

12º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS



TÍTULO DO TRABALHO:

O USO DE MODELOS PREDITIVOS DE MACHINE LEARNING EM POÇOS DE PETRÓLEO: OTIMIZAÇÃO EM PROCESSOS ANALÍTICOS NA SELEÇÃO DE UM NOVO POÇO DE PETRÓLEO

AUTORES:

Thiago Yeshua Damasceno Ferreira, Camilo Andrés Guerrero Martin, João Victor Araújo da Silva, Fernando Antônio da Silva Fernandes.

INSTITUIÇÃO:

UNIVERSIDADE FEDERAL DO PARÁ - UFPA

O USO DE MODELOS PREDITIVOS DE MACHINE LEARNING EM POÇOS DE PETRÓLEO: OTIMIZAÇÃO EM PROCESSOS ANALÍTICOS NA SELEÇÃO DE UM NOVO POÇO DE PETRÓLEO

Resumo

Este estudo se propõe a analisar e catalogar o desempenho de técnicas de aprendizado de máquina afim de determinar quais locais são mais propensos a fornecer maior lucro potencial e menor probabilidade de perda. Concentrando-se em propor formas mais otimizadas de manejar essas etapas de forma eficiente e econômica. O machine learning no mercado permite a análise de grandes volumes de dados de maneira mais precisa e rápida. Além disso, esses modelos ajudam na alocação de recursos, prevendo locais com melhores retornos para perfuração e, conseqüentemente, reduzindo custos e aumentando a eficiência operacional. O estudo foi conduzido utilizando a linguagem de programação Python, que é ideal para a implementação de machine learning. A metodologia envolveu o uso de métodos preditivos para treinar modelos a partir de séries de conjuntos de dados, visando descrever a correlação entre os recursos obtidos nas regiões analisadas e seus destinos. A técnica de Regressão Linear foi especialmente empregada para estudar as informações obtidas e antecipar resultados futuros. O processo incluiu desde o pré-processamento dos dados até a padronização das variáveis, permitindo prever quais poços têm maior número de reservas e calcular a quantidade necessária para o desenvolvimento de um poço. Os resultados indicam que o uso de machine learning na análise de dados de poços de petróleo oferece uma maneira eficiente e precisa de otimizar a seleção de locais para perfuração. O ensaio estatístico realizado para aferir as variáveis com maior poder preditivo resultou em conjuntos de dados claros e concisos, com resultados indiscutivelmente satisfatórios. A aplicação desses modelos preditivos não só agiliza o processo de análise como também melhora a precisão das previsões, reduzindo o risco de decisões baseadas em dados insuficientes ou imprecisos. Este estudo contribui significativamente para a literatura existente, demonstrando como os modelos preditivos de machine learning podem ser aplicados eficazmente na indústria do petróleo. A inovação reside na capacidade de processar e analisar grandes volumes de dados de maneira eficiente, oferecendo uma perspectiva mais detalhada e precisa sobre a viabilidade dos poços. Isso representa um avanço importante na otimização dos processos de seleção e desenvolvimento de poços, promovendo uma maior economia de custos e eficiência operacional. A integração dessas técnicas avançadas de análise de dados pode, portanto, revolucionar a forma como as empresas de petróleo tomam decisões estratégicas, aumentando a competitividade e sustentabilidade do setor.

Palavras-Chave: Machine Learning, Aprendizado de Máquina, Otimização, Automação.

Abstract

This study sets out to analyze and catalog the performance of machine learning techniques with the aim of determining which locations are most likely to provide the highest profit potential and lowest probability of loss. Focusing on more optimized proportions of managing these steps efficiently and cost-effectively. Machine learning on the market allows the analysis of large volumes of data more accurately and quickly. Furthermore, these models help in resource allocation, avoiding locations with better returns for profit and, consequently, reducing costs and increasing operational efficiency. The study was conducted using the Python programming language, which is ideal for implementing machine learning. The methodology involves the use of predictive methods to train models from a series of datasets, describing the transparency between the features obtained in the verified regions and their destinations. The Linear Regression technique was especially used to study the information obtained and anticipate future results. The process included everything from data pre-processing to the standardization of variables, allowing us to predict which wells have the greatest number of reserves and calculate the amount needed to develop a well. The results indicate that the use of machine learning in the analysis of oil well data offers an efficient and accurate way of optimizing the selection of locations for drilling. The statistical test carried out to compare the variables with the greatest predictive power drawn into clear and concise data sets, with essentially overwhelming indisputable results. The application of these predictive models not only speeds up the analysis process but also improves the accuracy of variations, reducing the risk of decisions based on insufficient or inaccurate data. This study contributes significantly to the existing literature by demonstrating how predictive machine learning models can be effectively applied in the petroleum industry. The innovation lies in the ability to analyze and analyze large volumes of data efficiently, offering a more specific and accurate perspective on the well view. This represents an important advance in optimizing well selection and development processes, promoting greater cost savings and operational efficiency. The integration of these advanced data analysis techniques could therefore revolutionize the way oil companies make strategic decisions, increasing the competitiveness and sustainability of the sector.

Keywords: Machine Learning, Optimization, Automation.

Introdução

Na perspectiva atual, classificar com exatidão o espaço que se adequa aos critérios exigidos para a perfuração de um poço de petróleo demanda enormes intervalos de tempo e recursos. Denominam-se “Avaliação de Formações” as 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 das suas reservas de óleo e gás (THOMAS, 2001). Em um ambiente onde informações, etapas analíticas e dados em larga escala são proeminentes, uma avaliação concisa é indispensável, entretanto, métodos clássicos aproximam-se cada vez mais ao arcaísmo, logo, a necessidade de lidar com tamanha fatura de elementos e noções de forma mais eficaz possível é indispensável.

O desejo de previsibilidade é recorrente em tal área de atuação, uma estratégia preditiva procura fazer o trabalho antecipadamente no projeto, a fim de gerar um maior entendimento do que precisa ser feito posteriormente. Dessa maneira, podendo chegar a um ponto onde a última parte do projeto pode ser estimada com um grau de precisão mais saudável (Martin, 2004). O propósito deste projeto é aplicar técnicas de aprendizado de máquina e catalogar seu desempenho em modelos preditivos para determinar através de conjuntos de dados recolhidos a partir de variadas regiões para discernir quais são propensas a fornecer maior potencial de lucro e menor probabilidade de perda.

Metodologia

Com técnicas de análises dados em mãos, o programa foi formado a partir de machine learnig para identificar e determinar quais delas são propensas a terem melhor resultado para perfuração, considerando maximizar o potencial lucro e minimizar a probabilidade de perda. A figura abaixo apresenta o modelo dos processos da formação do programa e python.



Figura 1. Etapas da formação do projeto

Conjunto de Dados

O Database foi coletado a partir de amostras de óleo de três regiões, a partir de uma avaliação de formação, foi realizado estudos geológicos, geofísicos, qualitativos e quantitativos. Cada amostra

12º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS

teve seu recolhimento com base em métodos de perfuração e extração, sendo ela selecionadas através de parâmetros e critérios específicos, como a qualidade do petróleo e volume das reservas. A coleta dos conjuntos de dados é proveniente de três pontos distintos, e a partir deles foi obtido três tabelas, com cada uma dividida em quatro subconjuntos, sendo eles “id”, “F0”, “F1”, “F2” e “Produto”. Cada subconjunto representa aspectos relacionados à amostra, sendo eles o identificador da amostra, F1, F2 e F3 como características físico-químicas das amostras, e o produto como derivados do petróleo e potencial econômico da reserva.

Tabela 1. Tabela Parcial dos Conjuntos de Dados Coletados a Partir da Terceira Região

id	f0	f1	f2	product
2	0.7057449842080644	-0.4978225001976334	-0.005876013693320655	3.179102583207246
3	1.3347112926051892	-0.3401642528583136	0.9991827365665829	26.95326103153969
4	1.0227322634721168	0.15199044457022945	5.001160160764809	134.76630515769844
5	-0.03217162450743976	0.13903270818447744	4.999414898074921	137.94540774090564

Resultados e Discussão

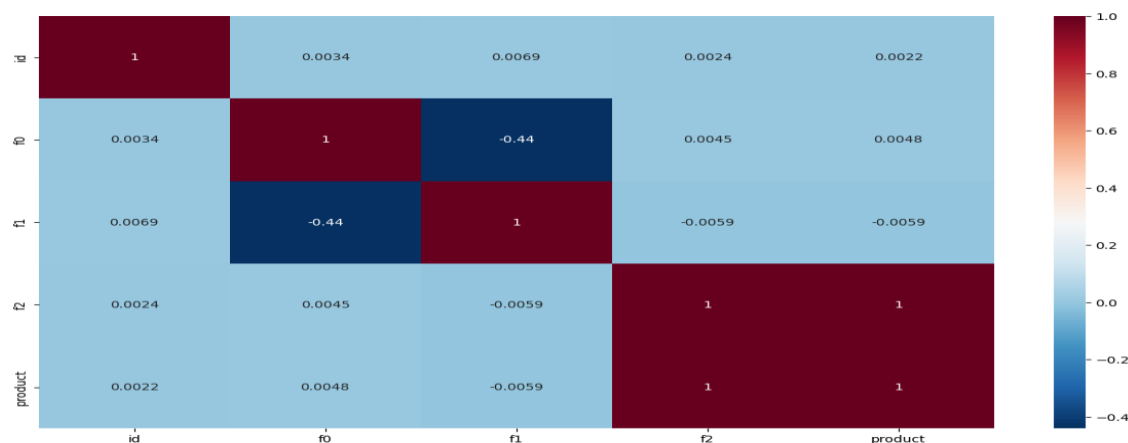


Figura 2. Terceiro Mapa de Correlação de Calor

A partir da biblioteca matplotlib, um ensaio estatístico para formular gráficos de correlação de calor dos conjuntos de dados. Onde foi mostrado que “F0” tem melhor correlação com “Produto” do que as outras colunas.

Podemos observar que temos um claro conjunto de dados. F2 geralmente tem maior correlação com a coluna de Produtos comparada as outras colunas. De todos os três conjuntos de dados, o último foi aquele no qual apresentou resultados mais agradáveis, onde temos uma correlação perfeita de 1. A distribuição dos valores nas colunas do *dataset* (Regiões) não são similares. DF0 e DF2 tem uma distribuição normal, enquanto DF1 tem uma distribuição senoidal.

Com o uso do método de Regressão Linear como técnica de análise de dados para modelar a relação entre os fatores e consequentemente poder receber o valor de dados necessários como MSE (Erro Quadrado Médio) e Valor Médio Previsto dos três conjuntos de dados por exemplo. O modelo atua a relação dos atributos de forma numérica, ou seja: explanatórios X e dependentes y. Age deste modo pois realiza ajustes matemáticos afim de minimizar erros. Resultado em três gráficos de erro residual distintos, onde o modelo três no demonstrou melhor desempenho.



Figura 3. Gráfico de Erro Residual do Terceiro Conjunto de Dados

Após a avaliação de nossos dados usando o modelo de Regressão Linear, podemos dizer claramente que as melhores pontuações são R3 e menor valor de RMSE que temos em nosso terceiro conjunto pode ser referido como mutabilidade de 99% do atributo dependente do produto pode ser explicado pelo modelo enquanto 1% da variabilidade ainda não é contabilizado. Entretanto, por outro lado, vemos que a pontuação de variância é enorme, o que nos diz que nos nossos dados há presença de uma grande dispersão de valores em relação ao valor médio.

Após realizadas as análises, foi observado que para ser rentável, um poço deve produzir no mínimo 111,11 mil barris do produto. Nas seções anteriores observamos que o valor médio da produção de todas as regiões era bastante baixo. Mas isto não significa necessariamente que os poços não sejam rentáveis, porque para verificar a rentabilidade dos poços, precisamos de saber a distribuição dos lucros dos poços, especialmente os poços com melhor desempenho.

Com o uso da função do lucro sem perdas completo, foi descrito que a melhor região para fazer dinheiro é a primeira, com o lucro obtido a partir das reservas sendo 27.27 milhões. Entretanto, não foi levado em consideração o cálculo de perdas, por isso não é possível ter 100% de certeza. Logo, para termos total ciência de todos os aspectos do projeto, foi realizado o cálculo dos riscos e lucros para cada região a partir do *bootstrapping*. Como resultado, as seguintes informações nos foram concebidas:

REGIÃO	LUCRO MÉDIO (mln USD)	INTERVALO DE CONFIANÇA DE 95%	RISCO(%)
Região 1	4.368	-1.325, 10	6.90%
Região 2	2.746	-2.412, 8	16.30%
Região 3	4.955	0.482, 9	1.70%

Tabela 2. Tabela de Lucros, Intervalo de Confiança e Riscos

Com base nas observações fornecidas, a terceira região aparenta ter o lucro mais elevado e o menor risco. As demais regiões têm lucros menores e riscos maiores, particularmente, a região dois tem lucro comparativamente baixo e riscos ligeiramente mais elevados. Com menor porcentagem de risco e maior lucro, a região 3 procede como mais promissora em comparação com as 1 e 2.

Conclusões

Um aspecto fundamental para o desenvolvimento de uma máquina é sua autonomia. Ser capaz de discernir correlações por meios de experiências em cadeia permite análises avançadas na qual o software aprende a realizar tomadas de decisões baseadas em dados. Este modelo estruturado através de aprendizado de máquina implementa ao setor da engenharia de petróleo a realização de análises em períodos de tempo mais curto, pois tais processos são constituídos de conjuntos de dados em larga escala, logo, se apto a cumprir-se de identificar padrões complexos que não são facilmente perceptíveis por métodos clássicos, é de fato uma grande otimização no processo, além de ao decorrer que mais informações são coletadas e estudadas, o modelo tende a contribuir para um ciclo contínuo de melhoria. O modelo também é capaz de avaliar riscos associados ao cenário da perfuração, auxiliando em tomadas de decisões mais seguras que previnem em possíveis perdas. Por fim, o algoritmo colabora em alocação de recursos ao prever os locais com melhor potencial para a perfuração, o que coopera na redução de custos e aumento de eficiência operacional.

O modelo tende a apresentar algumas limitações em relação a dados com tratamento insuficiente. O algoritmo trabalha com dados históricos para entender padrões e realizar previsões, quando as

informações fornecidas são limitadas ou más representadas, o desempenho do programa tende a cair. Em contrapartida, o programa apresentou boa generalização se adaptando adequadamente aos mais distintos cenários.

Agradecimentos

Agradeço primeiramente a Deus por me dar forças em momentos difíceis e me levantar e momentos mais obscuros para enfim resplandecer. Completa gratidão aos meus pais por me apoiarem e sonharem comigo na formação deste artigo. E por fim, agradeço a meu tutor por me conceber esta única oportunidade e a cada etapa me instruir pelo caminho correto.

Referências Bibliográficas

ANSELMO, FERNANDO. **MACHINE LEARNING NA PRÁTICA: MODELOS EM PYTHON**. 2020. <https://fernandoanselmo.orgfree.com>. acessado em: 14/06/2024

LIMA, A.; PAIVA, R.; MANZELA, A. **Inteligência Artificial e Indústria do Petróleo: Cenário e Possibilidades**. Rio de Janeiro, Revista de Engenharias da Faculdade Salesiana n. 16, p. 18-28, 2022.

Martin Fowler. **UML Essencial: Um Breve Guia para a Linguagem-Padrão de Modelagem de Objetos** - 3a Edição. Bookman, 2004.

Paixão GMM, Santos BC, Araujo RM, Ribeiro MH, Moraes JL, Ribeiro AL. **Machine Learning in Medicine: Review and Applicability**. Arq Bras Cardiol. 2022 Jan;118(1):95-102. English, Portuguese. doi: 10.36660/abc.20200596. PMID: 35195215; PMCID: PMC8959062.

SANTOS, H. G. DOS . et al.. **Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de São Paulo, Brasil**. Cadernos de Saúde Pública, v. 35, n. 7, p. e00050818, 2019.

SILVA, Edson de Araújo. **Uma abordagem baseada em Engenharia Dirigida por Modelos e Aprendizado de Máquina Aplicado a Robôs Móveis**. 2022.

THOMAS, JOSÉ EDUARDO. PETROBRAS. **Fundamentos de engenharia de petróleo**. 2.ed. Rio de Janeiro: Interciência: PETROBRAS, 2004. ISBN 85-7193-099-6.