

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

**Explorando Relações de Causalidade: Modelos Gráficos
Causais e o Framework DoWhy**

André Costa de Andrade

RELATÓRIO DE PROJETO FINAL II

CENTRO TÉCNICO CIENTÍFICO - CTC

DEPARTAMENTO DE INFORMÁTICA

Curso de Graduação em Ciência da Computação

Rio de Janeiro, junho de 2024



André Costa de Andrade

**Explorando relações de causalidade por meio da
convergência entre Aprendizado de Máquina e Inferência
Causal**

Relatório de Projeto Final II, apresentado ao programa Ciência da Computação da PUC-Rio como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Marcos Vianna Villas

Departamento de Informática - PUC-Rio

Rio de Janeiro, junho de 2024

Resumo

André Costa de Andrade; Marcos Vianna Villas; Explorando Relações de Causalidade: Modelos Gráficos Causais e o Framework DoWhy. Rio de Janeiro, 2024. 78. Relatório de Projeto Final de Graduação - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro

O objetivo foi explorar e aplicar técnicas de inferência causal e machine learning para entender melhor as relações entre variáveis em diferentes contextos. O método utilizado envolveu a construção de um grafo causal, o uso da ferramenta DoWhy para realização de inferência causal junto de modelos de machine learning e validação dos resultados. Inferência causal é a metodologia empregada para determinar relações de causa e efeito entre variáveis, diferentemente de simples associações estatísticas. Machine learning é uma área da inteligência artificial focada em treinar algoritmos para fazer previsões ou identificar padrões a partir de dados. DoWhy é uma ferramenta de código aberto que facilita a implementação de inferência causal, fornecendo um framework para definir, estimar e testar hipóteses causais. Os casos estudados foram Felicidade e Carros, onde analisamos os determinantes da felicidade e os fatores que influenciam o desempenho e a eficiência de veículos automotores. As conclusões foram que a aplicação combinada de inferência causal e machine learning, apoiada pelo DoWhy, oferece uma abordagem robusta para descobrir e validar relações causais em diversas áreas de estudo, fornecendo insights valiosos para decisões baseadas em dados.

Palavras-chave: Inferência Causal, Aprendizado de Máquina, DoWhy, Grafo Causal, Resultados Potenciais.

Abstract

André Costa de Andrade; Marcos Vianna Villas; Exploring Causal Relationships: Graphical Causal Models and the DoWhy Framework. Rio de Janeiro, 2024. 78. Final Undergraduate Project Report - Department of Informatics, Pontifical Catholic University of Rio de Janeiro

The objective was to explore and apply causal inference and machine learning techniques to better understand the relationships between variables in different contexts. The method used involved constructing a causal graph, utilizing the DoWhy tool for performing causal inference alongside machine learning models, and validating the results. Causal inference is the methodology employed to determine cause-and-effect relationships between variables, unlike simple statistical associations. Machine learning is a field of artificial intelligence focused on training algorithms to make predictions or identify patterns from data. DoWhy is an open-source tool that facilitates the implementation of causal inference, providing a framework to define, estimate, and test causal hypotheses. The case studies were Happiness and Cars, where we analyzed the determinants of happiness and the factors influencing the performance and efficiency of motor vehicles. The conclusions were that the combined application of causal inference and machine learning, supported by DoWhy, offers a robust approach to discovering and validating causal relationships in various fields of study, providing valuable insights for data-driven decisions.

Keywords: Causal Inference, Machine Learning, DoWhy, Causal Graph, Potential Outcomes.

Conteúdo

1	Motivação	2
2	Situação Atual	2
3	Objetivos	4
4	Plano de Ação	4
	4.1 Etapas	4
	4.2 Cronograma PF I	6
	4.3 Cronograma PF II	7
5	Inferência Causal	7
	5.1 Correlação não significa Causalidade	8
	5.2 Métodos clássicos para inferência causal	8
	5.3 Modelagem com grafos	11
	5.4 Modelo de <i>Potential Outcomes</i>	15
	5.5 Métricas	18
6	Aprendizado de Máquina	19
	6.1 Aprendizado Supervisionado	20
	6.2 Aprendizado Não-Supervisionado	20
	6.3 Algoritmos e métodos clássicos utilizados em Aprendi- zado de Máquina	21
	6.4 Aprendizado Profundo	24
7	Aprendizado de Máquina para Inferência Causal	25
	7.1 A Integração das Duas Áreas	25
	7.2 Métodos/Algoritmos	25
	7.3 Ferramentas	27
8	DoWhy	28
	8.1 Entendendo as capacidades da biblioteca	28
	8.2 Modelagem das Relações Causais	33
9	Primeiro Estudo de Caso - Felicidade	38
	9.1 Pré-processamento	40
	9.2 Criação do Grafo Causal	40
	9.3 Primeira Causal Query	42
	9.4 Segunda Causal Query	45
	9.5 Terceira Causal Query	51
	9.6 Quarta Causal Query	53
10	Segundo Estudo de Caso - Carros	54
	10.1 Pré-processamento	55
	10.2 Criação do Grafo Causal	56
	10.3 Primeira Causal Query	57
	10.4 Segunda Causal Query	61
	10.5 Terceira Causal Query	66
	10.6 Quarta Causal Query	67
11	Considerações Finais	69
	11.1 Comparação com o objetivo inicial	69
	11.2 Contribuições deste trabalho	70
	11.3 Dificuldades	70
	11.4 Oportunidades para trabalhos futuros	71

1 Motivação

Nos últimos tempos, houve grandes avanços na área de ciência de dados no que diz respeito a modelos de previsão de dados, devido a avanços em aprendizado de máquina. No entanto, é percebido que certos problemas atuais não podem apenas ser resolvidos apenas com predição. É preciso entender as relações de causalidade entre as variáveis presentes nos dados, pois o simples ato de prever resultados não proporciona uma compreensão profunda das dinâmicas subjacentes.

A busca por causalidade, também chamada de inferência causal, torna-se imperativa para extrair *insights* significativos e embasar decisões mais fundamentadas. Ao considerar o contexto mais amplo, a análise de causalidade vai além da mera identificação de correlações estatísticas. Envolve investigar as relações de causa e efeito que governam os fenômenos observados. Compreender tais relações é crucial para desenvolver intervenções eficazes, otimizar processos e antecipar possíveis impactos de mudanças nas variáveis de interesse.

A convergência entre aprendizado de máquina e inferência causal oferece uma oportunidade para o desenvolvimento de novos métodos para identificar e quantificar relações de causalidade. Esses métodos podem ser usados para resolver uma ampla gama de problemas, como por exemplo avaliação de políticas públicas, o desenvolvimento de novos tratamentos médicos e a prevenção de crimes. No entanto, a inferência de causalidade é um problema desafiador, pois é difícil controlar todos os fatores que podem influenciar uma relação de causalidade.

2 Situação Atual

Cao (2016) define a ciência de dados como um campo interdisciplinar que utiliza métodos e técnicas científicas para extrair *insights* e conhecimento de grandes volumes de dados. Além disso, a ciência de dados envolve a análise de dados de diferentes fontes, como bancos de dados,

redes sociais e sensores. Isso permite que as organizações tomem decisões mais informadas e baseadas em evidências.

Por sua vez, Mitchell (1997) define aprendizado de máquina como o campo de estudo que proporciona aos computadores a capacidade de aprender por meio da experiência, ou seja, o consumo constante de novos dados. O aprendizado de máquina é amplamente utilizado em áreas como reconhecimento de padrões, processamento de linguagem natural e análise preditiva.

A união desses dois campos, ciência de dados e aprendizado de máquina, tem impulsionado avanços significativos em diversas áreas, como medicina, finanças, marketing e transporte. Com a análise de dados e o aprendizado de máquina, é possível identificar padrões e tendências ocultas nos dados, bem como prever eventos futuros. No entanto, é importante ressaltar que apenas tentar prever eventos futuros não é suficiente. É necessário entender profundamente as relações de causalidade entre as variáveis para uma compreensão maior dos fenômenos, pois isso pode levar a intervenções mais efetivas.

Aristóteles, em sua obra "Metafísica" (Aristóteles, 2002), define causa como "aquilo sem o qual o efeito não pode existir", e o efeito, por sua vez, é resultado que ocorre devido à influência da causa. A Inferência causal é uma abordagem analítica que vai além das técnicas estatísticas convencionais. Enquanto a análise estatística padrão se concentra em avaliar parâmetros de uma distribuição com base em suas amostras, a inferência causal busca não apenas entender a probabilidade de eventos em condições estáticas, mas também compreender a dinâmica desses eventos sob condições em constante mudança.

Em outras palavras, a análise causal visa não apenas a inferir associações entre variáveis e estimar a probabilidade de eventos passados e futuros, mas também a compreender como os eventos evoluem em resposta a mudanças nas condições experimentais, como aquelas induzidas por tratamentos ou intervenções externas. Essa abordagem permite não apenas

avaliar a probabilidade de eventos sob condições específicas, mas também entender como essas probabilidades podem ser influenciadas por fatores externos e dinâmicas em evolução (Pearl, 2010).

3 Objetivos

O trabalho tem dois objetivos: explorar como o aprendizado de máquina pode ser utilizado na realização de inferência causal, abordando os métodos e algoritmos mais utilizados para tal. Após isso, serão escolhidos conjuntos de dados reais em que seja possível realizar análises causais usando o conteúdo estudado.

4 Plano de Ação

4.1 Etapas

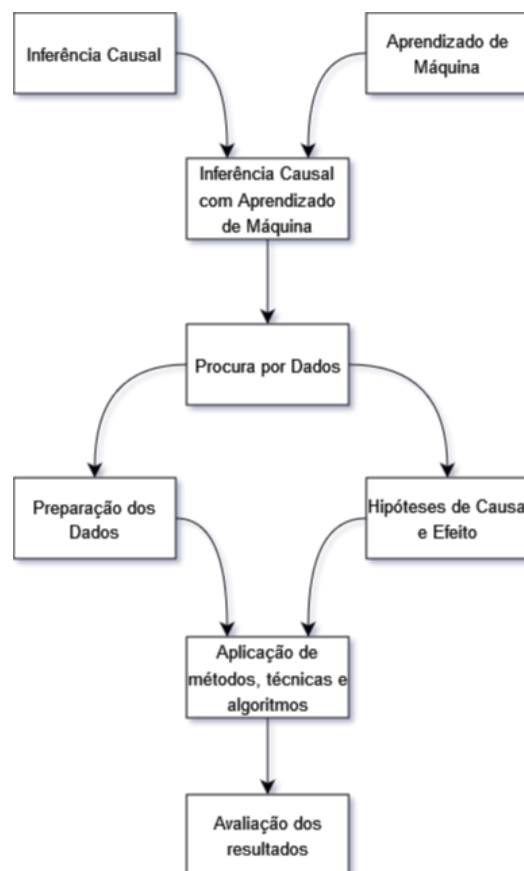


Figura 1: Plano de Ação

Estudo dos princípios da inferência causal. Isso inclui compreender os conceitos-chave, como causalidade, identificação de efeitos causais, variáveis de confusão e métodos estatísticos como experimentos controlados e modelos causais. Além disso, a pesquisa abordagens formais, como os grafos causais, para representar e analisar relações causais entre variáveis e o modelo de Resultados Potenciais, para calcular efeitos de tratamentos. O conteúdo estudado foi detalhado na seção 5.

Estudo dos princípios do aprendizado de máquina, que incluem algoritmos de classificação, regressão, clustering, redes neurais, árvores de decisão, entre outros. Também será abordado brevemente aprendizado profundo. Na seção 6 é discutido isso.

Estudo de como o aprendizado de máquina pode ser aplicados à inferência causal, além dos métodos e algoritmos especializados existentes. Estudo realizado na seção 8.

Após isso, serão selecionadas duas áreas ou domínios específicos onde os métodos e técnicas de inferência causal e aprendizado de máquina serão aplicados. Por exemplo, essas áreas podem ser a saúde pública e o marketing digital. A escolha dos domínios deve ser baseada na relevância e na disponibilidade de dados adequados para análise.

Na próxima etapa, serão pesquisadas na literatura hipóteses de causalidade específicas para cada uma das áreas selecionadas. Apenas exemplificando, em saúde pública, uma hipótese pode ser que a exposição a determinado poluente do ar cause um aumento na incidência de doenças respiratórias. Essas hipóteses servirão como base para a análise causal posterior.

Partindo para a parte prática, será feita a pesquisa, limpeza e preparação dos dados relevantes para testar as hipóteses de causalidade. Isso pode incluir a seleção de variáveis de interesse, o tratamento de dados ausentes e a normalização de dados. É fundamental garantir que os dados estejam prontos para serem usados em modelos de aprendizado de máquina e análises de causalidade.

Na penúltima etapa, os métodos e técnicas de inferência causal e aprendizado de máquina serão aplicados aos dados preparados. Isso pode incluir a construção de modelos de regressão, árvores de decisão, modelos de redes neurais ou a implementação de métodos de identificação de efeitos causais. A análise é realizada de acordo com as hipóteses formuladas anteriormente.

Na última etapa, os resultados obtidos são avaliados quanto à sua validade e efetividade na testagem das hipóteses de causalidade. Isso pode envolver a análise estatística dos resultados, a interpretação das relações causais identificadas e a discussão das implicações práticas. Também é importante considerar a robustez dos métodos utilizados e a possibilidade de viés nos resultados.

4.2 Cronograma PF I

O cronograma da figura 2 se refere tanto ao previsto, quanto ao que foi realmente realizado.

Atividades	Outubro			Novembro			Dezembro		
Estudo e pesquisa dos conceitos e fundamentos de Inferência Causal									
Estudo e pesquisa dos conceitos e fundamentos de Aprendizado de Máquina									
Estudo e pesquisa dos conceitos e fundamentos de Inferência Causal com Aprendizado de Máquina									

Figura 2: Cronograma de Outubro a Dezembro/2023

4.3 Cronograma PF II

Atividades	Março			Abril			Maio			Junho			Julho		
Definição de duas áreas onde serão aplicados os métodos e técnicas encontrados															
Identificação de hipóteses de causalidade para as áreas escolhidas															
Identificação e preparação dos dados para análises															
Aplicação dos métodos nos dados de interesse															
Avaliação dos resultados e efetividade das técnicas utilizadas															

Figura 3: Cronograma de Março a Julho/2024

5 Inferência Causal

Nessa seção é abordado um equívoco bem comum no estudo de causalidade, a armadilha da correlação. Em seguida, expus dois métodos clássicos utilizados por estatísticos para identificar causalidade em dados. Final-

mente, são abordados dois modelos amplamente utilizados para os estudos de problemas causais.

5.1 Correlação não significa Causalidade

Um dos princípios mais cruciais a serem lembrados ao analisar dados estatísticos é que correlação não implica causalidade (Huff, 1954). Correlação é uma relação existente entre fenômenos ou coisas, ou entre variáveis matemáticas ou estatísticas, que tendem a variar, estar associadas ou ocorrer juntas de uma maneira que não é esperada apenas com base no acaso (Merriam-Webster, 2023b), enquanto que causalidade é a relação entre eventos ou fenômenos regularmente correlacionados (Merriam-Webster, 2023a).

Isso quer dizer que mesmo que seja observado que um evento A e um evento B variem de forma semelhante, nada indica que um é causa do outro. Por exemplo, digamos que toda vez que o sol nasce um galo canta. Se coletarmos dados diariamente desse acontecimento, veremos que esses eventos sempre acompanham um ao outro, tendo então alguma correlação. No entanto, podemos dizer que o cantar do galo fez o sol nascer? Ou que o sol nascer, provocou o cantar do galo? Não é possível saber até ser feita uma análise minuciosa das variáveis que circundam todo esse sistema.

5.2 Métodos clássicos para inferência causal

Redes Bayesianas

Um método clássico que tenta lidar com causalidade são as Redes Bayesianas, as quais se caracterizam por serem modelos probabilísticos baseadas em grafos direcionados acíclicos (DAGs). Os nós de uma Rede Bayesiana representam variáveis de interesse e as conexões representam dependências informacionais ou causais entre as variáveis (Pearl, 2011). O grande objetivo das Redes Bayesianas é representar e resolver problemas que envolvem incerteza e dependência entre variáveis.

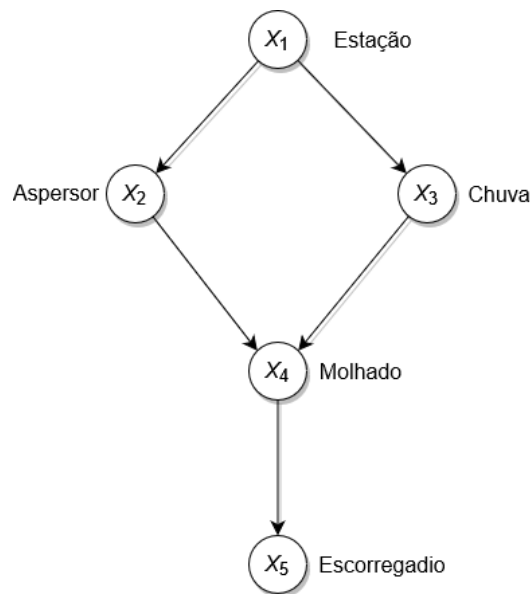


Figura 4: Uma Rede Bayesiana representando as influências causais entre 5 variáveis
 Fonte - Pearl (2011)

A figura 4 ilustra uma Rede Bayesiana simples, porém comum. Ela representa as relações causais entre a estação do ano (X_1), se está chovendo (X_3), se o aspersor está ligado (X_2), se o pavimento está molhado (X_4) e se o pavimento está escorregadio (X_5). Nesse exemplo, a ausência de uma conexão direta entre X_1 e X_5 , por exemplo, mostra que não existe uma influência direta entre a estação do ano e o quão escorregadio está o pavimento - a influência é mediada pelo nível de umidade do pavimento. A partir dela, algumas conclusões podem ser tiradas. Por exemplo, se o aspersor está ligado, então provavelmente o pavimento está molhado. De outra forma, se verificarmos que o pavimento está molhado, provavelmente isso quer dizer que o aspersor está ligado ou está chovendo.

Modelagem de equações estruturais

A Modelagem de Equações Estruturais (SEM) é uma técnica de análise de dados que envolve múltiplas variáveis. É um dos principais métodos estatísticos para testar, empiricamente, hipóteses sobre as relações entre variáveis mensuráveis e variáveis latentes (Hair Jr et al., 2021). Variáveis latentes são variáveis que só podem ser compreendidas conceitualmente e não podem ser mensuradas diretamente. Um exemplo disso é o conceito

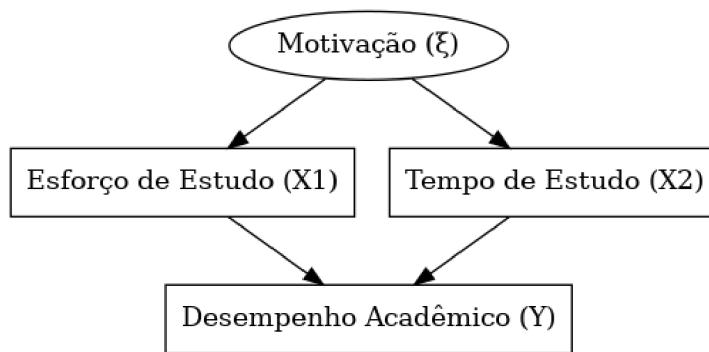


Figura 5: Exemplo de equação estrutural com 3 três variáveis mensuráveis e uma variável latente - Fonte: O próprio autor

de inteligência.

Nesse método, são utilizados diagramas chamados Modelos de Caminho para representar visualmente hipóteses e relações entre variáveis, que serão posteriormente examinadas quando a modelagem for aplicada.

A figura 5 ilustra um modelo de equações estruturais simples, porém comum. Ele representa as relações entre três variáveis mensuráveis: esforço de estudo (X_1), tempo de estudo (X_2), e desempenho acadêmico (Y), além de uma variável latente: motivação (ξ). Nesse exemplo, a variável latente motivação (ξ) influencia tanto o esforço de estudo (X_1) quanto o tempo de estudo (X_2), que por sua vez, influenciam o desempenho acadêmico (Y). A ausência de uma conexão direta entre motivação (ξ) e desempenho acadêmico (Y) indica que a influência de motivação no desempenho acadêmico é mediada pelo esforço e pelo tempo de estudo.

A partir desse modelo, algumas conclusões podem ser tiradas. Por exemplo, se a motivação de um estudante é alta, é provável que ele dedique mais esforço e tempo ao estudo, o que, conseqüentemente, resultará em um melhor desempenho acadêmico. Do mesmo modo, se verificarmos que o desempenho acadêmico de um estudante é alto, isso pode indicar que ele dedicou mais esforço e tempo ao estudo, o que pode ser influenciado por um alto nível de motivação.

5.3 Modelagem com grafos

Pearl (2009) propõe uma abordagem para modelagem de problemas causais por meio de Grafos Direcionados Acíclicos (DAGs), da mesma forma que fez nas Redes Bayesianas. Essa abordagem fornece uma representação visual e intuitiva das relações causais entre variáveis, facilitando a análise e interpretação das implicações causais em um sistema.

Nós e Arestas:

Cada nó do grafo representa uma variável no sistema que está sendo estudado. Essas variáveis podem ser observáveis ou latentes. As arestas direcionadas entre os nós representam relações causais. Se há uma seta de A para B, isso sugere que A causa B.

Do-Calculus:

O do-calculus (Pearl, 2012) é um framework matemático desenvolvido para permitir o raciocínio formal sobre relações causais e fazer inferências a partir de modelos causais. Nesse contexto, ele fornece um conjunto de regras e operações algébricas para manipular e derivar expressões envolvendo a distribuição probabilística de variáveis em um modelo causal. Através da aplicação das regras do do-calculus nos DAGs, é possível derivar fórmulas que permitem calcular distribuições causais e efeitos causais a partir de distribuições observacionais ou experimentais. Nesse framework, as perguntas podem ser formuladas com *do-expressions* do tipo $Q = P(y|do(x), z)$, onde se lê "A probabilidade de y dado que fazemos a intervenção $do(x)$ e dado o valor de z ".

No entanto, quando pesquisadores estão trabalhando com dados observacionais, é necessário utilizar alguns critérios de identificação de conjuntos de variáveis que, quando condicionados, permitem o cálculo do efeito causal de interesse. Considere um estudo observacional em que é desejado calcular o efeito de X em Y .

- Critério Backdoor: Um conjunto (S) é suficiente para estimar o efeito causal se duas condições forem satisfeitas:

1. Nenhum elemento do conjunto S é descendente de X .
2. Os elementos do conjunto S "bloqueiam" todos os caminhos de "backdoor" de X para Y , ou seja, todos os caminhos que terminam com uma aresta apontando para X .

Nesse contexto, "bloquear" se refere ao conceito de d-separação, onde um caminho c é bloqueado por um conjunto de nós S caso c contenha pelo menos um nó com aresta saindo que esteja em S , ou c contenha pelo menos um nó com colisão (Niu et al., 2023) que não esteja em S e não tenha nenhum descendente em S . Caso S bloqueie todos os caminhos de X para Y , pode-se afirmar que ele "d-separa X e Y ".

Caso o conjunto S satisfaça as condições do critério de backdoor relativo a X e Y então $P(y \mid \text{do}(x)) = \sum_s P(y \mid x, s)P(s)$. A expressão indica que, se conseguirmos identificar e controlar adequadamente o conjunto S que satisfaz o critério, então podemos estimar o efeito causal de X sobre Y através de observações de Y condicionadas a X e S , e usando a distribuição marginal de S .

Na figura 6 é possível ver um exemplo de grafo causal onde se aplica o critério backdoor.

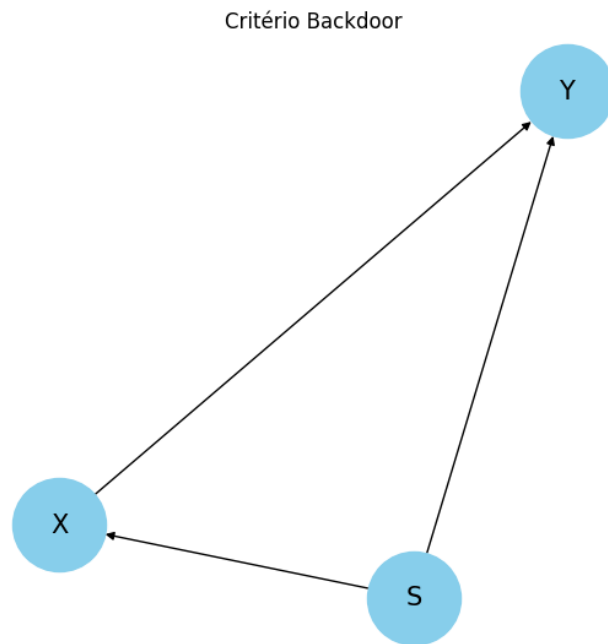


Figura 6: Para calcular o efeito de X em Y nesse caso, é necessário condicionar pelo nó S - Fonte: O próprio autor

- Critério Frontdoor: Uma variável Z satisfaz esse critério caso três condições sejam cumpridas:
 1. A variável Z intercepta todos os caminhos de X para Y .
 2. Não existe caminho de backdoor de X para Z .
 3. X bloqueia todos os caminhos de backdoor de Z para Y .

Caso essas condições seja atingidas, então $P(y \mid do(x)) = \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z) P(x')$. X' representa uma nova instanciiação ou um novo valor potencial de X . Assim, expressão representa a probabilidade de Y após a intervenção $do(x)$, calculada como a soma sobre Z , que é afetada por X , e então somando sobre todos os possíveis valores de X' , ponderada pela probabilidade de Y dado X' e Z , e a probabilidade de X' dado X . A variável X' é usada para abstrair os possíveis valores que X pode assumir quando consideramos o efeito de X sobre Y através do mediador Z .

Na figura 7 é possível ver um exemplo de grafo causal onde se aplica

o critério frontdoor.

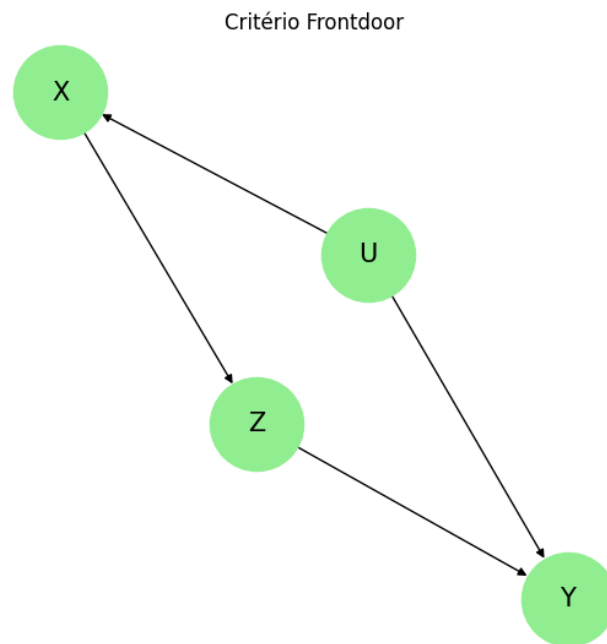


Figura 7: Para calcular o efeito de X em Y nesse caso, é necessário condicionar pelo nó Z - Fonte: O próprio autor

- Instrumental Variables (IV): Variáveis instrumentais (Greenland, 2000) são aquelas que afetam o tratamento X , no entanto, não afetam diretamente o resultado Y , exceto por sua relação com X . Para que uma variável Z seja considerada um instrumento válido, ela deve satisfazer duas condições principais:

1. Relevância: Z deve estar correlacionada com o tratamento X . Isso significa que Z deve fornecer variação suficiente em X para identificar o efeito causal.
2. Exogeneidade: Z deve ser independente dos fatores não observados que afetam Y . Em outras palavras, Z deve influenciar Y apenas através de sua relação com X , e não diretamente ou através de outras variáveis omitidas.

Essas condições garantem que Z pode ser usada para separar a variação em X que é exógena (não correlacionada com os fatores não

observados) da variação que é endógena. Utilizando Z como instrumento, podemos obter uma estimativa não viesada do efeito causal de X sobre Y .

A aplicação das variáveis instrumentais envolve geralmente dois estágios:

1. No primeiro estágio, X é regredido sobre Z (e possivelmente outras covariáveis) para obter os valores ajustados de X , que são chamados de \hat{X} .
2. No segundo estágio, Y é regredido sobre \hat{X} para estimar o efeito causal de X sobre Y .

Na figura 8 é possível ver um exemplo de grafo causal onde se aplica o critério IV.

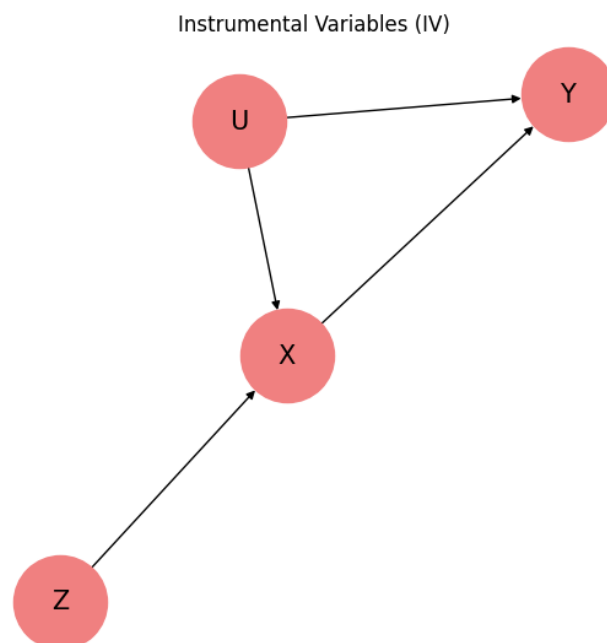


Figura 8: Para calcular o efeito de X em Y nesse caso, é necessário condicionar pelo nó Z - Fonte: O próprio autor

5.4 Modelo de *Potential Outcomes*

O Modelo de *Potential Outcomes* (Rubin, 2005), também conhecido como Modelo Causal de Rubin ou Modelo Causal de Neyman-Rubin, é uma es-

estrutura teórica para compreender e formalizar a causalidade em estudos observacionais. Foi desenvolvido pelos estatísticos Donald Rubin e Jerzy Neyman, e fornece uma maneira de expressar e analisar relações causais em situações em que experimentos randomizados controlados não são viáveis ou éticos.

A estrutura é especialmente relevante no contexto de estudos observacionais, onde o interesse geralmente é entender o efeito causal de um tratamento ou intervenção em uma variável de resultado. O conceito-chave no Modelo de Resultados Potenciais é a noção de resultados potenciais. Por exemplo, em um estudo sobre os efeitos de um medicamento para perda de peso, o resultado potencial para um indivíduo sob o tratamento seria o peso que ele perderia se tomasse o medicamento, enquanto o resultado potencial sob o controle seria o peso que ele teria se não tomasse o medicamento. No entanto, para cada indivíduo, apenas um dos resultados potenciais é observado na realidade, dependendo se ele recebeu ou não o tratamento. O outro resultado potencial permanece como um contrafactual não observado.

Componentes básicos da estrutura:

- Estrutura Causal: O mecanismo de atribuição de tratamento determina quais unidades recebem o tratamento e quais não, resultando nos respectivos resultados potenciais para cada unidade. Essa relação fundamental entre o mecanismo de atribuição e os resultados potenciais é a base da estrutura causal do modelo.
- Efeito Causal: O efeito causal de uma intervenção, que é o parâmetro de interesse primário a ser estimado, é a diferença média entre os resultados potenciais do grupo de tratamento e do grupo de controle. Obter uma estimativa não viesada desse efeito causal é o objetivo principal do Modelo de Resultados Potenciais.
- Suposições identificadoras : O modelo faz algumas suposições impor-

tantes, chamadas suposições identificadoras, que permitem a identificação e estimação válida de efeitos causais a partir de dados observacionais. Essas suposições são fundamentais, pois em estudos observacionais, a atribuição de tratamento não é controlada pelos pesquisadores, o que pode levar a vieses e confundimento. As suposições identificadoras ajudam a recriar as condições de um experimento randomizado, permitindo comparações válidas entre os grupos de tratamento e controle. As três suposições mais importantes são:

- Ignorabilidade: Condicionalmente às covariáveis observadas, a atribuição ao tratamento é independente dos resultados potenciais. Isso significa que, após controlar para as covariáveis, a atribuição de tratamento é essencialmente aleatória em relação aos resultados potenciais. Por exemplo, em um estudo sobre os efeitos de um programa de treinamento no desempenho de vendas, a suposição de ignorabilidade implica que, depois de considerar fatores como experiência prévia, idade e motivação dos vendedores, a atribuição ao programa de treinamento é efetivamente aleatória em relação aos seus desempenhos potenciais de vendas com ou sem o treinamento.
- Estabilidade da Unidade de Tratamento (SUTVA): O valor potencial de uma unidade não é afetado pelo tratamento atribuído a outras unidades. Isso implica que não há interferências entre unidades ou efeitos de "contágio". Por exemplo, em um estudo sobre os efeitos de uma vacina, a SUTVA pressupõe que o resultado potencial de um indivíduo vacinado não é influenciado pelo fato de outros indivíduos terem ou não sido vacinados.
- Positividade: Para cada conjunto de valores de covariáveis, há unidades nos grupos de tratamento e controle. Em outras palavras, para qualquer combinação de características observáveis, existe uma chance positiva de uma unidade receber ou não o tratamento.

Por exemplo, em um estudo sobre os efeitos de uma terapia médica, a suposição de positividade requer que, para cada conjunto de características do paciente (como idade, sexo, histórico médico), haja pacientes que receberam e não receberam a terapia.

- *Propensity Score*: O *propensity score* é uma ferramenta especialmente importante em estudos observacionais. É uma pontuação que representa a probabilidade condicional de receber um tratamento dado um conjunto de covariáveis observadas. Por exemplo, em um estudo sobre os efeitos de um novo medicamento para diabetes, o *propensity score* para um paciente poderia ser a probabilidade de receber o novo medicamento com base em fatores como idade, índice de massa corporal, nível de glicose no sangue e outras condições de saúde.

5.5 Métricas

Na investigação pela causalidade podem ser utilizadas algumas métricas dependendo do tipo de pesquisa sendo feita. Por exemplo, caso o objetivo seja estimar o efeito causal de uma intervenção em uma variável:

Average Treatment Effect (ATE)

O *Average Treatment Effect*, ou ATE, refere-se ao efeito médio do tratamento em uma amostra (Angrist and Imbens, 1995). O tratamento pode ser uma intervenção, uma exposição ou qualquer condição que está sendo aplicada aos participantes. O ATE é calculado como a diferença média entre os resultados para o grupo tratado e o grupo de controle. Em termos simples, o ATE mede o impacto médio global do tratamento na população.

$$ATE = E[Y_i^1 - Y_i^0]$$

Onde:

- Y_1 é o resultado potencial para as unidades sob tratamento.

- Y_0 é o resultado potencial para as unidades sob ausência de tratamento.
- A média é calculada sobre toda a população.

Conditional Average Treatment Effect (CATE)

O CATE, ou *Conditional Average Treatment Effect*, refere-se ao efeito médio do tratamento para uma subpopulação específica, condicionado a certas características ou variáveis (Abrevaya et al., 2015). Em outras palavras, o CATE explora como o impacto do tratamento pode variar dependendo das características individuais ou condições iniciais. O CATE é calculado como a diferença média condicional entre os resultados para o grupo tratado e o grupo de controle.

$$CATE(x_1) = E[Y_1 - Y_0 | X_1 = x_1]$$

Onde:

- Y_1 é o resultado potencial para as unidades sob tratamento.
- Y_0 é o resultado potencial para as unidades sob ausência de tratamento.
- X_1 são as características individuais das unidades.
- A média condicional é calculada para uma subpopulação específica caracterizada por $X_1 = x_1$.

6 Aprendizado de Máquina

Nesta seção, são apresentados os principais paradigmas do aprendizado de máquina, que incluem o Aprendizado Supervisionado e o Não-Supervisionado, assim como seus principais problemas e alguns algoritmos/métodos clássicos, tais como Regressão Linear, Árvore de Decisão e Rede Neural. Por fim, abordei brevemente o tema do Aprendizado Profundo. Os conceitos das subseções 6.1 a 6.3 foram retirados de (Müller and Guido, 2017).

6.1 Aprendizado Supervisionado

O aprendizado supervisionado é um paradigma do aprendizado de máquina, no qual queremos prever um resultado de acordo com alguma entrada, dado que também existam em mãos exemplos de entradas/saídas para alimentar o modelo de aprendizado de máquina. Assim, esse modelo depois de treinado, pode ser utilizado para fazer previsões em cima de entradas novas que ele nunca havia visto.

Classificação e Regressão

Os principais problemas tratados pelo aprendizado supervisionado são os de classificação e regressão.

Digamos que exista um conjunto de dados com entradas sobre três tipos de flores, Rosas, Tulipas e Petúnias. O objetivo final é: dizer, dado uma nova entrada, que tipo de flor ela é.

Esse é um exemplo de problema de Classificação. Essencialmente, o objetivo é treinar um modelo de aprendizado de máquina para que ele consiga rotular corretamente novas entradas.

Da mesma forma, agora digamos que exista um conjunto de dados com entradas sobre salário de pessoas dado o grau de educação, idade e profissão. O objetivo dessa vez é prever qual será o salário de uma pessoa dado essas informações.

Nesse caso, temos exemplo de um problema de Regressão. Aqui, o objetivo é treinar um modelo para prever um número contínuo, nesse caso o salário de uma pessoa.

6.2 Aprendizado Não-Supervisionado

No aprendizado não-supervisionado nós só possuímos um conjunto de dados sem uma saída específica, não existe instrução nenhuma para o algoritmo de aprendizado de máquina. Nesse caso, o objetivo é apenas

fornecer os dados para o algoritmo e em troca ele extrai conhecimento desses dados.

Transformações não-supervisionadas e Clusterização

Dois tipos clássicos de aprendizado não-supervisionado são as Transformações e Clusterização.

As transformações se caracterizam por serem algoritmos que criam uma nova representação dos dados com objetivo de torná-los mais compreensíveis do ponto de vista humano.

Uma aplicação comum desses algoritmos é a redução de dimensionalidade, que se trata de encontrar uma forma que sumariza as características essenciais de um conjunto de dados utilizando apenas as informações mais importantes dele. Outra aplicação comum é a extração de tópicos - onde, dado um conjunto de dados textuais, são extraídos os principais tópicos desses conjuntos.

Já os algoritmos de clusterização tem como objetivo agrupar itens distintos de acordo com algum critério de similaridade. Por exemplo, agrupar clientes de acordo com seus hábitos de compra.

6.3 Algoritmos e métodos clássicos utilizados em Aprendizado de Máquina

Regressão Linear

Regressão linear, ou Mínimos Quadrados Ordinários (MQO), é uma técnica simples para encontrar uma linha reta que melhor se ajusta a um conjunto de dados. Essa linha é usada para fazer previsões. A ideia é ajustar essa linha de modo a minimizar a diferença entre as previsões e os valores reais. Em outras palavras, a regressão linear procura os melhores valores para as variáveis da função dessa linha, de modo a tornar as previsões tão próximas quanto possível dos resultados reais no conjunto de treinamento. Apesar de ser simples, a regressão linear não tem muitos ajustes (parâmetros) para controlar a complexidade do modelo.

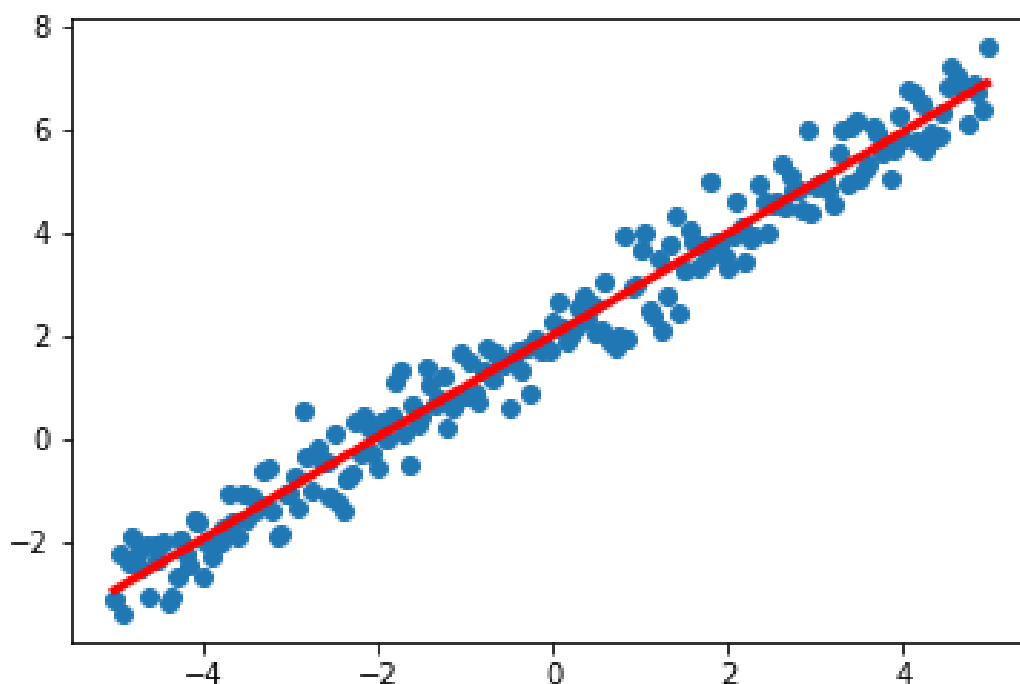


Figura 9: Exemplo de umas Regressão Linear básica - Fonte: O autor

Árvore de Decisão

Uma árvore de decisão é uma técnica em que você toma decisões com base em uma série de perguntas simples. Pode ser comparada a um jogo de "adivinhação", onde você faz perguntas em série para chegar a uma conclusão.

Na figura 10 está exemplificada uma árvore de decisão simples para prever a decisão de jogar tênis com base nas condições atmosféricas e na temperatura. A árvore começaria com uma pergunta como "Qual a condição climática?". Se a resposta for "Ensolarado", você faz outra pergunta, como "Qual a temperatura?" e assim por diante. Cada pergunta leva a um ramo diferente da árvore até que você chegue à classificação final: Joga tênis ou não.

Em termos mais técnicos, a árvore de decisão divide os dados em grupos com base em características relevantes, de modo que as decisões se tornam mais específicas à medida que você avança pela árvore. Essas divisões são feitas de maneira a minimizar a ambiguidade na classificação.

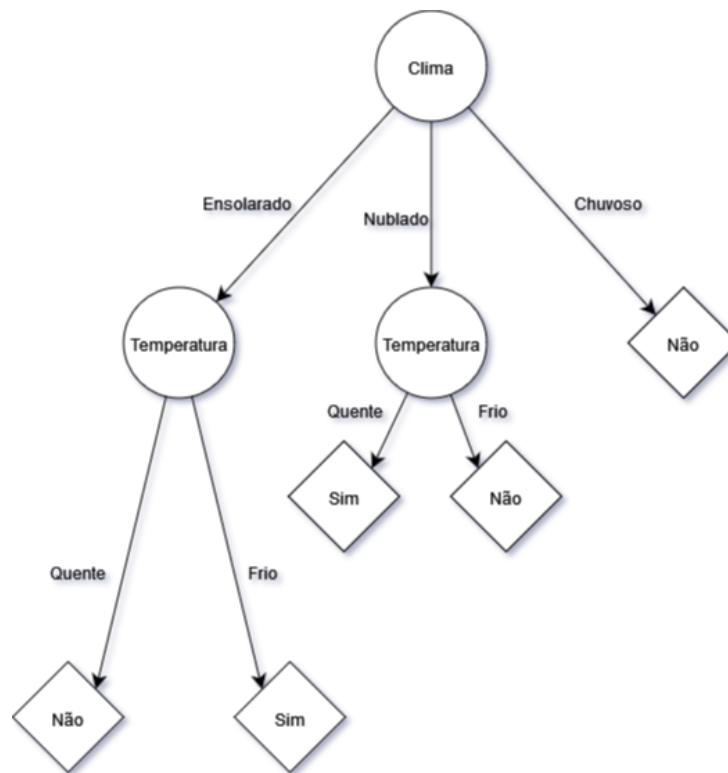


Figura 10: Exemplo de Árvore de Decisão - Fonte: O autor

Rede Neural

Segundo Gurney (2018), uma rede neural é como um conjunto interligado de elementos de processamento simples, chamados unidades ou nós, que são inspirados de maneira geral nos neurônios dos animais. A capacidade de processamento dessa rede é armazenada nas forças de conexão entre essas unidades, conhecidas como pesos. Esses pesos são ajustados através de um processo de adaptação ou aprendizado, onde a rede se aprimora com base em um conjunto de padrões de treinamento.

Uma rede neural é composta por várias camadas, sendo elas:

- Camada de Entrada: Recebe os dados ou características iniciais.
- Camadas Intermediárias (ou Ocultas): Realizam o processamento e aprendizado, identificando padrões e características nos dados.
- Camada de Saída: Produz a saída final da rede, que pode ser uma previsão ou classificação.

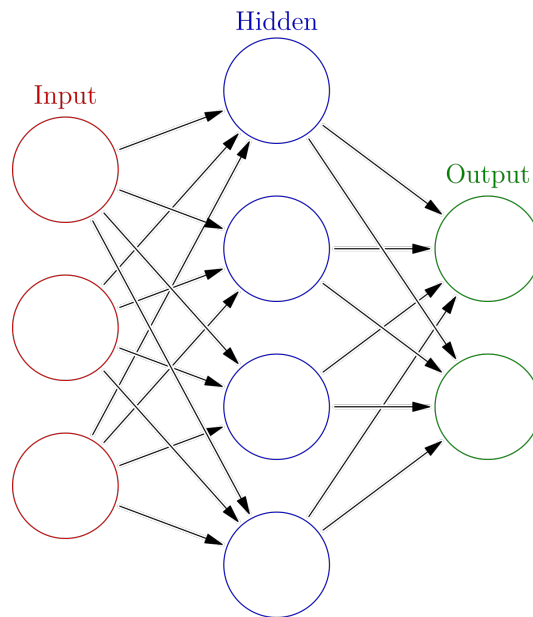


Figura 11: Exemplo de Rede Neural

Fonte: https://en.wikipedia.org/wiki/Artificial_neural_network acessado pela última vez em 08/06/2024

6.4 Aprendizado Profundo

O Aprendizado Profundo é uma subárea do Aprendizado de Máquina que se concentra no uso de redes neurais profundas para aprender e realizar tarefas complexas. O termo "profundo" refere-se ao uso de redes neurais com várias e várias camadas (também conhecidas como arquiteturas profundas).

A característica distintiva do Aprendizado Profundo é a capacidade de automaticamente aprender representações de dados em camadas hierárquicas. Isso significa que a rede neural pode aprender a identificar características ou padrões abstratos nos dados em diferentes níveis de complexidade. Quanto mais camadas a rede tiver, mais complexas podem ser as representações aprendidas.

O Aprendizado Profundo tem sido aplicado com sucesso em uma variedade de tarefas, incluindo reconhecimento de imagem (Liu et al., 2016), processamento de linguagem natural (Devlin et al., 2018), reconhecimento de voz (Yu and Deng, 2016), entre outras.

7 Aprendizado de Máquina para Inferência Causal

7.1 A Integração das Duas Áreas

A integração do Aprendizado de Máquina e Inferência Causal é uma área de pesquisa em crescimento. No mundo real, muitas vezes lidamos com dados complicados que têm muitas informações e variáveis conectadas entre si. O aprendizado de máquina é útil aqui porque consegue lidar com a alta dimensionalidade e encontrar padrões complicados nos dados. Quando juntamos o aprendizado de máquina com a inferência causal, conseguimos entender melhor como essas variáveis estão relacionadas.

7.2 Métodos/Algoritmos

Florestas Causais

As florestas causais são uma extensão das florestas aleatórias projetadas para estimar efeitos causais em vez de simplesmente prever resultados. Elas são utilizadas em cenários onde o interesse reside em entender o impacto de uma variável de tratamento em um resultado, enquanto se controla por variáveis de confusão (Athey and Wager, 2019). Suas principais características são:

- **Divisão Estratificada:** Em vez de dividir o espaço de covariáveis apenas para melhorar a previsão do resultado, as árvores nas florestas causais dividem o espaço para melhorar a estimativa do efeito do tratamento.
- **Balanceamento de Grupos:** Ao construir as árvores, a floresta causal se esforça para balancear os grupos de tratamento e controle dentro de cada folha da árvore para uma comparação justa.
- **Estimativa do Efeito:** Depois que a árvore é construída, cada folha contém subgrupos de unidades de tratamento e controle. A diferença média entre o resultado desses grupos é utilizada para estimar o efeito do tratamento dentro da folha.

Os efeitos do tratamento estimados em cada árvore são agregados para obter uma estimativa global robusta do efeito causal.

Aprendizado Duplamente Robusto

O Aprendizado Duplamente Robusto (Funk et al., 2011) é uma técnica de inferência causal que combina modelos de propensão e modelos de resultado para estimar efeitos causais. Sua principal vantagem é a robustez, visto que a estimativa do efeito causal é consistente se pelo menos um dos modelos estiver corretamente especificado. Suas principais características são:

- Combina dois tipos de modelos: modelos de propensão, que estimam a probabilidade de um tratamento ser recebido, e modelos de resultado, que estimam o resultado esperado sob diferentes condições de tratamento.
- Proporciona robustez contra especificações incorretas, garantindo que se um dos modelos for corretamente especificado, a estimativa do efeito causal será precisa.
- Utiliza técnicas estatísticas tradicionais, sendo relativamente simples de implementar em muitos contextos.

A estimativa do efeito causal no Aprendizado Duplamente Robusto é calculada combinando as previsões dos modelos de propensão e de resultado. A ideia é ajustar a estimativa do efeito causal com base na probabilidade de receber o tratamento e no resultado esperado, corrigindo qualquer viés potencial. Essencialmente, a abordagem calcula uma média ponderada das diferenças entre os resultados observados e esperados, ajustada pela probabilidade de tratamento.

Aprendizado Duplo de Máquina (ADM)

O Aprendizado Duplo de Máquina (Chernozhukov et al., 2017) é uma técnica que também utiliza dois modelos (de propensão e de resultado) para

corrigir vieses e melhorar a robustez das estimativas de efeitos causais. Ele é especialmente útil em situações com muitas variáveis e relações complexas.

- Também implementa um modelo de propensão e um de resultado, assim como no Aprendizado Duplamente Robusto.
- Utiliza uma abordagem de "residualização" ou "ortogonalização" para remover o viés das covariáveis nos modelos.
- Divide os dados em partes para evitar overfitting e garantir que as estimativas sejam não viesadas.

No Aprendizado Duplo de Máquina, os dados são divididos em partes para ajustar modelos de propensão e resultado separadamente. Os resíduos (diferenças entre valores observados e previstos) desses modelos são então usados para estimar o efeito causal.

Algoritmos de Meta-Aprendizado

A meta-aprendizagem treina modelos para melhorar suas capacidades de aprendizado em várias tarefas. Alguns meta-learners que podem ser utilizados para análise causal incluem X-Learner, T-Learner e S-Learner.

7.3 Ferramentas

Algumas ferramentas foram identificadas na literatura para realizar inferência causal com *machine learning*, sendo elas, a princípio, todas bibliotecas da linguagem de programação Python.

DoWhy

A biblioteca DoWhy (Sharma and Kiciman, 2020) é uma ferramenta poderosa desenvolvida pelo Microsoft Research da empresa Microsoft. A biblioteca fornece métodos para responder diversas perguntas sobre causalidade, fornecendo todo o aparato necessário. Atualmente, é a principal biblioteca para análise causal identificada, e será explicitada melhor na seção posterior.

EconML

A biblioteca EconML (Battocchi et al., 2019), também desenvolvida pelo Microsoft Research, fornece diversos métodos para estimar efeitos, como por exemplo algoritmos de Florestas Causais, *Double Machine Learning*, *Doubly Robust Learning* e Meta-Aprendizado.

CausalML

A biblioteca CausalML (Chen et al., 2020) é mantida e desenvolvida pelo time de desenvolvimento da empresa Uber, e também fornece diversos métodos para estimar efeitos. Ela fornece, por exemplo, algoritmos baseados em árvores, algoritmos de meta-aprendizado e algoritmos baseados em redes neurais.

Integração das bibliotecas

Como foi dito, a biblioteca DoWhy fornece os métodos para realizar os quatro passos do processo de análise causal. Apesar disso, ela fornece integração com as outras duas bibliotecas citadas para realização do terceiro passo, que é a estimativa do efeito. Isso abre diversas possibilidades, pois nos permite testar e utilizar algoritmos das três bibliotecas sem ter que fazer mudanças estruturais de código grandes.

8 DoWhy

A DoWhy funciona com base nos dois *frameworks* que foram citados nas seções 5.3 e 5.4: modelos causais gráficos e *potential outcomes*.

8.1 Entendendo as capacidades da biblioteca

Como foi especificado na seção anterior, a biblioteca DoWhy fornece métodos para realizar todos os passos de uma análise de causalidade. Na primeira coluna são as possíveis tarefas que o usuário pode realizar. Logo em seguida se encontra a modelagem do problema utilizando grafos e definição de métodos matemáticos. Finalmente, a realização da tarefa e testes de robustez.

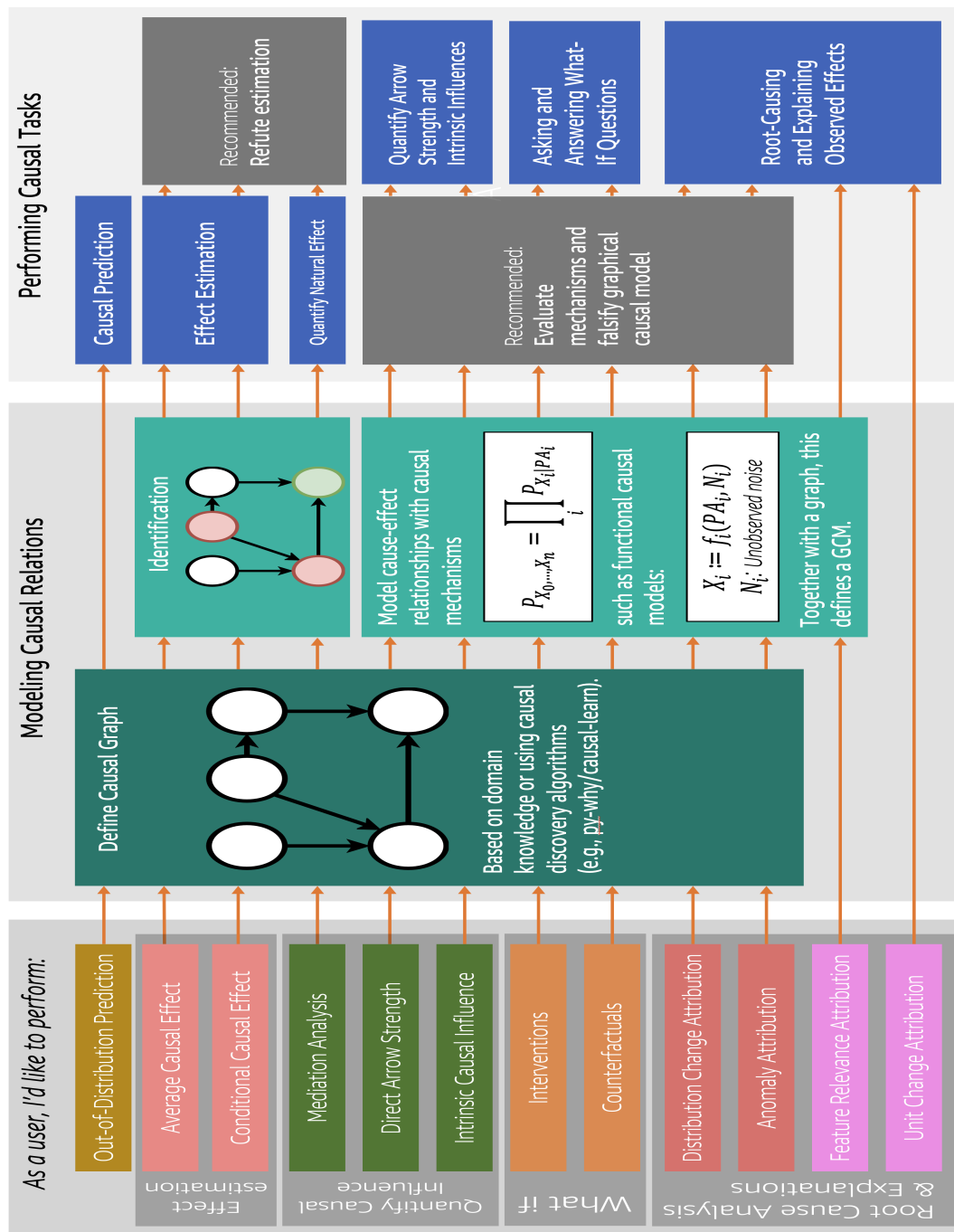


Figura 12: Passo a passo generalizado de utilização da biblioteca. Fonte: <https://github.com/py-why/dowhy?tab=readme-ov-file>. Acesso em 11/05/2023.

Funcionalidades

Na primeira coluna pode-se observar cada uma das análises possíveis de se realizar com a ferramenta.

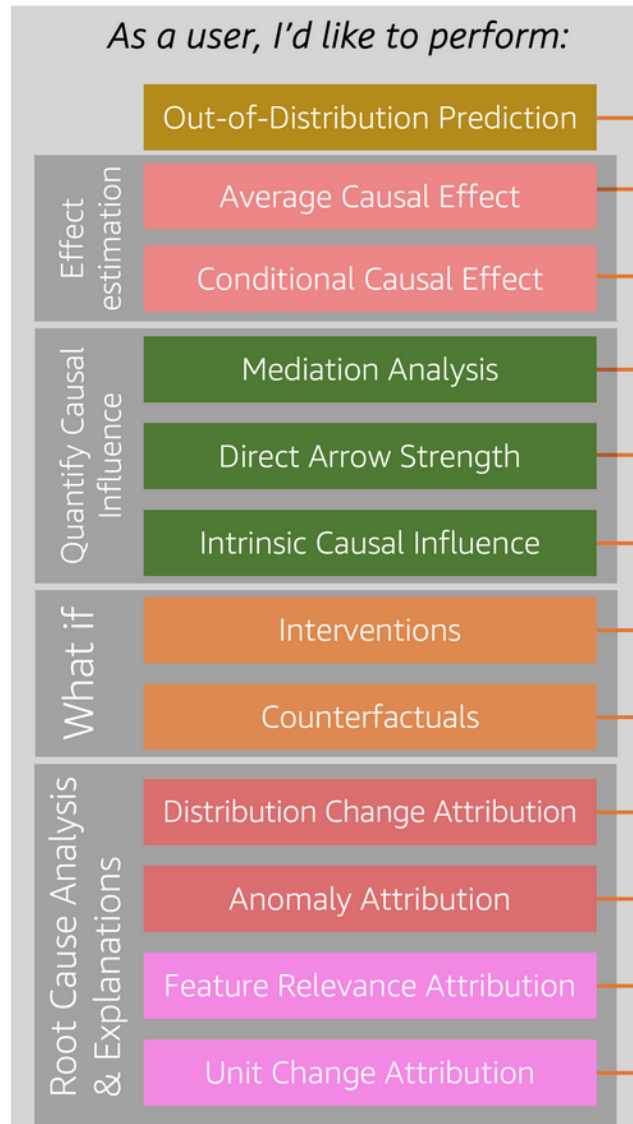


Figura 13: Funcionalidades

Out-of-Distribution Prediction

Tenta realizar previsões causais confiáveis mesmo em distribuições de dados ligeiramente diferentes daquelas usadas para treinar os modelos.

Effect Estimation

Estimar efeito causal de uma intervenção, ambos já citados em seção anterior.

- *Average Causal Effect*: Estimar o efeito causal médio de uma causa em um resultado para uma população específica.
- *Conditional Causal Effect*: Estimar o efeito causal médio de uma causa em um resultado para diferentes grupos dentro de uma população, ou seja, condicionado as diferentes características possíveis dentro dela.

Quantify Causal Influence

Esta seção abrange funcionalidades para quantificar e medir a influência causal de variáveis de entrada em uma variável de saída. Isso inclui:

- *Mediation Analysis*: Decompõe o efeito causal total em efeitos diretos e indiretos, permitindo entender quanto é mediado por outros nós do grafo causal.
- *Direct Arrow Strength*: Quantifica especificamente a força da relação causal direta entre um par de variáveis na estrutura causal. Este valor mede a quantidade de variância na variável alvo que pode ser diretamente atribuída à influência de uma variável específica, sem considerar influências indiretas ou mediadas por outras variáveis no modelo.
- *Intrinsic Causal Influence*: Calcula uma pontuação que representa quão direta e intrinsecamente um nó do grafo afeta outro, levando em conta a estrutura causal subjacente. Esta medida inclui tanto as influências diretas quanto as indiretas, proporcionando uma visão mais abrangente da importância de uma variável ao explicar a variância na variável alvo, considerando toda a rede de interações no modelo.

Essas métricas ajudam a quantificar e ranquear a importância causal das variáveis de entrada para uma variável de resultado específica.

What if

Esta seção permite a análise de cenários contrafactuais e intervenções hipotéticas por meio de:

- *Interventions*: Estima os efeitos causais de intervenções externas hipotéticas.
- *Counterfactuals*: Simula cenários "e se" alterando valores de variáveis de entrada e analisando as mudanças resultantes na variável de saída.

Essas funcionalidades permitem explorar resultados potenciais sob diferentes condições.

Root Cause Analysis & Explanation

Esta seção aborda a atribuição e explicação de modelos, identificando as causas fundamentais por trás de previsões e comportamentos:

- *Distribution Change Attribution*: Ajuda a identificar as razões pelas quais a distribuição de uma variável de interesse pode ter mudado ao longo do tempo.
- *Anomaly Attribution*: Identifica quais nós do grafo foram os principais responsáveis por gerar um valor ou observação atípica na variável de resultado, auxiliando na detecção de causas raiz
- *Feature Relevance Attribution*: Avalia quais características (ou "features") são mais relevantes para a previsão de um resultado.
- *Unit Change Attribution*: Análise que procura entender como mudanças em uma única observação podem afetar o resultado geral. Essa análise foca em medir o impacto da alteração de uma unidade individual no conjunto de dados como um todo.

Essas capacidades fornecem interpretabilidade e explicações baseadas em relações causais, ajudando a entender os motivos por trás de previsões e detectar causas raiz de comportamentos incomuns.

8.2 Modelagem das Relações Causais

Nessa etapa, o objetivo é construir um modelo formal que capture as relações de causa e efeito entre as variáveis de interesse, com base em um grafo causal.

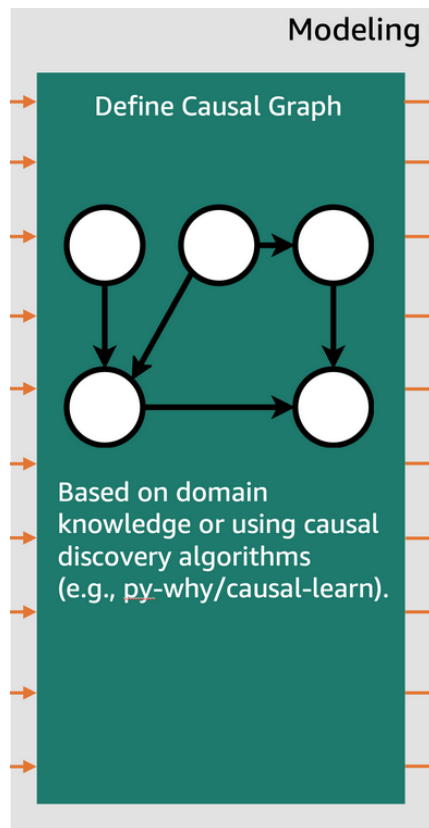


Figura 14: Etapa de criação do grafo causal

Definição do Grafo Causal

A construção do grafo causal do problema é extremamente importante, pois ele quem dita as relações causais entre as variáveis. Por isso ele deve ser construído, preferencialmente, por um especialista do domínio que está sendo estudado, de forma que as relações modeladas no grafo traduzam plenamente o que ocorre no mundo real. Como citado anteriormente, os grafos são direcionados e acíclicos, com cada aresta representando a causalidade de um nó com outro.

O contexto (Blöbaum et al. (2023)) do grafo causal definido na Figura 15 é sobre venda de um certo modelo de celular online. Como podemos

observar, a variável "Shopping Event?" (a qual indica se estava tendo uma promoção) influencia causalmente a variável "Page Views" (número de visualizações na página do produto), que por sua vez influencia causalmente "Sold Units" (número de unidades do produto vendidas). As arestas desse grafo são oriundas das hipóteses que alguém com conhecimento sobre esse domínio específico formulou, então por exemplo uma dessas hipóteses foi "Eventos promocionais (Shopping Event) atraem usuários a visitarem a página do produto (Page Views) devido a descontos ou ofertas."

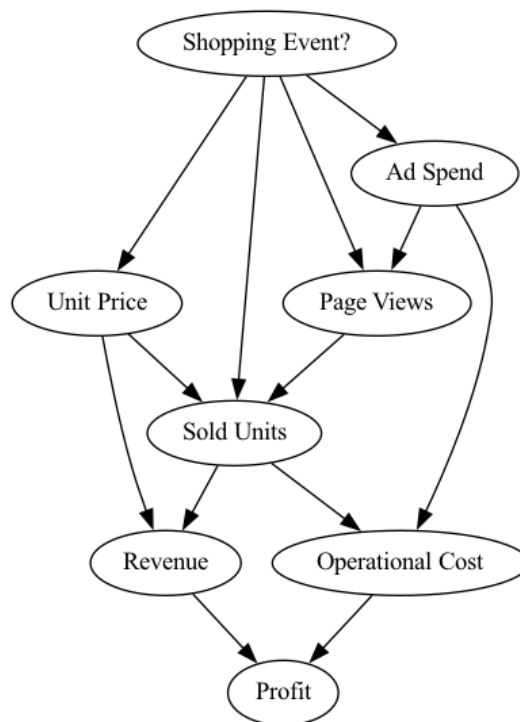


Figura 15: Exemplo de grafo causal. Fonte: <https://aws.amazon.com/pt/blogs/opensource/root-cause-analysis-with-dowhy-an-open-source-python-library-for-causal-machine-learning/>. Acesso em 19 abr. 2024

Identification/Cause-effect Modelling

Dependendo da análise que o usuário deseja realizar, dois caminhos diferentes são tomados aqui.

Para cálculo de "Average Causal Effect", "Conditional Causal Effect" e "Mediation Analysis", o usuário deve passar pela etapa denominada na Figura 12 por "Identification", que se trata da identificação da formulação de como será calculado o efeito causal de interesse. Aqui são utilizados os critérios mencionados na seção "Do-Calculus", backdoor, frontdoor e instrumental variables (IV).

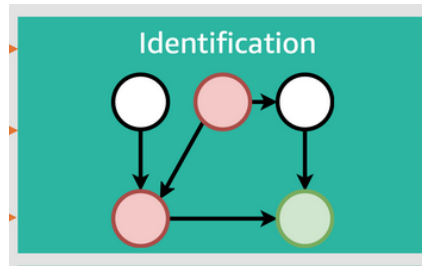


Figura 16: Etapa de identificação de como será estimado o efeito

Por outro lado, para as outras tarefas (exceto "Out-of-Distribution Prediction", "Unit Change Attribution" e as três citadas acima), o usuário deve construir um Graphical Causal Model (GCM) (Blöbaum et al., 2022) para o problema.

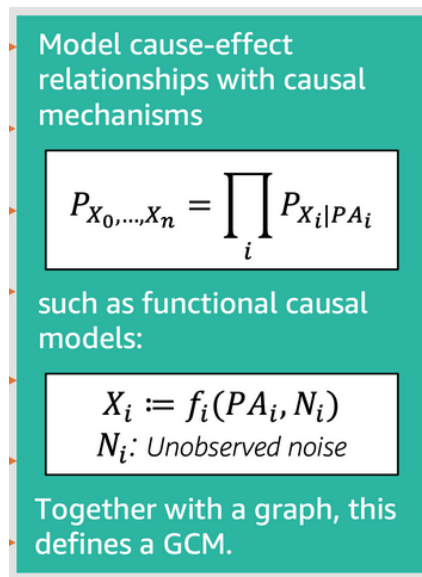


Figura 17: Etapa de criação do GCM

GCMs são a junção do grafo causal com um mecanismo causal para cada uma das variáveis. Um mecanismo causal define a distribuição condicional de uma variável dado seus pais (nós com aresta para a variável) no grafo, ou, em caso de nós raízes, simplesmente sua própria distribuição.

Realização das Tarefas de Interesse

Finalmente, o usuário pode realizar a tarefa que ele pretendia.

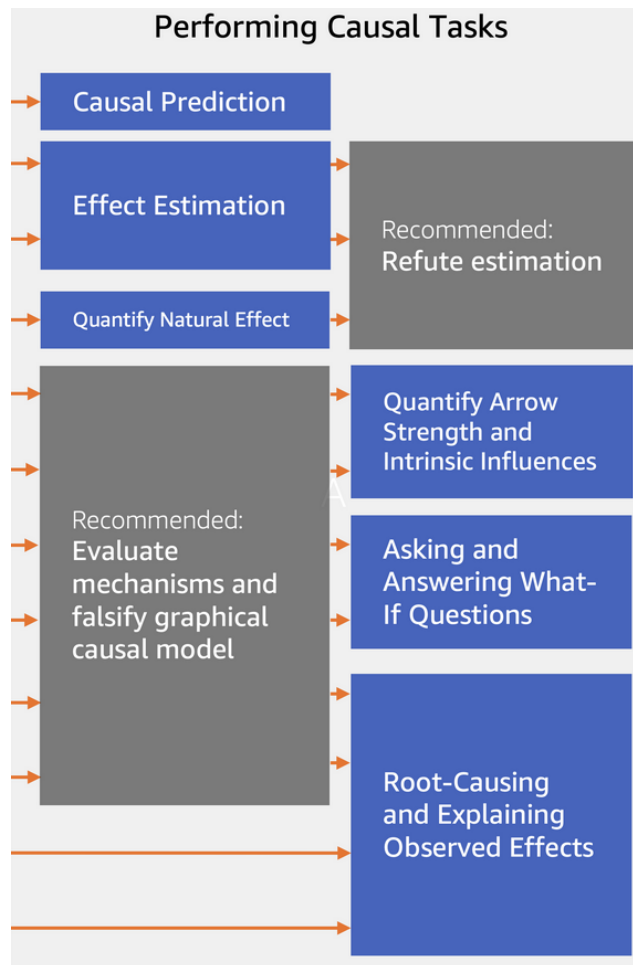


Figura 18: Realização da tarefa desejada e testes de robustez

Nessa etapa, vale a pena destacar que a DoWhy fornece aos usuários métodos para avaliar a robustez do modelo causal que o usuário criou aplicando um método de refutação. No caso de tarefas de estimação, ele pode utilizar alguns métodos para avaliar o resultado obtido, como por exemplo:

- Causa comum aleatória: É criada uma causa comum aleatória independente (uma variável que afeta tanto o tratamento, quanto o resul-

tado, mas não possui nenhuma relação causal com ambos) e é verificado como ela afeta o resultado da estimativa. O que se espera desse teste é que o novo valor estimado não mude, ou varie bem pouco.

- Utilização de apenas uma parte do conjunto de dados: Esse método testa a robustez da estimativa dividindo aleatoriamente os dados em dois subconjuntos e verificando se a estimativa causal é consistente entre os subconjuntos. Aqui também se espera que a nova estimativa varie bem pouco da estimativa original.
- Tratamento placebo: Esse método testa a robustez da estimativa causal simulando um "tratamento placebo". Isso significa que ele cria uma nova variável de tratamento artificial (placebo) que, por construção, não deveria ter nenhum efeito causal sobre o resultado. Em seguida, ele estima o efeito causal desse tratamento placebo no resultado. Se o efeito estimado for significativamente diferente de zero, isso pode indicar que há algum viés ou violação das suposições na estimativa original.

No contexto desses métodos de refutação, é importante explicitar os conceitos de "Hipótese nula" e "p-value". A hipótese nula (University, 2023) é a suposição inicial em um teste de hipótese estatístico. Ela geralmente assume que não há efeito ou relação entre as variáveis estudadas. Por exemplo, em um estudo clínico, a hipótese nula pode ser "o novo medicamento não tem efeito no tratamento da doença". O p-valor é a probabilidade de obter um resultado igual ou mais extremo do que o observado, assumindo que a hipótese nula seja verdadeira. Quanto menor o p-value, maior a evidência contra a hipótese nula. Existe um valor de referência convencional (geralmente 0,05) para ele. Se o p-value calculado for menor que esse valor de referência, rejeita-se a hipótese nula, caso contrário, não há evidências suficientes para rejeitar a hipótese nula. Portanto, esse valor é usado para avaliar se os resultados observados são estatisticamente significativos ou

não, fornecendo uma medida quantitativa da força da evidência contra a hipótese nula. Quanto menor ele for, mais improvável é que os resultados tenham ocorrido por acaso, caso a hipótese nula seja verdadeira. Em resumo, para os três testes mencionados, espera-se um p-value alto. No contexto destes testes, isso significa que, se algum deles tentasse introduzir um obstáculo na estimação, ele não rejeitaria a hipótese nula, mantendo, assim, o suporte à estimativa original.

Por outro lado, ao construir um Graphical Causal Model (GCM) usando a biblioteca DoWhy, ela também fornece métodos para testar a robustez do modelo (Eulig et al., 2023). Esses métodos envolvem calcular diversas estatísticas (mostradas expostas na subseção 9.4, na saída do Código 9) de acordo com o grafo causal, os mecanismos causais e dados utilizados. A partir disso, é possível ter alguma referência do quão bom é o modelo criado e as hipóteses utilizadas.

9 Primeiro Estudo de Caso - Felicidade

Para o primeiro estudo de caso foi utilizado o conjunto de dados *World Happiness Report 2015* (Gallup, 2015). Esse dataset baseia-se em uma série de fatores para calcular uma pontuação de felicidade para cada país avaliado. Os atributos utilizados para o estudo foram os seguintes:

- Economy (GDP per Capita): Indicador econômico que mede a produção média de bens e serviços por pessoa em um país em um ano. É calculado dividindo-se o PIB total do país pela sua população.
- Family: Média nacional das respostas binárias (0 ou 1) à pergunta "Se você estivesse em apuros, teria parentes ou amigos com quem pode contar para ajudá-lo sempre que precisar, ou não?"
- Healthy (Life Expectancy): Número médio de anos que uma pessoa pode esperar viver com boa saúde. Ao contrário da expectativa de vida simples, dados sobre a expectativa de vida saudável não estão

amplamente disponíveis, então foram utilizadas alguns métodos para calculá-lo. Para mais detalhes, consulte a fonte referenciada.

- Freedom: Média nacional das respostas para a pergunta "Você está satisfeito(a) ou insatisfeito(a) com a sua liberdade de escolher o que fazer da sua vida?"
- Trust (Government Corruption): Média nacional das respostas da pesquisa para duas perguntas na Gallup World Poll: "A corrupção está generalizada em todo o governo ou não?" e "A corrupção está generalizada dentro das empresas ou não?" A percepção geral é apenas a média das duas respostas binárias (0 ou 1).
- Generosity: Residual da regressão da média nacional de respostas à pergunta "Você doou dinheiro para uma instituição de caridade no mês passado?" em relação ao PIB per capita.
- Happiness Score: É a média nacional para a pergunta: "Por favor, imagine uma escada, com degraus numerados de 0 na parte inferior a 10 na parte superior. O topo da escada representa a melhor vida possível para você e a parte inferior da escada representa a pior vida possível para você. Em que degrau da escada você diria que pessoalmente se sente neste momento?"

Os detalhes de como foi coletado e/ou calculado cada uma das informações podem ser encontrados na referência.

As causal queries (perguntas causais) que serão respondidas são:

1. Qual o impacto médio do PIB per Capita na Pontuação de Felicidade?
2. Como cada variável afeta a Pontuação de Felicidade?
3. Qual a influência causal intrínseca de cada nó na Pontuação de Felicidade (a força de influência dos nós sem contar com o que cada um herda de seus pais)?

4. O que aconteceria caso ocorresse uma intervenção no PIB per Capita dos países?

As três primeiras perguntas foram utilizadas, pois são bem gerais para analisar problemas de causalidade, encontradas também em outros exemplos de aplicação. A última pergunta foi escolhida de modo a exemplificar as capacidades da biblioteca.

9.1 Pré-processamento

No Código 1 foi realizado o pré-processamento dos dados, onde foram removidas algumas colunas que não eram necessárias para o estudo de caso.

Código 1: Descartando as colunas que não serão utilizadas

```
dataset = dataset.drop(['Country',  
                        'Region',  
                        'Standard Error',  
                        'Dystopia Residual',  
                        'Happiness Rank'], axis=1)
```

Além disso, no Código 2 as colunas restantes foram renomeadas para facilitar o entendimento.

Código 2: Renomeação das colunas

```
dataset = dataset.rename(columns={'Happiness Score': 'Pontuação de  
Felicidade',  
                                'Economy (GDP per Capita)': 'PIB per Capita',  
                                'Family': 'Apoio social',  
                                'Health (Life Expectancy)': 'Expectativa de  
Vida Saudável',  
                                'Freedom': 'Liberdade',  
                                'Trust (Government Corruption)': 'Corrupção',  
                                'Generosity': 'Generosidade'})
```

9.2 Criação do Grafo Causal

Primeiramente, no Código 3 foi criado o grafo causal, o qual pode ser visto na Figura 19. As hipóteses foram retiradas da literatura, mais especificamente de Tay et al. (2015) e Diener et al. (1995)

Código 3: Criação do Grafo Causal

```
import networkx as nx
```

```

causal_graph = nx.DiGraph()

causal_graph.add_nodes_from(['Pontuação de Felicidade',
                              'PIB per Capita',
                              'Apoio social',
                              'Expectativa de Vida Saudável',
                              'Liberdade',
                              'Corrupção',
                              'Generosidade'])

causal_graph.add_edges_from([
    ('PIB per Capita', 'Pontuação de Felicidade'),
    ('PIB per Capita', 'Apoio social'),
    ('PIB per Capita', 'Expectativa de Vida Saudável'),
    ('Apoio social', 'Pontuação de Felicidade'),
    ('Expectativa de Vida Saudável', 'Pontuação de Felicidade'),
    ('Liberdade', 'Pontuação de Felicidade'),
    ('Generosidade', 'Pontuação de Felicidade'),
    ('Corrupção', 'Liberdade'),
    ('Corrupção', 'PIB per Capita'),
    ('Apoio social', 'Liberdade'),
    ('PIB per Capita', 'Generosidade'),
    ('PIB per Capita', 'Liberdade'),
])

```

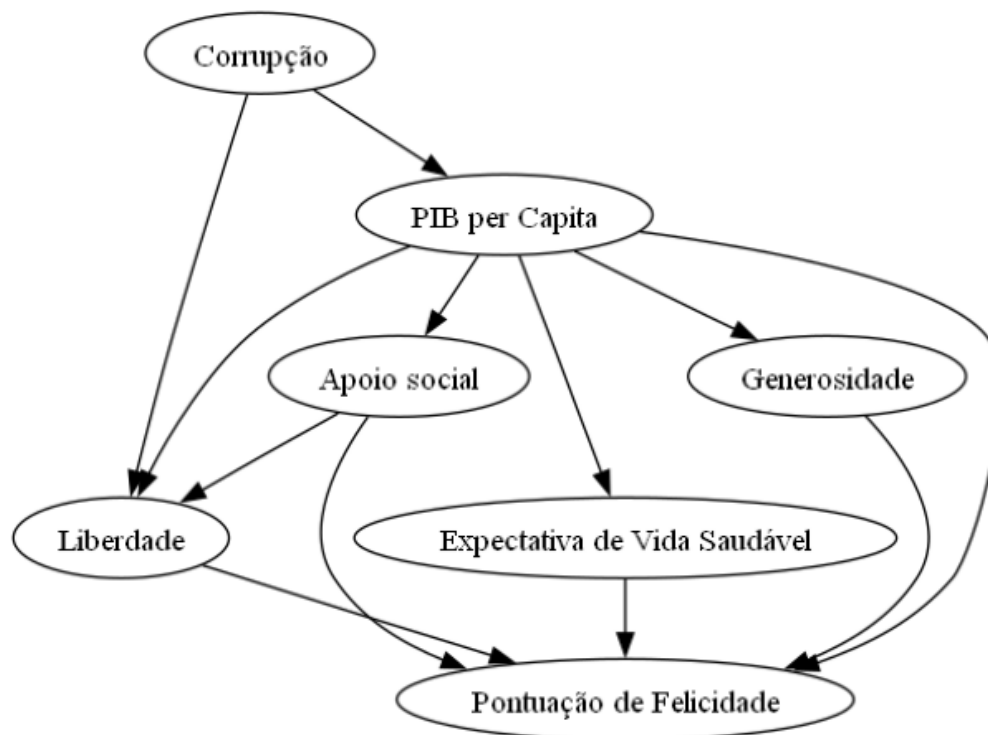


Figura 19: Grafo causal representando as relações das variáveis

9.3 Primeira Causal Query

Para responder a primeira causal query ("Qual o impacto médio do PIB per Capita na Pontuação de Felicidade?") o passo a passo seguido foi: identificação do estimando, estimação do efeito e, finalmente, teste de robustez do efeito estimado.

Criação do Modelo Causal e Identificação do Estimando

Após isso, criamos o modelo causal no Código 4 e identificamos o método utilizado para fazer o cálculo da relação que está sendo estudada no Código 5.

Código 4: Criação do modelo causal, especificando o tratamento como o PIB per Capita e a variável de resultado como a Pontuação de Felicidade

```
causal_model = CausalModel(  
    data=dataset,  
    treatment='PIB per Capita',  
    outcome='Pontuação de Felicidade',  
    graph=causal_graph  
)
```

Código 5: Identificação do estimando

```
identified_estimand =  
    causal_model.identify_effect(proceed_when_unidentifiable=True)  
print(identified_estimand)
```

```
Estimand type: EstimandType.NONPARAMETRIC_ATE  
  
### Estimand : 1  
Estimand name: backdoor  
Estimand expression:  
    d  
    _____(E[Pontuação de Felicidade|Corrupção])  
d[PIB per Capita]  
Estimand assumption 1, Unconfoundedness: If  $U \rightarrow \{\text{PIB per Capita}\}$  and  $U \rightarrow \text{Pontuação de Felicidade}$  then  $P(\text{Pontuação de Felicidade}|\text{PIB per Capita}, \text{Corrupção}, U) = P(\text{Pontuação de Felicidade}|\text{PIB per Capita}, \text{Corrupção})$   
  
### Estimand : 2  
Estimand name: iv  
No such variable(s) found!  
  
### Estimand : 3
```

```
Estimand name: frontdoor
No such variable(s) found!
```

Como pode-se ver, o estimando encontrado foi o "backdoor" e o subconjunto de variáveis que satisfazem esse critério é composto apenas pela variável Corrupção. Isso quer dizer que para encontrar o efeito do PIB per Capita na Pontuação de Felicidade, é necessário condicionar pela Corrupção.

Estimação do Efeito

Foi utilizada a técnica de Double Machine Learning no Código 6 para realizar a estimativa. Foi feito um pequeno ajuste fino nos modelos de regressão utilizado, definindo o número de estimadores como 1200 e profundidade máxima igual a 6. Já os parâmetros "treatment_value" e "control_value" são os da relação que estamos estudando, ou seja, queremos verificar qual o valor médio de "Pontuação de Felicidade" após aumentar o valor de "PIB per Capita" de 0.5 para 0.8.

Código 6: Estimando com Double Machine Learning

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.linear_model import LassoCV

dml_estimate = causal_model.estimate_effect(
    identified_estimand,
    method_name='backdoor.econml.dml.DML',
    method_params={
        'init_params': {
            'model_y': GradientBoostingRegressor(n_estimators=1200,
                                                  max_depth=6),
            'model_t': GradientBoostingRegressor(n_estimators=1200,
                                                  max_depth=6),
            'model_final': LassoCV(fit_intercept=False),
        },
        "fit_params": {}
    },
    treatment_value=0.8,
    control_value=0.5,
)

print(dml_estimate)
```

```
*** Causal Estimate ***

## Identified estimand
Estimand type: EstimandType.NONPARAMETRIC_ATE
```

```

### Estimand : 1
Estimand name: backdoor
Estimand expression:
      d
      _____(E[Pontuação de Felicidade|Corrupção])
d[PIB per Capita]
Estimand assumption 1, Unconfoundedness: If  $U \perp \{ \text{PIB per Capita} \}$  and  $U \perp \text{Pontuação de Felicidade}$  then  $P(\text{Pontuação de Felicidade} | \text{PIB per Capita}, \text{Corrupção}, U) = P(\text{Pontuação de Felicidade} | \text{PIB per Capita}, \text{Corrupção})$ 

## Realized estimand
b: Pontuação de Felicidade~PIB per Capita+Corrupção |
Target units: ate

## Estimate
Mean value: 0.582934763867024
Effect estimates: [[0.58293476]]

```

O resultado indica que aumentando o valor da variável de tratamento (PIB per Capita) de 0.5 para 0.8 causa um aumento de aproximadamente 0.58 na variável de resultado (Pontuação de Felicidade), sobre a distribuição de dados/população representada pelo conjunto de dados.

Teste de Robustez

Naturalmente, queremos testar a qualidade da nossa estimativa e a robustez do modelo causal criado. Para isso, utilizamos as funções do Código 7.

Código 7: Testes de robustez utilizados

```

refute_results_1 = causal_model.refute_estimate(
    identified_estimand,
    dml_estimate,
    method_name='random_common_cause'
)

refute_results_2 = causal_model.refute_estimate(
    identified_estimand,
    dml_estimate,
    method_name='data_subset_refuter',
    subset_fraction=0.9
)

refute_results_3 = causal_model.refute_estimate(
    identified_estimand,
    dml_estimate,
    method_name='placebo_treatment_refuter',
    placebo_type='permute',
    num_simulations=5
)

```



```
print(refute_results_1)
print(refute_results_2)
print(refute_results_3)
```

```
Refute: Add a random common cause
Estimated effect:0.582934763867024
New effect:0.5650160955065813
p value:0.72

Refute: Use a subset of data
Estimated effect:0.582934763867024
New effect:0.5955306096103797
p value:0.8

Refute: Use a Placebo Treatment
Estimated effect:0.582934763867024
New effect:-0.02833414031031505
p value:0.41053069776276974
```

Perceba que nos dois primeiros testes o valor da estimativa não difere muito, e no tratamento placebo a nova estimativa vai para perto de 0 (resultados positivos). Além disso, nos três testes o valor de p é relativamente alto considerando o valor de referência (0,05), não rejeitando a hipótese nula, portanto mostrando uma certa robustez do efeito original estimado.

Conclusões - Primeira Causal Query

Podemos concluir que, dado esse grafo causal e o conjunto de dados disponíveis, "PIB per Capita" tem uma relação causal positiva com "Pontuação de Felicidade" para o valor de tratamento estudado (0.5 para 0.8). Além disso, analisando a saída dos testes de robustez, pode-se perceber que a estimativa original se mantém relevante, pois os testes não conseguem refutá-la.

9.4 Segunda Causal Query

Para responder a segunda causal query ("Como cada variável afeta a Pontuação de Felicidade?"), foi necessário criar um Graphical Causal Model. No Código 8 é criado o GCM, os mecanismos causais são automaticamente atribuídos aos nós e, por fim, os modelos são ajustados aos dados.

Vale destacar que o grafo criado anteriormente é reutilizado na criação do GCM, não sendo necessário criar outro. Por último, o GCM é utilizado para responder a pergunta.

Código 8: Criação do Graphical Causal Model

```
causal_model_2 = gcm.StructuralCausalModel(causal_graph)
auto_assignment_summary =
    gcm.auto.assign_causal_mechanisms(causal_model_2, dataset)
gcm.fit(causal_model_2, dataset)
```

Perceba que a atribuição dos mecanismos causais das variáveis foi feita de forma automática por meio da funcionalidade `gcm.auto.assign_causal_mechanisms()`. Após isso, foi verificada a qualidade dos mecanismos causais atribuídos e do DAG no Código 9 com a função `gcm.evaluate_causal_model()`.

Código 9: Função para avaliar o modelo

```
print(gcm.evaluate_causal_model(causal_model_2, dataset))
```

```
Evaluated the Performance of Causal Mechanisms and Invertibility Assumption
Overall average KL divergence between the generated and observed distribution and
the graph structure. The results are as follows:
Evaluation of Causal Mechanisms

The used evaluation metrics are:

KL Divergence (only for root-nodes):
Evaluates the divergence between the generated and the observed distribution.

Mean Squared Error (MSE):
Evaluates the average squared differences between the observed values and the
conditional expectation of the causal mechanisms.

Normalized MSE (NMSE):
The MSE normalized by the standard deviation for better comparison.

R² Coefficient:
Indicates how much variance is explained by the conditional expectations of
the mechanisms. This can be misleading for nonlinear relationships.

F1 Score (only for categorical non-root nodes):
The harmonic mean of precision and recall indicating the quality of the
classifier model.
```

(Normalized) Continuous Ranked Probability Score (CRPS):

Generalizes the Mean Absolute Percentage Error to probabilistic predictions.

Provides insights into the accuracy and calibration of causal mechanisms.

NOTE: Every metric focuses on different aspects and might not consistently indicate good or bad performance. The CRPS will be mainly used to compare and interpret performance since it captures important properties for the causal model.

Node Evaluations

Node Corrupção

KL Divergence: 0.09617974229864792

Indicates an overall very good representation of the data distribution.

Node PIB per Capita

MSE: 0.14852767732085578

NMSE: 0.9580101099078047

R² Coefficient: 0.07947847071695016

Normalized CRPS: 0.5446782285623917

The estimated CRPS indicates only fair performance, which might also be due to a small signal-to-noise ratio.

Node Apoio Social

MSE: 0.043792114878004815

NMSE: 0.7759583743607756

R² Coefficient: 0.3948191779417355

Normalized CRPS: 0.4365319119645844

CRPS indicates only a fair performance, possibly due to a small signal-to-noise ratio.

Node Expectativa de Vida Saudável

MSE: 0.021013545691484625

NMSE: 0.5969596963940272

R² Coefficient: 0.6368364670177347

Normalized CRPS: 0.3201938481716197

CRPS indicates good performance.

Node Generosidade

MSE: 0.014227113131435428

NMSE: 0.9519505722985228
R² Coefficient: 0.08634398699394093
Normalized CRPS: 0.5361682579258404
CRPS indicates fair performance, possibly due to a small signal-to-noise ratio.

Node Liberdade

MSE: 0.014868037577313767
NMSE: 0.8324534514721975
R² Coefficient: 0.30333687031177725
Normalized CRPS: 0.47069942023357647
CRPS indicates fair performance, possibly due to a small signal-to-noise ratio.

Node Pontuação de Felicidade

MSE: 0.33295035728799177
NMSE: 0.5181711240186592
R² Coefficient: 0.7281434217329263
Normalized CRPS: 0.29285924311572825
CRPS indicates good performance.

Evaluation of Invertible Functional Causal Model Assumption

Node Pontuação de Felicidade:
Not rejected (p-value = 0.3812144046527506, alpha = 0.05).
The assumption might be valid.

Node PIB per Capita:
Rejected (p-value = 0.04094581455876223, alpha = 0.05).
The assumption might not be valid due to a hidden confounder or incorrect mechanism.

Node Apoio Social:
Rejected (p-value = 0.014325460679931457, alpha = 0.05).
The assumption might not be valid due to a hidden confounder or incorrect mechanism.

Node Expectativa de Vida Saudável:
Rejected (p-value = 0.02260661553081733, alpha = 0.05).
The assumption might not be valid due to a hidden confounder or incorrect mechanism.

Node Liberdade:
Not rejected (p-value = 1.0, alpha = 0.05).

```

The assumption might be valid.

Node Generosidade:
Not rejected (p-value = 1.0, alpha = 0.05).
The assumption might be valid.

Note: These results are based on statistical independence tests. An assumption not
being rejected does not mean it's correct, only that there is no evidence against
it.

Evaluation of Generated Distribution

Overall Average KL Divergence: 0.049347348392486855
Indicates an overall very good representation of the data distribution.

Evaluation of the Causal Graph Structure

+-----+
| Falsification Summary |
+-----+
| The given DAG is informative because 0/50 permutations lie in the Markov |
| equivalence class |
| of the given DAG (p-value: 0.00). |
| The given DAG violates 5/14 LMCs and is better than 92.0 percent of permuted |
| DAGs (p-value: 0.08). |
| Based on the significance level (0.2) and because the DAG is informative, |
| we do not reject the DAG. |
+-----+

Note

Always double-check model assumptions regarding graph structure and causal
mechanisms. These evaluations provide insights into the causal model but
shouldn't be overinterpreted since some relationships are hard to model.
Many algorithms are fairly robust against inaccuracies.

```

Resumindo, os resultados da avaliação sugerem que o modelo causal gerado representa razoavelmente bem a distribuição dos dados (KL Divergence de aproximadamente 0.04) e captura os mecanismos causais para alguns fatores (Por exemplo, Nó "Expectativa de Vida Saudável" com CRPS apresentando boa performance). No entanto, existem possíveis problemas com alguns pressupostos fundamentais do modelo e com a estrutura das relações causais para algumas variáveis (Por exemplo, o nó "Generosidade" com CRPS apresentando uma performance mediana), o que pode exi-

gir uma investigação mais aprofundada ou ajustes no modelo.

Respondendo à pergunta

A resposta de como cada variável afeta a Pontuação de Felicidade é basicamente a força de cada conexão no grafo para a variável. Para calcular isso, é utilizada a função `gcm.arrow_strength()`. Consulte Janzing et al. (2013) para mais detalhes sobre o método.

Código 10: Arrow Strengths

```
from dowhy.utils import plot

arrow_strengths = gcm.arrow_strength(causal_model_2,
                                     target_node='Pontuação de Felicidade')

plot(causal_graph=causal_graph,
     causal_strengths=arrow_strengths,
     figure_size=(10, 10),)
```

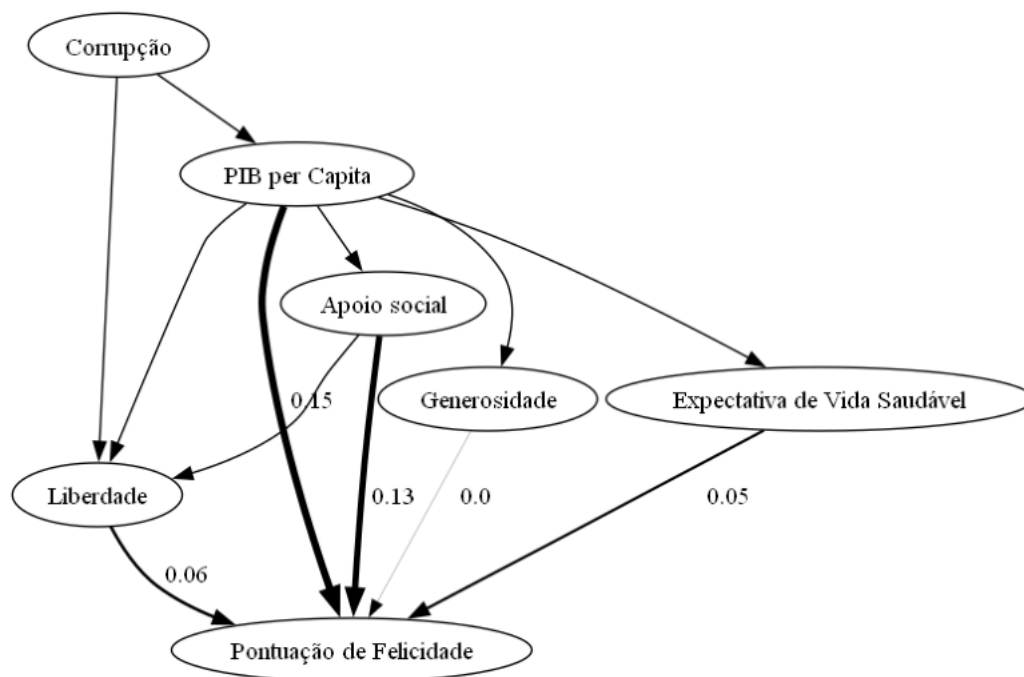


Figura 20: Grafo causal com a força de cada conexão com a variável Pontuação de Felicidade

No Código 11 é calculada a variância da Pontuação de Felicidade.

Código 11: Variância de Pontuação de Felicidade

```
dataset['Pontuação de Felicidade'].var()
```

```
1.31104820914295
```

No Código 12 é calculado a porcentagem que cada nó contribui para a variância da Pontuação de Felicidade a partir do resultado do `arrow_strengths()`.

Código 12: Porcentagem da variância de Pontuação de Felicidade que cada variável explica

```
arrow_strengths_percentages = {k: (v / dataset['Pontuação de Felicidade'].var()) * 100 for k, v in arrow_strengths.items()}  
print(arrow_strengths_percentages)
```

```
{('Apoio social', 'Pontuação de Felicidade'): 10.028142073906945,  
( 'Expectativa de Vida Saudável', 'Pontuação de Felicidade'): 4.174103116337675,  
( 'Generosidade', 'Pontuação de Felicidade'): 0.19808576118455476,  
( 'Liberdade', 'Pontuação de Felicidade'): 4.947312298258729,  
( 'PIB per Capita', 'Pontuação de Felicidade'): 11.688872351775801}
```

Conclusões - Segunda Causal Query

Perceba que as variáveis com mais influência sobre "Pontuação de Felicidade" são "PIB per Capita" e "Apoio social", com "Liberdade" e "Expectativa de Vida Saudável" vindo logo após. O número associado com cada conexão no grafo significa o aumento na variância de "Pontuação de Felicidade" que surgiria caso fosse removida a aresta, enquanto que no Código 12, é possível ver a porcentagem da variância de "Pontuação de Felicidade" que cada variável contribui. Esses valores são relativamente pequenos, o que sugere que a influência direta de cada variável sobre a "Pontuação de Felicidade" é bem limitada.

9.5 Terceira Causal Query

Perceba que na figura 20, "PIB per Capita" além de ter sua influência direta em "Pontuação de Felicidade", possui também diversas variáveis como mediadoras, além dele próprio ser um mediador para o nó "Corrupção". Dito isso, um fato importante de se investigar é a influência causal intrínseca de cada nó na "Pontuação de Felicidade", ou seja, a força de influência dos nós sem contar com o que cada um herda de seus pais. Isso é justamente

a terceira causal query que será respondida (Qual a influência causal intrínseca de cada nó na Pontuação de Felicidade?). Para mais detalhes sobre o método, consulte Janzing et al. (2024).

Para descobrir a influência causal intrínseca, usamos as seguintes funções do Código 13:

Código 13: Influência Causal Intrínseca

```
from dowhy.utils import bar_plot

iccs = gcm.intrinsic_causal_influence(causal_model_2,
    target_node='Pontuação de Felicidade', num_samples_randomization=500)
bar_plot(iccs, ylabel='Variance attribution', figure_size=(6, 2))
```

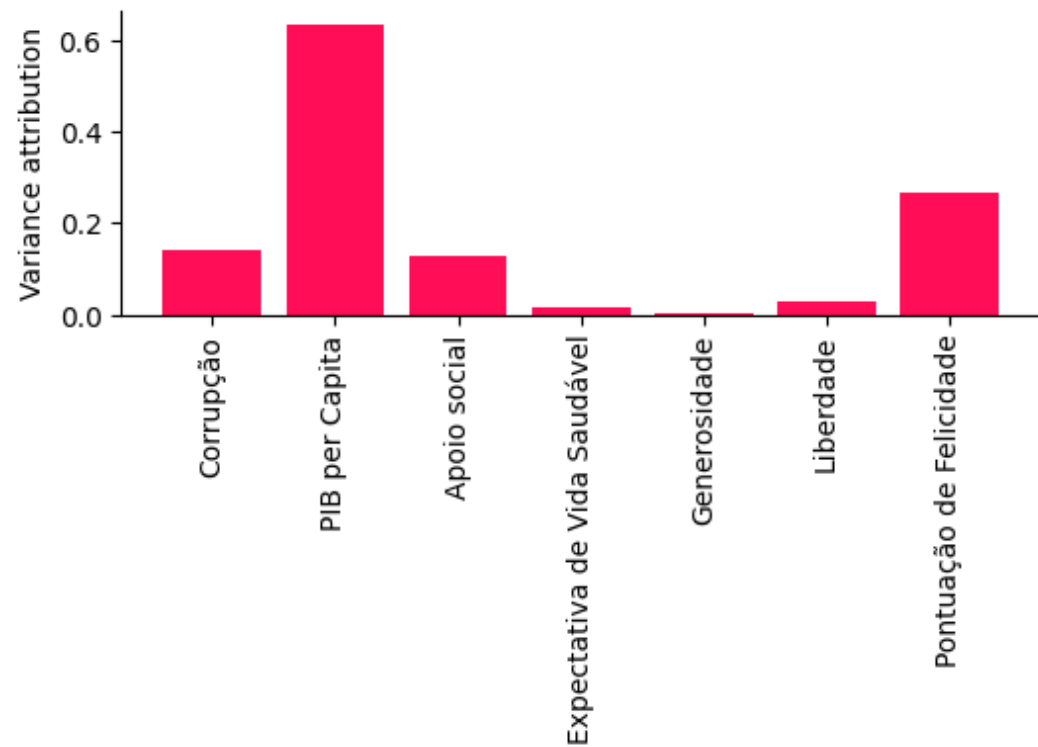


Figura 21: Influência causal intrínseca de cada nó com "Pontuação de Felicidade"

Código 14: Porcentagem da variância de Pontuação de Felicidade que cada variável explica intrinsecamente

```
iccs_percentages = {k: (v / dataset['Pontuação de Felicidade'].var()) *
    100 for k, v in iccs.items()}
print(iccs_percentages)
```

```
{'Corrupção': 11.56259901857358,
'PIB per Capita': 50.708133610646875,
```



```
'Apoio social': 11.159581962786017,  
'Expectativa de Vida Saudável': 1.2087433320901588,  
'Generosidade': 0.09294017264900535,  
'Liberdade': 2.3993318391938434,  
'Pontuação de Felicidade': 22.205226496407597}
```

Conclusões - Terceira Causal Query

Ao observar a Figura 21 pode-se perceber que "PIB per Capita" explica a maior parte da variância da "Pontuação de Felicidade", enquanto que as outras variáveis estão bem mais abaixo. É necessário destacar que, de acordo com os resultados, aproximadamente 22% da variância de "Pontuação de Felicidade" não pode ser explicada pelas variáveis no conjunto de dados disponíveis, indicando que possivelmente existem variáveis confundidoras que não estão no conjunto de dados, ou que existem poucas entradas nele.

9.6 Quarta Causal Query

De acordo com os resultados obtidos anteriormente, foi possível ver que PIB per Capita tem grande influência positiva sobre a Pontuação de Felicidade. Dito isso, a quarta causal query é "O que aconteceria caso ocorresse uma intervenção no PIB per Capita dos países?". É possível simular uma intervenção (Peters et al., 2017) nesse caso utilizando a função `gcm.interventional_samples()` do Código 15.

Código 15: Simulando a intervenção

```
intervention = gcm.interventional_samples(causal_model_2, {'PIB per  
Capita': lambda x: x + 0.2}, num_samples_to_draw=158)
```

O código simula uma intervenção nos dados aumentando o PIB per Capita dos países em 0,2. São extraídas 158 amostras ao todo (mesmo número de amostras no conjunto de dados original.) No Código 16 são calculadas as médias dos nós antes e depois da intervenção.

Código 16: Diferença das médias dos atributos antes e após a intervenção.

```

print("Média da Pontuação de Felicidade: {0:.2f} ->
      {1:.2f}".format(dataset['Pontuação de Felicidade'].mean(),
                      intervention['Pontuação de Felicidade'].mean()))
print("Média do PIB per Capita: {0:.2f} -> {1:.2f}".format(dataset['PIB
per Capita'].mean(), intervention['PIB per Capita'].mean()))
print("Média do Apoio social: {0:.2f} -> {1:.2f}".format(dataset['Apoio
social'].mean(), intervention['Apoio social'].mean()))
print("Média da Expectativa de Vida Saudável: {0:.2f} ->
      {1:.2f}".format(dataset['Expectativa de Vida Saudável'].mean(),
                      intervention['Expectativa de Vida Saudável'].mean()))
print("Média da Liberdade: {0:.2f} ->
      {1:.2f}".format(dataset['Liberdade'].mean(),
                      intervention['Liberdade'].mean()))
print("Média da Corrupção: {0:.2f} ->
      {1:.2f}".format(dataset['Corrupção'].mean(),
                      intervention['Corrupção'].mean()))
print("Média da Generosidade: {0:.2f} ->
      {1:.2f}".format(dataset['Generosidade'].mean(),
                      intervention['Generosidade'].mean()))

```

```

Média da Pontuação de Felicidade: 5.38 -> 6.07
Média do PIB per Capita: 0.85 -> 1.20
Média do Apoio social: 0.99 -> 1.12
Média da Expectativa de Vida Saudável: 0.63 -> 0.82
Média da Liberdade: 0.43 -> 0.47
Média da Corrupção: 0.14 -> 0.14
Média da Generosidade: 0.24 -> 0.26

```

Conclusões - Quarta Causal Query

Assim como esperado, é possível ver um aumento geral na média das colunas, demonstrando o impacto causal positivo que PIB per Capita possui. O impacto mais expressivo ocorreu justamente na Pontuação de Felicidade, mostrando um aumento de aproximadamente 0,69.

10 Segundo Estudo de Caso - Carros

Para o segundo estudo de caso foi utilizado o conjunto de dados *Auto-MPG* (Quinlan, 1993). Esse conjunto de dados baseia-se em diversos veículos e suas características:

- MPG: Eficiência de combustível medida em milhas por galão.
- Cylinders: Número de cilindros no motor.

- Displacement: Cilindrada do motor, indicando seu tamanho ou capacidade medida em polegadas cúbicas.
- Horsepower: Potência do motor em cavalos de potência.
- Weight: Peso do automóvel em libras.
- Acceleration: Tempo necessário para acelerar de 0 a 60 milhas por hora medido em segundos.

As causal queries que serão respondidas são:

1. Qual o impacto médio da cilindrada na eficiência de combustível?
2. Como cada variável afeta a eficiência de combustível?
3. Qual a influência causal intrínseca de cada nó na eficiência de combustível
4. Qual seria a eficiência de combustível de um certo veículo caso seu peso fosse diminuído?

As perguntas foram elaboradas seguindo a mesma lógica do primeiro estudo de caso - as três primeiras são perguntas bem gerais e a última foi escolhida de modo a mostrar as capacidades da biblioteca.

10.1 Pré-processamento

No pré-processamento dos dados foram removidas algumas colunas que não eram necessárias para o estudo de caso (Código 17).

Código 17: Remoção das colunas

```
dataset = dataset.drop(columns=['name', 'origin', 'model_year'])
```

Além disso, os valores da eficiência de combustível, cilindrada e peso foram convertidos para os padrões do Sistema Internacional de Unidades no Código 18.

Código 18: Criação do Grafo Causal

```
def mpg_to_kmpl (mpg) :
```

```

    return mpg * 0.425144

def in3_to_cm3(in3):
    return in3 * 16.387

def lbs_to_kg(lbs):
    return lbs * 0.453592

dataset['mpg'] = dataset['mpg'].apply(mpg_to_kmpl)
dataset['displacement'] = dataset['displacement'].apply(in3_to_cm3)
dataset['weight'] = dataset['weight'].apply(lbs_to_kg)

```

Por último, as colunas restantes foram renomeadas no Código 19 para facilitar o entendimento.

Código 19: Renomeação das colunas

```

dataset = dataset.rename(columns={'mpg': 'Km/l', 'cylinders':
    'Cilindros', 'acceleration': '0_96km/h (s)', 'horsepower': 'Potência
    (CV)', 'displacement': 'Cilindrada (cm3)', 'weight': 'Peso (kg)'})

```

10.2 Criação do Grafo Causal

Foram encontradas dificuldades para achar referências bibliográficas para esse caso de estudo. Muitas das fontes relevantes estavam atrás de *paywalls*¹, o que limitou o acesso a informações detalhadas. Portanto, o grafo causal neste caso foi construído no Código 20 sem referências bibliográficas sólidas, baseando-se principalmente no conhecimento geral sobre automóveis.

Código 20: Criação do Grafo Causal

```

import networkx as nx

G = nx.DiGraph()

nodes = ['Km/l', 'Cilindros', '0_96km/h (s)',
    'Potência (CV)', 'Cilindrada (cm3)', 'Peso (kg)']

G.add_nodes_from(nodes)

edges = [
    ("Cilindros", "Cilindrada (cm3)"),
    ("Cilindros", "Potência (CV)"),
    ("Cilindros", "Km/l"),
    ("Cilindros", "0_96km/h (s)"),
    ("Cilindrada (cm3)", "Peso (kg)"),
    ("Potência (CV)", "Peso (kg)"),
    ("Peso (kg)", "0_96km/h (s)"),
    ("Cilindrada (cm3)", "Km/l"),
    ("Cilindrada (cm3)", "Potência (CV)"),

```

¹Um *paywall* é um sistema que restringe o acesso a conteúdos digitais, como artigos acadêmicos ou livros online, permitindo que apenas usuários pagantes ou assinantes possam visualizar o conteúdo completo.

```

    ("Peso (kg)", "Km/l"),
]

G.add_edges_from(edges)

```

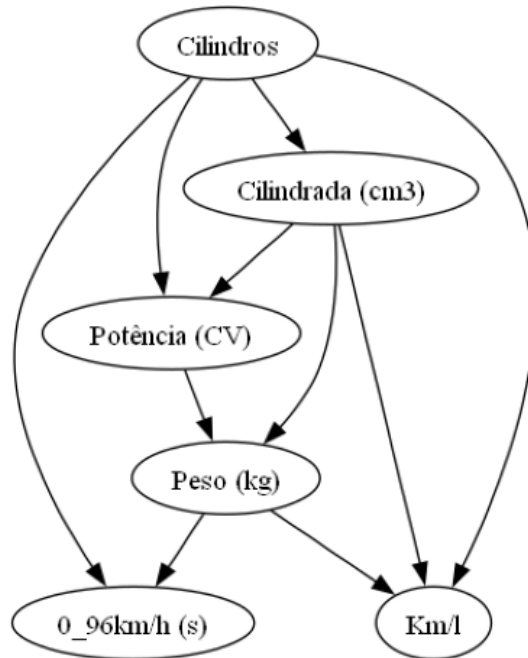


Figura 22: Grafo causal representando as relações das variáveis

10.3 Primeira Causal Query

Para responder a primeira causal query ("Qual o impacto médio da cilindrada na eficiência de combustível?") o passo a passo seguido foi: identificação do estimando, estimação do efeito e, finalmente, teste de robustez do efeito estimado.

Criação do Modelo Causal e Identificação do Estimando

A criação do modelo foi feita no Código 21, enquanto que a identificação do estimando foi feita no Código 22.

Código 21: Criação do modelo causal, especificando o tratamento como a Cilindrada (cm3) e a variável de resultado como a Km/l

```

causal_model = CausalModel(
    data=dataset,
    graph=G,
    treatment='Cilindrada (cm3)',
    outcome='Km/l'
)

```

```
)
causal_model.view_model()
```

Código 22: Identificação do estimando

```
identified_estimand =
    causal_model.identify_effect(proceed_when_unidentifiable=True)
print(identified_estimand)
```

```
Estimand type: EstimandType.NONPARAMETRIC_ATE

### Estimand : 1
Estimand name: backdoor
Estimand expression:
    d
    _____(E[Km/l|Cilindros])
d[Cilindrada (cm3)]
Estimand assumption 1, Unconfoundedness: If  $U \rightarrow \{Cilindrada (cm3)\}$  and  $U \rightarrow Km/l$  then
 $P(Km/l|Cilindrada (cm3), Cilindros, U) = P(Km/l|Cilindrada (cm3), Cilindros)$ 

### Estimand : 2
Estimand name: iv
No such variable(s) found!

### Estimand : 3
Estimand name: frontdoor
No such variable(s) found!
```

O estimando encontrado foi o "backdoor" e o subconjunto de variáveis que satisfazem esse critério é composto apenas pela variável Cilindros. Isso quer dizer que para encontrar o efeito da Cilindrada em Km/l, é necessário condicionar por Cilindros.

Estimação do Efeito

Foi utilizada a técnica de Double Machine Learning no Código 23 para realizar a estimativa, sendo feito um pequeno ajuste fino nos modelos de regressão utilizados, definindo o número de estimadores como 1200 e profundidade máxima igual a 6. Já os parâmetros "treatment_value" e "control_value" são os da relação que estamos estudando, ou seja, queremos verificar qual o valor médio de "Km/l" após aumentar o valor de "Cilindrada"

de 3185 para 3750. Esses valores foram selecionados a partir da média da coluna "Cilindrada", sendo 3185 a média e 3750 uma escolha de aumento arbitrária. Em resumo, queremos ver o que acontece com a eficiência de combustível quando o carro é mais potente.

Código 23: Estimando com Double Machine Learning

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.linear_model import LassoCV

dml_estimate = causal_model.estimate_effect(
    identified_estimand,
    method_name='backdoor.econml.dml.DML',
    method_params={
        'init_params': {
            'model_y': GradientBoostingRegressor(n_estimators=1200,
                                                  max_depth=6),
            'model_t': GradientBoostingRegressor(n_estimators=1200,
                                                  max_depth=6),
            'model_final': LassoCV(fit_intercept=False),
        },
        "fit_params": {}
    },
    control_value=3185,
    treatment_value=3750
)

print(dml_estimate)
```

```
*** Causal Estimate ***

## Identified estimand
Estimand type: EstimandType.NONPARAMETRIC_ATE

### Estimand : 1
Estimand name: backdoor
Estimand expression:
    d
    _____(E[Km/l|Cilindros])
d[Cilindrada (cm3)]
Estimand assumption 1, Unconfoundedness: If  $U \perp \{Cilindrada (cm3)\}$  and  $U \perp Km/l$  then
 $P(Km/l|Cilindrada (cm3), Cilindros, U) = P(Km/l|Cilindrada (cm3), Cilindros)$ 

## Realized estimand
b: Km/l~Cilindrada (cm3)+Cilindros |
Target units: ate

## Estimate
Mean value: -0.6988599237881171
```

```
Effect estimates: [[-0.69885992]]
```

O resultado indica que aumentando o valor da variável de tratamento (Cilindrada) de 3185 para 3750 causa uma redução de aproximadamente -0,69 na variável de resultado (Km/l), sobre a distribuição de dados/população representada pelo conjunto de dados.

Teste de Robustez

Naturalmente, queremos testar a qualidade da nossa estimativa e a robustez do modelo causal criado. Para isso, foram utilizados os seguintes testes no Código 24:

Código 24: Testes de robustez utilizados

```
refute_results_1 = causal_model.refute_estimate(  
    identified_estimand,  
    dml_estimate,  
    method_name='random_common_cause'  
)  
  
refute_results_2 = causal_model.refute_estimate(  
    identified_estimand,  
    dml_estimate,  
    method_name='data_subset_refuter',  
    subset_fraction=0.9  
)  
  
refute_results_3 = causal_model.refute_estimate(  
    identified_estimand,  
    dml_estimate,  
    method_name='placebo_treatment_refuter',  
    placebo_type='permute',  
    num_simulations=5  
)  
  
print(refute_results_1)  
print(refute_results_2)  
print(refute_results_3)
```

```
Refute: Add a random common cause  
Estimated effect:-0.6988599237881171  
New effect:-0.6970560671040105  
p value:1.0  
  
Refute: Use a subset of data  
Estimated effect:-0.6988599237881171  
New effect:-0.7477356078436184  
p value:0.4
```



```
Refute: Use a Placebo Treatment
Estimated effect:-0.6988599237881171
New effect:0.008644310505170262
p_value:0.23975006109347669
```

Perceba que nos dois primeiros testes o valor da estimativa não difere muito, e no tratamento placebo a nova estimativa vai para perto de 0 (resultados positivos). Além disso, nos três testes o valor de p é relativamente alto considerando o valor de referência (0,05), não rejeitando a hipótese nula, portanto mostrando uma certa robustez do efeito original estimado.

Conclusões - Primeira Causal Query

Podemos concluir que, dado esse grafo causal e o conjunto de dados disponíveis, "Cilindrada (cm3)" tem uma relação causal negativa com "Km/l" para o valor de tratamento estudado (3185 para 3750) - resultado que já era esperado, pois carros mais potentes tendem a ser menos eficientes. Além disso, analisando a saída dos testes de robustez, pode-se perceber que a estimativa original se mantém relevante, pois os testes não conseguem refutá-la.

10.4 Segunda Causal Query

Para responder a segunda causal query ("Como cada variável afeta a eficiência de combustível?"), foi criado um Graphical Causal Model no Código 25 e o utilizamos para responder a pergunta.

Código 25: Criação do Graphical Causal Model

```
causal_model_2 = gcm.StructuralCausalModel(G)
auto_assignment_summary =
    gcm.auto.assign_causal_mechanisms(causal_model_2, dataset)
gcm.fit(causal_model_2, dataset)
```

Foi verificada a qualidade dos mecanismos causais atribuídos e do grafo causal no Código 26.

Código 26: Função para avaliar o modelo

```
print(gcm.evaluate_causal_model(causal_model_2, dataset))
```

Evaluated the performance of the causal mechanisms and the invertibility assumption of the causal mechanisms and the overall average KL divergence between generated and observed distribution and the graph structure. The results are as follows:

==== Evaluation of Causal Mechanisms ====

The used evaluation metrics are:

- KL divergence (only for root-nodes): Evaluates the divergence between the generated and the observed distribution.
- Mean Squared Error (MSE): Evaluates the average squared differences between the observed values and the conditional expectation of the causal mechanisms.
- Normalized MSE (NMSE): The MSE normalized by the standard deviation for better comparison.
- R2 coefficient: Indicates how much variance is explained by the conditional expectations of the mechanisms. Note, however, that this can be misleading for nonlinear relationships.
- F1 score (only for categorical non-root nodes): The harmonic mean of the precision and recall indicating the goodness of the underlying classifier model.
- (normalized) Continuous Ranked Probability Score (CRPS): The CRPS generalizes the Mean Absolute Percentage Error to probabilistic predictions. This gives insights into the accuracy and calibration of the causal mechanisms.

NOTE: Every metric focuses on different aspects and they might not consistently indicate a good or bad performance.

We will mostly utilize the CRPS for comparing and interpreting the performance of the mechanisms, since this captures the most important properties for the causal model.

--- Node Cilindros

- The KL divergence between generated and observed distribution is 0.7352601343814152.

The estimated KL divergence indicates a good representation of the data distribution, but might indicate some smaller mismatches between the distributions.

--- Node Cilindrada (cm3)

- The MSE is 1044.4110895233946.
- The NMSE is 0.31170898764829935.
- The R2 coefficient is 0.9020035162196101.
- The normalized CRPS is 0.1743013961262155.

The estimated CRPS indicates a very good model performance.

--- Node Potência (CV)

- The MSE is 209.58636806231743.
- The NMSE is 0.3805675963835672.
- The R2 coefficient is 0.8527962549449253.
- The normalized CRPS is 0.2037887011409544.

The estimated CRPS indicates a good model performance.

--- Node Peso (kg)

- The MSE is 74171.0770529049.
- The NMSE is 0.32238730815166844.
- The R2 coefficient is 0.8953557899623084.
- The normalized CRPS is 0.17664858711933157.

The estimated CRPS indicates a very good model performance.

--- Node 0_96km/h (s)

- The MSE is 5.243354231627333.
- The NMSE is 0.834116265576472.
- The R2 coefficient is 0.2999428013866431.
- The normalized CRPS is 0.46421006140812004.

The estimated CRPS indicates only a fair model performance. Note, however, that a high CRPS could also result from a small signal to noise ratio.

--- Node Km/l

- The MSE is 3.1557797854893876.
- The NMSE is 0.5382065011597058.
- The R2 coefficient is 0.7090750139111679.
- The normalized CRPS is 0.29196714839624577.

The estimated CRPS indicates a good model performance.

==== Evaluation of Invertible Functional Causal Model Assumption ====

--- The model assumption for node Km/l is rejected with a p-value of 3.04629617503549e-05 (after potential adjustment) and a significance level of 0.05.

This implies that the model assumption might not be valid. This is, the relationship cannot be represent with this type of mechanism or there is a hidden confounder between the node and its parents.

--- The model assumption for node 0_96km/h (s) is not rejected with a p-value of 1.0 (after potential adjustment) and a significance level of 0.05.

This implies that the model assumption might be valid.

--- The model assumption for node Potência (CV) is not rejected with a p-value of 1.0 (after potential adjustment) and a significance level of 0.05.

This implies that the model assumption might be valid.

```

--- The model assumption for node Cilindrada (cm3) is not rejected with a p-value
of 1.0 (after potential adjustment) and a significance level of 0.05.
This implies that the model assumption might be valid.

--- The model assumption for node Peso (kg) is not rejected with a p-value of 1.0
(after potential adjustment) and a significance level of 0.05.
This implies that the model assumption might be valid.

Note that these results are based on statistical independence tests, and the
fact that the assumption was not rejected does not necessarily imply that it
is correct. There is just no evidence against it.

==== Evaluation of Generated Distribution ====
The overall average KL divergence between the generated and observed
distribution is 0.9928170019612677
The estimated KL divergence indicates a good representation of the data
distribution, but might indicate some smaller mismatches between the distributions
.

==== Evaluation of the Causal Graph Structure ====
+-----+
|                                     Falsificaton Summary                                     |
+-----+
| The given DAG is informative because 0 / 50 of the permutations lie in the
Markov |
| equivalence class of the given DAG (p-value: 0.00). |
| The given DAG violates 3/6 LMCs and is better than 92.0 percent of the permuted
DAGs (p-value: 0.08).|
| Based on the provided significance level (0.2) and because the DAG is
informative, we do not reject the DAG.|
+-----+

==== NOTE ====
Always double check the made model assumptions with respect to the graph
structure and choice of causal mechanisms.
All these evaluations give some insight into the goodness of the causal model,
but should not be overinterpreted, since some causal relationships can be
intrinsically hard to model. Furthermore, many algorithms are fairly robust
against misspecifications or poor performances of causal mechanisms.

```

Os resultados mostram um desempenho geral bom, representando bem a distribuição dos dados. As previsões para "Cilindrada (cm3)", "Potên-

cia (CV)” e ”Peso (kg)” foram bastante precisas, indicando que os modelos para essas variáveis estão bem ajustados. A previsão para ”0_96km/h (s)” foi menos precisa, sugerindo dificuldades em modelar essa variável. A suposição de que o modelo pode ser invertido foi rejeitada apenas para ”Km/l”, indicando possíveis limitações ou variáveis ocultas que não foram consideradas. A estrutura geral do grafo causal foi considerada adequada, representando bem as relações causais entre as variáveis.

Respondendo a pergunta

Utilizando o algoritmo de calculo de força das arestas no Código 27.

Código 27: Arrow Strengths

```
from dowhy.utils import plot

arrow_strengths = gcm.arrow_strength(causal_model_2, target_node='Km/l')
plot(causal_graph=causal_graph, causal_strengths=arrow_strengths,
     figure_size=(10, 10),)
```

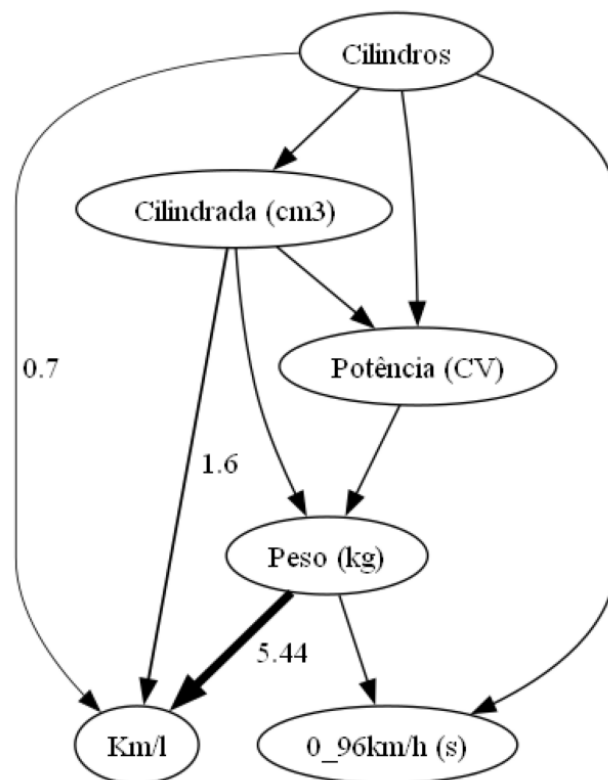


Figura 23: Grafo causal com a força de cada conexão com a variável Km/l

No Código 28 é calculada a variância de Km/l, e no Código 29 a porcen-

tagem da variância de Km/l que cada nó contribui.

Código 28: Variância de Km/l

```
dataset['Km/l'].var()
```

```
11.0107970182084
```

Código 29: Porcentagem da variância de Km/l que cada variável contribui

```
arrow_strengths_percentages = {k: (v / dataset['Km/l'].var()) * 100 for  
    k, v in arrow_strengths.items()}  
print(arrow_strengths_percentages)
```

```
{('Cilindrada (cm3)', 'Km/l'): 14.577345060134162,  
( 'Cilindros', 'Km/l'): 6.3863692407713835,  
( 'Peso (kg)', 'Km/l'): 49.454405125746774}
```

Conclusões - Segunda Causal Query

Perceba que, de acordo com o algoritmo, a variável com mais influência sobre "Km/l" é "Peso (kg)", com "Cilindrada (cm3)" em segundo e "Cilindros" vindo logo após.

10.5 Terceira Causal Query

Da mesma forma que o primeiro caso de estudo, existem diversas influências indiretas no modelo, sendo necessário verificá-las para entender melhor o que afeta a eficiência de combustível (terceira causal query; qual a influência causal intrínseca de cada nó na eficiência de combustível). Para isso, foram utilizadas as funções do Código 30.

Código 30: Calculo da Influência causal intrínseca de cada nó

```
iccs = gcm.intrinsic_causal_influence(causal_model_2, 'Km/l')  
print(iccs)  
  
from dowhy.utils import bar_plot  
  
bar_plot(iccs, ylabel='Variance attribution', figure_size=(6, 2))
```

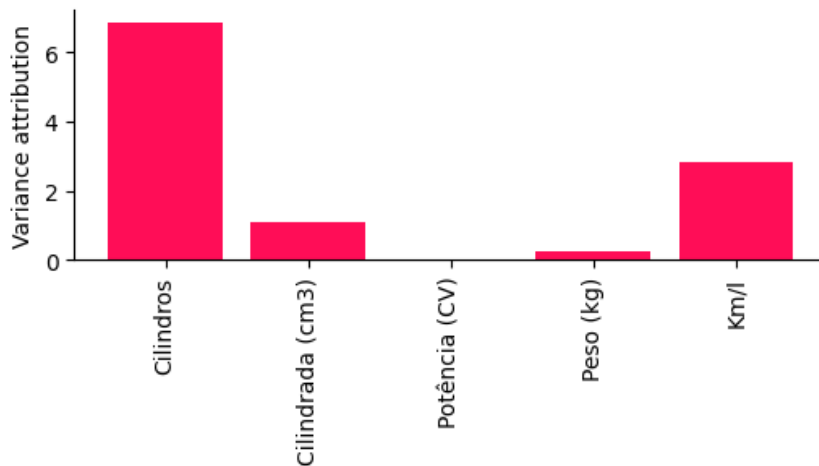


Figura 24: Influência causal intrínseca de cada nó

No Código 31 é calculada a porcentagem da variância de Km/l que cada nó contribui intrinsecamente.

Código 31: Porcentagem da variância de Km/l que cada variável explica intrinsecamente

```
iccs_percentages = {k: (v / dataset['Km/l'].var()) * 100 for k, v in
    iccs.items()}
print(iccs_percentages)
```

```
{'Cilindros': 62.442742766501766,
'Cilindrada (cm3)': 9.928501388375386,
'Potência (CV)': 0.1497300898575058,
'Peso (kg)': 2.159443793223195,
'Km/l': 25.84833432357178}
```

Conclusões - Terceira Causal Query

Perceba que "Cilindros" contribui para a maior parte da variância de "Km/l" considerando as influências diretas e indiretas do modelo; enquanto que em segundo lugar vem o ruído da própria variável. Por fim "Cilindrada" e "Peso" mostram uma pequena contribuição.

10.6 Quarta Causal Query

Um outro fato interessante de se investigar é a eficiência de combustível de um certo carro caso, por exemplo, ele fosse mais leve. Para responder a isso, é possível computar o contrafactual (algo que não aconteceu) dele

(quarta causal query; "qual seria a eficiência de combustível de um certo veículo caso seu peso fosse diminuído?").

No Código 32 é retirada uma instância aleatória do conjunto de dados.

Código 32: Pegando uma amostra aleatória do dataset

```
row = dataset.sample(1)

for column in row.columns:
    print(f"{column}: {row[column].values[0]}")
```

```
Cilindros: 4
Cilindrada (cm3): 1769.796
Potência (CV): 94.0
Peso (kg): 1079.095368
0_96km/h (s): 16.5
Km/l: 9.353168
```

O peso desse carro é 1079kg. O que aconteceria caso o peso dele fosse 1000kg mantendo suas outras características fixas? Para calcular isso, utilizamos a função `gcm.counterfactual_samples()` no Código 33.

Código 33: Computando o contrafactual

```
counterfactual = gcm.counterfactual_samples(causal_model_2, {'Peso (kg)': lambda x: 1000}, observed_data=row)
display(counterfactual)
```

```
Cilindros: 4
Cilindrada (cm3): 1769.796
Potência (CV): 94.0
Peso (kg): 1000
0_96km/h (s): 14.883473589890198
Km/l: 9.87273630129303
```

Conclusões - Quarta Causal Query

Com exceção de "0_96km/h (s)" e "Km/l", todas as outras variáveis se mantêm com valor fixo, pois o peso não as afeta. Veja que reduzindo o peso desse carro em 79kg, ele ganhou eficiência de combustível (aumento de aproximadamente 0,52km/l) e ganhou aceleração, conseguindo ir de 0 a 96km/h em menos tempo (redução de aproximadamente 1,62 segundos).

11 Considerações Finais

11.1 Comparação com o objetivo inicial

O trabalho teve dois objetivos principais: explorar como o aprendizado de máquina pode ser utilizado na realização de inferência causal, abordando os métodos e algoritmos mais utilizados para tal; e, em seguida, escolher conjuntos de dados reais em que fosse possível realizar análises causais usando o conteúdo estudado, com apoio de especialistas.

Em relação ao primeiro objetivo, foi realizada uma revisão abrangente dos métodos e algoritmos de aprendizado de máquina mais utilizados para inferência causal. Isso incluiu o estudo de técnicas como florestas causais, aprendizado duplo de máquina e aprendizado duplamente robusto. Para cada método foi descrito seus principais componentes e como cada um realiza a estimativa do efeito causal. Vale destacar que, havia expectativa de existir mais material sobre a utilização de aprendizado de máquina na exploração causal, no entanto, durante a elaboração do trabalho não foram encontradas muitas referências nesse sentido.

Para atingir o segundo objetivo, foram selecionados conjuntos de dados reais onde fosse possível aplicar as técnicas de inferência causal estudadas. Dois casos de estudo foram escolhidos: um relacionado à análise de felicidade e outro relacionado a dados de automóveis. Em ambos os casos, foram realizadas análises causais detalhadas, aplicando os métodos estudados para identificar relações causais e estimar efeitos causais relevantes. Esses casos de estudo permitiram demonstrar a aplicação prática das técnicas de inferência causal em contextos diferentes e demonstraram a utilidade dessas técnicas.

No entanto, um dos objetivos era colaborar com especialistas de dois domínios específicos para desenvolver hipóteses mais fundamentadas e robustas. Infelizmente, essa colaboração não foi possível devido à dificuldade em encontrar conjuntos de dados adequados para os casos de estudo

e à falta de tempo, uma vez que se trata de um trabalho de conclusão de curso. Essa limitação impactou a profundidade das hipóteses desenvolvidas e a robustez das análises causais realizadas.

Em resumo, o objetivo inicial de explorar o uso do aprendizado de máquina na inferência causal foi amplamente cumprido. Foram abordados e implementados métodos e algoritmos de forma detalhada, e análises causais foram realizadas em dois conjuntos de dados reais. A principal limitação foi a impossibilidade de colaboração com especialistas, o que teria enriquecido ainda mais as análises realizadas.

11.2 Contribuições deste trabalho

Este trabalho serve como um guia introdutório para qualquer pessoa que deseja começar a aplicar os conceitos e técnicas de investigação de causalidade em seus conjuntos de dados, sejam eles reais ou apenas como forma de aprendizado. As principais contribuições deste trabalho incluem:

- Apresentação da biblioteca DoWhy, destacando suas funcionalidades e como ela pode ser utilizada para realizar inferências causais de maneira sistemática e robusta.
- Apresentação de como construir grafos causais, demonstrando a modelagem das relações entre diferentes características de um conjunto de dados.
- Aplicação prática em conjuntos de dados reais, estimando efeitos, testando robustez de estimativas, calculando influência causal direta e intrínseca, simulando intervenções e contrafactuais.

11.3 Dificuldades

Na construção do trabalho foram encontradas algumas barreiras que impactaram a eficiência e a compreensão. As principais dificuldades enfrentadas são detalhadas a seguir:

- Documentação confusa da biblioteca DoWhy: A documentação é dispersa e carece de coesão, dificultando a compreensão e a utilização das funcionalidades da ferramenta. Algumas explicações são encontradas diretamente na documentação das funcionalidades, enquanto outras estão em exemplos aleatórios.
- Muita estatística e matemática: uma das dificuldades foi entender toda a matemática e estatística envolvida para poder compreender os resultados e as funcionalidades corretamente.
- Funcionalidades bem exigentes de poder computacional: por exemplo, para rodar os testes de robustez com os conjuntos de dados pequenos (200 a 300 entradas), demorou em média 6 minutos, o que demonstra a necessidade de recursos computacionais consideráveis.

11.4 Oportunidades para trabalhos futuros

- Colaboração Interdisciplinar: Embora este trabalho não tenha conseguido, seria interessante testar parcerias com especialistas de diferentes áreas para desenvolver hipóteses mais fundamentadas e validar a utilidade dos resultados.
- Contribuir para o desenvolvimento das ferramentas: As ferramentas citadas (DoWhy, EconML, CausalML) fazem parte de um ecossistema open-source de bibliotecas para aprendizado de máquina causal chamado PyWhy.
- Estudos de Caso em Novas Áreas: Aplicar os conceitos e métodos de investigação causal em novos domínios, como saúde, economia, educação, entre outros, para explorar as relações causais específicas de cada área.

Bibliografia

- Abrevaya, J., Hsu, Y.-C., and Lieli, R. P. (2015). **Estimating conditional average treatment effects**. *Journal of Business & Economic Statistics*, 33(4):485–505.
- Angrist, J. and Imbens, G. (1995). **Identification and estimation of local average treatment effects**.
- Aristóteles (2002). **Metafísica**. Loyola, São Paulo.
- Athey, S. and Wager, S. (2019). **Estimating treatment effects with causal forests: An application**. *Observational studies*, 5(2):37–51.
- Battocchi, K., Dillon, E., Hei, M., Lewis, G., Oka, P., Oprescu, M., and Syrgkanis, V. (2019). **EconML: A Python package for ML-Based heterogeneous treatment effects estimation**. *Version 0. x*.
- Blöbaum, P., Götz, P., Budhathoki, K., Mastakouri, A. A., and Janzing, D. (2022). **DoWhy-GCM: An extension of DoWhy for causal inference in graphical causal models**. *arXiv preprint arXiv:2206.06821*.
- Blöbaum, P., Budhathoki, K., and Götz, P. (2023). **Root Cause Analysis with DoWhy: An Open Source Python Library for Causal Machine Learning**. <https://aws.amazon.com/pt/blogs/opensource/root-cause-analysis-with-dowhy-an-open-source-python-library-for-causal-machine-learning/>. Acessado em: 19 de abril de 2024.
- Cao, L. (2016). **Data Science: A Comprehensive Overview**. *ACM Computing Surveys*, 1(1):8.
- Chen, H., Harinen, T., Lee, J.-Y., Yung, M., and Zhao, Z. (2020). **Causalm1: Python package for causal machine learning**. *arXiv preprint arXiv:2002.11631*.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W.

- (2017). **Double/debiased/neyman machine learning of treatment effects**. *American Economic Review*, 107(5):261–265.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). **Bert: Pre-training of deep bidirectional transformers for language understanding**. *arXiv preprint arXiv:1810.04805*.
- Diener, E., Diener, M., and Diener, C. (1995). **Factors predicting the subjective well-being of nations**. *Journal of personality and social psychology*, 69(5):851.
- Eulig, E., Mastakouri, A. A., Blöbaum, P., Hardt, M., and Janzing, D. (2023). **Toward Falsifying Causal Graphs Using a Permutation-Based Test**.
- Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., and Davidian, M. (2011). **Doubly robust estimation of causal effects**. *American journal of epidemiology*, 173(7):761–767.
- Gallup (2015). **World Happiness Report 2015**. <https://worldhappiness.report/ed/2015/>. Disponível em: <<https://worldhappiness.report/ed/2015/>>. Acesso em: 07 mai. 2024.
- Greenland, S. (2000). **An introduction to instrumental variables for epidemiologists**. *International Journal of Epidemiology*, 29(4):722–729.
- Gurney, K. (2018). **An introduction to neural networks**. CRC press.
- Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., Ray, S., Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., et al. (2021). **An introduction to structural equation modeling**. *Partial least squares structural equation modeling (PLS-SEM) using R: a workbook*, pages 1–29.
- Huff, D. (1954). **How to lie with statistics**. W.W. Norton & Company.
- Janzing, D., Balduzzi, D., Grosse-Wentrup, M., and Schölkopf, B. (2013). **Quantifying causal influences**.
- Janzing, D., Blöbaum, P., Mastakouri, A. A., Faller, P. M., Minorics, L., and Budhathoki, K. (2024). **Quantifying intrinsic causal contributions via structure preserving interventions**. In *International Conference on Artificial Intelligence and Statistics*, pages 2188–2196. PMLR.
- Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., and Ma, Y. (2016). **Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment**. In

- Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14*, pages 37–48. Springer.
- Merriam-Webster (2023a). **Causality**.
- Merriam-Webster (2023b). **Correlation**.
- Mitchell, T. M. (1997). **Machine Learning**. McGraw-Hill Science/Engineering/Math.
- Müller, A. and Guido, S. (2017). **Introduction to Machine Learning with Python: A Guide for Data Scientists**. O'Reilly Media.
- Niu, Y., Fan, Y., Gao, Y., and Li, Y. (2023). A causal inference method for improving the design and interpretation of safety research. *Safety science*, 161:106082.
- Pearl, J. (2009). **Causality**. Cambridge university press.
- Pearl, J. (2010). **Causal inference**. *Causality: objectives and assessment*, pages 39–58.
- Pearl, J. (2011). **Bayesian networks**.
- Pearl, J. (2012). **The do-calculus revisited**. *arXiv preprint arXiv:1210.4852*.
- Peters, J., Janzing, D., and Schölkopf, B. (2017). **Elements of causal inference: foundations and learning algorithms**. The MIT Press.
- Quinlan, R. (1993). **Auto MPG**. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5859H>.
- Rubin, D. B. (2005). **Causal inference using potential outcomes: Design, modeling, decisions**. *Journal of the American Statistical Association*, 100(469):322–331.
- Sharma, A. and Kiciman, E. (2020). **DoWhy: An end-to-end library for causal inference**. *arXiv preprint arXiv:2011.04216*.
- Tay, L., Kuykendall, L., and Diener, E. (2015). **Satisfaction and happiness—the bright side of quality of life**. *Global handbook of quality of life: Exploration of well-being of nations and continents*, pages 839–853.
- University, N. (2023). **Hypothesis Testing**. Disponível em: <<https://resources.nu.edu/statsresources/hypothesis>>. Acesso em: 03 mai. 2024.
- Yu, D. and Deng, L. (2016). **Automatic speech recognition**, volume 1. Springer.