curitiba, PR, 2016.

SCIVITTARO, W. B.; PARFITT, J. M. B. Arroz Irrigado por Aspersão no Rio Grande do Sul. **EMBRAPA - Sistemas de Produção** 24, n. 1, p. 17, 2011.

SHARMA, S.; MAJUMDAR, A. Unsupervised Detection of Non-Technical Losses via Recursive Transform Learning. **IEEE Transactions on Power Delivery**, v. 36, n. 2, p. 1241–1244, 2020.

SILVA, J. A. **Método para a Detecção de Perdas Não Técnicas usando Lógica Nebulosa como principal Agente Preditivo**. 2013. 83 p. Dissertação (Mestrado em Engenharia Elétrica)-Universidade Estadual de Campinas, 2013.

SOARES, H. **Capítulo IV: Máquinas Hidráulicas**. Juíz de Fora, 2012. Disponível em: <https://www.ufjf.br/engsanitariaeambiental/files/2012/09/HGHTEO__Cap4_Aula-1_-10012013_V2.pdf>. Acesso em: 7 out. 2020.

SOSBAI. **Arroz Irrigado: Recomendações técnicas da pesquisa para o Sul do Brasil**. Cachoeirinha, Brasil: Sociedade Sul-Brasileira de Arroz Irrigado (SOSBAI), 2018. 205 p.

SOUZA, N. M. **Tolerância a baixas temperaturas na fase de microsporogênese em genótipos de arroz irrigado**. 2015. 93 p. Dissertação (Mestrado em Produção Vegetal)-Universidade do Estado de Santa Catarina, Lages, SC, 2015.

STONE, L. F. Eficiência do Uso da Água na Cultura do Arroz Irrigado. **EMBRAPA - Documentos** 176, n. 1, p. 48, 2005.

TABBAL, D. F. et al. On-farm strategies for reducing water input in irrigated rice; case studies in the Philippines. **Agricultural Water Management**, v. 56, n. 2, p. 93–112, 2002.

TEHRANI, S. O.; MOGHADDAM, M. H. Y.; ASADI, M. **Decision Tree based Electricity Theft Detection in Smart Grid**. 2020 4th International Conference on Smart City, Internet of Things and Applications (SCIoT 2020). **Anais...Mashhad, Irã: IEEE**, 2020.

TERCIYANLI, E.; EMRE, T.; CALISKAN, S. **Score based non-technical loss detection algorithm for electricity distribution networks**. 2017 5th International Istanbul Smart Grid and Cities Congress and Fair (ICSG 2017). **Anais...Istambul, Turquia: IEEE**, 2017.

TOMA, R. N. et al. **Electricity Theft Detection to Reduce Non-Technical Loss using**

Support Vector Machine in Smart Grid. 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT 2019). **Anais...**Dhaka, Bangladesh: IEEE, 2019.

TRAPP, S. **Produção de arroz irrigado e soja em Santa Vitória do Palmar – RS.** 2018. 40 p. Trabalho de Conclusão de Curso (Graduação em Agronomia)-Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2018.

UBERTI, V. A. **Lógica Fuzzy para Avaliação de Eficiência Energética em Sistemas de Irrigação de Lavouras de Arroz.** 2017. 172 p. Dissertação (Mestrado em Engenharia Elétrica)-Universidade Federal de Santa Maria, Santa Maria, RS, 2017.

UMAR, H. A.; PRASAD, R.; FONKAM, M. **Assessing severity of non-technical losses in power using clustering Algorithms.** 2019 15th International Conference on Electronics, Computer and Computation (ICECCO 2019). **Anais...**Abuja, Nigéria: IEEE, 2019.

WANG, Z. et al. **Analysis of 10kV Non-technical Loss Detection with Data-driven Approaches.** 2019 IEEE Innovative Smart Grid Technologies Asia (ISGT Asia 2019). **Anais...**Chengdu, China: IEEE, 2019.

ZHANG, T.; GAO, R.; SUN, S. **Theories, Applications and Trends of Non-Technical Losses in Power Utilities Using Machine Learning.** 2018 2nd IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference (IMCEC 2018). **Anais...**Xiam, China: IEEE, 2018.

ZHAO, Y.; MA, X. **Study on credit evaluation of electricity users based on random forest.** 2017 Chinese Automation Congress (CAC 2017). **Anais...**Jinan, China: IEEE, 2017.

APÊNDICE A – CÓDIGO DO MODELO PROPOSTO PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM UNIDADES CONSUMIDORAS IRRIGANTES

Nesse Apêndice são apresentados os códigos de programação em linguagem *Python* 3.7 para a implementação da metodologia proposta para identificação de possíveis ocorrências de perdas não técnicas em unidades consumidoras irrigantes. Os scripts serão descritos de maneira individual nos itens A.1 a A.7, para cada estudo de caso (consumidor irrigante) avaliado.

As Figuras A.1 e A.2 ilustram respectivamente, a programação da inicialização das bibliotecas da linguagem *Python* e o pré-processamento e leitura do conjunto de dados de unidades consumidoras irrigantes do município de Uruguai/RS. Já às Figuras A.3 e A.4, apresentam os códigos referentes as etapas de agrupamento e classificação de dados da metodologia proposta para esse estudo.

Figura A.1 – *Script* para inicialização das Bibliotecas *Python* do algoritmo proposto para a metodologia.

```

3 ## Inicilização Bibliotecas Python
4 import matplotlib.pyplot as plt
5 import numpy as np
6 from pandas import ExcelWriter
7 from shapely.geometry import Point, Polygon
8 import itertools
9 import numpy as np
10 import matplotlib.pyplot as plt
11 from matplotlib.ticker import NullFormatter
12 import pandas as pd
13 import numpy as np
14 import matplotlib.ticker as ticker
15 from sklearn import preprocessing
16 import seaborn as sns
17 from geopy.geocoders import Nominatim
18 import folium
19 import pandas as pd
20 import shapefile as shp
21 from shapely.ops import cascaded_union
22 from sklearn.preprocessing import MinMaxScaler
23 import statistics
24 from statistics import mean
25 import geopandas

```

Figura A.2 – Código para a leitura do conjunto de dados das unidades consumidoras irrigantes.

```

27 ## Leitura do Conjuntos de Dados das UCs
28 data_1 = pd.read_excel('105_CONSULTA_IRRIG_NOTURNA_020221.xlsx')
29 data_1_irrigante = data_1[data_1['IND_IRRIGANTE'] == 'S']
30 data_1_irrigante_uruguaiana = data_1_irrigante[data_1_irrigante['DESC_MUNICIPIO'] == 'URUGUAIANA']
31 data_1_COD_INSTALACAO = data_1[['COD_INSTALACAO']]
32 data_1A = pd.read_csv('110_GRUPO_A_CONSUMOS_CONSULTA_IRRIG_NOTURNA_COMPLETO_090221.txt', delimiter="|")
33 data_1B = pd.read_csv('109_GRUPO_B_CONSUMOS_CONSULTA_IRRIG_NOTURNA_COMPLETO_090221.txt', delimiter="|")
34 data_1A_2 = data_1A[['COD_INSTALACAO', 'DT_REFERENCIA', 'CONSUMO']]
35 data_1A_3 = data_1A.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['CONSUMO'].sum().unstack(-1)
36 data_1B_2 = data_1B[['COD_INSTALACAO', 'DT_REFERENCIA', 'QTD_ENERGIA_FATURADA']]
37 data_1B_3 = data_1B.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['QTD_ENERGIA_FATURADA'].sum().unstack(-1)
38 data_A = data_1A_3.reset_index()
39 data_B = data_1B_3.reset_index()
40 data_A_1 = pd.merge(data_1, data_A, on='COD_INSTALACAO', how='inner')
41 data_B_1 = pd.merge(data_1, data_B, on='COD_INSTALACAO', how='inner')
42 data_A_1_municipios = data_A_1['DESC_MUNICIPIO'].value_counts()
43 data_B_1_municipios = data_B_1['DESC_MUNICIPIO'].value_counts()
44 data_A_1_COD_INSTALACAO = data_A_1['COD_INSTALACAO'].value_counts()
45 data_B_1_COD_INSTALACAO = data_B_1['COD_INSTALACAO'].value_counts()
46
47 data_A_2 = data_A_1[['COD_INSTALACAO', 'COD_GRUPO_TENSAO', 'DESC_MUNICIPIO', 'NUM_COORDENADA_GIS_X', 'NUM_COORDENADA_GIS_Y',
48                    201701, 201702, 201703, 201704, 201705, 201706, 201707, 201708, 201709, 201710,
49                    201711, 201712, 201801, 201802, 201803, 201804, 201805, 201806, 201807, 201808,
50                    201809, 201810, 201811, 201812, 201901, 201902, 201903, 201904, 201905, 201906,
51                    201907, 201908, 201909, 201910, 201911, 201912, 202001, 202002, 202003, 202004,
52                    202005, 202006, 202007, 202008, 202009, 202010, 202011, 202012]]
53
54 data_A_2.columns = ['COD_INSTALACAO', 'COD_GRUPO_TENSAO', 'DESC_MUNICIPIO', 'NUM_COORDENADA_GIS_X', 'NUM_COORDENADA_GIS_Y',
55                    '201701', '201702', '201703', '201704', '201705', '201706', '201707', '201708', '201709', '201710',
56                    '201711', '201712', '201801', '201802', '201803', '201804', '201805', '201806', '201807', '201808',
57                    '201809', '201810', '201811', '201812', '201901', '201902', '201903', '201904', '201905', '201906',
58                    '201907', '201908', '201909', '201910', '201911', '201912', '202001', '202002', '202003', '202004',
59                    '202005', '202006', '202007', '202008', '202009', '202010', '202011', '202012']
60
61 data_A_2_2017 = data_A_2[['201701', '201702', '201703', '201704', '201705', '201706', '201707', '201708', '201709', '201710',
62                          '201711', '201712']]
63 data_A_2['mean_2017'] = data_A_2_2017.mean(axis=1)
64 data_A_2_2018 = data_A_2[['201801', '201802', '201803', '201804', '201805', '201806', '201807', '201808',
65                          '201809', '201810', '201811', '201812']]
66 data_A_2['mean_2018'] = data_A_2_2018.mean(axis=1)
67 data_A_2_2019 = data_A_2[['201901', '201902', '201903', '201904', '201905', '201906',
68                          '201907', '201908', '201909', '201910', '201911', '201912']]
69 data_A_2['mean_2019'] = data_A_2_2019.mean(axis=1)
70 data_A_2_2020 = data_A_2[['202001', '202002', '202003', '202004', '202005', '202006',
71                          '202007', '202008', '202009', '202010', '202011', '202012']]
72 data_A_2['mean_2020'] = data_A_2_2020.mean(axis=1)
73 data_A_2_MEDIA = data_A_2[['201701', '201702', '201703', '201704', '201705', '201706', '201707', '201708', '201709', '201710',
74                          '201711', '201712', '201801', '201802', '201803', '201804', '201805', '201806', '201807', '201808',
75                          '201809', '201810', '201811', '201812', '201901', '201902', '201903', '201904', '201905', '201906',
76                          '201907', '201908', '201909', '201910', '201911', '201912']]
77 data_A_2['MEDIA'] = data_A_2_MEDIA.mean(axis=1)
78 data_A_2['STD_VALUE'] = data_A_2_MEDIA.std(axis = 1, skipna = True)
79 data_A_2['STD_2019'] = data_A_2_2019.std(axis = 1, skipna = True)
80 data_A_2 = data_A_2.dropna(subset=['MEDIA'])
81 data_A_2 = data_A_2.dropna(subset=['STD_VALUE'])
82 data_A_2 = data_A_2.dropna(subset=['STD_2019'])
83 data_A_2 = data_A_2.drop_duplicates(['COD_INSTALACAO'], keep='last')
84 data_A_2['NUM_COORDENADA_GIS_X'] = data_A_2['NUM_COORDENADA_GIS_X'] / 1000000
85 data_A_2['NUM_COORDENADA_GIS_Y'] = data_A_2['NUM_COORDENADA_GIS_Y'] / 1000000
86 data_A_2 = data_A_2[data_A_2['NUM_COORDENADA_GIS_Y'] != 0]
87 data_A_2 = data_A_2[data_A_2['NUM_COORDENADA_GIS_X'] != 0]

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.3 – Código para a etapa de agrupamento de dados.

```

89 ##Etapa de Agrupamento de Dados
90
91 from sklearn.cluster import KMeans
92 from sklearn.decomposition import PCA
93 from sklearn.preprocessing import scale
94 from sklearn.cluster import SpectralClustering
95 from numpy import unique
96 from numpy import where
97 from matplotlib import pyplot
98 from sklearn.metrics import silhouette_samples, silhouette_score
99 from kmeans_pytorch import kmeans, kmeans_predict
100 from imblearn.under_sampling import RandomUnderSampler
101 from k_means_constrained import KMeansConstrained
102 from pylab import plot, show, title, xlabel, ylabel
103 from pylab import legend
104 from pylab import axis
105
106 #### Definição das Classes de Perfil de Consumo
107 X = dfArroz[['mean_2019', 'STD_2019']]
108 nc = 4
109 smin = int(len(dfArroz)/nc)
110 smax = smin+1
111 model = KMeansConstrained(n_clusters=nc, size_min=smin, size_max=smax, random_state=42)
112 model.fit_predict(X)
113 yhat = model.labels_
114 X['c'] = yhat
115 dfArroz['c'] = yhat
116 dfArroz['Perfil de Consumo'] = yhat
117 Perfil = dfArroz['Perfil de Consumo'].value_counts()
118 dfArroz['Perfil_CONSUMO'] = dfArroz['c']
119 perfis = {0: 'Perfil_1', 1: 'Perfil_2', 2: 'Perfil_3', 3: 'Perfil_4'}
120 dfArroz.Perfil_CONSUMO = dfArroz.Perfil_CONSUMO.replace(perfis)
121 X_class_0 = dfArroz[dfArroz['c'] == 0] #maior
122 X_class_1 = dfArroz[dfArroz['c'] == 1] #minority class
123 X_class_2 = dfArroz[dfArroz['c'] == 2] #terceira maior
124 X_class_3 = dfArroz[dfArroz['c'] == 3] #Penultima
125 Describe0 = X_class_0[['mean_2019', 'STD_2019']].describe()
126 Describe1 = X_class_1[['mean_2019', 'STD_2019']].describe()
127 Describe2 = X_class_2[['mean_2019', 'STD_2019']].describe()
128 Describe3 = X_class_3[['mean_2019', 'STD_2019']].describe()
129 maximo_0 = X_class_0['mean_2019'].max()
130 maximo_1 = X_class_1['mean_2019'].max()
131 maximo_2 = X_class_2['mean_2019'].max()
132 maximo_3 = X_class_3['mean_2019'].max()
133 X_class_1.loc[(X_class_1.mean_2019 < maximo_0), 'c'] = 0
134 X_class_2.loc[(X_class_2.mean_2019 < maximo_1), 'c'] = 1
135 X_class_3.loc[(X_class_3.mean_2019 < maximo_2), 'c'] = 2
136 X_classes_todas = pd.concat([X_class_0, X_class_1, X_class_2, X_class_3])
137 X_classes_todas['PF'] = X_classes_todas['c']
138 X_classes_todas.PF = X_classes_todas.PF.replace(perfis)

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.4 – Código para a etapa de classificação de dados.

```

167 ## Random Forest
168
169 from sklearn.model_selection import train_test_split
170 X7 = dfArroz_2[['201901','201902','201903','201910','201911','201912']]
171 y7= dfArroz_2.c
172 X_train, X_test, y_train, y_test = train_test_split(X7,y7,test_size=0.20,random_state=42)
173 from sklearn.ensemble import RandomForestClassifier
174 from sklearn import tree
175 clf = RandomForestClassifier(max_features = 3, random_state = 42)
176 clf.fit(X_train, y_train)
177 y_pred = clf.predict(X_test)
178 print("\n")
179 print('Acuracia do Classificador RF nos dados de Treino: {:.2f}'
180       .format(clf.score(X_train, y_train)))
181 print('Acuracia do Classificador RF nos dados de Teste: {:.2f}'
182       .format(clf.score(X_test, y_test)))
183 from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
184 print("\n")
185 print("Matriz de Confusão:")
186 print(confusion_matrix(y_test,y_pred))
187 print("\n")
188 print(classification_report(y_test,y_pred))
189 print("Potuação de Acuracia:",accuracy_score(y_test, y_pred))
190 print("\n")
191 X_preditivo = teste_UC1[['202001','202002','202003','202010','202011','202012']]
192 max_preditivo = int(X_preditivo.max(axis=1))
193 d1 = int(X_preditivo['202001'])
194 d2 = int(X_preditivo['202002'])
195 d3 = int(X_preditivo['202003'])
196 d10 = int(X_preditivo['202010'])
197 d11 =int(X_preditivo['202011'])
198 d12 =int(X_preditivo['202012'])
199 X_preditivo0 = [d1,d2,d3,d10,d11,d12]
200 X_preditivo2 = [[d1,d2,d3,d10,d11,d12]]
201 max_preditivo2 = max(X_preditivo0)
202 classe_cod = clf.predict(X_preditivo)
203 if classe_cod == 3:
204     classePF = "Perfil 4"
205 if classe_cod == 2:
206     classePF = "Perfil 3"
207 if classe_cod == 1:
208     classePF = "Perfil 2"
209 if classe_cod == 0:
210     classePF = "Perfil 1"

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

A.1. Unidade Consumidora Irrigante UC-1

Nesta subseção de apêndices, serão apresentados os códigos para a etapa da estimativa de consumo de energia elétrica, e para a saída do modelo preditivo no estudo de caso da unidade consumidora irrigante UC-1.

Figura A.5 – Código da etapa de estimativa de consumo para o estudo de caso da UC-1.

```

140 ## Estimativa de Consumo de Energia Elétrica das UCs
141
142 teste_UC1 = X_classes_todas.loc[X_classes_todas['COD_INSTALACAO'] == UC1]
143 months = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
144 dfArroz_2020 = teste_UC1[['202001', '202002', '202003', '202004', '202005', '202006',
145                          '202007', '202008', '202009', '202010', '202011', '202012']]
146 teste_UC1['Area_2020'] = 249.36
147 teste_UC1['AMT_mca'] = 10
148 teste_UC1['VAZAO_2020'] = teste_UC1.Area_2020 * 1.5
149 teste_UC1['Potencia_Sistema_Irrigacao_2020'] = (teste_UC1.VAZAO_2020 * teste_UC1.AMT_mca
150          * 0.0098) / (0.65)
151 teste_UC1['Estimativa_CONSUMO_2020'] = teste_UC1.Potencia_Sistema_Irrigacao_2020 * (450)
152
153 a = teste_UC1.Estimativa_CONSUMO_2020
154 a = float(a)
155 Estimativa_2020 = [a, a, 0, 0,
156                   0, 0, 0, 0,
157                   0, 0, a, a]
158
159 class0 = X_classes_todas[X_classes_todas['c'] == 0]
160 class1 = X_classes_todas[X_classes_todas['c'] == 1]
161 class2 = X_classes_todas[X_classes_todas['c'] == 2]
162 class3 = X_classes_todas[X_classes_todas['c'] == 3]
163
164 dfArroz_2 = X_classes_todas.fillna(0)
165 teste_UC1 = teste_UC1.fillna(0)

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.6 – Código para saída do modelo aplicado ao estudo de caso da UC-1.

```

212 ## Saída do Modelo
213
214 print("                Relatório Final de Simulação                ")
215 print("Classe Prevista:", classePF)
216 print("Classe Original:", teste_UC1['PF'][teste_UC1.index].tolist())
217 print("\n")
218 if classe_cod[0] != int(teste_UC1.c):
219     if (classe_cod[0] < int(teste_UC1.c)):
220         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
221         print("Situação: Possível ocorrência de Perdas Não Técnicas")
222     if (classe_cod[0] > int(teste_UC1.c)):
223         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
224         print("Situação: Indicador Regular")
225 else:
226     print("Indicador 1: Não houve alteração de Classe de Perfil de Consumo")
227     print("Situação: Indicador Regular")
228 print("\n")
229
230 print("                Relatório Final de Simulação                ")
231 if (d11 or d12 or d1 or d2) < a*0.6:
232     if (d11 < a*0.6):
233         print("Indicador 2: Consumo do mês de Novembro menor do que o limite estimado")
234         print("Situação: Possível ocorrência de Perdas Não Técnicas")
235     if (d12 < a*0.6):
236         print("Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado")
237         print("Situação: Possível ocorrência de Perdas Não Técnicas")
238     if (d1 < a*0.6):
239         print("Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado")
240         print("Situação: Possível ocorrência de Perdas Não Técnicas")
241     if (d2 < a*0.6):
242         print("Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado")
243         print("Situação: Possível ocorrência de Perdas Não Técnicas")
244 else:
245     print("Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados")
246     print("Situação: Indicador Regular")

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

A.2. Unidade Consumidora Irrigante UC-2

Nesta subseção de apêndices, serão apresentados os códigos para a etapa da estimativa de consumo de energia elétrica, e para a saída do modelo preditivo no estudo de caso da unidade consumidora irrigante UC-2.

Figura A.7 – Código da etapa de estimativa de consumo para o estudo de caso da UC-2.

```

140 ## Estimativa de Consumo de Energia Elétrica das UCs
141
142 teste_UC2 = X_classes_todas.loc[X_classes_todas['COD_INSTALACAO'] == UC2]
143 months = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
144 dfArroz_2020 = teste_UC2[['202001', '202002', '202003', '202004', '202005', '202006',
145                          '202007', '202008', '202009', '202010', '202011', '202012']]
146 teste_UC2['Area_2020'] = 716.18
147 teste_UC2['AMT_mca'] = 10
148 teste_UC2['VAZAO_2020'] = teste_UC2.Area_2020 * 1.5
149 teste_UC2['Potencia_Sistema_Irrigacao_2020'] = (teste_UC2.VAZAO_2020 * teste_UC2.AMT_mca
150          * 0.0098) / (0.65)
151 teste_UC2['Estimativa_CONSUMO_2020'] = teste_UC2.Potencia_Sistema_Irrigacao_2020 * (450)
152
153 a = teste_UC2.Estimativa_CONSUMO_2020
154 a = float(a)
155 Estimativa_2020 = [a, a, 0, 0,
156                   0, 0, 0, 0,
157                   0, 0, a, a]
158
159 class0 = X_classes_todas[X_classes_todas['c'] == 0]
160 class1 = X_classes_todas[X_classes_todas['c'] == 1]
161 class2 = X_classes_todas[X_classes_todas['c'] == 2]
162 class3 = X_classes_todas[X_classes_todas['c'] == 3]
163
164 dfArroz_2 = X_classes_todas.fillna(0)
165 teste_UC2 = teste_UC2.fillna(0)

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.8 – Código para saída do modelo aplicado ao estudo de caso da UC-2.

```

212 ## Saída do Modelo
213
214 print("                Relatório Final de Simulação                ")
215 print("Classe Prevista:", classePF)
216 print("Classe Original:", teste_UC2['PF'][teste_UC2.index].tolist())
217 print("\n")
218 if classe_cod[0] != int(teste_UC2.c):
219     if (classe_cod[0] < int(teste_UC2.c)):
220         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
221         print("Situação: Possível ocorrência de Perdas Não Técnicas")
222     if (classe_cod[0] > int(teste_UC2.c)):
223         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
224         print("Situação: Indicador Regular")
225 else:
226     print("Indicador 1: Não houve alteração de Classe de Perfil de Consumo")
227     print("Situação: Indicador Regular")
228 print("\n")
229
230 print("                Relatório Final de Simulação                ")
231 if (d11 or d12 or d1 or d2) < a*0.6:
232     if (d11 < a*0.6):
233         print("Indicador 2: Consumo do mês de Novembro menor do que o limite estimado")
234         print("Situação: Possível ocorrência de Perdas Não Técnicas")
235     if (d12 < a*0.6):
236         print("Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado")
237         print("Situação: Possível ocorrência de Perdas Não Técnicas")
238     if (d1 < a*0.6):
239         print("Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado")
240         print("Situação: Possível ocorrência de Perdas Não Técnicas")
241     if (d2 < a*0.6):
242         print("Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado")
243         print("Situação: Possível ocorrência de Perdas Não Técnicas")
244 else:
245     print("Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados")
246     print("Situação: Indicador Regular")

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

A.3. Unidade Consumidora Irrigante UC-3

Nesta subseção de apêndices, serão apresentados os códigos para a etapa da estimativa de consumo de energia elétrica, e para a saída do modelo preditivo no estudo de caso da unidade consumidora irrigante UC-3.

Figura A.9 – Código da etapa de estimativa de consumo para o estudo de caso da UC-3.

```

140 ## Estimativa de Consumo de Energia Elétrica das UCs
141
142 teste_UC3 = X_classes_todas.loc[X_classes_todas['COD_INSTALACAO'] == UC3]
143 months = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
144 dfArroz_2020 = teste_UC3[['202001', '202002', '202003', '202004', '202005', '202006',
145                          '202007', '202008', '202009', '202010', '202011', '202012']]
146 teste_UC3['Area_2020'] = 92.36
147 teste_UC3['AMT_mca'] = 10
148 teste_UC3['VAZAO_2020'] = teste_UC3.Area_2020 * 1.5
149 teste_UC3['Potencia_Sistema_Irrigacao_2020'] = (teste_UC3.VAZAO_2020 * teste_UC3.AMT_mca
150          * 0.0098) / (0.65)
151 teste_UC3['Estimativa_CONSUMO_2020'] = teste_UC3.Potencia_Sistema_Irrigacao_2020 * (450)
152
153 a = teste_UC3.Estimativa_CONSUMO_2020
154 a = float(a)
155 Estimativa_2020 = [a, a, 0, 0,
156                   0, 0, 0, 0,
157                   0, 0, a, a]
158
159 class0 = X_classes_todas[X_classes_todas['c'] == 0]
160 class1 = X_classes_todas[X_classes_todas['c'] == 1]
161 class2 = X_classes_todas[X_classes_todas['c'] == 2]
162 class3 = X_classes_todas[X_classes_todas['c'] == 3]
163
164 dfArroz_2 = X_classes_todas.fillna(0)
165 teste_UC3 = teste_UC3.fillna(0)

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.10 – Código para saída do modelo aplicado ao estudo de caso da UC-3.

```

212 ## Saída do Modelo
213
214 print("                Relatório Final de Simulação                ")
215 print("Classe Prevista:", classePF)
216 print("Classe Original:", teste_UC3['PF'][teste_UC3.index].tolist())
217 print("\n")
218 if classe_cod[0] != int(teste_UC3.c):
219     if (classe_cod[0] < int(teste_UC3.c)):
220         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
221         print("Situação: Possível ocorrência de Perdas Não Técnicas")
222     if (classe_cod[0] > int(teste_UC3.c)):
223         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
224         print("Situação: Indicador Regular")
225 else:
226     print("Indicador 1: Não houve alteração de Classe de Perfil de Consumo")
227     print("Situação: Indicador Regular")
228 print("\n")
229
230 print("                Relatório Final de Simulação                ")
231 if (d11 or d12 or d1 or d2) < a*0.6:
232     if (d11 < a*0.6):
233         print("Indicador 2: Consumo do mês de Novembro menor do que o limite estimado")
234         print("Situação: Possível ocorrência de Perdas Não Técnicas")
235     if (d12 < a*0.6):
236         print("Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado")
237         print("Situação: Possível ocorrência de Perdas Não Técnicas")
238     if (d1 < a*0.6):
239         print("Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado")
240         print("Situação: Possível ocorrência de Perdas Não Técnicas")
241     if (d2 < a*0.6):
242         print("Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado")
243         print("Situação: Possível ocorrência de Perdas Não Técnicas")
244 else:
245     print("Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados")
246     print("Situação: Indicador Regular")

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

A.4. Unidade Consumidora Irrigante UC-4

Nesta subseção de apêndices, serão apresentados os códigos para a etapa da estimativa de consumo de energia elétrica, e para a saída do modelo preditivo no estudo de caso da unidade consumidora irrigante UC-4.

Figura A.11 – Código da etapa de estimativa de consumo para o estudo de caso da UC-4.

```

140 ## Estimativa de Consumo de Energia Elétrica das UCs
141
142 teste_UC4 = X_classes_todas.loc[X_classes_todas['COD_INSTALACAO'] == UC4]
143 months = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
144 dfArroz_2020 = teste_UC4[['202001', '202002', '202003', '202004', '202005', '202006',
145                          '202007', '202008', '202009', '202010', '202011', '202012']]
146 teste_UC4['Area_2020'] = 194.38
147 teste_UC4['AMT_mca'] = 10
148 teste_UC4['VAZAO_2020'] = teste_UC4.Area_2020 * 1.5
149 teste_UC4['Potencia_Sistema_Irrigacao_2020'] = (teste_UC4.VAZAO_2020 * teste_UC4.AMT_mca
150          * 0.0098) / (0.65)
151 teste_UC4['Estimativa_CONSUMO_2020'] = teste_UC4.Potencia_Sistema_Irrigacao_2020 * (450)
152
153 a = teste_UC4.Estimativa_CONSUMO_2020
154 a = float(a)
155 Estimativa_2020 = [a, a, 0, 0,
156                   0, 0, 0, 0,
157                   0, 0, a, a]
158
159 class0 = X_classes_todas[X_classes_todas['c'] == 0]
160 class1 = X_classes_todas[X_classes_todas['c'] == 1]
161 class2 = X_classes_todas[X_classes_todas['c'] == 2]
162 class3 = X_classes_todas[X_classes_todas['c'] == 3]
163
164 dfArroz_2 = X_classes_todas.fillna(0)
165 teste_UC4 = teste_UC4.fillna(0)

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.12 – Código para saída do modelo aplicado ao estudo de caso da UC-4.

```

212 ## Saída do Modelo
213
214 print("                Relatório Final de Simulação                ")
215 print("Classe Prevista:", classePF)
216 print("Classe Original:", teste_UC4['PF'][teste_UC4.index].tolist())
217 print("\n")
218 if classe_cod[0] != int(teste_UC4.c):
219     if (classe_cod[0] < int(teste_UC4.c)):
220         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
221         print("Situação: Possível ocorrência de Perdas Não Técnicas")
222     if (classe_cod[0] > int(teste_UC4.c)):
223         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
224         print("Situação: Indicador Regular")
225 else:
226     print("Indicador 1: Não houve alteração de Classe de Perfil de Consumo")
227     print("Situação: Indicador Regular")
228 print("\n")
229
230 print("                Relatório Final de Simulação                ")
231 if (d11 or d12 or d1 or d2) < a*0.6:
232     if (d11 < a*0.6):
233         print("Indicador 2: Consumo do mês de Novembro menor do que o limite estimado")
234         print("Situação: Possível ocorrência de Perdas Não Técnicas")
235     if (d12 < a*0.6):
236         print("Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado")
237         print("Situação: Possível ocorrência de Perdas Não Técnicas")
238     if (d1 < a*0.6):
239         print("Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado")
240         print("Situação: Possível ocorrência de Perdas Não Técnicas")
241     if (d2 < a*0.6):
242         print("Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado")
243         print("Situação: Possível ocorrência de Perdas Não Técnicas")
244 else:
245     print("Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados")
246     print("Situação: Indicador Regular")

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

A.5. Unidade Consumidora Irrigante UC-5

Nesta subseção de apêndices, serão apresentados os códigos para a etapa da estimativa de consumo de energia elétrica, e para a saída do modelo preditivo no estudo de caso da unidade consumidora irrigante UC-5.

Figura A.13 – Código da etapa de estimativa de consumo para o estudo de caso da UC-5.

```

140 ## Estimativa de Consumo de Energia Elétrica das UCs
141
142 teste_UC5 = X_classes_todas.loc[X_classes_todas['COD_INSTALACAO'] == UC5]
143 months = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
144 dfArroz_2020 = teste_UC5[['202001', '202002', '202003', '202004', '202005', '202006',
145                          '202007', '202008', '202009', '202010', '202011', '202012']]
146 teste_UC5['Area_2020'] = 399.81
147 teste_UC5['AMT_mca'] = 10
148 teste_UC5['VAZAO_2020'] = teste_UC5.Area_2020 * 1.5
149 teste_UC5['Potencia_Sistema_Irrigacao_2020'] = (teste_UC5.VAZAO_2020 * teste_UC5.AMT_mca
150          * 0.0098) / (0.65)
151 teste_UC5['Estimativa_CONSUMO_2020'] = teste_UC5.Potencia_Sistema_Irrigacao_2020 * (450)
152
153 a = teste_UC5.Estimativa_CONSUMO_2020
154 a = float(a)
155 Estimativa_2020 = [a, a, 0, 0,
156                   0, 0, 0, 0,
157                   0, 0, a, a]
158
159 class0 = X_classes_todas[X_classes_todas['c'] == 0]
160 class1 = X_classes_todas[X_classes_todas['c'] == 1]
161 class2 = X_classes_todas[X_classes_todas['c'] == 2]
162 class3 = X_classes_todas[X_classes_todas['c'] == 3]
163
164 dfArroz_2 = X_classes_todas.fillna(0)
165 teste_UC5 = teste_UC5.fillna(0)

```

Fonte: Elaborado pelo autor com o auxílio do Spyder®.

Figura A.14 – Código para saída do modelo aplicado ao estudo de caso da UC-5.

```

212 ## Saída do Modelo
213
214 print("                Relatório Final de Simulação                ")
215 print("Classe Prevista:", classePF)
216 print("Classe Original:", teste_UC5['PF'][teste_UC5.index].tolist())
217 print("\n")
218 if classe_cod[0] != int(teste_UC5.c):
219     if (classe_cod[0] < int(teste_UC5.c)):
220         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
221         print("Situação: Possível ocorrência de Perdas Não Técnicas")
222     if (classe_cod[0] > int(teste_UC5.c)):
223         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
224         print("Situação: Indicador Regular")
225 else:
226     print("Indicador 1: Não houve alteração de Classe de Perfil de Consumo")
227     print("Situação: Indicador Regular")
228 print("\n")
229
230 print("                Relatório Final de Simulação                ")
231 if (d11 or d12 or d1 or d2) < a*0.6:
232     if (d11 < a*0.6):
233         print("Indicador 2: Consumo do mês de Novembro menor do que o limite estimado")
234         print("Situação: Possível ocorrência de Perdas Não Técnicas")
235     if (d12 < a*0.6):
236         print("Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado")
237         print("Situação: Possível ocorrência de Perdas Não Técnicas")
238     if (d1 < a*0.6):
239         print("Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado")
240         print("Situação: Possível ocorrência de Perdas Não Técnicas")
241     if (d2 < a*0.6):
242         print("Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado")
243         print("Situação: Possível ocorrência de Perdas Não Técnicas")
244 else:
245     print("Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados")
246     print("Situação: Indicador Regular")

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

A.6. Unidade Consumidora Irrigante UC-6

Nesta subseção de apêndices, serão apresentados os códigos para a etapa da estimativa de consumo de energia elétrica, e para a saída do modelo preditivo no estudo de caso da unidade consumidora irrigante UC-6.

Figura A.15 – Código da etapa de estimativa de consumo para o estudo de caso da UC-6.

```

140 ## Estimativa de Consumo de Energia Elétrica das UCs
141
142 teste_UC6 = X_classes_todas.loc[X_classes_todas['COD_INSTALACAO'] == UC6]
143 months = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
144 dfArroz_2020 = teste_UC6[['202001', '202002', '202003', '202004', '202005', '202006',
145                          '202007', '202008', '202009', '202010', '202011', '202012']]
146 teste_UC6['Area_2020'] = 116.52
147 teste_UC6['AMT_mca'] = 10
148 teste_UC6['VAZAO_2020'] = teste_UC6.Area_2020 * 1.5
149 teste_UC6['Potencia_Sistema_Irigacao_2020'] = (teste_UC6.VAZAO_2020 * teste_UC6.AMT_mca
150          * 0.0098)/(0.65)
151 teste_UC6['Estimativa_CONSUMO_2020'] = teste_UC6.Potencia_Sistema_Irigacao_2020*(450)
152
153 a = teste_UC6.Estimativa_CONSUMO_2020
154 a = float(a)
155 Estimativa_2020 = [a, a, 0, 0,
156                   0, 0, 0, 0,
157                   0, 0, a, a]
158
159 class0 = X_classes_todas[X_classes_todas['c'] == 0]
160 class1 = X_classes_todas[X_classes_todas['c'] == 1]
161 class2 = X_classes_todas[X_classes_todas['c'] == 2]
162 class3 = X_classes_todas[X_classes_todas['c'] == 3]
163
164 dfArroz_2 = X_classes_todas.fillna(0)
165 teste_UC6 = teste_UC6.fillna(0)

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.16 – Código para saída do modelo aplicado ao estudo de caso da UC-6.

```

212 ## Saída do Modelo
213
214 print("                Relatório Final de Simulação                ")
215 print("Classe Prevista:", classePF)
216 print("Classe Original:", teste_UC6['PF'][teste_UC6.index].tolist())
217 print("\n")
218 if classe_cod[0] != int(teste_UC6.c):
219     if (classe_cod[0] < int(teste_UC6.c)):
220         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
221         print("Situação: Possível ocorrência de Perdas Não Técnicas")
222     if (classe_cod[0] > int(teste_UC6.c)):
223         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
224         print("Situação: Indicador Regular")
225 else:
226     print("Indicador 1: Não houve alteração de Classe de Perfil de Consumo")
227     print("Situação: Indicador Regular")
228 print("\n")
229
230 print("                Relatório Final de Simulação                ")
231 if (d11 or d12 or d1 or d2) < a*0.6:
232     if (d11 < a*0.6):
233         print("Indicador 2: Consumo do mês de Novembro menor do que o limite estimado")
234         print("Situação: Possível ocorrência de Perdas Não Técnicas")
235     if (d12 < a*0.6):
236         print("Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado")
237         print("Situação: Possível ocorrência de Perdas Não Técnicas")
238     if (d1 < a*0.6):
239         print("Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado")
240         print("Situação: Possível ocorrência de Perdas Não Técnicas")
241     if (d2 < a*0.6):
242         print("Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado")
243         print("Situação: Possível ocorrência de Perdas Não Técnicas")
244 else:
245     print("Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados")
246     print("Situação: Indicador Regular")

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

A.7. Unidade Consumidora Irrigante UC-7

Nesta subseção de apêndices, serão apresentados os códigos para a etapa da estimativa de consumo de energia elétrica, e para a saída do modelo preditivo no estudo de caso da unidade consumidora irrigante UC-7.

Figura A.17 – Código da etapa de estimativa de consumo para o estudo de caso da UC-7.

```

140 ## Estimativa de Consumo de Energia Elétrica das UCs
141
142 teste_UC7 = X_classes_todas.loc[X_classes_todas['COD_INSTALACAO'] == UC7]
143 months = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
144 dfArroz_2020 = teste_UC7[['202001', '202002', '202003', '202004', '202005', '202006',
145                          '202007', '202008', '202009', '202010', '202011', '202012']]
146 teste_UC7['Area_2020'] = 225.84
147 teste_UC7['AMT_mca'] = 10
148 teste_UC7['VAZAO_2020'] = teste_UC7.Area_2020 * 1.5
149 teste_UC7['Potencia_Sistema_Irrigacao_2020'] = (teste_UC7.VAZAO_2020 * teste_UC7.AMT_mca
150          * 0.0098) / (0.65)
151 teste_UC7['Estimativa_CONSUMO_2020'] = teste_UC7.Potencia_Sistema_Irrigacao_2020 * (450)
152
153 a = teste_UC7.Estimativa_CONSUMO_2020
154 a = float(a)
155 Estimativa_2020 = [a, a, 0, 0,
156                   0, 0, 0, 0,
157                   0, 0, a, a]
158
159 class0 = X_classes_todas[X_classes_todas['c'] == 0]
160 class1 = X_classes_todas[X_classes_todas['c'] == 1]
161 class2 = X_classes_todas[X_classes_todas['c'] == 2]
162 class3 = X_classes_todas[X_classes_todas['c'] == 3]
163
164 dfArroz_2 = X_classes_todas.fillna(0)
165 teste_UC7 = teste_UC7.fillna(0)

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.

Figura A.18 – Código para saída do modelo aplicado ao estudo de caso da UC-7.

```

212 ## Saída do Modelo
213
214 print("                Relatório Final de Simulação                ")
215 print("Classe Prevista:", classePF)
216 print("Classe Original:", teste_UC7['PF'][teste_UC7.index].tolist())
217 print("\n")
218 if classe_cod[0] != int(teste_UC7.c):
219     if (classe_cod[0] < int(teste_UC7.c)):
220         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
221         print("Situação: Possível ocorrência de Perdas Não Técnicas")
222     if (classe_cod[0] > int(teste_UC7.c)):
223         print("Indicador 1: Alteração de Classe de Perfil de Consumo")
224         print("Situação: Indicador Regular")
225 else:
226     print("Indicador 1: Não houve alteração de Classe de Perfil de Consumo")
227     print("Situação: Indicador Regular")
228 print("\n")
229
230 print("                Relatório Final de Simulação                ")
231 if (d11 or d12 or d1 or d2) < a*0.6:
232     if (d11 < a*0.6):
233         print("Indicador 2: Consumo do mês de Novembro menor do que o limite estimado")
234         print("Situação: Possível ocorrência de Perdas Não Técnicas")
235     if (d12 < a*0.6):
236         print("Indicador 2: Consumo do mês de Dezembro menor do que o limite estimado")
237         print("Situação: Possível ocorrência de Perdas Não Técnicas")
238     if (d1 < a*0.6):
239         print("Indicador 2: Consumo do mês de Janeiro menor do que o limite estimado")
240         print("Situação: Possível ocorrência de Perdas Não Técnicas")
241     if (d2 < a*0.6):
242         print("Indicador 2: Consumo do mês de Fevereiro menor do que o limite estimado")
243         print("Situação: Possível ocorrência de Perdas Não Técnicas")
244 else:
245     print("Indicador 2: Consumo dos meses de Novembro, Dezembro, Janeiro e Fevereiro dentro dos limites estimados")
246     print("Situação: Indicador Regular")

```

Fonte: Elaborado pelo o autor com o auxílio do Spyder®.