

Prazo de entrega no transporte rodoviário internacional de cargas

Giuliano Curtinaz^{1*}; Vanilde de Castro²

¹ GC Informática. Diretor. Rua Fab, 133/401 – Bairro Nossa Senhora das Graças 92110-030 Canoas, RS, Brasil

² Prof.^a Dr.^a Orientadora Convidada do curso de MBA Data Science e Analytics da ESALQ /USP - Piracicaba – SP, Brasil

*autor correspondente: giuliano@curtinaz.com.br

Prazo de entrega no transporte rodoviário internacional de cargas

Resumo

Diante dos desafios de previsibilidade de tempos de viagem rodoviárias internacionais, este trabalho teve como objetivo implementar algoritmos de Machine Learning em uma empresa brasileira de transporte rodoviário internacional no Mercosul, visando prever com precisão os prazos de entrega de todas as mercadorias em trânsito internacional. A técnica empregada consistiu em um estudo de caso com uma base de dados de aproximadamente 5.300 registros de viagens, em que diversos modelos de regressão foram treinados e validados em quatro estágios distintos: carregamento, ingresso em aduana, liberação de aduana e chegada ao cliente, os quais precedem a descarga e o término da viagem, simulando a disponibilidade progressiva de informações. Os resultados demonstraram a alta capacidade preditiva dos modelos de “ensemble”, como “LGBM”, “XGBoost” e “RandomForest”, que alcançaram um coeficiente de determinação (R^2) superior a 90% nas fases finais da viagem. A análise da acurácia em intervalos de predição confirmou a robustez dos modelos para o planejamento operacional, atingindo até 84% de acerto. Esses achados reforçam o uso dos modelos na prática, possibilitando melhor alocação de recursos e veículos, além de reduzir os custos na gestão de estoques. Concluiu-se que o aprendizado de máquina quando bem aplicado fornece aos transportadores um diferencial importante para se destacar no mercado elevando a confiabilidade de suas entregas em um contexto logístico desafiador.

Palavras-chave: Previsibilidade; Frota; Mercosul; Machine Learning.

Introdução

O transporte rodoviário internacional de cargas é essencial para a integração econômica dos países do Mercosul, formado por Brasil, Argentina, Uruguai e Paraguai. O Brasil, sozinho, responde por mais de 72% do Produto Interno Bruto total do bloco, evidenciando sua relevância para o comércio regional, que movimentou aproximadamente US\$ 2,7 trilhões em 2022 (Poder360, 2023). De acordo com a Associação Brasileira de Transportadores Internacionais [ABTI], esse setor não apenas facilita o fluxo de mercadorias, mas também fortalece o desenvolvimento econômico ao conectar indústrias e mercados em diferentes países (ABTI, 2021).

Apesar de sua importância, o transporte internacional enfrenta desafios significativos, como a instabilidade econômica e política na região, variações regulatórias entre os países, custos elevados, greves e atrasos logísticos. Durante a pandemia de COVID-19, essas dificuldades foram acentuadas, com aumento no tempo ocioso dos caminhões, custos logísticos e redução na previsibilidade das operações (ABTI, 2020). Tais desafios reforçam a urgência de implementar soluções capazes de otimizar a cadeia logística, minimizando os impactos decorrentes de imprevistos e assegurando maior eficiência operacional.

Nesse contexto de busca por maior integração e superação de gargalos logísticos, surgem projetos de infraestrutura de grande envergadura, como a Rota de Integração Latino-Americana (RILA), que visa conectar corredores bioceânicos. Tais iniciativas são frequentemente acompanhadas por narrativas de progresso e desenvolvimento. Contudo, a forma como esses empreendimentos são comunicados e percebidos pela sociedade e pelos envolvidos pode influenciar o debate público e as expectativas em torno de seus reais impactos. O estudo de Machado e Maciel (2024) sobre a cobertura jornalística da RILA, analisa os discursos sobre desenvolvimento veiculados pela mídia, investigando como as promessas e os desafios inerentes a tais projetos são construídos e disseminados.

O estudo de Machado e Maciel (2024) sobre projetos como a RILA afirma que a mídia geralmente foca muito em como essas grandes obras podem transformar a região e trazer benefícios econômicos no futuro. Entretanto, os problemas do dia a dia que os transportadores enfrentam na logística e nas operações nem sempre recebem a mesma atenção nessas reportagens. Tal fato reforça que não basta apenas construir estradas e pontes. Mas também é muito importante investir em tecnologias e ferramentas que ajudem a resolver as dificuldades e as dúvidas que aparecem todos os dias no transporte de cargas. Por exemplo, conseguir prever com boa precisão quanto tempo uma viagem vai levar continua sendo uma grande necessidade para as empresas, mesmo que se fale muito sobre o desenvolvimento que novas rotas podem trazer.

A cadeia logística internacional envolve diversos agentes, como fabricantes, transportadoras, agentes aduaneiros e clientes finais. A previsibilidade no transporte é um fator determinante para garantir a eficiência desse sistema, permitindo que as empresas planejem estoques e prazos de maneira precisa (Servos et al., 2019). No entanto, a complexidade dos dados envolvidos, que incluem informações de rastreamento, variações climáticas e regulatórias, torna difícil prever com exatidão os tempos de entrega (Tsolaki et al., 2022).

Nesse cenário, tecnologias como o “Machine Learning” têm se destacado como ferramentas poderosas para resolver problemas de previsão no transporte internacional. Algoritmos avançados conseguem analisar grandes volumes de dados históricos e em tempo real, identificando padrões e fornecendo previsões mais precisas sobre os tempos de entrega. Isso não apenas reduz os custos operacionais, mas também aumenta a confiabilidade das empresas junto aos seus clientes (Hastie et al., 2009). Portanto, a modernização tecnológica é imprescindível para superar os desafios enfrentados pelo transporte rodoviário internacional de cargas. A adoção de ferramentas baseadas em dados não apenas otimiza a logística, mas também fortalece a competitividade das empresas brasileiras no mercado global, contribuindo para o crescimento econômico do Mercosul (ABTI, 2020).

Este trabalho teve como objetivo implementar algoritmos de Machine Learning em uma empresa brasileira de transporte rodoviário internacional no Mercosul, visando prever com precisão os prazos de entrega de todas as mercadorias em trânsito internacional.

Material e Métodos

O presente trabalho configura-se como um estudo de caso, investigando em profundidade a problemática da previsão de tempos de entrega no contexto específico de uma empresa brasileira de transporte rodoviário internacional de cargas com atuação no Mercosul. Distanciando-se da busca por generalizações amplas, o foco reside na compreensão detalhada da aplicabilidade e dos desafios da inteligência artificial neste cenário particular. Esta abordagem foi selecionada por permitir uma análise profundada dos dados reais da empresa, assegurando a relevância prática dos achados e a possibilidade de sua aplicação direta nas operações cotidianas da transportadora rodoviária internacional. Além disso, o estudo de caso viabiliza a observação não apenas do desempenho técnico dos modelos de aprendizado de máquina, mas tem por objetivo trazer benefícios ao negócio, como a otimização da utilização da frota e o aumento da satisfação do cliente. Ao concentrar-se em um exemplo de um caso real, busca-se extrair aprendizados sobre os sucessos e dificuldades encontradas nos dados estudados, podendo ser valiosos para outras empresas do mesmo

segmento. Desta forma, este estudo visa oferecer uma visão detalhada de como técnicas de aprendizado de máquina podem auxiliar na logística rodoviária internacional.

A base de dados primária para este estudo foi extraída de uma consulta SQL parametrizada, consolidando aproximadamente 5.300 registros de transportes internacionais realizados pela empresa no período de 01 de janeiro de 2024 a 31 de março de 2025. Os dados referentes aos primeiros três meses de 2025 foram estrategicamente separados para constituir um conjunto de teste cego, servindo como uma alternativa para validação dos modelos de Machine Learning desenvolvidos nos algoritmos.

A base de dados abrange um conjunto diversificado de variáveis descritivas das operações, incluindo informações detalhadas de origem e destino (país, unidade federativa, cidade, códigos específicos), características da carga (peso bruto, tipo), detalhes da frota (terceirizada ou própria), registros temporais de etapas logísticas cruciais (exemplo: data de agendamento de coleta, tempos de entrada e liberação em aduanas nos países de origem e destino, tempos para chegada e entrega ao destinatário), distâncias (remetente-fronteira, fronteira-destinatário, remetente-destinatário), cidade da fronteira brasileira utilizada no transporte e coordenadas geográficas (latitude e longitude) dos pontos de coleta e entrega. Uma hipótese relevante levantada pela empresa, e considerada na análise, é a de que motoristas residentes nas cidades onde o cruze internacional ocorre poderiam, por interesse pessoal, prolongar sua estadia, impactando negativamente o tempo total da viagem.

Ao ser analisada detalhadamente, a base de dados original apresentou certas inconsistências e lacunas que demandaram tratamento específico. A quilometragem total, por exemplo, não era consistentemente registrada através de odômetros no início e fim da viagem, sendo, portanto, estimada com base em distâncias padrões entre as localidades. Similarmente, o horário exato de início das viagens era incerto, dispondo-se apenas do registro de agendamento da coleta. Para padronizar este aspecto, e considerando que a maioria dos carregamentos era agendada para as 08h00 e usualmente levava quatro horas para ser concluída, todas as viagens foram consideradas como iniciadas às 12h00min. Registros com ausência de informações críticas, como datas de chegada ao cliente, datas de entrega efetiva, ou datas de ingresso e liberação aduaneira, foram desconsiderados para não comprometer a integridade da modelagem.

Uma transformação fundamental aplicada à variável alvo, `tempo_total` (originalmente em horas), foi sua conversão para uma escala em dias, utilizando um “divisor tempo” de valor igual a 24. Esta normalização visou alinhar as predições com uma unidade temporal mais intuitiva para o planejamento logístico e agrupar os tempos em dias. As predições geradas pelos modelos de aprendizado de máquina também foram submetidas a esta mesma transformação. Adicionalmente, foi adicionada a coluna `tempo_mediano_anterior`,

representando a mediana histórica dos tempos de trânsito (em dias) por par origem-destino, ordenada pela data de agendamento de coleta, para servir como variável preditora. Outras variáveis criadas incluíram um indicador binário para motoristas residentes na cidade de fronteira e o dia da semana do início do transporte e da chegada à fronteira.

Para explorar estruturas latentes nos dados que pudessem auxiliar na segmentação das viagens e potencialmente melhorar as previsões, foram implementadas duas estratégias de clusterização. A primeira, baseada no algoritmo “K-Means”, utilizou um conjunto de variáveis numéricas relevantes (excluindo-se a variável alvo e informações temporais diretas como ano, mês e dia) após padronização “Z-score”. A determinação do número ótimo de “clusters” (k) para o “K-Means” foi guiada pelo método do cotovelo (Elbow Method) apresentado na Figura 1, que analisa a soma dos quadrados “intra-cluster” (WCSS) para diferentes valores de k. Uma segunda abordagem de clusterização envolveu o uso de um “Autoencoder”, uma rede neural não supervisionada, para aprender representações compactas dos dados de entrada. Essas preditoras codificadas foram, então, submetidas ao algoritmo “K-Means”, utilizando um número de clusters informado pelo k ótimo do “K-Means”. As colunas resultantes destas clusterizações, grupo_kmeans e grupo_autoencoder_kmeans, foram incorporadas como variáveis preditoras categóricas no algoritmo, sendo subsequentemente transformadas em variáveis “dummy” para os modelos de machine learning.

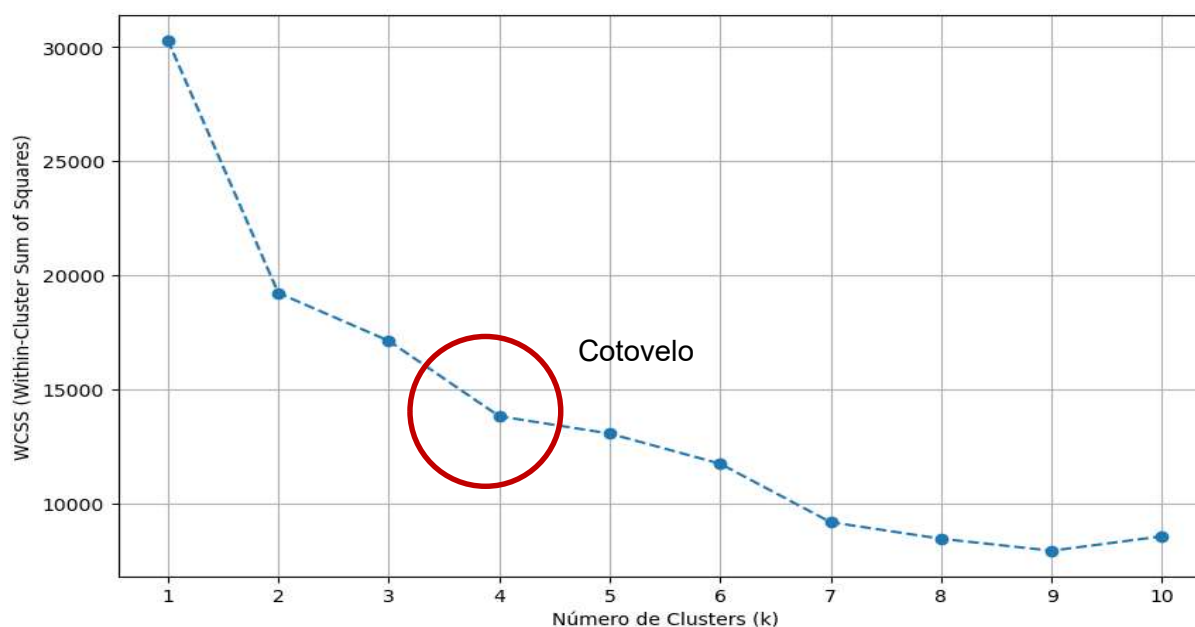


Figura 1. Detalhamento do Nº de Clusters e o Cotovelo (Método de Elbow)

Fonte: Dados originais da pesquisa

Todo o algoritmo foi construído em Python, o pré-processamento dos dados foi estruturado com parâmetros dinâmicos para simular diferentes estágios de disponibilidade de

informação ao longo de uma viagem, controlados pela variável `estado_da_viagem_do_modelo`. Os estados de viagem possíveis são de 1 a 4, onde cada estado implica a exclusão seletiva de informações que não estariam disponíveis operacionalmente (Tabela 1). No **Estado 1 ("Início de viagem")**, excluem-se todas as informações de eventos futuros à coleta. No **Estado 2 ("Veículo já ingressou na aduana de exportação")**, mantêm-se os dados até o ingresso na aduana de origem, excluindo-se os subsequentes. No **Estado 3 ("Veículo já liberou na aduana de importação")**, consideram-se os dados até a liberação na importação, removendo-se os relativos à chegada ao cliente. Finalmente, no **Estado 4 ("Veículo já chegou no cliente e aguarda descarga")**, a predição do tempo total ainda considera o histórico da viagem, pois pode ocorrer demora na descarga da mercadoria.

Tabela 1. Presença das variáveis preditoras em cada fase da viagem

Variável	Estado 1 Início Viagem	Estado 2 Ingressou Aduana Exp.	Estado 3 Liberou Aduana Imp.	Estado 4 Chegou Cliente
<code>dia_ingresso_expo</code>	X	✓	✓	✓
<code>horas_para_ingressar</code>	X	✓	✓	✓
<code>dia_liberacao_expo</code>	X	X	✓	✓
<code>horas_para_liberar_expo</code>	X	X	✓	✓
<code>horas_para_ingressar_impo</code>	X	X	✓	✓
<code>dia_ingresso_impo</code>	X	X	✓	✓
<code>dia_liberacao_impo</code>	X	X	✓	✓
<code>horas_para_liberar_impo</code>	X	X	✓	✓
<code>dia_chegada_cliente</code>	X	X	X	✓
<code>horas_para_chegar_cliente</code>	X	X	X	✓
			X	Ausente
			✓	Presente

Fonte: Dados originais da pesquisa

Independentemente do estado da viagem considerado, o pré-processamento envolveu tratamento de valores ausentes, conversão de tipos e remoção de registros com dados inválidos. Colunas com variância zero foram eliminadas. As variáveis categóricas remanescentes foram transformadas em variáveis “dummy” (“one-hot encoding” com “`drop_first=True`”), e os nomes das colunas foram sanitizados. A seguir Tabela 2 apresenta a cardinalidade das variáveis dumificadas.

Tabela 2. Cardinalidade das variáveis dumificadas

Variável Categórica	Valores Distintos
cod_destino	356
cod_origem	254
cidadeentrega_crt	136
ufcoleta	10
ufentrega	8
fronteira	8
dia_ingresso_impo	7
dia_liberacao_impo	7
dia_chegada_cliente	7
dia_ingresso_expo	7
dia_liberacao_expo	6
paisentrega	5
paiscoleta	5

Fonte: Dados originais da pesquisa

O conjunto de dados foi então dividido em treinamento (70%) e teste (30%), quando o parâmetro segregar dados de 2025 não foi ativado. Preditoras numéricas para modelos sensíveis à escala, como por exemplo peso bruto e distância, foram submetidas à padronização “Z-score”, com tratamento de valores nulos residuais.

Para a construção dos modelos preditivos, foi empregado um conjunto diversificado de técnicas de aprendizado de máquina para regressão. Este leque metodológico abrangeu desde métodos lineares, como Mínimos Quadrados Ordinários (OLS), até modelos baseados em árvores de decisão, incluindo o “DecisionTreeRegressor” (Hastie et al., 2009). Foram exploradas também abordagens de “ensemble” mais robustas, como “RandomForestRegressor” (Breiman, 2001), “XGBRegressor” (Friedman, 2001) e “LGBMRegressor”, que são otimizações de “Gradient Boosting”. A capacidade de aprendizado de redes neurais foi testada com uso do “MLPRegressor” e de uma Rede Neural Densa (Dense NN) implementada com Keras (Goodfellow et al., 2016). Complementarmente, aplicou-se “Support Vector Regression” (SVR) com kernel linear.

A fim de mitigar o risco de sobreajuste (“overfitting”) e obter uma estimativa mais robusta do desempenho dos modelos em dados novos, foi utilizada a técnica de validação cruzada K-Fold (Stone, 1974). Neste procedimento, o conjunto de treinamento foi dividido em K subconjuntos (“folds”) de tamanho aproximadamente igual. Os modelos foram iterativamente treinados em K-1 subconjuntos e avaliados no subconjunto restante, repetindo-se o processo K vezes, de forma que cada subconjunto serviu como conjunto de teste exatamente uma vez.

A otimização de hiper parâmetros, controlada pelo parâmetro “executar gridsearches aprofundados”, foi realizada dentro deste esquema de validação cruzada através de

“GridSearchCV”. Nos casos em que a busca exaustiva em grade não foi ativada, os modelos foram configurados com hiper parâmetros pré-definidos, baseados em conhecimento prévio de execuções anteriores.

A performance final de cada modelo foi medida utilizando o coeficiente de determinação (R^2) e a Raiz do Erro Quadrático Médio (RMSE), tanto para o conjunto de treinamento quanto para o de teste.

Adicionalmente, para fornecer uma perspectiva da precisão operacional do modelo na escala da variável alvo transformada (dias), calculou-se a “Acurácia em Intervalos de Predição”. Essa métrica representa o percentual de previsões que, após serem submetidas à mesma transformação de discretização temporal que a variável alvo, coincidem com o intervalo de tempo (dia) observado. Este tipo de avaliação é comum quando a saída de um modelo de regressão é mapeada para classes ou intervalos discretos para fins práticos. A discretização de saídas contínuas para avaliação compartilha princípios com as técnicas de discretização de variáveis preditoras amplamente discutidas na literatura (Dougherty, Kohavi, & Sahami, 1995).

Todos os resultados, parâmetros de configuração do algoritmo e os melhores hiper parâmetros de modelo (encontrados ou utilizados) foram registrados sistematicamente para análise comparativa e reprodutibilidade.

A avaliação do desempenho dos modelos de aprendizado de máquina foi conduzida através de quatro cenários distintos, correspondentes aos valores 1 a 4 do parâmetro criado no algoritmo `config_estado_da_viagem_do_modelo`. Estes estados simulam diferentes momentos da operação logística, caracterizados por uma progressiva disponibilidade de informações sobre o andamento da viagem.

Adicionalmente, foram exploradas quatro configurações principais de pré-processamento e dados para cada modelo:

Config A ("2 Clusters + Y Norm."): Utiliza duas preditoras de clusterização e normalização da variável alvo (Y). O R^2 de Treino, quando mencionado, refere-se exclusivamente a esta configuração.

Config B ("Sem Cluster + Y Original"): Não aplica a clusterização e utiliza a variável alvo em sua escala original.

Config C ("Base Treino 2024 - Teste 2025"): Avalia a generalização do modelo em dados exclusivamente de teste referente aos três primeiros meses de 2025, após treinamento em dados completos dos doze meses do ano de 2024. A ideia aqui foi descobrir se os modelos baseando-se exclusivamente em dados de 2024 poderiam prever os dados de 2025.

Config D ("Sem Preditor Tempo Mediano"): Esta configuração excluiu a variável `tempo_mediano_anterior`. Embora a construção desta preditora no algoritmo principal

tenha sido cuidadosamente desenhada para usar apenas dados históricos estritamente anteriores a cada viagem (evitando assim o vazamento da variável alvo), esta configuração permitiu testar a sensibilidade dos modelos à presença desta informação histórica e sua relevância preditiva.

Resultados e Discussão

Após sanitização da base de dados, 3.739 registros foram eleitos como válidos para o estudo, tal amostra é composta integralmente por viagens internacionais, todas com origem ou destino no Brasil. O país com maior número de viagens é a Argentina, representando 78% dos registros de transporte (Tabela 3), seguido pelo Uruguai, Chile e Paraguai. Em relação às demais características operacionais, 54% dos transportes foram realizados com veículos terceirizados. Do total movimentado, 95% ocorreu em carretas, sendo que 16% do volume refere-se a cargas químicas, enquanto o restante é considerado “carga geral”, normalmente paletizadas. A quantidade total de viagens para cada país, distância média, peso e tempos de liberação em aduana, assim como tempo total de viagem, podem ser observadas a seguir.

Tabela 3. Principais variáveis e indicadores da base de dados em estudo

	Argentina	Chile	Uruguai	Paraguai	Total
Quantidade de viagens	2.918	375	365	81	3.739
% Representativo	78%	10%	10%	2%	100%
Distância média (km)	1.687	2.909	1.110	1.297	1.745
Peso médio (Kg)	16.013	21.293	14.594	19.950	16.490
Tempo médio em horas para ingressar expo	64	98	72	95	69
Tempo médio em horas para liberar expo	5	6	7	19	6
Tempo médio em horas para ingressar impo	32	102	17	11	37
Tempo médio em horas para liberar impo	53	39	23	39	49
Tempo médio total (variável y do estudo)	203	293	157	189	208

Fonte: Resultados originais da pesquisa

Em seguida a avaliação do desempenho dos diversos modelos de aprendizado de máquina foi realizada considerando quatro cenários operacionalmente distintos, controlados pela variável `estado_da_viagem_do_modelo` (variando de 1 a 4), e quatro principais configurações de pré-processamento e dados (denominadas Configurações A, B, C e D).

A seguir, cada algoritmo é analisado individualmente, detalhando o impacto em R^2 das diferentes configurações e estados da viagem do veículo. Ao final foi discutida a métrica de "% Acurácia em Intervalos de Predição" e de RMSE (Raiz do erro quadrático médio) complementando a análise.

OLS “Selected Features” (Regressão de Mínimos Quadrados Ordinários)

A construção desse modelo foi motivada, pela sua simplicidade, levando em conta propositalmente somente variáveis originalmente numéricas da base de dados, não fazendo parte de sua fórmula colunas que sofreram dumerização. A expressão deste modelo no estado de viagem 3 é apresentada abaixo por meio da form. (1):

$$\begin{aligned} \text{tempo} \sim & \text{horas_para_ingressar} + \text{horas_para_liberar_expo} + \text{horas_para_ingressar_impo} \\ & + \text{horas_para_liberar_impo} + \text{mês} + \text{pesobruto} + \text{distancia} + \text{ano} \\ & + \text{tempo_mediano_anterior} \end{aligned} \quad (1)$$

O modelo OLS com “features” selecionadas, não sofre impacto entre a variações das configurações A e B (Figura 2), pois não leva em conta variáveis dumificadas. Este modelo demonstrou um desempenho respeitável saindo de um R^2 de Teste de 7% no Estado 1 alcançando 84% nos Estados 3 e 4. A avaliação na base de teste de 2025 (Config C) resultou em uma performance ligeiramente inferior, mas ainda robusta, com R^2 de Teste máximo de 78%. Na configuração D ao excluir o tempo_mediano_anterior também não gerou efeito significativo no R^2 . Isso indica que o OLS com features selecionadas é um modelo estável demonstrando pequena redução de performance ao se utilizar dados de teste exclusivos de 2024.

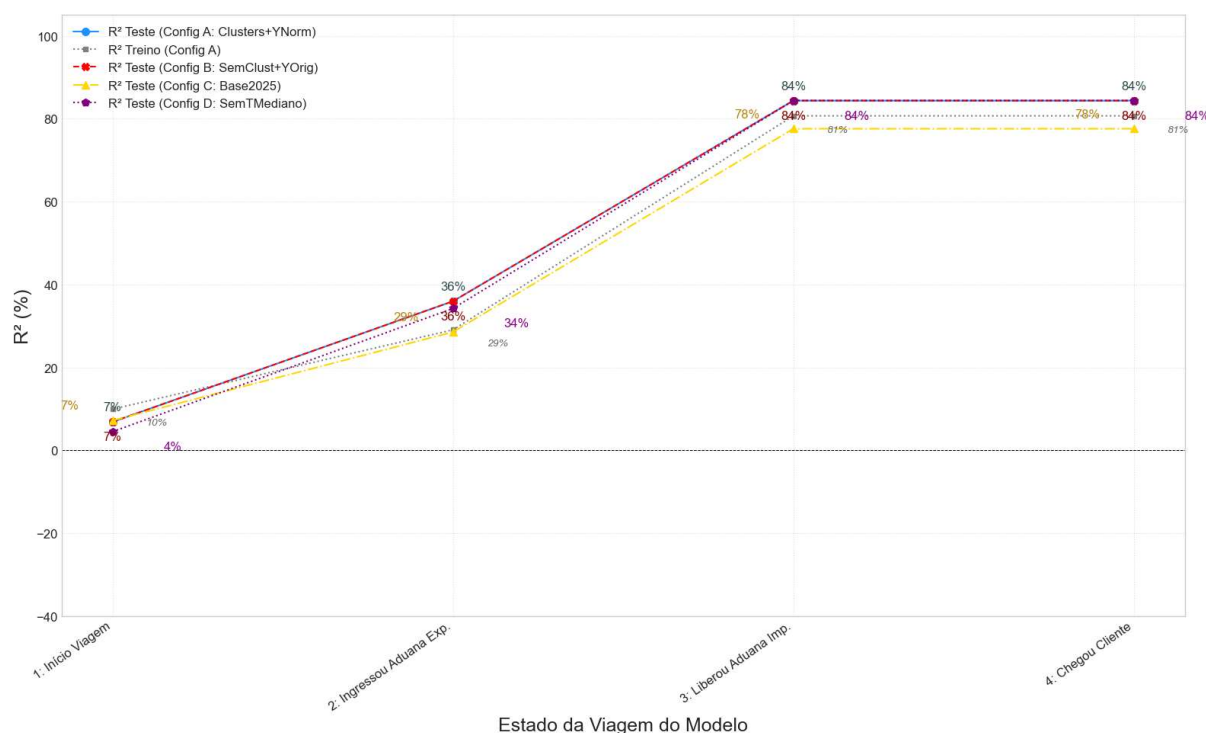


Figura 2. Comparativo R^2 usando OLS – Selected Features
Fonte: Resultados originais da pesquisa

OLS com Todas as Features (Stepwise)

Diferentemente do modelo anterior (OLS “Selected Features”), a ideia aqui foi desenvolver um modelo OLS completo utilizando todas as preditoras possíveis, tal qual os demais modelos. Além disso, foi utilizada uma técnica para seleção regressiva de features (“backward stepwise”), utilizando um limiar de p-valor de 0,05 para retenção de variáveis.

Em estados da viagem específicos, o modelo demonstrou alta capacidade preditiva (Figura 3). Na Configuração A e B o R^2 de Teste alcançou um expressivo valor de 88% no Estado 3 (“Liberou Aduana Imp.”). Para a Configuração C (teste em dados de 2025), o modelo também atingiu R^2 de Teste considerável de 84% e 83% nos Estados 3 e 4, respectivamente, indicando potencial de generalização para períodos futuros.

No entanto, o modelo apresentou instabilidade em outras situações. Na Configuração A, o R^2 de Teste no Estado 4 caiu para -84%, e R^2 de Teste iniciais (Estado 1) foi de -15% na Config B. Os R^2 de Treino demonstraram-se estáveis e elevados (94%) e a discrepância com os valores de teste em certos pontos sugere que o “overfitting” não foi completamente eliminado pela seleção “stepwise” em todos os cenários.

Embora o uso de “stepwise” com p-valor de 0,05 tenha refinado as variáveis preditoras, os resultados indicam que a complexidade dos dados de tempo de trânsito pode limitar a performance de modelos lineares, mesmo após otimização das features.

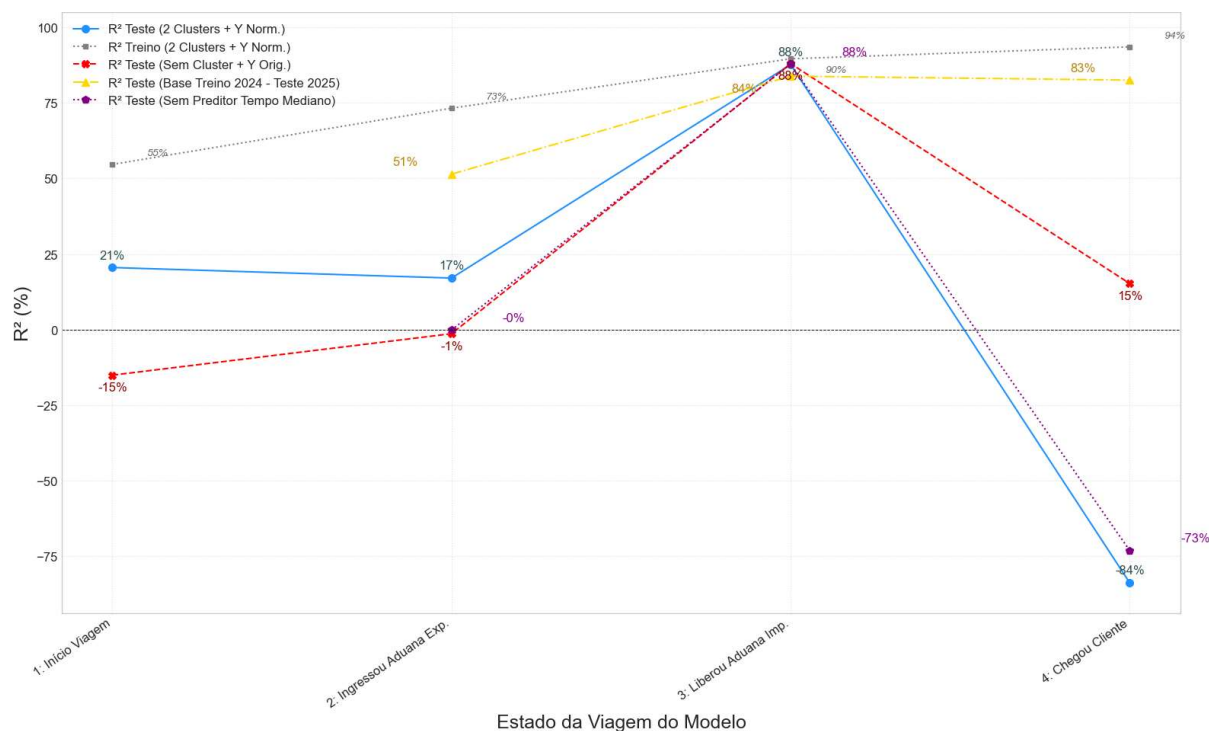


Figura 3. Comparativo R^2 usando OLS com Todas as Features

Fonte: Resultados originais da pesquisa

Shapiro-Francia e Distribuição dos resíduos nos modelos OLS

Além da análise de desempenho através de R^2 , foi investigada a normalidade dos resíduos dos dois modelos de OLS, a fim de validar suas premissas estatísticas. Conforme recomendado para amostras que possuem mais de 30 observações (Fávero; Belfiore, 2024), foi aplicado o teste de Shapiro-Francia, que resultou em um p-valor extremamente baixo ($p < 0,001$), levando à rejeição da hipótese nula de normalidade H_0 . Essa violação pode ser comum em estudos aplicados que utilizam regressão por mínimos quadrados ordinários conforme o autor supracitado. A não aderência dos resíduos à normalidade pode indicar erro na formulação dos modelos ou omissão de variáveis explicativas importantes. Tal constatação reforça a importância de explorar algoritmos de machine learning mais flexíveis, como os baseados em árvores ou redes neurais, que não dependem desta suposição.

A análise visual do histograma de resíduos do modelo OLS com features selecionadas (Figura 4) e do modelo OLS com Todas as Features (Figura 5), corroboram o resultado do teste de Shapiro-Francia, evidenciando o desvio da normalidade dos resíduos dos modelos. As distribuições apresentam uma forte assimetria, com uma cauda à direita. Este padrão indica que, embora os modelos errem por pouco na maioria das previsões (concentração de resíduos em torno de zero), eles tendem a cometer erros prevendo viagens cujo tempo (em horas) real é significativamente maior do que o previsto pelo algoritmo. A forma da distribuição, afastando-se da curva de sino simétrica, sugere que os modelos OLS possuem dificuldade em capturar toda a complexidade e variabilidade dos tempos de viagem.

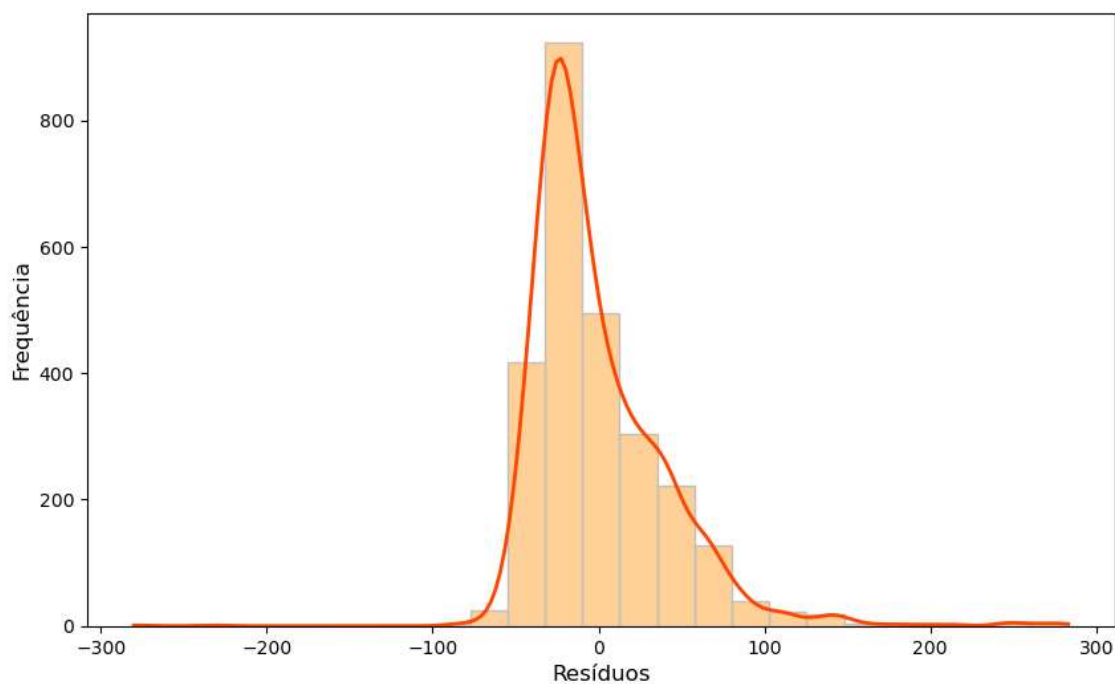


Figura 4. Histograma dos resíduos OLS Selected Features
Fonte: Resultados originais da pesquisa

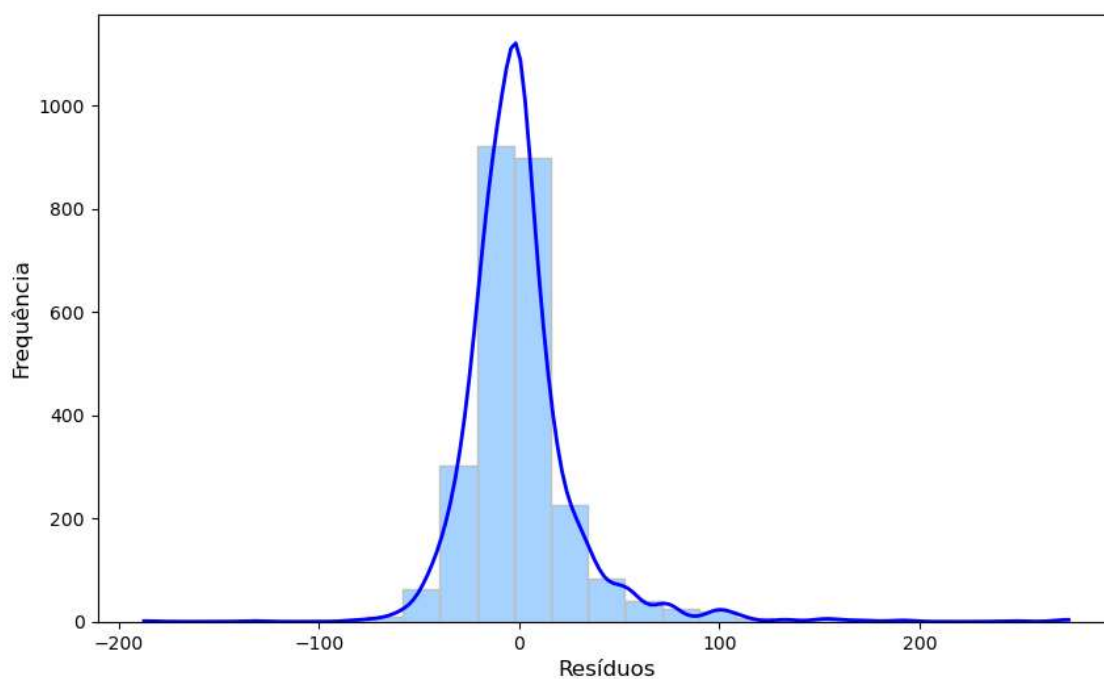


Figura 5. Histograma dos resíduos OLS com Todas as Features (Stepwise)
Fonte: Resultados originais da pesquisa

Mediana por Origem Destino (Modelo de referência)

O modelo de referência "Mediana Origem Destino", como esperado de um método não supervisionado simples, apresentou R^2 de Teste modestos (Figura 6). Nas Configs A ("2 Clusters + Y Norm.") e B ("Sem Cluster + Y Orig."), os resultados foram idênticos, com R^2 s de Teste em torno de 6% no Estado 2 e 12% nos Estados 3 e 4 (sem dados significativos para o Estado 1 na Config A). Quando avaliado na base de teste de 2025 (Config C), os R^2 s de Teste foram ligeiramente superiores, variando entre 14% (Estado 1) e 16% (Estado 2). Não houve dados para a Config D ("Sem Preditor Tempo Mediano"). Estes resultados confirmam o papel deste modelo como um ponto de referência simples, amplamente superado pelos modelos de machine learning mais complexos.

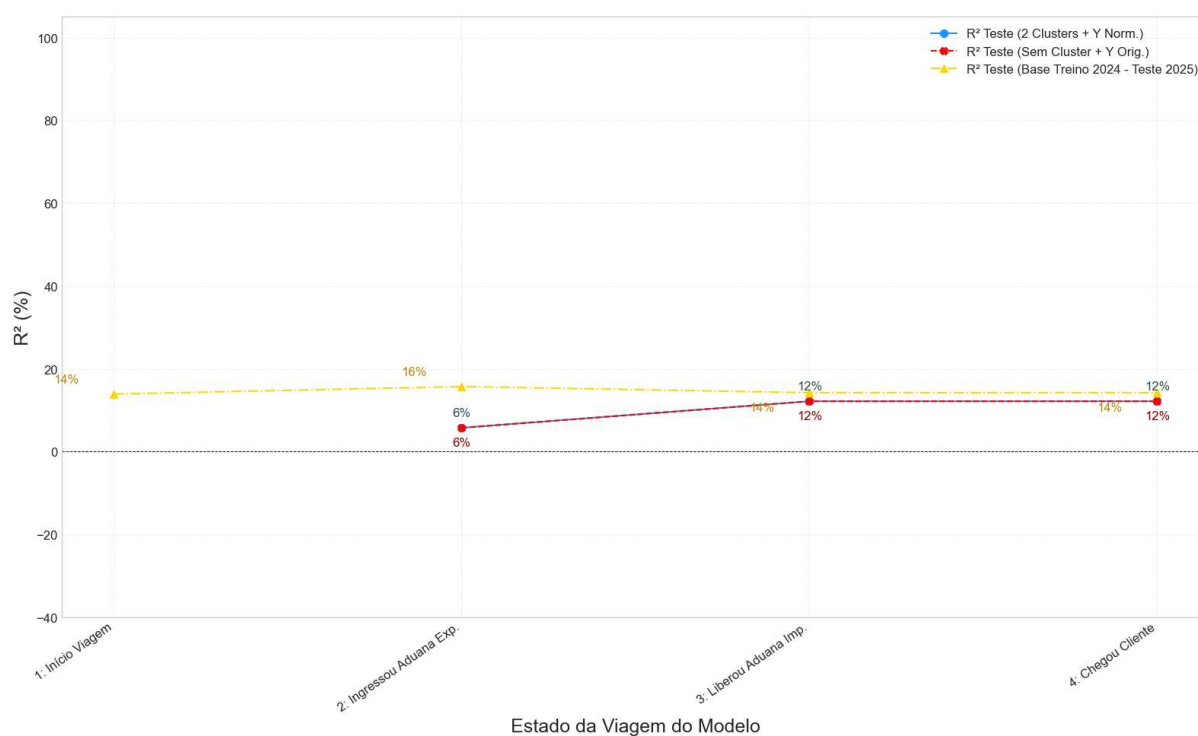


Figura 6. Comparativo R^2 usando Mediana por par Origem-Destino
Fonte: Resultados originais da pesquisa

DecisionTreeRegressor

O “DecisionTreeRegressor”, otimizado via “GridSearch”, demonstrou sensibilidade às diferentes configurações de pré-processamento. Para a configuração que inclui '2 Clusters + Y Norm.' (Config A), o R^2 de Teste evolui de 19% no Estado 1 para um bom 83% no Estado 4 (Figura 7), indicando que estas preditoras de pré-processamento são benéficas, especialmente nas etapas finais da viagem. O R^2 de Treino de 40% associado a esta Config A é consistentemente mais alto que seu R^2 de Teste 19% no Estado 1, sugerindo um certo grau de “overfitting”, embora a diferença diminua nos estados de viagem mais avançados. A remoção da clusterização e da normalização do Y (Config B: "Sem Cluster + Y Orig.") resultou em um R^2 de Teste muito similar aos da Config A (83% no Estado 4), sugerindo que, para este modelo específico, a clusterização e a normalização do Y não surtiram efeitos. Em contrapartida, a avaliação com a base de teste de 2025 (Config C) revelou uma queda acentuada no desempenho, com o R^2 de Teste máximo atingindo apenas 67% no Estado 4, o que aponta para uma dificuldade de generalização do modelo usando dados de treinamento apenas de 2024. A remoção da preditora de tempo mediano (Config D) mostrou um leve ganho no R^2 de Teste no Estado 1 (24%) e manteve resultados comparáveis às Configs A e B nas demais etapas, indicando que esta preditora específica pode não ser crucial ou até introduzir algum ruído para o “Decision Tree” neste contexto.

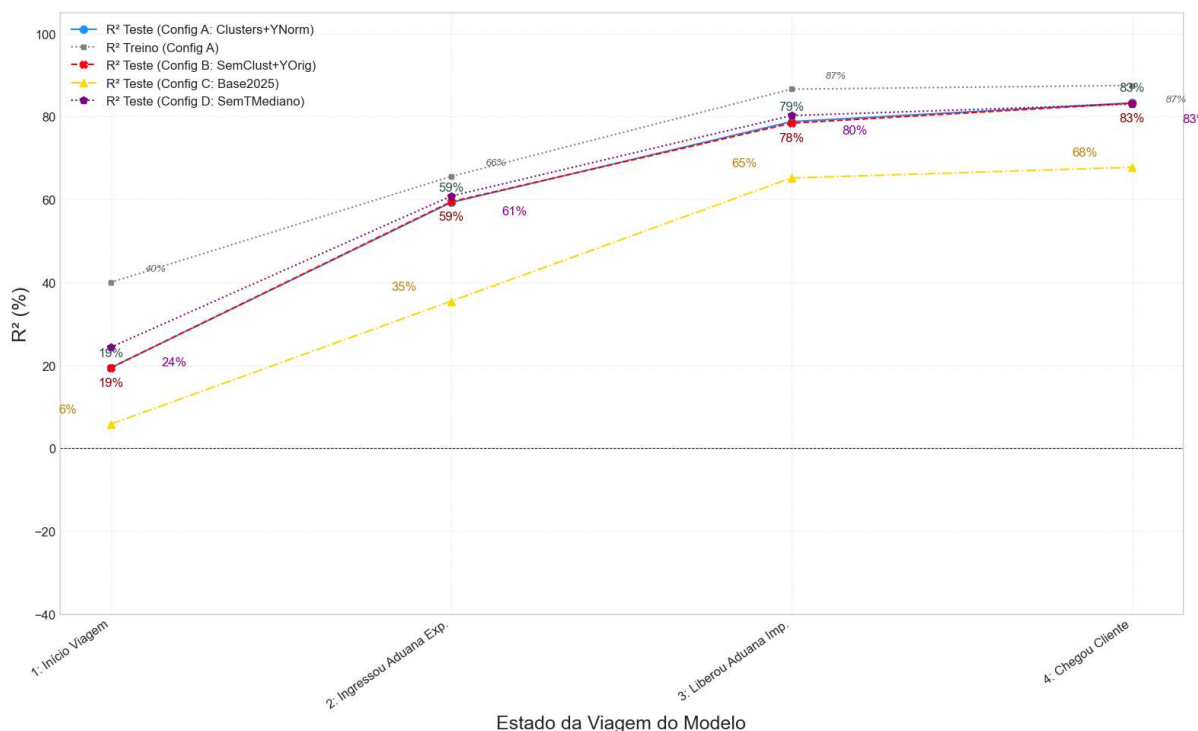


Figura 7. Comparativo R^2 usando DecisionTreeRegressor

Fonte: Resultados originais da pesquisa

RandomForestRegressor

O RandomForestRegressor, assim como o LGBM, mostrou-se um modelo de alta performance e considerável equilíbrio (Figura 8). Na Config A ("2 Clusters + Y Norm."), alcançou R^2 de Teste excelentes, variando de 37% no Estado 1 a 90% no Estado 4. Os R^2 s de Treino para esta Config A (97% no Estado 4) indicaram um bom aprendizado sem overfitting. A Config B ("Sem Cluster + Y Orig.") teve um impacto mínimo, com R^2 s de Teste muito próximos (90% no Estado 4). A avaliação na base de teste de 2025 (Config C) resultou em uma queda de performance, iniciando com R^2 de Teste negativo (-5%) no Estado 1 mas atingindo um máximo de 81% no Estado 4, demonstrando um bom resultado, mas evidencia sensibilidade à mudança no período dos dados. A Config D ("Sem Preditor Tempo Mediano") também manteve um desempenho forte sendo muito similar a Config A e B. Isso posiciona o RandomForest como uma excelente opção, não sendo muito sensível às variações de pré-processamento de cluster, normalização de Y, e à presença do preditor de tempo mediano, mas, como outros modelos, é afetado pela mudança na distribuição dos dados de teste para um novo ano.

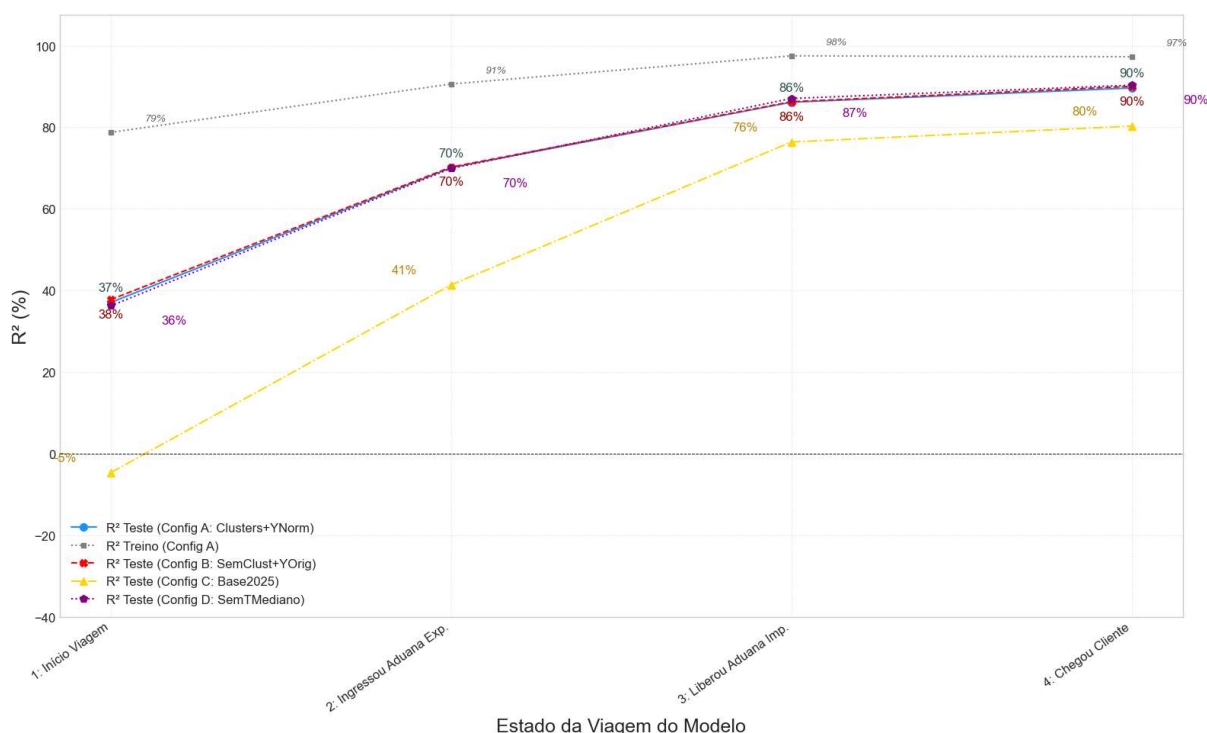


Figura 8. Comparativo R^2 usando RandomForestRegressor

Fonte: Resultados originais da pesquisa

XGBRegressor

O XGBRegressor ou XGBoost se consolidou como um dos modelos de melhor desempenho, comparável ao LGBM e RandomForest. Com a Config A ("2 Clusters + Y Norm."), entregou R^2 de Teste excelentes, progredindo de 34% no Estado 1 para surpreendentes 93% no Estado 4 (Figura 9). Os R^2 s de Treino para esta Config A (98% no Estado 4) foram altos e bem próximos em relação ao teste, indicando um bom aprendizado. A Config B ("Sem Cluster + Y Orig.") teve impacto mínimo, com R^2 de Teste quase idênticos (93% no Estado 4). A avaliação na base de teste de 2025 (Config C) demonstrou uma queda esperada na performance, com um R^2 de Teste máximo de 85% no Estado 4 – ainda um resultado muito bom, mas que reflete a dificuldade em generalizar para um período de dados fora da amostra. A Config D ("Sem Preditor Tempo Mediano") resultou na melhor performance entre todas as configurações para este modelo, atingindo 93% de R^2 de Teste no Estado 4. No geral, o XGBoost destacou-se com a mais alta performance entre os modelos e relativa insensibilidade às etapas de pré-processamento de cluster, e parece se beneficiar da remoção do preditor de tempo mediano.

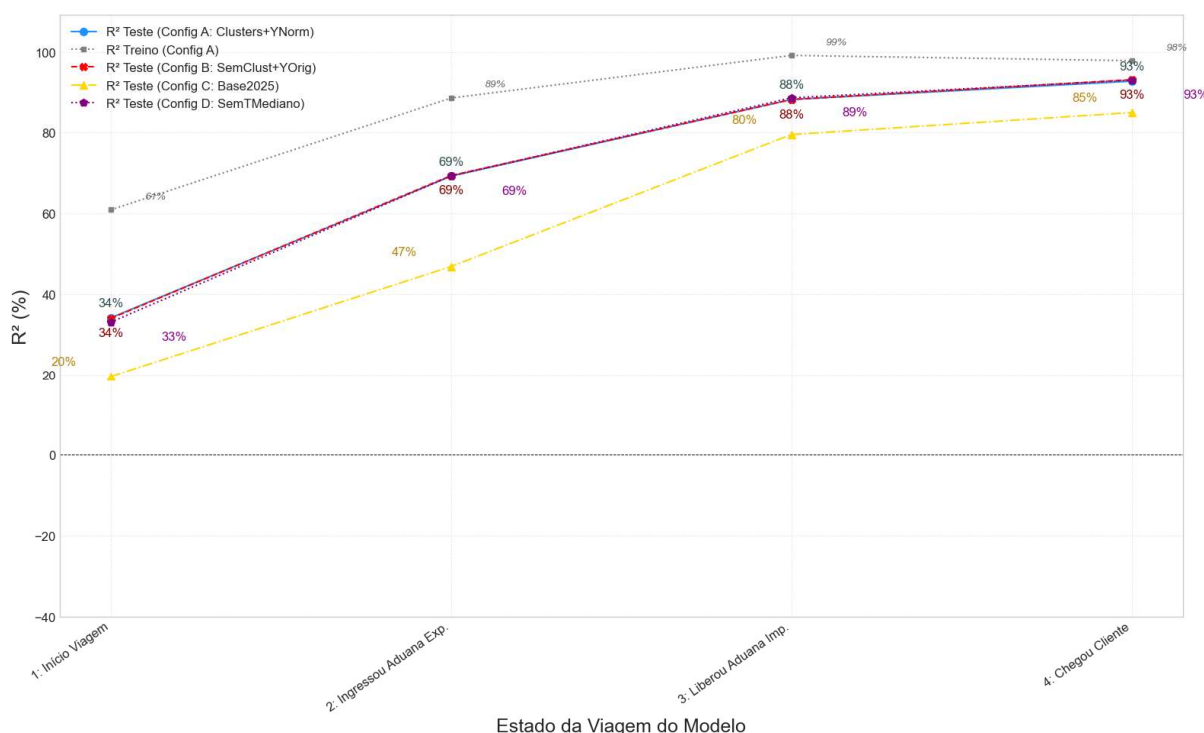


Figura 9. Comparativo R^2 usando XGBRegressor

Fonte: Resultados originais da pesquisa

LGBMRegressor

O LGBMRegressor apresentou performance igualável ao XGBRegressor, tanto na Config A ("2 Clusters + Y Norm.") como na Config B ("Sem Cluster + Y Orig.") (Figura 10). O R^2 de Treino e de Teste atingiram expressivos 93%. Os índices sugerem que o LGBM lida eficientemente com os dados mesmo sem as clusterizações e normalização do Y. A avaliação na base de teste de 2025 (Config C) resultou em uma queda na performance, atingindo um R^2 de Teste máximo de 85% no Estado 4 – ainda um resultado forte, mas que indica sensibilidade à mudança na distribuição dos dados. A Config D ("Sem Preditor Tempo Mediano") também sustentou uma performance elevada, comparável às Configs A e B, com 92% de R^2 de Teste no Estado 4. Estes resultados apontam o LGBM como um modelo poderoso para este problema, com as etapas de clusterização, normalização de Y e a presença do preditor de tempo mediano não se mostrando significativas para seu sucesso, embora a mudança de ano na base de teste (Config C) tenha um impacto negativo considerável.

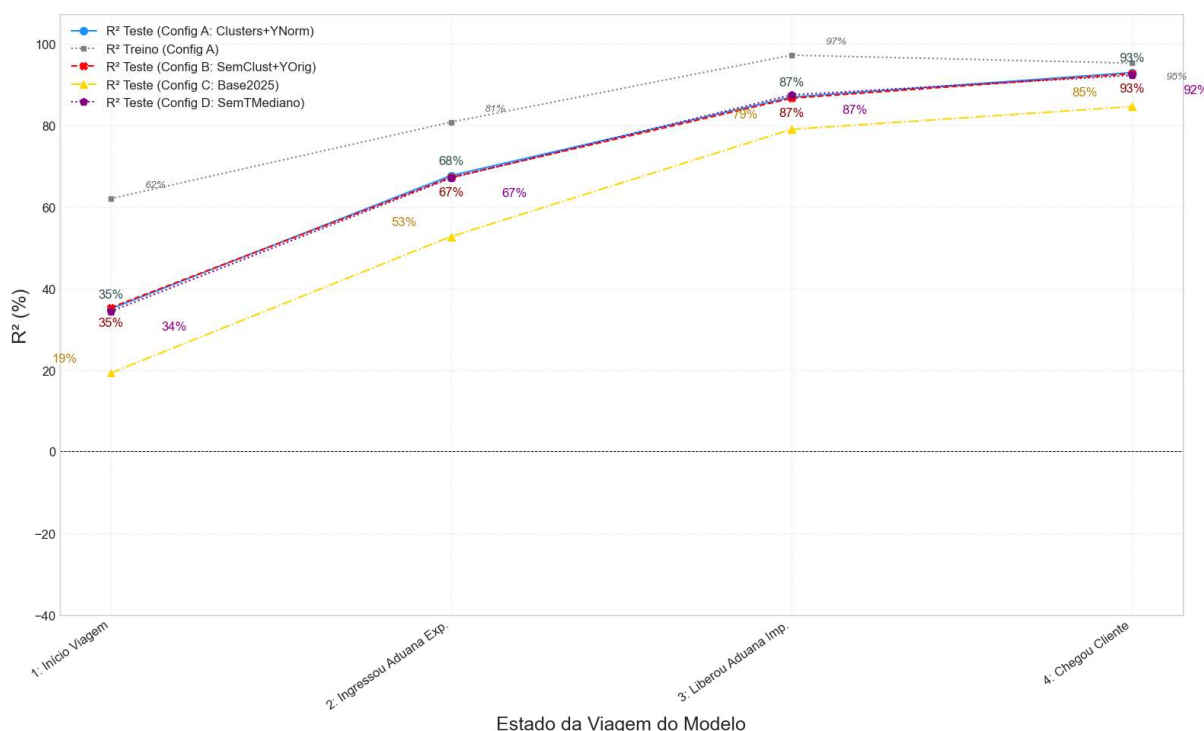


Figura 10. Comparativo R^2 usando LGBMRegressor
Fonte: Resultados originais da pesquisa

MLPRegressor

O MLPRegressor, otimizado com GridSearchCV, demonstrou boa capacidade de aprendizado na Config A ("2 Clusters + Y Norm."), com o R^2 de Teste progredindo de 21% no Estado 1 para 90% no Estado 4 (Figura 11). Os R^2 s de Treino para esta Config A (91% no Estado 3) indicaram um bom ajuste, sem overfitting excessivo. Em contraste, a Config B ("Sem Cluster + Y Orig.") resultou em uma degradação significativa da performance, com R^2 s de Teste consideravelmente mais baixos e irregulares (23% no Estado 4), sublinhando a importância da clusterização e normalização do Y para este modelo. A utilização da base de teste de 2025 (Config C) apresentou um desempenho inferior atingindo 86% no Estado 4. A Config D ("Sem Preditor Tempo Mediano") produziu um R^2 de Teste de 88% no Estado 4, muito similar à Config A, indicando que o impacto desta feature específica não foi determinante quando as outras otimizações estavam presentes.

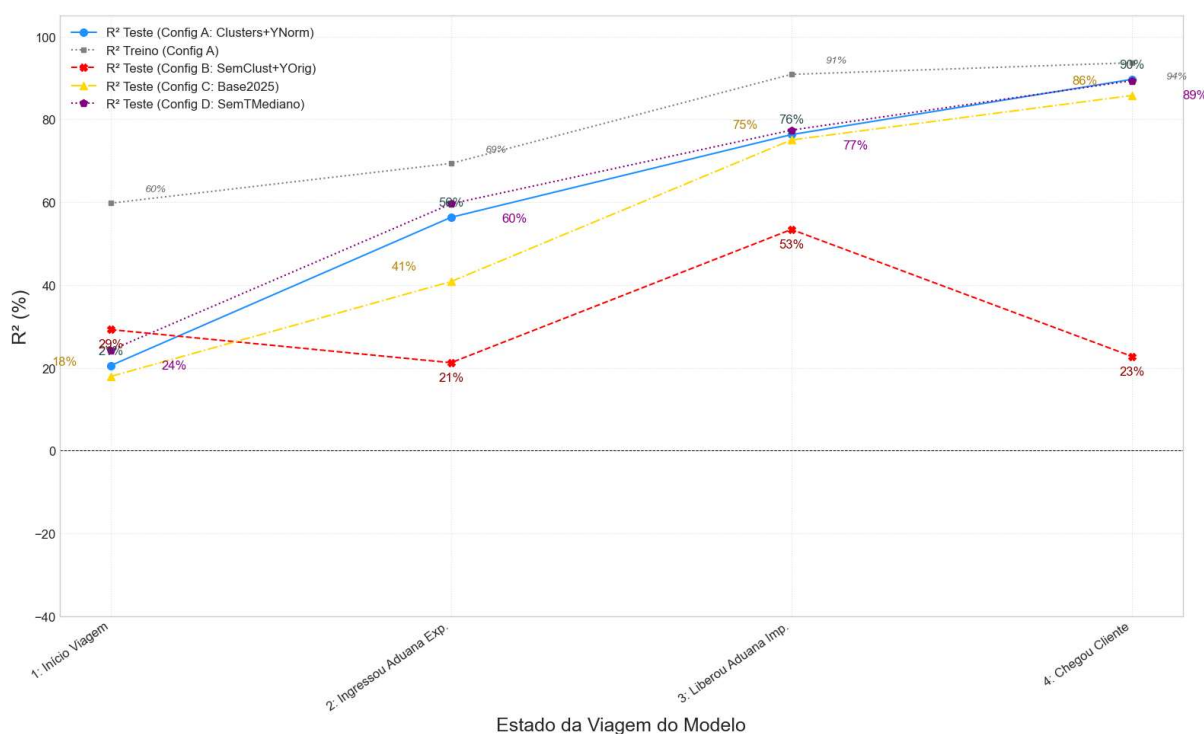


Figura 11. Comparativo R^2 usando MLPRegressor

Fonte: Resultados originais da pesquisa

Rede Neural Densa Keras

A Rede Neural Densa implementada com Keras exibiu uma variação de performance notável entre as configurações (Figura 12). Utilizando "2 Clusters + Y Norm." (Config A), o modelo alcançou um R^2 de Teste de 67% no Estado 3 ("Liberou Aduana Imp."), elevando-se para 75% no Estado 4 ("Chegou Cliente"). A Config B ("Sem Cluster + Y Orig.") melhorou o R^2 de Teste no Estado 3 para 79%, porém resultou em um R^2 de Teste insatisfatório de 44% no Estado 4, indicando excessiva especialização nos dados de treino. A transição para a base de teste de 2025 (Config C) inicialmente produziu R^2 de Teste negativo (-59% no Estado 1), mas demonstrou uma recuperação robusta, atingindo 74% no Estado 4. A ausência do preditor de tempo mediano (Config D) resultou em um desempenho similar a configuração A, com R^2 de Teste de 51% e 69% nos Estados 3 e 4, entretanto apresentou problema de overfitting no último estado da viagem.

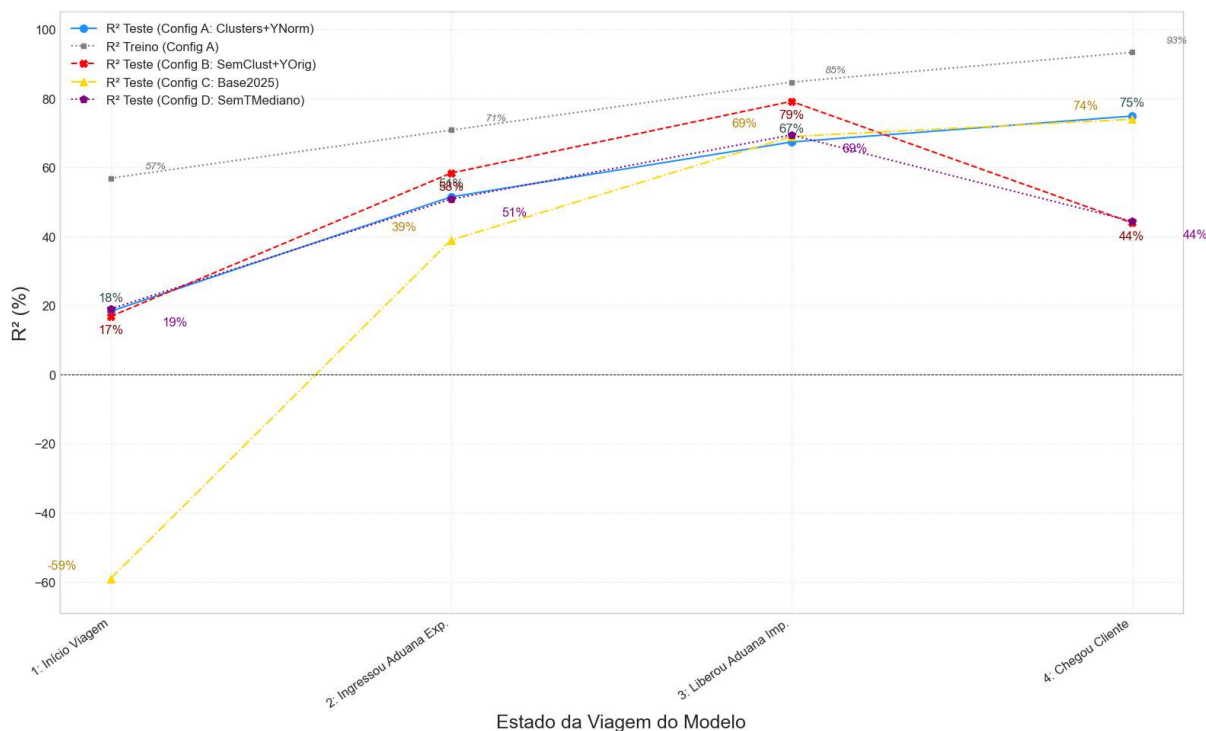


Figura 12. Comparativo R^2 usando Rede Neural Densa Keras
Fonte: Resultados originais da pesquisa

Support Vector Regressor

O SVR com kernel linear apresentou um comportamento misto (Figura 13). Na Config A ("2 Clusters + Y Norm."), o R^2 de Teste evoluiu de 18% no Estado 1 para um bom 89% no Estado 3, mas sofreu uma queda acentuada para 23% no Estado 4. Os R^2 s de Treino para esta Config A (92% no Estado 4) permaneceram altos, indicando overfitting na última etapa. A Config B ("Sem Cluster + Y Orig.") mostrou R^2 s de Teste inicialmente melhores no Estado 1 (28%) e manteve-se bem no Estado 3 (89%), mas também teve uma queda expressiva no Estado 4 (-6%). A base de teste de 2025 (Config C) foi particularmente incomum, iniciando com R^2 de Teste de -110%, mas, curiosamente, recuperando-se para 86% e 89% nos Estados 3 e 4, superando as outras configurações nessas etapas para dados de 2025. A Config D ("Sem Preditor Tempo Mediano") resultou em um padrão de R^2 de Teste muito similar à Config A, incluindo a queda no Estado 4 (23%). O SVR Linear parece ser sensível à etapa final da viagem, e seu desempenho pode variar consideravelmente com as features e a distribuição dos dados.

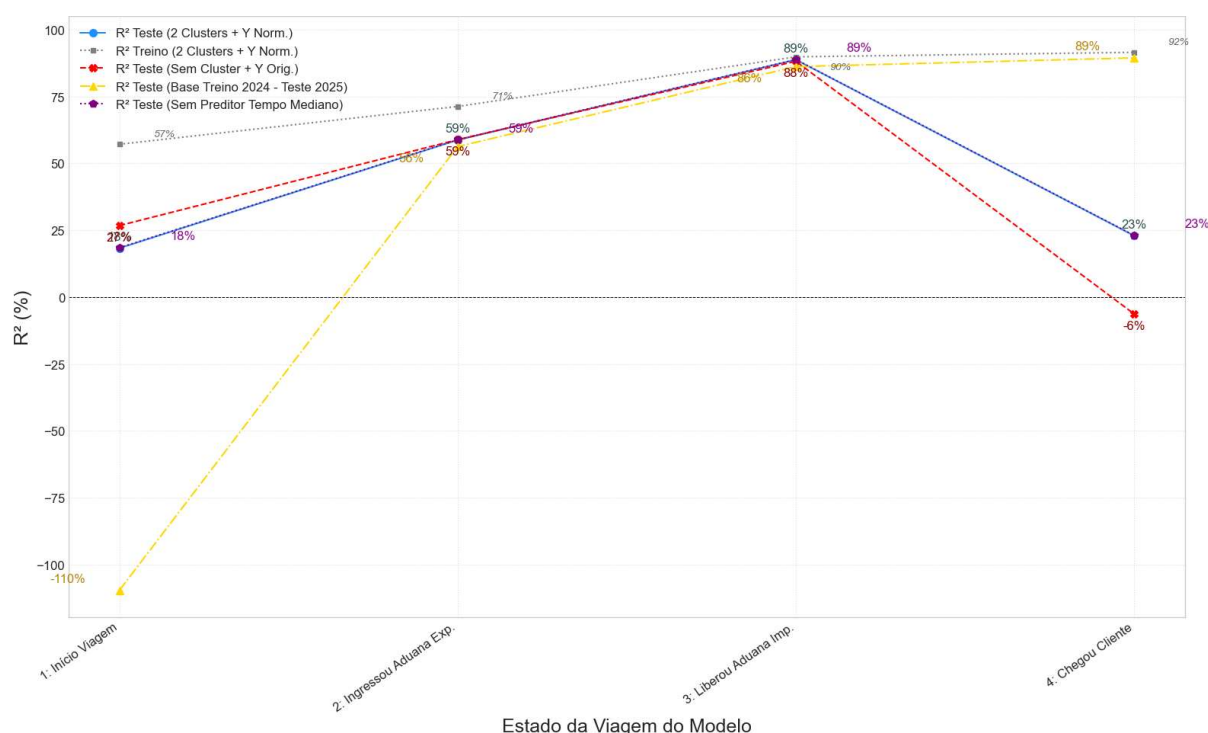


Figura 13. Comparativo R^2 usando Support Vector Regressor
Fonte: Resultados originais da pesquisa

% Acurácia em Intervalos de Predição

Essa métrica oferece uma importante visão sobre a utilidade operacional dos modelos, indicando a frequência com que as previsões de tempo das fases do transporte coincidiram com o intervalo temporal correto. A análise foi aprofundada ao comparar o desempenho do modelo XGBRegressor (Configuração A) sob três intervalos de tempo distintos: 24, 48 e 72 horas. Os resultados apresentados na Figura 14 revelam que em todos os intervalos a acurácia evolui de forma positiva à medida que a viagem avança e mais informações se tornam disponíveis. Os intervalos de tempo maiores resultam em uma maior acurácia. Na janela de 24 horas, a acurácia progride de 20% no "Início da Viagem" para 59% na etapa final "Chegou Cliente". Já no intervalo de 72 horas, o modelo demonstra uma robustez notável, partindo de 46% atingindo um índice de acerto de 84% no final da viagem.

Essa análise comparativa de parametrizações distintas fornece à empresa flexibilidade na implantação dos algoritmos, permitindo a escolha deliberada entre um cenário de maior precisão (janela menor) e um de maior confiabilidade (janela maior). O alto desempenho do modelo especialmente após a liberação da aduana (Estado 3), confirma que o modelo é uma ferramenta valiosa para otimizar a logística e a comunicação com o cliente.

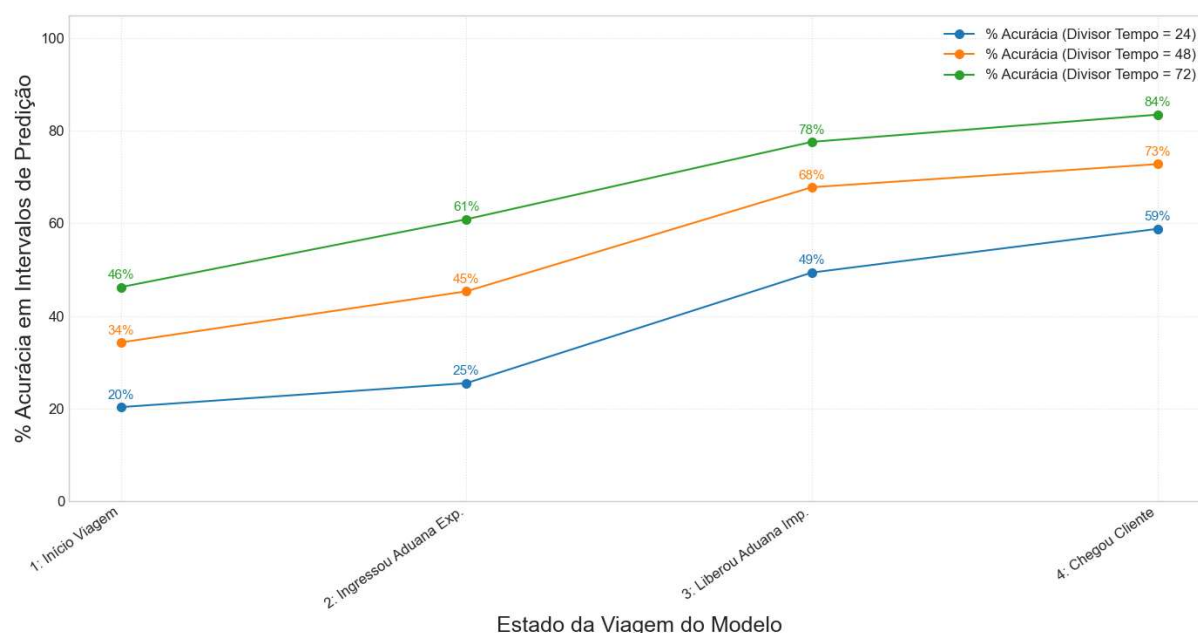


Figura 14. Comparativo do % Acurácia em Intervalos de Predição usando XGBRegressor
Fonte: Resultados originais da pesquisa

Os altos percentuais de acurácia em intervalos de predição, como os observados para o XGBRegressor, indicam que o modelo não só possui um bom ajuste aos dados (conforme refletido pelo R^2), mas também consegue prever com qualidade os tempos de transporte, o que é fundamental para o planejamento logístico da empresa. A análise desta métrica para

outros modelos e configurações revelaram padrões semelhantes embora com variações na magnitude do acerto.

A tabela 4 apresenta os valores encontrados de RMSE (Raiz do erro quadrático médio) encontrados para cada modelo desenvolvido. Esses valores indicam a magnitude do erro de previsão em dias.

Tabela 4. Demonstrativo dos RMSEs apresentados pelos modelos em cada configuração

Config	Estado	OLS Selected	OLS StepWise	Mediana	Decision Tree	Random Forest	XGB	LGBM	MLP	KERAS	SVR
A	1	4	3,7	6,8	3,7	3,3	3,3	3,3	3,7	3,7	3,7
A	2	3,3	3,8	4	2,7	2,3	2,3	2,4	2,8	2,9	2,7
A	3	1,6	1,4	3,7	1,8	1,5	1,4	1,4	1,9	2,3	1,3
A	4	1,6	5,4	3,7	1,6	1,3	1,1	1,1	1,3	2	5,1
B	1	4	4,4	6,8	3,7	3,3	3,3	3,3	3,5	3,8	3,5
B	2	3,3	4,2	4	2,7	2,3	2,3	2,4	3,7	2,7	2,7
B	3	1,6	1,4	3,7	1,8	1,5	1,4	1,5	2,7	1,8	1,4
B	4	1,6	3,7	3,7	1,6	1,3	1	1,1	3,5	3	4,1
C	1	3,3	79,8	3,2	3,3	3,5	3,1	3,1	3,1	4,3	5,5
C	2	2,9	2,4	3,1	2,7	2,6	2,5	2,3	2,6	2,7	2,3
C	3	1,6	1,4	3,1	2	1,6	1,5	1,6	1,7	1,9	1,3
C	4	1,6	1,4	3,1	1,9	1,5	1,3	1,3	1,3	1,7	1,1
D	1	4	12,9		3,6	3,3	3,4	3,3	3,6	3,7	3,7
D	2	3,4	4,2		2,6	2,3	2,3	2,4	2,6	2,9	2,7
D	3	1,6	1,4		1,8	1,4	1,3	1,4	1,9	2,2	1,3
D	4	1,6	5,2		1,6	1,2	1	1,1	1,3	3	5,1

Fonte: Resultados originais da pesquisa

Config A - Duas preditoras de clusterização e normalização da variável alvo (Y)

Config B - Sem clusterização e variável alvo em sua escala original.

Config C - Dados exclusivamente de teste, considerando os três primeiros meses de 2025, após treinamento com dados completos (os doze meses de 2024)

Config D - Exclusão da variável tempo_mediano_anterior

Estado 1 - Início da viagem

Estado 2 - Veículo já ingressou na aduana de exportação

Estado 3 - Veículo já liberou na aduana de importação

Estado 4 - Veículo já chegou no cliente e aguarda descarga

Conforme observado, modelos como LGBM, XGB e Random Forest demonstram os menores valores de RMSE, indicando maior precisão com erro mínimo de 1 dia e máximo de 3,5 dias, enquanto o OLS StepWise exibe os valores mais elevados, indicando maior variabilidade de erro médio (1,4 a 79,8 dias) sendo o modelo que apresenta o maior RSME dentre os demais modelos (RSME de 79,8 dias – Config. C estado 1).

A tabela 5 demonstra os resultados obtidos de R^2 em todos os modelos de algoritmos desenvolvidos nesta pesquisa.

Tabela 5. Demonstrativo dos R^2 s apresentados pelos modelos em cada configuração

Config	Estado	OLS Selected	OLS StepWise	Mediana	Decision Tree	Random Forest	XGB	LGBM	MLP	KERA S	SVR
A	1	7%	21%	-171%	19%	37%	34%	35%	21%	18%	18%
A	2	36%	17%	6%	59%	70%	69%	68%	56%	51%	58%
A	3	84%	88%	12%	79%	86%	88%	87%	76%	67%	89%
A	4	84%	-84%	12%	83%	90%	93%	93%	90%	75%	23%
B	1	7%	-15%	-171%	19%	38%	34%	35%	29%	17%	27%
B	2	36%	-1%	6%	59%	70%	69%	67%	21%	58%	59%
B	3	84%	88%	12%	78%	86%	88%	87%	53%	79%	88%
B	4	84%	15%	12%	83%	90%	93%	93%	23%	44%	-6%
C	1	7%	-545%	14%	6%	-5%	20%	19%	18%	-59%	-110%
C	2	29%	51%	16%	35%	41%	47%	53%	41%	39%	56%
C	3	78%	84%	14%	65%	76%	80%	79%	75%	69%	86%
C	4	78%	83%	14%	68%	80%	85%	85%	86%	74%	89%
D	1	4%	-882%		24%	36%	33%	34%	24%	19%	18%
D	2	34%	0%		61%	70%	69%	67%	60%	51%	58%
D	3	84%	88%		80%	87%	89%	87%	77%	69%	89%
D	4	84%	-73%		83%	90%	93%	92%	89%	44%	23%

Fonte: Resultados originais da pesquisa

Config A - Duas preditoras de clusterização e normalização da variável alvo (Y)

Config B - Sem clusterização e variável alvo em sua escala original.

Config C - Dados exclusivamente de teste, considerando os três primeiros meses de 2025, após treinamento com dados completos (os doze meses de 2024)

Config D - Exclusão da variável tempo_mediano_anterior

Estado 1 - Início da viagem

Estado 2 - Veículo já ingressou na aduana de exportação

Estado 3 - Veículo já liberou na aduana de importação

Estado 4 - Veículo já chegou no cliente e aguarda descarga

Os valores negativos apresentados na tabela 5 reforçam a ocorrência de overfitting em alguns cenários e modelos conforme já debatido anteriormente. Ao analisar o resultado geral de todos os modelos, compreende-se que enquanto o modelo de referência de mediana apresentou limitações severas em todos os estágios da viagem, os modelos de OLS exibiram um desempenho variável: o OLS com features selecionadas demonstrou estabilidade e melhores resultados (acima de 80%) nas etapas mais avançadas, ao passo que a versão com todas as features, mesmo após seleção stepwise, sofreu com overfitting em diversos cenários (-84%). Em contraste, os algoritmos de SVR e ensemble (como RandomForest, XGBoost e LGBM), demonstraram capacidade superior de aprendizado e generalização, gerando previsões mais precisas (acima de 85%) nos estágios mais avançados da operação logística. À medida que mais informações da viagem se tornavam disponíveis, observou-se uma

progressiva melhoria na performance de todos os modelos. Por exemplo, na Configuração A (tabela 5), o R^2 de Teste para modelos como RandomForest e LGBM evoluiu de cerca de 35% no Estado 1 para patamares de 87% no Estado 3, culminando com XGBoost e LGBM superando 90% no Estado 4. A análise das diferentes configurações de pré-processamento e bases de dados revelou detalhes importantes sobre a robustez e sensibilidade de cada algoritmo, destacando que não existe uma única abordagem ótima para todos os modelos, e a escolha da configuração ideal pode depender do algoritmo específico e das características dos dados de teste.

Ao efetuar a comparação dos resultados da presente pesquisa, com os achados de Servos et al. (2019), o qual analisou a previsão de tempos de viagem em transporte multimodal ocorrida da Alemanha para os Estados Unidos, observa-se que no estudo de Servos et al., o melhor entre os três algoritmos analisados (SVR, ExtraTrees e AdaBoost) foi o SVR, atingindo um erro médio absoluto (MAE) de 17 horas para um transporte de até 30 dias. Apesar de utilizarem métricas de avaliação dos modelos distintas em contextos heterogêneos, ambas as pesquisas apontam que o uso de Machine Learning para previsões de tempo de viagem eleva a precisão das previsões de entrega quando comparada aos métodos tradicionais de média ou mediana utilizadas normalmente no setor de transporte.

Considerações Finais

A pesquisa demonstrou a viabilidade e a alta capacidade preditiva dos algoritmos de aprendizado de máquina, confirmando que essas tecnologias podem oferecer soluções robustas para o transporte rodoviário do Mercosul. Os resultados indicam que os modelos de “ensemble”, com destaque para o “XGBoost”, se mostraram os algoritmos mais eficazes para a resolução do problema. A possibilidade de aplicar os algoritmos em várias fases do transporte conforme aplicados neste estudo (estados de viagem 1, 2, 3 e 4), permitirá à empresa informar e corrigir seus prazos de entrega ao longo do processo logístico, otimizar a alocação de recursos e elevar a satisfação de seus clientes.

Como limitações, reconhece-se que este trabalho se configura como um estudo de caso focado em uma única empresa, o que demanda cautela na generalização dos resultados para aplicação em todo o setor. Além disso, a qualidade dos dados, com certas informações sendo estimadas (como a quilometragem), pode influenciar o desempenho dos modelos.

Para trabalhos futuros, sugere-se a expansão da base de dados para incluir múltiplas transportadoras, o que permitiria a construção de modelos mais generalistas e robustos. A aplicação de um Modelo Linear Hierárquico (HLM) é também recomendado, pois pode capturar novas dependências entre os dados. Recomenda-se também a incorporação de

variáveis externas, como condições climáticas em tempo real, indicadores econômicos dos países envolvidos e informações sobre greves e bloqueios de rodovias por nevascas ou alagamentos, que poderiam enriquecer o poder preditivo dos algoritmos. Por fim, a implementação de um sistema de predição integrado à rotina da empresa e com um ciclo de retreinamento automatizado, representa o próximo passo lógico para transformar os achados desta pesquisa em uma ferramenta operacional valiosa.

Referências

Associação Brasileira de Transportadores Internacionais [ABTI]. 2020. Cenário do Transporte: edição nº 56. Disponível em: <<https://www.abti.org.br/images/cenario-do-transporte/pdfs/56.pdf>>. Acesso em: 16 março 2025.

Associação Brasileira de Transportadores Internacionais [ABTI]. 2021. Cenário do Transporte: edição nº 63. Disponível em: <<https://www.abti.org.br/images/cenario-do-transporte/pdfs/63.pdf>>. Acesso em: 16 mar. 2025.

Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5-32.

Dougherty, J.; Kohavi, R.; Sahami, M. 1995. Supervised and unsupervised discretization of continuous features. In: *International Conference on Machine Learning*, 12., 1995, Tahoe City, CA, USA. *Proceedings...* p. 194-202.

Fávero, L.P.; Belfiore, P. 2024. *Manual de Análise de Dados: Estatística e Machine Learning com Excel®, SPSS®, R® e Python®*. 2. ed. LTC, Rio de Janeiro, RJ, Brasil.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29(5): 1189-1232.

Goodfellow, I.; Bengio, Y.; Courville, A. 2016. *Deep Learning*. MIT Press, Cambridge, MA, EUA.

Hastie, T.; Tibshirani, R.; Friedman, J. 2009. *The elements of statistical learning: data mining, inference, and prediction*. 2. ed. Springer, New York, NY, EUA. Disponível em: <<https://web.stanford.edu/~hastie/ElemStatLearn/>>. Acesso em: 16 mar. 2025.

Machado, F.C.A.P.; Maciel, J.C. 2024. En los márgenes de ruta: discursos sobre desarrollo en la cobertura periodística de la RILA. *Interacciones (Campo Grande)* 24(3): 911-933.

Poder360. 2023. Brasil responde por 72% do PIB nominal do Mercosul. Disponível em: <<https://www.poder360.com.br/economia/brasil-responde-por-72-do-pib-nominal-do-mercosul/>>. Acesso em: 16 mar. 2025.

Servos, N.; Liu, X.; Teucke, M.; Freitag, M.. 2019. Travel time prediction in a multimodal freight transport relation using machine learning algorithms. In: *IEEE International Conference On Intelligent Transportation Systems*, 22., 2019, Auckland, Nova Zelândia. *Anais...* p. 1-6.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)* 36(2): 111–147.

Tsolaki, K.; Politis, S.; Tsalikis, E.; Charalambous, Y. 2022. The role of artificial intelligence in supply chain management: a comprehensive review. *Procedia Computer Science* 201: 333-340.