PRODUÇÃO ONLINE
REVISTA CIENTÍFICA ELETRÔNICA DE ENGENHARIA DE PRODUÇÃO
ISSN 1676-1901

BUSINESS INTELLIGENCE E MODELAGEM PREDITIVA APLICADOS À LOGÍSTICA AGRÍCOLA: UM ESTUDO DE CASO NA CADEIA CITRÍCOLA

BUSINESS INTELLIGENCE AND PREDICTIVE MODELING APPLIED TO AGRICULTURAL LOGISTICS: A CASE STUDY IN THE CITRUS SUPPLY CHAIN

Vinícius Sarandi de Souza* Email: viniciussarandi@usp.br
Alexandre Duarte** Email: alexandre.duarte@ufg.br
*Universidade de São Paulo (USP), Piracicaba, SP, Brasil.
**Universidade Federal de Goiás (UFG), Goiânia, GO / Universidade de São Paulo (USP), São Paulo, SP, Brasil.

Resumo: A logística agrícola na cadeia citrícola brasileira enfrenta desafios relacionados à dispersão geográfica, à dependência do transporte rodoviário e à variabilidade operacional. Nesse contexto, a falta de visibilidade estruturada sobre indicadores logísticos compromete o monitoramento e a tomada de decisão. Este estudo avalia como a aplicação de *Business Intelligence* (BI), associada à análise de indicadores e à modelagem preditiva, pode contribuir para a gestão logística em uma indústria citrícola. Trata-se de um estudo de caso com abordagem quantitativa e descritiva, utilizando dados operacionais extraídos do sistema ERP da empresa, tratados e analisados em ambiente analítico. Foram desenvolvidos indicadores-chave, como *Transit Time* e *Gate to Gate*, além de modelos de regressão linear para previsão desses indicadores. A solução foi avaliada por meio da comparação entre os processos de gestão anteriores e posteriores à implementação do painel, da análise do desempenho preditivo fora da amostra e de sua utilização em fóruns gerenciais. Os resultados evidenciam aumento da visibilidade operacional, maior padronização dos indicadores e redução do esforço manual de preparação e consolidação das informações. Os modelos apresentaram capacidade preditiva distinta, com maior estabilidade fora da amostra para o indicador *Gate to Gate*, enquanto o modelo de *Transit Time* apresentou maior dependência das características específicas das rotas. A integração entre BI e modelagem preditiva demonstrou potencial para ampliar simultaneamente a visibilidade, a capacidade de diagnóstico e a antecipação de desvios, contribuindo para a transição de uma gestão reativa para uma abordagem mais proativa e orientada por dados.

Palavras-chave: Apoio à decisão. Cadeia citrícola. *Business Intelligence*. Logística agrícola. Modelagem preditiva.

Abstract: Agricultural logistics in Brazil's citrus supply chain faces challenges related to geographic dispersion, dependence on road transportation, and operational variability. In this context, the lack of structured visibility into logistics indicators limits performance monitoring and managerial decision-making. This study evaluates how the integration of Business Intelligence (BI), logistics indicator analysis, and predictive modeling can contribute to logistics management in a citrus-processing company. This is a quantitative and descriptive case study based on operational data extracted from the company's ERP system and subsequently processed and analyzed in an analytical environment. Key performance indicators, including Transit Time and Gate-to-Gate, were developed, together with multiple linear regression models to predict these indicators. The solution was evaluated through comparisons between management processes before and after dashboard implementation, assessment of out-of-sample predictive performance, and observation of its use in managerial forums. The results indicate increased operational visibility, greater standardization of indicators, and reduced manual effort in preparing and consolidating information. The models exhibited different predictive capabilities, with greater out-of-sample stability for the Gate-to-Gate indicator, whereas the Transit Time model showed greater dependence on route-specific characteristics. The integration of BI and predictive modeling

demonstrated potential to simultaneously enhance operational visibility, diagnostic capability, and the anticipation of deviations, thereby supporting the transition from reactive management to a more proactive and data-driven approach.
Keywords: Agricultural logistics. Citrus supply chain. Business Intelligence. Decision support. Predictive modeling.

1 INTRODUÇÃO

A citricultura constitui uma das cadeias produtivas mais relevantes do agronegócio brasileiro, destacando-se pela expressiva participação do país na produção e exportação mundial de suco de laranja. O Brasil responde por cerca de 30% da produção global, consolidando-se como líder no mercado internacional e gerando impactos econômicos e sociais significativos (EMBRAPA, 2024; CNA, 2019).

Nesse contexto, a logística agrícola exerce papel estratégico ao assegurar o fluxo contínuo de matéria-prima entre as áreas produtoras e as unidades industriais. Sua eficiência impacta diretamente custos operacionais, competitividade e atendimento à demanda (Noro *et al.*, 2005). No cenário brasileiro, fatores como grandes distâncias, dependência do modal rodoviário e limitações de infraestrutura ampliam a complexidade da operação, elevando custos e variabilidade nos tempos logísticos (Faro, 2021).

Diante desse ambiente, o monitoramento de indicadores logísticos torna-se fundamental para a gestão eficiente das operações (Wajong, 2017). Indicadores como tempo de transporte, permanência em origem e tempo total de ciclo são amplamente utilizados para avaliar desempenho e identificar oportunidades de melhoria (Sharma *et al.*, 2021). Entretanto, a ausência de sistemas estruturados de informação ainda limita a visibilidade operacional e dificulta a tomada de decisão baseada em dados (Sharma *et al.*, 2018; Kamble *et al.*, 2020).

Nos últimos anos, abordagens baseadas em dados, como Business Intelligence (BI) e modelagem preditiva, têm se expandido como suporte à gestão operacional e estratégica. Essas ferramentas permitem integrar e analisar grandes volumes de dados, transformando-os em informações relevantes para a decisão, além de antecipar comportamentos operacionais e favorecer uma gestão mais proativa (Simão *et al.*, 2023; Sharda *et al.*, 2018).

Apesar desses avanços, a literatura ainda apresenta lacunas quanto à aplicação integrada de BI e modelagem preditiva na logística agrícola, sobretudo em

estudos empíricos que avaliem seus efeitos sobre o desempenho operacional (Routroy; Behera, 2017; Kamble *et al.*, 2020; Khandelwal *et al.*, 2021).

Diante desse contexto, este estudo busca responder à seguinte questão: como a aplicação integrada de Business Intelligence e modelagem preditiva pode contribuir para a melhoria do desempenho logístico em operações agrícolas? O objetivo é avaliar o impacto de um sistema de apoio à decisão baseado em dados na gestão logística de uma indústria citrícola, analisando seus efeitos sobre visibilidade operacional, padronização de indicadores e capacidade de antecipação de desvios logísticos.

Além de sua aplicação prática, este estudo contribui para a literatura de Engenharia de Produção em três dimensões. Primeiro, propõe e analisa empiricamente uma arquitetura integrada de monitoramento descritivo e modelagem preditiva aplicada à logística agrícola, articulando indicadores operacionais, visualização analítica e previsão de desempenho em um mesmo sistema de apoio à decisão.

Segundo, evidencia que diferentes indicadores logísticos apresentam estruturas explicativas distintas: enquanto o *Transit Time* é predominantemente condicionado por características estruturais das rotas, o *Gate to Gate* apresenta maior associação com variáveis contextuais e operacionais, como turnos, dias da semana e volume processado. Essa diferenciação contribui para o entendimento de que indicadores logísticos distintos requerem estratégias analíticas e gerenciais específicas.

Terceiro, o estudo demonstra empiricamente a importância de avaliar a capacidade de generalização dos modelos por meio de dados fora da amostra, evidenciando que um elevado ajuste no conjunto de treinamento não implica necessariamente maior desempenho preditivo em novas observações. Dessa forma, a contribuição do trabalho reside na proposição e na avaliação de uma abordagem integrada para transformar dados operacionais em mecanismos de monitoramento, diagnóstico e antecipação de desvios em uma cadeia logística agrícola.

2 REFERENCIAL TEÓRICO

2.1 Logística Agrícola

A logística agrícola é um dos principais determinantes da eficiência e competitividade das cadeias agroindustriais, especialmente em países de grande extensão territorial, como o Brasil (Andrade *et al.*, 2024). Características como dispersão geográfica das áreas produtivas, predominância do transporte rodoviário e influência de fatores sazonais ampliam a complexidade operacional e a variabilidade dos processos logísticos (Caixeta Filho, 2010; Garcia; Vieira Filho, 2019; Ehrig; Aurélio, 2026).

No contexto brasileiro, tais desafios são intensificados por limitações históricas de infraestrutura e pela concentração do escoamento da produção no modal rodoviário, o que aumenta a instabilidade dos tempos de transporte e a ocorrência de gargalos operacionais (Ribeiro *et al.*, 2020; Gonçalves *et al.*, 2024). Como consequência, a eficiência logística impacta diretamente custos, nível de serviço e capacidade de resposta organizacional (Ballou, 2010).

Em cadeias baseadas em produtos perecíveis, como a citricultura, essa variabilidade assume papel ainda mais crítico, pois atrasos, desbalanceamento de fluxos e falhas de sincronização podem comprometer a continuidade produtiva e gerar perdas relevantes (Boechat, 2015; Rodrigues *et al.*, 2006). Nesse cenário, a gestão eficiente da logística requer mecanismos que permitam maior controle, previsibilidade e coordenação das operações.

2.2 Indicadores logísticos

A mensuração do desempenho logístico é fundamental para compreender a eficiência das operações e orientar a tomada de decisão. Indicadores baseados em tempo, como tempos de ciclo, espera, carregamento e transporte, são amplamente utilizados para avaliar a utilização de recursos e identificar gargalos operacionais (Angelo, 2005; Ballou, 2010).

A análise estruturada desses indicadores permite decompor o processo logístico em etapas distintas, facilitando a identificação de fontes de variabilidade e

ineficiência. Tempos elevados ou inconsistentes podem indicar restrições de capacidade, falhas de coordenação ou problemas de sincronização entre agentes envolvidos (Hijjar, 2005; Viet *et al.*, 2018).

Na logística agrícola, indicadores temporais assumem relevância central por refletirem diretamente a dinâmica operacional de sistemas complexos e distribuídos. Assim, sua utilização sistemática constitui instrumento essencial para diagnóstico operacional, planejamento e melhoria contínua (Bejarano-Martínez; Caro, 2025; Vargas-Muñoz *et al.*, 2025).

2.3 Business Intelligence e visibilidade operacional

O Business Intelligence (BI) consolidou-se como abordagem fundamental para o suporte à decisão em ambientes organizacionais complexos. Ao integrar, processar e analisar dados de múltiplas fontes, o BI transforma grandes volumes de informação em conhecimento estruturado e acionável (Watson & Wixom, 2007; Turban *et al.*, 2011; Niu *et al.*, 2021).

Na logística, sua aplicação está associada ao aumento da visibilidade operacional, permitindo monitoramento contínuo, identificação de desvios e análise integrada do desempenho (Chen *et al.*, 2012; Jin; Kim, 2018; Sharma *et al.*, 2021). Essa capacidade é especialmente relevante em cadeias agroindustriais, marcadas por múltiplas etapas, dispersão geográfica e elevada variabilidade (Bhagat; Dhar, 2011; Khandelwal *et al.*, 2021).

Além disso, o BI contribui para a padronização das informações e para a criação de uma base única de dados, reduzindo inconsistências entre áreas e fortalecendo a transparência organizacional (Kimball; Ross, 2013). A visualização estruturada de indicadores possibilita análises comparativas, acompanhamento de metas e suporte mais ágil à tomada de decisão (Ni *et al.*, 2019).

Estudos recentes evidenciam que a adoção de abordagens baseadas em dados na gestão logística contribui para a redução de tarefas manuais, melhoria da qualidade da informação e aumento da eficiência operacional (Nascimento *et al.*, 2024). Em cadeias agroalimentares, a integração de dados logísticos favorece a

coordenação entre os elos da cadeia, reduz atrasos e melhora o desempenho das operações (Garcia *et al.*, 2012; Verdouw *et al.*, 2016; Yazdani *et al.*, 2021).

Nesse sentido, o BI pode ser compreendido como um elemento estruturante da transformação digital na logística, ampliando a capacidade de monitoramento, análise e suporte à decisão em ambientes operacionais complexos.

2.4 Modelagem preditiva aplicada à logística

A modelagem preditiva destaca-se como abordagem relevante para antecipação de comportamentos em sistemas logísticos. Entre as técnicas mais utilizadas, a regressão linear sobressai por sua simplicidade, interpretabilidade e capacidade de relacionar uma variável dependente a fatores explicativos (Montgomery *et al.*, 2012; Fávero; Belfiore, 2017).

Essa técnica permite identificar padrões, quantificar impactos operacionais e realizar previsões com base em dados históricos, desde que atendidos pressupostos estatísticos adequados (Morais *et al.*, 2020). Sua interpretabilidade favorece aplicações gerenciais, nas quais compreender fatores explicativos é tão importante quanto prever resultados (Mumtaz *et al.*, 2018).

Na logística, modelos preditivos têm sido aplicados para estimar tempos de transporte, otimizar rotas e melhorar o planejamento das operações (Oliveira *et al.*, 2020; Hardjomidjojo *et al.*, 2022; Nangpiire *et al.,* 2025). Estudos empíricos demonstram que a utilização de modelos estatísticos baseados em dados históricos contribui para aumentar a previsibilidade, reduzir atrasos e melhorar a coordenação entre as etapas da cadeia logística (Martins *et al.*, 2004; Wang *et al.*, 2016; Mahmud *et al.*, 2024; Haobijam; Seram, 2025).

Em cadeias agroindustriais, a aplicação de modelos preditivos permite incorporar variáveis operacionais relevantes — como distância, características da frota e condições de transporte — na estimativa de indicadores de desempenho, ampliando a capacidade de antecipação de desvios e de planejamento das atividades (Verdouw *et al.*, 2016; Rahman; Hye, 2021).

Além disso, a integração entre modelagem preditiva e sistemas analíticos potencializa o suporte à decisão, permitindo que gestores não apenas monitorem o

desempenho passado, mas também antecipem cenários futuros e adotem uma postura mais proativa na gestão das operações (Waller ; Fawcett, 2013; Kache; Seuring, 2017; Trivellas *et al.*, 2020).

2.5 Integração conceitual

A literatura revisada evidencia que a eficiência da logística agrícola está diretamente relacionada à capacidade de monitoramento e análise de indicadores operacionais, especialmente aqueles baseados em tempo, que refletem a dinâmica e a variabilidade das operações distribuídas geograficamente (Bejarano-Martínez; Caro, 2025; Vargas-Muñoz *et al.,* 2025). Entretanto, o acompanhamento desses indicadores, isoladamente, tende a produzir uma visão predominantemente retrospectiva do desempenho.

Nesse contexto, abordagens de *Business Intelligence* ampliam a visibilidade e a integração das informações, permitindo o acompanhamento estruturado do desempenho logístico (Yazdani *et al.*, 2021; Nascimento *et al.*, 2024). Complementarmente, a modelagem preditiva possibilita antecipar comportamentos operacionais e reduzir incertezas, contribuindo para o planejamento e a tomada de decisão (Mumtaz *et al.*, 2018; Trivellas *et al.*, 2020).

Apesar desses avanços, ainda são limitadas as evidências empíricas que analisem conjuntamente essas duas dimensões em operações agrícolas reais. Em particular, permanece pouco explorado como diferentes indicadores temporais podem apresentar estruturas explicativas distintas quando analisados por meio de modelos preditivos integrados a sistemas de monitoramento operacional (Routroy; Behera, 2017; Kamble *et al.,* 2020; Khandelwal *et al.*, 2021). Essa lacuna é relevante porque a simples disponibilização de indicadores e previsões não garante, por si só, uma compreensão adequada dos mecanismos associados ao desempenho logístico.

Dessa forma, este estudo propõe e avalia empiricamente um sistema integrado de apoio à decisão que combina *Business Intelligence*, indicadores logísticos e modelagem preditiva em uma operação de transporte agrícola da cadeia citrícola. A contribuição científica está, portanto, na análise integrada de três dimensões: (i) a visibilidade e a padronização do desempenho por meio de indicadores estruturados;

(ii) a identificação dos fatores associados à variabilidade de diferentes indicadores logísticos; e (iii) a avaliação da capacidade de generalização dos modelos para observações não utilizadas na estimação.

A partir dessa abordagem, o estudo busca verificar se indicadores logísticos de naturezas distintas apresentam estruturas explicativas diferentes e se a integração entre monitoramento descritivo e previsão pode ampliar a capacidade analítica da gestão logística. Assim, o trabalho utiliza essa integração para produzir evidências sobre a relação entre características estruturais da operação, condições contextuais e desempenho logístico em uma cadeia agrícola complexa.

3 MATERIAIS E MÉTODOS

3.1 Delineamento da pesquisa e coleta de dados

Esta pesquisa caracteriza-se como um estudo de caso aplicado, com abordagem quantitativa e descritiva, conduzido em uma empresa de grande porte do setor citrícola, localizada no interior do estado de São Paulo, dedicada à produção e ao processamento de laranja para fabricação de suco e derivados.

A natureza aplicada do estudo decorre da busca por solução para um problema prático relacionado à ausência de visibilidade estruturada e padronização de indicadores logísticos, condição recorrente em operações que ainda não consolidaram práticas analíticas baseadas em dados (Turban *et al.*, 2009).

Os dados utilizados são de natureza secundária, provenientes do sistema integrado de gestão empresarial da organização, responsável pelo registro das operações logísticas de transporte entre fazendas e unidades industriais.

A coleta ocorreu por meio de uma arquitetura corporativa baseada em repositório centralizado (data lake), no qual as informações operacionais são integradas e disponibilizadas por consultas estruturadas. Essa abordagem reduz extrações manuais, minimiza inconsistências e assegura maior confiabilidade, rastreabilidade e atualização dos dados, em consonância com boas práticas de governança analítica (Bonel, 2020).

Por questões de confidencialidade, os dados apresentados ao longo do estudo foram anonimizados, mantendo-se sua estrutura e características estatísticas, sem exposição de informações estratégicas da organização.

3.2 Caracterização das variáveis e base de dados

A base de dados é composta por registros operacionais de transporte de fruta, em que cada observação corresponde a uma viagem individual entre a fazenda de origem e a unidade industrial de destino.

As variáveis utilizadas incluem informações categóricas, temporais e numéricas, permitindo tanto a construção de indicadores logísticos quanto a modelagem preditiva dos processos analisados. A Tabela 1 apresenta a descrição das principais variáveis utilizadas no estudo.

Tabela 1 – Descrição das variáveis do *dataset*

Variável	Tipo	Descrição	Uso na análise
Safra	Categórica	Ano agrícola da operação	Segmentação temporal
Fábrica	Categórica	Unidade industrial de destino	Comparação entre unidades
Código do Imóvel	Categórica	ID da fazenda de origem	Identificação da origem
Nome do Imóvel	Categórica	Fazenda de origem	Contextualização
Docto. Frete	Identificador	Documento de transporte	Identificação da viagem
Lote	Identificador	Lote de carregamento	Rastreamento da carga
Data saída fazenda	Temporal	Saída do caminhão	Cálculo do transit time
PEX – Data Turno	Temporal	Data operacional do PEX	Controle operacional
Data e Hora PEX	Temporal	Confirmação do evento PEX	Cálculo do transit time
Qntd. de Caixas	Numérica	Volume líquido descarregado	Volume transportado
Qtd. Caixas Refugo	Numérica	Volume descartado	Indicador de qualidade
Tipo de Veículo	Categórica	Modal do veículo	Variável explicativa
Placa do Veículo	Identificador	Identificação do caminhão	Controle operacional
Placa da Carreta	Identificador	Identificação da carreta	Rastreamento
Peso Tara Balança	Numérica	Peso do veículo vazio	Controle de carga
Peso Bruto Balança	Numérica	Peso total carregado	Controle operacional
Data entrada fazenda	Temporal	Entrada do caminhão	Cálculo do “gate to gate”

Fonte: Dados da pesquisa.

A partir dessas variáveis, foram construídos indicadores logísticos baseados em tempo, com destaque para o *Transit Time* e o *Gate to Gate*. A Tabela 2 apresenta um recorte ilustrativo da base utilizada.

Tabela 2 – Exemplo de observações do *dataset*

Safra	Fábrica	ID imóvel	Imóvel	Doc. Frete	Lote	Data saída	Data Turno
2025	1000	30XXXX	FAZENDA X	610XXX	MCXX	02/07/2025 12:20	02/07/2025
2025	1000	30XXXX	FAZENDA Y	610XXX	MCXX	02/07/2025 12:20	02/07/2025
2025	1000	30XXXX	FAZENDA Z	610XXX	MCXX	03/07/2025 10:35	03/07/2025

Fonte: Dados da pesquisa (valores anonimizados).

3.3 Cenário operacional inicial (*As Is*)

Antes da implementação da solução proposta, o monitoramento logístico era conduzido por fluxos predominantemente manuais e descentralizados. Os dados eram extraídos do sistema corporativo e tratados em planilhas eletrônicas, exigindo atividades recorrentes de consolidação, padronização e validação.

Esse processo demandava elevado tempo operacional e estava sujeito a retrabalhos e inconsistências decorrentes de alterações nas bases e ausência de padronização. Além disso, a análise dos indicadores ocorria de forma estática, limitando a agilidade e a confiabilidade das informações utilizadas na tomada de decisão. Esse cenário foi adotado como base comparativa para avaliação dos impactos da solução proposta.

3.4 Construção do sistema de monitoramento logístico

A solução proposta consiste na implementação de um sistema de apoio à decisão baseado em *Business Intelligence*, estruturado a partir da integração, tratamento e modelagem dos dados operacionais.

Inicialmente, os dados passaram por etapas de preparação, incluindo remoção de duplicidades, padronização de chaves e harmonização de formatos, assegurando consistência e integridade das informações.

Com base nessa estrutura, foram construídos indicadores logísticos baseados em tempo, com destaque para o *Transit Time* (tempo de deslocamento entre fazenda e unidade industrial) e o *Gate to Gate* (tempo total do ciclo operacional na origem). A base final resultante desse processo é apresentada na Tabela 3.

Tabela 3 - Exemplo de estrutura do dataset utilizado no painel de “Business Intelligence”

Fazenda	Fábrica	Rota	Doc. Frete	Modal	Placa	GtG	TT
FAZENDA X	FABRICA A	Rota_01	610XX	RODO-TL	EVUXXXI	1,43	7,2
FAZENDA Y	FABRICA A	Rota_03	610XX	RODO-TL	EVYYYYI	1,43	7,2
FAZENDA Z	FABRICA C	Rota_05	610XX	RODO-TL	EXJZZZI	2,82	6,84

Fonte: Dados originais da pesquisa (valores anonimizados).

Para ambos, foram definidos parâmetros de referência operacionais: no caso do *Transit Time*, a partir da distância entre origem e destino e de uma velocidade média de referência; no *Gate to Gate*, com base no tempo padrão de carregamento e movimentação interna. Esses parâmetros possibilitam avaliar o desempenho observado em relação às metas operacionais.

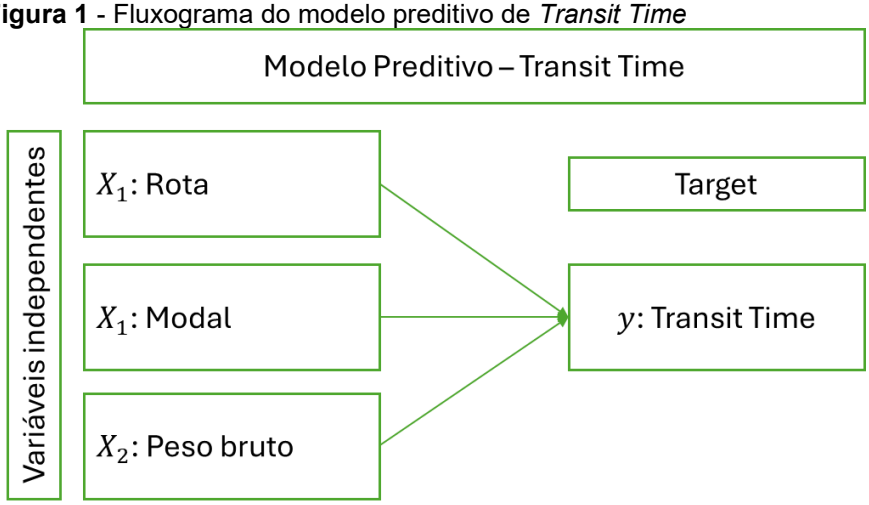
Cada registro representa uma operação logística individual, permitindo análises detalhadas por rota, unidade e período. A variável “Rota”, definida pela combinação entre origem e destino, foi utilizada como elemento-chave para segmentação e comparação do desempenho logístico.

O sistema analítico desenvolvido permite comparar valores observados e metas operacionais, identificar desvios significativos e realizar análises comparativas entre unidades, períodos e características operacionais. Dessa forma, o sistema de BI implementado consolida informações e estrutura um ambiente analítico voltado ao monitoramento contínuo e ao suporte à tomada de decisão.

3.5 Modelagem preditiva

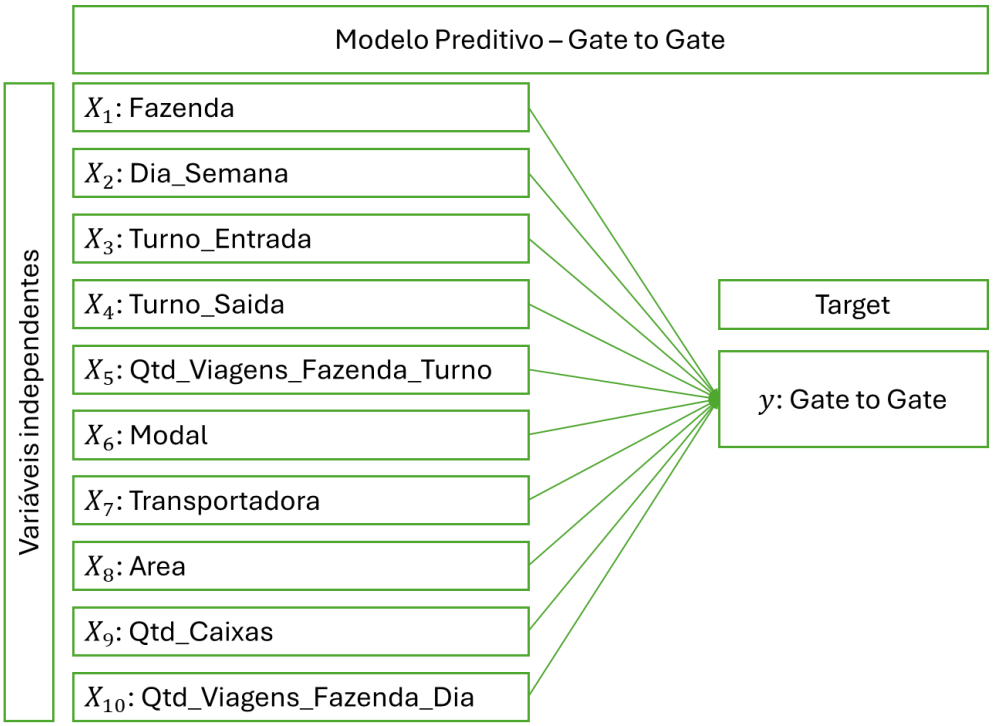
Com o objetivo de ampliar a capacidade analítica do sistema de Business Intelligence, foram desenvolvidos dois modelos de regressão linear múltipla para previsão de indicadores logísticos: um modelo para o *Transit Time* (TT) e outro para o *Gate to Gate* (GtG). Em ambos os casos, a variável dependente (y) corresponde ao tempo operacional, enquanto as variáveis independentes (X) representam características da rota, da operação, do veículo, da fazenda, do período e do volume transportado. As Figuras 1 e 2 apresentam os fluxogramas conceituais dos modelos.

Figura 1 - Fluxograma do modelo preditivo de *Transit Time*



Fonte: Dados originais da pesquisa.

Figura 2 - Fluxograma do modelo preditivo de *Gate to Gate*



Fonte: Dados originais da pesquisa.

A construção dos modelos foi realizada por meio de um procedimento sequencial de preparação dos dados, transformação da variável dependente, seleção de variáveis, diagnóstico de multicolinearidade, identificação de observações influentes e avaliação do desempenho preditivo. A análise foi realizada

separadamente para cada indicador logístico, respeitando as características específicas de cada processo operacional.

As variáveis categóricas foram convertidas em variáveis indicadoras (*dummies*), com a adoção de categorias de referência. Esse procedimento permitiu representar características qualitativas, como rota, fazenda, tipo de veículo, turno, dia da semana e transportadora, na estrutura de regressão linear, evitando a utilização direta de códigos numéricos arbitrários para representar categorias (Wooldridge, 2010).

O modelo de *Transit Time* considera como variáveis explicativas a Rota (X_1), o tipo de veículo (Modal) (X_2) e o Peso Bruto da carga (X_3). A variável dependente (y) corresponde ao tempo de viagem em horas.

Já o modelo de *Gate to Gate* utiliza como variáveis explicativas a Fazenda de origem (X_1), Dia da semana (X_2), Turno de entrada (X_3), Turno de saída (X_4), Quantidade de viagens da fazenda no turno (X_5), Tipo de veículo (X_6), Transportadora (X_7), Área da fazenda (X_8), Quantidade de caixas líquidas do lote (X_9) e Quantidade de viagens da fazenda por dia (X_{10}). A variável dependente (y) corresponde ao tempo total do ciclo operacional em horas.

3.5.1 Transformação da variável dependente

A adequação da variável dependente à estrutura do modelo linear foi avaliada por meio da transformação de Box-Cox (BOX; COX, 1964). O parâmetro de transformação λ foi estimado com base exclusivamente nos dados do conjunto de treinamento, evitando que informações do conjunto de teste fossem utilizadas durante a construção do modelo. A transformação foi definida pela Eq. (1).

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, \lambda \neq 0 \\ \ln(y), \lambda = 0 \end{cases} \quad (1)$$

Para o modelo de *Transit Time*, foi obtido λ = 0,79, enquanto para o modelo de *Gate to Gate* foi obtido λ = 0,65.

Os modelos foram estimados na escala transformada. Entretanto, para avaliação da capacidade preditiva, as previsões foram posteriormente convertidas novamente para a escala original da variável dependente por meio da transformação inversa de Box-Cox, dada pela Eq. (2).

$$\hat{y} = \begin{cases} \left(\lambda \hat{y}^{(\lambda)} + 1\right)^{1/\lambda}, \lambda \neq 0 \\ \exp(\hat{y}^{(\lambda)}), \lambda = 0 \end{cases} \quad (2)$$

Dessa forma, os indicadores de erro foram calculados tanto na escala transformada quanto na escala original, permitindo avaliar separadamente o ajuste estatístico do modelo e a utilidade prática das previsões em unidades originais de tempo.

3.5.2 Separação entre treinamento e teste

A avaliação preditiva foi realizada por meio da separação dos dados em conjuntos de treinamento e teste.

Para o modelo de Transit Time, após a aplicação dos critérios de disponibilidade e validade das variáveis, foram consideradas 21.115 observações elegíveis, particionadas em 16.747 observações de treinamento e 4.368 observações de teste. O conjunto de treinamento foi utilizado para estimar o parâmetro da transformação Box-Cox, realizar a seleção de variáveis e executar os procedimentos de diagnóstico e refinamento do modelo. O conjunto de teste permaneceu separado durante todo o processo de estimação e foi utilizado exclusivamente para avaliar a capacidade de generalização.

Para o modelo de Gate to Gate, após a aplicação dos critérios de disponibilidade e validade das variáveis, foram consideradas 22.021 observações elegíveis. Adotou-se uma divisão temporal, com ponto de corte em 30 de dezembro de 2025: as 19.615 observações anteriores ao corte foram destinadas ao treinamento e as 2.406 observações posteriores, ao teste.

As amostras analíticas diferem da base operacional de 23.446 viagens porque a disponibilidade e os critérios de validade das variáveis necessárias não são idênticos

para os dois indicadores. As exclusões pelo intervalo interquartil e pela distância de Cook descritas a seguir foram aplicadas somente ao conjunto de treinamento, preservando o conjunto de teste para a avaliação genuinamente fora da amostra.

A divisão temporal foi adotada para aproximar a avaliação das condições de aplicação operacional do modelo, simulando a utilização de informações históricas para previsão de operações futuras. Assim, o desempenho no conjunto de teste representa uma avaliação fora da amostra, realizada sobre observações não utilizadas no ajuste dos coeficientes.

3.5.3 Seleção de variáveis pelo método stepwise

A seleção das variáveis explicativas foi realizada exclusivamente no conjunto de treinamento, utilizando o procedimento *stepwise*. O processo foi baseado em um critério formal de significância estatística, com inclusão e exclusão iterativas de variáveis candidatas conforme seus respectivos *p-values* (Fávero; Belfiore, 2017).

O procedimento foi aplicado sobre as variáveis *dummy* e quantitativas disponíveis após a etapa de preparação dos dados. A cada etapa, variáveis que não atendiam ao critério estatístico definido foram removidas, enquanto variáveis com contribuição estatisticamente significativa foram mantidas.

Para o modelo de *Transit Time*, o procedimento *stepwise* selecionou inicialmente 31 variáveis explicativas, todas mantidas após a etapa subsequente de avaliação de significância. Para o modelo de *Gate to Gate*, foram inicialmente selecionadas 33 variáveis, das quais 32 permaneceram após a filtragem por significância estatística.

Após essa etapa, as variáveis selecionadas foram submetidas ao diagnóstico de multicolinearidade, conforme descrito na subseção seguinte. A relação completa das variáveis selecionadas, removidas e mantidas em cada etapa foi registrada durante o processo de modelagem, permitindo a rastreabilidade da especificação final dos modelos.

3.5.4 Diagnóstico de multicolinearidade e observações influentes

Após a seleção estatística das variáveis, a multicolinearidade foi avaliada por meio do fator de inflação da variância (*Variance Inflation Factor* – VIF). O VIF foi calculado para as variáveis explicativas presentes na especificação do modelo, sendo removidas iterativamente as variáveis que apresentavam maior contribuição para a colinearidade até que todos os valores permanecessem abaixo do limite adotado.

No modelo de TT, nenhuma variável adicional foi removida nessa etapa, e o maior VIF observado foi 3,868. No modelo de GtG, seis variáveis foram removidas devido à multicolinearidade, e o maior VIF do modelo final foi 4,813.

A identificação de observações potencialmente influentes foi realizada por meio da distância de Cook (Cook, 1977). As observações identificadas como influentes foram removidas do conjunto utilizado para o ajuste final, e o modelo foi reestimado após essa etapa. No modelo de TT, foram removidas 838 observações por esse critério. No modelo de GtG, foram removidas 820 observações.

Adicionalmente, no modelo de TT, foram removidas 702 observações do conjunto de treinamento pelo critério do intervalo interquartil, aplicado durante a etapa de tratamento da distribuição da variável dependente. Assim, o ajuste final do modelo de TT utilizou 15.207 observações. No modelo de GtG, não houve remoção por esse critério e, após a exclusão das 820 observações influentes, o ajuste final utilizou 18.795 observações.

3.5.5 Estimação dos modelos

Após a definição das variáveis finais, os modelos foram estimados por mínimos quadrados ordinários (OLS). Considerando a presença de heterocedasticidade identificada nos testes de diagnóstico, foram utilizados erros-padrão robustos do tipo HC3 (Davidson; Mackinnon, 1993).

O modelo de *Transit Time* pode ser representado genericamente pela Eq. (3), em que $TT_i^{(\lambda)}$ representa o *Transit Time* transformado por Box-Cox, α é o intercepto, X_{ij} representa as variáveis explicativas e ε_i corresponde ao termo de erro aleatório.

De forma análoga, o modelo de *Gate to Gate* pode ser representado genericamente pela Eq. (4), em que $GtG_i^{(\lambda)}$ representa o *Gate to Gate* transformado por Box-Cox.

$$TT_i^{(\lambda)} = \alpha + \sum_{j=1}^{k} \beta_j X_{ij} + \epsilon_i \quad (3)$$

$$GtG_i^{(\lambda)} = \alpha + \sum_{j=1}^{k} \beta_j X_{ij} + \epsilon_i \quad (4)$$

As variáveis explicativas finais foram constituídas por combinações de variáveis categóricas transformadas em *dummies* e variáveis quantitativas associadas ao volume operacional.

3.5.6 Avaliação dos modelos

Para o desenvolvimento e avaliação dos modelos preditivos, foram utilizados dados coletados entre abril de 2025 e fevereiro de 2026, abrangendo operações logísticas de transporte de fruta entre fazendas e unidades industriais da empresa.

A base contempla operações provenientes de 22 fazendas próprias, destinadas a 3 unidades industriais, envolvendo aproximadamente 500 veículos ativos, seis transportadoras e quatro tipos de veículos: Carreta, Rodo, Bitrem e Truck. Foram identificadas 23.446 operações logísticas, considerando documentos de frete distintos, cada um representando uma viagem individual.

Os registros incluem informações sobre origem, destino, veículo, modal, volume transportado e tempos operacionais. A Tabela 4 apresenta a distribuição das operações por transportadora e modal, enquanto a Tabela 5 exemplifica a estrutura do dataset utilizado na modelagem.

Tabela 4 - Distribuição das operações logísticas por transportadora e modal

Transportadora	Carreta	Rodo	Bitrem	Truck	Total
A	4.031	210	26	8	4.275
B	1.469	84	23	68	1.644
C	1.662	2.147	27	840	4.676
D	10.125	2.094	625	7	12.851

Total	17.287	4.535	701	923	23.446

Fonte: Dados originais da pesquisa (valores anonimizados).

Tabela 5 - Amostra das primeiras observações da base de dados.

Rota	Frete	Modal	Peso bruto	Dia	Turno entrada	Turno saída	Qtd. Caixas fazenda turno	GtG (min)	TT (min)
R01	610XX	RODO	93700	Quarta	Manhã	Tarde	1280	172	262
R01	611XX	RODO	93700	Quarta	Manhã	Tarde	1275	164	275
R12	612XX	TRUCK	73460	Quinta	Manhã	Tarde	980	85	200
R23	613XX	RODO	93000	Quinta	Tarde	Noite	1240	195	328

Fonte: Dados originais da pesquisa (valores anonimizados).

A qualidade de ajuste dos modelos foi avaliado pelo coeficiente de determinação R^2, calculado na escala transformada, enquanto o desempenho preditivo foi analisado com base no erro absoluto médio (MAE) e na raiz do erro quadrático médio (RMSE) (Fávero; Belfiore, 2017), vide Eq. (5) e Eq. (6). As métricas foram calculadas separadamente para os conjuntos de treinamento e teste, tanto na escala transformada por Box-Cox quanto na escala original após a aplicação da transformação inversa.

$$MAE = \frac{1}{n} + \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{6}$$

O R^2 foi utilizado como medida de ajuste e capacidade explicativa da variabilidade observada, enquanto o MAE e o RMSE foram empregados para avaliar a magnitude dos erros de previsão. A comparação entre os resultados de treinamento e teste permitiu avaliar a ocorrência de perda de desempenho fora da amostra e, consequentemente, a capacidade de generalização dos modelos.

Também foram realizados testes de diagnóstico envolvendo significância global do modelo, normalidade dos resíduos, heterocedasticidade, autocorrelação e multicolinearidade. A interpretação desses testes foi realizada de forma conjunta com

as métricas preditivas, evitando que a avaliação dos modelos dependesse exclusivamente de medidas de ajuste dentro da amostra (Turban *et al.*, 2009).

3.6 Cenário proposto (*To Be*) e validação

A validação da solução ocorreu em ambiente operacional, por meio do acompanhamento do uso do sistema em rotinas gerenciais, como reuniões de monitoramento logístico e fóruns com transportadoras.

A adoção do sistema permitiu centralizar informações, padronizar indicadores e ampliar a transparência entre os agentes envolvidos. Na comparação documentada do processo, a preparação manual das informações consumia aproximadamente uma hora antes de cada reunião; com a atualização automatizada e o acesso imediato ao painel, esse tempo foi praticamente eliminado, deslocando o esforço da consolidação para a análise dos resultados.

Não foram mantidos registros sistemáticos da frequência de acesso ao painel, do número de reuniões apoiadas ou da economia acumulada de horas. Por essa razão, a validação operacional é interpretada como evidência quantitativa de redução do esforço por ciclo de preparação e como evidência observacional de uso gerencial, sem atribuição causal dos resultados operacionais exclusivamente ao dashboard.

3.7 Ambiente computacional e reprodutibilidade

O painel foi desenvolvido no Microsoft Power BI Desktop (v. 2.157.879.0). A modelagem preditiva foi implementada em Python (v. 3.14.7), com apoio das seguintes bibliotecas principais: NumPy (v. 1.26.4), statsmodels (v. 0.14.6), pandas (v 2.2.3) e matplotlib (v 3.10.0). Para favorecer a rastreabilidade, esta seção explicita a plataforma computacional, enquanto as subseções anteriores documentam os critérios de filtragem, a separação das amostras, a transformação de Box-Cox, a seleção de variáveis, os diagnósticos e o cálculo das métricas.

3.7 Avaliação comparativa dos cenários

A avaliação dos resultados foi realizada por meio da comparação entre o cenário inicial (*As Is*) e o cenário proposto (*To Be*), considerando cinco critérios: eficiência operacional, padronização dos indicadores, produtividade, suporte à tomada de decisão e desempenho logístico.

Essa abordagem é consistente com a literatura, que destaca que iniciativas de *Business Intelligence* devem ser avaliadas com base em seus impactos sobre o desempenho organizacional e a qualidade do processo decisório (Turban *et al.*, 2009; Bonel, 2020).

4 RESULTADOS E DISCUSSÕES

4.1 Modelo preditivo – *Transit Time*

O modelo de *Transit Time* foi desenvolvido a partir da divisão dos dados em 16.747 observações de treinamento e 4.368 observações de teste. A transformação de Box-Cox, estimada no conjunto de treinamento, resultou em λ = 0,79.

Durante o processo de preparação e refinamento, foram removidas 702 observações pelo critério do intervalo interquartil e 838 observações identificadas como influentes pela distância de Cook. Após essas etapas, o modelo final foi estimado com 15.207 observações.

O procedimento de seleção *stepwise* resultou inicialmente em 31 variáveis explicativas. Todas permaneceram após a avaliação de significância estatística, e nenhuma variável adicional foi removida pela análise de VIF. O modelo final foi, portanto, composto por 31 variáveis explicativas, predominantemente associadas às categorias de rota, além da variável indicadora do tipo de veículo (“Veiculo_CARRETA”).

O modelo final apresentou R^2 de 0,84, indicando que aproximadamente 84% da variabilidade do Transit Time transformado foi explicada pelas variáveis incluídas na especificação final. O teste de significância global apresentou $p < 0.001$, indicando que o conjunto de variáveis explicativas possui capacidade estatisticamente significativa de explicar a variável dependente.

A maior parte das variáveis associadas às rotas apresentou significância estatística ao nível de 1%. Os coeficientes apresentaram magnitudes distintas entre as rotas, evidenciando diferenças sistemáticas no tempo de deslocamento associadas às características específicas dos trajetos. A variável “Veículo_CARRETA” apresentou coeficiente de −0,297, indicando diferença estatisticamente significativa em relação à categoria de referência, *ceteris paribus*.

A interpretação dos coeficientes indica que o *Transit Time* é fortemente condicionado pelas características específicas das rotas. As diferenças entre os coeficientes das categorias de rota evidenciam que determinados trajetos apresentam tempos de deslocamento sistematicamente superiores ou inferiores à categoria de referência.

O maior VIF observado no modelo final foi 3,868, indicando níveis de multicolinearidade inferiores ao limite adotado no procedimento de diagnóstico.

A Tabela 6 apresenta os coeficientes estimados para o modelo de *Transit Time*.

Tabela 6 - Resumo dos coeficientes estimados para o modelo final de “transit time”.

Coeficiente (Variável)	Valor	p-valor
R^2	0,84	-
Intercepto (β)	2,561	p < 0,01
Veiculo_CARRETA ($\alpha_{veicCARRETA}$)	-0,297	p < 0,01
ROTA_03 (α_{ROTA_03})	-1,654	p < 0,01
ROTA_04 (α_{ROTA_04})	-2,088	p < 0,01
ROTA_05 (α_{ROTA_05})	-2,077	p < 0,01
ROTA_06 (α_{ROTA_06})	-1,585	p < 0,01
ROTA_07 (α_{ROTA_07})	2,627	p < 0,01
ROTA_08 (α_{ROTA_08})	2,685	p < 0,01
ROTA_09 (α_{ROTA_09})	2,668	p < 0,01
ROTA_10 (α_{ROTA_10})	2,883	p < 0,01
ROTA_11 (α_{ROTA_11})	1,435	p < 0,01
ROTA_12 (α_{ROTA_12})	2,850	p < 0,01
ROTA_13 (α_{ROTA_13})	1,569	p < 0,01
ROTA_14 (α_{ROTA_14})	-0,265	p < 0,01
ROTA_15 (α_{ROTA_15})	-0,665	p < 0,01
ROTA_16 (α_{ROTA_16})	-0,859	p < 0,01
ROTA_17 (α_{ROTA_17})	1,678	p < 0,01
ROTA_18 (α_{ROTA_18})	1,975	p < 0,01
ROTA_19 (α_{ROTA_19})	1,488	p < 0,01
ROTA_20 (α_{ROTA_20})	1,301	p < 0,01
ROTA_21 (α_{ROTA_21})	1,290	p < 0,01
ROTA_22 (α_{ROTA_22})	1,333	p < 0,01
ROTA_23 (α_{ROTA_23})	0,832	p < 0,01
ROTA_24 (α_{ROTA_24})	1,077	p < 0,01
ROTA_25 (α_{ROTA_25})	1,207	p < 0,01
ROTA_27 (α_{ROTA_27})	0,777	p < 0,01

ROTA_28 (α_{ROTA_28})	0,693	p < 0,01
ROTA_29 (α_{ROTA_29})	0,563	p < 0,01
ROTA_30 (α_{ROTA_30})	0,984	p < 0,01
ROTA_31 (α_{ROTA_31})	1,357	p < 0,01
ROTA_32 (α_{ROTA_32})	0,512	p < 0,01
ROTA_33 (α_{ROTA_33})	0,576	p < 0,01

Fonte: Dados originais da pesquisa.

Os resultados do modelo de *Transit Time* permitem uma interpretação relevante para a análise do desempenho logístico agrícola. A predominância das variáveis associadas às rotas indica que o tempo de deslocamento é determinado principalmente por características estruturais do sistema de transporte. Esse resultado é consistente com a literatura que destaca a influência da distância, da configuração espacial da rede e das condições específicas dos trajetos sobre a variabilidade dos tempos de transporte (Oliveira *et al.*, 2020; Hardjomidjojo *et al.*, 2022; Nangpiire *et al.*, 2025).

Entretanto, os resultados também qualificam essa interpretação ao indicar que a rota, e não apenas uma medida agregada de distância, constitui uma importante unidade de diferenciação do desempenho. Isso sugere que características não observadas diretamente nas variáveis agregadas, como condições específicas do trajeto, restrições de circulação, padrões de tráfego e características operacionais recorrentes, podem estar incorporadas na diferenciação entre as categorias de rota.

A diferença entre o desempenho no treinamento e no teste também possui implicações metodológicas relevantes. Embora o modelo tenha apresentado elevado poder explicativo na amostra utilizada para estimação, a redução do R^2 para 0,47 demonstra que a capacidade de generalização é inferior ao ajuste dentro da amostra. Esse resultado reforça a importância, destacada na literatura sobre modelagem preditiva aplicada à logística, de avaliar modelos com observações não utilizadas na estimação, especialmente em ambientes operacionais caracterizados por heterogeneidade e variabilidade (Waller; Fawcett, 2013; Kache; Seuring, 2017).

Assim, o modelo de *Transit Time* apresenta maior contribuição para a identificação de padrões estruturais e para a comparação do desempenho entre rotas do que para a previsão altamente precisa de novas observações. Essa distinção é relevante porque demonstra que o valor de um modelo preditivo em sistemas

logísticos não deve ser avaliado apenas pelo seu R^2, mas também pela natureza da informação produzida e pela finalidade gerencial para a qual o modelo é utilizado.

Sob a perspectiva científica, a perda de desempenho fora da amostra indica que as variáveis indicadoras de rota absorvem heterogeneidades locais que não necessariamente se transferem para novas operações. O achado qualifica as aplicações de previsão de tempos e otimização de rotas discutidas por Oliveira *et al.* (2020), Hardjomidjojo *et al.* (2022) e Nangpiire *et al.* (2025): para a Engenharia de Produção, não basta obter elevado ajuste estatístico; é necessário alinhar a unidade de análise, a finalidade decisória e a validade externa do modelo. A incorporação futura de atributos observáveis do trajeto pode reduzir a dependência de categorias específicas e ampliar a transferibilidade das previsões.

4.1.1 Desempenho preditivo

A Tabela 7 apresenta o desempenho do modelo na escala transformada (Box-Cox – BC).

Tabela 7 - Resumo do desempenho do modelo de *Transit Time* na escala transformada

Amostra	R^2 BC	RMSE BC	MAE BC	RMSE Original	MAE Original
Treinamento	0,836	0,684	0,489	2,031	1,861
Teste	0,473	1,653	0,823	2,616	1,210

Fonte: Dados originais da pesquisa.

Os resultados evidenciam diferença relevante entre o desempenho dentro e fora da amostra. Embora o modelo apresente elevado poder explicativo no conjunto de treinamento, o desempenho no conjunto de teste é inferior, com R^2 = 0,473 na escala transformada.

Esse resultado indica que o modelo captura uma parcela relevante da estrutura observada nos dados de treinamento, mas apresenta perda de capacidade explicativa quando aplicado a observações não utilizadas na estimação. Portanto, o desempenho do modelo deve ser interpretado principalmente sob a perspectiva de sua capacidade de generalização, e não apenas pelo elevado R^2 obtido no ajuste.

Na escala original, o modelo apresentou RMSE de 2,616 e MAE de 1,210 no conjunto de teste. Em termos práticos, o erro absoluto médio representa uma

diferença média de aproximadamente 1,21 unidade de tempo em relação ao *Transit Time* observado, enquanto o RMSE evidencia a influência de erros de maior magnitude.

Esse resultado sugere que o modelo possui utilidade para estimativas e análises comparativas do *Transit Time*, embora sua aplicação operacional deva considerar a perda de desempenho observada fora da amostra.

4.1.2 Diagnóstico dos resíduos

Os testes indicaram presença de heterocedasticidade, com $p < 0{,}001$ no teste de Breusch-Pagan (Breusch; Pagan, 1979). Por essa razão, a inferência estatística foi realizada com erros-padrão robustos HC3.

O teste de normalidade indicou rejeição da hipótese de normalidade dos resíduos. Entretanto, considerando o tamanho da amostra e o objetivo predominantemente preditivo do estudo, a não normalidade dos resíduos não foi utilizada isoladamente como critério para rejeição do modelo.

O valor do teste de Durbin-Watson (Durbin; Watson, 1950) foi 2,009, próximo de 2, não indicando evidência relevante de autocorrelação serial dos resíduos no modelo de *Transit Time*.

A Tabela 8 resume os testes de validação do modelo de *Transit Time*.

Em conjunto, os resultados indicam que o modelo apresenta bom ajuste na amostra de treinamento e desempenho preditivo moderado fora da amostra, com a heterocedasticidade tratada por meio de erros robustos e ausência de evidência relevante de autocorrelação.

Tabela 8 - Resumo dos testes de validação do modelo de “transit time”

Teste	Estatística	Resultado
F-test	p < 0,01	Modelo globalmente significativo
Normalidade dos resíduos	p < 0,01	Resíduos não normais
Breusch-Pagan	p < 0,01	Heterocedasticidade presente
Durbin-Watson	2,007	Sem autocorrelação relevante
VIF (máx.)	3,868	Multicolinearidade em nível aceitável

Fonte: Dados originais da pesquisa.

Do ponto de vista gerencial, os resultados indicam que a gestão do *Transit Time* deve priorizar o monitoramento das rotas logísticas, principal fonte de variabilidade do indicador.

4.2 Modelo preditivo – *Gate to Gate*

O modelo de *Gate to Gate* foi desenvolvido utilizando uma divisão temporal dos dados, com data de corte em 30 de dezembro de 2025. Foram utilizadas 19.615 observações para treinamento e 2.406 observações para teste.

A transformação Box-Cox foi estimada no conjunto de treinamento, resultando em λ = 0,65. Diferentemente do modelo de Transit Time, não foram removidas observações pelo critério do intervalo interquartil. Entretanto, 820 observações influentes foram removidas pela distância de Cook antes da estimação final.

O procedimento *stepwise* selecionou inicialmente 33 variáveis. Após a etapa de significância estatística, permaneceram 32 variáveis. Posteriormente, seis variáveis foram removidas devido à multicolinearidade, resultando em uma especificação final composta por 26 variáveis explicativas.

As variáveis finais representam principalmente cinco dimensões operacionais: características da fazenda de origem; turno de entrada e saída; dia da semana; tipo de veículo; e volume de caixas processado no turno.

O modelo final apresentou R^2 = 0,709, indicando que aproximadamente 70,9% da variabilidade do Gate to Gate transformado foi explicada pelo conjunto de variáveis incluídas. O teste F apresentou $p < 0{,}001$, indicando significância estatística global do modelo.

A Tabela 9 apresenta os coeficientes estimados para o modelo do *Gate to Gate*.

Tabela 9 - Resumo dos coeficientes estimados do modelo “gate to gate”.

Coeficiente (Variável)	Valor	p-valor
R^2	0,709	-
Intercepto (β)	0,446	p < 0,01
FAZENDA_04 ($\alpha_{FAZENDA_04}$)	-1,309	p < 0,01
FAZENDA_06 ($\alpha_{FAZENDA_06}$)	-0,024	p < 0,01
FAZENDA_07 ($\alpha_{FAZENDA_07}$)	-0,217	p < 0,01
FAZENDA_09 ($\alpha_{FAZENDA_09}$)	-0,228	p < 0,01
FAZENDA_12 ($\alpha_{FAZENDA_12}$)	-0,649	p < 0,01
FAZENDA_13 ($\alpha_{FAZENDA_13}$)	0,284	p < 0,01

FAZENDA_14 ($\alpha_{FAZENDA_14}$)	-1,139	p < 0,01
FAZENDA_15 ($\alpha_{FAZENDA_15}$)	-0,999	p < 0,01
FAZENDA_17 ($\alpha_{FAZENDA_17}$)	-0,266	p < 0,01
FAZENDA_18 ($\alpha_{FAZENDA_18}$)	0,359	p < 0,01
FAZENDA_20 ($\alpha_{FAZENDA_20}$)	-0,052	p < 0,01
Dia da semana_seg ($\alpha_{DIA_SEGUNDA}$)	0,297	p < 0,01
Dia da semana_ter (α_{DIA_TERCA})	0,225	p < 0,01
Dia da semana_qua (α_{DIA_QUARTA})	0,207	p < 0,01
Dia da semana_qui (α_{DIA_QUINTA})	0,180	p < 0,01
Dia da semana_sex (α_{DIA_SEXTA})	0,146	p < 0,01
Turno_Saida_Manha ($\alpha_{TURNO_SAIDA_MANHA}$)	-0,893	p < 0,01
Turno_Saida_Madrugada ($\alpha_{TURNO_SAIDA_MADRUGADA}$)	4,742	p < 0,01
Turno_Saida_Tarde ($\alpha_{TURNO_SAIDA_TARDE}$)	1,242	p < 0,01
Turno_Saida_Noite ($\alpha_{TURNO_SAIDA_NOITE}$)	2,871	p < 0,01
Modal_RODO ($\alpha_{modalRODO}$)	0,053	p < 0,01
Turno_Entrada_Manha ($\alpha_{TURNO_ENTRADA_MANHA}$)	1,525	p < 0,01
Turno_Entrada_Tarde ($\alpha_{TURNO_ENTRADA_TARDE}$)	-1,294	p < 0,01
Turno_Entrada_Noite ($\alpha_{TURNO_ENTRADA_NOITE}$)	-3,983	p < 0,01
Qtd_Caixas_Fazenda_Turno ($\alpha_{QTD_CXS_FAZ_TURNO}$)	$1{,}211 \times 10 - 5$	p < 0,01

Fonte: Dados originais da pesquisa.

Entre os coeficientes estimados, destacam-se os efeitos associados aos turnos de saída. Em relação à categoria de referência, o turno da madrugada apresentou coeficiente de 4,74, enquanto o turno da noite apresentou coeficiente de 2,87 e o turno da tarde coeficiente de 1,24. Esses resultados indicam diferenças relevantes no tempo de permanência associadas ao momento de saída da operação.

As variáveis relativas ao turno de entrada também apresentaram efeitos expressivos. O turno de entrada pela manhã apresentou coeficiente de 1,53, enquanto o turno da tarde apresentou coeficiente de −1,29 e o turno da noite coeficiente de −3,98. A variável quantitativa “Qtd_Caixas_Fazenda_Turno” apresentou coeficiente positivo de $1{,}211 \times 10^{-5}$, indicando que o aumento do volume de caixas processado no turno está associado a aumento do tempo de permanência, mantendo-se constantes as demais variáveis.

O maior VIF do modelo final foi 4,813, observado para a variável “Turno_Entrada_Manha”, permanecendo abaixo do limite de 5 adotado no procedimento de controle da multicolinearidade.

Os resultados do modelo de *Gate to Gate* revelam uma estrutura explicativa distinta daquela observada para o *Transit Time*. Enquanto o modelo de *Transit Time* foi predominantemente associado às características das rotas, o modelo de *Gate to Gate* apresentou maior participação de variáveis relacionadas ao contexto

operacional, especialmente turnos, dia da semana, fazenda e volume de caixas processado.

Esse resultado sugere que o tempo de permanência é mais sensível à organização e à sincronização das atividades realizadas na origem. A associação positiva entre o volume de caixas processado no turno e o tempo de permanência é coerente com estudos que destacam a influência da intensidade operacional, da coordenação entre atividades e da capacidade de processamento sobre o desempenho de cadeias agroalimentares (Verdouw *et al.*, 2016; Rahman; Hye, 2021).

A influência dos turnos também apresenta implicações relevantes. As diferenças observadas entre os períodos de entrada e saída sugerem que a duração do processo não depende exclusivamente do volume de operações, mas também da forma como os recursos e as atividades são distribuídos ao longo do dia. Dessa forma, os resultados reforçam a importância de incorporar variáveis temporais e contextuais na modelagem do desempenho logístico agrícola.

A maior estabilidade entre os conjuntos de treinamento e teste, quando comparada ao modelo de Transit Time, constitui outro resultado relevante. O R^2 de 0,621 na escala transformada no conjunto de teste indica que as variáveis operacionais incluídas capturam padrões relativamente recorrentes do processo de permanência. Esse resultado está alinhado com a literatura que aponta a utilidade de dados históricos e variáveis operacionais para aumentar a previsibilidade de processos logísticos e apoiar decisões mais proativas (Martins *et al.*, 2004; Wang *et al.*, 2016; Mahmud *et al.*, 2024).

Entretanto, a presença de possível autocorrelação dos resíduos, indicada pelo valor de Durbin-Watson de 1,346, impõe cautela à interpretação inferencial dos coeficientes. Embora não impeça a utilização do modelo para fins preditivos, esse resultado sugere que parte da estrutura temporal dos dados pode não ter sido completamente capturada pelas variáveis explicativas. Esse aspecto reforça a necessidade de incorporar, em estudos futuros, estruturas temporais ou modelos capazes de representar explicitamente a dependência entre observações sucessivas.

Em conjunto, esses resultados contribuem para a compreensão de que indicadores logísticos temporalmente distintos podem responder a mecanismos operacionais diferentes. Essa diferenciação constitui uma das principais contribuições

analíticas do estudo, pois demonstra que uma estratégia única de monitoramento e previsão pode ser insuficiente para representar a complexidade do desempenho logístico agrícola.

Em diálogo com Verdouw *et al.* (2016) e Rahman e Hye (2021), o resultado especifica uma condição importante para o uso de dados históricos: a previsibilidade tende a ser maior quando as variáveis representam mecanismos operacionais recorrentes e potencialmente controláveis. Ao mesmo tempo, a possível autocorrelação mostra que a recorrência temporal ainda não foi totalmente representada. A implicação para a Engenharia de Produção é que modelos interpretáveis podem simultaneamente apoiar a previsão e revelar alavancas de intervenção, desde que a dependência temporal seja incorporada à validação e à especificação do modelo.

4.2.1 Desempenho preditivo

A Tabela 10 apresenta o desempenho do modelo de *Gate to Gate* nos conjuntos de treinamento e teste.

O modelo de Gate to Gate apresentou desempenho mais consistente entre os conjuntos de treinamento e teste. Na escala transformada, o R^2 passou de 0,709 no treinamento para 0,621 no teste.

Tabela 10 - Desempenho preditivo do modelo de "*Gate to Gate*".

Amostra	R^2 BC	RMSE BC	MAE BC	RMSE Original	MAE Original
Treinamento	0,709	0,828	0,670	1,985	1,686
Teste	0,621	0,854	0,672	1,189	0,944

Fonte: Dados originais da pesquisa.

No conjunto de teste, o MAE foi 0,944 e o RMSE foi 1,189 na escala original. Isso indica que o modelo apresentou erro absoluto médio inferior a uma unidade de tempo, embora os erros de maior magnitude tenham aumentado o RMSE.

A menor diferença entre os resultados de treinamento e teste, quando comparada ao modelo de *Transit Time*, sugere maior estabilidade preditiva do modelo de *Gate to Gate*. Esse comportamento pode estar relacionado à divisão temporal

adotada e à maior capacidade das variáveis operacionais incluídas no modelo de capturar padrões recorrentes do processo de permanência nas fazendas.

4.2.2 Diagnóstico dos resíduos

O teste de Breusch-Pagan indicou presença de heterocedasticidade, com $p < 0{,}001$. Assim, os erros-padrão HC3 foram utilizados para a inferência dos coeficientes.

O teste de normalidade rejeitou a hipótese de normalidade dos resíduos. Entretanto, como no modelo de *Transit Time*, essa característica foi interpretada considerando o tamanho amostral e o objetivo da modelagem.

O valor do teste de Durbin-Watson foi 1,346, inferior a 2. Diferentemente do modelo de *Transit Time*, esse resultado sugere a possibilidade de autocorrelação positiva dos resíduos. Portanto, não é metodologicamente recomendável afirmar que houve ausência de autocorrelação relevante para esse modelo.

Esse resultado deve ser considerado uma limitação da especificação atual, especialmente porque a divisão temporal dos dados pode preservar dependências associadas à sequência das operações. Embora a presença de autocorrelação não invalide o uso do modelo para previsão, ela indica que os erros podem não ser completamente independentes.

A Tabela 11 resume os resultados dos testes de validação do modelo de *Gate to Gate*.

Tabela 11 - Resumo dos testes de validação do modelo de “gate to gate”

Teste	Estatística	Resultado
F-test	p < 0,01	Modelo globalmente significativo
Normalidade dos resíduos	p < 0,01	Resíduos não normais
Breusch-Pagan	p < 0,01	Heterocedasticidade presente
Durbin-Watson	1,345	Possível autocorrelação dos resíduos
VIF (máx.)	4,813	Multicolinearidade em nível aceitável

Fonte: Dados originais da pesquisa.

Assim, o modelo de *Gate to Gate* apresentou bom desempenho preditivo fora da amostra, mas seus resultados inferenciais devem ser interpretados com cautela em razão da presença de heterocedasticidade e da possível autocorrelação dos resíduos.

Sob a perspectiva gerencial, os resultados evidenciam que a melhoria do indicador *Gate to Gate* depende principalmente da organização dos fluxos operacionais, com ênfase na gestão de turnos, sincronização de atividades e controle do volume de operações por período. Isso sugere que ações voltadas à programação logística e balanceamento de carga entre turnos podem gerar ganhos relevantes de eficiência.

4.3 Comparação do desempenho dos modelos

A Tabela 12 resume o desempenho preditivo dos dois modelos no conjunto de teste.

Tabela 12 - Resumo do desempenho preditivo dos modelos no conjunto de teste

Modelo	R^2 BC	RMSE BC	MAE BC	RMSE Original	MAE Original
Transit Time	0,473	1,653	0,823	2,616	1,210
Gate to Gate	0,621	0,854	0,672	1,189	0,944

Fonte: Dados originais da pesquisa.

A comparação dos resultados evidencia que o modelo de *Gate to Gate* apresentou maior capacidade de generalização. Na escala transformada, o modelo apresentou R^2 = 0,621, enquanto o modelo de *Transit Time* apresentou R^2 = 0,473.

O *Gate to Gate* também apresentou menores erros no conjunto de teste, com MAE de 0,944 e RMSE de 1,189, em comparação com MAE de 1,210 e RMSE de 2,616 para o *Transit Time*.

Embora o modelo de *Transit Time* tenha apresentado maior R^2 no conjunto de treinamento, essa vantagem não se manteve fora da amostra. A redução do R^2 entre treinamento e teste indica uma perda de desempenho mais acentuada, sugerindo menor capacidade de generalização para as observações utilizadas na avaliação.

Esse resultado é particularmente importante porque evidencia que o desempenho preditivo não deve ser avaliado exclusivamente pelo ajuste dentro da amostra. No presente estudo, a avaliação fora da amostra mostrou que o modelo de *Gate to Gate* foi mais estável e apresentou maior capacidade de previsão em dados não utilizados na estimação.

A diferença entre os modelos também pode ser interpretada à luz da natureza dos indicadores. O *Transit Time* é fortemente condicionado por características específicas das rotas, que podem apresentar grande heterogeneidade entre os diferentes trajetos. O *Gate to Gate*, por sua vez, apresenta maior relação com padrões operacionais recorrentes, como turnos, dias da semana, fazendas e volume processado por período.

Do ponto de vista gerencial, os resultados sugerem que o modelo de *Gate to Gate* apresenta maior potencial para apoiar previsões operacionais em novas observações, enquanto o modelo de *Transit Time* pode ser utilizado principalmente como ferramenta de análise comparativa entre rotas e identificação de padrões associados ao desempenho do transporte.

4.4 Integração com painel de *Business Intelligence*

Os resultados dos modelos preditivos evidenciam elevado potencial de integração ao painel de *Business Intelligence*, ampliando sua utilidade para além da análise descritiva. Atualmente, o painel permite acompanhar os valores realizados de *Transit Time* e *Gate to Gate*, comparando-os com metas operacionais e séries históricas.

Com os modelos estimados, torna-se possível incorporar ao ambiente analítico uma nova camada de informação: os valores previstos dos indicadores com base nas características operacionais de cada viagem. Essa abordagem permite evoluir de uma análise retrospectiva para uma análise orientada à antecipação de desvios.

No caso do *Transit Time*, a integração do modelo possibilita comparar o tempo realizado com o esperado para cada rota, facilitando a identificação de desvios e a priorização de ações corretivas. Para o *Gate to Gate*, a influência de variáveis ligadas à dinâmica operacional permite destacar contextos com maior propensão a gargalos.

A integração entre valores realizados, metas e previsões tende a qualificar significativamente a análise dos indicadores, permitindo uma atuação mais proativa na gestão logística.

Dessa forma, a contribuição científica da integração entre o painel de *Business Intelligence* e os modelos preditivos consiste na articulação de diferentes níveis de

análise do desempenho logístico. O painel permite observar o comportamento realizado dos indicadores e identificar padrões históricos, enquanto os modelos preditivos permitem estimar o desempenho esperado e identificar as variáveis associadas à sua variabilidade.

Essa integração evidencia que o apoio à decisão pode ser estruturado em três níveis complementares: monitoramento do desempenho observado, diagnóstico dos fatores associados aos desvios e antecipação de resultados futuros. No caso do *Transit Time*, essa abordagem permite comparar o tempo realizado com padrões esperados para diferentes rotas. No caso do *Gate to Gate*, permite identificar situações operacionais associadas a maior permanência e antecipar possíveis gargalos.

Assim, o estudo demonstra empiricamente que a integração entre BI e modelagem preditiva pode funcionar como uma arquitetura analítica para conectar dados operacionais, interpretação estatística e decisão gerencial. Essa perspectiva amplia a aplicação tradicional de dashboards, que frequentemente se limita à visualização retrospectiva de indicadores, ao incorporar mecanismos de diagnóstico e antecipação de desvios.

Em comparação com estudos que enfatizam os ganhos de visibilidade e eficiência associados ao BI na logística (Simão *et al.*, 2023; Nascimento *et al.*, 2024), os resultados acrescentam uma dimensão metodológica: a arquitetura analítica deve ser avaliada tanto pelas informações que oferece à decisão quanto pela validade fora da amostra de sua camada preditiva. A implicação científica é que transformação digital, neste contexto, não se resume à digitalização de relatórios, mas envolve a integração verificável entre monitoramento, explicação e previsão.

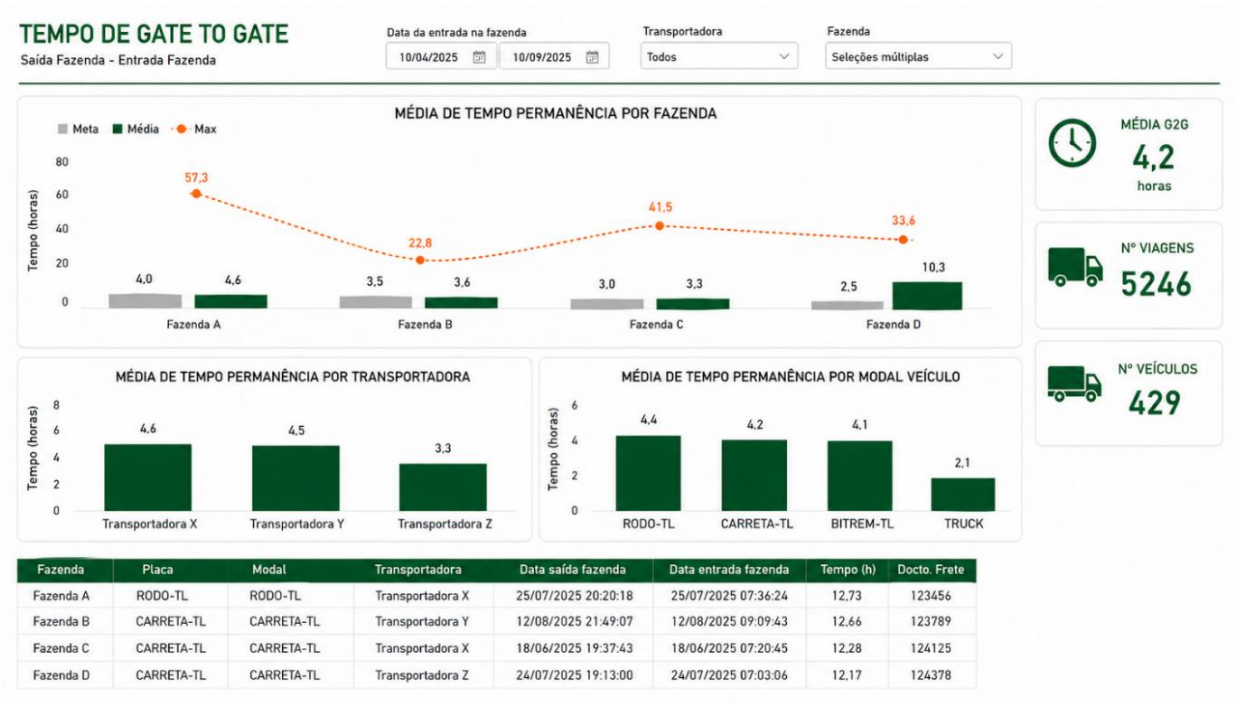
4.5 Painel de *Business Intelligence*

O painel de *Business Intelligence* foi desenvolvido para consolidar a análise dos indicadores logísticos, com foco em *Transit Time* e *Gate to Gate*. A ferramenta permite segmentação por período, fazenda, transportadora e unidade industrial, ampliando a visão sobre o desempenho operacional.

A análise do *Gate to Gate* (Figura 3) evidenciou variabilidade relevante entre fazendas, indicando diferenças no tempo de permanência e carregamento. Esse resultado converge com o modelo preditivo, que apontou influência significativa da fazenda e dos turnos operacionais.

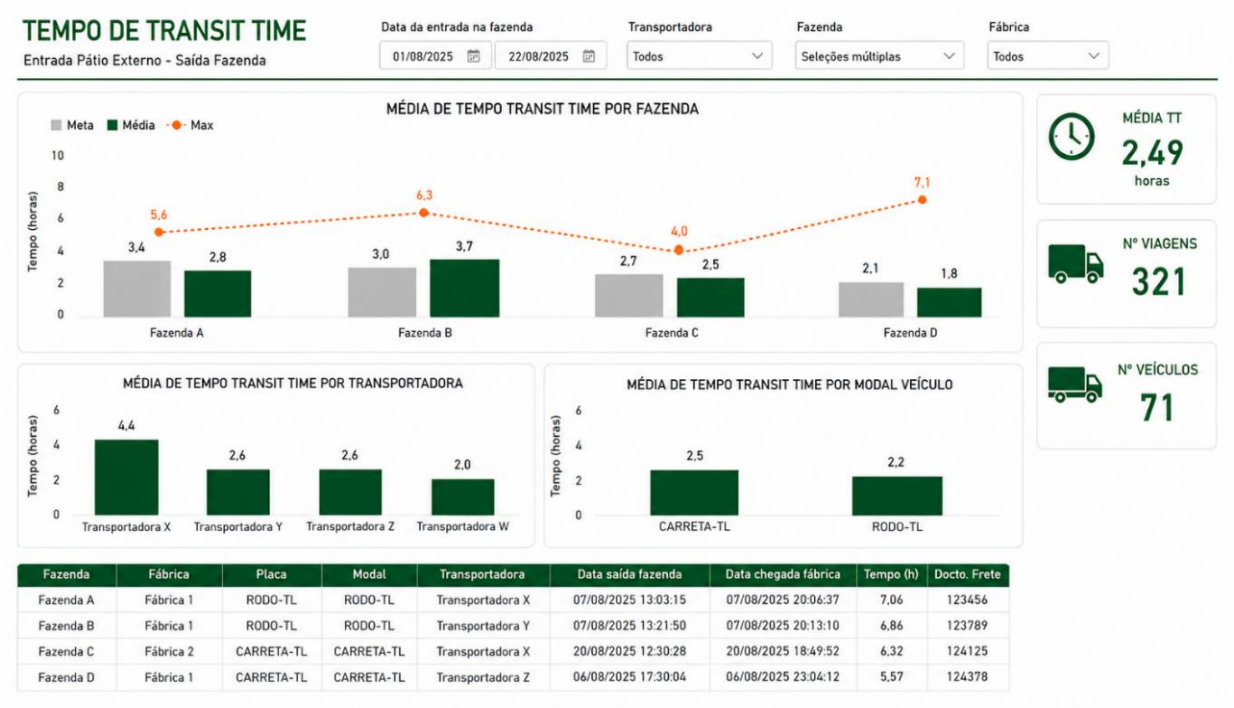
Já a análise do *Transit Time* (Figura 4) revelou diferenças entre rotas, transportadoras e tipos de veículo, reforçando os achados estatísticos sobre a predominância da variável rota na explicação do tempo de deslocamento.

Figura 3 - Painel de *Business Intelligence* do indicador *Gate to Gate*



Fonte: Dados originais da pesquisa.

Figura 4 - Painel de *Business Intelligence* do indicador *Transit Time*



Fazenda	Fábrica	Placa	Modal	Transportadora	Data saída fazenda	Data chegada fábrica	Tempo (h)	Docto. Frete
Fazenda A	Fábrica 1	RODO-TL	RODO-TL	Transportadora X	07/08/2025 13:03:15	07/08/2025 20:06:37	7,06	123456
Fazenda B	Fábrica 1	RODO-TL	RODO-TL	Transportadora Y	07/08/2025 13:21:50	07/08/2025 20:13:10	6,86	123789
Fazenda C	Fábrica 2	CARRETA-TL	CARRETA-TL	Transportadora X	20/08/2025 12:30:28	20/08/2025 18:49:52	6,32	124125
Fazenda D	Fábrica 1	CARRETA-TL	CARRETA-TL	Transportadora Z	06/08/2025 17:30:04	06/08/2025 23:04:12	5,57	124378

Fonte: Dados originais da pesquisa.

De forma integrada, o *dashboard* permite monitorar indicadores, identificar padrões e reconhecer gargalos logísticos, consolidando-se como instrumento de suporte à decisão.

4.6 Uso do *dashboard* na gestão logística: comparação "*As Is*" e "*To Be*"

A implementação do painel de *Business Intelligence* permitiu a reestruturação do processo de tomada de decisão na gestão logística citrícola, promovendo a transição de um modelo tradicional ("*As Is*") para um modelo orientado a dados ("*To Be*").

A Tabela 13 apresenta a sistematização dessa transição, evidenciando mudanças nas dimensões de eficiência, padronização, produtividade, suporte à decisão e desempenho operacional.

A análise comparativa evidencia ganho considerável de eficiência: o processo anterior exigia aproximadamente uma hora de preparação manual antes de cada reunião, enquanto o painel passou a disponibilizar os indicadores após a atualização automatizada. Isso representa economia aproximada de uma hora por ciclo de

preparação, além de reduzir inconsistências decorrentes da consolidação de múltiplas planilhas. Como a frequência de uso não foi registrada sistematicamente, não se estimou a economia acumulada no horizonte analisado.

Tabela 13 - Framework de transição do modelo de gestão “As Is” vs. “To Be”

Dimensão	**“*As-Is*” (processo anterior)**	**“*To-Be*” (orientado a dados)**
Eficiência	Relatórios manuais em Excel, com múltiplas bases e ~1h de preparação pré-reunião.	Atualização automatizada no *Power BI*, com acesso imediato aos indicadores.
Padronização	Planilhas com cálculos manuais, sujeitas a erros e inconsistências.	KPIs padronizados via modelo de dados e medidas DAX, garantindo uniformidade.
Produtividade	Tempo concentrado na preparação dos dados.	Foco deslocado para análise, com antecipação das reuniões em ~1h.
Suporte à decisão	Relatórios estáticos, com baixa flexibilidade analítica.	Dashboard interativo como fonte oficial, permitindo análises em tempo real.
Desempenho operacional	Monitoramento limitado e baixa agilidade na identificação de desvios.	Acompanhamento contínuo (*Transit Time* e *Gate to Gate*), facilitando controle e ações.

Fonte: Dados originais da pesquisa.

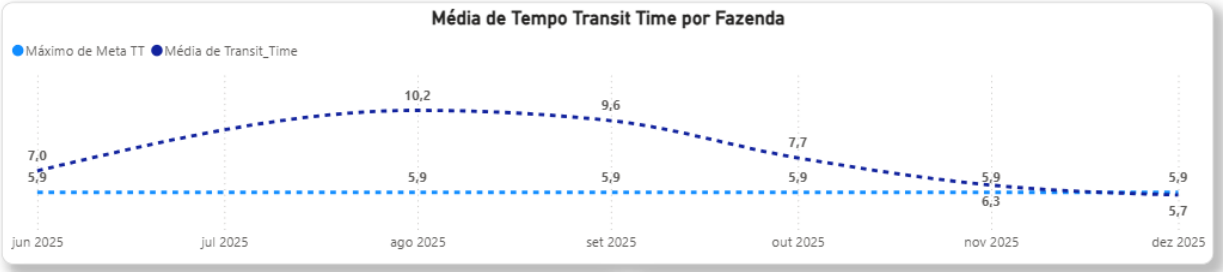
Também ocorreu mudança no foco das atividades, com menor esforço operacional e maior dedicação à análise e interpretação dos resultados. O dashboard passou a atuar como fonte oficial nas reuniões operacionais, substituindo relatórios estáticos por um ambiente analítico interativo.

Em relação ao desempenho, observou-se uma redução nos valores dos indicadores *Transit Time* (Figura 5) e *Gate to Gate* (Figura 6) no período posterior à implementação do painel.

Essa coincidência temporal não permite atribuir causalmente a redução exclusivamente ao uso do *dashboard*, uma vez que outros fatores operacionais podem ter influenciado os resultados. Entretanto, a evolução observada é consistente com a hipótese de que o aumento da visibilidade e do monitoramento sistemático pode favorecer a identificação de desvios e a adoção de ações corretivas.

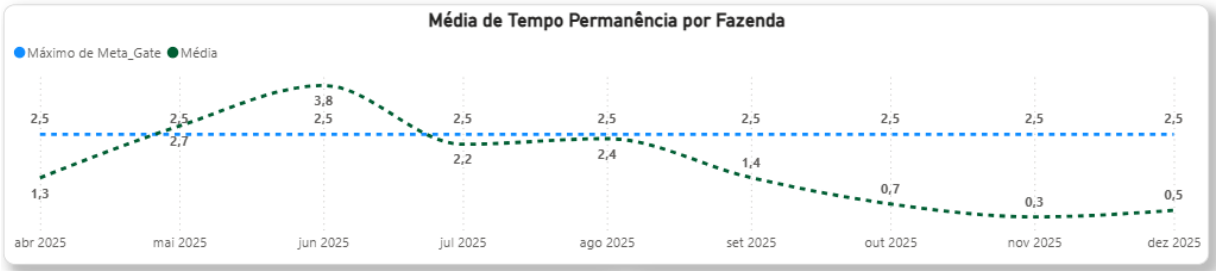
Embora não seja possível estabelecer causalidade direta, as evidências empíricas indicam que o uso sistemático de ferramentas analíticas contribui para o aprimoramento do controle operacional e para uma gestão mais proativa.

Figura 5 - Exemplo de evolução de *Transit Time*



Fonte: Dados originais da pesquisa.

Figura 6 - Exemplo de evolução de *Gate to Gate*



Fonte: Dados originais da pesquisa.

5 CONCLUSÕES

Este estudo avaliou a contribuição do *Business Intelligence* e da modelagem preditiva para a gestão logística agrícola em uma indústria citrícola. Os resultados demonstram que a integração entre visualização analítica, indicadores estruturados e modelos estatísticos pode ampliar a capacidade de monitoramento, diagnóstico e antecipação de comportamentos operacionais.

A primeira contribuição do estudo está relacionada à compreensão das diferentes estruturas associadas aos indicadores logísticos analisados. O modelo de *Transit Time* evidenciou forte predominância das variáveis associadas às rotas, indicando que o tempo de deslocamento é condicionado principalmente por características estruturais e específicas dos trajetos. Por outro lado, o modelo de *Gate to Gate* apresentou maior associação com variáveis contextuais e operacionais, como turnos, dia da semana, fazenda e volume de caixas processado. Essa diferença demonstra que indicadores logísticos distintos podem refletir dimensões diferentes do desempenho operacional e, portanto, requerem estratégias específicas de análise e modelagem.

A segunda contribuição está relacionada à avaliação da capacidade de generalização dos modelos. Embora o modelo de *Transit Time* tenha apresentado maior ajuste no conjunto de treinamento, o modelo de *Gate to Gate* apresentou desempenho superior no conjunto de teste e menor diferença entre os resultados dentro e fora da amostra. Esse resultado reforça a importância de avaliar modelos preditivos com dados não utilizados na estimação, evitando que o desempenho seja interpretado exclusivamente a partir do ajuste dentro da amostra. A comparação também demonstra que maior complexidade ou maior poder explicativo dentro da amostra não implica necessariamente melhor capacidade preditiva em novas observações.

A terceira contribuição refere-se à integração entre análise descritiva e preditiva em um sistema de apoio à decisão. O *Business Intelligence* permitiu consolidar e padronizar indicadores logísticos, reduzir processos manuais e ampliar a visibilidade operacional. A incorporação dos modelos preditivos adicionou uma camada analítica voltada à identificação de fatores associados à variabilidade dos indicadores e à estimativa de comportamentos esperados. Dessa forma, o estudo demonstra que a integração entre BI e modelagem preditiva pode ampliar a função dos sistemas analíticos, que deixam de atuar apenas como instrumentos de monitoramento retrospectivo e passam também a apoiar diagnóstico e antecipação de desvios.

Do ponto de vista da Engenharia de Produção, os resultados contribuem para a compreensão de como dados operacionais podem ser transformados em informação analítica e, posteriormente, em suporte estruturado à decisão em sistemas logísticos complexos. A principal contribuição está na demonstração empírica de que a integração entre essas abordagens permite relacionar diferentes níveis de análise (*e.g.*, desempenho observado, fatores associados à variabilidade e previsão de resultados) em uma operação agrícola real.

Do ponto de vista gerencial, a adoção integrada de BI e análise preditiva favorece a transição de uma gestão predominantemente reativa para uma abordagem mais orientada por dados. O painel fornece visibilidade contínua dos indicadores, enquanto os modelos permitem identificar fatores associados ao desempenho e apoiar a priorização de ações. No entanto, a utilização dos modelos deve considerar

as diferenças de capacidade de generalização observadas e as limitações identificadas nos diagnósticos dos resíduos.

Como limitações, destaca-se que o estudo foi conduzido como caso único em uma empresa e em uma cadeia produtiva específica, o que restringe a validade externa dos resultados. O horizonte de abril de 2025 a fevereiro de 2026 também pode não representar integralmente variações sazonais ou mudanças estruturais de longo prazo. A validação baseou-se em uma única separação entre treinamento e teste, sem validação cruzada ou repetição temporal, e a avaliação operacional do painel foi observacional, sem desenho causal ou registro sistemático de frequência de uso. Além disso, os modelos foram baseados em regressão linear e podem não capturar relações não lineares complexas; a possível autocorrelação no modelo de *Gate to Gate* recomenda cautela na interpretação inferencial dos coeficientes.

Como agenda para pesquisas futuras, recomenda-se replicar a análise em outras empresas e cadeias do agronegócio, ampliar o horizonte temporal e empregar validação cruzada compatível com a estrutura temporal dos dados. Sugere-se comparar a regressão linear com Random Forest, XGBoost e Gradient Boosting, avaliando se os ganhos preditivos compensam a menor interpretabilidade e utilizando técnicas de explicabilidade para preservar a utilidade gerencial. Também se recomenda testar modelos com dependência temporal explícita e avaliar a implementação completa da camada preditiva no ambiente de BI por meio de indicadores de uso, tempo economizado e efeitos operacionais acompanhados longitudinalmente.

Em síntese, a contribuição do estudo para a Engenharia de Produção consiste em demonstrar como uma arquitetura integrada de BI e modelagem preditiva pode apoiar decisões em sistemas logísticos complexos, articulando monitoramento, diagnóstico e antecipação. Ao transformar dados operacionais dispersos em indicadores padronizados, previsões avaliadas fora da amostra e informações acionáveis, a abordagem conecta apoio à decisão, transformação digital e gestão logística baseada em dados, oferecendo um referencial aplicável ao desenho e à avaliação de sistemas analíticos em operações produtivas.

AGRADECIMENTOS

Os autores agradecem à Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo pelo apoio institucional e infraestrutura disponibilizada.

REFERÊNCIAS

ANDRADE, Carolina Aparecida Roque de; GERST, Kevin Robert; GIMENEZ, Igor Maciesza. Logística 4.0 e suas Aplicações na Indústria Agrícola. **Revista do Encontro de Gestão e Tecnologia**, [s. l.], v. 1, n. 03, p. 41–47, 2024. DOI: 10.5281/zenodo.11075165. Disponível em: https://revista.fateczl.edu.br/index.php/engetec_revista/article/view/84. Acesso em: 16 abr. 2026.

ÂNGELO, L. B. **Indicadores de desempenho logístico**. Florianópolis: Grupo de Estudos Logísticos, 2005.

BALLOU, Ronald H. **Gerenciamento da cadeia de suprimentos/logística empresarial**. 5. ed. Porto Alegre: Bookman, 2006.

BALLOU, Ronald H. **Logística empresarial**: transportes, administração de materiais e distribuição física. São Paulo: Atlas, 1998.

BEJARANO-MARTÍNEZ, Carlos; CARO, Martha. Exploring Performance Evaluation Methodologies in Agri-food Supply Chains. **Operations and Supply Chain Management**: An International Journal, [s. l.], v. 19, n. 2, p. 226-247, 2026. DOI: 10.31387/oscm0650517.

BHAGAT, Deepak; DHAR, U. R. Agriculture supply chain management: A review. **IUP Journal of Supply Chain Management**, [s. l.], v. 8, n. 3, 2011.

BOECHAT, Cássio Arruda. A questão agrária e a crise da citricultura brasileira. **Revista Teoria e Cultura**, Juiz de Fora, v. 10, n. 2, p. 66-83, 2015. Disponível em: https://periodicos.ufjf.br/index.php/TeoriaeCultura/article/view/12259/6486. Acesso em: 16 abr. 2026.

BONEL, Cristiano. **Metodologia e engenharia de requisitos para projetos de business intelligence**: análise de dados levada a sério. Rio de Janeiro: Ed. do Autor, 2021.

BOWERSOX, Donald J.; CLOSS, David J.; COOPER, M. Bixby. **Gestão logística de cadeias de suprimento**. Porto Alegre: Bookman, 2006.

BOX, George E. P.; COX, David R. An analysis of transformations. **Journal of the Royal Statistical Society**: Series B (Methodological), [s. l.], v. 26, n. 2, p. 211-252, 1964. DOI: 10.1111/j.2517-6161.1964.tb00553.x.

BREUSCH, Trevor S.; PAGAN, Adrian R. A simple test for heteroscedasticity and random coefficient variation. **Econometrica**, [s. l.], v. 47, n. 5, p. 1287-1294, 1979.

CAIXETA FILHO, José Vicente. Logística para a agricultura brasileira. **Revista Brasileira de Comércio Exterior**, [s. l.], n. 103, p. 18-30, 2010. Disponível em: https://esalqlog.esalq.usp.br/upload/kceditor/files/2015/08/Logistica-para-agricultura-brasileira.pdf. Acesso em: 10 dez. 2025.

CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL (CNA). **A importância da citricultura nacional**. Brasília: CNA, 2019. Disponível em: https://www.cnabrasil.org.br/noticias/a-importancia-da-citricultura-nacional. Acesso em: 17 set. 2025.

COOK, R. Dennis. Detection of influential observation in linear regression. **Technometrics**, [s. l.], v. 19, n. 1, p. 15-18, 1977. DOI: 10.1080/00401706.1977.10489493.

DE BEM NORO, G.; DE OLIVEIRA, J. H. R.; WITTMANN, M.; ZAGO, C. A. A logística como fator determinante para o sistema de agronegócio (SAG) do Brasil. In: CONGRESSO BRASILEIRO DE CUSTOS, 12., 2005, Itajaí. **Anais...** Florianópolis: ABC, 2005. p. 1-10.

DURBIN, James; WATSON, Geoffrey S. Testing for serial correlation in least squares regression: I. **Biometrika**, [s. l.], v. 37, n. 3/4, p. 409-428, 1950. DOI: 10.2307/2332391.

EHRIG, Magnus Emanuel; AURÉLIO, Cristian Matheus. A Logística Brasileira: Desafios, Avanços E Perspectivas Para O Desenvolvimento Nacional. **RCMOS - Revista Científica Multidisciplinar O Saber**, Brasil, v. 1, n. 1, 2026. DOI: 10.51473/rcmos.v1i1.2026.2032.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA). **Citros**. Cruz das Almas: Embrapa Mandioca e Fruticultura, 2024. Disponível em: https://www.embrapa.br/mandioca-e-fruticultura/cultivos/citros. Acesso em: 17 set. 2025.

FARO, K. C. Entraves logísticos ao desenvolvimento agrícola brasileiro à luz da Nova Economia Institucional. **Diversitas Journal**, [s. l.], v. 6, n. 4, p. 4043-4066, 2021. DOI: 10.48017/dj.v6i4.1992.

FAWCETT, Stanley; CLINTON, Steven; BIROU, Laura. Evaluating the impact of logistics performance. **International Journal of Physical Distribution & Logistics Management**, [s. l.], v. 27, n. 9/10, 1997.

FÁVERO, Luiz Paulo; SANTOS, H. P.; BELFIORE, Patrícia; DUARTE, A.; COSTA, I. P. A.; TERRA, A. V.; MOREIRA, M. A. L.; TARANTIN JUNIOR, W.; SANTOS, M. A proposal for a new Python library implementing stepwise procedure. **Algorithms**, [s. l.], v. 17, n. 11, p. 502, 2024. DOI: 10.3390/a17110502.

GARCIA, Fernanda A.; SILVA, A. L.; PINHEIRO, R. G. D.; MACHADO, R. L. A framework for measuring logistics performance in the wine industry. **International Journal of Production Economics**, [s. l.], v. 135, n. 1, p. 284-298, 2012. DOI: 10.1016/j.ijpe.2011.08.003.

GARCIA, Junior Ruiz; VIEIRA FILHO, J. E. R. Logística e armazenamento. **Diagnóstico e desafios da agricultura brasileira**, Brasília, v. 1, p. 59-88, 2019. Disponível em: https://portalantigo.ipea.gov.br/agencia/images/stories/PDFs/livros/livros/191126_diagnostico_e_desafios_da_agricultura_brasileira.pdf. Acesso em: 16 abr. 2026.

HAOBIJAM, J. W.; SERAM, Devina. Supply Chain Optimization: AI in Agriculture Distribution. In: **Transforming Agriculture through Artificial Intelligence for Sustainable Food Systems**. Singapore: Springer Nature Singapore, 2025. p. 109-127. DOI: 10.1007/978-981-96-4795-8_7.

HARDJOMIDJOJO, Hartrisari; WIGUNA, B. P.; ARITONANG, I. D. R.; GURITNO, A. D. Sustainable Agro-industry Logistics Solutions using Spatial Analysis. **Operations and Supply Chain Management**, [s. l.], v. 15, n. 1, p. 41-55, 2022. DOI: 10.31387/oscm0480329.

HIJJAR, Mohammed. Desempenho logístico e suas métricas. **Revista Tecnologística**, São Paulo, 2005.

JIN, Dong-Hui; KIM, Hyun-Jung. Integrated understanding of big data, big data analysis, and business intelligence: A case study of logistics. **Sustainability**, [s. l.], v. 10, n. 10, p. 3778, 2018. DOI: 10.3390/su10103778.

KAMBLE, Sachin S.; GUNASEKARAN, Angappa; GAWANKAR, Shradha A. Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. **International Journal of Production Economics**, [s. l.], v. 219, p. 179-194, 2020. DOI: 10.1016/j.ijpe.2019.05.022.

KHANDELWAL, Chandni; JAIN, P.; SHARMA, A.; TYAGI, S. Agriculture supply chain management: a review (2010–2020). **Materials Today**: Proceedings, [s. l.], v. 47, p. 3144-3153, 2021.

KIMBALL, Ralph; ROSS, Margy. **The data warehouse toolkit**. 3. ed. New York: Wiley, 2013.

MAHMUD, Faiyaz Bin; ROUF, Md Jawad Bin; MAHIN, Mohammad Mynul Islam. Optimization of rice supply chain through multivariate forecasting using ensemble learning models. **Supply Chain Insider**, [s. l.], v. 14, n. 1, 2024. DOI: 10.5281/zenodo.13865817.

MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. **Introduction to linear regression analysis**. 5. ed. New Jersey: Wiley, 2012.

MUMTAZ, Ubaidullah; ALI, Yousaf; PETRILLO, Antonella. A linear regression approach to evaluate the green supply chain management impact on industrial

organizational performance. **Science of the Total Environment**, [s. l.], v. 624, p. 162-169, 2018. DOI: 10.1016/j.scitotenv.2017.12.089.

NANGPIIRE, C.; SALIFU, Z. N.; BEDUWA, E. Transport optimization practices in the supply chain of agro-firm companies: evidence in Ghana. **Journal of Future Sustainability**, [s. l.], v. 5, n. 1, p. 1-12, 2025. DOI: 10.5267/j.jfs.2025.1.001.

NASCIMENTO, F. L.; NOGUEIRA, S. M. S.; FERREIRA JUNIOR, L. D.; LORDELO, S. A. V. Aplicação do business intelligence em logística: uma revisão bibliométrica. **Revista Produção Online**, Florianópolis, v. 24, n. 1, 2024. DOI: 10.14488/1676-1901.v24i1.5113.

NI, Feng; ARNOTT, David; GAO, Shijia. The anchoring effect in business intelligence supported decision-making. **Journal of Decision Systems**, [s. l.], v. 28, n. 2, p. 67-81, 2019. DOI: 10.1080/12460125.2019.1620573.

NIU, Yanfang; YING, L.; YANG, J.; BAO, M.; SIVAPARTHIPAN, C. B. Organizational business intelligence and decision making using big data analytics. **Information Processing & Management**, [s. l.], v. 58, n. 6, p. 102725, 2021. DOI: 10.1016/j.ipm.2021.102725.

NORO, G. A.; SCHMITZ, J.; COSTA, V. Custos logísticos no agronegócio brasileiro. **Revista de Política Agrícola**, Brasília, v. 14, n. 3, 2005.

OLIVEIRA, Andréa Leda Ramos de; LOPES, R. L.; CAIXETA-FILHO, J. V. Logistical transportation routes optimization for Brazilian soybean: an application of the origin-destination matrix. **Ciência Rural**, Santa Maria, v. 51, n. 2, p. e20190786, 2020. DOI: 10.1590/0103-8478cr20190786.

RAHMAN, SM Toufiqur; HYE, Abdul. The Role Of Predictive Analytics In Enhancing Agribusiness Supply Chains. **Review of Applied Science and Technology**, [s. l.], v. 6, n. 1, p. 183-229, 2021. DOI: 10.63125/n9z10h68.

ROUTROY, Srikanta; BEHERA, Astajyoti. Agriculture supply chain: a systematic review of literature and implications for future research. **Journal of Agribusiness in Developing and Emerging Economies**, [s. l.], v. 7, n. 3, p. 275-302, 2017. DOI: 10.1108/JADEE-06-2016-0039.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business intelligence and analytics**: systems for decision support. 10. ed. Boston: Pearson, 2018.

SHARMA, Manoj Kumar; DHAKA, Vijaypal Singh; SHEKHAWAT, Rajveer Singh. Intelligent agro-food chain supply. In: **Internet of Things and Analytics for Agriculture**. Singapore: Springer Singapore, 2021. v. 3, p. 65-91. DOI: 10.1007/978-981-16-6210-2_4.

SILVA, João Carlos. Business intelligence como ferramenta de apoio à gestão logística. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, [s. l.], n. 11, 2023.

SIMÃO, João A. S.; SARTORI, G.; REIS, J. S. M.; OLIVEIRA, S. B. Os efeitos da utilização do Power BI no controle de processos de armazenagem em um centro de distribuição. **Revista Produção Online**, Florianópolis, v. 23, n. 1, p. 1-15, 2023. DOI: 10.14488/1676-1901.v23i1.4431.

SNEE, Ronald D. Discussion of developments in linear regression methodology: 1959-1982. **Technometrics**, [s. l.], v. 25, n. 3, p. 230, 1983. DOI: 10.2307/1268604.

TIPERLOG. **Relatório técnico**. [S. l.], 2005. Disponível em: https://pt.scribd.com/document/899878926/3. Acesso em: 16 abr. 2026.

TRIVELLAS, Panagiotis; MALINDRETOS, Georgios; REKLITIS, Panagiotis. Implications of green logistics management on sustainable business and supply chain performance: evidence from a survey in the greek agri-food sector. **Sustainability**, [s. l.], v. 12, n. 24, p. 10515, 2020. DOI: 10.3390/su122410515.

TURBAN, Efraim; SHARDA, Ramesh; DELEN, Dursun. **Decision support and business intelligence systems**. 9. ed. New Jersey: Prentice Hall, 2011.

TURBAN, Efraim; SHARDA, Ramesh; DELEN, Dursun; KING, David. **Business intelligence**: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.

VARGAS-MUÑOZ, Juan Camilo; SANCHEZ-NITOLA, Flor Angela; ADARME JAIMES, Wilson; RIOS, Richard. Enhancing Logistical Performance in a Colombian Citrus Supply Chain Through Joint Decision Making: A Simulation Study. **Logistics**, [s. l.], v. 9, n. 1, p. 30, 2025. DOI: 10.3390/logistics9010030.

VIET, Nguyen Quoc; BEHDANI, Behzad; BLOEMHOF, Jacqueline. Value of information to improve daily operations in high-density logistics. **International Journal on Food System Dynamics**, [s. l.], v. 9, n. 1, 2018. DOI: 10.18461/ijfsd.v9i1.911.

WAJONG, A. M. R. Logistics indicators could improve logistics performance of hospitals. **MATEC Web of Conferences**, [s. l.], v. 108, p. 17002, 2017. DOI: 10.1051/matecconf/201710817002.

WOOLDRIDGE, Jeffrey M. **Introductory econometrics**: a modern approach. 4. ed. [S. l.]: South-Western, 2010.

MORAIS, E. O. de; RIBEIRO, K. de L.; VELOSO, R. B.; VELOSO, M. das D. M. Aplicação de modelos de regressão linear e não linear para estimativa de volume de biomassa e estoque de carbono. **Brazilian Journal of Development**, [s. l.], v. 6, n. 7, p. 45621–45632, 2020. DOI: 10.34117/bjdv6n7-259.

YAZDANI, Morteza; GONZALEZ, Ernesto D. R. S.; CHATTERJEE, Prasenjit. A multi-criteria decision-making framework for agriculture supply chain risk management under a circular economy context. **Management Decision**, [s. l.], v. 59, n. 8, p. 1801-1826, 2021. DOI: 10.1108/MD-10-2018-1088.

Biografia do(s) autor(es)

Vinícius Sarandi de Souza
Especialista em Data Science & Analytics pela Universidade de São Paulo. Bacharel em Engenharia de Alimentos pelo Instituto Federal de Educação, Ciência e Tecnologia de São Paulo. Experiência como Pesquisador Bolsista da EMBRAPII (Empresa Brasileira Pesquisa e Inovação Industrial). Desenvolve pesquisas nas áreas de ciência de dados e tomada de decisão aplicadas à gestão da logística agrícola.

Alexandre Duarte
Professor Assistente no curso de Engenharia de Transportes da Faculdade de Ciências e Tecnologia (FCT) da Universidade Federal de Goiás (UFG). Doutorando em Engenharia de Transportes pela Escola Politécnica da Universidade de São Paulo. Mestre em Engenharia de Sistemas Logísticos pela Escola Politécnica da Universidade de São Paulo. Especialista em Data Science & Analytics pela Universidade de São Paulo. Pesquisador no Centro de Inovação em Sistemas Logísticos da Universidade de São Paulo e no Laboratório de Planejamento e Operação de Transportes da Escola Politécnica da Universidade de São Paulo. Desenvolve pesquisas nas áreas de logística e transportes, pesquisa operacional, modelagem matemática, otimização, ciência de dados e inteligência artificial.



Artigo recebido em: 28/04/2026 e aceito para publicação em: 30/08/2026
DOI: https://doi.org/10.14488/1676-1901.v26i4.5893