

OTIMIZAÇÃO DA BIOPROSPECÇÃO DE ATIVOS BIOLÓGICOS POR INTELIGÊNCIA ARTIFICIAL

OPTIMIZATION OF BIOPROSPECTING OF BIOLOGICAL ASSETS BY ARTIFICIAL INTELLIGENCE

OPTIMIZACIÓN DE LA BIOPROSPECCIÓN DE ACTIVOS BIOLÓGICOS MEDIANTE INTELIGENCIA ARTIFICIAL

 10.56238/revgeov17n4-029

Wilson Kume

Doutorando em Gestão da Inovação em Biotecnologia - PPGBIOTEC
Instituição: Universidade Federal do Amazonas
E-mail: wilson.kume@ufam.edu.br
Orcid: <https://orcid.org/0000-0003-1542-7720>

Junior de Jesus Costa

Doutorando em Gestão da Inovação em Biotecnologia – PPGBIOTEC
Instituição: Universidade Federal do Amazonas
E-mail: costa.junior@ufam.edu.br
Orcid: <https://orcid.org/0000-0003-4540-3375>

RESUMO

O presente estudo examina a identificação de ativos biológicos por meio da união estratégica entre dados ambientais georreferenciados e modelos preditivos, com base em inteligência artificial (IA). Seu objetivo principal é a otimização dos processos e a diminuição das despesas operacionais na bioprospecção tradicional. Para tal, parte-se do pressuposto de que a biodiversidade terrestre é um vetor de ativo de valor incalculável, evidenciando a urgência em se criar e operacionalizar metodologias úteis para lidar com essa complexidade, transpondo as limitações das metodologias tradicionais. Desse modo, aplicou-se o algoritmo *Random Forest* junto a *rasters* climáticos, *shapefiles* e dados do Sistema de Informação sobre a Biodiversidade Brasileira (SiBBR), visando mapear áreas de elevado potencial bioativo. A metodologia foi testada por meio de protótipo em R Shiny, que possibilitou simular cenários pragmáticos e mapear riscos operativos. Mesmo os resultados preliminares apontando para a necessidade de refino paramétrico, os achados ratificam o potencial das ferramentas de IA em servir como agentes catalisadores à bioeconomia sustentável na Amazônia.

Palavras-chave: Bioprospecção. Inteligência Artificial. *Random Forest*. Otimização de Prospeção.

ABSTRACT

The present study examines the identification of biological assets through the strategic union between georeferenced environmental data and predictive models, based on artificial intelligence (AI). Its main objective is the optimization of processes and the reduction of operating expenses in traditional bioprospecting. To this end, it assumes that terrestrial biodiversity is an asset vector of incalculable value, evidencing the urgency of creating and operationalizing useful methodologies to deal with this



complexity, overcoming the limitations of traditional methodologies. Thus, the Random Forest algorithm was applied to climate rasters, shapefiles and Brazilian Biodiversity Information System (SiBBr) data, aiming to map areas of high bioactive potential. The methodology was tested through a prototype in R Shiny, which made it possible to simulate pragmatic scenarios and map operational risks. Even though the preliminary results point to the need for parametric refining, the findings ratify the potential of AI tools to serve as catalytic agents for the sustainable bioeconomy in the Amazon.

Keywords: Bioprospecting. Artificial Intelligence. Random Forest. Prospecting Optimization.

RESUMEN

El presente estudio examina la identificación de activos biológicos mediante la unión estratégica entre los datos ambientales georreferenciados y los modelos predictivos, basados en inteligencia artificial (IA). Su principal objetivo es la optimización de procesos y la reducción de los gastos operativos en la bioprospección tradicional. Para ello, se basa en la suposición de que la biodiversidad terrestre es un vector activo de valor incalculable, lo que evidencia la urgencia de crear y operacionalizar metodologías útiles para afrontar esta complejidad, superando las limitaciones de las metodologías tradicionales. Así, el algoritmo *Random Forest* se aplicó a *rastres* climáticos, *shapefiles* y datos de Sistema Brasileño de Información sobre la Biodiversidad (SiBBr), con el objetivo de mapear áreas de alto potencial bioactivo. La metodología se probó mediante un prototipo en R Shiny, que permitió simular escenarios pragmáticos y cartografiar riesgos operativos. Aunque los resultados preliminares apuntan a la necesidad de refinamiento paramétrico, los hallazgos ratifican el potencial de las herramientas de IA para actuar como agentes catalíticos para la bioeconomía sostenible en el Amazonas.

Palabras clave: Bioprospección. Inteligencia Artificial. *Random Forest*. Optimización de Prospección.



1 INTRODUÇÃO

A diversidade de vida na Terra é um de seus mais valiosos patrimônios, especialmente em regiões tropicais como a Amazônia. Ainda assim, a dimensão extensa e a natureza intrincada dos dados nas áreas de biologia, química e ecologia fazem com que sua análise constitua um desafio científico de complexidade crescente. A procura por compostos com atividade biológica ou a identificação de elementos que podem ser utilizados em medicamentos, na agricultura ou na produção industrial vem passando por uma transição paradigmática, impulsionada pelo progresso da inteligência artificial (IA). Enquanto os métodos tradicionais de investigação exigem muito tempo e recursos financeiros, a IA oferece a possibilidade de converter dados em informações úteis na seleção de componentes que são verdadeiramente valiosos.

Neste contexto, emerge a proposição de investigar: investigar os mecanismos pelos quais os algoritmos de Machine Learning conseguem aprimorar a escolha e a priorização das áreas de prospecção de ativos biológicos. Essa abordagem acelera a identificação de ativos biológicos de maneira eficiente, economicamente viável e sustentável.

Na bioeconomia, a velocidade de uma descoberta configura-se como um determinante crítico da competitividade setorial, e os métodos convencionais de triagem biológica têm uma taxa de sucesso comercial bastante baixa em comparação com os altos investimentos que demandam. Por outro lado, técnicas de ML, como *Random Forest* e redes neurais, permitem detectar padrões ocultos em bancos de dados complexos, identificar variáveis com maior poder preditivo e gerar modelos preditivos aplicáveis diretamente na triagem *in silico* de ativos. A junção da bioprospecção com a inteligência artificial transforma a maneira de procurar novos ativos biológicos, ao mesmo tempo em que diminui a necessidade de abundância de recursos.

Este trabalho apresenta um modelo computacional prototipado em R e Shiny, componentes da plataforma Open Science, concebido como um *mockup* funcional para validar a metodologia proposta. A iniciativa não se limita à fundamentação teórica ou à revisão bibliográfica, mas busca uma abordagem prática que permita testar e demonstrar a aplicabilidade do método. Esse protótipo funcional viabiliza a experimentação de conceitos e a análise de sua viabilidade, explorando as condições ambientais mais favoráveis à atividade biológica de compostos e refinando critérios estratégicos, embora não se tenha o propósito de desenvolver um *software* completo.

A metodologia adotada enfatiza a aplicação de testes de validação, essenciais para garantir a robustez dos modelos desenvolvidos, além da identificação de outliers, que podem comprometer interpretações e conclusões. A integração de atributos físico-químicos, aplicações conhecidas e dados climáticos organizados em camadas espaciais não só permite treinar modelos de classificação, mas também minimizar vieses e aprimorar a confiabilidade das decisões estratégicas de coleta, análise e investimento. Portanto, o objetivo precípua não consiste em entregar um sistema operacional, mas sim



estabelecer um alicerce conceitual que garanta a consistência e pertinência do modelo sugerido nos padrões científicos.

Diante da urgência climática, crise sanitária global e da busca por alternativas sustentáveis aos recursos sintéticos, integrar ciência de dados à bioprospecção passou a ser de uma alternativa facultativa em um imperativo estratégico.

2 REFERENCIAL TEÓRICO

A fundamentação teórica deste estudo repousa sobre a intersecção entre diversidade biológica, inteligência artificial e modelagem preditiva, enquanto os aprimoramentos tecnológicos têm potencial para mudar os encaminhamentos da bioprospecção de ativos biológicos. É a intersecção entre a ciência de dados, geoinformação e bioeconomia que aqui estabelece uma abordagem do referencial teórico, fornecendo suporte metodológico para a construção de modelos mais eficientes, mais sustentados e estrategicamente orientados para a valorização da diversidade biológica.

2.1 PREDIÇÃO, BASES DE DADOS E DESENVOLVIMENTO TÉCNICO

Nesta seção do estudo, analisa-se a contribuição da inteligência artificial para acelerar e refinar a descoberta de compostos com potencial bioativo, ao reunir informações ambientais, químicas e biológicas, e integrá-las com técnicas avançadas de predição, ML e dados georreferenciados, obtém-se uma abordagem mais estratégica e eficiente na exploração da biodiversidade.

2.1.1 Biodiversidade e Ativos Biológicos na Amazônia

A exploração da rica biodiversidade da Amazônia, seus materiais sustentáveis e metabólitos secundários com usos farmacêuticos, cosméticos e agrícolas, requer métodos de triagem e previsão mais eficazes. *Shapefiles* são, então, constituídos por elementos vetoriais, que podem ser pontos, linhas ou polígonos, delimitando fronteiras, áreas de distribuição e registros de ocorrência. Os *rasters*, diferentemente, estruturam os dados em formatos de grade, nos quais cada ponto retém informações com referência geográfica, representando variações do meio ambiente, a exemplo de temperatura, altitude ou o tipo de vegetação existente (Santos, 2024). Tais formatos, quando combinados, potencializam a análise, permitindo a utilização de dados espaciais em modelos ecológicos, a previsão de áreas de ocorrência de espécies e o estudo dos impactos ambientais de modo mais rigoroso.

2.1.2 Modelagem Preditiva com IA

Segundo Prajapati et al. (2025), tem se mostrado uma estratégia eficaz, a classificação de áreas de interesse por algoritmos supervisionados como o Random Forest, com base na probabilidade de ocorrência de determinados recursos. O uso de algoritmos supervisionados, principalmente *Random*



Forest, destaca-se na classificação de áreas favoráveis à ocorrência de ativos específicos, conforme demonstram estudos recentes.

2.1.3 Implantação em R Shiny

Com o uso de pacotes como Shiny (Sievert, 2020), no ambiente integrado de desenvolvimento da linguagem R, RStudio, é possível desenvolver aplicações web interativas, para as análises preditivas, utilizando técnicas de *Machine Learning* (ML) como o *Random Forest*, que implementa um conjunto de árvores de decisão e tem sido amplamente utilizado em estudos de ML (Mishra et al., 2024). A combinação dessas ferramentas permite gerar visualizações como mapas de probabilidade, mapas de calor e cenários de coleta baseados em dados reais, tornando o processo analítico mais acessível e favorecendo a tomada de decisões por pesquisadores e gestores. Além disso, amplia a eficiência e a aplicabilidade das análises ambientais e biológicas.

2.1.4 Dados utilizados

Para os modelos preditivos poderem funcionar, são incorporadas variáveis ambientais, químicas e de biodiversidade, tanto de fontes públicas quanto de dados simulados, provenientes do Sistema de Informação sobre a Biodiversidade Brasileira (SiBBR). É imprescindível a integração dos modelos ambientais e biológicos com informações vetoriais espaciais, geográficas e climáticas para que se possa realizar uma modelagem geoespacial minuciosa e precisa, o que permitirá criar mapas preditivos e planejar de forma mais direcionada as estratégias de bioprospecção em campo.

2.2 APLICAÇÕES DE IA NA PROSPECÇÃO DE RECURSOS NATURAIS E EXPLORAÇÃO GEOESPACIAL

Análises que envolvem grandes volumes de dados, imagens e sinais geofísicos, em prospecção de áreas de interesse, beneficiam-se da utilização de inteligência artificial. Áreas de difícil acesso tiveram a identificação de sítios possibilitada pela utilização de imagens de satélite, sensores remotos e algoritmos de detecção em investigações arqueológicas sobre os geoglifos da Costa de Nasca e as construções do Vale de Ica (Karamitrou et al., 2022; Küçükdemirci; Sarris, 2022).

No setor mineral, modelos como redes neurais, árvores de decisão e florestas aleatórias são aplicados ao exame de dados geológicos, geofísicos e geoquímicos, visando a identificar padrões associados à mineralização (Hasan; Shafiq; Khatun, 2023). Outras abordagens incluem mineração de texto, modelagem conceitual e visualizações baseadas em grafos de conhecimento, que contribuem para o refinamento do processo de descoberta (Li; Chen; Liu, 2022). Técnicas como GNNWLR, GWLR e GWRF também aparecem em estudos recentes, sobretudo pela capacidade de captar variações espaciais e relações não lineares (Wang et al., 2024).



Na indústria de petróleo e gás, métodos supervisionados como Support Vector Machines e *Random Forest* são empregados na classificação de regiões geológicas com potencial para exploração, contribuindo para decisões mais precisas (Choubey; Karmakar, 2020).

Em estudos hidrológicos realizados na bacia de Fincha, na Etiópia, redes neurais e métodos hierárquicos analíticos foram empregados para identificar zonas de potencialidade de aquíferos. A integração entre IA e sistemas de informação geográfica tem fortalecido análises espaciais e apoiado a gestão de recursos hídricos (Tamiru; Wagari; Tadese, 2022).

No campo biológico, iniciativas como a da Basecamp Research utilizam algoritmos de ML para catalogar biodiversidade, prever funções proteicas e identificar novas moléculas, com foco em práticas éticas e parcerias internacionais (Johnson, 2025).

2.3 ASPECTOS OPERACIONAIS, ECONÔMICOS E DE GESTÃO

Os aspectos práticos da inteligência artificial na bioprospecção, nesta seção, são os impactos operacionais, as previsões financeiras e as orientações de gestão que tornam a operação mais eficiente, menos custosa e com decisões mais assertivas e sustentáveis.

2.3.1 Cenários de Redução de Prospecção com IA vs. Tradicional

Uma vantagem significativa em relação às abordagens tradicionais, para otimizar a celeridade e reduzir os custos na identificação de novos ativos biológicos, se dá pela da inteligência artificial (Alloun; Calvio, 2024). Ela ainda possibilita o monitoramento em tempo real das variáveis durante a coleta, elevando tanto a produção quanto o padrão (Pomeroy; Grile; Likozar, 2022), e estima uma diminuição nos custos operacionais de até 30% (Chan, W. K. C. et al., 2019).

A estimativa de demandas e estoque, juntamente ao setor logístico, a inteligência artificial preditiva é utilizada para planejar suprimentos e gerenciar etapas críticas da cadeia logística e de suprimentos (Joel et al., 2024). A adoção destas tecnologias requer um investimento de capital inicial e formação profissional das equipes. Em contextos de pouca disponibilidade de dados ou quando procedimentos simples atendem ao objetivo, permanece a vigência das metodologias convencionais de triagem.

2.3.2 Questões Operacionais

Na logística operacional e no planejamento de campanhas de prospecção, a IA melhora a coordenação entre equipes e otimiza a logística com apoio de análises preditivas e processamento, em tempo real Danach, Dirani e Rkein (2024). Segundo Rane, Choudhary e Rane (2024), dados históricos analisados por IA são empregados na seleção de áreas para as campanhas. O material orienta a distribuição dos recursos e o planejamento das saídas de campo. Os modelos preditivos podem simular



diversos cenários, com a possibilidade de propiciar uma estimativa orçamentária com maior acurácia. A análise preditiva também melhora a previsão de demanda, reduzindo os riscos de excesso ou falta de estoque, conforme apontado por Joel et al. (2024).

2.3.3 Aspectos Econômico-Gestores

A bioprospecção de ativos biológicos, direcionada por IA, otimiza a gestão econômica, em especial na avaliação de ROI, economia financeira, utilização de recursos e administração de riscos.

O ML possibilita observar através da realocação de portfólios em tempo real, um crescimento em torno de 12,5% no retorno sobre o investimento (ROI) (Ambuli et al., 2024; Mediansyah; Septian; Zikry, 2024). A geração de recursos extras e redução de riscos, proporcionada pela análise de grandes volumes de dados feita com apoio de IA torna as estratégias mais rápidas e adaptáveis às mudanças do mercado (Zhu; Hu; Luo, 2024). Na gerência de riscos, sem dúvida, os instrumentos como Redes Neurais Artificiais e Máquinas de Vetores de Suporte aproveitam dados passados para entender as tendências e prever movimentos (Faheem; Aslam; Kakolu, 2022).

Esses modelos preditivos possibilitam que as empresas ajam antes que ocorra um problema, reduzindo riscos oriundos da incerteza do mercado (Gupta; Sharma; Abubakar, 2018).

2.3.4 Impactos na Gestão

Além dos aspectos econômicos, os impactos gerenciais também merecem destaque. Segundo Boardman et al. (2018), modelagens preditivas, mesmo com limitações na precisão dos modelos, são capazes de reduções expressivas de custos operacionais em comparação com métodos empíricos convencionais, técnicas como *Random Forest* possibilitam a economia de recursos quando aplicadas com critério (Giannini et al., 2012).

3 METODOLOGIA

A abordagem utilizada neste trabalho une métodos de modelagem preditiva, fusão de dados ambientais e recursos computacionais para validar a proposta apresentada. O principal objetivo é criar um protótipo funcional que possa replicar cenários e testar a viabilidade da estratégia proposta.

Nota Técnica: Este estudo não deve ser considerado um projeto de desenvolvimento de software; trata-se, na verdade, de uma aplicação metodológica destinada a explorar e melhorar as habilidades prospectivas na detecção de ativos biológicos.

3.1 ABORDAGEM TÉCNICA E DESENVOLVIMENTO DE MODELO

A seguir, os processos técnicos que serviram para moldar o modelo preditivo são explicados, desde a escolha das variáveis e o treinamento com os algoritmos de ML até a adição de dados



geoespaciais, para validação da proposta e a simulação dos cenários de prospecção com maior acurácia.

3.1.1 Fontes de Dados e Inserção no Sistema

Os dados biológicos foram extraídos das bases de dados do SiBBr, com registros de espécies e seus atributos. Foram complementados com uma coluna que atribui a cada amostra um nível de interesse – de 1 a 10, com uma casa decimal – com base na atividade biológica e no potencial do ativo biológico identificado. Esse nível orienta a busca pelas áreas prioritárias.

Além disso, os *rasters* climáticos e ambientais, obtidos do Banco de Dados de Variáveis Ambientais para Modelagem de Espécies (AMBDATA) e *shapefiles* do Ministério do Meio Ambiente (MMA), foram processados utilizando o pacote *raster* no R. Passaram por etapas de filtragem, padronização e normalização, procedimentos que garantem a compatibilidade dos dados com o algoritmo preditivo, permitindo uma modelagem precisa e confiável.

3.1.2 Modelagem Preditiva

Foram utilizados modelos de *Random Forest*, que foram alimentados com dados ambientais, como precipitação, pressão de vapor, temperatura e uso do solo, além de técnicas para aumentar a precisão das previsões. A solidez do modelo foi testada utilizando um esquema de validação cruzada repetida (*Repeated K-Fold Cross-Validation*), onde os dados de treinamento são divididos em cinco subconjuntos (*folds*), de modo que o modelo seja testado em diversas amostras, repetindo esse procedimento três vezes para minimizar a variabilidade e melhorar a confiabilidade dos resultados (Wong; Yeh, 2020).

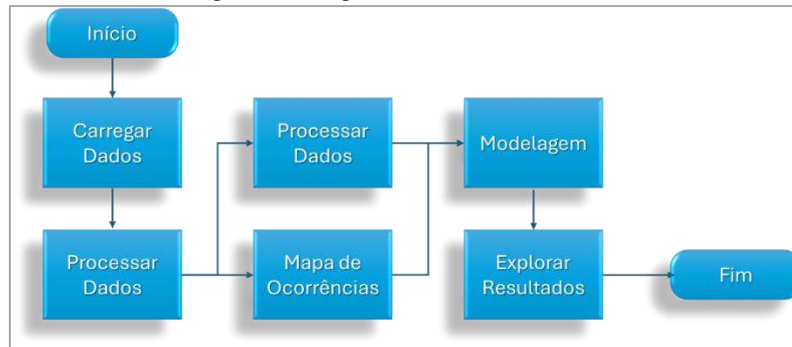
A literatura recomenda que, em cenários de baixa explicabilidade global e amostras reduzidas, observações cujo erro absoluto seja substancialmente inferior ao erro médio podem ser classificadas como previsões de alta confiabilidade local. Abordagem essa, consistente com recomendações metodológicas de avaliação de modelos que utilizam Raiz do Erro Quadrático Médio (RMSE) e Erro Médio Absoluto (MAE) como referências diretas para definir limites de erro aceitáveis (Chai; Draxler, 2014; Kuhn; Johnson, 2016; Willmott; Matsuura, 2005).

3.1.3 Implementação em R Shiny

Mesmo sendo somente um *mockup*, ele foi estruturado com interface amigável, contendo elementos de visualização de mapas e identificando as regiões com maior potencial para a presença de ativos biológicos, contemplando comandos para exportação de resultados (Jia et al., 2022).



Figura 1 - Diagrama de fluxo do Modelo



Fonte: Elaborado pelos Autores (2026).

3.2 ASPECTOS OPERACIONAIS E ESTIMATIVAS DE CENÁRIOS

Esta seção apresenta os componentes operacionais envolvidos na aplicação da inteligência artificial, abordando estimativas de cenários, redução de custos e otimização de recursos, com foco na eficiência estratégica das campanhas de prospecção de ativos biológicos.

3.2.1 Definição de Cenários

A inteligência artificial pode reunir dados operacionais e geológicos para criar cenários preditivos, apoiar a administração eficiente de recursos e orientar decisões em tempo real, resultando em eficiência e redução de custos operacionais (Gupta; Sharma; Abubakar, 2018; Monigha, 2024). A exemplo desse potencial, temos o RAP-Optimizer (otimizador de custos inteligente), que pode ter alcançado uma redução de 45% nos custos de servidores (Sathupadi et al., 2024) e de sistemas que, no tempo real, disponibilizam a estratégia a ser utilizada em função das variações de preços dos mercados (Verma et al., 2024). Comparações entre o custo atual e o custo gerado com IA indicam ser possível ter uma economia de até 40% no custo por mês no ambiente de nuvem (Sathupadi et al., 2024) e a redução do custo por barril de petróleo (Gupta; Sharma; Abubakar, 2018).

3.2.2 Avaliação de Risco e Gerenciamento

Dados históricos bem estruturados são a base. Dessa forma modelos preditivos de IA conseguem administrar riscos e alocar recursos com eficiência, da mesma forma que ferramentas de análise preditiva detectam ruídos, mapeiam gargalos e consideram variáveis externas. Já sistemas como o ARON (*Adaptive Resource Optimization Network*) se adaptam, ajustando a estratégia em conformidade com as flutuações e dinâmicas do mercado. O resultado pode chegar a 40% de ganho de eficiência (Verma et al., 2024).

Em situações voláteis e de falta de recursos, a adaptabilidade, oferecida pelos modelos preditivos, torna-se ainda mais significativa, já que permite que as decisões sejam rapidamente ajustadas frente a imprevistos, reduzindo a exposição a riscos que aparecem precisamente nessas



situações incertas (Zülfikaroğlu, 2024).

É importante que esses impasses sejam resolvidos cuidadosamente, para que as vantagens prometidas pela tecnologia se manifestem (Samarah et al., 2024; Yesufu; Alajlani, 2025).

3.2.3 Estrutura para a Comparação entre Modelos

Visando comparar a abordagem tradicional de busca biológica com a abordagem proposta baseada em *Random Forest*, foi elaborado um ábaco cruzado, que considera critérios organizados em três eixos: (1) eficiência preditiva; (2) relevância potencial dos achados; e (3) racionalização de custos e recursos logísticos. Cada critério foi avaliado qualitativamente por uma escala semiquantitativa (ex.: Baixo, Médio, Alto), utilizando correspondências visuais para facilitar sua representação gráfica.

Cabe destacar que o objetivo deste exercício não é realizar uma validação empírica exaustiva do modelo proposto, mas sim demonstrar sua viabilidade metodológica e argumentativa, conforme práticas já observadas em setores como petróleo e gás, onde a aplicação de IA em simulações controladas tem se mostrado eficaz na projeção de cenários e redução de custos operacionais (Gupta; Sharma; Abubakar, 2018; Sathupadi et al., 2024; Verma et al., 2024).

Adicionalmente, os critérios comparativos utilizados neste estudo foram inspirados em benchmarks e padrões recorrentes na literatura técnica, especialmente em aplicações de IA voltadas à otimização logística, estimativa de custos e gestão de recursos, com base em dados históricos e estudos de campo previamente documentados (Boardman et al., 2018; Chan, W. K. C. et al., 2019; Danach; Dirani; Rkein, 2024; Rane; Choudhary; Rane, 2024).

O ábaco comparativo elaborado indica que o modelo baseado em IA apresenta potenciais ganhos de eficiência frente à abordagem tradicional. Com relação às tentativas necessárias, o modelo de IA deve demonstrar um desempenho estimado superior ao método tradicional apresentando índices inferiores, na quantidade de incursões. Tais resultados refletem padrões relatados na literatura, que apontam que a sensibilidade e a acurácia de modelos *Random Forest* podem ultrapassar 90% em cenários com múltiplas variáveis ambientais (Cutler et al., 2007).

Figura 2 - Comparação de Métodos de Prospecção

Prospecção Tradicional	Alto	Empírico	Limitada	Indireto	Aleatória	Baixa	Alto	Meses	Alto
Modelo IA (Random Forest)	Baixo	Baseado em dados	Alta	Explícito	Elevada	Alta	Reduzido	Dias/Se m	Reduzido
Critério	Tentativas até achado relevante	Base de decisão	Abrangência espacial	Direcionamento por hotspots	Potencial de inovação disruptiva	Prospecção em áreas remotas	Custo por achado relevante	Tempo médio até decisão	Custo de oportunidade

Fonte: Elaborado pelos autores (2026).



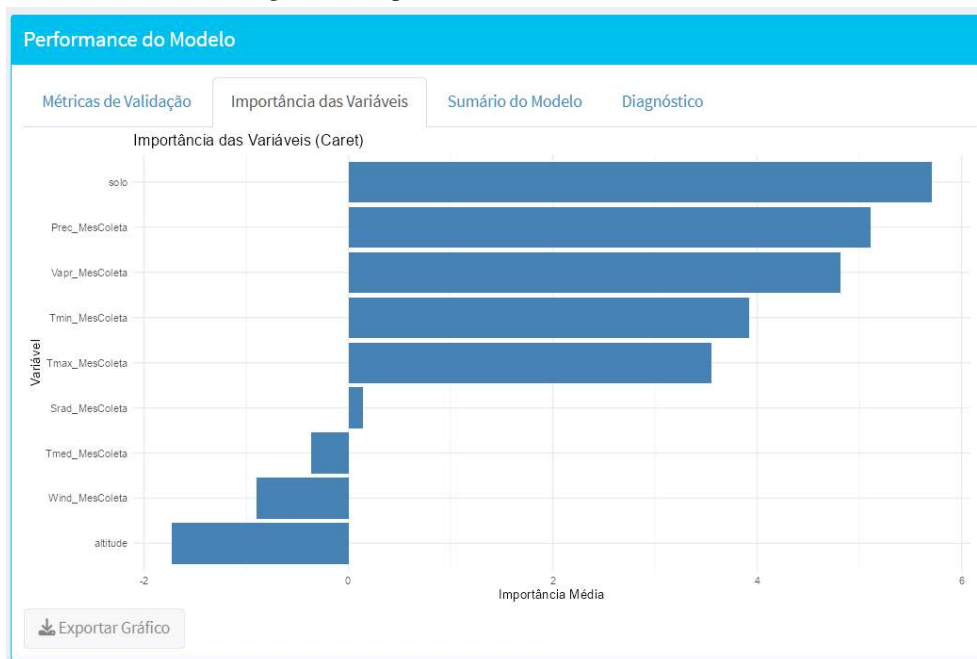
4 RESULTADOS E DISCUSSÕES

No recorte analisado neste estudo, foram utilizados 106 registros referentes ao composto Xantonas, modelados com 9 variáveis preditivas: altitude, Prec_MesColeta, solo, Srad_MesColeta, Tmax_MesColeta, Tmed_MesColeta, Tmin_MesColeta, Vapr_MesColeta e Wind_MesColeta.

O modelo de *Random Forest* foi ajustado com $n_{tree} = 500$ e $m_{try} = 3$, utilizando validação cruzada 5-fold repetida 3 vezes. As métricas de desempenho na validação cruzada indicaram RMSE = 2.4102, MAE = 1.956 e R-quadrado (Coeficiente de Determinação - R^2) = 0.0627, enquanto a variância explicada *Out-of-Bag* (OOB) do modelo final foi de -26,29%. Esses resultados sugerem que, embora o modelo capture parte da estrutura dos dados, ainda apresenta baixo poder explicativo global, reforçando a necessidade de refinamento paramétrico e ampliação da base de dados.

A Figura 3 apresenta a importância relativa das variáveis preditivas no modelo *Random Forest*, evidenciando que solo, precipitação e pressão de vapor foram os principais determinantes da atividade prevista.

Figura 3 – Importância das Variáveis Preditivas

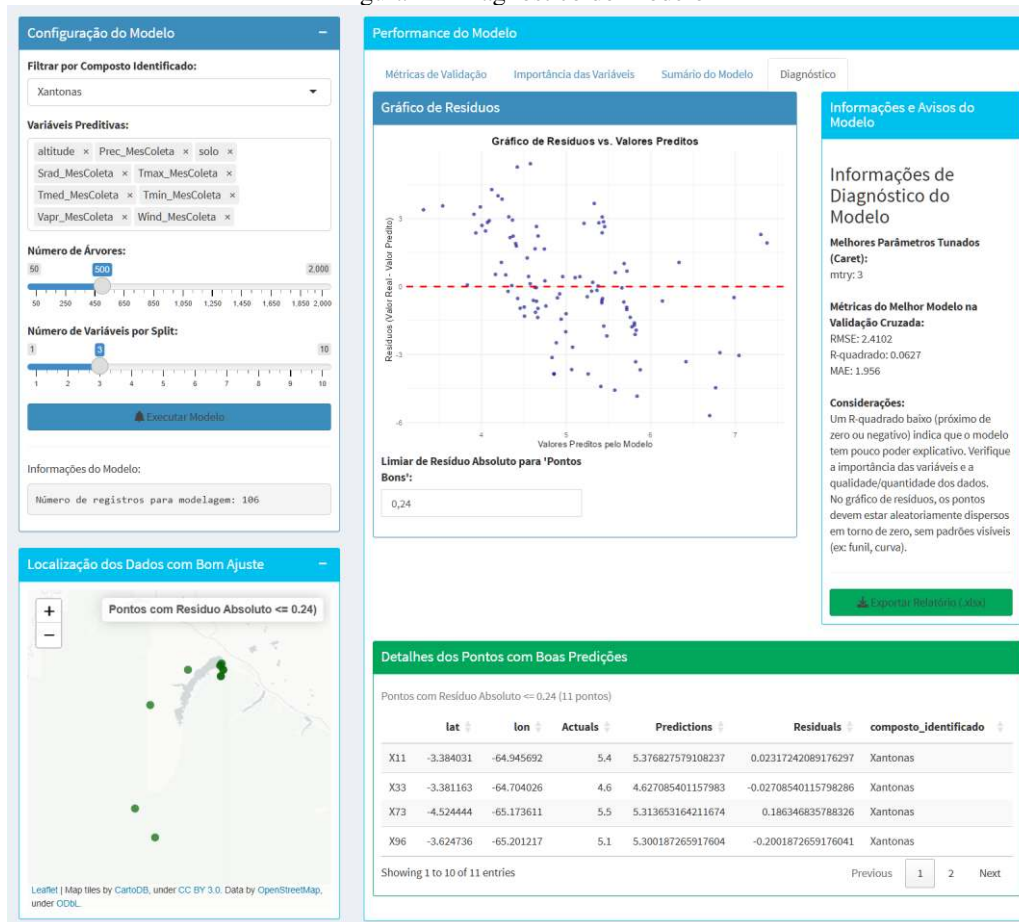


Fonte: Elaborado pelos autores (2026).

Para fins de diagnóstico, adotou-se um limiar de resíduo absoluto de 0,24, valor correspondente a aproximadamente 10% do RMSE (2,41) e 12% do MAE (1,96) do modelo. Esse valor decorre diretamente do critério metodológico previamente estabelecido, segundo o qual previsões cujo erro absoluto seja substancialmente inferior ao erro médio podem ser classificadas como de alta confiabilidade local.



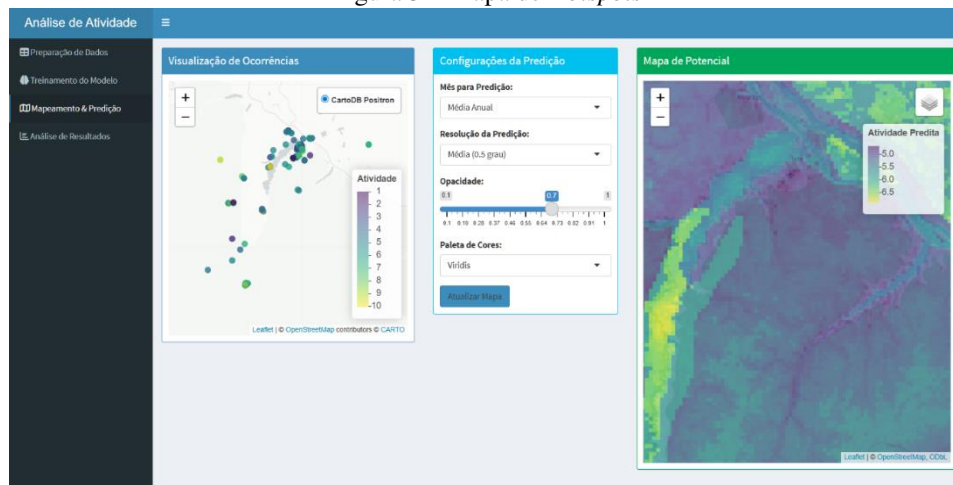
Figura 4 – Diagnóstico do Modelo



Fonte: Elaborado pelos autores (2026).

A Figura 5 apresenta o mapa de *hotspots* gerado pelo modelo, destacando as áreas com maior probabilidade de ocorrência de atividade biológica elevada. Observa-se a formação de gradientes espaciais coerentes com os padrões ambientais da região.

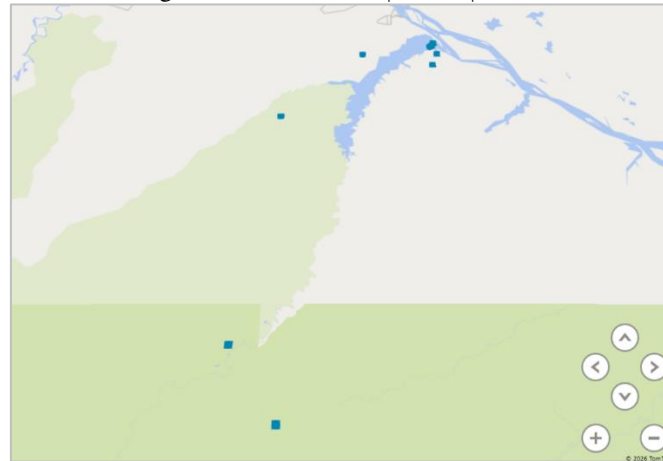
Figura 5 – Mapa de *Hotspots*



Fonte: Elaborado pelos autores (2026).

Sob esse critério conservador, foram identificados os pontos com $|\text{resíduo}| \leq 0,24$, os quais representam predições com erro muito inferior ao erro típico do modelo e, portanto, constituem candidatos prioritários para prospecção, mesmo diante do baixo número de amostras ($n = 106$) e da variância explicada reduzida. Esses pontos funcionam como ilhas de estabilidade preditiva, úteis para orientar ajustes futuros e para demonstrar que, apesar das limitações globais, o modelo apresenta potencial preditivo local válido.

Figura 6 – Pontos com $|\text{resíduo}| \leq 0,24$



Fonte: Elaborado pelos autores (2026).

As limitações observadas podem estar associadas à quantidade restrita de amostras, que pode não ser suficiente para capturar a complexidade das relações entre variáveis ambientais e a atividade. Além disso, a escolha de somente quatro variáveis ambientais pode limitar a explicação dos padrões observados, uma vez que fatores ecológicos, biológicos ou antrópicos não considerados podem influenciar significativamente os resultados. Para fortalecer o modelo, algumas estratégias podem ser adotadas:

- Expansão do conjunto de dados, tanto em quantidade e diversidade de amostras quanto em incorporação de variáveis adicionais (topográficas, uso e cobertura do solo, proximidade a corpos d'água, infraestrutura humana, variáveis biológicas e sociais).
- Configurar os hiperparâmetros do *Random Forest* para investigar diversas combinações de *mtry* e *ntree*, identificando a configuração mais apropriada.
- Testes com outros algoritmos de aprendizado, como modelos aditivos generalizados (GAMs), máquinas de vetores de suporte (SVMs) e abordagens baseadas em *boosting*, para avaliar métodos alternativos de modelagem preditiva, ajustes que podem contribuir significativamente para a melhoria do desempenho preditivo, possibilitando uma maior precisão e confiabilidade nas estimativas da atividade biológica.



5 CONCLUSÃO

A bioprospecção encontra na integração da inteligência artificial com dados espaciais, uma abordagem ajustada e conveniente para a busca de ativos biológicos relevantes. A metodologia racionaliza atividades que, na forma tradicional, exigiriam longos períodos para planejamento, altos custos e muitos recursos logísticos, auxiliando a identificação de áreas de maior potencial, reduzindo custos, encurtando prazos de análises e expandindo o alcance de coletas, tudo isso a favor de uma exploração mais efetivada da biodiversidade.

Por sua natureza interativa e escalável, a solução facilita o acesso a análises avançadas e organiza fluxos de trabalho que fortalecem a tomada de decisão em contextos de prospecção. Além disso, a metodologia empregada minimiza vieses na seleção de áreas prioritárias, promovendo avaliações mais objetivas e fundamentadas em padrões identificados pelo modelo preditivo, o que aumenta a confiabilidade dos resultados.

Diferentemente da abordagem tradicional, baseada predominantemente na experiência empírica de campo, o modelo orienta a prospecção para hotspots de biodiversidade inferidos a partir de variáveis ambientais como precipitação, altitude e cobertura vegetal. Embora a relevância absoluta de cada achado ainda não possa ser quantificada diretamente, a priorização de áreas com maior probabilidade de conter espécies endêmicas ou subexploradas se alinha aos objetivos estratégicos da bioeconomia.

Mesmo sem dados absolutos de custos, o ábaco comparativo indica que a racionalização das expedições proporcionada pelo modelo tem potencial para reduzir significativamente o custo por descoberta relevante e diminuir o tempo entre o planejamento e a execução de novas coletas. O custo de oportunidade associado a equipes especializadas, deslocamentos logísticos e tentativas malsucedidas pode, assim, ser substancialmente mitigado.

A pesquisa realizada não apenas confirma a viabilidade e a utilidade da abordagem proposta, como também estabelece uma base metodológica sólida para estudos futuros. Os resultados obtidos reforçam o potencial da inteligência artificial como ferramenta estratégica para ampliar a eficiência, a precisão e o alcance das iniciativas de bioprospecção voltadas à bioeconomia.

AGRADECIMENTOS

O presente trabalho foi realizado no âmbito do Programa de Pós-Graduação em Biotecnologia (PPGBIOTEC) da Universidade Federal do Amazonas, com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES).



REFERÊNCIAS

- ALLOUN, Wiem; CALVIO, Cinzia. Bio-Driven Sustainable Extraction and AI-Optimized Recovery of Functional Compounds from Plant Waste: A Comprehensive Review. **Fermentation**, [s. l.], v. 10, n. 3, p. 126, 2024.
- AMBULI, T.V *et al.* AI-Driven Financial Management Optimizing Investment Portfolios through Machine Learning. *In: 2024 7TH INTERNATIONAL CONFERENCE ON CIRCUIT POWER AND COMPUTING TECHNOLOGIES (ICCPCT)*, 2024, Kollam, India. **2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)**. Kollam, India: IEEE, 2024. p. 1822–1828. Disponível em: <https://ieeexplore.ieee.org/document/10672859/>. Acesso em: 21 jun. 2025.
- BOARDMAN, Anthony E *et al.* **Cost-benefit Analysis : Concepts and Practice**. [S. l.]: Cambridge University Press, 2018.
- CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, [s. l.], v. 7, n. 3, p. 1247–1250, 2014.
- CHAN, W. K. C. *et al.* Asset Optimization Using Integrated Modeling, Optimization, and Artificial Intelligence. [s. l.], 2019. Disponível em: <https://patents.google.com/patent/US20200387818A1/en>.
- CHOUBEY, Sachin; KARMAKAR, G. P. Artificial Intelligence Techniques and Their Application in Oil and Gas Industry. **Artificial Intelligence Review**, [s. l.], v. 54, n. 54, 2020.
- CUTLER, D. Richard *et al.* Random Forests for Classification in Ecology. **Ecology**, [s. l.], v. 88, n. 11, p. 2783–2792, 2007.
- DANACH, Kassem; DIRANI, Ali El; RKEIN, Hassan. Revolutionizing Supply Chain Management With AI: A Path to Efficiency and Sustainability. **IEEE Access**, [s. l.], v. 12, p. 188245–188255, 2024.
- FAHEEM, Muhammad Ashraf; ASLAM, Muhammad; KAKOLU, Sridevi. Artificial Intelligence in Investment Portfolio Optimization: A Comparative Study of Machine Learning Algorithms. **International Journal of Science and Research Archive**, [s. l.], v. 6, n. 1, p. 335–342, 2022.
- GIANNINI, Tereza C. *et al.* Desafios Atuais Da Modelagem Preditiva De Distribuição De Espécies. **Rodriguésia**, [s. l.], v. 63, n. 3, p. 733–749, 2012.
- GUPTA, Supriya; SHARMA, Abhishek; ABUBAKAR, Aria. Artificial Intelligence–Driven Asset Optimizer. *In: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION*, 2018, Dallas, Texas, USA. **SPE Annual Technical Conference and Exhibition**. Dallas, Texas, USA: SPE, 2018. p. D012S045R001. Disponível em: <https://onepetro.org/SPEATCE/proceedings/18ATCE/18ATCE/D012S045R001/213656>. Acesso em: 21 jun. 2025.
- HASAN, Sobuj; SHAFIQ, Sadman Bin; KHATUN, Lubna. EXPLORING THE POTENTIAL OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN MINERAL EXPLORATION: A REVIEW ARTICLE. **International Research Journal of Modernization in Engineering Technology and Science**, [s. l.], v. 5, n. 10, 2023.



JIA, Lihua *et al.* Development of interactive biological web applications with R/Shiny. **Briefings in Bioinformatics**, [s. l.], v. 23, n. 1, p. bbab415, 2022.

JOEL, Olorunyomi Stephen *et al.* Leveraging Artificial Intelligence for Enhanced Supply Chain Optimization: A Comprehensive Review of Current Practices and Future Potentials. **International Journal of Management & Entrepreneurship Research**, [s. l.], v. 6, n. 3, p. 707–721, 2024.

JOHNSON, Ben. AI biotechs launch bioprospecting expeditions with Indigenous groups, agree to share benefits. **Nature Biotechnology**, [s. l.], v. 43, n. 2, p. 154–156, 2025.

KARAMITROU, Alexandra *et al.* Towards the Use of Artificial Intelligence Deep Learning Networks for Detection of Archaeological Sites. **IOP Publishing**, [s. l.], v. 10, n. 4, p. 044001–044001, 2022.

KÜÇÜKDEMIRCI, Melda; SARRIS, Apostolos. GPR Data Processing and Interpretation Based on Artificial Intelligence Approaches: Future Perspectives for Archaeological Prospection. **Remote Sensing**, [s. l.], v. 14, n. 14, p. 3377, 2022.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. Corrected at 5th printing. New York: Springer, 2016.

LI, Shi; CHEN, Jianping; LIU, Chang. Overview on the Development of Intelligent Methods for Mineral Resource Prediction under the Background of Geological Big Data. **Minerals**, [s. l.], v. 12, n. 5, p. 616, 2022.

MEDIANSYAH, Iski; SEPTIAN, Firza; ZIKRY, Arief. Penerapan Whale Optimization Algorithm dalam Pengoptimalan Portofolio Investasi Menggunakan Model Prediktif Artificial Intelligence. **Jurnal Software Engineering and Computational Intelligence**, [s. l.], v. 2, n. 01, p. 50–58, 2024.

MISHRA, Manoranjan *et al.* Spatial analysis and machine learning prediction of forest fire susceptibility: a comprehensive approach for effective management and mitigation. **Science of The Total Environment**, [s. l.], v. 926, p. 171713, 2024.

MONIGHA, Adogioye. AI-Driven Cost Optimization in Oil and Gas Projects. **International Journal of Petroleum and Gas Engineering Research**, [s. l.], v. 7, n. 2, p. 17–32, 2024.

POMEROY, Brett; GRILC, Miha; LIKOZAR, Blaž. Artificial neural networks for bio-based chemical production or biorefining: A review. **Renewable and Sustainable Energy Reviews**, [s. l.], v. 153, p. 111748, 2022.

PRAJAPATI, Parixit *et al.* Deep Learning Approaches for Predicting Bioactivity of Natural Compounds. **The Natural Products Journal**, [s. l.], v. 15, 2025. Disponível em: <https://www.eurekaselect.com/238324/article>. Acesso em: 10 jun. 2025.

RANE, Nitin; CHOUDHARY, Saurabh; RANE, Jayesh. Artificial intelligence and machine learning for resilient and sustainable logistics and supply chain management. **SSRN Electronic Journal**, [s. l.], 2024. Disponível em: <https://www.ssrn.com/abstract=4847087>. Acesso em: 21 jun. 2025.

SAMARAH, Tariq *et al.* Intelligent Strategic Decision-Making for Optimized Project Management. *In*: 2024 INTERNATIONAL CONFERENCE ON DECISION AID SCIENCES AND APPLICATIONS (DASA), 2024, Manama, Bahrain. **2024 International Conference on Decision**



Aid Sciences and Applications (DASA). Manama, Bahrain: IEEE, 2024. p. 1–5. Disponível em: <https://ieeexplore.ieee.org/document/10836504/>. Acesso em: 21 jun. 2025.

SANTOS, Bruna Gava Dos. **Modelagem De Corredores Ecológicos No Estado Do Paraná E Análise Comparativa Com as Áreas Estratégicas Para a Conservação E Restauração Da Biodiversidade**. 2024. 48 f. - UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ, [s. l.], 2024. Disponível em: Acesso em: 10 jun. 2025.

SATHUPADI, Kaushik *et al.* RAP-Optimizer: Resource-Aware Predictive Model for Cost Optimization of Cloud AIaaS Applications. **Electronics**, [s. l.], v. 13, n. 22, p. 4462, 2024.

SIEVERT, Carson. **Interactive Web-Based Data Visualization with R, plotly, and shiny**. 1. ed. [S. l.]: Chapman and Hall/CRC, 2020. Disponível em: <https://www.taylorfrancis.com/books/9780429824210>. Acesso em: 10 jun. 2025.

TAMIRU, Habtamu; WAGARI, Meseret; TADESE, Bona. An Integrated Artificial Intelligence and GIS Spatial Analyst Tools for Delineation of Groundwater Potential Zones in Complex terrain: Fincha Catchment, Abay Basi, Ethiopia. **Air, Soil and Water Research**, [s. l.], v. 15, n. 15, p. 117862212110459–117862212110459, 2022.

VERMA, Pranay *et al.* ARON: Adaptive Resource Optimization Network for AI-Driven Business Management. *In*: 2024 4TH INTERNATIONAL CONFERENCE ON TECHNOLOGICAL ADVANCEMENTS IN COMPUTATIONAL SCIENCES (ICTACS), 2024, Tashkent, Uzbekistan. **2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)**. Tashkent, Uzbekistan: IEEE, 2024. p. 61–66. Disponível em: <https://ieeexplore.ieee.org/document/10840557/>. Acesso em: 21 jun. 2025.

WANG, Luoqi *et al.* Enhancing mineral prospectivity mapping with geospatial artificial intelligence: A geographically neural network-weighted logistic regression approach. **International Journal of Applied Earth Observation and Geoinformation**, [s. l.], v. 128, p. 103746–103746, 2024.

WILLMOTT, Cj; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate Research**, [s. l.], v. 30, p. 79–82, 2005.

WONG, Tzu-Tsung; YEH, Po-Yang. Reliable Accuracy Estimates from *k*-Fold Cross Validation. **IEEE Transactions on Knowledge and Data Engineering**, [s. l.], v. 32, n. 8, p. 1586–1594, 2020.

YESUFU, Lawal; ALAJLANI, Sami. Enhancing Strategic Decision-Making With AI: Unveiling the Untapped Potential. *In*: GÜNDAR, Altuğ (org.). **Advances in Finance, Accounting, and Economics**. [S. l.]: IGI Global, 2025. p. 415–436. Disponível em: <https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/979-8-3693-7036-0.ch018>. Acesso em: 21 jun. 2025.

ZHU, Xiaohui; HU, Chunhua; LUO, Ling. Research on Intelligent Portfolio Optimization Model Based on Big Data and Artificial Intelligence. *In*: 2024 INTERNATIONAL CONFERENCE ON INTERACTIVE INTELLIGENT SYSTEMS AND TECHNIQUES (IIST), 2024, Bhubaneswar, India. **2024 International Conference on Interactive Intelligent Systems and Techniques (IIST)**. Bhubaneswar, India: IEEE, 2024. p. 160–166. Disponível em: <https://ieeexplore.ieee.org/document/10734543/>. Acesso em: 21 jun. 2025.



ZÜLFIKAROĞLU, Sarp. AI-Driven Strategic Management and Decision Making for Energy Sector.
Next Frontier For Life Sciences and AI, [s. l.], v. 8, n. 1, p. 91, 2024.

