



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
CURSO DE ENGENHARIA DE PETRÓLEO

JOÃO VICTOR GONÇALVES CASTRO

**CARACTERIZAÇÃO DE ELETROFÁCIES UTILIZANDO APRENDIZADO DE
MÁQUINA: APLICAÇÃO DO ALGORITMO K-MEANS EM PERFIS GEOFÍSICOS**

FORTALEZA

2024

JOÃO VICTOR GONÇALVES CASTRO

CARACTERIZAÇÃO DE ELETROFÁCIES UTILIZANDO APRENDIZADO DE MÁQUINA: APLICAÇÃO DO ALGORITMO K-MEANS EM PERFIS GEOFÍSICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Petróleo do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia de Petróleo.

Orientador: Prof. PhD. Luis Glauber Rodrigues.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C351c Castro, João Victor Gonçalves.
Caracterização de eletrofácies utilizando Aprendizado de Máquina: aplicação do algoritmo K-Means em perfis geofísicos / João Victor Gonçalves Castro. – 2024.
72 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia de Petróleo, Fortaleza, 2024.
Orientação: Prof. Dr. Luis Glauber Rodrigues.

1. Algoritmos. 2. Eletrofácies. 3. Perfis. 4. Precisão. 5. Aprendizado. I. Título.

CDD 665.5092

JOÃO VICTOR GONÇALVES CASTRO

CARACTERIZAÇÃO DE ELETROFÁCIES UTILIZANDO APRENDIZADO DE MÁQUINA: APLICAÇÃO DO ALGORITMO K-MEANS EM PERFIS GEOFÍSICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Petróleo do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia de Petróleo.

Aprovada em: 10/10/2024.

BANCA EXAMINADORA

Prof. PhD. Luis Glauber Rodrigues (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Pedro Felipe Gadelha Silvino
Universidade Federal do Ceará (UFC)

Prof. Dr. Felipe Xavier Feitosa
Universidade Federal do Ceará (UFC)

A Deus.

Aos meus pais, Marlene e Antonio.

AGRADECIMENTOS

A Deus, pela minha vida e por iluminar meu caminho até aqui, auxiliando-me em ultrapassar qualquer obstáculo encontrado ao longo do curso.

Aos meus pais, Marlene e Antonio, e ao meu irmão, Felipe, por acreditarem nos meus estudos e dedicação, além do apoio incondicional desde sempre.

Ao Prof. PhD. Luis Glauber Rodrigues, pelos ensinamentos ao longo do curso e pela excelente orientação, não só neste trabalho, como em outros projetos acadêmicos.

Aos professores participantes da banca examinadora Pedro Felipe Gadelha Silvino e Felipe Xavier Feitosa pelo tempo, pelas valiosas colaborações e sugestões.

Aos meus amigos Gabriel, Aquécio, Paulo Matheus, Lucas, Clecio, Levy, Livia, Erika, Ana Alice, Giovanna e Alice por deixarem os períodos da faculdade mais leves e descontraídos.

À Escola Piloto de Engenharia de Petróleo, em especial ao amigo João Victor, a professora Narelle e ao Thiago Barbosa pelo incentivo ao desenvolvimento do projeto.

A todas as pessoas que passaram em minha vida no período de graduação e, de alguma forma, ajudaram-me ou colaboraram para concluir esta etapa.

RESUMO

A aplicação de algoritmos de machine learning tem se destacado como uma ferramenta eficaz e cada vez mais empregada em diversas áreas industriais, trazendo avanços significativos na análise e interpretação de dados complexos e executando tarefas autônomas de maneira a simular a capacidade e raciocínio humanos. Na indústria petrolífera, estas técnicas que utilizam o aprendizado de máquina têm desempenhado um papel muito importante no estudo de reservatórios, sobretudo por conta das incertezas associadas, o que acaba por possibilitar uma melhor compreensão das características dos poços e aprimorar as operações de exploração e produção nos campos. Nesse sentido, o presente trabalho visa à caracterização de eletrofácies em dois poços de petróleo localizados na parte emersa da Bacia Potiguar, no Campo de Jaçanã, por meio da aplicação do algoritmo de machine learning K-means para realizar a clusterização dos dados dos poços disponibilizados pela plataforma REATE, da ANP. A metodologia do projeto envolveu a seleção de um conjunto de perfis de poços, cujos dados foram submetidos a um processo de análise, seleção, tratamento e normalização para posterior utilização no aprendizado. Nesta etapa, os perfis de poços foram pré-processados para corrigir anomalias e padronizar as medições. Em seguida, aplicou-se o algoritmo K-means para realizar a clusterização dos dados, por meio da utilização de dados de um "poço-teste" e dos dados de cada "poço-objetivo", em dois casos diferentes, identificando padrões distintos que foram associados às diferentes eletrofácies. Os resultados obtidos foram promissores, apresentando uma alta correspondência com as condições reais observadas nos poços estudados. A análise de comparação de imagens por pixels feitas por uma das bibliotecas de programação aplicadas no modelo, utilizada para validar a precisão do método, revelou uma semelhança de 97,22 % e 98,13 % entre as eletrofácies reais e as previstas pelo modelo nos dois exemplos. Estes achados demonstram a eficácia do projeto e do uso do K-means na caracterização de eletrofácies, oferecendo uma ferramenta robusta para a análise geológica e contribuindo para a otimização das operações de exploração.

Palavras-chave: algoritmos; eletrofácies; perfis; precisão; aprendizado.

ABSTRACT

The application of machine learning algorithms has emerged as an effective and increasingly employed tool across various industrial sectors, bringing significant advancements in the analysis and interpretation of complex data and performing autonomous tasks in a way that simulates human capacity and reasoning. In the oil industry, these machine learning techniques have played a crucial role in reservoir studies, especially due to the uncertainties involved, allowing for a better understanding of well characteristics and improving exploration and production operations in the fields. In this context, the present work aims to characterize electrofacies in two oil wells located in the onshore portion of the Potiguar Basin, in the Jaçanã Field, through the application of the K-means machine learning algorithm to cluster the well data made available by the REATE platform of ANP. The project's methodology involved selecting a set of well logs, which were subjected to a process of analysis, selection, treatment, and normalization for later use in learning. At this stage, the well logs were pre-processed to correct anomalies and standardize measurements. Then, the K-means algorithm was applied to cluster the data using the data from a "test well" and data from each "target well" in two different cases, identifying distinct patterns that were associated with different electrofacies. The results obtained were promising, showing a high correspondence with the actual conditions observed in the studied wells. The pixel-by-pixel image comparison analysis performed by one of the programming libraries applied to the model, used to validate the method's accuracy, revealed a similarity of 97.22% and 98.13% between the actual and predicted electrofacies in the two examples. These findings demonstrate the effectiveness of the project and the use of K-means in electrofacies characterization, offering a robust tool for geological analysis and contributing to the optimization of exploration operations.

Keywords: algorithms; electrofacies; logs; accuracy; learning.

LISTA DE FIGURAS

Figura 1	– Mapa de localização da Bacia Potiguar	16
Figura 2	– Mapa estrutural da porção emersa da Bacia Potiguar	18
Figura 3	– Diagrama estratigráfico da Bacia Potiguar	19
Figura 4	– Seção transversal do Perfil esquemático da parte emersa da Bacia Potiguar ..	20
Figura 5	– Mapa de localização do Campo de Jaçanã	23
Figura 6	– Exemplificação de conjunto de dados para Aprendizado Semi-Supervisionado	26
Figura 7	– Ilustração simples do processo de Clusterização	27
Figura 8	– Exemplo da aplicação do K-Means sobre um conjunto de dados iniciais (a), mostrando algumas etapas do processo, como a definição dos clusters iniciais e seus respectivos centroides (b) e as seguintes iterações, segunda (c), terceira (d) e última (e), até convergir adequadamente	29
Figura 9	– Representação gráfica do número correto de clusters pelo Método do Cotovelo	30
Figura 10	– Representação simples da etapa de medição de perfis e propriedades do poço	31
Figura 11	– Representação da medida do perfil raio gama para diferentes litologias	33
Figura 12	– Esquematização básica de um dos tipos de equipamentos de medição do cáliper	34
Figura 13	– Representação do Perfil Cáliper de acordo com a formação	35
Figura 14	– Resposta do Perfil Neutrão para algumas das principais litologias	36
Figura 15	– Esquematização da ferramenta de medição do perfil de densidade	38
Figura 16	– Resposta do perfil de densidade para algumas litologias mais comuns	39
Figura 17	– Princípio de medição do tempo de trânsito (T.T.) em intervalo sônico usando um único transmissor (T) e dois receptores (R_1 e R_2)	40
Figura 18	– Valores de tempo de trânsito para diferentes litologias	41

Figura 19 – Valores de resistividade para diferentes litologias	42
Figura 20 – Coordenadas Geográficas dos poços estudados	45
Figura 21 – Comparação entre os dados dos perfis Cáliper (a) e DT (b) dos arquivos primário e secundário (1 e 2) do Poço 4-BRSA-924-RN	47
Figura 22 – Comparação entre os dados dos perfis NPHI (a) e RHOB (b) dos arquivos primário e secundário (1 e 2) do Poço 4-BRSA-423D-RN	48
Figura 23 – Comparação entre os dados dos perfis HMRS - Resistividade (a) e GR (b) dos arquivos primário e secundário (1 e 2) do Poço 4-BRSA-937D-RN	49
Figura 24 – Perfis do intervalo de estudo do Poço 4-BRSA-924-RN	51
Figura 25 – Perfis do intervalo de estudo do Poço 4-BRSA-937D-RN	52
Figura 26 – Perfis do intervalo de estudo do Poço 4-BRSA-423D-RN	52
Figura 27 – Representação da distribuição para a Regra dos 3 Sigmas	53
Figura 28 – Identificação dos <i>outliers</i> nos dados do Perfil de Cáliper do Poço de Treino ..	55
Figura 29 – Fluxograma de trabalho para obtenção do valor de semelhança entre a imagem gerada pelo modelo e a imagem real do poço	60
Figura 30 – Ilustração da correção da imagem retirada do perfil composto para o padrão da imagem do modelo (Poço 4-BRSA-423D-RN: 1450 - 1550 metros)	61
Figura 31 – Ilustração da correção da imagem retirada do perfil composto para o padrão da imagem do modelo (Poço 4-BRSA-937D-RN: 2200 - 2300 metros)	61
Figura 32 – Resultado da Predição do Perfil de Eletrofácies do Poço 4-BRSA-423D-RN (1450 - 1550 metros)	63
Figura 33 – Resultado da Predição do Perfil de Eletrofácies do Poço 4-BRSA-937D-RN (2200 - 2300 metros)	64
Figura 34 – Comparação entre os Perfis de Eletrofácies do Perfil Composto e do Aprendizado de Máquina do Poço 4-BRSA-423D-RN (1450 - 1550 metros)	65
Figura 35 – Comparação entre os Perfis de Eletrofácies do Perfil Composto e do Aprendizado de Máquina do Poço 4-BRSA-937D-RN (2200 – 2300 metros)	66

LISTA DE GRÁFICOS

Gráfico 1 – Histórico de Produção de Petróleo e Gás Natural do Campo de Jaçanã	23
Gráfico 2 – Histograma e Curva de Distribuição Normal dos dados definidos do Perfil Cáliper do poço de treino	54
Gráfico 3 – Curva da soma dos erros quadrados de acordo com o número de clusters para o caso com o poço 4-BRSA-423D-RN	58
Gráfico 4 – Curva da soma dos erros quadrados de acordo com o número de clusters para o caso com o poço 4-BRSA-937D-RN	59

LISTA DE TABELAS

Tabela 1 – Proporção de Reservas e Recursos de petróleo e gás natural na Bacia Potiguar	17
Tabela 2 – Quantidade de Arquivos Lógicos de cada poço	46
Tabela 3 – Identificação dos Intervalos de estudo de cada poço e sua identificação no Aprendizado de Máquina	50
Tabela 4 – Resultado numérico da comparação entre as imagens reais e preditas	66

LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
API	Grau API – Escala para medição da densidade de Fluidos
bb1	Barril de óleo
ENE	Leste–nordeste ou Lés-nordeste
km	Quilômetro
km ²	Quilômetro quadrado
mD	Millidarcy
m ³	Metro cúbico
NE	Nordeste
SW	Sudoeste
WSW	Oés-sudoeste

SUMÁRIO

1	INTRODUÇÃO	14
2	ÁREA DE ESTUDO	16
2.1	Bacia Potiguar	16
2.1.1	<i>Arcabouço Estrutural e Embasamento</i>	17
2.1.2	<i>Arcabouço Estratigráfico</i>	18
2.1.3	<i>Sistemas Petrolíferos</i>	20
2.2	Campo de Jaçanã	22
3	FUNDAMENTAÇÃO TEÓRICA	24
3.1	Machine Learning	24
3.1.1	<i>Tipos de Aprendizado de Máquina</i>	24
3.1.2	<i>Clusterização</i>	27
3.1.3	<i>Algoritmo K-Means</i>	28
3.2	Perfilagem de Poços	31
3.2.1	<i>Perfil Gamma-Ray (GR)</i>	32
3.2.2	<i>Perfil de Cáliper</i>	33
3.2.3	<i>Perfil Neutrão</i>	35
3.2.4	<i>Perfil Densidade (RHOB)</i>	37
3.2.5	<i>Perfil Sônico (DT)</i>	39
3.2.6	<i>Perfil de Resistividade</i>	41
4	MATERIAIS E MÉTODOS	43
4.1	Base de dados	43
4.2	Processamento dos dados	46
4.3	Aprendizado Não-Supervisionado	55
5	RESULTADOS E DISCUSSÕES	62
5.1	Resultados das predições	62
5.2	Comparações entre as eletrofácies	64

6	CONCLUSÃO	67
	REFERÊNCIAS	

1 INTRODUÇÃO

A caracterização de reservatórios exige a compreensão da complexidade do subsolo e a avaliação dos riscos envolvidos, utilizando uma variedade de abordagens e técnicas, como a física de rochas, inversão sísmica, modelagem geológica e geoestatística, além de ferramentas baseadas em inteligência artificial (Spikes *et al.*, 2017). Entre as técnicas utilizadas, a caracterização de eletrófacies se destaca como uma ferramenta essencial para entender as propriedades geológicas dos reservatórios. As eletrófacies representam diferentes zonas dentro do reservatório, identificadas com base em perfis de poços, permitindo uma compreensão detalhada das camadas subterrâneas e a capacidade do reservatório de armazenar e produzir hidrocarbonetos (Zahmatkesh *et al.*, 2021).

Apesar da importância da caracterização de eletrófacies, o processo de identificação e análise dessas zonas muitas vezes enfrenta desafios significativos, como a variabilidade geológica, dados limitados e/ou com elevado custo para obtenção e a complexidade dos reservatórios. Diante disso, tem-se a problemática: como é possível obter a caracterização de eletrófacies de um poço de petróleo de maneira eficiente e prática sem a necessidade de equipamentos de medição ou técnicas e softwares específicos de identificação?

Compreender e identificar as eletrófacies é um dos aspectos mais importantes para analisar e explorar áreas dentro de campos de petróleo com potencial para produção e que possuem adequada viabilidade econômica. Além disso, a aplicação de técnicas por algoritmos de aprendizado de máquina pode trazer inovações para a área de perfilagem, além de possibilitar análises precisas e detalhadas de dados geológicos.

O presente trabalho tem como objetivo analisar o uso de técnicas de aprendizado de máquina, especificamente o algoritmo de clusterização K-means, na caracterização de eletrófacies em poços de petróleo. Busca-se investigar como essa abordagem pode tornar mais eficiente e acessível e, também, melhorar o processo de identificação de eletrófacies, por meio da comparação dos resultados preditos com as condições reais observadas em reservatórios de petróleo. Para isso, será realizada uma análise detalhada dos perfis de poços, incluindo o tratamento e normalização de dados, além de uma comparação pixel a pixel entre as eletrófacies reais e preditas pelo modelo.

Este projeto está organizado de modo que, no tópico 2, as características e informações da área de estudo serão expostos e destacados. No tópico 3, será apresentada a fundamentação teórica sobre a perfilagem de poços e alguns conceitos e técnicas de aprendizado de máquina. No capítulo 4, a metodologia e os materiais utilizados ao longo do

desenvolvimento e implementação do projeto serão apresentados, assim como a aplicação do modelo. Finalmente, os tópicos 5 e 6 trarão os resultados e conclusões do estudo, respectivamente. As identificações das eletrofácies dos poços realizadas pela máquina serão apresentadas, além da demonstração da semelhança para com as distribuições reais de eletrofácies dos poços, tanto visual como quantitativamente.

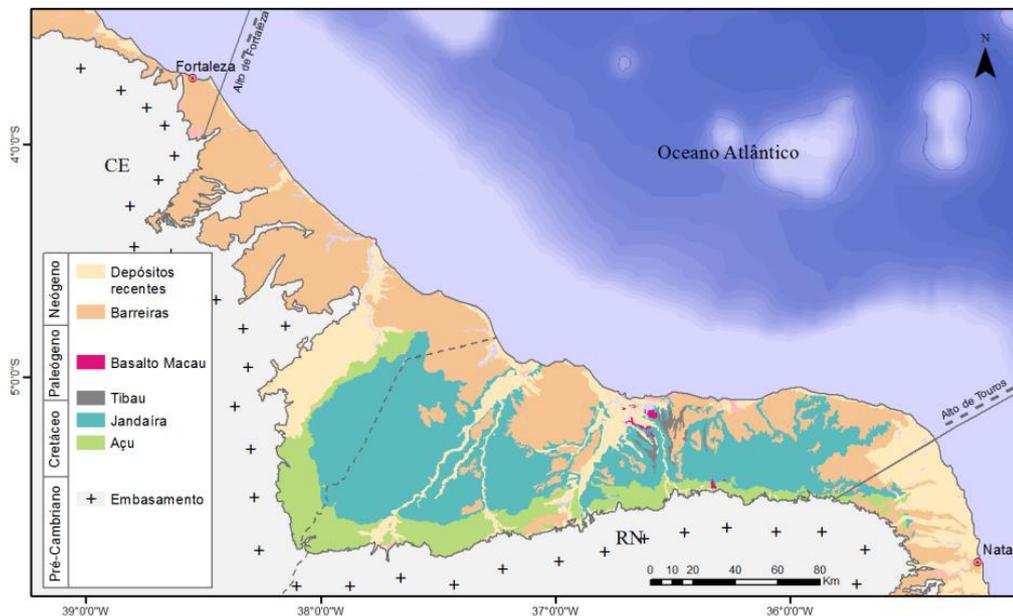
2 ÁREA DE ESTUDO

A região dos poços objetos de estudo compreende a parte emersa da Bacia Potiguar, em específico o Campo de Jaçanã. As características e aspectos geológicos de ambos serão abordados a seguir.

2.1 Bacia Potiguar

A Bacia Potiguar está localizada na porção mais oriental da região nordeste do Brasil, abrangendo os estados do Rio Grande do Norte e do Ceará. De acordo com Portella e Fabianovicz (2017), sua área sedimentar em terra firme é de cerca de 26.700 km², enquanto a porção submersa totaliza aproximadamente 195.400 km². Do ponto de vista geológico, faz fronteira a leste com a Bacia de Pernambuco-Paraíba através do Alto de Touros, a noroeste com a Bacia do Ceará pelo Alto de Fortaleza, e ao sul, está limitada por rochas do embasamento cristalino, como mostra a figura 1.

Figura 1 - Mapa de localização da Bacia Potiguar



Fonte: Portella e Fabianovicz (2017).

A exploração de hidrocarbonetos na Bacia Potiguar começou em 1949 com mapeamentos conduzidos pelo Conselho Nacional do Petróleo (CNP) (Portella; Fabianovicz, 2017). Em 1956, após a criação da Petrobras, foram perfurados dois poços estratigráficos. Na

década de 70, durante a primeira crise do petróleo, as atividades exploratórias foram retomadas (Femenick, 2014), resultando na descoberta dos campos de Ubarana (1973), Agulha (1975) e Mossoró (1979). Essas descobertas levaram a um aumento na exploração da bacia, o que resultou em importantes descobertas de campos entre 1980 e 1985, como Fazenda Belém, Alto do Rodrigues, Estreito, Fazenda Pocinho, Guamaré, Serraria, Lorena, Upanema e Canto do Amaro. Entre 1980 e 1990, a bacia teve seu maior investimento exploratório, com a perfuração de 675 poços, representando cerca de 50% do total de poços exploratórios perfurados na bacia.

Após a criação da Agência Nacional do Petróleo (ANP) em 1997, a Bacia Potiguar foi alvo de novos levantamentos e licitações, resultando em novas descobertas e a incorporação de pequenas acumulações pela Petrobras.

Até hoje, a exploração na Bacia Potiguar envolveu levantamentos gravimétricos, magnetométricos, dados sísmicos e a perfuração de 1.353 poços exploratórios (Travassos *et al.*, 2021). A porção terrestre tem 341,2 km de extensão e 19,03 km² de área cobertos por dados sísmicos 2D e 3D, respectivamente. A porção marítima possui 57.917,59 km de dados sísmicos 2D e 23.246,01 km² de dados 3D. Atualmente, a bacia conta com 85 campos em produção e desenvolvimento, com estimativas de reservas e recursos mostrados na tabela 1, de acordo com o Boletim Anual de Reservas atualizado em 1 de abril de 2024.

Tabela 1 – Proporção de Reservas e Recursos de petróleo e gás natural na Bacia Potiguar

	Petróleo (Milhões de m ³)				Gás Natural (Milhões de m ³)			
	Reservas 1P	Reservas 2P	Reservas 3P	Recursos Contingentes	Reservas 1P	Reservas 2P	Reservas 3P	Recursos Contingentes
Terra	31,3364	36,7420	38,3159	1,2035	3.883,00	4.158,00	4.238,00	47,17
Mar	0,6487	1,5469	1,5883	0,0668	525,00	886,00	970,00	344,05
Total	31,9851	38,2889	39,9042	1,2703	4.408,00	5.044,00	5.208,00	391,22

Fonte: Adaptado de SIGEP/SDP/ANP (2023).

Sabendo-se que:

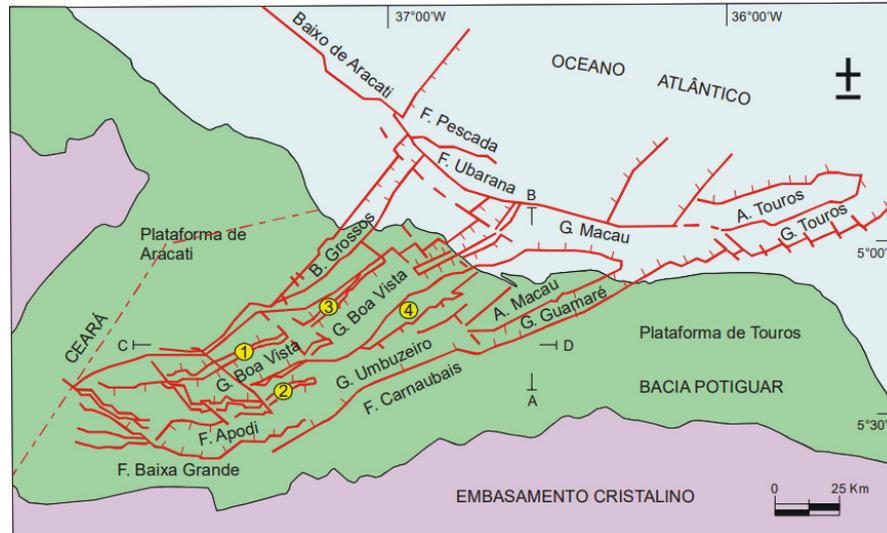
- Provadas (1P);
- Provadas e Prováveis (2P);
- Provadas, Prováveis e Possíveis (3P).

2.1.1 Arcabouço Estrutural e Embasamento

A Bacia Potiguar é parte integrante do Sistema de Riftes Cretáceos no Nordeste do Brasil. Sua formação e desenvolvimento está ligada ao processo de estiramento da crosta

terrestre, que levou à fragmentação do Supercontinente Gondwana durante o Mesozoico. Este processo culminou na separação das placas Sul-Americana e Africana e na formação do Oceano Atlântico (Bertani *et al.*, 1990).

Figura 2 - Mapa estrutural da porção emersa da Bacia Potiguar



Arcabouço estrutural da Bacia Potiguar. (1) Alto dos Canudos, (2) Alto de Quixabá, (3) Alto de Mossoró e (4) Alto da Serra do Carmo. A - alto, B - baixo, G - gráben, F - falha, — — — indicação aproximada dos perfis esquemáticos.

Fonte: Angelim *et al.* (2006)

A seção emersa do Rifte Potiguar estende-se na direção ENE-WSW, sendo delimitada pelas falhas de Carnaubais a leste e de Areia Branca a oeste. Esta estrutura é composta por três unidades principais: grábens, altos internos e plataformas do embasamento, conforme ilustrado na figura 2 (Bertani *et al.*, 1990). O embasamento da bacia é constituído por rochas pré-cambrianas da Província Borborema, formada pela fusão de blocos crustais arcaicos e proterozoicos, que são compostos por seqüências litoestratigráficas de rochas ígneas e metamórficas (Pedrosa Jr. *et al.*, 2010).

2.1.2 Arcabouço Estratigráfico

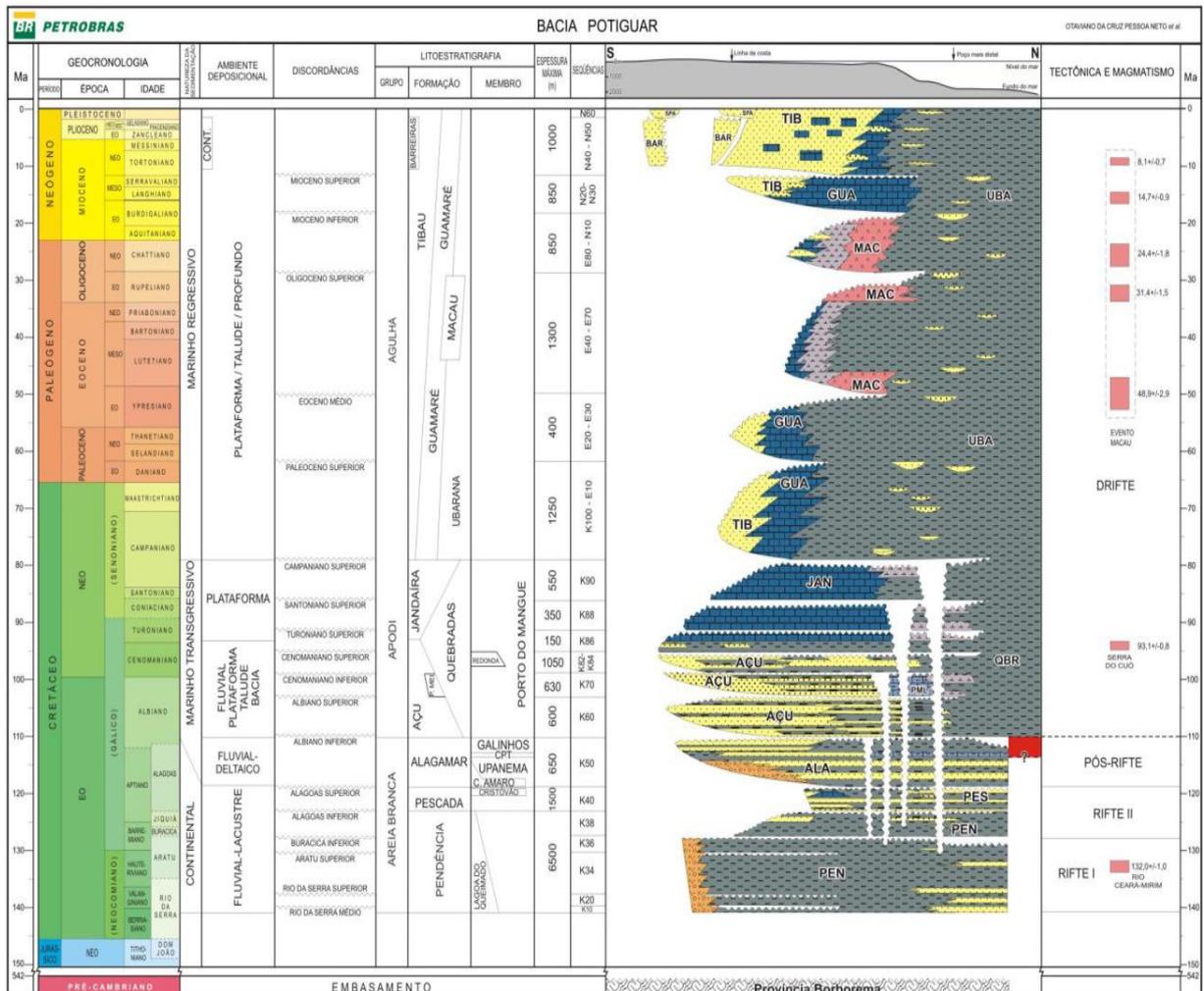
A subdivisão da Bacia Potiguar em 3 Superseqüências segue a evolução tecnoestratigráfica proposta por Pessoa Neto *et al.* (2007). Segundo ele, o preenchimento da bacia ocorreu conforme as diferentes fases de sua evolução tectônica: Rifte I, Rifte II, Pós-Rifte e Drifte, como mostra a carta estratigráfica na figura 3.

Na primeira superseqüência (Rifte I e II), no início, durante o Rifte I, entre o Neoberriasiano e o Eobarremiano, predominou-se um regime tectônico de estiramento crustal,

com o desenvolvimento de falhas normais de grande rejeito que definem hemi-grábens assimétricos e altos internos na direção geral NE-SW. Nesses grábens, foram depositados deltas lacustres prográdantes, associados a folhelhos ricos em matéria orgânica intercalados com turbiditos, agrupados na Formação Pendência (Bertani *et al.*, 1990).

Na fase seguinte, Rifte II, durante o Neobarremiano/Eo-Aptiano, instaurou-se um regime transcorrente/transformante ao longo da futura margem equatorial, provocando uma mudança na cinemática do rifte e deslocando o eixo de rifteamento para a parte submersa da Bacia Potiguar, enquanto a porção emersa possuiu por soerguimento e erosão. Nesse contexto, foram depositados arenitos grossos e pelitos na parte superior da Formação Pendência, ainda em um ambiente flúvio-deltaico-lacustre. Também foram formados leques aluviais e sistemas fluviais de alta energia da Formação Pescada, em sistemas deposicionais continentais (Travassos *et al.*, 2021).

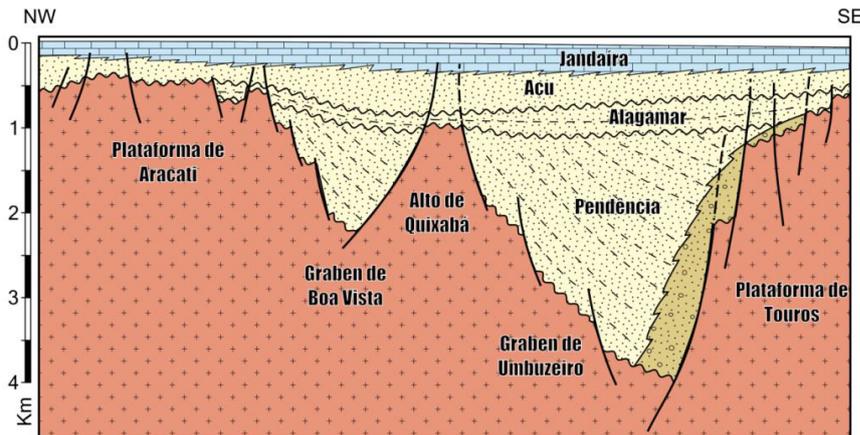
Figura 3 - Diagrama estratigráfico da Bacia Potiguar



Fonte: Pessoa Neto *et al.* (2007)

A bacia, na sua porção terrestre, apresenta uma assimetria evidente, com uma borda falhada no Sudeste (Sistema de Falhas de Carnaubais). Essa estrutura controla a presença de fanglomerados e um grande depocentro da Formação Pendência (Supersequência Rifte), conforme mostrado na figura 4.

Figura 4 – Seção transversal do Perfil esquemático da parte emersa da Bacia Potiguar



Fonte: Bertani *et al.* (1990)

Após a primeira supersequência, as fases seguintes tratam-se da transição e desenvolvimento tectônico em ambientes submersos. A Supersequência Pós-Rifte é marcada por um regime tectônico relativamente calmo e pela transição gradual de sistemas deposicionais continentais para marinhos. Seus sedimentos incluem os fandeltaicos e flúvio-deltaicos da Formação Alagamar (Membros Canto do Amaro e Upanema) e os folhelhos transitórios (Membro Galinhos), com um evento de máxima transgressão, conhecido como Camada Ponta do Tubarão (CPT).

Já a Supersequência Drifte, depositada entre o Albiano e o período Recente, durante a fase de subsidência termal, consiste em uma sequência flúvio-marinha transgressiva coberta por uma sequência clástica e carbonática regressiva, representada pelos sedimentos siliciclásticos das formações Açu (proximal) e Quebradas (distal) e pela plataforma carbonática da Formação Ponta do Mel (Portella; Fabianovicz, 2017).

2.1.3 Sistemas Petrolíferos

A Bacia de Potiguar, de acordo com a classificação de Morelato e Fabianovicz (2015), possui 3 sistemas petrolíferos atuantes: Pendência-Pendência (!), Pendência-Alagamar (!) e Alagamar-Açu (!).

As primeiras rochas geradoras comprovadas na Bacia de Potiguar são os folhelhos lacustres Neocomianos da Formação Pendência. A integração de dados geoquímicos e geotérmicos indica que a Formação Pendência atingiu maturidade a partir do Santoniano, sendo submetida a condições de temperatura bastante elevadas, o que a tornou uma potencial geradora de petróleo (Morelato; Fabianovicz, 2015). As rochas geradoras da Formação Pendência são compostas por folhelhos negros ricos em matéria orgânica, depositados em ambientes lacustres de água doce, flúvio-deltaicos e fandeltaicos (Pessoa Neto *et al.*, 2007).

Além disso, a geração de hidrocarbonetos na bacia também provém dos folhelhos e margas da Formação Alagamar da idade Aptiana (Portella; Fabianovicz, 2017). Os folhelhos da formação são lacustres cinza escuros a negros provenientes de ambiente flúvio-deltaico e marinho restrito (Pessoa Neto *et al.*, 2007).

Os hidrocarbonetos gerados pela Formação Alagamar exibem diferentes níveis de evolução térmica, indicando que se originaram a partir de rochas geradoras com variados graus de maturação. De acordo com estudos geoquímicos, na parte terrestre da bacia, as unidades geradoras da Formação Alagamar são imaturas. Contudo, na porção marinha, esta formação apresenta alta evolução geotérmica, alcançando o pico de maturação no Mioceno. Os óleos gerados na plataforma continental migraram lateralmente em direção à porção terrestre da bacia, seguindo os alinhamentos estruturais pré-existentes (Morelato; Fabianovicz, 2015).

Os principais reservatórios da Bacia de Potiguar são formados por arenitos flúvio-deltaicos e turbidíticos da Formação Pendência (Neocomiano), arenitos deltaicos da Formação Alagamar (Aptiano) e arenitos flúvio-eólicos da Formação Açú (Albiano-Cenomaniano), de acordo com Portella e Fabianovicz (2017). Os arenitos da Formação Pendência apresentam porosidades entre 18% e 24% e contêm óleo de 31 a 36° API, o que indica leve e boa qualidade. Os arenitos deltaicos da Formação Alagamar possuem uma porosidade de aproximadamente 12% e contêm óleo com densidades variando de 16 a 34° API, variando de óleo pesado a leve. Finalmente, os arenitos da Formação Açú têm porosidades de 19% a 30% e contêm óleo com densidades de 14 a 34° API, também variando de pesado a leve (Barbosa, 2013).

Com relação aos selos dos reservatórios da bacia, os selos das Formações Açú e Alagamar são proporcionados por níveis pelíticos, que são camadas de rochas sedimentares finas compostas por argila e silte, caracterizadas por baixa permeabilidade. Já os selos das acumulações da Formação Pendência, são formados por folhelhos associados estratigraficamente aos reservatórios e pelo fechamento de armadilhas devido a movimentações em falhas, que justapõem rochas reservatório com folhelhos de blocos baixos ou sequências superiores (Morelato; Fabianovicz, 2015).

O traqueamento na Formação Pendência possui uma grande variedade de cenários, incluindo blocos falhados rotacionados relacionados às falhas de borda dos grabens onshore e estruturas associadas a falhas antitéticas, gravitacionais e de transferência. Embora essas trapas tenham um forte componente estrutural, variações laterais de fácies e feições paleogeomórficas também podem contribuir para o aprisionamento do petróleo (Travassos *et al.*, 2021).

Na Formação Açu, as trapas podem ser estruturais, mistas ou paleogeomórficas. As trapas estruturais são formadas pela combinação do mergulho regional das camadas com o arrasto de falhas normais reativadas nas ombreiras dos grabens onshore. As trapas mistas têm um componente estrutural e importantes variações laterais de fácies. Por fim, as trapas paleogeomórficas são definidas por leques aluviais nos flancos de paleoaltos do embasamento (Travassos *et al.*, 2021).

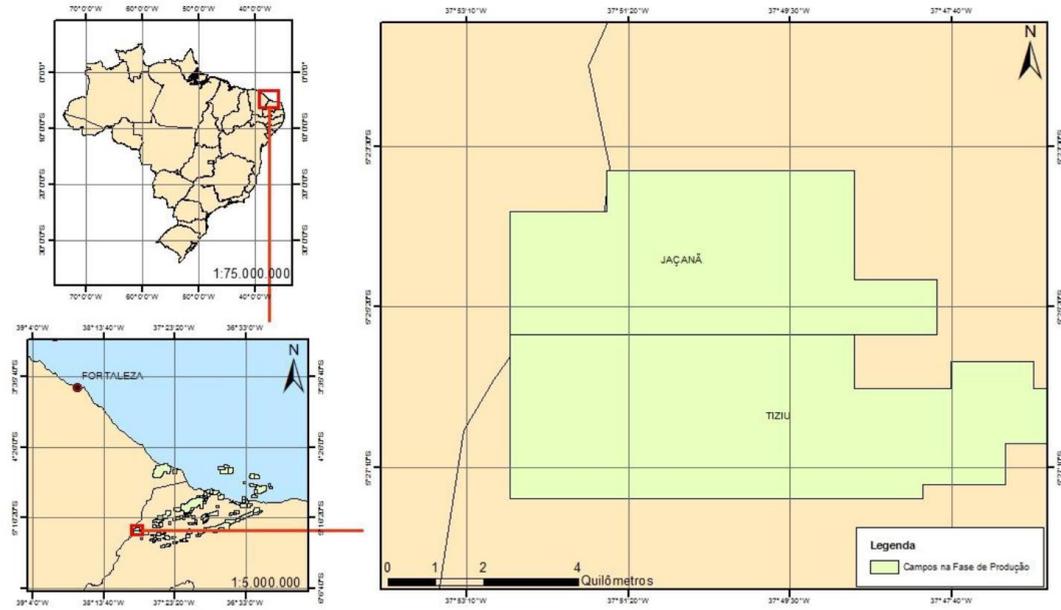
Na Formação Alagamar, as trapas são principalmente estruturas homoclinais truncadas por cânions submarinos e estruturas dômicas associadas a falhas transcorrentes. Além disso, na Formação Ubarana, podem ocorrer trapas estratigráficas, definidas pela intercalação de arenitos turbidíticos com folhelhos intraformacionais (INSIGHT, 2005).

2.2 Campo de Jaçanã

O Campo de Jaçanã está situado na porção emersa da Bacia Potiguar, nos municípios de Apodi e Governador Dix-sept Rosado, com uma área de desenvolvimento de 25,19 km², conforme ilustrado na figura 5. Localizado a aproximadamente 310 km a oeste de Natal, a capital do Estado do Rio Grande do Norte, o campo teve seus maiores picos de produção mensal de petróleo, fluido principal produzido, nos quatro primeiros anos, chegando a 910,7 bbl/dia em março de 2010, de acordo com os painéis dinâmicos da ANP.

Em se tratando da geologia da área do campo e das fontes de hidrocarbonetos, os principais reservatórios de Jaçanã são compostos por arenitos flúvio-deltaicos do Cretáceo inferior, pertencentes à Formação Pendência. Esses arenitos possuem uma porosidade média que varia entre 8% e 20% e permeabilidades que vão de 0,1 a 670 mD, estando saturados com óleo de 33 a 36° API, considerado leve e de qualidade. O principal mecanismo primário de produção é o influxo de água, com um aquífero de fundo atuando nas zonas mais rasas, e gás em solução nas zonas mais profundas. Não há injeção de fluidos no campo para fins de recuperação aprimorada (ANP, 2016).

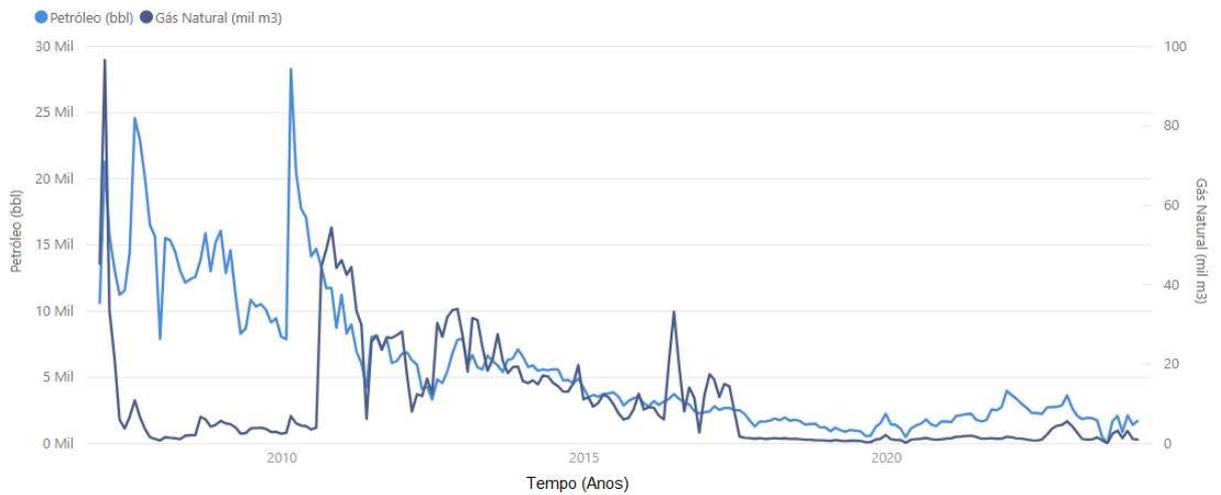
Figura 5 – Mapa de localização do Campo de Jaçanã



Fonte: ANP, 2016

A descoberta do campo em julho de 2005 marcou o início de estudos detalhados e análises minuciosas para avaliar o potencial do campo, bem como a viabilidade econômica da exploração, que culminaram na declaração de comercialidade em dezembro de 2006 (ANP, 2016). Logo após, deu-se início a produção no Campo de Jaçanã, em janeiro de 2007, apontando anos iniciais promissores, com bons volumes de produção de óleo até o final de 2010, como mostra o gráfico 1, seguidos de um gradativo declínio ao longo dos anos seguintes.

Gráfico 1 – Histórico de Produção de Petróleo e Gás Natural do Campo de Jaçanã



Fonte: ANP, 2024

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Machine Learning

Machine Learning, em português, aprendizagem de máquina, representa uma área em constante evolução de algoritmos computacionais desenvolvidos para emular a inteligência humana, aprendendo a partir do ambiente ao seu redor (El Naqa; Murphy, 2015). As aplicações deste ramo da computação podem ser observadas em vários setores da sociedade, por exemplo: na saúde, contribuindo em diagnósticos mais precisos, nas finanças, em previsões de mercado e preços, assim como na indústria do petróleo, em previsões de reservas e declínios de produção e na análise de dados sísmicos e geológicos.

O aprendizado de máquina é definido como a utilização de vários algoritmos para ensinar computadores a identificar padrões em dados a serem utilizados para previsões e previsões futuras, ou como uma verificação de qualidade para otimização de desempenho. Ele proporciona aos computadores a capacidade de aprender sem serem explicitamente programados (BELYADI; HAGHIGHAT, 2021).

Ao implementar o aprendizado de máquina, é essencial que se compreenda o contexto, o conhecimento prévio e os objetivos do problema a ser abordado. Mesmo com dados disponíveis, é fundamental que estes sejam selecionados, limpos, integrados e pré-processados (Chapman *et al.*, 2000). A seleção de características desempenha um papel crucial, pois influencia diretamente no sucesso ou fracasso do sistema de aprendizado de máquina. Também é importante considerar a sensibilidade ou proteção de certas características. Em seguida, deve-se tomar decisões sobre o modelo de aprendizado ou tipo de aprendizado de máquina a ser utilizado, treinando-o com dados e ajustando os parâmetros para alcançar um desempenho otimizado. Após a construção do sistema, a avaliação e interpretação dos resultados tornam-se determinantes para conclusões finais. A fase final consiste na implementação prática, seguida por ajustes contínuos no sistema de aprendizado de máquina para aprimorar suas variáveis (GEETHA; SENDHILKUMAR, 2023).

3.1.1 Tipos de Aprendizado de Máquina

Com relação as formas de aprendizagem, Geetha e Sendhilkumar (2023) classificam os tipos algoritmos em aprendizado de máquina clássico, aprendizado por esforço e aqueles baseados em redes neurais e aprendizado profundo. A categorização dos algoritmos

clássicos dá-se pela representação dos dois principais tipos de aprendizado, o supervisionado e o não supervisionado. O aprendizado semi-supervisionado também está incluindo na categoria clássica, visto que engloba os outros dois.

Aprendizado Supervisionado

No aprendizado supervisionado, a máquina é treinada para aprender uma função que relaciona entradas a saídas com base em exemplos rotulados. Essa abordagem é utilizada quando existem metas específicas a serem atingidas a partir de um conjunto conhecido de entradas. As tarefas mais comuns nesse contexto são “classificação”, que envolve a separação dos dados, e “regressão”, que ajusta os dados. Por exemplo, prever a classe ou sentimento de um texto, como em um tweet ou análise de produto, é um exemplo de aprendizado supervisionado. Isso implica que a máquina utiliza dados de treinamento com rótulos conhecidos para aprender a relação entre as entradas e as saídas desejadas (SARKER, 2021).

Aprendizado Não Supervisionado

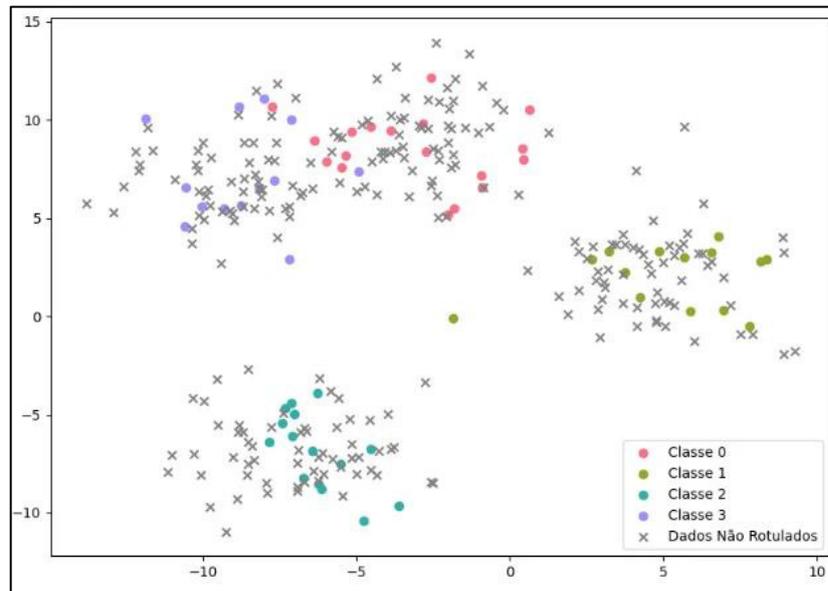
Neste tipo de aprendizado de máquina, ao contrário do supervisionado, os dados disponíveis são apenas de entrada, sem rótulos associados, ou seja, procura-se realizar a identificação de padrões neste conjunto de características. Além disso, devido à falta de saídas associadas, não é possível fazer comparações entre as previsões fornecidas pelo modelo e dados reais do estudo (BELYADI; HAGHIGHAT, 2021). Um exemplo de aprendizado não supervisionado é a aplicação de um algoritmo de agrupamento em um conjunto de dados com informações sobre clientes de uma loja (Entrada). O algoritmo, então, agrupa os clientes em um determinado número de grupos com características de comportamentos de compras semelhantes (Saída), como aqueles que fazem compras frequentes com valores baixos e outro grupo com clientes que voltam com menos frequência, mas cujas compras têm valores mais altos.

Aprendizado Semi-Supervisionado

Neste caso, a combinação dos outros dois tipos de aprendizagem clássica ocorre quando os dados disponíveis são rotulados e não rotulados. Entretanto, na maioria das vezes, esse tipo de configuração acontece em casos que os dados rotulados são escassos devido a

vários fatores, como custo e dificuldade de obtenção, enquanto os dados não rotulados são bem mais numerosos. A proposta do aprendizado semi-supervisionado é, então, empregar exemplos rotulados para adquirir informações iniciais sobre o problema e, posteriormente, utilizar essas informações para orientar o processo de aprendizado a partir de exemplos não rotulados (Bruce, 2001). A figura 6 mostra um exemplo simples de situação com alguns dados rotulados e o restante não rotulado, o que torna o aprendizado semi-supervisionado útil.

Figura 6 - Exemplificação de conjunto de dados para Aprendizado Semi-Supervisionado



Fonte: Almeida, 2023

Aprendizado por Esforço

É um algoritmo de aprendizagem que permite as máquinas interagirem com um ambiente ou contexto específico e melhorarem seu desempenho automaticamente. Isso ocorre devido aos feedbacks que a máquina recebe da iteração com o ambiente na forma de recompensas ou penalidades, de acordo com as sucessivas tomadas de decisão. O objetivo do aprendizado é, então, utilizar essas informações e maximizar a recompensa final ao longo do tempo, já que realiza um processo de tentativa e erro (Mohammed *et al*, 2016). O aprendizado por esforço é uma ferramenta muito empregada nos campos da automação e otimização, além da robótica e jogos.

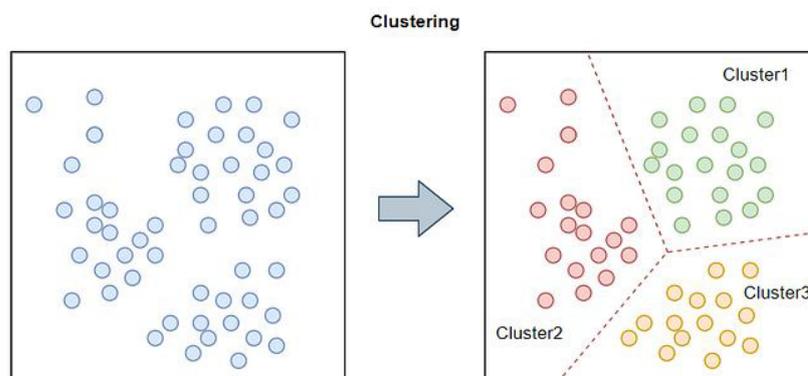
Aprendizado Profundo e Redes Neurais

O aprendizado profundo não é apenas um tipo de algoritmo, mas um subconjunto do aprendizado de máquina que se baseia no aprimoramento dos próprios métodos ao analisar os algoritmos computacionais. As redes neurais artificiais são a base deste aprendizado e se concentram em emular o processo de pensamento e aprendizado humano. Com os avanços tecnológicos e computacionais, o aumento da capacidade de processamento tem possibilitado a evolução das redes neurais, o que permite aos computadores a resolução e compreensão de problemas cada vez mais complexos. O aprendizado profundo possui vasta aplicabilidade em diferentes áreas, especialmente na resolução de casos de reconhecimento de padrões sem a necessidade de intervenção humana (GEETHA; SENDHILKUMAR, 2023).

3.1.2 Clusterização

A clusterização de dados é uma técnica bem comum de análise de dados no campo do aprendizado de máquina. É a principal abordagem utilizada de aprendizado não supervisionado e envolve a tarefa de agrupar um conjunto de dados em *clusters* (ou grupos), como mostra a figura 7, de modo que os itens dentro do mesmo grupo sejam mais semelhantes entre si do que itens de outros grupos (Raschka; Mirjalili, 2017). O objetivo principal da clusterização é identificar grupos naturais ou categorias, por meio da descoberta de estruturas ou padrões ocultos nos dados não rotulados disponibilizados (GEETHA; SENDHILKUMAR, 2023).

Figura 7 – Ilustração simples do processo de Clusterização



Fonte: Gomes, 2024

Devido à subjetividade da tarefa de agrupar os dados, vários métodos podem ser usados para realizar tal processo. Métodos Hierárquicos, Particionais, Baseados em Densidade

e Baseados em Grade são alguns dos exemplos, entretanto alguns desses métodos de clusterização incorporam conceitos de outras técnicas, o que acaba por dificultar, em alguns casos, a classificação de certo algoritmo em uma única categoria de método (Han; Kamber, 2001). O método *K-Means*, que será abordado no próximo tópico e utilizado neste trabalho, é uma das ferramentas mais utilizadas para agrupamento e é caracterizado como um Método Particional, em que os dados são divididos em um número pré-definido de *clusters*.

3.1.3 Algoritmo K-Means

O algoritmo *k-means* é um método de clusterização que agrupa dados não rotulados em *clusters*, com base na similaridade entre eles. O objetivo principal é dividir um conjunto de dados em *k* grupos (*clusters*) distintos, em que cada ponto de dados pertence ao *cluster* cujo centroide é mais próximo (GEETHA; SENDHILKUMAR, 2023).

O *k-means* busca minimizar a variância intra-cluster, ou seja, a soma dos quadrados das distâncias entre os pontos de dados e os centroides de seus *clusters* correspondentes. Ou seja, dado um conjunto de dados $X = \{x_1, x_2, \dots, x_n\}$, o objetivo é, então, dividir os dados em *k clusters* (C_1, C_2, \dots, C_k) de forma a minimizar a soma dos quadrados das distâncias euclidianas entre os dois pontos (OLIVEIRA, 2008). A função objetivo do *k-means* é definida como:

$$J(C_1, C_2, \dots, C_K, m_1, m_2, \dots, m_k) = \sum_{i=1}^k \sum_{j=1}^{|C_i|} \|x_{ij} - m_i\|^2 \quad (3.1)$$

Onde:

$|C_i|$: número de pontos no *cluster* C_i .

x_{ij} : *j*-ésimo ponto no *cluster* C_i .

m_i : centroide do *cluster* i , o qual é ponto médio de todos os pontos em C_i .

$\|x_{ij} - m_i\|^2$: distância euclidiana entre um ponto de um *cluster* e o centroide.

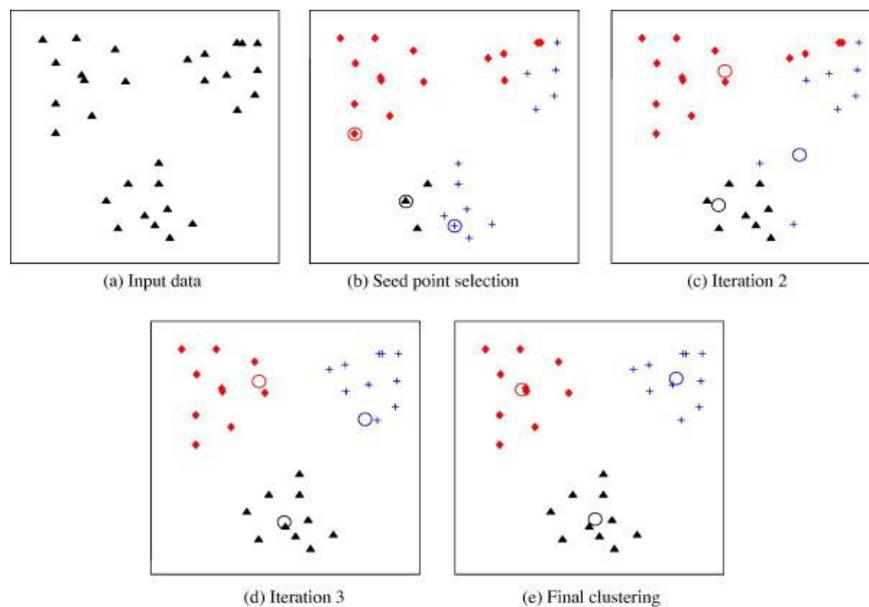
De acordo com (LACHI; ROCHA, 2005), o método do *K-Means* possui um conjunto de etapas seguidas pela ordem:

1. **Inicialização:** Definição de *k clusters* iniciais, bem como os seus centroides (pode ser determinado de forma aleatória ou usando algum método específico, como o Método do Cotovelo, o qual será abordado mais a frente).

2. **Atribuição:** Atribuição de cada ponto ao *cluster* cujo centroide é o mais próximo, de acordo com a distância euclidiana.
3. **Atualização do Centroide:** Os centroides são recalculados como os pontos médios dos pontos em cada *cluster*.
4. **Repetição:** Os passos 2 e 3 são retornados até que os centroides não mudem significativamente ou até que um número máximo de iterações seja atingido.

O algoritmo converge para uma solução, mas a solução final pode depender da escolha inicial dos centroides (LI; WU, 2012). Portanto, é comum executar o algoritmo várias vezes com diferentes inicializações e escolher a solução com a menor soma de quadrados. A figura 8 ilustra o processo de iteração do algoritmo até encontrar a solução mais viável de acordo com a otimização da função objetivo do *K-Means*.

Figura 8 - Exemplo da aplicação do *K-Means* sobre um conjunto de dados iniciais (a), mostrando algumas etapas do processo, como a definição dos *clusters* iniciais e seus respectivos centroides (b) e as seguintes iterações, segunda (c), terceira (d) e última (e), até convergir adequadamente.



Fonte: (JAIN, 2010)

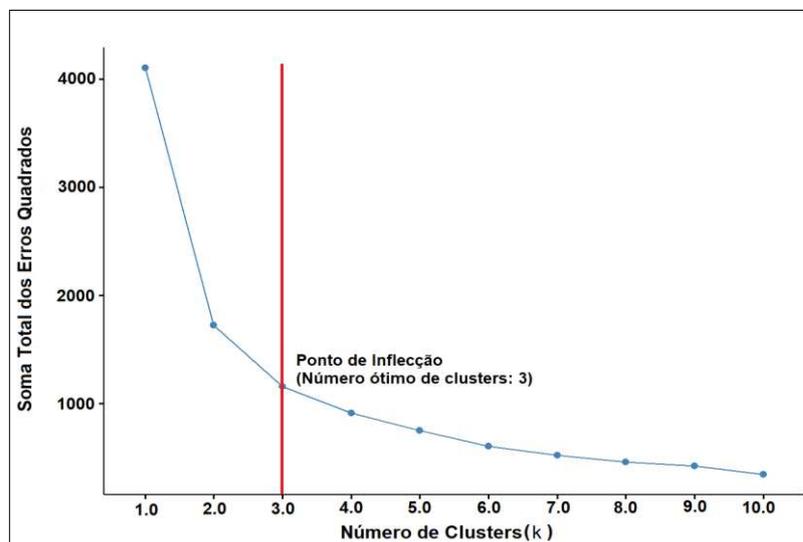
Método do Cotovelo

O método do cotovelo (Elbow Method) é uma técnica muito utilizada na análise de *clusters*, especialmente em algoritmos como o *k-means*, a fim de determinar o número ótimo de *clusters* (k) a serem escolhidos para um conjunto de dados. A ideia principal é avaliar a variabilidade intra-cluster em função do número de *clusters* e identificar o ponto onde o ganho na redução da variabilidade começa a diminuir, assemelhando-se a um "cotovelo" em um gráfico (BHOLOWALIA; KUMAR, 2014).

O procedimento geral para aplicar o método do cotovelo é indicado por Nainggolan *et al.* (2019) pelas seguintes etapas:

1. Execução do algoritmo de clusterização (por exemplo, *k-means*) para diferentes valores de k .
2. Para cada valor de k , a variabilidade intra-cluster (também chamada de inércia ou soma dos quadrados intra-cluster) é calculada.
3. O gráfico para identificação dos *clusters* é plotado (o eixo x representa os valores de k e o eixo y representa a variabilidade intra-cluster).
4. Identificação do ponto no gráfico onde a redução da soma dos erros quadrados começa a diminuir, formando um padrão semelhante a um cotovelo (Figura 9).

Figura 9 - Representação gráfica do número correto de *clusters* pelo Método do Cotovelo



Fonte: Autor (2024)

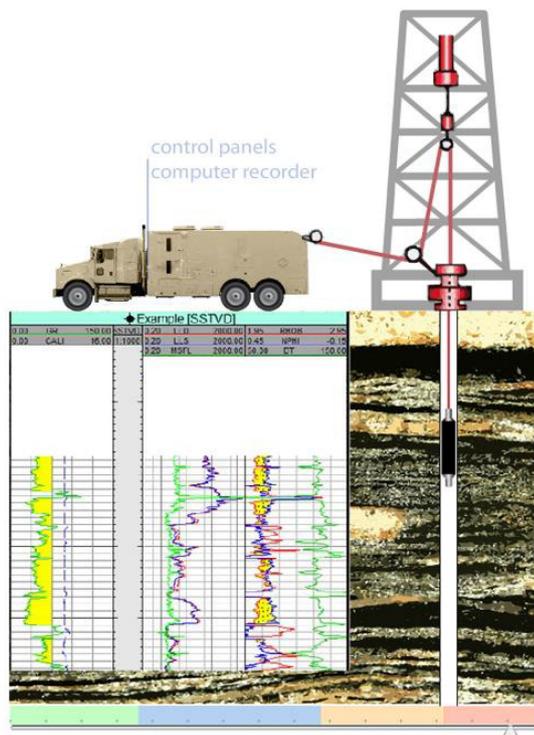
A interpretação do método do cotovelo requer alguma subjetividade e, em alguns casos, pode não haver um cotovelo claro no gráfico. Portanto, a escolha do número de *clusters* deve ser considerada juntamente com a compreensão do domínio do problema e outros métodos de validação de *clusters*.

3.2 Perfilaragem de Poços

Os perfis geofísicos de poços são medições feitas ao longo de um poço de perfuração utilizando instrumentação geofísica especializada, a qual depende das propriedades que se deseja medir (Rider, 2002). Eles fornecem informações sobre as características geológicas, estruturais e fluidas das formações rochosas atravessadas pelo poço. Portanto, visam avaliar e classificar reservatórios, assim como determinar a natureza dos possíveis fluidos presentes. Estas informações, como litologia, porosidade, saturação e permeabilidade de fluidos, são utilizadas para atividades relacionadas à exploração e produção de petróleo e gás.

Tipicamente, as medições desses perfis são realizadas progressivamente em uma direção, capturando amostras regularmente ao longo da profundidade. Os métodos para obter tais medições têm evoluído ao longo do tempo, assim como as ferramentas e técnicas utilizadas (Figura 10). Atualmente, nas operações de perfuração, várias ferramentas de medição são integradas em uma mesma sonda, incluindo os sensores MWD (Measurement While Drilling), os quais medem as propriedades mecânicas e parâmetros do poço (SAMPAIO, 2012), e as ferramentas de medição de propriedades geofísicas, conhecidas como LWD (Logging While Drilling) (Ellis e Singer, 2009).

Figura 10 - Representação simples da etapa de medição de perfis e propriedades do poço



Fonte: (OIL & GAS E&P – WELL LOGS, 2013)

Neste estudo, os principais perfis convencionais de alguns poços da Bacia Potiguar, fornecidos pelo site REATE da ANP, que será abordado no tópico 4, foram analisados e utilizados na aplicação do aprendizado de máquina. De acordo com os perfis de poços coletados, são apresentados os princípios básicos dos perfis de *gamma-ray* (GR), cáliber, neutrônico (NPHI), sônico (DT), densidade (RHOB) e resistividade.

3.2.1 Perfil Gamma-Ray (GR)

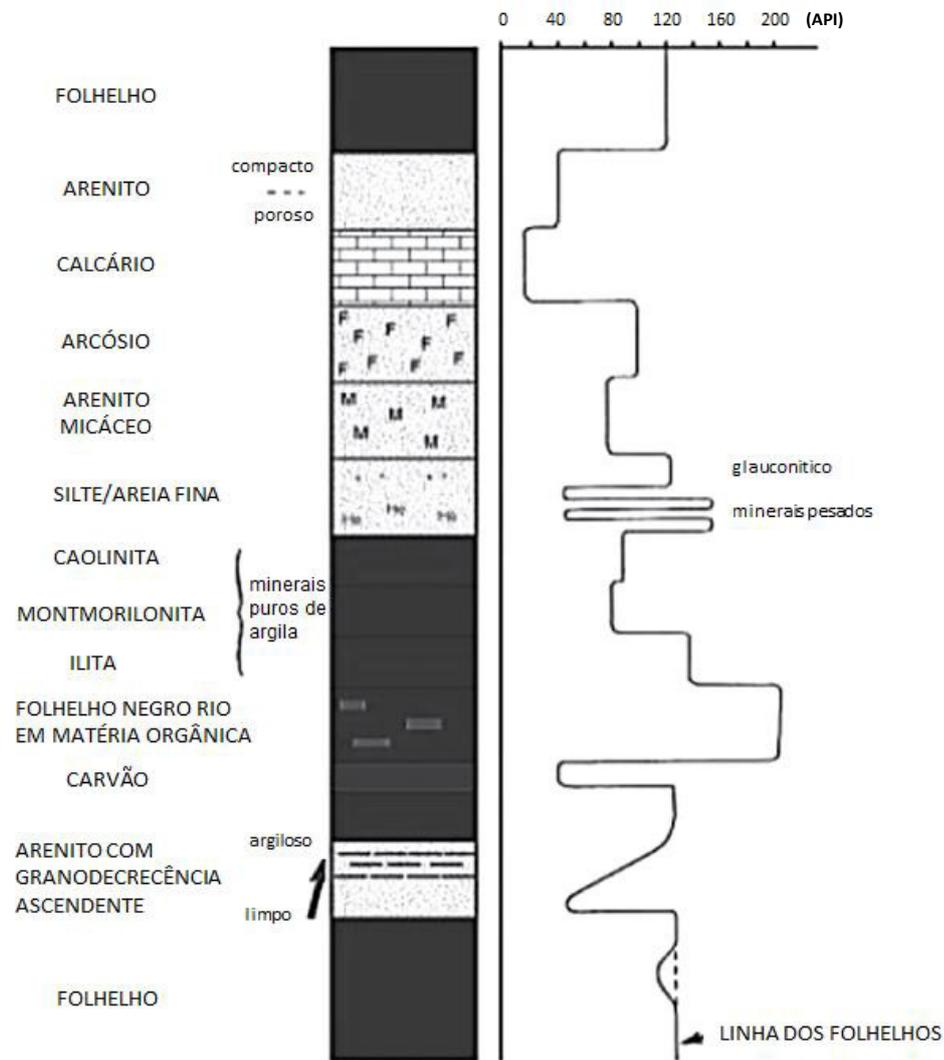
O perfil de raio gama (GR - *gamma-ray log*) é uma técnica de perfilagem utilizada na exploração e produção de hidrocarbonetos e mede a radioatividade natural das formações rochosas de acordo com a profundidade. A medição é realizada por uma sonda que é baixada ao longo do poço, a qual detecta variações na radioatividade que podem ser correlacionadas com diferentes tipos de litologias (Nazeer *et al.*, 2016).

Este perfil é registrado por um dispositivo conhecido como sonda de detectores de radiação, como o cintilômetro e, às vezes, o contador Geiger-Muller, que são componentes da ferramenta de raios gama. Esses aparelhos medem a radioatividade proveniente da desintegração de elementos radioativos, como potássio (K^{40}), tório (Th^{232}) e urânio (U^{238}), que ocorrem naturalmente nas rochas (Hurley, 1956). A ferramenta de raios gama pode detectar a radioatividade originada a poucos decímetros da parede do poço, fornecendo dados essenciais para a análise geológica (Rider, 2002).

Rochas ricas em argila, como folhelhos, geralmente apresentam altas leituras de GR devido ao maior conteúdo de minerais radioativos, enquanto rochas como arenitos e calcários apresentam leituras mais baixas, como mostra a figura 11 (Russel, 1944). Pontos altos no perfil GR indicam a presença de rochas argilosas ou folhelhos, que podem atuar como barreiras ou selos de hidrocarbonetos, enquanto pontos baixos sugerem a presença de arenitos ou calcários, que são tipicamente reservatórios potenciais devido à sua maior porosidade e permeabilidade.

Este contraste de valores do perfil *gamma-ray* observado entre as formações permite a identificação das interfaces entre diferentes camadas geológicas, sendo possível determinar a distribuição vertical das litologias e auxiliar na interpretação da estrutura do reservatório (Alvarez *et al.*, 2003). Além disso, o perfil GR é útil na correlação entre poços e na determinação da espessura e continuidade das zonas de interesse (Huang, 1983).

Figura 11 - Representação da medida do perfil raio gama para diferentes litologias



Fonte: Adaptado de Rider, 2002

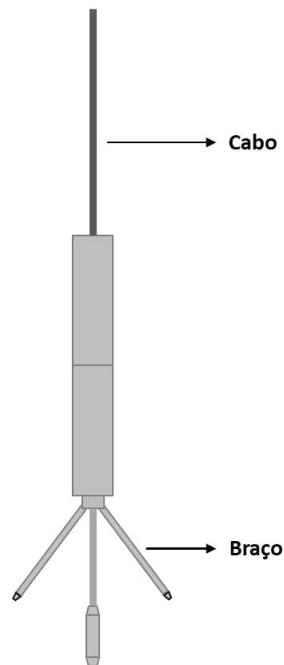
A unidade de medida utilizada para o perfil é o API (American Petroleum Institute) gama, que é uma unidade padronizada de radioatividade. A escala de medição do perfil GR normalmente varia de 0 a 200 API, embora em algumas circunstâncias, perfis com uma escala maior possam ser usados dependendo da radioatividade da formação rochosa (Santos, 2010).

3.2.2 Perfil de Cáliper

O perfil cáliper é uma técnica de medição utilizada para determinar o diâmetro de um poço ao longo de sua profundidade, fornecendo informações sobre a geometria do poço. O processo é realizado com uma ferramenta específica que é inserida no poço por meio de um cabo de perfilagem. Esta ferramenta possui braços extensíveis (*calipers*) que são pressionados

às paredes do poço, ajustando-se conforme às dimensões do mesmo (Figura 12). O número de braços pode variar, geralmente sendo três ou mais, dependendo do tipo de perfilagem e da precisão necessária. Conforme a ferramenta é descida ou puxada para cima ao longo do poço, os braços medem continuamente o diâmetro interno do poço (Santos, 2010).

Figura 12 – Esquematização básica de um dos tipos de equipamentos de medição do cáliper

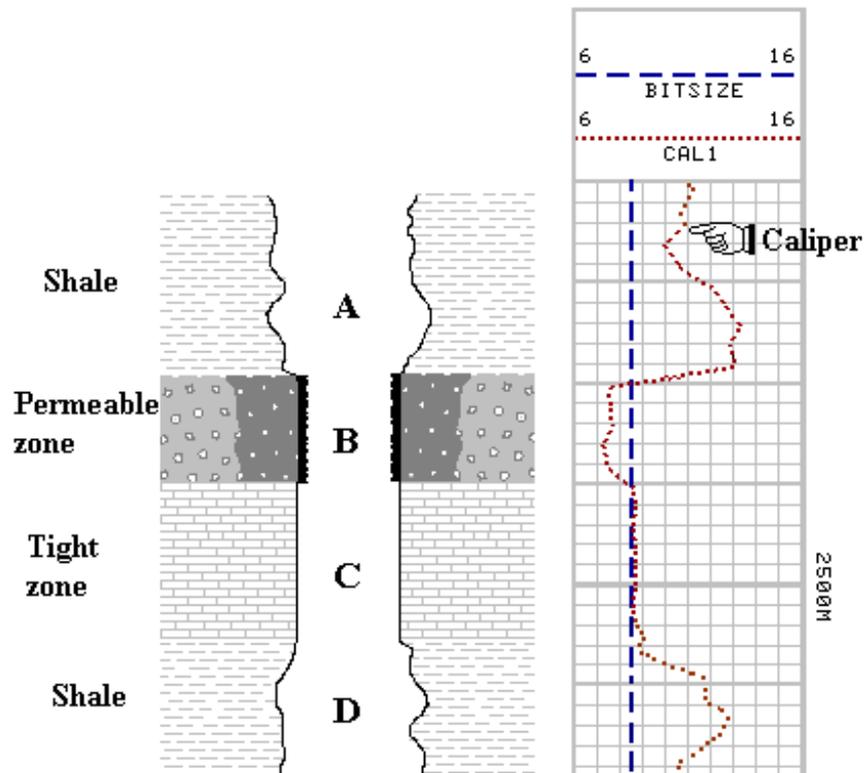


Fonte: Autor (2024)

A medição do diâmetro do poço auxilia na verificação da qualidade de outras medições realizadas durante o levantamento do poço, ou seja, de outros perfis (Timm, 2017). Este controle também avalia a precisão da perfuração e permite monitorar mudanças no diâmetro ao longo do tempo, o que ajuda na identificação de problemas ou fenômenos no poço.

A análise da integridade do poço feita a partir do perfil cáliper sofre influência de determinadas informações da formação, como a litologia e a textura e estrutura das rochas, como mostra a figura 13 (Serra, 1984). Localmente, dissoluções anteriores à perfuração, como cavernas cársticas ou dolinas, e rochas, como sal ou carbonatos, que podem ser dissolvidas pela lama, afetam o diâmetro do poço. Além disso, a compactação e a resistência ao desmoronamento das formações são fatores litológicos importantes. Já com relação à textura e estrutura, a porosidade e a permeabilidade das rochas também sofrem influência, as quais são determinantes para a aderência e espessura do reboco na etapa de cimentação (Santos, 2010).

Figura 13 – Representação do Perfil Cáliper de acordo com a formação



Fonte: Assaife, 2022

O perfil deve ser próximo ao tamanho da broca (*bit size*) para fornecer medições precisas e representativas do diâmetro do poço. A unidade de medida utilizada é geralmente polegadas (*in*) ou milímetros (*mm*), dependendo do sistema de unidades adotado.

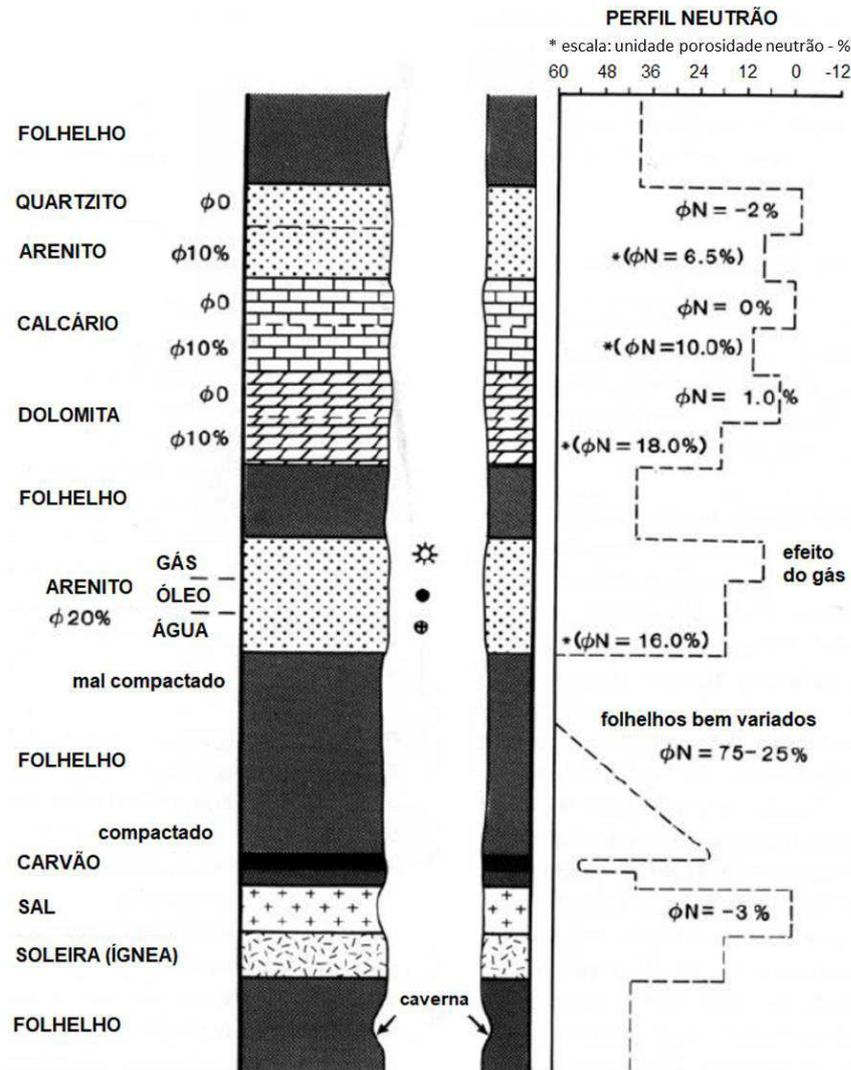
3.2.3 Perfil Neutrão (NPHI)

O perfil de porosidade neutrão (NPHI) fornece a medição contínua da reação das rochas ao bombardeamento de nêutrons. Ele, na prática, estima o teor de hidrogênio na formação, o qual reflete diretamente na porosidade da rocha (Rider, 2002). Os nêutrons são partículas subatômicas, eletricamente neutras e com massa semelhante à do hidrogênio. Através de uma fonte radioativa, nêutrons de alta energia são emitidos, os quais penetram profundamente na matéria, perfurando a eletrosfera e colidindo diretamente com os núcleos dos elementos presentes na rocha, de forma a interagir elástico e/ou inelasticamente (Santos, 2010).

Ao colidirem com átomos de tamanho semelhante, como o hidrogênio, os nêutrons perdem energia. Durante esses choques, a perda de energia permite que os nêutrons sejam capturados pelo núcleo do átomo, que então fica excitado e libera raios gama de alta energia. O

receptor da ferramenta conta os raios gama liberados em um determinado intervalo de tempo. A taxa de decaimento observada está relacionada tanto à captura de nêutrons quanto à quantidade de átomos de hidrogênio presentes na rocha da formação. (Rider, 2002).

Figura 14 – Resposta do Perfil Neutrão para algumas das principais litologias



Fonte: Adaptado de Rider, 2002

A figura 14 mostra alguns valores estimados de porosidade neutrão para diferentes litologias. A aplicação do perfil neutrão tem como objetivo principal determinar a porosidade, além de ser uma informação relevante para a discriminação de zonas de gás, em especial quando analisado junto com o perfil sônico. Geologicamente, pode identificar diferentes litologias, como evaporitos (Ex.: Camadas de sal), minerais hidratados (Ex.: Argilas hidratadas) e rochas vulcânicas (Ex.: Basalto) (Stevanato, 2011).

A interpretação do perfil também envolve o cruzamento com outros perfis para, assim, obter características mais visíveis e detalhadas. É necessário considerar, por exemplo, que as zonas de gás ou hidrocarbonetos possuem baixa porosidade, mas alta densidade. Além disso, argilas possuem água adsorvida, resultando em folhelhos com maior porosidade do que arenitos limpos. A presença de lama entre a sapata e a parede do poço ao redor do detector aumenta a quantidade de hidrogênio e diminui a resposta das formações. Para a identificação de rochas ígneas, elas mostram baixa porosidade no perfil de nêutrons, alta densidade e baixo raio gama (Nery, 1990).

3.2.4 Perfil Densidade (RHOB)

O perfil de densidade (RHOB – Bulk Density) trata-se de um registro contínuo da densidade das formações ao longo do poço. Este perfil é utilizado para a determinação da porosidade das rochas, a qual é calculada ao comparar a densidade medida com a densidade da matriz da rocha e a densidade dos fluidos presentes nos poros, como mostra a equação 3.1 (Rider, 2002).

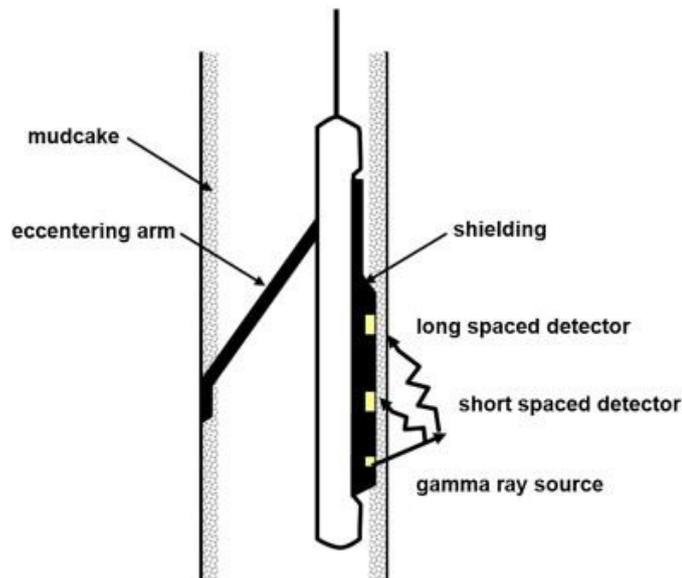
$$\phi (\text{porosidade}) = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f} \quad (3.1)$$

Em que:

- ρ_{ma} – Densidade da matriz (ou grão);
- ρ_b – Densidade da formação lida no perfil;
- ρ_f – Densidade do fluido nos poros.

O perfil de densidade tem o princípio de funcionamento baseado na técnica de espalhamento de Compton, em que uma fonte radioativa, geralmente Césio (Cs^{137}) ou Cobalto (Co^{60}), dentro da sonda emite raios gama de média e alta energia que penetram na formação. Esses raios gama são espalhados ao interagir com os elétrons das rochas e a quantidade de espalhamento é proporcional à densidade da formação (Keary *et al*, 2009). Os detectores presentes na sonda medem, então, os raios gama espalhados, permitindo a conversão desses dados em valores de densidade aparente, que são utilizados para determinar a densidade da formação rochosa e o tipo de litologia presente ao longo do poço (Ellis, 1988). A figura 15 mostra a ferramenta de medição de porosidade e o funcionamento da mesma.

Figura 15 – Esquemática da ferramenta de medição do perfil de densidade

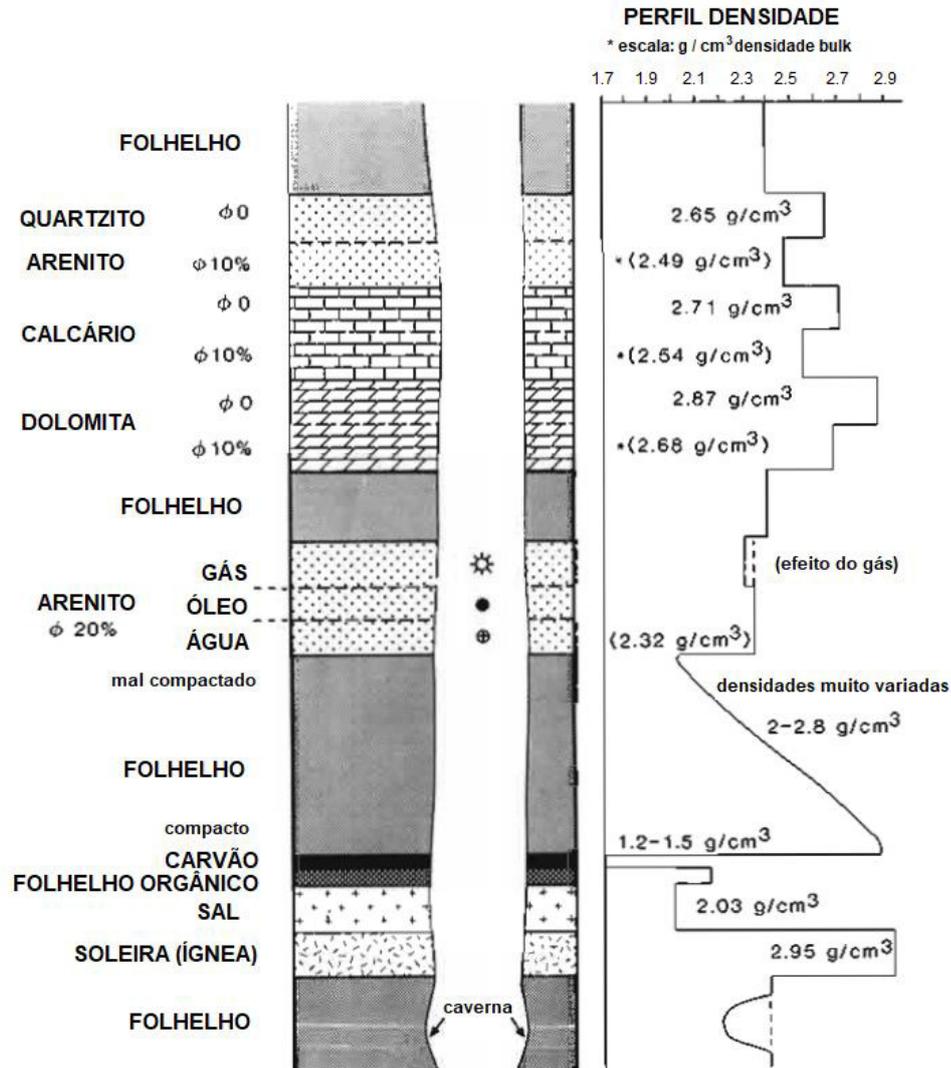


Fonte: Baker *et al.*, 2015

A interpretação dos dados medidos do perfil de densidade permite identificar diferentes litologias, calcular a porosidade e avaliar a presença de hidrocarbonetos. Por exemplo, uma densidade mais baixa pode indicar maior porosidade ou a presença de hidrocarbonetos, enquanto uma densidade mais alta pode sugerir rochas mais densas e compactas (Rider, 2002).

No entanto, ao observar a figura 16, a qual exibe alguns valores de densidade para diferentes litologias e suas variações, é possível perceber que os intervalos de densidades típicas apresentam proximidade umas com as outras. Por exemplo, folhelhos e arenitos podem exibir valores próximos dependendo do grau de compactação da formação. Por conta disso, a análise isolada do perfil de densidade dificilmente é usada como determinante para a indicação da maioria das litologias, sendo comumente combinada com o perfil de porosidade neutro e o perfil sônico para a identificação de zonas com gás e a determinação da impedância acústica, respectivamente (Stevanato, 2011).

Figura 16 – Resposta do perfil de densidade para algumas litologias mais comuns

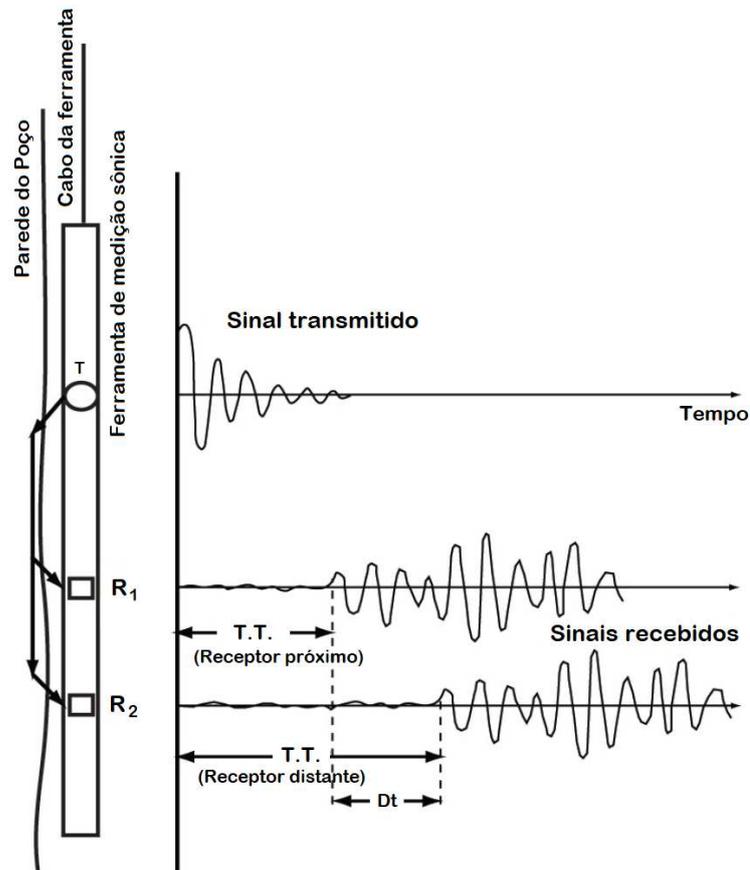


Fonte: Adaptado de Rider, 2002

3.2.5 Perfil Sônico (DT)

O perfil sônico (Delay Time – DT), ou perfil acústico, registra o intervalo de tempo entre a emissão de um pulso sonoro compressional por um transmissor (T), montado em um mandril no interior do poço, e a recepção desse pulso por dois receptores distintos (R_1 e R_2), localizados a distâncias fixas e pré-determinadas no mesmo mandril (Serra, 1984), como mostra a figura 17. Os receptores, geralmente espaçados cerca de 300 mm entre si, detectam o tempo de trânsito do pulso sonoro através do poço, lama e formação, permitindo a medida da capacidade da formação de transmitir ondas sonoras. A frequência dos pulsos ultrassônicos emitidos pela fonte varia entre 20-40 kHz e a unidade de medida do perfil é expressa em microssegundos por pé ($\mu\text{s}/\text{ft}$) (Keary *et al.*, 2009).

Figura 17 - Princípio de medição do tempo de trânsito (T.T.) em intervalo sônico usando um único transmissor (T) e dois receptores (R_1 e R_2)



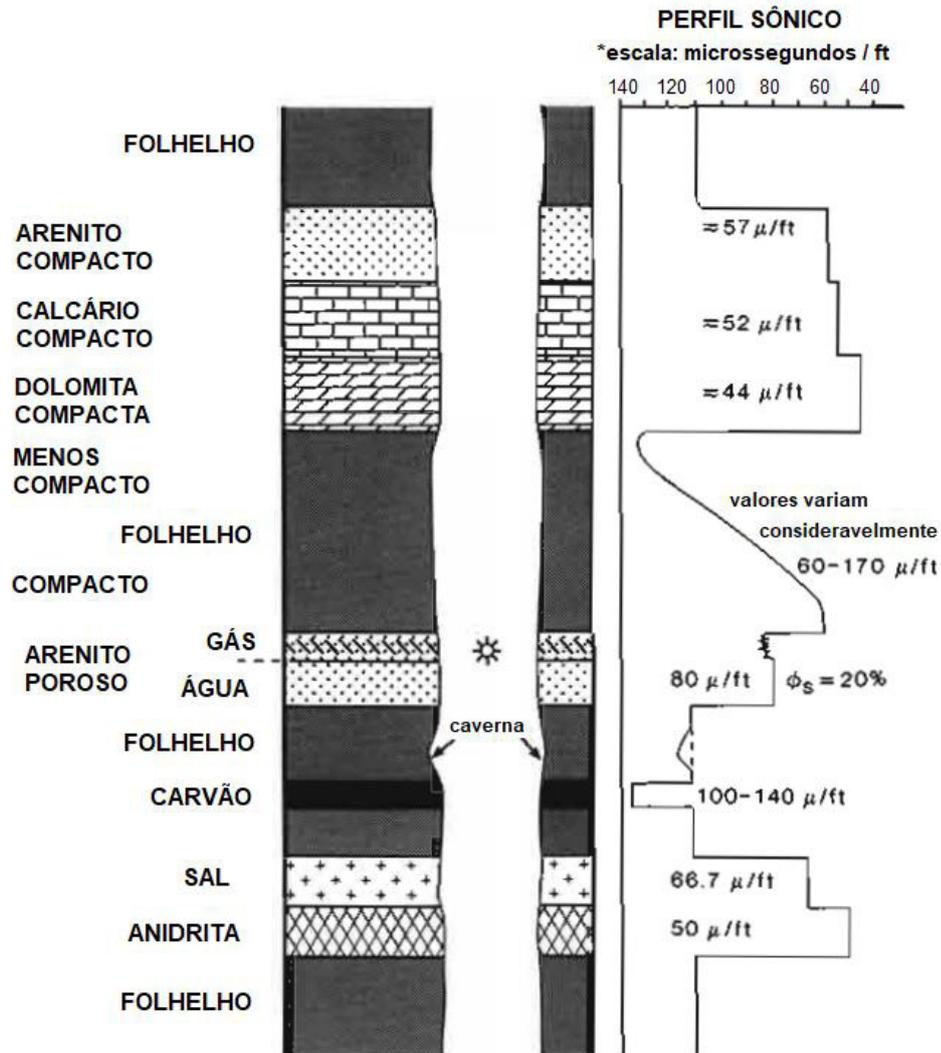
Fonte: Adaptado de Close *et al.*, 2009

A velocidade das ondas acústicas nas formações rochosas tem uma relação inversamente proporcional com os tempos de trânsito. Em outras palavras, em formações rochosas compactas, onde a velocidade da onda acústica é maior, os tempos de trânsito são menores. Em contrapartida, em formações com maior porosidade e saturação de fluidos nos seus poros, a velocidade da onda acústica é menor, resultando em tempos de trânsito maiores (Rider, 2002).

Qualitativamente, o perfil sônico é utilizado para calcular a porosidade da formação e é essencial na interpretação sísmica, fornecendo intervalos de velocidade e perfis de velocidade que podem ser calibrados com dados de seções sísmicas, o que permite melhorar a precisão dos modelos geológicos e sísmicos. Além disso, o perfil sônico é sensível a mudanças texturais tanto em arenitos quanto em folhelhos, auxiliando na identificação de litologias, apesar da alta variação entre as velocidades dentro de cada formação (Figura 18). Ele também pode

indicar a presença de falhas e fraturas na formação, proporcionando informações valiosas para a caracterização do reservatório (Stevanato, 2011).

Figura 18 – Valores de tempo de trânsito para diferentes litologias



Fonte: Adaptado de Rider, 2002

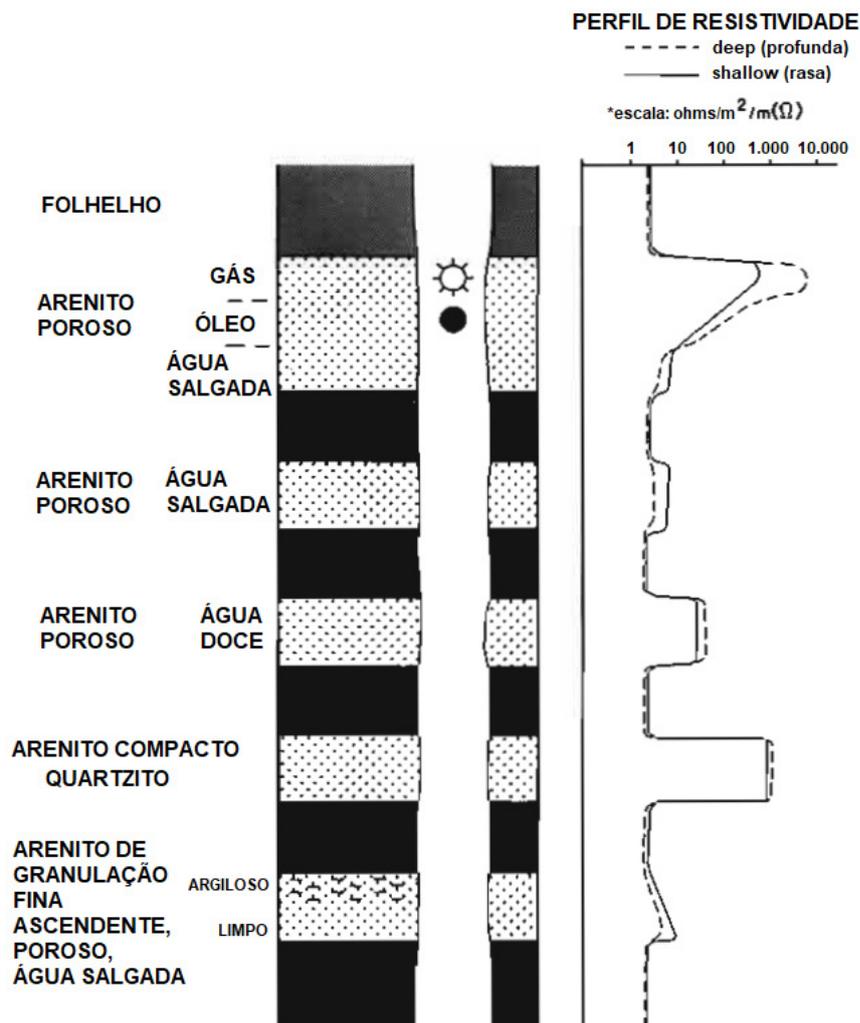
3.2.6 Perfil de Resistividade

O perfil de resistividade é um perfil elétrico utilizado para medir a resistência elétrica das formações ao longo do poço. O processo envolve a injeção de uma corrente elétrica nas formações circundantes e a medição da diferença de potencial resultante. Estas medições permitem identificar diferentes camadas de materiais, fluidos ou formações geológicas, como

rochas, água, hidrocarbonetos e cavidades, com base nas variações de resistividade, além de ajudar em informações sobre a litologia, textura e outras características das rochas (Rider, 2002).

De acordo com Santos (2010), o princípio de funcionamento da ferramenta de indução para medir o perfil de resistividade é baseado na geração de campos eletromagnéticos primários por uma bobina emissora, que induzem correntes elétricas alternadas na formação, conhecidas como correntes de Foucault. Estas correntes, por sua vez, geram campos magnéticos secundários que são captados por bobinas receptoras. A corrente alternada enviada para a bobina emissora cria um campo magnético concêntrico que induz essas correntes na formação, propagando o sinal. A bobina receptora capta tanto o campo magnético primário quanto os secundários, gerando uma corrente induzida em seu circuito. A partir do sinal recebido e da geometria de aquisição, calcula-se a impedância elétrica e a constante dielétrica da formação, estimando assim a resistividade.

Figura 19 – Valores de resistividade para diferentes litologias



Fonte: Adaptado de Rider, 2002

O principal objetivo do perfil de resistividade é detectar hidrocarbonetos, pois estes são maus condutores elétricos e, conseqüentemente, aumentam a resistividade da rocha medida, como pode ser observado na figura 19. Em contraste, quando os poros da rocha estão preenchidos com água, a resistividade diminui. Além disso, os valores de resistividade são influenciados pela litologia e pela granulometria da rocha (Rider, 2002). A resistividade permite identificar zonas com hidrocarbonetos e água, indicar áreas permeáveis e determinar a porosidade com base na resistividade. No entanto, as ferramentas de resistividade só são eficazes em poços que contêm lamas condutivas, as quais são geralmente misturadas com água salgada.

4 MATERIAIS E MÉTODOS

No âmbito deste trabalho, é imprescindível compreender a relevância da metodologia empregada, a qual desempenha um papel crucial na condução e execução da pesquisa. Este capítulo tem como propósito oferecer uma visão abrangente sobre a abordagem metodológica adotada e os materiais utilizados, destacando os processos essenciais relacionados à coleta, seleção e tratamento dos dados, bem como à aplicação do algoritmo *k-means* na caracterização das eletrofácies.

A solidez e compreensão dessa abordagem metodológica é fundamental para assegurar a validade e a confiabilidade dos resultados obtidos, o que contribui para a construção de uma base sólida para a análise e interpretação dos achados deste projeto.

4.1 Base de Dados

Fonte dos Dados

A fase inicial deste estudo envolveu a coleta criteriosa de dados, em que o site REATE da ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis) foi escolhido como fonte primária de informações dos poços. Essa seleção baseou-se na relevância e abrangência dos dados disponíveis no referido site para o escopo da pesquisa. Trata-se de um banco de dados de domínio público, disponível desde 2021, com dados técnicos de bacias sedimentares terrestres brasileiras e possui dados digitais de poços, dados sísmicos 2D e 3D

pós-stack (Técnica usada para identificar as características geológicas do subsolo), dados não-sísmicos, dados geoquímicos e estudos diversos.

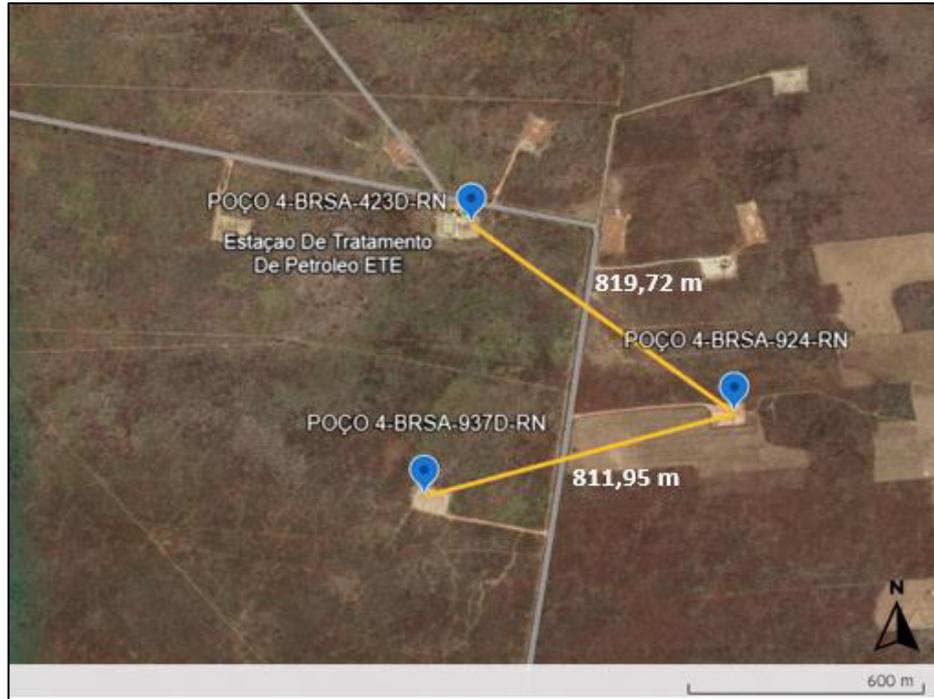
Poços Estudados

Para esse trabalho, foram determinados três poços presentes na porção emersa da Bacia Potiguar, são eles: Poço 4-BRSA-423D-RN, Poço 4-BRSA-924-RN e Poço 4-BRSA-937D-RN. Os três poços estão localizados no Campo de Jaçanã, todos situados no município de Apodi, no Rio Grande do Norte. A figura 20 mostra a disposição dos poços escolhidos de acordo com as coordenadas geográficas das sondas disponibilizadas nos perfis compostos dos mesmos. A visualização da localização dos poços no mapa e a distância entre eles foram obtidas com o auxílio do programa Google Earth.

A escolha dos poços foi fundamentada na premissa de que poços geograficamente próximos tendem a apresentar padrões e propriedades geológicas semelhantes. Contudo, é importante destacar que essa similaridade não é garantida, pois variáveis geológicas podem diferir significativamente, mesmo em áreas próximas. No caso específico analisado, os três poços apresentam formações geológicas similares, embora com diferentes históricos e potenciais de produção. Um dos poços possui um histórico comprovado de produção, enquanto os outros dois não tiveram confirmação de hidrocarbonetos ou não se mostraram economicamente viáveis para a produção, como mostram os relatórios finais dos poços.

Ao incorporar esses dados no processo de treinamento de modelos de aprendizado de máquina, a intenção é aproveitar a similaridade geológica identificada para prever com maior precisão as características geofísicas em novos locais, considerando as variações potenciais na produtividade.

Figura 20 – Coordenadas Geográficas dos poços estudados



Fonte: Autor, 2024

Arquivos dos Poços

Os documentos referentes aos poços, como perfis compostos, relatórios finais e acompanhamento geológico, são comumente disponibilizados no formato PDF. Essa escolha de formato visa proporcionar uma visualização estável e acessível desses documentos, facilitando a revisão e análise por parte dos interessados.

Por outro lado, os dados digitais de perfilagem destinados ao aprendizado e análise mais aprofundada são disponibilizados em arquivos DLIS (Digital Log Interchange Standard). O formato binário padronizado dos arquivos DLIS é especialmente projetado para representar uma variedade de dados geofísicos complexos (Faiz, 2023), oferecendo uma estrutura eficiente para a troca de informações entre diferentes sistemas e ferramentas na indústria de exploração e produção de petróleo e gás. Essa padronização contribui para a precisão e interpretação consistente desses dados ao longo do tempo, sendo particularmente valiosa em contextos de aprendizado e pesquisa na área.

A leitura eficiente dos arquivos com esse formato foi possível por meio da utilização da biblioteca **dlisio**, em Python. Essa biblioteca oferece ferramentas e funcionalidades específicas para trabalhar com os arquivos DLIS, o que permite a extração e interpretação de dados geofísicos complexos de maneira eficaz. A integração do **dlisio** proporcionou uma

abordagem programática para acessar e analisar as informações contidas no arquivo DLIS, facilitando assim a manipulação e utilização desses dados no contexto do projeto.

4.2 Processamento dos dados

Posteriormente ao estágio de escolha dos poços, a leitura e o processamento dos arquivos foram feitos por meio da biblioteca `dlisio` dentro do ambiente do Jupyter Notebook, o qual se mostrou uma ferramenta importante e de fácil compreensão e manuseio. Para esta etapa, foram realizados os seguintes passos: A exploração, seleção e o tratamento dos dados.

Exploração dos Arquivos dos Perfis Compostos dos Poços

Devido à possibilidade de um único arquivo DLIS físico conter vários arquivos lógicos, por conta da facilidade de acesso e manipulação e a fim de garantir a integridade dos dados, foi realizada a verificação do número de arquivos e o teor de cada informação para cada um dos poços estudados. Para isso, o arquivo foi, então, carregado e, por meio da função “load”, que faz parte da biblioteca `dlisio`, foi possível ver a quantidade de arquivos lógicos (`LogicalFile()`) existentes. A tabela 2 mostra esses dados para os poços.

Tabela 2 – Quantidade de Arquivos Lógicos de cada poço

Poço	Arquivos Lógicos	Nomes dos Arquivos
4-BRSA-423D-RN	2	LogicalFile(GEOLOAD.1), LogicalFile(GEOLOAD.2)
4-BRSA-924-RN	2	LogicalFile(GEOLOAD.1), LogicalFile(GEOLOAD.2)
4-BRSA-937D-RN	2	LogicalFile(GEOLOAD.1), LogicalFile(GEOLOAD.2)

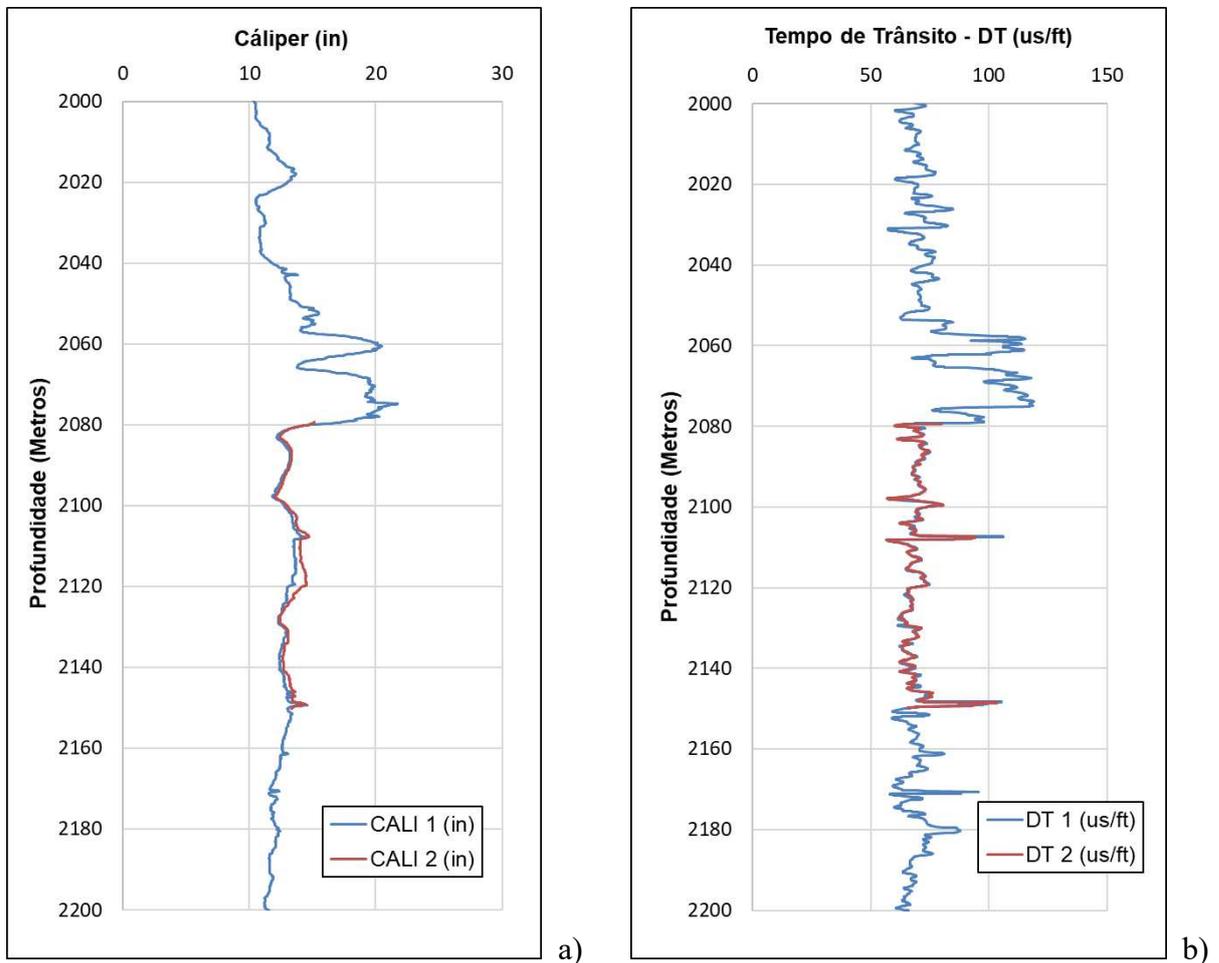
Fonte: Autor, 2024

Após observar a existência de dois arquivos lógicos para cada um dos poços, foi necessário visualizar o conteúdo de cada arquivo e definir aquele que será utilizado para o estudo, ou ainda, caso alguns dados dos dois arquivos sejam mais precisos e corretos, unir as

informações contidas neles em um só *DataFrame*, estrutura de dados bidimensional, tabular e mutável, a fim de processá-los posteriormente.

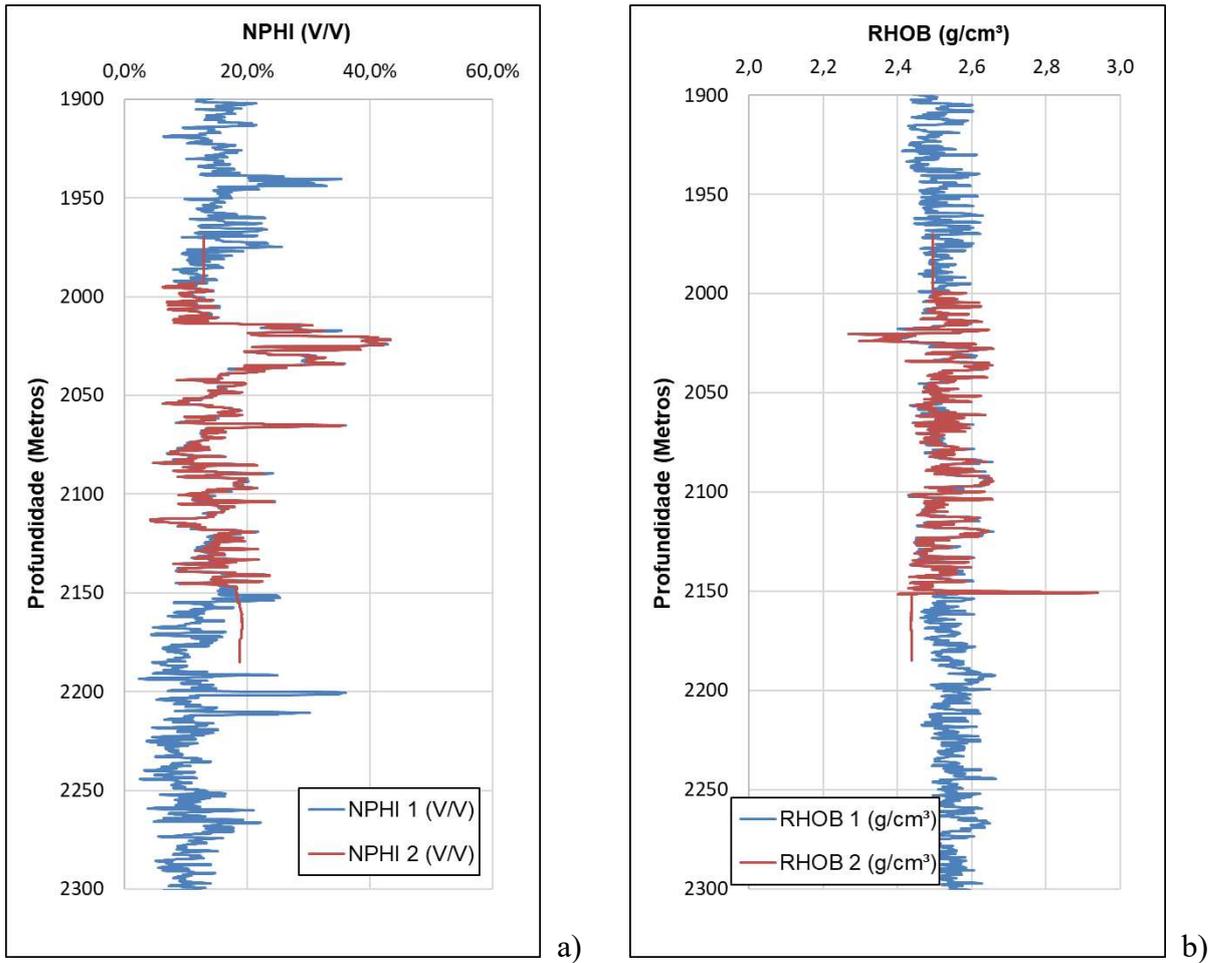
As figuras 21, 22 e 23 exibem a comparação visual entre alguns dos perfis presentes nos dados dos arquivos lógicos dos poços. É possível observar que todos os arquivos secundários apresentam menor intervalo que os arquivos principais, “LogicalFile (GEOLOAD.1)”, e, além disso, a maioria dos pontos estão sobrepostos ao primeiro arquivo de cada poço.

Figura 21 – Comparação entre os dados dos perfis Cáliper (a) e DT (b) dos arquivos primário e secundário (1 e 2) do Poço 4-BRSA-924-RN



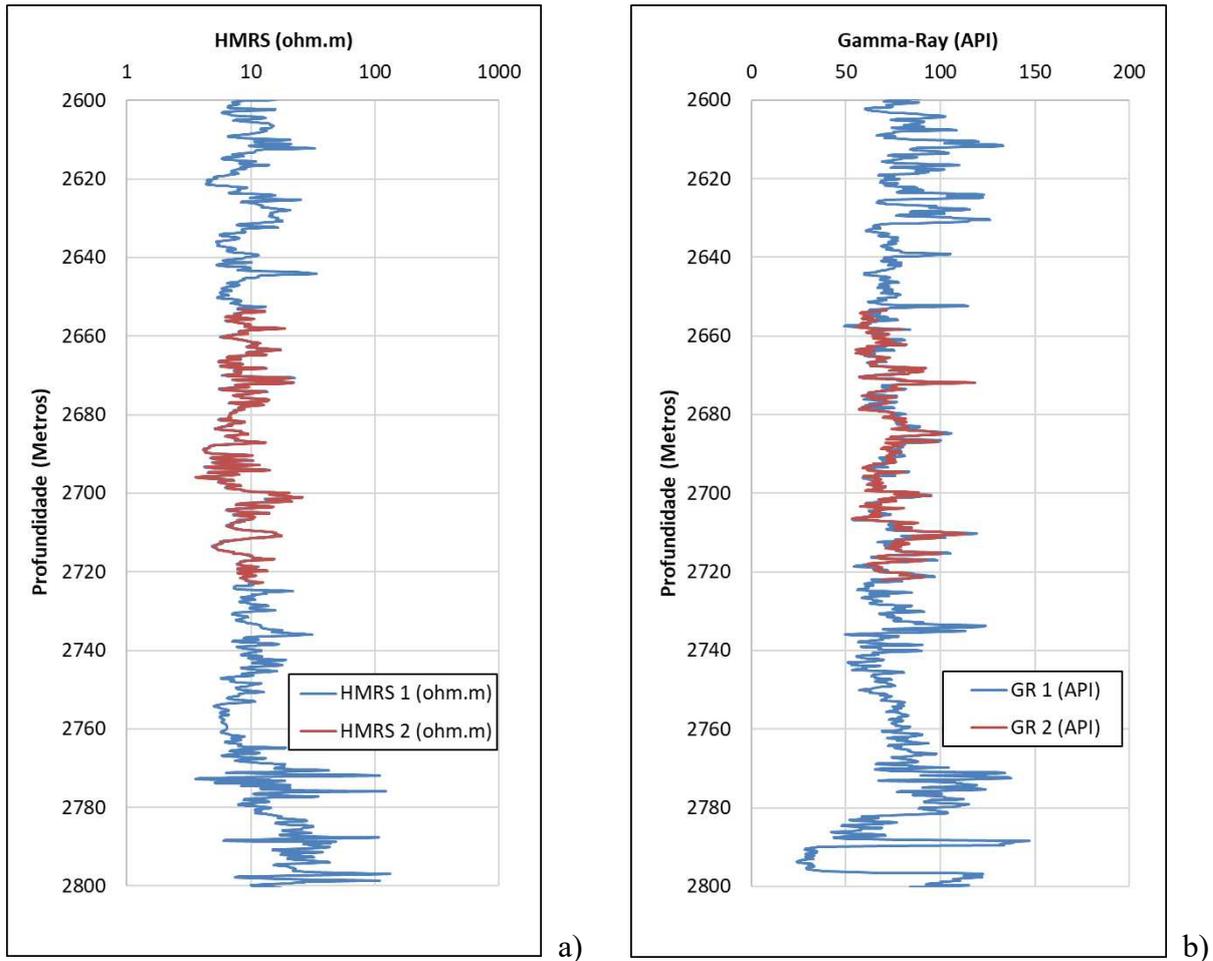
Fonte: Autor, 2024

Figura 22 – Comparação entre os dados dos perfis NPHI (a) e RHOB (b) dos arquivos primário e secundário (1 e 2) do Poço 4-BRSA-423D-RN



Fonte: Autor, 2024

Figura 23 – Comparação entre os dados dos perfis HMRS - Resistividade (a) e GR (b) dos arquivos primário e secundário (1 e 2) do Poço 4-BRSA-937D-RN



Fonte: Autor, 2024

Dessa forma, após essa verificação, o arquivo secundário do DLIS de cada poço foi desconsiderado e o estudo focou no arquivo principal completo, evitando problemas de dados escassos e descontinuação nas curvas dos perfis de poço.

Seleção dos dados

Com os arquivos para o processamento dos dados definidos, a próxima etapa tratou-se da seleção dos perfis que serão usados no estudo e posterior aprendizado de máquina para predição das eletrofácies. Os perfis foram escolhidos de acordo com a importância geológica em representar e caracterizar a formação e suas características, como foi mostrado em alguns itens do tópico 3, os quais mostram as informações e atributos que os principais perfis medem e fornecem. Sendo assim, os perfis determinados foram: Cáliper, Gamma-Ray, Tempo de Trânsito, Resistividade, Densidade e Neutrão.

Após a definição dos perfis, um *DataFrame* foi criado para cada poço com as informações destas curvas com o objetivo de analisar e tratar estes dados que serão usados em uma etapa posterior. Para isso, o módulo Pandas, biblioteca para manipulação de dados em Python, foi utilizado para o desenvolvimento desse processo. Os perfis são distribuídos em colunas nesta estrutura de dados de acordo com a profundidade e, então, os dados poderão ser modelados.

Tratamento dos dados

O primeiro passo do tratamento estava ligado ao entendimento dos dados, a fim de identificar problemas na qualidade dos mesmos, como valores nulos ou errados. Diante disso, foram observados valores nulos, sobretudo, no início e no final da medição dos perfis dos três poços, ou seja, em intervalos de profundidade sem interesse para o estudo e, com isso, estes pontos foram excluídos.

A partir disso, tem-se a fase seguinte, a qual incluiu a definição do intervalo de profundidade para o estudo, tanto para o poço de treino como para os poços de teste. A tabela 3 mostra as regiões escolhidas para cada poço e que serão utilizadas na etapa de clusterização.

Tabela 3 – Identificação dos Intervalos de estudo de cada poço e sua identificação no Aprendizado de Máquina

Poço	Intervalo de Profundidade (Metros)	Divisão dos Dados	Quantidade de dados do poço
4-BRSA-423D-RN	1450 - 1550	Poço de Teste	1000 linhas de dados
4-BRSA-924-RN	1100 - 1500	Poço de Treino	4000 linhas de dados
4-BRSA-937D-RN	2200 - 2300	Poço de Teste	1000 linhas de dados

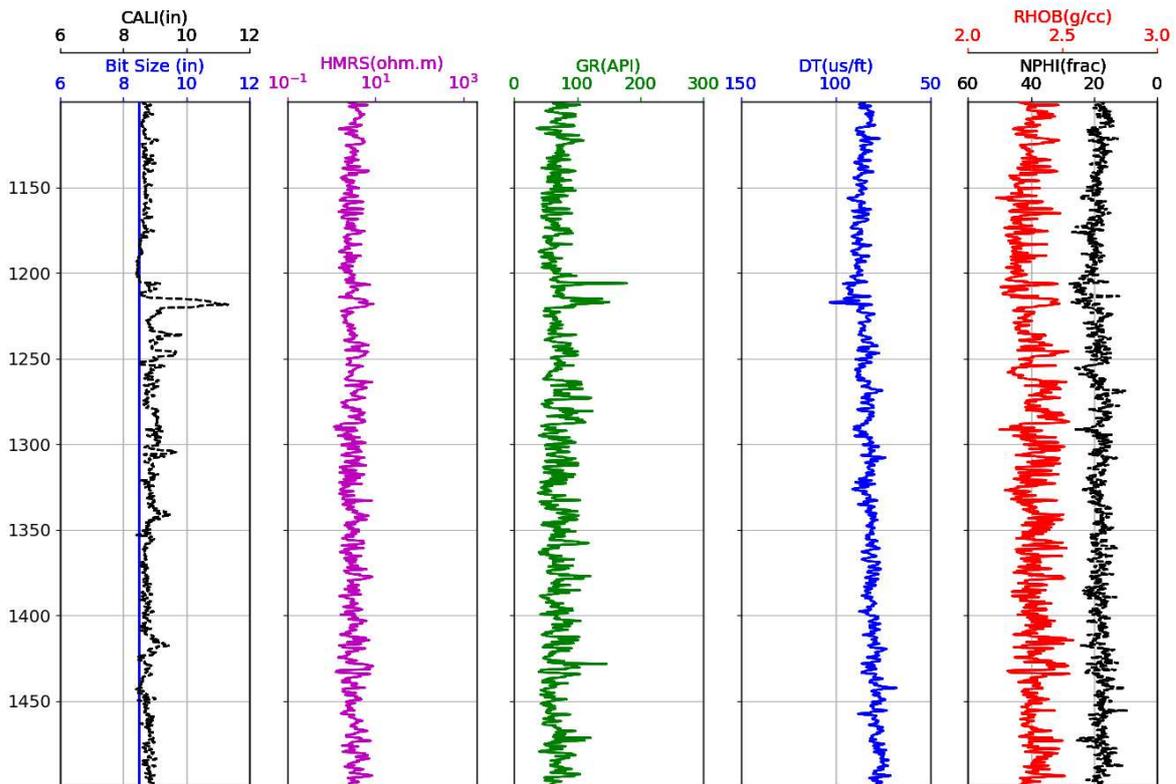
Fonte: Autor, 2024

A determinação dos intervalos de profundidade foi tomada de acordo com a formação da Bacia, a qualidade da medição do perfil de Cáliper e de regiões potenciais para o armazenamento ou estocagem de fluidos, no caso dos poços de teste.

Todos os intervalos de estudo dos poços são da Formação Pendência, uma vez que se busca representar características semelhantes da formação para os três poços. Além disso, os intervalos possuem medição aproximadamente constante do perfil Cáliper, ou seja, o diâmetro do poço segue com valores bem próximos do valor do tamanho da broca, o que significa regiões sem arrombamento do poço. Já em se tratando da escolha por regiões potenciais de reservatórios, os perfis compostos e os testes de formação disponíveis para estudo dos poços foram utilizados para a análise e definição de intervalos de profundidade adequados e de relevância para a estocagem. Com isso, foi verificada a existência de pontos ótimos para o estudo, sobretudo para o poço 4-BRSA-423D-RN, o qual, além de possuir registros de teste de formação, apresenta histórico de produção no intervalo selecionado, o que não acontece com os outros dois poços.

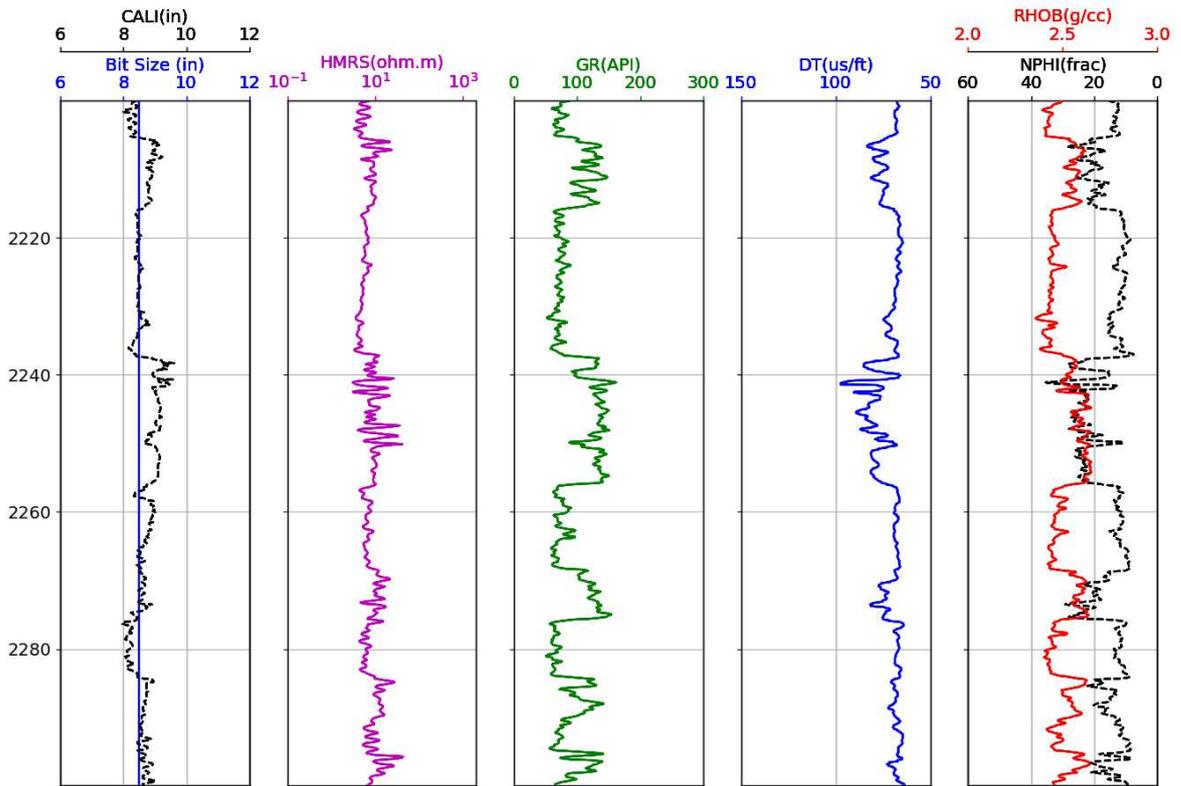
As figuras 24, 25 e 26 mostram a visualização dos intervalos de profundidade escolhidos para os três poços e seus respectivos perfis utilizados no estudo.

Figura 24 – Perfis do intervalo de estudo do Poço 4-BRSA-924-RN



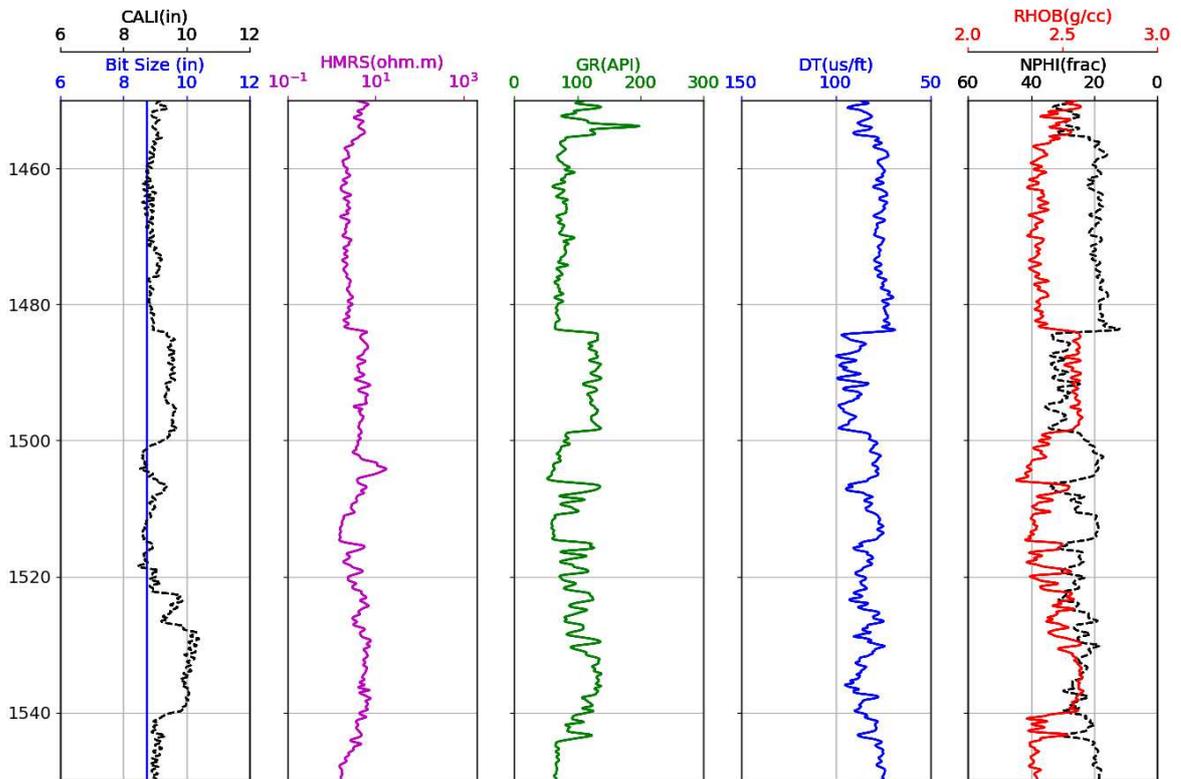
Fonte: Autor, 2024

Figura 25 – Perfis do intervalo de estudo do Poço 4-BRSA-937D-RN



Fonte: Autor, 2024

Figura 26 – Perfis do intervalo de estudo do Poço 4-BRSA-423D-RN

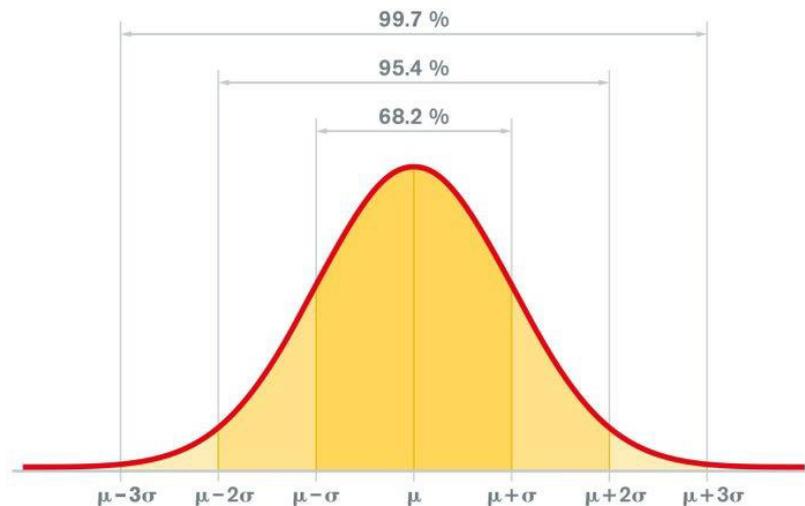


Fonte: Autor, 2024

Posterior a etapa de seleção do intervalo, foi necessário tratar os dados do poço de treino para que as informações enviadas à máquina para o aprendizado sejam adequadas e tragam confiabilidade para os resultados que serão obtidos. Dessa forma, o próximo passo com relação aos dados orientou-se para a análise dos *outliers* presentes nos pontos do intervalo escolhido do poço. Para o projeto, foi escolhido o Método de Desvio Padrão, também conhecido como a regra dos 3 sigmas.

Trata-se de uma técnica empírica de análise de dados que filtra os *outliers* usando intervalos baseados nos valores de desvio padrão e da média de uma série de dados (SEO, 2006). Tem-se que, do conjunto de dados, é retirada a média total e, a partir dela e do desvio padrão, encontra-se os intervalos de distribuição para 3 desvios (σ) de cada lado da média (μ), ou seja, ao somar ou diminuir os desvios padrões, como mostra a figura 27.

Figura 27 – Representação da distribuição para a Regra dos 3 Sigmas

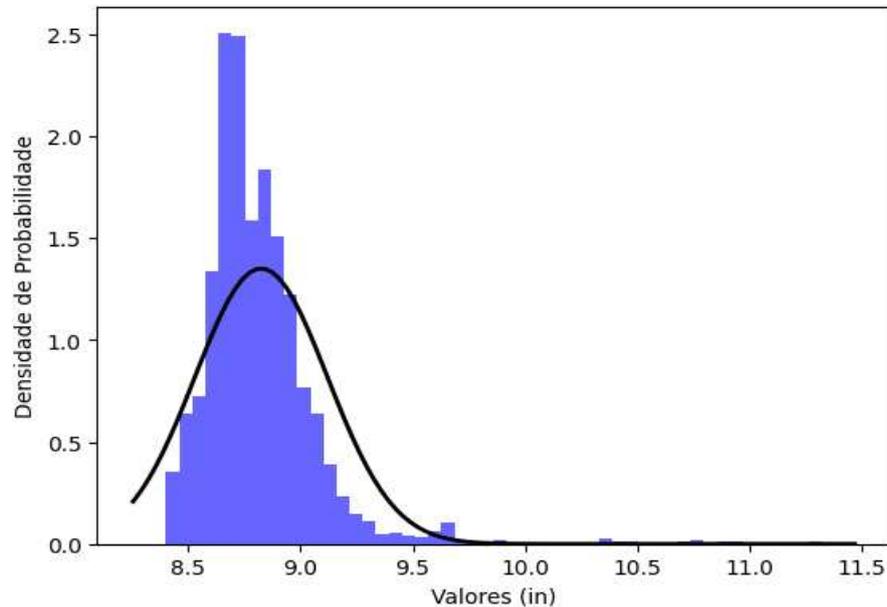


Fonte: Data Science Time, 2020

Esta regra estatística estima que, para um conjunto de dados normalmente distribuídos, aproximadamente 99,7% dos dados estão dentro do intervalo de até três desvios padrões além da média. A regra também é conhecida como 68–95–99,7, devido à distribuição dos dados estar dentro da faixa de 1, 2 e 3 desvios da média, respectivamente (Ross, 2004).

A utilização desse método de visualização e retirada de *outliers* se deu por conta dos dados do perfil de cáliper do intervalo definido para o poço de treino (1100 a 1500 metros) seguirem, aproximadamente, um padrão de distribuição normal, como mostra o gráfico 2, ou seja, a maioria dos valores segue próximo ao valor da média, o que seria representado pelo diâmetro da broca, e os demais dados seguirem dentro das faixas dos desvios padrões da média.

Gráfico 2 – Histograma e Curva de Distribuição Normal dos dados definidos do Perfil Cáliper do poço de treino



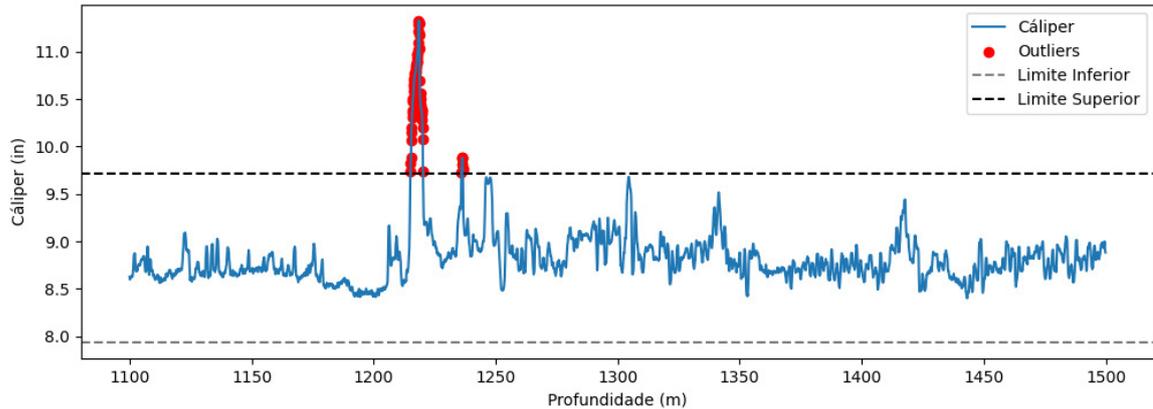
Fonte: Autor (2024)

Apenas o perfil cáliper foi analisado e modelado diante do método de desvio padrão, em razão de que se trata de um registro usado na identificação de informações pouco confiáveis (cáliper anormal, para mais ou menos, em geral, indica dados pouco confiáveis).

Após a observação visual da distribuição dos dados, tem-se que os valores fora do intervalo representado por 3 desvios padrões da média são considerados *outliers* e, dessa forma, foram retirados do conjunto de dados que será informado à máquina para realizar o aprendizado. Isso foi feito com o intuito de evitar erros de interpretação do modelo, garantir confiabilidade e precisão nos resultados, além de prevenir distorções e enviesamento.

Dessa forma, a quantidade de dados do perfil de treino, que antes era de 4.000 linhas de dados, passa a ter 3.942, já que 58 pontos do perfil cáliper se tratavam de *outliers*. Todos estes pontos, inclusive, eram valores maiores que 3 desvios padrões somados à média, como mostra a figura 28.

Figura 28 – Identificação dos *outliers* nos dados do Perfil de Cáliper do Poço de Treino



Fonte: Autor (2024)

4.3 Aprendizado Não-Supervisionado

Posterior à etapa de processamento e tratamento dos dados, tem-se a aplicação da técnica de agrupamento, o algoritmo *k-means*, a fim de organizar as observações amostrais de modo a formar grupos com alta similaridade entre os dados dentro de cada *cluster*, enquanto há diferenças significativas entre os outros *clusters*.

No contexto de dados geológicos, em que a localização espacial é crucial, é fundamental que os grupos identificados apresentem uma correlação espacial coerente, o que contribui para uma interpretação geológica adequada. Nesse caso, devido à proximidade espacial entre os poços selecionados para o estudo, o sentido geológico é mantido.

Dessa forma, o aprendizado não-supervisionado através do algoritmo *k-means* é empregado para criar agrupamentos que não apenas refletem a similaridade dos dados dos poços, mas também mantêm a relevância geológica ao considerar a distribuição espacial das observações.

Concatenação dos dados

Para encaminhar os dados para que o modelo realize o agrupamento correto, foi necessário, então, unir os dados com as informações do poço de treino e do poço-objetivo, o qual será o foco da análise da predição das eletrofácies pelo algoritmo de *machine learning*. Para o estudo, a divisão dos dados se aproximou de 80% das informações advindas do poço denominado de treino (Poço 4-BRSA-924-RN) e o restante, 20%, proveniente do poço de teste (Poço 4-BRSA-423D-RN e Poço 4-BRSA-937D-RN).

Após a retirada dos pontos *outliers* do poço de treino, a razão de dados de treino e teste passou de 4.000/1.000, para 3.942/1000, o que configura, de forma aproximada, 79,8% e 20,2% de dados de treino e teste, respectivamente. Essa proporção de divisão permite que o algoritmo seja executado com eficiência (Huč *et al.*, 2021).

A concatenação dos dados, entretanto, necessitou de um pequeno passo anterior, que se trata da identificação dos dados, distinguidos em treino e teste. Para isso, uma nova coluna foi adicionada em ambos os *DataFrames* com os dados de treino e teste. Nesta coluna foram atribuídos valores diferentes para cada conjunto de dados, por exemplo, no *DataFrame* com os dados do poço de treino, a coluna, que foi chamada ‘POÇO’, possui, em suas linhas, a identificação ‘Treino’, para que assim esses dados sejam diferenciados das outras informações do conjunto de teste. Portanto, a etapa de identificação dos dados antes da concatenação facilitou a manipulação e avaliação precisa dos resultados que serão gerados pelo modelo no agrupamento.

Após a realização da concatenação, o conjunto de dados contém informações de Cáliper, Gamma-Ray, Densidade, Neutrão, Tempo de Trânsito (Sônico), Resistividade e a Identificação do tipo de poço (Treino e Teste). Por outro lado, nem todas essas observações são adequadas para utilizar no aprendizado, já que a máquina tenta agrupar diferentes informações dos perfis em grupos (Abri *et al.*, 2020), o que levou ao descarte dos dados sobre a identificação entre treino e teste nesta etapa, pois serve apenas para distinguir o conjunto de dados. Além disso, o perfil de Cáliper, por se tratar de um registro para realizar controle de qualidade dos dados, também não foi usado no aprendizado, uma vez que a distribuição dos dados do perfil, em sua maioria, não afeta ou tem ligação com variações nas curvas dos outros perfis.

Dessa forma, um novo *DataFrame* foi criado, apenas com os perfis de Gamma-Ray, Densidade, Neutrão, Tempo de Trânsito (Sônico) e Resistividade. Este será informado à máquina para que o modelo realize a clusterização e forneça as informações de eletrofácies, que são a representação dos grupos gerados pelo algoritmo *k-means*.

Implementação da Normalização dos Dados

Ainda na etapa de processamento dos dados para a aplicação do algoritmo de clusterização, o conjunto de dados presente no *DataFrame* possui informações em diferentes grandezas, como o perfil de densidade, o qual pode variar de 1,5 g/cm³ a 3 g/cm³, e o perfil de resistividade, variando entre 0,1 ohm.m a 100 ohm.m. Estas diferenças entre as escalas dos perfis podem prejudicar o desempenho e estabilidade do modelo (Ambarwari *et al.*,

2020). Dessa forma, a etapa de normalização dos dados dos perfis foi adotada devido a sua importância em garantir que todas as informações tenham igual contribuição no aprendizado, além de trazer melhoria na qualidade e precisão dos resultados, sobretudo por se tratar de um algoritmo baseado nas distâncias entre os pontos (Starovoitov; Golub, 2021).

Para o projeto, a técnica de padronização de dados utilizada foi o *StandardScaler*, disponível dentro da biblioteca *scikit-learn* do Python. O método foi importado para o ambiente do Jupyter Notebook por meio do código e usado para normalizar o conjunto de dados dos perfis. A técnica padroniza as características por meio da remoção da média ($\mu = 0$) e escalonamento para a variância unitária ($\sigma = 1$) dos dados (“StandardScaler”, 2024), conforme a fórmula abaixo:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Em que:

- z – Dado normalizado;
- x – Dado original;
- μ – Média dos dados;
- σ – Desvio padrão dos dados.

O método de padronização *StandardScaler* foi usado nesta etapa em razão de preservar a estrutura estatística dos dados originais, pois não altera a forma da distribuição dos mesmos. Além disso, a normalização das informações traz consigo uma convergência mais expedita para os algoritmos de aprendizado de máquina, como o *k-means* (Punia *et al.*, 2021).

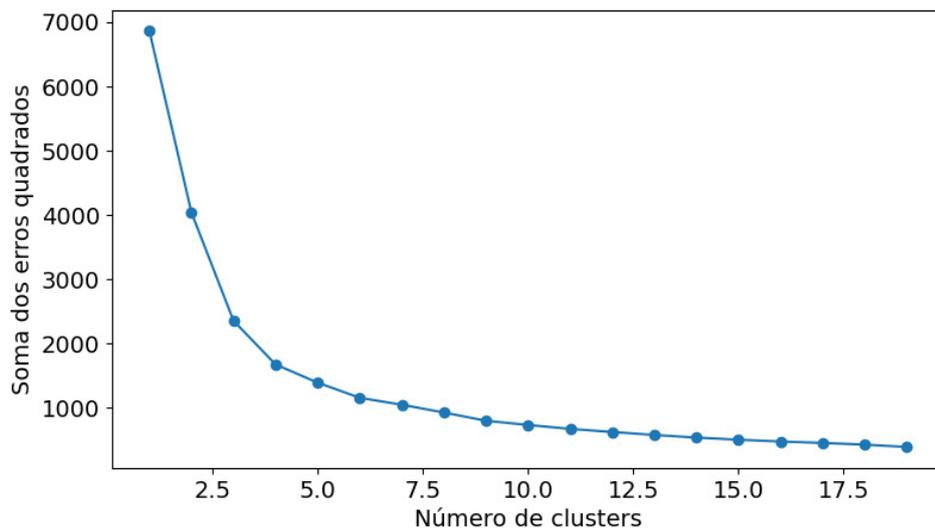
Número de *clusters* ideal

Após realizar a normalização dos dados, a próxima etapa tratou-se da definição do número ideal de grupos (*clusters*) que serão adotados pelo algoritmo no momento de aplicar o agrupamento. Para determinar a quantidade adequada, o método do cotovelo (*Elbow Method*), o qual já foi citado no tópico 3.1.3, foi utilizado para avaliar a qualidade dos agrupamentos e escolher o número ótimo de *clusters*, a fim de evitar grupos muito amplos e perder detalhes importantes ou levar ao *overfitting* (ajuste excessivo do modelo para os dados de treinamento e

consequentes previsões precisas para estes dados, porém apresenta queda na capacidade de fazer previsões para novos dados) e agrupamentos insignificantes.

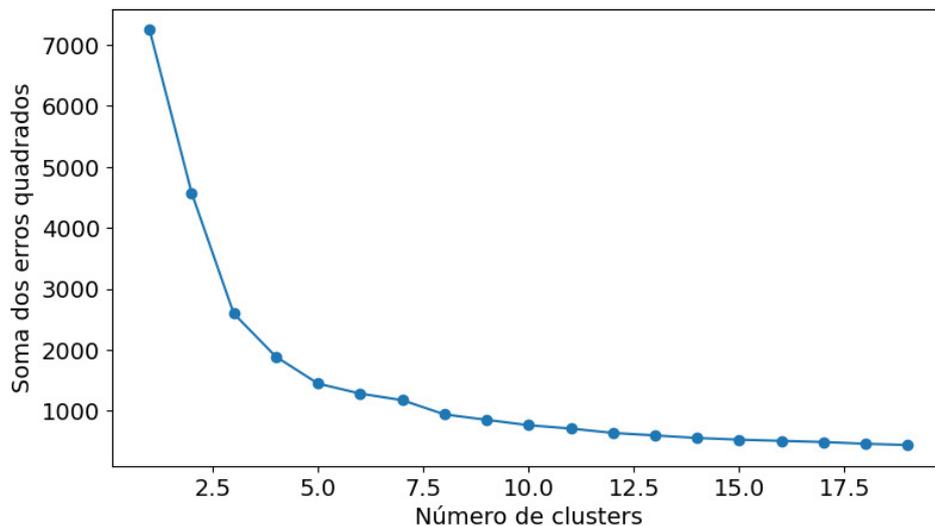
O método do cotovelo é avaliado graficamente ao observar o comportamento da inércia, que representa a soma dos quadrados das distâncias intra-cluster, em relação ao número de *clusters*. Para o projeto, foram analisados os dois gráficos para determinação do número de *clusters* das duas previsões, tanto para o poço 4-BRSA-423D-RN quanto para o poço 4-BRSA-937D-RN. Os gráficos 3 e 4 mostram o comportamento e a identificação visual dos pontos de inflexão das curvas da soma dos erros quadrados que representam o número ideal de grupos para cada poço-objetivo.

Gráfico 3 – Curva da soma dos erros quadrados de acordo com o número de clusters para o caso com o poço 4-BRSA-423D-RN



Fonte: Autor (2024)

Gráfico 4 – Curva da soma dos erros quadrados de acordo com o número de clusters para o caso com o poço 4-BRSA-937D-RN



Fonte: Autor (2024)

Pelos gráficos, foi possível observar que, a partir de 3 ou 4 *clusters*, o comportamento da curva se mantém aproximadamente constante e com pouca redução da soma dos erros quadrados ao aumentar o número de grupos. Dessa forma, foram definidos 3 *clusters* para realizar uma aplicação inicial do modelo sobre os dados. Esta escolha também leva em consideração a leitura dos perfis compostos dos poços-objetivos, uma vez que apresentam, nas faixas de profundidade selecionadas, apenas 2 eletrofácies dominantes, arenitos e folhelhos. O outro grupo, neste caso, seria destinado à ocorrência de possíveis informações divergentes que o modelo possa encontrar e dividir o conjunto de dados em mais um agrupamento.

Aplicação do algoritmo e Comparação dos resultados observados

Finalmente, a clusterização pelo algoritmo *k-means* foi realizada de acordo com o ajuste para o número de grupos definidos na etapa anterior. Os *clusters* formados para cada caso representam as eletrofácies preditas pelos perfis informados e utilizados no aprendizado. Os resultados do agrupamento são, então, recolhidos e comparados com os dados reais observados dos intervalos de profundidade definidos para os poços-objetivos. Ao comparar as respostas do algoritmo com os dados reais, é possível encontrar quaisquer diferenças e fazer alterações no modelo para que funcione melhor.

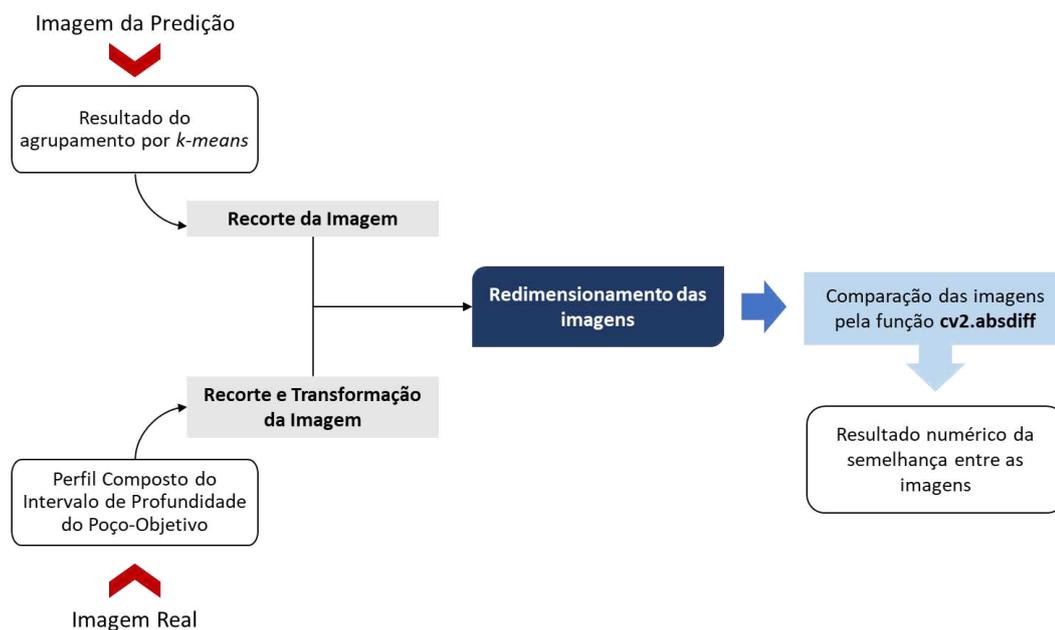
A análise comparativa entre os resultados obtidos pelo algoritmo e os dados reais foi realizada com uma função da biblioteca OpenCV, em Python, a qual realiza a comparação entre as imagens pixel por pixel. A utilização dessa função permitiu mensurar, com exatidão, as

diferenças presentes entre as imagens geradas pelo modelo e as imagens reais retiradas do perfil composto dos poços.

O método de comparação pixel a pixel consiste na verificação de cada ponto correspondente às duas imagens, realizando o cálculo da disparidade entre os valores de cor e intensidade de cada pixel. A utilização desse método junto à biblioteca OpenCV possibilitou o cálculo e uma avaliação detalhada e precisa das diferenças presentes nas imagens, proporcionando uma clareza referente ao local e à forma com que o modelo está divergindo da realidade.

A metodologia da etapa de comparação seguiu um fluxograma de trabalho específico, uma vez que diferentes tarefas são realizadas até que o resultado numérico seja obtido. A figura 29 mostra a sequência de passos realizados para alcançar o objetivo.

Figura 29 – Fluxograma de trabalho para obtenção do valor de semelhança entre a imagem gerada pelo modelo e a imagem real do poço

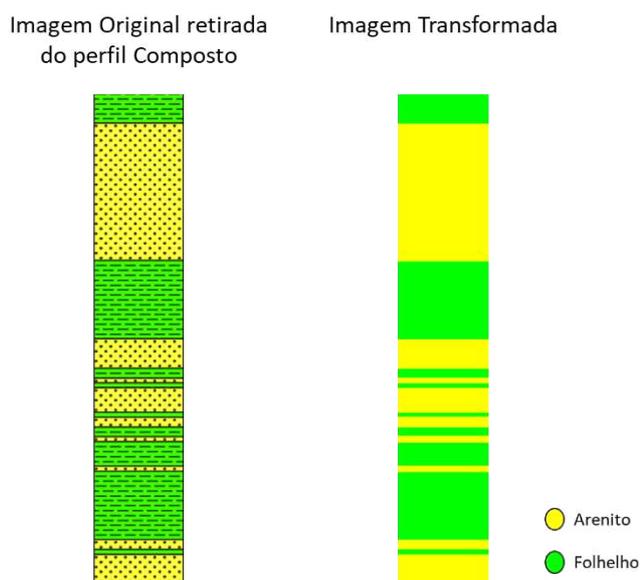


Fonte: Autor (2024)

Enquanto que a imagem gerada pela predição necessitou apenas de um recorte inicial, já que além da informação de eletrofácies, também apresenta as curvas dos perfis, a ilustração obtida pelo perfil composto também precisou ser transformada para o mesmo padrão da imagem do modelo. Esta transformação tratou-se da correção da representação visual de cores das eletrofácies da imagem para a mesma representação da imagem gerada pelo algoritmo.

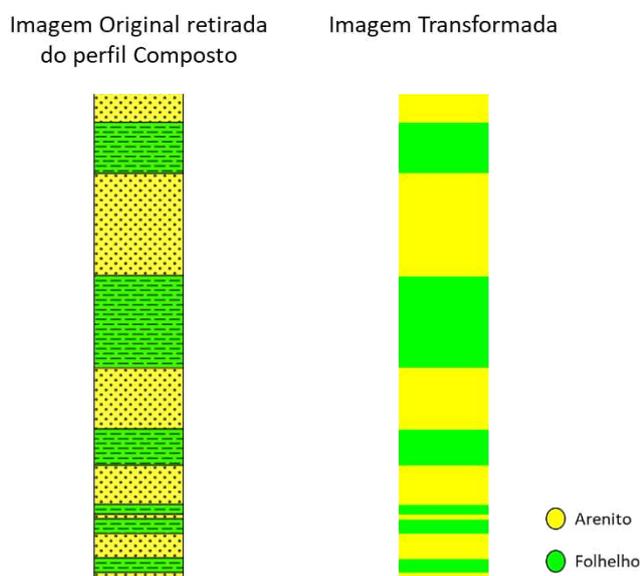
Essa mudança foi realizada por meio do aplicativo de edição e criação de imagens Paint, disponível no Windows. As figuras 30 e 31 mostram a transformação das imagens originais dos dois poços-objetivos para aquelas que serão adotadas corretamente na comparação.

Figura 30 – Ilustração da correção da imagem retirada do perfil composto para o padrão da imagem do modelo (Poço 4-BRSA-423D-RN: 1450 – 1550 metros)



Fonte: Autor (2024)

Figura 31 – Ilustração da correção da imagem retirada do perfil composto para o padrão da imagem do modelo (Poço 4-BRSA-937D-RN: 2200 – 2300 metros)



Fonte: Autor (2024)

A partir das imagens transformadas, foi necessário redimensionar as imagens para que apresentem a mesma dimensão, em pixels, já que a função `cv2.absdiff` somente aceita comparações com imagens de mesma altura e largura. Para isso, o próprio aplicativo Paint também realizou esta tarefa de edição de pixels. Por fim, foi realizada a comparação entre as figuras e obtido o valor representativo da semelhança, numa escala que vai de 0 a 100%.

5. RESULTADOS E DISCUSSÕES

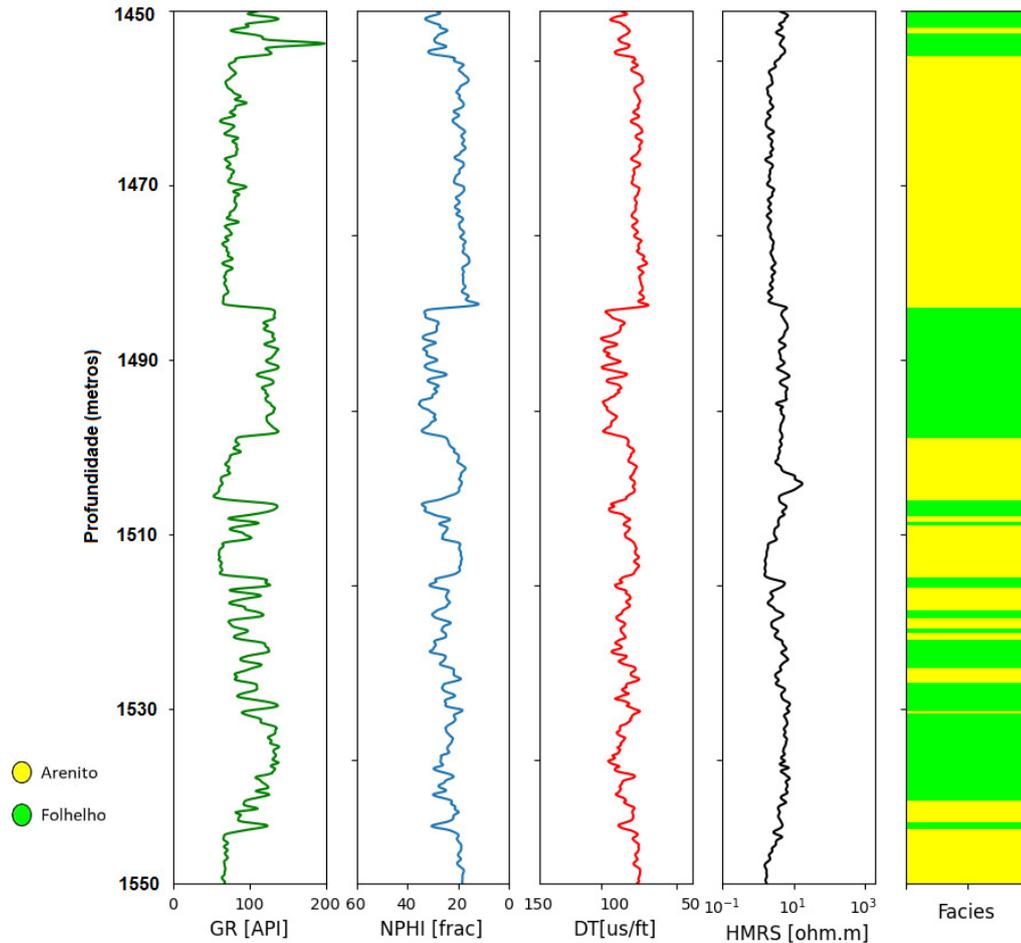
Neste tópico serão apresentados os resultados obtidos na predição de eletrofácies para os dois poços utilizados no projeto (Poço 4-BRSA-423D-RN e Poço 4-BRSA-937D-RN), destacando a precisão e a acurácia do modelo desenvolvido. Além disso, para cada poço, foi realizada a comparação entre os perfis de eletrofácies do perfil composto do poço e do perfil gerado a partir do algoritmo de clusterização, o que permitiu efetuar uma avaliação mais detalhada da performance do modelo.

5.1 Resultado das predições

As simulações feitas com o algoritmo *K-Means* fornecem os grupos divididos de acordo com os identificadores ou rótulos (*labels*) atribuídos a cada ponto de dado. Para o caso do projeto, a identificação dos *clusters* se refere ao tipo de eletrofície detectada, seja arenito, folhelho ou outra possível formação geológica. A identificação do cluster para cada linha de dado referente a profundidade gerou o perfil de eletrofácies completo para os poços analisados, como mostram as figuras 32 e 33.

As duas figuras apresentam os resultados obtidos da predição, além de exibir alguns dos outros perfis que foram utilizados no aprendizado, como gamma-ray, neutrão, sônico e resistividade.

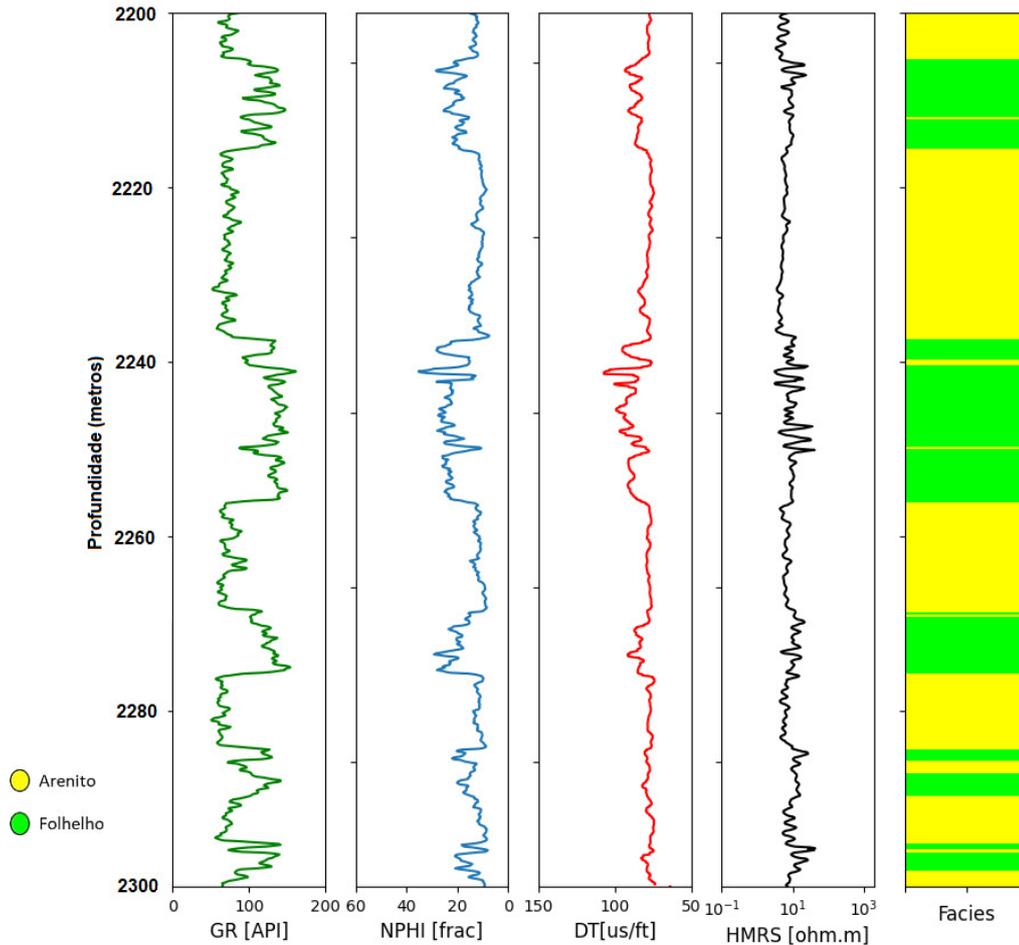
Figura 32 – Resultado da Predição do Perfil de Eletrofácies do Poço 4-BRSA-423D-RN
(1450 – 1550 metros)



Fonte: Autor (2024)

A análise visual revela que os resultados da predição seguem um padrão de características muito próximo das curvas dos perfis reais dos poços, sobretudo do perfil de gamma-ray. Este perfil, que mede a radioatividade natural das formações rochosas, é particularmente importante na identificação de eletrofácies devido à sua sensibilidade às variações litológicas (Thibal *et al.*, 1999). Dessa forma, a correspondência alta com o perfil de raios gama indica que o modelo conseguiu capturar precisamente as intercalações da litologia.

Figura 33 – Resultado da Predição do Perfil de Eletrofácies do Poço 4-BRSA-937D-RN (2200 – 2300 metros)



Fonte: Autor (2024)

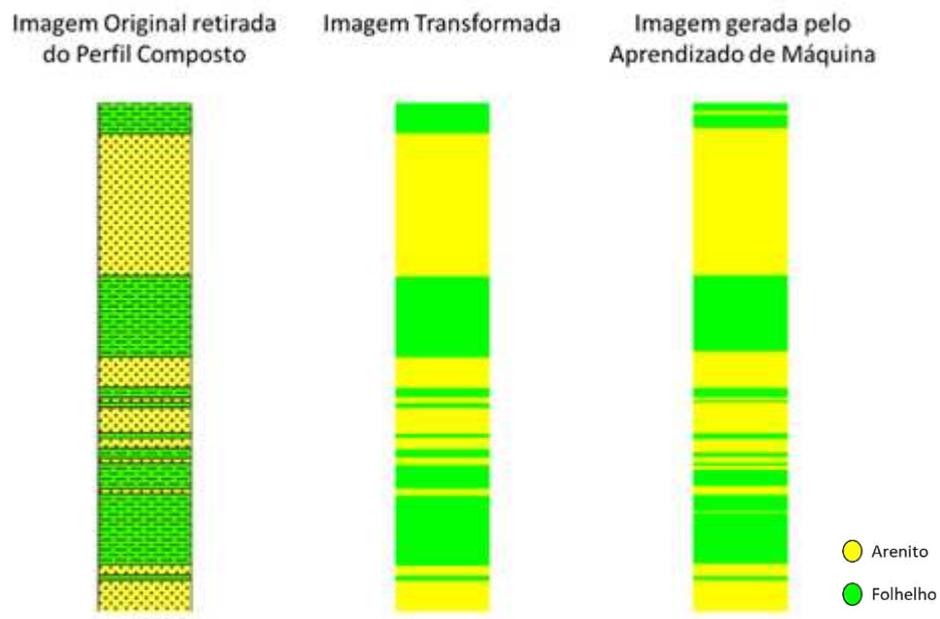
Além do perfil de gamma-ray, os outros perfis observados nas figuras também mostram uma boa concordância entre as características reais e a representação da predição. É possível observar que as variações nas curvas em determinados intervalos apresentam transformações correspondentes na identificação de eletrofácies no perfil gerado pelo modelo. O perfil de neutrão, que mede a porosidade da formação, e o perfil sônico, que avalia a velocidade de propagação das ondas sonoras nas rochas, ambos confirmam a consistência da predição das eletrofácies com as propriedades petrofísicas observadas. O perfil de resistividade, que reflete a resistência elétrica das formações, complementa a análise ao confirmar as zonas de maior e menor saturação de fluidos.

5.2 Comparações entre as eletrofácies

Após trazer os resultados da predição e realizar a análise inicial junto com os outros perfis de poço, foi feita a comparação visual entre as figuras preditas e as figuras reais retiradas

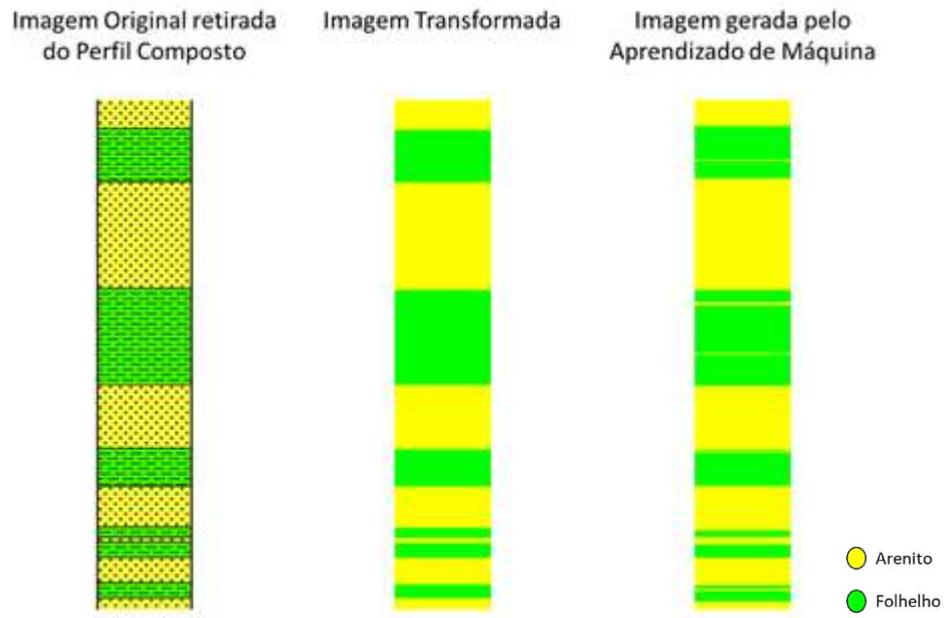
dos perfis compostos dos poços. Foi possível verificar que as previsões seguem um padrão de características muito próximo do perfil composto real, como mostram as figuras 34 e 35. Esta concordância visual é um indicativo da eficácia e precisão do modelo utilizado

Figura 34 – Comparação entre os Perfis de Eletrofácies do Perfil Composto e do Aprendizado de Máquina do Poço 4-BRSA-423D-RN (1450 – 1550 metros)



Fonte: Autor (2024)

Figura 35 – Comparação entre os Perfis de Eletrofácies do Perfil Composto e do Aprendizado de Máquina do Poço 4-BRSA-937D-RN (2200 – 2300 metros)



Fonte: Autor (2024)

Além da análise visual, a comparação também foi quantificada numericamente por meio da aplicação da função `cv2.absdiff` da biblioteca OpenCV. A semelhança entre as imagens dos perfis de eletrofácies preditos e reais foi calculada para os casos dos dois poços, resultando em valores superiores a 97%, como mostra a tabela 4. Este alto grau de similaridade confirma que o modelo conseguiu capturar com precisão as características das eletrofácies presentes nos dados reais, validando a metodologia empregada no projeto.

Tabela 4 – Resultado numérico da comparação entre as imagens reais e preditas

Poço	Intervalo de Profundidade (Metros)	Semelhança entre as imagens (%)
4-BRSA-423D-RN	1450 - 1550	97,22
4-BRSA-937D-RN	2200 - 2300	98,13

Fonte: Autor, 2024

A utilização da ferramenta para o cálculo da diferença absoluta entre as duas imagens pixel a pixel foi fundamental para obter um resultado numérico confiável, complementando a análise visual e corroborando a alta acurácia da predição das eletrofácies.

6. CONCLUSÃO

Este trabalho teve por objetivo desenvolver e aplicar o algoritmo de aprendizado de máquina *K-means* na caracterização de eletrofácies de poços de petróleo da Bacia Potiguar, a fim de implementar uma alternativa eficiente e inovadora na aquisição de informações geofísicas dos reservatórios. Para a construção do projeto, a metodologia adotada foi de fundamental importância, abrangendo diversas etapas e a aplicação de métodos e técnicas específicas para se alcançar um modelo ideal para este caso.

O processo iniciou-se com a escolha e tratamento dos dados, garantindo a qualidade e consistência das informações utilizadas no aprendizado de máquina. Em seguida, foram definidas as características mais relevantes para o modelo, como a escolha dos perfis de poços, que orientaram a clusterização das eletrofácies. Por fim, a etapa de leitura e comparação dos resultados foi realizada por meio da análise de imagens, permitindo validar a eficácia do modelo ao comparar as eletrofácies preditas com as observadas.

Com relação aos resultados obtidos na predição, é possível resumi-los de acordo com as seguintes considerações:

- Alta correspondência entre as eletrofácies preditas com as características observadas nos outros perfis de poço, em especial o perfil de gamma-ray;
- Semelhança elevada entre as eletrofácies reais e preditas dos dois casos (97,22 e 98,13%) quando realizada a comparação de pixels entre as imagens.

Apesar da elevada similaridade obtida, alguns desafios foram encontrados durante o desenvolvimento do projeto, em especial a identificação e tratamento adequado dos dados *outliers*. Esta etapa foi verificada inicialmente comparando dois métodos de identificação destes valores atípicos: Método do *Box Plot* e a Regra dos 3 sigmas (Utilizada no projeto). Ao aplicar o *box plot* nos dados do “poço de treino”, foi possível perceber uma quantidade considerável de valores considerados outliers, o que acabou por afetar o resultado das predições das eletrofácies. Por outro lado, a regra dos 3 sigmas apresentou menos outliers e alcançou melhores respostas ao final do projeto.

Como limitação do projeto, o tamanho do intervalo de profundidade estudado foi um fator a ser considerado. Além dos motivos já apresentados da definição das profundidades de estudo dos poços, o ponto observado foi o tamanho do intervalo, uma vez que, em testes iniciais, ao utilizar o poço completo ou recortes maiores (1000 ou 2000 metros), a semelhança

entre a predição e o real apresentou menores porcentagens, além da transformação da imagem do perfil composto real apresentar maiores dificuldades e se tornar mais complexa. Dessa forma, foram utilizados intervalos pequenos, mas com relativa importância e que forneciam visualizações mais precisas e maiores capacidades de comparação.

Para estudos posteriores, é possível implementar outros algoritmos de aprendizado não-supervisionado, ou até o uso de redes neurais, para realizar a comparação com o método *K-means*, a fim de encontrar melhores resultados.

Por fim, após a apresentação destas conclusões, os resultados promissores do projeto confirmam a eficiência e precisão da aplicação de modelos computacionais para processos e estudos importantes na área de reservatórios. Além disso, a capacidade do algoritmo em representar a realidade da formação do poço mostra o potencial de integração dessas tecnologias com métodos ou softwares tradicionais com o objetivo de melhorar cada vez mais a compreensão dos reservatórios e otimizar alguns dos processos e operações de campo, como a medição dos perfis geofísicos.

REFERÊNCIAS

ABRI, S.; ABRI, R; ÇETIN, S. Group-based Personalization Using Topical User Profile. Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization. New York, NY, USA: ACM, 2020.

ALMEIDA, M. **Machine Learning: o que é aprendizado semi-supervisionado**. Disponível em: <https://www.alura.com.br/artigos/machine-learning-aprendizado-semi-supervisionado> . Acesso em: 15 fev. 2024.

ALVAREZ, G.; SANSÓ, B.; MICHELENA, R.; JIMENEZ, J. Lithologic characterization of a reservoir using continuous-wavelet transforms. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, n. 1, p. 59–65, 1 jan. 2003.

AMBARWARI, A.; JAFAR ADRIAN, Q.; HERDIYENI, Y. Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification. **Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)**, v. 4, n. 1, p. 117–122, 9 fev. 2020.

ANGELIM, L.A.A.; MEDEIROS, V.C.; NESI, J.R. Programa Geologia do Brasil –PGB. Projeto Geologia e Recursos Minerais do Estado do Rio Grande do Norte. Mapa geológico do Estado do Rio Grande do Norte. Escala. 1:500.000. Recife: CPRM/FAPERN, 2006. 1mapa color.

ASSAIFE, R. F. Perfilagem durante a perfuração de poços de petróleo: estudo bibliográfico. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, p. 51–80, 31 mar. 2022.

AZANK, F. **Clustering — Conceitos básicos, principais algoritmos e aplicação**. Disponível em: <https://medium.com/turing-talks/clustering-conceitos-b%C3%A1sicos-principais-algoritmos-e-aplica%C3%A7%C3%A3o-ace572a062a9>. Acesso em: 14 fev. 2024.

BAKER, R. O.; YARRANTON, H. W.; JENSEN, J. **Practical Reservoir Engineering and Characterization**. [s.l.] Gulf Professional Publishing, 2015.

BARBOSA, S.G. **Bacia Potiguar**. Décima primeira rodada de licitações. Superintendência de Definição de Blocos, Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, 2013. Disponível em: https://www.gov.br/anp/pt-br/rodadas-anp/rodadas-concluidas/concessao-de-blocos-exploratorios/11a-rodada-licitacoes-blocos/arquivos/seminarios/bacia_potiguar.pdf. Acesso em: 23 mar. 2024.

BELYADI, H.; HAGHIGHAT, A. **Machine Learning Guide for Oil and Gas Using Python**. [s.l.] Gulf Professional Publishing, 2021.

BERTANI, R.T.; COSTA, I.G.; MATOS, R.M.D. Evolução tectono-sedimentar, estilo estrutural e hábitat do petróleo na Bacia Potiguar. In: RAJA GABAGLIA, G.P.; MILANI, E.J. (Coord.) **Origem e evolução de bacias sedimentares**. Rio de Janeiro: Petrobras. SEREC.CENSUD, 1990. p. 291-310.

BHOLOWALIA, P.; KUMAR, A. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. **International Journal of Computer Applications**, v. 105, n.

9, p. 17–24, 18 nov. 2014.

BRUCE, R. F. A Bayesian Approach to Semi-Supervised Learning. *NLPRS*, p. 57–64, 1 jan. 2001.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. **CRISP-DM 1.0: Step-by-step data mining guide**. SPSS, 2000.

CLOSE, D.I.; CHO, D.; HORN, F.; EDMUNDSON, H. The Sound of Sonic: A Historical Perspective and Introduction to Acoustic Logging. *CSEG RECORDER*, 2009.

DATA SCIENCE TEAM. **Calculadora de Regra Empírica — Matemática e Estatística — DATA SCIENCE**. Disponível em: <https://datascience.eu/pt/matematica-e-estatistica/calculadora-de-regra-empirica/>. Acesso em: 19 abr. 2024.

EL NAQA, I.; MURPHY, M. J. What Is Machine Learning? **Machine Learning in Radiation Oncology**, v. 1, n. 1, p. 3–11, 2015.

ELLIS, D. V. Gamma ray scattering measurements for density and lithology determination. *IEEE Transactions on Nuclear Science*, v. 35, n. 1, p. 806–811, 1988.

ELLIS, D. V.; SINGER, J. M. **Well Logging for Earth Scientists**. [s.l.] Springer, 2009.

FAIZ, S. **DLIS File Format - Well Log Data File**. Disponível em: <https://docs.fileformat.com/database/dlis/>. Acesso em: 22 maio 2024.

FEMENICK, T.R. Uma história do petróleo potiguar. *Jornal Tribuna do Norte*, Natal, 20 abr. 2014.

GEETHA, T. V.; SENDHILKUMAR, S. **Machine learning: concepts, techniques and applications**. 1. ed. Nova Iorque: Chapman Hall/CRC Press, 2023.

GOMES, P. C. T. **Clustering: O que é Cluster Analysis e quais suas Aplicações? - DataGeeks**. Disponível em: <https://www.datageeks.com.br/clustering/>. Acesso em: 21 jul. 2024.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. San Francisco: Morgan Kaufmann, 2001.

HUANG, Z. THE ANALYSIS AND DECONVOLUTION CORRECTION OF NATURAL GAMMA-RAY LOGS. *Chinese Journal of Geophysics*, 1 jan. 1983.

HUČ, A.; ŠALEJ, J.; TREBAR, M. Analysis of Machine Learning Algorithms for Anomaly Detection on Edge Devices. *Sensors*, v. 21, n. 14, p. 4946, 20 jul. 2021.

HURLEY, P. M. DIRECT RADIOMETRIC MEASUREMENT BY GAMMA-RAY SCINTILLATION SPECTROMETER. *Geological Society of America Bulletin*, v. 67, n. 4, p. 405-412, 1956.

INSIGHT, GEOLOGIA DO PETRÓLEO LTDA. **Petroleum systems of the Potiguar Basin**. 295 p. Relatório final da avaliação dos sistemas petrolíferos da bacia Potiguar em atendimento ao Ofício da ANP nº 150/SDT/2005, 2005.

- JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, jun. 2010.
- KEAREY, P.; BROOKS, M.; HILL, I. **Geofísica de exploração**. São Paulo: Oficina De Textos, 438 p., 2009.
- LACHI, R. L.; ROCHA, H. V. d. Aspectos básicos de clustering: conceitos e técnicas. [S.l.], 2005.
- LI, Y.; WU, H. A Clustering Method Based on K-Means Algorithm. **Physics Procedia**, v. 25, p. 1104–1109, 2012.
- MOHAMMED, M.; KHAN, M. B.; BASHIER, E. B. M. Machine Learning. CRC Press, 19 ago. 2016.
- MORELATTO, R.; FABIANOVICZ, R. **Sumário Geológico e Setores em Oferta - BACIA POTIGUAR**. Superintendência de Definição de Blocos, SDB; Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Décima Terceira Rodada De Licitações, Rio de Janeiro, 2015.
- NAINGGOLAN, R. et al. Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. **Journal of Physics: Conference Series**, v. 1361, p. 012015, nov. 2019.
- NAZEER, A.; ABBASI, S. A.; SOLANGI, S. H. Sedimentary facies interpretation of Gamma Ray (GR) log as basic well logs in Central and Lower Indus Basin of Pakistan. **Geodesy and Geodynamics**, v. 7, n. 6, p. 432–443, 1 nov. 2016.
- NERY, G. G. Perfilagem Geofísica em Poço Aberto. Salvador, BA, 231p., 1990.
- OLIVEIRA, T. B. S. d. Clusterização de dados utilizando técnicas de redes complexas e computação bioinspirada. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação, 2008.
- PEDROSA JR., N. C.; CASTRO, D. L. D.; MATOS, J. P. L. D. Assinaturas magnéticas e gravimétricas do arcabouço estrutural da bacia Potiguar emersa, NE do Brasil. **Revista Brasileira de Geofísica**, v. 28, n. 2, p. 265-278, 2010.
- PESSOA NETO, O.C.; SOARES, U.M.; SILVA, J.G.F.; ROESNER, E.H.; FLORENCIO, C.P.; SOUZA, C.A.V. Bacia Potiguar. **Boletim de Geociências da Petrobras**, v. 15, n. 2, p. 357-369, maio/nov. 2007.
- PORTELLA, A. Y.; FABIANOVICZ, R. **Bacia Potiguar – Sumário Geológico e Setores em Oferta**. ANP. Rio de Janeiro. 2017.
- PUNIA, S. K. et al. Performance Analysis of Machine Learning Algorithms for Big Data Classification. **International Journal of E-Health and Medical Communications**, v. 12, n. 4, p. 60–75, jul. 2021.
- RASCHKA, S.; MIRJALILI, V. **Python machine learning : machine learning and deep learning with Python, scikit-learn, and TensorFlow**. Birmingham (Uk): Packt Publishing, 2017.

RIDER, M. **The geological interpretation of well logs**. Sutherland: Rider-French Consulting. Copyright, 2002.

ROSS, S. M. **Introduction to probability and statistics for engineers and scientists**. [S.l.] Elsevier, p. 32-33, 2004.

RUSSELL, W. L. THE TOTAL GAMMA RAY ACTIVITY OF SEDIMENTARY ROCKS AS INDICATED BY GEIGER COUNTER DETERMINATIONS. **GEOPHYSICS**, v. 9, n. 2, p. 180–216, abr. 1944.

SAMPAIO, T. R. **Utilização de LWD na indústria do petróleo**. 2012. 58 f. Trabalho de Conclusão de Curso (Graduação em Geologia) – Instituto de Geociências, Universidade Federal da Bahia, Salvador, 2012.

SANTOS, R. A. Interpretação e modelagem de perfis geofísicos de poço em campos marginais. Trabalho de graduação - Universidade Federal da Bahia, Salvador, BA. 2010.

SARKER, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. **SN Computer Science**, v. 2, n. 3, p. 1–21, 2021.

SEO, S. A review and comparison of methods for detecting outliers in univariate data sets. Dissertação (Mestrado em Ciências) — University of Pittsburgh, Pennsylvania, 2006.

SERRA, O. **Fundamentals of Well-log Interpretation: The interpretation of logging data**. [s.l.] Elsevier Publishing Company, 1984.

SPIKES, K.; GOUVEIA, W.; NASSER, M. Introduction to this special section: Reservoir characterization. **The Leading Edge**, v. 36, n. 11, p. 885–885, 1 nov. 2017.

StandardScaler. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#> . Acesso em: 3 jul. 2024.

STAROVOITOV, V. V.; GOLUB, YU. I. Data normalization in machine learning. **Informatics**, v. 18, n. 3, p. 83–96, 30 set. 2021.

STEVANATO, Ana Carolina Ribeiro e Silva. **Análise petrofísica de reservatório**. Orientação de Alexandre Campana Vidal. Coorientação de Juliana Finoto Bueno. Campinas, SP: [s.n.], 67 f., 2011.

THIBAL, J.; ETCHECOPAR, A.; POZZI, J.; BARTHÈS, V.; POCACHARD, J. Comparison of magnetic and gamma ray logging for correlations in chronology and lithology: example from the Aquitanian Basin (France). **Geophysical Journal International**, v. 137, n. 3, p. 839–846, 1 jun. 1999.

TRAVASSOS, R. M.; SAUDINO, C.; FREITAS, I. A. **Sumário Geológico e Setores em Oferta - Bacia do Potiguar**. ANP. 2021.

ZAHMATKESH, I; KADKHODAIE, A.; SOLEIMANI, B.; AZARPOUR, M. Integration of well log-derived facies and 3D seismic attributes for seismic facies mapping: A case study from mansuri oil field, SW Iran. **Journal of Petroleum Science and Engineering**, v. 202, p. 108563, jul. 2021.