

Matheus Henrique Schaly

# Aplicação de Aprendizado de Máquina na Classificação de Litofácies

Florianópolis

2020

Matheus Henrique Schaly

# **Aplicação de Aprendizado de Máquina na Classificação de Litofácies**

Proposta de monografia submetida ao Programa de Graduação em Ciência da Computação para a obtenção do Grau de Bacharel.

Universidade Federal de Santa Catarina  
Departamento de Informática e Estatística  
Ciência da Computação

Orientador: Prof. Dr. Mauro Roisenberg

Florianópolis  
2020

## FOLHA DE APROVAÇÃO DE PROPOSTA DE TCC

|                                       |                                                           |
|---------------------------------------|-----------------------------------------------------------|
| <b>Acadêmico(s)</b>                   |                                                           |
| <b>Título do trabalho (subtítulo)</b> |                                                           |
| <b>Curso</b>                          | Ciência da Computação ou Sistemas de Informação /INE/UFSC |
| <b>Área de Concentração</b>           | (seguindo a classificação documentada no anexo 1)         |

### Instruções para preenchimento pelo **ORIENTADOR DO TRABALHO**:

- Para cada critério avaliado, assinale um X na coluna SIM apenas se considerado aprovado. Caso contrário, indique as alterações necessárias na coluna Observação.

| Critérios                                                                                                                                                                                                                                                                                        | Aprovado |         |     |               | Observação |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|---------|-----|---------------|------------|
|                                                                                                                                                                                                                                                                                                  | Sim      | Parcial | Não | Não se aplica |            |
| 1. O trabalho é adequado para um TCC no CCO/SIN (relevância / abrangência)?                                                                                                                                                                                                                      |          |         |     |               |            |
| 2. O título do trabalho é adequado?                                                                                                                                                                                                                                                              |          |         |     |               |            |
| 3. O tema de pesquisa está claramente descrito?                                                                                                                                                                                                                                                  |          |         |     |               |            |
| 4. O problema/hipóteses de pesquisa do trabalho está claramente identificado?                                                                                                                                                                                                                    |          |         |     |               |            |
| 5. A relevância da pesquisa é justificada?                                                                                                                                                                                                                                                       |          |         |     |               |            |
| 6. Os objetivos descrevem completa e claramente o que se pretende alcançar neste trabalho?                                                                                                                                                                                                       |          |         |     |               |            |
| 7. É definido o método a ser adotado no trabalho? O método condiz com os objetivos e é adequado para um TCC?                                                                                                                                                                                     |          |         |     |               |            |
| 8. Foi definido um cronograma coerente com o método definido (indicando todas as atividades) e com as datas das entregas (p.ex. Projeto I, II, Defesa)?                                                                                                                                          |          |         |     |               |            |
| 9. Foram identificados custos relativos à execução deste trabalho (se houver)? Haverá financiamento para estes custos?                                                                                                                                                                           |          |         |     |               |            |
| 10. Foram identificados todos os envolvidos neste trabalho?                                                                                                                                                                                                                                      |          |         |     |               |            |
| 11. As formas de comunicação foram definidas (ex: horários para orientação)?                                                                                                                                                                                                                     |          |         |     |               |            |
| 12. Riscos potenciais que podem causar desvios do plano foram identificados?                                                                                                                                                                                                                     |          |         |     |               |            |
| 13. Caso o TCC envolva a produção de um software ou outro tipo de produto e seja desenvolvido também como uma atividade realizada numa empresa ou laboratório, consta da proposta uma declaração (Anexo 3) de ciência e concordância com a entrega do código fonte e/ou documentação produzidos? |          |         |     |               |            |

| Avaliação                    | <input type="checkbox"/> Aprovado |        | <input type="checkbox"/> Não Aprovado |
|------------------------------|-----------------------------------|--------|---------------------------------------|
| <b>Professor Responsável</b> | (Nome)                            | (Data) | (Assinatura)                          |
| <b>Orientador externo</b>    | (Nome)                            | (Data) | (Assinatura)                          |

# Resumo

Aprendizado de máquina (AM) vem se tornando uma ferramenta cada vez mais importante em vários campos da ciência, neste trabalho aplicaremos AM no ramo das geociências. O objetivo do trabalho consiste em propor um modelo eficaz de AM, incluindo a parte de manipulação dos dados, para a classificação de litofácies em poços geológicos. Acreditamos que a sequência de padrões sedimentares possa ajudar no processo de classificação, e para isso poderia ser utilizado uma versão modificada de uma rede neural recorrente (RNN). A classificação acurada de litofácies é de grande importância para obter informações geológicas úteis para a exploração e produção de hidrocarbonetos. A classificação automática de litofácies torna o processo de estudo da litologia dos poços mais rápido e menos oneroso. A classificação de litofácies é realizada estudando as propriedades litológicas das rochas encontradas em poços, que são características dos sedimentos atuais acumulados em determinadas condições físicas e geográficas. As propriedades litológicas podem incluir raio gama, resistividade, efeito fotoelétrico, densidade de porosidade de nêutrons, porosidade de densidade de nêutrons média, entre outras. Dado um banco de dados contendo as características e a classificação das litofácies, é esperado que o modelo proposto consiga, de maneira eficaz, realizar automaticamente a classificação de tais litofácies. A eficácia do método será medida através das métricas de classificação, como acurácia, precisão, *recall* e *F1-score*.

**Palavras-chave:** aprendizado de máquina, classificação automática de litofácies.

# Sumário

|     |                                 |    |
|-----|---------------------------------|----|
| 1   | INTRODUÇÃO . . . . .            | 5  |
| 2   | OBJETIVOS . . . . .             | 7  |
| 2.1 | Objetivos específicos . . . . . | 7  |
| 2.2 | Escopo do trabalho . . . . .    | 7  |
| 3   | MÉTODO DE PESQUISA . . . . .    | 8  |
| 4   | PLANEJAMENTO . . . . .          | 9  |
| 4.1 | Cronograma . . . . .            | 9  |
| 4.2 | Custos . . . . .                | 9  |
| 4.3 | Recursos humanos . . . . .      | 9  |
| 4.4 | Comunicação . . . . .           | 9  |
| 4.5 | Riscos . . . . .                | 10 |
|     | REFERÊNCIAS . . . . .           | 11 |

# 1 Introdução

Há diversas definições existentes para o termo fácies. Definimos facies como qualquer parte restrita não comparável de uma unidade estratigráfica projetada que exibe caráter significativamente diferentes daqueles de outras partes da unidade (MOORE, 1949). Biofácies são fácies identificadas por características paleontológicas (conteúdo fóssil) sem levar em conta o caráter litológico. Litofácies são fácies identificadas com base em características litológicas (BOGGS, 2001). Usaremos litofácies como base de dados no presente trabalho.

A classificação de litofácies consiste em atribuir uma classe de rocha a uma amostra específica com base nas características medidas. A fonte ideal para classificação de litofácies são amostras de núcleo de rochas extraídas de poços. No entanto, devido aos custos associados, nem sempre as amostras de núcleo podem ser obtidas. Além disso, o método convencional é um processo tedioso e demorado, pois consiste em classificar litofácies manualmente por intérpretes humanos. Portanto, um método para classificar fácies a partir de medidas indiretas (por exemplo, gerar perfis utilizando cabo de aço) é necessário. Várias abordagens distintas para a questão da classificação de fácies utilizando dados de poços já foram propostas (BESTAGINI; LIPARI; TUBARO, 2017). Neste trabalho será investigado reservatórios de hidrocarbonetos que, dependendo do banco de dados utilizado, podem possuir como atributos: raios gama, resistividade, efeito fotoelétrico, densidade, entre outros. As classes das litofácies, também dependentes do banco de dados, podem ser: carvão, argila, silte, areia fina, areia média, areia grossa, arenito limpo, arenito sujo, dolomita, entre outras.

Nos últimos anos, AM se tornou uma ferramenta interdisciplinar cada vez mais importante, que avançou vários campos da ciência, como biologia, química, medica e farmacologia. Especificamente, o método de rede neural profunda (DNN) encontrou ampla aplicação. Enquanto a geociência foi mais lenta na adoção, a bibliometria mostra adoção do aprendizado profundo em todos os aspectos da geociência (DRAMSCH, 2020).

Aprendizado de máquina é profundamente enraizado em estatísticas aplicadas, criando modelos computacionais que utilizam de inferência e reconhecimento de padrões em vez de conjuntos explícitos de regras (DRAMSCH, 2020). Aprendizado de máquina é aqui definido como sendo o campo de estudo que fornece aos computadores a capacidade de aprender sem serem explicitamente programados (Samuel, 1959). Aprendizagem supervisionada consiste na tarefa de um algoritmo de AM em aprender uma função que mapeia uma entrada para uma saída com base em exemplos de pares de entrada e saída (RUSSELL et al., 2010). Uma função é inferida a partir de dados de treinamento rotulados que consistem em um conjunto de exemplos de treinamento (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

Aprendizado profundo é uma forma de AM que permite que os computadores aprendam com a experiência e entendam o mundo em termos de uma hierarquia de conceitos. A hierarquia de conceitos permite que o computador aprenda conceitos complicados construindo-os a partir de outros mais simples (GOODFELLOW; BENGIO; COURVILLE, 2016).

Recentemente, as técnicas de aprendizado profundo foram desenvolvidas e amplamente adotadas para extrair informações de vários tipos de dados. Considerando as diferentes características dos dados de entrada, existem vários tipos de arquiteturas para aprendizado profundo, como a RNN, rede neural convolucional (CNN), e DNN. Geralmente a CNN e a DNN não podem lidar com as informações temporais de entrada de dados. Portanto, em áreas de pesquisa que contêm dados sequencias, como texto, áudio e vídeo, RNNs são dominantes. Contudo, RNNs são incapazes de aprender as informações relevantes dos dados de entrada quando a lacuna de entrada é grande. Ao introduzir funções de portão na estrutura da célula de uma RNN a *long short-term memory* (LSTM) poderia lidar bem com o problema das dependências de longo prazo. Desde a introdução de RNN quase todos os resultados interessantes baseados em RNNs foram alcançados pela LSTM. A LSTM se tornou o foco do aprendizado profundo (Yu et al., 2019).

O problema de classificação automática de litofácies deve ser explorado a fim de diminuir os custos envolvidos na classificação manual de litofácies. Nossa solução ao problema de classificação de litofácies utilizará um algoritmo de aprendizado de máquina supervisionado. Acreditamos que a sequência de padrões sedimentares possa ajudar no processo de classificação. Portanto, podemos, mais especificamente, criar uma nova topologia de RNN ou LSTM que venha a considerar este aspecto sequencial do nosso banco de dados. Além disso, criaremos um pipeline para a manipulação dos dados para organizar e melhorar os dados de entrada ao modelo.

## 2 Objetivos

O objetivo do trabalho proposto consiste em desenvolver uma solução com melhor desempenho em relação ao estado da arte no contexto de classificação de litofácies a fim de reduzir os custos envolvidos na classificação manual de litofácies.

### 2.1 Objetivos específicos

Os objetivos específicos são:

- a) aprimorar ou reestruturar um algoritmo de AM já existente a fim torná-lo um algoritmo com bom desempenho para o problema de classificação supervisionada de litofácies;
- b) construir um pipeline para a manipulação dos dados que serão utilizados como input para o algoritmo proposto;

### 2.2 Escopo do trabalho

O critério de aceitação consiste em construir um algoritmo de AM, junto ao seu pipeline de manipulação de dados, que possua métricas de desempenho próximas aos do estado da arte. O critério de desempenho será medido por métricas de classificação, como acurácia, precisão, *recall* e *F1-score*.

O projeto será entregue dentro do prazo de acordo com o cronograma descrito na tabela 1.

O trabalho a princípio não visa a criação de um software para a utilização do público geral.



### 3 Método de pesquisa

Os procedimentos metodológicos para a realização da pesquisa serão os seguintes:

- a) estudo da fundamentação teórica através de livros e materiais online. Revisão do estado da arte por meio de artigos científicos;
- b) desenvolvimento da solução utilizando uma linguagem de programação, como Python, R ou Octave, com o auxílio de suas bibliotecas auxiliares voltadas à matemática e AM. Ao mesmo tempo também será utilizado um banco de dados contendo a classificação e as características de um conjunto de litofácies.
- c) comparação entre o desempenho alcançado pela solução proposta e o desempenho alcançado pelo estado da arte. O desempenho será medido por métricas de classificação, como acurácia, precisão, *recall* e *F1-score*;
- d) elaboração da monografia do trabalho de conclusão de curso. Incluindo entrega do relatório parcial em Projeto I, entrega final e defesa pública.

## 4 Planejamento

### 4.1 Cronograma

O cronograma é descrito pela tabela 1.

Tabela 1 – Planejamento das etapas do trabalho de conclusão de curso

| Etapas                                    | Meses |     |      |     |     |     |     |     |     |     |     |     |     |     |
|-------------------------------------------|-------|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                                           | 2020  |     | 2021 |     |     |     |     |     |     |     |     |     |     |     |
|                                           | nov   | dez | jan  | fev | mar | abr | mai | jun | jul | ago | set | out | nov | dez |
| Entrega da proposta completa              | 15/11 |     |      |     |     |     |     |     |     |     |     |     |     |     |
| Estudo da fundamentação teórica           | X     | X   | X    |     |     |     |     |     |     |     |     |     |     |     |
| Revisão do estado da arte                 |       |     | X    | X   | X   | X   |     |     |     |     |     |     |     |     |
| Desenvolvimento da solução                |       |     |      | X   | X   | X   | X   | X   | X   | X   |     |     |     |     |
| Desenvolvimento do relatório de Projeto I |       |     |      |     |     | X   | X   |     |     |     |     |     |     |     |
| Entrega do relatório de Projeto I         |       |     |      |     |     |     | X   |     |     |     |     |     |     |     |
| Redação do rascunho do TCC                |       |     |      |     |     |     |     | X   | X   | X   | X   | X   |     |     |
| Entrega do rascunho do TCC                |       |     |      |     |     |     |     |     |     |     |     | X   |     |     |
| Preparação da defesa pública              |       |     |      |     |     |     |     |     |     |     |     | X   | X   |     |
| Defesa pública                            |       |     |      |     |     |     |     |     |     |     |     |     | X   |     |
| Ajustes no relatório final do TCC         |       |     |      |     |     |     |     |     |     |     |     |     | X   | X   |

### 4.2 Custos

Não haverá custos envolvidos no desenvolvimento do projeto.

### 4.3 Recursos humanos

O tabela 2 indica quem fará parte no projeto.

Tabela 2 – Envolvidos no projeto

| Nome                    | Função      |
|-------------------------|-------------|
| Matheus Henrique Schaly | Autor       |
| Mauro Roisenberg        | Responsável |
| Mauro Roisenberg        | Orientador  |

### 4.4 Comunicação

A comunicação será realizada por meio de reuniões ocasionais entre o autor e orientador.

## 4.5 Riscos

Tabela 3 – Riscos do projeto

| Risco                        | Probabilidade | Impacto | Prioridade | Estratégia de resposta                             | Ações de prevenção                                                              |
|------------------------------|---------------|---------|------------|----------------------------------------------------|---------------------------------------------------------------------------------|
| Perda do código ou dos dados | baixa         | alto    | alta       | Recuperar versões anteriores                       | Utilizar ferramentas de controle de versionamento e realizar backups constantes |
| Alteração do cronograma      | média         | médio   | alta       | Redefinição do foco do trabalho ou mudança do tema | Monitoramento contínuo das informações obtidas com superiores imediatos         |

## Referências

BESTAGINI, P.; LIPARI, V.; TUBARO, S. A machine learning approach to facies classification using well logs. In: . [S.l.: s.n.], 2017. p. 2137–2142.

BOGGS, S. *Principles of Sedimentology and Stratigraphy*. Prentice Hall, 2001. ISBN 9780130996961. Disponível em: <<https://books.google.com.br/books?id=L4LBQgAACAAJ>>.

DRAMSCH, J. S. Chapter one - 70 years of machine learning in geoscience in review. In: MOSELEY, B.; KRISCHER, L. (Ed.). *Machine Learning in Geosciences*. Elsevier, 2020, (Advances in Geophysics, v. 61). p. 1 – 55. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0065268720300054>>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. (Adaptive Computation and Machine Learning series). ISBN 9780262035613. Disponível em: <<https://books.google.com.br/books?id=Np9SDQAAQBAJ>>.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning*. MIT Press, 2012. (Adaptive Computation and Machine Learning series). ISBN 9780262018258. Disponível em: <<https://books.google.com.br/books?id=maz6AQAAQBAJ>>.

MOORE, R. C. Meaning of facies. *Geol. Soc. America Mem*, v. 39, p. 1–34, 1949.

RUSSELL, S. et al. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2010. (Prentice Hall series in artificial intelligence). ISBN 9780136042594. Disponível em: <<https://books.google.com.br/books?id=8jZBksh-bUMC>>.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, v. 3, n. 3, p. 210–229, 1959.

Yu, Y. et al. A review of recurrent neural networks: Lstm cells and network architectures. *Neural Computation*, v. 31, n. 7, p. 1235–1270, 2019.