

Gaussian process models applied for monthly forecast coal price mineral : a case study of Mozambique

Alfeu .D. Martinho^{1,2}, Leonardo Goliatt¹, Henrique Hippert^{1,3}

¹Graduate Program in Computational Modeling, Federal University of Juiz de Fora
Rua José Lourenço Kelmer, s/n, Campus Universitário, São Pedro, 36036-330, Juiz de Fora, Minas Gerais, Brazil
alfeudiasm@gmail.com, leonardo.goliatt@ufff.edu.br, henrique.hippert@ufff.edu.br

²Dept. of Exact Sciences and Technology, University of Pínguê

Campus Universitário de Cambinde-EN nº. 07, Telefone 82 3074748; Telefax 25220421; TETE, Moçambique

³Dept. of Statistics, Federal University of Juiz de Fora

Abstract. The search for models capable of predicting price movements and economic variables is recurrent in the finance literature. The understanding of these prices' behavior is essential for proper inflation control and planning of production in exporting countries. Although the research in this regard is relatively vast, presenting studies on statistical or econometric learning models of time series. Some approaches have deserved greater prominence in modeling and their prediction, such as moving averages methods, exponential smoothing, seasonal ARIMA, Autoregressive Vector (VAR), and ARCH or GARCH, among others. However, due to the peculiar characteristics of the commodities market, such series often describe movements such as randomness, changing levels, explained by various market factors and exogenous aspects. Linear models, as previously mentioned, may not be entirely appropriate for capturing this nonlinear behavior and therefore requiring other modeling forms. This work is part of this context since it proposes the application of Gaussian process models as an alternative for modeling and forecasting the behavior of the monthly prices of mineral coal in Mozambique. The time series refers to the monthly sales of coal mineral between 2011 to March 2020. The data were collected on the website of the bank of Mozambique, in the external sector database. The proposed model is compared to the other models using proper metrics. Our findings show that the Gaussian processes model presented promising results in the forecast one month ahead. The results obtained of this research can provide useful predictions for the coal prices that can assist the treasury managers in previewing the economic performance and potentially improve Mozambique's performance on the global commodities market.

Keywords: Gaussian processes models, forecasting, mineral coal

1 Introdução

A procura por modelos com capacidade de prever movimentos de preços e variáveis econômicas é recorrente na literatura de finanças. A compreensão do comportamento desses preços é extremamente importante para um apropriado controle da inflação e planejamento da produção nos países exportadores. A literatura a esse respeito é relativamente vasta, apresentando maioritariamente estudos através de modelos estatísticos ou econométricos de aprendizagem de séries temporais. Algumas abordagens mereceram maior destaque na modelagem e previsão, como métodos de médias móveis, suavização exponencial, ARIMA simples ou sazonal, vetor autorregressivo (VAR) e ARCH ou GARCH, entre outros. No entanto, devido às características peculiares do mercado de commodities, tais séries frequentemente descrevem movimentos como aleatoriedade ou ruído, mudança de níveis, não estacionariedade, explicados por vários fatores de mercado e aspectos exógenos, tornando a previsão um desafio. Modelos lineares, como os mencionados anteriormente, podem não ser inteiramente apropriados para capturar esse comportamento não linear, demandando assim outras formas de modelagem. Atualmente, modelos de aprendizado de máquina têm sido explorados na previsão de preços de commodities obtendo bons resultados. Ouyang et al. [1], Ticlavilca and Feuz [2], e Wang [3] aplicaram a rede LSTM (Long and Short-Term Memory), a Regressão Bayesiana Multivariada (MVRVM) e modelos de aprendizado supervisionado (regressão logística e SVM), respectivamente, para a previsão de preços de commodities agrícolas. Huang and Wu [4] e Panella et al. [5] utilizaram Deep Multiple Kernel Learning e mistura da rede neural Gaussiana, respectivamente, para previsão

de commodities de energia. Abdullah and Zeng [6], aplicaram redes neurais artificiais (ANN-Q), para prever o preço do petróleo. Este trabalho se insere nesse contexto, tendo como objectivo modelar e prever o comportamento dos preços mensais do carvão mineral de Moçambique através dos modelos de processos gaussianos. Um método atrativo para modelar dados ruidosos de acordo com Sun et al. [7], é um método não paramétrico baseado em modelar as respostas observadas dos diferentes pontos de dados de treinamento como uma variável aleatória normal multivariada (mais detalhes em Rasmussen [8]). Sua aplicação, pode ser encontrada em diversos problemas de aprendizagem supervisionada (regressão e classificação) e também não supervisionada. O conhecimento dos valores futuros do preço do carvão mineral pode contribuir para tomada de decisões importantes como a expansão das áreas de exploração, suporte nas logística de produção, gestão dos reassentamentos da população nas zonas afectadas. Os dados analisados no mesmo, são referentes ao preço mensal de venda do carvão mineral, entre 2011 e Março de 2020 e, foram obtidos no site do banco de Moçambique, numa base de dados do sector externo. Mediante a comparação do modelo proposto em relação aos outros modelos: floresta aleatória (RF) e regressão de vector suporte(SVR), usando como critérios algumas medidas de desempenho, observou-se que o modelo de processos gaussianos obteve resultados promissores na previsão um mês a frente. Os resultados obtidos nesta pesquisa fornecem previsões que podem auxiliar aos gerentes do Tesouro a visualizar o desempenho econômico e potencialmente melhorar o desempenho de Moçambique no mercado global de commodities.

2 Materiais e Métodos

2.1 Área de estudo

Moçambique, é um país situado na África Oriental, apresenta cerca de 780.000 km² de superfície e 2500 km da linha da costa. Caracterizado por elevada diversidade geológica, com recursos naturais de extrema importância. A sua economia esteve, desde muito, assentada na exploração dos recursos naturais. São vários recursos naturais explorados, entre os quais se destaca o carvão mineral. O distrito de Moatize, mostrado em Fig. 1 localizado na Província de Tete, no centro de Moçambique, abriga o maior jazigo de carvão mineral, com uma capacidade de cerca de 23 bilhões de toneladas de reservas de carvão, que têm atraído investimentos de todo o mundo. Essa Mina, produz carvão metalúrgico e térmico e tem como mercado Ásia oriental, Américas, Índia e Europa, contribuindo para o desenvolvimento econômico do país.

2.2 Dados

Os dados consistem em uma série histórica mensal de um total de 111 observações, correspondente a aproximadamente 9 anos, referentes ao período entre 2011 e 2020. A variável em análise é o preço de venda do carvão mineral (P). Base de dados disponibilizada pelo banco de Moçambique, referente ao sector externo. Os dados de preços de venda do carvão observados foram particionados em duas partes, a parte de desenvolvimento do modelo (treinamento) e de avaliação do modelo (teste). O preço previsto foi considerado como uma função de subconjuntos finitos de observações anteriores de preços. O modelo preditivo tem o formato mostrado na eq. (1):

$$P_{t+1} = F(P_t, P_{t-1}, P_{t-2}, P_{t-3}, \dots, P_{t-r}) \quad (1)$$

onde $P_t, P_{t-1}, \dots, P_{t-r}$ são as entradas com r o histórico de referência e F é a função a estimar.

2.3 Modelos de Processos Gaussianos (GP)

Vamos considerar um grupo preditivo x constituído por d -variáveis. O objetivo de um método de aprendizado de máquina é representar a relação funcional, mostrada na eq. (2) entre o preditor d -dimensional $x \in \mathbb{R}^d$, que pode incluir termos autoregressivos ou realizações passadas de y Bukkapatnam and Cheng [10] e a variável resposta y ,

$$y = f(x) + \epsilon, \quad (2)$$

onde f é desconhecida e $\epsilon \sim N(0, \sigma^2)$. Encontrar uma melhor aproximação \hat{f} para a função f constitui o foco dos diferentes métodos de aprendizado de máquinas.

Um processo gaussiano, é uma coleção de variáveis aleatórias, onde qualquer conjunto finito dos quais tem uma distribuição Gaussiana conjunta, ou seja, da eq. (3),

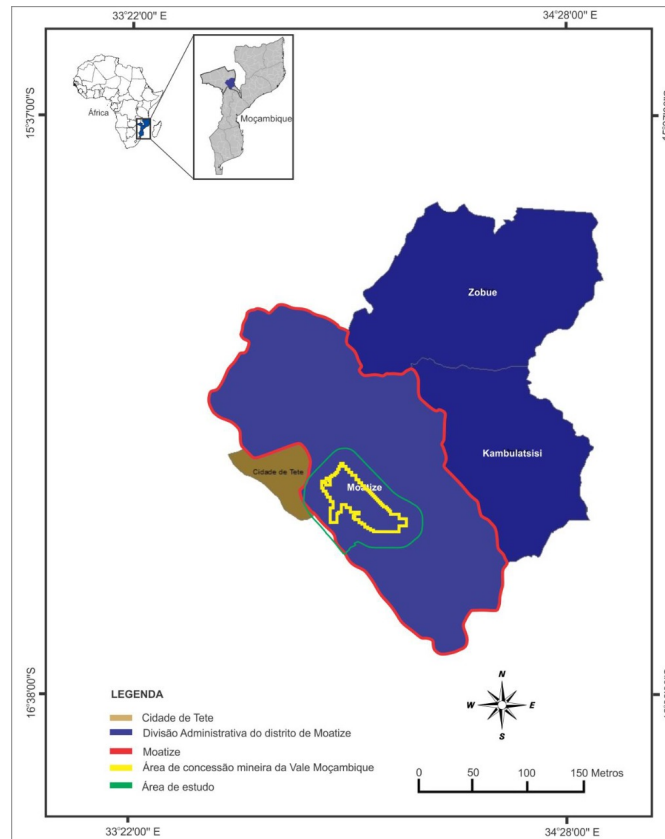


Figure 1. Mapa da localização geográfica da área de estudo, Macie [9].

$$f(x) \sim GP(m(x), k(x, x')), \quad (3)$$

onde $m(x) = E[f(x)]$ e $k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]$ são a média, que reflete o valor esperado da função na entrada x e, covariância, que modela a dependência entre os valores da função em diferentes pontos de entrada x e x' de f , que significa que pontos próximos têm a mesma distribuição.

Empiricamente, considera-se $m(x) = 0$ na eq. (3), o que faz com que a performance dos modelos de processos gaussianos dependa da escolha da função $k(x, x')$, conhecida como kernel. Um kernel apropriado é baseado em suposições como suavidade e padrões prováveis a serem esperados nos dados, popularmente é dado na forma da função base radial, definido como na eq. (4),

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{|x - x'|^2}{2\lambda^2}\right), \quad (4)$$

A adição e multiplicação de funções kernels simples são úteis na construção de uma variedade de funções kernels. Depois que um kernel é escolhido, podemos usar o Processo gaussiano para desenhar valores de função a priori, bem como valores de função condicionados a observações anteriores.

O treinamento do modelo de processos Gaussianos, consiste em aprender os hiperparâmetros a partir dos dados de treino, ou seja, supondo que uma forma de função kernel foi escolhida, mas depende de hiperparâmetros indeterminados, aprendemos esses hiperparâmetros a partir dos dados. E Numa estrutura de máxima verossimilhança, ajustamos os hiperparâmetros para maximizar a probabilidade logarítmica dos hiperparâmetros, eq. (5).

$$\log(p(y|x, \theta^*)) = -\frac{1}{1} y^T (k + \sigma_\epsilon^2 I)^{-1} y - \frac{1}{2} \log(k + \sigma_\epsilon^2 I) - \frac{n}{2} \log(2n) \quad (5)$$

com

$$\theta^* = \max_{\theta} \log(p(y|x, \theta)) \quad (6)$$

A eq. (5) tende a não ser côncavo com máximos locais múltiplos. Porém, vários métodos numéricos para estimativa dos hiperparâmetros ótimos, eq. (6) são relatados na literatura e já implementados em programas computacionais, como por exemplo o de gradientes conjugados .

A precisão da previsão depende dos hiperparâmetros da função de kernel. Por sua vez, a previsão consiste na construção da função a posteriori isto é, supondo que coletamos observações $D = X, y$ e queremos fazer previsões para novas entradas X^* desenhando f^* da distribuição posterior $p(f/D)$. Por definição, observações y e valores de função f^* seguem uma distribuição normal multivariada, com média

$$k(X_*, X)[k(X, X) + \sigma_{\epsilon}^2 I]^{-1} y$$

e matriz de covariância

$$k(X_*, X_*) - k(X_*, X)[k(X, X) + \sigma_{\epsilon}^2 I]^{-1} k(X, X_*)$$

que corresponde a um processo Gaussiano, definido na eq. (7)

$$f_*(x) \sim GP(m'(x), k'(x, x')) \quad (7)$$

onde média $m'(x) = k(x, X)[k(X, X) + \sigma_{\epsilon}^2 I]^{-1} y$ e kernel $k'(x, x') = k(x, x') - k(x, X)[k(X, X) + \sigma_{\epsilon}^2 I]^{-1} k(X, x')$. Ademais, tratando-se de modelos baseados em noção bayesiana na abordagem das distribuições de probabilidade, podemos construir intervalos de credibilidade sobre as previsões.

3 Discussão dos resultados

3.1 Preprocessamento dos dados

As séries que descrevem a variação de preços de commodities geralmente apresentam características peculiares como o ruído, a não-estacionariedade, Ouyang et al. [1]. Uma etapa de pré-processamento foi indispensável e, consistiu basicamente em:

- **Análise descritiva da série:** foi a etapa onde se fez o estudo das características importantes da série, através do método de decomposição sazonal, observou-se que a série tinha tendência, sazonalidade e ruído. A comparação de medidas descritivas em diferentes partições da série, o estudo da distribuição dos dados tanto com análises gráficas, histograma, assim como através de testes de normalidade formais e ainda testes estatísticos de estacionariedade, permitiram concluir sobre a não estacionariedade da série analisada.
- **Transformações:** Esta etapa consistiu basicamente em transformar a série dada em estacionária para a etapa da modelagem. Portanto, foi aplicada a transformação logarítmica para redução da assimetria da distribuição dos dados ou viés, A diferenciação, para a remoção da tendência, reescala dos dados aplicando o minmax, $y = \frac{x - \min}{\max - \min}$ e a transformação da série numa base de dados do tipo supervisionada. As bases de dados supervisionadas foram construídas com base o histórico de referência sendo 3, 6 e 9 meses atrás respectivamente.
- **Redução do ruído:** consistiu na aplicação do filtro média móvel de janela 3, 6 e 9, respectivamente, correspondente ao histórico, como descrito em Sagheer and Kotb [11] este filtro fornece uma média ponderada de pontos dos retardos da série, um intervalo de tempo correspondente ao número de retardos para gerar uma estimativa de suavização de uma série temporal. Esta etapa foi imperativamente incorporada para reduzir o ruído aleatório nos dados, mantendo a resposta de etapa mais nítida associada a dados brutos.

3.2 Resultados Experimentais

Para a comparação da performance do Modelo proposto (GP), dois métodos separados de aprendizado de máquina foram implementados na mesma estrutura computacional: Regressão de Vetor Suporte (SVR) e Floresta Aleatória (RF). Detalhes sobre estes modelos podem ser encontrado em Behnood et al. [12] e Hussain and Khan [13], respectivamente . Na modelagem, todos os modelos foram aplicados com a maioria dos seus parâmetros default, o GP foi alterado apenas na escolha da função kernel, que resultou da combinação aditiva de kernels simples que reproduzem características observadas da série na etapa do pré-processamento, tendência, sazonalidade e

ruído, o SVR nos parâmetros $C = 0.1$ e $epsilon = 0.005$ e o RF apenas no parametro $n - estimators = 500$. As métricas usadas neste artigo são mostradas na Table 1, onde R^2 é o coeficiente de determinação e $RMSE$ é Raiz do Erro Quadrático Médio, de acordo com Bukkapatnam and Cheng [10], onde O_i representa as observações e P_i os valores preditos. \bar{O} é a média das observações

Table 1. Métricas para a comparação da performance dos modelos.

Métrica	Expressão
R^2	$\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2}$
$RMSE$	$\frac{1}{N} \sqrt{\sum_{i=1}^N (O_i - P_i)^2}$

Os resultados na Tabela 2 mostram que todos os modelos nos diferentes periodos em análise, em geral, tiveram precisão de previsão adequada, baixo $RMSE$ e R^2 alto. Mas, o modelo GP, superou outros, ao obter menor $RMSE$ e maior R^2 .

Essa precisão adequada, pode estar relacionada a etapa de pré-processamento efectuada a serie para torná-la estacionária, como descrito em Bukkapatnam and Cheng [10], que obteve resultados similares e apontou a estacionariedade da serie como factor da variação na qualidade das previsões. Ademais, podemos observar ainda na Tabela 2, maior robustez do modelo GP, ao apresentar variações muito pequenas da precisão de previsão, comparando os históricos. A precisão das previsões, também contribuiu para um erro padrão quase nulo, no intervalo de credibilidade das previsões, ao nível de 95%, tornando as estimativas melhores.

A figura 2, mostra a Comparação de precisão para a previsão do preço de carvão um passo a frente, considerando diferentes periodos históricos analisados, podemos chegar a mesma conclusão obtida na análise dos resultados da Tabela 2, isto é, O GP resultou como o melhor modelo em relação aos outros. Mas importa referir que, o RF, também produziu resultados competitivos com o GP. Porém, podemos observar ainda na figura 2, uma ligeira redução da precisão de previsão em todos modelos, caracterizada pelo aumento de $RMSE$ ou diminuição do R^2 , com variação do histórico ou seja o aumento de informação, contribui para o aumento do ruído nos dados.

A Fig. 3 mostra o gráfico das previsões frente aos dados observados do conjunto de teste e o seu respectivo gráfico de residuos do modelo proposto com referência a 3meses anteriores, podemos observar que o modelo produziu estimativas acuradas do preço do carvão mineral um mês à frente, ao reproduzir as características da série de dados observados, como tendência, sazonalidade, e quaisquer outros com muita qualidade. Os resultados obtidos no mesmo, podem fornecer informações úteis que podem ajudar os gestores do tesouro a preverem e melhorar potencialmente o desempenho de Moçambique no mercado global de commodities.

Table 2. Comparação de desempenho dos modelos na previsão um passo a frente do preço do carvão mineral usando diferentes períodos de tempo como referência

	Métrica	03	06	09
GP	RMSE	0.0000011	0.0000023	0.0000021
	R^2	0.9997726	0.9994436	0.9993566
SVR	RMSE	0.1047485	0.113666	0.1173445
	R^2	0.9180106	0,9034555	0.8971015
RF	RMSE	0.0086859	0.0136677	0.0152437
	R^2	0.9994362	0.9986040	0.9982636

4 Conclusões

Os modelos de processos Gaussianos mostraram-se potenciais para realizar previsões de preços de carvão mineral de Moçambique. Conforme foi observado neste estudo, GP superou outros modelos, como SVR e RF,

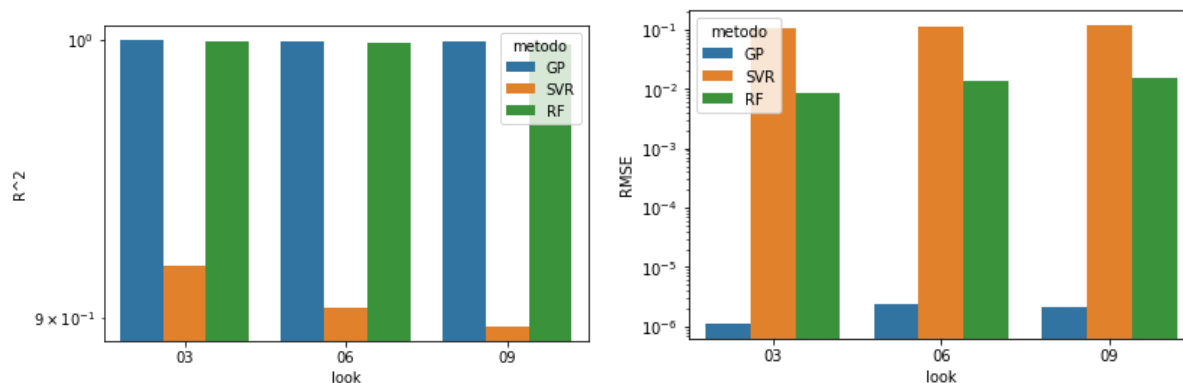


Figure 2. Comparação de precisão para previsão do preço de carvão , indicando que entre todos os modelos testados, O GP sustenta o menor $RMSE$ e maiores valores de predição de R^2

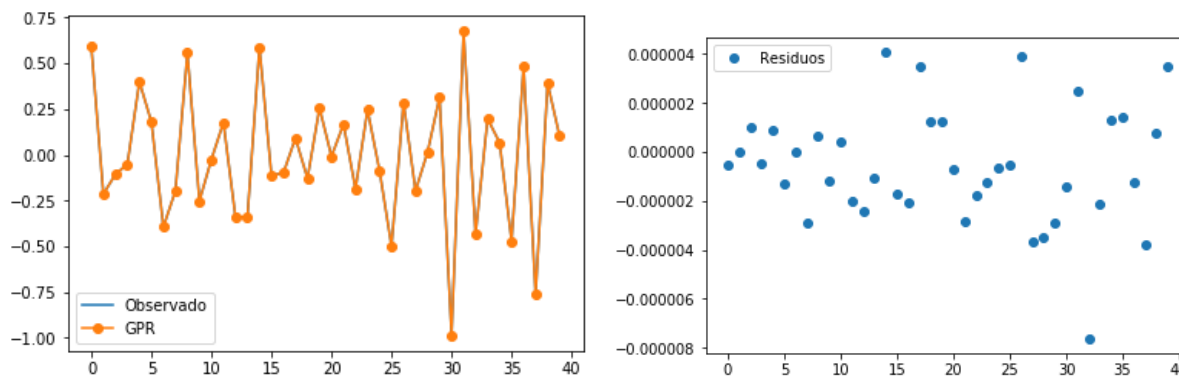


Figure 3. Comparação entre as series temporais,a previsão do preço de carvão frente aos dados observados e os resíduos pelo GP, com referência a 3meses anteriores, mostrando maior acurácia do mesmo

na previsão um passo a frente do preço de carvão, considerando diferentes períodos de tempo anteriores como referência. O modelo GP, mostrou ainda maior robustez, ao apresentar pequenas variações na precisão em diferentes períodos em análise. o modelo proposto pode ser útil para previsão não só do preço de carvão mineral, mas também de outras commodities, fornecendo informações úteis que podem ajudar os gestores do tesouro a preverem e melhorarem potencialmente o desempenho de Moçambique no mercado global de commodities. O modelo de processos gaussianos precisa da escolha da função do kernel e dos hiperparâmetros associados. Como trabalhos futuros, pretende-se aplicar uma abordagem híbrida do GP com estratégia evolutiva para selecção de função Kernel e otimização dos hiper-parâmetros associados, aplicar uma abordagem multivariadas,através da introdução de outras variáveis como de entradas, como por exemplo variáveis macroeconômicas (taxa de câmbio, de juros) teoricamente relacionadas a variação dos preços de commodities e outras formas de modelagem, como a estratégia online.

Agradecimentos. Este trabalho foi financiado pela Universidade Federal de Juiz de Fora (UFJF), pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e, pelo Programa de Formação de Professores de Educação Superior de Países Africanos (ProAfri) do Grupo Coimbra de Universidades Brasileiras (GCUB) – Código de Financiamento 001/2018.

Declaração de autoria. Os autores deste trabalho, confirmam serem as únicas pessoas responsáveis pela autoria do mesmo e, todo o material incluído como parte deste artigo é de propriedade (e autoria) dos autores ou têm a permissão dos proprietários já incluídos nele.

References

- [1] Ouyang, H., Wei, X., & Wu, Q., 2019. Agricultural commodity futures prices prediction via long- and short-term time series network. *Journal of Applied Economics*, vol. 22, n. 1, pp. 468–483.
- [2] Ticiavilca, A. M. & Feuz, D. M., 2010. Forecasting agricultural commodity prices using multivariate bayesian machine.
- [3] Wang, F., 2016. Forecasting agricultural commodity prices through supervised learning. vol. 2016.
- [4] Huang, S.-C. & Wu, C.-F., 2018. Energy commodity price forecasting with deep multiple kernel learning. *Energies*, vol. 11, pp. 3029.
- [5] Panella, M., Barcellona, F., & d'eclesia, R., 2012. Forecasting energy commodity prices using neural networks. *Advances in Decision Sciences*, vol. 2012.
- [6] Abdullah, S. N. & Zeng, X., 2010. Machine learning approach for crude oil price prediction with artificial neural networks-quantitative (ann-q) model. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.
- [7] Sun, A. Y., Wang, D., & Xu, X., 2014. Monthly streamflow forecasting using gaussian process regression. *Journal of Hydrology*, vol. 511, pp. 72 – 81.
- [8] Rasmussen, 2004. *Gaussian Processes in Machine Learning*. In: Bousquet O., von Luxburg U., Rätsch G. (eds) *Advanced Lectures on Machine Learning. ML 2003*. Volume Lecture Notes in Computer Science, vol 3176 of [8].
- [9] Macie, A. E. A., 2015. Mineração de carvão na bacia carbonífera de moatize, província de tete – noroeste de moçambique: Uma análise socioambiental. *Dissertação apresentada no Programa de Pós-Graduação em Mineralogia e Petrologia Instituto de Geociências da Universidade de São Paulo*.
- [10] Bukkapatnam, S. T. S. & Cheng, C., 2010. Forecasting the evolution of nonlinear and nonstationary systems using recurrence-based local Gaussian process models. , vol. 82, n. 5, pp. 056206.
- [11] Sagheer, A. & Kotb, M., 2019. Time series forecasting of petroleum production using deep lstm recurrent networks. *Neurocomputing*, vol. 323, pp. 203 – 213.
- [12] Behnood, A., Verian, K. P., & Gharehveran, M. M., 2015. Evaluation of the splitting tensile strength in plain and steel fiber-reinforced concrete based on the compressive strength. *Construction and Building Materials*, vol. 98, pp. 519 – 529.
- [13] Hussain, D. & Khan, A., 2020. Machine learning techniques for monthly river flow forecasting of hunza river, pakistan. *Earth Science Informatics*, pp. 1–11.