

Relatório Analítico

Construção de um Modelo com Regressão Linear

Guilherme Alves de Oliveira e Oliveira
João Pedro Lima Paulo
Lara Marina de Oliveira
Giulia Nogueira Lopes De Sá
Beatriz de Castilho Ferreira

1. Introdução

Objetivo: Desenvolver um modelo de Regressão Linear para prever preço de corridas com base na distância, utilizando método dos mínimos quadrados.

2. Metodologia

Dados utilizados:

- Variável independente (X): Distância (num_distancia_m_1), tempo (num_tempo_estimado_seg_1).
- Variável dependente (y): Preço (num_preco)

Modelo Estatístico:

- Regressão Linear simples:
 - Equação:
$$\text{Preço} = \theta^1 + \theta^2 * \text{Distância (m)}$$
- Método: Mínimos quadrados (implementação manual e via SKLearn)

Ferramentas:

- Bibliotecas: numpy, pandas, matplotlib, scikit-learn

3. Resultados

Coeficientes do modelo:

Parâmetro	Valor	Interceptação
Intercepto θ^1	4.05	Custo inicial
Distância θ^2	0.00174	Acréscimo de ~R\$1.74 por KM

Métricas de desempenho:

- R^2 (coeficiente de determinação) : 0,78
 - 78% da variabilidade do preço é explicada pela distância
- Erro Quadrático Médio (MSE) de 243.64:
 - correspondendo a um erro médio de aproximadamente R\$15.61 por previsão
- Equação final: $\text{Preço} = 4.05 + 0.00174 * \text{Distância (em metros)}$
- Intercepto:
 - 4.05 (equivalente a um custo base por corrida)
- Coeficiente angular:
 - 0.00174 (indicando acréscimo de ~R\$1.74 por km)

Limitações e Melhorias Potenciais

- Fatores Não Considerados:
 - O modelo atual não incorpora variáveis como demanda momentânea, tráfego, tipo de veículo ou tarifas dinâmicas
- Oportunidades de Aprimoramento:
 - Inclusão de mais variáveis preditoras (horário do dia, condições climáticas)
 - Coleta de maior volume de dados para reduzir variabilidade

O gráfico de dispersão mostra uma clara tendência linear positiva entre distância e preço das corridas, confirmando a relação direta modelada pela regressão ($R^2=0.75$). A concentração de dados abaixo de 40km sugere que a maioria das corridas analisadas são de curta/média distância.

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Grafico 4 - ScatterPlot - Preço pela Distancia):
plt.figure(figsize=(10,6))
sns.scatterplot(x='num_distancia_m_1', y='num_preco', data=df)
plt.title("Preço VS Distancia")
plt.show()
```

