

Avaliação de métodos de predição do nível global do mar: uma abordagem comparativa entre modelos menos complexos de Regressão Linear, KNN, MLP e Prophet

Ana Júlia Ribeiro da Silva - ana.silva30@unifesp.br
Universidade Federal de São Paulo, UNIFESP - Instituto de Ciência e Tecnologia
São José dos Campos, São Paulo. 2023

Resumo

Através de dados coletados de 1993 a 2021, este estudo não apenas busca aprimorar a previsão do nível do mar através de técnicas menos complexas de inteligência artificial, como também defende a importância de ações visando preservar os ecossistemas marinhos diante das mudanças climáticas assim como a remediação das mesmas.

Palavras-chave: nível do mar, regressão linear, global, KNN, MLP

I. INTRODUÇÃO

O fenômeno do aquecimento global, impulsionado principalmente pelas emissões antropogênicas de gases de efeito estufa, tem desencadeado mudanças climáticas em escala global. Entre as manifestações mais alarmantes, destaca-se o aumento do nível do mar, cujas implicações vão muito além da simples elevação das águas.

O aumento das temperaturas globais têm acelerado o derretimento das geleiras e das calotas polares, contribuindo diretamente para o aumento volume dos oceanos, alterações nas correntes oceânicas e na salinidade dos mares, fatores que, consequentemente, representa uma ameaça iminente para ecossistemas marinhos e comunidades costeiras.

A acidificação dos oceanos, decorrente da absorção de dióxido de carbono atmosférico, tem efeitos devastadores para organismos marinhos. Espécies essenciais para a pesca, como moluscos e crustáceos, enfrentam desafios crescentes à medida que seus habitats se tornam mais ácidos e hostis. Isso não apenas ameaça a sustentabilidade dos recursos pesqueiros, mas também compromete a segurança alimentar de comunidades

dependentes da pesca como fonte primária de nutrição. Além disso, o aumento da temperatura da água pode alterar os padrões de migração e distribuição de várias espécies marinhas. Muitos animais marinhos dependem de condições específicas de temperatura para reprodução e alimentação e, à medida que as águas aquecem, essas espécies podem ser forçadas a se mover para áreas diferentes em busca de condições mais favoráveis, o que pode impactar ainda mais nos ecossistemas locais.

A elevação global do nível do mar, além de afetar a vida marinha, intensifica a vulnerabilidade das comunidades costeiras, expondo-as a eventos climáticos extremos, como tempestades e inundações. Infraestruturas críticas, como portos, áreas urbanas litorâneas e instalações industriais, enfrentam riscos substanciais.

Diante desse complexo cenário, a abordagem deste estudo converge diretamente com dois dos dezessete Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos pela Assembleia Geral das Nações Unidas (ONU): ODS 13 - Ação Contra a Mudança Global do Clima e a ODS 14 - Vida na Água. O entendimento aprofundado sobre a emissão de gases, efeito estufa e aquecimento global, assim como do seu efeito no aumento do mar é fundamental para orientar ações que possam atenuar os efeitos prejudiciais sobre os ecossistemas marinhos, garantindo a sustentabilidade ambiental e o bem-estar de comunidades costeiras.

Este artigo visa contribuir para a compreensão aprimorada dos métodos de previsão do nível do mar, utilizando uma análise comparativa entre algoritmos como Regressão Linear, K-nearest Neighbors (KNN), Multilayer Perceptron (MLP) e Prophet. Ao compreender as

nuances dessas técnicas, espera-se fornecer insights valiosos para a tomada de decisões informadas por cientistas e pesquisadores, em sintonia com os desafios iminentes associados às mudanças climáticas e ao aumento do nível do mar.

II. TRABALHOS RELACIONADOS

A. Restauração mensal da pressão do nível do mar do Atlântico Norte até 1750 dC usando otimização de inteligência artificial

Este artigo apresenta uma técnica de otimização baseada em inspiração biológica para restaurar mensalmente a pressão do nível do mar no Atlântico Norte até 1750 dC. A abordagem utiliza um algoritmo evolutivo, o algoritmo Coral Reef Optimization (CRO), combinado com o método analógico (AM). A otimização melhora significativamente o processo de restauração, especialmente em regiões oceânicas mal amostradas, e reproduz variações realistas em padrões climáticos regionais. A integração de estatística, ciência da computação e inteligência artificial destaca o potencial dessa abordagem multidisciplinar para melhorar previsões climáticas e preencher lacunas em conjuntos de dados observacionais.

B. Simulação e análise da mudança no nível do mar da estação maregráfica usando modelos de redes neurais artificiais

Ao decorrer do estudo é possível compreender processos para a simulação e análise das mudanças no nível do mar em Kuala Terengganu, na costa leste da Península da Malásia, até 2020. Utilizando modelos de Rede Neural Artificial (RNA), incluindo a Rede Neural Feed Forward (FF-NN) e o modelo Autoregressivo Não Linear com Redes Neurais Exógenas (NARX-NN), os resultados indicam que ambos os modelos são eficazes na replicação dos padrões de níveis do mar. No entanto, o modelo NARX-NN, com arquitetura (5-6-1) e quatro seleções de lag, demonstrou o melhor desempenho, projetando um aumento de

aproximadamente 25,34 mm/ano no nível do mar em Kuala Terengganu.

C. Avaliação de splines de regressão adaptativa multivariada e rede neural artificial para previsão da tendência do nível médio do mar ao redor da costa do norte da Austrália

Este estudo aborda a avaliação da tendência do nível médio do mar na costa norte da Austrália, considerando o impacto das mudanças climáticas. Utilizando séries temporais de quatro locais costeiros, o artigo emprega modelos de inteligência artificial, como Splines de Regressão Adaptativa Multivariada (MARS) e Redes Neurais Artificiais (RNA), para previsões precisas. Ambos os modelos demonstram alta precisão ($R^2 > 0,98$) e baixo erro (RMSE < 27%). O modelo RNA, especificamente, destaca-se ao reproduzir perfis de surto associados a ciclones, apresentando melhor concordância com picos medidos e horários de chegada. O estudo enfatiza a eficácia da RNA na previsão do nível do mar e sugere sua aplicabilidade em outras regiões do mundo.

D. Aplicação de Rede Neural Artificial (RNA) para melhorar a previsão do nível do mar

O artigo propõe a aplicação de Redes Neurais Artificiais (RNA) para aprimorar a previsão do nível do mar, indo além da análise harmônica tradicional. Utilizando séries temporais de pressão atmosférica, vento e marés, juntamente com dados meteorológicos, a RNA é treinada para considerar a influência combinada das forças meteorológicas e astronômicas. Os resultados mostram uma redução significativa na margem de erro, de 26% para 12%, indicando a eficácia da RNA na previsão da variabilidade do nível do mar. A aplicação potencial dessa abordagem se estende a áreas como terminais portuários e marinas, proporcionando informações valiosas para melhorar a segurança da navegação em canais estuarinos.

E. Drivers de variabilidade do nível do mar usando redes neurais

O estudo aborda a necessidade crítica de compreender a variabilidade regional do nível do mar, especialmente em face das ameaças globais. Utilizando aprendizado de máquina, a rede neural LSTM é aplicada para identificar os principais fatores influenciadores na variabilidade costeira de alta frequência do nível do mar na região do Mar do Norte. Os resultados destacam a dominância do padrão de pressão e vento zonal como fatores significativos. A abordagem oferece potencial para aplicação em várias zonas costeiras, proporcionando insights valiosos para decisores globais e contribuindo para modelos de previsão futura e gestão costeira.

III. METODOLOGIA

A. Pré-processamento de dados

Visando realizar modelagem preditiva do nível global do mar, o processo é dividido em três etapas cruciais para garantir uma melhor qualidade e a eficácia da análise. As bibliotecas Pandas e NumPy são empregadas para a manipulação eficiente de dados tabulares e operações numéricas, enquanto Matplotlib e Seaborn são utilizadas para visualização de padrões temporais por meio de gráficos informativos. A biblioteca Scikit-Learn desempenha um papel crucial, proporcionando funcionalidades como a divisão do conjunto de dados em treino, validação e teste, além da normalização Min-Max para preparar os dados para a modelagem preditiva.

O estudo utiliza um conjunto de dados fornecido pelo Global Climate Change, NASA, que documenta a variação no nível global do mar mensalmente durante o período de 1993 a 2021. Os dados são provenientes de observações de satélites e compreendem diversas colunas, incluindo informações cruciais para análise, como a variação do nível do mar, ajustes isostáticos globais (GIA) aplicados ou não, desvios padrão e variações suavizadas.

O pré-processamento inicia-se com uma exploração detalhada dos dados, proporcionando uma compreensão inicial do conjunto. Durante essa etapa, são apresentadas informações essenciais, como a forma do conjunto de dados,

estatísticas descritivas e a presença de valores nulos. A identificação e contagem de valores ausentes são cruciais para entender a integridade do conjunto de dados, assim como a visualização e remoção de possíveis dados duplicados, garantindo que não haja redundâncias que possam afetar a modelagem.

A visualização de padrões temporais ocorre a partir da criação de gráficos de dispersão, permitindo a análise da distribuição de uma variável específica ao longo do tempo, permitindo a identificação de tendências ao longo dos anos, compreendendo o comportamento dos atributos e identificando possíveis *outliers*.

Para a organização dos dados para modelagem preditiva, realizou-se a separação de atributos (X) e alvo (Y), juntamente com a divisão em conjuntos de treino, validação e teste. Além disso, a normalização Min-Max é aplicada aos atributos para garantir uniformidade na escala, facilitando o treinamento dos modelos.

B. K-nearest Neighbors (KNN)

O algoritmo de Regressão k-Vizinhos Mais Próximos (KNN) é um método de aprendizado supervisionado que atua com base na similaridade entre instâncias de dados, onde instâncias semelhantes tendem a possuir respostas semelhantes. No contexto da regressão, método utilizado neste estudo, o KNN procura vizinhos mais próximos em termos de características para prever valores contínuos.

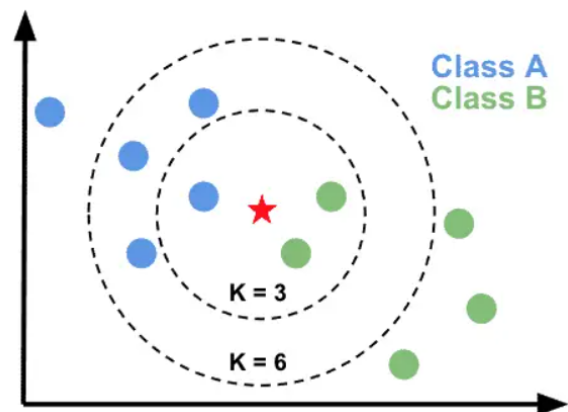


Figura 1 - Mecanismo do KNN. (Fonte: JC CHOUINARD)

Ao implementar o KNN, primeiro, um número "k" de vizinhos é escolhido e, em seguida, para uma instância de teste, o algoritmo identifica os "k" vizinhos mais próximos com base em alguma medida de distância (geralmente, a distância euclidiana). Para problemas de regressão, a previsão do valor para a instância de teste é calculada como a média dos valores alvo das instâncias vizinhas.

No código fornecido, a biblioteca scikit-learn é utilizada para implementar o KNN. O modelo é inicializado com $K=5$, número padrão indicando que serão considerados os 5 vizinhos mais próximos. Após o treinamento do modelo com o conjunto de treino, as previsões são geradas para o conjunto de testes. As métricas de desempenho, como o Erro Quadrático Médio (MSE) e o R2 Score, são então calculadas para avaliar a precisão do modelo.

C. Regressão linear simples

No contexto da previsão de séries temporais do nível global do mar, como abordado neste estudo, a regressão linear, técnica de modelagem estatística, procura ajustar uma linha reta aos dados de treinamento, onde a variável de saída (neste caso, a mudança no nível do mar) é modelada como uma combinação linear das variáveis de entrada.

A equação da regressão linear é uma representação matemática que descreve a relação linear entre uma variável dependente (ou resposta) e uma ou mais variáveis independentes (ou preditoras). A forma mais simples da equação para uma regressão linear univariada (com uma variável independente), mas o modelo pode ser facilmente aplicado para cenários multivariados. A ideia fundamental é ajustar a reta (ou hiperplano, no caso multivariado) aos dados representando a relação linear entre as variáveis da maneira mais precisa possível.

$$Y = \beta_0 + \beta_1 X + e$$

Y = β_0 + β_1 X + e
 Variável resposta Intercepto Coeficiente angular Variável explicativa Erro

Figura 2 - Atributos da representação matemática da regressão linear. (Fonte: Alura)

Para este estudo, a biblioteca scikit-learn é utilizada para implementar a regressão linear, onde o modelo é inicializado e treinado com os conjuntos de treino, e em seguida, são feitas previsões para o conjunto de teste. As métricas de desempenho, como o Erro Quadrático Médio (MSE) e o R2 Score, são calculadas para avaliar a qualidade do modelo.

Se torna relevante ressaltar que a Regressão Linear é uma abordagem simples, adequada para capturar tendências lineares nos dados. No entanto, ela pode ser limitada em capturar relações mais complexas que podem existir nos dados.

D. Multilayer Perceptron (MLP)

O modelo de Regressão MLP (Multilayer Perceptron) adotado neste estudo representa uma rede neural artificial composta por múltiplas camadas de neurônios, treinada usando o algoritmo de retropropagação (backpropagation), que ajusta iterativamente os pesos da rede para minimizar uma função de perda, medida da diferença entre as previsões do modelo e os valores reais.

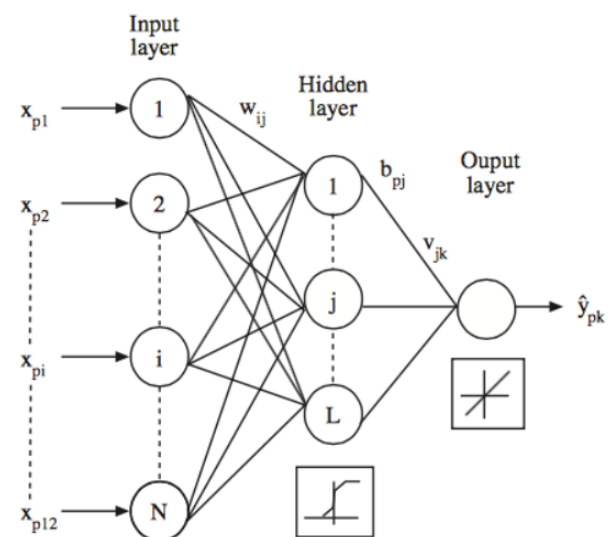


Figura 3 - Esquematização da MLP. (Fonte: Stack Exchange)

Durante a instância do modelo utilizando a classe MLP Regressor do scikit-learn, diversos parâmetros foram ajustados para otimizar o desempenho da rede. A arquitetura da rede é

configurada com três camadas ocultas contendo 100, 50 e 25 neurônios, respectivamente, proporcionando uma capacidade maior de aprender padrões complexos nos dados.

A função de ativação adotada é a ReLU (Rectified Linear Unit), conhecida por sua eficácia em lidar com não linearidades nos dados, e, além disso, é implementado também a regularização L2 para evitar overfitting, randomização dos pesos e estratégia adaptativa de aprendizado. O treinamento do modelo ocorre com os dados de treino, e, posteriormente, as previsões são realizadas utilizando os dados de teste. Ao calcular métricas de avaliação, como o Erro Quadrático Médio (MSE) e o R^2 Score, é possível avaliar o desempenho preditivo da regressão MLP neste contexto específico.

E. Prophet

O Prophet, desenvolvido pelo Facebook, é uma biblioteca projetada para facilitar a modelagem de séries temporais, especialmente aquelas que exibem padrões sazonais, feriados e outros efeitos temporais. O modelo adota uma abordagem aditiva, onde os diferentes componentes são somados para gerar a previsão total. Uma característica notável é a capacidade de incorporar automaticamente mudanças, adaptando-se dinamicamente a alterações nos padrões dos dados. Além disso, o Prophet lida de maneira robusta com dados ausentes e outliers, proporcionando uma modelagem mais resiliente.

Ainda que o código tenha sofrido algumas modificações para ir de encontro com o objetivo do estudo, a aplicação do modelo Prophet para previsão da variação do nível médio do mar segue uma metodologia específica proposta por Aran Pandey, disponibilizada na comunidade do site Kaggle.

Inicialmente, é realizada uma exploração dos dados para compreender a estrutura e características do conjunto. A seleção das colunas relevantes ('Year' e 'GMSL_noGIA') foi feita, e procedemos com a divisão dos dados em conjuntos de treinamento e teste. A escolha de focar na variável 'GMSL_noGIA' reflete a intenção de avaliar precisamente o aumento no

nível do mar relacionado ao efeito do aquecimento global, excluindo a influência do ajuste isostático global (GIA).

O ajuste isostático global refere-se às mudanças na crosta terrestre em resposta às variações de carga e processos geodinâmicos. Excluir esse ajuste permite uma análise mais direta das contribuições relacionadas ao aquecimento global na elevação do nível do mar.

Configuramos o modelo Prophet, especificando a inclusão de sazonalidade anual para capturar padrões recorrentes. A série temporal foi transformada, renomeando as colunas para 'ds' (representando o ano) e 'y' (representando a variação do GMSL). O modelo Prophet foi treinado utilizando os dados de treinamento, e as previsões foram geradas para 60 anos.

A visualização dos resultados foi realizada através de gráficos, que destacam as previsões, os intervalos de confiança e a tendência esperada ao longo do tempo, entretanto, não foi possível comparar esses resultados com dados reais para garantir a veracidade da predição.

IV. ANÁLISE EXPERIMENTAL

A. Conjunto de dados

O conjunto de dados analisado abrange observações e estimativas relacionadas à variação do nível médio global do mar (GMSL) ao longo do período de 1993 a 2021. Composto por 9 colunas e 1058 linhas, o conjunto fornece informações fundamentais, cada coluna neste conjunto de dados representa uma medida específica relacionada à variação do Nível Médio Global do Mar, considerando fatores como o ajuste isostático global (GIA), suavização por filtro gaussiano e remoção de sinal anual e semestral.

1. Year : representa o ano correspondente à observação;
2. TotalWeightedObservations: indica o número total de observações ponderadas para o respectivo ano.
3. GMSL_noGIA: refere-se à variação do Nível Médio do Mar (GMSL) em relação

- à referência média, sem a aplicação do Ajuste Isostático Global (GIA);
4. StdDevGMSL_noGIA: apresenta a estimativa da variação do desvio padrão do GMSL, sem considerar o Ajuste Isostático Global;
 5. SmoothedGMSL_noGIA: indica a variação suavizada do GMSL, utilizando um filtro tipo Gaussiano de 60 dias, sem aplicação do Ajuste Isostático Global;
 6. GMSL_GIA: Representa a variação do GMSL em relação à referência média, considerando o Ajuste Isostático Global (GIA);
 7. StdDevGMSL_GIA (Col7): Apresenta o desvio padrão da estimativa da variação do GMSL, considerando o Ajuste Isostático Global;
 8. SmoothedGMSL_GIA (Col8): Indica a variação suavizada do GMSL, utilizando um filtro tipo Gaussiano de 60 dias, considerando o Ajuste Isostático Global;
 9. SmoothedGMSL_GIA_sigremoved: apresenta a variação suavizada do GMSL, considerando o ajuste isostático global, com sinal anual e semestral removido.

```

RangeIndex: 1048 entries, 0 to 1047
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Year                                  1048 non-null   int64
1   TotalWeightedObservations            1048 non-null   float64
2   GMSL_noGIA                          1048 non-null   float64
3   StdDevGMSL_noGIA                    1048 non-null   float64
4   SmoothedGMSL_noGIA                  1048 non-null   float64
5   GMSL_GIA                            1048 non-null   float64
6   StdDevGMSL_GIA                      1048 non-null   float64
7   SmoothedGMSL_GIA                    1048 non-null   float64
8   SmoothedGMSL_GIA_sigremoved          1048 non-null   float64

```

Figura 4 - Colunas e suas devidas características

Para uma análise mais detalhada e abrangente, foram examinados individualmente os atributos (features) presentes no conjunto de dados. Essa abordagem permitiu uma compreensão mais minuciosa do comportamento específico de cada característica ao longo dos anos, contribuindo para identificar padrões, tendências e peculiaridades em suas variações, como exemplificado na figura a seguir:

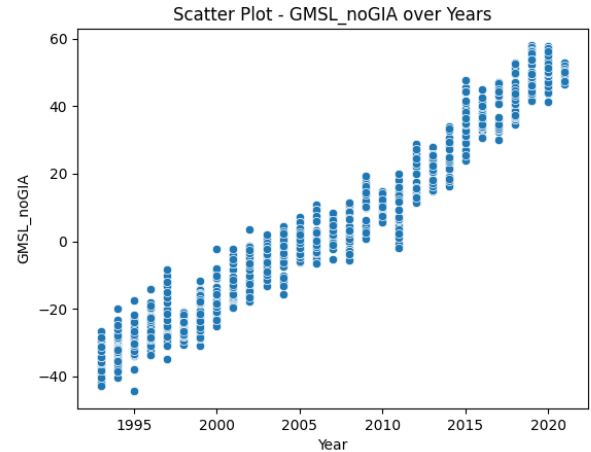


Figura 5 - Gráfico de dispersão demonstrando o nível médio do mar por ano

Após a fase inicial de exploração e entendimento do conjunto de dados, procedeu-se à divisão do conjunto em conjuntos de teste, validação e treinamento, destinando 20% dos dados para teste (210 instâncias), 16% para validação (168 instâncias) e a maior parcela, correspondente a 80%, para o conjunto de treinamento (670 instâncias), nos quais, posteriormente, foram normalizados.

Tamanho do Conjunto de Treinamento: 670
Tamanho do Conjunto de Validação: 168
Tamanho do Conjunto de Teste: 210

Figura 6 - Divisão dos dados para treinamento, validação e teste

B. Configuração do algoritmo e ambiente computacional

Em relação à configuração do algoritmo de maneira geral, algumas variações foram exploradas durante os testes, como a otimização do valor de K no algoritmo KNN, pequenas mudanças nas camadas da MLP assim como a adição da função $\alpha = 0,1$, que demonstrou uma redução significativa no erro. No entanto, de maneira geral, os padrões convencionais foram mantidos para preservar a integridade da análise. Essas variações específicas foram implementadas com o intuito de investigar potenciais melhorias nos resultados, mas respeitando as configurações gerais dos algoritmos para evitar mudanças que pudessem impactar negativamente a avaliação comparativa.

O estudo foi conduzido utilizando a linguagem de programação Python, versão 3.11.5, empregando o ambiente de desenvolvimento Anaconda, que possibilitou a implementação da biblioteca Fbprophet, compilado no Visual Studio Code (versão 1.82.2). A análise e processamento de dados foram realizados com o auxílio das bibliotecas Pandas, NumPy e Scikit-learn, além disso, a visualização de dados foi aprimorada com a utilização das bibliotecas Matplotlib e Seaborn. Além disso, o ambiente de desenvolvimento compreendeu um notebook em um sistema equipado com um processador Intel Core i7-8550U com 1.80GHz e 8 GB de RAM.

C. Critério de análise

No estudo realizado foi utilizado o Mean Squared Error (MSE) e o R2 Score como critérios de análise para avaliar o desempenho dos modelos de regressão. O MSE se baseia em uma métrica que quantifica a média dos quadrados dos erros entre as previsões do modelo e os valores reais, atribuindo pesos maiores a erros maiores, o que significa que valores mais altos indicam um ajuste menos preciso do modelo aos dados.

Paralelamente a isso, o R2 Score, ou Coeficiente de Determinação, indica a proporção da variabilidade nos dados, quando igual à 1 significa que o modelo explica perfeitamente a variabilidade dos dados, enquanto um R2 Score igual à 0 indica que o modelo não oferece precisão a uma previsão constante da média dos dados.

Além dos tradicionais indicadores de desempenho dos algoritmos, como o Mean Squared Error (Erro Quadrático Médio) e R2 Score, o tempo de execução foi incorporado como um parâmetro crucial. Utilizando a biblioteca 'time', foram calculados os tempos individuais de execução para cada algoritmo: KNN Regression, Linear Regression, MLP Regression e Prophet Prediction, ignorando o tempo de execução necessário para a etapa de pré-processamento de dados.

D. Resultado e discussão

Na análise comparativa entre os algoritmos de MLP, KNN e regressão linear,

observamos que eles apresentaram desempenhos semelhantes em termos de tempo de execução, com a regressão linear sendo a mais rápida, demandando apenas 0,001 segundo, enquanto o KNN é o modelo mais demorada, levando aproximadamente 5,8 segundos e a MLP se mantém mediana levando 2,5 segundos para executar. No que diz respeito à acurácia, a regressão linear se destaca com um coeficiente de determinação (R2 Score) de 0,999998, seguida pela MLP com 0,997 e o KNN com 0,991.

```
| Mean Squared Error: 5.88932820952381
| R2 Score: 0.9914142605434936
o de Execução do KNN Regression: 0.009678125381469727 segundos

ar Regression | Mean Squared Error: 0.001314252296583242
ar Regression | R2 Score: 0.9999980840212335
o de Execução da Linear Regression: 0.0019214153289794922 segundos

Regression | Mean Squared Error: 1.5856617941031437
Regression | R2 Score: 0.9976883477120038
o de Execução da MLP Regression: 2.527052402496338 segundos
```

Figura 7 - Comparação entre KNN, regressão linear e MLP, levando em conta os parâmetros MSE, R2 Score e tempo de execução

No entanto, ao analisar o erro, notamos diferenças mais significativas. Apesar dos esforços para otimizar o KNN, ele ainda apresenta um Mean Squared Error (MSE) de 5,8, contrastando com o MSE de 0,001 da regressão Linear e 1,5 da MLP.

No caso do modelo Prophet, embora não tenha sido realizada a implementação de outro algoritmo específico para previsões em anos posteriores, impossibilitando uma comparação direta de precisão, algumas características relevantes foram observadas. O Prophet mostrou-se como o algoritmo mais demorado, demandando cerca de 5,2 segundos para treinar o modelo e realizar a predição. Apesar de não ser possível validar sua precisão, o Prophet fornece parâmetros informativos, como 'yhat', 'yhat_lower' e 'yhat_upper'. O 'yhat' representa a previsão central, enquanto 'yhat_lower' e 'yhat_upper' indicam intervalos de confiança inferior e superior, respectivamente. Esses parâmetros oferecem uma estimativa do valor predito e a faixa de variação esperada, permitindo uma análise mais detalhada da incerteza nas previsões.

	ds	yhat	yhat_lower	yhat_upper
79	2020-12-31 00:00:00.000002021	54.828577	48.036168	61.582813
80	2021-12-31 00:00:00.000002021	58.851784	52.129621	65.361467
81	2022-12-31 00:00:00.000002021	62.873332	56.596978	69.815136
82	2023-12-31 00:00:00.000002021	66.893223	60.371912	73.396214
83	2024-12-31 00:00:00.000002021	70.940738	64.093109	77.808367
84	2025-12-31 00:00:00.000002021	74.963945	68.310131	82.307888
85	2026-12-31 00:00:00.000002021	78.985493	72.303724	85.929840
86	2027-12-31 00:00:00.000002021	83.005385	76.057329	89.914000
87	2028-12-31 00:00:00.000002021	87.052900	79.797557	93.569006
88	2029-12-31 00:00:00.000002021	91.076106	84.129128	97.454450
Tempo de Execução da Prophet Prediction: 3.683401346206665 segundos				

Figura 8 - Parâmetros do Prophet e seu tempo de execução

Além disso, o gráfico gerado pelo Prophet proporciona uma visualização do comportamento do atributo ao longo dos anos, contribuindo para uma compreensão mais aprofundada das tendências e padrões temporais identificados pelo modelo.

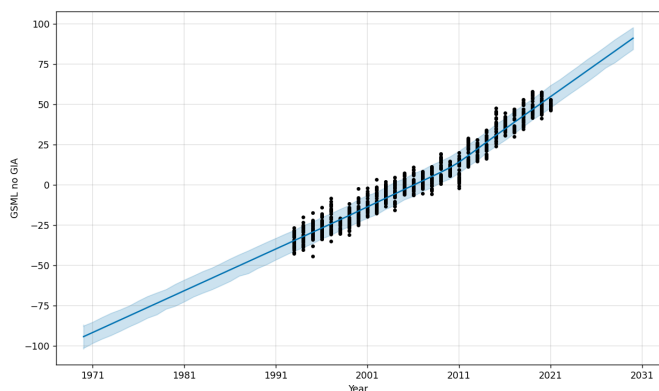


Figura 9 - Gráfico demonstrando a regressão de dez anos posteriores e valor de GSML_noGIA

A partir desses resultados, algumas observações e discussões foram feitas. Antes de qualquer coisa, sabe-se que o pré-processamento de dados é uma etapa fundamental em qualquer modelo preditivo, em estudos globais, como o realizado usando o conjunto de dados do Global Climate Change da NASA, a necessidade de uma análise mais aprofundada se torna evidente. Dados globalmente distribuídos podem apresentar características específicas que requerem uma abordagem cuidadosa, a variação na escala geográfica e a homogeneidade dos dados são fatores críticos a serem considerados durante o pré-processamento.

Embora os algoritmos menos complexos, como KNN, Regressão Linear e MLP em conjunto com as técnicas de pré-processamento empregadas tenham produzido resultados interessantes em termos de acurácia e erro baixo, é crucial

reconhecer suas limitações. Esses algoritmos podem não ser capazes de capturar nuances em dados complexos e altamente correlacionados, uma vez que a similaridade nos dados pode levá-los a um desempenho excessivamente otimista. A busca por uma maior precisão na previsão do nível do mar motiva a exploração de algoritmos mais complexos, principalmente quando cogitado a aplicação do modelo para dados regionais e, não necessariamente, apenas dados globais.

V. CONCLUSÃO

Diversos desafios foram identificados, incluindo a similaridade dos dados, o risco da ocorrência de *overfitting* e, consequentemente, a falta de generalização dos modelos, interferindo em sua precisão e impossibilitando a aplicação deles em diferentes regiões do globo.

Esses pontos indicam áreas de pesquisa promissoras para o futuro, destacando a importância de pesquisas futuras para aprimorar a compreensão e predição das mudanças no nível do mar. Isso inclui explorar abordagens mais avançadas de aprendizado de máquina como, por exemplo, redes neurais recorrentes para predição de séries temporais, sendo possível até mesmo realizar a comparação com o resultado do modelo Prophet.

É importante buscar por alternativas onde seja plausível ajustar os modelos para maior robustez, preparando-os para prever métricas de diferentes regiões do globo e examinar de perto as características dos dados. Por fim, o objetivo principal dos próximos passos baseia-se em desenvolver modelos mais precisos e aplicáveis, tanto em escalas globais como em escalas regionais.

VI. REFERÊNCIAS

CARLSTEDT, L. Drivers of sea level variability using neural networks. 2023. Disponível em: <<https://gupea.ub.gu.se/handle/2077/76416>>. Acesso em 1 dez. 2023

CHOUINARD, Jean. How to use k-Nearest Neighbors (KNN) with Python (Scikit-Learn

example). JC CHOUINARD, 2023. Disponível em:

<<https://www.jcchouinard.com/k-nearest-neighbor-s/>>. Acesso em: 1 dez. 2023

DUARTE, Ana. Desvendando a Regressão Linear. Alura, 2022. Disponível em: <<https://www.alura.com.br/artigos/desvendando-a-regressao-linear>>. Acesso em: 1 dez. 2023.

Climate Change: Vital Signs of the Planet. NASA. Disponível em: <<https://climate.nasa.gov/>>. Acesso em: 1 dez. 2023.

JAUME-SANTERO, F. et al. Monthly North Atlantic sea level pressure reconstruction back to 1750 CE using artificial intelligence optimization. Journal of climate, v. 35, n. 11, p. 3395–3410, 2022. Disponível em: <<https://journals.ametsoc.org/view/journals/clim/35/11/JCLI-D-21-0155.1.xml>>. Acesso em: 1 de dezembro

KOUSTUBHK. Global Sea Level | 1993 - 2021. Kaggle, 2021. Disponível em: <<https://www.kaggle.com/datasets/kkhandekar/global-sea-level-1993-2021/data>>. Acesso em: 26 nov. 2023

LINDSEY, R. Climate change: Global sea level. Climate GOV, 2022. Disponível em: <<https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level>>. Acesso em: 26 nov. 2023

MLP: Classification vs. Regression. Stack Exchange, 2015. Disponível em: <<https://stats.stackexchange.com/questions/149002/mlp-classification-vs-regression>>. Acesso em: 1 dez. 2023.

Objetivos de Desenvolvimento Sustentável. Nações Unidas Brasil. Disponível em: <<https://brasil.un.org/pt-br/sdgs>>. Acesso em: 26 nov. 2023.

PANDEY, Aran. Global Sea Level EDA + predictions. Kaggle, 2023. Disponível em:

<<https://www.kaggle.com/code/aranpandey/global-sea-level-eda-predictions>>. Acesso em: 6 dez. 2023.

RAJ, N.; GHARINEIAT, Z. Evaluation of multivariate adaptive regression splines and artificial neural network for prediction of mean sea level trend around northern Australian coastlines. Mathematics, v. 9, n. 21, p. 2696, 2021. Disponível em: <<https://www.mdpi.com/2227-7390/9/21/2696>>. Acesso em: 1 de dez. 2023

SIQUEIRA, B. V.; PAIVA, A. Using neural network to improve sea level prediction along the southeastern Brazilian coast. Ocean modelling, v. 168, n. 101898, p. 101898, 2021. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S1463500321001517>>. Acesso em: 1 dez. 2023