

Nome: Guilherme Mendonça Pinheiro

RA: 221024281

Relatório Seminário Trabalho Final

1. INTRODUÇÃO E OBJETIVO

No cenário atual, com o crescimento do e-commerce e dos serviços logísticos como Mercado Livre, Amazon e entre outros, prever o tempo de entrega de pacotes tornou-se um diferencial competitivo importante. Estimar com precisão o prazo de entrega permite o aprimoramento de processos, redução de atrasos e aumento da satisfação do consumidor final.

Como tal, neste trabalho temos como objetivo utilizar Regressão Linear Simples para prever o tempo estimado de entrega de pacotes, com base em atributos como o peso do produto e descontos aplicados. A análise foi realizada com base em um conjunto de dados extraídos da plataforma Kaggle, contendo mais de 10 mil registros de entregas reais.

2. DESCRIÇÃO DO CONJUNTO DE DADOS

O dataset contém 10.999 registros (linhas) e 12 colunas (atributos), com informações relacionadas a remessas de pacotes. As principais variáveis incluem:

- **Warehouse_block**: Bloco do armazém (categorias A-F)
- **Mode_of_Shipment**: Modal de envio (Flight, Road, Ship)
- **Customer_care_calls**: Nº de contatos ao SAC
- **Customer_rating**: Avaliação do cliente (1 a 5)
- **Cost_of_the_Product**: Custo do produto
- **Prior_purchases**: Compras anteriores
- **Product_importance**: Importância do item (low, medium, high)
- **Gender**: Gênero do cliente (M/F)
- **Discount_offered**: Desconto aplicado

- `Weight_in_gms`: Peso em gramas
- `Reached.on.Time_Y.N`: Variável binária indicando se chegou no prazo
- `Delivery_time_estimated`: Variável contínua criada artificialmente para representar o tempo de entrega em dias

3. LIMPEZA E TRATAMENTO DE DADOS

- Verificação de valores nulos ou inconsistentes
- Conversão e normalização de variáveis categóricas, quando necessário
- Criação da variável `Delivery_time_estimated`:
 - Fórmula: $\text{tempo} = (\text{peso em kg} \times 1.5) - (\text{desconto\%} \times 2) + \text{ruído aleatório}$
 - Resultados arredondados e limitados para evitar valores negativos

4. ANÁLISE EXPLORATÓRIA DOS DADOS

Foram realizadas estatísticas descritivas e visualizações, destacando:

- Distribuição do tempo estimado de entrega
- Relação entre peso e tempo de entrega
- Influência do desconto no tempo final

5. APLICAÇÃO DA TÉCNICA PREDITIVA: REGRESSÃO LINEAR SIMPLES

A regressão foi aplicada para prever `Delivery_time_estimated` usando como variável explicativa principal o `Weight_in_gms`. O modelo obteve:

- Coeficiente angular (β_1) indicando a variação no tempo a cada grama adicional
- R^2 (coeficiente de determinação) indicando o quanto da variação no tempo é explicada pelo peso

6. DISCUSSÃO DOS RESULTADOS

O modelo obteve um R^2 moderado, sugerindo que o peso do pacote tem impacto direto no tempo de entrega, mas não é o único fator. Isso é coerente com expectativas logísticas, já que outros elementos como distância, trânsito e disponibilidade de transporte também influenciam o prazo (Não temos essas outras variáveis disponíveis no conjunto de dados analisado).

A simplicidade do modelo facilita sua compreensão e aplicação em cenários reais com dados limitados, sendo uma base inicial útil para o desenvolvimento de modelos mais robustos no futuro.

7. CONSIDERAÇÕES FINAIS

Neste trabalho, usamos a Regressão Linear Simples para prever o tempo estimado de entrega de pacotes, considerando apenas o peso como fator. O modelo mostrou que, de forma geral, pacotes mais pesados tendem a demorar mais para serem entregues, o que faz sentido no mundo real.

Mesmo sendo um modelo simples, com apenas uma variável, ele conseguiu fazer boas previsões e pode ser útil em situações onde temos poucos dados disponíveis. No entanto, sabemos que outros fatores também influenciam o tempo de entrega, como a distância, o trânsito ou o tipo de transporte, mas essas informações não estavam no nosso conjunto de dados.

Para melhorar esse tipo de previsão no futuro, seria interessante testar modelos mais avançados e incluir mais variáveis, que representem melhor toda a situação logística.