
**Inferência e Aprendizagem Automática Causal na Ciência de Dados
Educativos: Uma Revisão da Literatura**

*Inference and Causal Machine Learning in Educational Data Science: A
Literature Review*

Nuno Castro Lopes¹

Resumo

A Ciência de Dados Educativos procura obter conhecimento e prever eventos futuros com base em dados educativos. No entanto, muitas análises ainda se limitam à descrição e previsão, sem identificar as causas subjacentes. Este artigo integra duas revisões da literatura: uma narrativa sobre métodos clássicos de inferência causal (*Propensity Score Matching*, Regressão Descontínua, Diferença-em-Diferenças e Controlo Sintético) e uma revisão de revisões sobre técnicas de Aprendizagem Automática Causal (CML), como a Transformação de Resultados, os *Meta-learners* e os algoritmos *Causal Tree* e *Causal Forest*. A análise evidencia o potencial destas metodologias para avaliar o impacto real das intervenções pedagógicas e personalizar o ensino a distância e e-learning, destacando a necessidade de maior formação técnica e literacia causal entre investigadores e analistas.

Palavras-chave: Ciência de Dados Educativos; Análise de Aprendizagem; Inferência Causal; Aprendizagem Automática Causal; Heterogeneidade dos Efeitos do Tratamento

Summary

Educational Data Science seeks to gain knowledge and predict future events based on educational data. However, many analyses are still limited to description and prediction, without identifying the underlying causes. This article integrates two literature reviews: a narrative on classical causal inference methods (*Propensity Score Matching*, Regression Discontinuity, Difference-in-Differences, and Synthetic Control) and a review of reviews on Causal Machine Learning (CML) techniques, such as Outcome Transformation, Meta-learners, and Causal Tree and Causal Forest algorithms. The analysis highlights the potential of these methodologies to assess the real impact of pedagogical interventions and personalize distance learning and e-learning, emphasizing the need for greater technical training and causal literacy among researchers and analysts.

Keywords: Education Data Science; Learning Analytics; Causal Inference; Causal Machine Learning; Treatment Effects Heterogeneity

¹ Universidade Aberta, Portugal , dadoseimpacto@gmail.com, <https://orcid.org/0000-0003-2448-9798>

1. Introdução

A Prospecção de Dados Educativos (Educational Data Mining, EDM) e a Analítica da Aprendizagem (Learning Analytics, LA) são duas áreas de investigação que se focam na recolha, análise e interpretação de dados relacionados com o ensino-aprendizagem, tais como, dados sociodemográficos, de práticas pedagógicas, políticas educativas, resultados escolares, dados comportamentais e cognitivos, entre outros. Apesar de existirem algumas diferenças entre as duas disciplinas, verifica-se que ambas utilizam técnicas usuais da ciência de dados (Romero & Ventura, 2020). Assim, podemos denominar o conjunto destas duas disciplinas como: Ciência de Dados Educativos.

O objetivo da Ciência de Dados Educativos é obter conhecimento e prever eventos futuros baseados em dados educativos. No entanto, importa destacar que a realização de previsões tem por base a análise de padrões na ligação entre os dados e não na relação de causalidade que pode existir entre estes (Cunningham, 2021; Pearl & Mackenzie, 2022).

Esta diferença entre análise de padrões e causalidade é particularmente relevante quando se trabalha com dados educativos, já que estes podem ser uma mistura de dados comportamentais e cognitivos complexos e multivariados. Com esta tipologia de dados, a simples verificação de correlações não é suficiente, sendo também, necessário compreender a razão dessas relações para que a informação obtida seja útil. A correlação entre dois atributos não determina a existência de causalidade. Para se aferir a causalidade entre as variáveis é necessário realizar uma análise mais profunda, de forma a determinar quais as causas e os efeitos.

Por exemplo, num ambiente virtual de aprendizagem, como no Moodle, podemos verificar uma relação entre o tempo despendido pelo aluno a consultar o material disponibilizado pelo docente e as notas obtidas no final do semestre. A existência de uma correlação entre estas duas variáveis não determina que a nota obtida no final do semestre seja causada pelo tempo gasto a consultar a matéria, é necessário ter em consideração outros indicadores que podem ter influenciado essa relação, como por exemplo, os conhecimentos e competências prévias ou a motivação para estudar determinados conteúdos.

A Ciência de Dados Educativos, apesar de ser ainda uma área de nicho, tem vindo a assumir um papel fundamental na compreensão e melhoria dos processos de ensino-aprendizagem, principalmente em contextos com grandes volumes de dados, como é o caso do e-learning. Se, por um lado, se recorre cada vez mais a técnicas de Ciência de Dados para descrever e prever fenómenos educativos, por outro, é essencial ir além da simples previsão. Nos contextos educativos, designadamente nos ambientes virtuais de aprendizagem, a compreensão das relações de causa e efeito é indispensável para fundamentar decisões pedagógicas e políticas baseadas em evidência.

Neste contexto, a inferência causal é a abordagem utilizada para determinar a relação de causa e efeito entre variáveis. O objetivo é verificar se uma variável independente (um tratamento ou outra intervenção/ação) causa alterações na variável dependente (resultado/efeito). De acordo com a formulação clássica de Donald Rubin (1974), um aspeto fundamental da inferência causal é a noção de resultados potenciais, que se referem aos valores que a variável dependente poderia assumir sob diferentes condições de tratamento para o mesmo sujeito. Para cada indivíduo, existe um resultado potencial associado a possíveis estados de tratamento. No entanto, surge o problema fundamental da inferência causal: a impossibilidade de observar simultaneamente os

mesmos indivíduos em dois estados. Assim, apenas um desses estados pode ser observado, enquanto o outro, denominado contrafactual, permanece desconhecido. O objetivo da inferência causal é estimar os resultados contrafactuais para determinar o efeito do tratamento. (Athey, Imbens & Wager, 2018; Brand, Zhou & Xie, 2023; Imbens, 2022; Lechner, 2023).

Assim, antes de se tentar inferir a causalidade entre variáveis, é importante implementarmos estratégias para compreender as possíveis relações que podem levar da causa ao efeito. Uma das representações formais destas relações foi desenvolvida por Judea Pearl e é denominada de “Structural Causal Models” (SCM). O SCM representa as relações entre variáveis utilizando, para tal, grafos acíclicos dirigidos (DAG). Esses grafos são diagramas que exibem as relações causais, com setas dirigidas da causa (X) para o efeito (Y). Para além disso, podem existir variáveis Z no diagrama. As variáveis Z são outras variáveis que podem, ou não, influenciar o processo $X \rightarrow Y$, podendo estas serem: intermediárias, colisores ou confundidores (Cunningham, 2021; Huntington-Klein, 2022; Pearl & Mackenzie, 2022).

No exemplo da relação entre o tempo despendido a consultar os materiais (X) e as notas finais (Y), podem existir variáveis a influenciar esta relação. A variação da qualidade do serviço de internet em diferentes regiões ou países, pode ser um colisor (Z1) que pode afetar o tempo de consulta dos materiais (X), mas não diretamente as notas finais (Y). Isto significa que o aluno poderia consultar menos os materiais apenas porque a qualidade do seu serviço de internet era menos fiável. Neste caso seria fundamental controlar o Z1, para comprovar se a relação entre X e Y existe. Os conhecimentos prévios do aluno (Z2), podem ser um confundidor, já que podem afetar quer o tempo gasto a consultar (X), quer diretamente os resultados (Y). Z2 não é uma causa direta da relação $X \rightarrow Y$, já que afeta ambas. A compreensão da matéria por parte do aluno (Z3) pode ser um mediador entre X e Y, isto é, o tempo gasto a consultar o material, pode afetar a qualidade da compreensão da informação, que por sua vez afeta os resultados escolares no final do semestre ($X \rightarrow Z3 \rightarrow Y$).

A explicação detalhada deste método ultrapassa os objetivos deste artigo, sendo o nosso propósito apresentar as principais técnicas de inferência causal aplicáveis à Ciência de Dados Educativos, com enfoque tanto nos métodos clássicos como nas abordagens emergentes de Aprendizagem Automática Causal (CML – *Causal Machine Learning*)

Assim, o presente trabalho concentra-se nos métodos de aferição da causalidade que podem integrar as ferramentas da Ciência de Dados Educativos, com especial destaque para o ensino a distância. O artigo tem por objetivo sintetizar e discutir as principais abordagens de inferência causal — tanto as clássicas como as emergentes no domínio da Aprendizagem Automática Causal.

A metodologia adotada combina uma revisão narrativa das abordagens clássicas com uma revisão de revisões dedicada às técnicas mais recentes de Aprendizagem Automática Causal e à sua aplicabilidade na Educação a Distância e no *e-learning*.

Esta estratégia metodológica é descrita em detalhe na secção seguinte.

2. Metodologia da Revisão

O presente artigo integra duas abordagens de revisão da literatura para sintetizar e discutir métodos de aferição de causalidade em contextos educativos, com destaque para o e-learning:

a) Uma revisão narrativa sobre os métodos clássicos de inferência causal aplicados a dados observacionais — *Propensity Score Matching*, Regressão Descontínua, Diferença-em-Diferenças e Controlo Sintético. Esta revisão foi desenvolvida de forma iterativa, acompanhando as necessidades de enquadramento teórico ao longo do trabalho, não tendo seguido um protocolo sistemático de pesquisa e seleção de estudos. A seleção das fontes privilegiou literatura de referência e trabalhos amplamente citados na área da inferência causal.

b) Uma síntese de revisões centrada nas técnicas de Aprendizagem Automática Causal orientadas para a análise da heterogeneidade dos efeitos do tratamento (HTE) e para o Uplift Modeling. As buscas foram realizadas nas bases de dados Scopus, Web of Science, ERIC e Google Scholar, recorrendo a uma combinação de descritores que articulou termos relacionados com aprendizagem automática e causalidade — “machine learning”, “analytics”, “data mining”, “artificial intelligence”, “AI”, combinados com “causality” ou “causal” — e com a Heterogeneidade dos Efeitos do Tratamento (“HTE”, “heterogeneous treatment effects”, “CATE”, incluído como descritor de pesquisa). Foram ainda incluídos termos associados ao contexto educativo e digital, como “distance education”, “online learning”, “learning analytics”, “education data mining”, “learning management system”, “Moodle” e “personalized learning”.

O processo de revisão desenvolveu-se em duas fases complementares. Numa primeira fase, procedeu-se à identificação e análise de revisões sobre técnicas de Aprendizagem Automática Causal orientadas para a exploração dos efeitos heterogêneos do tratamento. Não tendo sido identificadas revisões focadas especificamente no contexto educativo, esta análise incidiu sobre literatura de âmbito mais geral. Numa segunda fase, foram considerados estudos de aplicação em contexto educativo, utilizados para contextualizar a discussão. Dado o número reduzido e a natureza recente das revisões identificadas, não foi aplicado um processo formal de seleção em múltiplas fases, tendo sido incluídas todas as revisões consideradas relevantes para as questões de investigação, desde que publicadas em revistas científicas com revisão por pares e indexadas nas bases de dados consultadas, assegurando a credibilidade das fontes.

As questões de investigação que orientaram a extração da informação foram as seguintes: (i) o que são os efeitos heterogêneos do tratamento e o uplift modeling, (ii) que técnicas de Aprendizagem Automática Causal são utilizadas para a sua análise, (iii) em que contextos são aplicadas, (iv) que métodos são utilizados para a sua validação, (v) quais as vantagens e limitações destas abordagens e (vi) que definições são utilizadas para o conceito de causalidade em aprendizagem automática. O processo de análise envolveu a extração de informação relevante dos artigos, seguida de codificação temática, predominantemente de natureza conceptual, permitindo a identificação de cinco temas principais relacionados com fundamentos causais, aplicações, modelos e métodos, eficiência e validação. A análise seguiu uma abordagem iterativa, permitindo identificar lacunas e melhorar progressivamente a análise.

3. Técnicas Clássicas de Inferência Causal

Um conceito central na inferência causal é o ATE – Average Treatment Effect (Efeito Médio do Tratamento), que mede a diferença média entre os resultados obtidos pelos indivíduos que recebem uma intervenção e aqueles que não a recebem, considerando toda a população (Athey & Imbens, 2019; Brand, Zhou & Xie, 2023; Imbens, 2022; Lechner, 2023; Künzel et al., 2019).

Considera-se como método mais robusto para estimar esse efeito a realização de experiências aleatórias controladas, nas quais os participantes são aleatoriamente distribuídos em dois grupos: tratamento e controlo. Este método assegura a comparabilidade entre grupos e reduz os enviesamentos, permitindo atribuir diferenças de resultados ao efeito da intervenção.

Apesar de ser considerado o mais robusto, este tipo de desenho experimental geralmente não é viável nem ético em contextos educativos, nomeadamente na análise de ambientes virtuais de aprendizagem. A necessidade de intervenções com grupos específicos, a constituição prévia de turmas ou o início tardio da avaliação, já após o começo do processo de ensino-aprendizagem, tornam muitas vezes a experiência aleatória impossível.

Assim, os métodos quase-experimentais podem colmatar estas dificuldades, utilizando dados reais com base em informação observacional. Esta abordagem permite estimar o ATE em contextos reais, oferecendo uma alternativa metodologicamente sólida quando os ensaios controlados não são possíveis. Entre os métodos mais utilizados destacam-se o *Propensity Score Matching*, a Regressão Descontínua (*Regression Discontinuity*), a Diferença-em-Diferenças (*Difference-in-Differences*) e o Controlo Sintético (*Synthetic Control*), apresentados nas secções seguintes, não segundo um critério cronológico, mas antes organizados de forma a facilitar a compreensão das diferentes abordagens de inferência causal e das suas relações.

3.1. *Propensity Score Matching*

Propensity Score Matching, é uma técnica desenvolvida por Rosenbaum & Rubin (1983) para minimizar vieses que possam ocorrer na comparação entre grupos não equivalentes, tanto em estudos experimentais quanto na análise causal de dados observacionais. A distribuição aleatória dos participantes em grupos de tratamento/intervenção e os de controlo, muitas vezes, não é proporcional às covariáveis que podem afetar o resultado do tratamento/intervenção. Para minimizar este problema, a técnica gera uma “pontuação de propensão” que agrupa todas as potenciais covariáveis, possibilitando um equilíbrio na criação dos grupos de controlo. A realização desta técnica possibilita que a comparação seja realizada entre grupos de tratamento/intervenção e controlo com pontuação de propensão equivalentes. Assim, numa primeira fase é importante verificar quais as covariáveis que podem afetar os resultados do estudo (verificar quais os Z confundidores). A identificação destas covariáveis deve ser feita através da revisão da literatura especializada, por exemplo, nos estudos relacionados com o desempenho escolar, é comum considerar variáveis como o género, a educação dos encarregados de educação (principalmente da mãe), localização habitacional, rendimentos e composição do agregado, entre outros. Depois desta identificação, se se verificar que não existem desequilíbrios na distribuição das covariáveis entre os dois grupos, não existe necessidade da aplicação desta técnica. Caso se verifiquem esses desequilíbrios, pode-se recorrer a um conjunto de métodos para criação destas pontuações de propensão. O método mais comum é a utilização da

regressão logística, mas também se pode aplicar uma análise discriminante, um modelo probit, árvores de classificação, redes neuronais, árvores de decisão, máquinas de vetor suporte ou meta classificadores (Cunningham, 2021; Fan & Nowell, 2011; Huntington-Klein, 2022; Imbens, 2000; Rosenbaum & Rubin, 1983; Tu, 2019).

Por exemplo, Foo, Cheung e Chu (2021), tentaram analisar o efeito do ensino online, durante a pandemia COVID-19, num grupo de alunos do quarto ano de medicina. O estudo comparou os resultados obtidos por estes alunos com os do ensino presencial clássico. Para criar um grupo de controlo, utilizaram a metodologia *Propensity Score Matching*, para equilibrar as covariáveis relacionadas com o desempenho académico ocorrido no semestre anterior, o qual tinha sido em regime presencial para todos.

Noutro estudo, Wang, Wen & Rosé (2016) investigaram a relação entre comportamentos dos alunos em situações de debate num ambiente de aprendizagem online. Tratou-se de um curso de “Introdução à Psicologia”, oferecido pela plataforma MOOC Coursera. A massificação de utilizadores nestas plataformas permite a recolha big data educacional. Neste estudo os autores compararam os alunos que utilizaram estratégias cognitivas de ordem superior (como análise, síntese e avaliação da informação) e o foco nos materiais de aprendizagem disponibilizados. A pontuação de propensão foi calculada com base em variáveis relacionadas ao envolvimento prévio dos alunos noutros cursos na mesma plataforma. Desta forma, foi possível comparar os alunos que demonstravam estas estratégias cognitivas com aqueles que não as evidenciaram, tendo em conta que as variáveis relacionadas com o envolvimento prévio nos cursos estavam equilibradas em ambos os grupos.

3.2. Regressão Descontínua

A regressão descontínua é um método desenvolvido por Thistlethwaite e Campbell em 1960 (Thistlethwaite & Campbell, 2017). Estes autores, compararam a probabilidade de dois grupos de alunos receberem bolsas de estudo, tendo por base diferentes critérios de elegibilidade. Um dos grupos (constituído por 5126 alunos) tinha recebido anteriormente um certificado de mérito escolar, enquanto o outro grupo (2848 alunos) tinha apenas obtido uma carta de recomendação. A hipótese dos autores sugeria que o reconhecimento público obtido pelo certificado de mérito aumenta a probabilidade de um aluno receber bolsa de estudos numa universidade dos Estados Unidos e que o mesmo não ocorre quando apenas se obtém uma carta de recomendação. Assim, os autores utilizaram o facto de se ter “recebido o certificado de mérito” como causa (X), sendo o efeito (Y) a “probabilidade de receber bolsas de estudo”.

O valor de corte é um elemento fundamental na técnica de regressão descontínua. A lógica é que a elegibilidade para tratamento/intervenção é determinada por esse ponto de corte, que pode ser uma variável contínua, uma pontuação ou um índice. No estudo de Thistlethwaite e Campbell, o ponto de corte era a nota mínima para receber o certificado de mérito. Esta teoria postula que as observações que estão perto do valor de corte, dentro de um determinado limite superior e inferior (limite de banda) não são muito diferentes. Neste caso os alunos que tiveram notas um pouco inferiores ou um pouco superiores ao limite para receber certificado, não são muito díspares. Esta divisão entre grupo de tratamento/intervenção e grupo de controlo, com base no corte e dentro da largura de banda, assemelha-se a uma distribuição aleatória dos grupos. Para que ocorra esta distribuição semelhante à aleatoriedade na escolha, é fundamental que não seja possível sobreposição dos grupos. Neste método a escolha do limite de banda é fundamental: um limite de banda muito estreito incluirá poucos casos, mas muito semelhantes; o aumento da largura de banda proporciona um maior número de

indivíduos a participar no estudo, mas pode levar a uma menor a semelhança entre si (Cook, 2008; Huntington-Klein, 2022; Gopalan, Rosinger & Ahn, 2020; Thistlewaite & Campbell, 2017).

Apesar das suas vantagens, só alguns estudos sobre impacto de intervenções educativas podem ser realizados com esta técnica, já que obriga à existência de um corte claro que divida as observações dos grupos de tratamento/intervenção e controlo. Por exemplo, pode ser utilizado para avaliar o rendimento escolar dos alunos universitários bolsheiros ou não bolsheiros ou do ensino básico ou secundário beneficiários ou não do Serviço de Ação Social Escolar (SASE), entre outros exemplos. Mas a complexidade multifatorial de escolha dos grupos de intervenção da maioria das intervenções educativas, limita a utilização da regressão descontínua.

Um exemplo de utilização desta técnica no âmbito da Ciência de Dados Educativos, pode ser observada no estudo realizado por Heinesen (2018). Este investigador, analisou os dados de admissão dos estudantes dinamarqueses no ensino pós-secundário, entre os anos de 1994 e 2002. Heinesen pretendeu verificar se a escolha de um curso como primeira opção teve alguma influência em 3 indicadores: desempenho, conclusão do curso e ganhos pós curso. Assim, o corte correspondia à nota mínima de entrada em cada curso, a largura de banda dizia respeito aos alunos que entraram como primeira opção mas com uma nota perto do limite e os que tiveram de ir para um 2º curso porque a sua nota era ligeiramente inferior à mínima de entrada na primeira escolha.

Outro exemplo, pode ser encontrado no estudo de Vergolini & Zanini (2015) sobre a avaliação do impacto da Bolsa 5B aplicada na região de Trento, em Itália. Esta bolsa foi criada para apoiar alunos com notas finais do secundário iguais ou superiores a 93 (em 100) e com rendimentos familiares inferiores a 30 mil €. Para aferir se a obtenção de bolsa pode ter determinado a matrícula destes alunos no ensino superior e, principalmente, a escolha de universidades fora da região, foram comparados os estudantes com rendimentos baixos semelhantes (inferiores aos 30 mil €), mas que as suas notas de final do secundário estavam um pouco acima ou abaixo dos 93, dentro de uma determinada largura de banda.

3.3. Diferença em Diferenças

A diferença em diferenças (DiD) é uma das técnicas de investigação quase-experimentais mais clássicas e amplamente utilizadas nas ciências sociais. A sua origem remonta a 1855, quando o epidemiologista John Snow tentou provar que a cólera era transmitida pela água. Até essa data, acreditava-se que esta doença era transmitida pelo ar através de odores fétidos e venenosos (teoria miasmática). Esta crença fazia com que se impusesse quarentenas obrigatórias que não tinham qualquer efeito na diminuição da transmissão da doença. Snow e os outros médicos da época, não sabiam como responder a estas mortes constantes provocadas por vômitos e diarreias.

Depois de observar vários pacientes e analisar as suas características socioeconómicas, bem como as rotas de transmissão da doença, este epidemiologista colocou a hipótese da transmissão ocorrer através da água. Na época não se realizavam estudos clínicos controlados, mas mesmo que esse método já fosse utilizado, seria eticamente reprovável expor um dos grupos de investigação a água potencialmente contaminada. Assim, Snow enveredou por uma outra estratégia. Observou as empresas que abasteciam de água a cidade de Londres, numa época onde assolava uma epidemia de cólera. Verificou que todas as empresas até 1849 recolhiam a água perto dos esgotos

que desaguavam no rio Tâmsa, mas que após esta data, uma das empresas alterou o seu local de abastecimento, para um outro local mais a norte e afastado de qualquer descarga de detritos humanos. Se a teoria de Snow estivesse correta, as casas abastecidas por esta empresa, teriam taxas de mortalidade por cólera inferiores às outras. Assim, analisando dois momentos temporais 1849 e 1854, as casas abastecidas pela empresa que alterou o seu local de abastecimento, tiveram uma diminuição considerável no número de mortes comparativamente com as abastecidas por outras empresas. Importa referir que Snow verificou que existiam semelhanças entre as famílias analisadas, já que em muitas situações o abastecimento era realizado por uma empresa de um lado de uma rua e do outro por outra empresa, no mesmo bairro (Caniglia & Murray, 2020; Cunningham, 2021)

No séc. XX, a técnica DiD passou a ser utilizada frequentemente na econometria, nomeadamente para avaliação do impacto de políticas públicas. Um dos estudos clássicos, envolve a análise da introdução do salário mínimo na cidade de Portland e a comparação das alterações ocorridas com outras cidades localizadas no mesmo estado dos Estados Unidos (Oregon). Para aplicação do método é necessário ter dois grupos populacionais (g) e dois momentos temporais (t). Um grupo corresponde ao grupo de controlo (g0) e o outro ao de tratamento/intervenção (g1). Os períodos temporais correspondem a uma fase anterior (t0) e outra posterior (t1) à implementação do tratamento/intervenção (este também pode ser utilizado para comparar acontecimentos naturais ou ações humanas como: desastres, guerras, crises económicas, etc.). Para se calcular a diferença em diferenças, é necessário realizar uma dupla subtração: primeiro entre os períodos antes e depois no grupo de tratamento/intervenção (t1g1-t0g1) e no grupo de controlo (t1g0- t0g0); e uma segunda subtração entre a diferença entre essas duas diferenças ([t1g1- t0g1] - [t1g0- t0g0]).

Usualmente, para estimar o efeito do tratamento/intervenção, a maioria dos estudos utiliza um modelo de regressão, sendo a equação:

$$Y_i = \alpha + \beta g_i + \gamma t_i + \delta(g_i * t_i) + \epsilon_i \quad (1)$$

Em que:

Y_i é a variável estudada;

α é o valor esperado da variável estudada quando se analisa o grupo de controlo antes da mudança estudada (termo constante);

β é o impacto do grupo de intervencionado na variável estudada (diferença permanente entre comparação e intervenção)

γ é o impacto no segundo período (pós intervenção) na variável estudada (tendência no tempo comum entre comparação e intervenção);

δ é o impacto pós intervenção no grupo de intervenção em relação ao grupo de controlo sobre a variável estudada (efeito verdadeiro da intervenção);

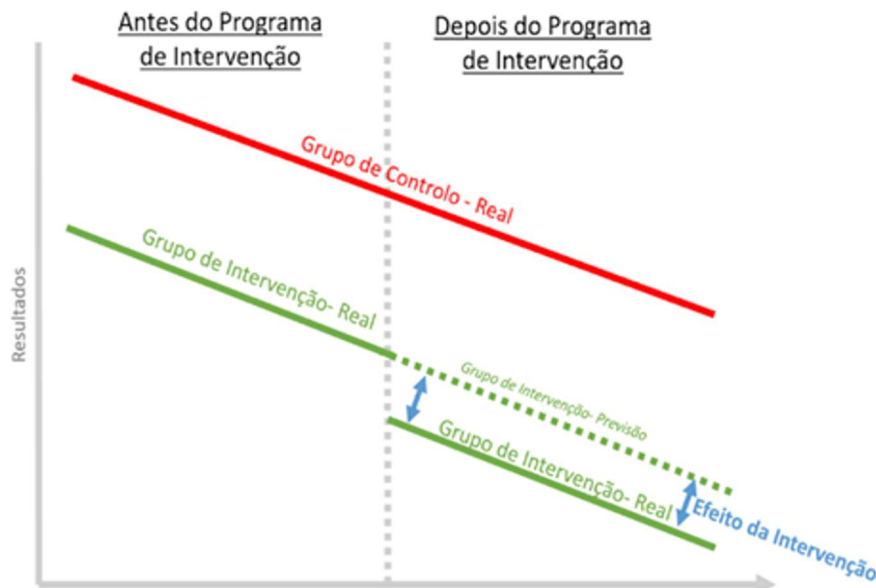
ε é o erro aleatório não observado (variáveis não incluídas que podem ser colisores ou confundidores)

Como podemos observar no exemplo da figura 1, o grupo de controlo (g0), apesar de ter resultados superiores na fase pré-intervenção (t0), comparativamente com o grupo de intervenção (g1), a trajetória (neste caso descendente) é semelhante em ambos os grupos. Esta metodologia postula que, se essa trajetória é semelhante pré-programa (t0), também o deveria continuar numa fase posterior (t1), caso não tivesse ocorrido o tratamento/intervenção (ver comparação entre grupo de controlo –Real e Grupo de

Intervenção-Previsão). Nesta lógica, o efeito da intervenção é a diferença entre a previsão e a realidade ocorrida no grupo de intervenção (g1) (Cunningham, 2021 ; Huntington-Klein, 2022; Gopalan, Rosinger & Ahn, 2020; Lechner, 2011).

Figura 1

Exemplo da aplicação da metodologia DiD na avaliação de um programa de intervenção.



Fonte: Elaboração própria com base na literatura

Para validar a metodologia DiD, é fundamental que exista uma trajetória semelhante entre g0 e g1 na fase t0. Para além desta premissa, pode-se também utilizar outras estratégias para validar o processo, tais como: 1) realizar um estudo placebo, onde g1 é substituído por um grupo que não foi intervencionado (g1-Fake) e verificar se o efeito é zero; 2) utilizar apenas o tempo t0, dividi-lo em dois supostos tempos t0 e t1-Fake e verificar se o efeito é zero; 3) realizar vários estudos semelhantes, variando os grupos de controlo (vários g0 com trajetórias semelhantes a g1 na fase t0) e verificar se o efeito é semelhante (Huntington-Klein, 2022).

Um exemplo de aplicação da técnica DiD na Ciência de Dados Educativos, pode ser verificada no estudo de López-Torres, Prior & Santín (2019), que tentaram analisar o efeito de um programa implementado em escolas do ensino básico e secundário na Catalunha, em Espanha. Esse programa foi implementado entre 2010/11 e 2013/14 e tinha por objetivo melhorar a qualidade do ensino e reduzir o absentismo escolar, principalmente em escolas localizadas em zonas com maior propensão para a existência de vulnerabilidades socioeconómicas e socioculturais. A participação dependia da escolha das direções das escolas e do cumprimento de um conjunto de pré-requisitos. O objetivo inicial dos autores era realizar um estudo tipo DiD tendo como períodos temporais antes e depois da implementação do programa e como grupos escolas intervencionadas com o programa e como controlo escolas que não usufruíram do programa. Os autores não conseguiram realizar o estudo seguindo a metodologia inicialmente pensada, porque não tinham dados suficientes para analisar a semelhança de trajetórias das escolas na fase pré- intervenção. Um estudo relativamente semelhante foi desenvolvido por Cortes, Moussa & Weinstein, (2013) que compararam os resultados escolares entre as escolas que atribuem bolsas de mérito (grupo de intervenção), com as escolas que não ofereciam essas bolsas (contrafactual).

A falta de uma trajetória semelhante entre grupos ou a dificuldade de avaliar esta trajetória, na fase pré-intervenção, limita a aplicação da metodologia DiD em muitas situações de análise do impacto de intervenções e ações educativas e/ou de aprendizagem.

3.4. Controlo Sintético

O Controlo Sintético é uma metodologia que pode, em parte, colmatar a dificuldade em se encontrar grupos de controlo com trajetórias semelhantes numa fase pré-intervenção, quando se deseja aplicar a metodologia DiD. Este método foi desenvolvido por Abadie & Gardeazabal (2003), inicialmente com o objetivo de tentar avaliar o impacto que o terrorismo do grupo ETA teve no desenvolvimento económico do País Basco. Para se compreender de que forma este acontecimento influenciou, durante anos, a economia desta região espanhola, seria necessário comparar esta com alguma região com características semelhantes e que não tivesse ocorrido o número de atentados que aí sucederam. Apesar da Catalunha ser uma região com muitas semelhanças com o País Basco, estes autores verificaram que esta não era suficientemente semelhante para ser aplicado do método DiD.

Assim, a proposta dos autores foi utilizar uma nova metodologia, que permitisse a criação de uma nova região contrafactual: um País Basco Sintético. Um contexto sintético, é um contexto desenvolvido artificialmente, através da combinação ponderada de covariáveis de outros contextos. Neste caso, o País Basco Sintético foi desenvolvido tendo por base a ponderação de alguns indicadores económicos relacionados com o PIB de outras regiões espanholas. Assim, para a construção do grupo sintético, é necessário primeiro escolher as covariáveis relacionadas com a variável dependente (neste caso o PIB) e escolher grupos que tenham semelhanças com o grupo de intervenção. Escolhidas esses grupos, realiza-se a ponderação dessas covariáveis através de métodos de otimização que minimizem as diferenças entre o grupo sintético e o grupo real de intervenção. A ideia é escolher um grupo de controlo que seja o mais parecido possível com o grupo de intervenção em termos dessas variáveis, de forma que quaisquer diferenças observadas entre os dois grupos após a intervenção possam ser atribuídas à própria intervenção. No estudo inicial de Abadie & Gardeazabal, na fase anterior ao início da campanha terrorista do grupo ETA (pré-tratamento), os valores dos indicadores relacionados com o PIB eram muito semelhantes no País Basco real e no sintético. O mesmo não foi observado na fase pós-tratamento, em que se observou uma significativa melhoria no PIB da região sintética, comparativamente com a realidade ocorrida no País Basco. Para validar esta metodologia, os autores realizaram o mesmo procedimento relativamente à região da Catalunha. Como esta região não sofreu o mesmo nível de terrorismo durante estes anos, este procedimento serviu de contrafactual para a validade da utilização desta técnica, neste estudo. Verificou-se assim, validade nesta metodologia, já que o percurso do PIB da Catalunha real foi muito semelhante ao ocorrido na Catalunha Sintética no período posterior ao início das atentados, contrariamente ao que aconteceu no País Basco. Mais tarde, outros estudos semelhantes foram realizados, como a avaliação do impacto no consumo de tabaco na Califórnia depois da implementação em 1989 de uma lei que aumenta os impostos sobre estes produtos (Abadie, Diamond & Hainmueller, 2010; Cunningham, 2021 ; Huntington-Klein, 2022).

Um exemplo do uso da técnica de Controle Sintético na Ciência de Dados Educativos, pode ser verificado no estudo de Birdsall (2017), que pretendeu verificar o efeito que pode ter na gestão escolar a forma de financiamento do sistema. Nos Estados Unidos existem dois sistemas de financiamento escolar, um baseado na valor-desempenho, em

que o financiamento depende dos resultados escolares em testes padronizados e na frequência dos alunos, e outro sistema baseado no valor-aluno, que tem em consideração a ponderação das características sociais, culturais e económicas dos alunos. Segundo este autor, este último sistema proporciona aos diretores escolares uma maior liberdade de gestão, permitindo decisões mais baseadas nas necessidades dos alunos e não tanto na obtenção (média) de resultados de desempenho. No estado do Texas, apenas no distrito escolar de Houston se aplica o sistema valor-aluno. Birdsall queria confirmar estudos anteriores, que não tendo contrafactual postulavam que os sistemas valor-aluno, a longo prazo, poderiam melhorar os resultados padronizados dos alunos, já que as escolas tinham maior autonomia para intervir com as comunidades menos favorecidas. O autor verificou que não poderia aplicar uma metodologia clássica de DiD neste estudo, já que o único distrito escolar com características semelhantes a Houston no Texas era Dallas, mas não se verificavam tendências semelhantes nos dois distritos escolares na fase pré-intervenção, ou seja, numa época em que os dois distritos ainda usavam o sistema valor-desempenho. Assim, a opção do autor foi criar um distrito escolar sintético semelhante ao de Houston, baseado num conjunto de covariáveis de 29 outros distritos escolares do Texas, que tivessem pelo menos mais de 25 mil alunos.

Várias técnicas foram, entretanto, desenvolvidas para melhorar o método de criação de grupos de controlo sintético. Uma dessas abordagens foi apresentada pela Google (Brodersen, Gallusser, Koehler, Remy & Scott, 2015), que utiliza modelos de séries temporais estruturais bayesianas para modelar os dados do grupo sintético e gerar previsões (Causallmpact). Segundo os autores, uma das vantagens desta abordagem é a capacidade de inferir impacto causal, mesmo em situações em que existem dados incompletos, com erros ou inconsistentes. Tendo em consideração o número avassalador de utilizadores dos produtos da Google, a empresa foi capaz de comparar os efeitos da aplicação do novo método com um teste tradicional A/B. Realizaram a alteração a uma determinada página, mas apenas numa determinada região. Nessa região, foi realizada uma alteração à página e foi avaliado o tráfego. Utilizando esta técnica, a Google criou um grupo sintético em que reconstruiu qual teria sido o tráfego da página caso não tivesse ocorrido a alteração. Para validar a técnica, realizou um teste A/B (analisando pré e pós de cada utilizador) entre a região em que alterou a página e uma região semelhante em que não realizou essa alteração. O estudo de validação confirmou a capacidade desta técnica para inferir causalidade.

Durante a Pandemia COVID-19 passaram dois grandes tufões nas Filipinas, o Goni e o Vamco. Durante esse período, a maioria dos alunos filipinos encontrava-se em casa em ensino remoto. Lagmay e Rodrigo, tentaram assim compreender de que forma a passagem destes tufões afetou o envolvimento dos alunos e dos docentes nas atividades de ensino a distância. Tendo por base um período pré tufões (de 9 de setembro a 28 de outubro de 2020), um durante a crise dos tufões (29 de outubro a 13 de novembro de 2020) e um numa fase pós tufões e até ao final do período letivo (de 14 de outubro a 23 de dezembro), os autores analisaram as atividades realizadas pelos alunos e professores de uma universidade de Manila na plataforma Moodle. Para realizar esta análise os autores criaram um contrafactual utilizando a técnica Causallmpact desenvolvida pela Google.

Ainda são poucos os estudos que utilizaram estes modelos de séries temporais estruturais bayesianas para modelar dados contrafactuais em estudos relacionados com Ciência de Dados Educativos.

4. Aprendizagem Automática Causal e Heterogeneidade dos Efeitos

Em diversos contextos, há uma necessidade crescente de compreender para além do ATE, os efeitos dos tratamentos tanto a nível individual quanto em determinados subgrupos. Nesse âmbito, surgiram duas correntes de investigação: uma focada na análise da Heterogeneidade do Efeito do Tratamento (HTE-*Heterogeneous Treatment Effects*) e outra no Uplift Modeling. Estas correntes, conceptualmente próximas, mas oriundas de tradições distintas: HTE é mais estudada na comunidade da estatística e econometria, enquanto a Uplift Modeling é mais relacionada com investigações na área do marketing e dos sistemas de informação (Rößler & Schoder, 2022; Zhang, Li e Liu, 2021). Apesar de algumas diferenças entre elas, ambas as abordagens têm como objetivo explorar e prever o efeito do tratamento que varia na população, seja em subgrupos ou a nível individual, comparando características semelhantes (covariáveis) nos grupos de tratamento e de controlo (contrafactual). Esse processo é denominado Efeito Médio Condicional do Tratamento (CATE- *Conditional Average Treatment Effect*). Calcular o CATE permite estimar o Efeito Individual do Tratamento (ITE-*Individual Treatment Effect*).

Para estimar o CATE, têm surgido, nos últimos anos, diversos métodos de Aprendizagem Automática Causal (CML), resultantes de investigações tanto nas comunidades de HTE quanto de *Uplift Modeling*. A Aprendizagem Automática Causal combina metodologias de Aprendizagem Automática com os princípios fundamentais da inferência causal (Lechner, 2023). Desta forma, os métodos de Aprendizagem Automática são utilizados para responder a questões causais utilizando grandes conjuntos de dados. Enquanto a Aprendizagem Automática tradicional realiza previsões baseando-se em padrões correlacionais para antecipar resultados futuros, a Aprendizagem Automática Causal procura discernir relações causais que expliquem como as variáveis interagem (Lechner, 2023).

Assim, muitos dos métodos de CML foram desenvolvidos especificamente para estimar o CATE. Diferentes autores têm agrupado estes métodos em diversas categorias (Caron, Baio & Manolopoulou, 2022; Devriendt, Moldovan & Verbeke, 2018; Gutierrez & Gérardy, 2017; Hu, 2023; Zhang, Li & Liu, 2021). Por um lado, temos os métodos que utilizam algoritmos comuns de Aprendizagem Automática supervisionada, os quais podem envolver adaptações ou transformações iniciais das variáveis, ou a combinação de métodos tradicionais através de configurações particulares. Por outro lado, existem métodos que desenvolveram novos algoritmos especificamente para estimar diretamente o HTE ou o *Uplift Modeling*.

4.1. Abordagens que adaptam algoritmos clássicos de Aprendizagem Automática

Para compreendermos melhor esta abordagem que adapta algoritmos clássicos de Aprendizagem Automática, importa entender brevemente como funciona a aprendizagem supervisionada.

Segundo Gama, Lorena, Faceli, Oliveira e Carvalho (2017), a Aprendizagem Automática consiste num processo em que o computador aprende a partir de exemplos anteriores, descobrindo padrões nos dados para fazer previsões sobre novos casos. Na aprendizagem supervisionada, tal como salienta Lechner (2023), o objetivo é encontrar uma relação entre um conjunto de variáveis explicativas (covariáveis, X_1, X_2, \dots, X_n) e uma variável de resultado (Y). Existem dois tipos principais de tarefas: classificação, quando se prevê categorias (como aprovado ou reprovado), e regressão, quando se

prevê valores contínuos (como a nota final de um estudante). De forma simples, o modelo é treinado com dados conhecidos, aprendendo a associar $X = (X_1, X_2, \dots, X_n)$ a Y , e depois é usado para prever Y em novos casos.

Por exemplo, num curso em educação a distância, um modelo de Aprendizagem Automática pode ser usado para prever se um estudante irá concluir o curso ou desistir, com base nos dados registados na plataforma de aprendizagem — como número de acessos semanais, tempo médio de permanência, realização de tarefas, participação em fóruns, visualização de vídeos, resultados das avaliações, juntamente com outros dados sociodemográficos do estudante. Neste caso, as covariáveis correspondem a X , e Y assumiria o valor 1 para conclusão e 0 para desistência.

Assim, existem abordagens que, tratando e organizando a forma como os dados são utilizados nos algoritmos de aprendizagem supervisionada, permitem ir além da simples previsão de resultados. Esses métodos possibilitam estimar relações causais e identificar como o efeito de uma intervenção pode variar entre diferentes indivíduos ou grupos, isto é, obter os Efeitos Heterogêneos do Tratamento (HTE).

4.1.1 Transformação de Resultados

A abordagem de transformação de resultados modifica o resultado observado (Y) num novo resultado transformado (Y^*), que incorpora, além da informação do efeito Y , também a informação da variável de tratamento (T) num só valor (Devriendt, Moldovan & Verbeke, 2018; Zhang, Li & Liu, 2021). Ao converter as variáveis T e Y numa única variável Y^* , podemos prever essa nova variável apenas com base nas covariáveis, tal como num problema clássico de Aprendizagem Automática. Assim, prever Y^* corresponde a estimar o CATE.

Um dos métodos mais referidos na literatura é o proposto por Jaskowski e Jaroszewicz (2012), sendo particularmente útil em casos de resultados binários. Esta abordagem considera que Y^* é igual a 1 se o resultado da intervenção for pelo menos tão bom quanto o resultado que teria ocorrido no grupo de controlo, caso soubéssemos o resultado em ambos os grupos. Na prática, $Y^* = 1$ quando há resultado positivo com intervenção ($T = 1$ & $Y = 1$) ou quando o resultado nulo decorre da ausência de intervenção ($T = 0$ & $Y = 0$); em todas as outras situações, $Y^* = 0$ (Devriendt, Moldovan & Verbeke, 2018; Gutierrez & Gerardy, 2017; Jaskowski & Jaroszewicz, 2012; Lopes & Cavique, 2023; Pinheiro & Cavique, 2022).

Desta forma, esta técnica permite transformar um problema causal num problema de previsão supervisionada, possibilitando a utilização de algoritmos clássicos de Aprendizagem Automática para estimar efeitos causais.

No âmbito da Ciência de Dados Educativos, este tipo de abordagem pode ser particularmente útil para identificar quais as estratégias pedagógicas que funcionam melhor para diferentes perfis de estudantes, contribuindo para uma maior personalização das intervenções educativas. Importa referir que um cientista de dados educativos, mesmo sem conhecimentos em programação, pode realizar este tipo de análise utilizando apenas uma folha de cálculo para preparar os dados e a ferramenta gratuita Orange Data Mining para aplicar os modelos. Por outro lado, quem possui conhecimentos em programação pode implementar esta abordagem através da transformação da variável de resultado, reformulando o problema para posterior aplicação de algoritmos de Aprendizagem Automática supervisionada, como modelos de classificação ou regressão, disponíveis em diferentes ambientes de programação.

4.1.2 Meta-learners

Relativamente à combinação de métodos tradicionais através de configurações particulares, surgem os Meta-learners ou meta-algoritmos, que são estruturas algorítmicas agnósticas projetadas para estimar o CATE utilizando qualquer algoritmo de Aprendizagem Automática supervisionado (aprendiz base) (Hu, 2023; Künzel et al., 2019). Estes métodos têm a capacidade de processar diferentes segmentos de dados e diferentes aspetos do problema para obter uma estimativa do CATE. Desta forma, são muito flexíveis, podendo adaptar-se a vários tipos de dados e ser integrados em sistemas de análise já existentes (Künzel et al., 2019). Os Meta-learners mais citados na literatura são o S-Learner, o T-Learner e o X-Learner (Devriendt, Moldovan & Verbeke, 2018; Gutierrez & Gerardy, 2017; Künzel et al., 2019).

O S-Learner é o mais simples destes métodos. Utiliza um único modelo que recebe como entrada as características do indivíduo (X) e a variável que indica se houve ou não intervenção (T). O modelo aprende a prever o resultado (Y) com base nessas informações. Por exemplo, num curso de educação a distância, o S-Learner pode ser usado para estimar o impacto de uma tutoria personalizada na taxa de conclusão. O modelo seria treinado com dados sobre os estudantes — como número de acessos à plataforma, tempo médio de estudo, realização de tarefas, participação em fóruns e notas — juntamente com a variável que indica se o estudante recebeu tutoria (T=1) ou não (T=0). Depois de treinado, o modelo pode estimar o resultado de cada estudante em dois cenários: com e sem tutoria; a diferença entre os dois valores representa o efeito individual da intervenção (Künzel et al., 2019).

O T-Learner utiliza dois modelos separados: um para o grupo de estudantes que participou na intervenção (T=1) e outro para o grupo que não participou (T=0). Cada modelo aprende a prever o resultado com base nas características dos estudantes de cada grupo. No mesmo exemplo anterior, seriam criados dois modelos — um baseado nos alunos com tutoria e outro nos alunos sem tutoria. Para estimar o efeito em novo estudante, calculam-se as previsões dos dois modelos e subtrai-se o resultado do grupo sem tutoria ao do grupo com tutoria (Künzel et al., 2019).

O X-Learner é considerado um método mais robusto, embora também mais complexo. Este método combina os princípios do T-Learner com um modelo adicional que estima a probabilidade de um estudante ter recebido a intervenção (propensão - ver ponto 3.1.). Em seguida, utiliza essas probabilidades para ponderar as estimativas dos modelos anteriores, tornando os resultados mais fiáveis — especialmente quando o número de estudantes com e sem intervenção é diferente. Em termos simples, o X-Learner procura equilibrar os grupos e melhorar a precisão das estimativas causais, mesmo quando os dados estão desequilibrados (Acharki et al., 2023)

Para além destes, outros métodos são referidos na literatura, como o R-Learner (Caron, Baio & Manolopoulou, 2020; Hu, 2023; Rößler & Schoder, 2022; Zhang, Li & Liu, 2021), o Multitask-Learner e o τ -Learner (Caron, Baio & Manolopoulou, 2020), geralmente aplicados a contextos mais complexos ou com múltiplas intervenções.

Na Ciência de Dados Educativos, estas abordagens permitem não apenas identificar quais as estratégias pedagógicas mais eficazes para diferentes perfis de estudantes, mas também apoiar a tomada de decisão docente, otimizar intervenções educativas e aprofundar a compreensão dos fatores que influenciam o sucesso na aprendizagem. Mesmo sem conhecimentos de programação, é possível realizar este tipo de análise com uma folha de cálculo de para preparação de dados e a ferramenta gratuita Orange

Data Mining para aplicar os modelos de forma visual e intuitiva. Para utilizadores com conhecimentos em programação, o pacote “*causalToolbox*” permite implementar os métodos S-Learner, T-Learner e X-Learner em linguagem R (Zhang, Li & Liu, 2021) existindo igualmente bibliotecas em Python, como “*CausalML*”, orientadas para inferência causal e análise de efeitos de tratamento.

4.2 Algoritmos específicos de Aprendizagem Automática Causal

Para além dos métodos que adaptam algoritmos clássicos de Aprendizagem Automática, também têm surgido algoritmos de Aprendizagem Automática Causal desenvolvidos especificamente para estimar diretamente o CATE. Nas revisões da literatura sobre métodos diretos, é descrito um conjunto muito alargado de algoritmos (Caron, Baio & Manolopoulou, 2022; Devriendt, Moldovan & Verbeke, 2018; Gutierrez & Gérardy, 2017; Hu, 2023; Zhang, Li & Liu, 2021). Zhang, Li e Liu (2021) agrupam estes métodos em quatro categorias principais: métodos de modelação de Uplift baseados em Máquinas de Vetores de Suporte, métodos baseados em Árvores, métodos baseados em Ensemble e métodos baseados em Deep Learning.

Entre estes, destacam-se dois algoritmos amplamente referidos na literatura mais recente: a *Causal Tree* (CT), baseada em árvores de decisão, e a *Causal Forest* (CF), baseada em métodos ensemble (Athey & Imbens, 2016; Wager & Athey, 2018).

A *Causal Tree* deriva das Árvores de Decisão tradicionais, que dividem os dados em ramos até formar grupos mais homogéneos. No entanto, ao contrário das árvores clássicas, o objetivo da *Causal Tree* não é apenas prever o resultado com precisão, mas identificar subgrupos que respondem de forma diferente à intervenção. O algoritmo divide os dados em dois conjuntos: um é usado para construir a árvore, estabelecendo as regras de segmentação, e o outro é utilizado para calcular o efeito médio da intervenção em cada nó terminal, comparando os resultados médios entre grupos de intervenção e de controlo. Esta separação é designada por abordagem “honesta”, pois evita enviesamentos que poderiam ocorrer se o mesmo conjunto de dados fosse usado para treino e validação (Lopes & Cavique, 2023)..

De forma análoga às Florestas Aleatórias (*Random Forests*), que combinam várias árvores de decisão para aumentar a precisão e reduzir o sobreajuste, as *Causal Forests* constroem múltiplas árvores causais e combinam as suas estimativas, obtendo resultados mais robustos e fiáveis sobre o efeito da intervenção (Wager & Athey, 2018). Este processo permite capturar de forma mais estável a heterogeneidade dos efeitos do tratamento, mesmo em amostras grandes e com elevado número de variáveis.

No contexto da Ciência de Dados Educativos, estes algoritmos podem ser aplicados para analisar como diferentes grupos de estudantes reagem a estratégias pedagógicas ou recursos digitais, identificando em que condições cada intervenção é mais eficaz. Contudo, ao contrário de métodos anteriores que podem ser utilizados sem recorrer a programação, a aplicação destes algoritmos requer conhecimentos técnicos de programação, uma vez que são executados através de código. Em linguagem R, podem ser utilizados os pacotes “*causalTree*” para a implementação da *Causal Tree* e “*grf*” para a *Causal Forest* (Zhang, Li & Liu, 2021), podendo também ser implementados em Python através de bibliotecas como “*EconML*” ou “*CausalML*”.

4.3 Métricas para avaliação dos modelos causais

A avaliação dos modelos de Aprendizagem Automática Causal apresenta desafios

distintos das abordagens tradicionais de Aprendizagem Automática supervisionada. Nos modelos preditivos clássicos, a precisão é avaliada comparando as previsões com os resultados observados, normalmente através de métricas como o erro médio ou a acurácia. No entanto, na inferência causal, este procedimento não é possível, uma vez que nunca se pode observar simultaneamente o resultado de um mesmo indivíduo com e sem tratamento — o chamado problema fundamental da inferência causal.

Para contornar esta limitação, os modelos causais são avaliados comparando grupos semelhantes que receberam tratamentos diferentes, e observando se o modelo é capaz de identificar corretamente os indivíduos que mais beneficiam da intervenção.

Entre as métricas mais utilizadas estão a curva de ganho incremental e o coeficiente Qini, que avaliam a capacidade do modelo em distinguir os grupos com maior efeito positivo da intervenção (Devriendt, Moldovan & Verbeke, 2018; Gutierrez & Gérardy, 2017; Olaya et al., 2020).

4.4. Aprendizagem Automática Causal no Ensino

A investigação aplicada à estimativa do CATE utilizando métodos de CML tem sido predominantemente desenvolvida no contexto do marketing, mas também têm surgido trabalhos na área da medicina personalizada e, de forma mais limitada, nas ciências sociais (Caron, Baio & Manolopoulou, 2022; Devriendt, Moldovan & Verbeke, 2018; Gutierrez & Gérardy, 2017; Hu, 2023; Olaya, Coussement & Verbeke, 2020; Rößler & Schoder, 2022; Zhang, Li & Liu, 2021). No entanto, a aplicação destes métodos à análise de problemas educativos ainda parece ser muito limitada.

Podemos destacar Olaya et al. (2020), que investigaram a Heterogeneidade do Efeito do Tratamento no abandono escolar de estudantes universitários, com base em terem ou não beneficiado da utilização de tutoriais de apoio no primeiro ano. Este estudo identificou que o método de Transformação de Resultados, com o algoritmo “Boosted Trees”, foi o mais eficaz, superando os métodos preditivos tradicionais.

Num estudo muito semelhante, Tanai & Ciftci (2023) analisaram o efeito de cursos preparatórios de Matemática e Inglês no abandono escolar, estimando uma redução entre 20% e 30% para os participantes desses cursos e identificando as características dos alunos que mais beneficiaram da intervenção. Nesta investigação, verificou-se que o método de Transformação de Resultados, com o algoritmo *Random Forests*, foi o mais eficaz.

Outra investigação foi desenvolvida por Deho et al. (2022), que compararam o desempenho de estudantes em aulas presenciais e online durante a pandemia de COVID-19. Os autores verificaram que o efeito médio foi inferior no desempenho das aulas online, e a análise heterogénea, desenvolvida com o algoritmo *Causal Tree*, indicou que o impacto negativo foi mais acentuado em alunos internacionais.

Por último, Leite et al. (2022) analisaram o efeito da utilização de sistemas de recomendação de vídeos no desenvolvimento de competências em álgebra. A análise da heterogeneidade, com o algoritmo *Causal Forest*, identificou que o sistema beneficiou especialmente os alunos com competências avançadas, de origem hispânica, beneficiários de apoios sociais e que participaram nas aulas de forma remota durante a pandemia de COVID-19.

Estes exemplos demonstram que os métodos de Aprendizagem Automática Causal

constituem uma ferramenta poderosa no âmbito da Ciência de Dados Educativos, permitindo analisar padrões causais complexos e identificar efeitos individuais ou em subgrupos específicos resultantes de ações pedagógicas, tanto no ensino presencial como no híbrido e a distância. Apesar deste potencial, a aplicação destas metodologias na área educativa ainda é incipiente e enfrenta desafios significativos, nomeadamente a disponibilidade e qualidade dos dados, em especial dados históricos sobre efeitos de intervenções anteriores, bem como questões relacionadas com a validação dos modelos e a capacidade técnica para aplicar estes métodos e interpretar corretamente os resultados.

5. Discussão

A integração da inferência causal nas ferramentas da Ciência de Dados Educativos representa um avanço significativo para a melhoria da qualidade na Educação a Distância e no e-Learning. Esta integração permite ir além do carácter meramente descritivo e preditivo, habitualmente associado às áreas de Educational Data Mining e à Learning Analytics, orientando essas abordagens para a compreensão dos fatores que efetivamente produzem impacto no desempenho e na aprendizagem.

Os métodos clássicos de inferência causal — como o *Propensity Score Matching*, a Regressão Descontínua, a Diferença-em-Diferenças e o Controlo Sintético — continuam a desempenhar um papel importante na investigação educativa, no âmbito das abordagens analisadas nesta revisão, oferecendo bases sólidas de validade estatística e rigor metodológico na identificação de efeitos causais. Estes métodos são particularmente úteis quando existem grupos de comparação bem definidos e dados estruturados sobre intervenções anteriores.

No entanto, a crescente complexidade dos contextos educativos e o volume de dados gerados em ambientes digitais requerem abordagens mais flexíveis e escaláveis. Assim, as abordagens de Aprendizagem Automática Causal apresentadas anteriormente, — incluindo métodos como a Transformação de Resultados, os Meta-learners e os algoritmos baseados em Árvores e Florestas Causais — ampliam a capacidade de análise em contextos caracterizados por grandes volumes de dados e múltiplas variáveis interdependentes, possibilitando uma análise das causas individuais ou condicionais e o desenvolvimento de ações pedagógicas mais personalizadas. Esta complementaridade é particularmente relevante em ambientes virtuais de aprendizagem, onde coexistem diversidade de perfis de estudantes, múltiplas estratégias pedagógicas e interações mediadas pela tecnologia.

Ao incorporar a lógica causal na Ciência de Dados Educativos, torna-se possível não apenas identificar correlações, mas compreender quais as ações, recursos ou condições que provocam mudanças efetivas no comportamento e nos resultados dos estudantes. Essa capacidade de estimar efeitos diferenciados (heterogéneos) constitui a base para uma personalização mais precisa da aprendizagem e para o desenvolvimento de sistemas de apoio à decisão docente baseados em evidência.

Estas metodologias podem também reforçar os processos de avaliação da qualidade em e-Learning, oferecendo indicadores causais de eficácia pedagógica que vão além das simples medidas de satisfação ou de sucesso global. Permitem compreender o que funciona, para quem e em que circunstâncias, apoiando uma gestão da qualidade mais adaptativa e contextualizada.

Ao privilegiar a explicação em detrimento da mera previsão, estas abordagens colocam a aprendizagem no centro do processo educativo mediado pela tecnologia, permitindo transformar a análise de dados em ações pedagógicas concretas — como ajustar estratégias de ensino, personalizar intervenções e melhorar a eficácia das práticas educativas. Dessa forma, promovem uma integração mais significativa entre pedagogia, dados e tecnologia.

6. Conclusão

Mais do que descrever e prever efeitos, a Ciência de Dados Educativos deve evoluir no sentido de se centrar na compreensão das causas. Esta transição, ainda em curso, representa um passo fundamental para consolidar esta área científica e, de forma pragmática, transformar os dados educativos em conhecimento sobre o impacto real das ações pedagógicas.

Apesar deste potencial, a aplicação da inferência causal na Ciência de Dados Educativos permanece reduzida, de acordo com os estudos analisados. A maioria da investigação e da prática continua fortemente focada em análises descritivas e preditivas, sobretudo na previsão do sucesso, insucesso ou desistência dos estudantes. O uso de metodologias quase-experimentais — embora amplamente reconhecidas, pelo menos nas suas formas mais clássicas, como o *Propensity Score Matching*, a Regressão Descontínua e a Diferença-em-Diferenças — continua a ser pouco explorado nestas áreas. Mesmo o método de Controlo Sintético, quer na sua versão original, quer nas variantes mais recentes, apresenta uma presença muito limitada nos estudos analisados.

Já as abordagens de Aprendizagem Automática Causal, desenvolvidas precisamente para lidar com grandes volumes de dados e múltiplos atributos — típicos de contextos de big data —, e que podem beneficiar da informação acumulada nos ambientes virtuais de aprendizagem para promover uma personalização mais eficaz baseada na causalidade, são pouco representadas na investigação analisada, sendo escassas as evidências da sua aplicação em contextos educativos. A revisão realizada confirma uma presença residual destas metodologias, limitada a poucos estudos isolados e, em grande parte, de carácter exploratório.

Tendo em consideração a análise realizada, conclui-se que para que esta transformação se concretize, torna-se pertinente investir na formação de profissionais e investigadores que atuam nas áreas de Educational Data Mining e Learning Analytics, bem como de analistas responsáveis pelo tratamento dos múltiplos dados gerados nos ambientes virtuais de aprendizagem. Revela-se igualmente importante articular este conhecimento técnico com uma compreensão pedagógica sólida, promovendo uma integração efetiva entre análise de dados e prática educativa. O domínio das metodologias clássicas de inferência causal e das mais recentes técnicas de Aprendizagem Automática Causal surge como uma competência relevante para que a Ciência de Dados Educativos — ou as suas designações clássicas — se afirme como um campo científico mais robusto e alinhado com as necessidades do ensino a distância.

A literacia causal assume igualmente um papel relevante para melhorar a qualidade da recolha e da gestão dos dados, permitindo que estes sejam concebidos e armazenados desde a origem segundo uma perspetiva orientada para variáveis de tratamento, covariáveis e efeitos esperados. A adoção desta perspetiva poderá permitir às

organizações educativas criar bases de dados mais versáteis e preparadas para análises causais, favorecendo avaliações mais rigorosas e a personalização das intervenções pedagógicas.

Importa, contudo, reconhecer algumas limitações desta revisão. Para além da escassez de estudos empíricos sobre estas abordagens em contextos educativos, verificou-se também uma falta de uniformização conceptual e terminológica entre diferentes correntes de investigação, nomeadamente entre as abordagens de *Uplift Modeling* e de *Heterogeneous Treatment Effects (HTE)*. Esta diversidade de designações e enquadramentos teóricos dificultou a sistematização dos conceitos e das abordagens analisadas, podendo igualmente ter limitado a identificação de todos os estudos relevantes, dada a utilização de diferentes terminologias para descrever conceitos semelhantes. Adicionalmente, a natureza predominantemente conceptual ou exploratória de parte da literatura analisada reflete o estado ainda emergente destas abordagens no contexto educativo.

Referências

- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1), 113–132. <https://doi.org/10.1257/000282803321455188>
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505. <https://doi.org/10.1198/jasa.2009.ap08746>
- Acharki, N., Lugo, R., Bertonecello, A., & Garnier, J. (2023). Comparison of meta-learners for estimating multi-valued treatment heterogeneous effects. In *Proceedings of the International Conference on Machine Learning* (pp. 91–132). PMLR. <https://proceedings.mlr.press/v202/acharki23a.html>
- Athey, S. & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360. <https://doi.org/10.1073/pnas.1510489113>
- Athey, S. & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- Athey, S., Imbens, G. W., & Wager, S. (2018). Approximate residual balancing: Debaised inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4), 597–623. <https://doi.org/10.1111/rssb.12268>
- Birdsall, C. (2017). The synthetic control method for comparative case studies: An application estimating the effect of managerial discretion under performance management. *International Public Management Journal*, 20(1), 25–53. <https://doi.org/10.1080/10967494.2015.1121178>
- Brand, J. E., Zhou, X., & Xie, Y. (2023). Recent developments in causal inference and

- machine learning. *Annual Review of Sociology*, 49, 81–110.
<https://doi.org/10.1146/annurev-soc-030420-015345>
- Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1), 247–274. <https://doi.org/10.1214/14-AOAS788>
- Caniglia, E. C., & Murray, E. J. (2020). Difference-in-difference in the time of cholera: A gentle introduction for epidemiologists. *Current Epidemiology Reports*, 7, 203–211. <https://doi.org/10.1007/s40471-020-00245-2>
- Caron, A., Baio, G., & Manolopoulou, I. (2022). Estimating individual treatment effects using non-parametric regression models: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3), 1115–1149. <https://doi.org/10.1111/rssa.12824>
- Cook, T. D. (2008). Waiting for life to arrive: A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2), 636–654. <https://doi.org/10.1016/j.jeconom.2007.05.002>
- Cortes, K. E., Moussa, W. S., & Weinstein, J. M. (2013). Educating bright students in urban schools. *Economics of Education Review*, 37, 286–297. <https://doi.org/10.1016/j.econedurev.2013.08.001>
- Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press. <https://mixtape.scunning.com/>
- Deho, O. B., Liu, L., Joksimovic, S., Li, J., Zhan, C., & Liu, J. (2022). Assessing the causal impact of online instruction due to COVID-19 on students' grades and its aftermath on grade prediction models. *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, 32–38. <https://doi.org/10.1145/3524458.3547270>
- Devriendt, F., Moldovan, D., & Verbeke, W. (2018). A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics. *Big Data*, 6(1), 13–41. <https://doi.org/10.1089/big.2017.0104>
- Fan, X., & Nowell, D. L. (2011). Using propensity score matching in educational research. *Gifted Child Quarterly*, 55(1), 74–79. <https://doi.org/10.1177/0016986210390635>
- Foo, C. C., Cheung, B., & Chu, K. M. (2021). A comparative study regarding distance learning and the conventional face-to-face approach during the COVID-19 pandemic. *BMC Medical Education*, 21(141). <https://doi.org/10.1186/s12909-021-02575-1>
- Gama, J., Lorena, A., Faceli, K., Oliveira, M., & Carvalho, A. (2017). *Extração de conhecimento de dados: Data mining* (3.^a ed.). Edições Sílabo.
- Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, 44(1), 218–243.

<https://doi.org/10.3102/0091732X20903302>

- Gutierrez, P., & Gérardy, J. Y. (2017). Causal inference and uplift modelling: A review of the literature. In *Proceedings of the 3rd International Conference on Predictive Applications and APIs* (pp. 1–13)..
- Heinesen, E. (2018). Admission to higher education programmes and student outcomes: Evidence from Denmark. *Economics of Education Review*, 63, 1–19. <https://doi.org/10.1016/j.econedurev.2018.01.002>
- Hu, A. (2023). Heterogeneous treatment effects analysis for social scientists: A review. *Social Science Research*, 109, 102810. <https://doi.org/10.1016/j.ssresearch.2022.102810>
- Huntington-Klein, N. (2022). *The effect: An introduction to research design and causality*. Chapman and Hall/CRC. <https://theeffectbook.net/>
- Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3), 706–710. <https://doi.org/10.1093/biomet/87.3.706>
- Imbens, G. W. (2022). Causality in econometrics: Choice vs chance. *Econometrica*, 90(6), 2541–2566. <https://doi.org/10.3982/ECTA21204>
- Jaskowski, M., & Jaroszewicz, S. (2012). Uplift modeling for clinical trial data. *Proceedings of the ICML Workshop on Machine Learning for Clinical Data Analysis*, 1–6.
- Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10), 4156–4165. <https://doi.org/10.1073/pnas.1804597116>
- Lagmay, E. A., & Rodrigo, M. M. T. (2021). Quantifying the impact of severe weather conditions on online learning during the COVID-19 pandemic. In *Artificial Intelligence in Education: 22nd International Conference, AIED 2021* (pp. 229–233). Springer. https://doi.org/10.1007/978-3-030-78292-4_42
- Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, 4(3), 165–224. <https://doi.org/10.1561/08000000014>
- Lechner, M. (2023). Causal machine learning and its use for public policy. *Swiss Journal of Economics and Statistics*, 159(1), 8. <https://doi.org/10.1186/s41937-023-00113-y>
- Leite, W. L., Kuang, H., Shen, Z., Chakraborty, N., Michailidis, G., D’Mello, S., & Xing, W. (2022). Heterogeneity of treatment effects of a video recommendation system for algebra. *Proceedings of the Ninth ACM Conference on Learning at Scale*, 12–23. <https://doi.org/10.1145/3491140.3528250>
- Lopes, N. C., & Cavique, L. (2023). Causal machine learning in social impact assessment. In *Philosophy of artificial intelligence and its place in society* (pp. 56–77). IGI Global Scientific Publishing

- López-Torres, L., Prior, D., & Santín, D. (2019). Assessing the effect of educational programs on public schools' performance. *Applied Economics*, 51(48), 5205–5226. <https://doi.org/10.1080/00036846.2019.1610712>
- Olaya, D., Coussement, K., & Verbeke, W. (2020). A survey and benchmarking study of multitreatment uplift modeling. *Data Mining and Knowledge Discovery*, 34, 273–308. <https://doi.org/10.1007/s10618-019-00670-y>
- Olaya, D., Vásquez, J., Maldonado, S., Miranda, J., & Verbeke, W. (2020). Uplift modeling for preventing student dropout in higher education. *Decision Support Systems*, 134, 113320. <https://doi.org/10.1016/j.dss.2020.113320>
- Pearl, J., & Mackenzie, D. (2022). *O livro do porquê: A nova ciência da causa e do efeito*. Temas e Debates.
- Pinheiro, P., & Cavique, L. (2022). Uplift modeling using the transformed outcome approach. In G. Marreiros, B. Martins, A. Paiva, B. Ribeiro, & A. Sardinha (Eds.), *Progress in artificial intelligence* (Vol. 13566, pp. 632–643). Springer. https://doi.org/10.1007/978-3-031-16474-3_51
- Rößler, J., & Schoder, D. (2022). Bridging the gap: A systematic benchmarking of uplift modeling and heterogeneous treatment effects methods. *Journal of Interactive Marketing*, 57(4), 629–650. <https://doi.org/10.1177/10949968221111083>
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701. <https://doi.org/10.1037/h0037350>
- Tanai, Y., & Ciftci, K. (2023). How to customize an early start preparatory course policy to improve student graduation success: An application of uplift modeling. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05607-9>
- Thistlethwaite, D. L., & Campbell, D. T. (2017). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Observational Studies*, 3(2), 119–128. <https://doi.org/10.1353/obs.2017.0011>
- Tu, C. (2019). Comparison of various machine learning algorithms for estimating generalized propensity scores. *Journal of Statistical Computation and Simulation*, 89(4), 708–719. <https://doi.org/10.1080/00949655.2019.1571059>
- Vergolini, L., & Zanini, N. (2015). Away, but not too far from home: The effects of financial aid on university enrolment decisions. *Economics of Education Review*, 49, 91–109. <https://doi.org/10.1016/j.econedurev.2015.08.003>
- Wager, S. & Athey, S. (2018). Estimation and inference of heterogeneous treatment

effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>

Wang, X., Wen, M. & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 398–407. <https://doi.org/10.1145/2883851.2883964>

Zhang, W., Li, J., & Liu, L. (2021). A unified survey of treatment effect heterogeneity modelling and uplift modelling. *ACM Computing Surveys*, 54(8), Article 162. <https://doi.org/10.1145/3466818>

Recebido 07/11/2025

Aceite 05/05/2026

Publicado 12/05/2026

Este artigo está disponível segundo uma licença [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)
