



Universidade Federal do Rio de Janeiro

Centro de Ciências Matemáticas e da Natureza

Instituto de Geociências

Departamento de Geologia

Sarah Beatriz de Oliveira Aleixo

**Aprendizado de Máquina Aplicado à Classificação de Fácies Através de Perfis de Poços
Convencionais e Perfis de Imagem do Campo de Atapu, Bacia de Santos**

Rio de Janeiro

2023

Sarah Beatriz de Oliveira Aleixo

Aprendizado de Máquina Aplicado à Classificação de Fácies Através de Perfis de Poços Convencionais e Perfis de Imagem do Campo de Atapu, Bacia de Santos

Trabalho de Conclusão de Curso em Geologia do Instituto de Geociências, Universidade Federal do Rio de Janeiro - UFRJ, apresentado como requisito necessário para obtenção do título de Bacharel em Geologia.

Orientador: Prof. Dr. Leonardo Fonseca Borghi de Almeida

Coorientador: Msc. Lucas Abreu Blanes de Oliveira

Rio de Janeiro

2023

ALEIXO, Sarah Beatriz de Oliveira

Aprendizado de Máquina Aplicado à Classificação de Fácies Através de Perfis de Poços Convencionais e Perfis de Imagem do Campo de Atapu, Bacia de Santos / Sarah Beatriz de Oliveira

Aleixo - - Rio de Janeiro: UFRJ/IGEO, 2023.

75p.: il.; 30 cm

Trabalho de Conclusão de Curso (Bacharelado em Geologia) - Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Geologia, 2023.

Orientador: Prof. Dr. Leonardo Fonseca Borghi de Almeida

Coorientador: Msc. Lucas Abreu Blanes de Oliveira

1. Geologia. 2. IGEO – Trabalho de Conclusão de Curso.

I. Leonardo Fonseca Borghi de Almeida. II. Universidade Federal do Rio de Janeiro, Instituto de Geociências, Departamento de Geologia. III. Título.

Sarah Beatriz de Oliveira Aleixo

Aprendizado de Máquina Aplicado à Classificação de Fácies Através de Perfis de Poços Convencionais e Perfis de Imagem do Campo de Atapu, Bacia de Santos

Trabalho de Conclusão de Curso em Geologia do Instituto de Geociências, Universidade Federal do Rio de Janeiro - UFRJ, apresentado como requisito necessário para obtenção do título de Bacharel em Geologia.

Aprovado em: 22.12.2023

Por:

Orientador: Prof. Dr. Leonardo Fonseca Borghi de Almeida (UFRJ)

D.Sc. Marco Antonio da Silva Braga (UFRJ)

M.Sc. Vinicius Carneiro Santana (Lagesed)

Rio de Janeiro

2023

Aos meus pais, pelo apoio contínuo ao longo da minha vida.

AGRADECIMENTOS

Primeiramente, gostaria de expressar minha profunda gratidão aos meus orientadores, Lucas Blanes e Leonardo Borghi, pela oportunidade de desenvolver este trabalho. Ambos foram essenciais na minha formação acadêmica e na elaboração deste projeto. Em especial ao Lucas, por toda a paciência, ensinamentos e incentivo ao longo dessa jornada.

Sou extremamente grata à pessoa que despertou meu desejo de estudar geologia, professora Fernanda Amante, do meu antigo e amado Colégio Pedro II. Muito obrigada por me mostrar o caminho das geociências. Agradeço ainda à Universidade Federal do Rio de Janeiro (UFRJ), onde vivi 6 anos marcantes aprendendo cada dia mais sobre o mundo da geologia. Agradeço a todos os professores da Geologia UFRJ, aos motoristas de campo e funcionárias da secretaria que contribuíram para minha formação e crescimento acadêmico. Também agradeço ao Laboratório de Geologia Sedimentar da UFRJ (Lagesed), em especial ao João Paulo e Eduardo Cristo, pelo suporte e disponibilização do software utilizado no trabalho.

Minha sincera gratidão às minhas queridas amigas Evelyn Luna, Luiza Nehrer, Matthielle Lima e Nathália Marconato. São mais de 10 anos de amizade com muitas risadas, companheirismo e apoio investido. Vocês foram essenciais para os momentos de distrações. Sou muito grata por ter vocês ao meu lado por tanto tempo, celebrando e acompanhando cada conquista.

Em seguida, estendo meus agradecimentos aos meus colegas de turma, Sofia d'Orsi, Gabriela Carvalho, Matheus Violante e Amanda Mendes, bem como à minha amiga de curso, Teresa Mourão, pelo apoio inestimável. Sobretudo a Teresa, que sempre me deu os melhores conselhos e foi minha parceira de projeto tanto no Brasil quanto na Noruega. Também não poderia deixar de agradecer meu parceiro Mateus Ferreira, que tanto me ajuda nos momentos mais difíceis da vida. Obrigada por estar ao meu lado. Cada dia tenho mais certeza de que fiz a escolha certa.

Por fim, mas não menos importante, agradeço aos meus pais, Francisca e Sebastião, por todo esforço, dedicação e apoio incansável durante meus 24 anos de vida. Minha mãe é minha maior torcida, e meu pai, o melhor conselheiro. Vocês são minha base, minha estrutura. É por vocês que desejo alcançar as maiores conquistas e os melhores sonhos que poderia sonhar. Este diploma é nosso. Obrigada por tudo.

RESUMO

ALEIXO, Sarah B. O. **Aprendizado de Máquina Aplicado à Classificação de Fácies Através de Perfis de Poços Convencionais e Perfis de Imagem do Campo de Atapu, Bacia de Santos.** 2023, p.75, Trabalho de Conclusão de Curso (Bacharelado em Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Atualmente, o aprendizado de máquina é amplamente empregado no setor de óleo e gás para classificação de fácies. No entanto, essa tarefa é geralmente realizada utilizando os perfis convencionais como variáveis de entrada, os quais nem sempre fornecerão informações úteis no processo de classificação, uma vez que respondem principalmente a variações composicionais das rochas, não abrangendo suas variações texturais e estruturais. Partindo deste princípio, o trabalho teve como objetivo avaliar a qualidade dos modelos de aprendizado de máquina treinados com variáveis de entrada provenientes de perfis convencionais, perfis de imagem e ambos, dentro de um intervalo pré-sal da Bacia de Santos. Para isso, foram utilizados dois algoritmos de aprendizado de máquina (*Self Organizing Maps* – SOM e CatBoost) a fim de garantir que os diferentes resultados obtidos nas análises estivessem relacionados aos dados de entrada (perfis convencionais e de imagem). Tanto os perfis convencionais (raios gama, densidade, fator fotoelétrico, nêutrons, geoquímico, sônico e ressonância magnética nuclear) quanto os perfis de imagem (acústica e resistiva) foram disponibilizados pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) e implementados no software *Interactive Petrophysics* (IP) para serem analisados através do SOM e da Análise Textural de Fácies. A metodologia adotada foi dividida em três etapas principais: (1) classificação das fácies sedimentares utilizando apenas os perfis convencionais de forma não supervisionada (SOM) e supervisionada (SOM e CatBoost); (2) classificação das fácies com base apenas nos perfis de imagem através da Análise Textural de Fácies; (3) análise integrada dos perfis convencionais juntamente com as informações texturais extraídas dos perfis de imagem. Essa última etapa foi realizada exclusivamente de forma supervisionada através do CatBoost. Visando garantir o controle de classificação, todas as etapas levaram em consideração dados de testemunho de rocha descritos por especialistas e cada modelo foi avaliado através de matrizes de confusão a fim de comparar as fácies descritas com as fácies preditas. Como resultado, foram obtidos (a) mapas auto-organizados para as análises supervisionadas e não supervisionadas utilizando o SOM; (b) matrizes de confusão para todas as análises supervisionadas (SOM e CatBoost); (c) perfis de fácies preditas para cada análise. Em geral, as análises supervisionadas através do SOM e do CatBoost utilizando apenas os perfis convencionais apresentaram resultados mais satisfatórios quando comparados às análises supervisionadas e não supervisionadas dos perfis de imagem. Fica evidente, ao final do estudo, que os perfis de imagem não contribuíram significativamente com as técnicas implementadas no trabalho, confundindo o algoritmo e trazendo ruído à classificação. Apesar do aprendizado de máquina auxiliar no desempenho do processo, ainda exige abordagens de processamento mais avançadas capazes de extrair informações relevantes dos perfis de imagem, além de um conhecimento profundo de especialistas.

PALAVRAS-CHAVE: aprendizado de máquina; perfis de poços; perfis de imagem; classificação de fácies; Bacia de Santos.

ABSTRACT

ALEIXO, Sarah B. O. **Machine Learning Applied to Facies Classification Using Conventional Well Logs and Image Logs from the Atapu Field, Santos Basin.** 2023, p.76, 2023, p.75, Trabalho de Conclusão de Curso (Bacharelado em Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Machine learning is currently widely used in the oil and gas industry for facies classification. However, this is usually done using conventional logs as input variables, which will not always provide useful information in the classification process, since they mainly respond to compositional variations in the rocks, and do not cover their textural and structural variations. Based on this principle, the aim of this work was to evaluate the quality of machine learning models trained with input variables from conventional profiles, image profiles and both, within a Pre-Salt interval in the Santos Basin. For this purpose, two machine learning algorithms were used (Self Organizing Maps - SOM and CatBoost) in order to ensure that the different results obtained in the analyses were related to the input data (conventional and image profiles). Both the conventional profiles (gamma rays, density, photoelectric factor, neutrons, geochemical, sonic and nuclear magnetic resonance) and the image profiles (acoustic and resistive) were provided by the National Petroleum, Natural Gas and Biofuels Agency (ANP) and implemented in the Interactive Petrophysics (IP) software to be analyzed using SOM and Facies Textural Analysis. The methodology was divided into three main stages: (1) classification of sedimentary facies using only conventional profiles in an unsupervised (SOM) and supervised method (SOM and CatBoost); (2) classification of facies based only on image profiles using Facies Textural Analysis; (3) integrated analysis of conventional profiles together with textural information extracted from image profiles. This last stage was performed exclusively in a supervised method using CatBoost. In order to ensure classification control, all stages took into account rock sample data described by experts and each model was evaluated using confusion matrices in order to compare the real facies with the predicted facies. As a result, we obtained (a) self-organizing maps for the supervised and unsupervised analyses using the SOM; (b) confusion matrices for all the supervised analyses (SOM and CatBoost); (c) predicted facies profiles for each analysis. In general, the supervised analyses using SOM and CatBoost using only the conventional profiles showed more satisfactory results when compared to the supervised and unsupervised analyses of the image profiles. It is clear at the end of the study that the image profiles did not contribute significantly to the techniques implemented in the work, confusing the algorithm and bringing noise to the classification. Although machine learning helps the performance of the process, it still requires more advanced processing approaches capable of extracting relevant information from the image profiles, as well as in-depth expert knowledge.

KEYWORDS: machine learning; conventional well logs; image logs; facies classification; Santos Basin.

LISTA DE FIGURAS

Figura 1: Mapa de localização da Bacia de Santos. Fonte: Castro (2019).....	17
Figura 2: (A) Gráfico percentual da produção total de petróleo equivalente nas Bacias de Santos, Campos e Solimões. (B) Gráfico percentual da produção total de petróleo equivalente na Bacia de Santos e Campos no Pré-sal. Todos os dados foram adquiridos no mês de junho de 2023. Fonte: modificado de ANP (2023).	18
Figura 3: Carta estratigráfica da Bacia de Santos com sistema petrolífero atuante na seção Pré-sal. Fonte: modificado de Moreira <i>et al.</i> (2007).....	19
Figura 4: (A) Mapa do Brasil em destaque para a Bacia de Santos. (B) Visão aproximada da Bacia de Santos com seus respectivos blocos, campos e parte do polígono do pré-sal. (C) Mapa da área denominada Entorno de Iara, com destaque para o Campo de Atapu e o poço de estudo (9-ATP-1-RJS). Fonte: autora.....	21
Figura 5: Petróleo Equivalente (boe/d) por Campo e Ambiente. Fonte: ANP (2023).....	22
Figura 6: Espectro de radiação gama e as respectivas janelas relacionadas ao Th, U e K. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).....	24
Figura 7: Exemplo de ferramenta de aquisição do perfil de densidade sendo adquirida junto a parede do poço. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).....	25
Figura 8: Representação esquemática do espectro de raios gama lido em três formações compostas por átomos de diferentes números atômicos. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).	26
Figura 9: Classificação dos nêutrons de acordo com suas energias e velocidades. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).....	27
Figura 10: Exemplo de ferramenta de aquisição do perfil de nêutrons, adquirida junto à parede do poço. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).....	28
Figura 11: Gráfico mostrando a relação entre (a) a razão das contagens dos detectores e a porosidade para diferentes tipos de rocha, (b) a razão das contagens dos detectores e a distância de migração e (c) a distância de migração e a porosidade para diferentes tipos de rocha. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).	29
Figura 12: Exemplo de espectro de radiação gama e sua decomposição em espectros elementares de referência para diversos elementos químicos. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).	30
Figura 13: Exemplo de ferramenta de aquisição dos perfis acústicos, sendo adquirida centralizada em relação ao poço. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).	31
Figura 14: Simulação 2D de uma onda acústica se propagando do poço para a formação. Em 90 μ s é possível observar que a frente de onda da formação faz um ângulo de 90° com a parede do poço, formando uma nova frente de onda que se propaga pelo fluido de perfuração do poço com a mesma velocidade da onda na formação. Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).....	32
Figura 15: Perfis de Imagem do poço 9-ATP-1-RJS. (A) Perfil de Imagem acústica. (B) Perfil de Imagem resistiva. Analisando os dois perfis é possível perceber que as ferramentas de imagem acústica retratam 100% da parede do poço, enquanto as ferramentas de imagem resistiva retratam uma porcentagem menor. Fonte: autora.	35

- Figura 16:** Exemplo de um self organizing map de Kohonen, pesquisador responsável pela criação do algoritmo. | Fonte: modificado de Castellani & Castellani (2003). 38
- Figura 17:** A primeira e a segunda árvore normalmente apresentam erros significativos. | Fonte: modificado de CatBoost (2018). 39
- Figura 18:** Repetição da construção das árvores até gerar um modelo adequado. | Fonte: modificado de CatBoost, (2018). 39
- Figura 19:** Histograma com a distribuição de dados para cada fácies. F1 – fácies 1; F2 – fácies 2; F3 – fácies 3; F4 – fácies 4; F5 – fácies 5; F6 – fácies 6; F7 – fácies 7; F8 – fácies 8. | Fonte: autora, a partir do software IP (2023). 40
- Figura 20:** Perfis convencionais utilizados no trabalho: Coluna 3 – Perfil de diâmetro do poço (CAL); Coluna 4 – Perfil de raios gama total (HSGR) e de raios gama espectral: potássio (amarelo), tório (bege) e urânio (marrom). ; Coluna 5 – Perfis geoquímicos: alumínio (cinza), ferro (marrom), cálcio (azul claro), magnésio (azul escuro), silício (amarelo) e enxofre (roxo); Coluna 6 – Perfis de densidade (RHOZ), nêutrons (TNPH) e fator fotoelétrico (PEFZ); Coluna 7 – Perfis sônico: vagarosidade da onda compressional (DTCO) e cisalhante (DTSM); Coluna 8 – Perfis de ressonância magnética nuclear: porosidade total (marrom), porosidade efetiva (oliva) e fluido livre (azul); Coluna 9 – Fácies. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 41
- Figura 21:** Perfis de imagem utilizados no trabalho: Coluna 3 – Perfil de diâmetro do poço (CAL); Coluna 4 – Perfil de raios gama total (HSGR) e de raios gama espectral: potássio (amarelo), tório (bege) e urânio (marrom); Coluna 5 – Perfil de imagem acústica; Coluna 6 – Perfil de imagem resistiva; Coluna 7 – Fácies. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 41
- Figura 22:** Fluxograma das três etapas do estudo. | Fonte: autora 43
- Figura 23:** Exemplo de um gráfico de aleatoriedade apresentando a relação entre o número de agrupamentos e a razão de aleatoriedade. | Fonte: Al-Baldawi (2020). 45
- Figura 24:** esquema do processo de transformação para frequência realizado pelo algoritmo. O perfil de imagem original é segmentado em uma série de mosaicos de altura e largura fixos. Em seguida, cada um desses mosaicos passa por uma transformação para o domínio da frequência. | Fonte: autora. 47
- Figura 25:** foi aplicada a transformada de frequência na figura original (A), gerando a figura (B) que ressalta, através da frequência, certas propriedades da imagem original. As cores mais claras representam valores mais altos, ou seja, onde há mais informações. Na figura (C) foi retirada as informações do ponto central do domínio da frequência, recriando a imagem original (D) sem a informação que foi retirada e ressaltando algumas texturas e estruturas da imagem original. | Fonte: modificado de Shree K. Nayar, Columbia University (2021). 48
- Figura 26:** (A) critério da duplicação, onde o tamanho do bin aumenta de forma duplicada. Bin 1 – tamanho 1:1; Bin 2 – tamanho 2:2; Bin 3 – tamanho 4:4 e assim por diante. (B) critério de incrementação, onde o tamanho do bin aumenta de 1 para 1. Bin 1 – tamanho 1:1; Bin 2 – tamanho 2:2; Bin 3 – tamanho 3:3 e assim por diante. | Fonte: modificado de Shree K. Nayar, Columbia University (2021). 49
- Figura 27:** Curva de bins apresentando o número 10 como o número de bins ideal para o perfil de imagem acústica. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 50
- Figura 28:** Curva de bins apresentando o número 10 como o número de bins ideal para o perfil de imagem resistiva. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 50

Figura 29: Exemplo de matriz de confusão. A matriz tem uma forma quadrada onde a coluna representa os valores reais e a linha representa o valor previsto do modelo. P – Positivo: o valor real foi positivo e o modelo previu um valor positivo. FP – Falso positivo: o valor real foi negativo e o modelo previu um valor positivo. FN – Falso negativo: o valor real foi positivo e o modelo previu um valor negativo. N – Negativo: O valor real é negativo e o modelo previu um valor negativo. | Fonte: modificado de Karimi (2021). 53

Figura 30: Gráfico de aleatoriedade adquirido através do método não supervisionado. Percebe-se uma queda significativa a partir do cluster 8. Visualização no software IP. | Fonte: autora, a partir do software IP (2023)... 54

Figura 31: mapa auto-organizado gerado pela análise não supervisionada dos perfis convencionais. Fácies 1 – bege; Fácies 2 – amarelo; Fácies 3 – marrom claro; Fácies 4 – marrom escuro; Fácies 5 – verde claro; Fácies 6 – verde escuro; Fácies 7 – azul claro; Fácies 8 – azul escuro. Visualização no software IP. | Fonte: autora, a partir do software IP (2023)..... 55

Figura 32: Coluna 6 – perfil de fácies descritas por especialistas; Coluna Resultado SOM – perfil de fácies realizado pelo SOM através do método não supervisionado. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 56

Figura 33: mapa auto-organizado gerado pela análise supervisionada dos perfis convencionais. Fácies 1 – bege; Fácies 2 – amarelo; Fácies 3 – marrom claro; Fácies 4 – marrom escuro; Fácies 5 – verde claro; Fácies 6 – verde escuro; Fácies 7 – azul claro; Fácies 8 – azul escuro. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 57

Figura 34: Coluna 6 – perfil de fácies descritas por especialistas; Coluna Resultado SOM – perfil de fácies realizado pelo SOM através do método supervisionado. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 57

Figura 35: matriz de confusão dos resultados obtidos pela análise supervisionada (SOM) através dos perfis convencionais. | Fonte: autora. 58

Figura 36: matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) através dos perfis convencionais. Nesse caso, o modelo não predisse a fácies 7 pois apresentou um peso muito baixo para ser treinada. | Fonte: autora. 59

Figura 37: mapa auto-organizado gerado pela análise não supervisionada do perfil de imagem acústica. Fácies 1 – vermelho; Fácies 2 – verde escuro; Fácies 3 – azul; Fácies 4 – amarelo; Fácies 5 – cinza; Fácies 6 – rosa; Fácies 7 – verde claro; Fácies 8 – marrom. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 60

Figura 38: Coluna Imagem Estática – perfil de imagem acústica; Coluna Análise Estatística – perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método não supervisionado; Coluna 6 – perfil de fácies descritas por especialistas; Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 60

Figura 39: mapa auto-organizado gerado pela análise não supervisionada do perfil de imagem resistiva. Fácies 1 – vermelho; Fácies 2 – verde escuro; Fácies 3 – azul; Fácies 4 – amarelo; Fácies 5 – cinza; Fácies 6 – rosa; Fácies 7 – verde claro; Fácies 8 – marrom. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 61

Figura 40: Coluna Imagem Estática – perfil de imagem resistiva; Coluna Análise Estatística– perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método não

supervisionado; Coluna 6 – perfil de fácies descritas por especialistas; Visualização no software IP. | Fonte: autora, a partir do software IP (2023)..... 61

Figura 41: mapa auto-organizado gerado pela análise supervisionada do perfil de imagem acústica. Nesse caso, o algoritmo distribuiu as 8 fácies em diversos grupos. Visualização no software IP. | Fonte: autora, a partir do software IP (2023)..... 62

Figura 42: Coluna Imagem Estática – perfil de imagem acústica; Coluna Análise Estatística – perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método supervisionado; Coluna 6 – perfil de fácies descritas por especialistas; Visualização no software IP. | Fonte: autora, a partir do software IP (2023)..... 62

Figura 43: mapa auto-organizado gerado pela análise supervisionada do perfil de imagem resistiva. Fácies 1 – vermelho; Fácies 2 – verde escuro; Fácies 3 – azul; Fácies 4 – amarelo; Fácies 5 – cinza; Fácies 6 – rosa; Fácies 7 – verde claro; Fácies 8 – marrom. Visualização no software IP. | Fonte: autora, a partir do software IP (2023). 63

Figura 44: Coluna Imagem Estática – perfil de imagem resistiva; Coluna Análise Estatística – perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método supervisionado; Coluna 6 – perfil de fácies descritas por especialistas; Visualização no software IP. | Fonte: autora, a partir do software IP (2023)..... 63

Figura 45: matriz de confusão dos resultados obtidos pela análise supervisionada (SOM) através do perfil de imagem acústica. | Fonte: autora. 64

Figura 46: matriz de confusão dos resultados obtidos pela análise supervisionada (SOM) através do perfil de imagem resistiva. | Fonte: autora. 65

Figura 47: matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) através do perfil de imagem acústica. O modelo não predisse a fácies 7 pois apresentou um peso muito baixo para ser treinada. | Fonte: autora..... 66

Figura 48: matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) através do perfil de imagem resistiva. O modelo não predisse a fácies 7 pois apresentou um peso muito baixo para ser treinada. | Fonte: autora..... 66

Figura 49: matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) dos perfis convencionais e perfil de imagem acústica. | Fonte: autora..... 67

Figura 50: matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) dos perfis convencionais e perfil de imagem resistiva. | Fonte: autora. 68

Figura 51: matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) dos perfis convencionais e perfis de imagem. | Fonte: autora. 68

LISTA DE TABELAS

Tabela 1: (A) Produção de barris de óleo por dia em cada bacia sedimentar brasileira. (B) Produção diária de gás natural em cada bacia sedimentar brasileira. Ambas as tabelas estão em ordem decrescente de produção de óleo e gás diária. Fonte: ANP (2023).	18
Tabela 2: Perfis convencionais utilizados como variável de entrada para o ajuste do mapa auto-organizado. Fonte: autora.....	44
Tabela 3: Critérios utilizados para treinar o SOM. Fonte: autora.	44
Tabela 4: critérios utilizados para clusterizar os dados. Fonte: autora.....	44
Tabela 5: Parâmetros utilizados no CatBoost para testar o modelo gerado através dos perfis convencionais. Fonte: autora.....	46
Tabela 6: Parâmetros utilizados na etapa Transformação para Frequência. Fonte: autora.	47
Tabela 7: Método e estatística de binarização utilizados nos perfis de imagem acústica e resistiva. Fonte: autora.....	49
Tabela 8: Método de classificação de bins utilizado nos perfis de imagem acústica e resistiva. Fonte: autora. .	50
Tabela 9: Critérios utilizados para treinar o SOM. Fonte: autora.	51
Tabela 10: Critérios utilizados para clusterizar os dados do perfil de imagem. Fonte: autora.....	51
Tabela 11: Precisão da matriz de confusão referente a cada análise supervisionada, com destaque em cinza para os melhores resultados de classificação entre todas as análises. Fonte: autora.....	69

SUMÁRIO

1. INTRODUÇÃO	15
1.1. Objetivos	15
2. FUNDAMENTOS TEÓRICOS	17
2.1. Área de Estudo.....	17
2.1.1. Contexto Geológico.....	19
2.1.2. Campo de Atapu	21
2.2. Perfis de Poços.....	22
2.2.1. Perfis de Raios Gama	23
2.2.2. Perfis de Densidade e Fator Fotoelétrico.....	24
2.2.3. Perfis de Nêutrons	27
2.2.4. Perfis Geoquímicos	29
2.2.5. Perfis Sônicos	31
2.2.6. Perfis de Ressonância Magnética Nuclear	33
2.2.7. Perfis de Imagem.....	33
2.2.7.1. Perfis de imagem resistiva.....	34
2.2.7.2. Perfis de imagem acústica.....	35
2.3. Aprendizado de Máquina.....	36
2.3.1. <i>Self Organizing Maps</i> (SOM).....	37
2.3.2. <i>Categorical Boosting</i> (Catboost).....	38
3. MATERIAIS E MÉTODOS.....	40
3.1. Análise dos Perfis Convencionais.....	43
3.2. Análise dos Perfis de Imagem	46
3.3. Análise Integrada dos Perfis Convencionais e de Imagem: CatBoost.....	52
4. RESULTADOS E INTERPRETAÇÃO	54
4.1. Perfis Convencionais	54
4.1.1. Análise Não Supervisionada: clusterização do mapa auto-organizado	54
4.1.2. Análise Supervisionada: calibração do mapa auto-organizado	56
4.1.3. Análise Supervisionada: CatBoost	58
4.2. Perfis de Imagem	59
4.2.1. Análise Não Supervisionada: clusterização do mapa auto-organizado	59
4.2.2. Análise Supervisionada: calibração do mapa auto-organizado	61

4.2.3. Análise Supervisionada: CatBoost	65
4.3. Análise Integrada dos Perfis Convencionais e de Imagem: CatBoost.....	67
4.4. Análise Comparativa.....	69
5. DISCUSSÃO.....	70
6. CONCLUSÃO E RECOMENDAÇÕES	71
7. REFERÊNCIAS	72

1. INTRODUÇÃO

A capacidade de previsão dos algoritmos de aprendizado de máquina desempenha um papel crucial na otimização da classificação de fácies na indústria de óleo e gás, já que acelera este processo e proporciona resultados confiáveis. Tradicionalmente, as variáveis de entrada dos modelos de classificação de fácies são os perfis convencionais, entregando informações composicionais das formações. Entretanto, essas variáveis podem não ser as mais adequadas no processo de classificação de fácies, principalmente para rochas carbonáticas do pré-sal da Bacia de Santos, onde informações texturais e estruturais são significativas para essa classificação.

O aprendizado de máquina é uma subárea da Inteligência Artificial e envolve principalmente a realização de previsões reconhecendo padrões em base de dados (Friedman *et al.*, 2001). Isso é alcançado através do treinamento do algoritmo de aprendizado de máquina utilizando métodos supervisionados e não supervisionados. Este campo é amplamente utilizado por estatísticos para ajudar na análise de um grande volume de dados e, recentemente, tem sido adotado cada vez mais por geocientistas para explorar diversos métodos de processamento e interpretação na classificação de fácies (Guarido, 2018).

Os perfis de imagem são pouco utilizados em problemas de classificação e podem trazer informações texturais e estruturais importantes na classificação de fácies carbonáticas do pré-sal. Esse pouco uso é devido ao fato de que o perfil de imagem é uma informação 2D, impossibilitando sua utilização direta nos algoritmos de classificação. É preciso, então, extrair variáveis através de transformações e análises estatísticas dessas imagens, o que não é trivial.

Nesse contexto, o presente estudo foi desenvolvido empregando métodos tanto supervisionados quanto não supervisionados, através de perfis convencionais e de imagem, para a classificação de fácies do poço 9-ATP-1-RJS, localizado no campo de Atapu, na Bacia de Santos. Como trata-se de um intervalo do pré-sal, com rochas altamente heterogêneas, a utilização de dados de perfis de imagem apresenta um grande potencial para fornecer informações texturais e estruturais, agregando na classificação de forma integrada com os perfis convencionais. A escolha dessa bacia se deu pelo fato ser uma das principais bacias produtoras de petróleo do Brasil (ANP, 2022), sendo muito importante para a economia do país.

1.1. Objetivos

O presente trabalho teve como objetivo principal analisar a eficácia dos modelos de aprendizado de máquina treinados com variáveis de entrada provenientes de perfis convencionais, perfis de imagem e uma combinação de ambos. Para isso, foram utilizados dois

algoritmos de aprendizado de máquina (SOM e CatBoost) para garantir que os resultados obtidos não fossem influenciados pelo tipo de algoritmo utilizado, mas sim pelos dados de entrada (perfis convencionais e de imagem). Com o intuito de alcançar o objetivo principal, foram estabelecidos os seguintes objetivos específicos:

- (a) comparar as fácies observadas em amostras de rochas e descritas por especialistas com os resultados obtidos por esses dois algoritmos;
- (b) compreender se a análise textural de fácies e o algoritmo SOM implementados no software IP são capazes de extrair informações relevantes para esse processo de classificação;
- (c) comparar o desempenho de classificação de fácies entre os resultados obtidos pelos perfis convencionais e de imagem.

2. FUNDAMENTOS TEÓRICOS

2.1. Área de Estudo

A Bacia de Santos está localizada na margem continental sudeste brasileira e é considerada a maior bacia sedimentar offshore do país, abrangendo uma área total de aproximadamente 350.000 km² e cota batimétrica de 3.000 m (Moreira *et al.*, 2007). A bacia estende-se de Cabo Frio até Florianópolis, englobando os litorais dos Estados do Rio de Janeiro, São Paulo, Paraná e Santa Catarina. Geologicamente, segundo Souza *et al.*, 2019, seus limites são definidos pela Bacia de Campos ao norte (por meio do alto estrutural de Cabo Frio), pela Bacia de Pelotas ao sul (por meio do alto de Florianópolis), a oeste pela Serra do Mar e a leste pelo limite oriental do Platô de São Paulo (Figura 1).

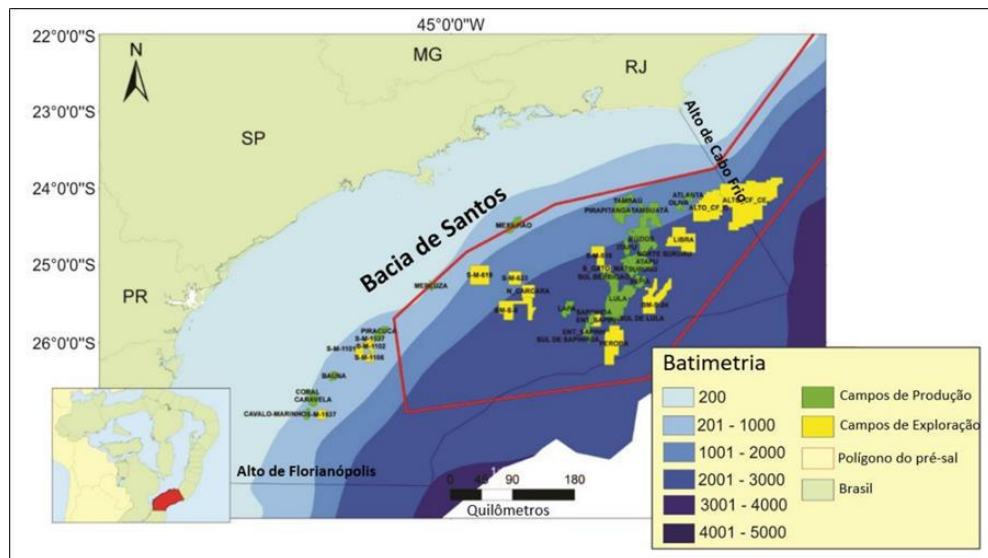


Figura 1: Mapa de localização da Bacia de Santos. O polígono vermelho indica o polígono do pré-sal. | Fonte: Castro (2019).

No que se refere ao histórico exploratório, em 1980 ocorreu a primeira descoberta de petróleo na Bacia de Santos, pela Pecten, em arenitos turbidíticos de talude da Formação Itajaí-Açu no campo de Merluza (Pereira *et al.*, 1990). Em 1997, a partir da criação da Lei do Petróleo, a bacia adquiriu novos investimentos exploratórios da Petrobras e demais empresas para fins de aquisição de dados geológicos e geofísicos, que favoreceram a descoberta de novas jazidas de óleo em Oliva e Atlanta; e de gás natural em Lagosta, Tambuatá, Tambaú, Mexilhão, Carapiá, Uruguá e Pirapitanga (ANP, 2017).

Em 2006 foi perfurado o primeiro poço exploratório na seção pré-sal da Bacia de Santos, comprovando a existência de um sistema petrolífero produtivo no prospecto de Tupi e criando expectativas para o novo *play* exploratório da Bacia de Santos: o pré-sal. A partir dessa descoberta, segundo a Agência Nacional do Petróleo, Gás e Biocombustíveis (ANP), o cenário

exploratório brasileiro se transformou, dando origem ao novo marco regulatório para as áreas do pré-sal da Bacia de Santos e Campos (Lei nºs 12.276/10, 12.304/10 e 12.351/10).

Atualmente, a Bacia de Santos é uma das principais bacias produtoras de óleo e gás no Brasil e foi responsável por cerca de 74% de todo o petróleo produzido no país em junho de 2023 (Figura 2A). No mesmo mês, a bacia teve uma produção de óleo na ordem de 2,5 milhões de barris por dia (bbl/d) e a produção diária de gás natural na ordem de 111,2 mil de m³ (ANP, 2023), ficando na frente da produção da Bacia de Campos (Tabela 1). Em relação ao pré-sal, a Bacia de Santos correspondeu a cerca de 96% de toda a produção de petróleo equivalente (Figura 2B).

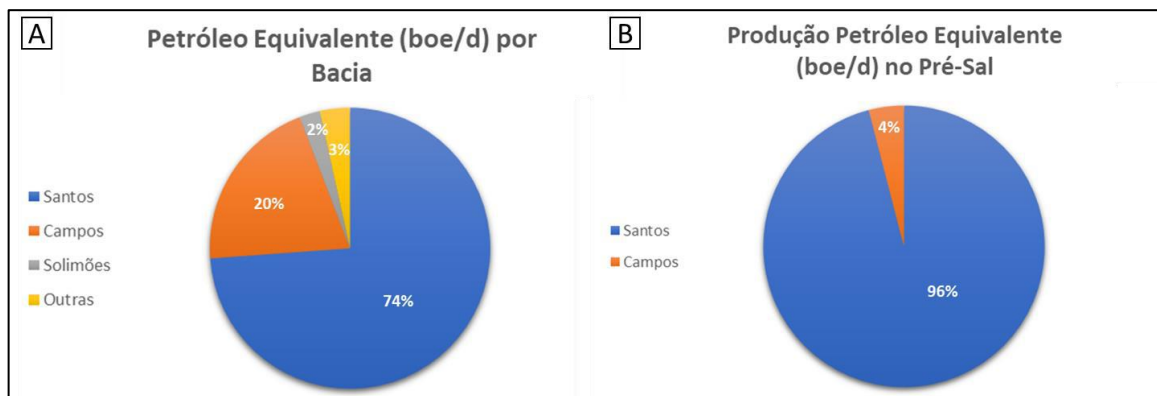


Figura 2: (A) Gráfico percentual da produção total de petróleo equivalente nas Bacias de Santos, Campos e Solimões. (B) Gráfico percentual da produção total de petróleo equivalente na Bacia de Santos e Campos no pré-sal. Todos os dados foram adquiridos no mês de junho de 2023. | Fonte: modificado de ANP (2023).

A Bacia		Óleo (bbl/d)	B Bacia		Gás Natural (mil m ³ /d)
Santos		2.482.263,66	Santos		111.252,74
Campos		796.089,11	Solimões		14.355,36
Potiguar		33.314,09	Campos		12.645,08
Recôncavo		17.062,68	Parnaíba		5.634,55
Espírito Santo		13.925,10	Recôncavo		2.447,88
Solimões		8.967,04	Camamu		1.810,74
Sergipe		5.956,92	Alagoas		1.383,96
Alagoas		1.789,46	Potiguar		1.069,20
Barreirinhas		51,20	Espírito Santo		743,86
Amazonas		0,00	Amazonas		700,79
Camamu		0,00	Tucano Sul		134,71
Paraná		0,00	Sergipe		50,81
Total		3.359.419,27	Total		152.257,89

Tabela 1: (A) Produção de barris de óleo por dia em cada bacia sedimentar brasileira. (B) Produção diária de gás natural em cada bacia sedimentar brasileira. Ambas as tabelas estão em ordem decrescente de produção de óleo e gás diária. | Fonte: ANP (2023).

Segundo os dados da ANP, são no total 38 campos descobertos, sendo 16 na fase de desenvolvimento, 18 na fase de produção e 4 em devolução. A Bacia de Santos também conta com 898 poços perfurados, entre eles 530 poços exploratórios (considerando poço pioneiro, estratigráfico, extensão, pioneiro adjacente, jazida mais rasa e jazida mais profunda). O presente trabalho foi elaborado sobre um poço especial de aquisição de dados de reservatório (ADR) de nome ANP 9-ATP-1-RJS, no campo de Atapu, sendo ele operado pela Petrobras (71%) em consórcio com a Shell (15%) e TotalEnergies (14%) (ANP, 2023).

2.1.1. Contexto Geológico

Do ponto de vista geológico, a Bacia de Santos corresponde a uma bacia de margem passiva gerada no Neocomiano durante o processo de ruptura do supercontinente Gondwana, resultando na abertura do Oceano Atlântico e, conseqüentemente, na separação dos continentes americano e africano (Mio *et al.*, 2005).

A história geológica da bacia pode ser segmentada em três estágios tectônicos: Rifte, Pós-Rifte e Drifte (Figura 3). O primeiro estágio iniciou-se a partir do Cretáceo Inferior com a separação das placas Sulamericana e Africana através de esforços extensionais progressivos e diacrônicos que favoreceram a formação de espaço de acomodação de sedimentos. O rifteamento da crosta continental também foi responsável pela formação do Platô de São Paulo, significativo na sedimentação dos evaporitos na fase Pós-Rifte (Moreira *et al.*, 2007).

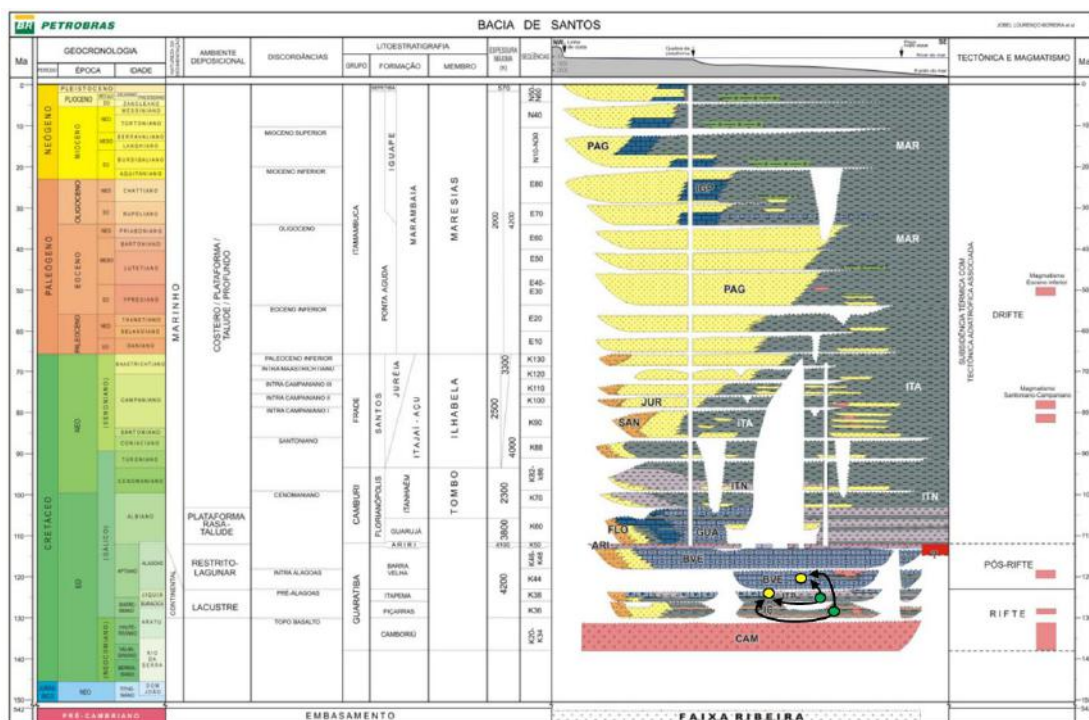


Figura 3: Carta estratigráfica da Bacia de Santos com sistema petrolífero atuante na seção pré-sal. Os círculos verdes representam as rochas geradoras (Itapema e Piçarras), enquanto os círculos amarelos representam as rochas reservatório (Barra Velha). | Fonte: modificado de Moreira *et al.* (2007).

Estratigraficamente, a fase Rifte é caracterizada por três formações do Grupo Guaratiba: Camboriú, Piçarras e Itapema. A Formação Camboriú compreende os basaltos toleíticos procedentes dos primeiros derrames de lava da Formação Serra Geral, que perduraram do Jurássico Superior ao Cretáceo Inferior. Estes ocorrem na Bacia do Paraná e estendem-se nas bacias marginais do sudeste do Brasil (Mohriak, 2012). Sobreposta a esses basaltos ocorre a Formação Piçarras, que nas porções proximais compreende os sedimentos de leques aluviais datados do Barremiano (conglomerados e arenitos polimíticos) e nas porções distais compreende rochas de ambientes lacustres, como arenitos, siltitos e folhelhos de composição talco estevensítica - silicato hidratado de magnésio/esmectita (Moreira *et al.*, 2007). Por fim, sobrepostos a Formação Piçarras, ocorrem sedimentos da Formação Itapema datados do Neobarremiano ao Eoaptiano, que são caracterizados por conglomerados e arenitos com sedimentos de leques aluviais na porção proximal. No restante da sequência ocorrem sedimentos depositados nas primeiras incursões marinhas na bacia, como intercalações de calcirruditos e folhelhos escuros.

Durante o segundo estágio tectônico, fase Pós-Rifte, ocorreu o aumento gradual das incursões marinhas que contribuíram para a formação do Oceano Atlântico Sul. A restrita circulação de água marinha, juntamente com o clima árido, criou condições ideais para acumulação e deposição de evaporitos durante o Aptiano (Souza *et al.*, 2019). Esse estágio é representado estratigraficamente pelas formações Barra Velha (carbonatos) e Ariri (evaporitos), ambas do Grupo Guaratiba. A Formação Barra Velha é subdivida em duas sequências depositadas em ambiente transicional entre continental e marinho raso desde o Eoaptiano (calcários microbiais, estromatólitos, laminitos e folhelhos) ao Neoaptiano (conglomerados, arenitos, calcários estromatolíticos e laminitos microbiais), sendo elas separadas por uma discordância de 117 Ma (Moreira *et al.*, 2007). Sobreposta a esses sedimentos ocorre a Formação Ariri, constituída de evaporitos depositados no Neoaptiano em um ambiente de golfo estreito e alongado (Riccomini *et al.*, 2012).

Por fim, a fase Drifte é caracterizada pela individualização das placas sulamericana e africana, promovendo a transição de crosta continental para crosta oceânica e o término do caráter evaporítico da bacia. No que se refere a estratigrafia, esta fase é constituída pelos Grupos Camburi, Frade e Itamambuca (Moreira *et al.*, 2007). O Grupo Camburi retrata uma fase transgressiva que se encerra a partir da deposição dos folhelhos anóxicos da transgressão Turoniana. Esta sequência de sedimentos representa todos aqueles depositados desde os leques aluviais até os pelitos e arenitos batiais depositados após a Formação Ariri até o topo do Cenomaniano (Moreira *et al.*, 2007). O Grupo Frade encontra-se sobreposto ao Grupo Camburi

e compreende uma fase de regressão marinha na qual o limite da plataforma avançou em torno de 200 km costa adentro devido ao soerguimento da Serra do Mar. Este Grupo compreende todos os sedimentos depositados desde os leques aluviais até os pelitos e arenitos batiais depositados do topo do Cenomaniano até o Cretáceo/Paleógeno. O Grupo Itamambuca, por sua vez, abrange os depósitos de leques aluviais proximais até pelitos e arenitos batiais, podendo ocorrer sedimentação carbonática próximo à quebra da plataforma. Este Grupo representa todos os sedimentos depositados após o Cretáceo/Paleógeno até os dias atuais (Moreira *et al.*, 2007).

2.1.2. Campo de Atapu

O Campo de Atapu, foco deste estudo, está inserido no polígono do pré-sal na porção central da Bacia de Santos, a cerca de 230 km da costa e a uma profundidade que varia de 2.000 a 2.300 m. Mais especificamente, este campo integra - juntamente com Norte de Berbigão, Norte de Sururu e Sul de Sururu - uma área denominada Entorno de Iara, localizada a norte do Campo de Sépia, a sul do Campo de Búzios e a sudeste do Campo de Itapu (Figura 4). A descoberta do Campo de Atapu se deu a partir da perfuração do poço 1-BRSA-1146-RJS, sendo sua comercialidade declarada em dezembro de 2014 (ANP, 2019).

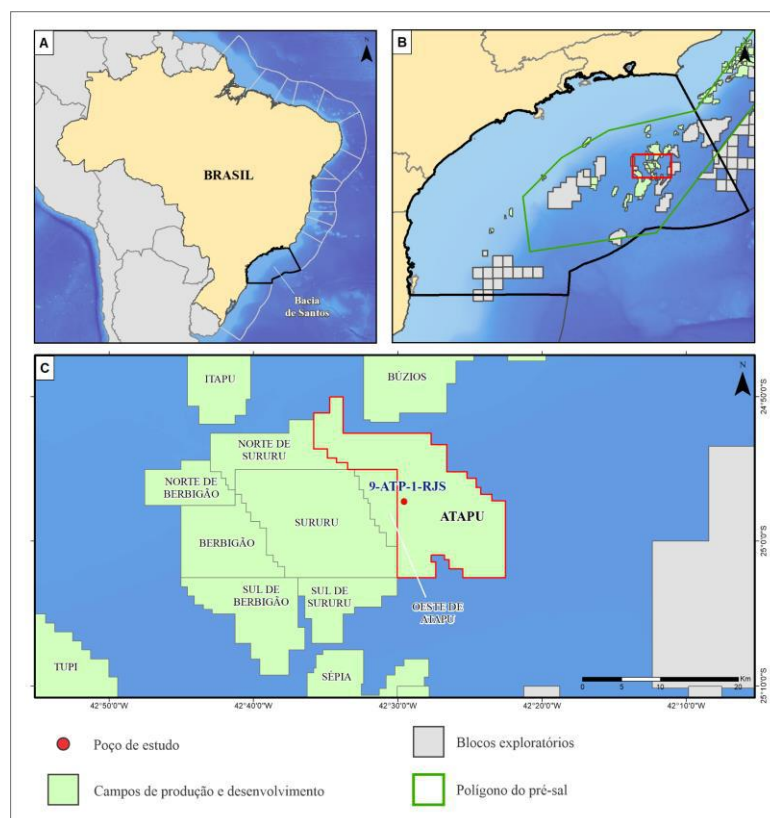


Figura 4: (A) Mapa do Brasil em destaque para a Bacia de Santos. (B) Visão aproximada da Bacia de Santos com seus respectivos blocos, campos e parte do polígono do pré-sal. (C) Mapa da área denominada Entorno de Iara, com destaque para o Campo de Atapu e o poço de estudo (9-ATP-1-RJS). | Fonte: autora.

O campo é operado pela Petrobras e está entre os 10 maiores campos produtores de petróleo equivalente (boe/d) do Brasil, tendo produzido cerca de 134 Mboe/d (mil barris de óleo equivalente por dia) no mês de junho de 2023 (Figura 5). Seu sistema petrolífero é definido por Itapema/Piçarras-Barra Velha e é atuante no pré-sal da Bacia de Santos, com geração de hidrocarbonetos em folhelhos lacustres da Formação Itapema e/ou Formação Piçarras e acumulação nos reservatórios carbonáticos das formações Itapema e Barra Velha (Figura 3). O reservatório apresenta óleo de boa qualidade e previsão de término da produção para 2045 (ANP, 2023).

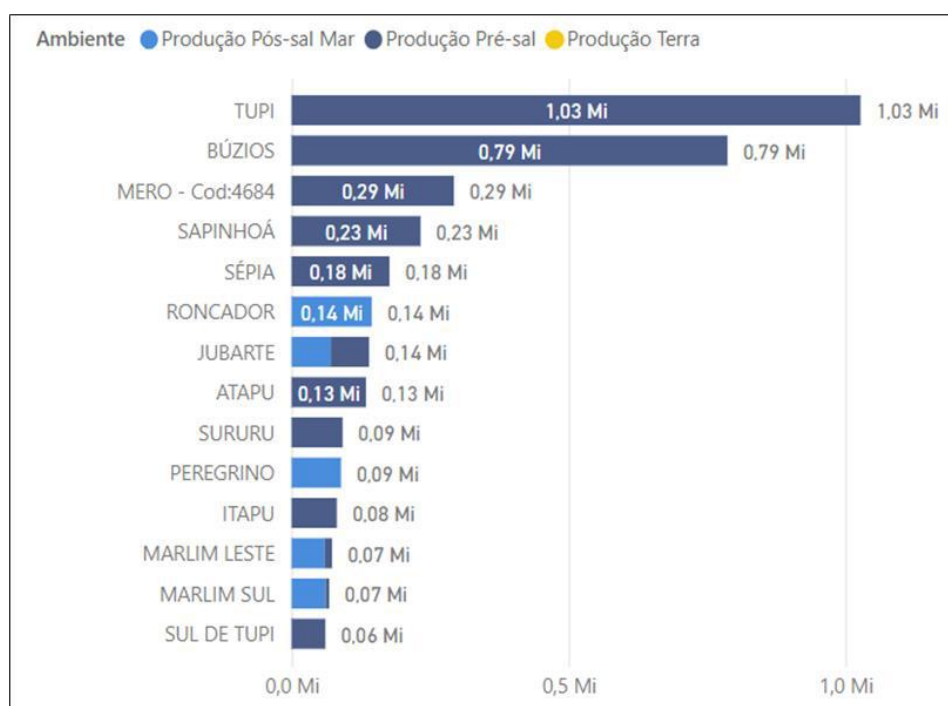


Figura 5: Petróleo Equivalente (boe/d) por Campo e Ambiente. | Fonte: ANP (2023).

2.2. Perfis de Poços

Os perfis de poços são registros detalhados das formações geológicas em subsuperfície na configuração de um perfil vertical (imagem visual em relação à profundidade). Esse levantamento de dados é obtido a partir de ferramentas de perfilagem que realizam medições de diferentes propriedades geofísicas das rochas à medida que se deslocam no interior dos poços. Segundo Oliveira (2022), as medições são adquiridas pelas ferramentas de perfilagem através da emissão de um sinal ou perturbação no meio rochoso por um transmissor, sendo esta perturbação posteriormente lida por um receptor após interagir com a formação. A partir da diferença entre o sinal emitido e o que foi lido, é possível estimar determinadas propriedades das rochas. Estes registros podem ser captados após a perfuração do poço (*wireline logging*),

através de ferramentas de perfilagem a cabo, ou durante a perfuração (*logging while drilling - LWD*), por ferramentas que se conectam à coluna de perfuração (Oliveira, 2022).

A perfilagem de poços auxilia em um melhor entendimento das camadas em subsuperfície, já que é possível obter diversas informações sobre as rochas e fluidos. Dito isto, as propriedades petrofísicas estimadas por meio da interpretação dos perfis (como porosidade, permeabilidade, densidade, saturação, fraturas, inclinação de camadas, etc.) permitem avaliar e estimar o valor comercial de um prospecto, assim como sua viabilidade econômica (Leite *et al.*, 2007).

No presente estudo foram utilizados sete perfis de poços (raios gama, densidade, fator fotoelétrico, nêutrons, geoquímico, sônico e ressonância magnética nuclear) que são afetados pela variação composicional da rocha e dois perfis de imagem que levam em consideração características resistivas e acústicas das formações. Os perfis serviram de base na identificação de estratos do poço 9-ATP-1-RJS através do aprendizado de máquina e posteriormente foram comparados com dados de testemunho de rocha descritos por especialistas para definir quais deles apresentaram o melhor desempenho na classificação litológica.

2.2.1. Perfis de Raios Gama

Os perfis de raios gama são amplamente utilizados no registro da radioatividade natural emitida pela formação, principalmente relacionada aos elementos instáveis ^{232}Th , ^{238}U e ^{40}K , geralmente presentes em rochas com grandes quantidades de argilominerais ou matéria orgânica. Esses tipos de perfis são obtidos através de ferramentas capazes de captar apenas a radiação natural emitida pelo meio, não disparando nenhum sinal para a formação (Lima, 2018). Espectros da radiação gama natural podem ser medidos por ferramentas modernas, sendo possível definir janelas de energia e calcular a fração relativa ao conteúdo de Th, U e K presentes na formação. Como consequência disso tem-se a melhora na diferenciação de folhelhos e detecção de arenitos arcóseos (Oliveira, 2022). Na Figura 6 é possível observar o espectro de radiação gama e as respectivas janelas relacionadas ao Th, U e K.

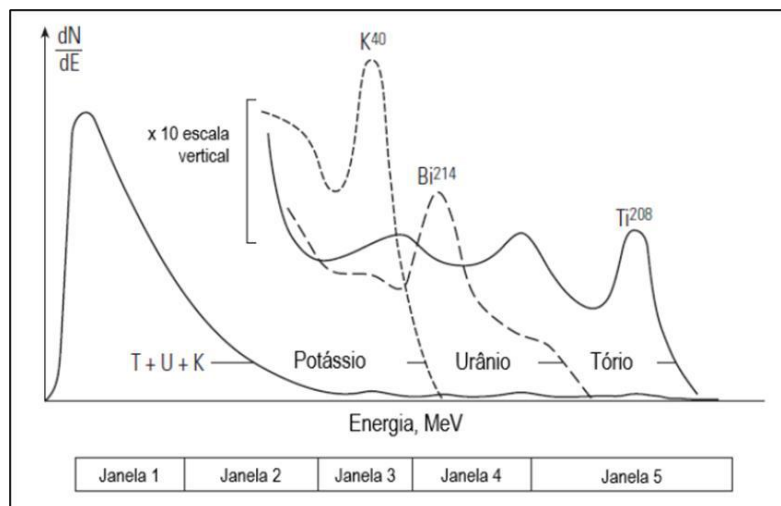


Figura 6: Espectro de radiação gama e as respectivas janelas relacionadas ao Th, U e K. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

Em ambientes siliciclásticos, os folhelhos que apresentam maior concentração de argilominerais manifestam alta radioatividade e, conseqüentemente, uma maior resposta no perfil. Já os carbonatos e evaporitos costumam apresentar menor radioatividade e menor resposta no perfil. Nos carbonatos, a alta ou baixa radiação gama está normalmente relacionada à quantidade de matéria orgânica. No caso dos evaporitos, estes costumam apresentar valores de radiação gama baixo, com exceção da silvinita e carnalita que apresentam potássio em sua composição. Além disso, rochas com baixa argilosidade também podem apresentar elevados teores de raios gama, visto que essa resposta pode ser causada por altas concentrações de feldspato potássico, micas, glauconita ou águas ricas em urânio, configurando erros de interpretações litológicas no perfil (Lima, 2018).

Dito isto, os perfis de raios gama são muito utilizados na identificação de rochas reservatório e não reservatório, interpretação de seqüências sedimentares, análise quantitativa de presença de argilominerais e identificação geral de litotipos (Lima, 2018).

2.2.2. Perfis de Densidade e Fator Fotoelétrico

Os perfis de densidade e fator fotoelétrico são adquiridos a partir de ferramentas que utilizam uma fonte de radiação gama, geralmente ^{137}Cs , e dois detectores de NaI(Tl): um posicionado mais próximo da fonte radioativa e outro mais distante (Ellis & Singer, 2007). A ferramenta emite uma radiação gama que interage com a formação, sendo então atenuada pelo espalhamento Compton que apresenta correlação com a densidade *bulk* da formação (medida da massa de um material por unidade de volume), como pode ser observado na equação (1) (Oliveira, 2022):

$$\Phi = \Phi_0 e^{-\rho_b \frac{Z}{A} N_0 \sigma x} \quad \text{Equação (1)}$$

O fluxo de radiação é definido por Φ , o número de densidade de elétrons de um material de densidade ρ_b é definido por $\rho_b \frac{Z}{A} N_0$, a seção de choque do espalhamento Compton é representado por σ e a distância entre a fonte radioativa e o detector da ferramenta por x , possibilitando o cálculo da densidade da formação (geralmente em g/cm^3). Além disso, a razão $\frac{Z}{A}$ é aproximadamente 0.5 para a maioria dos elementos químicos (Oliveira, 2022).

A ferramenta de densidade é corrida junto a parede do poço com o intuito de reduzir distorções ocasionadas pela presença de reboco ou fluido de perfuração caso haja arrombamento do poço (Figura 7). Dessa forma, os dois detectores mencionados previamente são essenciais na correção da leitura de densidade caso ocorra um desses dois casos. Mais especificamente, o detector mais próximo lerá uma maior quantidade de reboco e, conseqüentemente, sua leitura será mais afetada do que a do detector mais distante. A comparação entre as duas leituras gera um fator de correção que fornece a densidade real da formação (Oliveira, 2022).

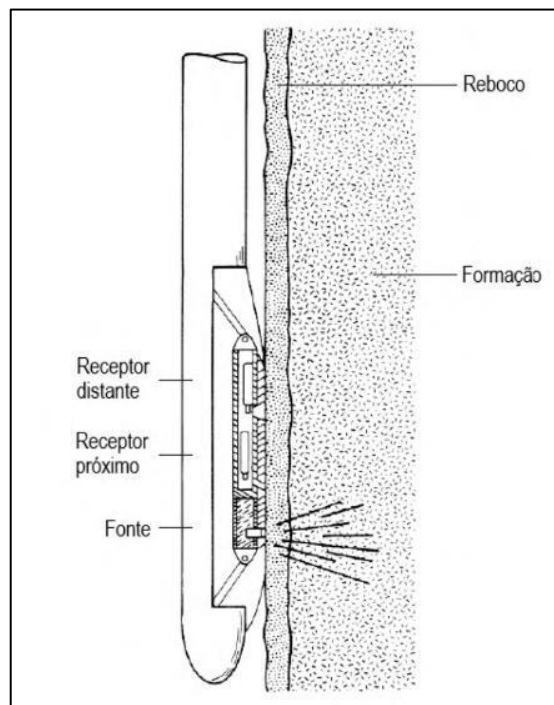


Figura 7: Exemplo de ferramenta de aquisição do perfil de densidade sendo adquirida junto a parede do poço. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

Em relação à finalidade do uso do perfil de densidade, pode-se dizer que este é usado para estimar a porosidade da formação através da equação (2):

$$\rho_b = \phi \rho_f + (1 - \phi) \rho_{ma} \quad \text{Equação (2)}$$

No qual ϕ é a porosidade, ρ_f é a densidade do fluido presente nos poros e ρ_{ma} é a densidade da matriz da rocha. Vale destacar que outros perfis de poço também são capazes de estimar as densidades da matriz e do fluido.

Quanto à absorção fotoelétrica (P_e), esta é proporcional ao número atômico Z do material que interage com a radiação (equação 3) e seu efeito é observado em baixas energias. Sendo assim, outra forma de representar a equação (1) é através da equação (4), onde a energia da radiação gama é representada por E e os coeficientes conhecidos e constantes definidos por a e b . É importante ressaltar que a equação (1) só é válida para radiação gama de alta energia. Com o uso de ferramentas de densidade mais atuais, é possível adquirir a leitura da radiação gama em altas e baixas energias (Figura 8), obtendo, então, o perfil de fator fotoelétrico.

$$P_e \equiv \frac{Z^{3.6}}{10} \quad \text{Equação (3)}$$

$$\Phi = \Phi_0 e^{-N_0 \rho b (a(E) P_e + b(E)) x} \quad \text{Equação (4)}$$

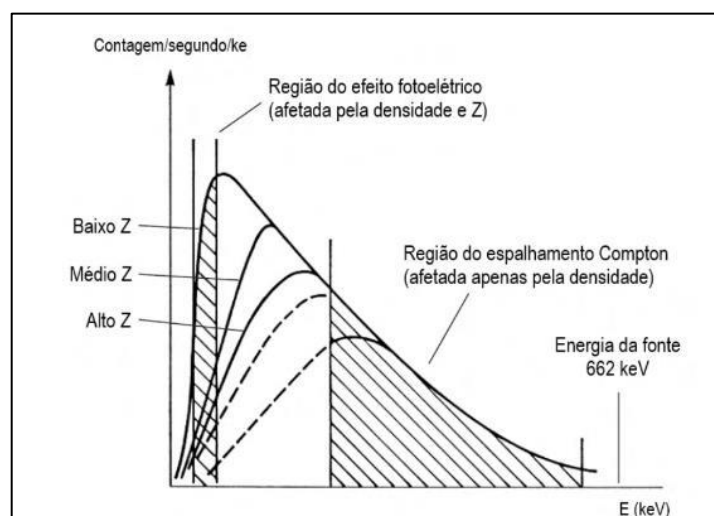


Figura 8: Representação esquemática do espectro de raios gama lido em três formações compostas por átomos de diferentes números atômicos. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

O perfil fotoelétrico estima a composição da matriz da rocha, já que está relacionado ao número atômico dos átomos que compõem a formação, e costuma não ser facilmente afetado pelas variações de porosidade (Ellis & Singer, 2007). Os valores de fator fotoelétrico normalmente encontrados em rochas variam entre 1 e 6, mas alguns materiais – como a barita (muito utilizada em fluidos de perfuração) – podem gerar valores anômalos.

2.2.3. Perfis de Nêutrons

Os perfis de nêutrons estão associados à análise de porosidade da rocha através de ferramentas localizadas junto a parede do poço que captam as interações entre os nêutrons e a formação; diferente da aquisição do perfil de densidade, no qual é capturado as interações entre radiação gama e formação (Oliveira, 2022). As fontes emitem nêutrons rápidos de alta energia, podendo esta ser uma fonte química (como AmBe) ou uma fonte eletrônica (como D-T). Esses nêutrons são então detectados próximos e/ou distantes da fonte radioativa e, uma vez que interagem com a formação, perdem velocidade até alcançarem as faixas epitermal ou termal de energia, como pode ser visto na Figura 9, e são capturados pelos átomos da formação (Ellis & Singer, 2007):

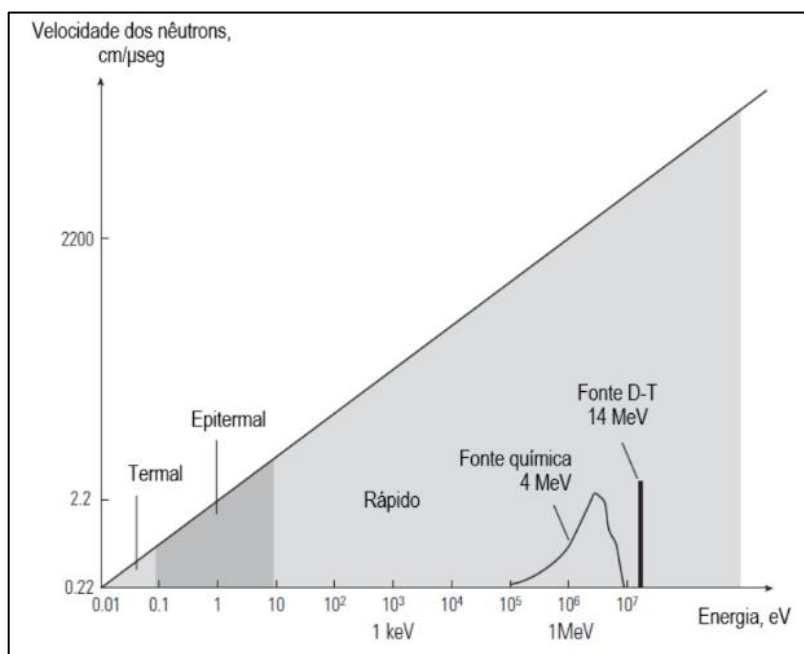


Figura 9: Classificação dos nêutrons de acordo com suas energias e velocidades. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

Segundo Oliveira (2022), existem três tipos de interações responsáveis pela desaceleração dos nêutrons de alta energia:

- I. Espalhamento elástico: se torna mais efetivo quando os nêutrons interagem com elementos cuja a massa está muito próxima à sua, como o hidrogênio por exemplo;
- II. Espalhamento inelástico: uma porção da energia do nêutron excita o núcleo do átomo que ele atinge, sendo capaz de produzir radiação gama;
- III. Captura radioativa: ocorre com nêutrons de baixa energia termal quando estes são absorvidos pelos átomos da formação, também produzindo radiação gama.

No que diz respeito a ferramenta de aquisição dos perfis de nêutrons (Figura 10), esta apresenta receptores que realizam medições em uma faixa de energia pré definida (termal ou epitermal) e, em seguida, contabilizam o número de nêutrons dentro dessa energia específica. A razão entre essas leituras está relacionada com a distância de migração dos nêutrons, sendo ela afetada pelo espalhamento elástico e pela captura radioativa. Dito isto, a razão entre as leituras dos detectores será afetada pela composição química da rocha e pelo índice de hidrogênio, sendo este último possuindo relação direta com a porosidade como pode ser observado na Figura 11 (Ellis & Singer, 2007).

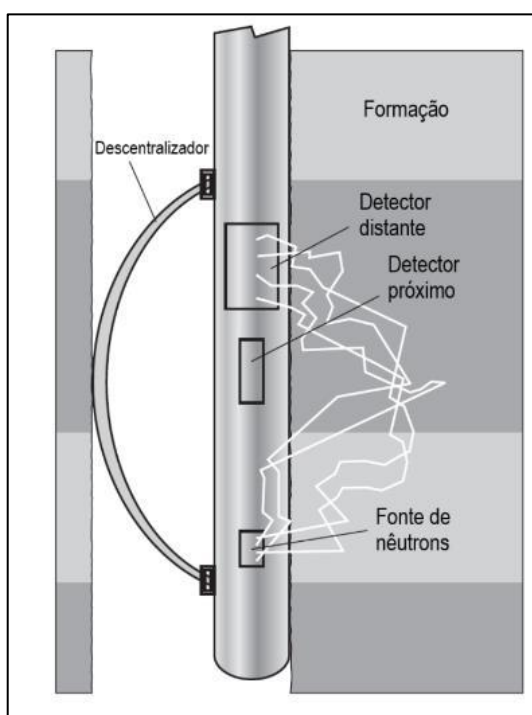


Figura 10: Exemplo de ferramenta de aquisição do perfil de nêutrons, adquirida junto à parede do poço. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

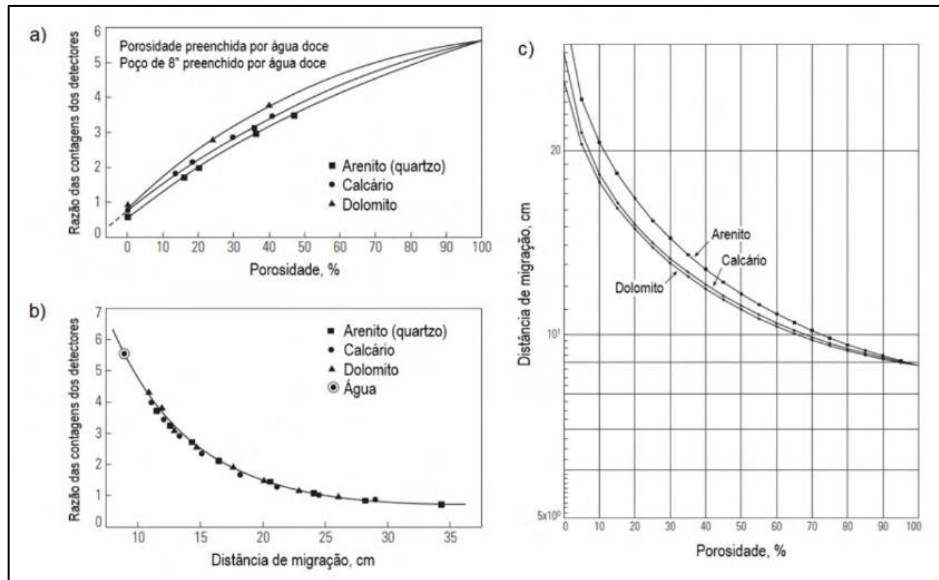


Figura 11: Gráfico mostrando a relação entre (a) a razão das contagens dos detectores e a porosidade para diferentes tipos de rocha, (b) a razão das contagens dos detectores e a distância de migração e (c) a distância de migração e a porosidade para diferentes tipos de rocha. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

É importante destacar que alguns fatores podem influenciar na leitura da ferramenta de nêutrons. Segundo SLB (2023), os principais fatores são: diâmetro do poço; salinidade; peso do fluido de perfuração; espessura do reboco; temperatura; pressão. Dessa forma, para mitigar os impactos na leitura, é necessário que o perfil de nêutrons passe por diversas correções ambientais.

2.2.4. Perfis Geoquímicos

Segundo Ellis & Singer (2007), os perfis geoquímicos são adquiridos através de ferramentas capazes de captar o espectro de radiação gama provocado pelas interações entre nêutrons emitidos e os elementos químicos da rocha. A leitura da radiação é feita através de instrumentos capazes de medir e detectar radiação gama, como cintilômetros de NaI(Tl), BGO ou LaBr₃(Ce) que geram um espectro com a soma das inúmeras emissões (Oliveira, 2022). A partir da decomposição desse espectro em espectros elementares de referência, tem-se as assinaturas típicas de cada elemento químico contido tanto na formação quanto no poço (Figura 12).

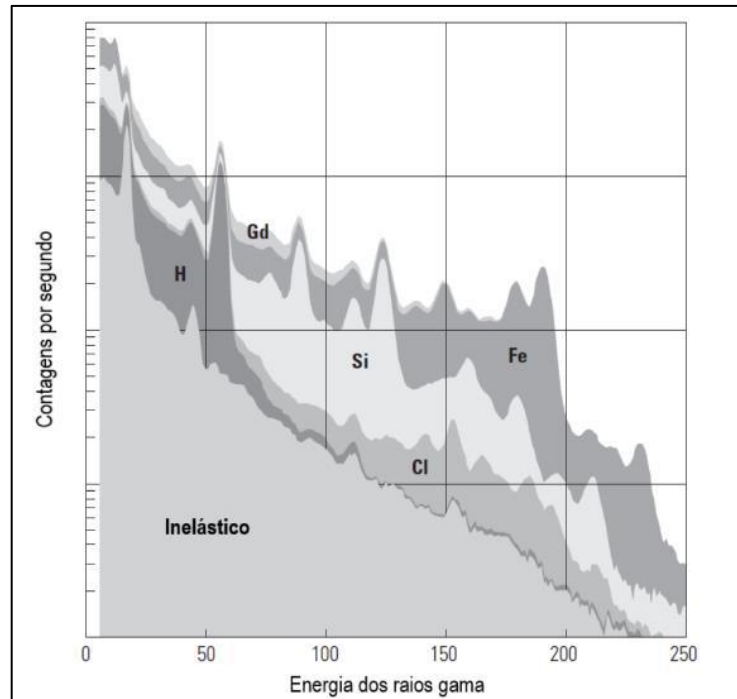


Figura 12: Exemplo de espectro de radiação gama e sua decomposição em espectros elementares de referência para diversos elementos químicos. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

Os pesos de cada elemento químico no espectro avaliado são representados pelos rendimentos ou *yields* (coeficientes associados aos espectros de referência). No entanto, segundo Hertzog *et al.* (1989), algumas questões impedem o uso direto dos rendimentos no cálculo das composições da massa dos elementos, dificultando a determinação da concentração dos elementos químicos no ambiente de perfilagem de poços. Alguns exemplos dessas questões são:

- I. O grau de resposta de cada elemento é intrínseco, então a quantidade de radiação emitida será diferente para o mesmo teor na rocha de dois elementos distintos;
- II. Elementos como O e C estão presentes no fluido e na matriz da rocha, então não é possível obter a contribuição exclusiva desses elementos na matriz de uma forma direta;
- III. Os rendimentos calculados envolvem respostas tanto de fluido quanto de matriz. Levando em consideração que a decomposição espectral é realizada de forma que a soma de todos os rendimentos seja igual à 1, fatores como porosidade e salinidade deveriam ser levados em consideração nos rendimentos de elementos da matriz (como Ca, Mg e Si) para uma mesma composição de formação.

Apesar da dificuldade em determinar a concentração dos elementos químicos, as questões mencionadas podem ser resolvidas através da remoção dos rendimentos correspondentes a elementos presentes em grande parte nos fluidos, como H, O, C e Cl (Hertzog

et al., 1989) e a posterior estimativa das concentrações dos elementos presentes na matriz da rocha.

2.2.5. Perfis Sônicos

Também conhecidos como “perfis acústicos”, os perfis sônicos levam em consideração as medidas das vagarosidades das ondas P (compressivas) e S (cisalhantes) propagadas na formação, sendo muito utilizados no cálculo de porosidade, identificação litológica, identificação e quantificação de gás, determinação das propriedades mecânicas da rocha, estimativa da pressão de formação, calibração sísmica e avaliação da cimentação de poços revestidos (Ellis & Singer, 2007). As ferramentas mais modernas utilizadas para sua aquisição são adquiridas centralizadas em relação ao poço (Figura 13) e podem incluir mais de dez receptores, o que reduz os efeitos de parede de poço e compromete a resolução vertical

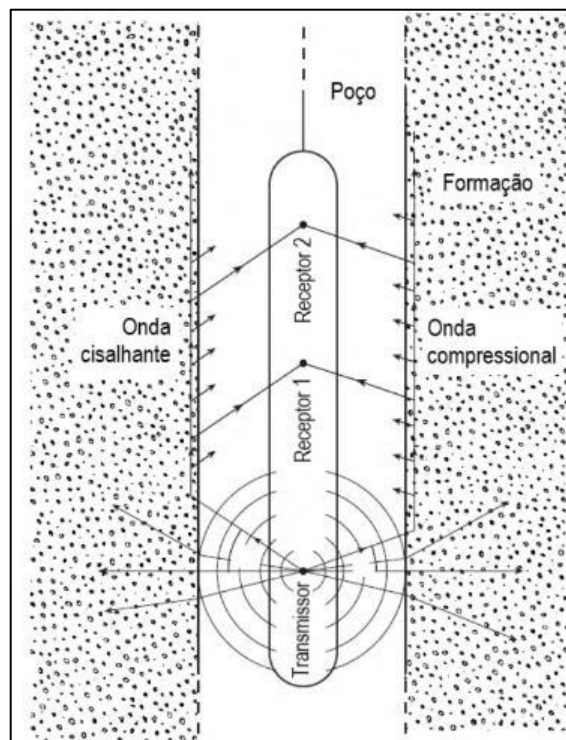


Figura 13: Exemplo de ferramenta de aquisição dos perfis acústicos, sendo adquirida centralizada em relação ao poço. | Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

Quanto a aquisição de dados, o transmissor emite uma onda acústica que se propaga no interior do poço até chegar na formação, onde a onda continua se propagando até formar uma frente de onda com um ângulo de 90° em relação à parede do poço (Figura 14). Distúrbios de pequena amplitude gerados na parede do poço produzem uma nova frente de onda que vai se propagar pelo fluido de perfuração do poço com uma velocidade semelhante à da onda que se

propaga na formação. A velocidade da onda acústica é obtida pela diferença do tempo de detecção da nova frente de onda entre os receptores.

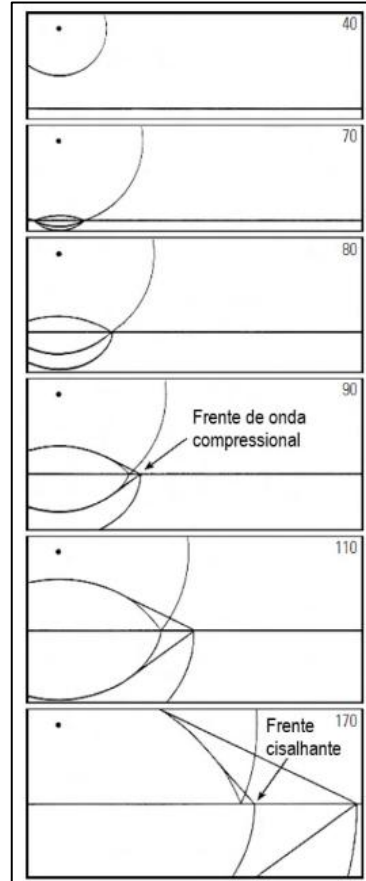


Figura 14: Simulação 2D de uma onda acústica se propagando do poço para a formação. Em 90 μs é possível observar que a frente de onda da formação faz um ângulo de 90° com a parede do poço, formando uma nova frente de onda que se propaga pelo fluido de perfuração do poço com a mesma velocidade da onda na formação.
| Fonte: Ellis & Singer (2007) e modificado por Oliveira (2022).

O cálculo da porosidade da formação pode ser feito através da vagarosidade da onda P utilizando a seguinte equação (Wyllie; Gregory; Gardner, 1956):

$$\Delta t = \phi \Delta t_f + (1 - \phi) \Delta t_{ma} \quad \text{Equação (5)}$$

Onde a porosidade é definida por ϕ , a vagarosidade do fluido presente nos poros por Δt_f e a vagarosidade da matriz da rocha por Δt_{ma} . É importante destacar que a equação (5) só é aplicável em formações altamente compactadas.

2.2.6. Perfis de Ressonância Magnética Nuclear

Os perfis de ressonância magnética nuclear são muito utilizados para estimar a porosidade total da formação independente da matriz. Além disso, também é possível analisar as propriedades dos fluidos presentes na formação, determinar o tamanho dos poros e calcular a permeabilidade absoluta da rocha de forma mais precisa, sendo tudo isso possível através da análise da distribuição do tempo de relaxação (T_2) medida em milissegundos (Oliveira, 2022).

Quanto à aquisição de dados, a ferramenta de ressonância magnética nuclear realiza as leituras muito próximas à parede do poço. Essas medições são significativamente afetadas pela invasão da lama de perfuração que, nos poços perfurados no pré-sal, é principalmente constituída de lama à base óleo (Coates *et al.*, 1999; SLB, 2015). O processo de leitura da ferramenta é feito através do experimento Carr-Purcell-Meiboom-Gill (Carr & Purcell, 1954; Meiboom; Gill, 1958), que pode ser descrito em três momentos: (a) polarização do momento magnético dos átomos de hidrogênio presentes na rocha a partir de um campo magnético constante B_0 ; (b) excitação dos átomos de hidrogênio com um campo magnético B_1 provocado por pulsos de rádio frequência oscilando na frequência ressonante dos átomos; e (c) remoção de B_1 e leitura do decaimento do sinal magnético emitido pelos átomos à medida que retornam ao seu estado original. A partir disso, o decaimento medido é modelado como a combinação linear de vários decaimentos exponenciais (Oliveira, 2022), gerando então o espectro de T_2 .

O T_2 é muito utilizado no estudo de reservatórios carbonáticos do pré-sal brasileiro, pois diferentes tipos de fluidos podem ser identificados e quantificados no espectro T_2 através de valores específicos. Ao analisar rochas carbonáticas, por exemplo, valores de T_2 abaixo de 100 ms são associados a fluidos irreduzíveis e retidos pela capilaridade do reservatório, enquanto valores de T_2 acima de 100 ms caracterizam fluidos livres e produzíveis (Chang *et al.*, 1994; Coates *et al.*, 1997; Straley *et al.*, 1997). É possível dizer, então, que a correlação entre esses tipos de fluidos está diretamente relacionada com o tamanho dos poros e, conseqüentemente, com a permeabilidade da rocha.

2.2.7. Perfis de Imagem

De uma maneira geral, os perfis de imagem representam fotografias em alta resolução da parede do poço levando em consideração características resistivas e acústicas das formações. Informações como inclinação de camadas, ocorrência de fraturas, desmoronamentos, direção de tensões *in situ*, litotipos, discordâncias, camadas delgadas e zonas com porosidade vugular são alguns dados que podem ser obtidos através da aquisição de perfis de imagem (Félix *et al.*, 2013). Tais informações são muito utilizadas em conjunto com os perfis petrofísicos

caracterizados desde o Capítulo 2.2.1 ao Capítulo 2.2.6 deste texto, sendo importantes para identificar se o reservatório possui condições de viabilidade econômica ou não, além de estimativas de volume e aproveitamento do reservatório.

Nos próximos Capítulos serão detalhados dois tipos de perfis de imagem utilizados no poço 9-ATP-1-RJS.

2.2.7.1. Perfis de imagem resistiva

Esses tipos de perfis medem as características resistivas de um determinado intervalo rochoso a partir da passagem de corrente elétrica. Estas informações são obtidas através de medições microrresistivas de altíssima resolução muito próximas da parede do poço, fornecendo informações detalhadas da formação (Félix *et al.*, 2013).

Quanto aos dados, estes são adquiridos através de uma ferramenta descida a cabo que, ao atingir a profundidade de interesse, pressiona seus patins contra a parede do poço. Ao ser posicionada, a ferramenta percorre o poço emitindo corrente elétrica pela sua parede e com uma velocidade determinada de acordo com a resolução desejada. Quanto maior a velocidade de subida da ferramenta menor será a resolução do perfil, pois serão registradas menores quantidades de pulsos elétricos pelo sensor (Félix *et al.*, 2013). Por fazerem medições muito próximas da parede do poço, a imagem resistiva pode ser adquirida em poços perfurados tanto com fluido base óleo ou base água.

Após a aquisição de dados, estes são processados por programas especializados que convertem as informações em imagens gráficas, facilitando o processo de interpretação (Félix *et al.*, 2013). Como resultado, de uma maneira geral, feições de baixa resistividade (folhelhos ou fraturas preenchidas por fluido condutivo) costumam apresentar cores escuras no perfil, enquanto feições de alta resistividade (arenitos e calcários saturados por hidrocarbonetos) apresentam cores que variam entre o marrom, amarelo e branco (Hurley, 2004). A Figura 15B representa um exemplo de um perfil de imagem resistiva.

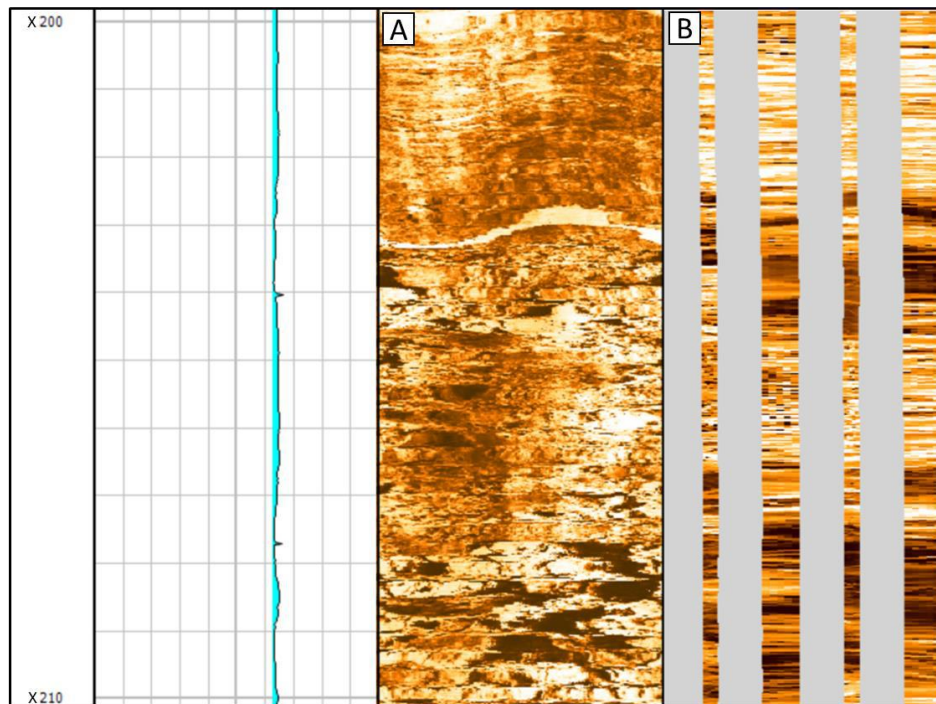


Figura 15: Perfis de Imagem do poço 9-ATP-1-RJS. (A) Perfil de Imagem acústica. (B) Perfil de Imagem resistiva. Analisando os dois perfis é possível perceber que as ferramentas de imagem acústica retratam 100% da parede do poço, enquanto as ferramentas de imagem resistiva retratam uma porcentagem menor. | Fonte: autora.

2.2.7.2. Perfis de imagem acústica

Esses perfis identificam as propriedades da formação levando em consideração a amplitude e o tempo de trânsito das reflexões de pulsos acústicos emitidos na parede do poço (Félix *et al.*, 2013). A ferramenta responsável pela aquisição de dados desce no poço até atingir a profundidade desejada e, durante o trajeto, é importante que esteja centralizada para não interferir no tempo de trânsito. Um dispositivo presente na ferramenta, conhecido como “transdutor”, é ligado e rotacionado, emitindo pulsos acústicos na formação e registrando seus reflexos.

O processamento de dados é feito por programas especializados e como resultado tem-se a imagem acústica, gerada pela amplitude e tempo de trânsito dos pulsos sonoros ao serem refletidos pela formação (Félix *et al.*, 2013). Em geral, formações com baixa amplitude ou tempo de trânsito prolongado (folhelhos, dissoluções, desmoronamentos e fraturas preenchidas por fluidos) são imageadas em cores escuras, enquanto feições com alta amplitude ou tempo de trânsito curto (arenitos e carbonatos) são imageados em tons de marrom, amarelo e branco (Hurley, 2004).

As ferramentas de aquisição de perfis acústicos também não possuem restrição quanto à resistividade do fluido presente no poço. Além disso, as ferramentas de imagem acústica conseguem retratar 100% da parede do poço, em contraste com as ferramentas de imagem

resistiva que normalmente retratam uma porcentagem menor da parede do poço devido à disposição dos seus receptores de dados (Félix *et al.*, 2013). Na Figura 15 é possível observar essa diferença entre os dois perfis.

2.3. Aprendizado de Máquina

O aprendizado de máquina pode ser definido como um campo de estudo que instrui os computadores a tratar os dados de forma mais eficiente e adquirir conhecimento sem a necessidade da definição de modelos de forma explícita (Mahesh, 2018). Uma das suas principais vantagens é que a partir do momento em que o algoritmo aprende a trabalhar com os dados, a tarefa passa a ser efetuada de forma automática pelos computadores.

Os algoritmos de aprendizado de máquina estão incorporados em numerosas atividades que fazem parte do nosso cotidiano. Eles são poderosas ferramentas para identificar correlações e padrões estruturais em grandes bases de dados, sendo posteriormente utilizados para fazer previsões com novos valores de entrada (Witten & Frank, 2005). Existem diferentes algoritmos para solucionar problemas de dados e sua escolha depende do tipo de problema a ser resolvido, do número de variáveis, do tipo de modelo mais adequado e assim por diante (Mahesh, 2018). No presente estudo foram utilizados dois tipos de algoritmos de aprendizado de máquina que serão melhor detalhados nos Capítulos 2.3.1. e 2.3.2.

No que diz respeito ao aprendizado de máquina aplicado a perfis de poço, pode-se dizer que nos últimos anos sua aplicação tem aumentado consideravelmente na indústria de óleo & gás (Oliveira, 2022). As aplicações de seus algoritmos, nesse caso, podem ser divididas em três tipos: previsão, interpretação e imputação de dados. Segundo Oliveira (2022), o primeiro tipo calcula os parâmetros petrofísicos de interesse a partir de perfis de poço e aprendizado de máquina. Em geral, esses modelos fazem o uso dos perfis de poços como entrada e os parâmetros petrofísicos medidos em laboratório como saída. O segundo tipo pode ser dividido em classificação litológica (aplica o método supervisionado quando há amostras de rocha disponíveis e o método não supervisionado quando não há amostras de rocha) ou petrofísica (identifica automaticamente as zonas de interesse dividindo o poço em aquíferos, zonas produtivas e zonas portadoras de hidrocarbonetos). Por último, a imputação de dados utiliza o aprendizado de máquina para simular os dados de perfis em intervalos onde apresentaram baixa qualidade de aquisição (Oliveira, 2022).

Dois métodos de aprendizado de máquina são amplamente utilizados no contexto de resolução de problemas de geociências (Lima, 2019) e ambos foram empregados neste estudo: aprendizado supervisionado e aprendizado não supervisionado. Segundo Bishop (2006), o

aprendizado supervisionado é utilizado quando seu propósito é prever uma variável de saída com base em um conjunto de variáveis de entrada. Ocorre, então, um processo de calibração conhecido como “treinamento” no qual o algoritmo realiza o ajuste de uma série de parâmetros específicos para identificar como as variáveis de entrada se relacionam entre si, a fim de produzir estimativas mais precisas das variáveis de saída. Já o aprendizado não supervisionado é aquele cuja a base de dados não apresenta variáveis de saída. Neste caso, o algoritmo ajustará uma série de parâmetros para subdividir as instâncias em grupos que compartilham alguma semelhança entre si e reduzir a dimensionalidade das variáveis de entrada (Bishop, 2006).

2.3.1. *Self Organizing Maps (SOM)*

Esse tipo de algoritmo, proposto por Kohonen (1997), é treinado de forma não supervisionada e tem como princípio o agrupamento e a visualização de dados. Seu principal objetivo é transformar um espaço de entrada complexo em um espaço de saída mais simples, preservando as relações de semelhança entre os dados e gerando um mapa auto-organizado que pode ser usado tanto em problemas de aprendizado de máquina supervisionado e não supervisionado (Asan & Ercan, 2012).

O sistema do SOM é fundamentado no princípio do aprendizado competitivo. Os neurônios, também conhecidos como “nós”, competem para determinar qual deles será ativado em resposta a um conjunto específico de entradas, sendo então designado como o “vencedor” (*winner*). Os neurônios vencedores têm a capacidade de reduzir a atividade dos seus vizinhos (inibições laterais) a fim de dar um feedback negativo a eles (Asan & Ercan, 2012).

Kohonen configurou a rede SOM em duas camadas de neurônios: a camada de entrada e a camada de saída (Figura 16). Os neurônios da camada de entrada representam as variáveis contidas nos dados de entrada, sendo cada um deles representado por um vetor de entrada dimensional cujos elementos indicam os valores dos atributos de um determinado conjunto de dados. A camada de entrada dos neurônios está diretamente ligada à camada de saída, fazendo com que o SOM seja uma rede completamente conectada (Asan & Ercan, 2012). A camada de saída representa uma visualização de baixa dimensão dos dados. Seus neurônios são dispostos sob a forma de uma estrutura bidimensional organizada em linhas e colunas. O número de neurônios na camada de saída indica o número máximo de agrupamentos de dados, o que induz a precisão do SOM.

Após o treinamento feito de forma não supervisionada, os neurônios do mapa auto-organizado resultante podem ser agrupados de forma supervisionada ou não supervisionada através de algoritmos de agrupamento de dados (processos de cluterização).

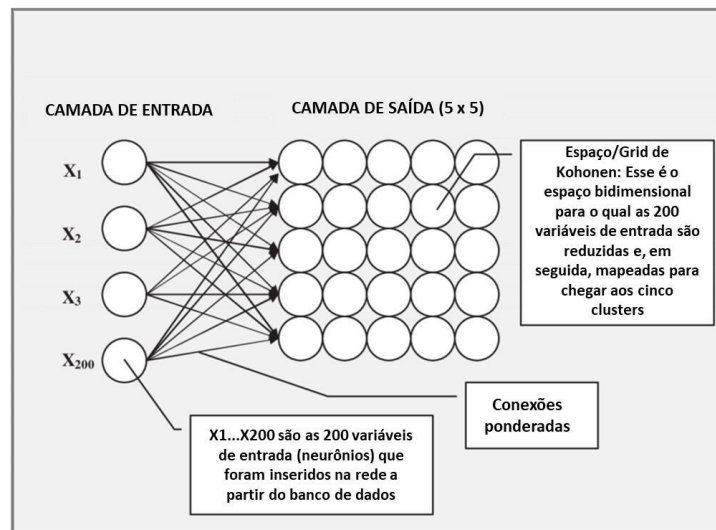


Figura 16: Exemplo de um self organizing-map de Kohonen, pesquisador responsável pela criação do algoritmo. | Fonte: modificado de Castellani & Castellani (2003).

2.3.2. *Categorical Boosting (Catboost)*

CatBoost é um algoritmo de aprendizagem de máquina supervisionado que utiliza a técnica de *gradient boosting* para classificar os dados através de árvores de decisão. Essas árvores utilizadas pelo *CatBoost* também são conhecidas como *symmetric trees* (árvores simétricas), pois utilizam o mesmo critério de divisão de dados em todos os níveis da árvore. Uma série de modelos de árvores são construídos dividindo o conjunto de dados em instâncias semelhantes e cada árvore subsequente é construída com o objetivo de minimizar a diferença entre os dados preditos e os dados de treinamento, dada por uma função custo, a cada interação, melhorando a precisão do algoritmo como um todo (Cao *et al.*, 2021).

O *gradient boosting* é conhecido como uma das abordagens mais eficazes para a criação de modelos de conjunto. A combinação do *gradient boosting* com árvores simétricas fornece excelentes resultados em várias aplicações que envolvem dados estruturados (Ershov, 2018). Essas técnicas são empregadas para construir modelos de conjuntos de maneira interativa. Na primeira interação, o algoritmo aprende a primeira árvore com o objetivo de reduzir o erro de treino, conforme ilustrado na Figura 17A. Geralmente, esse primeiro modelo apresenta um erro considerável.

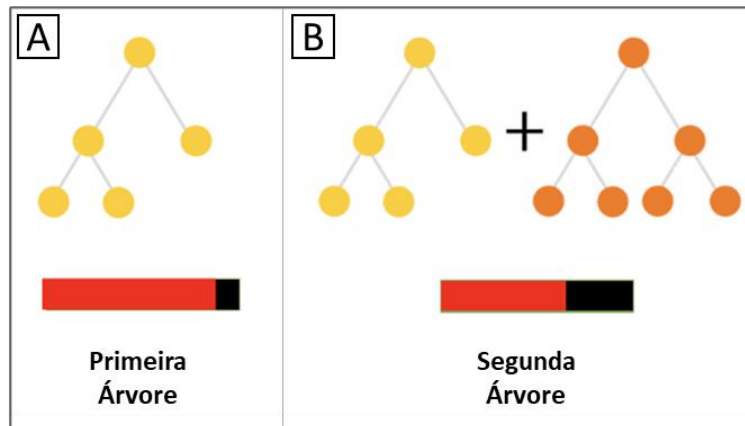


Figura 17: A primeira e a segunda árvore normalmente apresentam erros significativos. | Fonte: modificado de Ershov (2018).

Na Figura 17B é possível observar a segunda interação, na qual o algoritmo aprende mais uma árvore para reduzir o erro cometido pela primeira. O algoritmo repete este procedimento até construir um modelo de boa qualidade, como pode ser visto na Figura 18.

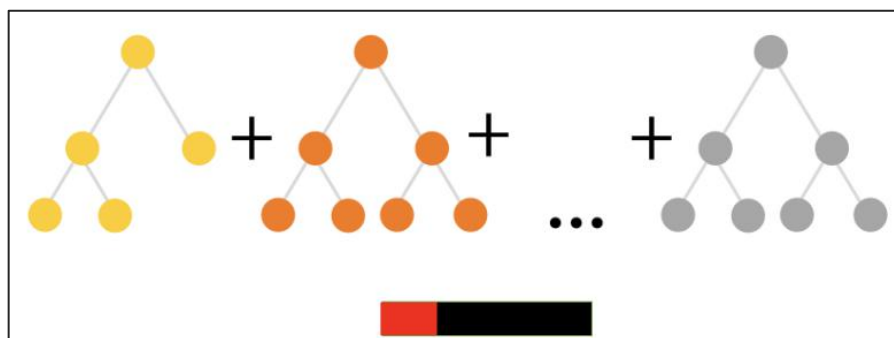


Figura 18: Repetição da construção das árvores até gerar um modelo adequado. | Fonte: modificado de Ershov (2018).

Existem diversos tipos de árvore de decisão no campo de aprendizado de máquina, mas o *CatBoost* utiliza o modelo de árvores simétricas devido às várias vantagens que elas podem oferecer, tais como: esquema de ajuste simples (o treinamento das árvores é direto e facilita a configuração do modelo); implementação eficiente na unidade central de processamento (CPU); capacidade de criar modelos muito rápidos. Essa estrutura de árvore funciona como uma forma de regularização, o que pode proporcionar benefícios em termos de qualidade de previsão para várias tarefas e minimizar *overfitting* (Ershov, 2018).

3. MATERIAIS E MÉTODOS

Os dados utilizados nesta pesquisa foram fornecidos pela ANP e consistem em perfis adquiridos por sete ferramentas consideradas convencionais (raios gama, densidade, fator fotoelétrico, nêutrons, geoquímico, sônico e ressonância magnética nuclear) e perfis adquiridos por dois tipos de ferramentas de imagem (perfil de imagem acústica e resistiva) da empresa SLB. O perfil de resistividade também foi adquirido; porém, como ele responde principalmente aos fluidos presentes na formação, ele não foi utilizado. Além disso, também foi fornecida a descrição de oito fácies sedimentares identificadas no intervalo de interesse através de análises visuais de testemunho, lâminas petrográficas de plugues amostrados a cada 30 cm no testemunho, análises de composição mineral desses plugues e análises estratigráficas realizadas por especialistas. Esta descrição foi empregada como referência para a comparação final entre as classificações derivadas dos perfis convencionais e dos perfis de imagem. A distribuição dos dados para cada uma dessas fácies pode ser observada no histograma da Figura 19 e é importante destacar que não foi realizada nenhuma ponderação para lidar com um possível desbalanceamento da base de dados. As Figuras 20 e 21 apresentam os perfis utilizados no trabalho.

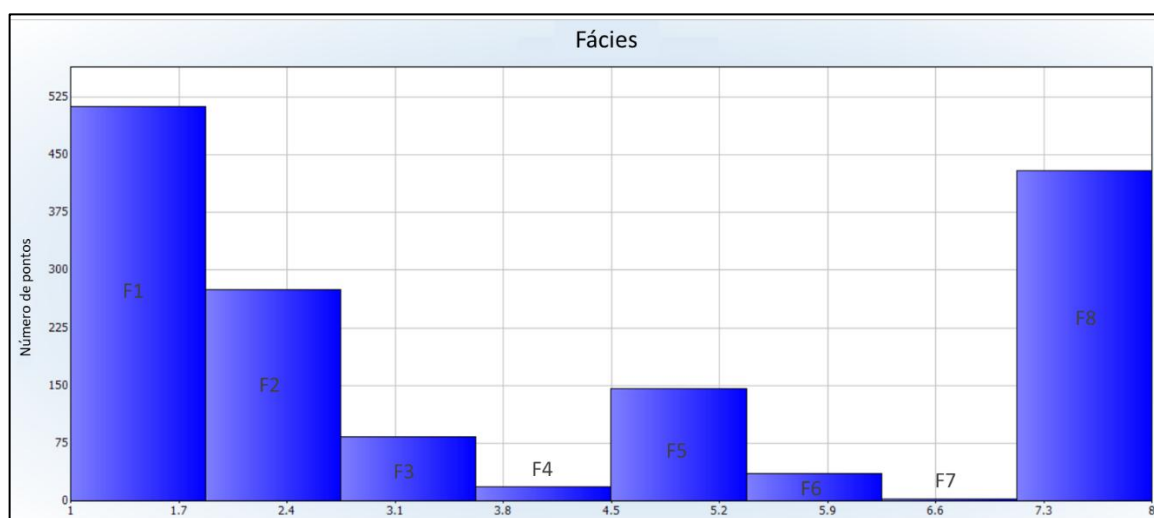


Figura 19: Histograma com a distribuição de dados para cada fácies. F1 – fácies 1; F2 – fácies 2; F3 – fácies 3; F4 – fácies 4; F5 – fácies 5; F6 – fácies 6; F7 – fácies 7; F8 – fácies 8. | Fonte: autora, a partir do software IP (2023).

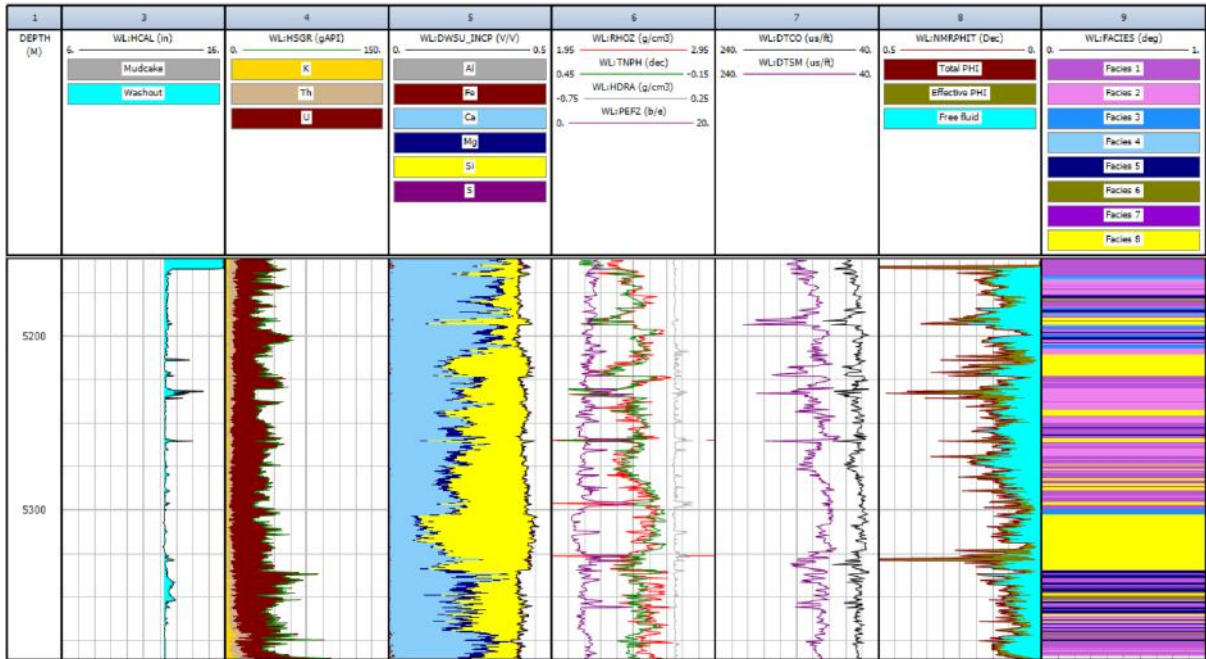


Figura 20: Perfis convencionais utilizados no trabalho: Coluna 3 – Perfil de diâmetro do poço (CAL); Coluna 4 – Perfil de raios gama total (HSGR) e de raios gama espectral: potássio (amarelo), tório (bege) e urânio (marrom); Coluna 5 – Perfis geoquímicos: alumínio (cinza), ferro (marrom), cálcio (azul claro), magnésio (azul escuro), silício (amarelo) e enxofre (roxo); Coluna 6 – Perfis de densidade (RHOZ), nêutrons (TNPH) e fator fotoelétrico (PEFZ); Coluna 7 – Perfis sísmicos: vagarosidade da onda compressional (DTCO) e cisalhante (DTSM); Coluna 8 – Perfis de ressonância magnética nuclear: porosidade total (marrom), porosidade efetiva (oliva) e fluido livre (azul); Coluna 9 – Fácies. | Fonte: autora, a partir do software IP (2023).

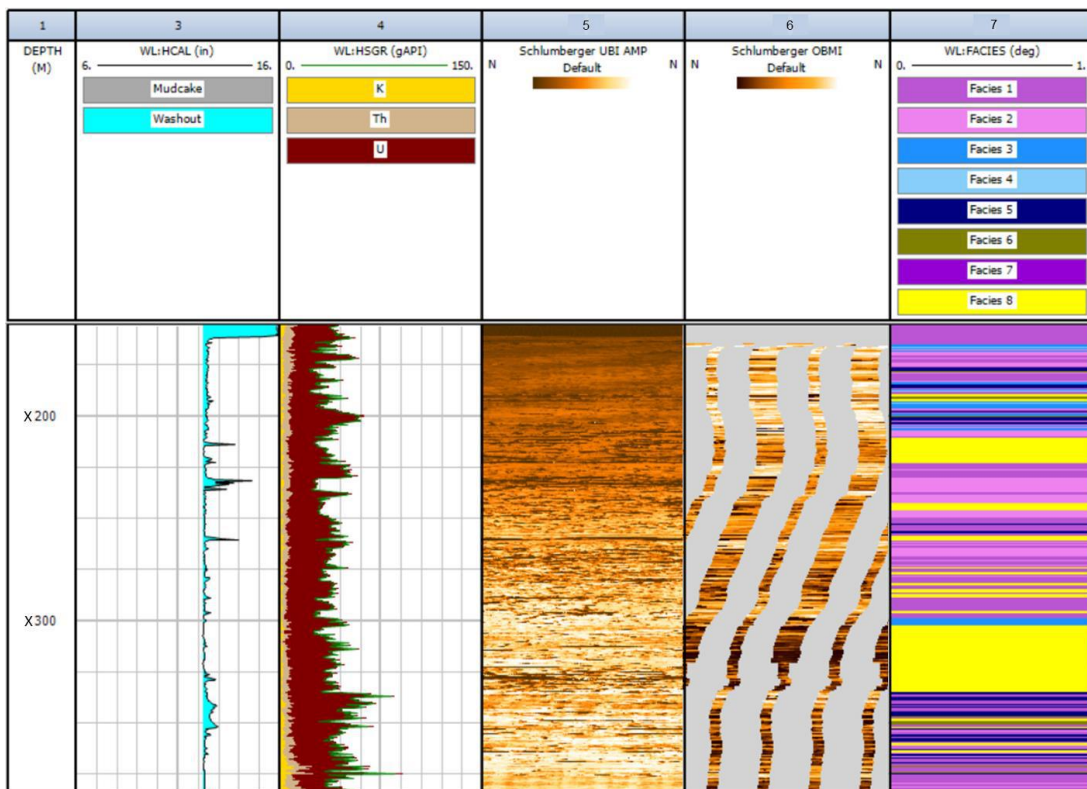


Figura 21: Perfis de imagem utilizados no trabalho: Coluna 3 – Perfil de diâmetro do poço (CAL); Coluna 4 – Perfil de raios gama total (HSGR) e de raios gama espectral: potássio (amarelo), tório (bege) e urânio (marrom); Coluna 5 – Perfil de imagem acústica; Coluna 6 – Perfil de imagem resistiva; Coluna 7 – Fácies. | Fonte: autora, a partir do software IP (2023).

O trabalho foi majoritariamente desenvolvido utilizando as ferramentas do software *Interactive Petrophysics* (IP) versão 2023 (5.1.2), fornecido pelo Laboratório de Geologia Sedimentar da UFRJ. Os métodos desenvolvidos buscaram realizar a clusterização de perfis de poços, tanto de forma supervisionada e não supervisionada, e compará-la com os dados de fácies descritas em rocha.

O método adotado pode ser subdividido em três análises principais:

- I) **Análise de perfis convencionais:** essa etapa teve como principal objetivo a classificação das fácies sedimentares com base apenas nos perfis convencionais, de forma não supervisionada e supervisionada. A forma não supervisionada utilizou o algoritmo SOM disponibilizado pelo IP, e a forma supervisionada utilizou os algoritmos SOM e CatBoost. Os modelos foram avaliados através de matrizes de confusão, comparando as fácies reais, isto é, aquelas oriundas da descrição e interpretação dos testemunhos, e as fácies previstas interpretadas pelos algoritmos.
- II) **Análise de perfis de imagem:** essa etapa consistiu na classificação de fácies sedimentares com base apenas nos perfis de imagem, fazendo uso do módulo Análise Textural de Fácies presente no IP. Esse módulo inclui as etapas de transformação para frequência e análise estatística antes da aplicação dos algoritmos de aprendizado de máquina. Essas etapas buscam extrair dos perfis de imagem parâmetros texturais que podem então ser incorporados nos algoritmos. Assim como na análise de perfis convencionais, os algoritmos foram treinados de forma não supervisionada (SOM) e supervisionada (SOM e CatBoost), e os resultados também foram avaliados através de matrizes de confusão.
- III) **Análise integrada de perfis convencionais e de imagem:** essa etapa consistiu na classificação de fácies sedimentares com base nos perfis convencionais e nas informações texturais extraídas pelos perfis de imagem. Ela foi feita de forma supervisionada utilizando apenas o algoritmo CatBoost, e os resultados foram avaliados através de matrizes de confusão.

A utilização de dois algoritmos de aprendizado de máquina (SOM e CatBoost) buscou garantir que os diferentes resultados obtidos nas diversas análises estavam relacionados aos diferentes dados de entrada utilizados (perfis convencionais e de imagem), e não a possíveis dificuldades de um ou outro algoritmo de se ajustar aos dados de entrada. As etapas do projeto

foram resumidas no fluxograma apresentado na Figura 22 e abordadas detalhadamente nos capítulos seguintes.

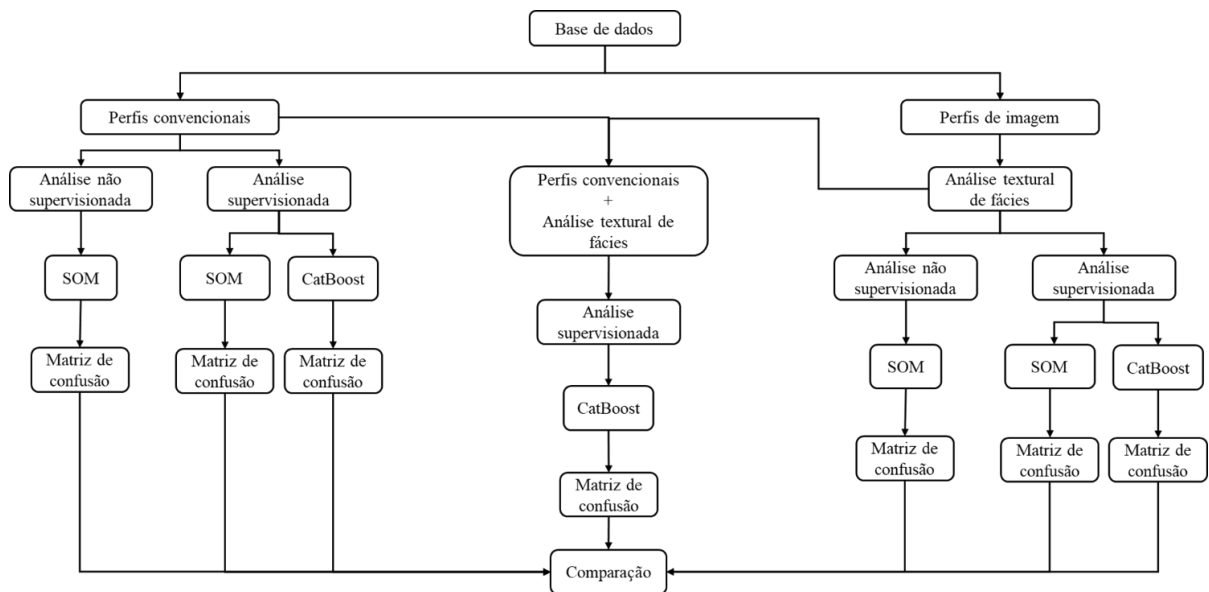


Figura 22: Fluxograma das três etapas do estudo. | Fonte: autora.

3.1. Análise dos Perfis Convencionais

3.1.1. Treinamento do SOM: ajuste do mapa auto-organizado

Inicialmente, os neurônios do mapa auto-organizado foram treinados para se ajustarem ao espaço de variáveis n-dimensional dado pelos perfis convencionais, onde cada perfil foi considerado uma variável. Os perfis utilizados são apresentados na Tabela 2. Nessa etapa, foi possível configurar o tamanho, a forma e os parâmetros de treino do SOM, bem como determinar o número de iterações de treinamento a serem realizadas. Os diferentes parâmetros e geometrias capazes de serem ajustados nessa etapa são: geometria do SOM (quadrado, hexagonal ou esférico), tamanho do mapa (quantidade de neurônios), número de iterações de treino, e taxa de aprendizagem inicial. Os critérios utilizados na construção do SOM estão descritos na Tabela 3 e foram os parâmetros padrão do IP, uma vez que buscou-se fazer uma análise comparativa das técnicas dos algoritmos utilizados.

Perfil	Descrição	Ferramenta
DWAL_INCP	Perfis geoquímicos	Ferramenta geoquímica
DWFE_INCP		
DWCA_INCP		
DWWMG_INCP		
DWSI_INCP		
RHOZ	Perfil de densidade	Ferramenta de densidade
PEFZ	Perfil de fator fotoelétrico	
HSGR	Perfil de raios gama total	Ferramenta de raios gama total
K	Perfil de concentração de potássio	Ferramenta de raios gama espectral
Th	Perfil de concentração de tório	
U	Perfil de concentração de urânio	
TNPH	Perfil de nêutrons	Ferramenta de porosidade de nêutrons
DTCO	Perfil de vagarosidade da onda compressional	Ferramenta acústica
DTSM	Perfil de vagarosidade da onda cisalhante	
NMRPHIT	Perfil de porosidade total	Ferramenta de ressonância magnética nuclear
NMRPHIE	Perfil de porosidade efetiva	
NMRFF	Perfil de fluido livre	

Tabela 2: Perfis convencionais utilizados como variável de entrada para o ajuste do mapa auto-organizado. | Fonte: autora.

Geometria do SOM	Tamanho do Mapa (número de neurônios)	Número de Iterações de Treino	Taxa de Aprendizagem Inicial
Esférico	642	60000	0.1

Tabela 3: Critérios utilizados para treinar o SOM. | Fonte: autora.

O SOM foi treinado apenas uma vez utilizando os dados de poços como entrada. Após o treinamento, foi gerado um mapa auto-organizado conforme os parâmetros descritos na Tabela 3.

3.1.2. Análise Não Supervisionada: clusterização do mapa auto-organizado

Nessa etapa, empregou-se o método de clusterização hierárquica para agrupar os neurônios treinados na etapa 3.1.1 em conjuntos de similaridade, identificando os padrões existentes entre eles sem a necessidade de uma pré-determinação sobre como esses padrões deveriam ser criados. A Tabela 4 apresenta os parâmetros utilizados para clusterizar os neurônios treinados:

Métrica de Clusterização	Número de Grupo de Clusters	Curva Utilizada para Ordenar os Valores
Menor Soma do Quadrado das Distâncias Internas ao Cluster	8	Todos os Perfis

Tabela 4: Critérios utilizados para clusterizar os dados. | Fonte: autora.

O número ideal de agrupamentos (clusters) pode ser determinado com base na razão de aleatoriedade para um determinado número de clusters, calculando primeiramente o número médio de profundidade por cluster, ou seja, a espessura média de uma camada de grupo. Em seguida, é calculada a espessura média teórica, assumindo que os clusters são atribuídos aleatoriamente em cada nível de profundidade. A razão de aleatoriedade é a relação entre as duas espessuras, dado pelas equações:

$$\text{Espessura aleatória} = \sum p_i / (1 - p_i), \quad \text{Equação (6)}$$

$$\text{Razão de aleatoriedade} = \frac{\text{Espessura média}}{\text{Espessura aleatória}} \quad \text{Equação (7)}$$

Onde p_i é a proporção de níveis de profundidade atribuídos ao i -ésimo agrupamento.

A razão de aleatoriedade é plotada num gráfico em relação ao número total de clusters para avaliação visual (Figura 23). Quanto mais elevado for o valor (picos mais altos), menos aleatórios são os agrupamentos – ou seja, dados mais estruturados. Além disso, quando a linha do gráfico apresenta uma queda acentuada, isso indica um aumento significativo na aleatoriedade e o grupo imediatamente anterior à essa queda seria uma escolha apropriada de grupos de clusters.

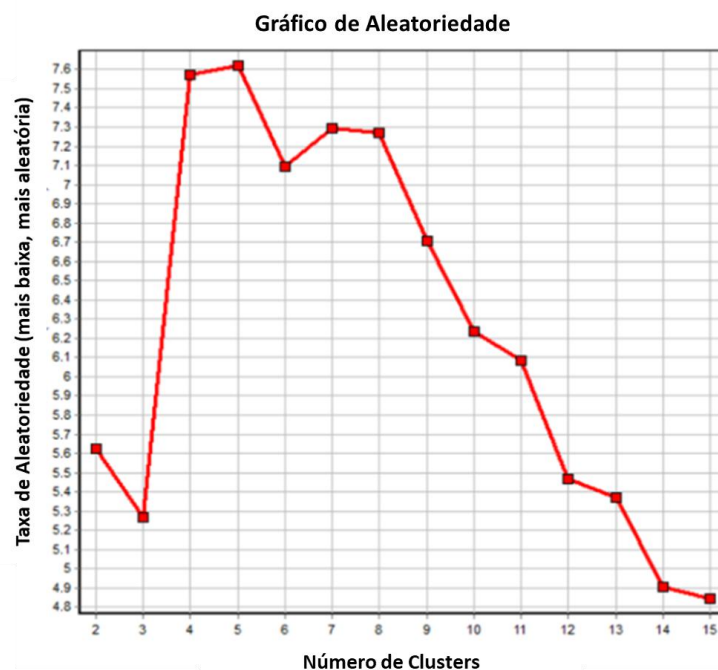


Figura 23: Exemplo de um gráfico de aleatoriedade apresentando a relação entre o número de agrupamentos e a razão de aleatoriedade. | Fonte: Al-Baldawi (2020).

Em um problema não supervisionado de classificação de fácies, seria desejado que o número ideal de clusters, dado pela menor aleatoriedade, fosse igual ou muito próximo ao número de fácies descrito. Caso contrário, existiria um indicativo muito forte de que os perfis utilizados no treinamento não seriam capazes de representar as fácies observadas.

3.1.3. Análise Supervisionada: calibração do mapa auto-organizado

Na análise supervisionada, o SOM treinado na etapa 3.1.1 foi calibrado utilizando a curva de fácies como variável de saída, enquanto os perfis convencionais foram utilizados como variáveis de entrada para o algoritmo.

3.1.4. Análise Supervisionada: *CatBoost*

Nessa etapa foi empregado o algoritmo *CatBoost* como uma redundância para avaliar a capacidade dos modelos de aprendizado de máquina de prever as fácies observadas no poço. Porém, o *CatBoost* permite realizar o treinamento apenas de forma supervisionada. O *CatBoost* recebeu diretamente os perfis convencionais como entrada (Tabela 2), sem a necessidade de um treinamento prévio, como é o caso do SOM.

A Tabela 5 apresenta os parâmetros padrão utilizados pelo algoritmo, uma vez que o objetivo é comparar os resultados e não otimizar os algoritmos.

Parâmetros do <i>CatBoost</i>	
Eval_metric	MultiClass
Iterations	500
Depth	5
Min_data_in_leaf	1
Learning_rate	1
Max_leaves	32

Tabela 5: Parâmetros utilizados no *CatBoost* para testar o modelo gerado através dos perfis convencionais. |
Fonte: autora.

3.2. Análise dos Perfis de Imagem

É importante destacar que antes de aplicar os métodos supervisionado e não supervisionado através do módulo Análise Textural de Fácies, foi preciso processar as imagens adquiridas pelas ferramentas para orientar as leituras com o norte magnético.

3.2.1. Análise Textural de Fácies: transformação para frequência

Nessa etapa, os dados dos perfis de imagem foram carregados e associados a uma ferramenta de imagem apropriada para sua análise. O intervalo de estudo foi estabelecido como a profundidade do topo igual a 5165 m e a da base igual a 5720 m, segundo as profundidades onde havia descrição de fácies. Após essas definições, o algoritmo segmentou as imagens em uma série de mosaicos de altura e largura fixas (Tabela 6) e foi realizada, então, uma transformação para o domínio da frequência para cada um desses mosaicos, conforme esquematizado na Figura 24. No domínio da frequência, o objetivo do algoritmo é ressaltar as diferentes estruturas e texturas das imagens, conforme exemplificado na Figura 25.

	Ferramenta de Imagem	Profundidade do Topo (m)	Profundidade da Base (m)	Largura (amostras)	Altura (amostras)	Altura (m)
Perfil de Imagem Acústica	Schlumberger UBI AMP	5165	5720	180	60	0.3048
Perfil de Imagem Resistiva	Schlumberger OBMI	5165	5720	5	60	0.3048

Tabela 6: Parâmetros utilizados na etapa Transformação para Frequência. | Fonte: autora.

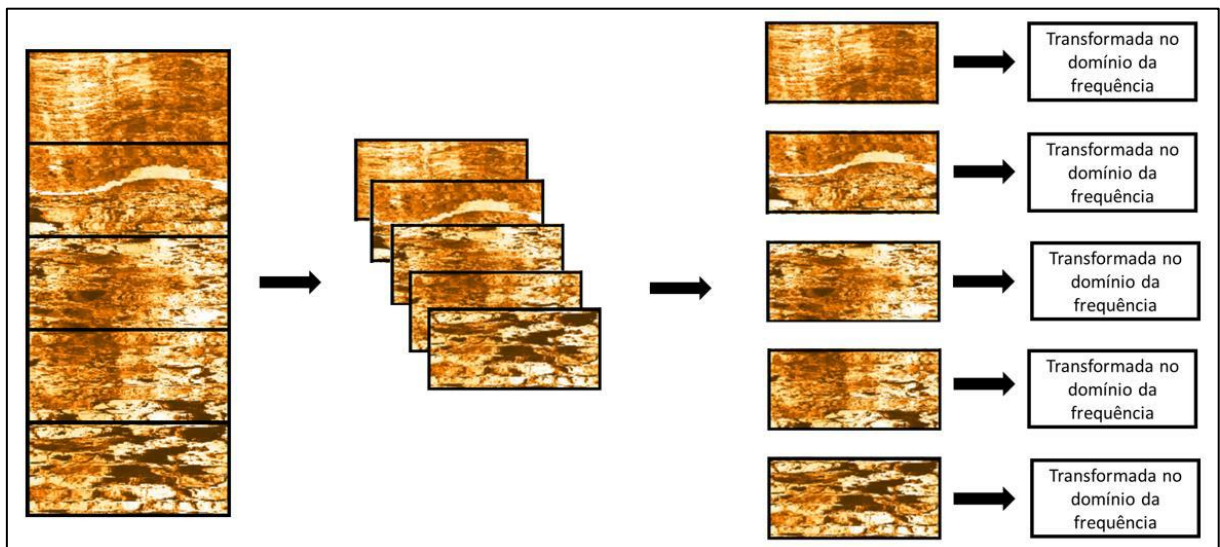


Figura 24: Esquema do processo de transformação para frequência realizado pelo algoritmo. O perfil de imagem original é segmentado em uma série de mosaicos de altura e largura fixos. Em seguida, cada um desses mosaicos passa por uma transformação para o domínio da frequência. | Fonte: autora.

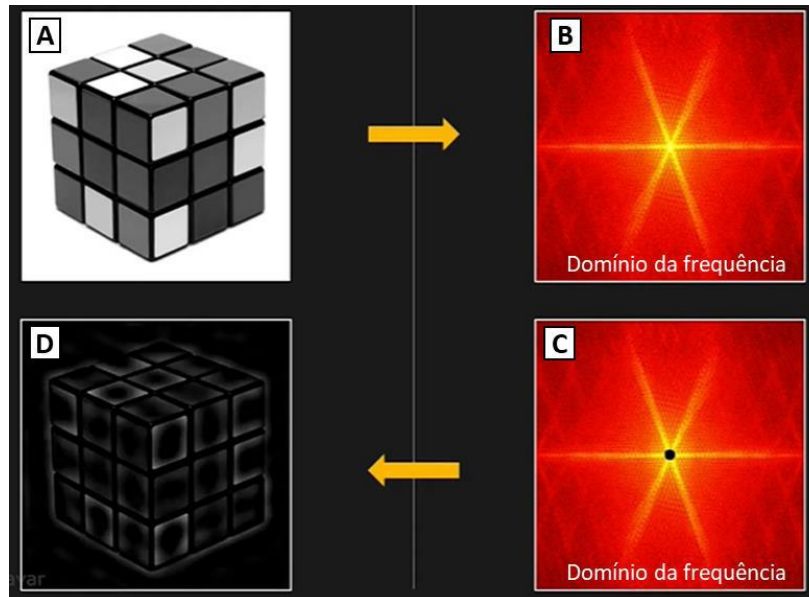


Figura 25: Exemplo de como funciona o domínio da frequência. Foi aplicada a transformada de frequência na figura original (A), gerando a figura (B) que resalta, através da frequência, certas propriedades da imagem original. As cores mais claras representam valores mais altos de frequência, ou seja, onde há mais informações. Na figura (C) foi retirada as informações do ponto central do domínio da frequência, recriando a imagem original (D) sem a informação que foi retirada e ressaltando algumas texturas e estruturas da imagem original. |

Fonte: modificado de Shree K. Nayar (2021).

3.2.2. Análise Textural de Fácies: análise estatística

Após realizar a transformação para o domínio da frequência, o algoritmo classificou o quanto de informação foi armazenada nesse domínio. Para isso, os mosaicos gerados na etapa anterior foram divididos em uma série de bins que aumentam de tamanho gradativamente. Esses bins podem aumentar seguindo o método de binarização, no qual pode ser utilizado o critério de duplicação (o tamanho do bin duplica até chegar no final do mosaico) ou incrementação (o tamanho do bin aumenta de 1 em 1). A Figura 26 exemplifica esse método que indica a taxa em que o bin irá aumentar dentro do mosaico.

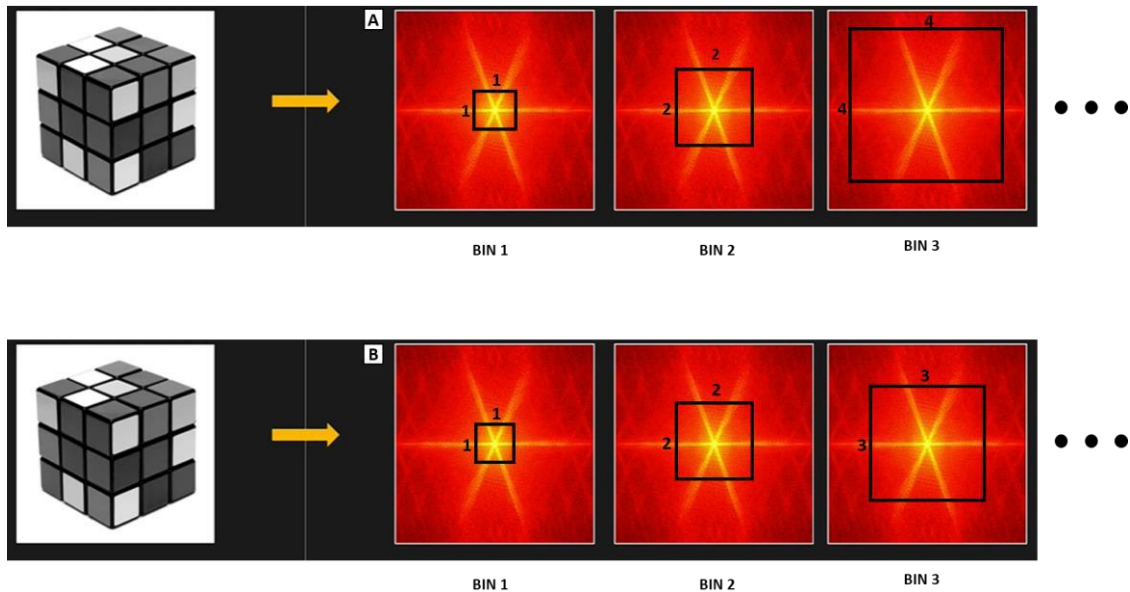


Figura 26: (A) critério da duplicação, onde o tamanho do bin aumenta de forma duplicada. Bin 1 – tamanho 1:1; Bin 2 – tamanho 2:2; Bin 3 – tamanho 4:4 e assim por diante. (B) critério de incrementação, onde o tamanho do bin aumenta de 1 para 1. Bin 1 – tamanho 1:1; Bin 2 – tamanho 2:2; Bin 3 – tamanho 3:3 e assim por diante. | Fonte: modificado de Shree K. Nayar (2021).

Em seguida, o algoritmo calculou a estatística de todos os valores dentro de cada bin utilizando os seguintes critérios: soma dos valores absolutos, média dos valores absolutos, o valor máximo ou o desvio padrão. A Tabela 7 apresenta o método e a estatística de binarização utilizados pelo algoritmo em cada perfil de imagem no presente trabalho:

	Método de Binarização	Estatística de Bins
Perfil de Imagem Acústica	Incrementação	Soma dos Valores Absolutos
Perfil de Imagem Resistiva	Duplicação	Soma dos Valores Absolutos

Tabela 7: Método e estatística de binarização utilizados nos perfis de imagem acústica e resistiva. | Fonte: autora.

Após este cálculo, o algoritmo classificou os bins em um gráfico de acordo com as estatísticas dos bins (desvio padrão ou valor máximo) ou componentes principais de todos os mosaicos. Os critérios escolhidos para essa classificação estão apresentados na Tabela 8. O gráfico gerou uma curva organizando os valores em ordem decrescente e, a partir dele, foi possível estabelecer um limite de número de bins onde se tem a maior parte das informações relevantes para serem introduzidas no SOM. Através da análise dos gráficos (Figuras 27 e 28), optou-se por escolher o número de bins igual a 10, tanto no perfil de imagem acústica quanto no perfil de imagem resistiva. Dessa forma, se interpretou que a maior parte das informações está antes do bin 10, guardando informações importantes quanto às texturas e estruturas

presentes nos mosaicos dos perfis de imagem. O restante dos bins provavelmente não estão associados a informações relevantes, estando mais suscetíveis a ruído.

Método de Classificação	
Perfil de Imagem Acústica	Estatística de Bins / Desvio Padrão
Perfil de Imagem Resistiva	Estatística de Bins / Desvio Padrão

Tabela 8: Método de classificação de bins utilizado nos perfis de imagem acústica e resistiva. | Fonte: autora.

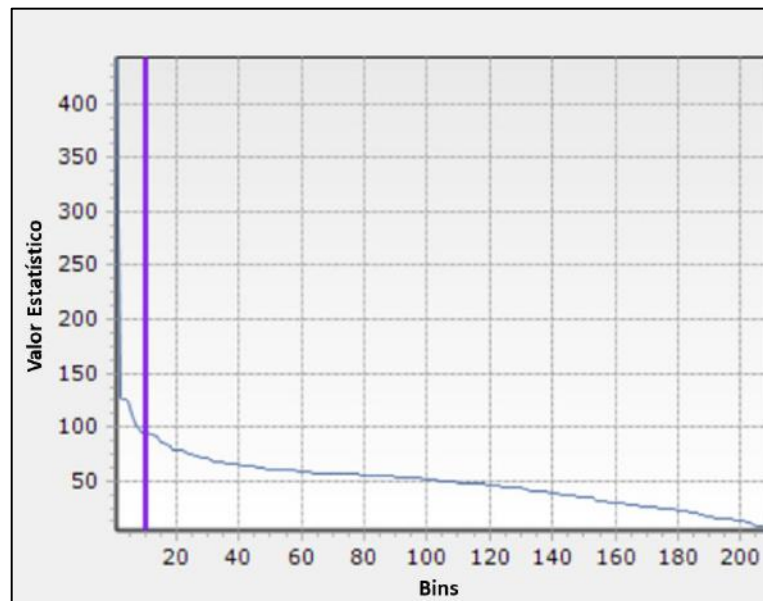


Figura 27: Curva de bins apresentando o número 10 como o número de bins ideal para o perfil de imagem acústica. | Fonte: autora, a partir do software IP (2023).

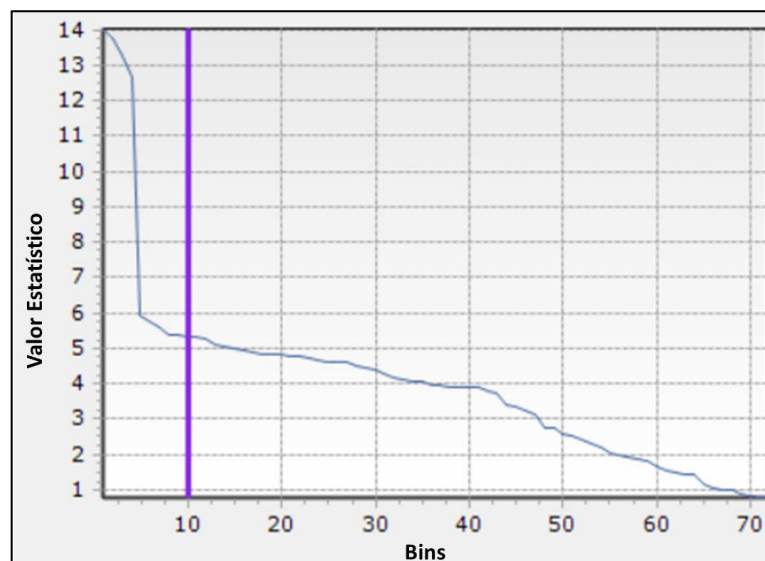


Figura 28: Curva de bins apresentando o número 10 como o número de bins ideal para o perfil de imagem resistiva. | Fonte: autora, a partir do software IP (2023).

3.2.3. Análise Textural de Fácies: treinamento do SOM

Os Bins extraídos da etapa anterior serviram de entrada para gerar um mapa auto-organizado. Seu ajuste foi similar às configurações da etapa 3.1.1, com exceção do número de neurônios (Tabela 9). Também foi possível definir a geometria do SOM, tamanho do mapa e números de interações de treino.

Geometria do SOM	Tamanho do Mapa (número de neurônios)	Número de Iterações de Treino
Esférico	362	60000

Tabela 9: Critérios utilizados para treinar o SOM. | Fonte: autora.

3.2.4. Análise Não Supervisionada: clusterização do mapa auto-organizado

Depois do treinamento do mapa, os neurônios resultantes foram agrupados para representar as fácies finais. Nessa etapa, o agrupamento foi feito através da clusterização hierárquica, no qual a métrica de agrupamento e o número de grupo de clusters foram especificados conforme a Tabela 10:

Métrica de Clusterização	Número de Grupo de Clusters
Menor Soma do Quadrado das Distâncias Internas ao Cluster	8

Tabela 10: Critérios utilizados para clusterizar os dados do perfil de imagem. | Fonte: autora.

Conforme a etapa 3.1.2., o número ideal de agrupamentos também foi definido pelo gráfico de aleatoriedade, como exemplificado na Figura 23. Além disso, por ser uma análise não supervisionada, espera-se que o número de grupo de clusters seja igual ao número de fácies descrito.

3.2.5. Análise Supervisionada: calibração do mapa auto-organizado

Na análise supervisionada foi preciso inserir uma curva de fácies como saída para calibrar o mapa auto-organizado, enquanto os Bins extraídos da etapa 3.2.2. foram utilizados como variáveis de entrada. Dito isso, o número de fácies de saída foi igual ao número de fácies descritas por especialistas.

3.2.6. Análise Supervisionada: CastBoost

Assim como na etapa 3.1.4., foi empregado o algoritmo *CatBoost* para avaliar a eficácia dos modelos de aprendizado de máquina na predição das fácies observadas no poço. Os parâmetros utilizados pelo algoritmo são os mesmos apresentados na Tabela 5 e o *CatBoost* recebeu os Bins extraídos da análise textural de fácies como entrada.

3.3. Análise Integrada dos Perfis Convencionais e de Imagem: CatBoost

Nessa etapa foi realizada uma análise supervisionada para classificar as fácies sedimentares, utilizando os dados dos perfis convencionais e as informações extraídas da análise textural de fácies. Essa análise foi feita através do *CatBoost* e os parâmetros utilizados pelo algoritmo são os mesmos apresentados na Tabela 5.

3.4. Matriz de Confusão

Ao final de todas as etapas foram elaboradas matrizes de confusão, através do Python, para avaliar os modelos de classificação feitos pelo SOM e pelo *CatBoost*. Em problemas de classificação, é possível que apenas a precisão global dos dados previstos pelo modelo não seja suficiente para sua correta avaliação. Isso se dá pelo fato de que frequentemente as classes (ou no presente caso, as fácies) não são balanceadas, ou seja, não existe a mesma quantidade de classes na base de treinamento. Sendo assim, um modelo pode apresentar alta precisão global apenas por acertar a classe mais frequente, errando todas as demais.

Para melhor avaliar os resultados de um modelo de classificação, a matriz de confusão (Figura 29) é uma ferramenta mais adequada. Ela é uma matriz quadrada com número de linhas e colunas iguais ao número de classes, onde as colunas representam os valores reais e as linhas os valores preditos (ou vice-versa). Um bom modelo é capaz de ter alta precisão em todas as classes, gerando uma matriz de confusão com altos valores em sua diagonal. Caso o modelo não seja capaz de prever uma determinada classe, isso será apresentado na forma de falsos positivos ou falsos negativos.

		Valor Real	
		Positivo	Negativo
Valor Predito	Positivo	P	FP
	Negativo	FN	N

Figura 29: Exemplo de matriz de confusão. A matriz tem uma forma quadrada onde a coluna representa os valores reais e a linha representa o valor previsto do modelo. P – Positivo: o valor real foi positivo e o modelo previu um valor positivo. FP – Falso positivo: o valor real foi negativo e o modelo previu um valor positivo. FN – Falso negativo: o valor real foi positivo e o modelo previu um valor negativo. N – Negativo: O valor real é negativo e o modelo previu um valor negativo. | Fonte: modificado de Karimi (2021).

4. RESULTADOS E INTERPRETAÇÃO

4.1. Perfis Convencionais

4.1.1. Análise Não Supervisionada: clusterização do mapa auto-organizado

Através da análise do gráfico de aleatoriedade, foi definido o número 8 como número ideal de clusters (Figura 30), coincidente com o número de fácies observado em dados de testemunho. Observando o gráfico, fica evidente que ocorre uma queda significativa a partir do oitavo cluster, o que justificou a escolha desse número para o agrupamento. O resultado da clusterização hierárquica gerou um mapa auto-organizado que agrupou os neurônios treinados em oito clusters distintos (que posteriormente foram considerados fácies), organizando-os em conjuntos de similaridade (Figura 31). Cada hexágono no mapa representa um neurônio treinado e, quando cores semelhantes estão próximas umas das outras, significa que os neurônios representam características semelhantes das variáveis de entrada.

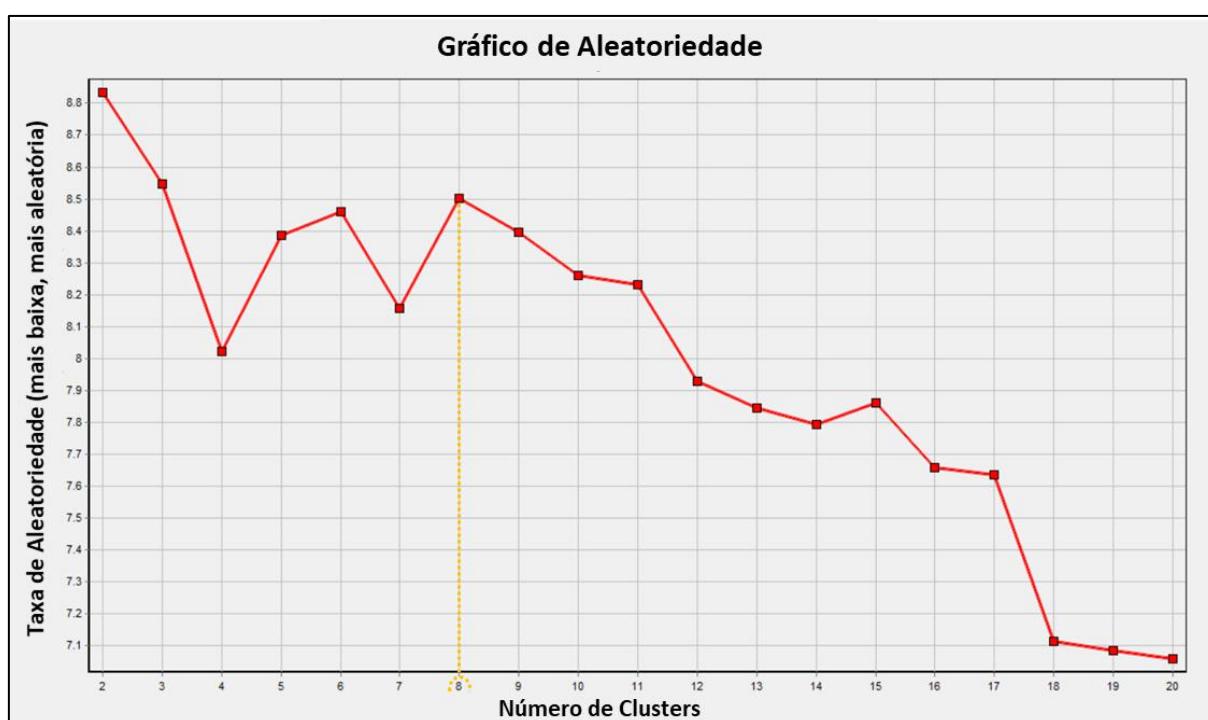


Figura 30: Gráfico de aleatoriedade adquirido através do método não supervisionado. Percebe-se uma queda significativa a partir do cluster 8. | Fonte: autora, a partir do software IP (2023).

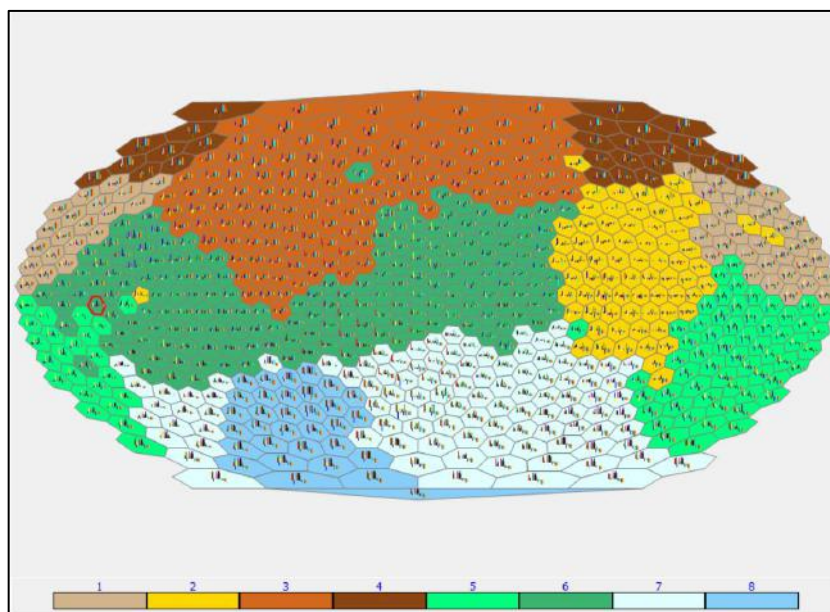


Figura 31: Mapa auto-organizado gerado pela análise não supervisionada dos perfis convencionais. Cluster 1 – bege; Cluster 2 – amarelo; Cluster 3 – marrom claro; Cluster 4 – marrom escuro; Cluster 5 – verde claro; Cluster 6 – verde escuro; Cluster 7 – azul claro; Cluster 8 – azul escuro. | Fonte: autora, a partir do software IP (2023).

A Figura 32 apresenta o resultado da clusterização hierárquica em visualização de perfil. À esquerda está representado o perfil descrito por especialistas, enquanto à direita encontra-se o perfil de fácies gerado pelo método não supervisionado do mapa auto-organizado (perfil predito). As cores foram ajustadas às cores das fácies descritas para fins de comparação.

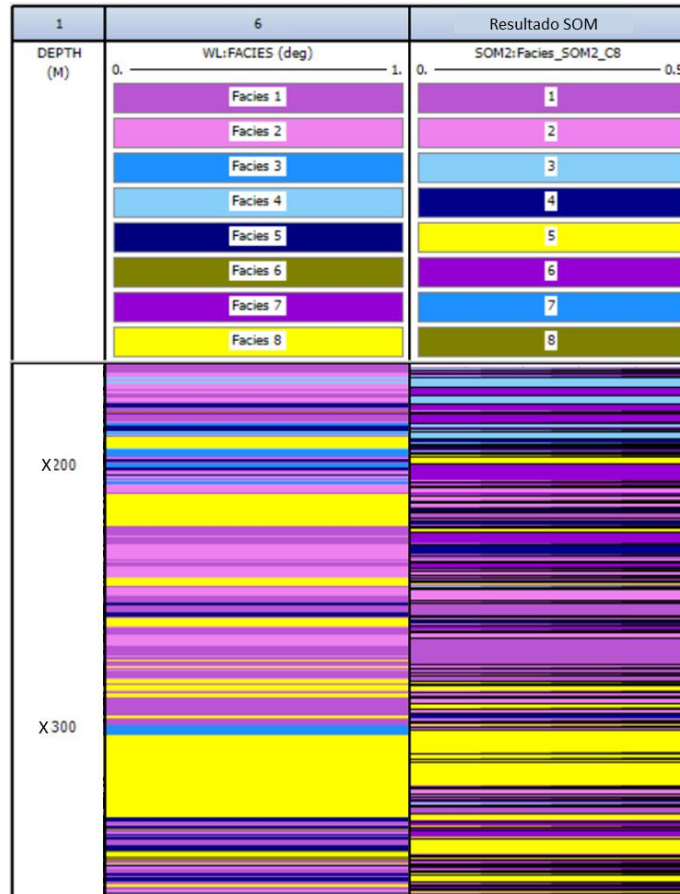


Figura 32: Coluna 6 – perfil de fácies descritas por especialistas; Coluna Resultado SOM – perfil de fácies realizado pelo SOM através do método não supervisionado. | Fonte: autora, a partir do software IP (2023).

4.1.2. Análise Supervisionada: calibração do mapa auto-organizado

A análise supervisionada também resultou um mapa auto-organizado (Figura 33) e um perfil de fácies preditas (Figura 34). Como foi utilizada a curva de fácies como variável de saída, o número de fácies identificado pelo algoritmo foi semelhante ao número descrito por especialistas. Comparando os perfis das Figuras 32 e 34, percebe-se que a análise supervisionada gerou classificações mais satisfatórias que a análise não supervisionada.

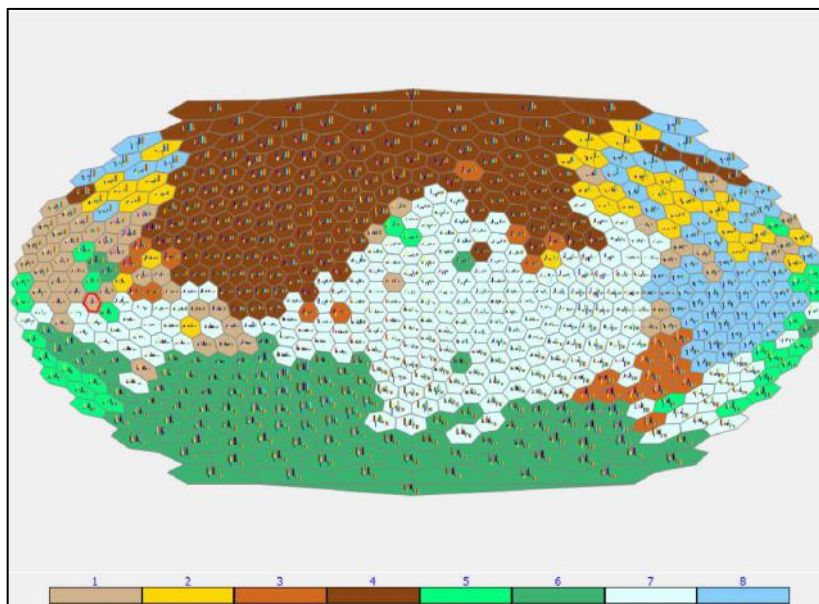


Figura 33: Mapa auto-organizado gerado pela análise supervisionada dos perfis convencionais. Fácies 1 – bege; Fácies 2 – amarelo; Fácies 3 – marrom claro; Fácies 4 – marrom escuro; Fácies 5 – verde claro; Fácies 6 – verde escuro; Fácies 7 – azul claro; Fácies 8 – azul escuro. | Fonte: autora, a partir do software IP (2023).

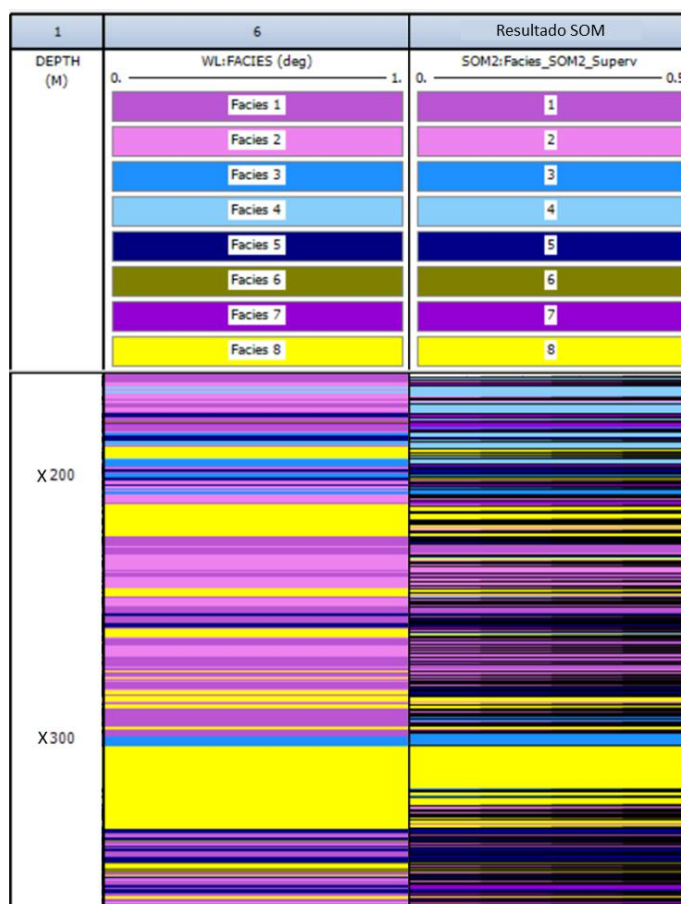


Figura 34: Coluna 6 – perfil de fácies descritas por especialistas; Coluna Resultado SOM – perfil de fácies realizado pelo SOM através do método supervisionado. | Fonte: autora, a partir do software IP (2023).

É interessante observar que o mapa auto-organizado gerado pela análise supervisionada, submetido à uma supervisão forçada para obter uma representação ideal das fácies descritas,

apresenta cores mais dispersas do que o mapa da análise não supervisionada. Isso é reflexo da dificuldade do algoritmo SOM de calibrar neurônios que ao mesmo tempo representem as variáveis de entrada (perfis convencionais) e honrem as fácies descritas. Em uma situação ideal, o mapa da análise supervisionada ficaria parecido com o da análise não supervisionada e os clusters de cores representariam as fácies de maneira satisfatória.

Com o intuito de comparar os resultados obtidos nesta etapa com a análise supervisionada realizada pelo CatBoost, foi gerada uma matriz de confusão (Figura 35). Nessa matriz observa-se uma precisão global de 0.5342 e fica evidente que as fácies 1 e 8 foram classificadas com maior precisão pelo algoritmo SOM. Nota-se que quanto mais escuro o tom de azul, melhor é a predição.

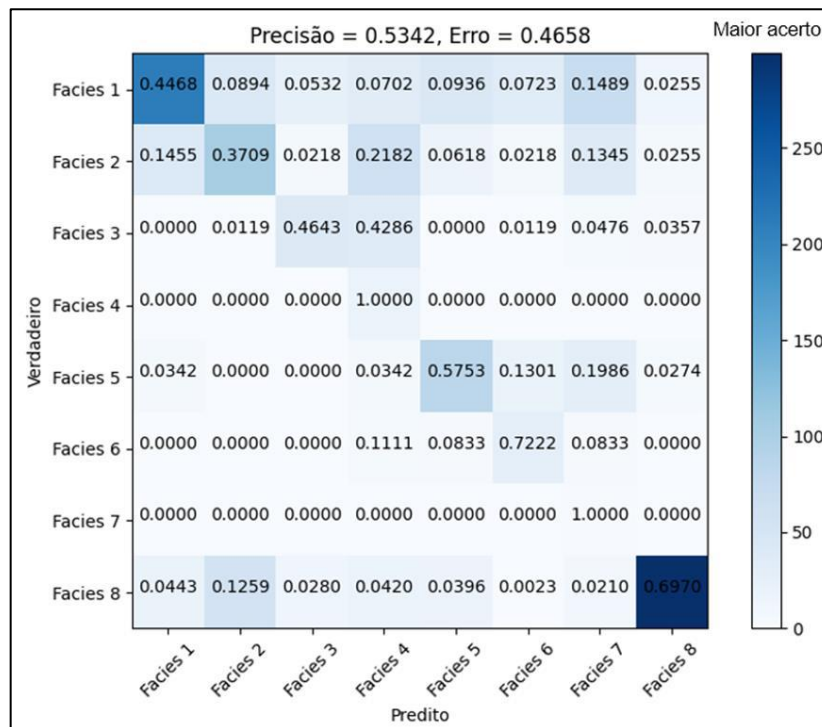


Figura 35: Matriz de confusão dos resultados obtidos pela análise supervisionada (SOM) através dos perfis convencionais. | Fonte: autora.

4.1.3. Análise Supervisionada: CatBoost

A utilização do CatBoost teve como objetivo assegurar que os resultados encontrados estivessem diretamente ligados aos dados de entrada e não com a capacidade do algoritmo de aprendizado de máquina. Dessa forma, os resultados do CatBoost serão apresentados apenas na forma de matriz de confusão, com o objetivo de compará-los com os resultados obtidos na análise supervisionada realizada pelo algoritmo SOM. A matriz representada na Figura 36

apresenta uma precisão global de 0.6278 e um melhor desempenho do CatBoost em classificar as fácies 1, 5 e 8 quando comparado ao resultado do SOM supervisionado.

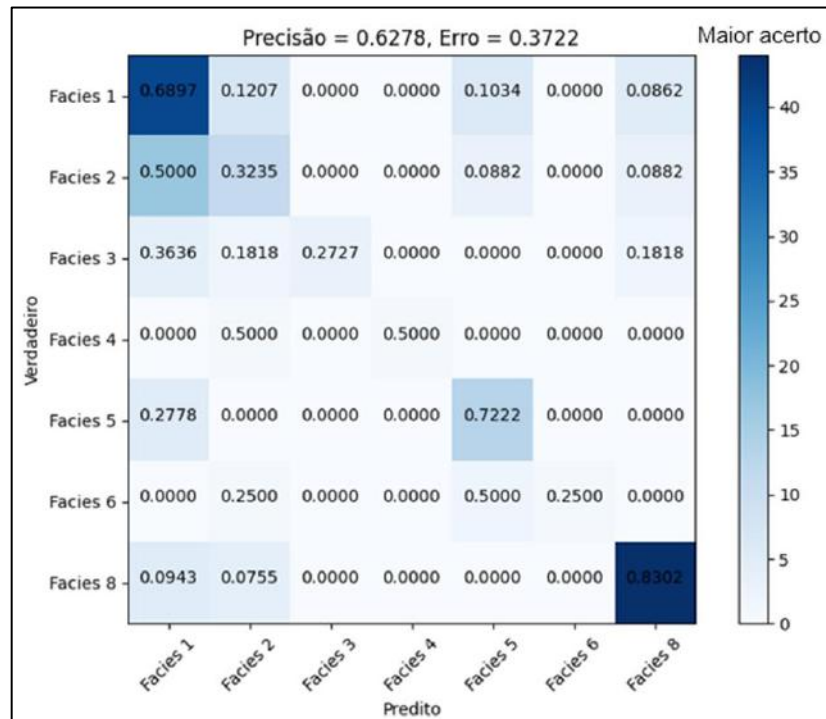


Figura 36: Matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) através dos perfis convencionais. Nesse caso, o modelo não predisse a fácies 7 pois apresentou um peso muito baixo para ser treinada. | Fonte: autora.

4.2. Perfis de Imagem

4.2.1. Análise Não Supervisionada: clusterização do mapa auto-organizado

A clusterização hierárquica dos perfis de imagem acústica e resistiva gerou um mapa auto-organizado com oito clusters distintos para cada perfil (Figuras 37 e 39). As fácies preditas podem ser observadas em perfil nas Figuras 38 e 40. Nota-se que a classificação de fácies não foi tão efetiva quanto a classificação dos perfis convencionais observadas no capítulo 4.1.1. Tanto a Figura 38 quanto a Figura 40 apresentam o perfil de análise estatística de bins utilizado como entrada para gerar o mapa auto-organizado.

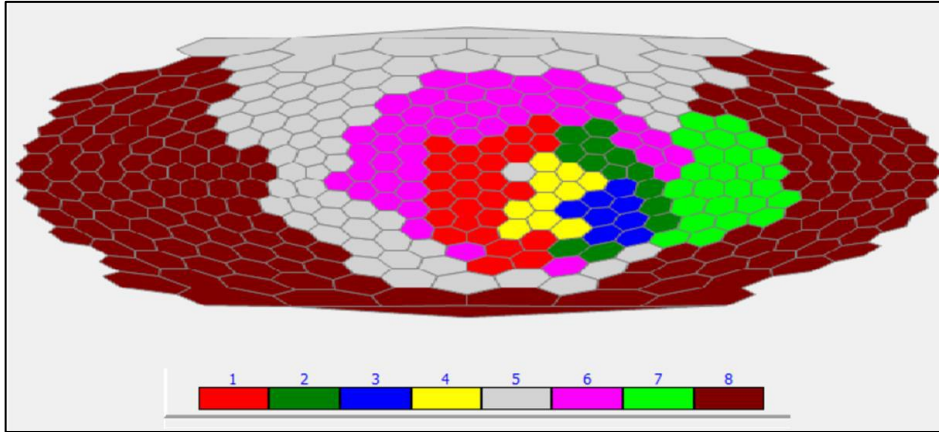


Figura 37: Mapa auto-organizado gerado pela análise não supervisionada do perfil de imagem acústica. Cluster 1 – vermelho; Cluster 2 – verde escuro; Cluster 3 – azul; Cluster 4 – amarelo; Cluster 5 – cinza; Cluster 6 – rosa; Cluster 7 – verde claro; Cluster 8 – marrom. | Fonte: autora, a partir do software IP (2023).

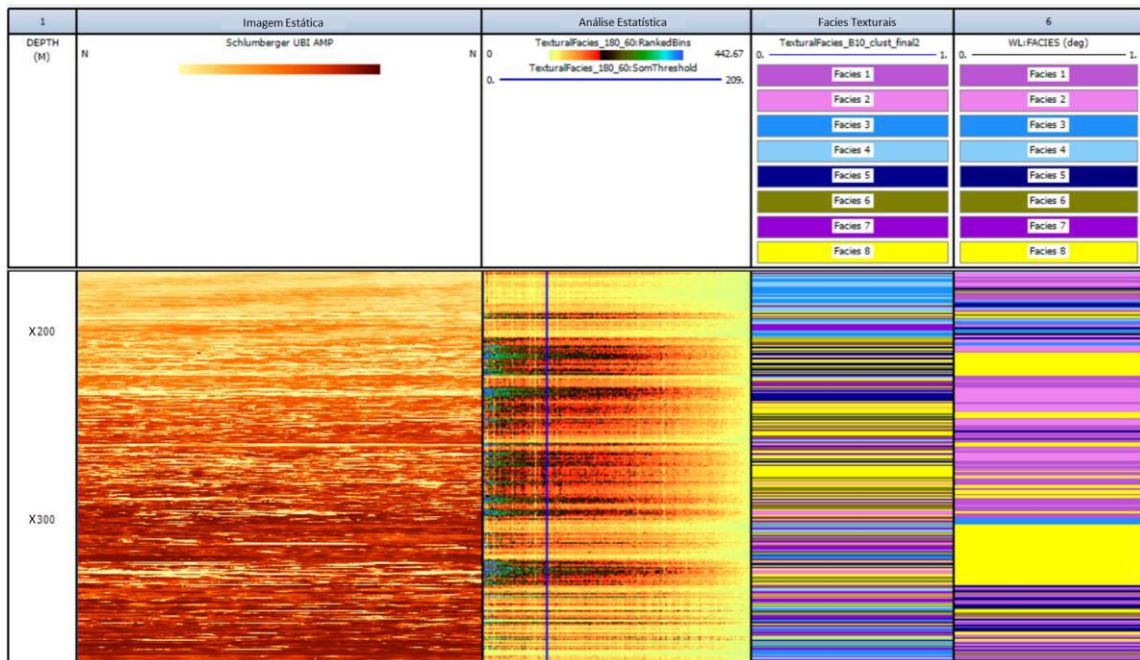


Figura 38: Coluna Imagem Estática – perfil de imagem acústica; Coluna Análise Estatística – perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método não supervisionado; Coluna 6 – perfil de fácies descritas por especialistas. | Fonte: autora, a partir do software IP (2023).

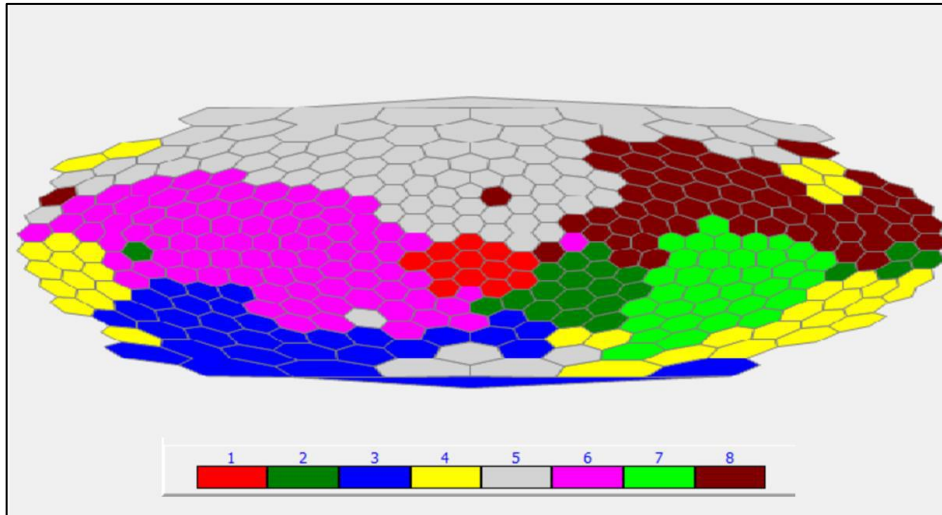


Figura 39: Mapa auto-organizado gerado pela análise não supervisionada do perfil de imagem resistiva. Cluster 1 – vermelho; Cluster 2 – verde escuro; Cluster 3 – azul; Cluster 4 – amarelo; Cluster 5 – cinza; Cluster 6 – rosa; Cluster 7 – verde claro; Cluster 8 – marrom. | Fonte: autora, a partir do software IP (2023).

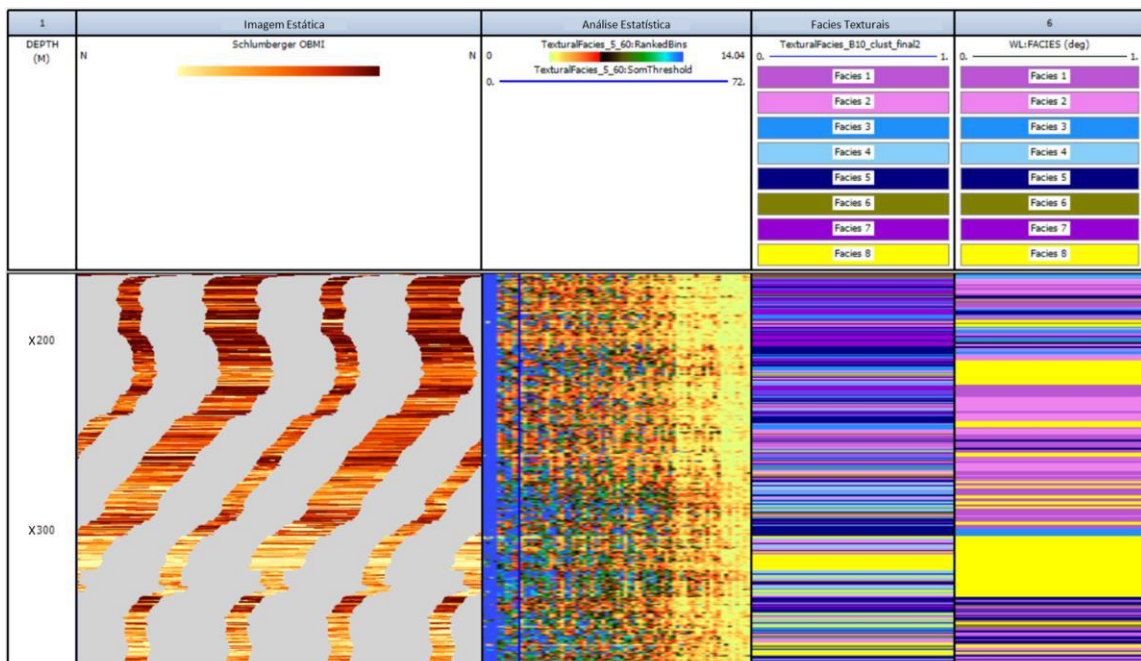


Figura 40: Coluna Imagem Estática – perfil de imagem resistiva; Coluna Análise Estatística – perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método não supervisionado; Coluna 6 – perfil de fácies descritas por especialistas. | Fonte: autora, a partir do software IP (2023).

4.2.2. Análise Supervisionada: calibração do mapa auto-organizado

A análise supervisionada também resultou um mapa auto-organizado (Figuras 41 e 43) e um perfil de fácies (Figuras 42 e 44) para cada perfil de imagem. Assim como na análise supervisionada dos perfis convencionais, o número de fácies identificado pelo algoritmo foi semelhante ao número descrito por especialistas, já que foi utilizada uma curva de fácies como saída para calibrar o mapa auto-organizado. No caso do resultado supervisionado do perfil de

imagem acústica (Figuras 41 e 42), o algoritmo não foi capaz de realizar uma boa calibração, dividindo as 8 fácies em diversos grupos.

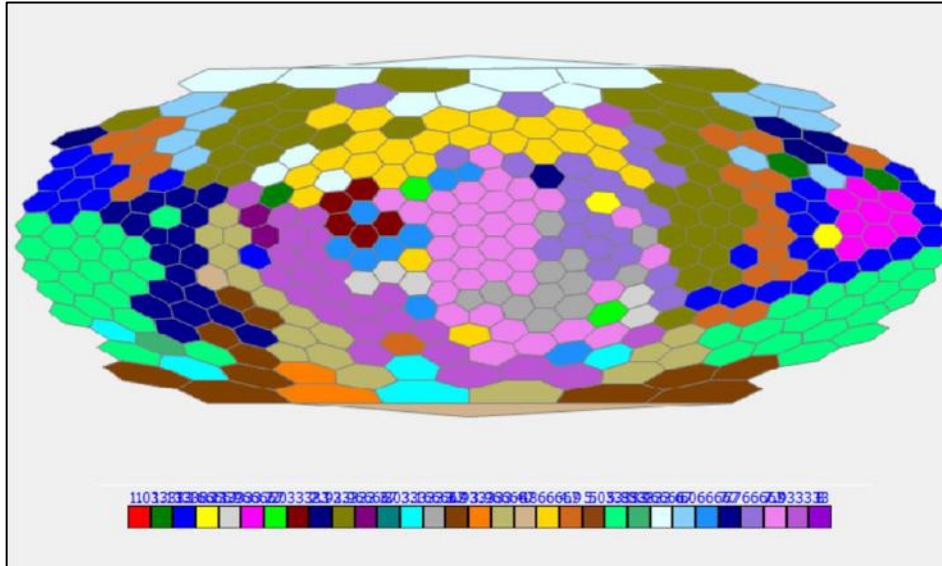


Figura 41: Mapa auto-organizado gerado pela análise supervisionada do perfil de imagem acústica. Nesse caso, o algoritmo distribuiu as 8 fácies em diversos grupos. | Fonte: autora, a partir do software IP (2023).

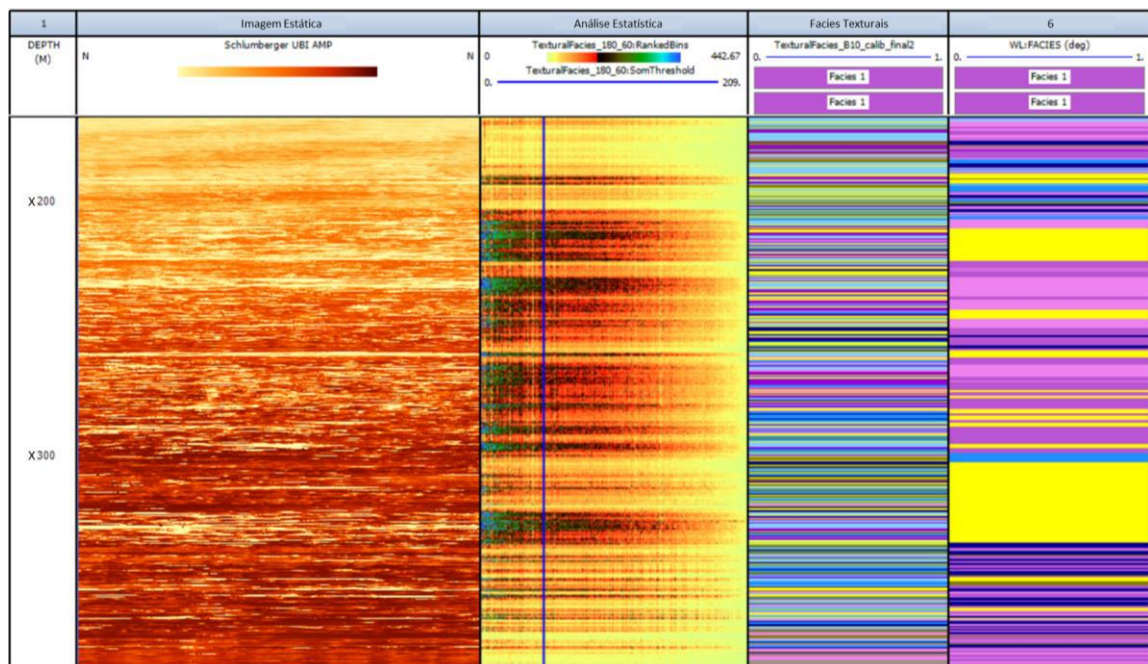


Figura 42: Coluna Imagem Estática – perfil de imagem acústica; Coluna Análise Estatística – perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método supervisionado; Coluna 6 – perfil de fácies descritas por especialistas. | Fonte: autora, a partir do software IP (2023).

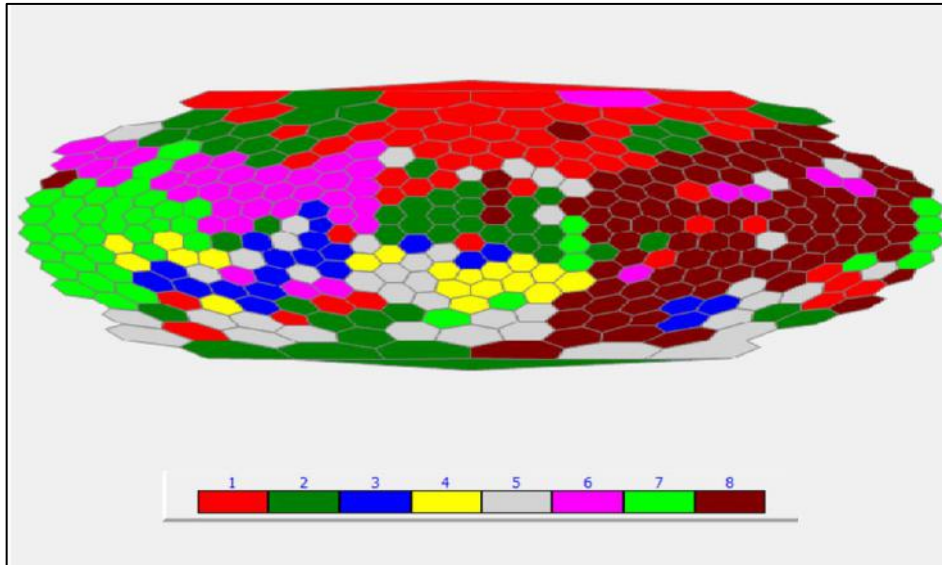


Figura 43: Mapa auto-organizado gerado pela análise supervisionada do perfil de imagem resistiva. Fácies 1 – vermelho; Fácies 2 – verde escuro; Fácies 3 – azul; Fácies 4 – amarelo; Fácies 5 – cinza; Fácies 6 – rosa; Fácies 7 – verde claro; Fácies 8 – marrom. | Fonte: autora, a partir do software IP (2023).

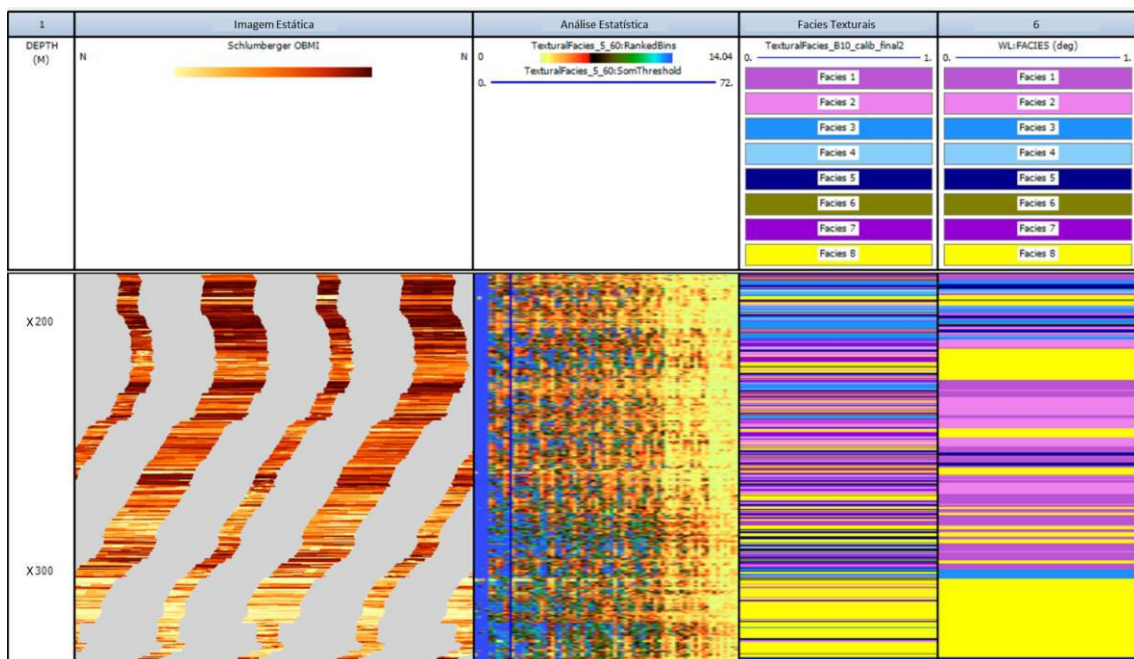


Figura 44: Coluna Imagem Estática – perfil de imagem resistiva; Coluna Análise Estatística – perfil da análise estatística de bins; Coluna Facies Texturais – perfil de fácies realizado pelo SOM através do método supervisionado; Coluna 6 – perfil de fácies descritas por especialistas. | Fonte: autora, a partir do software IP (2023).

Os grupos de fácies do mapa auto-organizado gerado pela análise supervisionada dos perfis de imagem acústica e resistiva ficam mais dispersos quando comparados ao mapa gerado pela análise não supervisionada destes perfis de imagem. Essa diferença também pode ser analisada com os perfis convencionais, que apresentaram melhor agrupamento de dados nas

análises não supervisionadas quando comparadas as análises supervisionadas. Isso é um reflexo da dificuldade do mapa auto-organizado supervisionado de classificar as fácies. Além disso, os perfis convencionais apresentaram melhores resultados de classificação quando comparado aos perfis de imagem.

Da mesma forma que a análise supervisionada dos perfis convencionais, foram geradas matrizes de confusão para comparar os resultados supervisionados gerados pelo SOM e pelo CatBoost. Ao analisar a matriz do perfil de imagem acústica (Figura 45) com o perfil de imagem resistiva (Figura 46), ambas geradas pelo resultado do SOM, é possível dizer que a primeira matriz indica que o modelo não previu os resultados tão bem quanto aqueles gerados pelo perfil de imagem resistiva. Além disso, a matriz referente ao perfil de imagem acústica apresenta uma precisão global igual a 0.1747, enquanto a matriz referente ao perfil de imagem resistiva indica uma precisão global de 0.3570.

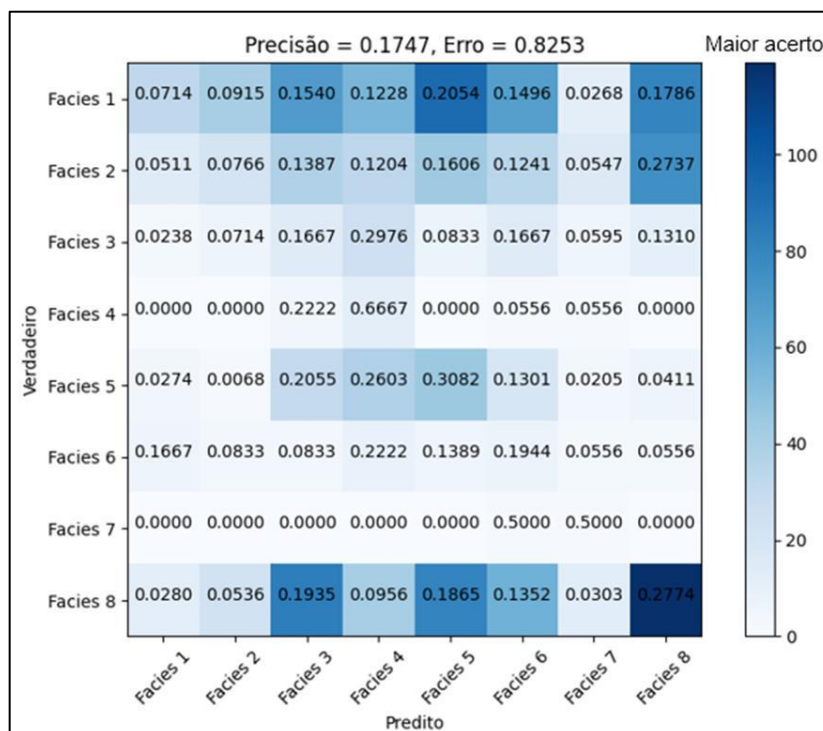


Figura 45: Matriz de confusão dos resultados obtidos pela análise supervisionada (SOM) através do perfil de imagem acústica. | Fonte: autora.

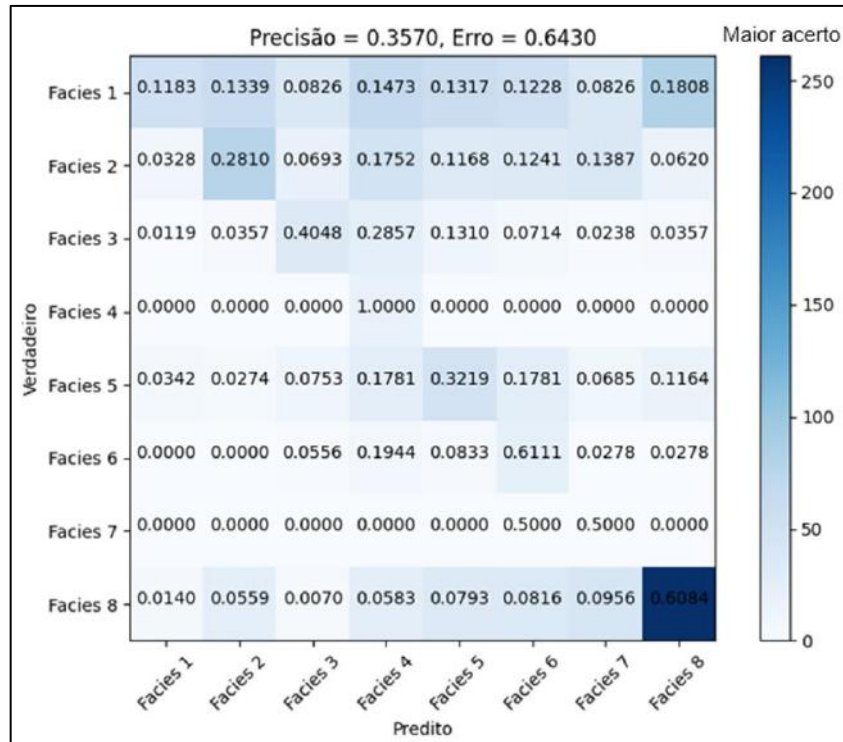


Figura 46: Matriz de confusão dos resultados obtidos pela análise supervisionada (SOM) através do perfil de imagem resistiva. | Fonte: autora.

4.2.3. Análise Supervisionada: CatBoost

Assim como no capítulo 4.1.3, o resultado da análise supervisionada realizada pelo CatBoost será apresentado na forma de matriz de confusão, tanto para os perfis de imagem acústica quanto para os de imagem resistiva. No caso das matrizes referentes ao perfil de imagem acústica, a matriz associada ao CatBoost (Figura 47) com precisão global igual a 0.3333 realizou melhores predições para as fácies 1, 2 e 8 quando comparado ao modelo associado ao SOM. Já a matriz referente ao perfil de imagem resistiva (Figura 48), com precisão global igual a 0.4722, demonstra que obteve melhores predições para as fácies 1 e 8 quando comparada ao modelo associado ao SOM.

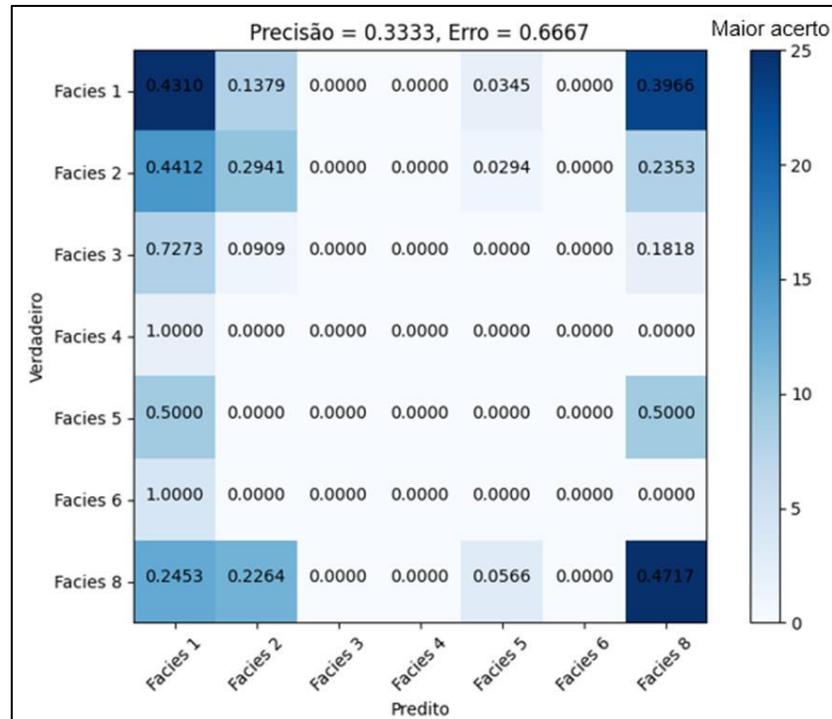


Figura 47: Matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) através do perfil de imagem acústica. O modelo não predisse a fácies 7 pois apresentou um peso muito baixo para ser treinada. | Fonte: autora.

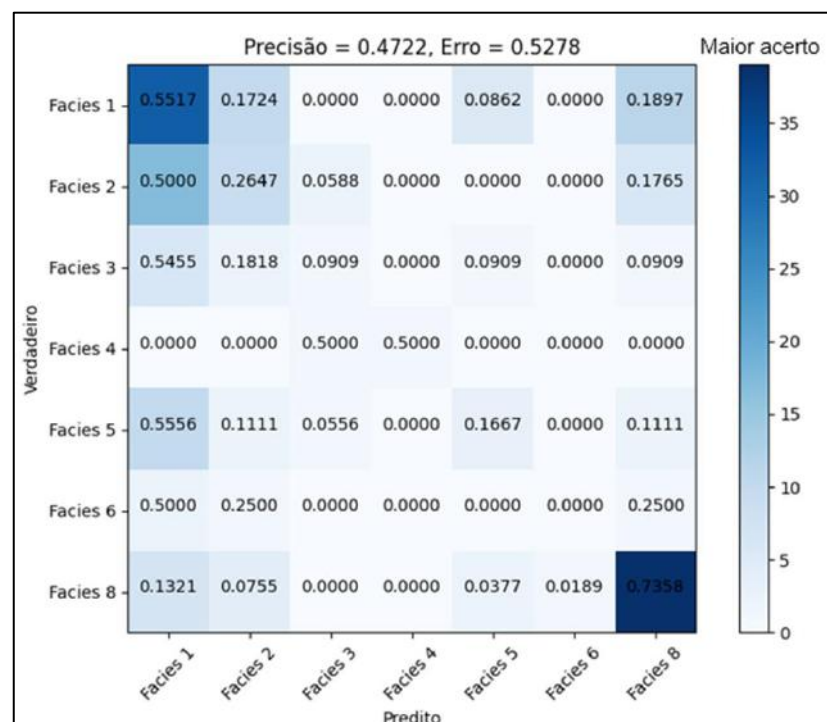


Figura 48: Matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) através do perfil de imagem resistiva. O modelo não predisse a fácies 7 pois apresentou um peso muito baixo para ser treinada. | Fonte: autora.

4.3. Análise Integrada dos Perfis Convencionais e de Imagem: CatBoost

Três matrizes de confusão foram geradas para a análise integrada dos perfis convencionais e de imagem. A primeira matriz (Figura 49) é referente análise conjunta dos dados dos perfis convencionais e de imagem acústica através do Catboost, apresentando boa predição para as fácies 1, 2, 4, 5 e 8 e com precisão global igual a 0.6278. A segunda matriz (Figura 50) está associada a análise dos dados dos perfis convencionais e de imagem resistiva, apresentando boas predições para as fácies 1, 2, 5 e 8 e precisão global igual a 0.5667. Já a terceira matriz (Figura 51), referente a análise dos dados dos perfis convencionais e de imagem (acústica e resistiva), também apresenta bons resultados para as fácies 1, 2, 5 e 8 e precisão global igual a 0.6444. Comparando os três resultados, a terceira matriz foi a que apresentou as melhores predições. É interessante observar que nenhum dos algoritmos treinados de forma integrada apresentaram resultados muito superiores ao treinado apenas com perfis convencionais.

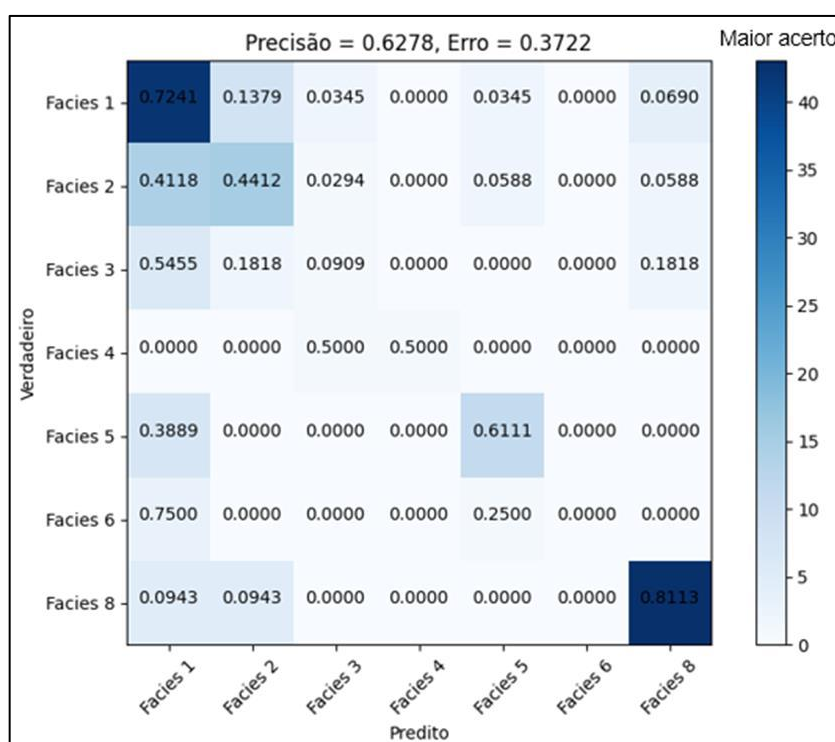


Figura 49: Matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) dos perfis convencionais e perfil de imagem acústica. | Fonte: autora.

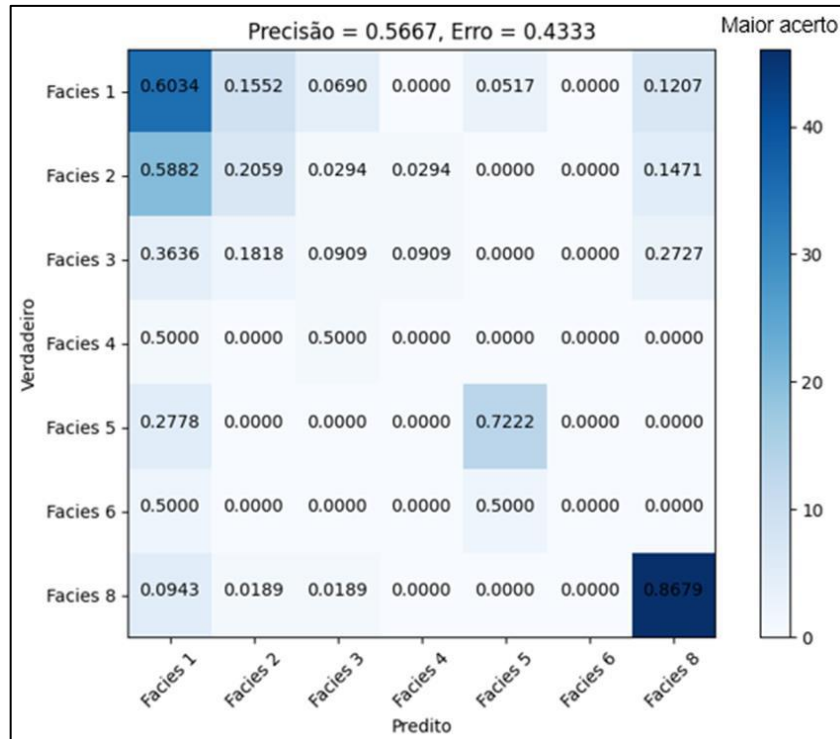


Figura 50: Matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) dos perfis convencionais e perfil de imagem resistiva. | Fonte: autora.

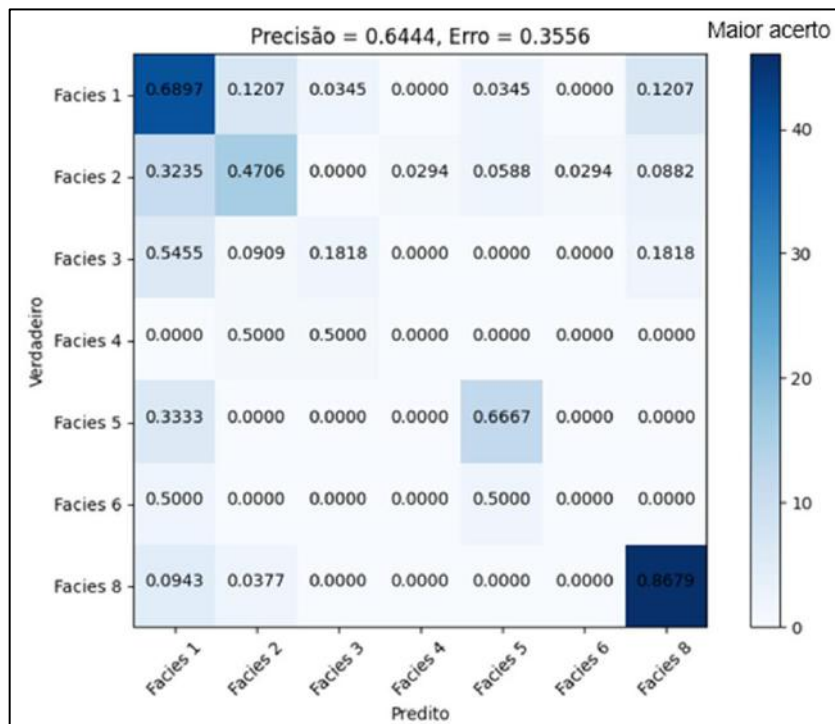


Figura 51: Matriz de confusão dos resultados obtidos pela análise supervisionada (CatBoost) dos perfis convencionais e perfis de imagem. | Fonte: autora.

4.4. Análise Comparativa

A Tabela 11 consolida e compara os resultados das análises supervisionadas apresentados anteriormente pelas matrizes de confusão. Com este comparativo, fica demonstrado que os perfis convencionais sozinhos utilizando o SOM obtiveram as melhores classificações de fâcies quando comparados às análises integradas e, principalmente, às análises utilizando somente os perfis de imagem. Incluir os dados da análise textural de fâcies junto aos perfis convencionais não melhorou significativamente o poder de predição do algoritmo.

Perfis	Análise Supervisionada	Fácie 1	Fácie 2	Fácie 3	Fácie 4	Fácie 5	Fácie 6	Fácie 7	Fácie 8	Precisão
Convencionais	SOM	0.4468	0.3709	0.4643	1	0.5753	0.7222	1	0.6970	0.5342
	CatBoost	0.6897	0.3235	0.2727	0.5	0.7222	0.2500	-	0.8302	0.6278
Imagem Acústica	SOM	0.0714	0.0766	0.1667	0.6667	0.3082	0.1944	0.5	0.2774	0.1747
	CatBoost	0.4310	0.2941	0	0	0	0	-	0.4717	0.3333
Imagem Resistiva	SOM	0.1183	0.2810	0.4048	1	0.3219	0.6111	0.5	0.6084	0.3570
	CatBoost	0.5517	0.2647	0.0909	0.5	0.1667	0	-	0.7358	0.4722
Convencionais + Imagem Acústica	CatBoost	0.7241	0.4412	0.0909	0.5	0.6111	0	-	0.8113	0.6278
Convencionais + Imagem Resistiva	CatBoost	0.6034	0.2059	0.0909	0	0.7222	0	-	0.8679	0.5667
Convencionais + Imagem Acústica e Resistiva	CatBoost	0.6897	0.4706	0.1818	0	0.6667	0	-	0.8679	0.6444

Tabela 11: Precisão da matriz de confusão referente a cada análise supervisionada, com destaque em cinza para os melhores resultados de classificação entre todas as análises. | Fonte: autora.

Apesar dessa análise integrada apresentar boa precisão em comparação com as análises utilizando apenas os perfis convencionais e os perfis de imagem, poucas fâcies obtiveram melhores predições com o CatBoost. Além disso, os resultados das análises dos perfis convencionais sozinhos utilizando o CatBoost também foram melhores que àquelas utilizando somente os perfis de imagem. No entanto, ao compararmos os resultados entre os perfis convencionais, é perceptível que o SOM teve um desempenho melhor que o CatBoost.

5. DISCUSSÃO

A partir dos resultados, é possível observar uma diferença entre as análises não supervisionadas e supervisionadas utilizando o SOM. Percebe-se, a partir dos perfis de fácies preditos (Figuras 32, 34, 38, 40, 42, 44), que os resultados das análises não supervisionadas não foram satisfatórios e os mapas auto-organizados gerados pela análise supervisionada (Figuras 33, 41 e 43) apresentaram muita dispersão quando comparado aos mapas auto-organizados não supervisionados (Figuras 31, 37 e 39). Essa circunstância sugere que a forma em que o SOM foi treinado não aparenta ser muito adequada para classificação de fácies no pré-sal da Bacia de Santos. Além disso, os resultados das matrizes de confusão do CatBoost não ficaram muito diferentes dos do SOM. Isso mostra que (1) os diferentes tipos de algoritmos não impactaram na qualidade da classificação e (2) que as variáveis de entrada utilizadas não foram suficientes para a adequada classificação das fácies do pré-sal, reflexo da heterogeneidade desses carbonatos.

Partindo para as diferenças entre os modelos utilizando os perfis convencionais, de imagem e ambos, observando a Tabela 11, nota-se que os modelos utilizando apenas os perfis de imagem apresentaram resultados piores que aqueles usando apenas os perfis convencionais. Enquanto isso, os modelos utilizando ambos os perfis não apresentaram resultados consideravelmente melhores que apenas os dos perfis convencionais. Novamente, os resultados das matrizes de confusão do CatBoost não ficaram muito diferentes dos do SOM. Isso mostra que (1) os diferentes tipos de algoritmos não impactaram na qualidade da classificação e (2) que os atributos gerados pela análise textural de fácies dos perfis de imagem não trouxeram informações relevantes para o processo de classificação e possivelmente adicionaram ruído aos modelos.

Por fim, apesar de apresentarem as melhores predições, os modelos usando apenas os perfis convencionais através do SOM não apresentaram resultados muito satisfatórios. Isso evidencia a real necessidade de incorporar variáveis de entrada que contenham informações relevantes quanto a textura e estrutura da rocha. Embora os perfis de imagem possam desempenhar um papel relevante neste processo, os métodos empregados no trabalho não foram suficientes para esse tipo de classificação. Além disso, conforme o histograma apresentado na Figura 19, as fácies não foram balanceadas para compor o mesmo número de amostras, dificultando a classificação realizada pelo algoritmo. Embora todos os algoritmos tenham sido treinados com a base de dados desbalanceada, isso apresentou um desafio para o algoritmo ao tentar classificar as fácies menos frequentes, como a fácies 7.

6. CONCLUSÃO E RECOMENDAÇÕES

Diante dos resultados obtidos na capacidade do algoritmo de aprendizagem de máquina para classificação de fácies no intervalo pré-sal da Bacia de Santos, é evidente que o SOM revelou limitações para resolver esse tipo de problema da forma em que foi treinado. Embora os perfis convencionais sozinhos utilizando esse algoritmo tenham apresentado os melhores resultados, ainda estão longe de serem considerados adequados. Isso evidencia que, por si só, não conseguem capturar uma carga textural para essa classificação que seria importante no contexto do pré-sal.

Surpreendentemente, os perfis de imagem, que naturalmente poderiam ser esperados para fornecer informações texturais, não contribuíram significativamente com as técnicas implementada no trabalho, possivelmente confundindo os algoritmos e trazendo ruído à classificação. Além disso, todas as abordagens que utilizaram somente os perfis de imagem apresentaram piores resultados quando comparado as abordagens usando somente os perfis convencionais. Isso demonstra que, pelo menos para o caso dos carbonatos do pré-sal da Bacia de Santos, a análise textural de fácies implementada no IP, que utiliza o espectro da frequência e a análise estatística, não está sendo capaz de extrair informações relevantes para o processo de classificação de fácies.

Assim, a inclusão de perfis de imagem na classificação de fácies no pré-sal revela-se uma tarefa não trivial e exige abordagens de processamento mais avançadas capazes de extrair informações relevantes desses perfis de imagem. Técnicas de *deep learning* com redes neurais, por exemplo, podem ser mais adequadas para enfrentar esse tipo de desafio. Em última análise, a pesquisa destaca a complexidade e a necessidade de aprimoramento nas técnicas de classificação de fácies no contexto específico do pré-sal, buscando abordagens inovadoras e mais eficazes para superar as limitações identificadas.

7. REFERÊNCIAS

- AL-BADALWI, B.A.H., 2021. **3D Reservoir rock type model based on cluster analysis technique of rumaila formation in the Ahdeb Oil Field, Central Iraq**. Journal of Petroleum Research and Studies, p.49-73.
- ANP, 2023. **Painel Dinâmico da Produção de Petróleo e Gás Natural**. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Acesso em 07 de setembro de 2023. Disponível em: <https://app.powerbi.com/view?r=eyJrIjoiNzVmNzI1MzQtNTY1NC00ZGVhLTk5N2ItNzBkMDNhY2IxZTIxIiwidCI6IjQ0OTlmNGZmLTI0YTtytNGI0Mi1iN2VmLTEyNGFmY2FkYzIxMyJ9>
- ANP, 2022. **Bacia de Santos: sumário geológico e setores em oferta**. Superintendência de Avaliação Geológica e Econômica (SAG).
- ANP, 2019. **Sumário Atapu e Oeste de Atapu (AIP)**.
- ANP, 2017. **Bacia de Santos: sumário geológico e setores em oferta. Superintendência de Definição de Blocos (SDB)**.
- ASAN, U.; Ercan, S., 2012. **An introduction to Self-Organizing Maps**. Atlantis Press Book, cap.14.
- BISHOP, C.M., 2006. **Pattern recognition and machine learning**. [S.I.]: Springer p.738.
- CAO, W.; Wang, K.; Gan, H.; Yang, M., 2021. **User online purchase behavior prediction based on fusion model of CatBoost and Logit**. Journal of Physics, conference series 2003(2021)012011.
- CARR, H.Y.; Purcell, E.M., 1954. **Effects of diffusion on free precession in nuclear magnetic resonance experiments**. Physical Review, v.94, n.3, p.630–638.
- CASTELLANI, B.; Castellani, J., 2003. **Data Mining: Qualitative Analysis with Health Informatics Data**. Qualitative Health Research, v.13, n.07, p.1005-1018.
- CASTRO, T.M., 2019. **Avaliação dos reservatórios carbonáticos do pré-sal no Campo de Búzios, Bacia de Santos**. Dissertação de Mestrado, Programa de Pós-Graduação em Dinâmica dos Oceanos e da Terra, Universidade Federal Fluminense, 182p.
- CHANG, D.; Vinegar, H.; Morriss, C; Straley, C., 1994. **Effective porosity, producible fluid and permeability in carbonates from NMR logging**. SPWLA 35th Annual Logging Symposium. Tulsa, Oklahoma, USA: Society of Petrophysicists and WellLog Analysts.
- COATES, G.R.; Marschall, D.; Mardon, D.; Galford, J., 1997. **A new characterization of bulk-volume irreducible using magnetic resonance**. SPWLA 38th Annual Logging Symposium. Houston, Texas, USA: Society of Petrophysicists and Well-Log Analysts.
- COATES, G.R.; Xiao, L.; Prammer, M.G., 1999. **NMR logging - principles and applications**. Houston: Halliburton Energy Services.

ELLIS, D.V.; SINGER, J. M., 2007. **Well logging for earth scientists**. Second. [S.l.]: Springer, 692 p.

ERSHOV, V., 2018. **CatBoost Enables Fast Gradient Boosting on Decision Trees Using GPUs**. Disponível em: <https://catboost.ai/news/catboost-enables-fast-gradient-boosting-on-decision-trees-using-gpus>

FÉLIX, F.; Nascimento, E.; Cláudio, B., 2013. **Perfis de imagem de poços de petróleo**. Cadernos de Graduação – Ciências Exatas e Tecnológicas, v.01, n.16, p.61-78.

FRIEDMAN, J.; Hastie, T.; Tibshirani, R., 2001. **The elements of statistical learning**, v.01. Springer series in statistics New York.

GUARIDO, M., 2018. **Machine learning in geoscience: facies classification with features engineering, clustering, and gradient boosting trees**. Crewes Research Report 30: p.13.1–13.23.

HERTZOG, R.; Colson, L.; Seeman, B.; O'brien, M.; Scott, H.; Mckeon, D.; Wraight, P.; Grau, J.; Ellis, D.; Schweitzer, J.; Herron, M., 1989. **Geochemical logging with spectrometry tools**. SPE Formation Evaluation, v.04, p.153–162.

HURLEY, Neil.; Asquith, G.; Rygowski, D., 2004. **Basic Well LogAnalysis**. Second Edition. Tulsa, Oklahoma: The American Association of Petroleum Geologists, cap.09, p.151–163.

KARIMI, Z., 2021. **Confusion Matrix**. Universidade de Kharazmi.

KOHONEN, T., 1997. **Self-Organizing Maps**, Springer Verlag (Berlim), 2nd edition.

LEITE, M., Carrasquilla, A.; Silva, J.; Lobo, L; Curtinhas, R.; Batista, S., 2007. **Aplicação da perfilagem geofísica em projetos de perfuração de poços de petróleo**. 10th International Congress of the Brazilian Geophysical Society.

LIMA, L.M.G., 2018. **Análise integrada de sísmica e poços da porção noroeste do Bloco de Libra, Bacia de Santos**. Dissertação de Mestrado, Programa de Pós-Graduação em Análise de Bacias e Faixas Móveis, Universidade do Estado do Rio de Janeiro, 98p.

LIMA, R.A.P., 2019. **Machine learning applications for geoscience problems**. Tese de Doutorado, Programa de Pós-Graduação em Geociências, Universidade de Oklahoma, 250p.

MAHESH, B., 2018. **Machine learning algorithms – A review**. International Journal of Science and Research, v.09.

MEIBOOM, S.; Gill, D., 1958. **Modified spin-echo method for measuring nuclear relaxation times**. Review of Scientific Instruments, v.29, n.8, p.688–691, 1958.

MIO, E.; Chang, H.K.; Corrêa, F.S., 2005. **Integração de métodos geofísicos na modelagem crustal da Bacia de Santos**. Revista Brasileira de Geofísica, v.23(3).

MOHRIAK, W.U., 2012. **Bacias de Santos, Campos e Espírito Santo**. Geologia do Brasil, Beca Edições, p.481-496.

MOREIRA, J.L.P.; Madeira, C.V.; Gil, J.A.; Machado, M.A.P., 2007. **Bacia de Santos. Boletim de Geociências Petrobras**, v.15, n.02, p.531-549.

NAYAR, S.K., 2021. **Image Filtering in Frequency Domain | Image Processing II**. Youtube, 2021. Disponível em: <https://www.youtube.com/watch?v=OOu5KP3Gvx0&t=9s>

OLIVEIRA, L.A.B., 2022. **Modelagem geoquímica e mineralógica dos reservatórios carbonáticos do pré-sal da Bacia de Santos através de perfis de poços e inteligência artificial**. Dissertação de Mestrado, Universidade de São Paulo, 180p.

OLIVEIRA, L.A.B.; Gonçalves, L.; Santos, B.C.C.; Trevizan, W.A.; Victor, R.A., 2022. **Improving the calculation of the petrophysical properties in vugular carbonates using logs and rock samples: a case study in a Brazilian pre-salt well**. SPWLA 63rd Annual Logging Symposium.

PEREIRA, M.J.; MACEDO, J.M., 1990. **A Bacia de Santos: perspectivas de uma nova província petrolífera na plataforma continental sudeste brasileira**. Boletim Geociências da Petrobras, v.4, p.3-11.

RICCOMINI, C.; Sant'Anna, L.G.; Tassinari, C.C.G., 2012. **Pré-sal: geologia e exploração**. Revista USP, n.95, p.33-42.

SLB., 2023. **Neutron Porosity**. Disponível em: https://glossary.slb.com/en/terms/n/neutron_porosity.

SLB., 2015. **Schlumberger wireline services catalog**, p. 66. Disponível em: https://www.slb.com/~media/Files/evaluation/catalogs/2015_wireline_services_catalog.pdf.

SOUZA, L.S.; Sgarbi, G.N.C., 2019. **Bacia de Santos no Brasil: geologia, exploração e produção de petróleo e gás natural**. Boletim de Geologia, 41(1), p.175-195.

STRALEY, C.; Rossini, D.; Vinegar, H.; Tutunjian, P.; Morriss, C., 1997. **Core analysis by low-field NMR**, Log Analyst, 38(2), pp. 84–93.

WITTEN, I.H.; Frank, E., 2005. **Data mining: practical machine learning tools and techniques**. Second. [S.I.]: Morgan Kaufmann Publishers, p.525.

WYLLIE, M.R.J.; Gregory, A.R.; Gardner, L.W., 1956. **Elastic wave velocities in heterogeneous and porous media**. Geophysics, v.21, n.01, p.41–70.