

Resultados de IA e ML no Projeto

1. Introdução

Neste projeto, utilizamos algoritmos de aprendizado de máquina para prever o preço de corridas com base em dados históricos. Ao analisar informações como distância, horário, tipo de dia e condições estimadas de tráfego, conseguimos estimar o custo de uma corrida com muito mais precisão, adaptando os valores de acordo com o serviço selecionado.

2. Dados Utilizados

Trabalhamos com uma base de dados chamada 'base_corridas.csv', que reúne informações históricas de corridas, incluindo:

- ano, mês, hora: informações temporais da corrida;
- lat_origem, lng_origem, lat_destino, lng_destino: localização geográfica da origem e destino;
- trafego_estimado: categoria de tráfego no momento da corrida;
- tipo_dia: se a corrida foi em dia útil ou fim de semana;
- productid: tipo de serviço (ex: UberX, UberBlack etc.);
- price: valor da corrida.

Para garantir uma análise precisa, calculamos a distância entre a origem e o destino usando a fórmula geodésica (geopy).

3. Algoritmos Utilizados

Optamos pelo algoritmo Random Forest Regressor, uma técnica que combina diversas árvores de decisão para obter previsões mais precisas e reduzir o risco de overfitting. Esse modelo é especialmente eficaz em problemas que envolvem variáveis contínuas e categóricas.

4. Pré-processamento dos Dados

Para garantir que os dados estivessem prontos para o modelo, realizamos as seguintes etapas:

- Cálculo da distância entre origem e destino (geopy.geodesic);
- Separação dos dados por tipo de serviço (productid);
- One-Hot Encoding para variáveis categóricas (trafego_estimado e tipo_dia);
- Separação dos dados em treino e teste (80/20);
- Padronização das colunas de entrada;
- Criação de um modelo específico para cada tipo de serviço.

5. Avaliação dos Modelos

Os modelos foram avaliados com base em três métricas principais: MAE (erro absoluto médio), RMSE (erro quadrático médio) e R^2 (coeficiente de determinação).

Abaixo, apresentamos os resultados obtidos:

Serviço	MAE	RMSE	R^2
Flash	3.50	6.31	0.88
99pop	3.02	6.33	0.89
UberX	3.23	5.82	0.89
Black	4.51	6.98	0.94
Comfort	3.00	4.83	0.98
Taxi	1.89	3.38	0.96

6. Conclusões

Os modelos de IA foram capazes de prever o preço das corridas com boa precisão. O modelo para o serviço Comfort se destacou, com um R^2 de 0.98, indicando uma excelente capacidade preditiva. No geral, os resultados mostram que os modelos são eficazes para prever valores com base em características como distância, horário e tipo de dia.