



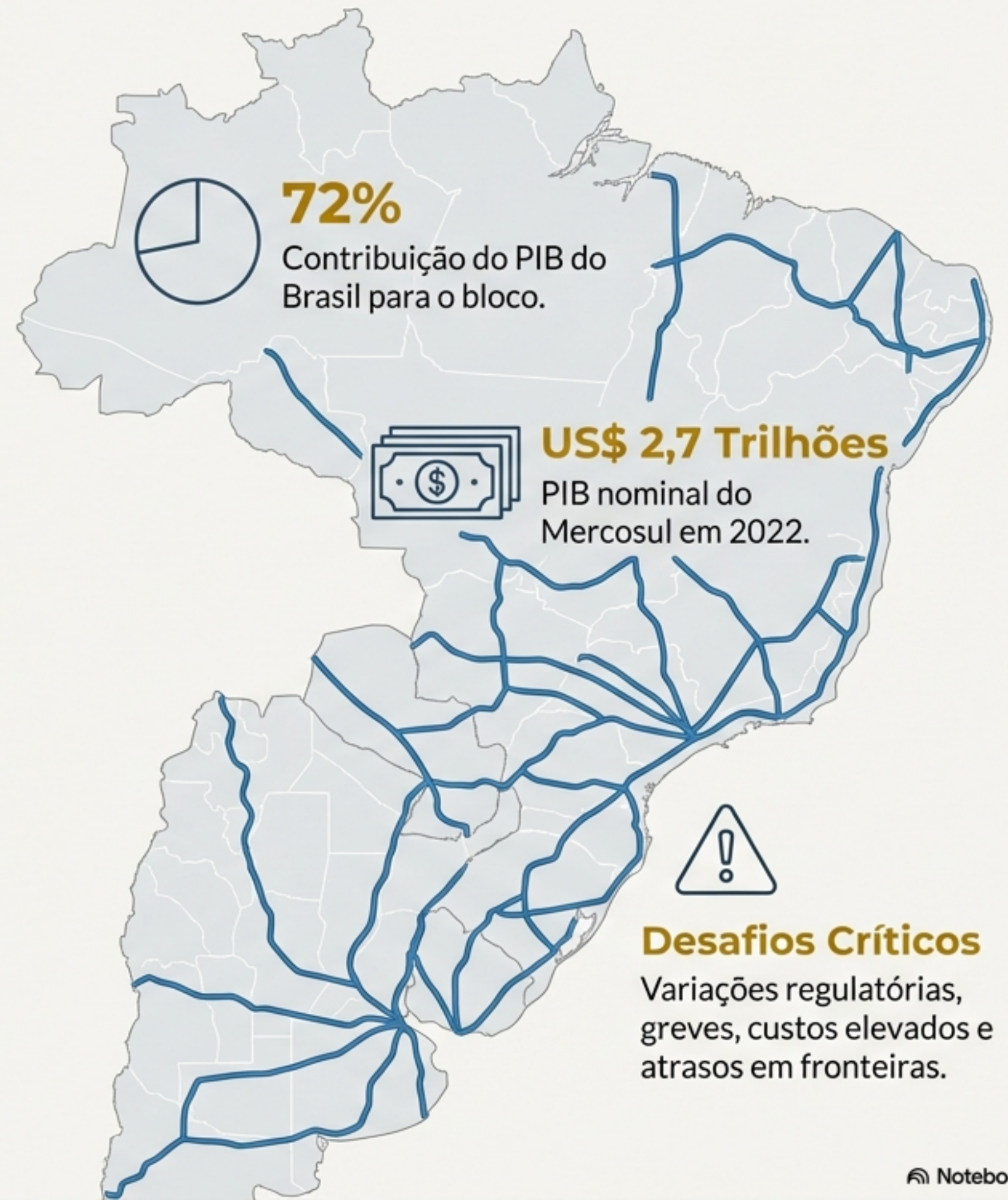
Do Caos à Previsibilidade: Como a Ciência de Dados está Redefinindo a Logística no Mercosul.

Um estudo de caso sobre a aplicação de Machine Learning para prever prazos de entrega no transporte rodoviário internacional.

Baseado no Trabalho de Conclusão de Curso (TCC) de Giuliano Curtinaz, MBA em Data Science e Analytics, ESALQ/USP.

O transporte rodoviário é a espinha dorsal de um mercado de US\$ 2,7 trilhões, mas opera sob constante incerteza.

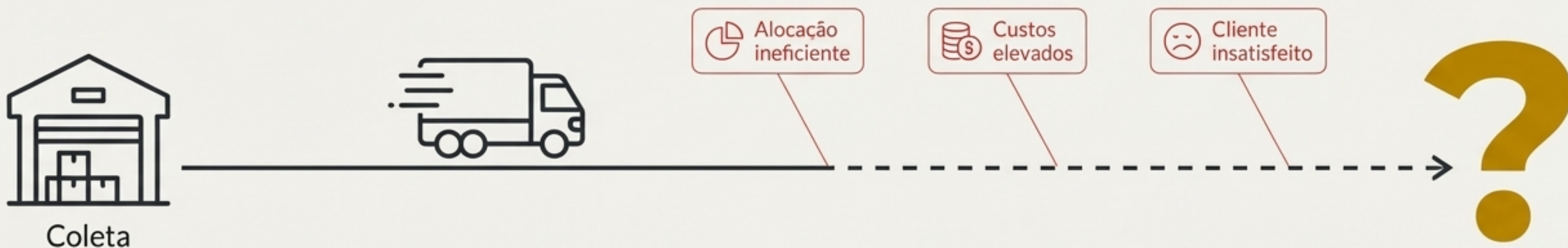
O Mercosul depende do transporte rodoviário para sua integração econômica. No entanto, instabilidade, burocracia aduaneira e desafios logísticos tornam a previsibilidade um dos maiores gargalos operacionais do setor, impactando custos e a competitividade.



A incapacidade de prever prazos de entrega com precisão era um desafio operacional crítico para o nosso negócio.

O Problema

A falta de previsibilidade resultava em má alocação de frota, custos de estoque elevados para os clientes e uma comunicação reativa, minando a satisfação e a eficiência.



O Objetivo do Projeto

Implementar algoritmos de Machine Learning para prever com alta precisão o tempo total de viagem, desde a coleta até a entrega final, utilizando dados operacionais reais da empresa.

A base da nossa análise foi um conjunto de dados rico, extraído de 3.739 viagens internacionais reais.

Consulta SQL a partir de registros de transporte da empresa, cobrindo o período de janeiro de 2024 a março de 2025.



Desenvolvemos um modelo que atualiza suas previsões em quatro estágios cruciais da viagem.

Em vez de uma única previsão estática, o modelo foi projetado para se adaptar, **recalculando o prazo de entrega** à medida que novas informações se tornam disponíveis, simulando a operação em tempo real.



Fomos além dos dados brutos, criando features inteligentes para capturar padrões ocultos.

Engenharia de Variáveis



tempo_mediano_anterior

Uma baseline histórica, representando a mediana de tempo de trânsito para cada par origem-destino.



Indicadores Binários

Variáveis para capturar hipóteses de negócio, como "motorista residente na cidade de fronteira".



Variáveis Temporais

Dia da semana do início da viagem e da chegada à fronteira.



Clusterização para Segmentação

Utilizamos K-Means e Autoencoders para agrupar viagens com características similares, transformando esses grupos em novas variáveis preditoras para os modelos.

Dados Brutos

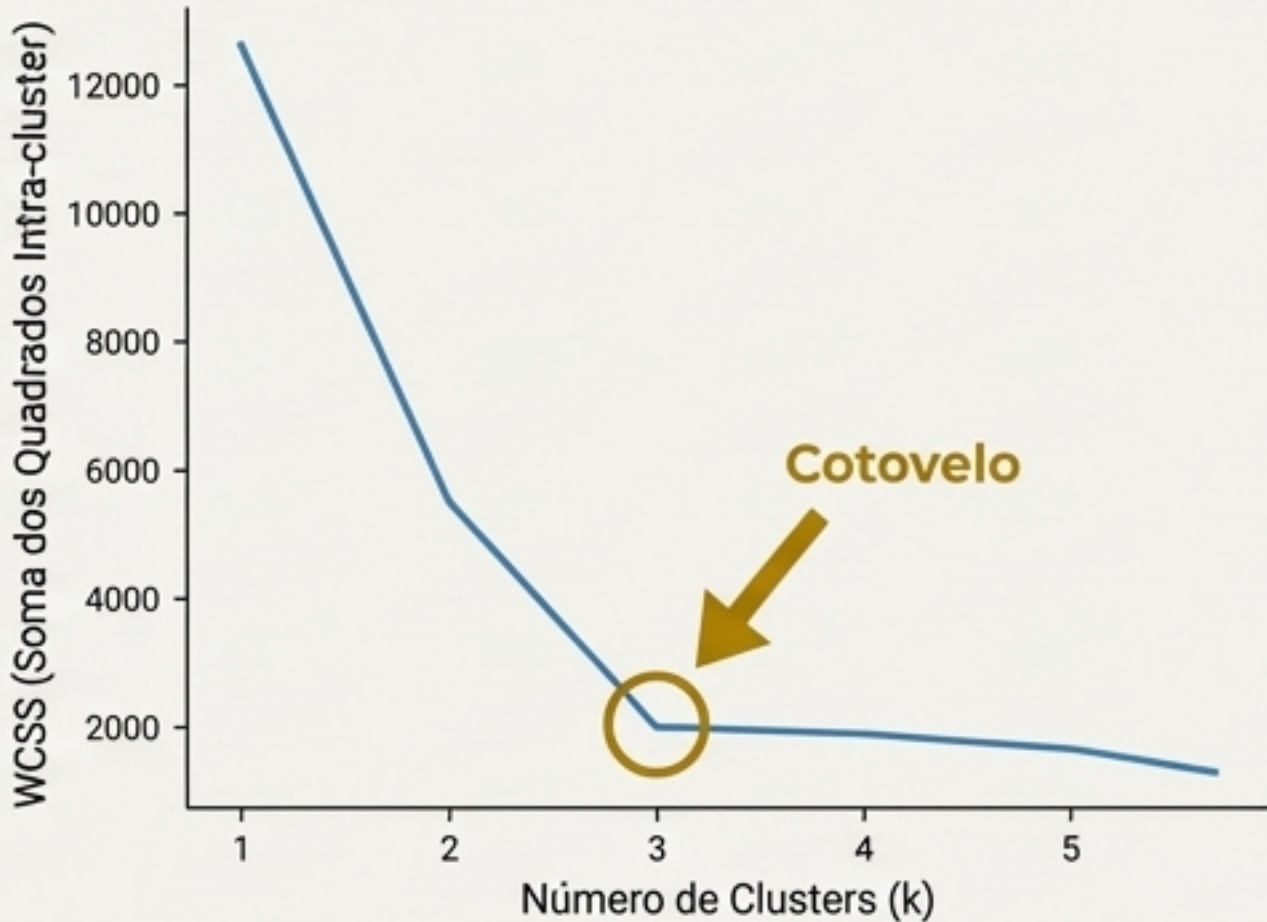
Origem	Destino	Peso
CRG	DES	200
LEO	DET	750
REL	CAS	600
GAS	LIL	700
...	...	...
REL	CAS	750



Features Enriquecidas

Origem	Destino	Peso	tempo_mediano_anterior	grupo_kmeans
CRG	DES	250	83	1
LEO	DET	750	87	1
REL	CAS	600	144	2
GAS	LIL	750	87	3
...	...	...	...	...
REL	CAS	700	136	4

Método do Cotovelo (Elbow Method)



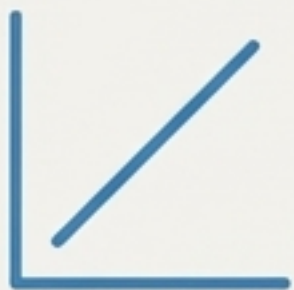
Submetemos os dados a um rigoroso processo de seleção, testando um espectro de algoritmos de Machine Learning.

Para garantir que encontraríamos a solução mais robusta e precisa, comparamos o desempenho de diferentes classes de modelos, desde baselines simples até arquiteturas de ponta.



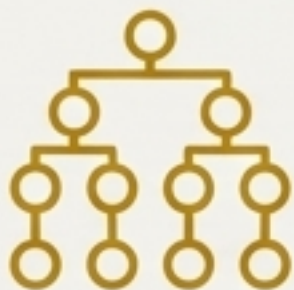
Modelo de Referência

- Mediana por Origem-Destino



Modelos Lineares

- Regressão Linear (OLS) com features selecionadas
- Regressão Linear (OLS) com todas as features (stepwise)



Modelos baseados em Árvores e Ensemble

- Decision Tree
- RandomForest
- XGBoost
- LGBM



Redes Neurais

- MLPRegressor
- Rede Neural Densa (Keras)



Máquinas de Vetores de Suporte

- Support Vector Regression (SVR)

Os resultados foram claros: modelos de ensemble superaram massivamente os métodos lineares em estabilidade e precisão.

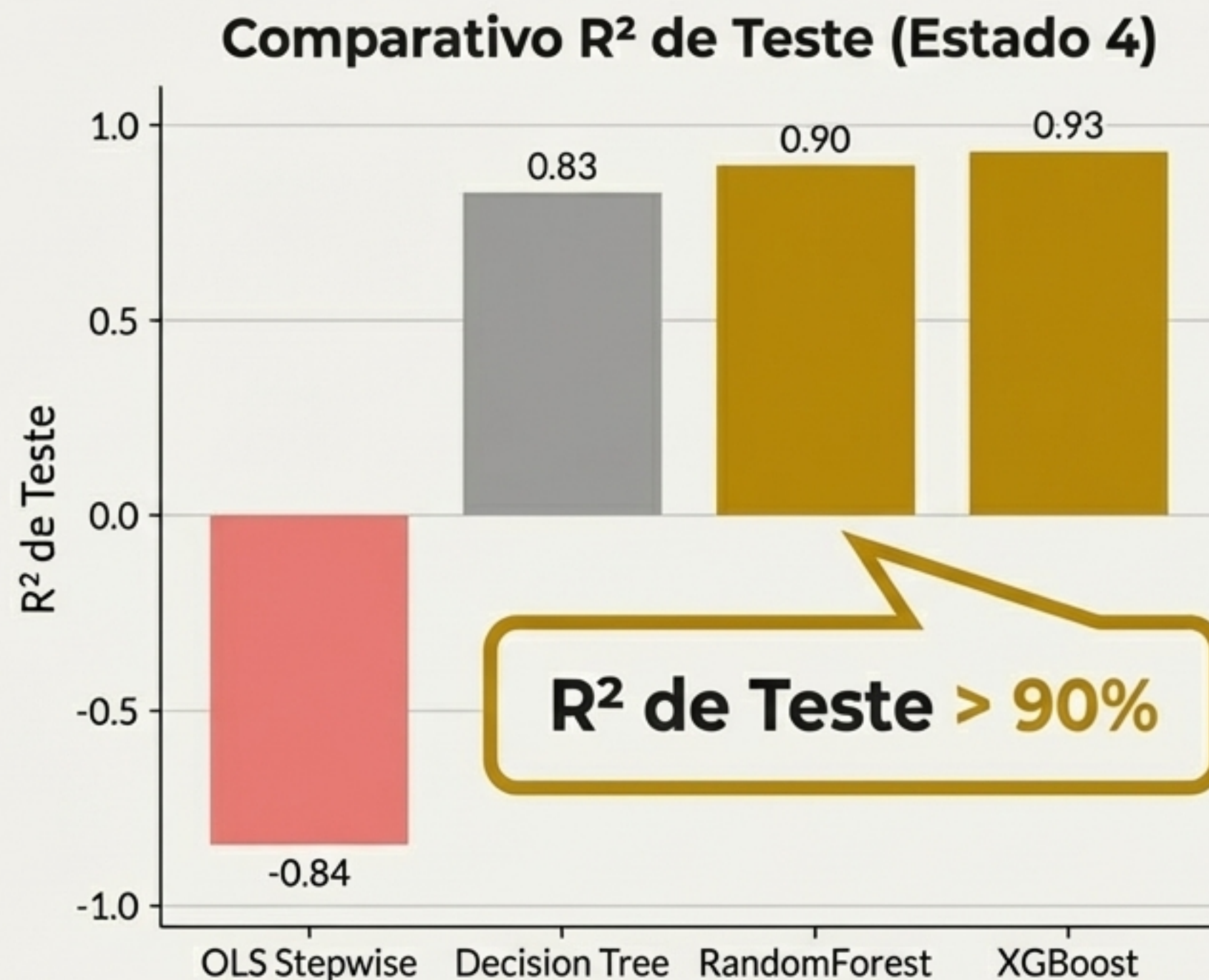
Análise Comparativa

Modelos OLS

Apresentaram instabilidade e alto risco de overfitting. O R^2 de teste chegou a -84% no Estado 4 em uma das configurações, indicando um péssimo ajuste a novos dados.

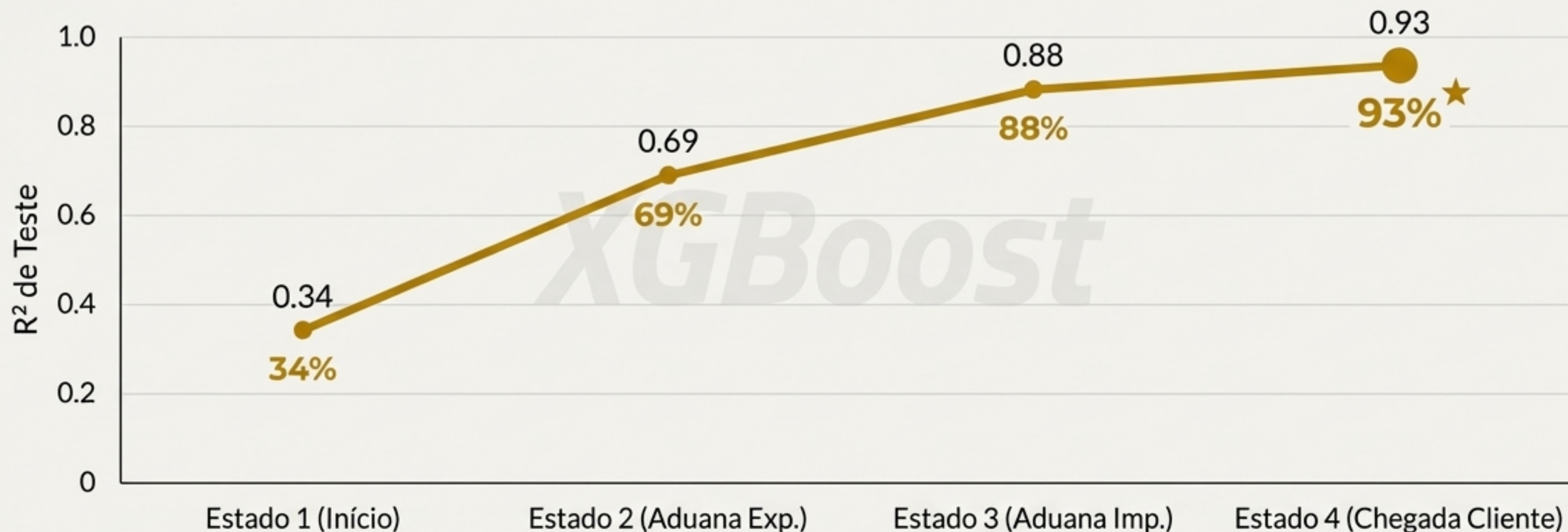
Modelos Ensemble (RandomForest, LGBM, XGBoost)

Demonstraram alta performance e robustez em todas as fases da viagem, aprendendo com a complexidade dos dados sem sofrer com overfitting.



XGBoost consolidou-se como o modelo campeão, alcançando um R^2 de 93% na fase final da viagem.

Evolução da Performance do XGBoost (R^2 de Teste)



O modelo se mostrou robusto e com alta performance, sendo relativamente insensível às variações de pré-processamento, o que o torna ideal para implementação em produção.

Além da precisão estatística, o modelo entrega acurácia operacional para o planejamento logístico.

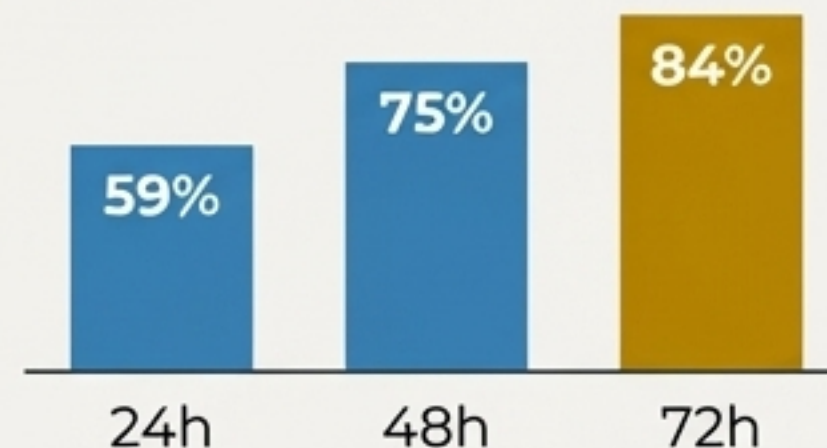
84%

de **acerto** na previsão do prazo de entrega dentro de uma janela de **72 horas** na fase final da viagem.

Medimos a porcentagem de vezes que a previsão do modelo acertou o dia da entrega dentro de uma janela de tempo específica. Este é o indicador que realmente importa para a comunicação com o cliente e alocação de recursos.



Acurácia por Janela de Tempo



Uma análise detalhada das métricas confirma a superioridade e a consistência dos modelos de ensemble.

Os valores de R^2 (coeficiente de determinação) e RMSE (raiz do erro quadrático médio, em dias) mostram não apenas quais modelos se ajustam melhor aos dados, mas também a magnitude **média** dos erros de previsão.

	Test R^2				Test RMSE (em dias)			
	OLS Selected	RandomForest	XGBoost	LGBM	OLS Selected	RandomForest	XGBoost	LGBM
Estado 1	7%	37%	34%	35%	4.0	3.3	3.3	3.3
Estado 2	36%	70%	69%	68%	3.3	2.3	2.3	2.4
Estado 3	84%	86%	88%	87%	1.6	1.5	1.4	1.4
Estado 4	84%	90%	93%	93%	1.6	1.3	1.1	1.1

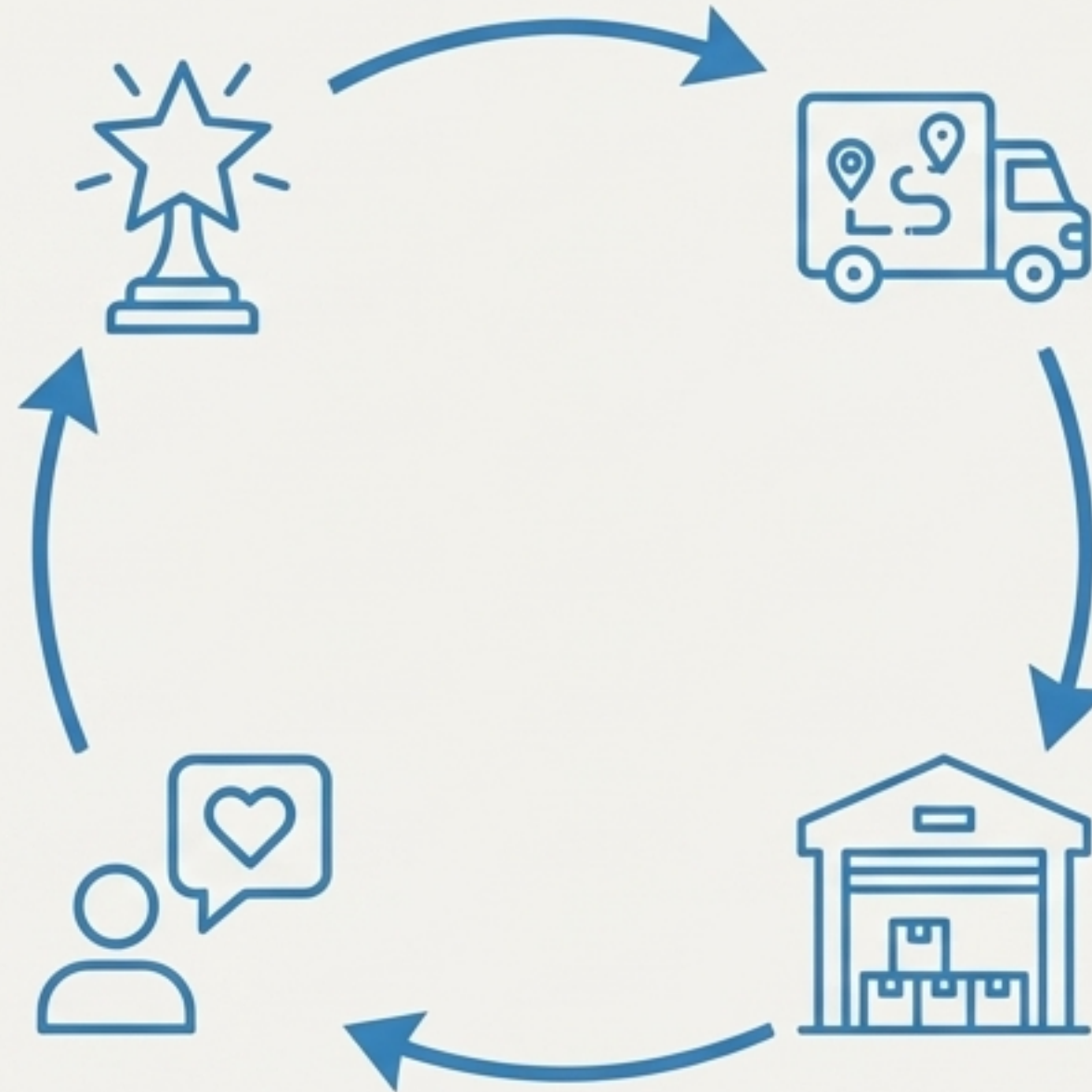
Maior Instabilidade:
OLS com RMSE de até 79.8 dias no pior cenário.

Menor Erro de Previsão:
XGBoost e LGBM com RMSE de apenas 1.1 dia na fase final.

O resultado é uma transformação: saímos da incerteza para uma vantagem estratégica e operacional.

Diferencial Competitivo:
A confiabilidade nas entregas torna-se um fator chave para se destacar no desafiador mercado do Mercosul.

Satisfação do Cliente:
A comunicação proativa sobre os prazos de entrega eleva a confiança e a qualidade do serviço percebida.



Otimização de Frota:
Previsões precisas permitem um melhor planejamento de alocação de veículos e motoristas, reduzindo o tempo ocioso.

Gestão de Estoques:
Clientes podem gerenciar seus estoques com mais eficiência, reduzindo custos e evitando rupturas.

Este projeto não é uma prova de conceito; é uma solução em produção, gerando valor real hoje.

Status do Projeto



Sistema implantado e operacional desde 15/10/2025.

Funcionalidades Ativas



Todas as viagens internacionais agora possuem previsão de descarga gerada pelo modelo.



Envio automatizado de notificações via WhatsApp para o cliente com a previsão de entrega.



Painel para medir a acurácia do modelo em tempo real.



Disponibilidade futura de veículos calculada com base nas previsões de ML.



Acurácia do Modelo
(Últimos 30 dias)



85%

Média Recente

A excelência do trabalho foi reconhecida com a nota máxima e uma indicação ao prêmio de melhor TCC do curso.



Nota final: 9,50 (Aprovado)



Indicação ao Prêmio de Melhor TCC

“O trabalho foi indicado pelo seu processo cuidadoso e por resultar em uma solução de impacto, representando um dos melhores projetos do MBA em Data Science & Analytics (ESALQ/USP).”



Machine Learning para previsão do Prazo de Entrega no Transporte Rodoviário Internacional de Cargas

giuliano@curtinaz.com.br

<http://giuliano.curtinaz.com.br>