

Geração de Recomendações Contrafactuais com Validação de Consistência para Melhoria da Eficiência de Condução em Frotas

Miguel Francisco¹, Carnot Braun¹, Bruno Bigatto², Eduardo Martinho²,
Leandro Villas¹, Allan M. de Souza¹

¹Instituto de Computação – Universidade Estadual de Campinas (IC - UNICAMP)

²MagisDev

{m218427, c255785}@dac.unicamp.br

{lvillas, allanms}@unicamp.br

Abstract. *Driving behavior directly impacts fuel consumption and emissions in fleet operations. While machine learning models are effective in classifying efficiency from telemetry data, they do not provide actionable guidance for improvement. This work proposes a framework that combines supervised learning, counterfactual explanations, and topographic-similarity filtering to generate actionable recommendations. Counterfactuals are generated using DiCE, restricted to behavioral features, and filtered to retain only recommendations whose route is topographically comparable to the original trip. Out of 395 candidate counterfactuals produced over eight intra-segment efficiency transitions, 165 were retained after topographic filtering, corresponding to a global retention rate of 41.77%, with per-transition rates ranging from 26.67% to 56.00%. The method enables not only the identification of inefficient driving patterns but also the generation of practical and domain-consistent improvements.*

Resumo. *O comportamento de condução impacta diretamente o consumo de combustível e as emissões em operações de frotas. Embora modelos de aprendizado de máquina sejam eficazes na classificação da eficiência a partir de dados de telemetria, eles não fornecem orientações práticas para melhoria. Este trabalho propõe um framework que combina aprendizado supervisionado, contrafactuais e filtragem por semelhança topográfica para gerar recomendações acionáveis. Os contrafactuais são gerados com DiCE, restritos a variáveis comportamentais, e filtrados para manter apenas recomendações cuja rota seja topograficamente comparável à viagem original. Dos 395 contrafactuais candidatos produzidos nas oito transições de eficiência intra-rote, 165 foram retidos após a filtragem topográfica, correspondendo a uma taxa de retenção global de 41,77%, com taxas por transição entre 26,67% e 56,00%. O método proposto permite não apenas identificar padrões ineficientes, mas também sugerir melhorias práticas e coerentes com o domínio do problema.*

1. Introdução

O setor de transporte desempenha um papel significativo nas emissões de dióxido de carbono (CO_2) e no agravamento das mudanças climáticas [Braun et al. 2025]. No Brasil, a forte dependência do modal rodoviário na cadeia logística intensifica esse cenário, uma vez que grandes frotas de veículos de médio e grande porte, majoritariamente movidos

a diesel, representam uma parcela expressiva do impacto ambiental associado à queima de combustíveis fósseis [Ritchie et al. 2023]. Nesse contexto, complementar a estratégias macro de gerência de tráfego e re-roteamento ciente de contexto [de Souza et al. 2017, de Souza et al. 2020], a gestão dos perfis de condução emerge como fator determinante, com relação direta com a eficiência energética do veículo, os custos operacionais e as emissões de CO_2 .

Abordagens baseadas em *Machine Learning (ML)* têm sido amplamente empregadas na classificação do comportamento dos motoristas a partir de dados de telemetria. Entretanto, embora modelos como redes neurais e *ensembles* apresentem elevado desempenho preditivo, eles operam como *caixas-pretas* e, mesmo apoiados por métodos de XAI como LIME e SHAP, limitam-se a explicações descritivas, sem fornecer orientações acionáveis sobre como melhorar o comportamento de condução. Nesse contexto, abordagens contrafactuais surgem como alternativa ao identificar modificações mínimas capazes de alterar a decisão do modelo, aproximando explicabilidade de ação.

Diferentemente de outras abordagens, o *framework* proposto restringe as recomendações a variáveis comportamentais efetivamente controláveis pelo condutor (como esforço de subida, uso de marcha reduzida e freio motor) e incorpora uma etapa de filtragem por semelhança topográfica entre a rota da instância original e as rotas dos contrafactuais recuperados, retendo apenas recomendações compatíveis em extensão e perfil de relevo. Dessa forma, o principal diferencial deste trabalho reside na integração entre modelagem preditiva, geração de recomendações contrafactuais e filtragem por comparabilidade topográfica, proporcionando não apenas a identificação de condutores ineficientes, mas também a definição de trajetórias concretas e personalizadas para melhoria do desempenho energético.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os principais trabalhos relacionados. A Seção 3 detalha a metodologia proposta. A Seção 4 apresenta e discute os resultados experimentais. Por fim, a Seção 5 apresenta as conclusões e direções para trabalhos futuros.

2. Trabalhos Relacionados

Recentemente, diversos estudos têm explorado as capacidades de modelos de IA na construção de perfis distintos de condutores [Payyanadan and Angell 2022], por meio do desenvolvimento de modelos preditivos e da aplicação de métodos de clusterização para segmentar diferentes estilos de direção. A partir dessa segmentação, tais trabalhos investigam também estratégias de recomendação voltadas à seleção de comportamentos mais eficientes, com o objetivo de otimizar o desempenho energético e operacional [Nasr Azadani and Boukerche 2022, Tselentis and Papadimitriou 2023].

Nesse contexto, a classificação do comportamento do condutor desempenha papel fundamental na gestão de frotas e na eficiência energética em redes veiculares. Bouhsissin et al. [Bouhsissin et al. 2023] evidenciam que fatores humanos estão associados a até 95% dos incidentes viários, motivando o uso de técnicas de aprendizado não supervisionado para segmentação de condutores. Mohammadnazar et al. [Mohammadnazar et al. 2024] aplicam métodos como *k-medoids* e agrupamento hierárquico em dados naturalísticos, ou seja, dados coletados durante a operação real dos veículos, sem controle experimental, identificando perfis agressivos associados a aumentos de até 23% no consumo de combustível. Ashqar et al. [Ashqar et al. 2024] utilizam *Random Forest* combinado com técnicas de clusterização para analisar o impacto da variabilidade da condução no con-

sumo, enquanto Ping *et al.* [Ping et al. 2019] combinam *spectral clustering* e redes LSTM para modelagem microscópica do comportamento veicular. No contexto de veículos pesados, Mumcuoglu *et al.* [Mumcuoglu et al. 2023] demonstram que a classificação precisa da eficiência requer o isolamento explícito de variáveis exógenas, como carga e inclinação da via.

A transição da análise descritiva para a intervenção ocorre por meio de sistemas de recomendação, frequentemente apoiados por técnicas de XAI. Vyas *et al.* [Vyas et al. 2021] propõem o sistema *DriveBFR*, que utiliza Modelos Ocultos de Markov e otimização multiobjetivo para recomendar limites de velocidade. Alam *et al.* [Alam et al. 2025] integram métodos como *SHAP* e *LIME* em redes neurais profundas para quantificar o impacto de variáveis dinâmicas nas emissões de CO_2 . De forma análoga, Braun *et al.* [Braun et al. 2025] aplicam *LIME* para avaliar a relevância de atributos em modelos de previsão de emissões. Apesar desses avanços, a maioria das abordagens baseadas em XAI permanece limitada a explicações retrospectivas, isto é, à identificação das variáveis mais influentes na decisão do modelo, sem fornecer direções explícitas de ação para modificar o comportamento do sistema.

Nesse cenário, métodos baseados em explicações contrafactuais têm emergido como uma alternativa promissora, ao fornecer recomendações acionáveis capazes de alterar a decisão do modelo [Mothilal et al. 2020]. Diferentemente das abordagens tradicionais, tais métodos permitem identificar modificações mínimas nas variáveis de entrada que resultam em mudanças desejadas na saída do modelo.

Paralelamente, técnicas de redução de dimensionalidade como UMAP [McInnes et al. 2018], combinadas com modelos substitutos locais (e.g., kNN), têm sido empregadas para visualizar regiões de decisão de modelos complexos preservando a estrutura local dos dados.

A Tabela 1 sintetiza os principais trabalhos analisados. Observa-se que, embora existam avanços na classificação de perfis, recomendação e interpretabilidade, há uma lacuna significativa na integração de recomendações contrafactuais com restrições de domínio que assegurem a comparabilidade contextual das recomendações e visualização do espaço de decisão do modelo.

Tabela 1. Comparação entre trabalhos relacionados na análise de condução, recomendação e explicabilidade.

Trabalho	Class.	Rec.	XAI	CF	Top.	Vis.
Mohammadnazar et al. (2024)	✓	✗	✗	✗	✗	✗
Mumcuoglu et al. (2023)	✓	✗	✗	✗	✓	✗
Ping et al. (2019)	✓	✗	✗	✗	✗	✗
Vyas et al. (2021)	✓	✓	✗	✗	✗	✗
Alam et al. (2025)	✓	✗	✓	✗	✗	✗
Braun et al. (2025)	✓	✗	✓	✗	✗	✗
Mothilal et al. (2020)	✗	✓	✓	✓	✗	✗
Este trabalho	✓	✓	✓	✓	✓	✓

Class.: Classificação Rec.: Recomendação CF: Contrafactual Top.: Restrições de Domínio
(Comparabilidade Topográfica) Vis.: Visualização do Espaço de Decisão

Diante dessas limitações, este trabalho propõe um *framework* integrado que combina classificação supervisionada, geração de explicações contrafactuais e imposição de restrições de domínio baseadas em semelhança topográfica entre rotas, garantindo a comparabilidade contextual das recomendações. Adicionalmente, introduz-se uma abordagem de visualização baseada em projeção UMAP e aproximação local das fronteiras de decisão por meio de um modelo kNN treinado sobre as predições do modelo original. Essa estratégia permite analisar, de forma interpretável, o deslocamento das instâncias no espaço de decisão, evidenciando não apenas quais variáveis são relevantes, mas também como modificá-las para alcançar estados mais eficientes de condução.

3. Descrição do *Framework*

Esta seção descreve o processo adotado neste trabalho, desde o tratamento dos dados até a preparação das entradas utilizadas nos modelos de aprendizado de máquina. O fluxo foi estruturado para permitir o processamento de múltiplas bases de telemetria, a construção de variáveis representativas do comportamento de condução e a definição de um critério consistente de eficiência energética. A visão geral do *framework* proposto é apresentada na Figura 1.

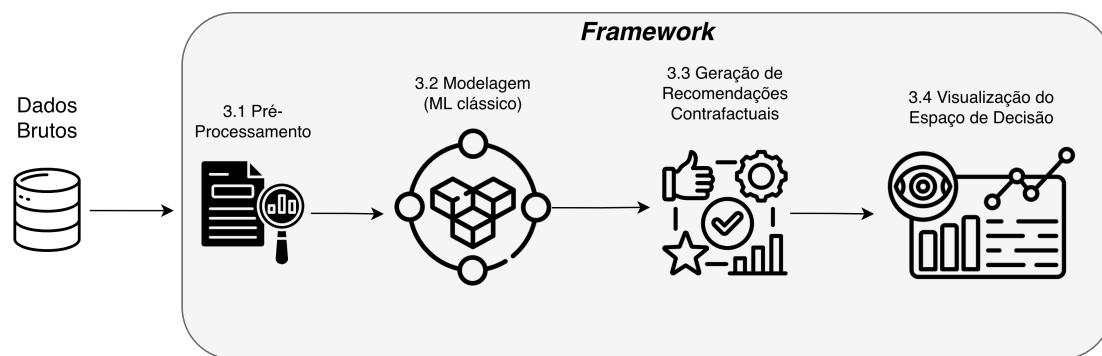


Figura 1. Visão geral do fluxo metodológico adotado no trabalho.

3.1. Pré-processamento

Os dados utilizados neste trabalho são provenientes de telemetria veicular coletada em diferentes tipos de operação (rodoviária, urbana e logística), disponibilizados no contexto de uma parceria com a empresa MagisDev. Para analisar a estrutura dos dados, foi aplicada uma técnica de redução de dimensionalidade, projetando as instâncias em um espaço bidimensional.

A Figura 2 mostra que as classes apresentam uma organização espacial estruturada, com regiões relativamente bem separadas. Esse comportamento sugere que as características extraídas são informativas para a distinção entre perfis de eficiência, justificando o uso de modelos de aprendizado supervisionado.

Cada registro da base representa uma rota individual, definida como o intervalo entre o início e o término da operação do veículo, incluindo tempos de operação do motor, distâncias percorridas e consumo de combustível. Por questões de confidencialidade, variáveis, identificadores e fontes externas foram anonimizados.

A eficiência de condução foi definida de forma relativa ao tipo de veículo, evitando o uso de limiares absolutos. Para cada rota, calcula-se inicialmente a eficiência real

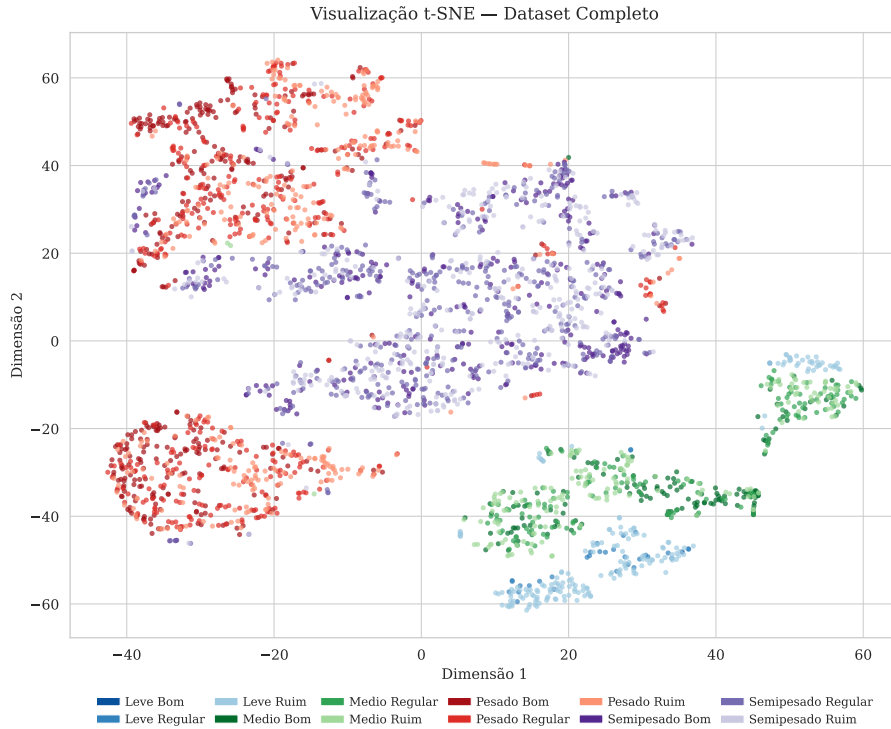


Figura 2. Projeção bidimensional do conjunto de dados utilizando t-SNE. As cores indicam as classes de eficiência e porte do veículo.

como a razão entre distância percorrida e consumo de combustível. Esse valor é então comparado a uma referência esperada para o respectivo perfil de veículo, resultando em uma medida de eficiência relativa.

Para a formulação da variável alvo, o problema foi modelado considerando a combinação entre o porte do veículo (categorizado por um índice $p \in \{0, 1, 2, 3\}$, representando, respectivamente, Leve, Médio, Pesado e Semi-Pesado) e o nível de eficiência relativa da rota (x). Primeiramente, define-se o nível de desempenho $E(x) \in \{0, 1, 2\}$ por meio de uma função baseada em intervalos:

$$E(x) = \begin{cases} 0 \text{ (Bom)} & \text{se } x \geq \tau_1 \\ 1 \text{ (Regular)} & \text{se } \tau_2 \leq x < \tau_1 \\ 2 \text{ (Ruim)} & \text{se } x < \tau_2 \end{cases} \quad \forall x \in X$$

em que τ_1 e τ_2 são limiares operacionais anonimizados por questões de confidencialidade.

Em seguida, a classe final de eficiência $C(x, p)$ é definida matematicamente por meio do produto cartesiano entre as categorias, mapeado para um único inteiro de 0 a 11:

$$C(x, p) = 3p + E(x)$$

Dessa forma, o modelo engloba 12 classes únicas, permitindo ao algoritmo de classificação distinguir comportamentos específicos, como "Leve Ruim" ($C = 2$) ou "Médio Regular" ($C = 4$).

As variáveis originais de telemetria apresentam alta correlação, principalmente por dependerem diretamente da duração da rota. Para reduzir esse efeito, foram construídas novas características baseadas em relações proporcionais entre grandezas físicas. Esse processo permite capturar o comportamento de condução de forma independente do tempo total de operação.

Foram definidas métricas associadas ao estilo de condução e ao esforço imposto ao veículo, incluindo medidas relacionadas ao comportamento em trechos de aclave, ao uso do motor, ao aproveitamento de energia durante o deslocamento e à frequência de utilização de determinados regimes de operação. Essas variáveis representam aspectos diretamente influenciáveis pelo condutor e, por esse motivo, são utilizadas posteriormente como base para a geração de recomendações. O conjunto final empregado nos modelos é composto por variáveis acionáveis que descrevem proporções de tempo em diferentes faixas de rotação do motor segregadas por perfil de trecho, complementadas pelo porte do veículo, tratado como atributo imutável do contexto operacional.

Após a etapa de engenharia de características, os dados foram preparados para a modelagem. Variáveis categóricas foram convertidas para representações numéricas por meio de codificação ordinal. Em seguida, os atributos numéricos foram normalizados utilizando uma transformação Min-Max, garantindo que todas as variáveis estivessem em uma mesma escala. Esse procedimento contribui para a estabilidade do treinamento dos modelos e para a consistência na etapa de geração de contrafactuais.

Todo o processo foi aplicado individualmente a seis bases distintas, bem como a um conjunto consolidado contendo a união de todas elas. As bases diferem em porte dos veículos (de leves a extra-pesados) e tipo de operação (rodoviária, urbana, logística e agrária), permitindo avaliar o comportamento dos modelos tanto em cenários controlados quanto em situações mais variadas.

3.2. Modelagem da Classificação de Eficiência

A etapa de modelagem foi conduzida com o objetivo de avaliar diferentes algoritmos de aprendizado supervisionado na tarefa de classificação dos perfis de eficiência. Foram considerados modelos clássicos amplamente utilizados na literatura, incluindo *XGBoost*, *Random Forest*, *kNN*, *Decision Tree*, *AdaBoost*, *Support Vector Machine* e Regressão Logística.

Os experimentos foram realizados sobre cada base processada individualmente, bem como sobre um conjunto consolidado contendo todas as bases combinadas. Para cada configuração, foram executadas quatro repetições independentes, utilizando diferentes sementes aleatórias, com o objetivo de reduzir a variabilidade dos resultados e obter estimativas mais robustas de desempenho.

Em cada execução, os dados foram particionados em conjuntos de treinamento e teste, mantendo uma proporção de 80% e 20%, respectivamente. As métricas de avaliação consideradas incluem acurácia, *F1-score* macro, precisão macro e *recall* macro. Os resultados finais foram obtidos a partir da média e do desvio padrão dessas métricas ao longo das diferentes execuções.

Embora os modelos baseados em *ensembles* de árvore tenham apresentado desempenho equivalente na tarefa de classificação, o *XGBoost* foi selecionado para a etapa subsequente de geração de recomendações por manter desempenho competitivo com o melhor classificador e por integrar-se nativamente ao método contrafactual adotado, que requer uma estrutura compatível para inspeção do espaço de decisão.

3.3. Geração de Recomendações Contrafactuais

A etapa de recomendação foi baseada na geração de exemplos contrafactuais a partir do modelo *XGBoost* previamente treinado. Um exemplo contrafactual corresponde a uma versão modificada de uma instância original, com pequenas alterações nas variáveis de entrada capazes de levar o modelo a prever uma classe distinta. No contexto deste trabalho, essas modificações representam ajustes no comportamento de condução que poderiam promover uma transição para uma classe de maior eficiência.

Para a geração foi adotado o método DiCE (*Diverse Counterfactual Explanations*) na variante `kdtree`, que recupera, para uma instância de origem, vizinhos reais do conjunto de dados que já são classificados pelo modelo na classe alvo. Diferentemente de variantes baseadas em gradiente ou perturbação estocástica, essa abordagem não constrói novas combinações sintéticas no espaço de características: cada contrafactual proposto é um registro efetivamente observado no histórico de telemetria, o que assegura por construção a viabilidade física das relações derivadas e dispensa um filtro adicional de consistência interna entre variáveis.

O processo foi configurado para promover transições entre classes adjacentes de eficiência dentro de cada porte, considerando os dois saltos de Ruim para Regular e de Regular para Bom. Combinadas com os quatro portes considerados, essas transições resultam em oito cenários distintos. Para cada cenário foram selecionadas 10 instâncias do conjunto de teste, totalizando 80 instâncias-base. A seleção foi conduzida em duas etapas: primeiro, foram retidas apenas as instâncias corretamente classificadas pelo modelo na classe de origem; em seguida, dentre essas, foi amostrada uma subseleção determinística com base na distância de cada instância ao centroide de sua classe, garantindo cobertura desde casos típicos (próximos do centroide) até casos atípicos (mais afastados). Para cada instância foram gerados até 5 contrafactuais, com limite de tempo de 400 segundos por instância.

A geração dos contrafactuais foi restrita ao subconjunto de variáveis acionáveis previamente definido, mantendo o porte do veículo como atributo imutável. Essa restrição evita que o método proponha alterações em atributos contextuais que não podem ser modificados pelo condutor, garantindo que as recomendações permaneçam efetivamente acionáveis.

Após a geração, é aplicada uma etapa adicional de filtragem por semelhança topográfica entre a rota da instância original e as rotas correspondentes aos contrafactuais recuperados. Apenas os contrafactuais cuja rota apresente algum grau de semelhança em termos de extensão e perfil de relevo em relação à rota original são retidos como recomendações válidas. Esse filtro garante que a transição sugerida seja contextualmente comparável à viagem de origem, evitando recomendações associadas a rotas estruturalmente distintas e dando origem à distinção entre contrafactuais gerados e contrafactuais válidos reportada na avaliação.

3.4. Visualização do Espaço de Decisão

Para complementar a análise das recomendações geradas, foi adotada uma estratégia de visualização do espaço de decisão do modelo. Para isto, as características foram projetadas em um espaço bidimensional utilizando o método *UMAP*.

A projeção foi realizada considerando conjuntamente instâncias do conjunto de teste e os exemplos contrafactuais gerados. Esse procedimento garante que todos os pon-

tos sejam representados em um mesmo espaço latente, permitindo a comparação direta entre o estado original e o estado modificado após a aplicação das recomendações.

A visualização resultante possibilita a análise do posicionamento das instâncias no espaço de decisão, bem como a verificação de se as recomendações geradas são capazes de deslocar esses pontos para regiões associadas a classes de maior eficiência. Dessa forma, torna-se viável avaliar não apenas a mudança de classe prevista pelo modelo, mas também a consistência espacial dessa transição em relação às regiões de decisão aproximadas.

Embora essa projeção não preserve integralmente as distâncias no espaço original, ela mantém as relações de vizinhança entre as instâncias, sendo, portanto, adequada para uma análise qualitativa do comportamento do modelo. Essa abordagem constitui um método complementar de validação das recomendações, contribuindo para a interpretação e a compreensão dos resultados obtidos.

Algorithm 1 Geração e Visualização de Recomendações Contrafactuais

Require: Dados de teste X_{test} , modelo treinado f_{xgb} , variáveis acionáveis F_c

Ensure: Projeção bidimensional com instâncias e contrafactuais válidos

- 1: Selecionar subconjunto $X_s \subset X_{test}$ (predição correta + cobertura típico-atípico)
 - 2: **for all** $x \in X_s$ **do**
 - 3: Definir classe alvo y_{target}
 - 4: Gerar contrafactuais CF_x com DiCE (`kdtree`), variando apenas F_c
 - 5: Filtrar CF_x por semelhança topográfica com a rota de x
 - 6: **end for**
 - 7: Construir $X_{vis} = X_{test} \cup CF$
 - 8: Obter rótulos $\hat{y} = f_{xgb}(X_{vis})$ ▷ rótulos do modelo original
 - 9: Ajustar UMAP em X_{vis}
 - 10: Projetar instâncias e contrafactuais: $Z = UMAP(X_{vis})$
 - 11: Treinar classificador kNN usando (Z, \hat{y})
 - 12: Visualizar regiões de decisão, instâncias e contrafactuais
-

O Algoritmo 1 resume o fluxo completo. Como o UMAP realiza uma transformação não linear do espaço de características, inviabilizando a aplicação direta do XGBoost nesse novo espaço, treina-se um classificador kNN sobre as projeções rotuladas pelas predições do XGBoost. O kNN atua, assim, como modelo auxiliar que aproxima as fronteiras de decisão do modelo original no espaço bidimensional, fornecendo uma ponte interpretável para a análise visual.

4. Resultados

Esta seção apresenta os resultados obtidos a partir da avaliação do *framework* proposto, considerando dois aspectos principais: (i) o desempenho dos modelos de classificação na tarefa de identificação de perfis de eficiência e (ii) a qualidade das recomendações contrafactuais geradas.

A avaliação dos modelos é conduzida por meio de métricas clássicas de classificação, incluindo acurácia, F1-score, precisão e recall, permitindo analisar a capacidade de generalização em diferentes cenários operacionais. Já as recomendações contrafactuais são avaliadas com base em sua aplicabilidade, considerando a taxa de sucesso na recuperação de contrafactuais válidos após a etapa de filtragem por semelhança topográfica. Por fim, uma análise qualitativa no espaço de decisão é realizada para investigar o comportamento das recomendações e sua relação com a estrutura dos dados.

4.1. Avaliação dos Modelos

A Tabela 2 apresenta o desempenho dos três melhores modelos para cada base avaliada. Os resultados correspondem à média e ao desvio padrão das métricas ao longo de diferentes execuções, permitindo uma análise mais robusta da performance.

Tabela 2. Desempenho dos três melhores modelos por base (média \pm desvio padrão, em %).

Base	Modelo	Accuracy	F1	Precision	Recall
Todas Frotas	RandomForest	67.53 \pm 0.21	63.95 \pm 1.28	69.80 \pm 0.28	61.97 \pm 1.06
	XGBoost	63.05 \pm 0.34	61.02 \pm 1.38	65.93 \pm 0.57	59.28 \pm 1.29
	DecisionTree	59.76 \pm 0.29	55.35 \pm 0.60	54.86 \pm 0.55	56.22 \pm 0.92
Frota 1	RandomForest	64.82 \pm 0.93	67.37 \pm 0.84	69.05 \pm 0.93	66.53 \pm 0.86
	XGBoost	60.91 \pm 0.36	64.29 \pm 1.17	65.18 \pm 1.27	63.91 \pm 1.17
	DecisionTree	56.19 \pm 0.56	58.90 \pm 1.04	58.81 \pm 1.13	59.12 \pm 0.95
Frota 2	RandomForest	67.28 \pm 0.68	68.58 \pm 1.01	71.89 \pm 1.66	66.72 \pm 1.15
	XGBoost	65.16 \pm 1.01	68.97 \pm 1.49	74.51 \pm 1.34	66.37 \pm 1.86
	KNN	60.08 \pm 0.53	55.37 \pm 2.83	57.58 \pm 4.28	54.90 \pm 2.87
Frota 3	RandomForest	64.25 \pm 0.31	60.95 \pm 1.28	66.32 \pm 0.34	58.88 \pm 1.43
	XGBoost	63.99 \pm 0.26	60.82 \pm 1.02	65.23 \pm 0.56	58.97 \pm 1.18
	SVM	60.34 \pm 0.26	47.13 \pm 0.73	47.51 \pm 0.84	47.98 \pm 0.70
Frota 4	XGBoost	87.54 \pm 1.21	52.55 \pm 6.57	66.44 \pm 15.14	48.32 \pm 4.76
	RandomForest	87.46 \pm 1.17	47.23 \pm 8.18	62.02 \pm 21.61	43.66 \pm 5.28
	SVM	86.33 \pm 1.50	30.89 \pm 0.29	28.78 \pm 0.50	33.33 \pm 0.00
Frota 5	RandomForest	62.92 \pm 0.73	61.24 \pm 0.53	62.77 \pm 0.22	60.65 \pm 0.53
	XGBoost	62.33 \pm 1.11	61.05 \pm 0.81	61.87 \pm 0.72	60.70 \pm 0.77
	KNN	54.55 \pm 0.89	51.35 \pm 0.92	54.37 \pm 0.92	50.92 \pm 0.76

De forma geral, os algoritmos baseados em *ensembles* de árvores dominaram as classificações. O Random Forest destacou-se por apresentar o melhor desempenho na base consolidada (Todas Frotas) e na maioria das frotas individuais, demonstrando grande capacidade de generalização. O XGBoost manteve-se altamente competitivo, figurando no topo em quase todos os cenários e liderando a Frota 4. Em contrapartida, modelos mais simples ou baseados em distância (Decision Tree, KNN e SVM) apareceram apenas de forma esporádica e com métricas inferiores.

A análise por frota reflete a complexidade introduzida pelo mapeamento em 12 categorias, com as acurácias estabilizando na faixa de 60% a 67% para a maioria das bases. A Frota 4 apresenta um comportamento atípico: embora alcance alta acurácia ($\sim 87\%$), seus valores de F1-score e *Recall* sofrem quedas drásticas (chegando a $\sim 30\%$ e 33% no SVM, respectivamente), evidenciando um cenário de forte desbalanceamento de perfis.

No cenário consolidado (Todas Frotas), o Random Forest provou ser o modelo mais robusto para lidar com a alta variabilidade dos dados combinados (67,53% de acurácia), justificando seu destaque frente aos demais métodos testados.

4.2. Avaliação das Recomendações Contrafactuais

A geração de recomendações foi conduzida sobre o modelo XGBoost treinado com o conjunto comportamental de 19 *features*: 18 variáveis acionáveis (*F01–F18*), que representam proporções de tempo em faixas operacionais por tipo de perfil de rota, e o porte do veículo, tratado como atributo imutável do contexto. Foram definidas oito transições incrementais (ruim \rightarrow regular \rightarrow bom) intra-porte, totalizando quatro portes \times dois saltos de classe. Para cada transição foram selecionadas 10 instâncias do conjunto de teste com cobertura de diversidade — filtradas por predição correta do modelo e amostradas por `np.linspace` sobre a distância ao centroide da classe de origem, garantindo que o conjunto incluía desde casos típicos até atípicos. Para cada instância foram gerados até 5 contrafactuais pelo método `kdtree` do DiCE, com tempo limite de 400 segundos por instância.

Diferentemente de métodos baseados em gradiente ou perturbação estocástica, o `kdtree` do DiCE retorna como contrafactuais registros reais do *dataset* classificados pelo modelo na classe alvo. Essa característica implica que os CFs gerados satisfazem por construção as restrições físicas do domínio (somadas de frações de tempo, limites do *MinMaxScaler*) e pertencem efetivamente à classe desejada, dispensando etapas adicionais de validação de consistência interna entre variáveis. Sobre o conjunto de candidatos retornados é então aplicada uma filtragem por semelhança topográfica em relação à rota da instância original, restando apenas os contrafactuais associados a rotas comparáveis em extensão e perfil de relevo. A métrica relevante passa a ser, portanto, a *taxa de retenção topográfica*: a proporção de contrafactuais gerados que sobrevivem a esse filtro.

A Tabela 3 apresenta os resultados por transição. No total, foram produzidos 395 contrafactuais candidatos — 50 por transição, exceto em *leve_ruim* \rightarrow *leve_regular*, em que um *timeout* reduziu o conjunto a 45, atribuível à baixa densidade de vizinhos válidos no *kdtree* para o porte Leve, que é subrepresentado no conjunto consolidado. Após a filtragem topográfica, 165 contrafactuais foram retidos como válidos, correspondendo a uma taxa de retenção global de 41,77%, com variação por transição de 26,67% (*leve_ruim* \rightarrow *leve_regular*) a 56,00% (*semipesado_ruim* \rightarrow *semipesado_regular*).

Tabela 3. Avaliação dos exemplos contrafactuais por tipo de transição.

Métrica	Leve Reg.	Leve Ruim	Médio Reg.	Médio Ruim	Pesado Reg.	Pesado Ruim	Semi. Reg.	Semi. Ruim	Total
	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	
	Leve Bom	Leve Reg.	Médio Bom	Médio Reg.	Pesado Bom	Pesado Reg.	Semi. Bom	Semi. Reg.	
Contrafactuais gerados	50	45	50	50	50	50	50	50	395
Contrafactuais válidos	17	12	22	25	22	17	22	28	165
Taxa de retenção	34.00%	26.67%	44.00%	50.00%	44.00%	34.00%	44.00%	56.00%	41.77%

A *direção* das recomendações foi analisada por meio do Δ médio assinado (CF – Original) calculado para cada par (*feature*, transição) sobre o conjunto completo de 395 recomendações. A Figura 3 apresenta o *heatmap* resultante: valores negativos (em azul) correspondem a *features* que os contrafactuais tendem a reduzir em relação à instância original, enquanto valores positivos (em vermelho) indicam as *features* que o *framework* tende a aumentar ao migrar uma viagem para a classe de eficiência superior.

A Figura 3 revela que o deslocamento comportamental se concentra em um pequeno subconjunto de *features*: *F01* é a variável mais sistematicamente reduzida, com magnitudes crescentes conforme o porte (atingindo $\Delta \approx -0,12$ nos portes pesados),

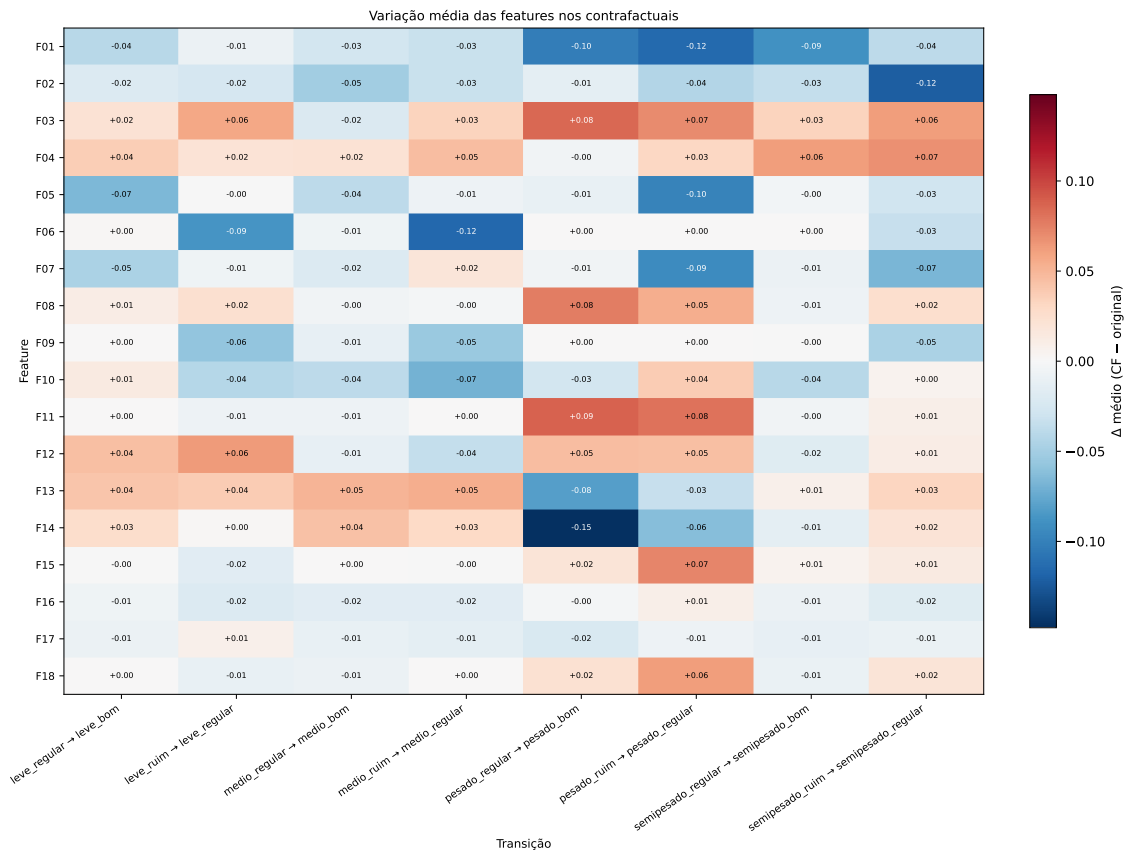


Figura 3. Variação média das *features* nos contrafactuais (Δ médio de CF – Original) para cada uma das 18 *features* acionáveis (F01–F18) sobre as oito transições avaliadas. Tons azulados indicam *features* reduzidas nos CFs; tons avermelhados indicam *features* aumentadas.

tendo *F03* como contrapartida positiva consistente em todas as transições. *F14* e *F11* apresentam comportamento específico dos portes mais pesados, sugerindo um padrão de reajuste distinto, enquanto *F15–F18* permanecem praticamente inertes ($|\Delta| \leq 0,02$). De forma geral, as recomendações são simultaneamente operacionalmente válidas — por serem instâncias reais do histórico classificadas na classe alvo — e interpretáveis, convergindo para um padrão direcional consistente em torno de poucas *features*-chave.

4.3. Análise no Espaço de Decisão

As Figuras 4a e 4b apresentam a projeção das instâncias no espaço bidimensional obtido por meio do UMAP, juntamente com as fronteiras de decisão aproximadas por um classificador kNN ajustado a partir das predições do modelo XGBoost. Nessas representações, os pontos correspondem a amostras do conjunto de teste, enquanto os marcadores destacados indicam instâncias originais e suas respectivas recomendações contrafactuais.

Observa-se que as classes se organizam em regiões relativamente bem definidas, com áreas de transição onde ocorre sobreposição entre perfis de eficiência. Esse comportamento é esperado, dado que pequenas variações no comportamento de condução podem resultar em mudanças graduais na eficiência.

Ao analisar as trajetórias contrafactuais, nota-se que as recomendações promovem deslocamentos direcionados no espaço, conectando regiões de menor para maior

eficiência. Na transição de Ruim para Regular (Figura 4a), os deslocamentos tendem a ser mais curtos e consistentes, indicando que pequenas alterações nas variáveis comportamentais já são suficientes para promover melhoria. Em contraste, na transição de Regular para Bom (Figura 4b), observa-se maior dispersão nas trajetórias, sugerindo a necessidade de ajustes mais específicos e localizados para atingir a classe alvo.

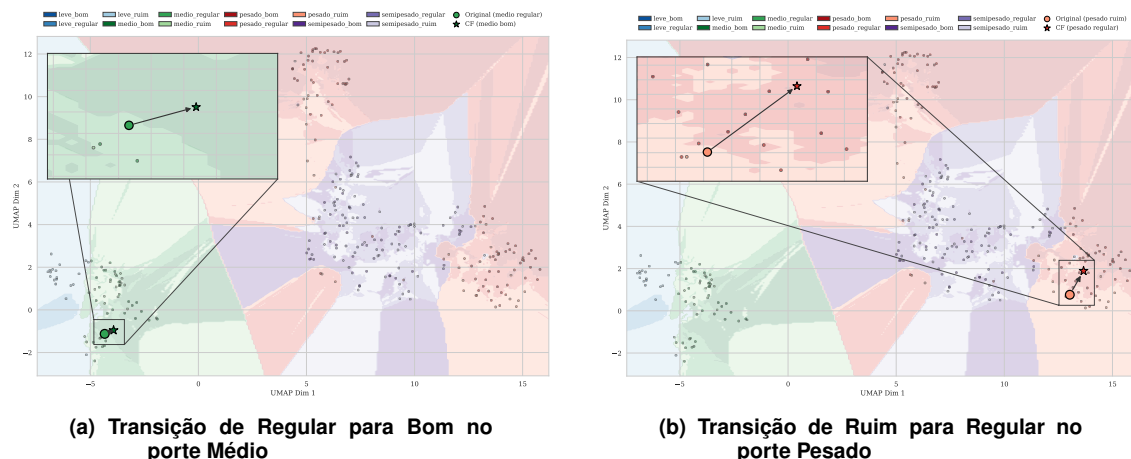


Figura 4. Projeção UMAP com fronteiras aproximadas por kNN (a partir do XG-Boost). Regiões indicam classes, pontos representam o teste e estrela a representação.

Esse comportamento é consistente com a estrutura observada no espaço de decisão: transições Ruim→Regular tendem a ocorrer entre regiões mais amplas e heterogêneas, enquanto Regular→Bom tende a se concentrar em regiões mais densas e localizadas. Isso indica que, embora a melhoria inicial exija mudanças relativamente simples, ela é mais suscetível a deslocamentos para rotas estruturalmente distintas da original, ao passo que as melhorias em níveis mais altos de eficiência tendem a recrutar contrafactuais associados a rotas mais próximas da viagem de origem em termos topográficos.

Outro aspecto relevante é que os deslocamentos não são aleatórios, mas seguem direções coerentes com a estrutura local do espaço, aproximando as instâncias de regiões densas da classe alvo. Isso sugere que o método é capaz de explorar padrões já presentes nos dados, em vez de propor soluções artificiais ou isoladas.

A análise também evidencia que regiões próximas às fronteiras de decisão apresentam maior sensibilidade, o que impacta diretamente a qualidade dos contrafactuais gerados. Nessas áreas, pequenas variações podem levar a mudanças de classe, mas também aumentam a probabilidade de o vizinho recuperado corresponder a uma rota estruturalmente distinta da original, reforçando a importância do filtro de semelhança topográfica.

De forma geral, os resultados mostram que o *framework* proposto é capaz de gerar recomendações coerentes com a estrutura do espaço de decisão e alinhadas com padrões reais dos dados. No entanto, a diferença observada entre as transições evidencia uma limitação importante: a geração de contrafactuais em regiões mais distantes ou menos estruturadas tende a produzir soluções menos comparáveis à viagem de origem. Isso justifica a etapa de filtragem topográfica e indica que abordagens futuras podem se beneficiar da incorporação dessas restrições de comparabilidade diretamente no processo de geração. Além disso, melhorias no método podem envolver estratégias que considerem explicitamente a geometria local do espaço de decisão, reduzindo a recuperação de vizinhos pouco contextualmente compatíveis e aumentando a eficiência do processo.

5. Conclusão

Este trabalho apresentou um *framework* para análise e melhoria do comportamento de condução em frotas, combinando modelagem preditiva, geração de recomendações contrafactuais e filtragem por semelhança topográfica entre rotas. A abordagem, baseada em dados reais de telemetria, mostrou-se capaz de classificar perfis de eficiência de forma consistente em diferentes cenários e de gerar recomendações eficazes para promover a melhoria do desempenho.

A etapa de filtragem topográfica mostrou-se essencial para garantir a comparabilidade contextual das recomendações, retendo apenas contrafactuais associados a rotas estruturalmente próximas da viagem de origem e aumentando sua aplicabilidade prática. Além disso, a análise no espaço de decisão indicou que as recomendações promovem deslocamentos coerentes em direção a regiões de maior eficiência.

Como trabalhos futuros, pretende-se explorar diferentes estratégias de geração de contrafactuais, incorporar restrições diretamente no processo de otimização e avaliar o método em cenários mais diversos, incluindo sua validação em ambientes reais de operação.

6. Agradecimentos

Este projeto foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei nº 8.248, de 23 de outubro de 1991, no âmbito do PPI-SOFTEX, coordenado pela Softex e publicado Arquitetura Cognitiva (Fase 3), DOU 01245.003479/2024-10

Referências

- Alam, G. M. I., Arfin Tanim, S., Sarker, S. K., Watanobe, Y., Islam, R., Mridha, M., and Nur, K. (2025). Deep learning model based prediction of vehicle co2 emissions with explainable ai integration for sustainable environment. *Scientific Reports*, 15(1):3655.
- Ashqar, H. I., Obaid, M., Jaber, A., Ashqar, R., Khanfar, N. O., and Elhenawy, M. (2024). Incorporating driving behavior into vehicle fuel consumption prediction: Methodology development and testing. *Discover Sustainability*, 5(1):344.
- Bouhsissin, S., Sael, N., and Benabbou, F. (2023). Driver behavior classification: A systematic literature review. *IEEE Access*, 11:14128–14153.
- Braun, C., Jarczewski, R. O., Villas, L., Moura, P., and de Souza, A. M. (2025). Modelagem preditiva e interpretação de emissões de dióxido de carbono em máquinas agrícolas com técnicas de ia. In *Workshop de Computação Urbana (CoUrb)*, pages 279–292. SBC.
- de Souza, A. M., Brennand, C. A. R. L., Yokoyama, R. S., Donato, E. A., Madeira, E. R. M., and Villas, L. A. (2017). Traffic management systems: A classification, review, challenges, and future perspectives. *International Journal of Distributed Sensor Networks*, 13(4).
- de Souza, A. M. et al. (2020). Safe and sound: Driver safety-aware vehicle re-routing based on spatiotemporal information. *IEEE Transactions on Intelligent Transportation Systems*.
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection. *arXiv preprint arXiv:1802.03426*.

- Mohammadnazar, A., Khattak, Z. H., and Khattak, A. J. (2024). Assessing driving behavior influence on fuel efficiency using machine-learning and drive-cycle simulations. *Transportation Research Part D: Transport and Environment*, 126:104025.
- Mothilal, R. K., Sharma, A., and Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. *Proceedings of FAT**.
- Mumcuoglu, M. E., Farea, S. M., Unel, M., Mise, S., Unsal, S., Yilmaz, M., and Koprubasi, K. (2023). Fuel consumption classification for heavy-duty vehicles: a novel approach to identifying driver behavior and system anomalies. In *2023 AEIT International Conference on Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6. IEEE.
- Nasr Azadani, M. and Boukerche, A. (2022). Driving behavior analysis guidelines for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):6027–6045.
- Payyanadan, R. P. and Angell, L. S. (2022). A framework for building comprehensive driver profiles. *Information*, 13(2):61.
- Ping, P., Qin, W., Xu, Y., Miyajima, C., and Takeda, K. (2019). Impact of driver behavior on fuel consumption: Classification, evaluation and prediction using machine learning. *IEEE access*, 7:78515–78532.
- Ritchie, H., Rosado, P., and Roser, M. (2023). Data page: Per capita CO₂ emissions. <https://ourworldindata.org/grapher/co-emissions-per-capita>. Parte da publicação "CO₂ and Greenhouse Gas Emissions". Dados adaptados do Global Carbon Project e várias fontes. Acesso em: 31 mar. 2025.
- Tselentis, D. I. and Papadimitriou, E. (2023). Driver profile and driving pattern recognition for road safety assessment: Main challenges and future directions. *IEEE Open Journal of Intelligent Transportation Systems*, 4:83–100.
- Vyas, J., Das, D., and Chaudhury, S. (2021). Drivebfr: driver behavior and fuel-efficiency-based recommendation system. *IEEE Transactions on Computational Social Systems*, 9(5):1446–1455.