



MÉTODO DE AVALIAÇÃO DE DESEMPENHO DE MODELOS DE APRENDIZADO SUPERVISIONADO POR DIFERENTES VARIÁVEIS

João P. do N. SILVA¹; Diego SAQUI²; Allan A. PEREIRA³

RESUMO

Este artigo aborda a relevante questão da otimização de hiperparâmetros em algoritmos de inteligência artificial supervisionados, com foco especial em métodos como o *GridSearch*. Exploramos a busca pelos melhores hiperparâmetros, tanto com base em uma única métrica quanto em uma análise de múltiplas métricas. Ao considerar a complexidade dessa tarefa, oferecemos insights valiosos sobre o aprimoramento do desempenho e eficiência dos modelos de IA supervisionados, impulsionando avanços nas aplicações práticas dessas tecnologias. Os resultados obtidos fornecem uma base sólida para o desenvolvimento de soluções mais eficazes e precisas em diversas áreas de pesquisa e aplicações do mundo real.

Palavras-chave:

Inteligência Artificial; Otimização; Hiperparâmetros; Avaliação de Modelos.

1. INTRODUÇÃO

Algoritmos de Inteligência Artificial (IA) tem se tornado amplamente utilizados em diversas áreas, desde previsão de dados meteorológicos até diagnóstico médico (JUNG, 2021). A expansão desses algoritmos levou a uma variedade de métodos e aplicabilidades, incluindo o Aprendizado de Máquina (ML), uma subárea da IA que visa melhorar o desempenho dos modelos por meio de experiência com dados. No contexto do ML supervisionado, os modelos são treinados com instâncias rotuladas, permitindo que aprendam com exemplos e façam previsões precisas em novos dados. No entanto, para alcançar a máxima eficiência desses modelos, é crucial considerar a escolha adequada dos hiperparâmetros. Hiperparâmetros são parâmetros externos ao próprio modelo, que precisam ser ajustados manualmente antes do treinamento. Eles influenciam o processo de treinamento e podem afetar significativamente o desempenho final do algoritmo. Neste projeto, propomos uma estrutura metodológica para analisar e identificar a influência dos hiperparâmetros no desempenho dos algoritmos de ML, com o objetivo de aprimorar a acurácia e a eficiência desses modelos em suas diversas aplicações. Conforme destacado por Quaresma (2018, p. 69-84), a diversidade de algoritmos de IA e suas aplicações resultou em modelos desenvolvidos para distintas finalidades, cada um com seus próprios parâmetros, métricas e objetivos específicos. Dentre esses modelos, o Aprendizado de Máquina supervisionado se destaca como uma abordagem que requer a otimização cuidadosa de hiperparâmetros para obter um desempenho máximo. No entanto, a

¹ Discente do Bacharelado em Ciência da Computação, IFSULDEMINAS – Campus Muzambinho. E-mail: jpons1118 @ gmail.com

² Orientador, IFSULDEMINAS – Campus Muzambinho. E-mail: diego.saqui@muz.ifsuldeminas.edu.br

³ Coordenador, IFSULDEMINAS – Campus Muzambinho. E-mail: allan.pereira@ifsuldeminas.edu.br

escolha adequada dos hiperparâmetros pode ser um desafio, como menciona Bashir (2020, p. 347-358), que alerta para o risco de *overfitting*, quando o modelo se ajusta muito bem aos dados de treinamento, mas não generaliza bem para novos dados, e *underfitting*, quando o modelo é muito simples para capturar a complexidade dos dados, resultando em um desempenho fraco. Neste contexto, nosso estudo propõe uma estrutura de metodologias para identificar e analisar como diferentes hiperparâmetros afetam o desempenho dos algoritmos de ML supervisionado, a fim de melhorar a eficiência desses modelos. Com base nos resultados obtidos, buscamos oferecer insights valiosos para aprimorar as aplicações práticas dessas técnicas em diferentes domínios, minimizando o risco de consequências drásticas.

2. FUNDAMENTAÇÃO TEÓRICA

A IA é uma área da ciência da computação que visa desenvolver modelos e algoritmos capazes de simular o pensamento humano, permitindo que as máquinas realizem tarefas complexas e aprendam a partir dos dados. Ela abrange uma variedade de técnicas, como Aprendizado de Máquina e Redes Neurais Artificiais, e tem sido amplamente aplicada em diversos setores, desde assistentes virtuais e veículos autônomos até diagnósticos médicos e previsões meteorológicas.

Com o avanço tecnológico, o desenvolvimento de modelos de IA supervisionados têm sido especialmente relevantes, onde algoritmos são treinados com instâncias rotuladas para fazer previsões precisas em novos dados. O entendimento adequado das métricas de avaliação é fundamental para garantir que os modelos atinjam o desempenho desejado em aplicações do mundo real. Os hiperparâmetros são parâmetros externos ao modelo de Aprendizado de Máquina supervisionado que desempenham um papel fundamental na otimização do desempenho do algoritmo. Diferentemente dos parâmetros internos do modelo, que são aprendidos durante o treinamento, os hiperparâmetros devem ser definidos antes do processo de treinamento e podem afetar significativamente a capacidade do modelo de generalizar e fazer previsões precisas em novos dados. A escolha adequada dos hiperparâmetros é essencial para evitar problemas como *overfitting*, quando o modelo se ajusta demais aos dados de treinamento, ou *underfitting*, quando o modelo não é capaz de capturar a complexidade dos dados. A análise criteriosa e a busca pelos melhores hiperparâmetros são elementos cruciais para o aprimoramento da eficiência e acurácia dos modelos de ML supervisionado, impactando diretamente em suas aplicações práticas em diferentes domínios.

3. MATERIAL E MÉTODOS

Para realizar este estudo, utilizou-se o *Google Colab*⁴ como plataforma de desenvolvimento

⁴ Disponível em: <<https://acesse.one/BNpYR>>. Último acesso em 10/08/2023

e o dataset "*Iris*" disponível no *Kaggle*⁵, que consiste em um conjunto de dados amplamente utilizado na área de Aprendizado de Máquina. O dataset contém 150 instâncias, cada uma com 6 atributos: *Id*, *SepalLengthCm* (comprimento da sépala em cm), *SepalWidthCm* (largura da sépala em cm), *PetalLengthCm* (comprimento da pétala em cm), *PetalWidthCm* (largura da pétala em cm) e *Species* (espécie pertencente ao indivíduo). O atributo "*Species*" é o rótulo esperado para o problema de classificação supervisionada, onde o modelo deve aprender a classificar cada instância corretamente em uma das espécies presentes no dataset. Após a preparação dos dados e a seleção dos algoritmos de IA para treinamento, utilizou-se o *GridSearch*⁶ para otimizar os hiperparâmetros de diferentes algoritmos, como o *Random Forest*⁷, *Decision Tree*⁸, *Perceptron*⁹, *Linear Regression*¹⁰ e *KNN*¹¹. O *GridSearch* é uma técnica que busca as melhores combinações de hiperparâmetros em um espaço pré-definido, a fim de encontrar os valores que maximizam o desempenho do modelo. Inicialmente, o *GridSearch* foi aplicado para encontrar os conjuntos de hiperparâmetros que maximizam o desempenho de cada modelo em relação às métricas selecionadas. Em seguida, ele foi expandido para otimizar conjuntos de hiperparâmetros simultaneamente, visando encontrar as combinações ideais que proporcionam um desempenho ainda mais aprimorado dos modelos.

Para analisar os resultados obtidos, foram aplicados o teste de *Friedman* e o Teste de *Nemenyi*. O teste de *Friedman* é uma análise estatística não paramétrica que verifica se há diferenças significativas nos desempenhos dos algoritmos em relação às métricas avaliadas. Se os resultados de um grupo diferem dos demais, o teste de *Nemenyi* é aplicado para realizar comparações múltiplas entre os algoritmos e gerar uma matriz de comparação dos resultados. Essa matriz fornece informações valiosas sobre quais algoritmos apresentaram diferenças estatisticamente significativas em relação ao desempenho.

4. RESULTADOS E DISCUSSÃO

Ao aplicar o *GridSearch*, os algoritmos *KNeighborsClassifier* e *RandomForestClassifier* demonstraram melhor adaptação ao problema de classificação multiclasse, obtendo resultados mais expressivos, enquanto o *DecisionTreeClassifier* apresentou bons resultados tanto em datasets de classificação binária quanto multiclasse. Já o *Perceptron*, de aplicação exclusiva em classificação binária, apresentou desempenho inferior nos testes com dataset multiclasse. Adicionalmente, a aplicação do *GridSearch* no algoritmo *KNeighborsClassifier* resultou em uma melhoria

⁵ Disponível em: <<https://www.kaggle.com/datasets/uciml/iris>>. Último acesso em 10/08/2023

⁶ Disponível em: <<https://urx1.com/LvRa9>>. Último acesso em 10/08/2023

⁷ Disponível em: <<https://urx1.com/sQAbw>>. Último acesso em 10/08/2023

⁸ Disponível em: <<https://ury1.com/QKkLc>>. Último acesso em 10/08/2023

⁹ Disponível em: <<https://ury1.com/vJdj6>>. Último acesso em 10/08/2023

¹⁰ Disponível em: <<https://11nq.com/TB2mN>>. Último acesso em 10/08/2023

¹¹ Disponível em: <<https://urx1.com/RWoAe>>. Último acesso em 10/08/2023

significativa nas métricas de *Accuracy* e *Recall*, conforme indicado na tabela 1.

<i>Algoritmos</i>	<i>Perceptron</i>	<i>Decision Tree</i>	<i>Logistic Regression</i>	<i>Random Forest</i>	<i>KNN</i>
<i>Accuracy</i>	0.63	0.9	0.96	0.96	0.93
<i>Recall</i>	0.66	0.92	0.96	0.94	0.93
<i>Precision</i>	0.40	0.92	0.96	0.97	0.92
<i>F1_Score</i>	0.48	0.91	0.96	0.95	0.92

Tabela 1: Métricas por classificador

A estatística de *Friedman* obtida foi de 8.2, com um valor p de 0.042. Esses valores são menores que o alfa padrão de 0,05 indicando que há evidências estatísticas suficientes para rejeitar a hipótese nula de que não há diferença significativa entre os algoritmos em termos de desempenho geral, logo existe uma diferença significativa entre um algoritmo e os demais ao comparar suas métricas. Em seguida foi utilizado o método de *Nemenyi* para verificar quais dos algoritmos apresentam uma maior diferença em seus resultados ao serem comparados, onde gerou uma matriz de comparação de similaridade entre os algoritmos onde quanto maior o valor, maior sua similaridade, como pode ser observado na tabela 2. Dessa forma podemos concluir que as duplas de algoritmos *KNN* com *Logistic Regression* ou com *Decision Tree* apresentam maior similaridade, porém ao serem comparados aos demais, não são iguais a um nível de significância de 5%, principalmente com o perceptron, logo a aplicação do *GridSearch* é capaz de melhorar as métricas de algoritmos de ML supervisionados quando aplicado nos algoritmos corretos.

<i>Algoritmos</i>	<i>KNN</i>	<i>Logistic Regression</i>	<i>Decision Tree</i>	<i>Perceptron</i>
<i>KNN</i>	1	0.75	0.57	0.39
<i>Logistic Regression</i>	0.75	1	0.11	0.05
<i>Decision Tree</i>	0.57	0.11	1	0.9
<i>Perceptron</i>	0.39	0.05	0.9	1

Tabela 2: Tabela gerada pelo teste de *Nemenyi*

REFERÊNCIAS

BASHIR, Daniel et al. An information-theoretic perspective on overfitting and underfitting. In: Australasian Joint Conference on Artificial Intelligence. Springer, Cham, 2020. p. 347-358.

JUNG, Jinha et al. The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. *Current Opinion in Biotechnology*, v. 70, p. 15-22, 2021.

QUARESMA QUARESMA, Alexandre. Inteligências artificiais e os limites da computação. *PAAKAT: revista de tecnologia y sociedad*, v. 8, n. 15, p. 69-84, 2018.