CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA

SOUZA - CEETEPS

Faculdade de Tecnologia Dr. Archimedes Lammoglia – FATEC/Indaiatuba

ALOCAÇÃO ESTRATÉGICA DE RECURSOS PARA MAXIMIZAR O IMPACTO DO PROGRAMA ESPORTE CIDADÃO

FATEC-INDAIATUBA

Pedro Henrique Bonomo Santos

Profª Sérgio Gustavo Pereira Medina

Indaiatuba

Junho/ 2025

**RESUMO**

Durante o semestre, a pesquisa cujo título é **Alocação estratégica de recursos para maximizar o impacto do programa Esporte Cidadão** avançou em direção ao seu objetivo geral, que é investigar por meio de técnicas de ciência de dados, como o programa pode ser aprimorado para maximizar seus efeitos perante a sociedade e a educação. Dentre os **objetivos específicos, destacam-se:**

* Identificar recursos que estão sendo sobrecarregados e recursos que estão sendo mal utilizados;
* Levantar dados relevantes para o problema;
* Elaborar um modelo analítico ou preditivo com base nos dados coletados;
* Analisar os resultados obtidos para tomar uma decisão sobre qual melhor abordagem.

A pesquisa será conduzida com base na abordagem metodológica **Design Science Research (DSR)**, que orienta a construção e avaliação de artefatos voltados à resolução de problemas reais. O DSR é especialmente útil em contextos educacionais com uso de tecnologias, permitindo rigor nos ciclos de desenvolvimento e validação de soluções aplicadas (ANGELUCI et al., 2021). Nesse sentido, busca-se desenvolver um modelo de alocação estratégica de recursos para o programa Esporte Cidadão. Para o processamento e exploração dos dados, serão utilizados o ambiente de programação **Python** (com bibliotecas como matplotlib, pandas e scikit-learn) e a plataforma visual **Orange**, que permitiu a aplicação de algoritmos de aprendizado de máquina de forma intuitiva e interativa. Serão exploradas técnicas como (ainda vou ver as melhores técnicas a serem utilizadas), conforme discutido por (adicionar artigo relevante).

O **Esporte Cidadão** é um programa público de formação esportiva voltado a crianças e adolescentes, com foco em promover saúde, disciplina e inclusão social por meio da prática regular de esportes. Ele funciona em diferentes polos espalhados pela cidade, atendendo centenas de participantes gratuitamente e oferecendo modalidades esportivas variadas. A correta alocação de recursos, como instrutores, equipamentos e horários de quadra, é essencial para garantir o impacto positivo do programa.

Para apoiar a análise dos dados e o desenvolvimento do modelo proposto, serão utilizadas duas ferramentas principais: o ambiente de programação **Python**, com bibliotecas como *pandas*, *matplotlib* e *scikit-learn*, e a plataforma visual **Orange Data Mining**. O Python foi responsável pela preparação dos dados e execução dos modelos, enquanto o Orange, por sua interface gráfica intuitiva baseada em blocos, facilitou a prototipação rápida e visualização dos resultados.

**OBJETIVOS**

* 1. **Objetivo Geral**

O objetivo central desta pesquisa é **investigar, por meio de técnicas de ciência de dados e da aplicação do método Design Science Research (DSR), como otimizar a alocação de recursos do programa Esporte Cidadão**, visando maximizar seu impacto social e educacional.

A partir disso, busca-se desenvolver um **modelo analítico ou preditivo** que auxilie os gestores do programa a tomar decisões mais assertivas quanto à distribuição de recursos, pensando em fatores como demanda, participação, evasão e impacto gerado nas comunidades.

Esse objetivo geral está alinhado com a crescente demanda por **gestão pública baseado em dados e evidências**, e com a necessidade de tornar as políticas públicas direcionadas e mais eficientes às reais necessidades da população.

* 1. **Objetivos Específicos**
     1. **Identificar recursos que estão sendo sobrecarregados e recursos que estão sendo mal utilizados**

O primeiro objetivo específico busca realizar um diagnóstico detalhado sobre a atual utilização dos recursos do programa. Isso inclui **analisar a distribuição de professores, equipamentos, espaços físicos, horários e demais recursos disponíveis.**

A partir dos dados coletados, será possível verificar quais recursos estão **sobrecarregados** e quais estão sendo **subutilizados ou ociosos,** prejudicando a efetividade do programa.

Essa análise inicial é **fundamental** para fornecer uma visão clara da situação, identificando os aspectos a serem melhorados.

* + 1. **Levantar dados relevantes para o problema**

Este objetivo tem como foco específico a **coleta e organização dos dados necessários para o desenvolvimento da pesquisa.** Isso inclui a definição de quais variáveis serão analisadas (como frequência, número de vagas, dados sociodemográficos, histórico de evasão etc.). Além da coleta, esta etapa a **limpeza e preparação** dos dados, garantindo que estejam organizados e prontos para serem utilizados para análises posteriores.

* + 1. **Elaborar um modelo analítico ou preditivo com base nos dados coletados**

Após o levantamento e tratamento de dados, o próximo objetivo será a elaboração de um **modelo que permita analisar padrões e prever cenários futuros.** Esse modelo poderá ser um modelo **preditivo de evasão** (utilizando técnicas de machine learning como árvore de decisão ou regressão logística), ou **um modelo de clusterização** (como **k-means**), para identificar perfis de participantes com características semelhantes. O objetivo visa criar um artefato que auxilie a gestão a visualizar cenários e tomar decisões com base em dados concretos.

* + 1. **Analisar os resultados obtidos para tomar uma decisão sobre qual melhor abordagem**

O próximo passo é realizar uma análise detalhada dos resultados gerados pelo modelo, comparando diferentes técnicas e métricas de desempenho.

A ideia é avaliar qual abordagem gera os **melhores resultados possíveis** para os responsáveis pelo programa, considerando não apenas as estatísticas, mas também a facilidade de aplicação com contexto real.

Ao final desta etapa, será possível apresentar **recomendações práticas**, apontando ações de realocação de recursos, intervenções estratégicas ou até mesmo ajustes no próprio funcionamento do programa, com o intuito de maximizar seu impacto social.

**INTRODUÇÃO**

Em programas sociais como o *Esporte Cidadão*, a eficiência na alocação de recursos é primordial para garantir a efetividade necessária e o impacto positivo pretendido. No entanto, a ausência de métodos baseados em evidências para orientar uma melhor distribuição, pode comprometer a efetividade educacional e o alcance social. Nesse contexto, surge a **pergunta-chave** que orienta essa pesquisa: **Como alocar estrategicamente os recursos do programa Esporte Cidadão, com base em dados, para maximizar seu impacto social e educacional?**

A resposta a essa questão requer o uso de métodos sistemáticos que associem análise de dados à construção de soluções práticas. Por isso, esta pesquisa se fundamenta na metodologia **Design Science Research (DSR)**, que propõe a criação e avaliação de artefatos voltados à resolução de problemas reais em contextos específicos (ANGELUCI et al., 2021). Essa abordagem é especialmente apropriada em estudos aplicados à educação com o uso de tecnologias, pois garante rigor científico ao desenvolvimento de soluções práticas.

Além disso, o uso de **ciência de dados** em políticas públicas tem se mostrado uma estratégia promissora para embasar decisões com base em evidências. Segundo Han, Pei e Kamber (2022), o uso de modelos preditivos e análises exploratórias permite extrair conhecimento útil de grandes volumes de dados, auxiliando gestores a direcionar ações de forma mais eficiente.

A justificativa para a realização deste trabalho reside na importância social do *Esporte Cidadão* como política pública voltada à formação cidadã por meio do esporte, e na oportunidade de **otimizar seu impacto** por meio da inteligência analítica. Ao identificar recursos mal distribuídos ou sobrecarregados, e propor um modelo baseado em dados, esta pesquisa pode contribuir com uma gestão mais eficaz e orientada a resultados. Mas afinal, **o que é o esporte?**

Segundo Valdir José Barbanti (2006), *“esporte é uma atividade competitiva institucionalizada que envolve esforço físico vigoroso ou o uso de habilidades motoras relativamente complexas, por indivíduos, cuja participação é motivada por uma combinação de fatores intrínsecos e extrínsecos”*.

**FUNDAMENTAÇÃO TEÓRICA**

* 1. **O que é o Machine Learning**

*Machine Learning*(ML) é uma ramificação da Inteligência Artificial (IA) focada em permitir que as máquinas aprendam como os humanos, realizem tarefas de forma autônoma e melhorem precisão e desempenho por meio da experiência e exposição. Segundo Lopes et al. (2021), ML pode ser:

“Machine learning é uma subárea da inteligência artificial que compreende algoritmos que aprendem padrões a partir de dados, com o objetivo de realizar previsões ou classificações, sem serem explicitamente programados para isso. ”

Isso se baseia em uma visão clássica e consolidada da área amplamente reconhecida pelos pesquisadores. A ideia central por trás do machine learning é:

*Um algoritmo aprende com os dados quando sua performance em uma tarefa melhora com a experiência.* **Mitchell (1997)** no livro *Machine Learning*.

* 1. **Machine Learning na Gestão de Programas Esportivos**

O uso de **machine learning (ML)** em contextos sociais e educacionais tem se expandido consideravelmente, inclusive em iniciativas esportivas voltadas à inclusão e formação cidadã. A capacidade dos algoritmos de ML de identificar padrões a partir de grandes volumes de dados permite construir modelos preditivos para auxiliar na tomada de decisões estratégicas (LOPES et al., 2021).

Em programas como o *Esporte Cidadão*, essa abordagem pode ser utilizada, por exemplo, para **prever a evasão de participantes**, detectar padrões de frequência, ou ainda identificar fatores que impactam diretamente na eficácia do programa. Tais informações são valiosas para gestores que precisam distribuir recursos de forma eficiente, seja em relação a professores, espaços ou horários.

A literatura já apresenta diversos estudos em que algoritmos como **árvore de decisão**, **k-means** e **regressão logística** são aplicados com sucesso na análise de dados educacionais e esportivos, oferecendo insights relevantes para a formulação de políticas públicas baseadas em evidências (HAN; PEI; KAMBER, 2022). A presente pesquisa busca aplicar esses métodos dentro da realidade do programa *Esporte Cidadão*, considerando suas especificidades e desafios.

**1.3 Design Science Research como Método para Construção de Soluções Educacionais**

A metodologia **Design Science Research (DSR)** tem se consolidado como uma abordagem robusta para pesquisas aplicadas que visam criar e validar soluções práticas para problemas reais. Diferentemente de métodos puramente descritivos ou explicativos, o DSR está centrado no desenvolvimento de **artefatos** — que podem ser modelos, sistemas, ferramentas ou processos — capazes de intervir de forma eficaz em determinado contexto (DRESCH; LACERDA; JÚNIOR, 2015).

Originalmente mais comum nas áreas de sistemas de informação e engenharia, o DSR tem ganhado destaque também em estudos na educação, especialmente em projetos que envolvem tecnologias digitais, dados e inovação social. Angeluci et al. (2021) mostram que o método é útil em contextos educacionais justamente por permitir que o pesquisador **iterativamente projete, teste e refine** soluções com base em ciclos de avaliação e validação junto aos usuários e ao ambiente real.

O processo metodológico do DSR envolve etapas como:

1. Identificação e contextualização do problema;
2. Levantamento teórico e análise de soluções existentes;
3. Concepção e construção do artefato;
4. Avaliação do artefato no contexto aplicado;
5. Reflexão e aprendizado com os resultados.

No contexto desta pesquisa, o uso do DSR permite articular ciência de dados e educação para **criar um modelo preditivo e analítico aplicável ao programa Esporte Cidadão**, com o objetivo de otimizar a alocação de recursos e maximizar o impacto social. Isso garante tanto a **relevância prática** quanto o **rigor metodológico**, dois pilares fundamentais dessa abordagem (HEVNER et al., 2004).

* 1. **Ferramentas de Análise de Dados: Pyhton e Orange**

Para o desenvolvimento desta pesquisa, a escolha das ferramentas de análise de dados foi fundamental para garantir a eficiência e a qualidade na manipulação, processamento e modelagem das informações coletadas. Duas ferramentas principais serão utilizadas: **Python** e **Orange Data Mining.**

O **Python** é uma linguagem de programação amplamente reconhecida por sua versatilidade e robustez em aplicações de ciências de dados. Segundo VanderPlas (2016), **Software Engineer na Google**, o Python oferece uma ampla gama de bibliotecas especializadas para análise estatística, aprendizado de máquina e visualização de dados, como **pandas, matplotlib, seaborn e scikit-learn.** Sua sintaxe simples e objetiva facilita a manipulação de grandes volumes de dados e a implementação de algoritmos de machine learning.

Além do Python, será utilizada a plataforma **Orange Data Mining**, uma ferramenta open source que oferece uma interface gráfica baseada em blocos visuais, permitindo a aplicação de algoritmos de mineração de dados de forma intuitiva. Conforme Demšar et al. (2013), o Orange facilita a construção de fluxos de trabalho para análise exploratória, classificação, regressão e clusterização, mesmo para usuários com menor familiaridade com programação.

A combinação dessas duas ferramentas, permitirá realizar desde a limpeza e preparação dos dados, passando pela **construção dos modelos preditivos**, até a **visualização dos resultados** de maneira eficiente e compreensível. O uso do Python proporcionará maior controle técnico sobre os algoritmos, enquanto o Orange oferece agilidade na prototipagem de soluções, favorecendo a experimentação de diferentes técnicas de machine learning.

* 1. **Integração entre Design Science Research, Ciência de Dados e Programas Educacionais.**

A integração entre **Design Science Research (DSR)**, **ciência de dados** e **programas educacionais esportivos**, como o Esporte Cidadão, representa uma abordagem inovadora e necessária para a gestão pública baseada em evidências. Enquanto o DSR fornece uma estrutura metodológica sólida para a **construção e avaliação de soluções práticas**, a ciência de dados oferece ferramentas para **coletar, organizar, analisar e interpretar grandes volumes de informações**, transformando dados brutos em conhecimento útil para a tomada de decisão (ANGELUCI et al., 2021; HAN; PEI; KAMBER, 2022).

O uso de **machine learning** dentro desse contexto amplia as possibilidades de análise, permitindo a construção de modelos preditivos que podem **identificar fatores de risco**, como a evasão de participantes, e **orientar a realocação eficiente de recursos**. As ferramentas tecnológicas utilizadas, como o **Python** e o **Orange**, viabilizam a execução dessas análises de forma eficiente e acessível, mesmo em ambientes com limitações de infraestrutura tecnológica.

Ao combinar essas três dimensões — **método científico (DSR)**, **técnicas analíticas (ciência de dados e machine learning)** e **contexto social-educacional (programa Esporte Cidadão)** —, esta pesquisa busca não apenas gerar conhecimento acadêmico, mas também **oferecer uma solução prática que possa ser aplicada diretamente na gestão pública**, contribuindo para o aumento da efetividade e do impacto social do programa.

**METODOLOGIA**

A presente pesquisa adota a abordagem metodológica **Design Science Research (DSR)**, por sua capacidade de orientar a construção e avaliação de soluções práticas para problemas reais, especialmente em contextos educacionais e sociais que envolvem o uso de tecnologia e análise de dados. O DSR é caracterizado por um processo iterativo que envolve a identificação de um problema, o desenvolvimento de um artefato, sua avaliação e a reflexão sobre os resultados (DRESCH; LACERDA; JÚNIOR, 2015).

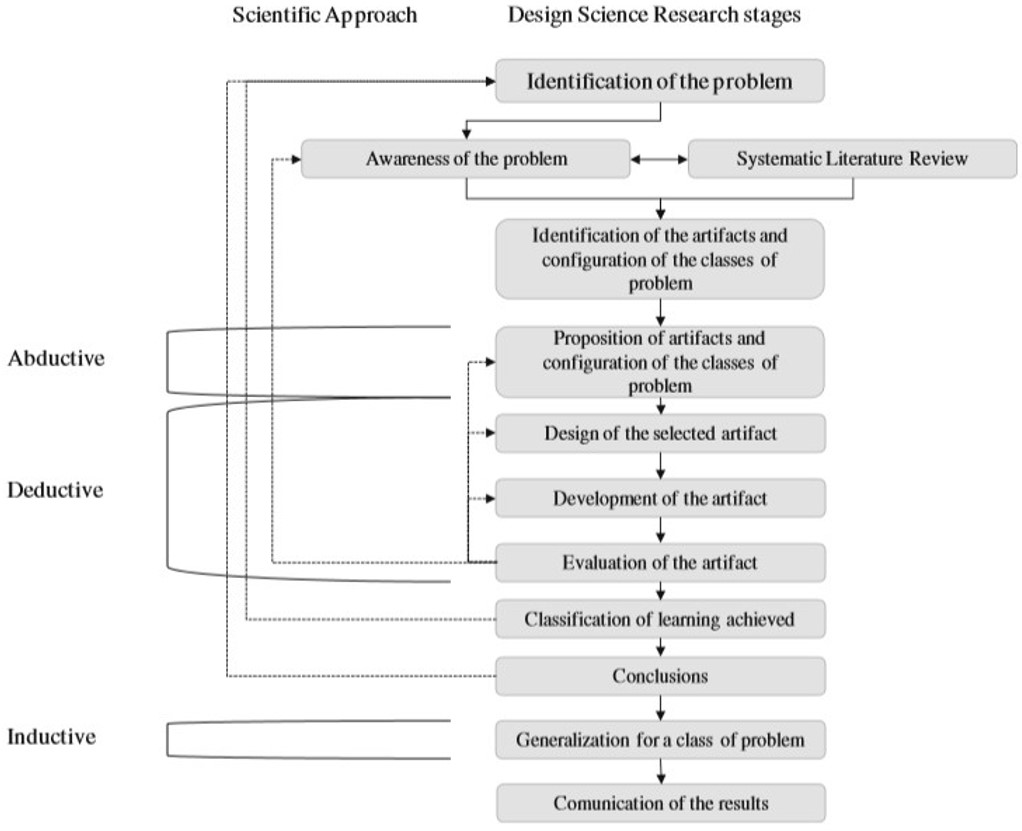
No contexto desta pesquisa, o problema central é a **distribuição ineficiente de recursos no programa Esporte Cidadão**, que pode impactar negativamente sua efetividade social e educacional. O artefato a ser desenvolvido é um **modelo analítico ou preditivo**, baseado em técnicas de **ciência de dados e machine learning**, com o objetivo de apoiar a tomada de decisão por parte dos gestores do programa.

* 1. **Etapas do Design Science Research (DSR)**

O processo metodológico irá seguir as principais etapas propostas por Dresch, Lacerda e Júnior (2015), adaptadas ao contexto da pesquisa:

1. **Identificação do problema**: análise preliminar da situação atual do programa Esporte Cidadão, com foco nos desafios relacionados à alocação de recursos.
2. **Levantamento teórico**: revisão da literatura nas áreas de esporte educacional, gestão de recursos públicos, ciência de dados e Design Science Research, para fundamentar teoricamente o desenvolvimento do artefato.
3. **Coleta e organização dos dados**: levantamento de informações relevantes sobre o programa, como número de participantes por modalidade e polo, frequência, histórico de evasão, disponibilidade de professores e infraestrutura.
4. **Desenvolvimento do artefato (modelo analítico/preditivo)**: construção de um modelo utilizando técnicas de machine learning para análise e previsão de padrões relacionados à evasão e ao uso dos recursos.
5. **Avaliação do artefato**: análise da precisão, aplicabilidade e utilidade prática do modelo, considerando os resultados obtidos nas simulações e testes.
6. **Reflexão e recomendações**: interpretação dos resultados, identificação de melhorias e proposição de recomendações práticas para a gestão do programa.

Figura 1: Design Research Science (DSR)



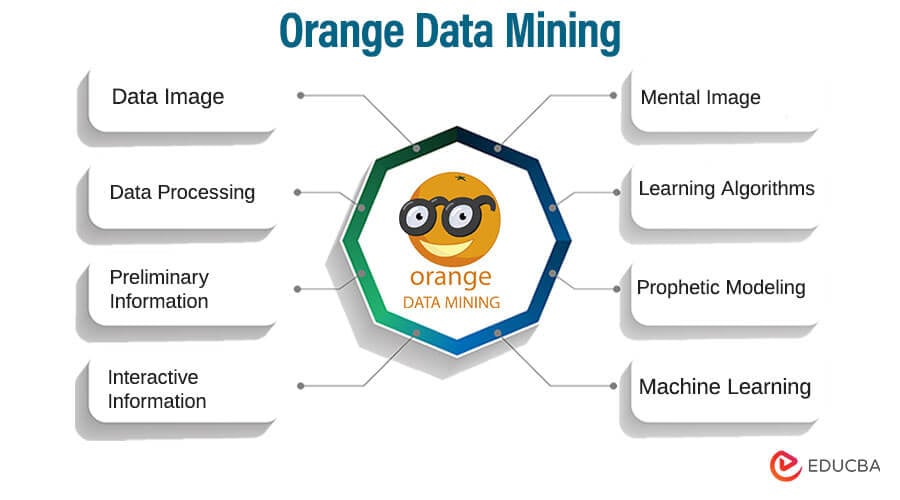
Fonte: Scielo.br - Design Science Research in practice: review of applications in Industrial Engineering

* 1. **Ferramentas utilizadas**

A implementação do modelo será realizada com o apoio de duas principais ferramentas:

* **Python**: linguagem de programação amplamente utilizada na ciência de dados, com bibliotecas como *pandas*, *matplotlib*, *seaborn* e *scikit-learn*, utilizadas para manipulação dos dados, construção dos modelos e análise dos resultados.
* **Orange Data Mining**: plataforma de mineração de dados com interface gráfica, que permite a aplicação de algoritmos de aprendizado de máquina de forma visual e interativa, facilitando a prototipagem e a comparação entre diferentes técnicas.

Figura 2: Orange Data Mining



Fonte: Github – Deep Learning Collab

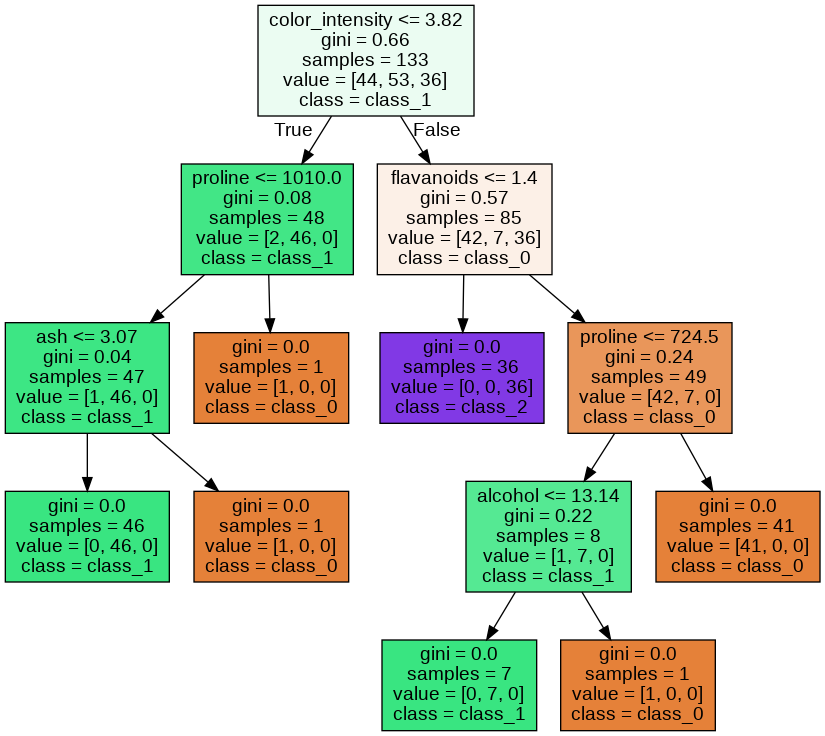
* 1. **Técnicas de Análise de Dados**

Serão utilizadas técnicas de **análise exploratória de dados (EDA)** para identificação de padrões iniciais, além de algoritmos de **machine learning**, como:

* **Árvores de decisão**: para classificação e identificação de variáveis que influenciam a evasão;
* **Clusterização (k-means)**: para segmentação dos participantes com base em características semelhantes;
* **Regressão logística**: para prever a probabilidade de evasão com base em múltiplas variáveis.

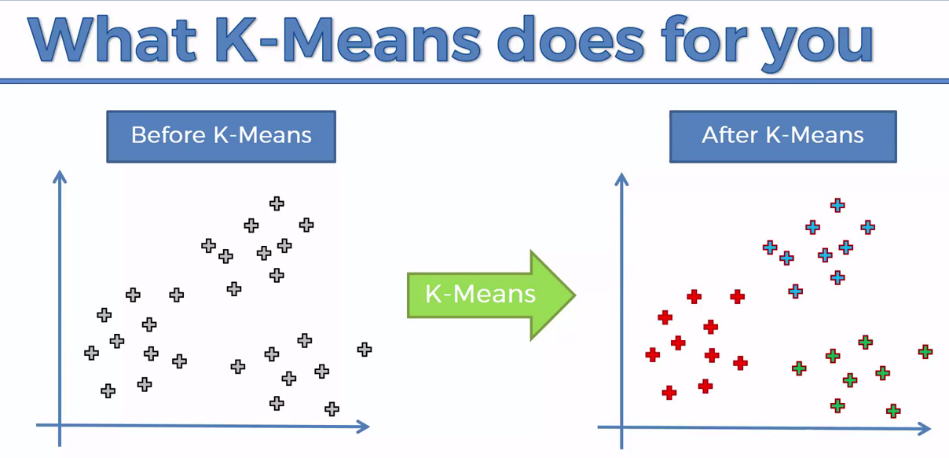
A escolha final dos algoritmos foi baseada na **acurácia**, **interpretabilidade** e **aplicabilidade prática dos resultados**, conforme diretrizes de Han, Pei e Kamber (2022).

Figura 3: Árvore de decisão



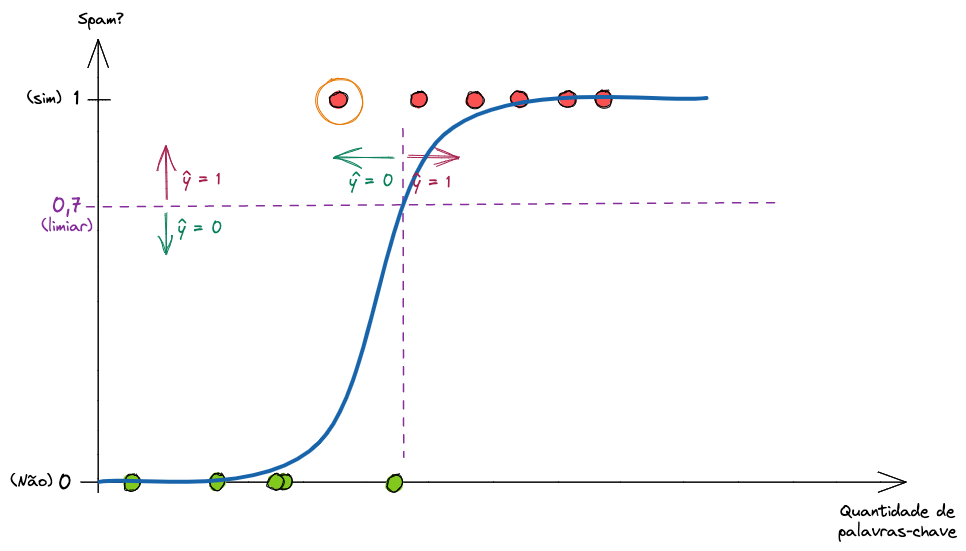
Fonte: Sigmoidal – Entendendo as árvores de decisão em Machine Learning

Figura 4: Clusterização (K-means)



Fonte: machinelearning.org.in – Machine Learning Tutorials, Courses and Certifications

Figura 5: Regressão Logística



Fonte: Brains Dev – Modelos de classificação: Regressão Logística

Com isso, espera-se construir um modelo capaz de apoiar os gestores do programa na tomada de decisões mais eficientes.

**Referências Bibliográficas**

ANGELUCI, A.; SILVA, M. P.; BARBOSA, J. L. V.; MOURA, H. F. de. **Design Science Research in practice: review of applications in Industrial Engineering**. Revista Produção Online, 2021.

BARBANTI, V. J. **Treinamento esportivo: estruturação e planejamento**. São Paulo: Manole, 2006.

DRESCH, A.; LACERDA, D. P.; JÚNIOR, J. J. C. **Design Science Research: método de pesquisa para avanço da ciência e tecnologia**. Porto Alegre: Bookman, 2015.

HAN, J.; PEI, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. 4. ed. Cambridge: Morgan Kaufmann, 2022.

LOPES, F. M.; LUDERMIR, T. B.; OLIVEIRA, W. R. de. **Aprendizado de Máquina**. Rio de Janeiro: LTC, 2021.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

VANDERPLAS, J. **Python Data Science Handbook: Essential Tools for Working with Data**. 1. ed. Sebastopol: O’Reilly Media, 2016.

DEMŠAR, J. et al. **Orange: Data Mining Toolbox in Python**. Journal of Machine Learning Research, v. 14, p. 2349-2353, 2013.

ORANGE DATA MINING. FAQ – Education & tutorials. Disponível em: <https://oldorange.biolab.si/faq/>. Acesso em: 20 jun. 2025

AJDA. 10 Tips and Tricks for Using Orange. Orange Data Mining Blog, 17 out. 2016. Disponível em: <https://orangedatamining.com/blog/10-tips-and-tricks-for-using-orange/> Acesso em: 20 jun. 2025

SUARDI, Mauro. Machine Learning for beginners with Orange Data Mining. Eni digiTALKS (Medium), 26 mar. 2025. Disponível em: <https://medium.com/eni-digitalks/machine-learning-for-beginners-with-orange-data-mining-0690372533b9> Acesso em: 20 jun. 2025

OMEGA TUTORIALS. Getting Started with Machine Learning Using Orange Data Mining. Omics Tutorials. Disponível em: <https://omicstutorials.com/getting-started-with-machine-learning-using-orange-data-mining/> Acesso em: 20 jun. 2025

PATEL, Radhey. Introduction to Orange Tool. Medium, 29 set. 2021. Disponível em: <https://medium.com/@radheypatel272/introduction-to-orange-tool-2a5b416c07d9> Acesso em: 20 jun. 2025 [iwconnect.com+4medium.com+4medium.com+4](https://medium.com/%40radheypatel272/introduction-to-orange-tool-2a5b416c07d9?utm_source=chatgpt.com).

ŻERKOWSKI, Adam. Unlocking the Power of Data with Orange Data Mining: A Beginner’s Guide. Medium, data? Disponível em: <https://medium.com/@zerkowski.adam/unlocking-the-power-of-data-with-orange-data-mining-a-beginners-guide-367a2c6f07f3> Acesso em: 20 jun. 2025

BATISTA, Bruno. Machine Learning sem código. Usando Orange Data Mining para criar um… Ensina.AI, Medium, data? Disponível em: <https://medium.com/ensina-ai/machine-learning-sem-c%C3%B3digo-636d1a8f9081> Acesso em: 20 jun. 2025

ERDEN, Caner. Orange Data Mining Tool and Association Rules. Medium (TDS Archive), 11 mai. 2020. Disponível em: <https://medium.com/data-science/orange-data-mining-tool-and-association-rules-caa3c728613d9> Acesso em: 20 jun. 2025

IWCONNECT. Mastering Data Mining with Orange: A Comprehensive Guide. IWConnect, 24 jan. 2020. Disponível em: <https://iwconnect.com/data-mining-using-orange/> Acesso em: 20 jun. 2025

CAPURSO, Mario. Learning Data Science visually with Orange Data Mining. LinkedIn Pulse, data? Disponível em: <https://www.linkedin.com/pulse/learning-data-science-visually-orange-mining-mario-capurso> Acesso em: 20 jun. 2025

RIBEIRO, F. N.; ARAÚJO, M.; GONÇALVES, P.; BENE­VENUTO, F.; GON­ÇALVES, M. A. SentiBench: a benchmark comparison of state-of-the-practice sentiment analysis methods. ArXiv, 6 dez. 2015