Estimativa do perfil de velocidade cisalhante (VS) a partir de métodos de aprendizado de máquina e rede neural na Formação Barra Velha, Campo de Búzios, Bacia de Santos

José Augusto Vitorino Dias[1], Mário Martins Ramos[1], Antonio Fernando Menezes Freire[1,2], Wagner Moreira Lupinacci[1,2]

[1]Grupo de Interpretação Exploratória e Caracterização de Reservatórios (GIECAR/UFF)
[2]Instituto de Ciência e Tecnologia de Geofísica do Petróleo (INCT-GP)

Copyright 2024, SBGf - Sociedade Brasileira de Geofísica

Este texto foi preparado para a apresentação no IX Simpósio Brasileiro de Geofísica, Salvador, 08 a 10 de outubro de 2024. Seu conteúdo foi revisado pelo Comitê Técnico do X SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Resumo

O objetivo deste trabalho é estimar a velocidade cisalhante em poços do Campo de Búzios, na Bacia de Santos, a partir de perfis de poços. Para isto, é realizada uma comparação de diferentes métodos de aprendizado de máquina. Como curvas de entrada, foi utilizado velocidade compressional (VP), densidade (RHOB) e porosidade total (NMRP1). A análise foi conduzida a partir da curva de VS medida, com o objetivo de comparar precisão e erros utilizando como métrica o coeficiente de determinação (R^2), além da curva de VP por ajuste de curva. Os métodos apresentaram resultados com média acima de 70% de acerto. A partir de análises das métricas e visualmente, observamos que o método MultiLayer Perceptron obteve o melhor resultado.

Introdução

O perfil de velocidade cisalhante (VS) é importante para estudos de AVO (amplitude versus offset), inversão elástica e análises 4D. Essas análises são importantes na caracterização de reservatórios, pois permitem uma melhor avaliação de litologias, identificação do tipo de fluido presente na formação e maior precisão nas modelagens 3D de propriedades petrofísicas.

O perfil VS nem sempre está disponível, principalmente, quando comparado ao perfil de velocidade compressional (VP). Então, têm-se utilizado diferentes métodos para a estimativa do perfil VS, desde abordagens empíricas (Greenberg e Castagna, 1992) até o uso de regressão por meio de técnicas de machine learning (Anemangely et al., 2019).

Foi demonstrado por Han et al. (1986) que a velocidade compressional varia de acordo com a quantidade de argila presente no reservatório, sendo que quanto maior a quantidade de argila, menor a velocidade compressional. Isso possibilitou o desenvolvimento de fórmulas empíricas para calcular as velocidades de cisalhamento (VS) e compressional (VP) para diferentes pressões do reservatório, utilizando o teor de argila e a porosidade como variáveis.

Um método de estimativa da VS a partir da VP e do conhecimento dos constituintes da rocha foi proposto por Greenberg e Castagna (1992). Ao determinar os volumes, os valores de VP para cada mineral constituinte da rocha, a porosidade e os volumes de fluidos presentes é possível obter a VS. Contudo, é difícil obter perfis de volumes de minerais constituintes das rochas (perfis de mineralogia), o que impacta diretamente na precisão da estimativa de VS.

A aplicação de técnicas de machine learning, permite que possa ser utilizado mais perfis para a previsão de VS, o que pode superar o problema de não conseguirmos ter uma boa estimativa de mineralogia ao longo do intervalo de interesse. Anemangely et al. (2019) aplicando o método LSSVM, Algoritmo de Otimização Cuco, Algoritmo PSO e Algoritmo Genético alcançou resultados com um altíssimo coeficiente de determinação (r^2) acima de 90%.

SOUZA, G. (2024) utilizou métodos baseados em árvore de decisões para predizer velocidade cisalhante e densidade. O melhor algoritmo, do trabalho, foi o *CatBoost Regressor* com R-quadrado (R^2) em 91% na predição e 90% na validação cruzada.

Neste estudo, o foco é a estimativa do perfil de velocidade cisalhante (VS) na Formação Barra Velha, localizada na Bacia de Santos. Essa formação é composta por depósitos carbonáticos que foram depositados após o processo de rifteamento da bacia (Moreira et al., 2007). É importante destacar que a Formação Barra Velha está situada dentro do pré-sal, uma área de significativo interesse para a indústria petrolífera brasileira.

Metodologia/ Problema Investigado

Os métodos de aprendizado de máquina utilizados foram: os *based-tree*, *XGBoost*, *CatBoost* e *LightGBM* além do método de rede neural MultiLayer Perceptron (MLP).
Os métodos *based-tree* são conhecidos por sua interpretabilidade e capacidade de lidar com dados categóricos, enquanto a *MultiLayer Perceptron* são eficazes na classificação e regressões de dados em espaços de alta dimensão.

De acordo com Quinlan (1986), a Árvore de Decisão é um método de aprendizado de máquina menos complexo em comparação com outros, mas ainda assim capaz de lidar com problemas de alta complexidade. Árvores de decisão podem resolver problemas de predileção de forma simples e compreensível, pois permitem uma visualização simplificada da classificação, sem a necessidade de um grande conhecimento matemático. O método de Árvore de Decisão teve seu início na década de 1960, desenvolvido por Hunt e Szymanski com o objetivo de modelar o processo de aprendizagem humano.

Os algoritmos de Boosting são classificadores que começaram a ser desenvolvidos por Leo Breiman. Estes algoritmos enfatizam o treinamento em amostras que foram erroneamente classificadas, atribuindo uma ponderação diretamente associada ao erro (Sammut e Webb, 2010). Dentro dessa família existem os classificadores *CatBoost*, *XGBoost*, *LightGBM*..

A Rede Neural Artificial (RNA) é um dos algoritmos de aprendizado de máquina mais amplamente utilizados, tanto para regressão linear quanto não linear. Inspirado no sistema de neurônios do cérebro humano (Russell e Norvig, 2002), o modelo de RNA consiste em três grupos principais de camadas: a camada de entrada, uma ou várias camadas ocultas, e a camada de saída. Cada camada é composta por uma matriz de dados unidimensional que contém diversos nós, também conhecidos como neurônios ou perceptrons.

Os algoritmos foram desenvolvidos em Python, aproveitando uma variedade de bibliotecas para facilitar o processo de desenvolvimento e análise de dados. As principais bibliotecas empregadas incluem o NumPy, que oferece recursos para manipulação eficiente de arrays e cálculos numéricos, o Matplotlib, utilizado para visualização de gráficos e representação dos resultados, e o *Scikit-learn* (Pedregosa, F, 2011), reconhecido por sua robustez em aprendizado de máquina e mineração de dados.

Neste trabalho, utilizamos dados de 8 poços fornecidos pela Petrobrás. A metodologia aplicada para treino e teste consistiu em 6 poços para treinamento dos modelos e 2 para teste.

Primeiro, foi realizadas visualizações estatísticas de Correlação de Pearson, que mede o grau de correlação entre duas variáveis e de Q-Q, que compara os quantis da distribuição resultando se ele segue uma distribuição normal e onde ele se afasta da simetria e um gráfico de distribuições bivariadas pareadas no conjunto de dados. Esse passo foi feito a partir de uma concatenação dos 8 poços.

Com base nessas visualizações, identificamos as curvas mais correlacionadas com a curva alvo (VS) (fig. 1 e 2), destacando-se VP, RHOB (densidade) e NMRP1 (porosidade total). É importante destacar a relevância das curvas selecionadas para o modelo de regressão. Por exemplo, a densidade (RHOB) e as curvas de porosidade são cruciais para prever VS, fornecendo informações essenciais sobre a composição da rocha e a porosidade.

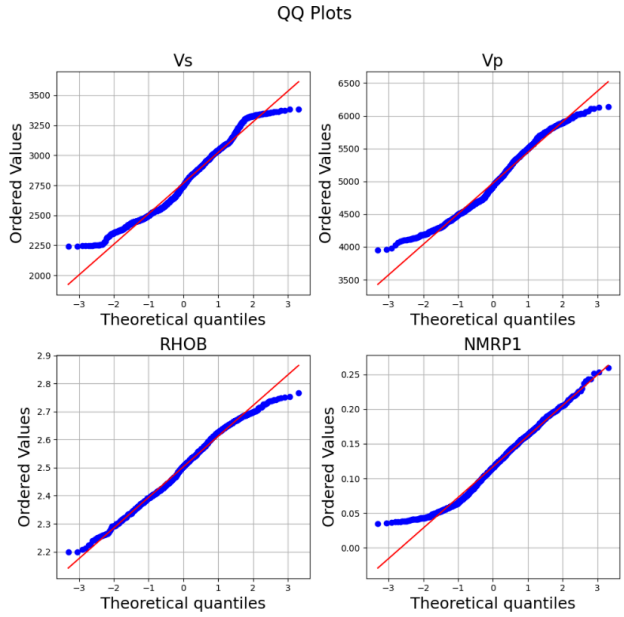

Figura 1: Q-Q Plot

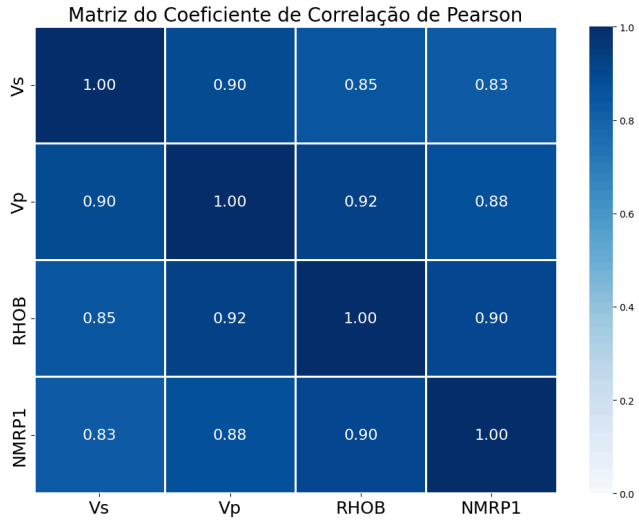

Figura 2: Coeficiente de Pearson

Para a validação dos modelos, foi utilizado o coeficiente de determinação (R^2) como métrica. Esse coeficiente mede a proporção da variância dos dados que é explicada pelo modelo. O valor de R^2 varia de 0 a 1, sendo 0 quando o modelo não explica nenhuma variância dos dados e 1 quando o modelo explica toda a variância dos dados. A fórmula de R^2 está expressa na equação 1, onde y são os valores reais e ŷ são os previsto.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Equação 1: Coeficiente de Determinação

Além do R², foi calculado a VS dos poços de testes a partir do DTS, pois era a curva que tínhamos disponível, para servir como comparação de dispersão. A velocidade cisalhante é calculada a partir da curva de DTS utilizando a equação 2, onde 304800 é um fator de conversão entre m/s e µs/ft.

$$VS = \frac{1}{DTS} \times 304800$$

Equação 2: Cálculo da VS a partir do DTS

Foi feito, também, uma regressão linear para achar valores de VS a partir da curva de VS Para encontrar os coeficientes foi utilizado o método de otimização numérico Algoritmo Broyden-Fletcher-Goldfarb-Shanno (BFGS). O BFGS determina a direção da descida pré-condicionando o gradiente com informações de curvatura,isso é feito melhorando gradualmente uma aproximação à matriz hessiana da função de perda, obtida apenas a partir de avaliações de gradiente (ou avaliações de gradiente aproximadas) através de um método secante generalizado (Dennis, J. E., Jr., 1983). Seguindo a equação 3, onde A é o intercepto e B inclinação da regressão foram dados pelo método de BFGS e o VP é do poço respectivo.

$$VS = A + B * VP$$

Equação 3: Cálculo de VS por regressão linear

Esse resultado foi feito para os dois poços de teste, separadamente.

Resultados

Ao comparar os gráficos de dispersão e R², no poço Búzios 4 (fig.3), foi observado que o método de MultiLayer Perceptron apresentou um desempenho superior para a predição de VS.

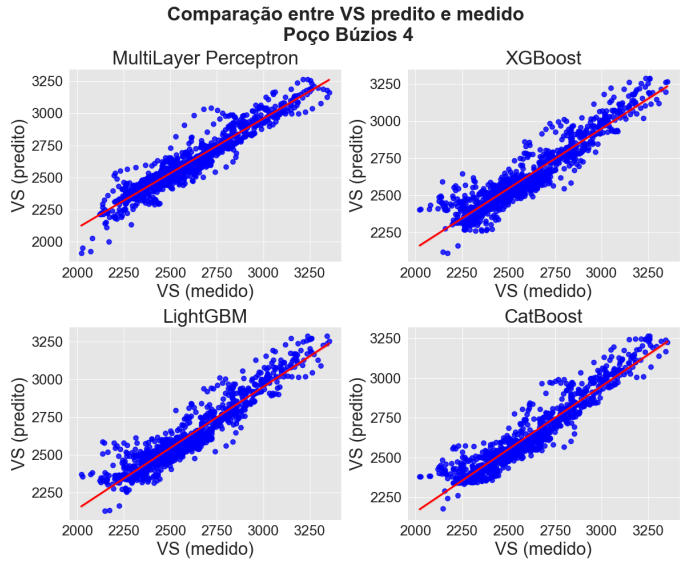

Figura 3: Gráfico de dispersão do perfil de VS estimado no poço Búzios 4.

Para o poço BRSA 1195, tivemos que o método XGBoost teve um melhor desempenho, com R² acima de 80%, os outros métodos ficaram com R² acima de 70%.

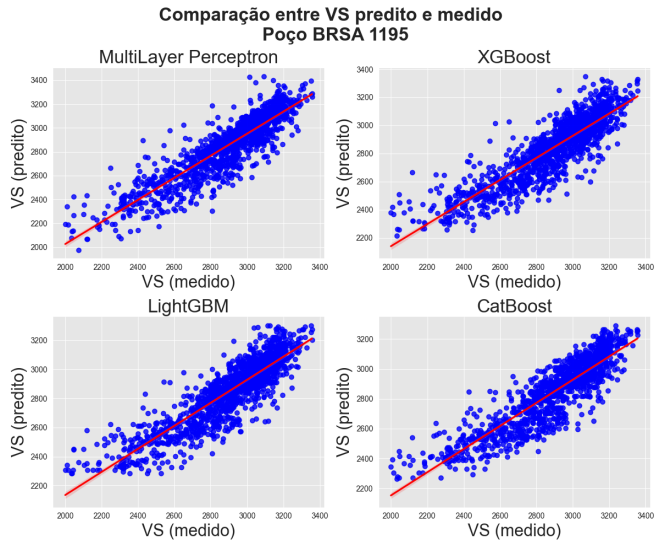

Figura 4: Gráfico de dispersão do perfil de VS estimado no poço Búzios 8.

Na análise visual do poço Búzios 4 (fig. 5), percebemos que o MLP performou bem em todo poço, seguindo as mesmas tendências do VS medido e do método empírico. Os outros métodos erraram mais no começo do poço, entre 5360 metros a 5400 metros. Nessa altura existe presença de rochas sedimentares e brechas, o que pode ter interferido no resultado.

No poço BRSA 1195 (fig.6), no geral, todos os métodos tiveram uma boa performance no geral, porém a rede neural ainda com resultado melhor.

Conclusões

Os resultados deste estudo, fundamentados em pesquisas anteriores, demonstraram que a previsão de VS utilizando métodos de inteligência artificial alcança resultados significativamente positivos. Além de boas métricas, a análise visual das curvas previstas em comparação com aquelas calculadas a partir do DTS revela um erro pequeno, destacando ainda mais a eficácia das redes neurais. Esses achados sugerem que o modelo desenvolvido não apenas oferece resultados confiáveis, mas também se ajusta de forma precisa aos dados disponíveis.

Para futuros trabalhos, é preciso expandir o conjunto de dados disponíveis, pois isso pode proporcionar uma base mais robusta para treinar e avaliar modelos de previsão de VS. Além disso, buscar uma precisão ainda maior no ajuste de hiperparâmetros dos modelos é essencial para garantir resultados ainda mais confiáveis e generalizáveis.

Agradecimentos

Os autores deste trabalho expressam gratidão à Petrobras e ao Instituto Nacional de Ciência e Tecnologia - Geofísica do Petróleo (INCT - GP) pelo apoio e incentivo fornecidos, fundamentais para a realização deste estudo.

Referências

Almeida, F. F. M. de, 1969. Diferenciação tectônica da Plataforma Brasileira (Tectonic differentiation of the Brazilian Platform). Proceedings 33rd. Brazilian Geological Congress, Salvador (BA), Brazil, vol. 1: 29-46.

Arora et al. (06 co-authors), 1999. 2D geoelectrical model for the Parnaíba conductivity anomaly, north-northeast Brazil and its tectonic implications. Tectonophysics, 302: 57-69.

Brocher, Thomas M., 2005. Empirical Relations between Elastic Wavespeeds and Density in the Earth's Crust. In: Bulletin of the Seismological Society of America.

Bott, M. H. P., 1982. Stress based tectonic mechanisms at passive continental margins. In: Scrutton, R. A. (Ed.), Dynamics of Passive Margins. AGU Geodynamic Series, Washington (DC), USA, vol. 6: 147-153.

Dennis, J. E., Jr.; SCHNABEL, R. B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, 1983.

Chandler, V. W., Kiski, H. S., Hinze, W. J. & Braile, L. W., 1981. Analysis of multisource gravity and magnetic anomaly data sets by moving-window applications of Poisson's theorem. Geophysics, 46: 30-39.

G. Souza. 2024. Shear Velocity and Density Prediction: A Machine Learning Approach in Subsalt Santos Basin.

Figueiredo, M. C. H., 1980. Geochemistry of high-grade metamorphic rocks, Northeastern Bahia, Brazil. Ph. D. Thesis, Department of Geology of the University of Western Ontario, Canada.

Pedregosa, F et.al, 2011. Scikit-learn: Machine Learning in {P}ython. Journal of Machine Learning Research

Turner, F. J. & Werhoogen, J., 1960. Igneous and metamorphic petrology. 2nd ed., McGraw Hill, New York, USA.

Russell, S.J. and Norvig, P. (2002) Artificial Intelligence: A Modern Approach. Second Edition, Prentice Hall, Englewood Cliffs.

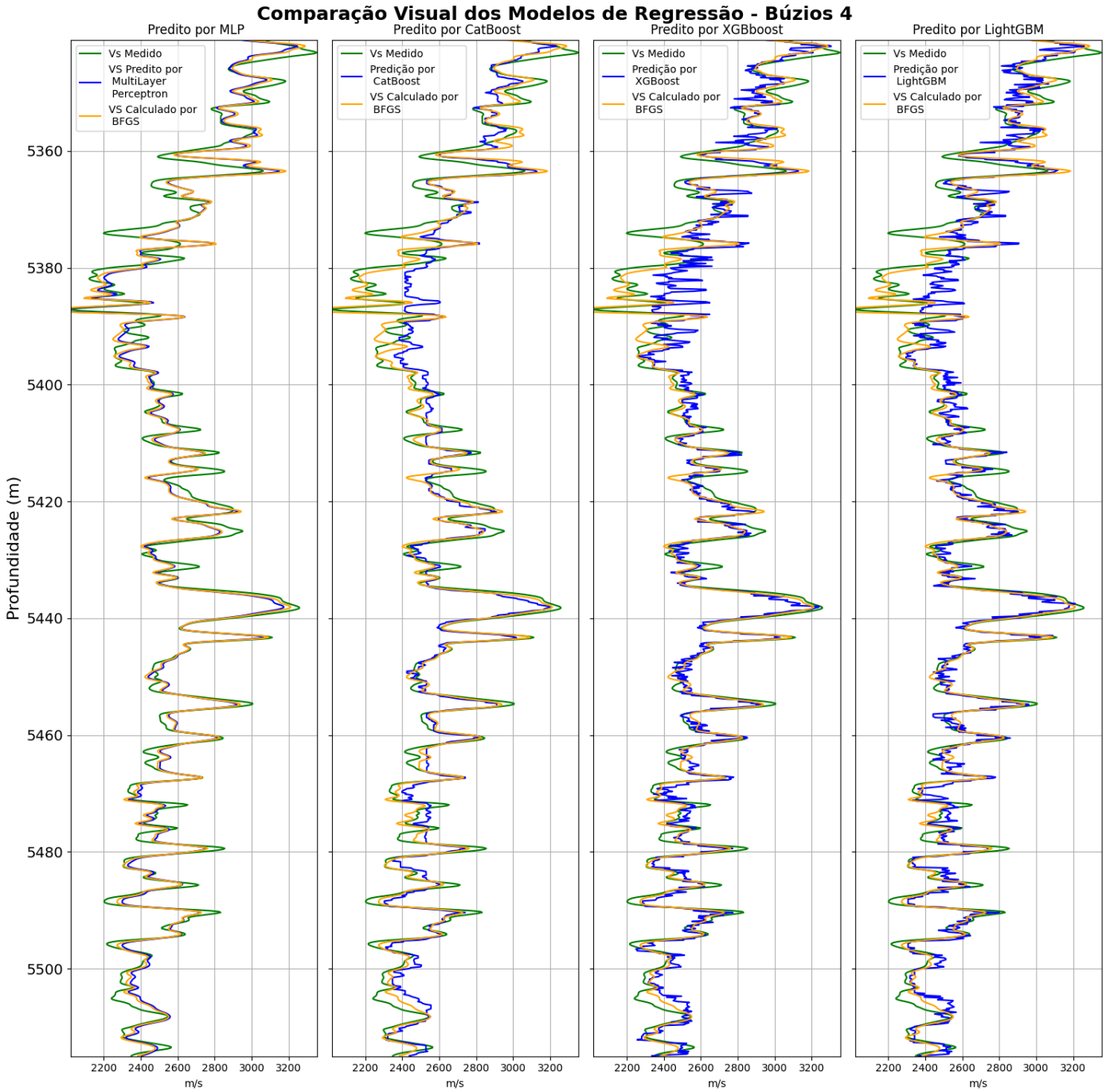

Figura 5: *Comparação visual entre os modelos de aprendizado de máquinas com o VS medido no poço Búzios 4*

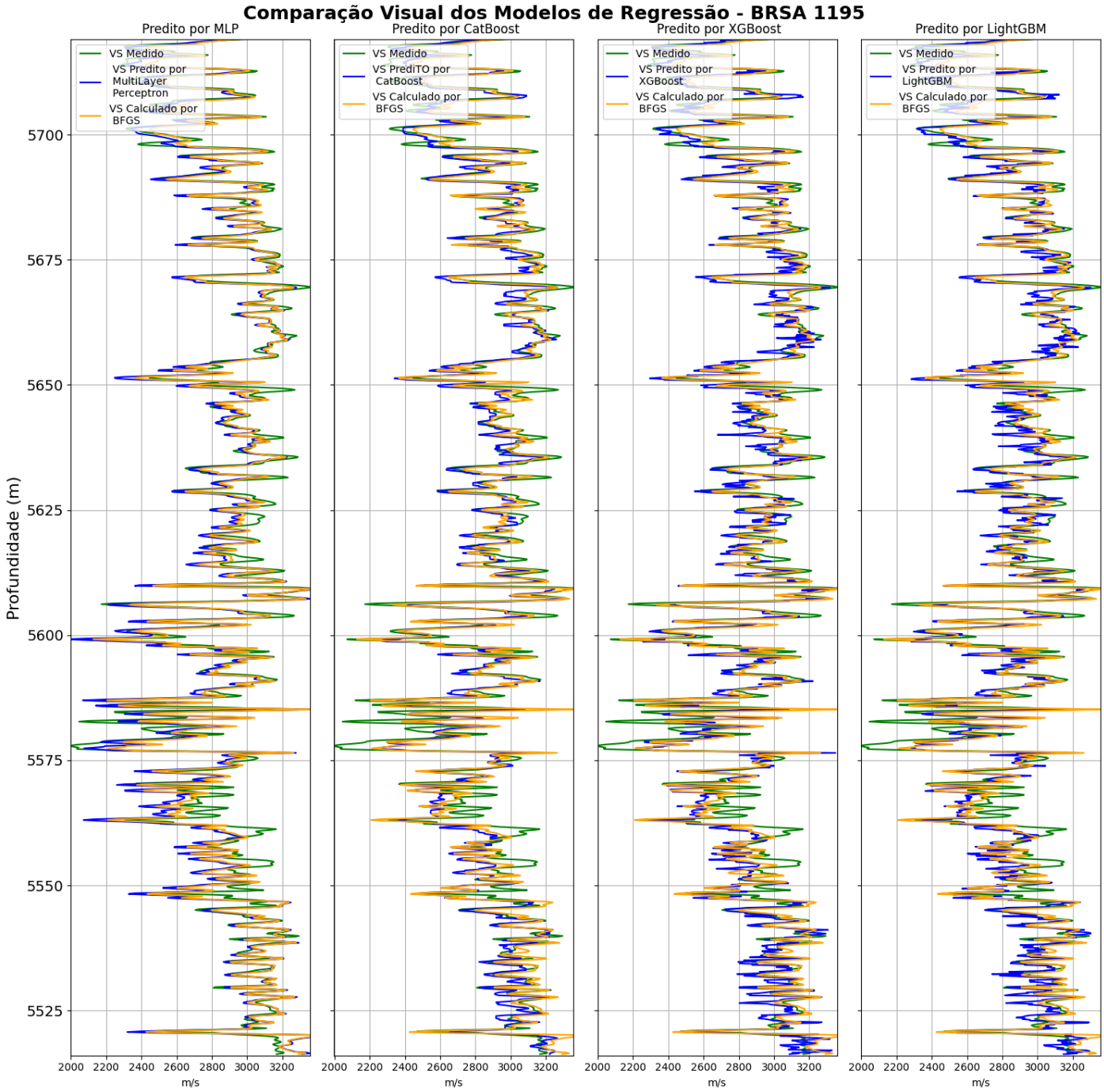

Figura 6: *Comparação visual entre os modelos de aprendizado de máquinas com o VS medido no poço BRSA 1195*