

Avaliação de Modelos de Aprendizado de Máquina aplicados a Predição da Produtividade Pesqueira Artesanal na Média Bacia Araguaia-Tocantins

Luna Loyolla¹, Saullo Andrade¹, Vanessa Oliveira¹, Marcela Souza¹,
Hugo Kuribayashi¹, Keid Sousa¹, Cristiane Cunha¹

¹Universidade Federal do Sul e Sudeste do Pará (Unifesspa) - Marabá, PA - Brasil

{luna.loyolla, saullo.guilherme7, vanessa.soliveira}@unifesspa.edu.br

{marcela.alves, hugo, keid.sousa, crisvieira.cunha}@unifesspa.edu.br

Abstract. *The Araguaia-Tocantins basin, known for its expansive physical extent and biodiversity, is vital to the subsistence of local populations; however, it lacks sufficient studies to aid in the formulation of effective public policies for the region. This research evaluates several machine learning (ML) models, focusing on the prediction of artisanal fishery productivity, using data analysis techniques. The models are assessed using the mean absolute error (MAE) and the root mean square error (RMSE), as well as the coefficient of determination (R^2). The results indicate that the Extreme Gradient Boosting (XGB) model demonstrated the best performance, achieving an average R^2 of 0.9993, suggesting its promising potential for future applications in the management of fishery resources within the basin.*

Resumo. *A Bacia Araguaia-Tocantins, caracterizada por sua vasta extensão física e biodiversidade, é essencial para a subsistência das populações locais, mas carece de estudos que ajudem na formulação de políticas públicas eficazes para a região. Este trabalho avalia diversos modelos de Aprendizado de Máquina (AM), com foco na predição da produtividade pesqueira artesanal, utilizando técnicas de análise de dados. Os modelos são avaliados usando as métricas de Mean Absolute Error (MAE) e Root Mean Square Error (RMSE), além do coeficiente de determinação (R^2). Os resultados indicam que o modelo Extreme Gradient Boosting (XGB) apresentou os melhores desempenhos, alcançando um R^2 médio de 0,9993, indicando ser promissor para aplicações em previsões futuras na gestão de recursos pesqueiros na bacia.*

1. Introdução

A bacia do Araguaia-Tocantins, uma das maiores bacias hidrográficas da América do Sul, abrange uma vasta área de aproximadamente 767.000 km^2 , situada inteiramente em território brasileiro. Esta bacia é caracterizada por sua biodiversidade, sendo lar de distintas formações ecológicas que vão desde savanas do Cerrado até florestas tropicais, e abriga dois dos principais rios da região: o Tocantins, famoso por sua ictiofauna, e o Araguaia, que possui uma das mais biodiversas planícies alagadas do mundo [Pelicice et al. 2021].

Apesar de seu valor ecológico incalculável, a bacia enfrenta sérios riscos ambientais provocados por atividades humanas, como a construção de grandes empreendimentos, a expansão da agricultura intensiva e a exploração mineral, que têm levado

a mudanças drásticas na cobertura do solo e no regime hidrológico, ameaçando a integridade dos ecossistemas locais, que são essenciais para a sustentabilidade da região [Winemiller et al. 2016, Nepstad et al. 2014].

Neste contexto é importante destacar que a atividade pesqueira na região, em especial artesanal e de pequena escala, representa um motor de desenvolvimento econômico e um fator de segurança alimentar para a população ribeirinha. O rio Araguaia, por exemplo, congrega cerca de 40% dos pescadores artesanais do Brasil. Ainda assim, a pesca de pequena escala não possui dados abundantes disponíveis para a bacia Araguaia-Tocantins, ao considerar informações como a diversidade de espécies exploradas, ou mesmo, em relação as quantidades capturadas, além de instrumentos e estratégias empregadas pelos pescadores da região [Prysthon et al. 2022]. Essa lacuna compromete a formulação de políticas públicas e estratégias eficazes de manejo e conservação. Além disso, em face das crescentes demandas por segurança alimentar e econômica, torna-se imperativo realizar estudos que estimem a produção de pescado na região [Isaac et al. 2015].

Embora a literatura relacionada apresente estudos voltados à formulação de estratégias para estimar a produtividade de recursos pesqueiros, a bacia Araguaia-Tocantins é limitada em termos de estudos que realizem abordagens preditivas. Uma forma de explicar tal comportamento se dá pelas características peculiares da pesca artesanal de pequena escala, que é exercida por um grande número de trabalhadores, com modalidades de captura muito variadas, que exploram amplas áreas e desembarcam suas capturas em locais dispersos [Prysthon et al. 2022]. Consequentemente, além dos desafios relacionados a coleta de dados, há desafios relacionados ao tratamento e identificação de variáveis e preditores que possam ser aplicados para a estimativa da produção pesqueira na região.

Técnicas de Aprendizado de Máquina (AM) tem sido aplicadas em várias áreas de estudo, devido a sua capacidade de generalizar, identificar padrões e fazer previsões. No contexto da bacia Araguaia-Tocantins, estes métodos podem ser aplicados para estimar a produção futura e os efeitos de mudanças nas práticas de pesca, possibilitando a criação de cenários que orientem políticas públicas e estratégias de manejo sustentável. Neste sentido, o principal objetivo deste trabalho é avaliar o desempenho de algoritmos de AM para predição da produção de pescado na bacia Araguaia-Tocantins. Assim, como objetivo, espera-se identificar modelos que tenham a capacidade de fazer previsões com base em variáveis diversas, como tamanho de embarcação (m), número de pescadores, espécie capturada (kg), entre outros atributos preditores, além de identificar as principais variáveis decisórias que influenciam a produtividade pesqueira na bacia Araguaia-Tocantins.

O restante deste trabalho está organizado da seguinte forma: A Seção 2 apresenta os trabalhos relacionados, enquanto a Seção 3 detalha os materiais e métodos empregados no desenvolvimento deste trabalho. Além disso, a Seção 4 descreve os resultados obtidos, enquanto que a Seção 5 encerra este trabalho com as conclusões obtidas.

2. Trabalho Relacionados

Após revisão de literatura, não foram encontrados estudos que tratem do mesmo objetivo deste trabalho, que destina-se a avaliação de métodos de AM para prever a produtividade pesqueira artesanal na bacia Araguaia Tocantins. No entanto, foram encontrados trabalhos relacionados que abordam pontos convergentes em termos de variáveis coletadas ou em termos da aplicação de métodos de estimativa de produtividade.

Em [Prysthon et al. 2019], os autores avaliam e compararam a produtividade e a composição das capturas de pesca da Usina Hidrelétrica de Tucuruí, propondo sugestões para um melhor manejo dos recursos pesqueiros. O estudo apresenta o registro da produção pesqueira por espécies ou grupos de peixes, utilizando o índice de Captura por Unidade de Esforço (CPUE) (kg/h) como principal indicador de abundância e produtividade. O estudo revela variáveis de interesse para a pesca artesanal na região, que além da CPUE envolvem o tipo e o peso (kg) de espécie capturada.

Em [Lourenço 2021], um estudo realiza uma análise da produção pesqueira artesanal na região do médio Madeira. O estudo considera dados sobre a composição específica das espécies de peixes capturadas, CPUE (kg/h), quantidades de captura, dados biométricos, além de informações sobre espécies de interesse comercial e suas receitas. A principal conclusão do estudo é que a implementação de um método de monitoramento mais robusto é crucial para melhorar a coleta de dados sobre as espécies capturadas, como a influência de variáveis ambientais, tais como cota do rio e distância dos locais de captura, garantindo assim a sustentabilidade dos estoques pesqueiros na região.

O trabalho de [Nunes 2015] apresenta um estudo voltado a estimativas de produção para pescarias artesanais na bacia do rio Xingu. O método adotado baseia-se na métrica de CPUE (kg/h), e avaliou diferentes variáveis de esforço, como dias de pesca, número de pescadores, e comprimento da embarcação, por meio de modelos estatísticos, incluindo as análises de regressão linear simples. Os resultados mostraram que o número de pescadores se ajustou melhor aos dados observados, concluindo que esta abordagem oferece uma representação mais precisa da abundância e produtividade das pescarias na região.

Em [Dutra 2023], os autores buscam analisar a importância dos peixes detritívoros na pesca artesanal dos rios Negro e Tapajós, considerando a co-gestão em áreas protegidas, sazonalidade e a influência da distância de centros urbanos. O trabalho emprega uma análise de regressão linear mista, que foi realizada utilizando variáveis resposta como a captura total, captura relativa e CPUE (kg/h), enquanto as variáveis independentes incluíram a distância de centros urbanos e as variações sazonais nos rios Tapajós e Negro.

Embora nenhum dos trabalhos relacionados apresentem objetivos similares aos propostos neste trabalho, tais estudos destacam a importância de iniciativas de monitoramento e estimativas de produtividade da pesca artesanal na região amazônica. Além disso, estes trabalhos sugerem a adoção, enquanto variáveis dependentes peso total (kg) e CPUE (kg/h), e enquanto variáveis preditoras, peso por espécie (kg), tamanho da embarcação (m), número de pescadores, distância de centros urbanos, entre outras variáveis.

3. Materiais e Métodos

3.1. Caracterização da Área

A Bacia do Araguaia-Tocantins está localizada na região que compreende os territórios dos Estados de Goiás, Mato Grosso, Tocantins, Maranhão, Pará e Distrito Federal, abrangendo 409 municípios ao todo. Caracterizada por suas águas claras e de baixa concentração de sólidos ou outros nutrientes dissolvidos, sua variabilidade socioeconômica e ambiental se dá devido ao fato da região possuir ampla diversidade de biomas tais quais Amazônia, Caatinga, Cerrado, Mata Atlântica e Pantanal.

Nessa perspectiva, a biodiversidade dos rios e corpos de água que compõem a Bacia Araguaia-Tocantins revelam uma alta heterogeneidade ictiofaunística, essencial para as atividades econômicas da região, contribuindo significativamente para o Produto Interno Bruto (PIB) local. Desta forma, as comunidades ribeirinhas dependem fortemente das águas da bacia para suas atividades pesqueiras, as quais envolvem tanto a subsistência quanto o abastecimento do comércio de pescado local [FAPESPA 2024].

Outro ponto crucial sobre a caracterização da Bacia Araguaia-Tocantins é a profunda crise socioambiental que ocorre na região devido a construção de grandes usinas hidrelétricas, avanço do agronegócio e aumento das ocorrências de garimpo ilegal [Pelicice et al. 2021]. Tais fatores tem contribuído para a degradação das áreas florestais e para a contaminação das águas afetando a fauna aquática local. As alterações causadas pelos impactos da exploração humana na região da bacia Araguaia-Tocantins tem ameaçado as populações ribeirinhas que dependem da pesca artesanal para sobrevivência. Esse fenômeno negativo está associado a inúmeras ações humanas indevidas, que afetam significativamente a qualidade da água e, consequentemente, a biodiversidade da ictiofauna da região [Santana et al. 2014].

3.2. Estratégia de Coleta e Seleção de Dados

O dataset utilizado neste trabalho é baseado em dados de desembarque auto-informados pelos pescadores da região, coletados pelo programa de Monitoramento Adaptativo Pesqueiro (MAP) [omitido para revisão]. Este programa busca promover uma melhor compreensão da dinâmica socioeconômica e ambiental da pesca na região e inclui dados coletados entre 2017 e 2020. Cada desembarque é composto por diversos Registros de Captura (RC), que incluem dados sobre a captura, a espécie pescada, a quantidade de biomassa e outras informações relevantes. O dataset abrange 7 municípios do Estado do Pará e várias comunidades situadas nas margens dos rios que compõem a bacia Araguaia-Tocantins. No melhor do conhecimento dos autores, esta é a única base de dados construída e disponível para a análise da dinâmica pesqueira na região.

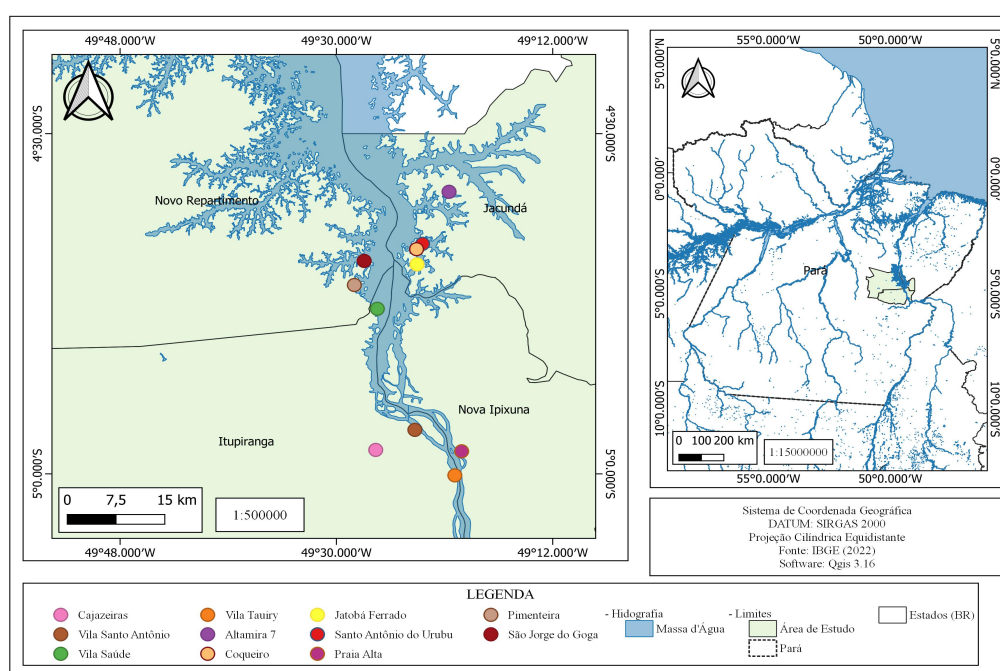
Este estudo está centrado nas comunidades da região denominada “Entorno do Pedral do Lourenção” (conforme representado pela Figura 1). Essa escolha reflete uma estratégia de manejo que prioriza uma abordagem territorial, ao invés de tratar a bacia como um espaço homogêneo em termos produtivos. Além disso, a escolha por essas localidades se justifica pela necessidade de compreender melhor a dinâmica da pesca artesanal nesta região, que é significativamente influenciada por fatores socioeconômicos e ambientais locais. Por outro lado, ao focar nesta área menos estudada, o trabalho busca preencher lacunas de conhecimento sobre a produtividade pesqueira local e oferecer subsídios para o manejo e a conservação sustentáveis das comunidades desta região. A Tabela 1 demonstra o apresenta uma síntese dos RCs coletados em cada comunidade do território.

Ademais, o dataset é composto originalmente de 52 atributos, que são classificados em variáveis de natureza temporal, produtiva, social, geográfica, instrumental, de peso e rendimento. Além disso, há atributos que descrevem as embarcações utilizadas pelos pescadores, as estratégias de pesca, os instrumentos utilizados e as espécies de peixes coletadas nos desembarques.

Tabela 1. Registros de captura em comunidades da Bacia Araguaia-Tocantins.

Território	Município	Comunidades	RCs
Entorno do Pedral do Lourenção	Itupiranga	Cajazeiras, Sto Antônio, Saúde e Tauiry	2942
	Jacundá	Altamira-07, Coqueiro, Jatobá Ferrado e Santo Antônio do Urubu	1542
	Nova Ipixuna	Praia Alta	137
	Novo Repartimento	Pimenteira e São Jorge do Goga	171

Figura 1. Localização das comunidades no Entorno do Pedral do Lourenção.



3.3. Pré-Processamento de Dados

O pré-processamento dos dados coletados incluiu a remoção de registros com valores ausentes, em particular para os atributos que registram o tamanho da embarcação, além de valor (R\$) e peso (kg) por espécie capturada. Atributos como o nome dos pescadores ou o nome científico da espécie capturada foram removidos, devido sua baixa utilidade nas tarefas de regressão. A variável peso total (kg) foi preservada durante o pré-processamento, enquanto outras variáveis relacionadas, como peso vendido (kg) e valor de venda total (R\$), foram removidas. Além disso, esta etapa incluiu a substituição de *outliers* pela média dos valores correspondentes no atributo, assim como a padronização dos atributos numéricos com média de 0 e desvio padrão de 1.

Além disso, parte das variáveis nominais foi transformada em atributos numéricos binários. Elementos enumeráveis como o tipo de embarcação (rabeta, canoa a remo, voadeira, etc.), os instrumentos utilizados (tarrafa, malhadeira, caniço, etc.), as estratégias de pesca (parada, ceva, caceia, etc.) e os rios onde as capturas foram realizadas (Tocantins, Araguaia, Itacaiúnas, Sororó, etc.) passaram por este tipo de transformação. Esse processo resultou na criação de novas colunas, em que cada valor categórico foi representado por 0 (“falso”) ou 1 (“verdadeiro”). A Tabela 2 apresenta as variáveis numéricas geradas

a partir da etapa de pré-processamento que totalizam 76 atributos.

Tabela 2. Variáveis numéricas restantes após o pré-processamento de dados.

territorio	cidade	especie
comunidade	nomePopular	comprador
ano	mes	qtdDias
gelo	outrosGastos	numeroPescadores
valorVenda	pesototal	CPUE
tamanhoEmbarcacao	potenciaEmbarcacao	especie.precio
especie.pesoTotal	tipoEmbarcacao_barco_com_motor_de_centro	tipoEmbarcacao_canoa
tipoEmbarcacao_canoa_a_remo	tipoEmbarcacao_rabeta	tipoEmbarcacao_rabeta_
tipoEmbarcacao_rabeta_canoa_a_remo	tipoEmbarcacao_rabeta_e_conoa_a_remo	tipoEmbarcacao_voadeira
rio_araguaia	rio_araguaia_tocantins	rio_araras
rio_cajazeiras	rio_itacaiunas	rio_itacaiunas-vermelho
rio_itacaiunas_sororo	rio_sororo	rio_tocantins
rio_vermelho	instrumento_anzol	instrumento_caniço
instrumento_carretilha	instrumento_espinhel	instrumento_linha
instrumento_linha de mão	instrumento_malhadeira	instrumento_malhadeira_linha
instrumento_malhadeira_tarrafa	instrumento_nao_informado	instrumento_outro
instrumento_pinda	instrumento_tarrafa	estrategia_baticao
estrategia_bloqueio	estrategia_boia	estrategia_caceia
estrategia_caceia_ceva	estrategia_caceia_parada	estrategia_canico
estrategia_canico_ceva	estrategia_canico_ceva	estrategia_ceva
estrategia_de_mao	estrategia_de_mao_ceva	estrategia_de_mao_canico
estrategia_de_mao_canico_ceva	estrategia_lance_rapido	estrategia_lance_rapido_parada
estrategia_moco	estrategia_nao_informado	estrategia_outra
estrategia_parada	estrategia_parada_caceia	estrategia_parada_canico
estrategia_parada_ceva	estrategia_parada_lance_rapido	estrategia_rede_parada

3.4. Desenho Experimental

A Figura 2 ilustra o fluxo do desenho experimental utilizado neste estudo, detalhando as etapas e processos envolvidos na pesquisa. O primeiro passo denota coleta de dados por meio do MAP. Em seguida, foi realizada a seleção e o pré-processamento dos dados, onde foram filtrados e organizados para melhor análise, considerando os RCs da região do “Entorno do Pedral do Lourenção”. Após essa etapa, aplicaram-se algoritmos de AM para as tarefas de regressão, que incluiu o ajuste dos hiperparâmetros destes modelos. Finalmente, a análise dos resultados foi realizada, avaliando as métricas de desempenho dos algoritmos, permitindo a identificação dos melhores modelos.

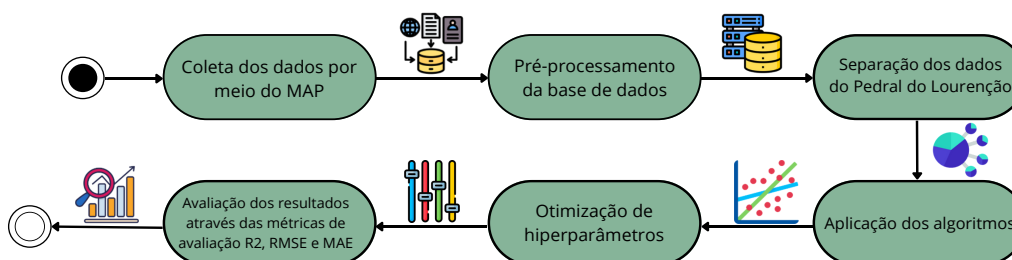


Figura 2. Representação do desenho experimental adotado neste trabalho.

A metodologia adotada neste trabalho envolveu a avaliação de modelos baseados em *Extreme Gradient Boosting* (XGB), *Random Forest* (RF), *Multi-layer Perceptron* (MLP) e *Support Vector Machine* (SVM). A adoção destas técnicas foram baseadas nas observações presentes na literatura relacionada, que destacam a eficácia de diversos

modelos em tarefas de regressão. As técnicas MLP e SVM se diferenciam pois se baseiam em redes neurais, utilizando camadas de neurônios e pesos para obter a melhor saída, e distanciamento entre vetores, buscando um hiperplano ótimo para os dados de treinamento, respectivamente. Apesar de XGB e RF utilizarem árvores de decisão, a primeira utiliza de árvores criadas sequencialmente para minimizar erros das anteriores, enquanto a outra demonstra seu funcionamento por meio de votações entre diversas árvores.

Para treinamento e teste dos modelos os dados foram divididos em partes de 80% e 20%, respectivamente. Ademais, no sentido de buscar a melhor divisão de dados para treino e teste a técnica de validação cruzada foi utilizada com K folds ($K = 5$), em conjunto com a otimização de hiperparâmetros do modelo por meio de *GridSearch*. Essa otimização é essencial e envolve a seleção de valores ideais que maximizam o desempenho dos modelos avaliados no desenho experimental.

3.5. Métricas de Avaliação

Com o intuito de mensurar o desempenho, as métricas adotadas neste trabalho são: Coeficiente de Determinação (R^2), onde valores mais próximos de 1 representam um modelo mais explicativo aos dados de teste, o *Root Mean Square Error* (RMSE) que é mais sensível à grandes erros, e o *Mean Absolute Error* (MAE) que penaliza erros de forma linear, independentemente de sua grandeza, conforme:

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (1b)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (1c)$$

Embora a quantidade de comunidades e algoritmos avaliados seja considerável, tais indicadores permitem que os algoritmos sejam quantificados e avaliados conforme o desempenho do modelo. Dessa forma, é possível identificar os melhores estimadores para cada localidade e possibilitar previsões futuras mais assertivas.

4. Resultados e Discussão

Todos os modelos foram implementados utilizando Python 3.11, com o uso das bibliotecas *sklearn* e *xgboost*. Estas bibliotecas foram utilizadas em diversas funções, que incluem tarefas pré-processamento de dados, seleção de algoritmos, mecanismos de validação cruzada e otimização de hiperparâmetros. Adicionalmente, a Tabela 3 apresenta os hiperparâmetros aplicados nos modelos testados em cada comunidade.

Os resultados das métricas de avaliação aplicadas aos modelos em cada comunidade são apresentadas na Tabela 4. Os resultados revelam que o XGB superou os demais modelos, destacando-se em todas as comunidades analisadas, apresentando um R^2 próximo de 1 e erros de previsão (RMSE e MAE) consideravelmente baixos. Esses resultados sugerem que o XGB possui uma capacidade promissora de ajustar-se ao conjuntos de dados do MAP e de minimizar erros de previsão. Em comunidades onde a heterogeneidade das práticas de pesca é alta e onde os dados disponíveis podem variar significativamente, o XGB mostrou um desempenho superior, possivelmente devido à sua flexibilidade em capturar essas dinâmicas. No entanto, é essencial reconhecer que, apesar do desempenho do XGB, a dependência de dados auto-relatados pelos pescadores

Tabela 3. Relação de hiperparâmetros ótimos encontrados no *GridSearch*.

Modelo	Melhores hiperparâmetros
Multi-Layer Perceptron	batch_size = 32, early_stopping = False, hidden_layer_sizes = (100,), max_iter = 300
Random Forest	max_depth = None, min_samples_leaf = 1, min_samples_split = 2, n_estimators = 100
SVM	kernel = “linear”
XGBoost	colsample_bytree = 0.8, learning_rate = 0.1, max_depth = 6, n_estimators: 200, subsample = 0.8

pode introduzir viés, e a inclusão de variáveis adicionais, como condições ambientais ou flutuações sazonais, poderia potencialmente aprimorar estas previsões.

Tabela 4. Síntese de resultados das métricas de avaliação.

Comunidade	Algoritmo	R ²	RMSE	MAE
ALTAMIRA 07	MLP	0.99478 ± 1.26E-3	38.3288 ± 8.5259	25.37415 ± 6.4479
	Random Forest	0.9985 ± 9.08E-4	18.80192 ± 6.3428	4.25824 ± 0.5408
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
CAJAZEIRAS	MLP	0.9911 ± 8.47E-3	42.10811 ± 4.7386	24.25029 ± 1.7799
	Random Forest	0.99855 ± 9.05E-4	19.03953 ± 5.958	4.28929 ± 0.5532
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
COQUEIRO	MLP	0.9936 ± 3.49E-3	36.12081 ± 8.6329	30.58191 ± 11.1093
	Random Forest	0.99853 ± 9.02E-4	18.96939 ± 5.9828	4.26776 ± 0.5059
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
JATOBÁ FERRADO	MLP	0.98962 ± 1.07E-2	41.00479 ± 13.116	26.81184 ± 5.92
	Random Forest	0.99859 ± 9.29E-4	18.77358 ± 6.1897	4.22936 ± 0.5558
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
PIMENTEIRA	MLP	0.99435 ± 1.71E-3	42.72586 ± 9.0841	28.91286 ± 7.7976
	Random Forest	0.99849 ± 8.51E-4	19.28428 ± 6.1863	4.34325 ± 0.4674
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
PRAIA ALTA	MLP	0.99354 ± 3.54E-3	36.02935 ± 9.8778	25.01933 ± 2.4181
	Random Forest	0.99856 ± 8.64E-4	18.8705 ± 6.1567	4.33234 ± 0.4613
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
SANTO ANTONINO	MLP	0.99288 ± 3.88E-3	41.2541 ± 8.5745	24.62877 ± 4.2373
	Random Forest	0.99848 ± 9.30E-4	18.95758 ± 6.3252	4.2086 ± 0.385
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
SANTO ANTÔNIO DO URUBU	MLP	0.99372 ± 2.87E-3	36.98002 ± 4.8703	25.60482 ± 6.193
	Random Forest	0.99853 ± 9.65E-4	18.85325 ± 6.1956	4.27814 ± 0.4685
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
SÃO JORGE DO GOGA	MLP	0.99474 ± 2.50E-3	37.29549 ± 8.6272	20.94623 ± 0.3839
	Random Forest	0.99851 ± 8.86E-4	18.9757 ± 6.08	4.27042 ± 0.5085
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
SAÚDE	MLP	0.99357 ± 5.10E-3	33.42123 ± 6.9192	24.27904 ± 4.1525
	Random Forest	0.99855 ± 8.80E-4	18.96777 ± 6.3086	4.2711 ± 0.4035
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181
TAUIRY	MLP	0.99174 ± 5.94E-3	40.56804 ± 4.7353	23.47183 ± 4.7025
	Random Forest	0.99862 ± 9.03E-4	18.47693 ± 5.9512	4.27809 ± 0.5087
	SVM	0.95684 ± 1.06E-2	106.00305 ± 13.38	45.1624 ± 2.1877
	XGBoost	0.99939 ± 4.42E-4	11.82038 ± 4.4458	4.20899 ± 0.5181

Por outro lado, o SVM se destacou como o modelo com os piores resultados em

todas as comunidades, apresentando métricas inferiores em comparação aos demais algoritmos. Essa performance insatisfatória sugere que sua capacidade de definir hiperplanos é inadequada para lidar com os valores mais complexos encontrados em diferentes contextos. Para melhorar o desempenho do SVM, a aplicação de uma técnica de redução de dimensionalidade poderia ser considerada, ajudando a simplificar a estrutura dos dados e potencialmente aumentando a sua eficácia. Ainda assim, MLP e RF demonstraram resultados mais consistentes e precisos em comparação ao SVM, variando significativamente conforme as comunidades analisadas. Ambas as abordagens indicam capacidade em modelar a complexidade dos dados, refletindo em métricas de desempenho mais favoráveis.

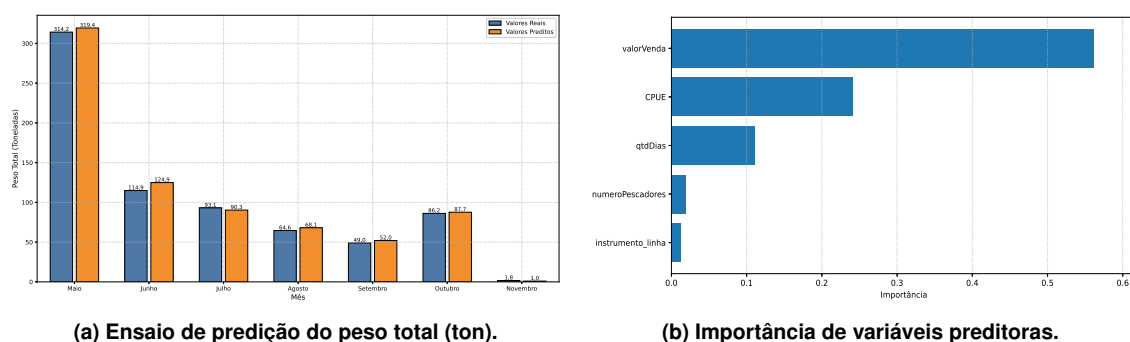


Figura 3. Ensaio de predição da variável peso total (kg) e gráfico de importância relativa das variáveis predictoras na estimativa da produtividade pesqueira.

De modo complementar, a Figura 3a apresenta um ensaio do processo de predição, no qual é realizada a previsão da variável peso total (ton) considerando os últimos meses do dataset. Para garantir a validade do modelo, esses meses foram excluídos do conjunto de treinamento, permitindo assim que a predição fosse feita com dados que não tinham sido utilizados na construção do modelo. Posteriormente, os valores previstos são comparados com os dados reais observados nesses meses, de tal forma que o R^2 médio obtido foi de 0.999688, no período considerado, evidenciando a capacidade preditiva do XGB.

Adicionalmente, a Figura 3b apresenta uma representação das 05 (cinco) variáveis mais importantes obtidas pelo XGB (melhor modelo), com o intuito de observar quais características são mais decisórias na predição da biomassa capturada. Neste contexto, o valor de venda corrobora com 56% para o modelo calcular o peso capturado, sendo seguido por: CPUE com 24% e a quantidade de dias com 11%, totalizando 91% de peso para a determinação do valor. A presença da CPUE, juntamente com variáveis temporais e o número de pescadores, condizem com os preditores utilizados em outros trabalhos da literatura, reforçando a relevância dessas variáveis na estimativa da produtividade pesqueira. Por outro lado, a variável valorVenda é aquela com a maior importância relativa calculada pelo modelo. Sua análise pode indicar um foco majoritário na captura de biomassa de baixo valor (R\$/kg) em grandes quantidades, em detrimento de capturas que exigem mais esforço e tempo (CPUE), mas que oferecem um maior valor agregado.

5. Conclusões

A bacia do Araguaia-Tocantins tem enfrentado sérios desafios ambientais e sociais que comprometem sua biodiversidade e a qualidade de vida das populações locais. Diante desta problemática, este trabalho buscou avaliar o desempenho de algoritmos de AM na

previsão da produtividade pesqueira artesanal na bacia do Araguaia-Tocantins. A metodologia envolveu a aplicação de técnicas como XGB, RF, MLP e SVM para realizar análises de regressão. Os resultados obtidos demonstraram a eficácia dos modelos na previsão da produtividade, com especial destaque ao modelo de XGB, que apresentou os melhores valores de R^2 , MAE e RMSE.

Estudos futuros devem avaliar tarefas de regressão separadas por espécie, uma vez que diferentes espécies podem apresentar padrões de captura, comportamento e resposta a variáveis ambientais distintas. Além disso, é possível investigar o efeito de variáveis socioeconômicas e ecológicas que influenciam a pesca. Esses esforços podem contribuir para a construção de políticas públicas mais eficazes e estratégias de manejo que promovam a preservação dos valiosos recursos pesqueiros da bacia do Araguaia-Tocantins.

6. Agradecimentos

Os autores agradecem o financiamento parcial desse estudo à Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Referências

- Dutra, M. C. F. (2023). Importância dos Peixes Detritívoros na Pesca Artesanal nos Rios Negro e Tapajós, Amazônia Brasileira. Master's thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS.
- FAPESPA (2024). Dashboard do Produto Interno Bruto (PIB) dos 144 Municípios do Estado do Pará.
- Isaac, V. J. et al. (2015). Food Consumption as an Indicator of The Conservation of Natural Resources in Riverine Communities of the Brazilian Amazon. *Anais da Academia Brasileira de Ciências*, 87(4):2229–2242.
- Lourenço, I. H. (2021). Análise da Produção Pesqueira Artesanal na Região do Médio Madeira. Master's thesis, Universidade Federal do Amazonas, Humaitá, AM.
- Nepstad, D. et al. (2014). Slowing Amazon Deforestation through Public Policy and Interventions in Beef and Soy Supply Chains. *Science*, 344(6188):1118–1123.
- Nunes, J. L. G. (2015). Estimador de Produtividade para as Pescarias Artesanais no Rio Xingu. Master's thesis, Universidade Federal do Pará, Belém, PA.
- Pellicce, F. M. et al. (2021). Large-scale Degradation of the Tocantins-Araguaia River Basin. *Environmental Management*, 68(4):445–452.
- Prysthon, A. et al. (2019). The Fishing Productivity Assessment Upstream and Downstream of Tucuruí Hydroelectric Dam, Tocantins-Araguaia Basin, Brazil. *Int. Journal of Advanced Engineering Research and Science*, 6(4):85–92.
- Prysthon, A. et al. (2022). *A Pesca Artesanal no Rio Araguaia, Tocantins, Brasil: Aspectos Tecnológicos e Socioeconômicos*. Editora EdEP.
- Santana, A. C. d. et al. (2014). Influência da barragem de Tucuruí no desempenho da pesca artesanal, estado do Pará. *Revista de Economia e Sociologia Rural*, 52:249–266.
- Winemiller, K. O. et al. (2016). Balancing Hydropower and Biodiversity in the Amazon, Congo, and Mekong. *Science*, 351:128–129.