

**CLASSIFICAÇÃO DE LITOFÁCIES ATRAVÉS DA  
ANÁLISE AUTOMÁTICA DE PERFIS ELÉTRICOS DE  
POÇOS DE PETRÓLEO DA AMAZÔNIA**



JOACIR MARQUES OLIVEIRA JÚNIOR

**CLASSIFICAÇÃO DE LITOFÁCIES ATRAVÉS DA  
ANÁLISE AUTOMÁTICA DE PERFIS ELÉTRICOS DE  
POÇOS DE PETRÓLEO DA AMAZÔNIA**

Dissertação apresentada ao Programa de Pós-  
-Graduação em Informática do Instituto de  
Computação da Universidade Federal do  
Amazonas como requisito parcial para a  
obtenção do grau de Mestre em Informática.

ORIENTADORES: EULANDA MIRANDA DOS SANTOS E JOSÉ REGINALDO  
HUGHES CARVALHO

Campus Universitário Senador Arthur Virgílio Filho

Janeiro de 2014

© 2014, Joacir Marques Oliveira Júnior.  
Todos os direitos reservados.

Oliveira Júnior, Joacir Marques

Classificação de Litofácies Através da Análise Automática de  
Perfis Elétricos de Poços de Petróleo da Amazônia / Joacir Marques  
Oliveira Júnior. — Campus Universitário Senador Arthur Virgílio  
Filho, 2014

xvi, 63 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal do Amazonas  
Orientador: Eulanda Miranda dos Santos e José Reginaldo  
Hughes Carvalho

1. Computação — Teses. 2. Visão Computacional — Teses.  
I. Orientadora. II. Título.

# Agradecimentos

Nada neste mundo se conquista sozinho. Definitivamente em tudo que se conquista, em algum momento houve a contribuição de alguém ou mesmo de várias pessoas possibilitando o feito. Esta dissertação de mestrado não é exceção. Apesar da alegria em concluir este trabalho, um pouco de tristeza surge em mim por estar fechando este ciclo. A experiência que tive nos últimos anos me possibilitou conhecer um pouco melhor a mim mesmo. Me deixou descobrir um pouco mais dos meus limites. De estudar com os colegas nos fins de semana para poder atingir boas notas, de conhecer pessoas maravilhosas entre colegas e professores. Me fez ver o quão excelente é nossa Universidade Federal do Amazonas, em especial o Instituto de Computação onde com frequência vi colegas sendo destaques de conferências internacionais. Foi uma honra estar com vocês e agradeço à Deus por me permitir estar aqui com saúde. Aos meus pais que não mediram esforços para me dar a melhor educação possível. Aos doutores José Raimundo e meu amigo Ricardo Belém que me concederam a carta de recomendação a esta instituição. À Petrobras por meio dos meus gerentes Clelson Cabral e Jorge Oliveira que me permitiram dividir meu tempo entre o meu trabalho e os estudos. Ao gerente José Mafra e ao consultor Paulo Soeiro pela disponibilização dos dados para a pesquisa. A minha querida orientadora Dra. Eulanda Miranda. Sou especialmente grato à senhora por todo o aprendizado que me proporcionou. Uma grande parcela do meu crescimento como profissional e pesquisador eu devo à senhora e mais uma vez agradeço à Deus por tê-la posto no meu caminho. Ao meu co-orientador Dr. José Reginaldo por toda a atenção, conselhos e incentivos. A todos os meus professores e colegas e por último a minha amada esposa Leyne Marques. Obrigado meu amor por estar comigo nestes 18 anos. Você é a melhor parte de mim. Eu te amo e agradeço mais uma vez a Deus. Obrigado senhor por ter posto tantas pessoas maravilhosas em minha vida.



# Resumo

Dentro das várias etapas que são necessárias até o petróleo ser comercializado, a análise de perfis elétricos representa papel de grande importância para se estimar a capacidade produtiva de um poço. A análise hoje é semi-automatizada e ocorre da seguinte forma: geólogos especialistas analisam gráficos de curvas gerados por um sistema, para então, realizar a caracterização do reservatório com base nas análises. Um dos objetivos dessa análise é a classificação de litofácies. Litofácies são unidades litológicas (rochas) que caracterizam o ambiente de formação e aspectos composicionais das rochas. Para que se forme um reservatório de petróleo, um conjunto de tipos de rochas precisa estar presente, sendo este um dos principais motivos para a classificação de litofácies. Esta dissertação de mestrado investiga o uso de técnicas de classificação automática aplicadas ao problema de classificação de litofácies. Serão investigados os seguintes cinco métodos de classificação: Support Vector Machines, k Vizinhos Mais Próximos, Multilayer Perceptron e Regressão Logística. A base de dados investigada é composta por amostras de perfis de três poços de uma reserva da Amazônia. Será ainda comparado o desempenho de classificadores individuais frente à combinação dos mesmos através do voto majoritário. Por fim, iremos verificar se o treinamento de um poço pode ser aproveitado para outro por meio de classificadores individuais e combinados por voto majoritário. Para obter essas respostas, fizemos dois tipos de testes. No primeiro, treinamos e testamos os classificadores individualmente e combinados dentro do mesmo poço. Os resultados apresentados mostraram que Support Vector Machines foi superior em dois dos três poços, enquanto Multilayer Perceptron, superou os demais métodos no terceiro poço. No segundo tipo de testes, treinamos com dados de um poço e testamos com dados de outro poço, simulando uma situação mais próxima do problema real que seria de calibrar os classificadores de uma reserva com um poço pioneiro e a partir daí replicar nos poços vizinhos. Nestes testes, a combinação de classificadores se mostrou a melhor solução em 4 das 6 combinações possíveis. Nas duas demais combinações, a combinação por voto majoritário alcançou o segundo melhor resultado. Vale dizer ainda que na média simples o sistema de votação majoritário, foi a melhor opção para classificar as litofácies. Nossos resultados indicam que combinar classificadores em um sistema de votação majoritário apre-

sentado desempenho superior ao uso de classificadores individuais, além de apresentar maior estabilidade.

**Palavras-chave:** Visão Computacional, Perfilagem, Reconhecimento de Padrões, kNN, SVM, Litofácies.



# Abstract

Among several steps which are necessary for the commercialization of oil, the analysis of well logs plays an important role to estimate the capacity of a well. Traditionally, this analysis is conducted in a semi-automated process which generates graphs of curves used by human experts to analyze and make the reservoir characterization. One goal of this analysis is to classify lithofacies. Lithofacies are lithological units(rocks) that characterize the environment and compositional aspects of the rocks. In order to characterize an oil reservoir, a set of classes of sedimentary rocks occur. This is which is the major reason for the classification of lithofacies. This master thesis investigates the use of automatic classification techniques applied to the problem of classification of lithofacies. The following five classification methods are investigated: Support Vector Machines, k-Nearest Neighbor, Multilayer Perceptron and Logistic Regression. The database investigated consists of samples from three oil wells of the same reservoir in the Amazon State. In addition, the performance of individual classifiers are compared to the combination of the same five classifiers through majority voting. Finally, we will verify whether or not individual classifiers, or ensemble of classifiers, may train using data obtained from one well and accurately classify data from other wells. In order to get these answers, we have run two series of experiments. First, we trained classifiers and test classifiers individually and combined within the same oil well. The obtained results show that Support Vector Machines achieved the best results in two of the three wells, while Multilayer Perceptron outperformed the other methods in the third well. In the second series of experiments, we trained classifiers with data from a well and then with data from another well, simulating a situation closer to a real application, since we may use a manually classified database to train a classifier, or ensemble of classifiers, in order to learn the pattern of the reservoir. Then, data from other wells of the same reservoir may be automatically classified. In this test, the ensemble of classifiers outperformed individual classifiers in 4 of the 6 possible combinations. In the two other combinations, the combination by majority vote was the second best. It is also worth saying that in average, ensemble of classifiers was the best option to classify lithofacies. Our results indicate that combining classifiers in a system of majority voting, shows a better performance and better stability of the results.

**Keywords:** Computer Vision, Well Logging, Pattern Recognition, kNN, SVM, Lithofacies.

# Lista de Figuras

|      |  |    |
|------|--|----|
| 2.1  | Torre de destilação que separa os compostos do petróleo. . . . .   | 7  |
| 2.2  | Formação de petróleo. Na figura podemos ver o processo migratório da rocha mãe ou rocha geradora até a rocha reservatório. . . . .   | 9  |
| 2.3  | Pré-Sal. . . . .   | 11 |
| 2.4  | Sísmica marítima utilizando canhões de pressão. . . . .  | 12 |
| 2.5  | Resultado da seção sísmica. Cabe ao geofísico analisar a imagem interpretando as formações. . . . .  | 12 |
| 2.6  | Exemplo de fluido de perfuração, o qual entre outras funções, ajuda a remover os fragmentos das rochas resultantes da perfuração. . . . .  | 13 |
| 2.7  | Diferentes tipos de broca. Cada uma é adequada a uma fase de perfuração e ao tipo de rocha que está sendo perfurado. . . . .   | 14 |
| 2.8  | Sonda terrestre para perfurações em terra. . . . .   | 15 |
| 2.9  | Diversos tipos de sonda marítima para perfurações no mar. A complexidade aumenta com a profundidade a ser atingida. . . . .  | 15 |
| 2.10 | Perfis elétricos coletados durante a perfuração de um poço. . . . .  | 21 |
| 2.11 | Amostras de testemunho retirados de um poço. Apesar de ser a melhor representação do perfil de um poço, é um processo longo e caro, sendo feito portanto somente em alguns poços estratégicos. . . . . | 22 |
| 2.12 | Ferramentas de perfuração utilizadas para extração das características de um poço. . . . .   | 23 |
| 2.13 | Porosidade e permeabilidade. . . . .   | 24 |
| 2.14 | Acumulações de Petróleo. . . . .   | 24 |
| 3.1  | Ilustração sobre o funcionamento de SVM aplicado à classificação de rochas sedimentares. . . . .   | 27 |
| 3.2  | Exemplo ilustrativo do funcionamento de kNN, bem como o processo de ajuste do parâmetro k. . . . .   | 28 |
| 3.3  | Exemplo ilustrativo do funcionamento de Multilayer Perceptron. . . . .   | 29 |
| 3.4  | Ilustração dos cinco Classificadores em conjunto por meio do Voto majoritário. . . . .   | 32 |

|     |  |    |
|-----|--|----|
| 4.1 | Distribuição quantitativa das quatro classes encontradas nos três poços. . . . .                           | 37 |
| 4.2 | Exemplos visuais de rochas representadas na base de dados investigada nos experimentos. . . . .            | 38 |
| 4.3 | Gráfico do desempenho dos classificadores treinados e testados no mesmo poço.                              | 40 |
| 4.4 | Gráfico do desempenho dos classificadores treinados em um poço e testado nos demais. . . . .               | 42 |
| 4.5 | Box-plot para índice de acertos dos classificadores . . . . .  | 43 |
| 4.6 | Desempenho Individual dos Classificadores e Combinados em Conjunto Utilizando Votação Majoritária. . . . . | 44 |

# Lista de Tabelas

|     |  |    |
|-----|--|----|
| 1.1 | Distribuição da Produção de Petróleo e Gás Natural no Brasil por Estado (2011).                              | 2  |
| 2.1 | Diversos tipos de compostos de carbonetos de acordo com a temperatura de ebulição. . . . .                   | 8  |
| 2.2 | Principais Tipos de perfis elétricos e suas propriedades. . . . .  | 19 |
| 4.1 | Ajuste de parâmetros nos classificadores. . . . .  | 39 |
| 4.2 | Exemplo de amostras da base de dados . . . . .   | 39 |
| 4.3 | Taxa de acerto dos métodos de classificação investigados. . . . .  | 40 |
| 4.4 | Taxa de acerto de poço para poço dos métodos de classificação investigados. . .                              | 41 |
| 4.5 | Resumo das Estatísticas Descritivas . . . . .  | 42 |
| 4.6 | Desempenho de poço para poço dos métodos de classificação investigados comparados ao Random Forrest. . . . . | 45 |



# Sumário

|  |             |
|--|-------------|
| <b>Agradecimentos</b>                        | <b>v</b>    |
| <b>Resumo</b>                                | <b>vii</b>  |
| <b>Abstract</b>                              | <b>ix</b>   |
| <b>Lista de Figuras</b>                      | <b>xi</b>   |
| <b>Lista de Tabelas</b>                      | <b>xiii</b> |
| <b>1 Introdução</b>                          | <b>1</b>    |
| 1.1 Motivação . . . . .                      | 2           |
| 1.2 Objetivos . . . . .                      | 4           |
| 1.2.1 Geral . . . . .                        | 4           |
| 1.2.2 Específicos . . . . .                  | 4           |
| 1.3 Organização do Documento . . . . .       | 5           |
| <b>2 Petróleo, o Ouro Negro</b>              | <b>7</b>    |
| 2.1 Noções de Geologia de Petróleo . . . . . | 8           |
| 2.1.1 Origem do Petróleo . . . . .           | 8           |
| 2.1.2 Migração do Petróleo . . . . .         | 9           |
| 2.2 Prospecção de Petróleo . . . . .         | 10          |
| 2.3 Métodos geológicos . . . . .             | 10          |
| 2.4 Métodos Sísmicos . . . . .               | 11          |
| 2.5 Perfuração . . . . .                     | 13          |
| 2.6 Avaliação de Formações . . . . .         | 14          |
| 2.6.1 Testemunhagem . . . . .                | 16          |
| 2.6.2 Amostras de Calha . . . . .            | 17          |
| 2.6.3 Perfilagem Elétrica . . . . .          | 17          |

|          |  |           |
|----------|--|-----------|
| 2.6.4    | Fundamentos de Perfilagem Elétrica . . . . .   | 18        |
| <b>3</b> | <b>Fundamentação Teórica e Trabalhos Relacionados</b>  | <b>25</b> |
| 3.1      | Reconhecimento de Padrões . . . . .  | 25        |
| 3.2      | Máquinas de Vetores de Suporte - SVM . . . . .   | 26        |
| 3.3      | k-Vizinhos mais Próximos - kNN . . . . .   | 26        |
| 3.4      | Redes Neurais Artificiais (RNA) - Multilayer Perceptron . . . . .  | 28        |
| 3.5      | Regressão Logística . . . . .  | 30        |
| 3.6      | Árvores de Decisão - J48 . . . . .   | 30        |
| 3.7      | Conjunto de Classificadores . . . . .  | 30        |
| 3.7.1    | Votação Majoritária . . . . .  | 31        |
| 3.8      | Trabalhos Relacionados . . . . .   | 31        |
| <b>4</b> | <b>Experimentos e Resultados</b>   | <b>35</b> |
| 4.1      | Protocolo Experimental . . . . .   | 35        |
| 4.1.1    | Base de Dados . . . . .  | 36        |
| 4.1.2    | Ajuste de Parâmetros . . . . .   | 37        |
| 4.1.3    | Normalização da Base . . . . .   | 38        |
| 4.1.4    | Desempenho Individual e Combinado com Treino e Teste no Mesmo Poço . . . . .   | 39        |
| 4.1.5    | Desempenho poço a poço dos classificadores. . . . .  | 41        |
| 4.1.6    | Análise Estatística dos Resultados dos experimentos de Poço a Poço. . . . .  | 42        |
| 4.2      | Resumo dos resultados . . . . .  | 44        |
| <b>5</b> | <b>Conclusões e Trabalhos Futuros</b>  | <b>47</b> |
| 5.1      | Trabalhos Futuros . . . . .  | 47        |
| 5.2      | Publicações Obtidas . . . . .  | 48        |
|          | <b>Referências Bibliográficas</b>  | <b>49</b> |
|          | <b>Apêndice A Artigo publicado no Congresso Latino-Iberoamericano de Investiga-<br/>ção Operativa e Simpósio Brasileiro de Pesquisa Operacional no ano de<br/>2012</b> | <b>51</b> |



# Capítulo 1

## Introdução

A constatação de que a Amazônia é uma região petrolífera aconteceu no dia 13 de março de 1955 [Garcia, 2008]. A euforia tomou conta principalmente do povo do Amazonas quando jorrou petróleo no poço pioneiro batizado de NO-1-AZ, a uma profundidade de 2744 metros e distante 125 quilômetros de Manaus. Essa notícia logo ganhou cenário internacional. Na época acreditava-se que em apenas três anos o Brasil estaria livre da importação de combustíveis, tão grande era o otimismo dos engenheiros da Petrobras. A perspectiva de produção do poço pioneiro NO-1-AZ depois dos primeiros testes realizados era de 600 a 700 barris por dia. Essa perspectiva de produção era suficiente para estimular mais pesquisas e a abertura de novos poços na região. Atualmente, momento em que a produção de petróleo declina mundialmente e seus preços disparam no mercado internacional, a indústria do petróleo brasileira cresce vigorosamente. Hoje o Brasil produz mais de 2,5 milhões de barris por dia e o estado do Amazonas é responsável por cerca de 107.496 barris diários (Tabela 1.1). Segundo o plano de negócios da Petrobras [Petrobras, 2011], a empresa espera duplicar suas reservas provadas até 2020.

Apesar de um momento promissor da produção de petróleo nacional, a produção de petróleo ainda dispense investimentos de cifras milionárias. Antes de um poço começar a produzir, existe um longo e caro processo que envolve o estudo e a análise de dados geológicos e geofísicos. Somente depois de uma exaustiva análise que favoreça a perspectiva positiva de produção em escala comercial é que os engenheiros decidem pela perfuração de um poço. Mesmo depois de perfurado, há ainda a necessidade de se confirmar a viabilidade econômica de exploração da reserva. Uma vez que se iniciou a perfuração de um poço, são descidas ainda durante a perfuração algumas ferramentas com o objetivo de avaliar melhor as características da formação. Essa avaliação é fundamental para justificar a continuidade da exploração da reserva. Se este for o poço pioneiro, provavelmente testes mais caros, porém mais precisos, serão realizados como a testemunhagem e amostra lateral. Esse último mais

**Tabela 1.1.** Distribuição da Produção de Petróleo e Gás Natural no Brasil por Estado (2011).

| Estado              | Petróleo (bbl/d) | Gás Natural (Mm <sup>3</sup> d) | Produção Total (boe/d) |
|---------------------|------------------|---------------------------------|------------------------|
| Rio de Janeiro      | 1.551.597        | 25.360                          | 1.711.114              |
| Espírito Santo      | 319.518          | 11.901                          | 394.373                |
| Amazonas            | 35.110           | 11.508                          | 107.496                |
| Bahia               | 42.880           | 6.565                           | 84.172                 |
| São Paulo           | 42.225           | 5.169                           | 74.739                 |
| Rio Grande do Norte | 58.639           | 1.701                           | 69.338                 |
| Sergipe             | 42.365           | 2.428                           | 57.640                 |
| Alagoas             | 5.198            | 1.539                           | 14.878                 |
| Ceará               | 7.209            | 86                              | 7.752                  |
| <b>Total geral</b>  | <b>2.104.743</b> | <b>66.257</b>                   | <b>2.521.502</b>       |

barato que o primeiro.

Esta fase é denominada avaliação de formações. Trata-se de atividades e estudos que visam definir em termos qualitativos e quantitativos o potencial de uma jazida petrolífera, isto é, sua capacidade produtiva e a valoração de suas reservas de óleo e gás. Essa avaliação é baseada principalmente nos dados coletados na perfilagem de um poço aberto. A perfilagem é uma imagem visual, em relação à profundidade, de uma ou mais características ou perfis das rochas perfuradas tais como resistividade elétrica, potencial eletroquímico natural, tempo de trânsito das ondas mecânicas, radioatividade natural ou induzida. Com perfilagem busca-se identificar propriedades das rochas e fluidos medindo informações elétricas, acústicas, termais e radioativas sendo possível extrair dados importantes das formações pelas quais o poço atravessa como a litologia (tipo de rocha), espessura, porosidade, presença de óleo, gás ou água. Dependendo da análise dos perfis e das informações complementares, pode-se decidir pelo abandono do poço ou, no caso de se provar interesse econômico, os testes de formação para se estimar a capacidade produtiva do poço são iniciados.

Na análise de perfis, há oito principais tipos de perfis elétricos: Raios Gama (GR), Neutrônico (NPHI), Indução (ILD), Sônico (DT), Densidade (RHOB), Caliper (Cal), Micro SFL (MSFL) e Índice Fotoelétrico (PEF). A descrição de cada perfil é mostrada na Tabela 2.2.

## 1.1 Motivação

A identificação de litofácies de um reservatório de petróleo é um processo intensivo que envolve o gasto de uma quantidade considerável de tempo por parte de um especialista experiente. O problema se torna muito mais difícil à medida que aumenta o número de perfis

(medidas de determinadas propriedades da formação geológica) simultâneos a serem analisados [Cunha, 2002].

Com base na análise destes perfis, poderá ser criada uma imagem a cada metro de várias características do poço como tipo de rocha, fluidos, características da formação, etc. É um processo que requer bastante habilidade e paciência dos geólogos e geofísicos e sua automatização trará por certo grandes ganhos para o estudo de reservatórios. Além disso, a caracterização de um reservatório é de fundamental importância para justificar investimentos da continuidade de exploração de um poço ou o seu abandono.

A classificação de perfis de poços de petróleo pode claramente ser tratada como um problema de reconhecimento automático de padrões. Não é surpresa a existência de trabalhos na literatura que realizam a classificação dos perfis de poços com ferramentas de reconhecimento de padrões Cunha [2002], Flexa et al. [2004], Wong et al. [2005] e An-nan & Lu [2009] sendo a maioria deles utilizando redes neurais e Máquinas de Vetores de Suporte (do Inglês, SVM).

Muitos trabalhos foram realizados com o intuito de analisar automaticamente estes perfis. No entanto, devido a heterogeneidade das diversas bacias sedimentares com características peculiares, o aprendizado em um estudo de uma bacia muitas vezes demonstrando taxas de acerto elevadas, quando levado a bacias diferentes, estes mesmos bons resultados não se repetem. O conhecimento do especialista geofísico é crucial nesse momento. Não somente o seu conhecimento acadêmico, mas principalmente o seu conhecimento da região em estudo. É ele que terá condições, portanto, de avaliar se um poço possui características semelhantes, avaliando assim se os padrões verificados em um poço se repetem em um poço distinto. É nesse ponto que esse trabalho se destaca. Iremos trabalhar sempre com um poço chave analisado por um especialista geofísico. O aprendizado do mesmo será propagado a novos poços com características semelhantes quanto à formação geológica.

Outro diferencial neste trabalho, que será explicado mais a frente, é a combinação dos classificadores em voto majoritário. Foram testados 5 classificadores que individualmente mostraram menos estabilidade do que os mesmos 5 em conjunto. Sendo mais específico, quando combinados em conjunto, os classificadores se mantiveram em taxas de acerto elevadas para todos os cenários testados sendo que, na média, a combinação atingiu a mais alta taxa de acerto.

Diante desse contexto, este trabalho aplica métodos de classificação ao problema de identificação automática de litofácies utilizando dados de perfis de três poços de petróleo da Amazônia da mesma reserva. Como já foi dito, há muitos trabalhos publicados utilizando técnicas de reconhecimento de padrões. Neste trabalho, além de comparar o desempenho individual de cinco classificadores, é apresentado um sistema de votação majoritária onde escolhe-se a classe com maior número de votos dos classificadores a partir da combinação

dos mesmos classificadores listados abaixo:

1. Máquinas de Vetores de Suporte(SVM);
2. Multilayer Perceptron(MLP);
3. Regressão Logística;
4. k-Vizinhos Mais Próximos(kNN);
5. Árvore(J48).

## 1.2 Objetivos

### 1.2.1 Geral

O objetivo geral desta dissertação é apresentar uma análise comparativa entre métodos de classificação aplicados ao problema de classificação automática de litofácies utilizando dados de diferentes poços de petróleo da mesma reserva.

### 1.2.2 Específicos

Os objetivos específicos são:

- Comparar classificadores e apontar o melhor método para a classificação automática de litofácies utilizando dados de perfis de poços de petróleo divididos nos seguintes estudos:
  - Comparar classificadores dentro do mesmo poço;
  - Comparar classificadores dentro de poços diferentes;
  - Comparar classificadores individuais e em conjuntos;
  - Constatar se o treinamento em um poço pode ser aproveitado em um poço distinto com características semelhantes.
- Combinar classificadores e apontar o melhor método para a classificação automática de litofácies utilizando dados de perfis de poços de petróleo divididos nos seguintes estudos:
  - Combinar classificadores dentro do mesmo poço;
  - Combinar classificadores dentro de poços diferentes;
  - Combinar classificadores individuais e em conjuntos;

- Constatar se o treinamento em um poço pode ser aproveitado em um poço distinto com características semelhantes.

## **1.3 Organização do Documento**

O restante deste documento está organizado da seguinte forma: no capítulo 2 são descritos vários conceitos importantes sobre a origem e exploração do petróleo. No capítulo 3 é descrita a fundamentação teórica necessária para a compreensão de reconhecimento de padrões, classificadores e conjunto de classificadores, bem como os trabalhos relacionados.

No capítulo 4, são apresentados a base de dados, normalização da base e descrição dos experimentos realizados com os classificadores através de testes individuais e em conjunto por meio de votação majoritária.

No capítulo 5, finalizamos com as conclusões deste trabalho.

Assim, no próximo capítulo, serão introduzidos os conceitos sobre a origem do petróleo e sua exploração.

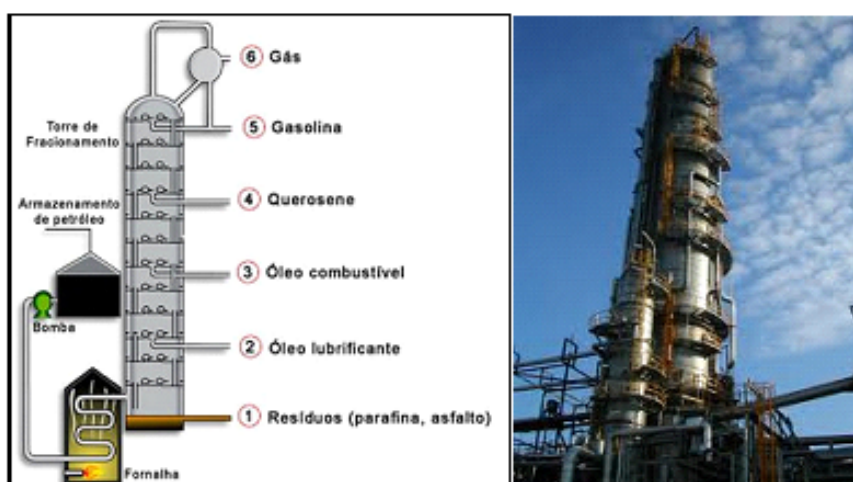


## Capítulo 2

# Petróleo, o Ouro Negro

Segundo Thomas et al. [2001], a palavra Petróleo vem do latim pedra(pedra) e oleum(óleo), o petróleo no estado líquido é uma substância oleosa, inflamável, menos densa que a água, com cheiro característico e cor variando entre o negro e o castanho claro. O petróleo é formado por centenas de compostos químicos, sendo os mais importantes para a geração de derivados, os hidrocarbonetos (moléculas de hidrogênio e carbono). Os derivados do petróleo como querosene, gasolina, lubrificante e outros, são obtidos de acordo com a faixa de ebulição por meio de uma torre de destilação ou fracionamento (Figura 2.1).

Através da torre de destilação (Figura 2.1), em uma escala crescente de temperatura, é possível separar cada composto de acordo com a Tabela 2.1.



**Figura 2.1.** Torre de destilação que separa os compostos do petróleo.

As características do petróleo diferem para cada reservatório, podendo ser observados óleos mais viscosos, densos, pretos, com bastante gás ou pouco gás. Muitas vezes também

**Tabela 2.1.** Diversos tipos de compostos de carbonetos de acordo com a temperatura de ebulição.

| Fração  | Temperatura de ebulição (°C) | Composição aproximada  | Usos  |
|---|------------------------------|------------------------|---|
| Gás residual e Gás liquefeito de petróleo - GLP | Até 40                       | $C_1 - C_2, C_3 - C_4$ | gás combustível, gás combustível engarrafado, uso doméstico e industrial. |
| Gasolina  | 40 - 175                     | $C_5 - C_{10}$         | combustível de automóvel, solvente.                                       |
| Querosene                                       | 175 - 235                    | $C_{11} - C_{12}$      | iluminação, combustível de aviões a jato.                                 |
| Gasóleo leve                                    | 235 - 305                    | $C_{13} - C_{17}$      | diesel, fornos.   |
| Gasóleo pesado                                  | 305 - 400                    | $C_{18} - C_{25}$      | combustível, matéria-prima para lubrificantes.                            |
| Lubrificantes                                   | 400 - 510                    | $C_{26} - C_{38}$      | óleos lubrificantes.  |
| Resíduo   | Acima de 510                 | $C_{38}+$              | asfalto, piche, impermeabilizantes.                                       |

é encontrado apenas gás em um reservatório. Entretanto, além de quantidades menos significativas de enxofre, nitrogênio, oxigênio e metais, uma grande quantidade de hidrogênio e carbono é sempre presente em torno de 11 a 14% de hidrogênio e entre 83 a 87% de carbono. Pela alta quantidade de hidrogênio e carbono, esses compostos são conhecidos como hidrocarbonetos.

## 2.1 Noções de Geologia de Petróleo

### 2.1.1 Origem do Petróleo

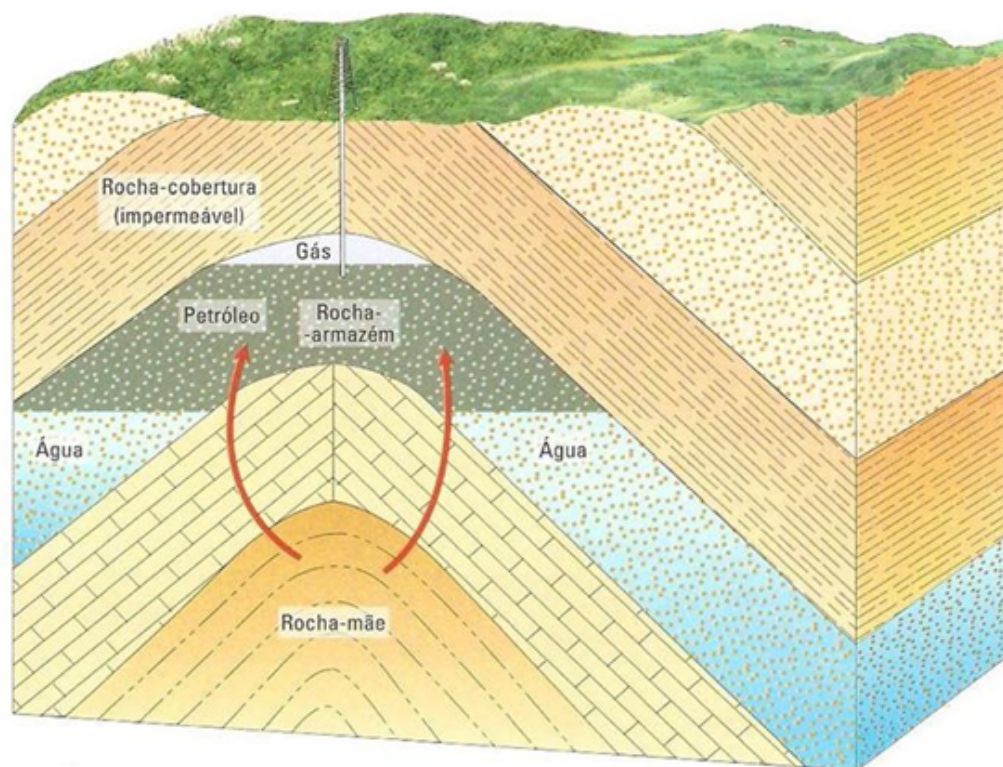
O petróleo tem origem na matéria orgânica depositada junto com os sedimentos. A matéria orgânica marinha é basicamente originada de microorganismos e algas que formam o fitoplâncton e não pode sofrer processo de oxidação [Thomas et al., 2001]. Assim, para que óleo ou gás seja formado, há a necessidade de condições não oxidantes, geralmente encontrados em compostos de baixa densidade, não permitindo a ação de água circulante. Ainda segundo Thomas et al. [2001], a interação destes fatores - matéria orgânica, sedimento e condições termoquímicas apropriadas - é fundamental para o início da cadeia de processos que leva à formação do petróleo. O tipo de matéria orgânica irá determinar por exemplo o tipo de hidrocarboneto gerado, óleo ou gás. Por exemplo, matéria orgânica proveniente de vegetais lenhosos poderá dar origem à gás enquanto que matérias orgânicas provenientes de



fitoplâncton submetidas a condições térmicas adequadas, poderá vir a gerar a hidrocarboneto líquido.

### 2.1.2 Migração do Petróleo

Uma vez que o processo de geração de petróleo tenha ocorrido dentro das condições ideais para a formação hidrocarbonetos líquidos ou gasosos, para que um reservatório seja formado e possa ser extraído é necessário ainda que este óleo ou gás migre para uma rocha com condições de armazenar este hidrocarboneto. Estas rochas são conhecidas como rochas reservatório ou rochas armazém (Figura 2.2). Os hidrocarbonetos tendem a migrar sempre para cima e irão mover-se nestas rochas até encontrar outra camada rochosa impermeável que interrompa esse fluxo. Estas rochas impermeáveis são frequentemente denominadas de rocha cobertura e dependendo de sua formação é dito que criou-se uma formação geológica conhecida como trapa(armadilha).



**Figura 2.2.** Formação de petróleo. Na figura podemos ver o processo migratório da rocha mãe ou rocha geradora até a rocha reservatório.

## 2.2 Prospecção de Petróleo

A descoberta de uma jazida de petróleo é um processo de um longo estudo por parte de geólogos e geofísicos. Hoje, várias ferramentas da Tecnologia da Informação auxiliam essa tarefa. Graças ao acúmulo de conhecimento por parte dos geólogos, dos geofísicos e a estas ferramentas cada vez mais sofisticadas, a taxa de acerto para perfuração de poços de petróleo pode chegar a 100% como no caso da Bacia de Santos<sup>1</sup>. É preciso explicar que esse é um caso excepcional do pré-sal. A camada de pré-sal é uma camada muito extensa e de baixíssima permeabilidade formando uma grande cobertura que impediu que óleo gerado a milhares de metros, fugisse e portanto ficou preso por milhares de anos (Figura 2.3). Essas reservas eram consideradas inatingíveis e de alto risco. Essa percepção mudou graças à evolução tecnológica e também do alto valor do petróleo que compensa, enfim, os custos de extração.

Achar reservas como essas, parte de um estudo conhecido como prospecção de petróleo que tem como objetivos fundamentais:

- Localizar dentro de uma bacia sedimentar as situações geológicas que tenham condição para acumulação de petróleo;
- Verificar qual, dentre estas situações, possui mais chance de conter petróleo.

Esse estudo é feito através de métodos sísmicos e geológicos explicados nas próximas seções. São investimentos que se tornam pequenos quando comparados ao alto custo de perfuração de um poço, justificando os mesmos.

## 2.3 Métodos geológicos

Ao se estudar se existem condições de formação de uma reserva em uma determinada região, os métodos geológicos constituem-se do primeiro passo para confirmar a existência da reserva de óleo ou gás. O geólogo irá elaborar um mapa de geologia de superfície. Este mapa é feito através de rochas que podem ser observadas na superfície. Os mapas geológicos, que indicam as áreas potencialmente interessantes são continuamente construídos e atualizados pelos exploracionistas. Nestes mapas, as áreas compostas por rochas ígneas e metamórficas são praticamente eliminadas, como também pequenas bacias com espessura sedimentar muito reduzida ou sem estruturas favoráveis à acumulação [Thomas et al., 2001].

---

<sup>1</sup><http://fatosedados.blogspot.com.br/historico-monitor-fatos-e-dados/destaques-internacionais-16/>

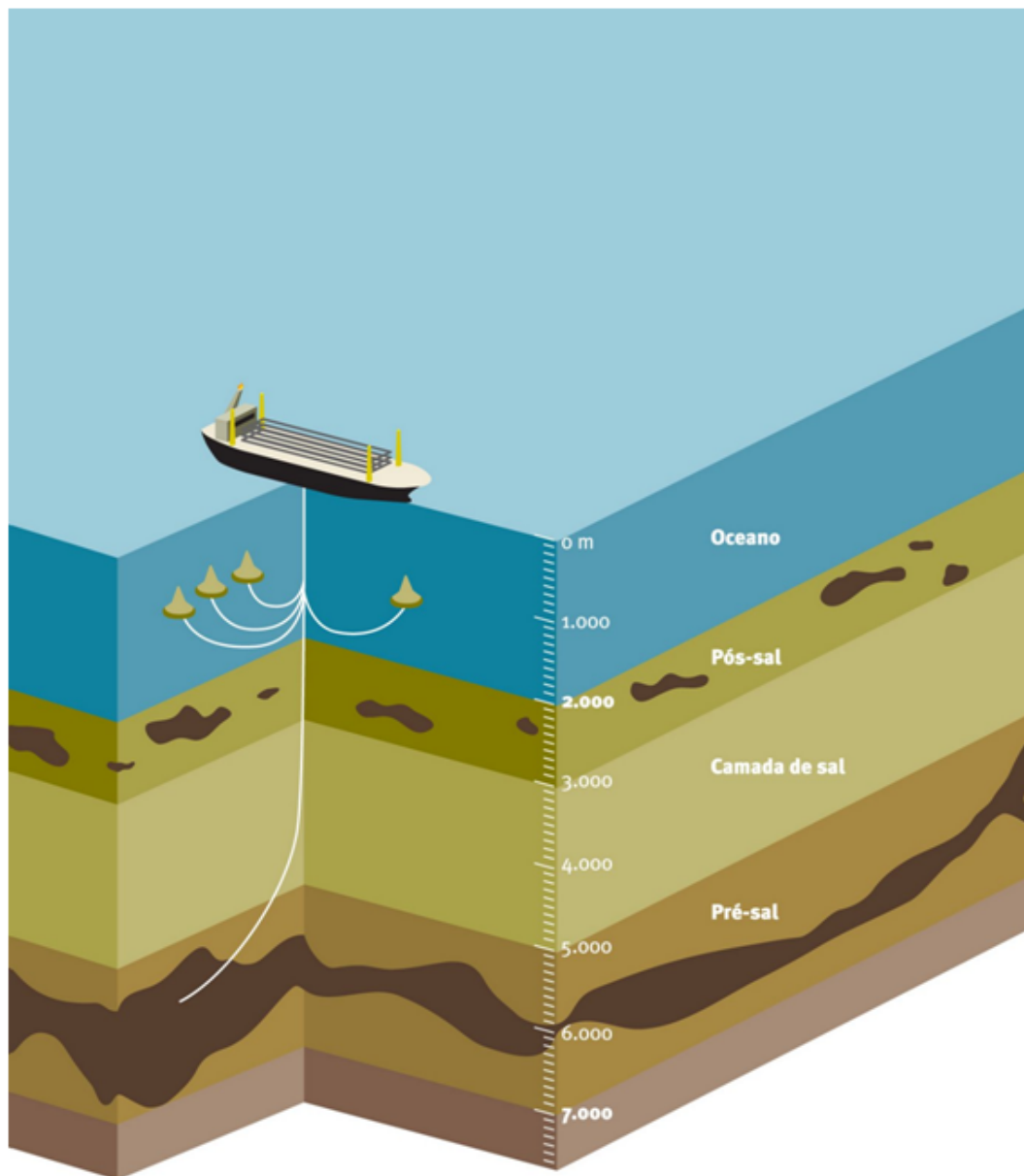


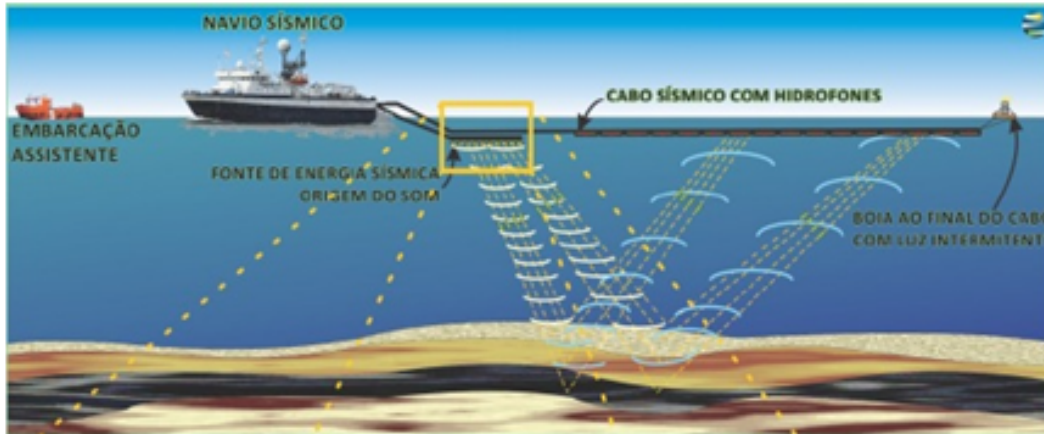
Figura 2.3. Pré-Sal.

## 2.4 Métodos Sísmicos

Os métodos geológicos mais utilizados para estudar as condições de formação de uma reserva de petróleo são os métodos sísmicos de reflexão. Esses métodos fornecem alta definição das feições geológicas em subsuperfície propícias à acumulação de hidrocarboneto, a um custo relativamente baixo.

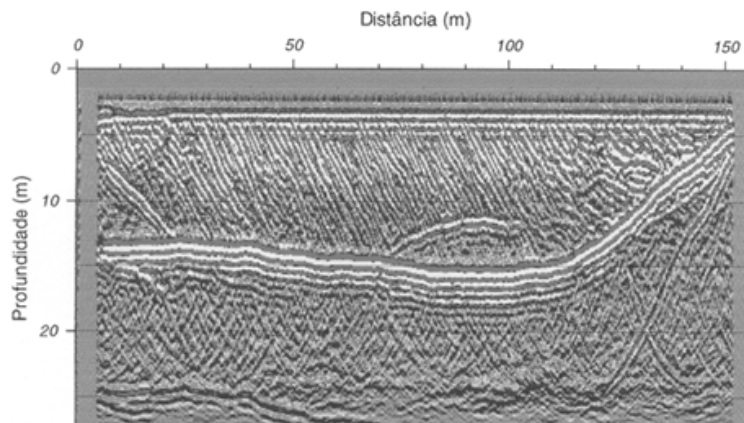
O levantamento sísmico inicia-se com geração de ondas através de fontes artificiais como dinamites e vibradores em terra e canhões de ar comprimido no levantamento marí-

timo (Figura 2.4). Cada uma dessas fontes irá emitir um sinal que irá refletir nas camadas geológicas, cujo retorno irá variar de acordo com cada formação. Esses sinais trazem informações valiosas que serão captadas pelos geofones quando em terra e pelos hidrofones quando em água.



**Figura 2.4.** Sísmica marítima utilizando canhões de pressão.

O resultado desse processo será uma seção sísmica (Figura 2.5). Ela é composta por amostras dos valores das amplitudes, espaçadas lateralmente ou em função da distância entre estações receptoras ou em função da distância entre cabos, e verticalmente pela razão da amostragem temporal [Thomas et al., 2001].



**Figura 2.5.** Resultado da seção sísmica. Cabe ao geofísico analisar a imagem interpretando as formações.

O próximo passo será processar estes dados através de aplicativos específicos para a indústria do petróleo. O objetivo é eliminar ruídos e produzir imagens da subsuperfície de maneira que facilite a interpretação por geólogos e geofísicos. Estes profissionais interpretam estas imagens procurando situações que favoreçam a acumulação de hidrocarbonetos.

Assim, diante destes dados, cabe a estes profissionais a responsabilidade de uma das decisões mais cruciais de uma empresa de petróleo: a perfuração de um poço.

## 2.5 Perfuração

A perfuração de um poço é a fase mais cara do processo de exploração de petróleo. Todo o estudo descrito anteriormente tem por objetivo que a perfuração seja bem sucedida e que se ache reservatório de óleo ou gás que possam ser comercializados. A perfuração de um poço de petróleo é realizada através de uma sonda. Na perfuração rotativa, as rochas são atravessadas pela ação da rotação e peso aplicados a uma broca existente na extremidade de uma coluna de perfuração a qual consiste basicamente de comandos (tubos de paredes espessas) e tubos de perfuração (tubos de paredes finas). Os fragmentos das rochas são removidos continuamente através de um fluido de perfuração(Figura 2.6) ou simplesmente lama, uma vez que sua base principal é a argila[Thomas et al., 2001].



**Figura 2.6.** Exemplo de fluido de perfuração, o qual entre outras funções, ajuda a remover os fragmentos das rochas resultantes da perfuração.

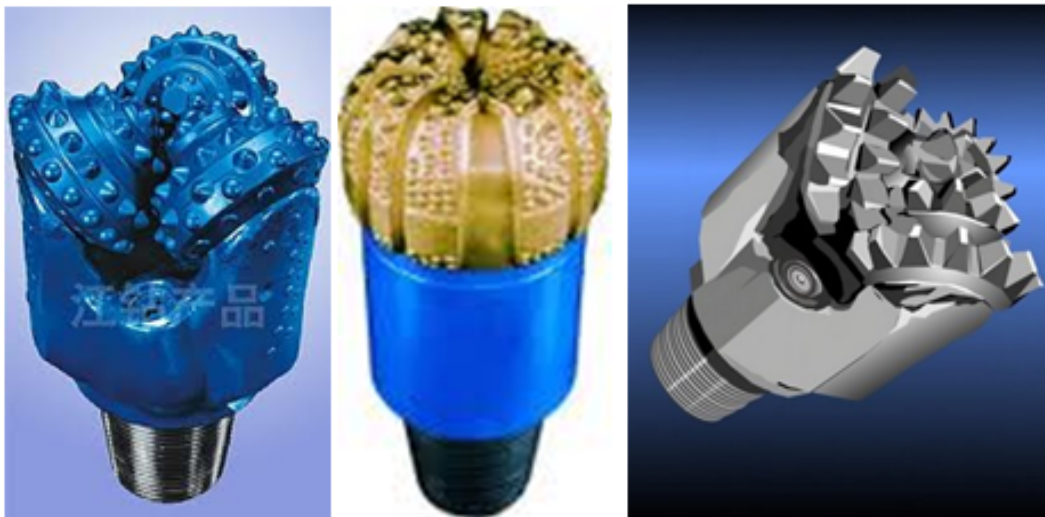
Esses fluidos têm como função:

- Manter as pressões de formação sob controle;
- Carrear os cascalhos até a superfície;
- Manter a estabilidade mecânica do poço;
- Resfriar a broca;



- Transmitir força hidráulica até a broca;
- Manter os cascalhos em suspensão quando sem circulação.

O espaço entre os tubos de revestimento e as paredes do poço (anular) é cimentado com a finalidade de isolar as rochas atravessadas, permitindo então o avanço da perfuração com segurança. Sem esses cuidados, pode acontecer o desabamento da parede do poço. Poderíamos nos perguntar se antes da cimentação esse desabamento não ocorre. A resposta é sim, poderia acontecer, no entanto, como citado anteriormente, um dos objetivos do fluido é manter as pressões de formação sobre controle apenas provisoriamente claro, até que a cimentação ocorra. Após as operações de cimentação, a coluna de perfuração é novamente descida ao poço, tendo na sua extremidade uma nova broca de diâmetro menor do que a do revestimento para o prosseguimento da perfuração. Assim, um poço de petróleo é perfurado em diversas fases, caracterizadas pelos diferentes diâmetros das brocas(Figura 2.7).

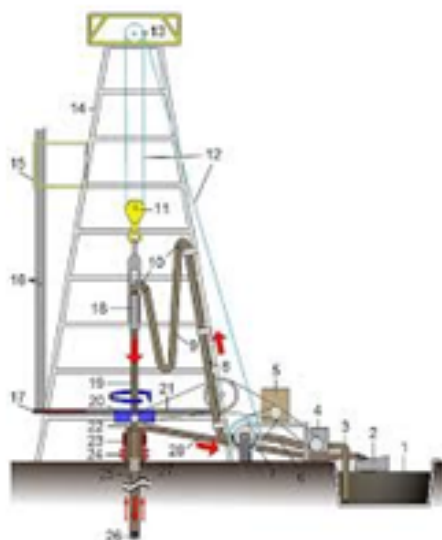


**Figura 2.7.** Diferentes tipos de broca. Cada uma é adequada a uma fase de perfuração e ao tipo de rocha que está sendo perfurado.

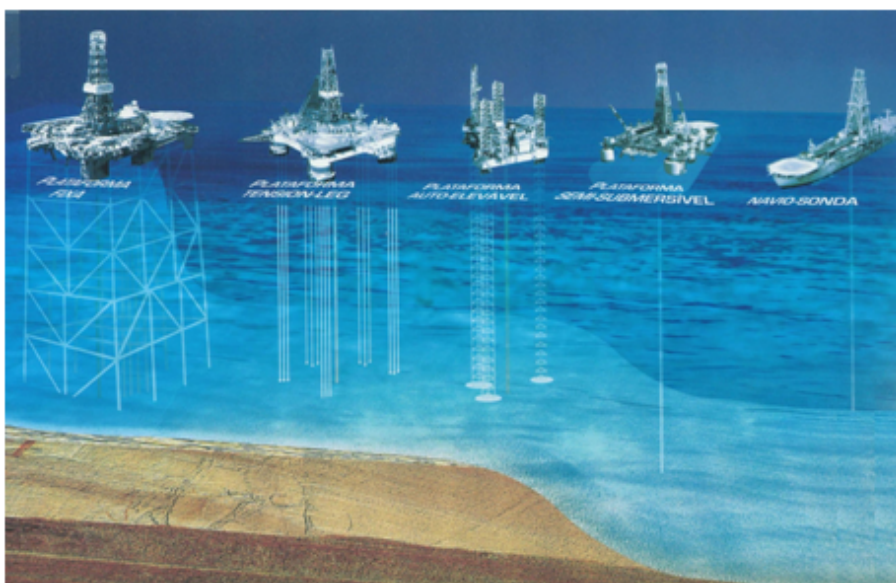
Para iniciar um poço de petróleo, necessita-se selecionar um tipo de sonda (Figura 2.8) ou plataforma compatível com as características gerais da região onde será perfurado o poço. Os serviços realizados nas áreas marítimas (Figura 2.9) são denominados (Off Shore). Já os serviços realizados nas áreas terrestres são denominados (On Shore).

## 2.6 Avaliação de Formações

A avaliação de formações envolve atividades e estudos que visam definir em termos qualitativos e quantitativos o potencial de uma jazida petrolífera, isto é, sua capacidade produtiva e



**Figura 2.8.** Sonda terrestre para perfurações em terra.



**Figura 2.9.** Diversos tipos de sonda marítima para perfurações no mar. A complexidade aumenta com a profundidade a ser atingida.

a valoração de suas reservas de óleo e gás [Thomas et al., 2001]. Essa avaliação é baseada principalmente nos dados coletados na perfilagem de um poço aberto, que segundo [Thomas et al., 2001] é uma imagem visual (Figura 2.10), em relação à profundidade, de uma ou mais características ou propriedades das rochas perfuradas tais como resistividade elétrica, potencial eletroquímico natural, tempo de trânsito das ondas mecânicas, radioatividade natural ou induzida. Ou ainda, segundo An-nan & Lu [2009], perfilagem é o método de identificar propriedades das rochas e fluidos medindo informações elétricas, acústicas, térmicas e

radioativas. Portanto, com a perfilagem é possível extrair dados importantes das formações pelas quais o poço atravessa como a litologia (tipo de rocha), espessura, porosidade, presença de óleo, gás ou água. Porém, a perfilagem sozinha não é suficiente para a análise. São necessárias informações auxiliares como as amostras de calhas que são resíduos das rochas trazidas à superfície e que ajudam na identificação das litofácies. Dependendo da análise dos perfis e das informações complementares, pode se decidir pelo abandono do poço ou no caso de se provar interesse econômico, os testes de formação para se estimar a capacidade produtiva do poço são iniciados.

### 2.6.1 Testemunhagem

Uma das técnicas utilizadas para extrair características do reservatório é a testemunhagem, que é uma das mais antigas e ainda praticadas. Entretanto, testemunhar todos os poços em um campo muito grande pode ser economicamente inviável, além disso, o tempo consumido pode ser muito grande.

Segundo [Thomas et al., 2001], a testemunhagem é o processo de obtenção de uma amostra real de rocha de subsuperfície, chamado testemunho (ver Figura 2.11), com alterações mínimas nas propriedades naturais da rocha. Com a análise deste testemunho obtém-se informações importantes sobre geologia, engenharia de reservatórios, completação e perfuração, tais como litologia, textura, porosidade, permeabilidade, saturação de óleo e água.

O processo de obtenção de uma amostra de testemunho, como já foi dito, é caro e demorado. Geralmente ele é feito em poços chaves como o poço pioneiro. Ele servirá também para confirmar os dados obtidos pela perfilagem elétrica. Os testemunhos são um volume cilíndrico de rocha que mantém os constituintes em sua estrutura original, como em profundidade, permitindo estudar os meios porosos onde estão armazenados os hidrocarbonetos. A partir deles ainda se tira amostras de lâminas para análise microscópica de granulometria onde se estuda a forma, tamanho, espaçamento dos grãos. Irá se verificar por exemplo a permeabilidade e porosidade da rocha.

Existem três métodos de retirada de testemunho:

- Testemunhagem com barrilete convencional: é o processo mais caro pois interrompe a perfuração. É necessário que se retire toda a coluna de perfuração e se coloque uma broca oca própria (barrilete) para criar o cilindro típico da rocha de testemunho. Depois sobe-se novamente essa broca já com a amostra e pode-se retornar com a broca de perfuração;
- Testemunhagem a cabo: Uma forma de se reduzir os custos da testemunhagem é substituir o barrilete convencional, que é fixo à coluna de perfuração, por um barrilete



móvel preso a um cabo. Assim, o barrilete pode ser levantado até a superfície sem a necessidade de se retirar toda coluna;

- Amostra Lateral (Testemunhagem lateral): é utilizada para se testemunhar alguma formação já perfurada. É feita com cilindros ociosos, presos a um cabo de aço e um canhão e são arremessados contra a parede do poço para retirada de amostras de rocha.

### 2.6.2 Amostras de Calha

A amostra de calha é utilizada juntamente com os perfis elétricos para ajudar na identificação das litofácies. Essa amostra contém os resíduos das formações rochosas, durante a perfuração, que permanecem na broca quando ela é levantada do interior do poço para a superfície.

### 2.6.3 Perfilagem Elétrica

Após a perfuração de uma fase do poço, são descidas várias ferramentas (Figura 2.12) com a finalidade de medir algumas características físicas que irão permitir a caracterização e a avaliação comercial. Essas características são os perfis elétricos.

Com base na análise dos perfis é possível decidir quais intervalos de poços são de interesse econômico potencial para se executar os testes de formação. Se não houver intervalos de interesse, o poço é abandonado. Os testes de formação têm sido amplamente utilizados na indústria petrolífera para se estimar a capacidade produtiva do poço [Thomas et al., 2001]. Uma das mais importantes características que irá determinar o interesse de um intervalo é o tipo de rocha encontrado e suas características como porosidade e permeabilidade. Neste contexto porosidade diz respeito a pequenos espaços dentro das rochas, onde pode ser acumulado algum tipo de líquido ou gás, enquanto que permeabilidade é a capacidade que o conteúdo dos poros tem em se mover.

Na Figura 2.13 podemos ver alguns exemplos de rochas com boa porosidade e permeabilidade e também com porosidade ruim e permeabilidade ruim. Isso deve-se ao fato da existência de acumulações de petróleo depender das características e do arranjo de certos tipos de rochas sedimentares no subsolo. Conforme pode ser observado na Figura 2.14, basicamente, é preciso que existam rochas geradoras que contenham a matéria-prima que se transforma em petróleo e rochas reservatório, ou seja, aquelas que possuem espaços vazios, chamados poros, capazes de armazenar o petróleo. É necessária a ocorrência de rochas selantes conhecidas como trapas, com baixa permeabilidade que não deixam o petróleo escapar. A ausência de qualquer um desses elementos impossibilita a existência de uma acumulação petrolífera.

Outro fator de influência é a localização de cada uma dessas rochas. Fica claro portanto, a importância da identificação das litofácies (tipos de rocha) na análise de um poço. É preciso ressaltar que a Figura 2.14 retrata de maneira simplista um reservatório de petróleo. No Brasil, por exemplo, conforme relatos colhidos por geólogos da Petrobras, dificilmente iremos encontrar rochas geradoras (folhelhos) perto de rochas reservatórios, como por exemplo arenito. Na maioria dos casos, as rochas geradoras estão a muitos quilômetros e quantidades muito pequenas de óleo ou gás foram através de milhares de anos, se deslocando e se acumulando em um reservatório.

## 2.6.4 Fundamentos de Perfilagem Elétrica

A maioria dos perfis utiliza a resistividade como forma de avaliar o tipo sedimento ou fluido existente. A resistividade é simplesmente a capacidade, neste caso da rocha, em impedir ou deixar passar algum tipo de sinal. Então, podemos dizer que a intensidade de retorno desse sinal irá variar com o tipo rocha ou fluido existente. A porosidade da rocha poderá influenciar significativamente na variação do sinal uma vez que uma rocha mais porosa será capaz de conter mais água por exemplo, o que diminuiria a resistividade. Podemos dizer, portanto, que a resistividade de uma rocha  $R_0$  varia na razão direta da resistividade da água e inversa da porosidade. A resistividade de uma rocha completamente saturada é dada por:

$$R_0 = \frac{aR_w}{\phi^m}, \quad (2.1)$$

onde  $a$  é o coeficiente litológico,  $m$  é o coeficiente de cimentação e  $\phi$  o de porosidade. Lembrando que a cimentação como já falado, é a parede feita para revestir o poço impedindo que ele desmorone. No caso da rocha conter qualquer tipo de fluido, como óleo, água ou gás, sua resistividade irá variar consideravelmente.

### 2.6.4.1 Tipos de perfis

Os principais tipos de perfis elétricos são: Raios Gama (GR), Neutrônico (NPHI), Indução (ILD), Sônico (DT), Densidade (RHOB), Caliper. A descrição de cada um deles é mostrada na Tabela 2.2.

Pela análise da Tabela 2.2, a principal propriedade encontrada a partir dos perfis é a porosidade de uma rocha definida pela relação entre o volume de espaços de uma rocha e pelo seu volume total. É justamente dentro desses espaços que poderemos encontrar óleo, gás ou água. Outra característica importante das rochas é sua permeabilidade definida pela medida da capacidade de circulação de um fluido através de uma rocha, sem alterar a sua estrutura interna. Portanto, para uma rocha ter uma boa permeabilidade, é necessário que

**Tabela 2.2.** Principais Tipos de perfis elétricos e suas propriedades.

| Perfil                    | Propriedades   |
|---------------------------|--|
| Raios Gama (GR)           | Detecta a radioatividade total da formação geológica. Utilizado para a identificação da litologia, a identificação de minerais radioativos e para o cálculo do volume de argilas ou argilosidade.  |
| Neutrônico (NPHI)         | São utilizados para estimativas de porosidade, litologia e detecção de hidrocarbonetos leves ou gás.   |
| Indução (ILD)             | Fornece leitura aproximada da resistividade da rocha contendo hidrocarbonetos, através da medição de campos elétricos e magnéticos induzidos nas rochas.   |
| Sônico (DT)               | Mede a diferença nos tempos de trânsito de uma onda mecânica através das rochas - É utilizado para estimativas de porosidade, correlação poço a poço, estimativas do grau de compactação das rochas ou estimativas das constantes elásticas, detecção de fraturas e apoio à sísmica para a elaboração do sismograma sintético. |
| Densidade (RHOB)          | Além da densidade das camadas, permite o cálculo da porosidade e a identificação das zonas de gás.   |
| Caliper                   | Fornece o diâmetro do poço. É aplicado no cálculo do volume de cimento para tampões ou cimentação do revestimento, apoio a operações de teste de formação, controle de qualidade de perfis e indicações das condições do poço em um determinado intervalo.   |
| Potencial Espontâneo (SP) | É o registro da diferença de potencial entre um eletrodo móvel descido dentro do poço e outro fixo na superfície. Este perfil permite determinar as camadas permoporosas, calcular a argilosidade das rochas, determinar a resistividade da água da formação e auxiliar na correlação de informações com poços vizinhos.       |

seus poros não estejam isolados e nem sejam tão pequenos ao ponto de não permitirem que estes fluidos circulem dentro da mesma. No entanto, a principal característica que pode ser identificada a partir dos perfis elétricos são as litofácies que atravessam o poço.

A análise dos perfis é sempre feita em conjunto para evitar erros de interpretação. Mudanças de valores significantes são sempre confirmadas em outro perfil. Por exemplo, a leitura de porosidade de uma rocha pelo perfil Sônico (DT) varia normalmente em torno de 9 a 27, sendo 27 uma rocha muito porosa com espaços internos onde poderia por exemplo existir óleo. Valores maiores que 27, poderiam indicar uma ruptura do poço e seria interessante se checar o perfil Caliper. O Caliper é conhecido como perfil que mede a qualidade da perfilagem feita em um poço. Isto acontece porque ao se perfurar um poço, há a preocupação de que as paredes do mesmo não desabem. Para que isto não aconteça, a medida que o poço é perfurado, é injetado um fluido com densidade própria para que sustente a parede de um poço antes que se possa fazer a cimentação do mesmo. Caso a parede do poço desabe em algum ponto, irá criar uma zona de arrombamento comprometendo a leitura de outros perfis do poço. Perfis por exemplo que lêem a porosidade de uma rocha, ao realizar a leitura em uma zona onde existe um buraco provocado pela queda da parede do poço naquele local, irá indicar valores altos de porosidade, mas que na verdade indicam um buraco provocado por desabamento ou até mesmo uma caverna. O perfil Caliper irá confirmar essa informação.

A análise destes perfis é de responsabilidade do geofísico, que é o profissional que estuda as estruturas e os processos dinâmicos da Terra, além de analisar as forças físicas que atuam no planeta. Este trabalho investiga a classificação de litofácies via técnicas de reconhecimento de padrões. O próximo capítulo descreve a fundamentação teórica necessária à compreensão deste estudo, bem como alguns trabalhos relacionados.

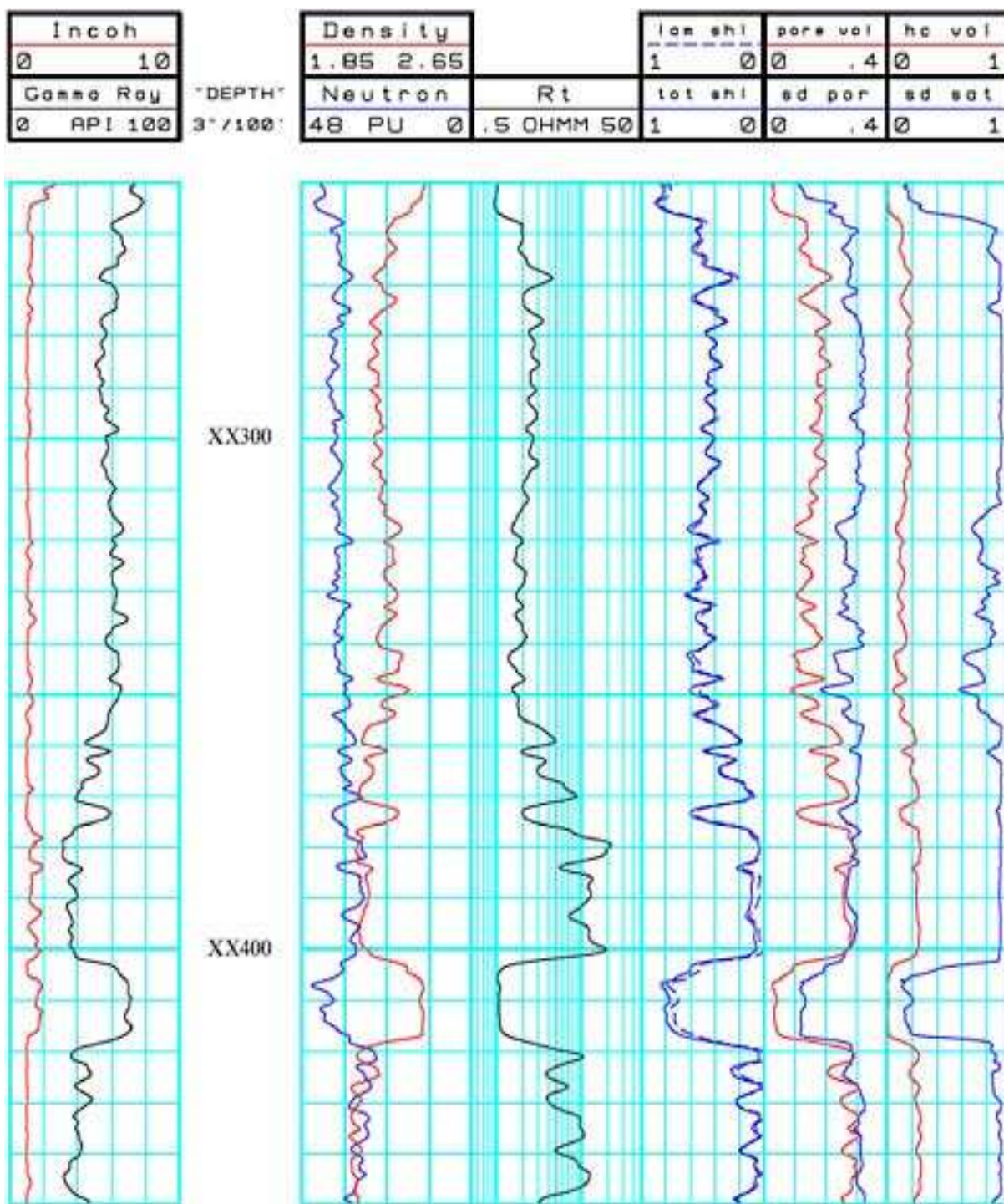


Figura 2.10. Perfis elétricos coletados durante a perfilagem de um poço.



**Figura 2.11.** Amostras de testemunho retirados de um poço. Apesar de ser a melhor representação do perfil de um poço, é um processo longo e caro, sendo feito portanto somente em alguns poços estratégicos.



(a) Ferramentas de perfilagem sendo descidas no poço.



(b) Exemplos de ferramentas de perfilagem.

**Figura 2.12.** Ferramentas de perfilagem utilizadas para extração das características de um poço.

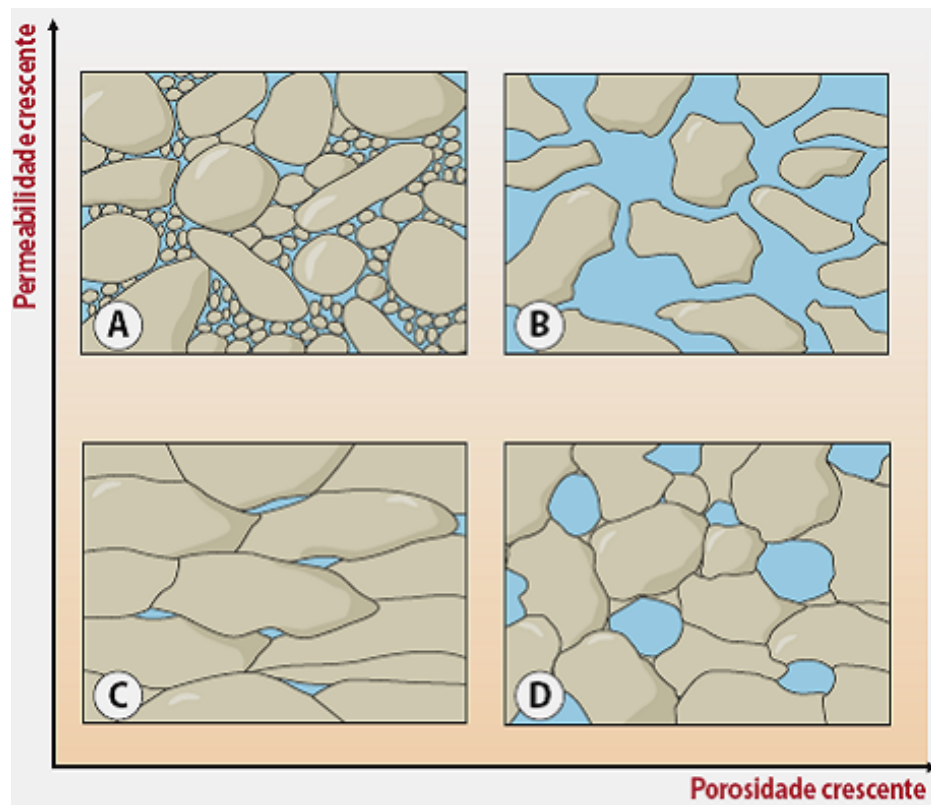


Figura 2.13. Porosidade e permeabilidade.

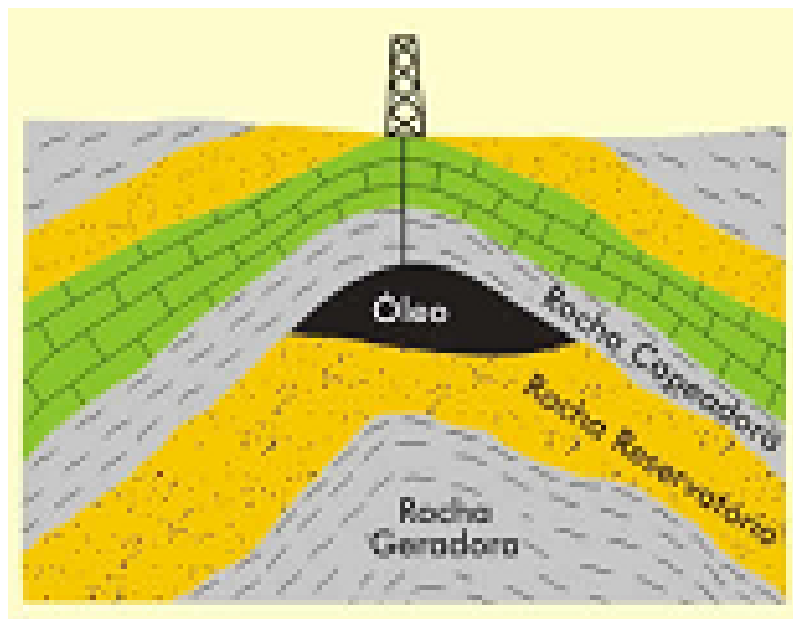


Figura 2.14. Acumulações de Petróleo.



## Capítulo 3

# Fundamentação Teórica e Trabalhos Relacionados

Conforme mencionado na introdução, neste trabalho serão utilizadas ferramentas de Reconhecimento de Padrões para a classificação automática de litofácies. Essas ferramentas são descritas em poucos detalhes na próxima seção. Em seguida, alguns trabalhos publicados na literatura relacionada ao objeto desta dissertação também são descritos e analisados.

### 3.1 Reconhecimento de Padrões

O reconhecimento automático de padrões é normalmente dividido em três etapas [K.Jain et al., 2000]: filtragem da entrada, extração/seleção de características e classificação. Na primeira etapa, dados irrelevantes, redundantes ou distorcidos, são eliminados. A segunda etapa, por sua vez, consiste na análise das características de entrada a fim de extrair e/ou selecionar características úteis para o processo de reconhecimento [Wang et al., 2003]. Por fim, na etapa de classificação, o objeto em análise é atribuído à uma das classes envolvidas no problema.

Nesta última etapa, métodos de classificação são investigados e comparados para que o melhor classificador para o problema seja identificado. São muitos os exemplos de classificadores que existem na literatura. Alguns dos mais utilizados são: classificadores estatísticos, redes neurais, árvores de decisão, SVM, kNN, entre outros.

Conforme mencionado anteriormente, cinco métodos de classificação são investigados neste trabalho: SVM, kNN, redes neurais do tipo Multilayer Perceptron, Árvores do tipo Tree(J48) e Regressão Logística. Além de serem comparados individualmente, esses quatro classificadores são combinados em um conjunto de classificadores heterogêneos por meio de voto majoritário. Portanto, as próximas seções descrevem resumidamente o funcionamento

dos cinco métodos de classificação investigados, bem como o método de combinação de classificadores.

## 3.2 Máquinas de Vetores de Suporte - SVM

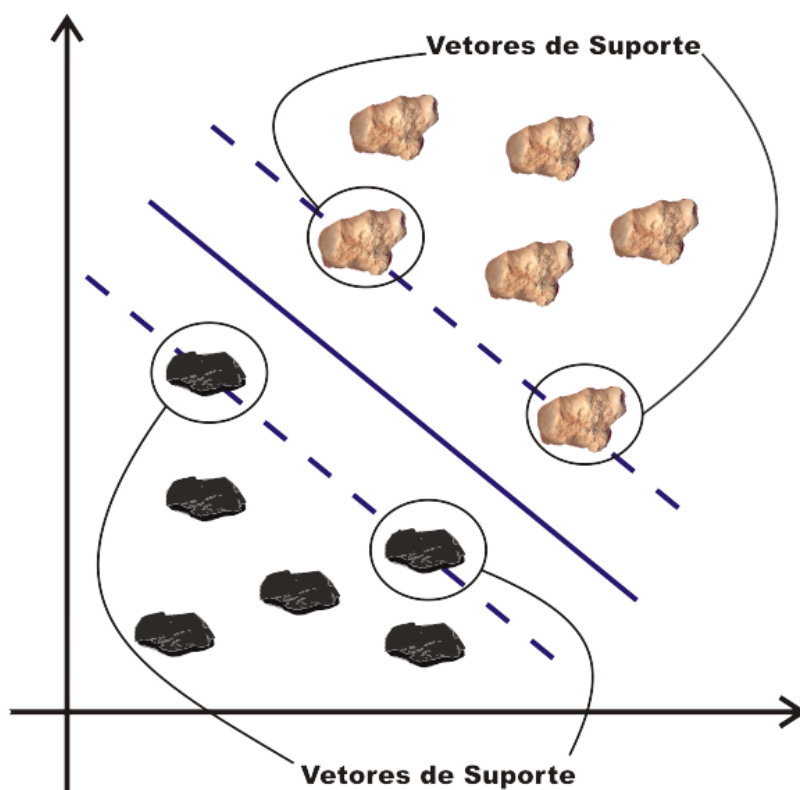
SVM é um tipo de classificador baseado na teoria da aprendizagem estatística proposto por Vapnik [1999] e outros pesquisadores. A teoria de Vapnik é baseada em uma forte fundamentação matemática para estimação de dependências e prever o aprendizado a partir de conjuntos de dados finitos. O objetivo do SVM é minimizar tanto o erro de treinamento quanto a complexidade do modelo gerado durante o treinamento.

Basicamente, o funcionamento de SVM pode ser descrito da seguinte forma: dadas duas classes e um conjunto de instâncias de treinamento cujas amostras pertencem a essas classes, SVM constrói um hiperplano que divide o espaço de características em duas regiões, maximizando a margem de separação entre as mesmas. Esse hiperplano é conhecido como hiperplano de separação ótima. As amostras desconhecidas (exemplos de teste) são então mapeadas para esse mesmo espaço, e atribuídas a uma das classes [Alpaydim, 2010]. A Figura 3.1 ilustra o funcionamento de SVM. Dadas duas classes de rochas (calcaranito e folhelho), SVM define o hiperplano de separação ótima (reta não pontilhada) que maximiza a margem entre as duas classes (linhas pontilhadas). As amostras que encontram-se na fronteira (destacadas com círculos) são chamadas vetores de suporte e constituem os dados mais importantes para a tomada de decisão de SVM.

Em certas situações, amostras conhecidas como *outliers*, ou seja, que fogem muito ao padrão das demais amostras, podem afetar o hiperplano ótimo. O parâmetro  $C$  é um parâmetro de regularização, cujo valor positivo é escolhido pelo usuário atuando como uma função de penalidade, prevenindo que *outliers* afetem o hiperplano ótimo. Quanto maior o valor de  $C$ , menor a tolerância à violação das restrições de separação. Para achar o melhor valor de  $C$ , recomenda-se através de validação cruzada, variar de 1 até 1000 até encontrar o melhor valor.

## 3.3 k-Vizinhos mais Próximos - kNN

A idéia de vizinho mais próximo (kNN) foi primeiro introduzida por Cover & Hart [1967], no qual a regra de decisão é designar uma amostra ainda não classificada para a mesma classificação do conjunto de pontos mais próximos já classificados. Os autores provaram que quando a quantidade de dados se aproxima do infinito, a classificação baseada nos vizinhos



**Figura 3.1.** Ilustração sobre o funcionamento de SVM aplicado à classificação de rochas sedimentares.

mais próximos é delimitada por duas vezes a taxa de erro assintótica como a regra de Bayes, independentemente da distância métrica aplicada.

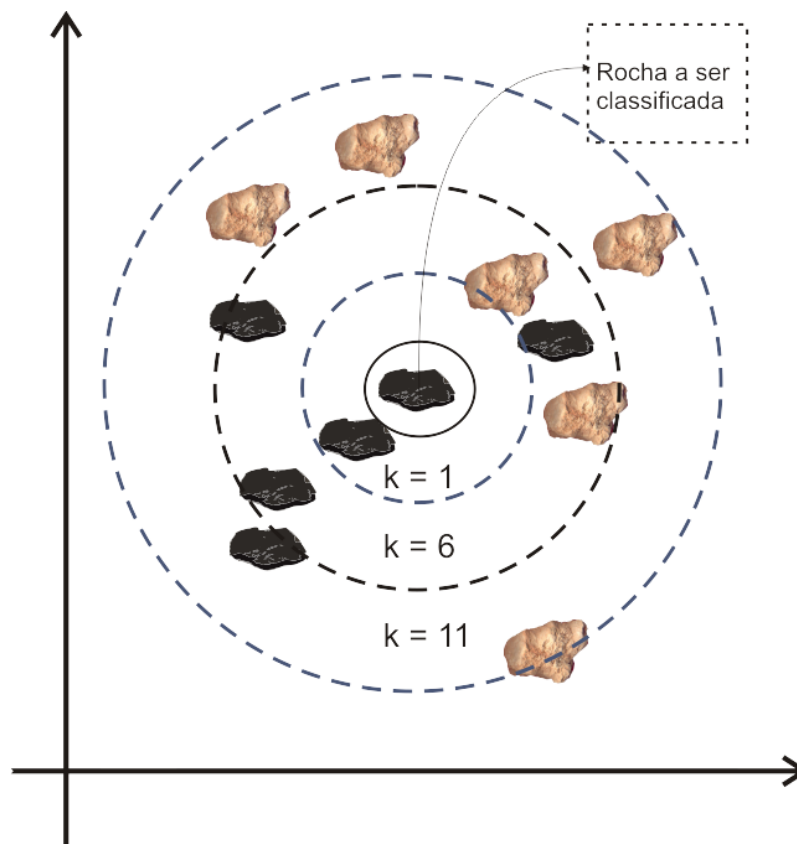
Em outras palavras, kNN é uma técnica de classificação baseada em instâncias, isto é, consiste em atribuir a classe de cada elemento desconhecido (amostras de teste) a partir da classe majoritária obtida entre os seus vizinhos mais próximos identificados no conjunto de treinamento [Henke et al., 2011]. A definição de vizinhança é feita segundo uma medida de similaridade que normalmente é uma medida de distância calculada no espaço de características. Uma das formas mais utilizadas para medir a distância é calculada pela distância a Euclidiana que é descrita pela seguinte fórmula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (3.1)$$

Sendo  $x$  e  $y$  duas amostras, representadas por características  $x_i$  e  $y_i$ , respectivamente.

Uma maneira de ilustrar o método é proposto na Figura 3.2. A instância mais ao centro da figura é utilizada para demonstrar como novas instâncias são classificadas, trata-se de uma instância da classe folhelho. Na figura, quando  $k=1$ , a instância será classificada

corretamente. Aumentando o valor de  $k$  para 6, novamente a instância será classificada corretamente, pois, dentre suas seis instâncias mais próximas, quatro instâncias são da classe folhelho e apenas duas são da classe anidrita, isto é, folhelho é a classe majoritária. Por fim, se o valor atribuído a  $k$  for 11, a instância será classificada incorretamente, pois, a classe majoritária (seis votos) será a classe anidrita. Portanto, o classificador  $k$ NN precisa do ajuste do parâmetro  $k$ , que define a quantidade de vizinhos que contribuirão para a definição das classes das amostras desconhecidas. A escolha do valor de  $k$  é feita repetindo o experimentos com diversos valores para o mesmo de 1 até um número que não haja mais melhora na taxa de acerto ou que a mesma comece a decrescer.

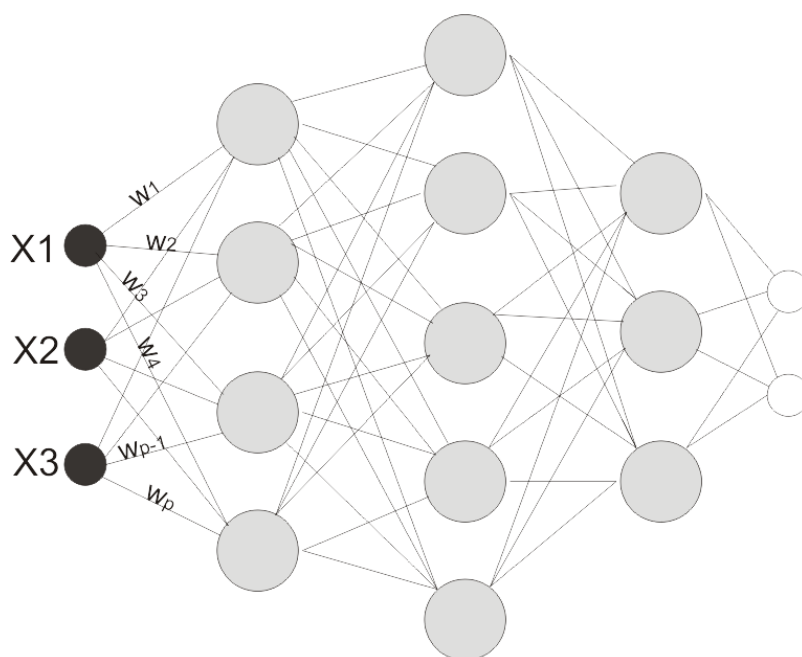


**Figura 3.2.** Exemplo ilustrativo do funcionamento de  $k$ NN, bem como o processo de ajuste do parâmetro  $k$ .

### 3.4 Redes Neurais Artificiais (RNA) - Multilayer Perceptron

*Multilayer Perceptron* é uma rede neural com as seguintes características principais:

- Múltiplas entradas, múltiplas saídas e múltiplas camadas;
- Entradas e saídas podem ser tanto analógicas quanto digitais;
- Não há restrição de separabilidade entre classes.



**Figura 3.3.** Exemplo ilustrativo do funcionamento de Multilayer Perceptron.

Redes Neurais Artificiais (RNA) são técnicas computacionais que propõem um modelo matemático que simula a rede de neurônios de um cérebro humano. Uma RNA procura por relacionamentos, constrói modelos automaticamente e os corrige de modo a diminuir seu próprio erro [Tafner et al., 1995].

Segundo Cunha [2002], o neurônio artificial foi projetado para imitar as características de um neurônio biológico. O neurônio artificial possui várias entradas ( $X_1, X_2, \dots, X_p$ ), que podem ser estímulos do sistema ou saídas de outros neurônios. Cada entrada é multiplicada por um peso correspondente ( $W_1, W_2, \dots, W_p$ ), gerando entradas ponderadas. Logo após, todas as entradas ponderadas são somadas e o valor resultante da soma será comparado com um valor limite para ativação do neurônio (função de ativação). Caso o valor da soma alcance o valor limite para ativação do neurônio, ele se ativará, caso contrário, ficará inativo. A saída no neurônio ( $Y$ ) é o valor processado pela função de ativação.

Dentre os principais parâmetros que devem ser ajustados durante o treinamento de uma RNA do tipo Multilayer Perceptron, destacam-se: quantidade de camadas escondidas, quantidade de neurônios em cada camada escondida, tipo de função de ativação, etc.

### 3.5 Regressão Logística

Regressão Logística em estatística é um tipo de análise de regressão usada para prever o resultado de uma variável categórica dependente, com base em uma ou mais variáveis de precisão. As probabilidades que descrevem o possível resultado de um único ensaio são modeladas como uma função de variáveis explanatórias, utilizando uma função logística.

### 3.6 Árvores de Decisão - J48

Outro classificador que testamos nesse trabalho foi o algoritmo J48 que permite a criação de modelos de decisão em árvore. Seu algoritmo utiliza a técnica gulosa para induzir árvores de decisão para posterior classificação. O modelo de árvore de decisão é construído pela análise dos dados de treino e o modelo utilizado para classificar dados ainda não classificados. O J48 gera árvores de decisão, em que cada nó da árvore avalia a existência ou significância de cada atributo individual. As árvores de decisão são construídas do topo para a base, através da escolha do atributo mais apropriado para cada situação. Uma vez escolhido o atributo, os dados de treino são divididos em sub-grupos, correspondendo aos diferentes valores dos atributos e o processo é repetido para cada sub-grupo até que uma grande parte dos atributos em cada sub-grupo pertençam a uma única classe. A indução por árvore de decisão é um algoritmo que habitualmente aprende um conjunto de regras com elevada acuidade.

### 3.7 Conjunto de Classificadores

O objetivo de combinar classificadores é aumentar a taxa de acerto em relação a classificadores utilizados individualmente. Até mesmo juntar um único tipo de classificador com parâmetros diferentes pode melhorar melhores taxas como descrito por Altınçay [2007] onde o autor demonstra que um conjunto de classificadores kNN obtém melhores taxas quando comparado ao mesmo classificador sozinho. Outra motivação para se utilizar conjuntos de classificadores é dada por Santos [2008]: conjuntos de classificadores tentam superar um classificador robusto e bem parametrizado, por classificadores relativamente simples e combinados. A vantagem citada pela autora vai além de se obter melhores taxas de acerto. Simplificar a parametrização de classificadores complexos como o SVM é um grande ganho. Podemos citar, por exemplo, que os melhores parâmetros de um classificador, identificados em um reservatório, podem ser diferentes em outros reservatórios, e nem sempre geólogos especializados em perfis terão familiaridade com a parametrização de classificadores.

### 3.7.1 Votação Majoritária

Conjuntos de classificadores podem ser gerados de diversas formas, tais como: manipulando os dados de treinamento, os atributos dos dados, as classes do problema e os membros do conjunto. Esta última estratégia possibilita a criação de conjuntos de classificadores heterogêneos, por meio do uso de diferentes métodos de classificação, diferentes arquiteturas ou parâmetros de um mesmo classificador. Neste trabalho, cinco tipos diferentes de métodos de classificação são combinados em um conjunto de classificadores heterogêneos, por meio de voto majoritário, conforme descrito a seguir.

Existem várias estratégias para se agrupar classificadores em conjunto. A mais simples e mais popular é a da votação majoritária [Santos, 2008]. Nesse método, escolhe-se a classe com maior número de votos dentre os classificadores membros. Seu funcionamento é ilustrado na Figura 4.6. Os classificadores que compõem o conjunto investigado neste trabalho são:

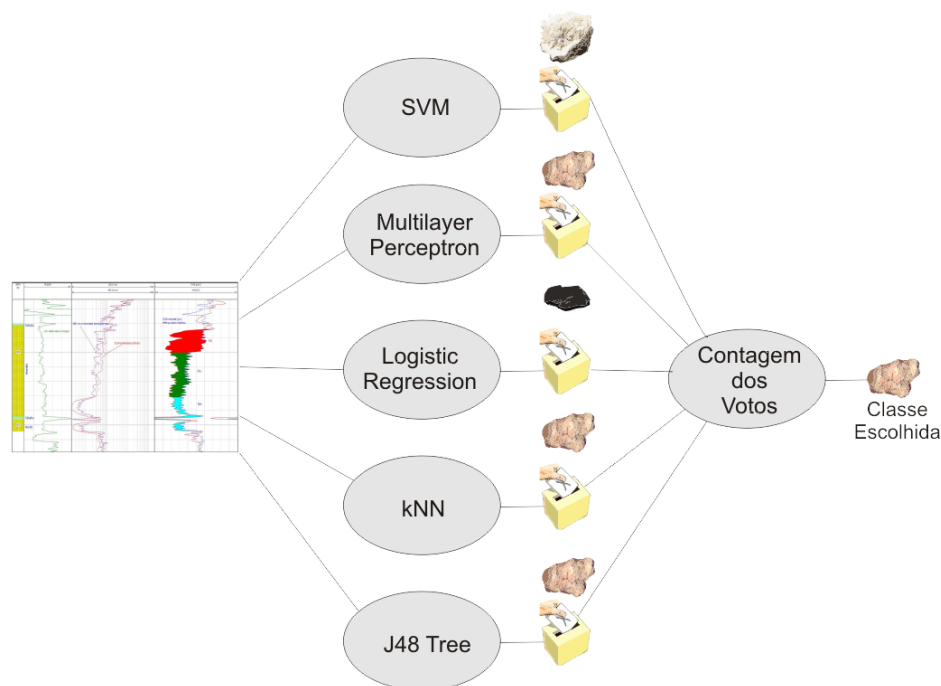
1. SVM;
2. Multilayer Perceptron;
3. Regressão Logística;
4. kNN;
5. Tree (J48).

A expectativa é que elevadas taxas de classificação sejam obtidas com a automatização do processo de classificação. Isso traria uma economia enorme de tempo e subsídios preciosos para justificar um projeto de exploração de um poço de petróleo ou ainda o cancelamento do mesmo.

Algumas tentativas de automatização desse processo têm sido publicadas na literatura. Na próxima seção são discutidos alguns desses trabalhos.

## 3.8 Trabalhos Relacionados

Cunha [2002] utiliza além dos perfis, testemunhos (Figura 2.11) de poços aplicando redes neurais para a classificação de litofácies. Em seu trabalho, Cunha utilizou uma base de 8 poços de petróleo e atingiu uma taxa de acerto de 80% e relata que uma solução para o problema só foi possível após a incorporação de uma estratégia para agrupamento prévio das litofácies e tratamento de padrões problemáticos (regiões de conhecimento incerto nos conjuntos de treinamento e de teste). A autora relata que a principal diferença de seu trabalho



**Figura 3.4.** Ilustração dos cinco Classificadores em conjunto por meio do Voto majoritário.

em relação a trabalhos prévios utilizando redes neurais está na maior quantidade de dados. No treinamento foi utilizado um simulador de Redes Neurais, o SNNS (*Stuttgart Neural Network Simulator*) e a arquitetura utilizada nos experimentos foi Multilayer Perceptron com algoritmo de treinamento Backpropagation. Nesse trabalho, apesar de terem sido utilizados dados de 08 poços, todos pertenciam à mesma reserva e, portanto, não há como avaliar o comportamento do classificador treinado em um poço e testado em outro, porque esse aspecto não foi investigado pela autora.

Flexa et al. [2004] propuseram uma metodologia baseada na técnica da análise discriminante para a identificação de litologias nos perfis de poços de duas reservas. Os autores pesquisaram a associação entre as variações das propriedades físicas e os tipos litológicos em profundidade. Os perfis geofísicos utilizados nesse trabalho foram raios gama (GR), sônico (DT), porosidade neutrônica (PHIN), densidade (RHOB) e resistividade (ILD). Segundo Flexa et al. [2004], a análise discriminante é uma técnica da estatística multivariada com grande eficiência na solução de problemas geofísicos. A técnica é utilizada nos casos em que se deseja realizar a separação estatística entre dois ou mais grupos de dados, que devem ser definidos previamente pela caracterização de suas variáveis discriminantes. Entretanto, no trabalho de Flexa et al. [2004], a técnica de análise discriminante não é comparada com outros classificadores. Os autores consideram que bons resultados foram obtidos, porém o desempenho real não foi apresentado. Além disso, não foi investigado o uso de conhecimento



obtido em um poço, para classificar amostras de outros poços.

An-nan & Lu [2009] propuseram um trabalho de identificação de litofácies a partir de dados de perfilagem utilizando DE-SVM com uma taxa de acerto de 90%. DE é o acrônimo de diferença evolutiva, um algoritmo baseado em algoritmos genéticos. Os autores observam que SVM tem sido usado com sucesso para identificação de litologias, mesmo com uma amostra pequena de dados, mas que no entanto, SVM não fornece um método de seleção para o melhor parâmetro de kernel  $\sigma$  e o fator de penalização C, que podem afetar a precisão de aprendizagem e previsão. O artigo apresentou um método de otimização baseado na diferença aritmética-evolutiva (DE) para selecionar os parâmetros acima, chegando assim ao melhor kernel e parâmetro de penalização C para o problema de identificação de litologia. O melhor valor encontrado para C, foi de 263.2 e para o parâmetro de kernel  $\sigma$  foi 11.5. Porém, a base utilizada era composta por dados de três poços adjacentes, não sendo possível tirar conclusões sobre a eficiência do método para poços diferentes. Além disso, como o objetivo principal dos autores foi desenvolver um método para encontrar os melhores parâmetros para SVM, nesse trabalho não houve comparações entre classificadores como kNN e Redes Neurais.

Wong et al. [2005] investigaram o desempenho do classificador SVM para a caracterização de reservatórios como permeabilidade e porosidade. Estas características são fundamentais para a localização de hidrocarbonetos, pois conforme já discutido, sem uma boa porosidade não há como a rocha conter fluidos e sem uma permeabilidade adequada, o fluido não tem como escoar entre os poros da rocha, podendo assim, ser retirado. Nos experimentos, o classificador SVM teve uma taxa de erro de 2.6% enquanto o classificador BPNN (*Backpropagation Neural Networks*) atingiu uma taxa de erro maior em 7.4%.

Al-Anazi & Gates [2010a] comentam sobre a demanda de novas técnicas de predição de litologia em reservatórios que sejam mais eficientes que os métodos estatísticos. Os autores citam a aplicação de técnicas como redes neurais e mais atualmente, a aplicação de SVM com bons resultados mesmo em reservatórios heterogêneos. Eles comparam então estes três métodos e revelam melhores resultados com SVM frente à análise estatística discriminante e redes neurais. Os experimentos foram realizados a partir de dois poços provenientes de reservatórios distintos. O autor relata que o desempenho do classificador SVM variou de acordo com o valor do parâmetro de penalidade C. O método de validação cruzada foi utilizado para encontrar o melhor parâmetro. Em um outro trabalho mais recente, Al-Anazi & Gates [2010b] aplicam SVM em um reservatório bastante heterogêneo com dados de três poços para classificar eletrofácies e prever distribuição de permeabilidade e novamente comparam o desempenho de SVM com redes neurais e com a técnica de análise discriminante, e novamente confirmam a superioridade de SVM frente a estas duas outras técnicas.

Existem muitos trabalhos sobre classificação de litofácies utilizando o reconhecimento de padrões. Os mais recentes têm demonstrado a superioridade do classificador SVM frente a classificadores como redes neurais e análise estatística discriminante ([Al-Anazi & Gates, 2010a], [Wong et al., 2005]). Os trabalhos em sua maioria têm utilizado uma base muito pequena, quase sempre de somente um poço, impossibilitando conclusões sobre o comportamento dos classificadores no caso de serem utilizados em um novo poço. A exceção foram os trabalhos de Al-Anazi & Gates [2010a] e Flexa et al. [2004], sendo que neste último não ficou claro se os testes foram executados em bases separadas ou se as bases dos dois reservatórios foram trabalhadas como uma só base. O autor também não revelou o real desempenho da sua classificação. Já no trabalho de Al-Anazi & Gates [2010a], não foram realizados experimentos para cada base separadamente e não foi verificado se o aprendizado de um classificador a partir de dados de uma reserva apresenta desempenho similar quando aplicado a uma outra reserva. Outro ponto observado é que nenhum trabalho comparou o desempenho do classificador kNN para o problema de classificação de litofácies a partir de dados de perfis elétricos.

O presente trabalho propõe responder às perguntas que foram deixadas pelos trabalhos anteriores utilizando uma base de dados de poços heterogêneos. A partir dessa base, serão comparados o desempenho dos classificadores SVM, kNN, Multilayer Perceptron, Regressão Logística e Árvores de decisão em diversos contextos como quando no mesmo poço e em poços diferentes e se o desempenho do aprendizado com dados do mesmo poço é maior do que o desempenho do aprendizado que utilize dados de poços diferentes em sua base de treinamento. É ainda estudado neste trabalho a combinação destes classificadores por meio de voto majoritário nos mesmos contextos abordados por classificadores individualmente. Na classificação por votação majoritária, escolhe-se a classe com maior número de votos dos classificadores. Os resultados obtidos com nossos experimentos são descritos no próximo capítulo.

# Capítulo 4

## Experimentos e Resultados

Este capítulo descreve os experimentos e os resultados obtidos. Inicialmente, descrevemos os detalhes do protocolo experimental utilizado nos experimentos. Em seguida, os classificadores Multilayer Perceptron, SVM, kNN, Árvore (J48) e Regressão Logística são comparados quanto ao seu desempenho individual. Por fim, os classificadores são combinados em um sistema de votação majoritária, cujo o resultado será apresentado no final.

Todos os passos acima são repetidos em dois contextos diferentes. No primeiro, o treino e teste é realizado dentro do mesmo poço. No segundo contexto, que simula uma situação real de aplicação da classificação automática, uma base de um poço é utilizada para treinamento, enquanto o teste é feito com uma base composta por dados de outro poço.

A base de dados foi fornecida pela Petrobras e está dividida em três partições, sendo que cada partição corresponde a dados de um poço diferente, porém da mesma reserva. Portanto, a base utilizada nos experimentos possui amostras de 03 poços da mesma reserva e que foram classificadas em litofácies por um geólogo especialista. Com a base de treinamento de um poço, iremos comparar o desempenho dos classificadores na classificação de um poço diferente, porém dentro da mesma reserva. Será respondido assim, se o classificador com melhor desempenho em um poço, generaliza seu conhecimento quando o aprendizado for medido em dados de um poço diferente da mesma reserva. Será necessário ainda comparar os mesmos resultados utilizando dados de mais de um poço.

### 4.1 Protocolo Experimental

Nesta seção são descritos detalhes sobre a base de dados, ajuste de parâmetros de classificadores e normalização da base de dados.

### 4.1.1 Base de Dados

Conforme mencionado no início deste capítulo, os experimentos foram realizados com uma base de dados composta por 03 poços, identificados por w6, w7 e w8, onde:

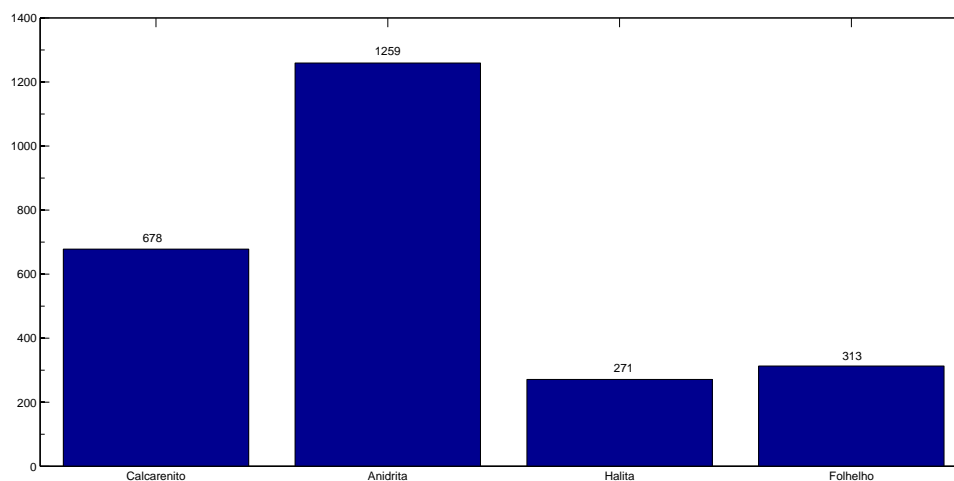
- w6 - constituído por 815 leituras de perfis;
- w7 - constituído por 808 leituras de perfis;
- w8 - constituído por 900 leituras de perfis;

Os perfis lidos, comuns aos três poços para cada registro de leitura, foram DT(Sônico), GR (Raios Gama), LLD e RHOB. Apesar de existirem outros perfis, mantivemos para este trabalho apenas os que se repetiram em todos os poços para que pudessem ser feitas as devidas comparações de desempenho dos classificadores utilizando sempre os mesmos perfis. Em trabalhos anteriores, testamos vários perfis para um poço classificado em fluidos (água, óleo e gás) no qual conseguimos diminuir o número de perfis lidos através de uma seleção de atributos [de Oliveira Júnior et al., 2012]. Essa seleção proporcionou ainda um incremento no desempenho da acurácia dos classificadores. A base rotulada em fluidos foi a primeira fornecida pela Petrobras. Isso justificou o trabalho inicial feito com essa base. Embora tenhamos atingido neste trabalho melhores taxas de acerto ao selecionar os perfis (seleção de características), esse método não se justificou na prática quando por meio de pesquisas mais recentes, entendemos melhor o relacionamento entre os perfis na análise de um poço.

É papel do geofísico decidir quais perfis utilizar. Muitas vezes a utilização de um perfil secundário é feita para confirmar ou mesmo rejeitar a análise de um perfil isolado. Existem perfis por exemplo que medem a porosidade de uma rocha. Quanto mais porosa uma rocha, melhor sua capacidade de reter fluidos. No entanto, valores muito elevados da porosidade, podem indicar um buraco ou caverna em um trecho. Podem mesmo indicar que houve um desabamento da parede de um poço, o que poderia dar uma análise falsa daquele trecho. Existe um perfil que é conhecido como perfil de qualidade que é o caliper. O caliper irá medir a cimentação do poço. Ele poderá indicar por exemplo se houve desabamento em um trecho. É um típico perfil que é utilizado apenas quando se deseja confirmar alguma leitura que cause desconfiança ao geofísico. Quando uma leitura de um perfil gerar dúvidas ao geofísico, é importante que perfis como o caliper estejam disponíveis.

Os dados estão divididos em quatro classes de litofácies: calcarenito, anidrita, folhelho e halita, e foram rotulados por um especialista geofísico. A Tabela 4.2 mostra um exemplo das amostras da base de dados e a Figura 4.1 mostra a distribuição das 4 classes em todos os 03 poços da base de dados.

Na Figura 4.2 podemos ver amostras reais dos tipos de rocha que estamos analisando.



**Figura 4.1.** Distribuição quantitativa das quatro classes encontradas nos três poços.

Nossos experimentos foram divididos em duas séries principais. Na primeira série, os métodos de classificação são comparados usando três diferentes bases de dados através da estratégia de validação cruzada *k-fold cross-validation* onde o valor escolhido de *k* foi 10. Cada base de dados é um conjunto de amostras obtidas a partir de um poço, ou seja, três poços diferentes são investigados. Na segunda série, o objetivo é o de simular uma situação real, onde os dados a partir de um poço são usados para treinar um classificador, enquanto que na fase de teste, o classificador treinado tem para atribuir classes de amostras provenientes de outros poços. Deste modo, estas experiências foram realizadas usando uma base de dados, como o conjunto de treino, enquanto que as restantes bases de dados são utilizadas como conjuntos de testes separados. Este processo foi repetido três vezes, a fim de permitir que cada base de dados para ser utilizado como a formação e conjunto de ensaio.

## 4.1.2 Ajuste de Parâmetros

Conforme mostra o Capítulo 3, a maioria dos classificadores necessita de ajuste de parâmetros. Portanto, foi realizada uma etapa de ajuste. Em todos os classificadores utilizamos como método a busca exaustiva<sup>1</sup> onde no caso do classificador SVM, como kernel escolhemos o PUK (Pearson VII Universal Kernel) que se mostrou melhor do que os kernels tradicionais. O parâmetro de regularização *C* teve no valor 110, a sua melhor taxa de sucesso encontrada em 89,312%. Para o classificador *Multilayer Perceptron*, o parâmetro *learning rate* foi ajustado para 0,17 e o parâmetro *momentum* foi ajustado para 0,1 atingindo sua melhor taxa de acerto em 87,715%. Para o classificador de Regressão Logística e Árvore (J48)

<sup>1</sup>Consiste em testar vários valores até encontrar os parâmetros que permitiram o melhor desempenho.

não houve ajuste de parâmetros. O resumo dos parâmetros pode ser visto na Tabela 4.1. Para essa fase de ajuste, foram usados os dados do poço w6 e os melhores parâmetros encontrados foram repetidos para os demais poços.

### 4.1.3 Normalização da Base

Como pode ser observado na Tabela 4.2, os valores obtidos para cada perfil estão em escalas diferentes. Essa não normalização dos dados pode influenciar negativamente o desempenho dos classificadores, especialmente kNN, que é baseado em medida de distância. Diante desse



(a) Folhelho



(b) Calcarenito



(c) Anidrita



(d) Halita

**Figura 4.2.** Exemplos visuais de rochas representadas na base de dados investigada nos experimentos.

**Tabela 4.1.** Ajuste de parâmetros nos classificadores.

| Classificadores     | Parâmetros de Ajuste                 |
|---------------------|--------------------------------------|
| SVM                 | C = 110; kernel = Puk                |
| kNN                 | k = 1                                |
| Árvore (J48)        | Default                              |
| Redes Neurais (MLP) | learning rate = 0,17; momentum = 0,1 |
| Regressão Logística | Default                              |

**Tabela 4.2.** Exemplo de amostras da base de dados

| n°  | Depth(m)  | Cal    | DT      | GR       | ILD     | MSFL    | NPHI    | Pef    | RhOB   |
|-----|-----------|--------|---------|----------|---------|---------|---------|--------|--------|
| 1   | 2435,0471 | 6,7227 | 56,2182 | 106,4375 | 27,6483 | 35,1158 | 10,8416 | 3,168  | 2,6629 |
| 2   | 2435,1995 | 6,7227 | 57,9455 | 146,5    | 25,7482 | 53,9805 | 12,8258 | 3,1602 | 2,6624 |
| 278 | 2477,262  | 6,7099 | 71,9797 | 75,8699  | 0,6001  | 1,6795  | 11,4636 | 2,9233 | 2,5086 |
| 283 | 2478,0239 | 6,582  | 68,1109 | 63,4688  | 5,0932  | 7,0145  | 13,3006 | 2,5742 | 2,4555 |

fato, um processo de normalização dos dados foi realizado. Aksoy & Haralick [2000] apresentam um estudo comparativo entre seis procedimentos de normalização que possibilitam que os dados fiquem dentro de um intervalo [0,1]. Neste trabalho foram utilizados os dois primeiros procedimentos investigados em Aksoy & Haralick [2000]. No primeiro procedimento de pré-processamento é feito um escalonamento linear normalizando os dados, sendo que é dado um limite inferior  $l$  e um limite superior  $u$  para uma característica  $x$ ,

$$\tilde{x} = \frac{x - l}{u - l}, \quad (4.1)$$

resultando em  $\tilde{x}$  dentro de um intervalo [0,1].

No segundo procedimento de pré-processamento os dados são padronizados, consistindo em transformar uma característica  $x$  para valores observados com média zero e desvio padrão igual a 1 como:

$$\tilde{x} \cong \frac{\mu - l}{\sigma}, \quad (4.2)$$

onde  $\mu$  e  $\sigma^2$  são a média da amostra e a variância igual a 1 da característica  $x$ , respectivamente.

#### 4.1.4 Desempenho Individual e Combinado com Treino e Teste no Mesmo Poço

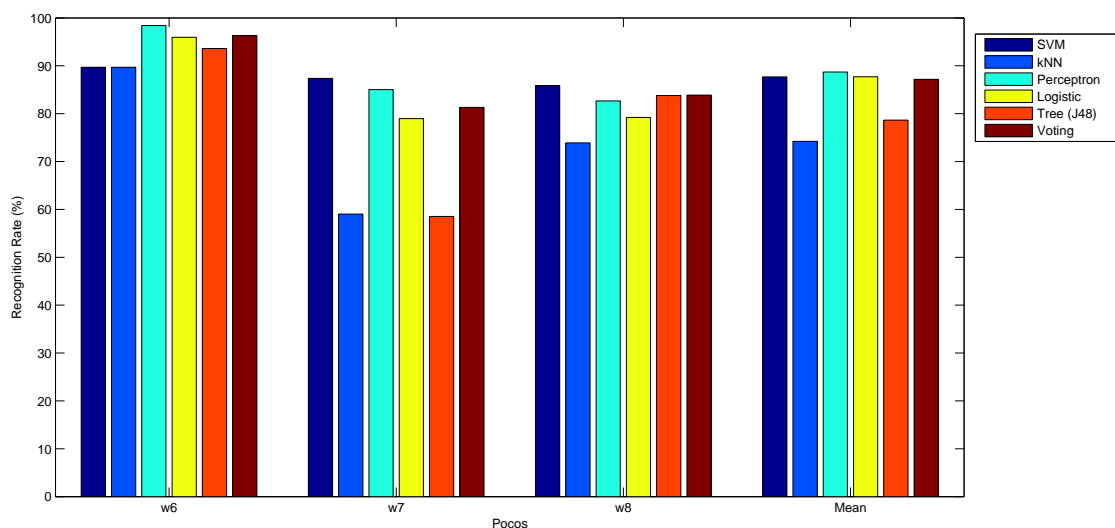
Inicialmente testamos os classificadores individualmente para cada poço. O classificador Multilayer Perceptron atingiu a melhor taxa de acerto em 98,40% para o poço w6. Para os demais poços, o classificador SVM obteve os melhores resultados sendo 87,38% para o poço

w7 e 85,89% para o poço w8. Os piores classificadores foram o kNN e Regressão Logística que além de apresentarem as piores médias de taxas de acerto para todos os poços, atingiram os piores resultados individualmente para o poço w7 com taxas de acerto de 59,03% para kNN e 58,54% para o classificador de Regressão Logística. É interessante notar que mesmo com dois classificadores com taxas de acerto tão baixas, os valores para o sistema de votação majoritária (voting, na Tabela 4.3) atingiu ainda 78,96%. Todos estes valores podem ser vistos em detalhe na Tabela 4.3 e no gráfico representado na Figura 4.3.

**Tabela 4.3.** Taxa de acerto dos métodos de classificação investigados.

| Well  | SVM           | kNN    | Perceptron    | Regressão Logística | Tree(J48) | Voting |
|-------|---------------|--------|---------------|---------------------|-----------|--------|
| w6    | 89,68%        | 89,68% | <b>98,40%</b> | 95,94%              | 93,61%    | 97,54% |
| w7    | <b>87,38%</b> | 59,03% | 85,02%        | 78,96%              | 58,53%    | 78,96% |
| w8    | <b>85,89%</b> | 73,89% | 82,67%        | 79,22%              | 83,78%    | 83,22% |
| Média | 87,68%        | 74,20% | <b>88,70%</b> | 87,71%              | 78,64%    | 87,17% |

No sistema de votação majoritária, o conjunto de classificadores não apresentou melhor desempenho em nenhum dos poços, no entanto, é importante observar que o conjunto de classificadores apresentou taxas muito próximas dos melhores classificadores para os poços w6 e w8 e na média das classificações. No poço w7, os classificadores kNN e Tree(J48) prejudicaram o desempenho do sistema de votação, embora como já dito antes, sua taxa de acerto atingiu 78,96% que representa 21,42% a mais que o pior classificador.



**Figura 4.3.** Gráfico do desempenho dos classificadores treinados e testados no mesmo poço.



Os experimentos realizados mostram claramente que é difícil indicar um único método de classificação como sendo o melhor, já que o desempenho dos classificadores varia de uma base para outra, mas que, conforme indica a literatura, SVM mostrou-se estar entre os melhores resultados, o que mostra que SVM pode até não alcançar o melhor resultado, mas provavelmente estará entre os melhores.

Esses resultados motivaram a investigação em poços diferentes mostrados na próxima subseção.

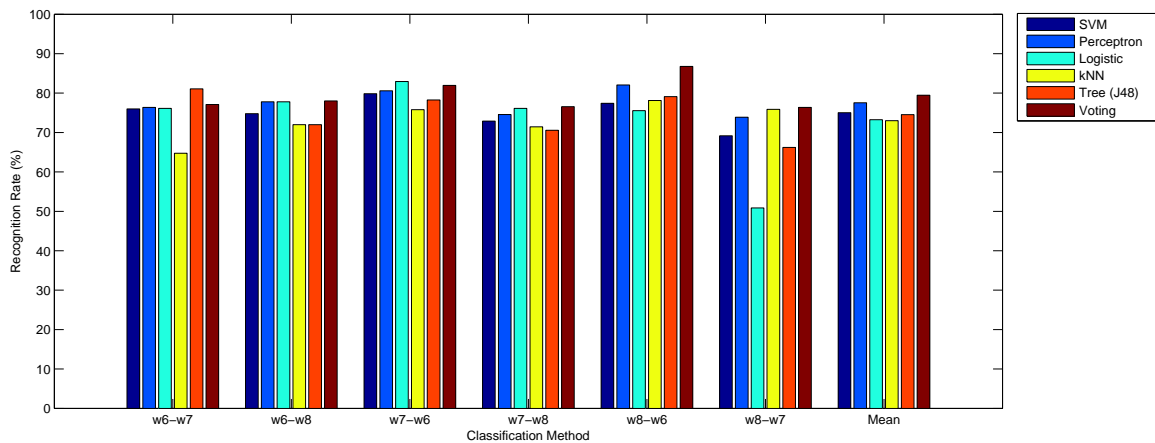
#### 4.1.5 Desempenho poço a poço dos classificadores.

Nesta subseção testamos o desempenho de cada classificador treinando em um poço e testando nos demais. O mesmo se repetiu para os classificadores combinados por votação majoritária. O desempenho dos mesmos pode ser visto na Tabela 4.4 e no gráfico mostrado na Figura 4.4.

**Tabela 4.4.** Taxa de acerto de poço para poço dos métodos de classificação investigados.

| Well  | SVM    | kNN    | Perceptron | Regressão Logística | Tree(J48)     | Voting        |
|-------|--------|--------|------------|---------------------|---------------|---------------|
| w6-w7 | 75,99% | 64,73% | 76,36%     | 76,11%              | <b>81,06%</b> | 77,10%        |
| w6-w8 | 74,78% | 72,00% | 77,78%     | 77,78%              | 72,00%        | <b>78,00%</b> |
| w7-w6 | 79,85% | 75,79% | 80,59%     | <b>82,92%</b>       | 78,26%        | 81,94%        |
| w7-w8 | 72,89% | 71,44% | 74,56%     | 76,11%              | 70,56%        | <b>76,56%</b> |
| w8-w6 | 77,39% | 78,13% | 82,06%     | 75,55%              | 79,11%        | <b>86,75%</b> |
| w8-w7 | 69,18% | 75,87% | 73,89%     | 50,87%              | 66,21%        | <b>76,36%</b> |
| Média | 75,01% | 72,99% | 77,56%     | 73,24%              | 74,53%        | <b>79,45%</b> |

Diferente dos experimentos feitos onde treino e teste pertenciam ao mesmo poço, o sistema de votação majoritária apresentou taxas de acerto superiores aos demais classificadores em 4 dos 6 testes realizados. Nos demais, ele apresentou a segunda melhor taxa de acerto, além de na média ter sido também a melhor. É importante ressaltar que estes testes simulam o que acontece no mundo real da indústria petrolífera. Isso se explica porque em bacias diferentes, a leitura de valores similares dos perfis pode apresentar classificações distintas. Por isso é importante o conhecimento da reserva por um geofísico para que este profissional possa fazer uma primeira classificação de um poço dando um padrão que possa ser aprendido pelos classificadores. Assim, a automatização deste processo, precisaria sempre de um poço onde seria feito o primeiro estudo pelo geofísico. Este poço serviria assim para calibrar os classificadores e a partir deste, classificar poços vizinhos com características similares poupando o profissional de classificar todos os poços e limitando-se apenas a poços chaves.



**Figura 4.4.** Gráfico do desempenho dos classificadores treinados em um poço e testado nos demais.

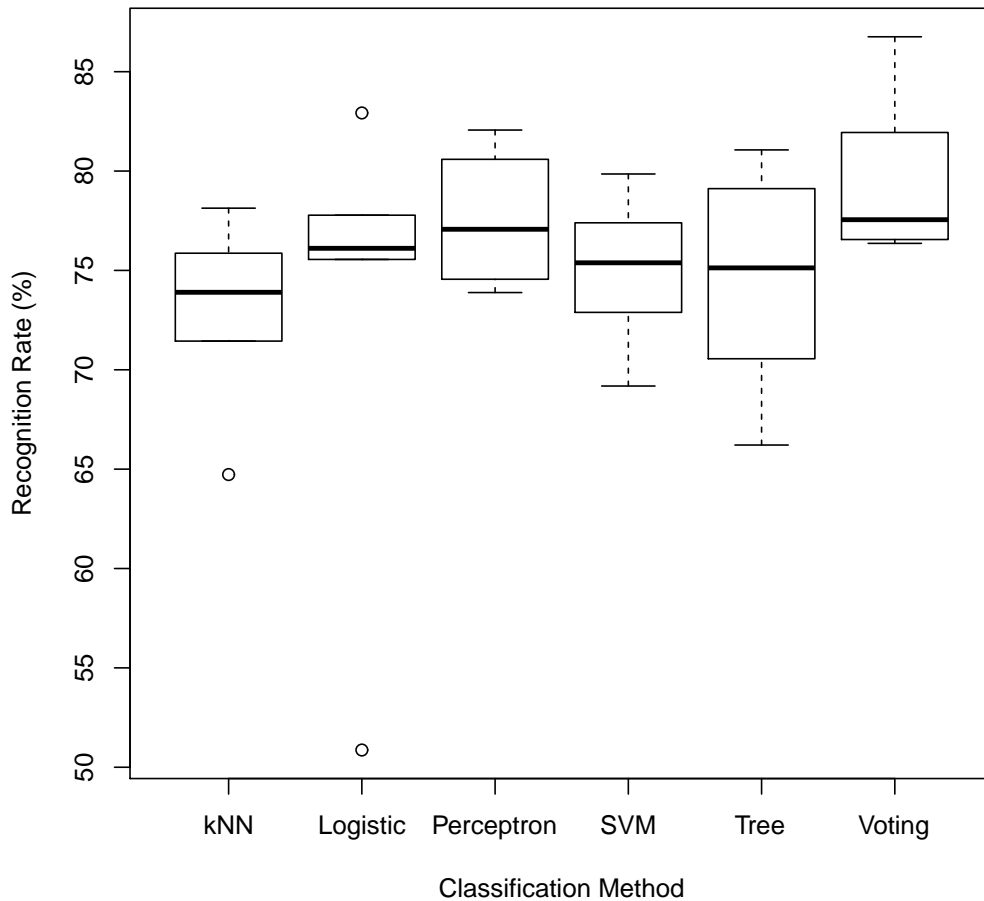
**Tabela 4.5.** Resumo das Estatísticas Descritivas

| Classificador       | Média  | Mediana | D.Padrão | Mínimo | Máximo | 1.Quartil | 2.Quartil | Coef.Assim |
|---------------------|--------|---------|----------|--------|--------|-----------|-----------|------------|
| KNN                 | 72,99% | 73,90%  | 4,78     | 64,73  | 78,13% | 71,58%    | 75,85%    | -0,0863    |
| Regressão Logística | 73,22% | 76,11%  | 11,28    | 50,87  | 82,92% | 75,69%    | 77,36%    | 0,4940     |
| Perceptron          | 77,54% | 77,07%  | 3,27     | 73,89  | 82,06% | 75,01%    | 79,89%    | 0,1556     |
| SVM                 | 75,01% | 75,38%  | 3,70     | 69,18  | 79,85% | 73,36%    | 77,04%    | -0,0999    |
| Tree(J48)           | 74,53% | 75,13%  | 5,81     | 66,21  | 81,06% | 70,92%    | 78,90%    | -0,0545    |
| Votação             | 79,45% | 77,55%  | 4,13     | 76,36  | 86,76% | 76,69%    | 80,96%    | 0,5962     |

Na próxima seção, iniciamos uma análise estatística destes dados.

#### 4.1.6 Análise Estatística dos Resultados dos experimentos de Poço a Poço.

São apresentados na Tabela 4.5 algumas estatísticas como média, mediana, desvio padrão, valor mínimo e máximo, e coeficiente de assimetria. Estes valores são referentes as taxas de acerto obtidos por cada um dos classificadores. Verificamos nessa tabela que o classificador Regressão Logística apresentou o menor índice de acertos em 50.87% caracterizado como um outlier no Gráfico 4.5, porém possui a menor variabilidade em relação aos valores centrais localizados entre o primeiro e terceiro quartil. O classificador Regressão Logística possui também o maior desvio padrão demonstrando uma grande variação da taxa de acerto. Os classificadores combinados em conjunto por meio de um sistema de votação majoritária apresentou mediana superior aos demais, mas similar ao classificador Perceptron, com 50% dos dados acima de 77%, indicando maior índice de acertos atingindo 86.76%, e seus valores centrais localizados entre 76% e 81%. O classificador kNN possui a menor mediana dentre os classificadores e também apresenta valores extremos. O classificador Perceptron



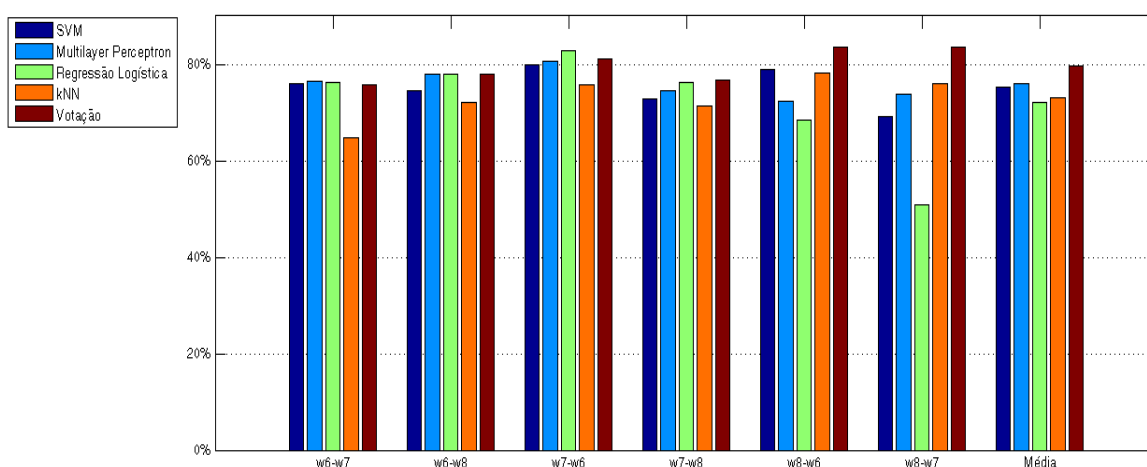
**Figura 4.5.** Box-plot para índice de acertos dos classificadores

possui valores centrais entre 75% e 79%, indicando margem de acertos um pouco inferior aos classificadores combinados por meio de votação. O classificador Tree(J48) apresentou uma maior variabilidade em relação aos valores centrais. Quanto à assimetria, os classificadores Regressão Logística, Perceptron e Votação apresentaram assimetria positiva, isto é, cauda longa à direita. Os demais classificadores apresentaram assimetria negativa.

Segundo o teste de Friedman [Conover, 1999], teste estatístico não paramétrico que faz uma comparação entre os métodos (classificadores) e testa a hipótese nula ( $H_0$ ), de que não há diferença entre os métodos, contra a hipótese alternativa ( $H_1$ ) de que pelo menos um dos métodos difere dos demais, observamos que o sistema de Votação e o classificador Perceptron não apresentam diferença significativa, no entanto, o sistema de Votação difere dos demais. O p-valor com resultado igual a 0,0053 sugere que devemos rejeitar a hipótese de igualdade entre os métodos.

## 4.2 Resumo dos resultados

A combinação de classificadores por voto majoritário, se mostrou inferior a classificadores individuais quando a base de treinamento e de testes pertencem ao mesmo poço. Porém, quando usamos uma base de treinamento de um poço para classificar dados de poços distintos, a combinação de classificadores apresentou desempenho superior a todos os classificadores individuais para os testes entre os poços w8-w6 e w8-w7. Na média dos testes realizados entre todos os poços, a combinação de classificadores mais uma vez apresentou superioridade.



**Figura 4.6.** Desempenho Individual dos Classificadores e Combinados em Conjunto Utilizando Votação Majoritária.

A votação apresentou menos variação na taxa de acerto (veja Figura 4.5), isto é, a taxas de voting são todas acima de 75% de acerto, enquanto bons métodos como MLP e SVM podem ser ótimos em algumas situações, também podem ser medianos em outras. Isso quer dizer que é melhor usar a combinação, pois, dessa forma, teremos uma taxa de acerto mais próxima da média e menos dependente da base de treinamento/teste.

É importante destacar que a estratégia de combinação de classificadores é baseada na hipótese de que classificadores fracos podem ser combinados a fim de produzir decisões mais corretas. Neste trabalho, porém, os classificadores combinados não são fracos. Todos os cinco membros do conjunto foram calibrados para produzirem decisões individuais otimizadas. Além disso, conjuntos heterogêneos são menos comuns na literatura [Santos, 2008]. Diante desse contexto, foi realizado um experimento adicional para compararmos o desempenho obtido pelo conjunto dos 5 cinco classificadores heterogêneos ao desempenho de um conjunto de classificadores homogêneos, criado por meio da tradicional técnica *Ran-*

*dom Forest*. Esta técnica gera  $n$  diferentes bases de treinamento por meio de seleção aleatória de amostras e de atributos. Cada base é então usada para treinar uma Árvore de Decisão. Portanto, é gerada uma floresta com  $n$  árvores.

A Tabela 4.6 mostra o resultado da comparação entre *Random Forest* e o nosso conjunto de 5 classificadores heterogêneos. É possível observar nessa tabela que o conjunto heterogêneo superou o método *Random Forest*. Portanto, nesse caso, a diversidade de opinião gerada por diferentes algoritmos de aprendizagem de máquina foi mais importante para aumentar a taxa de acerto do sistema do que a diversidade gerada por mudanças nas amostras e nos atributos.

**Tabela 4.6.** Desempenho de poço para poço dos métodos de classificação investigados comparados ao Random Forrest.

| Well  | SVM    | kNN    | Perceptron | Regressão Logística | Tree(J48) | Random Forrest | Voting |
|-------|--------|--------|------------|---------------------|-----------|----------------|--------|
| w6-w7 | 75,99% | 64,73% | 76,36%     | 76,11%              | 81,06%    | 69,68%         | 77,10% |
| w6-w8 | 74,78% | 72,00% | 77,78%     | 77,78%              | 72,00%    | 76,67%         | 78,00% |
| w7-w6 | 79,85% | 75,79% | 80,59%     | 82,92%              | 78,26%    | 80,34%         | 81,94% |
| w7-w8 | 72,89% | 71,44% | 74,56%     | 76,11%              | 70,56%    | 72,11%         | 76,56% |
| w8-w6 | 77,39% | 78,13% | 82,06%     | 75,55%              | 79,11%    | 74,57%         | 86,75% |
| w8-w7 | 69,18% | 75,87% | 73,89%     | 50,87%              | 66,21%    | 66,83%         | 76,36% |
| Média | 75,01% | 72,99% | 77,56%     | 73,24%              | 74,53%    | 73,37%         | 79,45% |

No próximo capítulo faremos as considerações finais e conclusões de trabalho.



# Capítulo 5

## Conclusões e Trabalhos Futuros

Neste trabalho foram comparados os classificadores Multilayer Perceptron, SVM, kNN, Árvore (J48) e Regressão Logística para o problema de classificação de litofácies por meio de perfis elétricos. Depois de comparados individualmente, os classificadores foram combinados em um sistema de votação majoritária. Simulamos duas situações hipotéticas. Na primeira, treinamos e testamos utilizando dados no mesmo poço. Para este cenário, os classificadores Perceptron e SVM apresentaram o melhor desempenho. O sistema de votos majoritário não se mostrou como melhor opção, porém foi mais estável que classificadores como Tree(J48) e kNN. No segundo cenário, testamos os classificadores em um ambiente que reflete melhor a realidade, onde usamos um poço para o treinamento e o teste foi realizado em um poço distinto. Nesse ambiente, o conjunto de classificadores utilizando voto majoritário, se mostrou a melhor opção entre 4 de 6 testes e nos dois testes restantes, teve o segundo melhor desempenho. Nos testes estatísticos, confirmamos seu melhor desempenho quando analisando todos os dados provando que classificadores combinados incrementam as taxas de sucesso além de dar mais estabilidade à classificação. Esse comportamento é vantajoso pois é mais próximo da realidade da análise de geólogos diferentes para o mesmo poço. Como algumas rochas possuem características muito próximas, é normal que haja diferença entre a análise de um geólogo e outro. Em entrevista com um geólogo da Petrobras, ao comentar sobre as diferenças entre taxas de erro próximas a 20%, o mesmo confidenciou que essas taxas são próximas as diferenças de análise entre colegas geofísicos que analisaram o mesmo poço. Portanto, mais importante que atingir taxas elevadas de acerto, a estabilidade do desempenho dos classificadores é fundamental para a confiança em sua classificação.

### 5.1 Trabalhos Futuros

Dando continuidade a estes estudos, é necessário comparar ainda o aprendizado de:

- Poços *onshore* (em terra) com poços *offshore* (no mar);
- Poços na mesma reserva como neste trabalho, porém mais distantes;
- Poços de reservas distintas;
- Poços na mesma reserva, porém sendo *offshore*.

Somando-se aos itens anteriores, precisaríamos aumentar a base de estudo para que se tenha mais confiabilidade nos dados e ainda que mais classes de rocha sejam estudadas. O incremento dos estudos acima iria nos proporcionar mais respostas que não pudemos dar neste trabalho.

## 5.2 Publicações Obtidas

Durante este mestrado, foram obtidas duas publicações relacionadas a este trabalho. O primeiro artigo foi publicado no Congresso Latino-Iberoamericano de Investigación Operativa e Simpósio Brasileiro de Pesquisa Operacional sob o título de "Seleção de Características para Classificação de Fluidos por Meio da Análise de Perfis Elétricos de Poços de Petróleo da Amazônia"[de Oliveira Júnior et al., 2012]. No trabalho citado, nossa meta foi identificar os melhores perfis para a classificação de fluidos por meio de testes exaustivos com diversas combinações de perfis buscando uma seleção de características que promovesse incremento no desempenho dos classificadores. Nossos resultados indicavam que o classificador com melhor desempenho foi o kNN, e o melhor subconjunto de características foi composto pelos perfis DT, GR, ILD, MSFL, NPHI e RHOB, sendo que o número de perfis foi reduzido enquanto ocorreu o aumento na taxa de acerto do classificador. Como já mencionado, depois de mais estudos e entrevistas com geofísicos, entendemos que esse tipo de seleção de características não é interessante, pois cabe ao geofísico identificar em que momento um perfil deveria ser descartado. A experiência e conhecimento da reserva pelo geofísico é fundamental para a tomada desse tipo de decisão. Por último, publicamos um artigo na conferência *The International Joint Conference on Neural Networks* sob o título "*Ensemble of Heterogeneous Classifiers Applied to Lithofacies Classification Using Logs from Different Wells*"[de Oliveira Júnior et al., 2013].



# Referências Bibliográficas

- Aksoy, S. & Haralick, R. M. (2000). Effects of feature normalization on image retrieval. *IAPR International Conference on Pattern Recognition*, 22:563 – 582.
- Al-Anazi, A. & Gates, I. (2010a). On the capability of support vector machines to classify lithology from well logs. *Natural Resources Research*, 19:125 – 139.
- Al-Anazi, A. & Gates, I. (2010b). A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Engineering Geology*, 114(3–4):267 – 277.
- Altınçay, H. (2007). Ensembling evidential k-nearest neighbor classifiers through multi-modal perturbation. *Applied Soft Computing*.
- An-nan, J. & Lu, J. (2009). Studying the lithology identification method from well logs based on de-svm. Em *Control and Decision Conference, 2009. CCDC '09. Chinese*, pp. 2314 –2318.
- Conover, W. J. (1999). *Practical Nonparametric Statistics*. 3rd ed. edição.
- Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21 –27.
- Cunha, E. S. (2002). Identificação de litofácies de poços de petróleo utilizando um método baseado em redes neurais artificiais. Dissertação de mestrado, Universidade Federal de Campina Grande.
- de Oliveira Júnior, J. M.; dos Santos, E. M.; Carvalho, J. E. T. R. H. & de Vasconcelos Marques, L. A. (2013). Ensemble of heterogeneous classifiers applied to lithofacies classification using logs from different wells. *The International Joint Conference on Neural Networks*.
- de Oliveira Júnior, J. M.; dos Santos, E. M. & Carvalho, J. R. (2012). Seleção de características para classificação de fluidos por meio da análise de perfis elétricos de poços de

- petróleo da amazônia. *Congresso Latino-Iberoamericano de Investigación Operativa e Simpósio Brasileiro de Pesquisa Operacional*.
- Flexa, R. T.; Andrade, A. & Carrasquilla, A. (2004). Identificação de litotipos nos perfis de poço do campo de namorado (bacia de campos, brasil) e do lago de maracaibo (venezuela) aplicando estatística multivariada. *Revista Brasileira de Geociências*, 34:571–578.
- Garcia, E. N. (2008). *A PETROBRAS NA AMAZÔNIA: a história da pesquisa de combustíveis fósseis*. Norma Editora, Manaus, 1 edição.
- Henke, M.; Santos, C.; Nunan, E.; Feitosa, E.; dos Santos, E. & Souto., E. (2011). *Aprendizagem de Máquina para Segurança em Redes de Computadores: Métodos e Aplicações*.
- K.Jain, A.; Duin, R. P. W. & Mao, J. (2000). Statistical pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:4–37.
- Petrobras (2011). *Conheça nosso plano de negócios 2011-2015*.
- Santos, E. M. (2008). Static and dynamic overproduction and selection of classifier ensembles with genetic algorithms.
- Tafner, M. A.; de Xerez, M. & Filho, I. W. R. (1995). *Redes Neurais Artificiais: introdução e princípios de neurocomputação*.
- Thomas, J. E.; Triggia, A. A. & Correia, C. A. (2001). *Fundamentos de Engenharia de Petróleo*. Interciência.
- Vapnik, V. N. (1999). *The Nature of Statistical Learning Theory (First Edition)*. Springer, 2 edição.
- Wang; Xuechuan; Paliwal & K., K. (2003). Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition*, 33:2429–2439.
- Wong, K.; Ong, Y.; Gedeon, T. & Fung, C. (2005). Reservoir characterization using support vector machines. Em *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, volume 2, pp. 354 –359.

## **Apêndice A**

**Artigo publicado no Congresso Latino-Iberoamericano de Investigación Operativa e Simpósio Brasileiro de Pesquisa Operacional no ano de 2012**

# Seleção de Características para Classificação de Fluidos por Meio da Análise de Perfis Elétricos de Poços de Petróleo da Amazônia

Joacir M. de Oliveira Júnior<sup>1</sup>, Eulanda M. dos Santos<sup>1</sup>, José Reginaldo Carvalho<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Informática  
Universidade Federal do Amazonas (UFAM) – Manaus – AM

{joacir,emsantos,reginaldo}@dcc.ufam.edu.br

**Abstract.** *Among several steps which are necessary in order to lead oil to be marketed, the analysis of well logs plays an important role to estimate the capacity of a well. One goal of this analysis is to classify fluids into oil, gas or water. This analysis is traditionally conducted as a semi-automated process which has a system to generate graphs of curves. We propose in this work to carry out well logs classification as an automatic pattern recognition process. In addition, we compare two classifiers in order to identify the best method for automatic analysis of well logs data which are obtained from the same reservoirs. Finally, we conduct an exhaustive feature subset selection process focusing on pointing out the best feature set for fluids classification. Our results indicate that kNN achieved the best recognition rates, while the feature subset composed of six features, namely DT, GR, ILD, MSFL, NPFI e RHOB, is the most relevant information for fluids classification. Our feature subset selection strategy helped to reduce the number of features while the recognition rate was increased.*

**Resumo.** *Dentro das várias etapas que são necessárias até o petróleo ser comercializado, a análise de perfis elétricos representa papel de grande importância para se estimar a capacidade produtiva de um poço. Um dos objetivos da análise é a classificação de fluidos em água, óleo, gás e não reservatório. Essa análise atualmente é semi-automatizada, onde um sistema gera gráficos de curvas a partir dos quais geólogos com conhecimentos específicos, analisam e realizam a caracterização do reservatório. Outro ponto importante para os geólogos é a escolha dos perfis a serem utilizados como determinantes à classificação. Neste trabalho busca-se realizar o reconhecimento automático dos padrões destes dados, demonstrar e comparar classificadores capazes de realizar análise automática de dados de perfis de poços de petróleo com índice de acerto igual ou superior à análise de especialistas humanos e identificar os melhores perfis para a classificação de fluidos por meio de testes exaustivos com diversas combinações de perfis. Nossos resultados indicam que o classificador com melhor desempenho foi o kNN, e o melhor subconjunto de características é composto pelos perfis DT, GR, ILD, MSFL, NPFI e RHOB, sendo que o número de perfis foi reduzido enquanto ocorreu o aumento na taxa de acerto do classificador.*

## 1. Introdução

A constatação de que a Amazônia é uma região petrolífera ocorreu no dia 13 de março de 1955, segundo relatos históricos de Garcia (2008). O poço pioneiro foi batizado de NO-1-AZ e localizava-se a 125 quilômetros de Manaus, a uma profundidade de 2.744 metros. Esse evento produziu muita euforia, especialmente na população do estado do Amazonas, pois a perspectiva de produção do poço NO-1-AZ depois dos primeiros testes realizados era de 600 a 700 barris por dia. Essa perspectiva de produção era suficiente para estimular mais pesquisas e a abertura de novos poços na região. Além disso, na época acreditava-se que em apenas três anos o Brasil estaria livre da importação de combustíveis, tão grande era o otimismo dos engenheiros da Petrobras.

Atualmente, a indústria de petróleo brasileira cresce vigorosamente. Conforme mostra a Tabela 1, o Brasil produz mais de 2,5 milhões de barris por dia e o estado do Amazonas é responsável por cerca de 107.496 barris diários. Segundo o plano de negócios da Petrobras (2011), a empresa espera duplicar suas reservas provadas até 2020. A realidade da indústria de petróleo brasileira é diferente da situação da indústria mundial, pois a produção de petróleo declina mundialmente e seus preços disparam no mercado internacional. Portanto, a pesquisa na área de petróleo apresenta elevada demanda no Brasil.

**Tabela 1. Distribuição da Produção de Petróleo e Gás Natural no Brasil por Estado. Ano de 2011. FONTE: ANP/SDP/SIGEP. Siglas: bbl = oil barrel, Mm<sup>3</sup> = metros cúbicos, boe = barris de óleo equivalente.**

| Estado              | Petróleo (bbl/d) | Gás Natural (Mm <sup>3</sup> /d) | Produção Total (boe/d) |
|---------------------|------------------|----------------------------------|------------------------|
| Rio de Janeiro      | 1.551.597        | 25.360                           | 1.711.114              |
| Espírito Santo      | 319.518          | 11.901                           | 394.373                |
| Amazonas            | 35.110           | 11.508                           | 107.496                |
| Bahia               | 42.880           | 6.565                            | 84.172                 |
| São Paulo           | 42.225           | 5.169                            | 74.739                 |
| Rio Grande do Norte | 58.639           | 1.701                            | 69.338                 |
| Sergipe             | 42.365           | 2.428                            | 57.640                 |
| Alagoas             | 5.198            | 1.539                            | 14.878                 |
| Ceará               | 7.209            | 86                               | 7.752                  |
| <b>Total geral</b>  | <b>2.104.743</b> | <b>66.257</b>                    | <b>2.521.502</b>       |

Antes de um poço começar a produzir, existe um longo e caro processo que envolve o estudo e a análise de dados geológicos e geofísicos. Esta fase é denominada **avaliação de formações**. Trata-se de atividades e estudos que visam definir em termos qualitativos e quantitativos o potencial de uma jazida petrolífera, isto é, sua capacidade produtiva e a valoração de suas reservas de óleo e gás (Thomas et al., 2001). Essa avaliação é baseada principalmente nos dados coletados na perfilagem de um poço aberto. Conforme mostra a Figura 1 e segundo Thomas et al. (2001), a perfilagem é uma imagem visual em relação à profundidade, de uma ou mais características ou perfis das rochas perfuradas tais como resistividade elétrica, potencial eletroquímico natural, tempo de trânsito das ondas mecânicas, radioatividade natural ou induzida. Outra definição pode ser encontrada em (An-nan and Lu, 2009). Esse autor afirma que perfilagem é o método que busca identificar propriedades das

rochas e fluidos medindo informações elétricas, acústicas, termais e radioativas. Portanto, com a perfilagem é possível extrair dados importantes das formações pelas quais o poço atravessa como a litologia (tipo de rocha), espessura, porosidade, presença de óleo, gás ou água. Dependendo da análise dos perfis e das informações complementares, pode se decidir pelo abandono do poço ou no caso de se provar interesse econômico, os testes de formação para se estimar a capacidade produtiva do poço são iniciados.

A classificação de perfis de poços de petróleo pode claramente ser tratada como um problema de reconhecimento automático de padrões. Não é surpresa a existência de trabalhos na literatura que realizam a classificação dos perfis de poços com ferramentas de reconhecimento de padrões (Cunha, 2002), (Flexa et al., 2004), (Wong et al., 2005) e (An-nan and Lu, 2009). Porém, apesar da possibilidade de resolução de diferentes problemas de classificação, tais como: litologia, porosidade, fluido, etc., a maioria dos trabalhos publicados na literatura realiza classificação de litologia, também conhecida como classificação de litofácies. Uma das principais razões desse consenso é que a classificação de litologia pode ser caracterizada apenas com a leitura dos perfis, fato que torna essa tarefa mais simples de ser automatizada do que a classificação de fluidos, por exemplo, que é mais complexa por depender da análise de um especialista (geólogo) com conhecimento sobre a reserva. Fluidos são classificados em óleo, gás ou água.

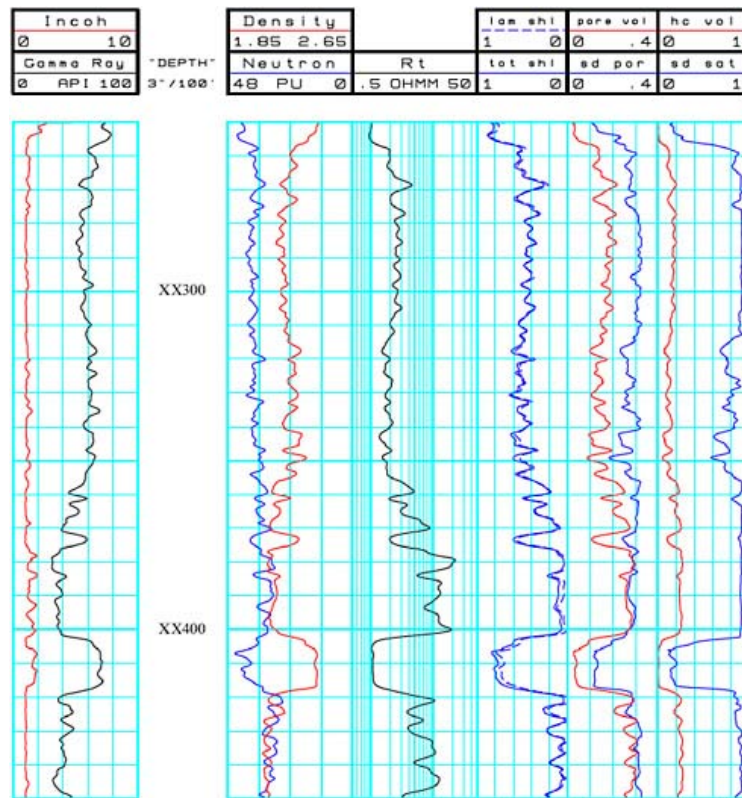
Porém, a identificação de fluidos de um reservatório de petróleo é um processo intensivo que envolve o investimento de uma quantidade considerável de tempo por parte de um especialista experiente e com conhecimento da reserva em estudo. O problema se torna muito mais difícil à medida que aumenta o número de perfis simultâneos a serem analisados (Cunha, 2002). É papel do geólogo determinar quais os melhores perfis a serem analisados para a reserva e fazer a caracterização do reservatório que é de fundamental importância para justificar investimentos da continuidade de exploração de um poço ou o seu abandono. Portanto, a automação do processo de seleção de perfis e da identificação de fluidos pode acelerar a etapa de caracterização de reservatórios.

A seleção de perfis para a classificação de fluidos pode ser visto como um processo de otimização conhecido na literatura como seleção de características ou atributos. Segundo Kudo and Sklansky (2000), a seleção de características contribui para: (1) reduzir os custos de extração de características; (2) aumentar a acurácia da classificação; e (3) aumentar a confiança da estimativa de desempenho. O primeiro aspecto é uma consequência da redução da dimensão do espaço de entrada obtida por meio da eliminação de características irrelevantes. Os dois últimos aspectos estão relacionados com o fato de que métodos de classificação são ajustados para tomar decisões com base nas características mais relevantes. Logo, características irrelevantes ou redundantes podem confundir o classificador.

O processo de seleção de subconjuntos de perfis pode ser realizado por algoritmos de busca, normalmente não exaustivos, especialmente quando há muitos perfis a serem testados. Dada uma amostra  $\mathcal{X}$ , representada por um vetor de características (perfis), onde  $\mathcal{X} = \{x_1, x_2, \dots, x_d\}$ , algoritmos de busca são usados com frequência quando  $\mathcal{X}$  possui muitos perfis devido à complexidade exponencial que

uma busca exaustiva demandaria, uma vez que, sendo  $\mathcal{P}(\mathcal{X})$  o conjunto de todos os subconjuntos possíveis dos perfis de  $\mathcal{X}$ , o tamanho de  $\mathcal{P}(\mathcal{X})$  é  $2^d$ , onde  $d$  representa o número de perfis em  $\mathcal{X}$ .

Na análise de perfis, entretanto, há oito principais tipos de perfis elétricos: Raios Gama (GR), Neutrônico (NPHI), Indução (ILD), Sônico (DT), Densidade (RHOB), Caliper (Cal), Micro SFL (MSFL) e Índice Fotoelétrico (PEF). A descrição de cada perfil é mostrada na Tabela 2. Portanto, um processo de busca exaustiva pode ser utilizado devido ao pequeno número de perfis que representam cada amostra  $\mathcal{X}$ . É importante também destacar que esses perfis são os mesmos utilizados para a classificação de litologias na maioria dos trabalhos publicados na literatura. Apesar do papel fundamental da etapa de seleção de características em problemas de classificação, normalmente não é realizada uma análise de impacto de cada perfil nos trabalhos relacionados com o estudo de perfis elétricos (Cunha, 2002), (Flexa et al., 2004), (Wong et al., 2005) e (An-nan and Lu, 2009). Logo, todos os perfis são utilizados para treinar os algoritmos de classificação.



**Figura 1. Perfis elétricos coletados durante a perfilagem de um poço.**

Diante desse contexto, este trabalho aplica métodos de classificação ao problema de identificação automática de fluidos utilizando dados de perfis de um poço de petróleo da Amazônia. Para que a taxa de acerto dos métodos de classificação seja maximizada, é realizada uma etapa de seleção de características para o problema investigado. O processo de otimização é realizado de forma exaustiva devido ao pequeno número de perfis investigados, isto é, oito perfis. Os objetivos deste trabalho são: (1) identificar os perfis mais relevantes para a tarefa de classificação

**Tabela 2. Principais Tipos de perfis elétricos e suas propriedades.**

| Perfil                    | Descrição  |
|---------------------------|--|
| Raios Gama (GR)           | Detecta a radioatividade total da formação geológica. Utilizado para a identificação da litologia, a identificação de minerais radioativos e para o cálculo do volume de argilas ou argilosidade.  |
| Neutrônico (NPHI)         | São utilizados para estimativas de porosidade, litologia e detecção de hidrocarbonetos leves ou gás.   |
| Indução (ILD)             | Fornece leitura aproximada da resistividade da rocha contendo hidrocarbonetos, através da medição de campos elétricos e magnéticos induzidos nas rochas.   |
| Micro SFL (MSFL)          | Similar ao ILD.  |
| Índice Fotoelétrico (PEF) | É uma medida complementar de ferramentas de medição de densidade que registra a absorção de raios gama.  |
| Sônico (DT)               | Mede a diferença nos tempos de trânsito de uma onda mecânica através das rochas.   |
| Densidade (RHOB)          | Além da densidade das camadas, permite o cálculo da porosidade e a identificação das zonas de gás.   |
| Caliper                   | Fornece o diâmetro do poço. É aplicado no cálculo do volume de cimento para tampões ou cimentação do revestimento, apoio a operações de teste de formação, controle de qualidade de perfis e indicações das condições do poço em um determinado intervalo. |

automática de fluidos; (2) mostrar que métodos de classificação automática podem alcançar índice de acerto igual ou superior à análise de especialistas humanos; e (3) comparar dois métodos de classificação, kNN (*k Nearest Neighbors*) e SVM (*Support Vector Machines*), a fim de também identificar o método com maior taxa de classificação correta para o problema.

O restante deste artigo está organizado da seguinte forma: na seção 2, é feito um levantamento do estado da arte através da descrição das soluções propostas nos principais trabalhos encontrados na literatura para o problema de análise de perfis elétricos. Na seção 3 é descrita a metodologia que será empregada no desenvolvimento do trabalho, bem como a base de dados, normalização dos dados, descrição dos experimentos realizados e dos resultados obtidos com os classificadores kNN e SVM. A seção 3.2 descreve a análise de características realizada através de um método exaustivo o qual seleciona os principais perfis que permitem atingir melhores taxas de acerto na classificação por fluidos. As discussões sobre os resultados obtidos são apresentadas na seção 4. Por fim, a seção 5 resume as conclusões deste trabalho.

## 2. Trabalhos Relacionados

Algumas tentativas de automatização do processo de análise de perfis elétricos têm sido publicadas na literatura. Nesta seção são discutidos alguns desses trabalhos. Conforme mencionado na introdução, a maioria dos trabalhos realiza classificação de litologia.

Cunha (2002) utilizou dados de perfis e de testemunhos de poços para treinar redes neurais para a classificação de litofácies. A rede neural foi treinada com dados



de 8 poços de petróleo e atingiu uma taxa de acerto de 80%. A autora relata que uma solução para o problema só foi possível após a incorporação de uma estratégia para agrupamento prévio das litofácies e tratamento de padrões problemáticos (regiões de conhecimento incerto nos conjuntos de treinamento e de teste). Outro ponto a destacar é que a autora utilizou apenas redes neurais, que têm apresentado um desempenho inferior quando comparadas com SVM (Al-Anazi and Gates, 2010a), (Wong et al., 2005), por exemplo.

Flexa et al. (2004) propuseram uma metodologia baseada na técnica de análise discriminante para a identificação de litologias nos perfis de poços. Os autores pesquisaram a associação entre as variações das propriedades físicas e os tipos litológicos em profundidade. Os perfis geofísicos utilizados nesse trabalho foram raios gama (GR), sônico (DT), porosidade neutrônica (PHIN), densidade (RHOB) e resistividade (ILD). Segundo Flexa et al. (2004), a análise discriminante é uma técnica da estatística multivariada com grande eficiência na solução de problemas geofísicos. A técnica é utilizada nos casos em que se deseja realizar a separação estatística entre dois ou mais grupos de dados, que devem ser definidos previamente pela caracterização de suas variáveis discriminantes. Entretanto, no trabalho de Flexa et al. (2004), a técnica de análise discriminante não é comparada com outros classificadores. Os autores consideram que bons resultados foram obtidos, porém o desempenho real não foi apresentado.

Wong et al. (2005) investigaram o desempenho do classificador SVM para a caracterização de reservatórios como permeabilidade e porosidade. Essas características são fundamentais para a localização de hidrocarbonetos, pois, sem uma boa porosidade não há como a rocha conter fluidos e sem uma permeabilidade adequada, o fluido não tem como escoar entre os poros da rocha, podendo assim, ser retirado. Nos experimentos, o classificador SVM teve uma taxa de erro de 2.6% enquanto o classificador baseado em rede neural do tipo *Backpropagation* (BPNN) atingiu uma taxa de erro superior, 7.4%. Apesar desse trabalho não ter como objetivo a classificação de litofácies, é realizada a classificação de características das rochas como porosidade e permeabilidade, em que os mesmos perfis utilizados para classificar litologias e fluidos são empregados.

An-nan and Lu (2009) propuseram um trabalho de identificação de litologias a partir de dados de perfilagem utilizando DE-SVM com uma taxa de acerto de 90%. DE é o acrônimo de diferença evolutiva, um algoritmo baseado em algoritmos genéticos. Os autores observam que SVM tem sido usado com sucesso para identificação de litologias, mesmo com uma amostra pequena de dados. O artigo apresentou um método de otimização baseado na diferença aritmética-evolutiva (DE) para selecionar os parâmetros. Além disso, como o objetivo principal dos autores foi desenvolver um método para encontrar os melhores parâmetros para SVM, nesse trabalho não houve comparações entre classificadores como kNN e redes neurais.

Al-Anazi and Gates (2010a) comentam sobre a necessidade de novas técnicas de predição de litologia em reservatórios que sejam mais eficientes que os métodos estatísticos. Os autores citam a aplicação de técnicas como redes neurais e, mais atualmente, a aplicação de SVM com bons resultados mesmo em reservatórios heterogêneos. Eles comparam então esses três métodos e revelam melhores resultados

com SVM frente à análise estatística discriminante e redes neurais. Os experimentos foram realizados a partir de dois poços provenientes de reservatórios distintos. Em um outro trabalho mais recente, Al-Anazi and Gates (2010b) aplicam SVM em um reservatório bastante heterogêneo com dados de três poços para classificar eletrofácies e prever distribuição de permeabilidade. Novamente, os autores comparam o desempenho de SVM com redes neurais e com a técnica de análise discriminante, e também confirmam a superioridade de SVM frente às duas outras técnicas.

Conforme pode ser observado nesta seção, existem diversos trabalhos sobre classificação de litofácies utilizando ferramentas de reconhecimento de padrões. Durante as pesquisas de trabalhos de percepção de padrões utilizando perfis elétricos, não foram encontrados trabalhos sobre a classificação de fluidos. Como foi mencionado na introdução, o motivo é que a classificação de fluidos em óleo, gás ou água exige do geólogo um profundo conhecimento da área de exploração. Outro fato que pode ser observado, é que os trabalhos mais recentes para o problema de classificação de litofácies têm demonstrado a superioridade do classificador SVM frente a classificadores como redes neurais e análise estatística discriminante ((Al-Anazi and Gates, 2010a), (Wong et al., 2005)). Porém, kNN não foi investigado. Por fim, nenhum trabalho relacionado nesta seção realizou a seleção dos perfis elétricos. Essa seleção pode contribuir para aumentar a taxa de acerto do classificador, conforme será mostrado na próxima seção.

### **3. Experimentos e Resultados**

Esta seção descreve os experimentos e os resultados obtidos. Inicialmente, porém, descrevemos os detalhes do protocolo experimental utilizado nos experimentos. Em seguida, é feita uma análise de características através de um método exaustivo para descobrir quais perfis são mais decisivos para a classificação de fluidos. Por fim, são apresentados os resultados da comparação entre os classificadores SVM e kNN.

#### **3.1. Protocolo Experimental**

Os experimentos foram realizados com uma base de dados composta por 283 amostras de perfis de um único poço, sendo que cada amostra representa uma profundidade específica do poço. Foi realizada uma etapa de ajuste de parâmetros, onde no caso de kNN, o melhor valor de  $k$  encontrado foi  $k=1$ . Para o classificador SVM, o parâmetro de regularização  $C$  foi variado de 0.001 a 1000 sem alteração na classificação. O kernel escolhido deste classificador foi o polinomial, sendo que grau do polinômio=2 obteve melhores taxas de acerto. Os dados estão divididos em quatro classes de fluidos: água, gás, petróleo e não reservatório, e foram rotulados por um especialista geofísico. A Tabela 3 mostra um exemplo das amostras da base de dados.

Como pode ser observado na tabela 3, os valores obtidos para cada perfil estão em escalas diferentes. Essa não normalização dos dados pode influenciar negativamente o desempenho dos classificadores, especialmente kNN, que é baseado em medida de distância. Diante desse fato, um processo de normalização dos dados foi realizado. Aksoy and Haralick (2000) apresentam um estudo comparativo entre seis procedimentos de normalização que possibilitam que os dados fiquem dentro

**Tabela 3. Exemplo de amostras da base de dados**

| n°  | Depth     | Cal    | DT      | GR       | ILD     | MSFL    | NPHI    | Pef    | RhOB   |
|-----|-----------|--------|---------|----------|---------|---------|---------|--------|--------|
| 1   | 2435,0471 | 6,7227 | 56,2182 | 106,4375 | 27,6483 | 35,1158 | 10,8416 | 3,168  | 2,6629 |
| 2   | 2435,1995 | 6,7227 | 57,9455 | 146,5    | 25,7482 | 53,9805 | 12,8258 | 3,1602 | 2,6624 |
| 278 | 2477,262  | 6,7099 | 71,9797 | 75,8699  | 0,6001  | 1,6795  | 11,4636 | 2,9233 | 2,5086 |
| 283 | 2478,0239 | 6,582  | 68,1109 | 63,4688  | 5,0932  | 7,0145  | 13,3006 | 2,5742 | 2,4555 |

de um intervalo  $[0,1]$ . Neste trabalho foram utilizados os dois primeiros procedimentos investigados em Aksoy and Haralick (2000). No primeiro procedimento de pré-processamento é feito um escalonamento linear normalizando os dados, sendo que é dado um limite inferior  $l$  e um limite superior  $u$  para uma característica  $x$ ,

$$\tilde{x} = \frac{x - l}{u - l}, \quad (1)$$

resultando em  $\tilde{x}$  dentro de um intervalo  $[0,1]$ .

No segundo procedimento de pré-processamento os dados são padronizados, consistindo em transformar uma característica  $x$  para uma variável aleatória com média zero e variância igual a 1 como:

$$\tilde{x} \cong \frac{\mu - l}{\sigma}, \quad (2)$$

onde  $\mu$  e  $\sigma^2$  são a média da amostra e a variância igual a 1 da característica  $x$ , respectivamente.

A estratégia de validação cruzada conhecida como *leave-one-out* foi utilizada para a divisão dos dados em base de treinamento e teste. Essa estratégia consiste em treinar o classificador com todas as amostras, exceto uma amostra que é usada para o teste. Em seguida, uma nova amostra é selecionada para teste, enquanto as demais são usadas para treinamento. Esse processo é repetido até que todas as amostras sejam usadas como amostra de teste. O resultado da classificação é obtido por meio do cálculo da razão entre o número de acertos e o número total de amostras. *Leave-one-out* foi utilizado devido à pequena quantidade de amostras na base de dados. Os parâmetros de SVM e kNN também foram ajustados.

Os métodos de normalização são comparados na tabela 4, com base na taxa de acerto dos classificadores kNN e SVM. Como pode ser observado na tabela 4, o classificador kNN foi mais beneficiado pela normalização dos dados do que SVM. O método kNN melhorou seu desempenho com ambos os procedimentos de normalização, fato que não foi observado com o classificador SVM, que não alterou a taxa de acerto de 94.34%. Esses resultados preliminares indicam que kNN, com  $k=1$ , superou SVM (polinomial de grau 2) em taxa de acerto. Entretanto, a maioria das classificações incorreta, tanto para SVM quanto kNN, ficou dentro da zona de transição entre água, óleo e gás. É possível que a própria classificação feita pelo especialista esteja incorreta, pois o limite entre essas zonas se confunde. A comparação entre kNN e SVM é discutida com mais detalhes na seção 3.2.

**Tabela 4. Comparação dos classificadores kNN e SVM com e sem normalização.**

| Classificador | Sem Normalização | Normalização Tipo 1 | Normalização Tipo 2 |
|---------------|------------------|---------------------|---------------------|
| SVM           | 94.34%           | 94.34%              | 94.34%              |
| kNN           | 95.75%           | 96.46%              | 96.11%              |

**Tabela 5. Taxas de erro de perfis analisados individualmente com o classificador kNN com k=1.**

| Perfis Utilizados | Cal   | DT    | GR    | <b>ILD</b>   | <b>MSFL</b>  | NPHI | PEF   | RHOB  |
|-------------------|-------|-------|-------|--------------|--------------|------|-------|-------|
| Tx de Erro        | 67.12 | 49.82 | 55.12 | <b>19.43</b> | <b>28.62</b> | 48.4 | 57.24 | 43.46 |

### 3.2. Análise de características

A seleção de características no contexto de classificadores baseados em percepção de padrões tem três objetivos: (1) reduzir os custos de extração de características, (2) aumentar a acurácia da classificação e (3) aumentar a confiança da estimativa de desempenho (Kudo and Sklansky, 2000).

Para descobrir qual ou quais as melhores combinações de perfis para a classificação, testamos todas as combinações possíveis para os 9 perfis, desconsiderando a profundidade uma vez que a mesma varia para cada poço e reserva e tende mais a ser prejudicial do que benéfica para a classificação, quando comparado com mais de um poço. No trabalho de Wong et al. (2005) a profundidade também foi retirada pois o autor trabalhou com reservatórios heterogêneos. Restaram assim 8 perfis (Cal, DT, GR, ILD, MSFL, NPHI, PEF e RHOB) conforme mostrado na Tabela 5, e estes geraram 244 combinações distintas. Combinações onde os perfis eram os mesmos e apenas a ordem era distinta, foram rejeitadas por não apresentarem alteração na taxa de acerto.

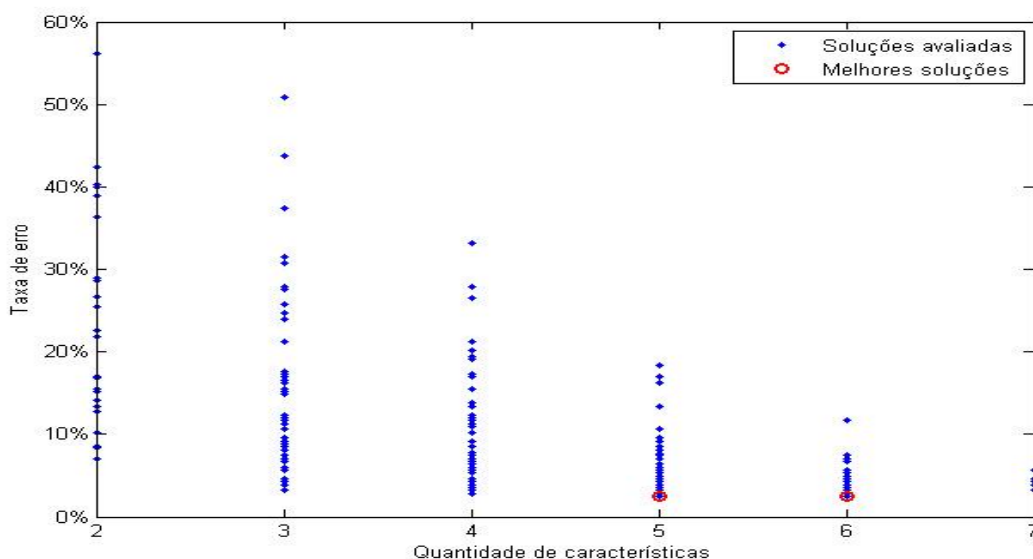
Inicialmente, o impacto individual de cada perfil foi analisado, ou seja, o classificador foi treinado com apenas uma característica por vez. As taxas de erro obtidas estão resumidas na Tabela 5. É possível observar nessa tabela que os perfis ILD e MSFL produziram os melhores resultados individuais, sendo que o classificador obteve taxas de erro iguais a 19.43% e 28.62%, respectivamente para cada perfil. Por outro lado, o perfil CAL foi o menos relevante para a classificação de fluidos, dado que o classificador apresentou o pior resultado (67.12% de taxa de erro) ao ser treinado apenas com esse perfil. Entretanto, conforme esperado, apenas um perfil não fornece informação suficiente para o classificador produzir elevada capacidade de generalização. Logo, a seleção de subconjuntos de perfis busca aumentar o poder de discriminação das informações fornecidas para treinar o método de classificação.

Na tentativa de identificar o melhor subconjunto de características, foram feitos testes exaustivos com todas as combinações de perfis, com o objetivo de aumentar a taxa de acerto e diminuir o número de características. Essa tarefa seria proibitiva caso o número de perfis fosse muito elevado. Os resultados dessa busca são sintetizados na Tabela 6. Pode ser observado que as melhores taxas de acerto foram obtidas com a seguinte combinação de perfis (DT, GR, ILD, MSFL, NPHI, RHOB), tanto com SVM quanto com kNN. Portanto, o número de características foi reduzido, de 08 para 06 perfis, enquanto a taxa de acerto aumentou em 0.34%

**Tabela 6. Melhores Combinações de Perfis.**

| Classif. | Perfis utilizados             | Tx. Acerto | Normalização |
|----------|-------------------------------|------------|--------------|
| SVM      | GR, ILD, MSFL, NPHI, RHOB     | 94.34%     | tipo 2       |
| SVM      | DT, GR, ILD, NPHI, RHOB       | 94.34%     | tipo 2       |
| SVM      | DT, ILD, MSFL, PEF, RHOB      | 92.22%     | tipo 2       |
| SVM      | DT, GR, ILD, MSFL, NPHI, RHOB | 94.69%     | tipo 1       |
| SVM      | DT, GR, ILD, MSFL, NPHI, RHOB | 94.69%     | tipo 2       |
| kNN      | GR, ILD, MSFL, NPHI, RHOB     | 96.81%     | tipo 2       |
| kNN      | DT, GR, ILD, NPHI, RHOB       | 97.17%     | tipo 2       |
| kNN      | DT, ILD, MSFL, PEF, RHOB      | 95.40%     | tipo 2       |
| kNN      | DT, GR, ILD, MSFL, NPHI, RHOB | 97.52%     | tipo 1       |
| kNN      | DT, GR, ILD, MSFL, NPHI, RHOB | 97.52%     | tipo 2       |

com SVM e 1.06% com kNN, confirmando a importância do processo de seleção de características. Outro fator de destaque é que combinações com 05 perfis também apresentaram bons resultados. Os seguintes subconjuntos de cinco perfis produziram os melhores resultados: (DT, GR, ILD, NPHI, e RHOB) com SVM - reduziu o número de características e manteve a taxa de acerto do classificador; e (GR, ILD, MSFL, PEF e RHOB) com kNN que também reduziu o número de características e aumentou a taxa de acerto. Embora as combinações sejam diferentes, os perfis GR, ILD e RHOB estão presentes nas duas combinações. O perfil ILD também foi o perfil mais relevante individualmente. A Figura 2 mostra o comportamento das soluções avaliadas durante o processo de busca.



**Figura 2. Evolução do Método Exaustivo.**

Analisando a matriz de confusão na Tabela 7 e a Figura 1 concluímos que a maioria dos erros ocorreu dentro das zonas de transição entre não reserva e água. Conforme mencionado anteriormente, é possível que os próprios rótulos dos dados nessas zonas de transição estejam incorretos, pois o limite entre essas zonas se confunde.

**Tabela 7. Matriz de confusão.**

|                  | Não Reservatório | Gás | Óleo | Água |
|------------------|------------------|-----|------|------|
| Não Reservatório |                  | 1   |      |      |
| Gás              | 1                |     | 1    |      |
| Óleo             |                  | 1   |      | 2    |
| Água             | 1                |     |      |      |

#### 4. Discussão

Respondemos com sucesso à seguinte pergunta:

- É eficaz o aprendizado dentro de um poço de petróleo para as demais profundidades do mesmo poço?

A resposta é sim, o classificador obtém uma alta taxa de acerto para as demais profundidades de um poço, além de utilizar uma representação reduzida dos dados, obtida por meio do processo de seleção de subconjuntos de perfis. No entanto, devido ao fato da base pertencer a apenas um poço, não foi possível verificar se o aprendizado gerado com dados de um único poço é suficientemente representativo para prever água, gás ou óleo de poços distintos, na mesma reserva e em diferentes reservatórios.

Foram entrevistados 03 geólogos sobre essa questão, sendo dois da Petrobras e 01 de uma empresa de perfilagem dos Estados Unidos. Todos dizem que a classificação de fluidos em água, óleo ou gás, não pode ser feita somente com os dados da perfilagem e que necessitam de um grande conhecimento da reserva pelos geólogos. Isso acontece porque, em muitos casos, os mesmos parâmetros que em uma reserva indicam a presença de óleo, em outra reserva podem indicar a presença de gás ou nenhuma presença de fluidos. Assim, o aprendizado de um poço, dificilmente poderia ser aproveitado para outros poços em reservas diferentes. Porém, para poços vizinhos ou com características similares do reservatório, o aprendizado poderia ser generalizado. Para responder esta pergunta com confiabilidade, serão necessários testes com bases de outros poços da mesma reserva.

#### 5. Conclusão

Neste trabalho foi realizado um estudo de seleção de características para o problema de classificação de fluidos por meio de perfis elétricos. A seleção dos perfis foi realizada através de um método exaustivo em que foram testadas todas as combinações possíveis (excluindo as mesmas combinações em ordens diferentes). Os testes foram feitos com somente 1 perfil até os 8 perfis combinados. Nossos resultados indicaram que 6 perfis (DT, GR, ILD, MSFL, NPHI, RHOB) são as informações mais relevantes tanto para o aprendizado do classificador kNN quanto de SVM, provando assim a importância da escolha das melhores características.

Foi também apresentada uma comparação entre os classificadores kNN e SVM. Em nossos experimentos realizados com uma pequena base de 283 amostras, kNN obteve melhores taxas de acerto. A comparação foi feita com dados de apenas um poço por meio da técnica *leave-one-out* para garantirmos uma melhor confiabilidade dos dados apresentados. No entanto, para a classificação de fluidos, é possível que o aprendizado obtido com dados de um poço somente seja relevante

para poços vizinhos ou poços com as mesmas características da base de treino. Ao se trabalhar em uma nova reserva, será necessário que o geólogo analise o novo poço para que poços vizinhos possam ser classificados automaticamente. Nossa proposta como trabalho futuro é incorporar à base dados de outros poços da mesma reserva e propor um sistema em que o geólogo não precise classificar todos os poços e sim, apenas classifique um poço para que os poços vizinhos a este possam ser classificados automaticamente.

## Referências

- Aksoy, S. and Haralick, R. M. (2000). Effects of feature normalization on image retrieval. *IAPR International Conference on Pattern Recognition*, 22:563 – 582.
- Al-Anazi, A. and Gates, I. (2010a). On the capability of support vector machines to classify lithology from well logs. *Natural Resources Research*, 19:125 – 139.
- Al-Anazi, A. and Gates, I. (2010b). A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Engineering Geology*, 114(3–4):267 – 277.
- An-nan, J. and Lu, J. (2009). Studying the lithology identification method from well logs based on de-svm. In *Control and Decision Conference, 2009. CCDC '09. Chinese*, pages 2314 –2318.
- Cunha, E. S. (2002). Identificação de litofácies de poços de petróleo utilizando um método baseado em redes neurais artificiais. Master's thesis, Universidade Federal de Campina Grande.
- Flexa, R. T., Andrade, A., and Carrasquilla, A. (2004). Identificação de litotipos nos perfis de poço do campo de namorado (bacia de campos, brasil) e do lago de maracaibo (venezuela) aplicando estatística multivariada. *Revista Brasileira de Geociências*, 34:571–578.
- Garcia, E. N. (2008). *A PETROBRAS NA AMAZÔNIA: a história da pesquisa de combustíveis fósseis*. Norma Editora, Manaus, 1 edition.
- Kudo, M. and Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1):25 – 41.
- Petrobras (2011). Conheça nosso plano de negócios 2011-2015.
- Thomas, J. E., Triggia, A. A., and Correia, C. A. (2001). *Fundamentos de Engenharia de Petróleo*. Interciência.
- Wong, K., Ong, Y., Gedeon, T., and Fung, C. (2005). Reservoir characterization using support vector machines. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, volume 2, pages 354 –359.