

# Análise de Dados de E-Commerce para previsão de entregas de pedidos

Lucas Auada Braga<sup>1</sup>, Vitor Kenzo Koga Onoue<sup>1</sup>

<sup>1</sup>Faculdade de Computação e Informática - Universidade Presbiteriana Mackenzie (UPM) - São Paulo – SP – Brasil

10403286@mackenzista.com.br, 10402362@mackenzista.com.br

**Abstract.** *This work proposes the development of a machine learning model to predict delivery delays in e-commerce orders. Using historical order data, the model can estimate whether a delivery will be on time or delayed. The proposal aims to help retailers anticipate logistical problems and improve customer experience, directly impacting on the company's reputation.*

**Resumo.** *Este trabalho propõe o desenvolvimento de um modelo de aprendizado de máquina para prever atrasos em entregas de pedidos de e-commerce. Utilizando dados históricos de pedidos, o modelo é capaz de estimar se uma entrega será realizada no prazo ou sofrerá atrasos. A proposta visa auxiliar lojistas a anteciparem problemas logísticos e melhorarem a experiência do cliente, com impacto direto na reputação da empresa.*

## 1. Introdução

A entrega pontual de produtos é um aspecto essencial para a satisfação de consumidores em plataformas de e-commerce. Com o aumento das compras online, prever possíveis falhas logísticas se tornou estratégico para empresas que desejam manter a competitividade no mercado.

A predição de atrasos pode ajudar empresas a tomarem decisões proativas, como ajustar o método de envio ou informar o cliente com antecedência. Isso reduz impactos negativos na reputação e melhora o relacionamento com o consumidor.

O objetivo deste projeto é desenvolver um modelo preditivo, utilizando técnicas de aprendizado de máquina, capaz de prever se um pedido será entregue dentro do prazo ou sofrerá atrasos com base em dados históricos.

O projeto selecionado pertence à área de Inteligência Artificial e tem como foco o uso de modelos supervisionados para resolver um problema de predição no contexto de logística em e-commerce.

## 2. Descrição do problema

O desafio consiste em prever, no momento da realização de um pedido, se a entrega ocorrerá no prazo ou não. Essa previsão deve considerar variáveis como localização do cliente, tipo de produto, prazo estimado de entrega, entre outros. O modelo deve aprender com padrões anteriores para oferecer previsões precisas.

### **3. Aspectos éticos e responsabilidade**

A utilização de Inteligência Artificial exige responsabilidade em relação à privacidade dos dados dos clientes e à transparência do modelo preditivo. É essencial garantir que os dados utilizados estejam anonimizados. A explicabilidade do modelo também é importante, principalmente para justificar decisões automatizadas.

### **4. Descrição do conjunto de dados**

Foi escolhido um conjunto de dados de E-commerce no Kaggle, que é uma plataforma de ciência de dados, onde sua comunidade disponibiliza datasets. O dataset escolhido é composto por pedidos feitos na Olist Store. Os dados já estão anonimizados.

O dataset é composto por um conjunto de arquivos .csv, que se conectam através de colunas de identificação única. Apenas alguns campos e alguns arquivos serão usados. Serão usados os campos: status do pedido, data de compra, data de entrega estimada, data de entrega (dia que chegou para o cliente de fato), estado do cliente (Minas Gerais, por exemplo), valor do pedido e valor do frete.

A preparação dos dados e a análise exploratória dos dados está contida no repositório do projeto no Github, que pode ser acessado através do link: [https://github.com/VitorOnoue/ai\\_college](https://github.com/VitorOnoue/ai_college).

### **5. Metodologia e resultados esperados**

A abordagem deste projeto consiste no uso de técnicas de aprendizado de máquina para resolver um problema de classificação: prever se um pedido será entregue no prazo ou com atraso, com base em informações do pedido, do cliente, do valor do frete e da compra. As etapas envolvem a coleta/seleção de dados, limpeza e preparação de dados, treino da modelagem e avaliação do modelo.

Na etapa de coleta/seleção dos dados, foi utilizado o conjunto de arquivos disponibilizados no Kaggle. Os diversos .csv's foram combinados em um único, contendo apenas campos que vamos usar no modelo.

Na etapa de limpeza e preparação, foram tratados os valores ausentes e a criação de variáveis novas, como por exemplo, a quantidade de dias de atraso (que foi criada a partir da diferença entre a data estimada e a data real de entrega).

Na etapa de treino da modelagem, será utilizado algoritmos de classificação como a Regressão Logística utilizando o framework scikit-learn do Python. Após o treino, chegamos na etapa de avaliação, onde iremos aplicar métricas de acurácia, precisão e F1-score, por exemplo.

Esperamos obter um modelo capaz de prever se um pedido será entregue a tempo para o cliente, ou se haverá algum atraso, além de identificar quais são variáveis que possuem mais impacto nessa situação.

## 6. Referências

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly.

Kaggle - Olist E-commerce Dataset. Disponível em: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>

Chawla, N. V., et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*.