

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

Natalia Bastos de Sousa

**UMA METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO
TÉCNICAS EM UNIDADES CONSUMIDORAS IRRIGANTES
DE CAFÉ DO ESTADO DE SÃO PAULO**

Santa Maria, RS

2022

Natalia Bastos de Sousa

**UMA METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS
EM UNIDADES CONSUMIDORAS IRRIGANTES
DE CAFÉ DO ESTADO DE SÃO PAULO**

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

Orientador: Prof. Daniel Pinheiro Bernardon

Santa Maria, RS

2022

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Sousa, Natalia

UMA METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO
TÉCNICAS EM UNIDADES CONSUMIDORAS IRRIGANTES DE CAFÉ DO
ESTADO DE SÃO PAULO / Natalia Sousa.- 2022.

148 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, 2022

1. Perdas não Técnicas 2. Consumidores Rurais 3.
Irrigantes de Café 4. Aprendizado de Máquina 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.

Declaro, NATALIA SOUSA, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Dissertação) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

Natalia Bastos de Sousa

**UMA METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS
EM UNIDADES CONSUMIDORAS IRRIGANTES
DE CAFÉ DO ESTADO DE SÃO PAULO**

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

Aprovado em 29 de Julho de 2022:

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

Luciano Lopes Pfitscher, Dr. Eng. (UFSC)

Nelson Knak Neto, Dr. Eng. (UFSM)

Santa Maria, RS

2022

AGRADECIMENTOS

Primeiramente, meus agradecimentos são destinados aos meus pais, por todo amor, paciência, confiança e direcionamentos me dados, que me trouxeram onde estou, e continuarão ajudando a trilhar meu caminho. Sempre serei grata ao meu pai, João Gabaia, e a minha mãe, Ana Cristina, por deixarem que eu tome minhas próprias decisões, mesmo sendo contrárias aos seus desejos e opiniões, por ser meu exemplo de força, bondade, amor e tolerância.

À minhas irmãs Natasha, Natali e Natássia, sou grata pelas nossas diferenças, pois suas visões de mundo ajudam a expandir o meu mundo. Agradeço por nos entender e aceitar, por me incentivarem e por acreditarem em mim, quando eu mesma duvido, por me deixarem assistir aos jogos e *Esquecera de mim* e *Legalmente Loira*, pelas milésimas vezes, só porque me faz bem assistir, mesmo quando querem assistir a qualquer outra coisa. Aos meus gatos Fergus e Momo, e minha cachorra Pandora, que sempre aquecem meu coração, e me dão saudades quando não estamos juntos em Parnaíba, casa dos meus pais e sede familiar. Agradeço, também a minha tia, Alice Quinn, e à sua família na Irlanda, pelo apoio e incentivo.

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.

Ao meu Orientador, Prof. Dr. Daniel Pinheiro Bernardon, pela orientação, paciência e confiança para realização deste trabalho.

Aos meus colegas, Henrique, Pedro e Daniel Lemes, pelo apoio e amizade, aprendizagem e troca de ideias que foram essenciais no desenvolvimento do trabalho. A todos os professores que durante minha vida acadêmica dividiram seus conhecimentos comigo. Aos amigos e colegas do Centro de Excelência em Energia e Sistemas de Potência (CEESP) pelo bom convívio e compartilhamento de conhecimentos.

A Professora Dr^a Luciana Marini Köpp da Universidade Federal de Pelotas (UFPEL), pela paciência e compartilhamento de conhecimentos ao decorrer deste trabalho.

Por último, e, não menos importante, agradeço a Deus, pela família que Ele me deu, saúde e por ter me dado força durante todo o período de Graduação.

If you live by comparing yourself to others
You forget to count your blessing
(Peniel Shin - Thank You)

RESUMO

UMA METODOLOGIA PARA IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM UNIDADES CONSUMIDORAS IRRIGANTES DE CAFÉ DO ESTADO DE SÃO PAULO

AUTORA: Natalia Bastos de Sousa

ORIENTADOR: Prof. Dr. Daniel Pinheiro Bernardon

No âmbito da rede rural, vários fatores podem influenciar na dificuldade por parte das distribuidoras em localizar unidades consumidoras fraudulentas ou irregulares que possam estar causando perdas não técnicas, como por exemplo, a dificuldade em correlacionar os consumos de seus clientes irrigantes com ocorrências de consumos divergentes do esperado, devido ao conhecimento técnico necessário a respeito das características de cada tipo de lavoura. Dessa forma, essa dissertação tem como objetivo apresentar uma metodologia de detecção de possíveis ocorrências de perdas não técnicas para unidades consumidoras irrigantes de café clientes da CPFL Paulista. Para isso foi realizado o estudo das características fenológicas do cafeeiro, atributos de manejo do sistema irrigante, e de dados de agrometeorologia essenciais para as estimativas de consumo. Mapeando os períodos esperados de manejo do sistema de irrigação, foi empregado a técnica de clusterização por Kmeans para agrupar as unidades consumidoras, gerando classes de perfis de consumo com base em seus históricos de faturamento de energia e o período de registro da energia faturada (mês). Essas relações de classes e consumo (mensais), são utilizadas como entrada da implementação de um modelo *Random Forest*, utilizado como o modelo preditivo que classifica as unidades com base em consumos mais atuais e compara a classe predita com a classe armazenada na base de dados. A identificação de clientes que possam ser fraudulentos ocorre se a classe predita for de nível inferior à classe esperada para o mesmo mês observado. Complementando a metodologia, uma vez que unidades consumidoras de classes inferiores não teriam mudança de Perfil de Consumo na ocorrência de menor consumo de energia, foi utilizado o método de estimativa de consumo para detectar possíveis ocorrências de perdas não técnicas, comparando o consumo real com o estimado. O modelo preditivo classificatório mostrou-se capaz de detectar as mudanças de Perfis de Consumo de maneira satisfatória, e o modelo de estimativa mostra que, para as unidades consumidoras da zona cafeeira de São Paulo, o método de irrigação ainda é utilizado de maneira complementar à produtividade da fazenda.

Palavras-chave: Perdas não Técnicas; Consumidores Rurais; Irrigantes de Café; Aprendizado de Máquina

ABSTRACT

A METHODOLOGY FOR IDENTIFYING NON-TECHNICAL LOSSES IN IRRIGANT COFFEE CONSUMERS IN THE STATE OF SÃO PAULO

AUTHOR: Natalia Bastos de Sousa

ADVISOR: Prof. Dr. Daniel Pinheiro Bernardon

Within the rural network, several factors can influence the difficulty on the part of distributors to locate fraudulent or irregular consumer units that may be causing non-technical losses, such as, for example, the difficulty in correlating the consumption of their irrigating customers with occurrences of divergent consumption. than expected, due to the necessary technical knowledge about the characteristics of each type of crop. Thus, this dissertation aims to present a methodology for detecting possible occurrences of non-technical losses for coffee irrigating consumer units that are CPFL Paulista's clients. For this, the study of the phenological characteristics of the coffee tree, management attributes of the irrigation system, and agrometeorological data essential for consumption estimates was carried out. Mapping the expected periods of irrigation system management, the clustering by Kmeans technique was used to group the consumer units, generating classes of consumption profiles based on their energy billing history and the period of billed energy registration (month). These classes and consumption (monthly) relationships are used as input to the implementation of a Random Forest model, used as the predictive model that classifies the units based on the most current consumptions and compares the predicted class with the class stored in the database. . The identification of customers that may be fraudulent occurs if the predicted class is of a lower level than the expected class for the same observed month. Complementing the methodology, since consumer units of lower classes would not change their Consumption Profile in the event of lower energy consumption, the consumption estimation method was used to detect possible occurrences of non-technical losses, comparing the real consumption with the estimated. The classification predictive model was able to detect changes in Consumption Profiles in a satisfactory way, and the estimation model shows that, for consumer units in the coffee growing area of São Paulo, the irrigation method is still used in a complementary way to the farm productivity.

Keywords: Non-Technical Losses; Rural Consumers; Coffee Irrigators; Machine Learning.

LISTA DE FIGURAS

Figura 1 – Perdas de Energia: média de 2010 – 2012.	30
Figura 2 - Perdas nas redes T&D de energia elétrica (% da produção) – Mundo, Brasil etc. (2014).	31
Figura 3 – Perdas Totais de Países.	32
Figura 4 – Perdas sobre a Energia Injetada (2020).....	32
Figura 5 - Perdas sobre a Energia Injetada (2008-2020).	33
Figura 6 – Demonstração de Perdas no Setor Elétrico.	35
Figura 7 – Participação de PNT Reais / PNT Brasil (2020).	36
Figura 8 – Participação das PNT regulatória sobre a Parcela B da tarifa (2020).	39
Figura 9 – Distribuição do Consumo de Energia Elétrica por Classe de Consumo, ano de 2020.	46
Figura 10 – Mapa das Regiões Produtoras de Café no Brasil.	51
Figura 11 – Representação Gráfica da estimativa de custos de produção de café.	52
Figura 12 – Período Fenológico do Café arábica, abrangendo as seis fases, durante 24 meses.	56
Figura 13 – Representação da Capacidade de Armazenamento de Água nos Poros do Solo ..	59
Figura 14 – Exemplo de Balanço Hídrico Climático de Mococa, SP.	62
Figura 15 – Exemplos de Áreas de Café com Irrigação por Pivô.	64
Figura 16 – Supervisionado x Não Supervisionado.	71
Figura 17 - Exemplo de aplicação de Cluster por Particionamento.	73
Figura 18 - Exemplo de aplicação de Cluster Hierárquico.....	73
Figura 19 - Exemplo de aplicação de Cluster Baseado em grade.	74
Figura 20 – Fluxograma da Metodologia Proposta.	76
Figura 21 - Carta Geográfica do Brasil ao Milionésimo	80
Figura 22 – Leitura do arquivo shapefile.	81
Figura 23 – Área de aplicação da metodologia.	86
Figura 24 - Desvio Padrão vs Média de Consumo para cada UC (72 meses).	90
Figura 25 – Máximo Consumos de Energia Registrado por Mês.	91
Figura 26 – Total de Energia Consumida por Período.	92
Figura 27 - Mapa de São Paulo e Centroides de Áreas de Plantio de Café.....	93
Figura 28 – Centroide de Áreas de Plantio de Café e Consumidores Irrigantes.	93
Figura 29 – Exemplo 1 de Relação Centroide x UC não correspondentes.....	94

Figura 30 - UC e Centroide de Área resultantes do pré-processamento.	95
Figura 31 – Resultado do Método <i>Elbow</i> para definir k.	98
Figura 32- Resultado da Etapa de Clusterização por Perfil de Consumo.	98
Figura 33 – N° de amostras por Cluster Gerado.	99
Figura 34 – Representação Gráfica dos Clusters de Perfil de Consumo.	99
Figura 35- Ilustração de Classificador de <i>Random Forest</i>	101
Figura 36 – Dados de Entrada do Modelo RF.	103
Figura 37 - Relatório Final de Simulação do Modelo RF.	105
Figura 38 – Pesos Atribuídos para Cada Classe.	106
Figura 39 – Exemplo de Balanço Hídrico.	111
Figura 40 – Caso 1: Resultado da Simulação do Modelo RF.	114
Figura 41 - Caso 1 validação: Resultado da Simulação do Modelo RF.	114
Figura 42 - Caso 1 Consumo Alterado: Resultado da Simulação do Modelo RF.	115
Figura 43 - Caso 2: Resultado da Simulação do Modelo RF.	116
Figura 44 - Caso 2 Consumos Alterados: Resultados da Simulação do Modelo RF.	117
Figura 45 - Caso 3: Resultado da Simulação do Modelo RF.	118
Figura 46 – Caso 1 da Estimativa: Atributos e Características.	119
Figura 47 – Caso 1 da Estimativa: Atributos para Cálculo de Estimativa de Consumo.	120
Figura 48 - Caso 1 da Estimativa: Resultado do BH.	120
Figura 49 - Caso 1 da Estimativa: Resultado de Energia Consumida.	121
Figura 50 - Caso 1 Estimativa de Consumo: Relatório Final de Simulação.	122
Figura 51 - Caso 2 da Estimativa: Atributos e Características.	123
Figura 52 - Caso 2 da Estimativa: Distância da Fonte de Água.	124
Figura 53 - Caso 2 da Estimativa: Resultado do BH.	125
Figura 54 - Caso 2 da Estimativa: Resultado de Energia Consumida.	125
Figura 55 - Caso 2 Estimativa de Consumo: Relatório Final de Simulação.	126

LISTA DE TABELAS

Tabela 1 – Participação das PNT em relação ao Brasil e Participação sobre a Parcela B.	39
Tabela 2- Distribuidoras com % de participação de PNT sobre a tarifa acima da média nacional.....	40
Tabela 3 – Valores médios mensais de coeficiente de cultura (Kc) para o café (Coffea arábica).	63
Tabela 4 – Atributos selecionados para o conjunto de dados de entrada da distribuidora	78
Tabela 5 - Descrição dos Componentes Gráficos.....	81
Tabela 6 – Descrição dos Atributos da Entrada de CAD.	82
Tabela 7 – Descrição dos Atributos da Entrada de Fenologia e Mapeamento Tecnológico. ...	83
Tabela 8 - Descrição dos Atributos da Entrada de Fenologia e Mapeamento Tecnológico.....	84
Tabela 9 - Descrição dos Atributos da Entrada de Áreas.	85
Tabela 10 – Exemplificação da Estrutura de dados de consumo e cadastrais.....	89
Tabela 11 – N° de Casos Por Faixa de Distâncias.....	95
Tabela 12 - Matriz de Confusão.	103
Tabela 13 – Médias máximas e Mínimas de cada Classe.....	112
Tabela 14 – Caso 1: Dados de Consumo.....	113
Tabela 15 - Caso 1: Dados de Perfil de Consumo.....	113
Tabela 16 – Caso 1: Novos Dados de Entrada e Variação.	115
Tabela 17 - Caso 2: Dados de Consumo.	116
Tabela 18 - Caso 2: Dados de Perfil de Consumo.....	116
Tabela 19 - Caso 3: Dados de Consumo e Variação Percentual.	118
Tabela 20 - Caso 3: Dados de Perfil de Consumo.....	118
Tabela 21 - Caso 1 Método de Estimativa: Consumos Mensais.	122
Tabela 22 - Caso 2 Método de Estimativa: Consumos Mensais.	126

SUMÁRIO

1. INTRODUÇÃO.....	21
1.1 MOTIVAÇÃO E JUSTIFICATIVA.....	23
1.2 OBJETIVO GERAL.....	25
1.3 OBJETIVOS ESPECÍFICOS.....	25
1.4 CONTRIBUIÇÕES DA DISSERTAÇÃO.....	26
1.5 ORGANIZAÇÃO DOS CAPÍTULOS.....	27
2. REVISÃO BIBLIOGRÁFICA.....	28
2.1 PERDAS DE ENERGIA ELÉTRICA - CONTEXTUALIZAÇÃO.....	28
2.1.1 <i>Perdas de energia: Cenário Global x Brasil</i>	29
2.1.2 <i>Perdas de Energia Elétrica na rede de Distribuição no Brasil</i>	32
2.1.3 <i>Perdas Não Técnicas (PNT) no Brasil</i>	33
2.1.4 <i>Impacto das Perdas nas Tarifas de Energia Elétrica</i>	36
2.2.1 <i>Métodos Indiretos</i>	40
2.2.2 <i>Métodos Diretos</i>	41
2.3 TRABALHOS RELACIONADOS À DETECÇÃO E CÁLCULO DE PNT.....	42
2.4 CARACTERIZAÇÃO DO CONSUMO DE ENERGIA NO BRASIL.....	45
2.4.1 <i>Classes de Consumo e suas Subclasses</i>	46
2.4.2 <i>Classe de Consumidores Rurais e Atividades de Irrigação</i>	48
2.4.3 <i>Desconto na Tarifa de Unidades Consumidoras Rurais Irrigantes</i>	49
2.5 LAVOURAS DE CAFÉ.....	50
2.5.1 <i>Fenologia do Café</i>	54
2.5.2 <i>Irrigação do Café</i>	57
2.5.3 <i>Manejo do Sistema Irrigante</i>	58
2.5.4 <i>Métodos de Manejo do Sistema Irrigante</i>	60
2.5.5 <i>Principais Sistemas de Irrigação do Cafeeiro</i>	63
2.6 AGROMETEOROLOGIA.....	66
2.7 BIBLIOTECAS PYTHON E ALGORITMOS UTILIZADOS.....	68
2.7.1 <i>Ambiente Spyder e Python</i>	68
2.7.2 <i>Pacotes Python Utilizados</i>	69
2.7.3 <i>Algoritmos de Aprendizado de Máquina Utilizados</i>	70
2.8 RESUMO DO CAPÍTULO.....	74
3. METODOLOGIA PARA DETECÇÃO DE PNT EM UNIDADES CONSUMIDORAS IRRIGANTES DE CAFÉ.....	76
3.1 CONJUNTO DE DADOS DE ENTRADA.....	77
3.1.1 <i>Dados de Entrada – Consumo e Cadastrais</i>	77
3.1.2 <i>Dados de Entrada – Dados de Desnível</i>	79
3.1.3 <i>Dados de Entrada – Dados de CAD</i>	82
3.1.4 <i>Dados de Entrada – Dados de Fenologia e Mapeamento Tecnológico</i>	83
3.1.5 <i>Dados de Entrada – Dados Meteorológicos</i>	83
3.1.6 <i>Dados de Entrada – Dados de Áreas</i>	84
3.2 LIMITAÇÕES ENCONTRADAS E CONSIDERAÇÕES.....	86
3.3 PRÉ-PROCESSAMENTO DE DADOS.....	87
3.4 APLICAÇÃO DO ALGORITMO DE CLUSTERIZAÇÃO E PREDIÇÃO.....	96
3.4.2 <i>Resultado da Clusterização – Geração de Classes de Perfil de Consumo</i>	97
3.4.3 <i>Random Forest – RF</i>	100

3.4.4	<i>Modelo RF Classificatório</i>	101
3.4.5	<i>Métricas de Avaliação do Modelo RF Classificatório e Resultados</i>	103
3.5	METODO DE ESTIMATIVA DE CONSUMO DAS UC IRRIGANTES	106
3.5.1	<i>Equações Utilizadas para Estimativa de Consumo</i>	106
3.5.2	<i>Exemplo de Cálculo do Balanço Hídrico</i>	109
4.	ESTUDO DE CASO	112
4.1	ESTUDO DE CASO - APLICAÇÃO DO MODELO RF	113
4.2	ESTUDO DE CASO – ESTIMATIVA DE CONSUMO	119
5.	CONCLUSÕES	127
5.1	SUGESTÕES DE CONTINUIDADE	128
5.2	PUBLICAÇÕES RELACIONADAS AO TRABALHO	129
	REFERÊNCIAS BIBLIOGRÁFICAS	130
	APÊNDICE – CODIGO DE IMPLEMENTAÇÃO DA METODOLOGIA	137

1. INTRODUÇÃO

Todo ano o sistema de potência enfrenta grandes perdas de energia que representam em redução dos ativos das distribuidoras de energia, uma vez que, essas perdas referem-se à energia elétrica que não foi comercializada. As perdas na Distribuição podem ser definidas como a diferença entre a energia elétrica adquirida pelas distribuidoras e a faturada aos seus consumidores, podem ser técnicas ou não técnicas (ANEEL, 2020). As perdas técnicas (PT) são inerentes à atividade de distribuição de energia elétrica, sendo a parcela que está associada às características de carregamento e configuração das redes das concessionárias de distribuição. Já as perdas não técnicas (PNT) se originam principalmente nos furtos, fraudes, erros de leitura, medição e faturamento, essas perdas são apuradas pela diferença entre as perdas totais e as perdas técnicas (ANEEL, 2020).

Em geral as PNT, incluem roubo de eletricidade, incluindo adulteração de medidores para confirmar que o medidor registrou uma leitura de consumo mais baixa, grampear linhas de transmissão, organizar leituras falsas, roubar energia ignorando os medidores ao fazer conexões ilegais, ignorar contas não pagas, falhas nos medidores de energia ou fornecimento não medido, erros e atrasos na leitura e faturamento, desvios de linhas não autorizados, não pagamento por parte dos clientes, inadequação e imprecisões na leitura dos contadores, faturamento impreciso da unidade consumidora, etc. (CHAUHAN, 2015).

Entre os tipos de PNT listados, Madrigal (2017) destaca que é importante indicar que também pode haver perdas em processos de faturamento e cobrança, mas que esses prejuízos são classificados como financeiros e dependem da eficiência administrativa da distribuidora. Além das questões econômicas, as PNT impactam no próprio planejamento e operação da rede, uma vez que, investimentos em expansão e equipamentos podem ser necessários para suprir um consumo não previsto e/ou não faturado, e que consumidores ilegais podem interferir na segurança e qualidade de energia da rede (EVALDT, 2014).

O relatório de Perdas de Energia Elétrica na Distribuição da ANEEL (2021) mostra que no ano de 2020 as perdas totais na distribuição representaram aproximadamente 14,8% do mercado consumidor. Em específico, as perdas não técnicas chegaram à marca de 37,9 TWh. Uma vez que as perdas reais das distribuidoras não são totalmente repassadas para a tarifa, esse valor representa para as concessionárias energia que foi comprada, gerada, mas não faturada. Em 2020 os custos referentes as perdas não técnicas representaram aproximadamente R\$ 8,6 bilhões (ANEEL, 2020).

Embora as distribuidoras de energia apresentem soluções para minimizar o impacto das PNT realizando campanhas educativas de conscientização, inspeções nos medidores de unidades consumidoras suspeitas, implementação de medidores eletrônicos, entre outras atividades. Não há como determinar de uma maneira totalmente confiável, quais unidades consumidoras estão de fato gerando PNT sem uma inspeção *in loco* (ZUEGE, 2018). Existem diversos fatores que dificultam a prática da fiscalização local (*in loco*), como os elevados custos com equipes de manutenção, 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. Desse modo, essa prática acaba apresentando uma resposta de baixa eficácia para o problema de detecção de perdas não técnicas (EVALDT, 2018; GHORI et al., 2020). Considerando que, inspecionar todos os consumidores suspeitos de gerar perdas é praticamente inviável, uma das soluções para este problema é analisar os registros de consumo e medição destes consumidores a fim de encontrar automaticamente o conhecimento necessário para executar uma ação corretiva ou preventiva (CASTRO; FERRARI, 2016).

Na literatura, são propostos vários métodos para a estimativa de perdas de energia na rede de distribuição, concentrados, em sua maioria, em dois tipos de métodos: clássicos aprimorados e métodos baseados em técnicas de inteligência artificial, como, por exemplo, lógica *fuzzy*, redes neurais, algoritmos genéticos e *clustering* (GRIGORAS; SCARLATACHE, 2013). Contudo são escassos os trabalhos envolvendo especificamente clientes de distribuidoras cadastrados como irrigantes.

Este Trabalho propõe uma metodologia empregando clusterização (agrupamento) de dados de consumo de energia elétrica das unidades consumidoras objeto de estudo, de maneira a encontrar os períodos característicos de queda e aumento de consumo devido à utilização do sistema de irrigação e ao seu desligamento. Mapeando essas ocorrências dentro de uma faixa de período mensal com os consumos anteriores dos clientes irrigantes, gerando classes de Perfis de Consumo baseadas em faixas do consumo faturado de energia, onde é possível comparar os consumos da mesma faixa de período dos anos anteriores com o ano vigente e encontrar possíveis divergências entre o consumo esperado e o consumo real. Para encontrar tais ocorrências, foi utilizado o método *Random Forest* para determinar se houve mudanças de Perfis de consumo. De maneira auxiliar, foi feita a metodologia de estimativa de consumo para fazendas de café da região de estudo (Alta Mogiana) para comparar os consumos reais de energia elétrica das unidades consumidoras com o consumo de energia estimado dos clientes que, no processo de clusterização, foram classificados na classe mais

inferior, de maneira que o menor faturamento de energia, ou ausência de faturamento de energia, não resultaria em mudança de perfil de consumo detectável pelo modelo *Random Forest*. Durante o período de elaboração dessa dissertação essa metodologia de estimativa de consumo foi utilizada de maneira complementar, como descrita, contudo, espera-se poder utilizá-la para alimentar o modelo preditivo.

O presente trabalho foi realizado com dados da Companhia Paulista de Força e Luz (CPFL) juntamente ao projeto de perdas não técnicas de irrigantes, durante o período de 2019 a 2022, estando ainda em implementação.

1.1 MOTIVAÇÃO E JUSTIFICATIVA

Os altos percentuais de PNT vêm causando muitos impactos nos sistemas de distribuição. Entre estes, destacam-se as consequências na tarifa de energia elétrica, afetando diretamente o consumidor na arrecadação de tributos; afetando o Estado devido aos impostos que não são arrecadados desta parte de energia consumida irregularmente; e ainda os impactos na qualidade do serviço prestado pela concessionária (SAMPAIO, 2012). Logo, com a perda de receita as distribuidoras sofrem redução de investimentos em infraestrutura, que poderiam ser usados para melhorar os índices de qualidade de energia. Além do aspecto financeiro, as ligações clandestinas prejudicam a integridade física da rede elétrica, pois causa sobrecargas e desligamentos indevidos no sistema de distribuição, além do uso irracional dos recursos naturais, da poluição e degradação ambiental (VIDINICH; NERY. 2009).

Na maioria das empresas de distribuição a identificação e atuação sobre as PNT são realizadas através de busca exaustiva, ou seja, seleciona-se uma área geográfica para a "operação de varredura" na qual todos os pontos de conexão de rede são inspecionados em busca de situações irregulares (ARANHA NETO, 2012). Se considerarmos o âmbito rural, localização do objeto de estudo desta dissertação, essas unidades consumidoras que estão mais afastadas são as que mais apresentam ineficiências para este tipo de operação, tanto ao considerarmos o tempo de atuação que seria requerido aos agentes de inspeção, quanto aos custos, limitações de acesso, entre outros fatores.

Diante do contexto apresentado, a pesquisa apresentada nesta dissertação refere-se às ocorrências de PNT em unidades consumidoras rurais irrigantes produtoras de café do estado de São Paulo. Dentro da área rural, nas fazendas, basicamente a energia é utilizada em atividades, como: (i) Abastecimento das dependências; (ii) Iluminação; (iii) Bombeamento de

água e (iv) Suporte de equipamentos de produção. Além disso, sabe-se que no setor rural a maior parcela de consumo de energia elétrica refere-se à utilização de maquinário, por exemplo, para preparo do solo, aplicação de fertilizantes, irrigação e combate a pragas. Uma fazenda de café apresenta as atividades citadas, logo, a parcela de consumo de energia elétrica, estudada nesta pesquisa, são as de atividade de irrigação e aplicação de fertirrigação (em que são utilizadas as bombas de água do sistema irrigante para aplicação de fertilizantes).

Para a realização do estudo de possíveis ocorrências de perdas não técnicas nas unidades consumidoras irrigantes de café, é apresentado o estudo do mapeamento tecnológico das fazendas de café no Estado de São Paulo, realizado com objetivo de identificar características e variáveis direta ou, indiretamente relacionadas ao manejo dos sistemas de irrigação. A escolha do tipo de cultura foi motivada pelo acesso aos dados históricos de consumo de unidades consumidoras rurais do Estado de São Paulo (disponibilizados pela CPFL Paulista) que é o segundo maior produtor de café, perdendo apenas para Minas Gerais, possuindo diversos municípios presentes em regiões cafeeiras de destaque do Brasil, destacando-se a região cafeeira de Alta Mogiana com quinze municípios paulistas fazendo pertencentes a mesma, região que corresponde à metade da produção de café do Estado de São Paulo, sendo esta uma das regiões de produção do grão com cultivo cuidadoso e focado na qualidade, uma região que implementa irrigação a mais de 20 anos (AGROMOGIANA, 2018).

Em geral, os trabalhos sobre PNT apresentam seu valor calculado a partir da estimativa de PT, por métodos indiretos que se caracterizam pelo uso de medidas da rede, dados de equipamentos e comparação entre sistemas, que são métodos de baixo grau de precisão. Métodos indiretos com maior grau de conhecimento do sistema exploram dados tais como parâmetros elétricos e curvas de carga do sistema, via aplicativos de fluxo de potência, e oferecem maiores possibilidades de identificar perdas com precisão (FERREIRA, 2015). Já os estudos que apresentam métodos diretos calculam diretamente as PNT por utilização e dados históricos ou dados obtidos em tempo real (FERREIRA; CURADO, 2015).

Segundo Rodrigues (2018), os métodos que utilizam de dados históricos, de maneira geral, empregam uma base de dados e geram um padrão de informações que serve para identificar possíveis fraudes de energia, em que a etapa inicial de preparação dos dados pode utilizar diferentes técnicas tais como utilização de redes neurais artificiais ou *clustering*. Dentro dos estudos de PNT que utilizam os métodos descritos, são escassos os trabalhos focados nos consumidores rurais, onde a maioria faz uma análise da rede de distribuição de estudo como um todo. Em especial, pesquisas que tratam dos consumidores, classificados no

banco de dados das concessionárias como irrigantes, com o objetivo de detectar PNT na área rural são inusuais.

Com isso, este trabalho busca apresentar uma metodologia de detecção de PNT, se diferenciando ao utilizar dados do mapeamento tecnológico, e, também, empregando dados de agrometeorologia e processamento de imagens para estimativa do consumo de energia elétrica das lavouras.

1.2 OBJETIVO GERAL

O objetivo principal deste trabalho é desenvolver uma metodologia para detecção de possíveis ocorrências de PNT em unidades consumidoras irrigantes de café, que são atendidas pela distribuidora de São Paulo, CPFL Paulista. A metodologia emprega um modelo preditivo classificatório que utiliza técnica de clusterização e algoritmo Random Forest para encontrar unidades consumidoras que apresentam consumos divergentes do esperado e um modelo de estimativa de consumo para os irrigantes de café, que será utilizado de maneira complementar.

1.3 OBJETIVOS ESPECÍFICOS

Como objetivos específicos deste trabalho podem ser citados:

- Fazer o mapeamento tecnológico dos sistemas de irrigação aplicados às lavouras de café;
- Identificar as variáveis meteorológicas que influenciam no manejo do sistema de irrigação;
- Realizar Associação de unidades consumidoras às áreas de plantio previamente identificadas por meio de processamento de imagens de satélite;
- Selecionar atributos da base de dados das unidades consumidoras que possam auxiliar na identificação de ocorrência de PNT;
- Gerar novos atributos na base de dados que possam auxiliar no processo de clusterização das unidades consumidoras;
- Comparar o consumo de energia elétrica de lavouras de café;
- Empregar técnicas de inteligência artificial para implementar um modelo preditivo que acuse casos de possível ocorrência de PNT;

- Validar por meio de métricas a técnica implementada de maior acuracidade na detecção de PNT.

1.4 CONTRIBUIÇÕES DA DISSERTAÇÃO

Segundo a revisão bibliográfica realizada, a maioria dos trabalhos encontrados que utilizam inteligência artificial na detecção de PNT são normalmente realizados para os consumidores de áreas urbanas, apresentando um conjunto de dados fornecido por uma distribuidora, no qual tais consumidores estão classificados como comercial ou residencial, ou ainda, residencial ou industrial, e com base em técnicas de mineração de dados e técnicas de inteligência artificial são feitas classificações ou detecção de outliers (clientes fraudulentos). Desta forma, são inusuais as metodologias que avaliam conjunto de dados de consumidores rurais, mais especificamente os irrigantes, que geralmente utilizam de tarifa beneficiária noturna (tarifa verde).

Diante disso, como contribuição deste trabalho destaca-se:

- Desenvolvimento de uma metodologia que detecte possíveis ocorrências de PNT em irrigantes rurais da zona cafeeira, delimitando e auxiliando a área de busca dos agentes de inspeção;
- Mapeamento do comportamento de consumo de energia elétrica dos clientes irrigantes de café, auxiliando a distribuidora na detecção de períodos de alto consumo para a rede rural de estudo;
- Análise da eficiência no manejo do sistema de irrigação para as lavouras de café estudadas, disponibilizando à concessionária, material para campanhas de conscientização de seus clientes que utilizam o benefício da tarifa verde.
- Apresentação de metodologia implementando dados agrometeorológicos de processamento de imagens.

1.5 ORGANIZAÇÃO DOS CAPÍTULOS

Este trabalho está organizado em cinco capítulos, referências bibliográficas e anexos. O capítulo 1 compõe a introdução quanto a temática do trabalho, apresentando uma visão geral sobre PNT no Brasil.

No capítulo 2 é apresentada o referencial teórico e pesquisa bibliográfica necessários para a elaboração do projeto descrito neste trabalho, começando numa breve contextualização de perdas no cenário mundo e Brasil, e o impacto das perdas nas tarifas de energia. São apresentados os métodos de detecção de PNT existentes na bibliografia atual e alguns trabalhos relacionados a temática abordada. No mesmo capítulo, é feita a caracterização do consumo de energia no Brasil, suas classes de consumo e subclasses, e comenta-se sobre a classe tarifaria que tem direito ao desconto de irrigantes, e como se dá o direito ao seu uso. Também é apresentada a pesquisa bibliográfica a respeito dos agricultores de café, descrevendo suas principais características, tanto quanto ao tipo de cultura, tipo de sistema irrigante e métodos de manejo. Ao fim deste capítulo é brevemente apresentado o conceito de agrometeorologia.

O capítulo 3 compõe a metodologia implementada para alcançar o objetivo deste trabalho, apresentando: os dados de entrada, essenciais ao projeto, seus atributos e como se deu o pré-processamento de cada um de maneira que pudessem ser utilizados pela metodologia; as limitações e dificuldades encontradas; a biblioteca e conceito dos algoritmos utilizados; por fim é apresentado a metodologia de estimativa de consumo.

No capítulo 4 são apresentados os estudos de casos e os resultados da implementação da metodologia desenvolvida de modelo preditivo classificatório e do método de estimativa de consumo para unidades consumidoras irrigantes de café.

O capítulo 5 é composto pelas conclusões e considerações obtidas neste trabalho, as sugestões de trabalhos futuros e as publicações desenvolvidas ao longo do mestrado.

Nos Anexos estão partes de código do algoritmo implementado e fluxograma da metodologia.

2. REVISÃO BIBLIOGRÁFICA

Neste capítulo, são contextualizados os referenciais teóricos fundamentais para o desenvolvimento do trabalho, incluindo o estudo de métodos de detecção de PNT encontrados na bibliografia. São contextualizados também, a modalidade de tarifação beneficiário para unidades consumidoras irrigantes, as principais variáveis meteorológicas para a agricultura, focando nas características das lavouras de café e suas influências no manejo do sistema irrigante.

2.1 PERDAS DE ENERGIA ELÉTRICA - CONTEXTUALIZAÇÃO

As perdas de energia elétrica referem-se à energia que foi gerada e passa pelas linhas de transmissão (rede básica) e redes da distribuição, mas não chega a ser comercializada, seja por motivos técnicos ou comerciais. O transporte de energia, resulta, inevitavelmente, tanto na rede básica quanto na rede de distribuição, em perdas técnicas relacionadas a transformação de energia elétrica em energia térmica nos condutores (efeito joule), perdas nos núcleos dos transformadores, perdas dielétricas etc. As perdas comerciais, ou PNT, decorrem principalmente pelas seguintes causas: furto, proveniente de ligação clandestina ou desvios diretos da rede; fraudes de energia, como adulteração no medidor; erros de medição e de faturamento (ANEEL, 2022).

As perdas técnicas na Distribuição (DT) são calculadas conforme as regras definidas no módulo 7 do Prodist. Resumidamente, o sistema de distribuição é dividido de acordo com os segmentos de rede (alta, média e baixa tensão), transformadores, ramais de ligações e medidores. Aplicam-se então os modelos específicos para cada um desses segmentos, utilizando-se informações simplificadas das redes e equipamentos existentes, como por exemplo, comprimento e bitola dos condutores, potência dos transformadores e energia fornecida às unidades consumidoras. Com base nessas informações, estima-se o percentual de perdas técnicas eficientes relativas à energia injetada na rede (ANEEL, 2022). Os cálculos de perdas técnicas são efetuados em cada segmento do sistema, de forma a permitir modelagem adequada e maior precisão nos resultados. Os segmentos de interesse para concessionárias de distribuição são: rede de alta tensão (69 a 138 kV) incluindo os ramais de subtransmissão, transformação AT/MT, rede primária (MT), transformador de distribuição (MT/BT), rede secundária, ramal de ligação e medidores de

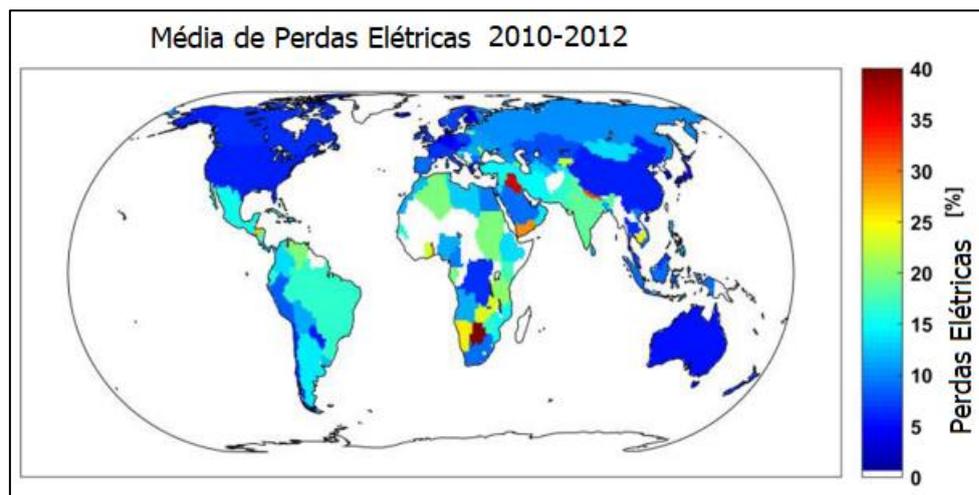
energia. Logo, as perdas técnicas podem ser estimadas utilizando os modelos matemáticos utilizados nos cálculos de fluxo de potência onde é necessário ter conhecimento das cargas e características do sistema elétrico (PENIN, 2008).

Segundo a regulamentação, os custos das perdas técnicas são considerados na tarifa de energia elétrica por serem inevitáveis em qualquer rede de distribuição no mundo, representando um custo para o setor elétrico. Os valores regulatórios das perdas técnicas das distribuidoras são calculados pela ANEEL, que observa os níveis eficientes dessas perdas conforme as características da rede elétrica de cada concessionária.

2.1.1 Perdas de energia: Cenário Global x Brasil

Realizando rápido levantamento de perdas de energia no cenário global, em 2016 as perdas de energia agregadas da rede básica e de distribuição chegaram à 19% na Índia e 16% no Brasil (WORLD BANK, 2022). Já no Haiti, Iraque e República do Congo as perdas chegaram a 50%, ou seja, de toda a energia gerada nesses três países, apenas metade foi de fato faturada. Em contrapartida, nos países mais desenvolvidos as perdas se mostram menores, a serem citados: os Estados Unidos tiveram perdas de 6% em 2016, a Alemanha teve 5% de perdas, e a Cingapura apresentou apenas 2%. De acordo com dados do site World Bank, as Perdas Elétricas, observadas em diversos países e economias nas redes de Transmissão e Distribuição (T&D), variam entre 3% e 70% do total da energia gerada. A Figura 1 ilustra a média no período 2010 – 2012 de perdas de energia no cenário global em relação com a energia total gerada, mostrando que as menores porcentagens ocorrem nos países mais desenvolvidos. O Brasil atingindo a marca entre 15% e 18%, na América do Sul, o país só perde para a Venezuela (20%) em termos de maiores índices de perdas elétricas na rede T&D (SADOVSKAIA, 2019).

Figura 1 – Perdas de Energia: média de 2010 – 2012.



Fonte: Adaptado de K.Sadovskaia et al. (2019).

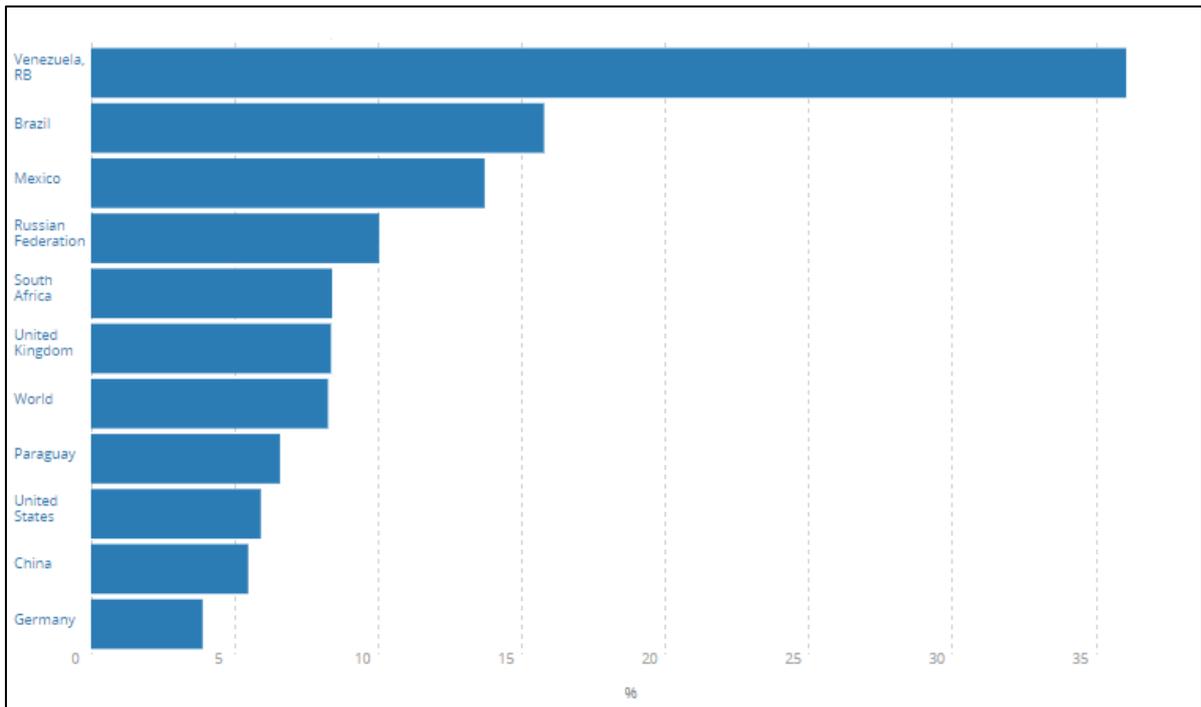
Com esses números pode-se observar que quanto menor é a distância a ser transmitida a energia mais eficiente se torna a entrega. Principalmente ao se considerar a distribuição de energia em longas distâncias para muitos clientes rurais dispersos. O que não é uma realidade para o Sistema Integrado do Brasil (SIN), onde as linhas de transmissão percorrem longas distâncias até os centros urbanos, além das distâncias do ponto da rede de distribuição até os consumidores rurais. Fazendo com que a rede apresente vários pontos susceptíveis a ocorrências de perdas por fraudes, furtos, além das perdas inerentes ao sistema (efeito joule).

Entre os parâmetros que descrevem as perdas de energia em T&D, no estudo analisando dados globais, Sadovskaia (2019) cita: (i) Densidade Populacional; (ii) Nível de Urbanização; (iii) Área do País; (iv) PIB (GDP) per capita; (v) Índice de percepção de corrupção; (vi) Temperatura. Destacando que os parâmetros que mais mostraram ligação direta com perdas de energia foram o PIB e o Índice de percepção de corrupção, já a Área do País não se mostra como um grande fator influenciador uma vez que a Rússia possui mais de 16,38 milhões de km² e, observando pelo gráfico da Figura 1, nota-se uma baixa porcentagem de perdas na rede T&D. Mais uma vez, diferente da realidade do Brasil que possui uma área de 8,51 milhões de km² e percentual de perdas maior de que a Rússia.

De acordo com os dados do cenário global mais recentes (2014), Figura 2, podemos ver que o comportamento observado nos dados de 2010-2012 se mantém para certos países. Enquanto o percentual de perdas nas redes de T&D no mundo foi de 8%, o Brasil apresentou o segundo maior índice (15,57%) dentro dos países da América do Sul,

atrás da Venezuela, enquanto países como Paraguai, Estados Unidos, China e Alemanha apresentaram índices menores do que o mundial.

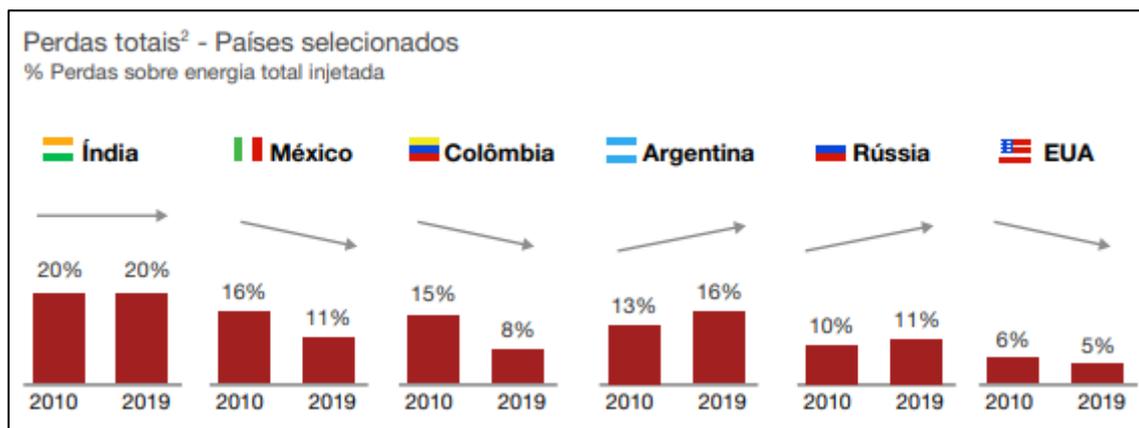
Figura 2 - Perdas nas redes T&D de energia elétrica (% da produção) – Mundo, Brasil etc. (2014).



Fonte: Adaptado de IEA *Statistics* (2022).

Segundo relatório da *Strategy&* (2020) o Brasil teve pouca melhora no índice de PNT reais na última década. Desde 2010, os índices reais se mantiveram próximos a 15%. No cenário global, alguns países, como México e Colômbia, evoluíram positivamente esta métrica no mesmo período, saindo de um índice parecido com o brasileiro em 2010 de 16% e 15% para 11% e 8%, respectivamente. Como referência de um país de alta renda, os Estados Unidos alcançaram em 2019 uma taxa de PNT de apenas 5%. A Figura 3 ilustra os índices de perdas dos países citados, e de alguns outros, mostrando a evolução de 2019 em comparação com o ano de 2010.

Figura 3 – Perdas Totais de Países.



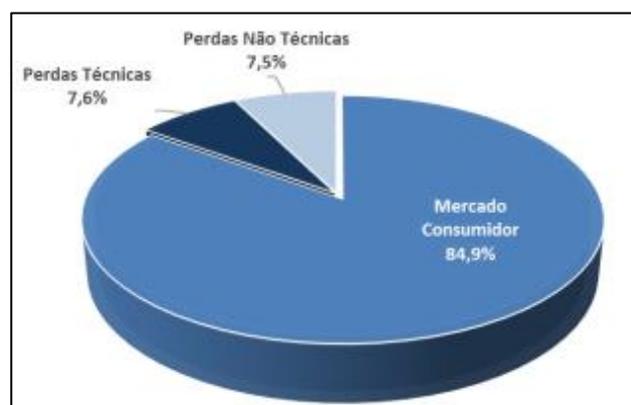
Fonte: Strategy& (2020).

2.1.2 Perdas de Energia Elétrica na rede de Distribuição no Brasil

As perdas totais na distribuição representaram aproximadamente 14,8% do mercado consumidor em 2020. Essas perdas representam mais do que o consumo de energia elétrica das regiões Norte e Centro-Oeste em 2018 (ANEEL, 2021a). Em montantes de energia, as perdas técnicas da distribuição corresponderam a cerca de 38,8 TWh e as perdas não técnicas 37,9 TWh em 2020. Nesse mesmo ano o Brasil apresentou total de energia injetada igual a 517,48 TWh.

A Figura 4 ilustra os percentuais de participação das perdas. Como comentado anteriormente, o repasse tarifário está previsto nos contratos de concessão (considerando os níveis eficientes das perdas), contudo as perdas reais das distribuidoras não são totalmente repassadas para a tarifa.

Figura 4 – Perdas sobre a Energia Injetada (2020).



Fonte: (ANEEL, 2021a.).

Ao comparar os valores das perdas do ano de 2020 com os anos anteriores, nota-se que nos últimos cinco anos esse valor vem crescendo, Figura 5, enquanto as perdas técnicas praticamente se mantêm constantes, a perdas não técnicas aumentaram.

Figura 5 - Perdas sobre a Energia Injetada (2008-2020).



Fonte: (ANEEL, 2021a).

Os custos dessas perdas técnicas, obtido ao multiplicar o montante pelo preço médio da energia nos processos tarifários, sem considerar tributos, é da ordem de R\$ 8,5 bilhões. Já os custos das perdas na rede básicas considerados na tarifa foram de aproximadamente R\$ 1,8 bilhão. Ambos os valores são receitas que poderiam ser utilizadas em investimentos na rede, melhorando a qualidade de energia e serviço, resultando também na redução no valor da tarifa para os consumidores. Ou seja, a perda de energia elétrica atualmente registrada não é um problema exclusivo das concessionárias do serviço público de eletricidade, mas de toda a sociedade, haja vista que de uma forma ou de outra os seus custos são transferidos via tarifa, subsídios etc. (VIEIRALVES, 2005).

2.1.3 Perdas Não Técnicas (PNT) no Brasil

As perdas não técnicas, também chamadas de perdas comerciais, é a energia que não é faturada pela distribuidora, apesar de ser injetada no sistema de distribuição. De acordo com Diaz (2015) as PNT estão diretamente ligadas à gestão comercial e ocorrem de duas formas distintas em cada distribuidora: perdas ligadas a variáveis que estão sob o controle das distribuidoras de energia, ou considerando variáveis que estão fora do controle das distribuidoras. Do ponto de vista das variáveis que estão sob controle das distribuidoras, destacam-se:

- Falta e erros de medição;
- Falhas no cadastramento;
- Erros no faturamento;
- Fraude interna;
- Erro no faturamento da iluminação pública.

Fora do controle da distribuidora, destacam-se:

- Fraudes e furtos de energia elétrica.

Segundo a Associação Brasileira de Distribuição de Energia Elétrica (ABRADEE) (2017), define-se furto e fraude como:

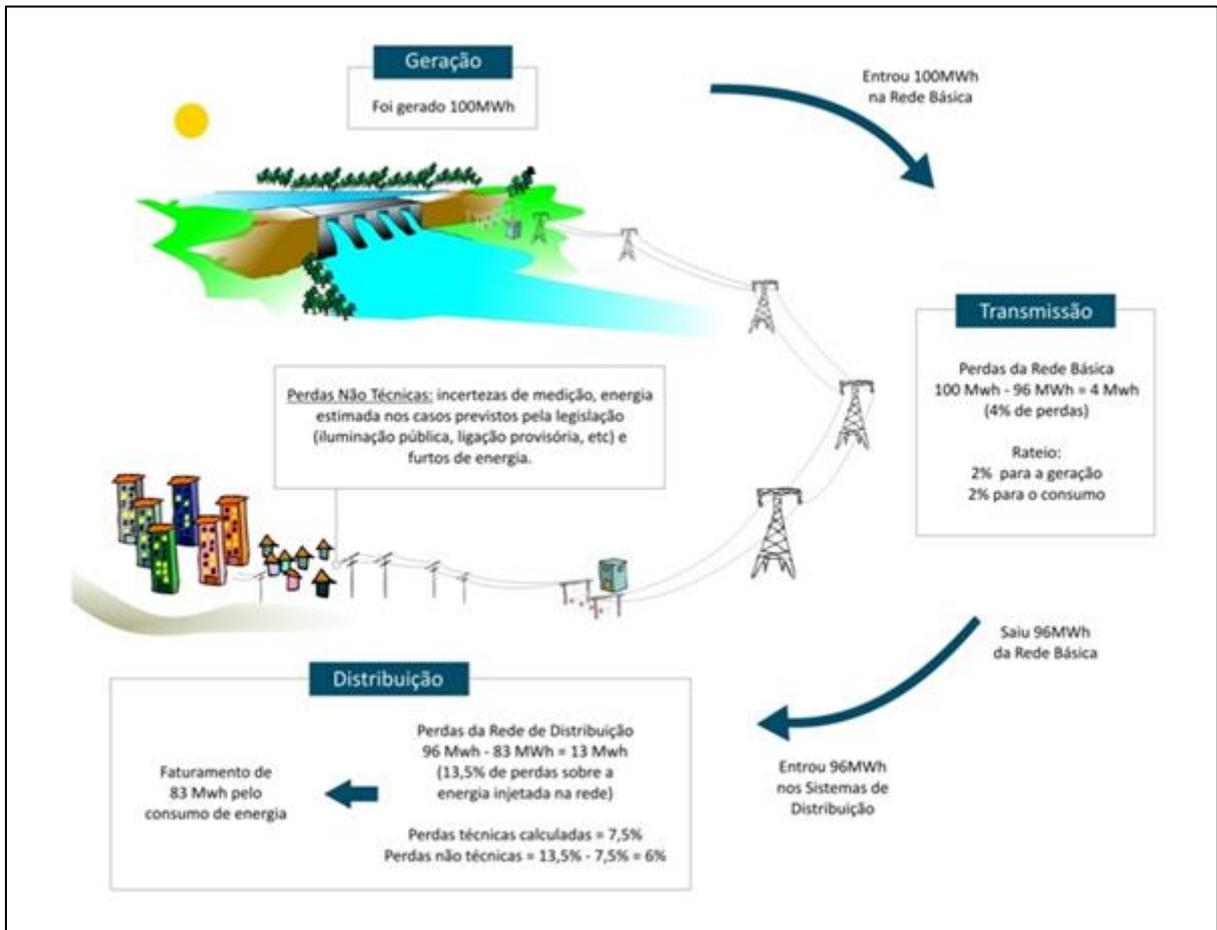
- Furto: pode ser definido como o desvio de energia elétrica do sistema de distribuição da concessionária de energia, realizado de forma ilegal para atender um consumidor ilícito.
- Fraude: caracteriza-se como uma alteração dos condutores de energia elétrica a fim de beneficiar o consumidor, legalmente registrado na distribuidora, para que este pague um valor menor do que o seu consumo.

Alguns exemplos de fraude são: (i) troca nas ligações da medição que fazem o disco girar para trás; (ii) seccionamento ou abertura da alimentação do potencial dos medidores (ou abertura dos terminais de prova); (iii) travamento ou bloqueio do disco do medidor; (iv) manuseio ou violação do registrador; (v) troca de engrenagens do registrador do medidor; (vi) inserção de diodos – eletrônica de controle da onda; (vii) bobina de potencial violada; (viii) lâmina da chave de aferição aberta; (ix) alargamento dos polos da chave de aferição; (x) interrompendo o contato elétrico; (xi) verniz isolante nos polos da chave de aferição.

As distribuidoras de maior porte são responsáveis por quase a totalidade dos montantes das PNTs no Brasil devido ao tamanho do mercado e à maior complexidade de se combater as perdas. Os níveis de PNT dependem da gestão das concessionárias, das características socioeconômicas (como apontado do estudo de Sadovskaia et. al. e como é realizado no cálculo de complexidade das distribuidoras realizado pela ANEEL) e de aspectos comportamentais existentes em cada área de concessão (ANEEL, 2020). A Figura 6 ilustra como as perdas ocorrem no sistema elétrico, onde no exemplo entra-se com 96MWh na

rede de distribuição, mas é faturada apenas 83MWh, ocorrendo o decréscimo pelas perdas técnicas e não técnicas.

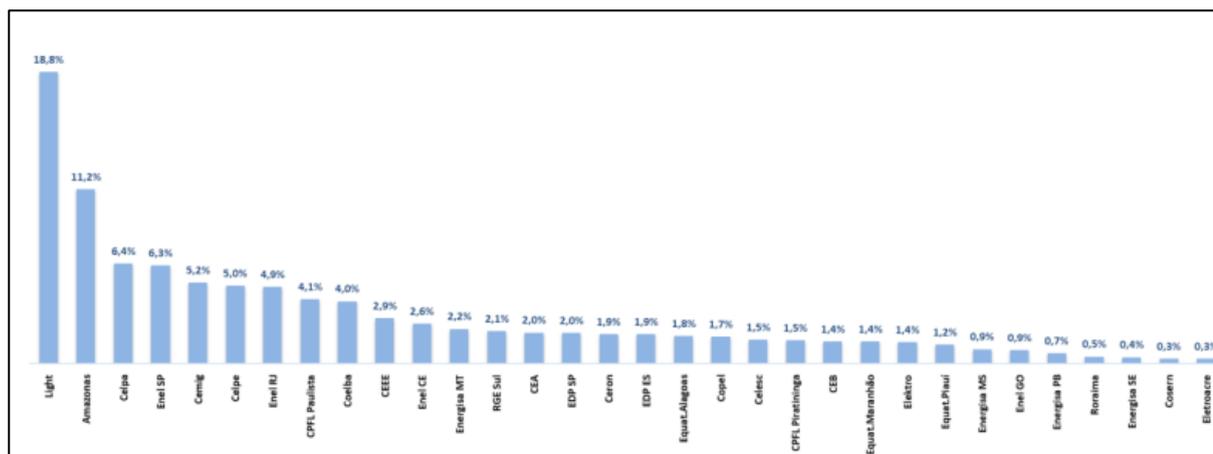
Figura 6 – Demonstração de Perdas no Setor Elétrico.



Fonte: ANEEL (2022).

O cenário regional de perdas não técnicas no Brasil é bastante desigual, onde a maior parte do percentual de PNT (70%) se concentrou em seis estados no ano de 2019, o equivalente ao montante de R\$ 4,7 bilhões. Já no ano de 2020, pelo relatório da ANEEL, 10 distribuidoras de energia corresponderam, juntas, à 68,9% das PNT do país. A Figura 7 apresenta a participação das concessionárias nas PNT observadas em 2020 em relação às PNT totais no Brasil.

Figura 7 – Participação de PNT Reais / PNT Brasil (2020).



Fonte: ANEEL (2021a)

2.1.4 Impacto das Perdas nas Tarifas de Energia Elétrica

No Brasil, a regulação do setor de distribuição de energia elétrica é atribuída à ANEEL. Para aplicar a regulação, desde a promulgação da Lei das Concessões de 1995, é utilizado o modelo de regulação econômica variante do modelo inglês de preço-teto, *Price Cap*, chamado de modelo de receita-teto, *Revenue Cap*. Neste modelo a ANEEL determina os preços a serem praticados pelas empresas que cobrem os custos de capital e de operação e manutenção considerados eficientes (ABRADEE, 2022). De maneira similar a que ocorre para as distribuidoras e transmissoras, os mecanismos de regulação das empresas são basicamente a revisão tarifária, que incide periodicamente a cada cinco anos, e o reajuste tarifário anual, que basicamente é uma correção monetária das tarifas praticadas, tais mecanismos dependem do contrato de concessão das empresas, de maneira que não são uniformes para todas as empresas.

Essencialmente, espera-se que o preço da energia, definido pela revisão tarifária, seja suficiente para arcar com os custos de operação e expansão de todos os elementos elétricos que compõem o sistema, desde a usina geradora até o ramal de ligação dos consumidores de baixa tensão. Além destes custos existem também os encargos e os impostos. Resumidamente, a tarifa de energia elétrica dos consumidores cativos (consumidores que só podem comprar energia de sua distribuidora local) é constituída por:

- Custos com a aquisição de energia elétrica;
- Custos relativos ao uso do sistema de distribuição;
- Custos relativos ao uso do sistema de transmissão;

- Perdas técnicas e não técnicas;
- Encargos diversos e impostos.

Os custos com a aquisição de energia são aqueles decorrentes da contratação de montantes de energia por meio dos leilões regulados. A empresa distribuidora compra uma quantidade de energia que considera suficiente para o atendimento do seu mercado cativo. Os custos com energia são alocados na chamada Tarifa de Energia (TE) e repassados integralmente aos consumidores. Os custos relativos ao uso do sistema de distribuição estão inseridos na Tarifa de Uso do Sistema de Distribuição (TUSD), como as despesas de capital e os custos de operação e manutenção das redes de distribuição. Muitos encargos setoriais também estão inseridos na TUSD, assim como os custos relativos do sistema de transmissão, que são arrecadados por meio da Tarifa de Uso do Sistema de Transmissão (TUST) (ABRADEE, 2021).

Cabe à ANEEL garantir aos consumidores o pagamento de uma tarifa justa pela energia fornecida e, ao mesmo tempo, preservar o equilíbrio econômico-financeiro das concessionárias, para que possa prestar o serviço com qualidade pactuada. Com isso, no cálculo da tarifa, são inclusos custos que a distribuidora tem, comentados anteriormente, que podem ser divididos em três custos distintos: Energia Gerada, Transporte de energia até as unidades consumidoras (UC), e encargos setoriais. Os encargos setoriais e tributos não são criados pela ANEEL, e são instituídos por lei.

Os custos relativos às perdas elétricas dividem-se em dois: perdas técnicas e perdas não técnicas. Todos os consumidores pagam pelas técnicas de energia ocasionadas pelo seu próprio consumo (perdas técnicas). Já referente às PNT, no Brasil, dependendo da área de concessão, os custos tocantes à PNT correspondem por boa parte do custo da energia. Existem métodos regulatórios que determinam qual o nível máximo de PNT que as distribuidoras podem repassar às tarifas, definido pela ANEEL.

Para fins de cálculo tarifário, os custos da distribuidora são classificados em dois tipos:

- Parcela A: Compra de Energia, transmissão e Encargos Setoriais; e
- Parcela B: Distribuição de Energia.

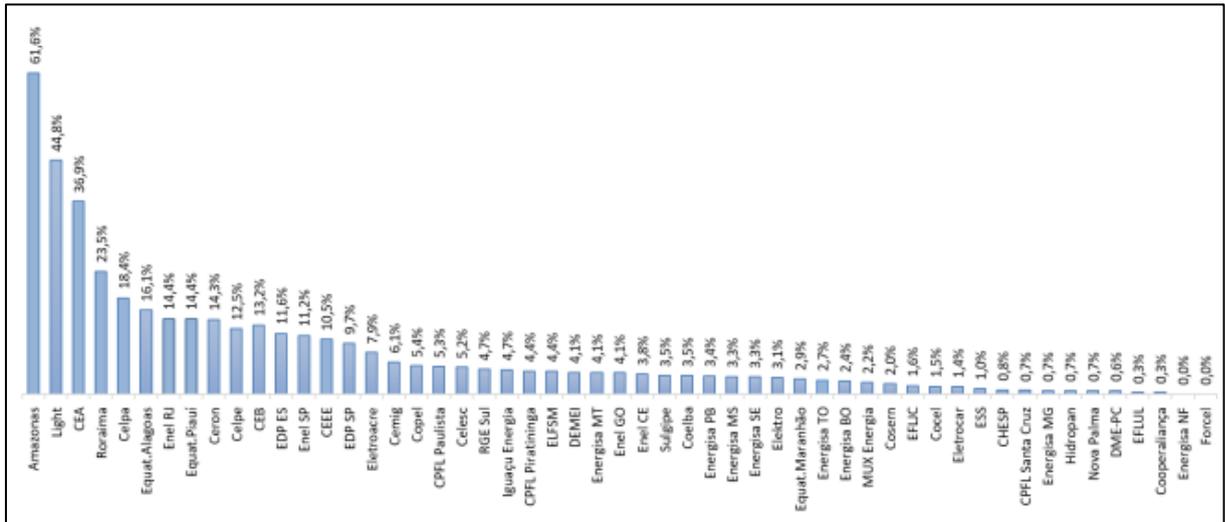
A parcela A é composta por custos não-gerenciáveis em que a empresa concessionária apenas cobra o consumidor final os valores necessários para ressarcir o

valor que foi gasto, podemos citar os custos de compra de energia, encargos setoriais e encargos de transmissão. A parcela B refere-se aos valores necessários à cobertura de pessoal, de material e outras atividades vinculadas diretamente à operação e manutenção dos serviços de distribuição, bem como os custos de depreciação dos investimentos realizados pela concessionária, são custos gerenciáveis. Dessa maneira os custos referentes às perdas estão na parcela B.

Como mencionado anteriormente, as PNT são apuradas pela diferença entre as e perdas totais e as perdas técnicas. As perdas técnicas são calculadas conforme as regras definidas no Módulo 7 do Prodist, e o limites regulatórios de PNT são calculados conforme as regras definidas no Submódulo 2.6 do Proret. Resumidamente, a ANEEL define os valores regulatórios das PNT por critérios de eficiência, normalmente esses valores são inferiores aos valores praticados pelas concessionárias de distribuição. Se a ANEEL observa ineficiência de gestão por parte das concessionárias, ela limita o repasse das PNT para a conta de energia (ANEEL, 2022). A Figura 8 apresenta, por concessionária os valores das perdas que constam nas tarifas e sua representação em relação à parcela destinada aos custos dos serviços de distribuição (parcela B).

A Tabela 1 apresenta o percentual de participação das concessionárias com maiores índices de PNT, somando 68,8% do total de PNT no Brasil (2020), presentes na Figura 7, juntamente com o percentual dos custos de PNT presente na tarifa do consumidor destas concessionárias, observa-se que 7 distribuidoras apresentam custos de PNT maiores do que 10%, sendo que a média nacional para o ano de 2020 foi de aproximadamente 9,8%, segundo relatório da ANEEL.

Figura 8 – Participação das PNT regulatória sobre a Parcela B da tarifa (2020).



Fonte: ANEEL (2021a).

Tabela 1 – Participação das PNT em relação ao Brasil e Participação sobre a Parcela B.

Distribuidora	% Part. PNT no Brasil (2020)	% Part. PNT sobre Parcela B
Light	18,80	44,8
Amazonas	11,20	61,6
Celipa	6,40	18,4
Enel SP	6,30	11,2
Cemig	5,20	6,1
Celpe	5,00	12,5
Enel RJ	4,90	14,4
CPFL Paulista	4,10	5,3
Coelba	4,00	3,5
CEEE	2,90	10,5

Fonte: Adaptado de ANEEL (2020).

A Tabela 2 apresenta as distribuidoras de energia com maior percentual de repasse de PNT à tarifa de energia para o consumidor final. Observa-se que as distribuidoras desta tabela atendem Brasília DF e os estados: Amazonas; São Paulo; Roraima; Pará; Alagoas; Rio de Janeiro; Piauí; Pernambuco; Rio Grande do Sul; Amapá; Rondônia; Espírito Santo.

Tabela 2- Distribuidoras com % de participação de PNT sobre a tarifa acima da média nacional.

Distribuidora	% Part. PNT sobre Parcela B
Amazonas	61,6
Light	44,8
CEA	36,9
Roraima	23,5
Celpe	18,4
Equat. Alagoas	16,1
Enel RJ	14,4
Equat. Piauí	14,4
Ceron	14,3
Celpe	12,5
CEB	13,2
EDP ES	11,6
Enel SP	11,2
CEEE	10,5
EDP SP	9,7

Fonte: Adaptado da ANEEL (2020).

2.2 MÉTODOS DE DETECÇÃO DE PNT

Os métodos de detecção de PNT podem ser divididos em duas classes principais: métodos indiretos e métodos diretos. Os métodos indiretos buscam estimar as perdas técnicas e obter as perdas não técnicas através da diferença entre o total de perdas e o estimado de perdas técnicas, enquanto os métodos diretos buscam detectar diretamente as PNT (FERREIRA et al. 2014).

2.2.1 Métodos Indiretos

De acordo com Ferreira (2015), os métodos indiretos são aqueles que calculam as PNT a partir da estimação das perdas técnicas. A eficiência dos métodos indiretos está relacionada ao grau de conhecimento do sistema. Métodos com baixo grau de eficiência fazem pouco uso do conhecimento da rede, ou seja, as perdas técnicas são estimadas a partir de medidas brutas disponíveis, dado de equipamentos e comparação com outros sistemas. Por outro lado, métodos com alto grau de conhecimento do sistema exploram dados, via aplicativos como fluxo de potência, e oferecem maiores possibilidades de se

identificar perdas com grande precisão. Outra maneira de analisar estes dados é pelo caso híbrido, neste modelo, quando se possui alto grau de conhecimento de apenas parte da rede utilizam-se outras metodologias.

2.2.2 Métodos Diretos

Os métodos diretos podem utilizar dados históricos a fim de definir padrões de consumo e detectar comportamentos anômalos, ou utilizar dados em tempo real ou quase tempo real para detectar a ocorrência de perdas não técnicas no sistema (FERREIRA et. al., 2014). Ferreira (2015) e Curado (2015) dividem estes métodos em duas classes: métodos que utilizam dados históricos, e métodos que utilizam dados obtidos em tempo real.

- Métodos que utilizam dados históricos:

Os métodos que utilizam dados históricos, de maneira geral empregam uma base de dados e geram um padrão de informações que serve para identificar possíveis fraudes de energia. A etapa inicial de preparação dos dados pode utilizar diferentes técnicas, tais como utilização de redes neurais artificiais ou *clustering* (RODRIGUES, 2018). Métodos que utilizam dados históricos possuem estrutura similar entre si, sendo divididos em três etapas: preparação, classificação e averiguação. A preparação consiste na criação de perfis de consumo de clientes similares levando em conta diversos fatores como dia da semana, temperatura, época do ano, região, entre outros. Essa etapa varia de acordo com o método utilizado, mas em geral tem como desvantagens o grande número de dados a serem agrupados e processados. Na classificação utilizam-se técnicas para comparar os perfis de consumo e identificar perfis irregulares que podem estar associados a consumidores fraudulentos. Por fim, na averiguação os consumidores classificados como suspeitos na etapa de classificação são examinados a fim de detectar o problema e tomar as medidas corretivas necessárias (FERREIRA et. al., 2014).

- Métodos que utilizam dados obtidos em tempo real:

Esses métodos não utilizam dados históricos, são caracterizados por utilizar as medidas obtidas da rede e, utilizando teorias de circuitos elétricos,

detectar PNT. Por exemplo, balanço de energia, técnicas de ondas viajantes e estimação de estado (RODRIGUES, 2018).

Considerando os dois métodos apresentados, esse trabalho abrange a técnica de métodos diretos utilizando dados históricos.

Muitas empresas têm elaborado seus próprios métodos para localizar os pontos críticos de PNT em seus sistemas, e assim definir o plano de ação a ser seguido para a redução dessas perdas. Um dos métodos adotados pelas empresas tem sido o investimento em maneiras cada vez mais eficazes de cadastro de clientes e na regularização desses consumidores (OLIVEIRA, 2009).

2.3 TRABALHOS RELACIONADOS À DETECÇÃO E CÁLCULO DE PNT

Existem vários trabalhos que se propõem a fazer detecção de PNT na distribuição de energia elétrica, que tem ganhado maior enfoque com o aumento de recursos tecnológicos e computacionais disponíveis. Nesta seção são apresentadas algumas técnicas utilizadas para detecção de PNT encontradas na literatura.

Utilizando a metodologia que usa do conhecimento das perdas globais e técnicas, Oliveira et. al. (2001) apresentam uma metodologia para cálculo de perdas técnicas nos sistemas de distribuição, e utilizando do balanço de energia, sugere-se a possibilidade de se determinar PNT. Méffe (2002), propõe uma metodologia que considera a utilização de dados de medições na rede para reajustar uniformemente as curvas de cargas típicas dos alimentadores, calculando as perdas técnicas com maior precisão. Celik (1999) destaca a importância de se considerar as características reais dos alimentadores, resultando num cálculo de fluxo de potência mais confiável. O estudo de Celik permite a obtenção de curvas de perdas ao longo de um determinado período e para um determinado componente, aumentando a precisão da obtenção de índices de perdas técnicas.

Outras metodologias utilizadas para detectar PNT utilizam de estimadores de estado das redes de distribuição, Méffe (2002), Cruz, Quintero e Pérez (2006), Donadel et. al. (2009), visando obter um modelo da rede mais aproximado da realidade. Ferreira Neto (2011) utiliza esse método combinado com métodos estatísticos, destacando o método dos mínimos quadrados ponderados, introduzindo pseudo medidas a partir de dados de faturamento dos consumidores, para estimar o estado da rede em um sistema de detecção

de PNT. Outro exemplo de técnicas estatísticas, Zarco P e Expósito (2001), utilizam o estimador de estado de mínimo erro absoluto ponderado e da mínima mediana ponderada de quadrado dos resíduos.

Alguns estudos mais recentes apresentam métodos de aperfeiçoar o sistema de distribuição, modernizar a atualizar os sistemas de medição, de maneira a diminuir os custos dom PNT, a serem citados: Yurtseven (2015), Never (2015), Lewis (2015), Chaunan (2015), Giraldo (2015), Bezerra et. al. (2018). Dentre os métodos utilizados estão análises de pontos específicos da rede de distribuição ou de regiões, como no estudo de Faria et. al (2014) e Madrigal et. al. Já o estudo de Madrigal et. al. (2017), apresenta uma metodologia que permite estimar o número de furtos e fraudes em uma distribuidora, assim determinando o montante de PNT, fazendo uso da análise de regressão linear múltipla, encontrando modelos matemáticos que descrevem o comportamento de roubo e fraude diante de determinadas variáveis, permitindo estimar em quais setores há maior probabilidade dessas anomalias.

Os estudos citados acima tratam da detecção do montante de PNT, alguns estudos destacam a detecção de unidades fraudulentas associadas a ocorrência de PNT, como os estudos de Faria et. al., onde se incorpora o espaço geográfico à análise do problema de perdas comerciais (PNT), essas perdas são modeladas como eventos pontuais considerando variáveis socioeconômicas que auxiliem na explicação da disposição espacial das mesmas. Tais variáveis são utilizadas em um modelo aditivo generalizado associado a um estudo casos-controles que realiza a comparação entre dois processos espaciais pontuais. Um deles é representado pelo conjunto dos casos que são as Unidades Consumidoras (UCs) flagradas em irregularidades. O outro processo é representado pelo conjunto de controles que é composto por UCs regulares obtidas aleatoriamente a partir da base de clientes da concessionária. O modelo possui a vantagem de permitir a estimação da distribuição espacial das perdas comerciais utilizando covariáveis socioeconômicas que caracterizam as sub-regiões em que ocorrem essas perdas.

Silva e Scarpel (2007), propõem a utilização do algoritmo de Máquina de Vetores de Suporte (SVM), para a detecção de fraudes na distribuição de energia elétrica, baseado nos dados fornecidos por uma empresa do ramo. Com objetivo de criar um modelo de classificação que possa rotular os clientes da empresa cujas UCs foram estudadas, classificando-as como honestas ou fraudulentas; Todesco et. al. (2007), propõem o desenvolvimento de uma aplicação para identificação de UCs fraudadoras, no sentido de melhorar o processo de inspeção de consumidores apontados como suspeitos, esse estudo

foi realizado com consumidores residenciais e industriais; Ferreira (2008), apresenta um estudo de utilização de aprendizado de máquina para identificação de PNT sobre os dados de inspeções de uma concessionária, ao aplicar quatro tipos diferentes de técnicas de aprendizado de máquina sobre diferentes conjuntos de dados comparando os resultados para selecionar a melhor estratégia, os algoritmos utilizados foram o SVM (*Support Vector Machine*), Redes Neurais Artificiais, Árvore de decisão (C4.5) e *Näive Bayes*.

Outros estudos fazem uso de técnicas de inteligência artificial para detecção de PNT: Tehrani et. al. (2012), apresenta um modelo de detecção de perdas não técnicas baseado nos métodos de *Decision Tree*, *Random Forest* e *Gradient Boosting*; Monteiro e Maciel (2018), apresentam técnicas com o objetivo de executar a tarefa de agrupamento de dados, utilizando os algoritmos de clusterização, *K-Means* e *K-Medoids* 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.

Sobre estudos focados em consumidores de áreas rurais, cita-se: Evaldt (2018), que apresentada uma proposta de identificação de perdas não técnicas em alimentadores rurais contendo estações de bombeamento para irrigação de lavouras de arroz utilizando Redes Neurais, utilizando como entradas as correlações dos padrões de consumo de energia elétrica, juntamente com as características da área irrigada e condições climáticas do período de irrigação. Para os estudos de casos, foram considerados um conjunto de UCs 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; Diaz (2015) apresenta o estudo de perdas elétricas em média e baixa tensão em cooperativas que compõem o sistema da FECOERUSC, propondo uma nova metodologia baseada em teoria de fluxo de potência capaz de classificar e estimar com maior precisão as perdas técnicas e a existência de PNT nas redes das cooperativas de eletrificação rural. Nesse estudo utiliza-se um sistema de medição de dois níveis, comparando a energia consumida pelo cliente final com a que está sendo injetada no transformador de distribuição; Eichkoff (2021) é proposta uma metodologia para detectar possíveis ocorrências de PNT em UCs alocadas em áreas rurais com cultivo de arroz irrigado, para isso, relacionando o consumo de energia elétrica com as características das lavouras e dos sistemas de irrigação, empregando técnicas de agrupamento e classificação de dados.

Pela literatura estudada observa-se que a estimação das PNT a partir da diferença entre perdas globais e técnicas é uma maneira clássica de se obter um indicativo da

ocorrência do problema na rede, contudo diversos estudos estão sendo desenvolvidos utilizando técnicas de inteligência artificial, com diferentes algoritmos, e diferentes opções de conjunto de dados, alguns classificados como fraudulentos, outros sem a previa identificação. O mais usual de ser encontrados na literatura são conjuntos de dados com UCs residenciais e, ou, comerciais, sendo o estudo com consumidores rurais, em específico, ainda pouco abordado.

2.4 CARACTERIZAÇÃO DO CONSUMO DE ENERGIA NO BRASIL

Analisando o consumo de energia elétrica no Brasil para o ano de 2020, como já mencionado, o consumo total foi de 475 TWh, 1,4% menor do que o consumo do ano anterior. A distribuição regional mostra-se, ainda, concentrada no Sudeste, mas vem se modificando dentre as outras regiões. No geral, analisando os consumos de cada região do Brasil para o período de 2012 a 2020, a região Sudeste apresenta média de participação de 50,5%, as regiões Sul, Nordeste, Centro-Oeste e Norte, apresentam média de participação de: 17,9%; 17,1%; 7,5% e 7%, respectivamente.

O consumo de energia por classe apresentou redução, em relação a 2019, em algumas das classes de consumo. As classes de consumo eram definidas na Resolução Normativa ANEEL n° 414, de 2010, sendo essa revogada recentemente pela Resolução Normativa ANEEL n° 1.000, de 7 de dezembro de 2021. No que se define quais as classes de consumo, a resolução de n° 1.000, define que a distribuidora deve classificar a unidade consumidora para fins de aplicação tarifária de acordo com a atividade comprovadamente exercida, a finalidade de utilização da energia elétrica e o atendimento aos critérios dispostos e na legislação, em uma das seguintes classes tarifárias:

- i. Residencial;
- ii. Industrial;
- iii. Comércio, serviços e outras atividades;
- iv. Rural;
- v. Poder público;
- vi. Serviço público, e;
- vii. Consumo próprio.

As classes que apresentaram redução de consumo no ano de 2020, foram: Industrial, com redução de 0,8%; Comercial, com redução de -10,4%; Poder Público,

redução de -19%; Iluminação pública. Redução de -2,4%, e; Consumo próprio, com redução de -3,7%. Já as classes de consumo Residencial e Serviço Público apresentaram crescimento de 3,8% e 2,4%, respectivamente. A classe de consumo Rural apresentou o maior crescimento de consumo para o ano de 2020, comparado com 2019, com crescimento de 7,1% (ANEEL, 2021a).

A distribuição do consumo de energia para o ano de 2020 é apresentada na

Figura 9. Observa-se que os maiores consumos foram das classes Industrial e Residencial, seguidas pela Comercial. O consumo de energia da classe Rural foi de 6,5%, menor do que as classes citadas, porém apresentando o crescimento mencionado.

Figura 9 – Distribuição do Consumo de Energia Elétrica por Classe de Consumo, ano de 2020.



Fonte: ANEEL, 2021a.

2.4.1 Classes de Consumo e suas Subclasses

As classes de consumo apresentam subclasses definidas pela ANEEL, que são essenciais para aplicação de tarifas, subsídios ou benefícios para as unidades consumidoras. A seguir são apresentadas as subclasses de cada classe de consumo, segundo a resolução normativa N° 1.000 da ANEEL.

- Classe Residencial
 - i. Residencial;
 - ii. Residencial baixa renda;

- iii. Residencial baixa renda indígena;
- iv. Residencial baixa renda quilombola;
- v. Residencial baixa renda benefício de prestação continuada da assistência social; e
- vi. Residencial baixa renda multifamiliar.
- Industrial
 - i. Industrial; e
 - ii. Transporte de resultantes do seu processamento, caracterizado como atividade de suporte e sem fim econômico.
- Comercial
 - i. Comercial
 - ii. Serviços de transporte, exceto tração elétrica;
 - iii. Serviço de comunicações e telecomunicações;
 - iv. Associações e entidades filantrópicas;
 - v. Templos religiosos;
 - vi. Administração condominial: instalações de uso coletivo de prédio ou conjunto de edificações, incluindo a iluminação das vias internas;
 - vii. Iluminação em vias: solicitada por quem detenha concessão ou autorização para administração de vias de titularidade da União ou dos Estados;
 - viii. Semáforos, radares e câmeras de monitoramento de trânsito, solicitados por quem detenha concessão ou autorização para controle de trânsito; e
 - ix. Outros serviços e atividades.
- Poder Público
 - i. Poder público federal;
 - ii. Poder público estadual ou distrital; e
 - iii. Poder público municipal.
- Iluminação Pública
- Serviço Público
 - i. Água, esgoto e saneamento; e
 - ii. Tração elétrica.

- Consumo Próprio
 - i. Estação de recarga de veículos elétricos; e
 - ii. Outras atividades.
- Rural
 - i. Agropecuária rural;
 - ii. Agropecuária urbana;
 - iii. Residencial rural;
 - iv. Cooperativa de eletrificação rural
 - v. Agroindustrial;
 - vi. Serviço público de irrigação rural;
 - vii. Escola agrotécnica;
 - viii. Aquicultura;

2.4.2 Classe de Consumidores Rurais e Atividades de Irrigação

De acordo com o Art. 184. Da resolução N° 1.000 da ANEEL, devem ser classificadas na classe rural unidades consumidoras em que sejam desenvolvidas as atividades dispostas nas seguintes subclasses apresentadas no item anterior, segundo as atividades dispostas, apresentadas na resolução. Das atividades descritas as que incluem o fornecimento de energia elétrica para serviço de bombeamento de água destinada à atividade de irrigação, são as subclasses: Aquicultura, Agropecuária rural; Agropecuária urbana; e Serviço Público de Irrigação (irrigação em área rural, sem fins lucrativos, explorado por entidade pertencente ou vinculada à administração direta, indireta ou fundações de direito público da União, dos Estados, Distrito Federal ou Municípios).

Para a unidade consumidora classificada na classe rural aplicam-se as tarifas homologadas para o grupo A, já para o grupo B, as tarifas homologadas do subgrupo B2. A unidade consumidora da classe rural tem direito ao benefício tarifário de redução nas tarifas aplicáveis ao consumo (TUSD em R\$/MWh e TE em R\$/MWh) destinado às atividades de irrigação e de aquicultura desenvolvidas em um período diário contínuo de 8 horas e 30 minutos (ANEEL, 2021), horário reservado. O benefício tarifário depende da comprovação pelo consumidor da existência do licenciamento ambiental e da outorga do direito de uso de recursos hídricos, quando exigido em legislação federal, estadual, distrital ou municipal. Sendo assim, a aplicação dos benefícios tarifários se destina apenas para as seguintes cargas:

- I. Aquicultura: cargas utilizadas no bombeamento para captação de água e nos tanques de criação, no berçário, na aeração, e na iluminação; e
- II. Irrigação: cargas utilizadas no bombeamento para captação de água e adução, na injeção de fertilizantes na linha de irrigação, na aplicação da água no solo pelo uso de técnicas específicas e na iluminação dos locais de instalação desses equipamentos.

Desse modo, o estudo apresentado neste trabalho irá envolver unidades consumidoras da classe rural, e subclasses agropecuária rural e urbana classificadas pela distribuidora de energia como pertencentes a modalidade tarifária Irrigação – período reservado.

2.4.3 Desconto na Tarifa de Unidades Consumidoras Rurais Irrigantes

O benefício de horário reservado é um desconto aplicado às unidades consumidoras (UC) do grupo A, das subclasses rural irrigante ou aquicultura, que busca incentivar o uso de tecnologias de irrigação (ou de aquicultura) ao mesmo tempo em que promove o consumo de energia elétrica nos horários de menor demanda, como o período noturno. Enquanto as UC buscam reduzir custos operacionais do seu sistema e aumentar a produtividade, o sistema torna mais tênue a sazonalidade diária da carga total. Desse modo, o horário reservado é o período 8 horas e 30 minutos do dia que abrange toda a madrugada, em que a carga destinada à irrigação ou aquicultura recebe um desconto na tarifa de acordo com a região em que se localiza e o grupo tarifário a que pertence, segundo regulamentação REN nº 1.000/2022 (ANEEL, 2022).

O desconto de irrigação no período de horário reservado (irrigante noturno) é calculado a partir da tarifa (TUSD + TE + Bandeiras Tarifárias) sem impostos multiplicada pelo consumo em horário reservado. Essa conta resulta em uma “base de cálculo”. A segunda variável que define a porcentagem de desconto aplicada sobre a base de cálculo depende do local em que a operação de irrigação acontece e a tensão de fornecimento. A região de estudo, municípios de São Paulo, apresentam redução de 60% na tarifa para o Grupo B e de 70% para o Grupo A.

Dentre as vantagens do manejo do sistema de irrigação no período noturno podemos citar, além do desconto concedido: (i) menores desvios de água devido a menor velocidade do vento; (ii) umidade relativa do ar mais alta o que diminui a perda de água;

(iii) como não há sol no período reservado tem-se diminuição da perda de água por evaporação.

2.5 LAVOURAS DE CAFÉ

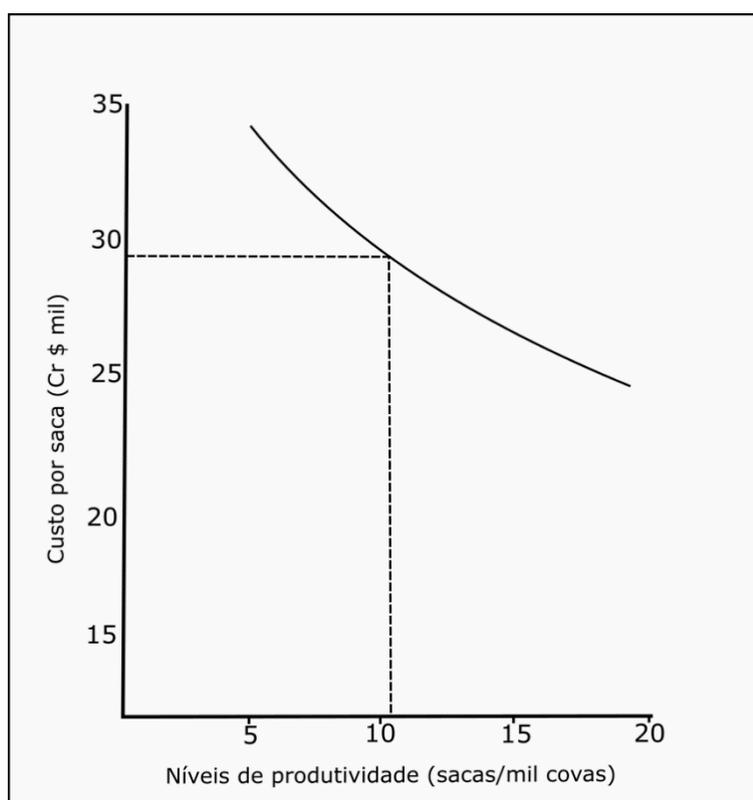
No estado de São Paulo o cafeeiro teria sido implementado em 1790, através do Vale do Paraíba, vindo do Rio de Janeiro. Caminhando para oeste, a cultura cafeeira, depois de conquistar todo o Vale do Paraíba até as vizinhanças de São Paulo, passou para Campinas, que se tornou então grande centro de irradiação: daí partiu o café para desbravar sertões nas zonas Mogiana, Paulista, Sorocabana e Araraquarense (GRANER; GODOY JÚNIOR, 1967). O gênero *Coffea* compreende várias espécies heterogêneas, entre as quais se destacam as espécies *C. arábica* (café arábica) e *C. canéfora* (café robusta), que são as únicas de maior cotação no mercado internacional.

O café arábica, com qualidade superior à do robusta, é produzido principalmente no continente americano, no Quênia e em alguns outros países de menor expressão econômica, enquanto o Robusta é proveniente principalmente da Costa do Marfim, Uganda, Java, União Indiana e Angola. O café plantado no Brasil é em sua maior parte arábica, existindo poucas áreas com o café robusta (Espírito Santo, norte do Rio de Janeiro, Vale do Rio Doce em Minas Gerais, Mato Grosso, Rondônia, sul da Bahia e Acre) (POTAFOS, 1986). A Figura 10 apresenta o mapa das zonas cafeeiras no Brasil, percebe-se que a maior concentração de plantações de café arábica se encontra no Sudoeste: Minas Gerais e São Paulo.

lavoura estar muito ou pouco produtiva. Por sua vez, os gastos variáveis, com fertilizantes, defensivos e outros insumos, e as práticas como a irrigação, influenciam bastante na produtividade e, por isso, não podem ser descuidados (MATIELLO, 2021).

Desse modo, a unidade (saca) produzida é mais inflacionada nas lavouras com produtividade baixa, como pode ser observado no gráfico de custos de produção de café, apresentado na Figura 11.

Figura 11 – Representação Gráfica da estimativa de custos de produção de café.



Fonte: Adaptado de POTAFOS (1986).

Com a Figura 11, podemos destacar que embora as despesas totais cresçam nas propriedades com níveis de produtividade mais altos, o custo por saca produzida decai com o aumento da produtividade, logo os cafeicultores buscam alcançar o ponto no gráfico em que os investimentos realizados na lavoura tenham retorno no valor da saca.

Dentre os fatores que influenciam sobre a produtividade dos cafezais no Brasil podem ser reunidos em 3 categorias principais (POTAFOS, 1986):

- i. Fatores econômicos-conjunturais: A conjuntura econômica determina condições favoráveis ou desfavoráveis em relação aos preços do café, à disponibilidade e custo dos insumos e da mão-de-obra, ao custo e à

disponibilidade de crédito etc. O balanço destes fatores influi sobre a tomada de decisão dos produtores, que passam a adotar em maior ou menor grau os investimentos nos tratos das lavouras.

- ii. Fatores climáticos: Quanto as condições climáticas, destaca-se o efeito das geadas, secas e veranicos, com ação bastante conhecida no meio cafeeiro. Estes fatores têm mostra mostrado ação mais marcante na determinação da produtividade da lavoura cafeeira, seja a nível regional ou nível global. O ciclo bienal de produção do cafeeiro no Brasil é um fenômeno muito importante que atua sobre a produtividade em determinados anos. Ele ocorre principalmente em função do cultivo das lavouras a pleno sol, que condiciona altas produções num ano, com o conseqüente esgotamento da planta, que assim não tem boa vegetação para voltar a produzir bem no ano seguinte. O ciclo bienal é mais pronunciado em lavouras malnutridas, maltratadas, sendo comum em áreas de “cerrado”, onde em casos graves, a lavoura leva de 2 a 3 anos para se recuperar após uma boa produção.
- iii. Manejo da cultura: Dentre os fatores de ordem técnica pode-se enumerar como os principais influentes na produtividade cafeeira: problemas de solo, sistema de cultivo, cultivares, espaçamento, adubação e calagem, manejo da plantação (práticas culturais), combate a pragas e doenças, proteção contra ventos e irrigação.

A cafeicultura é uma atividade cara, de alto custo de implantação, com retornos que acontecem três a quatro anos após o plantio. Em todo o país, há variações nos modelos de produção que vão desde as lavouras tradicionais até as propriedades tecnificadas, verdadeiras empresas que trabalham com técnicas de alto nível, dispondo de ferramentas da administração e da informática para estarem ligadas ao mercado e às bolsas (EPAMIG, 2008). Mais especificamente, o Estado de São Paulo, região da aplicação deste estudo, é o segundo maior produtor de café arábica do país, com produção esperado para o ano de 2021, de 4.007 milhões de sacas da espécie, segundo relatório da CONAB (2021).

O relatório da CONAB, descreve a safra paulista de café em 2021 com expectativa prevista de redução no volume obtido em comparação ao exercício passado, de modo que os efeitos da bienalidade negativa já eram esperados, perfazendo menor rendimento nas lavouras. No entanto, a diminuição foi acima do esperado também pelas intempéries climáticas registradas ao longo do ciclo, com presença marcante de períodos de baixa

incidência de chuvas, pouca disponibilidade hídrica nos solos, aumento das temperaturas provocando maiores taxas de evapotranspiração e até ocorrência de geadas em situações pontuais. Com isso, percebe-se que certas variáveis meteorológicas influenciam na produtividade do café.

O desenvolvimento da cultura do café e seus fatores relativos à produção, são afetados, dentre outras coisas, pelos elementos climáticos ocorridos durante as diferentes fases fenológicas da cultura. Dessa maneira, a ocorrência e distribuição das chuvas, bem como a ocorrência de extremos de temperaturas máxima e mínima, além da umidade do ar, vento e radiação solar, afetam o desenvolvimento da cultura (CAMARGO, 1993).

Assim como o desenvolvimento vegetativo e reprodutivo, a incidência das principais pragas e doenças na cafeicultura também é influenciada por fatores climáticos, principalmente a temperatura e precipitação. Outros fatores como manejo da lavoura, nutrição e carga pendente também são importantes para a ocorrência de pragas e doenças no cafeeiro. Dessa maneira, o monitoramento das condições climáticas, associado a levantamentos de incidência de pragas e doenças, pode ser uma importante ferramenta de auxílio visando tomada de decisão em lavouras cafeeiras e minimizar prejuízos decorrentes das condições meteorológicas adversas (CAMARGO, 1993).

A região cafeeira de São Paulo é uma região em que foi-se buscando o crescimento na produtividade investindo em tecnologia, manejo e maquinário, implementando ações como: o a eliminação de cafezais improdutivos, o emprego do plantio adensado, a condução com poda (safra zero), nutrição pautada pelas análises de solo e folha, implantação de cafeeiros adaptados à mecanização e o uso da irrigação em áreas com déficit hídrico – fazem a diferença na colheita. Um fato que certamente contribuiu para os resultados atuais foi a mecanização (TOLEDO FILHO, 2013).

2.5.1 Fenologia do Café

A fenologia pode ser definida como o estudo dos eventos periódicos da vida da planta em função da sua reação às condições do ambiente. Na cultura do cafeeiro arábica, nas condições de cultivo da maioria das regiões do Brasil, podem ser distinguidas as fases preparativa e construtiva durante seu ciclo fenológico. No período seco, associado a baixas temperaturas, ocorre a fase preparativa, que não se manifesta claramente por caracteres externos. Durante o período quente e chuvoso predomina a fase construtiva, onde se manifestam as atividades de crescimento dos ramos, folhas, gemas, flores e frutos.

Durante a fase construtiva existe uma competição entre o crescimento vegetativo e o processo de frutificação, evidenciada pelos anos alternados de grandes e pequenas produções (CAMARGO; CAMARGO, 2001). Essas são informações essenciais para determinação do período de irrigação do cafeeiro, em que espera-se que os agricultores de café utilizem seus sistemas irrigantes para atender as necessidades hídricas da sua área de plantio.

Diversas formas de definir e esquematizar a sequência dos estádios fenológicos do cafeeiro arábica já foram propostas (CAMARGO, 1998). Um modelo mais simples e racional foi apresentado por Camargo e Camargo (2001), válido para as cultivares de café Catuaí e Mundo Novo (ambos são café arábica). Nesse modelo, a fenologia do cafeeiro arábica foi definida e esquematizada para as condições tropicais do Brasil e está relacionada com as condições Agrometeorológicas de cada ano. A esquematização das diferentes fases fenológicas do cafeeiro permite identificar quais delas são mais exigentes em água, facilmente disponível no solo, e as que necessitam da ocorrência de um pequeno estresse hídrico para condicionar uma florada abundante (CAMARGO; CAMARGO, 2001; CAMARGO et al., 2001).

Na Figura 12 é apresentado um esquema das fases fenológicas do café arábica, segundo Camargo e Camargo (2001). O ciclo fenológico, para as condições tropicais no Brasil, foi subdividido em seis fases bem distintas, sendo duas vegetativas e quatro reprodutivas, a saber: (i) vegetação e formação de gemas foliares; (ii) indução e maturação das gemas florais; (iii) florada; (iv) granação dos frutos; (v) maturação dos frutos; e (vi) repouso e senescência dos ramos terciários e quaternários.

Figura 12 – Período Fenológico do Café arábica, abrangendo as seis fases, durante 24 meses.

Ano 1											
Período vegetativo											
Set.	Out.	Nov.	Dez.	Jan.	Fev.	Mar.	Abr.	Maió	Jun.	Jul.	Ago.
Vegetação e formação das gemas florais						Indução e maturação das gemas florais					
											Repouso
Ano 2											
Período reprodutivo											
Set.	Out.	Nov.	Dez.	Jan.	Fev.	Mar.	Abr.	Maió	Jun.	Jul.	Ago.
Florada, chumbinho e expansão dos frutos				Granação dos frutos			Maturação dos frutos			Repouso, senescência dos ramos 3° e 4°	
Período reprodutivo (novo período vegetativo)										Autopoda	

Fonte: Camargo e Camargo (2001).

A primeira fase, conhecida por “vegetação e formação das gemas vegetativas”, ocorre normalmente de setembro a março (CAMARGO, 1985b).

A segunda fase, a da “indução, diferenciação, crescimento e dormência das gemas florais”, é caracterizada por dias curtos, indo normalmente de abril a agosto. Essas gemas florais, depois de terem completado o desenvolvimento, entram em dormência e ficam prontas para a antese, quando acontece um aumento substancial de seu potencial hídrico, causado por chuva ou irrigação (GOUVEIA, 1984). Nos dois meses finais – julho e agosto –, as gemas dormentes produzem um par de folhas pequenas, etapa essa que separa o primeiro ano fenológico do segundo (EMBRAPA, 2009).

A terceira fase, que vai de setembro a dezembro, inicia-se com a “florada”, depois de um aumento do potencial hídrico nas gemas florais maduras (choque hídrico), e vai até a “expansão dos frutos”. Uma florada principal acontece quando se verifica um período de restrição hídrica seguido de chuva, irrigação ou mesmo de um acentuado aumento da umidade relativa do ar (CAMARGO; FRANCO, 1981; RENA; MAESTRI, 1985).

A quarta fase, a da “granação”, ocorre de janeiro a março do ano seguinte, quando há formação dos grãos. Nessa fase, um estresse hídrico pode ser prejudicial, produzindo frutos mal granados, que causam os defeitos preto, verde e ardido, como também o chochamento de grãos (MEIRELES et al., 2004c). A fase de “maturação dos frutos”, que se verifica normalmente de abril a junho, depende da periodicidade da cultivar e da

acumulação de energia solar, ou seja, do somatório de ETp, de aproximadamente 700 mm, após a florada principal. Nessa etapa, deficiências hídricas moderadas beneficiam a qualidade do produto (EMBRAPA, 2009).

A sexta fase, a do “repouso e senescência”, ocorre em julho e agosto, sendo caracterizada pela autopoda do cafeeiro, em que muitos ramos produtivos (terciários e quaternários) secam e morrem, limitando o crescimento vegetativo (EMBRAPA, 2009).

Segundo Camargo (1974), o cafeeiro, para vegetar e frutificar normalmente, necessita encontrar umidade suficiente no solo durante o período de vegetação e frutificação, que vai de setembro/outubro a abril/maio. Na fase de colheita e repouso, de junho a setembro, a umidade pode reduzir bastante e aproximar-se do ponto de murchamento, sem, com isso, trazer grandes prejuízos para a planta. É o que acontece nas regiões cafeeiras da Mogiana, em São Paulo, e no sul de Minas Gerais. Uma deficiência hídrica no período de julho a agosto, fase anterior à antese, pode se tornar benéfica, favorecendo uma florada mais uniforme já nas primeiras chuvas de setembro.

A exigência hídrica do cafeeiro arábica varia bastante, de acordo com suas fases fenológicas. Logo, pela descrição da fenologia do café e suas necessidades hídricas, percebe-se que, na ausência de chuvas, é esperado manejo do sistema irrigante no período: a partir de fevereiro do primeiro ano fenológico antes de julho; setembro a dezembro; janeiro a março do segundo ano fenológico. Toda a cafeicultura comercial do Brasil apresenta o florescimento na primavera, a frutificação no verão, a maturação no outono e a colheita no inverno. Ainda, segundo relatório da SENAR de 2016, para uma floração uniforme e abundante da lavoura, recomenda-se que provoque o estresse das plantas suspendendo a irrigação, no período de junho a setembro, e após a floração, que se reinicie o fornecimento de água.

2.5.2 Irrigação do Café

As mudanças climáticas globais, o aumento eminente da temperatura do ar e as mudanças no regime hídrico, ocasionam mudança no zoneamento climático para a cultura do café (JESUS FILHO et al., 2012). Tornando-se adequado a utilização de irrigação para sustentar a planta de cafeeiro em períodos de maior demanda hídrica (PIZZETA et al., 2016), evitando estresses que possa comprometer a produtividade e a qualidade do café. Visto que a irrigação é fundamental para promover a maximização da viabilidade

econômica da cultura, principalmente em regiões marginais, para a produção do cafeeiro arábica quanto ao déficit hídrico (SANTINO e FERNANDES, 2000).

Os déficits hídricos podem levar à queda de produtividade do cafeeiro, embora seus efeitos dependam da duração e da intensidade da deficiência hídrica e do estágio fenológico em que a planta se encontra (EMBRAPA, 2009). Segundo Camargo (1977), analisando dados comparativos do balanço hídrico climatológico de várias regiões produtoras do Brasil, declarou que a cafeicultura pode suportar deficiências hídricas de até 150 mm por ano, principalmente se esse período não se prolongar até o mês de setembro.

A irrigação já é uma realidade na cafeicultura brasileira, e ocupa área significativa entre as que são irrigadas no país. Para que um projeto de irrigação atinja seus objetivos, é necessário que o mesmo seja bem dimensionado e o manejo da irrigação eficiente. O manejo da irrigação tem, por um lado, o compromisso com a produtividade do cafeeiro, e por outro, o uso eficiente da água e da energia. Diversos enfoques foram adotados ao longo da evolução da cafeicultura irrigada no Brasil. Inicialmente, a única preocupação era o aumento da produtividade; atualmente, a modernização da agricultura mundial agregou o conceito da sustentabilidade, com a integração de todos os fatores que interferem na produção. Nesse contexto, é fundamental a utilização eficiente da água e a conservação do ambiente, os grandes desafios da cafeicultura irrigada (FOLEGATTI; FERNANDES, 2013).

2.5.3 Manejo do Sistema Irrigante

O manejo da irrigação é um conjunto de ações capaz de controlar a disponibilidade de água e satisfazer as necessidades hídricas das plantas, maximizando a produtividade e qualidade da produção com menor impacto ao meio ambiente. Logo, deve-se colher informações para saber o momento, a quantidade e o modo de aplicar água nas lavouras para obter o melhor aproveitamento (SALVADOR, 2021).

Segundo Folegatti et. al. (2013), o manejo da irrigação envolve três etapas distintas e complementares:

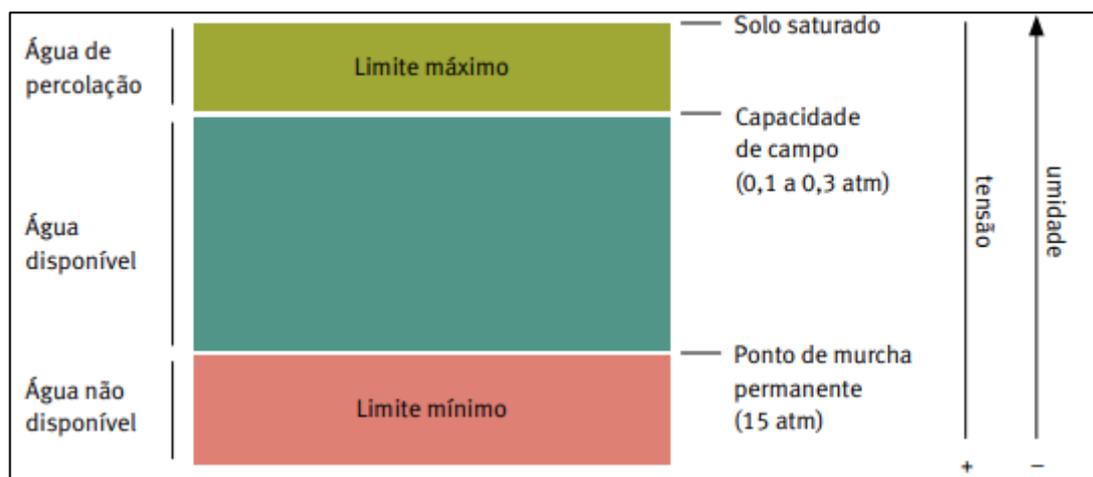
- Definição da época e da lâmina de água a ser aplicada;
- Definição das metas de eficiência na aplicação da água e de ajuste do funcionamento do mesmo para atingir esses limites;
- Manutenção preventiva e corretiva dos equipamentos e sistemas para que funcionem bem.

Também as características, funcionamento e potencialidades de um sistema de irrigação são aspectos que não podem ser negligenciados. Tradicionalmente, dividem-se os métodos de manejo da irrigação em três grupos:

- Os que se baseiam unicamente no conhecimento do estado hídrico do solo, ou seja, no manejo da água útil do solo;
- Os que se baseiam no conhecimento do estado hídrico da planta;
- Os que se baseiam na variação do clima.

Uma variável importante para o manejo da irrigação é o solo e suas características de capacidade de armazenamento de água. Antes de qualquer programa de manejo de irrigação, é aconselhável que se retirem amostras do solo, para que em laboratório obtenha a curva de retenção da água em camadas de 20 cm em 20 cm, até a profundidade efetiva do sistema radicular cafeeiro (60 cm). Pode-se afirmar que água disponível para as plantas é a retida entre a capacidade de campo (CC) e o ponto de murcha permanente (PMP). (FOLEGATTI; FERNANDES, 2013), de acordo com o esquema apresentado na Figura 13. Sabe-se que quantidade de água retida em um tipo de solo é influenciada por atributos de solo, especialmente pela textura, estrutura, mineralogia e teor de matéria orgânica. A variabilidade espacial do solo é uma das principais dificuldades para obtenção de informações robustas (SALVADOR, 2021).

Figura 13 – Representação da Capacidade de Armazenamento de Água nos Poros do Solo



Fonte: Adaptado de CaféPoint (FOLEGATTI; FERNANDES, 2013).

2.5.4 Métodos de Manejo do Sistema Irrigante

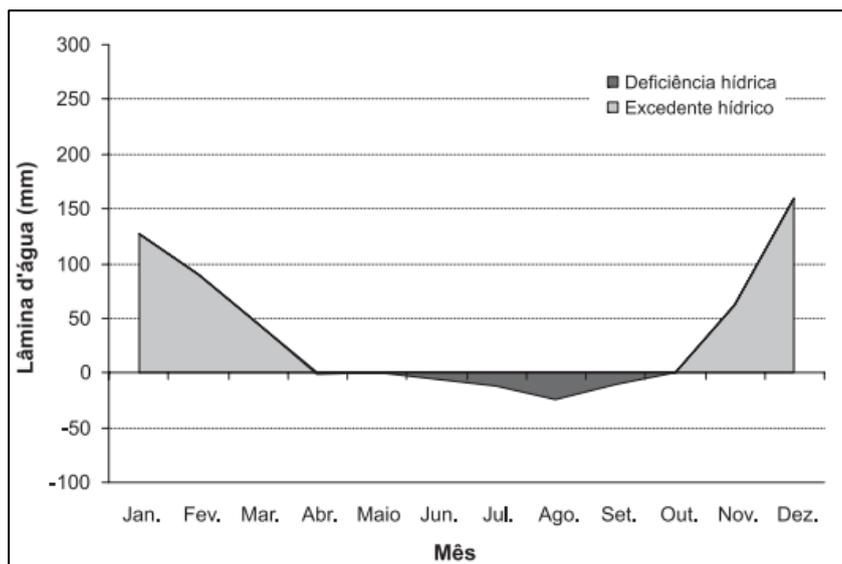
A determinação da época da irrigação e da quantidade de água a ser aplicada pode ser feita com o monitoramento da planta, do solo ou do clima, que são descritos como:

- **Monitoramento da planta:** O manejo pelo monitoramento da planta é o controle ideal, sob o ponto de vista do rigor científico, mas as implicações técnicas e operacionais dificultam sua utilização nas condições de campo. Diferentes métodos permitem o monitoramento das plantas, como: (i) medições do potencial hídrico; (ii) medições da temperatura das folhas, dentre outros. Cada uma dessas técnicas apresenta limitações de aplicabilidade em campo, em razão da pequena disponibilidade de informações dos limites e dos índices recomendáveis e dos problemas operacionais.
- **Monitoramento do solo:** constitui uma metodologia usual para manejo da irrigação, de modo independente ou associado a outros métodos de controle. Uma vez definidos os limites da CC e PMP e a estratégia de uso da água disponível, o momento de ser irrigar e a quantidade de água a ser aplicada são determinados mediante a avaliação do teor de água no solo. Esta determinação pode ser realizada de forma direta, com a retirada de amostras de solo em várias profundidades e diferentes locais, definindo, posteriormente, a umidade existente pelo método padrão, secagem e estufa. Por questões operacionais, a determinação do teor de água no solo para efeitos de manejo da irrigação pode ser de forma indireta, por meio de diversos equipamentos disponíveis no mercado. Os mais utilizados são tensiômetros, métodos eletrométricos, sonda de nêutrons, TDR (sigla do nome em inglês para Reflectometria de Domínio do Tempo), métodos amplamente utilizados na pesquisa e divulgados na literatura sobre irrigação.
- **Monitoramento do clima:** A metodologia de monitoramento do clima é uma das maneiras mais simples e operacionais de manejar a irrigação do cafeeiro. As variáveis meteorológicas são utilizadas para determinar indiretamente a necessidade de água da cultura. Trata-se de um balanço hídrico, onde se avalia a entrada de água (irrigação e chuva) e a saída (evapotranspiração, percolação e escoamento superficial). A

evapotranspiração pode ser obtida por meio de medições diretas no campo (lisímetros) ou por estimativas baseadas em dados climáticos ou da cultura. As medições diretas no campo são muito caras e são utilizadas principalmente para calibrar os métodos de estimativa de evapotranspiração, com base em dados climáticos. Calcula-se a evapotranspiração da cultura (ET_c) do café multiplicando-se a evapotranspiração potencial de referência (ET_o) por um coeficiente de cultura (K_c), que incorpora os efeitos do tipo de cultura e sua fase de desenvolvimento, frequência e forma de aplicação da água. Existem vários métodos que podem ser utilizados para a estimativa da evapotranspiração, como os de Camargo, Penman, Penman-Monteith, Radiação Solar, Hargreaves, Linacre, Blaney Criddle, dentre outros.

Em relação ao método de manejo pelo monitoramento do clima, o balanço hídrico, trata-se de uma forma de avaliar os períodos com excedentes e deficiências hídricas ao longo do ano, e pode ser seriado ou sequencial, normalmente utilizado no monitoramento agrometeorológico, usando uma das seguintes bases: diária, decendial, semanal e mensal. Especificamente no caso do café, o monitoramento agrometeorológico vem sendo feito desde 2002 para algumas regiões dos estados de Minas Gerais, São Paulo e Paraná, empregando-se o balanço hídrico sequencial decendial de Thornthwaite e Mather (1955), (MEIRELES et al., 2002, 2003, 2004a, 2004b, 2005). O monitoramento consiste na apresentação e na análise dos balanços hídricos sequenciais decendiais de algumas localidades. Esses balanços foram simulados utilizando-se o método proposto por Thornthwaite e Mather (1955), nos quais se destacam os períodos com excedentes e deficiências hídricas, além do armazenamento de água no solo (EMBRAPA, 2007). A Figura 14 apresenta um exemplo de balanço hídrico, onde ocorre excedente hídrico no período chuvoso da região de Mococa (segunda quinzena de outubro à primeira quinzena de abril) e déficit hídrico durante o período de seca da região (segunda quinzena de abril à primeira quinzena de outubro).

Figura 14 – Exemplo de Balanço Hídrico Climático de Mococa, SP.



Fonte: EMBRAPA, 2007.

O ET_c para a cultura do café, pode ser determinada pela equação (1), onde o K_c é uma variável adimensional que varia de acordo com a fase fenológica da cultura, com o manejo, com o estande das plantas, e também, com as condições climáticas e épocas de plantio, já que os parâmetros locais (temperatura, umidade relativa, vento, radiação solar) e as variações fisiológicas e aerodinâmicas da cultura influenciam diretamente na evapotranspiração (MUDRIK, 2005).

$$ET_c = K_c \times ET_o \quad (1)$$

A evapotranspiração de referência (ET_o), que expressa a influência do clima sobre a cultura de referência, é determinada por métodos que se baseiam em dados climáticos, utilizando desde equações simples, com base somente em valores médios de temperatura do ar, até equações mais complexas, baseadas em modelos físicos, passando por métodos evaporimétricos (MAROUELLI et. al, 2011). Segundo Salvador (2021), há a necessidade das informações meteorológicas completas para se obter um nível de precisão aceitável quando se faz o manejo de irrigação com base nesses indicadores,

Neste trabalho será considerado que as lavouras utilizam o método de manejo pelo monitoramento do clima e os valores de K_c utilizados para fins de cálculos de ET_c são dados resultante da pesquisa realizada por Rocha et. al (2016), em que, no período de janeiro de 2008 a dezembro de 2011, mensurou-se, na área experimental da Embrapa Cerrados, a demanda hídrica de um cultivar de café e oferecendo como saída dados médios anuais de K_c . Embora esse estudo seja para uma região específica e um cultivar

específico do café arábica, escolheu-se utilizar esses valores pois se assemelham à demais dados encontrados na literatura e são compatíveis com as fases fenológicas dos cafeeiros descritas por Camargo (1988). Os valores são apresentados na Tabela 3. Já os valores de ETo podem ser facilmente adquiridos no site do INMET na página de Balanço Hídrico Climatológico Mensal da Sisdagro (Sistema de Suporte à Decisão na Agropecuária).

Tabela 3 – Valores médios mensais de coeficiente de cultura (Kc) para o café (*Coffea arábica*).

Fase fenológicas do cafeeiro		Mês	Kc
1º ano	2º ano		
V+F	G	Jan.	1,18
		Fev.	1,2
		Mar.	1,2
		Abr.	1,12
I+Mg	M	Maió	1,07
		Jun.	0,85
		Jul.	0,87
		Ago.	0,88
V+F	R+Fp	Set.	0,98
		Out.	1,21
		Nov.	1,23
		Dez.	1,23
	C+E		

Fonte: Adaptado de Rocha et. al. (2016).

Onde: Vegetação (V), formação de gemas florais (F), Indução (I), Maturação das gemas florais (Mg), Granação (G), Maturação dos frutos (M), Repouso (R) e Folhas pequenas (Fp).

2.5.5 Principais Sistemas de Irrigação do Cafeeiro

Os sistemas de irrigação mais utilizados são os de aspersão (convencional, em malha e pivô central) e os localizados (gotejamento e modificados). Em função de aspectos relacionados ao consumo de energia, exigência de mão de obra e outros de natureza operacional, os sistemas mais viáveis de irrigação por aspersão têm sido o convencional (principalmente do tipo malha) e o pivô central. Com relação à irrigação localizada, os sistemas mais empregados são o de gotejamento – por suas características técnicas que permitem uma irrigação com grande precisão, economia de água e energia –,

e as fitas de polietileno – sistema também conhecido como “tripa”, principalmente pelo menor custo de implantação.

Basicamente, os sistemas de irrigação para o cafeeiro podem ser divididos em dois grandes grupos: sistemas de irrigação com aplicação da água em área total e com aplicação localizada da água, molhando apenas partes da área onde a cultura está implantada (LIMA; FERNANDES, 2013):

- Pivô central: Os primeiros pivôs centrais utilizados na cafeicultura foram adaptados de outras lavouras, com irrigação em área total; ou seja, tanto nas linhas de café, quanto nas entrelinhas. Apesar de viabilizar a cafeicultura empresarial nas regiões de cerrado, o sistema pivô central convencional apresentava o inconveniente da aplicação de grandes volumes de água e com irrigação das entrelinhas do café, exigindo controle mais intensivo do mato. A partir desta tecnologia, pesquisadores e consultores brasileiros desenvolveram uma técnica extremamente interessante para a irrigação do cafeeiro, com o pivô central e plantio realizado em círculo e localização dos emissores sobre as linhas do café, denominados Lepa (sigla em inglês para *Low Energy Precision Application*); ou seja, aplicação precisa de água, com baixo consumo de energia, já que a pressão requerida é menor (SANTINATO; FERNANDES, 2002). Uma característica desse tipo de sistema de irrigação é a plantação em áreas circulares, como ilustrado na Figura 15.

Figura 15 – Exemplos de Áreas de Café com Irrigação por Pivô.



Fonte: Fazenda Café Vô Luiz (2022).

- Aspersão convencional e em malha: Existem diferentes tipos de sistema de aspersão, quanto à manipulação de seus componentes (tubos, conexões e aspersores), tais como: portátil, semifixo e fixo. Uma inovação que tem apresentado bons resultados na irrigação do cafeeiro é uma adaptação no sistema convencional de aspersão: o sistema de aspersão em malha, no qual os ramais são conectados nas extremidades, para aumentar a uniformidade da pressão na rede hidráulica. Este sistema tem como características principais: (i) utilização de tubos de PVC de baixo diâmetro que constituem as linhas laterais ligadas em “malha” e enterradas; (ii) baixo consumo de energia, de 0,6 a 1,10 CV/ha; (iii) adaptação a qualquer formato de terreno; (iv) possibilidade de divisão da área em várias subáreas; (v) facilidade de operação e manutenção; (vi) possibilidade de fertirrigação; (vii) baixo custo de instalação e manutenção. Como limitações, este sistema apresenta: (i) dificuldade de automação; (ii) maior dependência de mão de obra; (iii) abertura de grande número de valetas para enterrar os tubos em malha (DRUMOND; FERNANDES, 2004).
- Autopropelido convencional: Consiste num aspersor tipo canhão, montado em veículo equipado com mecanismo autopropulsor que, por meio da movimentação hidráulica de um carretel, enrola um cabo de aço capaz de deslocar o sistema ao longo da faixa de solo a ser irrigada.
- Carretel enrolador: Mais moderno que o anterior, irriga uma faixa de terra continuamente, pelo deslocamento de um canhão hidráulico montado sobre rodas, conectado na extremidade de uma mangueira. Neste equipamento, apenas a mangueira é movimentada.
- Gotejamento: Efetua a irrigação sobre o solo na área de maior absorção das raízes do cafeeiro, com gotejadores de pequena vazão (1 a 10 litros/hora); porém, capazes de irrigar com alta frequência (até mesmo várias vezes ao dia), mantendo a umidade do solo na zona radicular próxima à capacidade de campo, condição que facilita a absorção de água pelo cafeeiro (Santinato e Fernandes, 2012). A água é pressurizada pelo sistema de bombeamento, e, antes que chegue aos gotejadores, passa por um sistema de filtragem, prática essencial ao bom funcionamento deste sistema. Apresenta, como principais vantagens: (i) alta uniformidade de aplicação de água, de 90 a

95%; (ii) redução de gastos de água, energia e mão de obra; (iii) aplicação eficiente de defensivos e fertilizantes; (iv) possibilidade de automação total do sistema. Como desvantagens, podem ser citados: (i) alto custo de implantação; (ii) riscos de entupimento de emissores se mal manejados; (iii) concentração do sistema radicular em apenas parte do solo, onde se forma o “bulbo molhado” do gotejador; (iv) possíveis danos e avarias das linhas laterais por trabalhadores (enxadas) e animais (roedores).

Resumidamente, a recomendação do método de irrigação depende de fatores como: a topografia da área; disponibilidade de água na propriedade; porosidade do solo; nível tecnológico da fazenda; condições climáticas como pluviosidade, temperatura e velocidade do vento. O manejo da irrigação passa pelo gerenciamento de todo o sistema, que inclui a manutenção dos equipamentos, a estratégia de horários de irrigação, acompanhamento de custos e tarifas e a definição de quando e quanto irrigar. Para realizar a irrigação devem ser considerados o déficit hídrico, o turno da rega e a operacionalização da irrigação, onde é possível avaliar a eficiência do sistema e fazer ajustes necessários (SENAR, 2016).

2.6 AGROMETEOROLOGIA

Segundo Smith (1975), a “Meteorologia Agrícola tem por objetivo colocar a ciência da Meteorologia à serviço da Agricultura em todas suas formas e facetas, para melhorar o uso da terra, para ajudar a produzir o máximo de alimentos, e a evitar o abuso irreversível dos recursos da terra”.

A meteorologia agrícola, também conhecida como agrometeorologia, é o ramo da meteorologia que estuda a influência das condições meteorológicas nas atividades agropecuárias. Em razão da sua interação com as diversas áreas ligadas à agronomia, a agrometeorologia tem papel fundamental tanto no planejamento agrícola como nas tomadas de decisão. No planejamento agrícola faz-se presente quando se define o zoneamento agroclimático, época mais adequada de semeadura e planejamento topo e microclimático, baseados no clima local e no balanço hídrico, que define as estações seca e úmida. Já nas tomadas de decisão, a agrometeorologia é usada para mapear épocas de semeadura, colheita, irrigação, preparo do solo, controle fitossanitário e para adotar

medidas contra eventos adversos, para isso, usando-se das condições do tempo vigente e previstas, que definem as condições térmicas e de disponibilidade hídrica mais atuais (SENTELHAS; ANGELOCCI, 2009).

Carlos (2021) cita que a agrometeorologia possui três principais pilares relacionados a sua aplicação:

- Propósito estratégico: associado ao planejamento, com zoneamentos e planejamentos topo climáticos;
- Propósito tático: associado à tomada de decisão, visando uma maior flexibilização dos sistemas de produção em função do acompanhamento das condições meteorológicas;
- Obtenção de resiliência: fortalecimento dos sistemas agrícolas por meio da diversificação e das estratégias de manejo dos riscos, como rotação de cultura, quebra-ventos, cobertura do solo e irrigação.

A agricultura é um sistema tecnológico artificial desenvolvido pelo homem com o objetivo de se obter alimento, fibra, energia em quantidade suficiente para garantir sua subsistência por um certo período. As plantas foram gradativamente sendo domesticadas até permitir que extensas áreas fossem cobertas com indivíduos com mesma composição genética. Nessa condição, a interação com o ambiente depende do estágio de desenvolvimento das plantas (fases fenológicas) visto que, num dado instante, todos os indivíduos daquela comunidade têm a mesma idade, com porte semelhante, e com a mesma suscetibilidade aos rigores impostos pelo meio. Quanto mais homogênea for a população de plantas, maior será sua suscetibilidade às condições ambientais (PEREIRA; ANGELOCCI; SENTELHAS, 2007).

A gestão do balanço hídrico no solo é uma boa estratégia de manejo de riscos, é favorecido com o uso de técnicas agrícolas de cobertura do solo, que aumentam a capacidade do solo em reter e armazenar água e conseqüentemente favorecem uma maior produtividade agrícola. O ritmo da disponibilidade de energia e de água de uma região determina o seu potencial de produtividade agrícola (PEREIRA; ANGELOCCI; SENTELHAS, 2007).

A disponibilidade de água depende do balanço entre chuva e evapotranspiração, sendo esta última dependente das condições da superfície (tipo de cobertura, tipo de solo) e da demanda atmosférica (disponibilidade energética, umidade do ar, e velocidade do vento). A disponibilidade hídrica no solo pode ser quantificada pelo balanço hídrico climatológico, evidenciando as flutuações temporais de períodos com excedente e com

deficiência, permitindo planejamento das atividades agrícolas, visando minimizar perdas. As exigências hídricas das culturas e sua relação com as condições ambientais embasam o suporte ao planejamento e quantificação da irrigação. Das atividades econômicas, a agricultura é sem dúvida aquela com maior dependência das condições do tempo e do clima (PEREIRA; ANGELOCCI; SENTELHAS, 2007).

2.7 BIBLIOTECAS PYTHON E ALGORITMOS UTILIZADOS

Nesta seção serão apresentados os algoritmos e ferramentas utilizados no desenvolvimento da metodologia. A linguagem de programação utilizada foi Python, e o ambiente de desenvolvimento utilizado para programação foi *Spyder*.

2.7.1 Ambiente Spyder e Python

Spyder é um poderoso ambiente científico escrito em Python, para Python e projetado por e para cientistas, engenheiros e analista de dados. Ele apresenta uma combinação exclusiva da funcionalidade avançada de edição, análise, depuração e criação de perfil de uma ferramenta de desenvolvimento abrangente com exploração de dados, execução interativa, inspeção profunda e recursos de visualização de um pacote científico. Além disso, o Spyder oferece integração integrada com muitos pacotes científicos populares, incluindo Numpy, SciPy, Pandas, IPython, QtConsole, Matplotlib, SymPy e muitos mais. Além de seus muitos recursos integrados (ANACONDA ORG, 2022).

Python é uma linguagem de programação interpretada, orientada a objetos, multiplataforma, de propósito geral, gratuita e de código aberto. Foi criada no Instituto Nacional de Pesquisa em Matemática e Ciência da Computação, por Guido van Rossum, na Holanda, baseada em uma linguagem chamada ABC, que tem como base BASIC e Pascal. O Python foi concebido para ser uma linguagem de programação altamente legível e por sua simplicidade, é de fácil aprendizado (TELLES, 2008).

De acordo com McKinney (2012), no início dos anos 2000, houve um aumento crescente na adoção da linguagem Python para fins de computação científica, tanto nos meios acadêmicos como na indústria. Nos últimos anos, esta linguagem vem se tornando uma alternativa para tarefas de análise de dados, principalmente devido a seu crescente número de bibliotecas para fins científicos, como o Pandas e NumPy.

2.7.2 Pacotes Python Utilizados

- **Pandas**

A biblioteca Pandas transforma o Python em um ambiente rico para análise de dados. Ela oferece funções e estruturas de dados valiosas no intuito de facilitar o trabalho com dados estruturados, disponibilizando métodos para junção de dados, limpeza e transformação, por exemplo. Outra facilidade que a biblioteca provém é a capacidade de manipular dados de uma forma flexível como em planilhas eletrônicas ou bancos de dados relacionais (MCKINNEY, 2012). Dentre os tipos de estruturas de dados no Pandas, dois se destacam, e foram utilizados no algoritmo deste trabalho, são o tipo: *Series* e *Dataframe*.

Series, trata-se de uma estrutura de dados em objeto, parecido com um vetor de apenas uma dimensão que contém um vetor de dados e outro vetor chamado de índice. A estrutura também pode ser vista como um dicionário de tamanho fixo. Por padrão, o vetor índice do objeto mostra o número do índice de cada posição do vetor de dados. Porém, é possível informar uma *label* para o índice e acessar as posições do vetor de dados através desta *label* especificada pelo usuário (MCKINNEY, 2012).

Já o *DataFrame*, é uma estrutura que se assemelha a uma planilha eletrônica, de forma tabular, contendo uma coleção ordenada de colunas que podem assumir diversos tipos de dados, nos quais possuem um índice para cada linha e coluna. Uma forma de criar um *DataFrame* pode ser declarando um dicionário com listas de igual tamanho (MCKINNEY, 2012). Também, é possível importar arquivos CSV, TXT, EXCEL dentro de um *DataFrame*. Neste trabalho, todas as bases de dados de entrada, descritas na seção anterior, foram exportadas para o formato *DataFrame*, para serem feitas as leituras e manipulações desejadas.

- **NumPy**

NumPy (abreviação de *Numerical Python*) fornece uma interface eficiente para armazenar e operar em *buffers* de dados densos. De certa forma, os *arrays* NumPy são como o tipo de lista integrado do Python, mas os *arrays* NumPy fornecem armazenamento e operações de dados muito mais eficientes à medida que os *arrays* aumentam de

tamanho. As matrizes NumPy formam o núcleo de quase todo o ecossistema de ferramentas de ciência de dados em Python (VANDERPLAS, 2017). De acordo com McKinney, 2012, algumas características da ferramenta são: (i) O objeto ndarray que é um vetor multidimensional rápido e eficiente; (ii) Funções para executar operações matemáticas entre vetores; (iii) Suporte a Transformações de Fourier, Álgebra Linear e geração de números aleatórios.

- ***Shapefile***

A biblioteca Python *Shapefile* (PyShp) fornece suporte de leitura e gravação para o formato Esri *Shapefile*. O formato *Shapefile* é um popular formato de dados vetoriais do Sistema de Informações Geográficas criado pela Esri. A Esri descreve os formatos de arquivo shp e shx. No entanto, um terceiro formato de arquivo chamado dbf também é necessário, e deve estar presente na pasta de arquivos (CURATED PYTHON, 2022). Essa biblioteca foi utilizada para ler os dados de entrada de solos e desnível. Sendo o arquivo de desnível o mais pesado dentre todos os dados utilizados neste trabalho, com 1, 29 GB, representando a etapa mais desafiadora de processamento.

- **Scikit-Learn**

Scikit-learn é uma biblioteca Python com documentação incomparável projetada para fornecer um API consistente para dezenas de algoritmos. Ele é contruído sobre – e é ele próprio – um componente central da pilha científica Python, ou seja, NumPy, SciPy, pandas e matplotlib. O Scikit-learn cobre as áreas de: (i) classificação; (ii) regressão; (iii) agrupamento; (iv) redução de dimensionalidade; (v) seleção de modelo; e, (vi) pré-processamento.

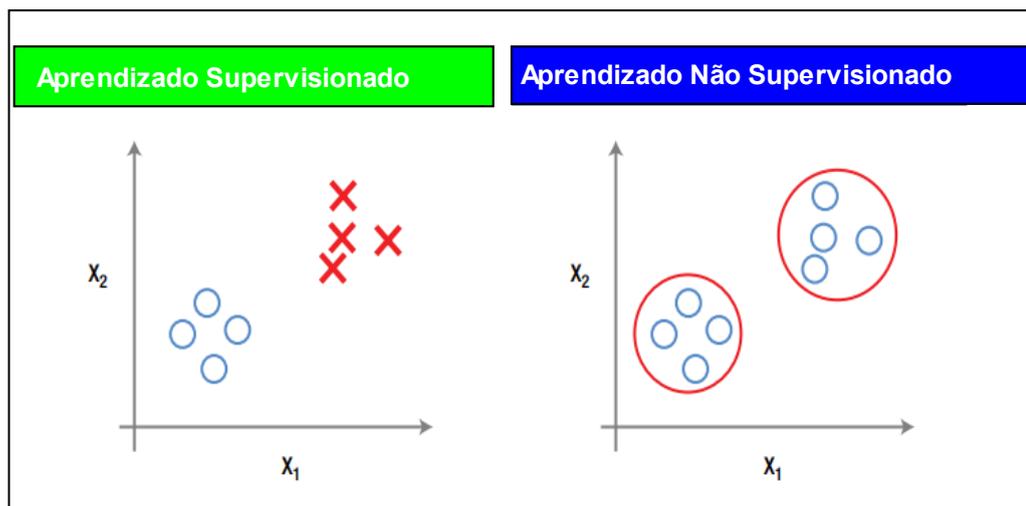
2.7.3 Algoritmos de Aprendizado de Máquina Utilizados

Como comentado anteriormente, foram utilizados, basicamente, duas técnicas de aprendizado de máquina, uma do tipo não-supervisionado e outra do tipo supervisionado: clusterização (agrupamento) e classificação, respectivamente. Nesta seção serão descritas estas técnicas.

- **Aprendizado Supervisionado x Não Supervisionado**

Algoritmos de aprendizado de máquina são divididos entre supervisionado e não supervisionado. A Figura 16 apresenta os dois tipos de aprendizado e suas principais diferenças. Como o nome sugere, algoritmos de aprendizado supervisionado requerem supervisão para treinar o modelo. Essa supervisão geralmente é necessária no caso de classificação onde temos dados rotulados nos quais treinamos o modelo para que ele preveja os rótulos dos dados não vistos. Aqui, a supervisão se dá por meio do rótulo fornecido com cada observação (ou seja, supervisionando o processo de aprendizagem). Exemplos incluem classificação para preditores discretos e regressão para os contínuos (HAROON, 2017).

Figura 16 – Supervisionado x Não Supervisionado.



Fonte: Elaboração Própria, adaptado de Haroon, (2017, p.133).

Os algoritmos de aprendizado não supervisionados não requerem supervisão dos dados durante o treinamento do modelo. Um excelente exemplo disso é o agrupamento, que descobre os rótulos sem qualquer supervisão necessária. Esses rótulos descobertos se tornam a base para classificar quaisquer novos dados não vistos. Outro exemplo de aprendizagem não supervisionada é a regra de associação, que envolve os conceitos de complemento e substituto. Complemento refere-se a um fenômeno pelo qual se um comprador compra X, então, com um alto grau de certeza, ele também comprará Y.

Substituto refere-se a um comportamento pelo qual um comprador comprará X ou Y (HAROON, 2017).

- **Clusterização**

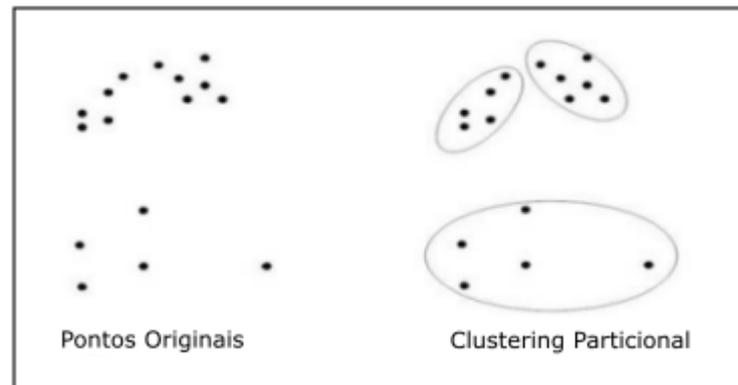
Segundo Haroon, 2017, a análise de cluster refere-se ao agrupamento de observações para que os objetos dentro de cada cluster compartilhem propriedades semelhantes e as propriedades de todos os clusters sejam independentes umas das outras. Os algoritmos de cluster geralmente otimizam maximizando a distância entre os clusters e minimizando a distância entre os objetos em um cluster. A análise de cluster não é concluída em uma única iteração, mas passa por várias iterações até que o modelo convirja. Convergência de modelo significa que as associações de cluster de todos os objetos convergem e não mudam a cada nova iteração.

Alguns algoritmos de *clustering* não pedem o número de clusters/componentes e apresentam o número de clusters que estatisticamente fazem mais sentido. No entanto, uma grande quantidade de algoritmos de *clustering* solicita ao usuário antecipadamente o número de clusters/componentes que ele deseja na saída, esse modelo foi o utilizado neste trabalho. É importante entender que o agrupamento, assim como a classificação, é usado para segmentar os dados; no entanto, esses grupos não são definidos anteriormente no conjunto de dados de treinamento (ou seja, são dados não rotulados).

Haroon, 2017 define os diferentes algoritmos que implementam diferentes técnicas para o cálculo de clusters. As técnicas são:

- **Particionamento:** Agrupa dados em um determinado número de clusters enquanto otimiza para o objetivo (por exemplo, distância), a Figura 17, exemplifica essa técnica.

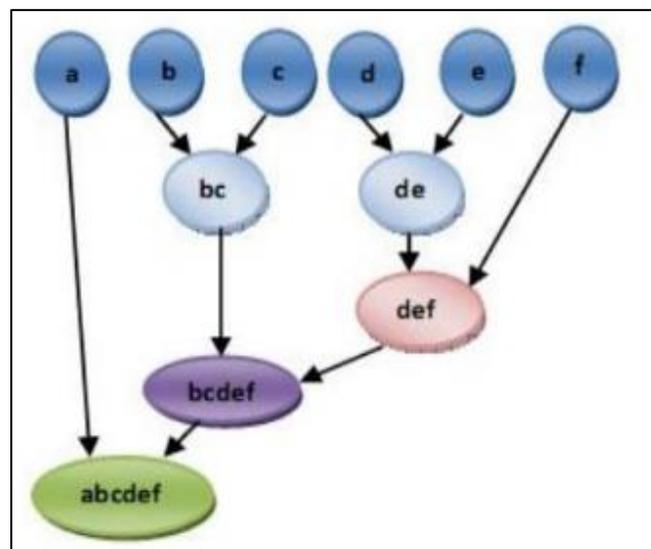
Figura 17 - Exemplo de aplicação de Cluster por Particionamento.



Fonte: Haq, Huarong e Khattak, 2017.

Hierárquico: Agrupa dados em uma hierarquia de clusters. Essas hierarquias são formadas de cima para baixo ou de baixo para cima, Figura 18.

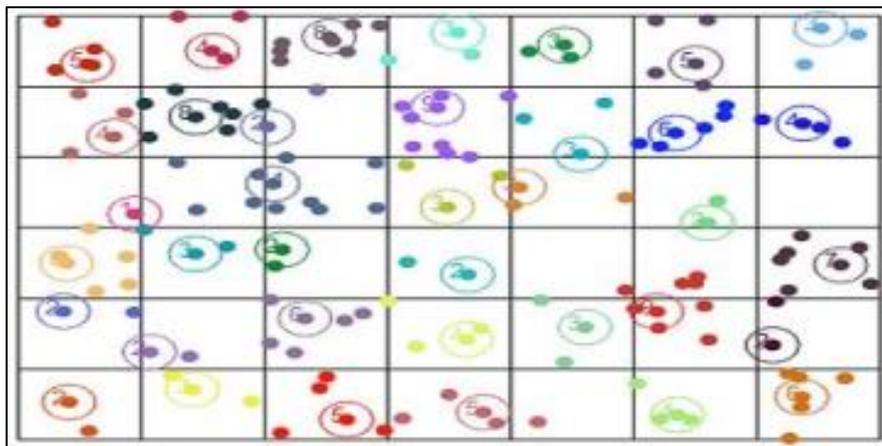
Figura 18 - Exemplo de aplicação de Cluster Hierárquico.



Fonte: Haq, Huarong e Khattak, 2017.

- Baseado em grade: divide os dados em células hiper retângulos, descarta células de baixa densidade e combina células de alta densidade para formar clusters, Figura 19.

Figura 19 - Exemplo de aplicação de Cluster Baseado em grade.



Fonte: Haq, Huarong e Khattak, 2017.

Um bom algoritmo de agrupamento atende ao requisito de semelhança dentro do cluster e dissimilaridade entre clusters, pode lidar com o conjunto de dados de treinamento que é de alta dimensão e afetado por ruído e discrepâncias (outliers).

2.8 RESUMO DO CAPÍTULO

Este capítulo apresentou uma revisão dos bibliográfica dos tópicos relacionados à dissertação, desde a definição de PNT aos estudos existentes relacionados ao tema. Tais estudos tratam da estimação dos montantes dessas perdas, identificação de ocorrência de PNT na rede e alguns estudos já buscam classificar unidades consumidoras das distribuidoras de estudo como sendo fraudulentas ou não. Empregando diversas técnicas de aprendizado de máquina como, por exemplo, *fuzzy*, *VSM*, *clustering*, *Decision Tree*, *Random Forest*, *K-Means*, etc. Contudo esses estudos, em sua maioria, são realizados para zonas urbanas.

De maneira a identificar quais seriam as variáveis cadastrais dentro da base de dados da CPFL dos clientes rurais irrigantes, após a revisão de PNT, é apresentada a caracterização de consumo de energia do Brasil, apontando quais são as subclasses das unidades consumidoras que podem ser classificadas como irrigantes dentro da classe de consumo rural. Esses clientes podem utilizar de desconto na tarifa de energia, se utilizarem seus sistemas de irrigação dentro do período indicado destinado à essa atividade. Logo o estudo busca aplicar uma metodologia classificatória preditiva de detecção de possíveis ocorrências de PNT para as unidades consumidoras cadastradas

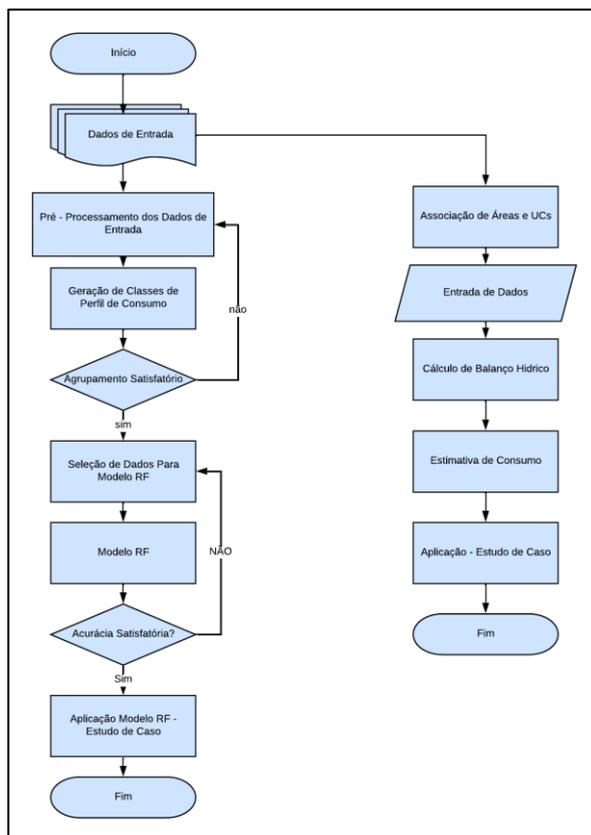
como irrigantes beneficiárias do desconto na tarifa, identificando padrões de consumo das unidades consumidoras e desvios nesses padrões de consumo.

3. METODOLOGIA PARA DETECÇÃO DE PNT EM UNIDADES CONSUMIDORAS IRRIGANTES DE CAFÉ

Neste capítulo é apresentada a metodologia proposta para detecção de PNT em unidades consumidoras rurais irrigantes de café no estado de SP. São apresentados o conjunto de dados utilizados na implementação e os atributos mapeados que influenciam no consumo de energia elétrica desses consumidores em específico. São descritas brevemente, também, as tecnologias empregadas na metodologia e as métricas de avaliação utilizadas para validação dos resultados obtidos.

O fluxograma da metodologia é apresentado na Figura 20, onde apresenta-se as duas ramificações do estudo apresentado neste trabalho, focado para irrigantes de café: além do modelo classificatório preditivo utilizando-se de *Kmeans* e *Random Forest*; são consideradas variáveis de agrometeorologia e de processamento de imagens para desenvolvimento da metodologia de detecção de PNT por estimativa de consumo.

Figura 20 – Fluxograma da Metodologia Proposta.



Fonte: Elaborado pelo Autor.

Como apresentado na Figura 20, tem-se duas ramificações do de aplicabilidade de metodologia para detecção de PNT:

- i. Uma (ramificação principal, a esquerda) referente à aplicação de aprendizado de máquina e geração de classes de perfis de consumo para serem usadas no modelo preditivo, onde a possível ocorrência de PNT é detectada pelas divergências de perfis de consumo entre os períodos de safra (comparando os perfis preditos num período do ano vigente com o perfil de consumo do mesmo período do ano anterior)
- ii. Já na outra ramificação aplica-se a metodologia do balanço hídrico em um subconjunto de unidades consumidoras para detectar PNT, pois, dentre as classes de perfis de consumo geradas sempre iremos ter a classe mais inferior, onde consumos menores do que o esperado não irão resultar em mudança de classe. Desse modo é aplicado a metodologia de balanço hídrico para essas UCs, onde é possível estimar a energia elétrica que seria necessária para irrigar sua área plantada nas condições de evapotranspiração do mês de análise.

3.1 CONJUNTO DE DADOS DE ENTRADA

Foram utilizados conjuntos de dados diversos para implementação do algoritmo, a serem citados: consumo de energia, dados cadastrais, dados agrometeorológicos (dados fenológicos, dados característicos do sistema irrigante, dados climáticos, dados de áreas de lavouras). Nesta seção são apresentados esses conjuntos de dados e as variáveis utilizadas.

3.1.1 Dados de Entrada – Consumo e Cadastrais

Um dos conjuntos de dados utilizados como entrada neste trabalho foram fornecidos por uma empresa de distribuição de energia elétrica que atua no estado de São Paulo. A distribuidora além de possuir dados históricos de consumo de energia elétrica de seus clientes também possui dados cadastrais de natureza comercial. Os registros dos dados foram disponibilizados em arquivos no formato CSV. Os atributos selecionados

dessa base de dados para compor as variáveis de entrada do algoritmo da metodologia proposta são:

- Consumo histórico: composto por 72 (6 anos, escolha própria do autor) meses de leituras das UC de estudo, disponibilizado em KWh;
- Data de referência: data referente a leitura do consumo de energia, formato AAAAMM;
- Descrição do município: nome do município da UC, disponibilizado em caixa alta sem caracteres especiais;
- Grupo de tensão: classificação do grupo de tensão da UC (A ou B);
- Número de Coordenadas GIS X e Y: coordenadas geográficas do ponto de distribuição que atende a UC;
- Descrição do Alimentador: classificação do alimentador que atende a UC;
- Código de Instalação: código único registrado da UC.

Como comentado, esses dados foram extraídos de três arquivos disponibilizados: um arquivo de consumo histórico de UC da classe de consumo A (108973 linhas e 4 colunas), um arquivo de mesma natureza para UC da classe de consumo B (2228395 linhas e 9 colunas), e um arquivo com dados cadastrais comerciais das UC (2731 linhas e 101 colunas). As linhas do último arquivo referem-se ao número de UC irrigantes cadastrados na modalidade tarifária de período reservado noturno. Os atributos e descrição são apresentados na Tabela 4.

Tabela 4 – Atributos selecionados para o conjunto de dados de entrada da distribuidora

Atributo Selecionado	Descrição	Natureza do dado
COD_ALIMENTADOR	Atributo que indica qual alimentador atende o consumidor.	Qualitativo - nominal
COD_GRUPO_TENSAO	Atributo que indica qual grupo de tensão o consumidor está cadastrado.	Qualitativo - nominal
COD_INSTALACAO	Código único referente a UC cadastrada.	Qualitativo - numérico
CONSUMO	72 linhas de consumo	Quantitativo – numérico [KWh]

DT_REFERENCIA	Data de leitura do consumo associado ao período, ano e mês da leitura.	Qualitativo – numérico
LAT	Coordenada de latitude do ponto de distribuição que atende o consumidor. Formato grau decimal.	Qualitativo - numérico
LONG	Coordenada de longitude do ponto de distribuição que atende o consumidor. Formato grau decimal.	Qualitativo - numérico
DESC_MUNICIPIO	Nome do município em que o consumidor está localizado.	Qualitativo - nominal

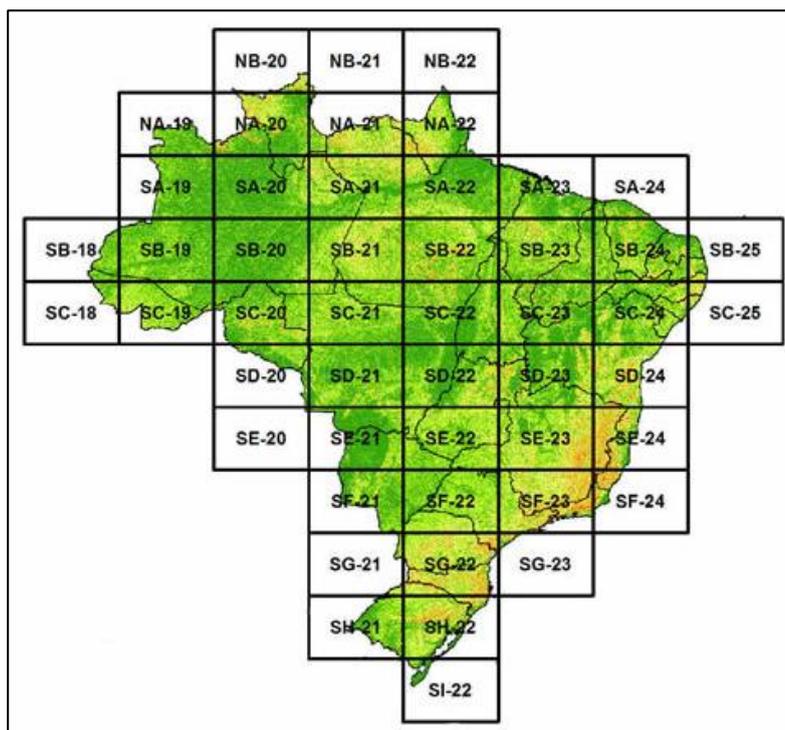
Fonte: Elaborado pelo Autor.

3.1.2 Dados de Entrada – Dados de Desnível

Os dados de desnível do relevo foram adquiridos pela plataforma do Serviço Geológico do Brasil, onde é possível baixar, em arquivo “.zip”, dados de declividade em percentual do relevo brasileiro de cada uma das folhas presente na Carta Geográfica do Brasil ao Milionésimo. Na Figura 21, é apresentada as folhas, pode-se identificar a área de estudo como estando na folha SF-22. O arquivo desses dados foi utilizado como entrada, para associar desníveis de solo às UCs, utilizando o procedimento de pontos inseridos em um polígono, de maneira que, para cada ponto de coordenada geográfica, das UCs das bases de dados, irá ser verificado qual polígono ela está inserida. Uma vez que cada polígono possui um valor de desnível (DN), a UC será associada ao valor de DN do polígono a que pertence.

O formato dessa entrada é em *shapefile* contendo a folha de interesse para aplicação do estudo em São Paulo, folha SF-22 na Figura 21.

Figura 21 - Carta Geográfica do Brasil ao Milionésimo



Fonte: Serviço Geológico do Brasil (2020).

- **Arquivo *Shapefile***

O formato *shapefile*, utilizado como entrada, é um formato de armazenamento de dados de vetor da Esri (empresa americana especializada na produção de soluções para a área de informações geográficas) para armazenar a posição, a forma e os atributos de feições geográficas. É armazenado como um conjunto de arquivos relacionados e contém uma classe de feição. Os *shapefile* normalmente contêm feições grandes com muitos dados associados e foi historicamente utilizado em aplicativos de desktop GIS como ArcMap. O caminho primário para disponibilizar dados de *shapefile* para outros visualizarem por um navegador da web e adicioná-lo a um arquivo .zip, transferi-lo e publicar uma camada de feição hospedada. O arquivo .zip deve conter pelo menos os componentes dos arquivos “.shp”, “.shx”, “.dbf” e “.prj” do *shapefile* (ArcGis Online, 2022).

O arquivo de solos contém como feições polígonos de solos, onde é possível, com dados de latitude e longitude de pontos desejados, visualizar onde esses pontos se encontram, em que polígono, e, dessa forma, identificar qual o tipo de solo associado ao ponto desejado. A Tabela 5 apresenta os atributos e sua descrição, e a Figura 22 apresenta

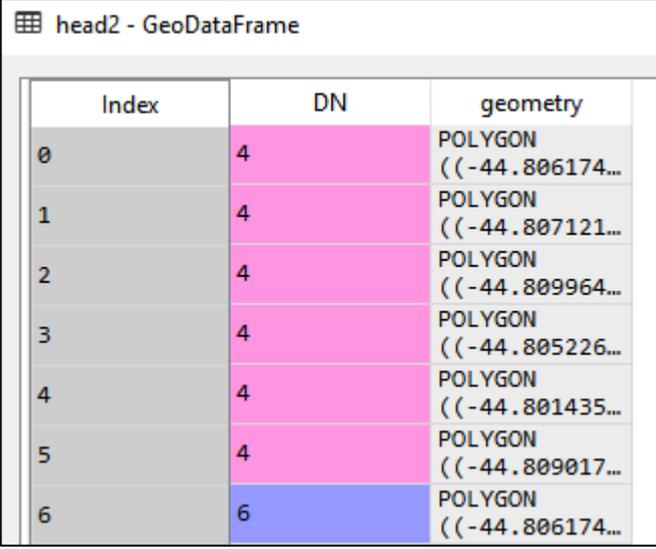
a leitura do *shapefile* pelo algoritmo, onde é possível visualizar as informações descritas na tabela.

Tabela 5 - Descrição dos Componentes Gráficos.

Nome do Atributo	Descrição	Natureza do dado
<i>index</i>	Posição de cada valor da tabela, n° de polígonos na tabela (1,704,190)	Qualitativa - Nominal
DN	Percentual de desnível da área contida no polígono. (de 1 a 6)	Quantitativo - Numérico
<i>geometry</i>	Feição do <i>shapefile</i> : polígono de solos	Quantitativo – Coordenadas de Polígonos

Fonte: Elaborado pelo Autor.

Figura 22 – Leitura do arquivo *shapefile*.



Index	DN	geometry
0	4	POLYGON ((-44.806174...
1	4	POLYGON ((-44.807121...
2	4	POLYGON ((-44.809964...
3	4	POLYGON ((-44.805226...
4	4	POLYGON ((-44.801435...
5	4	POLYGON ((-44.809017...
6	6	POLYGON ((-44.806174...

Fonte: Elaborado pelo Autor.

3.1.3 Dados de Entrada – Dados de CAD

A capacidade de água disponível (CAD ou AWC) dos solos no Brasil pode ser adquirida numa base de dados disponibilizada e realizada pela ANA e pela Universidade Federal do Paraná (UFPR), como uma das atividades do Atlas Irrigação: Uso da Água na Agricultura Irrigada. A umidade retida pelo solo é utilizada como critério no dimensionamento de projetos de irrigação e em aplicações diversas nas áreas da agronomia e hidrologia, que incluem estudos de: balanço hídrico, disponibilidade de água às plantas, infiltração, condições de drenagem, condutividade hidráulica, estresse hídrico e movimento de solutos no solo (Metadados INDE, 2022).

Tradicionalmente, a capacidade de campo (CC) e o ponto de murcha permanente (PMP) são considerados como os limites máximo e mínimo, respectivamente, de umidade retida nos solos que se encontra disponível para as plantas. A partir destes limites pode-se estimar a capacidade de água disponível no solo. Define-se a capacidade de água disponível como sendo o volume máximo de água no solo que está disponível para as plantas quando a umidade do solo se encontra na capacidade de campo, desconsiderando-se o volume que fica indisponível na condição de umidade de murcha permanente (ANA; UFPR, 2021).

O arquivo mais recente data de 2021, é um documento digital em formato de tabela, possui um modelo de mapa vetorial, contudo, para este trabalho, o arquivo de entrada utilizado foi no formato planilha “.xlsx”. A descrição dos atributos encontrados na planilha é apresentada na Tabela 6.

Tabela 6 – Descrição dos Atributos da Entrada de CAD.

Nome do Atributo	Descrição	Natureza do dado
Código	Código de identificação do município.	Qualitativo - Numérico
Município	Identificação do município.	Qualitativo - Nominal
UF	Sigla da unidade federativa referente ao estado a que o município pertence.	Qualitativo - Nominal
CAD ou AWC	Valor em m^3/m^3 ou % da capacidade de água disponível mapeada para o município.	Quantitativo - Numérico

Fonte: Elaborado pelo Autor.

3.1.4 Dados de Entrada – Dados de Fenologia e Mapeamento Tecnológico

Os dados de entrada referentes à fenologia foram adquiridos pela pesquisa bibliográfica do autor, apresentada no capítulo anterior. As informações foram armazenadas em planilha formato “.xlsx”, juntamente com os dados do mapeamento tecnológico referentes ao sistema irrigante mapeado para a região de estudo, Tabela 7.

Os dados de Kc são os apresentados na revisão bibliografia, e serão associados a cada mês de referência. A profundidade da raiz da planta de café é utilizada para calcular o CAD da cultura, dado pela equação (2), neste trabalho foi utilizado o valor de profundidade igual a 500 mm, um valor encontrado também na bibliografia. O tipo de irrigação mapeado para a região de mogiana foram do tipo gotejamento, os dados de eficiência e pressão de serviço são atribuídos por tipo de sistema irrigante.

$$CAD_{cultura} = CAD \times Profundidade \quad (2)$$

Tabela 7 – Descrição dos Atributos da Entrada de Fenologia e Mapeamento Tecnológico.

Nome do Atributo	Descrição	Natureza do dado
Cultura	Identificação do tipo de cultura – (no caso de estudo: Café)	Qualitativo - Nominal
Kc_Jan....Kc_Dez. Kc_Jan_1...Kc_Dez_1	24 colunas de valores de Kc da cultura de Café para o período bienal.	Quantitativo – Numérico
Profundidade	Valor da profundidade da raiz da planta de café.	Quantitativo – Numérico
Tipo_Irrigacao	Tipo de sistema de irrigação.	Qualitativo - Nominal
EFICIENCIA	Eficiência do tipo de sistema de irrigação.	Quantitativo – Numérico [%]
Pressão_De_Servico	Pressão de serviço do sistema irrigante.	Quantitativo – Numérico [mca]

Fonte: Elaborado pelo Autor.

3.1.5 Dados de Entrada – Dados Meteorológicos

Os dados de entrada de natureza meteorológica foram adquiridos na plataforma CIIAGRO do Portal Agrometeorológico e Hidrológico do Estado de São Paulo, onde é

possível acessar as informações do monitoramento climático, com dados de precipitação pluviométrica e evapotranspiração, que serão utilizados nos cálculos de estimativa de consumo, esses dados podem ser adquiridos por município e período desejado. Foram extraídas as informações para os municípios da região de estudo: Cristais Paulista, Pedregulho, Franca, Ribeirão Corrente, Jeriquara e Patrocínio Paulista. Os dados foram armazenados em planilha “.xlsx”. A descrição dos atributos desse dado de entrada é apresentada na Tabela 8.

Tabela 8 - Descrição dos Atributos da Entrada de Fenologia e Mapeamento Tecnológico.

Nome do Atributo	Descrição	Natureza do dado
MES	12 linhas referentes ao mês.	Qualitativo - Numérico
ETo	Valores de evapotranspiração potencial medido para cada mês do período de estudo.	Quantitativo - Numérico
PRECIPITACAO	Valores de precipitação medido para cada mês do período de estudo.	Quantitativo – Numérico [mm]
DESC_MUNICIPIO	Nome do município em que o consumidor está localizado.	Qualitativo - nominal

Fonte: Elaborado pelo Autor.

3.1.6 Dados de Entrada – Dados de Áreas

Junto ao estudo de estimativa de consumo de consumidores irrigantes e identificação de possíveis ocorrências de PNT, está sendo realizado o estudo de identificação de áreas de lavouras e o tipo de plantio da área. O dado de entrada de áreas de cultivo de café utilizado nesse trabalho é um resultado desse estudo, onde são fornecidas as coordenadas de latitude e longitude dos centroides das áreas de plantio identificadas, a área em hectares (ha), o tipo de cultura, a as UC mais próximas dessa área identificada (utilizando as coordenadas geográficas do cadastro das UC), e o município onde a área está localizada, descritos na Tabela 9.

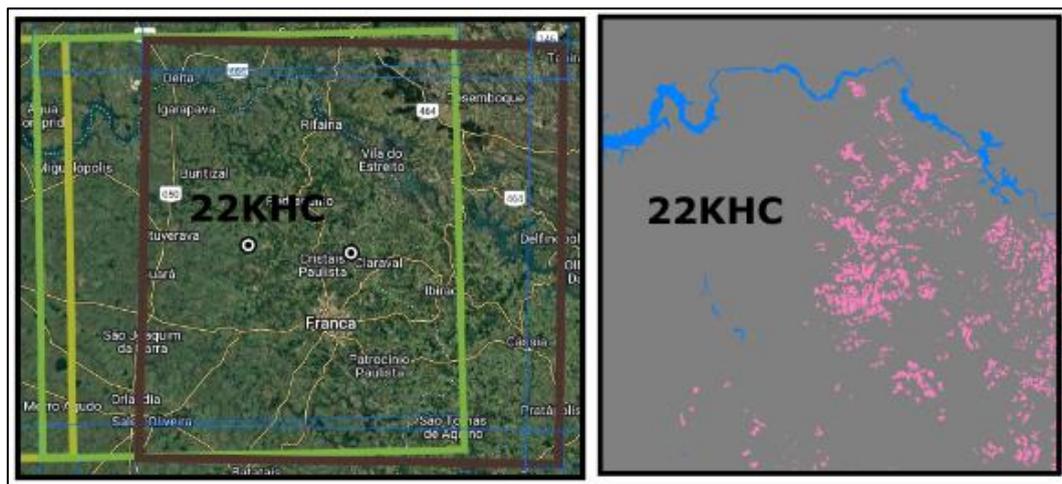
Tabela 9 - Descrição dos Atributos da Entrada de Áreas.

Nome do Atributo	Descrição	Natureza do dado
UC_1, UC_2, UC_3	3 colunas com o código de identificação das 3 UC mais próximas do centroide da área.	Qualitativo - Numérico
Distância_UC1, Distância_UC2, Distância_UC3	Distâncias medida das coordenadas de latitude e longitude das UC até o centroide da área.	Quantitativo - Numérico
point_lavoura	Valores de latitude e longitude do centroide da área identificada.	Quantitativo – Numérico
DESC_MUNICIPIO	Nome do município em que o consumidor está localizado.	Qualitativo - Nominal
Área	Valor calculado do talhão da área de cultivo identificada.	Quantitativo – Numérico [ha]
Cultivar	Tipo de planta cultivada	Qualitativo - Nominal

Fonte: Elaborado pelo Autor.

Para aquisição dos dados de áreas foram adquiridas, pela equipe de processamento de imagens, cenas referentes à região de plantio de café de alta tecnologia, apresentada na revisão bibliográfica, a região de mogiana em São Paulo. Onde, foram identificadas áreas de plantio e classificadas pelo seu tipo de cultura. Para este trabalho selecionou-se apenas as assinaladas como café. Como o processamento de imagens não é o foco deste trabalho, não é detalhado o processo, a Figura 23 apresenta a cena (22KHC) selecionada e o resultado na identificação de cultivo do café.

Figura 23 – Área de aplicação da metodologia.



Fonte: Elaborado pelo Autor.

3.2 LIMITAÇÕES ENCONTRADAS E CONSIDERAÇÕES

Algumas considerações foram realizadas antes de dar início a etapa de pré-processamento dos dados, na medida que algumas limitações foram encontradas na modelagem e aquisição dos dados de entrada. A metodologia propõe o cálculo da estimativa de consumo mensal de consumidores, para tanto, utiliza-se de dados do manejo do sistema irrigante característico para o tipo de lavoura identificado, conseguinte, os melhores resultados de estimativa seriam obtidos com dados pontuais de cada UC cadastrada. Contudo, como não é possível identificar os dados individuais de cada fazenda de café, o estudo utiliza de dados considerados comuns da área geográfica de estudo (mapeados pela revisão bibliográfica e estudo da região), além de valores médios para alguns atributos, que serão apresentados posteriormente.

Desse modo, para identificação de possível ocorrência PNT é utilizado um porcentual de variação de consumo real com o estimado, considerado aceitável pelo autor.

Na etapa de associação de áreas são feitas as seguintes considerações: (i) A área mais próxima de uma coordenada de UC não é 100% assegurada de que pertence à esse irrigante, pois a distribuição rural no Brasil não segue padrões facilmente identificáveis de propriedade e proprietário quando as coordenadas de UC estão muito próximas; (ii) áreas identificadas podem não estão associadas a nenhuma UC cadastrada na base de dados da CPFL; (iii) duas coordenadas, ou mais, do ponto de distribuição das UC podem pertencer

a um mesmo proprietário de fazenda, ou seja, uma área identificada pode possuir mais de uma UC cadastrada; (iv) Uma centroide não é necessariamente uma fazenda, dois ou mais centroides podem compor uma única fazenda, cujos talhões foram divididos no algoritmo de processamento de imagens para calcular as áreas.

Essas considerações foram levadas em conta na etapa de pré-processamento e de metodologia para associação de UC com áreas de plantio, fazendo-se um breve estudo, utilizando o Google Earth, para identificar uma distância mínima característica para fazendas de café a ser considerada na associação de UC e centroide de área. Essa correta alocação é importante, pois UCs com áreas maiores tenderão a ter maior consumo faturado de energia elétrica proveniente do seu manejo do sistema de irrigação, logo, UCs com áreas menores terão menor consumo faturado de energia elétrica.

3.3 PRÉ-PROCESSAMENTO DE DADOS

Para a etapa de pré-processamento dos dados de entrada, foram eliminados dados com erros de cadastro, nulos, ou com valores NaN. Nesta etapa, também, foram gerados novos atributos com base em atributos já existentes para compor a metodologia aplicada, como por exemplo dados de média e desvio padrão de consumo de energia, para posterior utilização. A etapa de pré-processamento foi realizada utilizando-se do software *Spyder*, plataforma de desenvolvimento em linguagem de programação *python*, para o pré-processamento dos dados de desnível de relevo da região de estudo, utilizou-se do ArcGis. Nesta seção são descritos os processos aplicados por dado de entrada.

- **Consumo e Cadastrais**

O pré-processamento de dados de consumo e cadastrais teve como objetivo identificar UC que apresentassem dados de cadastro nulos, ausentes ou incorretos, além de possibilitar possíveis a identificação de correlações de atributos presentes na base de dados fornecida pela distribuidora, melhorando a percepção nas informações dos dados.

A base de dados fornecida pela distribuidora possui todos os consumidores cadastrados como irrigantes de período reservado, ou seja, irrigantes com cuja tarifa é menor, como incentivo ao manejo do sistema de irrigação no horário reservado (8h30 do dia que abrange toda a madrugada).

O total de amostras é de 2731 UC, localizados em 189 municípios do estado. O primeiro passo desta etapa foi de selecionar, pelo algoritmo, apenas os municípios da região de estudo (Cristais Paulista; Pedregulho; Franca; Ribeirão Corrente; Jeriquara; Patrocínio Paulista), o que totaliza 65 UC para aplicação, logo, da base de dados cadastrais foram selecionadas 65 UC e seus respectivos atributos de latitude, longitude, código de instalação, código de grupo de tensão, código do alimentador, além do atributo de pré-seleção: município (DESC_MUNICIPIO).

Referente aos dados de histórico de consumo, trata-se de uma base de dados contendo 108973 linhas de código de UC, juntamente com colunas de consumos desde o ano de 2003 a janeiro de 2021, e sua respectiva coluna de data de referência, no formato:200301, 200302...202101. O dado de consumo é em kWh. O pré-processamento desses dados foi realizado para selecionar apenas as UC de interesse (65 amostras) e seus respectivos consumos dentro de um período de 72 meses, ou seja, foi utilizado 6 anos de consumo (201501 - 202101). Destaca-se que os consumos faturados na base de dados estão “deslocados” +1, ou seja, na análise de curva de consumo x consumo esperado, consumo x variáveis climáticas, devemos considerar que o consumo registrado no mês de janeiro, por exemplo, refere-se ao montante de energia elétrica utilizado no mês anterior: dezembro do ano antecedente.

Entre as 65 UC de estudo, uma foi eliminada no processo de associação do código de UC da base de dados cadastrais com as bases de consumo, significando que esta UC não estava presente no arquivo de consumo histórico do grupo A de tensão, nem no do grupo B, restando 64 UC. Dentre essas UC, 26 são do grupo de tensão A e 38 do grupo de tensão B.

Considerando a seleção de amostras de UC desejadas, pelo município, e a associação de seus consumos históricos, como sendo a primeira etapa do pré-processamento da base de dados de consumo e cadastrais. A segunda etapa trata-se da geração de novos atributos que possam melhor expressar o comportamento dos dados, e que possam ser utilizados na aplicação de técnicas de aprendizado de máquina da metodologia para formar agrupamentos de consumidores semelhantes. Dessa forma, escolheu-se gerar quatro novos atributos baseados nos consumos: média móvel, desvio padrão móvel, mínimo de consumo registrado móvel e máximo de consumo registrado móvel, para um período de 4 meses. Esses quatro novos atributos serão utilizados como indicadores de perfil de consumo das UC. O próximo passo, portanto, foi de selecionar os 73 meses escolhidos. Sendo assim, depois de inserir os novos atributos, são removidos da

nova base de dados os consumos anteriores ao ano de 2015. Resultando em uma nova estrutura de dados contendo 4672 linhas (64 UC x 73 meses) e 13 colunas, essa planilha de dados é no formato de linhas, seguindo a estrutura apresentada na Tabela 10.

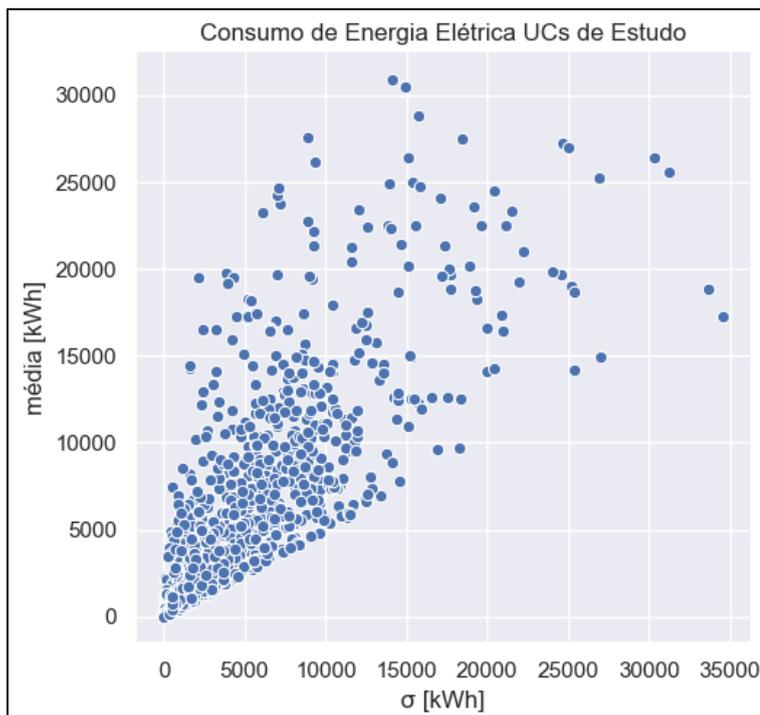
Tabela 10 – Exemplificação da Estrutura de dados de consumo e cadastrais.

COD_INSTALACAO	Atributos de Cadastro		DT_REFERENCIA	CONSUMO	Media	Std	Min	Max
xxxxx123			201501					
xxxxx123			201502					
xxxxx123			201503					
xxxxx123			201504					
...			
xxxxx123			202101					
12xxx45			201501					
12xxx45			201502					
12xxx45			201503					

Fonte: Elaborado pelo Autor.

Utilizando os atributos de média móvel de consumo e desvio padrão móvel do período definido (201501....202101) é possível representar graficamente os dados, Figura 24. Verifica-se que a média de consumo das UC se manteve abaixo de 15 GWh, enquanto tem-se poucas ocorrências de UCs com desvio de consumo de energia elétrica acima de 15 GWh.

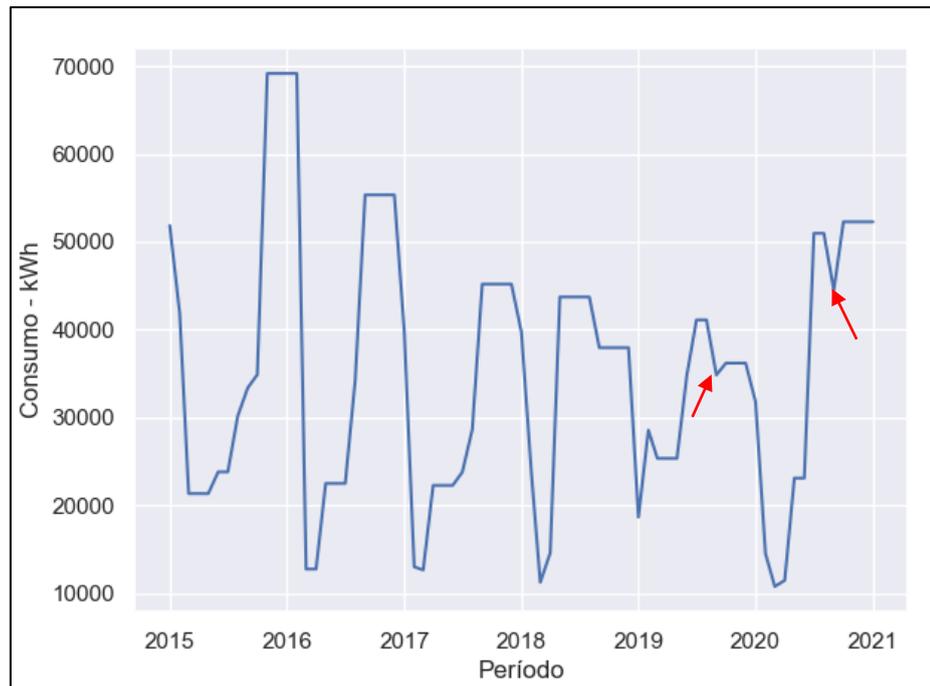
Figura 24 - Desvio Padrão vs Média de Consumo para cada UC (72 meses).



Fonte: Elaborado pelo Autor.

Quanto ao atributo referente ao máximo de consumo de energia pelas UC no período de 4 meses, a Figura 25 apresenta quais foram os valores de máximos móvel de consumo de cada mês (72). Observa-se que nos anos de 2015 e 2016 teve-se um alto consumo de energia, decaindo no ano de 2017, o ano de 2020 (especificamente o mês 5) o maior valor de energia consumida por uma das UC foi a menor dentre os valores. Também, indicados na figura estão o mês 09 de 2019 e 2020, percebe-se que esse mês, dentro da janela móvel de 4 meses marcou mudanças no consumo de energia.

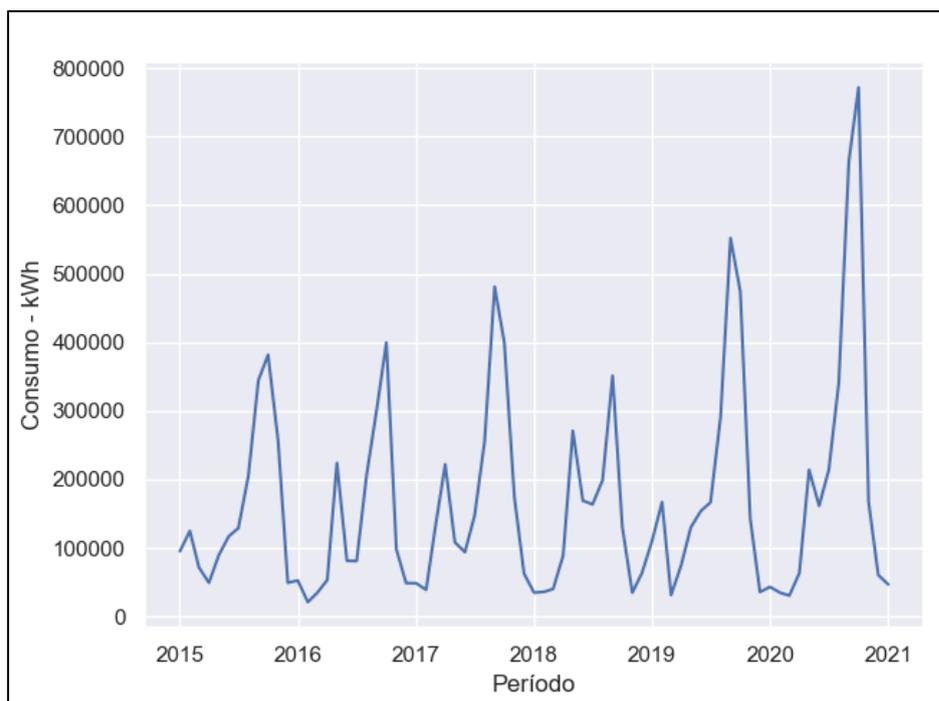
Figura 25 – Máximo Consumos de Energia Registrado por Mês.



Fonte: Elaborado pelo Autor.

Também foi plotado o gráfico que apresenta o total de energia elétrica que foi consumida em cada mês pelas 64 UC, Figura 26. Nesta figura podemos observar que ocorre picos de consumo de energia entre os meses 09 e 10 de cada ano, variando em amplitude de consumo, e, no ano de 2020 foi quando ocorreu o maior consumo de energia na região. Finalizado os estudos gráficos dos dados de consumo, termina-se a etapa de pré-processamento desse dado de entrada.

Figura 26 – Total de Energia Consumida por Período.



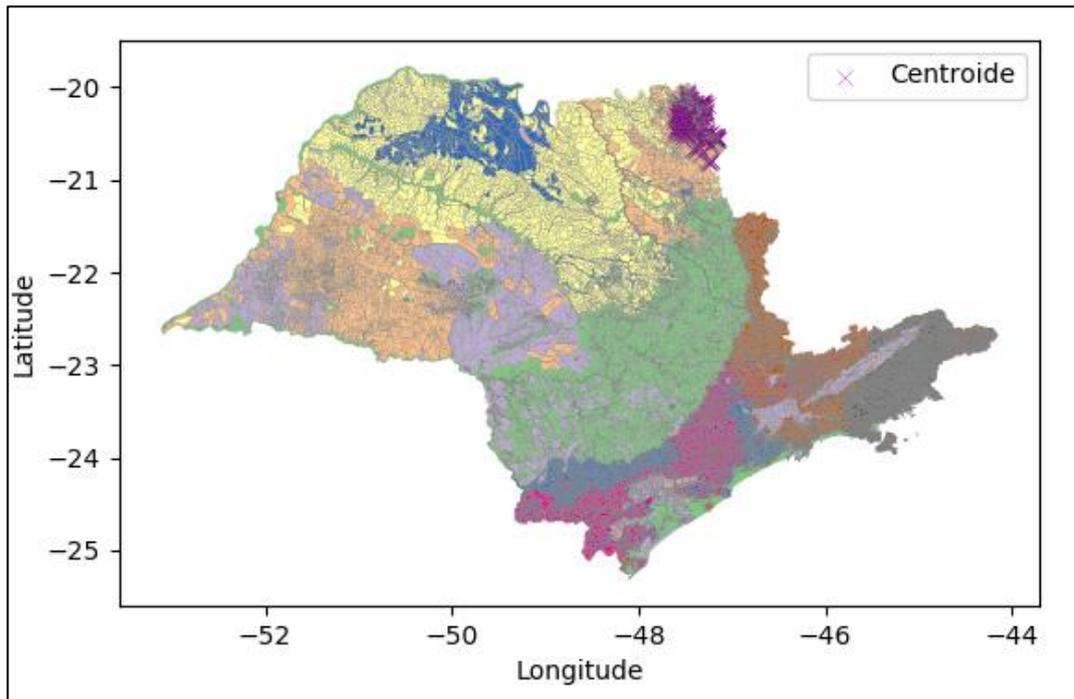
Fonte: Elaborado pelo Autor.

- **Dados de Áreas**

Os dados de áreas adquiridos são provenientes do estudo da zona rural de São Paulo, fornecendo as informações apresentadas na Tabela 9. Nata etapa de pré-processamento desses dados foram utilizados os atributos de descrição de município e do tipo de cultivar, selecionando as áreas presentes apenas na região de estudo e do tipo café. O resultado é apresentado na Figura 27, totalizando 472 centroides. Plotando o resultado juntamente com as coordenadas geográficas das UC irrigantes, Figura 28, verificou-se que há certas localidades que possuem vários centroides e nenhuma UC (próximo à fronteira com Minas Gerais) e duas UC, de mesma coordenada, consideravelmente afastadas de possíveis áreas de plantio de café, que podem ser descartadas (assinalada em vermelho).

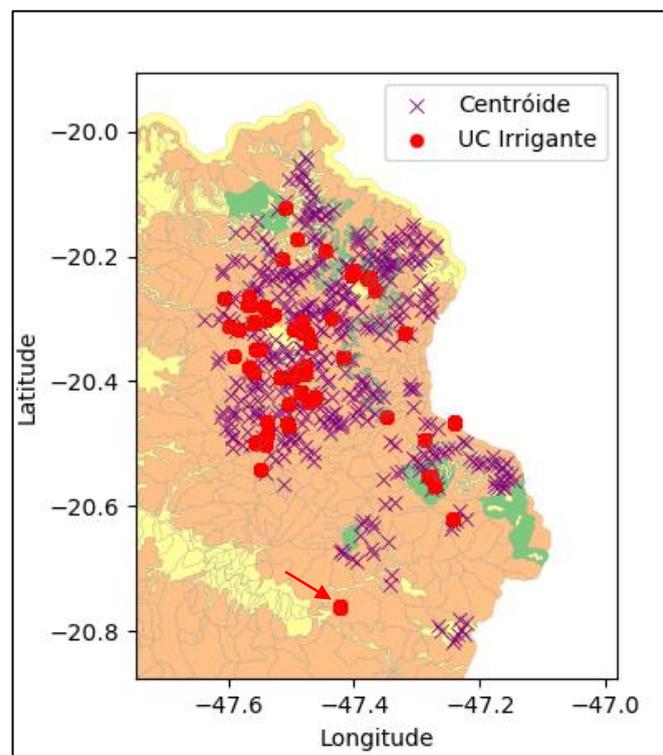
Com isso, o segundo passo do pré-processamento é filtrar os centroides, que possivelmente pertencem à alguma UC, delimitando a região de estudo. Para isto, foram importados no Google maps as coordenadas de ambos dados, com o objetivo de inspecionar a relação entre os dois, e, portanto, encontrar os centroides que estejam muito distantes das unidades consumidoras cadastradas, bem como as unidades consumidoras sem proximidade com as áreas encontradas.

Figura 27 - Mapa de São Paulo e Centroides de Áreas de Plantio de Café.



Fonte: Elaborado pelo Autor.

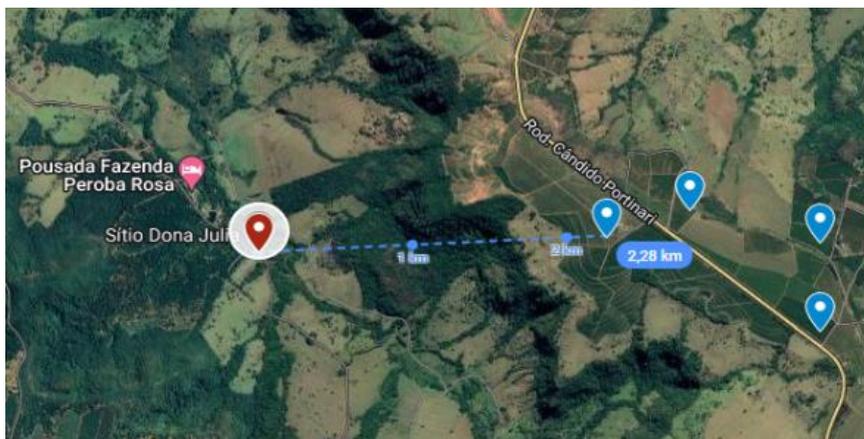
Figura 28 – Centroide de Áreas de Plantio de Café e Consumidores Irrigantes.



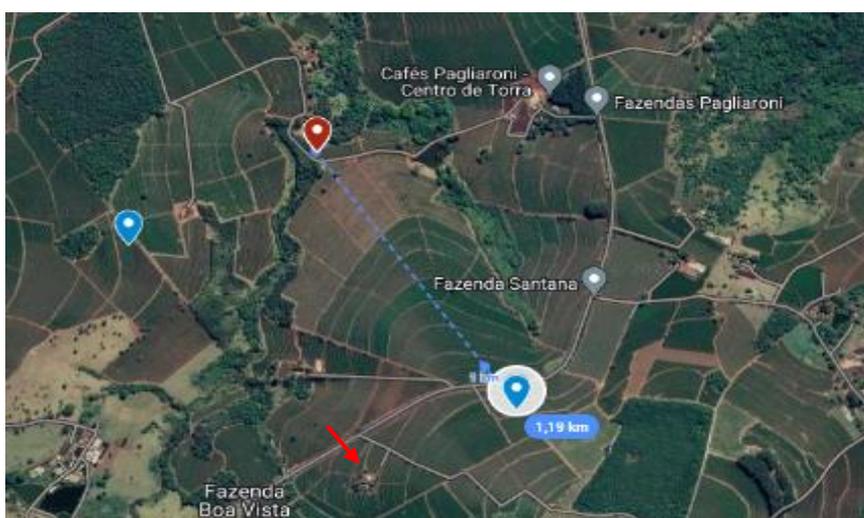
Fonte: Elaborado pelo Autor.

Pela análise das imagens, verificou-se que os centroides de áreas e as coordenadas de UC devem estar, no máximo, 1,10 km distantes entre si. Se a UC mais próxima de um centroide de área ultrapassa a faixa de valor, muito provavelmente, não existe correlação entre os dois, a Figura 29a exemplifica um dos casos encontrados, onde a UC do irrigante mais próximo da lavoura, detectada pelo processamento de imagens, ultrapassa os 2 km, além de possuir obstáculos de área verde entre os dois, e na Figura 29b, ocorre que o centroide de áreas é pertencente à um agricultor que não possui cadastro na distribuidora (assinalado em vermelho), contudo será associado a UC mais próxima que está a 1,19 km de distância. Dessa forma, foram eliminadas dos dados quaisquer relações de centroide x UC que ultrapasse o valor determinado pelo estudo da região.

Figura 29 – Exemplo 1 de Relação Centroide x UC não correspondentes.



- a) Caso 1: UC mais próxima de centroides distantes e com delimitação de área verde.



- b) Caso 2: Centroide mais próximo de UC cadastrada que provavelmente pertencente a outro proprietário, não pertencente a base de dados da distribuidora.

Fonte: Elaborado pelo Autor.

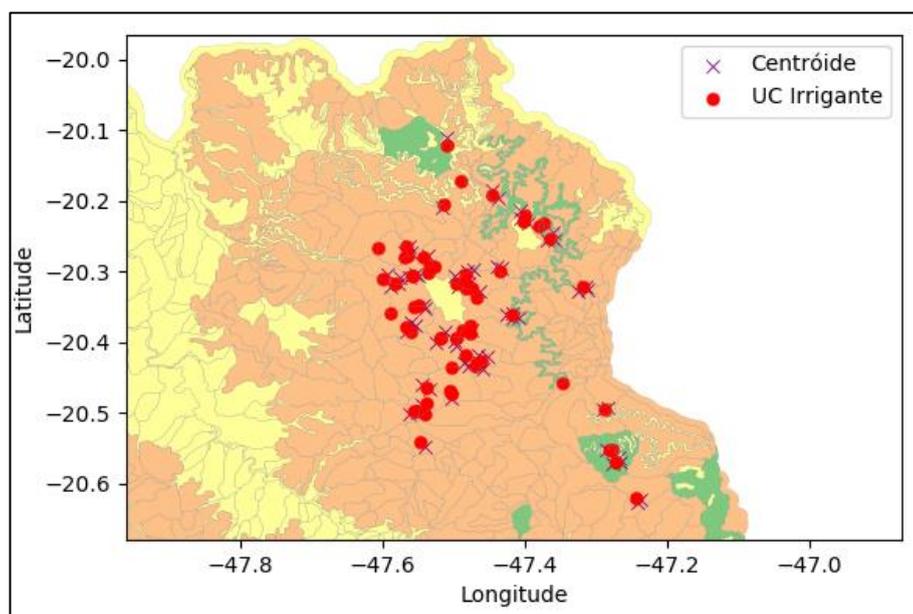
Eliminando das amostras de UC x Áreas as UC muito distantes de qualquer centroide, resulta em 436 valores. Desse resultado, tem-se as distâncias entre centroide e UC mais próxima, descritas na Tabela 11. Desse modo, as áreas de interesse são as 70, dentro do limite estabelecido $D < 1,2$ km. O total de UC de estudo, no final desta etapa de pré-processamento, foi de 43, e o total de centroides igual a 70, a distribuição espacial é apresentada. A Figura 30, apresenta geograficamente o resultado desta etapa, onde é possível visualizar a distribuição de centroide de área e UC a serem utilizadas na aplicação da metodologia.

Tabela 11 – N° de Casos Por Faixa de Distâncias.

Distâncias entre Área e UC	N° de Casos	Decisão
$D > 10$ km	16	Eliminar
$5 \text{ km} < D < 10$ km	99	Eliminar
$1,2 \text{ km} < D < 5$ km	251	Eliminar
$D < 1,2$ km	70	Manter
Total	436	

Fonte: Elaborado pelo Autor.

Figura 30 - UC e Centróide de Área resultantes do pré-processamento.



Fonte: Elaborado pelo Autor.

3.4 APLICAÇÃO DO ALGORITMO DE CLUSTERIZAÇÃO E PREDIÇÃO

Finalizando a etapa de pré-processamento dos dados de entrada, tem-se um novo conjunto de dados para ser utilizado na implementação do modelo proposto neste trabalho. Nesta seção serão apresentados os resultados da aplicação dos algoritmos e ferramentas utilizados no desenvolvimento dessa metodologia, descritos no referencial deste trabalho. Referentes aos algoritmos de aprendizado de máquina, foram utilizados: *Kmeans* como método de clusterização; e *Random Forest* para predição de classificações.

3.4.1 Clusterização por K-Means

O K-means foi implementado neste trabalho para gerar clusters que indicassem os perfis no consumo de energia elétrica de cada UC, durante o período dos 72 meses desejados, sendo utilizados os indicadores inseridos no banco de dados na etapa de pré-processamento, apresentados anteriormente: Média, Máximo, Mínimo e Desvio. Outra implementação do K-means foi para agrupar os centroides de áreas com as coordenadas das UC, gerando clusters de áreas.

O algoritmo de clusterização K-means pode ser também chamado de K-médias, segundo Jain et. al (1999) o algoritmo K-means é popular devido a sua facilidade de implementação e sua ordem de complexidade $O(Nkt)$, onde N é o número de padrões, k é o número de grupos e t é o número de iterações. De acordo com Fontana e Naldi (2009), K-means utiliza o conceito de centroides como protótipos representativos dos grupos, onde o centroide representa o centro de um grupo, sendo calculado pela média de todos os objetos do grupo.

O agrupamento K-means particiona o espaço de dados na representação de células de Voronoi. Essa transformação divide as observações de dados em k-clusters em que cada uma das observações pertence ao cluster com a média mais próxima (HAROON, 2017). O agrupamento K-means é feito da seguinte forma:

1. Os k-centroides são escolhidos aleatoriamente;
2. Cada uma das observações é vinculada ao centroide mais próximo;
3. Novos centroides para cada um dos clusters são recalculados por tomando a média das observações dentro de cada cluster;
4. O passo 2 é repetido.

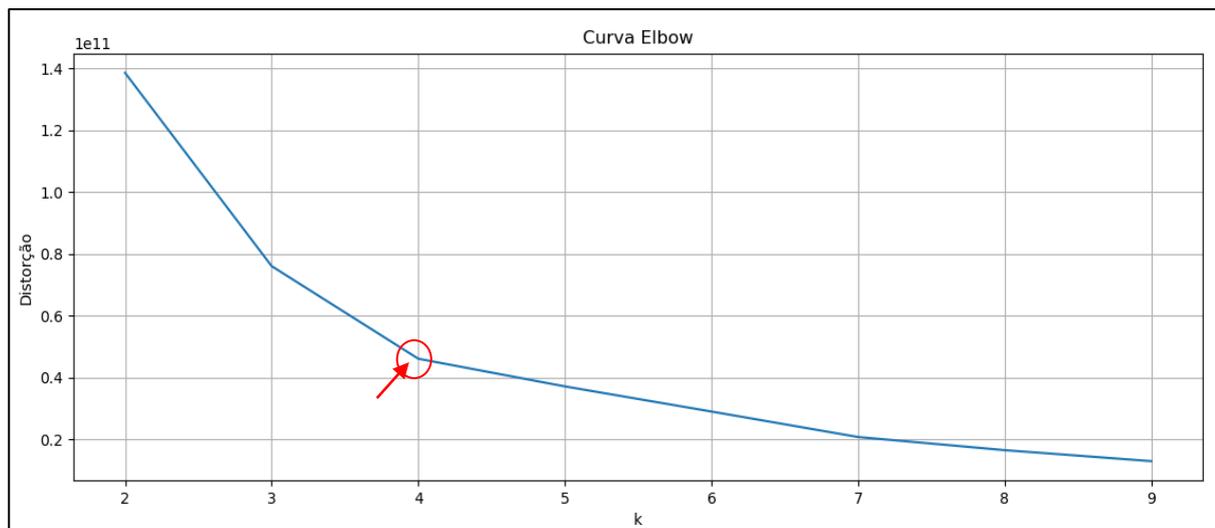
Nota-se que as etapas 1 e 3 se assemelham, a diferença é que na etapa 1 os centroides são escolhidos aleatoriamente, enquanto na etapa 3 eles são calculados pela média das observações dentro de cada um desses grupos. Os passos 2 e 4 são uma repetição do mesmo passo. As etapas 3 e 4 são iteradas até que os clusters convirjam, ou seja, as associações de cluster permanecem constantes para todas as observações no conjunto de dados. Isso maximiza a distância entre os clusters e minimiza a distância entre as observações dentro de cada cluster.

O agrupamento K-means nem sempre encontra a configuração mais ideal enquanto minimiza a função objetivo global. O algoritmo de agrupamento é altamente sensível a como os centros de agrupamento são inicialmente selecionados. Jain et. al. (1999) afirmam que, um dos maiores problemas do algoritmo K-means é que o mesmo é sensível à seleção da partição inicial (k) e pode convergir a um mínimo local do valor da função de critério se a partição inicial não for devidamente escolhida. Dentre os métodos de escolha do valor de k, temos o método Elbow (cotovelo).

No método *Elbow*, a porcentagem de variância é dada em função do número de clusters. Ele determina quanta variância marginal é contribuída por um cluster recém-adicionado. Um ponto de interesse é que o primeiro cluster explicará a variância máxima com ganhos marginais caindo na adição de cada novo cluster. O “Cotovelo” será o ponto em que um novo cluster resultará em uma queda marginal considerável, sendo esse o número ótimo de clusters. Esse método foi aplicado no conjunto de dados para definir o número ótimo de clusters para gerar os Perfis de Consumo, proposto na metodologia para agrupar unidades consumidoras com curva de consumo semelhantes.

3.4.2 Resultado da Clusterização – Geração de Classes de Perfil de Consumo

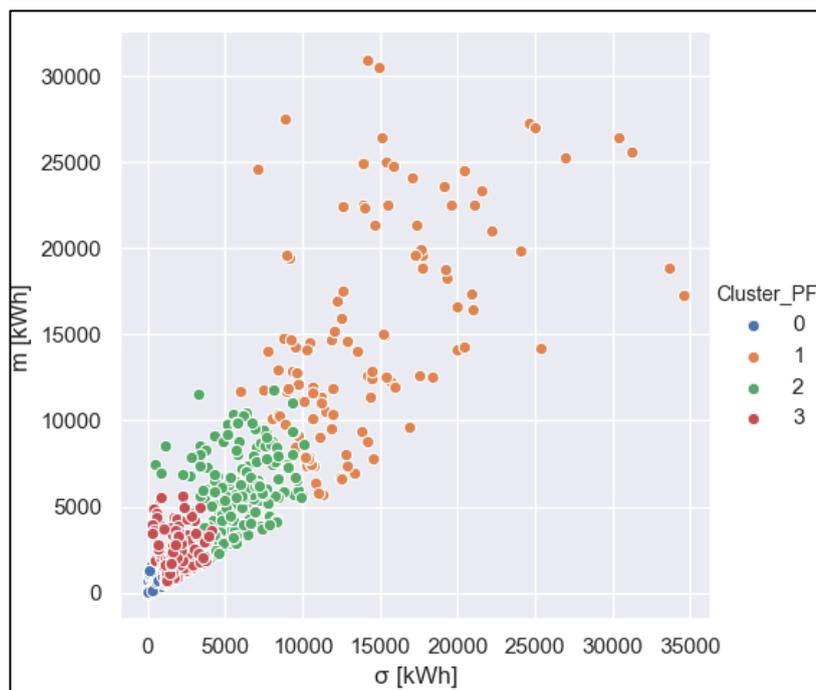
O código utilizado para clusterização das UC e o seu resultado é apresentado nesta seção. Foi utilizado o método *Elbow* para definir o melhor valor de k para o conjunto de dados, e assim o utilizar na aplicação do algoritmo K-means. Figura 31, ilustra o resultado da curva de *Elbow*, onde o ponto de inflexão da curva ocorre para $k = 4$.

Figura 31 – Resultado do Método *Elbow* para definir k.

Fonte: Elaboração Própria.

Com k definido como sendo 4, serão gerados 4 Perfis de consumo utilizando como dados de entrada (X) os valores de média de consumo móvel, máximo móvel, mínimo móvel e desvio móvel, a saída esperada (Y) é uma matriz NumPy de inteiros variando de 0 a 3 (4 classes), onde a ordem do *index* da matriz é a mesma ordem dos dados à que são referentes, em X. O resultado é apresentado na Figura 32.

Figura 32- Resultado da Etapa de Clusterização por Perfil de Consumo.



Fonte: Elaboração Própria.

Podemos perceber pela Figura 32, que poucas UCs foram agrupadas na Classe 0, e as UCs com médias de consumo maiores foram agrupadas na Classe 1. Enquanto, que, UCs com consumos intermediários foram agrupadas nas Classes 2 e Classe 3. Ressaltando que, uma UC pode ter seus valores de consumo no mês “x” em uma classe, e no outro mês apresentar o comportamento de outra classe. A ideia é, posteriormente, utilizar as classes dos meses de interesse do ano anterior (2020), e comparar com o resultado preditivo ao ano de estudo (2021), identificando possíveis divergências.

A Figura 33 apresenta como se deu a distribuição de amostras da base de dados por cluster.

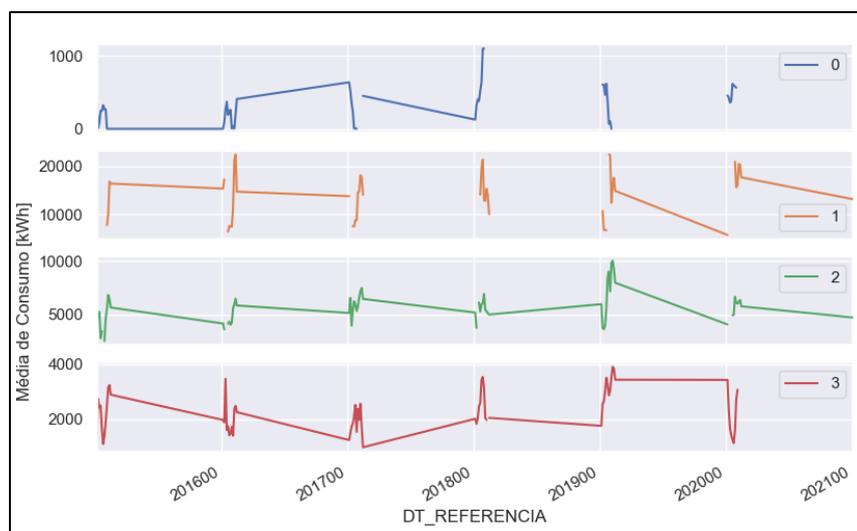
Figura 33 – N° de amostras por Cluster Gerado.

Classes	N° de UCs
0	1916
3	643
2	433
1	147

Fonte: Elaboração Própria, *Dataframe* auxiliar gerado pelo algoritmo.

Na Figura 34, são ilustrados os gráficos que representam o perfil de consumo das UCs que as integram, com valores de média de consumo e data de referência.

Figura 34 – Representação Gráfica dos Clusters de Perfil de Consumo.



Fonte: Elaboração Própria.

Pode-se perceber, pela Figura 34, que: a Classe 0 engloba as UCs que por algum motivo não tiveram consumo registrado; a Classe 1 também apresenta alguns meses sem registro do faturamento de energia, contudo, com média móvel com valores maiores do que a Classe 1; a Classe 2 apresenta consumos com amplitude menores do que a Classe 3, e, em certos meses.

3.4.3 *Random Forest* – RF

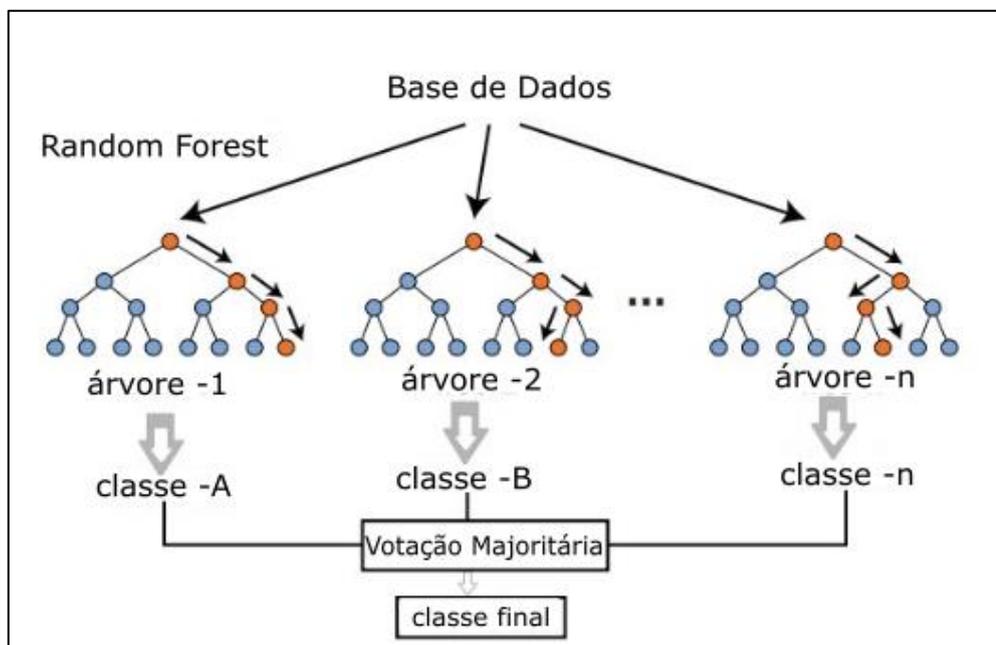
O algoritmo de florestas aleatórias foi desenvolvido pelo falecido professor de Berkeley Leo Breiman e Adele Cutler. *Random Forest* (RF) geram sua sequência de modelos treinando-os em subconjuntos de dados. Os subconjuntos são extraídos aleatoriamente do conjunto de treinamento completo. Uma maneira pela qual o subconjunto é selecionado é amostrar aleatoriamente linhas com substituição da mesma maneira que o algoritmo de agregação *bootstrap* de Brieman. O outro elemento aleatório é que os conjuntos de treinamento para as árvores individuais no conjunto de florestas aleatórias não incorporam todos os atributos, mas também recebem um subconjunto aleatório dos atributos (BOWLES, 2015).

RF é uma combinação de ensacamento (*bagging*) e uma modificação de seleção de atributos aleatórios para os aprendizes de base de árvore binária. Essas diferenças podem não parecer substanciais, mas dão às florestas aleatórias características de desempenho diferentes de ensacamento e aumento de gradiente. Alguns resultados sugerem que RF têm uma vantagem com amplos espaços de atributos esparsos, como ocorre em problemas de mineração de texto. RF são um pouco mais fáceis de paralelizar do que o aumento de gradiente porque os alunos de base individuais podem ser treinados independentemente um do outro, enquanto, com o aumento de gradiente, cada aluno de base precisa dos resultados dos anteriores (BOWLES, 2015). Diferenças como essas significam que você pode querer experimentar as duas florestas aleatórias, além do aumento de gradiente, se precisar extrair o máximo de desempenho possível dos dados.

Na classificação de árvores de decisão, diferentes sub-árvores podem ter muitas semelhanças estruturais que podem resultar em saídas de previsão fortemente correlacionadas entre si. O classificador de RF reduz essa correlação entre as sub-árvores limitando os recursos em cada ponto de divisão. Assim, em vez de escolher uma variável de todas as variáveis disponíveis, a floresta aleatória procura a variável que minimizará o erro de uma amostra aleatória limitada de recursos (HAROON, 2017).

A Figura 35 ilustra o RF e as sub-árvores de decisão geradas para a predição, resumidamente, o resultado de todas as sub-árvores são combinados para chegar ao resultado final. Uma vantagem de utilizar o algoritmo RF é que, além dele ser muito poderoso, ele retorna de maneira muito compreensiva a importância atribuída para cada variável independente na hora de realizar as predições. O RF foi utilizado na metodologia para realizar as predições de classificação das UC, tendo como alvo as classes geradas na etapa de clusterização.

Figura 35- Ilustração de Classificador de *Random Forest*.



Fonte: Adaptado de Haroon (2017, p.191).

3.4.4 Modelo RF Classificatório

Com a etapa de agrupamento por classes de Perfil de Consumo finalizada, é implementado o modelo RF da metodologia. Como apresentado na pesquisa bibliográfica, podemos dizer que os meses de interesse para a irrigação do café ocorre em três etapas: primeiro ano fenológico (fevereiro até antes de julho); setembro e dezembro de ambos os anos fenológicos; e, segundo ano fenológico (janeiro a março). Unindo essa informação com o da análise gráfica da curva de consumo do conjunto de dados, em que ocorre picos de consumo nos meses de setembro e outubro e as mudanças observadas que ocorrem após o mês de setembro. Decidiu-se utilizar apenas esse período crucial para a análise de divergência de consumo.

Dessa maneira, no modelo RF foram utilizadas como entradas os consumos de 4 meses de cada UC para o ano de 2020: setembro a dezembro (variável X), e o objetivo da predição classificatória seria o Perfil de Consumo observado nesses meses. Para isso, considerou-se que:

1. O Perfil de Consumo de uma UC nesses 4 meses deve ser o mesmo, ou seja, não ocorrerá mudança de classe do mês 09 para o mês 10 se não houver mudanças significativas no consumo de energia elétrica;
2. Considerando o item acima, o valor de Y, do modelo RF, é o Perfil de Consumo do mês 11, onde os quinze últimos dias do mês de outubro e os primeiros 15 dias do mês de novembro estão faturados, época em que esperamos que se dê início à irrigação.
3. A própria base de dados não pode ser considerada como ideal, ou seja, não pode ser considerado que não está ocorrendo PNT em UCs da base de dados usada de entrada no modelo RF;
4. Achar melhor acurácia no modelo RF, contudo, levar em conta que desvios de consumo estão ocorrendo, o que pode diminuir a eficiência do modelo RF.

A Figura 36 apresenta o *DataFrame* utilizado como entrada no modelo RF, percebe-se o exemplo de UC que mantém certa média de consumo nos meses apresentados (*index 0*), e um exemplo em que a UC teve alto consumo apenas nos meses de setembro e outubro, e o faturamento para os outros meses foi nulo (*index 7*). O desejado é que o modelo RF detecte essas variâncias, quando forem inseridas novas entradas, por exemplo, os consumos para o ano de 2021 (aplicados no estudo de casos).

Figura 36 – Dados de Entrada do Modelo RF.

Index	202009	202010	202011	202012
0	741.654	873.504	741.402	664.089
1	5295	16319	6621	1169
2	16537	30345.3	24906.8	5081.08
3	9437	14761	315	458
4	0	0	0	0
5	0	0	0	0
6	28820	46771	17493	1241
7	30779	48585	1	0
8	0	0	0	0
9	7393	7683	694	194
10	1492.88	1456.21	2045.33	459.984
11	5123.16	6756.27	925.596	1091.75

Fonte: Elaboração Própria.

3.4.5 Métricas de Avaliação do Modelo RF Classificatório e Resultados

Quando um modelo classificatório é elaborado, ele precisa ser testado e avaliado para que se tenha uma dimensão do quão assertivo é o modelo para predição. Para tanto, métricas de avaliação são utilizadas para avaliação dos modelos de classificação de dados como acurácia, precisão e F-score (MARIANO, 2021). Uma das maneiras de representar os resultados de um método é através da matriz de confusão, mostrada na Tabela 12, basicamente, a diagonal principal irá nos mostrar quantas amostras de um conjunto de dados foram corretamente classificadas.

Tabela 12 - Matriz de Confusão.

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

Fonte: Elaboração Própria.

Para calcular a taxa de acertos (acurácia) e de erros, as classes são associadas conforme as equações (2) e (3), respectivamente.

$$Acurácia = \frac{VP + VN}{VP + FN + FP + VN} \quad (2)$$

$$Erro = \frac{FP + FN}{VP + FN + FP + VN} \quad (3)$$

Outra métrica importante é a sensibilidade, que avalia a capacidade do método em detectar resultados classificados como positivos (MARIANO, 2021), dado por (4).

$$Sensibilidade = \frac{VP}{VP + FN} \quad (4)$$

A Precisão, que avalia a quantidade de verdadeiros positivos sobre a soma de todos os valores positivos (MARIANO, 2021), é dada por (5).

$$Precisão = \frac{VP}{VP + FP} \quad (5)$$

Por fim, a última métrica utilizada como avaliativa do modelo RF, é o *F-score* (F1), que é a média harmônica calculada com base na precisão e sensibilidade (MARIANO, 2021), dado por (6).

$$F1 = 2 * \frac{Precisão * Sensibilidade}{Precisão + Sensibilidade} \quad (6)$$

Dessa forma, o modelo RF foi avaliado seguindo as métricas apresentadas. O melhor resultado foi de 69% de acurácia para os dados de teste e de 100% para os dados de treino. A Figura 37 apresenta o relatório final de simulação da implementação do modelo RF, onde o conjunto de dados (43 x 4) foi particionado em 70-30, para treino e teste, respectivamente. Foram utilizadas 13 amostras de UCs para o teste do modelo, em que: 6 eram pertencentes à Classe 0; as Classe 2 e Classe 3 tiveram apenas 2 amostras cada; e a Classe 1 teve 3 amostras.

Figura 37 - Relatório Final de Simulação do Modelo RF.

```

Acuracia do Classificador RF nos dados de Treino: 1.00
Acuracia do Classificador RF nos dados de Teste: 0.69

Matriz de Confusão:
[[5 0 1 0]
 [0 1 2 0]
 [0 0 2 0]
 [1 0 0 1]]

          precision    recall  f1-score   support

     0       0.83        0.83        0.83         6
     1       1.00        0.33        0.50         3
     2       0.40        1.00        0.57         2
     3       1.00        0.50        0.67         2

 accuracy          0.69         13
 macro avg          0.81         13
 weighted avg       0.83         13

Potuação de Acuracia: 0.6923076923076923

```

Fonte: Elaboração Própria.

Analisando a diagonal principal, podemos verificar que: das 6 amostras pertencentes a Classe 1, foram corretamente classificadas 5 amostras e 1 foi erroneamente classificada como pertencente à Classe 2: a Classe 1, teve 3 amostras classificadas erroneamente; a Classe 2 teve todas as amostras classificadas corretamente; a Classe teve apenas 1 de suas amostras classificadas corretamente.

Como estamos lidando com classes desbalanceadas, foi utilizado o balanceamento de classes na implementação do RF. Sendo assim, as classes são automaticamente ponderadas inversamente proporcionais à frequência com que aparecem nos dados (7), classes maiores apresentam pesos menor, e classes menores apresentam pesos maiores.

$$w_j = \frac{n}{kn_j} \quad (7)$$

Onde:

- w_j é o peso atribuído para a classe j ;
- n é o número de amostras total (dados de teste);
- n_j é o número de amostras para a classe j ;
- k é o número total de classes.

Os pesos atribuídos para cada classes são apresentados na Figura 38.

Figura 38 – Pesos Atribuídos para Cada Classe.

```
In [162]: 13/(4*np.bincount(y_test))
Out[162]: array([0.54166667, 1.08333333, 1.625      , 1.625      ])
```

Fonte: Elaboração Própria.

3.5 METODO DE ESTIMATIVA DE CONSUMO DAS UC IRRIGANTES

Além do modelo RF para detecção de possíveis ocorrências de PNT por divergências de Perfil de Consumo, durante a época de irrigação, também é feita a estimativa de consumo das UC, utilizando os dados de entrada descritos anteriormente: solo; CAD; fenologia e sistema irrigante; meteorológicos; declividade; e, área. De maneira que, presumindo-se que a unidade consumidora é eficiente em seu manejo de sistema irrigante, temos uma margem de consumo mensal que seria o esperado para uma dada área e condições climáticas.

3.5.1 Equações Utilizadas para Estimativa de Consumo

Nesta seção serão apresentadas as equações que foram utilizadas na estimativa de consumo.

- **Lâmina de Irrigação (LI)**

O valor de LI dado em mm é a quantidade de água que deve ser reposta ao solo para uma dada cultura, a equação que determina seu valor é apresentada em (8).

$$LI = ETc - P_p \quad (8)$$

Onde:

- ETc é a evapotranspiração da cultura, dada por $ET_o \times K_c$, apresentado em (1);
- P_p é a precipitação pluviométrica (chuva) dada em mm;
- Para um estudo mensal, realizado neste trabalho, os valores são mensais, é utilizado, então, o acumulado no mês de chuvas e ET_o .

O valor de P_p é proveniente dos dados meteorológicos de entrada.

- **Lâmina de Irrigação Bruta (LI_B)**

É o valor calculado acima dividido pela eficiência do sistema de irrigação utilizado (9). Como resultado do estudo e mapeamento tecnológico da cultura do café, determinou-se que o sistema utilizado é o de gotejamento, cuja eficiência pode ser de 75% a 90%, utilizou-se o valor de 75%.

$$LI_B = \frac{LI}{\eta_{irrig}} \quad (9)$$

Onde:

- η_{irrig} é a eficiência do sistema de irrigação.

- **Tempo de Irrigação (t)**

O tempo de irrigação é determinado pela capacidade de infiltração do solo em relação à quantidade de água que deve ser repostada solo (10). CAD é a sigla de Capacidade de Água Disponível no solo. Na prática a CAD do solo é o máximo de água que o solo pode reter levando em conta suas características físicas. Isso significa que cada tipo e textura de solo possuem CADs diferentes, por esse motivo foram definidos, no algoritmo implementado, os CADs do solo de cada UC utilizando o *Shapefile* descrito da seção de Dados de Entrada.

$$t = \frac{LI_B}{CAD_{solo}} \quad (10)$$

Onde:

- CAD_{solo} é definido no algoritmo por meio dos dados de entrada do *Shapefile* e coordenadas de latitude e longitude das UC.

- **Vazão (Q)**

Conceitua-se Vazão como a quantidade de fluido que escoar através de um conduto num determinado período. Para o cálculo da vazão são necessários dados referentes ao desnível do solo, pressão de serviço do sistema irrigante e a área que deve ser irrigada. Novamente, o valor de desnível do solo para a localização das UC foi adquirido via *Shapefile* (descrita nos Dados de Entrada). Para cada coordenada de latitude e longitude

foi associada um valor de desnível em percentagem. O valor de pressão de serviço foi resultado da pesquisa bibliográfica, definido como 10 m.c.a (metros por coluna de água). O valor de área para cada UC foi definido ao final da etapa de pré-processamento dos dados de área, constituindo um *Dataframe* com a informação.

Antes de calcular Q necessita-se calcular a altura geométrica, perdas de carga e altura manométrica, utilizando os dados de desnível e pressão de serviço. A equação da altura geométrica (H_g) é apresentada em (11).

$$H_g = \frac{\Delta L * L}{100} \quad (11)$$

Onde:

- ΔL é o percentual do desnível do solo;
- L é a distância até a fonte de água;

A distância da fonte de água foi considerada como sendo o valor do reservatório da fazenda pertencente a UC até o centro da área de irrigação, esse valor foi definido manualmente como sendo 380 m, que é um valor médio observado pelo estudo dos pontos do conjunto de dados final da etapa de pré-processamento.

Definindo-se o valor de H_g , podemos calcular as perdas de carga (h_s), dada pela equação (12) (10% do valor de H_g).

$$h_s = 0,1 * H_g \quad (12)$$

O próximo passo é calcular a altura manométrica total, definida por (13). A altura manométrica (AMT) representa a energia que a bomba deverá transmitir a água para transportar uma determina vazão entre as tubulações de sucção e recalque.

$$AMT = H_g + h_s + P_s \quad (13)$$

Onde:

- P_s é o valor da Pressão de serviço (igual à 10 m.c.a).

Com os valores acima definidos, podemos calcular o valor de Q utilizando (14).

$$Q = \frac{LI_B * Area * AMT}{t} \quad (14)$$

- **Potência do Sistema de Irrigação**

A estimativa do consumo de energia elétrica em UCs irrigantes, inicia-se a partir do cálculo da estimativa da Potência Instalada do Sistema de Irrigação (P). Dada pela equação (15).

$$P = \frac{Q * AMT * 0,0098}{\eta} \quad (15)$$

Onde:

- η é a eficiência do sistema motor-bomba, que depende da qualidade dos equipamentos utilizados, período de uso etc.

Foi adotado um valor baixo para η levando-se em conta que a maioria dos agricultores não investem no sistema irrigante, curva custo x produtividade. Apesar de todos os benefícios e do alto investimento realizado pelos agricultores, nos sistemas irrigantes, a maioria deles não dá a devida importância ao manejo da irrigação. Segundo Silveira e Stone (1994), umas das causas a serem citadas para explicar tal atitude, são: (i) Custo da água, onde na maioria das situações, a água nada custa ao produtor irrigante; (ii) Custo da energia elétrica, com o desconto na tarifa para os irrigantes de período reservado, o custo da energia elétrica não é tão alto, o que pode girar em torno de 10% do custo total da produção. Dessa forma, o valor definido para as simulações foi de 45%.

- **Energia consumida pelo Sistema de Irrigação**

A estimativa da energia elétrica consumida pelo sistema de irrigação (E) é dada por (16).

$$E = P * t \quad (16)$$

3.5.2 Exemplo de Cálculo do Balanço Hídrico

Nesta seção é apresentado um exemplo de cálculo de balanço hídrico. O objetivo é chegar aos valores de Déficit hídrico, que seria o valor em mm da quantidade de água que deve ser reposta ao solo.

Pode-se calcular a estimativa de consumo, utilizando a LI_B, apresentada na seção anterior, contudo, foi utilizado os valores de déficit. O balanço hídrico também foi utilizado para indicar os meses em que foram necessários utilizar irrigação, e os meses em que não seria esperado consumo de energia pelo manejo do sistema, pois teve-se

excedente hídrico (não sendo necessário repor água ao solo). Nesta etapa, utiliza-se os valores de ETc e P_p , os valores de chuva e ETo utilizados foram do ano de 2021.

Dessa maneira, chegamos aos valores de Déficit e Excedente, seguindo os seguintes passos:

1° Passo: Calcula-se o Valor de ETc ;

2° Passo: Calcula-se o valor de $P_p - ETc$ (coluna P – ETc na Figura 39)

3° Passo: Calcula-se o Negativos Acumulados (NAC), onde são somados todos os valores de negativos de $P_p - ETc$;

4° Passo: Calcula-se o valor de ARM, definido como $CAD * e^{\frac{NAC}{CAD}}$, para todos os meses, onde o NAC for zero, o resultado será o próprio valor de CAD;

5° Passo: Calcula-se a Alteração, definida como o ARM do mês atual menos o ARM do mês anterior, nesse caso o valor para janeiro (2021) pode ser definido utilizando o mês de dezembro (2021), supondo-se que a região de estudo tem o mesmo comportamento de Clima esperado para o mês de dezembro do ano anterior (2020);

6° Passo: Calcula-se o ETr, definido como igual ao ETc , se Precipitação (P_p) $\geq ETc$, ou, igual a Precipitação (P_p) + Alteração, se Precipitação (P_p) $< ETc$;

7° Passo: Calcula-se o Déficit hídrico, definido por $ETc - ETr$;

8° Passo: Calcula-se o Excedente hídrico, onde é igual a zero se $ARM > CAD$, e igual ao valor de $P_p - ETc$ menos a Alteração, para os casos em que $ARM = CAD$.

Seguindo os passos, com dados meteorológicos para o município de Jeriquara no ano de 2021, tem-se o resultado apresentado na Figura 39. Podemos ver que os meses em que seria necessário utilizar do sistema de irrigação foram de abril a setembro.

Figura 39 – Exemplo de Balanço Hídrico.

MÊS	ETo	ETc	PRECIPITACAO	P-ETc	NAC	ARM	ALTERACAO	ETr	Deficit	Excedente
1	131.99	171.587	209.53	37.943	0	43.8553	4.89822	171.587	0	33.0448
2	114	148.2	197.86	49.66	0	43.8553	0	148.2	0	49.66
3	122.17	158.821	167.14	8.319	0	43.8553	0	158.821	0	8.319
4	93.96	122.148	14.73	-107.418	-107.418	3.78681	-40.0685	54.7985	67.3495	0
5	80.48	104.624	0.5	-104.124	-211.542	0.352488	-3.43432	3.93432	100.69	0
6	68.13	88.569	1.01	-87.559	-299.101	0.0478692	-0.304618	1.31462	87.2544	0
7	64.81	84.253	0	-84.253	-383.354	0.00700983	-0.0408594	0.0408594	84.2121	0
8	97.61	126.893	0.25	-126.643	-509.997	0.000390459	-0.00661938	0.256619	126.636	0
9	135.11	175.643	32.52	-143.123	-653.12	1.49363e-05	-0.000375523	32.5204	143.123	0
10	131.16	170.508	170.43	-0.078	-653.198	1.49098e-05	-2.65418e-08	170.43	0.078	0
11	103.41	134.433	310.64	176.207	0	43.8553	43.8553	134.433	0	132.352
12	106.38	138.294	133.1	-5.194	-5.194	38.9571	-4.89822	137.998	0.295784	0

Fonte: Elaboração Própria.

4. ESTUDO DE CASO

4.1 CARACTERIZAÇÃO DO ESTUDO DE CASO

Com os atributos e equações, descritos nas seções anteriores, definidos, foi realizado três estudos de caso utilizando o modelo RF classificatório, de modo que, o objetivo é entrar com dados de consumo mais atuais (ano de 2021) e comparar o valor da Classe prevista pelo modelo com o valor da Classe originária da UC, resultante da etapa de clusterização. Se os valores de consumo de energia da UC irrigante não forem muito diferentes do que foi consumido no ano anterior, ambas classes (Prevista e Original) deverão ser a mesma. Caso contrário, se o perfil de consumo de energia da UC mudar de maneira significativa, ambas as classes deverão ser diferentes.

Segundo a Figura 32, a ordem decrescente da relação de Média x Desvio é dada por: Classe 1 – Classe 2 – Classe 3 – Classe 0. Ou seja, valores maiores de médias e desvios serão da Classe 1, e os menores pertencerão a Classe 0, como apresentado na Tabela 13.

Tabela 13 – Médias máximas e Mínimas de cada Classe.

Classe	Média Max [kWh]	Média Min [kWh]
1	27252.25	5672.25
2	11827.25	3072.75
3	5629.50	684.50
0	1275.50	9.75

Fonte: Elaboração Própria.

Os dados de consumo utilizados como entrada foram referentes aos meses 09 ao 12 (setembro a dezembro). As UCs escolhidas para o estudo de caso foram selecionadas aleatoriamente.

Posteriormente à aplicação do modelo RF, foi escolhida uma UC dentre as três apresentadas no modelo RF, para ser feito o estudo de estimativa de consumo, onde é utilizado indicadores auxiliares para indicar ocorrência de PNT com base no comparativo do consumo esperado e consumo real.

4.2 ESTUDO DE CASO - APLICAÇÃO DO MODELO RF

Os resultados da aplicação do modelo RF são apresentados nesta seção.

- **Estudo de Caso 1**

A UC do caso 1 apresenta os valores de consumo descritos na Tabela 14. E os valores de Classe de Perfil de Consumo apresentados na Tabela 15. Logo, se os valores de entrada de consumo se mantiverem próximos dos valores registrados no ano de 2020, espera-se que a Classe não se altere, se os consumos forem divergentes significativamente, espera-se o modelo RF acuse uma Classe diferente. Com os valores de consumo registrados em 2021, apresentados na Tabela 14.

Tabela 14 – Caso 1: Dados de Consumo.

Caso 1 - Dados de Consumo			
Modelo RF		Dado de Entrada	
DT_REFERENCIA	CONSUMO	DT_REFERENCIA	CONSUMO
202009	16537	202109	19386
202010	30345,3	202110	23910
202011	24906,8	202111	19843
202012	5081,08	202112	4376

Fonte: Elaboração Própria.

Tabela 15 - Caso 1: Dados de Perfil de Consumo.

Caso 1 - Dados de Classe de Perfil de Consumo	
Resultado Clusterização	
DT_REFERENCIA	Classe
202009	2
202010	1
202011	1
202012	1

Classe Esperada: 1

Fonte: Elaboração Própria.

O relatório de final de simulação para a UC 1 com os valores da tabela é apresentado na Figura 40. Com os consumos inseridos, não houve mudança de Classes entre o previsto e o original da UC.

Figura 40 – Caso 1: Resultado da Simulação do Modelo RF.

```

Estudo de Caso 1
#####
Dados de Consumo de entrada:
  CONSUMO_9  CONSUMO_10  CONSUMO_11  CONSUMO_12
    19386.0    23910.0    19843.0    4376.0

Classe Original (consumos de 2020): 1
Classe Prevista (consumos de 2021): [1]

Situação: Não Houve Mudanças de Perfil

```

Fonte: Elaboração Própria.

Para validação do modelo, foram utilizados como entrada os consumos do ano de 2020, da Tabela 14, verificando se o modelo está classificando corretamente, Figura 41.

Figura 41 - Caso 1 validação: Resultado da Simulação do Modelo RF.

```

Estudo de Caso 1
#####
Dados de Consumo de entrada:
DT_REFERENCIA  202009  202010  202011  202012
                16536.989  30345.273  24906.758  5081.077

Classe Original (consumos de 2020): 1
Classe Prevista (consumos de 2021): [1]

Situação: Não Houve Mudanças de Perfil

```

Fonte: Elaboração Própria.

Para simulação de possível ocorrência de PNT foram inseridos alteração nos valores de consumo de 2021. Desse modo, os valores de Consumo de 2021 foram multiplicados por 65%. A variação em percentual do novo valor em relação ao consumo do ano de 2020 também foi calculado, Tabela 16.

Tabela 16 – Caso 1: Novos Dados de Entrada e Variação.

Caso 1 - Dados de Consumo					
Modelo RF		Dado de Entrada		Novos Dados de Entrada	
DT_REFERENCIA	CONSUMO	DT_REFERENCIA	CONSUMO	CONSUMO 0,65	% variação
202009	16537	202109	19386	12600,9	-23,80177783
202010	30345,3	202110	23910	15541,5	-48,78449051
202011	24906,8	202111	19843	12897,95	-48,21514606
202012	5081,08	202112	4376	2844,4	-44,01977532

Fonte: Elaboração Própria.

O resultado para esta simulação é ilustrado no relatório final de simulação da Figura 42. Percebe-se que o Modelo RF detectou a mudança como possível ocorrência de PNT, como esperado.

Figura 42 - Caso 1 Consumo Alterado: Resultado da Simulação do Modelo RF.

```

Estudo de Caso 1
#####
Dados de Consumo de entrada:
  CONSUMO_9  CONSUMO_10  CONSUMO_11  CONSUMO_12
    12600.9    15541.5    12897.95    2844.4

Classe Original (consumos de 2020): 1
Classe Prevista (consumos de 2021): [2]

Situação: Possível Ocorrência de Perdas Não Técnicas

```

Fonte: Elaboração Própria.

- **Estudo de Caso 2**

A UC do caso 2 apresenta os valores de consumo descritos na Tabela 17. E os valores de Classe de Perfil de Consumo apresentados na Tabela 18. Nesse caso espera-se uma mudança de Classe de Perfil de Consumo sem ser necessário alterar os valores (simulação de PNT), pois a Classe de origem é intermediária e os consumo de 2020 consideravelmente mais altos do que os do ano de 2021.

Tabela 17 - Caso 2: Dados de Consumo.

Caso 2 - Dados de Consumo			
Modelo RF		Dado de Entrada	
DT_REFERENCIA	CONSUMO	DT_REFERENCIA	CONSUMO
202009	3143,36	202109	1511
202010	5956,4	202110	3527
202011	1333,67	202111	738
202012	100	202112	290

Fonte: Elaboração Própria.

Tabela 18 - Caso 2: Dados de Perfil de Consumo.

Caso 2 - Dados de Classe de Perfil de Consumo	
Resultado Clusterização	
DT_REFERENCIA	Classe
202009	3
202010	3
202011	3
202012	3

Classe Esperada: 3

Fonte: Elaboração Própria.

O relatório final de simulação do Caso 2 é apresentado na Figura 43. De fato, houve a mudança de classe indicando PNT.

Figura 43 - Caso 2: Resultado da Simulação do Modelo RF.

```

Estudo de Caso 2
#####
Dados de Consumo de entrada:
  CONSUMO_9  CONSUMO_10  CONSUMO_11  CONSUMO_12
    1511.0    3527.0    738.0     290.0

Classe Original (consumos de 2020): 3
Classe Prevista (consumos de 2021): [0]

Situação: Possível Ocorrência de Perdas Não Técnicas

```

Fonte: Elaboração Própria.

A sensibilidade de alteração de consumo foi simulada para a UC deste caso, encontrando-se o limite de 0,68 de variação de consumos do observado em 2020, para que

não ocorra sinalização de PNT, se a UC tiver o consumo reduzido em 33% então já é o suficiente para que o modelo indique situação de possível PNT, o resultado para ambas as ocorrências é apresentado na Figura 44.

Figura 44 - Caso 2 Consumos Alterados: Resultados da Simulação do Modelo RF.

```

Estudo de Caso 2
#####
Dados de Consumo de entrada:
DT_REFERENCIA      202009      202010      202011      202012
0                   2137.48752  4050.34928  906.89696   68.0

Classe Original (consumos de 2020): 3
Classe Prevista (consumos de 2021): [3]

Situação: Não Houve Mudanças de Perfil

```

a) Não ocorrência de PNT.

```

Estudo de Caso 2
#####
Dados de Consumo de entrada:
DT_REFERENCIA      202009      202010      202011      202012
0                   2106.05388  3990.78532  893.56024   67.0

Classe Original (consumos de 2020): 3
Classe Prevista (consumos de 2021): [0]

Situação: Possível Ocorrência de Perdas Não Técnicas

```

b) Ocorrência de PNT.

Fonte: Elaboração Própria.

- **Estudo de Caso 3**

A UC do caso 3 apresenta os valores de consumo descritos na Tabela 19. Os valores de Consumo do ano de 2021 desta UC são maiores do que do ano anterior, 2020, a coluna de percentual de variação apresenta o quão maior foram. Dessa maneira, espera-se que a Classe de Perfil de Consumo altere-se para uma Classe superior. Este caso é interessante, pois o baixo consumo de 2020 pode ser uma possível ocorrência de PNT, como indicado no relatório final de simulação. Os valores de Classe de Perfil de Consumo esperado são apresentados na Tabela 20.

Tabela 19 - Caso 3: Dados de Consumo e Variação Percentual.

Caso 3 - Dados de Consumo				
Modelo RF		Dado de Entrada		
DT_REFERENCIA	CONSUMO	DT_REFERENCIA	CONSUMO	% variação
202009	5295	202109	27456	418,5269122
202010	16319	202110	14053	16,12467089
202011	6621	202111	13427	102,7941399
202012	1169	202112	1647	40,88964927

Fonte: Elaboração Própria.

Tabela 20 - Caso 3: Dados de Perfil de Consumo.

Caso 3 - Dados de Classe de Perfil de Consumo	
Resultado Clusterização	
DT_REFERENCIA	Classe
202009	3
202010	2
202011	2
202012	2

Classe Esperada: 2

Fonte: Elaboração Própria.

O relatório final de simulação do Caso 3 é apresentado na Figura 45. Ocorre mudança de Classe, que, por ser uma classe superior, pode ser um indicador de PNT no ano anterior.

Figura 45 - Caso 3: Resultado da Simulação do Modelo RF.

```

Estudo de Caso 3
#####
Dados de Consumo de entrada:
  CONSUMO_9  CONSUMO_10  CONSUMO_11  CONSUMO_12
0   27456.0    14053.0    13427.0    1647.0

Classe Original (consumos de 2020): 2
Classe Prevista (consumos de 2021): [1]

Situação: Possível Ocorrência de Perdas Não Técnicas
Consumo de 2021 maior do que 2020

```

Fonte: Elaboração Própria.

4.3 ESTUDO DE CASO – ESTIMATIVA DE CONSUMO

Para a estimativa de energia consumida pelo sistema irrigante foram utilizadas duas UCs de municípios diferentes, logo seus valores de ETo e P_P serão diferentes. Serão apresentados os valores dos atributos descritos na seção de Método de Estimativa de Consumo de UC Irrigante.

Nesta aplicação, se a energia consumida real for menor do que à estimada, ou dentro da faixa de valor definida, será indicada situação de possível ocorrência de PNT. Não é objetivo deste trabalho avaliar os consumos registrados maiores do que o esperado, contudo, também é indicação que poderia resultar em uma vistoria por parte da distribuidora, pois a UC pode estar utilizando a ligação que seria exclusiva da irrigação para outros fins, sendo faturado um valor a ser pago menor devido ao desconto.

- **Estudo de Caso 1**

A UC deste caso é a mesma UC do Caso 3 do modelo RF. A UC está localizada no município de Jeriquara e seus atributos são descritos na Figura 46.

Figura 46 – Caso 1 da Estimativa: Atributos e Características.

	28
COD_INSTALACAO	
CAD[%]	0.0877106
DN	3
Lat	-20.3031
Long	-47.5546
DESC_MUNICIPIO	Jeriquara
AREA	331
Profundidade	500
Tipo_Irrigacao	Gotejamento
EFICIENCIA	75
Pressão_De_Servico	10
L	380
Hg	11.4
AMT	22.54
LAMINA	43.8553
CULTURA	Cafe
Kc_Jan	1.18
Kc_Fev	1.2
Kc_Mar	1.2
Kc_Abr	1.12
Kc_Mai	1.07
Kc_Jun	0.85
Kc_Jul	0.87
Kc_Ago	0.88
Kc_Set	0.98
Kc_Out	1.21
Kc_Nov	1.23
Kc_Dez	1.23

Fonte: Elaboração Própria.

Os valores de: CAD; DN (desnível); Area; Profundidade; Tipo de Irrigação; Eficiência; e, Pressão de Serviço; foram definidos pela base de dados de entrada. Os valores de: H_g, AMT e Lâmina, foram calculados no algoritmo. O Valor de Lâmina, neste caso é dado pelo CAD percentual, cujo valor é apresentado na figura, multiplicado pela profundidade do sistema radicular do café, esse valor refere-se a lâmina de água armazenado no solo, e será utilizado como CAD_{solo}. O valor de L é definido como 380 metros. A parte do código, em que os valores são calculados e inseridos no *DataFrame*, é apresentado na Figura 47.

Figura 47 – Caso 1 da Estimativa: Atributos para Cálculo de Estimativa de Consumo.

```
Dados_SP_caso['L'] = 380
Dados_SP_caso['Hg'] = (Dados_SP_caso.L * Dados_SP_caso.DN)/100
Dados_SP_caso['AMT'] = Dados_SP_caso.Hg + (Dados_SP_caso.Hg *0.1) + Dados_SP_caso.Pressão_De_Servico
Dados_SP_caso['LAMINA'] = Dados_SP_caso['Profundidade'] * Dados_SP_caso['CAD[%]']
```

Fonte: Elaboração Própria.

O resultado do balanço hídrico (BH) é apresentado na Figura 48, onde ocorre déficit hídrico nos meses 05 a 09 (maio a setembro). E na Figura 49, na coluna Energia tem-se o resultado do quanto de energia elétrica (kWh) é esperada para irrigação da área associada à UC de 331 hectares, com as dadas condições de E_{To} e Precipitação pluviométrica, e demais características determinadas.

Figura 48 - Caso 1 da Estimativa: Resultado do BH.

MES	E _{To}	E _{Tc}	PRECIPITACAO	P-E _{Tc}	NAC	ARM	ALTERACAO	E _{Tr}	Deficit	Excedente
1	131.99	155.748	209.53	53.7818	0	43.8553	0	155.748	0	53.7818
2	114	134.52	197.86	63.34	0	43.8553	0	134.52	0	63.34
3	122.17	144.161	167.14	22.9794	0	43.8553	0	144.161	0	22.9794
4	93.96	110.873	14.73	-96.1428	-96.1428	4.897	-38.9583	53.6883	57.1845	0
5	80.48	94.9664	0.5	-94.4664	-190.609	0.568119	-4.32888	4.82888	90.1375	0
6	68.13	80.3934	1.01	-79.3834	-269.993	0.0929638	-0.475155	1.48516	78.9082	0
7	64.81	76.4758	0	-76.4758	-346.468	0.0162548	-0.076709	0.076709	76.3991	0
8	97.61	115.18	0.25	-114.93	-461.398	0.00118262	-0.0150722	0.265072	114.915	0
9	135.11	159.43	32.52	-126.91	-588.308	6.54741e-05	-0.00111714	32.5211	126.909	0
10	131.16	154.769	170.43	15.6612	0	43.8553	43.8552	154.769	0	-28.194
11	103.41	122.024	310.64	188.616	0	43.8553	0	122.024	0	188.616
12	106.38	125.528	133.1	7.5716	0	43.8553	0	125.528	0	7.5716

Fonte: Elaboração Própria.

Figura 49 - Caso 1 da Estimativa: Resultado de Energia Consumida.

Deficit	Excedente	Lamina_bruta	Tempo_Irrigacao	Vazao	Potencia	Energia
0	53.7818	0	nan	0	nan	nan
0	63.34	0	nan	0	nan	nan
0	22.9794	0	nan	0	nan	nan
57.1845	0	128.19	2.92303	14516.1	8016.23	23431.7
90.1375	0	125.955	2.87206	14516.1	8016.23	23023.1
78.9082	0	105.845	2.41349	14516.1	8016.23	19347.1
76.3991	0	101.968	2.32509	14516.1	8016.23	18638.5
114.915	0	153.24	3.49421	14516.1	8016.23	28010.4
126.909	0	169.213	3.85844	14516.1	8016.23	30930.1
0	-28.194	0	nan	0	nan	nan
0	188.616	0	nan	0	nan	nan
0	7.5716	0	nan	0	nan	nan

Fonte: Elaboração Própria.

Os maiores valores de consumo esperados são para os meses de agosto e setembro. Dessa forma, o algoritmo compara com o consumo real, do ano de 2021. Como comentado, os registros de consumo na base de dados da distribuidora seguem um determinado período de data de leitura, onde o comum é que no registro de um mês esteja faturado parte do consumido do mês vigente e parte de consumo do mês anterior. Sendo assim, para comparar o consumo estimado com o real foi feita a seguinte consideração:

$$Consumo_i = \frac{Consumo_i}{30} * 15 + \frac{Consumo_{i+1}}{30} * 15 \quad (17)$$

O resultado de (17) para a UC do caso 1 é apresentado na Tabela 21. Os consumos a serem comparados serão os da última coluna. O resultado do relatório final de simulação é apresentado na Figura 50, onde ocorre acusação de PNT.

Tabela 21 - Caso 1 Método de Estimativa: Consumos Mensais.

Caso 1 Método de Estimativa - Dados de Consumo			
Consumo Real		Consumo Para Comparação	
DT_REFERENCIA	CONSUMO	DT_REFERENCIA	CONSUMO
202101	NaN	202101	0
202102	1143	202102	2094
202103	3045	202103	4909
202104	6774	202104	3889
202105	1005	202105	2463
202106	3921	202106	8239
202107	12557	202107	16514
202108	20471	202108	23963,5
202109	27456	202109	20754,5
202110	14053	202110	13740
202111	13427	202111	7537
202112	1647	202112	823,5

Fonte: Elaboração Própria.

Figura 50 - Caso 1 Estimativa de Consumo: Relatório Final de Simulação.

```

Estudo de Caso 1: Estimativa de Consumo
#####
Mês: 4
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas:
Mês: 5
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas:
Mês: 6
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas:
Mês: 7
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas:
Mês: 8
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas:
Mês: 9
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas:

```

Fonte: Elaboração Própria.

- **Estudo de Caso 2**

Para o Caso 2, escolheu-se uma UC cuja Classe de Perfil de Consumo é fosse da Classe 0, de maneira que as divergências de consumo real com esperado, fosse detectado pelo método de estimativa, pois por pertencer a classe de nível mais baixo, consumos menores do que o esperado não seriam acusados por mudança de Classe de Perfil de Consumo. Dessa maneira, fazendo o estudo de estimativa de consumo para as UCs da Classe 0, podemos identificar UCs com baixo consumo que podem estar irregulares.

A UC deste caso é localizada no município de Pedregulho, seus atributos são descritos na Figura 51.

Figura 51 - Caso 2 da Estimativa: Atributos e Características.

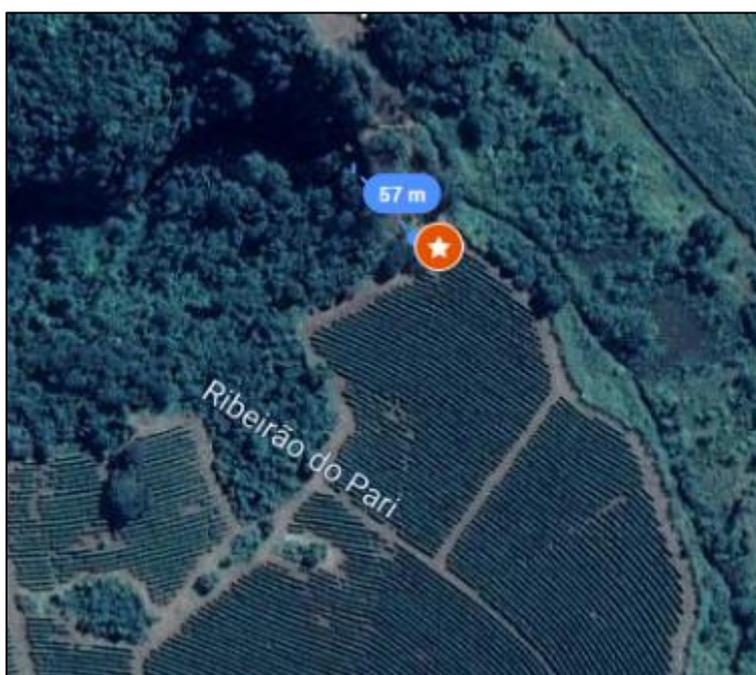
	61
COD_INSTALACAO	
CAD[%]	0.0877106
DN	3
Lat	-20.2099
Long	-47.5171
DESC_MUNICIPIO	Pedregulho
AREA	54.43
Profundidade	500
Tipo_Irigacao	Gotejamento
EFICIENCIA	75
Pressão_De_Servico	10
L	40
Hg	1.2
AMT	11.32
LAMINA	43.8553
CULTURA	Cafe
Kc_Jan	1.18
Kc_Fev	1.2
Kc_Mar	1.2
Kc_Abr	1.12
Kc_Mai	1.07
Kc_Jun	0.85
Kc_Jul	0.87
Kc_Ago	0.88
Kc_Set	0.98
Kc_Out	1.21
Kc_Nov	1.23
Kc_Dez	1.23

Fonte: Elaboração Própria.

Como comentado anteriormente, o atributo de distância do corpo hídrico (L). é dado pela distância entre a fonte de água e o reservatório, quanto mais distantes maior será a potência necessária para transportar água. Como não foi possível identificar fontes de

água menos proeminente, para algumas UCs o valor de L foi definido como 380 metros, casos em que a UC tem seu próprio reservatório e localizada distante de rios. Contudo, a UC do Caso 2 está ao lado de um rio identificável e sem reservatório na área de plantio, logo, o valor de L foi definido como sendo 40 metros, distância entre o rio e o ponto de medição da UC, Figura 52.

Figura 52 - Caso 2 da Estimativa: Distância da Fonte de Água.



Fonte: Google maps (2022).

O resultado do balanço hídrico (BH) para a UC do município de Pedregulho é apresentado na Figura 53. Percebe-se que, no município ocorre déficit hídrico nos meses 03 a 09 (março a setembro).

Figura 53 - Caso 2 da Estimativa: Resultado do BH.

MES	ETo	ETc	PRECIPITACAO	P-ETc	NAC	ARM	ALTERACAO	ETr	Deficit	Excedente
1	87.34	155.748	173.23	17.4818	0	43.8553	0	155.748	0	17.4818
2	108.27	134.52	161.04	26.52	0	43.8553	0	134.52	0	26.52
3	120.17	144.161	114.04	-30.1206	-30.1206	22.0669	-21.7884	135.828	8.33218	0
4	92.21	110.873	44.71	-66.1628	-96.2834	4.88133	-17.1856	61.8956	48.9772	0
5	82.94	94.9664	8.13	-86.8364	-183.12	0.673916	-4.20741	12.3374	82.629	0
6	68.14	80.3934	16.77	-63.6234	-246.743	0.157961	-0.515955	17.286	63.1074	0
7	66.96	76.4758	0	-76.4758	-323.219	0.0276197	-0.130341	0.130341	76.3455	0
8	95.18	115.18	0	-115.18	-438.399	0.001998...	-0.0256216	0.0256216	115.154	0
9	136.94	159.43	40.64	-118.79	-557.189	0.000133...	-0.00186492	40.6419	118.788	0
10	123.02	154.769	277.37	122.601	0	43.8553	43.8552	154.769	0	78.746
11	119.16	122.024	335.55	213.526	0	43.8553	0	122.024	0	213.526
12	116.87	125.528	143.77	18.2416	0	43.8553	0	125.528	0	18.2416

Fonte: Elaboração Própria.

Na Figura 54, a coluna Energia mostra o resultado da estimativa do consumo de energia elétrica esperada para irrigação da área associada à UC (em kWh), para as dadas condições de ETo e Precipitação pluviométrica. Os consumos a serem comparados são apresentados na Tabela 22.

Figura 54 - Caso 2 da Estimativa: Resultado de Energia Consumida.

Deficit	Excedente	Lamina_bruta	Tempo_Irrigacao	Vazao	Potencia	Energia
0	17.4818	0	nan	0	nan	0
0	26.52	0	nan	0	nan	0
8.33218	0	40.1608	0.915757	2387.04	662.023	606.252
48.9772	0	88.2171	2.01155	2387.04	662.023	1331.69
82.629	0	115.782	2.64009	2387.04	662.023	1747.8
63.1074	0	84.8312	1.93434	2387.04	662.023	1280.58
76.3455	0	101.968	2.32509	2387.04	662.023	1539.27
115.154	0	153.573	3.50181	2387.04	662.023	2318.28
118.788	0	158.386	3.61157	2387.04	662.023	2390.94
0	78.746	0	nan	0	nan	0
0	213.526	0	nan	0	nan	0
0	18.2416	0	nan	0	nan	0

Fonte: Elaboração Própria.

Tabela 22 - Caso 2 Método de Estimativa: Consumos Mensais.

Caso 2 Método de Estimativa - Dados de Consumo			
Consumo Real		Consumo Para Comparação	
DT_REFERENCIA	CONSUMO	DT_REFERENCIA	CONSUMO
202101	100	202101	0
202102	100	202102	100
202103	100	202103	100
202104	100	202104	100
202105	100	202105	488
202106	876	202106	591,5
202107	307	202107	203,5
202108	100	202108	100
202109	100	202109	100
202110	100	202110	687,5
202111	1275	202111	687,5
202112	100	202112	50

Fonte: Elaboração Própria.

O resultado do caso 2 é apresentado na Figura 55, onde ocorre acusação de PNT, para a UC. Percebe-se pela Tabela 22, que, nos primeiros cinco meses do ano de 2021, a UC apresentou parcela de consumo mensal inferior a 100 kWh.

Figura 55 - Caso 2 Estimativa de Consumo: Relatório Final de Simulação.

```

Estudo de Caso 2: Estimativa de Consumo
#####
Mês: 4
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas
Mês: 5
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas.
Mês: 6
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas.
Mês: 7
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas
Mês: 8
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas
Mês: 9
Indicador 1: Consumo observado menor do que o Estimado
Situação: Possível Ocorrência de Perdas Não Técnicas

```

Fonte: Elaboração Própria.

5. CONCLUSÕES

O presente trabalho apresentou uma metodologia para detecção de perdas não técnicas para consumidores rurais irrigantes de café, utilizando dados históricos de consumo de 72 meses para gerar classes de perfil de consumo por quadrimestres. Obtendo-se para cada janela de 4 meses a média de consumo esperada para as unidades consumidoras do trabalho, sendo possível identificar o comportamento esperado para o consumidor específico comparando com o seu perfil de consumo para o mesmo período de anos anteriores. Contudo, para UCs com baixo consumo de energia registrados, uma metodologia mais específica é necessária, pois a possível ocorrência de PNT desta unidade não seria detectável por meio da inteligência artificial aplicada por meio de modelo classificatório. Com isso, a metodologia de estimativa de consumo aplicada neste trabalho auxilia na identificação de divergências de consumo, utilizando de dados agrometeorológicos que possam fundamentar o consumo registrado.

Com o desenvolvimento do trabalho, observou-se as particularidades do agricultor produtor de café no estado de São Paulo, uma localidade propícia ao desenvolvimento da cultura do café, quanto ao manejo do sistema irrigante e consumo de energia elétrica, entre elas, a de que a água, como insumo da agricultura irrigada contribui com um elevado custo de produção, principalmente devido ao dispêndio com bombeamento. Dessa forma, do ponto de vista do agricultor, o manejo do sistema irrigante não pode ocasionar reduções na produtividade em decorrência da falta ou do excesso de água aplicada. Sendo que a aplicação de pouca água (irrigação com déficit) acaba sendo um desperdício, tendo em vista que a produção seria afetada e não poderia obter o benefício esperado ao fim da safra. Contudo, é uma prática ainda em uso, compondo um desafio para a metodologia apresentada. Também, pode-se concluir que o manejo da irrigação visa maximizar a produtividade da cultura e melhorar a qualidade do produto, ao mesmo tempo que busca minimizar o custo da água e energia. Apesar de oferecer benefícios e do alto investimento realizado pelos agricultores, a maioria não dá a devida importância ao manejo da irrigação.

Por fim, a metodologia desenvolvida e apresentada neste trabalho foi aplicada em estudos de casos, em que se obteve o resultado esperado de indicação de PNT, tanto na aplicabilidade do modelo RF preditivo, quanto na aplicabilidade do modelo de balanço hídrico nas UCs de consumos menores (classes inferiores). A ocorrência de indicador de

PNT nesses casos auxilia na rotina de inspeção da concessionária, para verificar se a UC está utilizando o ramal de ligação corretamente.

5.1 SUGESTÕES DE CONTINUIDADE

A sugestão de continuidade do trabalho tem como objetivo o melhoramento e aperfeiçoamento do desempenho da metodologia, abordando os seguintes pontos:

- Aprimoramentos na etapa de clusterização dos dados, buscando elevar a precisão do modelo e sensibilidade na detecção de mudanças de classe;
- Simulações com Classes balanceadas, de maneira a mitigar os problemas de simulação referentes a classes desbalanceadas, comparando com a metodologia de pesos, encontrando o melhor resultado;
- Aplicação do estudo com dados referentes a identificação de ocorrências de PNT;
- Identificar as distâncias ao corpo hídrico para fontes de água menores;
- Realizar Clusters de áreas de plantio, para agrupar UCs e áreas, de maneira a estimar o consumo esperado para cada cluster, eliminando assim o problema de exatidão da associação área de plantio x UC.
- Automatizar o processo de cálculo de balanço hídrico, gerando um modelo IA que irá usar como entrada os consumos reais e comparar com os estimados;
- Expandir a metodologia para outra região cafeeira, como por exemplo, Marília em São Paulo;

5.2 PUBLICAÇÕES RELACIONADAS AO TRABALHO

Os trabalhos publicados durante a elaboração desta dissertação, utilizados como base para a metodologia implementada foram:

SOUSA, 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.

EICHKOFF, H. S.; BERNARDON, D. P.; SOUZA, N. B.; MARCOLIN, P.; MADALAZ, 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.

REFERÊNCIAS BIBLIOGRÁFICAS

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.

ALBUQUERQUE, P. E. P. **Estratégias de Manejo de Irrigação: Exemplos de Cálculo**. Ministério da Agricultura, Pecuária e Abastecimento. ISSN 1518-4269. Sete Lagoas, MG, setembro de 2010.

ALBUQUERQUE, P. E. P.; GUIMARÃES, D. P. **Gestão da água na agricultura: coeficientes de cultivo (Kc) e de tanque classe A (Kp): um exemplo metodológico de estimativa do Kc da cultura do milho e do Kp para o Estado de Minas Gerais**. *Item*, Brasília, v. 63, p. 12-19, 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/metodologiadistribuicao/asset_publisher/e2INtBH4EC4e/content/perdas/654800>. Acesso em: mar. 2021.

ANEEL. **Tarifas Consumidores**. Brasília, 2016. Disponível em: <<https://www.aneel.gov.br/tarifasconsumidores/>>. Acesso em: set. 2021.

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.

ANEEL. **Resolução Normativa 1000/2021**. Brasília, Brasil: Agência Nacional de Energia Elétrica. Acesso em: abril 2020.

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.

ARCANJO, K. E.; RAMOS, F. S. **Impacto de variáveis socioeconômicas sobre as perdas comerciais de energia elétrica**. RBEE. 22 p. JEL: D00, C2. Disponível em <

https://www.researchgate.net/publication/336551649_Impacto_de_variaveis_socioeconomicas_sobre_as_perdas_comerciais_de_energia_eletrica> Acesso em: 2019

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.

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.

BOWLES, A. *Machine Learning in Python: Essential Techniques for Predictive Analysis*. ISBN: 978-1-118-96174-2. 360 p. Abril de 2015.

CAMARGO, A. P. **O clima e a cafeicultura no Brasil. Informe Agropecuário**. Belo Horizonte – MG, v. 11 n. 126, p.13-26, 1985.

CAMARGO, A. P.; SENTELHAS, P. C. **Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no estado de São Paulo, Brasil**. *Revista Brasileira de Agrometeorologia*, Santa Maria, v.5, n.1, p.89-97,1997.

CAMARGO, A. P. D.; M. B. P. D. CAMARGO. **Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil**. 2001.

CARVALHO, L. G; RIOS, G. F. A.; MIRANDA, W. L.; CASTRO NETO, P. **Evapotranspiração de referência: uma abordagem atual de diferentes métodos de estimativa**. *Pesquisa Agropecuária Tropical*, 2021, 41(3), 456-465. <https://dx.doi.org/10.5216/pat.v41i3.12760>.

CHAVES, A. C; TAVARES, A. et. al. **As Perdas Não Técnicas No Setor de Distribuição Brasileiro: Uma Abordagem Regulatória**. CPFL Energia. 257 p. 2018.

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.

CHAUHAN, A. A.; **Non-Technical Losses in Power System and Monitoring of Electricity Theft Over Low-Tension Poles**. Second International Conference on

Advances in Computing and Communication Engineering, 2015. IEEE. DOI 10.1109/ICACCE.2015.106.

CHEN, D. et al. **Métricas de Avaliação em Machine Learning: Classificação**. 2020. Disponível em: <<https://medium.com/kunumi/metricas-de-avaliacao-em-machine-learningclassificacao-49340dadb198>>. Acesso em: jan. 2021.

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

CONAB, **Acompanhamento da safra brasileira**. Café, v.4 –Safra 2017, n.4 - Quarto Levantamento, Brasília – DF, p. 1- 84, dez.2017.

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

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

ESTEVES, B. S; SILVA, D. G. et al. **Irrigação por Gotejamento**. Manual Técnico, n. 32, ISSN 1983-5671. Niteroi, RJ, 2012.

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 NonTechnical Losses Detection. **IEEE Transactions on Smart Grid**, v. 10, n. 3, p. 3226–3235, 2019.

FERNANDES, A. L.; LIMA, L. A. **Características dos Principais Sistemas de Irrigação do Cafeeiro. Condução da Lavoura**. Disponível em <<https://www.esalq.usp.br/visaoagricola/sites/default/files/va12-conducao-da-lavoura03.pdf>> Acesso em: março 2020.

FERNANDES, A. L.; LIMA, L. A. **Irrigação do cafeeiro: quando, quanto e por que se deve utilizar?**. Disponível em <<https://www.esalq.usp.br/visaoagricola/sites/default/files/va12-conducao-da-lavoura04.pdf>> Acesso em: março 2020.

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.

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.

HAQ, E.; HUARONG, X.; KHATTAK, M. I. **A Review of Various Clustering Techniques**. International Journal of Engineering and Applied Computer Science (IJEACS). Volume 2: Issue:01, January, 2017. ISBN: 978-0-9957075-2-8.

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.

HAROON, D. Python **Machine Learning Case Studies: Five Case Studies for the Data Scientist**. 204 p. ISBN: 9781484228234. Outubro de 2017.

KHAFRAF, N. A.; JALILI, M.; SOKOLOWSKI, P. **Optimal Clustering of Energy Consumers based on Entropy of the Correlation Matrix Between Clusters**. Electrical and Biomedical Engineering RMIT University, Melbourne, Australia. Disponível em <<https://doi.org/10.48550/arXiv.1903.01159>> Acesso em: 2019.

KRISTINA S.; Dmitrii B.; et. al. **Power transmission and distribution losses – A model based on available empirical data and future trends for all countries globally**. International Journal of Electrical Power & Energy Systems, Volume 107, 2019, Pages 98-109, ISSN 0142-0615. Disponível em <<https://doi.org/10.1016/j.ijepes.2018.11.012>> Acesso em: jan. 2022.

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.

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.

MARIN, F. **Balanco Hídrico**. Notas de Aula. Universidade De São Paulo. Disponível em: <https://edisciplinas.usp.br/pluginfile.php/5067838/mod_resource/content/0/LEB306_Marin%20Topico%208%20-%20Balanco%20Hidrico.pdf>. Acesso em: nov. 2021.

MAROTTA, V. **Aprendizado não supervisionado com o kMeans**. Viçosa, 2019. Disponível em: <<https://www.researchgate.net/publication/334574592>>. Acesso em: 13 mai. 2022.

MEIRELES, E. J. L. et al. **Fenologia do cafeeiro: condições agrometeorológicas e balanço hídrico do ano agrícola 2004–2005**. Brasília - DF: Embrapa Café, Documentos, n. 5. 2009. 128 p.

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.

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) - Universidade Estadual Paulista, Ilha Solteira, RS, 2009.

ORGANIZAÇÃO INTERNACIONAL DO CAFÉ. **Relatório Sobre o Mercado do Café**. Agosto de 2021.

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, A. R; ANGELOCCI L. R.; SENTELHAS, P. C. **Meteorologia Agrícola**. Universidade de São Paulo. Notas de Aula, Escola Superior de Agricultura Luiz de Queiroz. DEpartamento de Ciências Exatas. Piracicaba, SP, fevereiro de 2007.

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.

PORRAS, A. J; RIVERA, O. H. et. al. Identification of Non-Technical Electricity Losses in Power Distribution Systems by Applying Techniques of Information Analysis and Visualization. **IEEE Latin America Transactions**, v 13, n 3, março de 2015.

RAJVANSHI, S.; CHAUHAN, A. **Non Technical Losses in Power System: A Review**. 2013 International Conference on Power, Energy and Control (ICPEC). IEEE, 2013.

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.

RAMOS, H. E; MEDEIROS, T. T. et. al. **Análise das Variáveis Agrometeorológicas Para a Cultura do Café em Muniz Freire-ES entre 2016 e 2017.** Anais...Simpósio de Pesquisas dos Cafés do Brasil, ISSN: 1984-9249. Vitória – ES, outubro de 2019.

RODRIGUES, A. C. **Deteção e Identificação de Perdas Não-Técnicas em Sistemas de Distribuição via Estimação de Estados e Testes Geometricos.** 2018. 194 p. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis, SC, 2018.

RONCHI, C. P., et al. **Respostas ecofisiológicas de cafeeiros submetidos ao déficit hídrico para concentração da florada no Cerrado de Minas Gerais.** Pesquisa Agropecuária Brasileira, v.50, n.1, p.24-32. 2015.

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.

SARNIGHAUSEN. V. C; GOMES, F. G. et al. **Estimativa da Evapotranspiração de Referência para Botucatu -SP Por Meio de Regressão.** Revista Brasileira de Climatologia. ISSN: 2237-8643. Revista Eletrônica. Ano 17, v. 28, 2021.

SENAR. **Irrigação: Fertirrigação e Reuso de efluentes.** Brasília, 2019. Serviço Brasileiro de Aprendizagem Rural, Coleção SENAR n. 254. 2019.

SHARMA, S.; MAJUMDAR, A. **Unsupervised Detection on Non-Technical Losses via Recursive Transform Learning.** IEEE Transactions on Power Delivery, vol. 36, no. 2, pp. 1241-1244, April 2021, doi: 10.1109/TPWRD.2020.3029439.

SOUZA, A. J.; CAMILO, W. R. **Irrigação na cafeicultura: seis sistemas possíveis.** Disponível em: < <https://www.cafepoint.com.br/noticias/tecnicas-de-producao/irrigacao-na-cafeicultura-seis-sistemas-possiveis-95464n.aspx>>. Acesso em: 23 fev. 2020.

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.

THORNTHWAITE, C.W., MATHER, J.R. **The water balance.** Publications in Climatology, v.8, n.1, 1955. 104p.

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.

VIEIRALVES, X. E. **Proposta de uma Metodologia para Avaliação das Perdas Comerciais dos Sistemas Elétricos. O Caso Manaus.** 157 p. Dissertação (Mestrado em Planejamento de Sistemas) – Universidade Estadual de Campinas, Campinas, SP, 2005.

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.

ZUEGE, T. J. **Aplicação de Técnicas de Mineração de Dados para Detecção de Perdas Comerciais na Distribuição de Energia Elétrica**. 94 p. Monografia (Departamento de Engenharia da Computação) – Universidade do Vale do Taquari, Lajeado, RS, 2018.

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 – CODIGO DE IMPLEMENTAÇÃO DA METODOLOGIA.

No apêndice é apresentado partes do código de programação em *Python* utilizado para implementação da metodologia e aplicação dos estudos de casos. O código é apresentado por partes.

- Código Parte 1 - Importação de bibliotecas e pacotes.

```

Created on Fri Jan 7 00:55:35 2022
Natalia Bastos de Sousa
@author: rhyo_
"""

import matplotlib.pyplot as plt
import numpy as np
from pandas import ExcelWriter
from shapely.geometry import Point, Polygon
import itertools
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.ticker import NullFormatter
import pandas as pd
import numpy as np
import matplotlib.ticker as ticker
from sklearn import preprocessing
import seaborn as sns
from geopy.geocoders import Nominatim
import folium
import pandas as pd
import shapefile as shp
from shapely.ops import cascaded_union
from sklearn.preprocessing import MinMaxScaler
import statistics
from statistics import mean
import geopandas
import math
import matplotlib.patches as patches
from k_means_constrained import KMeansConstrained
from sklearn.cluster import KMeans
from scipy.cluster.vq import kmeans,vq

from sklearn.ensemble import RandomForestClassifier
from sklearn import tree
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.model_selection import train_test_split

```

- Código Parte 2 – Leitura de Dados de Entrada: nessa parte é feita a leitura dos dados necessários via ambiente IDE Spyder. São dados da pesquisa de mapeamento tecnológico, dados de CAD, *Shapefiles*, e dados das unidades consumidoras.

```

MT_dados = pd.read_excel('MT_Dados_v3.xlsx')
CAD_dados = pd.read_excel('CAD_AWC_media_municipal2.xlsx')
Dados_4_cenas_Uruguaiana = pd.read_excel('Dados_4_cenas_Uruguaiana.xlsx')
centroide_Todas = pd.read_excel('centroides_todos.xlsx')
data_1A = pd.read_csv('110_GRUPO_A_CONSUMOS_CONSULTA_IRRIG_NOTURNA_COMPLETO_090221.txt',
                    delimiter="|")
data_1B = pd.read_csv('109_GRUPO_B_CONSUMOS_CONSULTA_IRRIG_NOTURNA_COMPLETO_090221.txt',
                    delimiter="|")
file = pd.read_excel('2381_PNT_IRRIGACAO_SAO_PAULO_CADASTRO_050422_3.xlsx')
data_2021 = pd.read_excel('2341_1085_CADASTRO_MED_071221.xlsx')
data_2022_52 = pd.read_excel('2352_1085_CADASTRO_MED_240122.xlsx')
data_2022_53 = pd.read_excel('2353_1085_CONSUMOS_MED_240122.xlsx')
data_2022_54 = pd.read_excel('2354_1085_CONSUMOS_GRA_MED_240122.xlsx')
bh = pd.read_excel('bhTeste.xlsx')
bh_2 = pd.read_excel('bhTeste2.xlsx')
areas = pd.read_excel('output_22KHC_Cafe.xlsx')

crs = {'init': 'epsg:4326'}

states = geopandas.read_file('ANMS2010_03_solos15000\ANMS2010_03_solos15000.shp')
type(states)
states.crs
desnivel_SP = geopandas.read_file('Desnivel_SP\Desnivel_SP.shp')
type(desnivel_SP)
desnivel_SP.crs
SP_Cafe1 = geopandas.read_file(r'C:\Users\rhyo_\Google Drive\Pós\Dissertação\Dados CPFL')
type(SP_Cafe1)
UBC_v2 = geopandas.read_file(r'C:\Users\rhyo_\Google Drive\Pós\Dissertação\Dados CPFL\Mc')
type(UBC_v2)

```

- Código Parte 3 – Primeiras manipulações nos dados de consumo. Nessa parte é selecionado as unidades consumidoras dos municípios de interesse.

```

##### CONSUMO #####

municipios_SP = file['DESC_MUNICIPIO'].value_counts().reset_index()

file_1_cafe = file[file['DESC_MUNICIPIO'].isin(['CRISTAIS PAULISTA', 'PEDREGULHO',
                                             'FRANCA', 'RIBEIRAO CORRENTE', 'JERIQUARA', 'PATROCINIO PAULISTA'])]

file_1_cana = file[file['DESC_MUNICIPIO'].isin(['GUAIRA', 'BARRETOS', 'ITUVERAVA'])]

file_2_cafe = file_1_cafe[['COD_INSTALACAO', 'COD_GRUPO_TENSAO', 'DESC_MUNICIPIO',
                          'NUM_COORDENADA_GIS_X', 'NUM_COORDENADA_GIS_Y', 'COD_ALIMENTADOR']]

meses = pd.date_range(start='1/1/2015', end='1/1/2021', freq='MS')

```

- Código Parte 4 – Nessa seção do código são inseridos os novos dados de consumo médio móvel (CONSUMO_mean_12) e desvio (CONSUMO_std_12) para a janela (rolling) de 4 meses. Os dados fornecidos pela distribuidora são divididos em Grupo A e Grupo B, dessa forma, a manipulação dos dados neste ponto ainda é realizada de maneira individual.

```

data_1A_2 = data_1A[['COD_INSTALACAO', 'DT_REFERENCIA', 'CONSUMO']]
data_1A_2['CONSUMO_mean_12'] = data_1A_2.groupby('COD_INSTALACAO')['CONSUMO'].rolling(4).mean().reset_index(0, drop=True)
data_1A_2['CONSUMO_std_12'] = data_1A_2.groupby('COD_INSTALACAO')['CONSUMO'].rolling(4).std().reset_index(0, drop=True)

data_1A_3 = data_1A.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['CONSUMO'].sum().unstack(-1).reset_index()
data_1A_4 = data_1A_3.copy()
data_1A_5 = data_1A_4.iloc[:, 1:]
data_1A_melt = pd.melt(data_1A_4, id_vars='COD_INSTALACAO', value_vars=data_1A_5, var_name='DT_REFERENCIA', value_name='CONSUMO')
data_1A_melt['CONSUMO_mean_12'] = data_1A_melt.groupby('COD_INSTALACAO')['CONSUMO'].rolling(4).mean().reset_index(0, drop=True)
data_1A_melt['CONSUMO_std_12'] = data_1A_melt.groupby('COD_INSTALACAO')['CONSUMO'].rolling(4).std().reset_index(0, drop=True)
data_1B_2 = data_1B[['COD_INSTALACAO', 'DT_REFERENCIA', 'QTD_ENERGIA_FATURADA']]
data_1B_2['CONSUMO_mean_12'] = data_1B_2.groupby('COD_INSTALACAO')['QTD_ENERGIA_FATURADA'].rolling(4).mean().reset_index(0, drop=True)
data_1B_2['CONSUMO_std_12'] = data_1B_2.groupby('COD_INSTALACAO')['QTD_ENERGIA_FATURADA'].rolling(4).std().reset_index(0, drop=True)
data_1B_3 = data_1B.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['QTD_ENERGIA_FATURADA'].sum().unstack(-1).reset_index()
data_1B_2 = data_1B_2.rename(columns={'QTD_ENERGIA_FATURADA': 'CONSUMO'})
data_1B = data_1B.rename(columns={'QTD_ENERGIA_FATURADA': 'CONSUMO'})
data_1B_3 = data_1B.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['CONSUMO'].sum().unstack(-1).reset_index()
data_1B_4 = data_1B_3.copy()
data_1B_5 = data_1B_4.iloc[:, 1:]
data_1B_melt = pd.melt(data_1B_4, id_vars='COD_INSTALACAO', value_vars=data_1B_5, var_name='DT_REFERENCIA', value_name='CONSUMO')

```

- Código Parte 5 – Neste ponto do código são unidos em um único *DataFrame* as planilhas de grupo A e grupo B.

```

data_A2 = data_1A_melt.reset_index(drop = True)
data_B2 = data_1B_melt.reset_index(drop = True)
data_A_1 = pd.merge(file_2_cafe, data_A2, on = 'COD_INSTALACAO', how='inner')
data_B_1 = pd.merge(file_2_cafe, data_B2, on = 'COD_INSTALACAO', how='inner')
df_union_all = pd.concat([data_A_1, data_B_1 ], sort=True).reset_index(drop = True)

```

- Código Parte 6 – Nessa etapa são inseridos os outros dois novos atributos de consumo máximo e consumo mínimo móvel da janela de 4 meses, mais uma manipulação no atributo de DT_REFERENCIA, para que possamos selecionar apenas os anos de interesse usando fórmulas matemáticas.

```
df_union_all['max_12'] = df_union_all['CONSUMO'].rolling(window=4).max()
df_union_all['min_12'] = df_union_all['CONSUMO'].rolling(window=4).min()

stdMean_union_cafe_2 = df_union_all.copy()

stdMean_union_cafe_2['DT_REFERENCIA_2'] = stdMean_union_cafe_2['DT_REFERENCIA']/100
stdMean_union_cafe_2 = stdMean_union_cafe_2.reset_index().drop(["index"], axis=1)
```

- Código Parte 7 – Nessa parte do código é feita a manipulação para selecionar o período de 73 meses (2015-2021). Para isso, foi utilizada a lógica: $201501/100 = 2015.01$. Dessa forma podemos pegar a parte inteira (2021.01) e excluir todos os index do *DataFrame* cuja parte inteira seja menor do que o ano de interesse ($2021 - 6 = 2015$).

```
maxi = stdMean_union_cafe_2['DT_REFERENCIA_2'].max()
indexss = []

for i in range(len(stdMean_union_cafe_2)):
    a = int(stdMean_union_cafe_2['DT_REFERENCIA_2'][i])
    if a < int(maxi) - 6:
        indexss.append(stdMean_union_cafe_2.index[i])

stdMean_union_cafe_2 = stdMean_union_cafe_2.drop(stdMean_union_cafe_2.index[indexss])
stdMean_union_cafe_2 = stdMean_union_cafe_2[stdMean_union_cafe_2.DT_REFERENCIA_2 != 2021.02]

stdMean_union_cafe_2 = stdMean_union_cafe_2.reset_index().drop(["index"], axis=1)

cod_instalacao = stdMean_union_cafe_2.COD_INSTALACAO.value_counts().reset_index()
cod_instalacao = cod_instalacao.rename(columns = {"index":"COD_INSTALACAO"})
cod_instalacao_2 = cod_instalacao.iloc[:,0]
```

- Código Parte 8 – Leitura de Áreas: Nessa seção são feitas as manipulações do *DataFrame* gerado pela leitura da base de dados das áreas. São selecionados os municípios de interesse, a área é transformada de km² para hectares. São gerados pontos utilizando a biblioteca *geopandas* para poder plotar os pontos por georreferenciamento.

```
##### Leitura das áreas #####

data_SP_todas = areas.copy()
municipios_SP_2 = data_SP_todas['DESC_MUNICIPIO'].value_counts().reset_index()
data_SP_todas_2 = data_SP_todas[data_SP_todas['DESC_MUNICIPIO'].isin(['Cristais Paulista',
                                                                    'Pedregulho', 'Franca',
                                                                    'Ribeirão Corrente', 'Jeriquara', 'Patrocínio Paulista'])]
data_SP_todas_2['Área'] = data_SP_todas_2['Área']/10000

gdfsaopaulo = geopandas.GeoDataFrame(data_SP_todas_2, geometry=geopandas.points_from_xy(data_SP_todas_2.Long,
                                                                                       data_SP_todas_2.Lat))
gdfsaopauloAreas = gdfsaopaulo['Área'].value_counts()
gdfsaopaulo = gdfsaopaulo.rename(columns = {"UC_1": "COD_INSTALACAO"})

gdfUCs = geopandas.GeoDataFrame(stdMean_union_cafe_2, geometry=geopandas.points_from_xy(stdMean_union_cafe_2.Long,
                                                                                       stdMean_union_cafe_2.Lat))

gdfUCs2 = pd.merge(gdfUCs, cod_instalacao_2, on = 'COD_INSTALACAO', how='inner')
gdfUCs3 = gdfUCs2[['COD_INSTALACAO', 'Lat', 'Long', 'DESC_MUNICIPIO', 'geometry']].drop_duplicates()
```

- Código Parte 9 – Nesta parte do código são unidas, pelo COD_INSTALACAO, os *DataFrames* de consumo e áreas, eliminadas o conjunto de áreas x UC que estão distantes entre si mais de 1,2 km.

```
dadosCentroUC = pd.merge(cod_instalacao_2, gdfsaopaulo, on = 'COD_INSTALACAO', how='inner')
dadosCentroUC = dadosCentroUC[['COD_INSTALACAO', 'Número', 'Imagem_Base', 'Data', 'Dimensão',
                              'Comparação_Círculo', 'Distância_UC1_km', 'point_lavoura', 'Lat', 'Long',
                              'DESC_MUNICIPIO', 'Área', 'Cultivar', 'Distância_Corpo_Hídrico_km',
                              'geometry']]

teste = dadosCentroUC[dadosCentroUC['Distância_UC1_km'] < 1.2]

dadosCentroUC_2 = dadosCentroUC.sort_values(['Distância_UC1_km'], ascending=False).reset_index().drop(["index"], axis=1)
teste_2 = teste[['COD_INSTALACAO', 'DESC_MUNICIPIO', 'Lat', 'Long']]

teste_3 = geopandas.GeoDataFrame(teste_2, geometry=geopandas.points_from_xy(teste_2.Long, teste_2.Lat))

teste_4 = teste_3[['COD_INSTALACAO']].drop_duplicates()
gdfUCsNova = pd.merge(gdfUCs, teste_4, on = 'COD_INSTALACAO', how='inner').drop_duplicates()
gdfUCsNova2 = gdfUCsNova[['COD_INSTALACAO', 'DESC_MUNICIPIO', 'Lat', 'Long']].drop_duplicates().reset_index().drop(["index"], axis=1)
```

- Código Parte 10 – Nesta seção do código é inserido o novo atributo de Classe de Perfil de Consumo, após verificar o número de k ideal, pelo método de *Elbow*.

```
X = gdfUCsNova[['CONSUMO_mean_12', 'CONSUMO_std_12', 'max_12', 'min_12']].copy()
X = X.fillna(0)
data = X
distorsions = []
for k in range(2, 10):
    k_means = KMeans(n_clusters=k, random_state=42)
    k_means.fit(X)
    distorsions.append(k_means.inertia_)

#fig = plt.figure(figsize=(15, 5))
#plt.plot(range(2, 10), distorsions)
#plt.grid(True)
#plt.title('Curva Elbow')

centroids,_ = kmeans(data,4)
# associar cada amostra para um cluster
idx,_ = vq(data,centroids)

gdfUCsNova['Cluster_PF'] = idx
Perfis2 = gdfUCsNova['Cluster_PF'].value_counts().reset_index()
Perfis2 = Perfis2.rename(columns = {"index":"Classes", "Cluster_PF":"Nº de UCS"})
```

- Código Parte 11 – Associação de valores de CAD e desnível: Nesta etapa escolheu-se exportar a planilha com valores associados de desnível e CAD, pois a leitura do *shapefile* contendo os polígonos de desnível mostrou-se demorada, exigindo muito processamento. Dessa forma, pode-se gerar a associação aos desníveis uma vez e salvar como uma nova base de dados em Excel, para realizar a leitura no próprio algoritmo posteriormente.

```

data_CAD_2 = CAD_dados.copy()
data_CAD_2.columns = ['DESC_MUNICIPIO', 'UF', 'CAD[%]']
data_CAD_2_2 = data_CAD_2.loc[(data_CAD_2['UF'] == 'SP')]
data_CAD_3 = data_CAD_2_2[['DESC_MUNICIPIO', 'CAD[%]']]
data_CAD_3.DESC_MUNICIPIO = data_CAD_3.DESC_MUNICIPIO.str.upper()
data_CAD_4=data_CAD_3.copy()

cols = data_CAD_4.select_dtypes(include=[np.object]).columns
data_CAD_4[cols] = data_CAD_4[cols].apply(lambda x: x.str.normalize('NFKD').str.encode('ascii',
    errors='ignore').str.decode('utf-8'))

gdfUCs5 = pd.merge(data_CAD_4, gdfUCs3, on ='DESC_MUNICIPIO', how='inner')
data_union_12 = gdfUCs5.copy()
data_union_13 = data_union_12.tail()

def get_seaname(long,lat):
    pnt = Point(long,lat)
    for i,j in enumerate(desnivel_SP.geometry):
        if pnt.within(j):
            return desnivel_SP.DN.iloc[i]

data_union_12['DN'] = data_union_12.apply(lambda x: get_seaname(x['Long'], x['Lat']), axis=1)

with pd.ExcelWriter('dadosSP.xlsx') as writer:
    data_union_12.to_excel(writer,sheet_name='SP')

```

- Código Parte 12 - Etapa de Preparação de dados para a implementação do modelo RF e testes.

```

#### Dados para modelo RF#####

gdfUCsNova3 = pd.merge(gdfUCsNova, teste_4, on ='COD_INSTALACAO', how='inner')
gdfUCsNova4 = gdfUCsNova3[gdfUCsNova3['DT_REFERENCIA_2']>2020]
gdfUCsNova4 = gdfUCsNova4[gdfUCsNova4['DT_REFERENCIA_2']<2021]
gdfUCsNova5 = gdfUCsNova4.copy()
gdfUCsCONSUMO = gdfUCsNova5.CONSUMO.fillna(0)
gdfUCsNova5['CONSUMO'] = gdfUCsCONSUMO.values
gdfUCsNova6 = gdfUCsNova5.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['CONSUMO'].sum().unstack(-1).reset_index()
gdfUCsNova4_2 = gdfUCsNova4[['COD_INSTALACAO', 'DT_REFERENCIA', 'Cluster_PF']]
gdfUCsNova4_3 = gdfUCsNova4_2.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['Cluster_PF'].sum().unstack(-1).reset_index()
gdfUCsNova4_4 = gdfUCsNova4_3.iloc[:,9:]
gdfUCsNova4_5 = gdfUCsNova4_4.iloc[:,2]
gdfUCsNova7 = gdfUCsNova5.groupby(['COD_INSTALACAO', 'DT_REFERENCIA'])['CONSUMO_mean_12'].sum().unstack(-1).reset_index()
gdfUCsNova6_3 = gdfUCsNova6.iloc[:,9:]

dados_teste = pd.merge(file_1_cafe, teste_4, on ='COD_INSTALACAO', how='inner')

```

- Código Parte 13 – Modelo RF: o código para implementação do modelo RF é apresentado, foi especificado que os dados estivessem no estado randômico 42, 70% dos dados foram utilizados como treino e 30% utilizados para teste.

```
X7 = gdfUCsNova6_3.fillna(0)
Dados_Entrda_RF = gdfUCsNova6_3.fillna(0)
y7= gdfUCsNova4_5
X_train, X_test, y_train, y_test = train_test_split(X7,y7,test_size=0.30, random_state=42)
clf = RandomForestClassifier(n_estimators=50, random_state=42, class_weight = 'balanced')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)

print("\n")
print('Acuracia do Classificador RF nos dados de Treino: {:.2f}'
      .format(clf.score(X_train, y_train)))
print('Acuracia do Classificador RF nos dados de Teste: {:.2f}'
      .format(clf.score(X_test, y_test)))

print("\n")
print("Matriz de Confusão:")
print(confusion_matrix(y_test,y_pred))
print("\n")
print(classification_report(y_test,y_pred))
print("Potuação de Acuracia:",accuracy_score(y_test, y_pred))
print("\n")
```

- Código Parte 14 – Nessa parte do código é apresentado como foram realizados os estudos de casos para o modelo RF, inserindo o código da UC de interesse, são armazenados em variáveis auxiliares os consumos de 2020, 2021 e a Classe de Perfil de Consumo. No caso apresentado, foi simulado um consumo menor (consumo real multiplicado por 0,68) para a UC.

```

##### Estudo de Caso 1
X_teste_37661850_20 = gdfUCsNova4_3.loc[gdfUCsNova4_3['COD_INSTALACAO'] == 19699700].reset_index()
X_teste_37661850_21 = dados_teste.loc[dados_teste['COD_INSTALACAO'] == 19699700].reset_index()
X_teste_37661850_21 = dados_teste.loc[dados_teste['COD_INSTALACAO'] == 19699700].reset_index()

X_preditivo = X_teste_37661850_21[['CONSUMO_9', 'CONSUMO_10', 'CONSUMO_11', 'CONSUMO_12']]
X_preditivo_2 = gdfUCsNova6.loc[gdfUCsNova6['COD_INSTALACAO'] == 19699700].reset_index()
X_preditivo_3 = X_preditivo_2[['202009', '202010', '202011', '202012']]

classe_cod = clf.predict(X_preditivo*0.68)
print("\n")
print("Estudo de Caso 1")
print("#####")
print("Dados de Consumo de entrada:")
print(X_preditivo*0.68)
print("\n")
print("Classe Original (consumos de 2020):", int(X_teste_37661850_20[202010]))
#print("Relatório Final de Simulação")
print("Classe Prevista (consumos de 2021):", classe_cod)
print("\n")
if int(X_teste_37661850_20[202010]) == classe_cod:
    print("Situação: Não Houve Mudanças de Perfil")
if int(X_teste_37661850_20[202010]) == 1 and classe_cod == 2:
    print("Situação: Possível Ocorrência de Perdas Não Técnicas")

if int(X_teste_37661850_20[202010]) == 1 and classe_cod == 3:
    print("Situação: Possível Ocorrência de Perdas Não Técnicas")

if int(X_teste_37661850_20[202010]) == 1 and classe_cod == 0:
    print("Situação: Possível Ocorrência de Perdas Não Técnicas")

```

- Código Parte 15 – A preparação dos dados para a simulação de estimativa de consumo é apresentada nesta parte do código, onde são lidas as informações da base de dados do mapeamento tecnológico das lavouras de café, realizado pela autora. Neste caso foi selecionado o município de Jeriquara, após escolheu-se uma UC no município para realizar a simulação do estudo de caso.

```

MT_dados_2 = MT_dados[['CULTURA', 'Kc_Jan', 'Kc_Fev', 'Kc_Mar', 'Kc_Abr', 'Kc_Mai', 'Kc_Jun', 'Kc_Jul',
                    'Kc_Ago', 'Kc_Set', 'Kc_Out', 'Kc_Nov', 'Kc_Dez', 'Kc_Jan_1', 'Kc_Fev_1',
                    'Kc_Mar_1', 'Kc_Abr_1', 'Kc_Mai_1', 'Kc_Jun_1', 'Kc_Jul_1', 'Kc_Ago_1',
                    'Kc_Set_1', 'Kc_Out_1', 'Kc_Nov_1', 'Kc_Dez_1', 'Profundidade', 'Tipo_Irrigacao',
                    'Tempo_Irrigacao', 'EFICIENCIA', 'Pressão_De_Servico', 'LAMINA', 'VAZAO']]

dadosSP_1= dadosSP[['COD_INSTALACAO', 'CAD[%]', 'DN']]
Dados_SP_3 = pd.merge(dadosSP_1, gdfteste, on = 'COD_INSTALACAO', how='inner')
4
Dados_SP_4 = pd.merge(Dados_SP_3, MT_dados_2, on = 'CULTURA', how='inner')

Dados_SP_5 = Dados_SP_4[Dados_SP_4['DESC_MUNICIPIO'].isin(['Jeriquara'])]

Dados_SP_caso = Dados_SP_5.loc[Dados_SP_5['COD_INSTALACAO'] == 19699697]

Dados_SP_caso2 = Dados_SP_caso[['COD_INSTALACAO', 'CAD[%]', 'DN', 'Lat', 'Long', 'DESC_MUNICIPIO',
                                'AREA', 'Profundidade', 'Tipo_Irrigacao', 'EFICIENCIA',
                                'Pressão_De_Servico', 'L', 'Hg', 'AMT', 'LAMINA', 'CULTURA',
                                'Kc_Jan', 'Kc_Fev', 'Kc_Mar', 'Kc_Abr', 'Kc_Mai', 'Kc_Jun',
                                'Kc_Jul', 'Kc_Ago', 'Kc_Set', 'Kc_Out', 'Kc_Nov', 'Kc_Dez']]

```

- Código Parte 16 – Antes do início do cálculo do balanço hídrico e estimativa de consumo, insere-se os últimos atributos necessários. Utilizando os dados de diferentes bases de dados, que foram apresentadas na metodologia.

```

Dados_SP_caso['L'] = 380

Dados_SP_caso['Hg'] = (Dados_SP_caso.L * Dados_SP_caso.DN)/100

Dados_SP_caso['AMT'] = Dados_SP_caso.Hg + (Dados_SP_caso.Hg *0.1) + Dados_SP_caso.Pressão_De_Servico

Dados_SP_caso['LAMINA'] = Dados_SP_caso['Profundidade'] * Dados_SP_caso['CAD[%]']

ETC = bh['ETO'].values * Dados_SP_5[['Kc_Jan', 'Kc_Fev', 'Kc_Mar', 'Kc_Abr', 'Kc_Mai', 'Kc_Jun', 'Kc_Jul',
                    'Kc_Ago', 'Kc_Set', 'Kc_Out', 'Kc_Nov', 'Kc_Dez', 'Kc_Jan_1', 'Kc_Fev_1',
                    'Kc_Mar_1', 'Kc_Abr_1', 'Kc_Mai_1', 'Kc_Jun_1', 'Kc_Jul_1', 'Kc_Ago_1',
                    'Kc_Set_1', 'Kc_Out_1', 'Kc_Nov_1', 'Kc_Dez_1']].T.values

bh['ETC'] = ETC.T

CAD = Dados_SP_caso['Profundidade'][28] * Dados_SP_caso['CAD[%]'][28]

bh['P-ETC'] = bh['PRECIPITACAO'].values - bh['ETC'].values
bh['ETC-P'] = bh['ETC'].values - bh['PRECIPITACAO'].values

```

- Código Parte 17 – O cálculo do balanço hídrico e estimativa de consumo, são realizados seguindo o código apresentado, o passo a passo foi descrito na metodologia. Onde ao final, teremos a coluna Energia com a energia estimada do sistema irrigante para a UC de estudo.

```

for i in range(len(bh)):
    if bh['P-ETC'][i]>0:
        bh['NAC'][i] = 0
    if bh['P-ETC'][i]<0:
        bh['NAC'][i] = bh['P-ETC'][i] + bh['NAC'][i-1]
for i in range(len(bh)):
    bh['ARM'][i] = CAD *math.exp(bh['NAC'][i]/CAD)
for i in range(len(bh)):
    if i==0:
        bh['ARM'][i-1] = bh['ARM'].tail(1)
    bh['ALTERACAO'][i] = bh['ARM'][i] - bh['ARM'][i-1]
for i in range(len(bh)):
    if bh['PRECIPITACAO'][i]>= bh['ETC'][i]:
        bh['ETr'][i] = bh['ETC'][i]
    if bh['PRECIPITACAO'][i] < bh['ETC'][i]:
        bh['ETr'][i] = bh['PRECIPITACAO'][i] + abs(bh['ALTERACAO'][i])
for i in range(len(bh)):
    if bh['ARM'][i] < CAD:
        bh['Excedente'][i] = 0
    if bh['ARM'][i] == CAD:
        bh['Excedente'][i] = bh['P-ETC'][i] - bh['ALTERACAO'][i]
for i in range(len(bh)):
    bh['Deficit'][i] = bh['ETC'][i] - bh['ETr'][i]
for i in range(len(bh)):
    if bh['ETC-P'][i]<0:
        bh['Lamina_bruta'][i] = 0
    if bh['ETC-P'][i]>0:
        bh['Lamina_bruta'][i] = bh['ETC-P'][i]/(Dados_SP_caso['EFICIENCIA']/100)
for i in range(len(bh)):
    if bh['Lamina_bruta'][i]<0:
        bh['Tempo_Irrigacao'][i] = 0
    if bh['Lamina_bruta'][i]>0:
        bh['Tempo_Irrigacao'][i] = bh['Lamina_bruta'][i]/CAD
for i in range(len(bh)):
    if bh['ETC-P'][i]<0:
        bh['Vazao'][i] = 0
    if bh['ETC-P'][i]>0:
        bh['Vazao'][i] = (bh['Lamina_bruta'][i] * Dados_SP_caso['AREA'])/bh['Tempo_Irrigacao'][i]
for i in range(len(bh)):
    if bh['ETC-P'][i]<0:
        bh['Vazao'][i] = 0
    if bh['ETC-P'][i]>0:
        bh['Potencia'][i] = (bh['Vazao'][i]* Dados_SP_caso['AMT']*0.0098)/0.40
for i in range(len(bh)):
    if bh['ETC-P'][i]<0:
        bh['Vazao'][i] = 0
    if bh['ETC-P'][i]>0:
        bh['Energia'][i] = bh['Potencia'][i]*bh['Tempo_Irrigacao'][i]

```

- Código Parte 18 – A última parte do algoritmo é a construção do relatório final de simulação, onde os consumos estimados são comparados com o consumo real (Consumo_mensal_3), composto pelos consumos registrados na data de leitura, e o real modificado (Consumo_mensal_comp), composto pela média de dois meses e suas primeiras e últimas quinzenas.

```

print("\n")
print("Estudo de Caso 1: Estimativa de Consumo")
print("#####")
bh.fillna(0)
for i in range(len(bh)):

    if bh['Energia'][i]>Consumo_mensal_Comp['CONSUMO'][i]:
        print('Mês:',i+1)
        print("Indicador 1: Consumo observado menor do que o Estimado")
        print("Situação: Possível Ocorrência de Perdas Não Técnicas:")
    if bh['Energia'][i]<Consumo_mensal_Comp['CONSUMO'][i]:
        print("Indicador 2: Consumo observado maior do que o Estimado")
        var = ((Consumo_mensal_Comp['CONSUMO'][i] - bh['Energia'][i])/Consumo_mensal_Comp['CONSUMO'][i])*100
        print("Variação:", var)

Consumo_mensal_3.fillna(0)
for i in range(len(bh)):

    if bh['Energia'][i]>Consumo_mensal_3['CONSUMO'][i]:
        print('Mês:',i+1)
        print("Indicador 1: Consumo observado menor do que o Estimado")
        print("Situação: Possível Ocorrência de Perdas Não Técnicas:")
    if bh['Energia'][i]<Consumo_mensal_3['CONSUMO'][i]:
        print("Indicador 2: Consumo observado maior do que o Estimado")
        var = ((Consumo_mensal_3['CONSUMO'][i] - bh['Energia'][i])/Consumo_mensal_3['CONSUMO'][i])*100
        print("Variação:", var)

```