

GEOGRAFIA: Ambiente, Educação e Sociedades GeoAmbES



ARTIGO

APLICAÇÃO DE MACHINE LEARNING NA DETERMINAÇÃO DO MODELO IDEAL PARA ABERTURA DE ACADEMIAS

*Aplicación del aprendizaje automático para
determinar el modelo ideal para la apertura de
gimnasios.*

*Applying machine learning to determine the
ideal model for opening gyms.*

Gustavo Tavares Ramalho

Pós Graduado do curso em DATA SCIENCE E
ANALYTICS – 2024 do MBA USP/ESALQ.

Orcid: <https://orcid.org/>

E-mail: Gustavo.ramalho14@hotmail.com

Lisleandra Machado

Doutora em Engenharia e professora Instituto
Federal de Educação, Ciência e Tecnologia do
Sudeste de Minas Gerais - Campus Santos
Dumont.

Orcid: <https://orcid.org/0000-0002-7761-8023>

E-mail: lisleandra.machado@ifsudestemg.edu.br

Leonardo Amorim de Araújo

Doutor em Engenharia dos Transportes pela UFRJ
e professor do Instituto Federal do Sudeste de
Minas Gerais.

Orcid: <https://orcid.org/0000-0003-2722-7531>

E-mail: leonardo.araujo@ifsudestemg.edu.br

Como citar este artigo:

RAMALHO, Gustavo Tavares; MACHADO,
Lisleandra; ARAÚJO, Leonardo Amorim de.
Aplicação de machine learning na
determinação do modelo ideal para abertura de
academias. **GEOGRAFIA: Ambiente,
Educação e Sociedades** – GeoAmbES,
jul./dez. v. 4, n. 8, p. 138-154, 2025.

Disponível em:

<https://periodicos.unemat.br/index.php/geoambes>

Volume 4, Número 8 (2025)
ISSN 25959026

APLICAÇÃO DE MACHINE LEARNING NA DETERMINAÇÃO DO MODELO IDEAL PARA ABERTURA DE ACADEMIAS

Aplicación del aprendizaje automático para determinar el modelo ideal para la apertura de gimnasios.

Applying machine learning to determine the ideal model for opening gyms.

Resumo

Este texto tem o objetivo identificar necessidades dos consumidores por uma academia, visando estratégias que apoiem a tomada de decisões na abertura de novas academias. A pesquisa de abordagem exploratória e qualitativa com abordagem híbrida de Machine Learning (ML). A coleta foi realizada por meio de questionário, aplicado, abordando aspectos como atividades físicas, horários, faixa de preços e serviços adicionais. Esses resultados fornecem uma base sólida para que empreendedores possam planejar academias que atendam de forma mais precisa às demandas do mercado, desde a escolha de localização até a definição de preços e serviços oferecidos.

Palavras-chave: Academias. Consumidor. Análise de Clusters.

Abstract

This text aims to identify consumer needs for a gym, focusing on strategies to support decision-making in the opening of new gyms. The research employs an exploratory and qualitative approach with a hybrid Machine Learning (ML) methodology. Data was collected through a questionnaire covering aspects such as physical activities, schedules, price range, and additional services. These results provide a solid foundation for entrepreneurs to plan gyms that more accurately meet market demands, from location selection to pricing and service offerings.

Keywords: Gyms. Consumer. Cluster Analysis.

Resumen

Este texto busca identificar las necesidades de los consumidores de gimnasios, centrándose en estrategias para apoyar la toma de decisiones en la apertura de nuevos gimnasios. La investigación emplea un enfoque exploratorio y cualitativo con una metodología híbrida de aprendizaje automático (ML). Los datos se recopilaban mediante un cuestionario que abarcó aspectos como actividades físicas, horarios, rango de precios y servicios adicionales. Estos resultados proporcionan una base sólida para que los emprendedores planifiquen gimnasios que satisfagan con mayor precisión las demandas del mercado, desde la selección de la ubicación hasta los precios y la oferta de servicios.

Palabras clave: Gimnasios. Consumidor. Análisis de clústeres.

Introdução

A prática de atividade física vem adquirindo um grande significado na vida das pessoas, motivadas pelos benefícios que o exercício traz para a saúde física e mental. Além de auxiliar no controle do estresse e na manutenção do peso ideal, o exercício regular promove um estilo de vida mais saudável e ativo (Nahas, 2013). Com o aumento dos problemas de saúde relacionados ao sobrepeso, a procura por academias cresceu significativamente. A prática de atividade física tem inúmeros benefícios para os praticantes, dentre eles os essenciais são a redução de peso, o aumento da massa e da força muscular, a melhora da sensibilidade à insulina, da capacidade cardiorrespiratória, da pressão arterial, da flexibilidade e do equilíbrio (Coelho e Burini, 2009).

Por isso a prática de exercícios físicos é importante não só para estética, mas como também para prevenir e tratar diversas doenças metabólicas e intervir positivamente na capacidade funcional dos praticantes. Esses benefícios, amplamente comprovados por estudos científicos, incentivam pessoas de todas as idades e sexos a adotarem a prática regular de atividade física com foco na saúde (Fialho, 2014; Santos, 2013).

Nos últimos anos, o setor de academias e estúdios de treinamento experimentou um crescimento expressivo. Segundo o SEBRAE (2023), há mais de 55.000 estabelecimentos registrados no país na categoria CNAE 9313-1/00 (atividades de condicionamento físico). No entanto, apesar desse crescimento, o setor enfrenta uma taxa de mortalidade significativa, com 48,60% das empresas fechando nos últimos cinco anos (SEBRAE, 2023).

Um dos principais fatores que contribuem para essa alta taxa de mortalidade é a falta de planejamento adequado, gestão empresarial eficiente e conhecimento sobre o comportamento empreendedor (SEBRAE, 2014). Muitos gestores de academias possuem formação predominantemente em educação física, sem o preparo necessário em áreas como empreendedorismo e gestão, o que resulta na incapacidade de manter o negócio sustentável. Diante desse cenário competitivo, torna-se crucial para os empreendedores do setor adotarem ferramentas gerenciais que permitam prever, planejar e controlar suas operações. Isso inclui a oferta de

serviços diferenciados e a adoção de estratégias para manter-se competitivo no mercado.

Uma dessas ferramentas é o ML, ou, Aprendizado de Máquina, uma área da inteligência artificial (IA) que desenvolve algoritmos capazes de gerar conhecimento a partir de dados históricos. O ML tem a capacidade de identificar padrões em grandes conjuntos de dados e fazer previsões futuras, utilizando técnicas avançadas de análise (Amaral, 2016). Embora o conceito de aprendizado de máquina pareça recente, ele já vinha sendo aplicado em contextos científicos e empresariais desde antes da popularização da IA, na década de 1950 (Mackenzie, 2015).

Apesar da relevância do ML ser uma importante ferramenta para obter conhecimento, não existe um único algoritmo capaz de definir todos os comportamentos e definição dos problemas. É importante conhecer os conceitos dos algoritmos, avaliá-los e aplicá-los, e por fim, avaliarmos posteriormente (Monard & Baranauskas, 2003).

O ML pode ser dividido em três tipos principais: aprendizado supervisionado, aprendizado não-supervisionado e aprendizado reforçado (SHIN et al., 2019).

No aprendizado supervisionado é preciso pares de entrada-saída e o modelo tenta descobrir as relações entre as variáveis de entrada, também conhecidas como variáveis independentes e as variáveis de saída (variáveis dependentes) existentes (Müller & Guido, 2017). Já aprendizagem não supervisionada não é fornecido qualquer tipo de conjunto de informação de entrada, ele aprende sozinho as entradas. A função mais usual dessa aprendizagem é o agrupamento (Russel & Norvig, 2013). Por fim, na aprendizagem por reforço o sistema recebe informação quando o conjunto de informação não está certo, a informação correta não é fornecida a ele. O sistema aprende sozinho por meio de recompensas ou punições (Russel & Norvig, 2013).

Materiais e Métodos

Este projeto caracteriza-se como uma pesquisa aplicada, de abordagem exploratória e natureza qualitativa, utilizando uma combinação de técnicas de aprendizado de máquina (ML) supervisionado e não supervisionado. O objetivo principal é analisar as preferências do público-alvo em relação a academias, identificando padrões de comportamento e preferências nos dados coletados.

A análise foi conduzida com o uso da linguagem Python, utilizando o ambiente de desenvolvimento Spyder, fornecido pelo MBA da USP/Esalq. A coleta de dados foi realizada por meio de um questionário contendo 15 perguntas aplicadas ao público-alvo, distribuído por canais digitais, como WhatsApp e Instagram, entre os dias 5 e 19 de jun. de 2024. O questionário foi elaborado na plataforma Microsoft Forms e resultou em 158 respostas únicas, consideradas suficientes para o escopo da pesquisa.

Coleta de Dados

As perguntas do questionário foram formuladas para capturar características e preferências relevantes para o planejamento de academias, como o tipo de atividade física preferida, frequência ideal, faixa de preço, e horários mais convenientes. Não foram coletadas informações pessoais que pudessem identificar os respondentes, garantindo anonimato e conformidade com diretrizes éticas de proteção de dados.

Após a coleta, os dados foram carregados em formato Excel e tratados para análise. Foram realizadas tarefas de limpeza e preparação, como a renomeação de colunas e a remoção de espaços e caracteres especiais, visando facilitar a manipulação dos dados.

Análise Estatística

Inicialmente, foi aplicado o Teste Qui-quadrado para verificar a associação entre variáveis categóricas (Siegel & Castellan JR., 2006). Esse teste permite identificar variáveis significativamente associadas e selecionar as categorias mais importantes para os clientes na escolha de academias, utilizando tabelas de contingência e frequências observadas.

Em seguida, aplicamos a **Análise de Correspondência Múltipla (MCA)**, uma técnica de modelagem multivariada voltada para a exploração de relações entre variáveis categóricas. A MCA transforma dados categóricos em componentes principais, facilitando a visualização de padrões e associações em um espaço de menor dimensão (Fávero & Belfiore, 2024). Essa técnica é amplamente utilizada em segmentação de mercado e análise de perfis de consumidores.

Com os resultados da MCA, autovalores e coordenadas padronizadas das categorias foram obtidos, medindo a importância relativa de cada componente

principal (Hair, 2009). Para determinar o número ideal de grupos, utilizamos o método do cotovelo, que calcula a soma das distâncias quadráticas para diferentes números de clusters. Em seguida, aplicamos o algoritmo **K-Means**, uma técnica de aprendizado não supervisionado amplamente utilizada para agrupar indivíduos com preferências semelhantes (Fávero, 2021). A segmentação resultante permite uma análise mais detalhada dos grupos de clientes e suas preferências.

Modelagem Preditiva

Para avaliar a importância das variáveis e realizar classificações, foi utilizado o **Random Forest Classifier**, um método de aprendizado supervisionado que combina várias árvores de decisão para melhorar a acurácia das previsões (Câmera, 2020). Além disso, utilizamos a **validação cruzada K-Fold** para avaliar a performance do modelo de forma robusta, dividindo o conjunto de dados em K subconjuntos e testando o modelo em diferentes combinações de treino e teste (Oliveira, 2017).

Visualização dos Resultados

Os resultados da MCA e da clusterização foram visualizados por meio de mapas perceptuais, gráficos de dispersão (scatter plots) e gráficos de barras (bar plots), facilitando a interpretação das características predominantes em cada grupo de clientes. Essas visualizações ajudam a identificar os fatores prioritários na escolha de academias e fornecem insights valiosos para o planejamento estratégico.

Resultados e Discussões

A partir dos dados coletados, foi possível identificar algumas tendências preliminares no mercado de academias. Contudo, para uma análise mais aprofundada e a determinação de um modelo ideal para a abertura de uma academia, submetemos a base de dados a um processo de ML. Adotou-se uma abordagem exploratória, de natureza qualitativa, uma abordagem híbrida de ML, envolvendo técnicas de aprendizado não supervisionado e supervisionado para identificar padrões e agrupar as preferências dos participantes.

Análise Qui-quadrado

Aplicamos o teste qui-quadrado para verificar a relação entre as variáveis independentes (como Gênero, Idade, Proximidade etc.) e a variável dependente "Atividade.Principal" identificar quais são **mais importantes** na escolha de academias pelos clientes. Escolhemos como variável dependente a questão “Qual a principal atividade que você busca em uma academia?” devido a atividade oferecida pelas academias ser um fator determinante para adesão. Dentre as variáveis as que apresentaram maior associação significativa, ou seja, com p-valores menores que 0,05, foram: (i) Diferenciais: $p = 3.4348e-12$; (ii) Funcionamento.Fds: $p = 7.0725e-0$; (iii) Gênero: $p = 0.000568$; (iv) Idade: $p = 0.000393$; e (v) Proximidade: $p = 3.4722e-21$.

Podemos observar que as 05 variáveis possuem uma associação significativa com a atividade principal que os clientes buscam nas academias e são importantes para compreender as preferências dos clientes.

Realizamos uma identificação dos itens mais importantes das variáveis significativas do DataFrame e o filtramos para conter apenas as linhas que atendem a pelo menos 80% dos itens mais importantes identificados. A partir deste momento o `df_filtrado` passou a ter 78 linhas.

Análise De Correspondência Múltipla

A seguir aplicamos a técnica de análise de correspondência Múltipla ao conjunto de dados filtrados e utilizamos 3 dimensões, transformamos as variáveis categóricas em componentes principais e reduzimos a dimensionalidade dos dados. Geramos os autovalores para explicar no conjunto de dados quanta variação é explicada por cada uma das dimensões.

Porcentagem de variação explicada por cada dimensão:

Dimensão 1: 37.38%

Dimensão 2: 32.42%

Dimensão 3: 30.20%

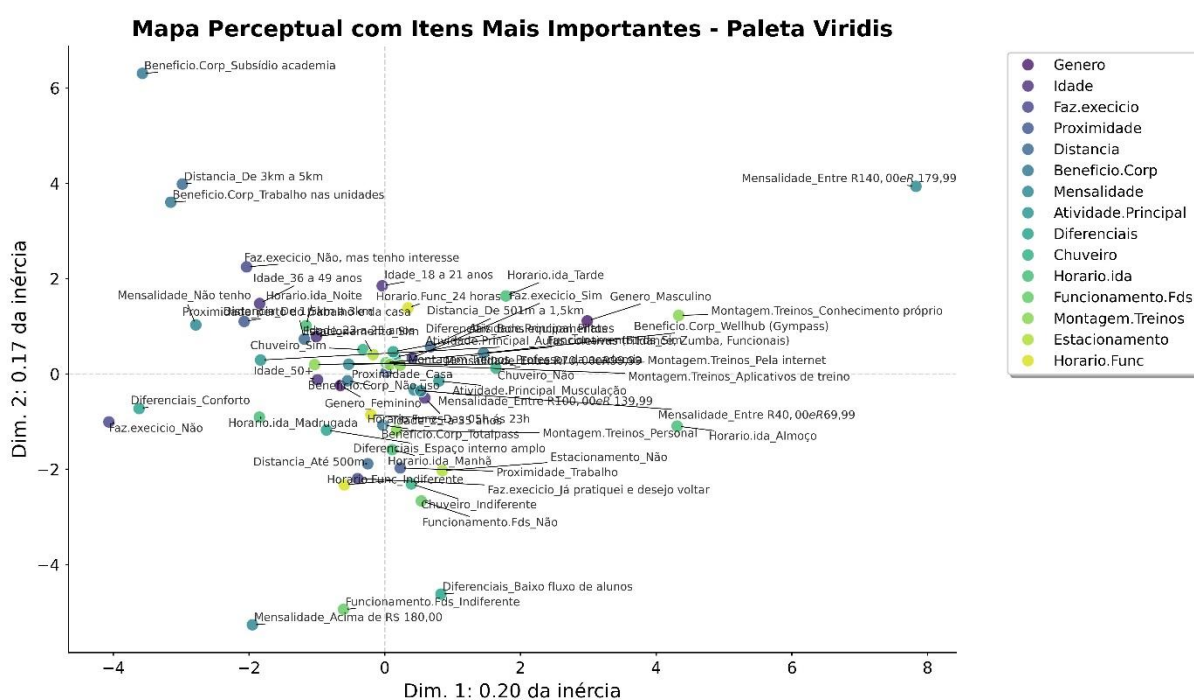
As três dimensões escolhidas representam diferentes aspectos da variação nos dados, as duas primeiras dimensões juntas explicam cerca de 69.80% da variação (37.38% + 32.42%), o que significa uma boa representação dos dados se focar nelas,

porém optamos por seguir com 3 dimensões visto que apresenta um valor significativo de 30,20% e pode fornecer insights adicionais na análise.

Mapa Perceptual

A partir dos dados da MCA plotamos o mapa perceptual, apresentando como as variáveis se agrupam nas dimensões identificadas, permitindo visualizar como as diferentes características das academias são percebidas pelos clientes.

Figura 1 – Mapa perceptual (itens importantes)



Fonte: Resultados originais da pesquisa

As dimensões resultantes explicam parte significativa da variabilidade dos dados, permitindo identificar padrões de comportamento e perfis de clientes. A Dimensão 1 explica 20% da variação e está associada a fatores como proximidade da academia e uso de benefícios corporativos, e a Dimensão 2 explica 17% da variação, relacionada com a prática ou não de exercícios físicos. Juntas, essas dimensões explicam **37%** da variação total dos dados, oferecendo uma visão substancial das interrelações entre as variáveis.

Padrões Identificados

Observamos diversos agrupamentos e correlações entre as variáveis relacionadas às preferências dos usuários de academias. Esses agrupamentos indicam como diferentes fatores influenciam as escolhas dos clientes.

- (i) Benefícios Corporativos e Distância: Verificou-se que usuários que recebem subsídios corporativos tendem a optar por academias localizadas próximas ao ambiente de trabalho, diferenciando-se dos demais perfis de usuários.
- (ii) Atividades Físicas e Proximidade: Observou-se que indivíduos que praticam musculação preferem academias com horários mais flexíveis, como aquelas que funcionam 24 horas, e que estejam próximas de suas rotinas diárias, destacando a importância da conveniência.
- (iii) Mensalidade e Benefícios: Usuários que pagam mensalidades mais altas estão geralmente associados a subsídios corporativos e à conveniência de localização. Em contrapartida, aqueles que pagam valores mais baixos buscam academias com características diferenciadas, como maior foco em conforto ou serviços adicionais.

Com base nas correlações e agrupamentos identificados, foi possível delinear quatro principais perfis de usuários:

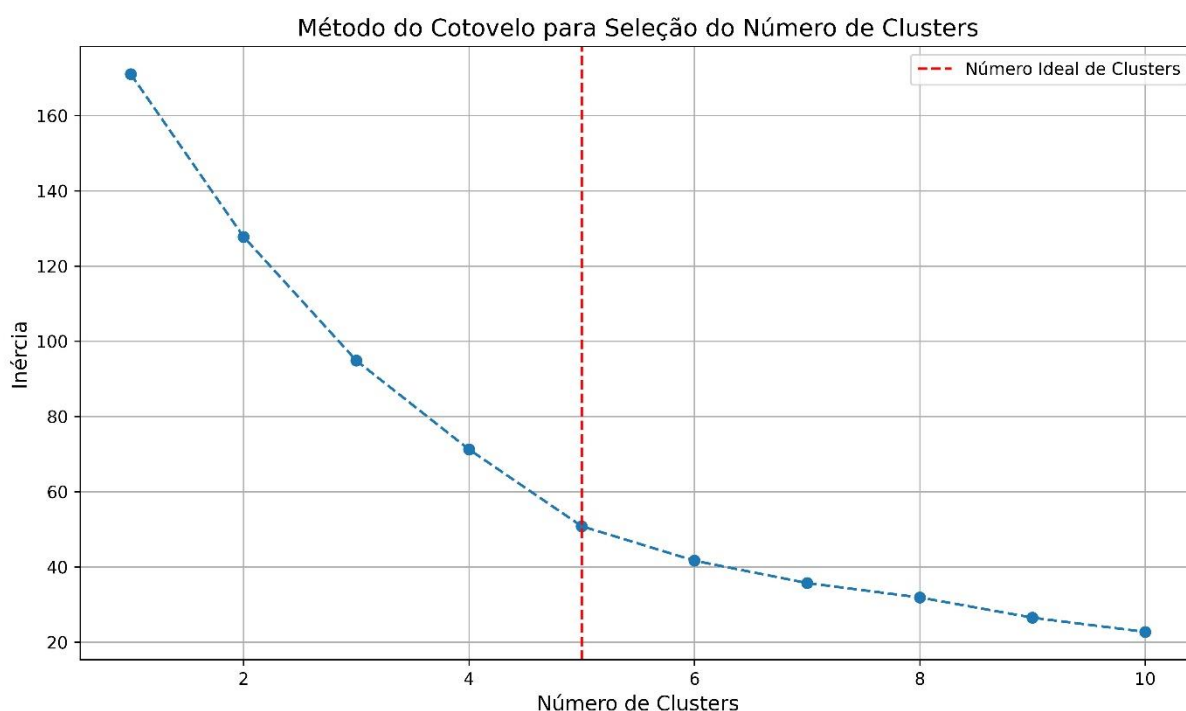
- (i) Usuários Corporativos: Estes indivíduos tendem a preferir academias localizadas próximas ao trabalho e que ofereçam subsídios corporativos. Geralmente, estão dispostos a pagar mensalidades mais altas em troca de conveniência e benefícios.
- (ii) Jovens Ativos: Este grupo valoriza a variedade de atividades oferecidas pela academia, como Pilates, além de buscar horários de funcionamento mais flexíveis para acomodar suas rotinas.
- (iii) Homens Focados em Musculação: Indivíduos deste perfil têm preferência por academias que oferecem musculação como atividade principal, além de priorizarem estabelecimentos com funcionamento 24 horas e serviços de treinamento personalizado.
- (iv) Usuários Menos Ativos: Este grupo apresenta menor engajamento com atividades físicas e não prioriza a proximidade da academia ou diferenciais

de conforto. Seus critérios de escolha são menos influenciados por esses fatores.

Análise de Cluster com K-means

Ato contínuo, optamos por realizar uma análise de cluster, onde aplicamos o algoritmo **K-Means** para agrupar os clientes com base em suas preferências. Primeiramente, normalizamos as coordenadas geradas pela Análise de Correspondência Múltipla, garantindo que todas as variáveis sejam escaladas de forma igual. A seguir utilizamos o **método do cotovelo** para determinar o número ideal de clusters.

Figura 2 – Método do cotovelo

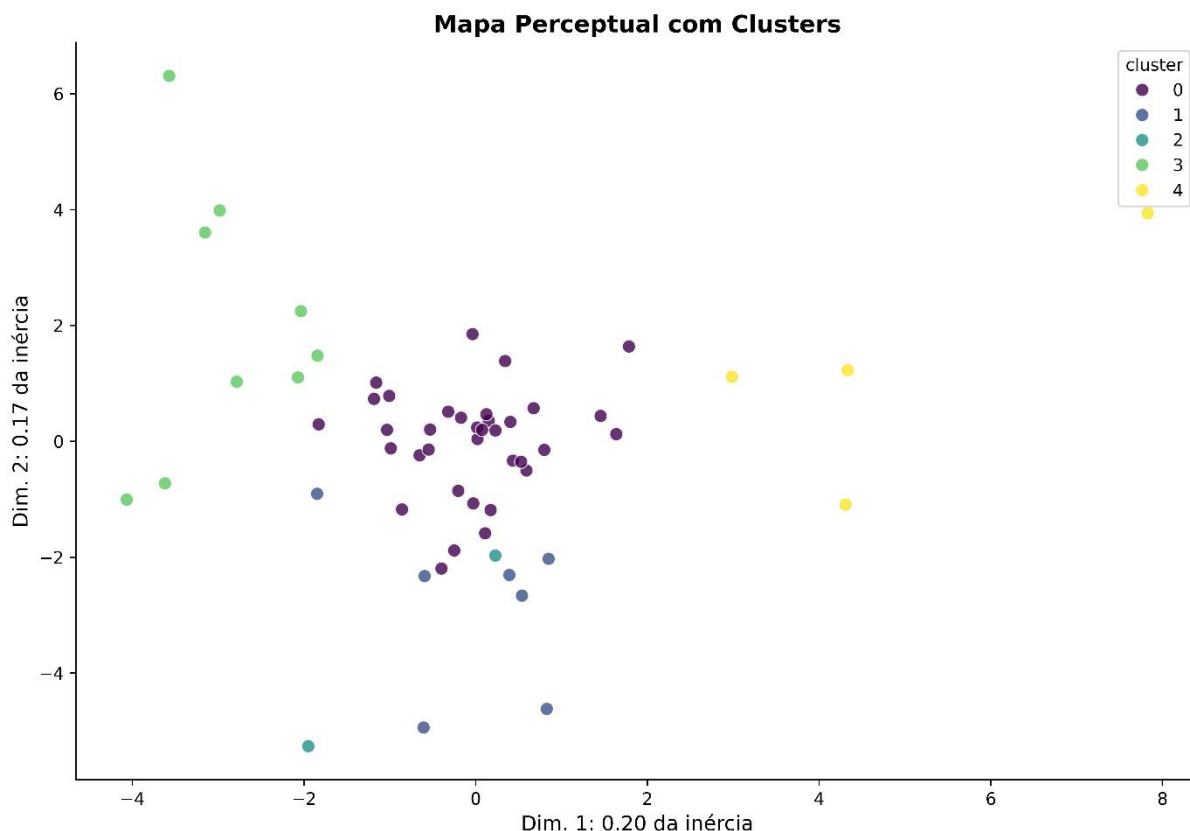


Fonte: Resultados originais da pesquisa

Com base no método do cotovelo identificamos que o número de clusters mais apropriado para segmentar os dados de maneira eficiente é **5**, permitindo uma boa separação dos dados onde cada cluster representa um grupo de clientes com características semelhantes. A seguir aplicamos o K-means para identificar grupos de usuários com preferências semelhantes.

Após a definição do número de clusters e aplicação do K-means foi elaborado um mapa perceptual com clusters que permite a visualização das preferências dos indivíduos agrupados em diferentes clusters.

Figura 3 – Mapa Perceptual com Clusters



Fonte: Resultados originais da pesquisa

Visualização dos Clusters

Os clusters estão representados por diferentes cores, e no gráfico temos cinco clusters identificados pelas cores:

- Cluster 0 (Roxo): este cluster é o mais densamente agrupado no centro do gráfico, indicando uma alta similaridade entre os indivíduos deste grupo em relação às preferências. A proximidade dos pontos sugere que os indivíduos desse cluster têm respostas muito consistentes ou similares.
- Cluster 1 (Azul): foi distribuído de maneira mais dispersa ao redor do cluster roxo. Os pontos são mais espaçados, indicando que este grupo possui

preferências que variam mais do que o Cluster 0, mas ainda apresentam uma certa similaridade entre si.

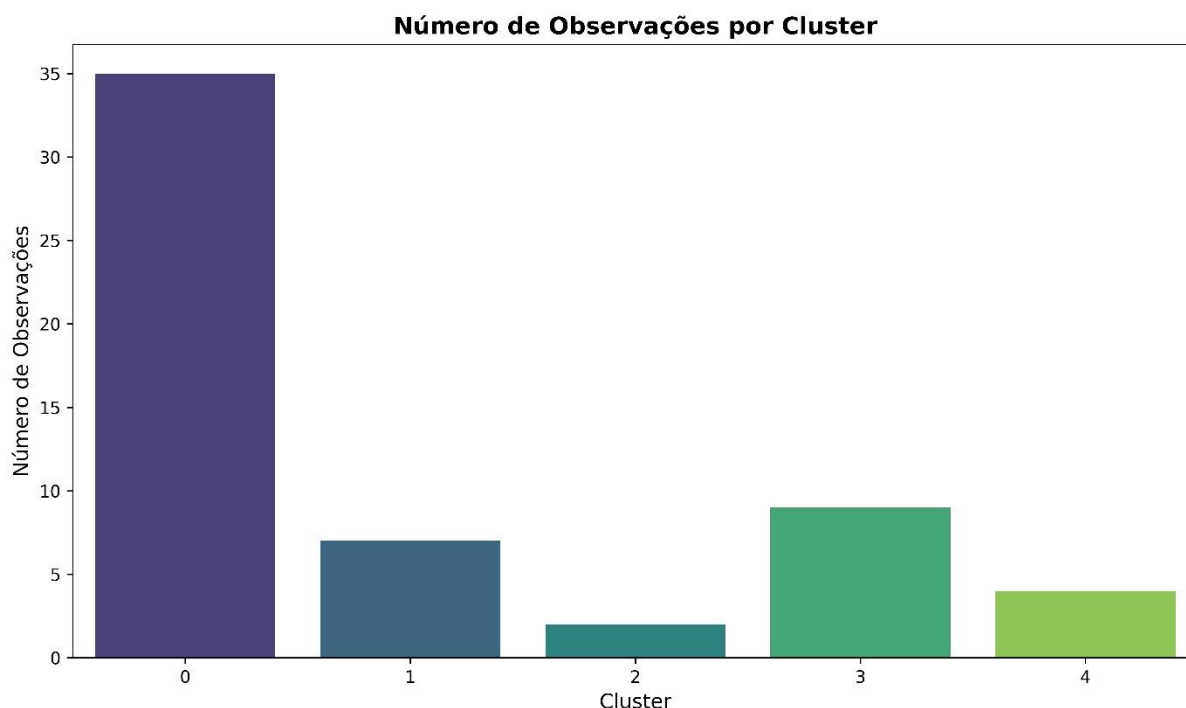
- Cluster 2 (Verde-claro): este grupo está relativamente mais afastado do centro do gráfico, especialmente nas regiões superiores. Isso indica que as preferências das pessoas deste grupo divergem significativamente das de outros clusters. Essas preferências podem ser menos comuns ou mais específicas em relação às variáveis analisadas.

- Cluster 3 (Verde-escuro): Similar ao Cluster 2 em termos de distribuição espacial, com uma maior dispersão na região inferior do gráfico. Isso também sugere preferências distintas em comparação com o restante dos clusters.

- Cluster 4 (Amarelo): Este cluster é notável por sua posição isolada à direita do gráfico, sugerindo que os indivíduos neste grupo possuem um conjunto de preferências claramente diferenciado dos demais grupos. A distância entre os pontos desse cluster e dos outros clusters indica uma grande divergência em suas respostas.

Adicionalmente, elaboramos uma visualização da proporção de observações em cada cluster em relação ao total de observações:

Figura 4 – Número de Observações por Clusters



Fonte: Resultados originais da pesquisa

Com base na análise dos clusters gerados, o Cluster 0 (Roxo) destaca-se como o grupo mais densamente agrupado e consistente em suas preferências, indicando uma forte correlação entre os indivíduos deste grupo. Assim, este cluster será utilizado como referência para delinear as características principais do modelo ideal de academia, visando atender de forma eficaz as necessidades e expectativas dos usuários mais representativos. A compreensão das preferências deste grupo permitirá a formulação de estratégias de marketing e desenvolvimento de serviços que se alinhem às expectativas dos clientes, promovendo a fidelização e a satisfação.

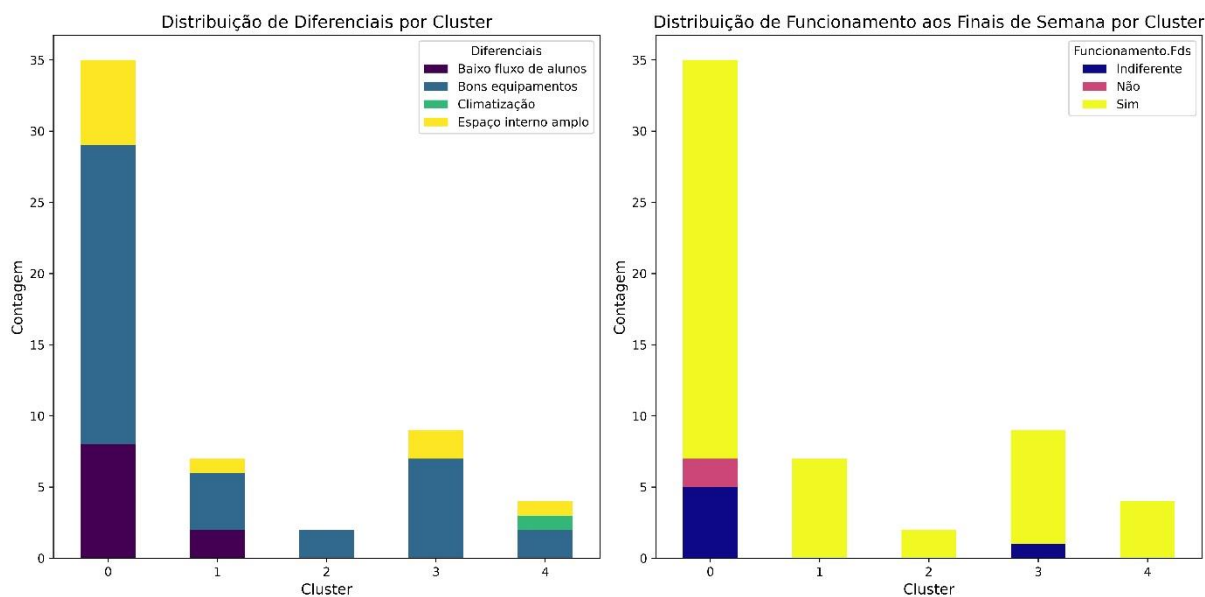
Avaliação do Modelo Random Forest Classifier

Aplicamos a técnica de Random Forest Classifier, dividindo o conjunto de dados em 70% para treinamento e 30% para teste. O modelo alcançou uma acurácia de 89% no conjunto de teste, demonstrando um bom desempenho na classificação correta dos clientes em seus respectivos clusters. Além disso, a acurácia média obtida por meio de validação cruzada foi de 84%, com um desvio padrão de 0,06. Esse resultado indica uma performance consistente e estável, com baixa sensibilidade à divisão dos dados, o que reforça a estabilidade e capacidade de generalização do modelo.

Distribuição das variáveis por cluster

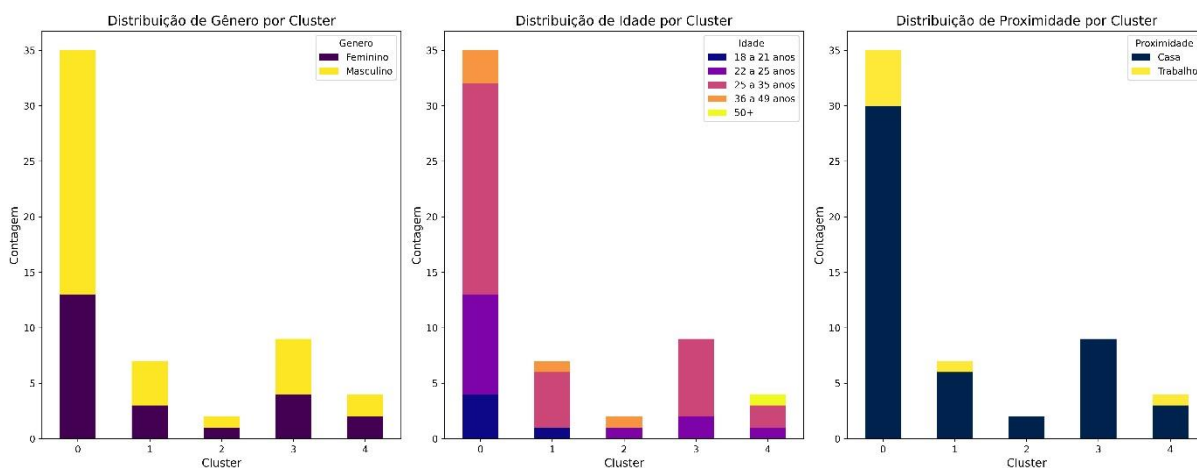
E por fim, realizamos a elaboração de gráficos de barras empilhadas para visualizar a distribuição de variáveis categóricas em cada cluster, facilitando a interpretação de como os clusters diferem entre si em termos das características analisadas. As figuras 5, 6 e 7 são as variáveis que apresentaram maior associação significativa no teste do qui-quadrado maiores que 0,05, foram: (i) Diferenciais: $p = 3.4348e-12$; (ii) Funcionamento.Fds: $p = 7.0725e-0$; (iii) Gênero: $p = 0.000568$; (iv) Idade: $p = 0.000393$; e (v) Proximidade: $p = 3.4722e-21$. Os gráficos das demais variáveis foram elaborados e encontram-se no Apêndice B.

Figura 5 – Distribuição da variável Diferenciais e Funcionamento.FDS por Cluster



Fonte: Resultados originais da pesquisa

Figura 6 – Distribuição da variável Gênero, Idade e Proximidade por Cluster



Fonte: Resultados originais da pesquisa

Considerações finais

A pesquisa apresentou um panorama detalhado sobre o comportamento e preferências dos consumidores ao escolherem academias para a prática de atividades físicas, utilizando uma abordagem híbrida de ML. A análise dos dados identificou variáveis significativas que influenciam as escolhas dos consumidores, como

benefícios corporativos, proximidade da academia, e tipos de atividades físicas preferidas.

Através da aplicação de técnicas como Análise de Correspondência Múltipla (MCA), K-Means Clustering e Random Forest, foi possível segmentar os consumidores em perfis distintos e compreender suas necessidades específicas. Por exemplo, o perfil dos "Usuários Corporativos" destaca a importância da localização próxima ao ambiente de trabalho e os subsídios corporativos como fatores decisivos.

Com base na pesquisa, o modelo ideal de academia deve priorizar a musculação como principal atividade, contar com bons equipamentos, e funcionar aos finais de semana. A localização deve ser estrategicamente escolhida, preferencialmente em bairros residenciais e próximos a centros corporativos, idealmente em um raio de até 1,5 km dos usuários. A faixa de preço ideal para a mensalidade varia entre R\$70 e R\$139,99, sendo que academias que aceitam benefícios corporativos podem cobrar um valor mais alto. Adicionalmente, a academia deve operar em horários comerciais para atender a todos os perfis de clientes.

A utilização de ML demonstrou ser uma ferramenta eficaz para identificar padrões de comportamento e otimizar a tomada de decisões empresariais no setor fitness. A análise dos clusters gerados forneceu insights valiosos para o desenvolvimento de estratégias de marketing e planejamento de novos empreendimentos, garantindo maior competitividade e alinhamento com as expectativas dos clientes.

Os resultados reforçam a importância de um planejamento estratégico embasado em dados para a abertura de novas academias, permitindo que empreendedores atendam de forma precisa às demandas do mercado e promovam a fidelização dos clientes.

Referências

AMARAL, F. **Aprenda Mineração de Dados**: Teoria e prática. São Paulo: Alta Books, 2016.

CÂMARA, Gilberto; LIMA, Eduardo. (2020). **Introdução ao Aprendizado de Máquina**: Análise de Dados e Inteligência Artificial com Python. São Paulo: Editora Ciência Moderna.

COELHO, Christianne de Faria; BURINI, Roberto Carlos. Atividade física para prevenção e tratamento das doenças crônicas não transmissíveis e da incapacidade funcional. **Revista de Nutrição** [online], Campinas, v. 22, p.937-946, dez. 2009. Disponível em: <https://www.scielo.br/j/rn/a/3CfMRjMyHsMGzBxKRM6jtWQ/?lang=pt>. Acesso em: 29 de março de 2024.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados**: estatística e machine learning com Excel®, SPSS®, Stata®, R® e Python®. 2 ed. Rio de Janeiro: LTC, 2024.

FIALHO, A. et al. O imaginário coletivo de estudantes de educação física sobre vida saudável. **Revista Brasileira de Ciências do Esporte**, v. 36, n. 3, p. 626-631, jul./set. 2014. Disponível em: <https://www.scielo.br/j/rbce/a/Qtwrf3HCQdKJQrszMby7td/abstract/?lang=pt>. Acesso em 31 de mar. 2024.

HAIR, Joseph F. et al. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

LIMA, Érika D. C.; FILHO, Carlos A. P. **Introdução ao Aprendizado de Máquina**. São Paulo: Novatec, 2018.

MACKENZIE, A. (2015). **A produção de previsão: o que o aprendizado de máquina quer?** **Jornal Europeu de Estudos Culturais**, 18 (4-5), 429-445. Disponível em: <https://doi.org/10.1177/1367549415577384>. Acesso em 31 de mar. 2024.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. **Conceitos sobre aprendizado de máquina**. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>. Acesso em 31 de mar. 2024.

MÜLLER, A. C., & GUIDO, S. (2017). **Introduction to Machine Learning with Python** (D. Schanafelt & Judy McConville (Eds.); 1st ed.). O'Reilly Media, Inc. https://doi.org/10.1007/978-3-030-36826-5_10. Acesso em 02 de abr. 2024.

NAHAS MV. **Atividade física, saúde e qualidade de vida**: conceitos e sugestões para um estilo de vida ativo. 6. ed. Londrina: Midiograf; 2013. Acesso em 31 de mar. 2024.

OLIVEIRA, P. B. **Análise de dados e modelagem estatística**. 3. ed. São Paulo: Editora Atlas, 2017.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013. Acesso em 02 de abr. 2024.

SANTOS, A. E. et al. **Fatores que levam as pessoas a procurarem a academia para a prática de exercício físico**. EfDeportes, Buenos Aires, 2013. Disponível em: < <https://www.efdeportes.com/efd180/fatores-que-levam-a-academia.htm> >. Acesso em 31 de mar. 2024.

SEBRAE - SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS. **Tudo o que você precisa saber para criar o seu plano de negócio**. Disponível em: <https://sebrae.com.br/sites/PortalSebrae/artigos/como-elaborar-um-plano-de-negocio,37d2438af1c92410VgnVCM100000b272010aRCRD>. Acesso em: 29 de mar. 2024.

SEBRAE - SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS. **Sebrae em Dados** - Academias e Studios. Disponível em: <https://www.sebraepr.com.br/comunidade/artigo/sebrae-em-dados-academias-e-studios>. Acesso em: 29 de mar. 2024.

SEBRAE - SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS DE SÃO PAULO. Causa Mortis: **O sucesso e o fracasso das empresas nos primeiros cinco anos de vida**. [São Paulo]: Unidade Inteligência de Mercado, 2014. Disponível em: https://sebrae.com.br/Sebrae/Portal%20Sebrae/UFs/SP/Pesquisas/causa_mortis_2014.pdf. Acesso em: 29 de mar. 2024.

SHIN, J.; BADGWELL, T. A.; LIU, K. H. & LEE, J. H. (2019). Reinforcement Learning – Overview of recent progress and implications for process control. *Computers and Chemical Engineering*, 127, 282–294. <https://doi.org/10.1016/j.compchemeng.2019.05.029>. Acesso em 02 de abr. 2024.

SIEGEL, Sidney; CASTELLAN JR., N. John. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed, 2006.

Recebido: 20/11/2025

Aprovado: 11/12/2025

Publicado: 30/12/2025