

CONFIABILIDADE OPERACIONAL NO TRANSPORTE PÚBLICO: CAUSAS, IMPACTOS E PREVISÃO COM INTELIGÊNCIA ARTIFICIAL

Kaio Gefferson de Almeida Mesquita

Centro Universitário Fametro

Área Temática: Tecnologia em Engenharia de Tráfego e Transporte

Área de Conhecimento: Ciências Tecnológicas

Encontro Científico: VI Encontro de Experiências Docentes

RESUMO

Introdução: A baixa confiabilidade operacional no transporte público urbano compromete a atratividade do sistema, gerando atrasos, irregularidade nos intervalos entre veículos e variabilidade dos tempos de viagem. Esses fatores afetam diretamente a percepção dos usuários e reduzem a acessibilidade, sobretudo entre a população mais dependente do transporte coletivo. **Objetivo:** Mensurar, modelar e prever a confiabilidade operacional do transporte público de Fortaleza, identificando suas causas e impactos a partir da integração de diferentes fontes de dados. **Métodos:** Foram utilizadas bases de GPS, bilhetagem eletrônica e GTFS, processadas em etapas de validação, análise exploratória, cálculo de indicadores operacionais (pontualidade, headway e tempo de viagem), modelagem preditiva (Regressão Ridge e Random Forest) e integração entre confiabilidade e demanda observada. **Resultados:** Verificou-se que 42% das viagens apresentaram atraso superior a 3 minutos, com coeficientes de variação de headways acima de 35% em corredores troncais e desvio-padrão médio de 9 minutos no tempo de viagem em linhas de maior extensão. A modelagem preditiva evidenciou melhor desempenho do Random Forest ($R^2=0,9774$), destacando o headway anterior e a carga de validações embarcadas como variáveis de maior relevância. **Considerações finais:** A confiabilidade deve ser tratada como indicador central de qualidade no transporte público, e o uso de inteligência artificial mostrou potencial para antecipar falhas operacionais, subsidiando políticas públicas mais eficientes e orientadas à equidade na mobilidade urbana. **Palavras-chave:** Confiabilidade; Transporte Público; Inteligência Artificial; *Headway*.

INTRODUÇÃO

A confiabilidade operacional é amplamente reconhecida como um dos principais atributos de qualidade no transporte público urbano, estando diretamente relacionada à regularidade e à previsibilidade das viagens. Atrasos, partidas irregulares e a variabilidade no tempo de viagem comprometem a experiência dos passageiros, resultando em tempos de espera elevados e, frequentemente, superestimados pelos usuários. Esse fenômeno impacta

diretamente a atratividade do transporte coletivo frente ao transporte individual motorizado, configurando-se como um desafio estrutural para a mobilidade urbana (VAN OORT, 2011).

Diversos estudos internacionais têm destacado a importância da confiabilidade na avaliação da qualidade percebida do transporte coletivo. Cats e Ferranti (2016) ressaltam que a irregularidade operacional é um dos fatores mais relevantes na decisão do usuário sobre permanecer ou migrar para outros modos de transporte. Braga *et al.* (2020) reforçam que a confiabilidade não deve ser tratada apenas como um indicador técnico, mas como elemento central na formulação de políticas de transporte que visem eficiência e justiça social. Pereira *et al.* (2019), ao analisarem a experiência de usuários em sistemas urbanos, evidenciam que a superestimação do tempo de espera está diretamente associada à irregularidade de headways, afetando negativamente a percepção do serviço.

No Brasil, entretanto, ainda são escassos os trabalhos que mensuram a confiabilidade operacional a partir da integração de diferentes fontes de dados e em contextos de elevada demanda, como é o caso de Fortaleza. A disponibilidade de dados massivos — registros de GPS, bilhetagem eletrônica e arquivos GTFS — cria uma oportunidade singular para analisar, de forma robusta, os fatores que afetam a confiabilidade e suas consequências para a experiência do usuário.

Diante desse contexto, o presente estudo busca mensurar, modelar e prever a confiabilidade operacional do transporte público de Fortaleza por meio da integração de múltiplas bases de dados, associando a análise de indicadores tradicionais (pontualidade, regularidade de *headways* e tempo de viagem) à aplicação de técnicas de inteligência artificial. O objetivo central é identificar os padrões de irregularidade, compreender sua relação com a demanda observada e propor um caminho metodológico capaz de subsidiar políticas públicas mais eficientes e orientadas à equidade na mobilidade urbana.

METODOLOGIA

Este estudo adota uma abordagem quantitativa, exploratória e aplicada, com foco na análise da confiabilidade operacional do transporte público urbano por meio de dados empíricos de GPS e bilhetagem eletrônica, utilizando técnicas de ciência de dados e aprendizado de máquina. A estrutura metodológica foi concebida com base em boas práticas de modelagem preditiva (James *et al.*, 2013; Deisenroth *et al.*, 2020), visando a reprodutibilidade, a robustez estatística e a aplicabilidade dos resultados. As etapas foram organizadas em cinco

blocos principais, conforme ilustrado na Figura 1:

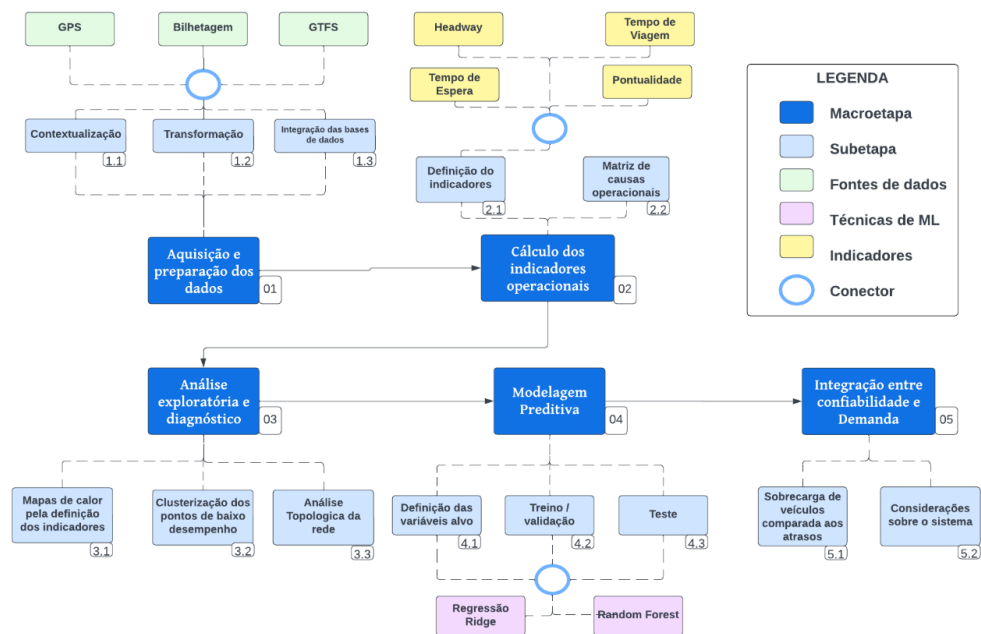


Figura 1 – Metodologia proposta

A primeira etapa é composta pela (i) aquisição e preparação dos dados: Integração de registros de localização (GPS), bilhetagem eletrônica e dados de referência operacional (GTFS), com formatação das colunas e remoção de inconsistências. Seguido da etapa (ii) Cálculo de Indicadores Operacionais: Geração de métricas empíricas de desempenho (headway, tempo de viagem, tempo de espera, variabilidade, lentidão) a partir da estruturação dos dados. A terceira etapa proposta consiste na (iii) Análise Exploratória e Diagnóstico Estatístico: Aplicação de técnicas estatísticas descritivas, boxplots, histogramas e análise de séries temporais para identificar padrões e anomalias. Após elaboração das análises exploratórias dos dados, apoia-se na (iv) Modelagem Preditiva e Avaliação: Construção de modelos de regressão, utilizando algoritmos supervisionados de aprendizado de máquina com validação cruzada. E por fim, última etapa composta pela (v) Integração entre Confiabilidade e Demanda: Análise dos impactos da variabilidade operacional sobre a demanda observada, utilizando dados de bilhetagem para explorar correlações.

A base de dados utilizada neste estudo foi composta por registros operacionais extraídos de sistemas embarcados de monitoramento da frota (GPS), dados de bilhetagem eletrônica (SBE) e General Transit Feed Specification (GTFS). O conjunto de dados abrangeu um período amostral contínuo de 1 semana do mês de novembro de 2023, totalizando aproximadamente 4 milhões de registros diários, com arquivos individuais de cerca de 350 MB, o que exigiu estratégias específicas de processamento e otimização computacional.

Para a análise cruzada com a bilhetagem, foram utilizados dados agregados por parada, horário e veículo, considerando a frequência de validações e a estimativa de ocupação. A integração futura com o GTFS permitirá expandir as análises para pontualidade e trajetos planejados. Este tratamento inicial viabilizou a extração de atributos confiáveis para o cálculo de indicadores operacionais, apresentado na seção seguinte.

Com base nos indicadores operacionais extraídos, desenvolveu-se uma etapa exploratória de modelagem preditiva com o objetivo de estimar o desempenho do sistema sob diferentes condições operacionais. Foram selecionadas as seguintes variáveis independentes: (i) Hora do dia (transformada em variável numérica cíclica); (ii) Dia da semana (como proxy para padrões sazonais curtos); (iii) Informações espaciais de localização; (iv) Distância percorrida desde o início da viagem (proxy para tempo de operação); e (v) Headway programado (podendo ser comparado com o real). A variável dependente foi a confiabilidade (numérica contínua entre 0 e 100%), calculada a partir dos registros de GPS.

Foram implementados dois modelos preditivos de base (Géron, 2019): (i) Regressão Linear Regularizada (Ridge): com validação cruzada para escolha do parâmetro de regularização, servindo como baseline; e (ii) *Random Forest Regressor*: visando capturar relações não lineares e interações entre variáveis. Ambos os modelos foram ajustados com validação cruzada K-Fold ($k=5$), e avaliados com as métricas: (i) RMSE (*Root Mean Squared Error*); (ii) MAE (*Mean Absolute Error*); e R^2 (Coeficiente de Determinação).

RESULTADOS E DISCUSSÃO

A etapa inicial consistiu na avaliação da consistência e da complementaridade das três bases de dados utilizadas: GTFS, GPS e bilhetagem eletrônica. O GTFS apresentou 13.175 viagens programadas associadas ao serviço ativo selecionado, distribuídas em mais de 6.600 paradas cadastradas, configurando uma boa referência para a análise. A base de GPS contabilizou, em média, 98 registros válidos por viagem no recorte analisado, cobrindo de forma satisfatória os corredores estruturais da rede, ainda que com lacunas pontuais de registros nulos. A bilhetagem eletrônica apresentou milhares de validações no período considerado, com média de 900 mil registros por dia, reforçando a relevância dessa fonte como indicador da demanda real. No que se refere ao desempenho operacional, a Figura 2 apresenta o *boxplot* da velocidade por hora do dia a partir dos registros de GPS. Nota-se uma redução significativa das velocidades médias durante os períodos de maior fluxo de passageiros, especialmente entre 6h–8h e 17h–19h, quando predominam congestionamentos e maiores tempos de embarque.

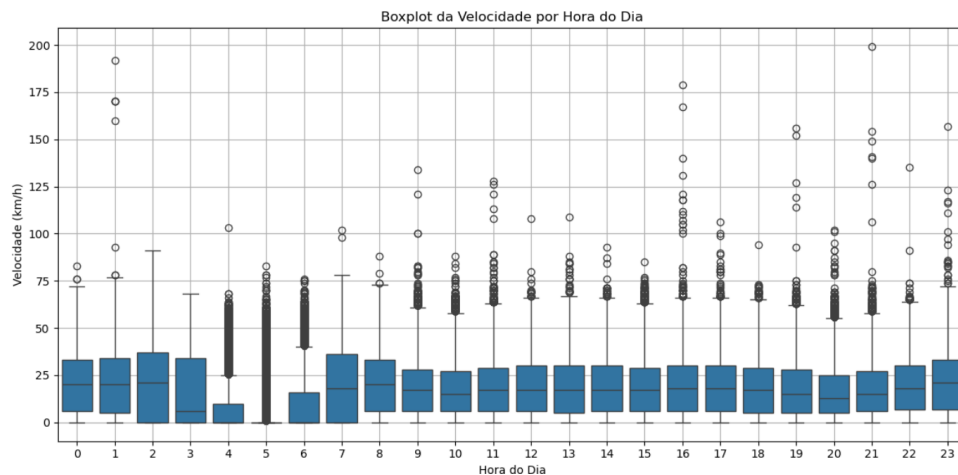


Figura 2 – *Boxplot* de velocidades por hora do dia

A análise de pontualidade mostrou que aproximadamente 42% das viagens iniciaram com atraso superior a 3 minutos em relação ao horário programado, caracterizando baixa aderência ao planejamento definido no GTFS. A avaliação da regularidade operacional evidenciou coeficientes de variação de *headways* acima de 35% em corredores troncais, o que indica significativa irregularidade nos intervalos entre veículos, sobretudo nos horários de pico. A espacialização dos resultados é apresentada na Figura 3, que mostra a diferença média entre *headways* reais e programados por shape de rota.

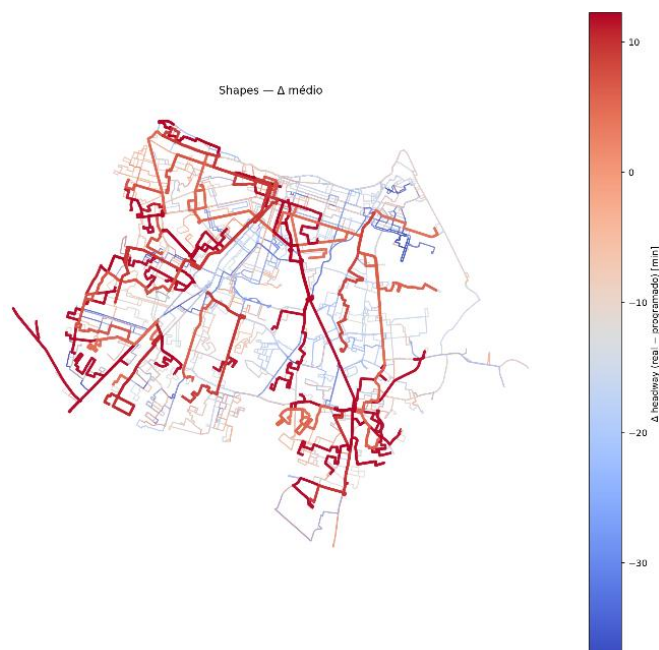


Figura 3 – Análise espacial da variação do *headway* por rota

Os trechos destacados em vermelho concentram as maiores discrepâncias, chegando a valores superiores a 10 minutos acima do programado, enquanto alguns trechos em azul apresentam *headways* menores do que o previsto, indicando operação concentrada em determinadas rotas. Complementarmente, a Figura 4 apresenta o mapa da velocidade média

estimada por ponto GPS. Nota-se um predomínio de valores entre 10 e 25 km/h na maior parte da rede, com alguns corredores alcançando médias superiores a 40 km/h. As áreas mais críticas, em vermelho, coincidem com zonas de maior adensamento urbano e tráfego viário, reforçando a influência dos congestionamentos na redução da confiabilidade operacional.

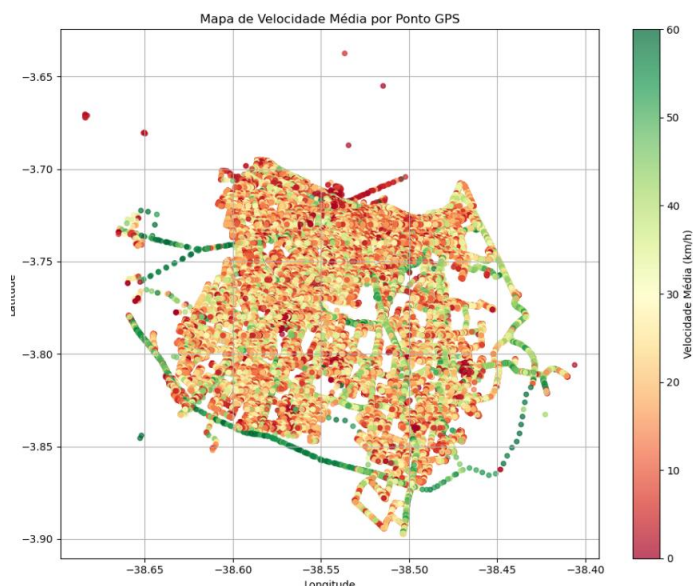


Figura 4 – Distribuição da velocidade média

Para avaliar a capacidade preditiva dos dados, foram implementados dois modelos supervisionados: Regressão Ridge e Random Forest, ambos tendo como variável alvo a confiabilidade do sistema (conforme expresso no capítulo de método). O desempenho obtido, resumido na Tabela 1, evidencia altos níveis de acurácia, com valores de R^2 superiores a 0,96.

Tabela 1 – Resumo dos indicadores da modelagem supervisionada

Modelo	RMSE	MAE	R^2
Ridge	0,4001	0,1440	0,9295
Random Forest	0,3440	0,0189	0,9774

O cruzamento dos indicadores operacionais com os registros de bilhetagem permitiu evidenciar que a demanda se concentra justamente nos pontos de maior irregularidade do sistema. Em paradas de terminais, por exemplo, observou-se média superior a 200 validações por hora, com *headway* teórico de 6 minutos, mas *headway* real chegando a 15 minutos entre veículos consecutivos. Esse descompasso gera tempos de espera excessivos, percebidos de forma ainda mais severa pelos usuários do que o atraso absoluto da viagem (Tabela 2).

Tabela 2 – Resumo dos indicadores de confiabilidade e demanda

Infraestrutura	Validações média por hora	Headway teórico	Headway real	Impacto percebido
Paradas em Terminais	> 220	6 min	15,4 min	Tempo de espera elevado
Paradas em Corredores Exclusivos	150	8,5 min	12,1 min	Maior frustração em picos
Paradas em Linhas de longa extensão (>10 km)	100	10,2 min	18 min	Variabilidade impactando acessibilidade de pessoas

CONSIDERAÇÕES FINAIS

O estudo evidenciou que a confiabilidade operacional é um dos principais desafios do transporte público urbano de Fortaleza, com impactos diretos sobre a percepção de qualidade do serviço e a atratividade do sistema. A integração entre dados de GPS, bilhetagem eletrônica e GTFS possibilitou uma análise abrangente, revelando que 42% das viagens iniciaram com atraso superior a três minutos, que os coeficientes de variação de *headways* superaram 35% em corredores troncais e que o tempo de viagem apresentou desvio-padrão médio de 9 minutos em linhas de maior extensão. Esses resultados reforçam que a irregularidade não ocorre de forma aleatória, mas concentra-se nos pontos e horários de maior demanda, ampliando os efeitos negativos sobre os usuários mais dependentes do transporte coletivo. A aplicação de modelos preditivos demonstrou o potencial da inteligência artificial em antecipar falhas operacionais, com destaque para o Random Forest, que alcançou R^2 de 0,9774. Conclui-se que a confiabilidade deve ser tratada como indicador central no planejamento do transporte público, orientando ajustes operacionais e subsidiando políticas públicas mais eficientes. Além disso, o trabalho contribui ao propor uma metodologia replicável em outras cidades brasileiras, fortalecendo a integração entre ciência de dados e gestão da mobilidade urbana.

REFERÊNCIAS

- BRAGA, C. K. V.; LOUREIRO, C. F. G.; PEREIRA, R. H. M. Analisando a variabilidade de estimativas de acessibilidade por transporte público a partir de dados de GPS. *Transportes*, São Paulo, v. 28, n. 5, p. 169–185, 2020. Disponível em: <https://www.transportes.org.br>. Acesso em: 26 nov. 2024.
- CATS, O.; JENELIUS, E. Reliability engineering and transportation: Analysing public transport performance with automated data. *Journal of Advanced Transportation*, v. 48, n. 6, p. 472–492, 2014. DOI: <https://doi.org/10.1002/atr.1257>.
- DEISENROTH, M. P.; FAISAL, A. A.; ONG, C. S. (2020). *Mathematics for Machine Learning*. Cambridge University Press.
- GÉRON, A. (2019) *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media. v 1. 856 p.
- JAMES, G., WITTEN, D., HASTIE, T., & TIBSHIRANI, R. (2013). *An introduction to statistical learning: with applications in R* (Vol. 103). New York: springer.
- PEREIRA, R. H.; BANISTER, D.; SCHWANEN, T.; WESSEL, N. (2019). Distributional effects of transport policies on inequalities in access to opportunities in Rio de Janeiro. *Journal of Transport and Land Use*, 12(1), 741-764.
- VAN OORT, N. *Service Reliability and Urban Public Transport Design*. PhD Thesis. Delft University of Technology, 2011.