



UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E
PESQUISA COORDENAÇÃO DE PESQUISA

PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO
CIENTÍFICA – PIBIC

**CARACTERIZAÇÃO DE ELETROFÁCEIS POR MEIO
DE INTELIGÊNCIA ARTIFICIAL COM ABORDAGEM
SUPERVISIONADA**

Relatório Final
Período da bolsa: de 09/23 a 08/24

Este projeto é desenvolvido com bolsa de iniciação científica

PIBIC/COPES

Sumário

1.	Introdução.....	3
2.	Objetivos.....	3
2.1	Objetivo geral.....	3
2.2	Objetivos específicos	3
3.	Metodologia (Incluir fluxograma e explicações).....	4
3.1	Conjunto de dados (falar sobre o AGP).....	4
3.2	Curvas de perfis petrofísicos escolhidas	4
3.3	Tratamento dos dados	5
3.3.1	Conversão da unidade de medida de profundidade	6
3.3.2	Remoção registros que não possuem todas as curvas escolhidas	6
3.3.3	Manutenção de apenas uma curva de resistividade profunda	7
3.3.4	Padronização do nome das curvas de resistividade profunda	7
3.3.5	Aplicação dos mnemônicos das curvas.....	7
3.3.6	Adição da curva DCAL	8
3.3.7	Reclassificação da litologia metamórfica	8
3.4	Algoritmo.....	10
3.5	Algoritmo.....	10
4.	Resultados e discussões.....	10
4.1	Dados desbalanceados e sem unidades geológicas	11
4.2	Y	12
4.3	Z.....	12
4.4	W	12
5.	Conclusões	12
6.	Perspectivas de futuros trabalhos	12
7.	Referências bibliográficas	12
8.	Outras atividades	12
9.	Justificativa	12

1. Introdução

2. Objetivos

2.1 Objetivo geral

O objetivo geral deste plano de trabalho é, por meio dos perfis petrofísicos, identificar as eletrofáceis de poços de petróleo da bacia Sergipe/Alagoas com o método de inteligência artificial conhecido como Floresta Aleatória.

2.2 Objetivos específicos

- Realizar o estado da arte sobre a identificação de eletrofáceis utilizando ferramentas com métodos de inteligência artificial (Florestas Aleatórias) com abordagem supervisionada;
- Levantar dados rotulados de perfis e definir as curvas que serão utilizadas para treinamento e teste da inteligência artificial com suporte de especialistas;
- Escolher o algoritmo de inteligência artificial para o treinamento com os dados levantado dos poços;
- Realizar o teste e treinamento do algoritmo de identificação das eletrofáceis a partir dos dados rotulados;
- Obter uma inteligência artificial treinada em uma base de dados referente a

algum campo de petróleo da bacia Sergipe/Alagoas;

- Propagar as eletrofácies para poços que tenham as mesmas unidades estratigráficas, pertencentes, ou não, ao campo de petróleo utilizando a inteligência artificial.

3. Metodologia (Incluir fluxograma e explicações)

3.1 Conjunto de dados (falar sobre o AGP)

No desenvolvimento deste trabalho, foram utilizados arquivos de dados no formato Digital Log Interchange Standard (DLIS) referentes a poços da bacia Sergipe/Alagoas. Estes arquivos contêm registros de perfis petrofísicos dos poços de petróleo. Eles podem ser acessados de forma gratuita por meio do seguinte hiperlink: [ANP-TERRESTRE \(cprm.gov.br\)](http://cprm.gov.br/ANP-TERRESTRE).

Tabela <num_tabela> - Nomes dos poços

Nome ANP	Nome BR	Nome abreviado
1-BRSA-459-SE	1-NCL-2-SE	P_459
1-BRSA-551-SE	1-FSG-1-SE	P_551
1-BRSA-574-SE	1-FSJQ-1-SE	P_574
1-BRSA-595-SE	1-FSG-2-SE	P_595
1-BRSA-605-SE	1-FSJQ-2-SE	P_605
1-BRSA-643-SE	1-CBO-1-SE	P_643
1-BRSA-645-SE	1-SIB-1-SE	P_645
1-BRSA-659-SE	1-POI-2-SE	P_659
1-BRSA-689-SE	1-POI-1-SE	P_689
1-BRSA-696-SE	1-FSJQ-3-SE	P_696
1-BRSA-698-SE	1-DP-2-SE	P_698

Fonte: Autor

3.2 Curvas de perfis petrofísicos escolhidas

Para o trabalho em questão, foram escolhidas 11 curvas de perfis petrofísicos:

- Profundidade (TDEP): mede a distância medida a partir da superfície

até um determinado ponto do poço.

- *Bit size* (BS): diâmetro da broca utilizada para perfurar o poço.
- Caliper (CAL): diâmetro do poço perfurado.
- Diferença entre CAL e BS (DCAL): diferença entre o tamanho do poço perfurado e o tamanho da broca que perfurou o poço.
- *Gamma ray* (GR): detecta a radioatividade total da formação geológica.
- Resistividade profunda (RESO): mede a resistividade das rochas.
- Sônico (DT): detecta uma relação direta entre o tempo de propagação do som e a porosidade das rochas.
- Densidade (RHOB): detecta os raios gama defletidos pelos elétrons orbitais dos componentes das rochas.
- Correção da densidade (DRHO)
- Neutrão (NPHI): detecta a porosidade das rochas.
- Fator fotoelétrico (PE): mede a absorção fotoelétrica de raios gama de baixa energia.

<Explicar que CAL e BS não são variáveis discriminatórias>

3.3 Tratamento dos dados

Antes de aplicar o algoritmo de Floresta Aleatória, é fundamental que os dados passem por um processo de tratamento para que eles possam ser analisados pelos especialistas e utilizados pelo algoritmo. Dessa forma, o tratamento foi realizado por meio do seguinte procedimento:

- I. Conversão da unidade de medida de profundidade de polegada para metro; (ver tópico 3.3.1)
- II. Remoção de registros que não possuíam todas as curvas escolhidas; (ver tópico 3.3.2)
- III. Manutenção de apenas uma curva de resistividade profunda por poço; (ver tópico 3.3.3)
- IV. Padronização do nome das curvas de resistividade profunda; (ver tópico 3.3.4)

- V. Aplicação dos mnemônicos das curvas para padronizar os nomes; (ver tópico 3.3.5)
- VI. Adição da curva DCAL em poços que ainda não a possuem; (ver tópico 3.3.6)
- VII. Reclassificação da litologia metamorfica em xisto. (ver tópico 3.3.7)

3.3.1 Conversão da unidade de medida de profundidade

Para manter o padrão usual da indústria de petróleo do Brasil, a unidade de medida de profundidade de 7 dos 11 poços passou por um processo de conversão de polegada (pol) para metro (m). A tabela <num_tabela> apresenta a unidade de medida de profundidade de cada poço antes da padronização para metro.

Tabela <num_tabela> - Unidade de medida de profundidade antes da padronização para metro

Poço	Unidade
P_459	pol
P_551	pol
P_574	m
P_595	m
P_605	pol
P_643	pol
P_645	m
P_659	m
P_689	pol
P_696	pol
P_698	pol

Fonte: Autor

3.3.2 Remoção registros que não possuem todas as curvas escolhidas

Em determinados poços, o arquivo dlis organizava os registros de perfilagem

em diversos quadros de dados, com diferentes intervalos de profundidade. Dentre esses diversos intervalos, alguns não possuíam registro de todas as curvas de perfis petrofísicos escolhidas para o trabalho. Para garantir que o algoritmo seja treinado com todo o conjunto de variáveis definido, foi realizada uma filtragem, preservando apenas as amostras que possuíam todas as curvas requisitadas para o estudo. A tabela <num_tabela> mostra quais poços passaram por essa etapa de filtragem.

3.3.3 Manutenção de apenas uma curva de resistividade profunda

Em certos poços, foram identificadas múltiplas curvas de resistividade profunda. Nesses casos, foi imprescindível manter apenas uma curva e eliminar as demais, a fim de evitar que a duplicação dessa característica influenciasse negativamente o processo de aprendizado do modelo. A tabela <num_tabela> apresenta as curvas de resistividade profunda originalmente presentes em cada poço antes do processamento desta etapa. Nos casos em que a filtragem foi necessária, priorizou-se a manutenção da curva RT, com a subsequente remoção das demais.

3.3.4 Padronização do nome das curvas de resistividade profunda

De acordo com o especialista Luiz Henrique Vandelli, existem inúmeros mnemônicos de curvas de resistividade profunda que poderiam ser utilizadas no estudo. Conforme acontece em companhias de perfilagem, existem dois princípios de funcionamento, por indução e laterolog. O princípio da indução engloba as curvas ILD, RILD, IEL, AIT90, AHT90, RT90, AT90, AO90, RT, AF90, AHF90, AFH90. Enquanto o laterolog possui as curvas LLD, RLLD, HDRS, HLLD, LL7 E RLL7.

Com o objetivo de padronizar a nomenclatura deste perfil, foi acordado utilizar o termo RESD para se referir a uma curva de resistividade profunda. A tabela <num_tabela> apresenta a nomenclatura dessas curvas antes e depois da padronização em cada um dos poços.

3.3.5 Aplicação dos mnemônicos das curvas

Assim como acontece com os perfis de resistividade profunda, existem

curvas equivalentes para outros perfis utilizados, com uma variação de nomenclatura. Desse modo, a fim de padronizar os nomes dessas variáveis, duas ou mais curvas equivalentes tiveram seus nomes renomeados para um nome padrão escolhido.

Na tabela <num_tabela> estão apresentados os conjuntos de curvas equivalentes e o nome padrão que foi definido para ser utilizado neste trabalho.

3.3.6 Adição da curva DCAL

Para fins de controlar a qualidade dos dados utilizados, foi adicionado o perfil de controle DCAL, que é o resultado da diferença entre o valor de CAL e BS. Por meio do DCAL é possível identificar se uma determinada região de um poço tem reboco (DCAL negativo) ou arrombamento (DCAL positivo). Para evitar que regiões que possuam elevado grau de reboco ou arrombamento impactem negativamente o modelo, os registros de dados utilizados neste trabalho possuem o DCAL dentro do intervalo [-1, 1.5].

3.3.7 Reclassificação da litologia metamórfica

Segundo o especialista Vitor Hugo Simon, geólogo consultor do projeto, as litologias presentes no Embasamento Fraturado da Sub-bacia de Sergipe, na região do Alto de Aracaju são: principalmente rochas metamórficas, xisto e filito, e secundariamente rochas graníticas.

Durante o pré-processamento dos dados, foi identificado no poço P_459, um tipo de litologia descrita como “metamórfica não identificada”, referente ao Embasamento da Sub-bacia de Sergipe. Sendo esse um termo muito genérico para classificar rochas e com a intenção de obter uma caracterização mais precisa das litologias, as amostras de “metamórfica não identificada” foram submetidas a um algoritmo de classificação litológica para se obter uma caracterização mais precisa do tipo litológico.

Para realizar a reclassificação, foi utilizado um modelo de classificação de floresta aleatória, treinado com as propriedades de perfis das amostras litológicas de xisto e filito, presentes nos poços P_689 e P_659. Após o treinamento, o modelo

foi capaz de rotular as amostras de “metamórfica não identificada” como uma dessas duas categorias.

No conjunto de dados em estudo, havia 540 amostras de xisto e 202 de filito, nos poços P_689 e P_659 respectivamente. Dessa maneira, o algoritmo de classificação foi treinado e testado com um total de 642 registros.

A divisão das amostras em treinamento e teste foi realizada separando 2/3 do total para treinamento e 1/3 para teste, conforme recomendado na literatura. (FACELI et al. 2011) Os resultados da divisão são apresentados na tabela <num_tabela>.

Tabela <num_tabela> - Divisão das amostras de xisto e filito em conjuntos de treino e teste

Litologia	Número de amostras	Amostras para treinamento	Amostras para teste
Filito	202	132	70
Xisto	540	362	178
Total	642	494	248

Fonte: Autor

Uma vez que o algoritmo foi treinado, a sua acurácia pôde ser testada com o conjunto de teste. Isto é, após aprender com as amostras para treinamento, o modelo foi testado ao classificar, em xisto ou filito, a litologia correspondente aos 248 registros de propriedades de perfis petrofísicos separados para teste. A acurácia do modelo foi de 99.59%.

Em seguida, após as etapas de treinamento e teste, o algoritmo foi utilizado para categorizar as amostras de “metamórfica não identificada” em xisto ou filito, com base em suas respectivas propriedades. Dentre os 585 registros de metamórfica que existiam no conjunto de dados, 573 foram reclassificados como xisto e apenas 12 como filito.

De acordo com os especialistas, não foi observado, até então, no intervalo estratigráfico do EMB de um poço, a existência de mais de um tipo litológico no intervalo normalmente perfurado. Desse modo, todos os registros de “metamórfica

não identificada” foram convertidos em xisto, e os filitos foram considerados como erros de classificação.

3.4 Algoritmo

Neste trabalho foi empregado o método de *random forest*, algoritmo baseado em árvore com abordagem supervisionada, para caracterizar eletrofáceis. Um modelo *random forest* de classificação é uma coleção de árvores de decisão, da qual a previsão final é obtida através de uma votação majoritária entre as previsões realizadas por cada uma das árvores individuais que compõem o algoritmo. (Shalev-Shwartz, S.; Ben-David, S. 2014) Desse modo, a classe mais frequentemente prevista pelas árvores será a previsão final do modelo.

3.5 Algoritmo

Neste trabalho foi empregado o método de *random forest*, algoritmo baseado em árvore com abordagem supervisionada, para caracterizar eletrofáceis. Um modelo *random forest* de classificação é uma coleção de árvores de decisão, da qual a previsão final é obtida através de uma votação majoritária entre as previsões realizadas por cada uma das árvores individuais que compõem o algoritmo. (Shalev-Shwartz, S.; Ben-David, S. 2014) Desse modo, a classe mais frequentemente prevista pelas árvores será a previsão final do modelo.

4. Resultados e discussões

Esta seção é destinada à análise dos resultados obtidos com o modelo de aprendizado de máquina para caracterização de eletrofáceis. Neste trabalho foram treinados quatro modelos diferentes de classificação, com as seguintes características de dados de entrada:

- I. Dados desbalanceados e sem unidades geológicas como variáveis preditivas; (seção <>)
- II. Dados desbalanceados e com unidades geológicas como variáveis preditivas; (seção <>)

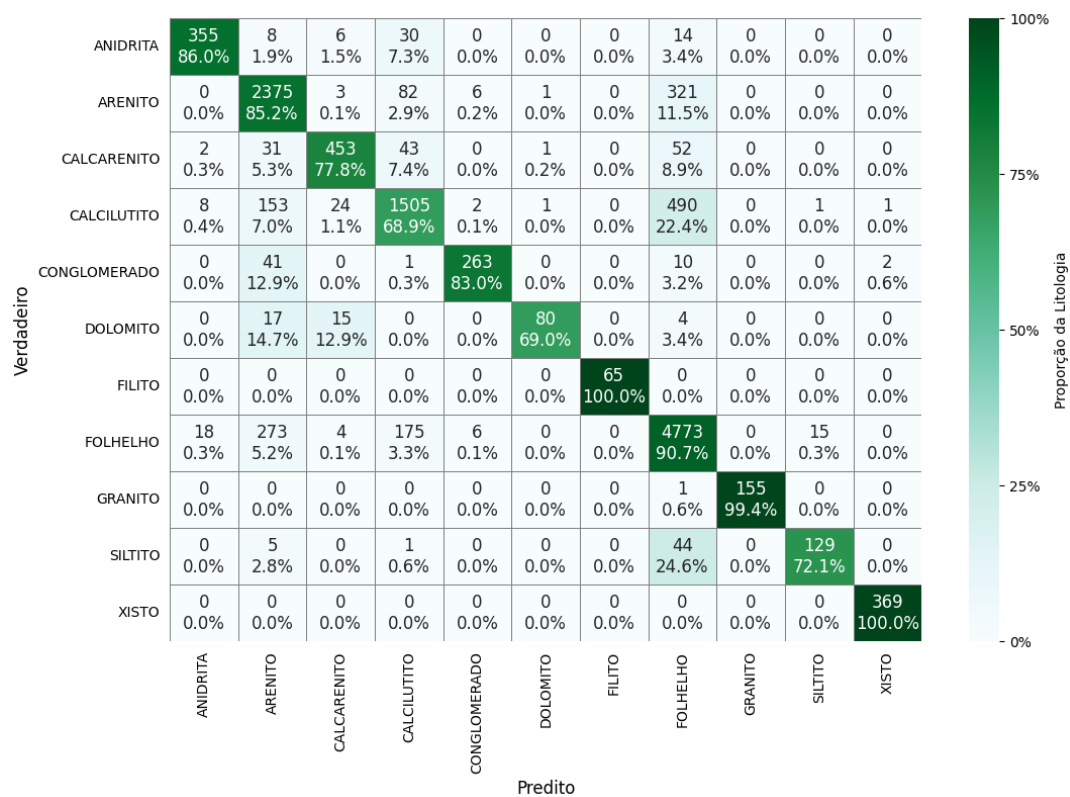
- III. Dados balanceados com *undersampling* e sem unidades geológicas como variáveis preditivas; (seção <>)
- IV. Dados balanceados com *undersampling* e com unidades geológicas como variáveis preditivas. (seção <>)

4.1 Dados desbalanceados e sem unidades geológicas

O primeiro modelo foi treinado apenas com as curvas de perfis petrofísicos discriminatórios como variáveis preditivas, ou seja, não se utilizou das unidades geológicas na etapa de treinamento. Ademais, os dados não foram balanceados, de modo que determinados tipos litológicos tiveram um número elevado de amostras de treinamento e teste. A tabela <num_tabela> mostra o número de registros de cada tipo litológico no conjunto de teste.

A figura <num_figura> é a matriz de confusão do primeiro modelo treinado, do qual teve seu desempenho testado com o seu conjunto de teste. No eixo y encontram-se as classes litológicas verdadeiras e no eixo x as preditas. Cada célula da matriz possui dois valores: a quantidade absoluta e a porcentagem de amostras da litologia da linha (verdadeira) que foram classificadas como a litologia da coluna (predita).

Figura <num_tabela> - Matriz de confusão do teste do primeiro modelo



Fonte: Autor

4.2 Y

4.3 Z

4.4 W

5. Conclusões

6. Perspectivas de futuros trabalhos

7. Referências bibliográficas

8. Outras atividades

9. Justificativa