

Global Solution - Inteligência Artificial e Computacional

Gabriel Gonçalves - 93069
Daniel Okudeira Carapeto - 93180

Novembro de 2024

Resumo

Este relatório descreve o desenvolvimento de um sistema de Machine Learning para prever valores de reembolso associados a estações de carregamento de veículos elétricos. Foram explorados vários modelos de aprendizado supervisionado, como regressão linear, árvores de decisão e algoritmos baseados em ensemble (XGBoost, LightGBM), para identificar a solução mais eficaz. O modelo XGBoost apresentou o melhor desempenho, com um coeficiente de determinação (R^2) de 0,883 e um erro médio quadrático (MSE) de 5.820.779. O sistema pode ser aplicado para otimizar o planejamento de incentivos e subsídios.

1 Introdução

Com o aumento da adoção de veículos elétricos, o incentivo à instalação de estações de carregamento desempenha um papel fundamental na transição energética. Este projeto visa desenvolver um sistema preditivo para auxiliar na estimativa de valores de reembolso com base em características das estações, como custo, localização e marca.

O sistema utiliza um conjunto de dados real contendo informações sobre 717 estações de carregamento e aplica técnicas de aprendizado de máquina para prever o valor do reembolso. Isso possibilita otimizar a alocação de recursos e reduzir o tempo necessário para análises manuais.

2 Metodologia

2.1 Descrição do Dataset

O conjunto de dados utilizado possui 717 registros e as seguintes colunas principais:

- **Cost of Charging Station:** Custo da estação de carregamento.
- **Rebate Amount:** Valor do reembolso.
- **City, County:** Localização da estação.
- **Charging Station Brand:** Marca da estação.
- **Month, Year:** Data de instalação.

2.2 Pré-processamento

O pré-processamento envolveu:

- **Normalização:** As variáveis numéricas foram normalizadas usando o `StandardScaler`.
- **Codificação:** As variáveis categóricas (*City*, *County*, *Charging Station Brand*) foram codificadas com *OneHotEncoder*.
- **Divisão de Dados:** O dataset foi dividido em 80% para treino e 20% para teste.

2.3 Modelos Testados

Os seguintes modelos de aprendizado supervisionado foram testados:

1. **Regressão Linear:** Modelo base para comparação.
2. **Árvore de Decisão:** Para capturar relações não lineares.
3. **Random Forest:** Conjunto de árvores para maior robustez.
4. **XGBoost:** Algoritmo de *Gradient Boosting* eficiente.
5. **LightGBM:** Variante leve do *Gradient Boosting*.
6. **Rede Neural:** Para capturar padrões complexos nos dados.

3 Resultados

Os resultados foram avaliados com base nas métricas de **Erro Médio Quadrático (MSE)** e **Coefficiente de Determinação (R^2)**. A tabela a seguir resume os resultados obtidos:

Modelo	MSE	R^2
XGBoost	5.820.779	0.883
LightGBM	9.446.493	0.811
Random Forest	10.510.300	0.789
Regressão Linear	12.109.290	0.757
Rede Neural	12.612.910	0.747
Árvore de Decisão	16.663.950	0.666

Tabela 1: Resultados dos Modelos Testados

O modelo **XGBoost** apresentou o melhor desempenho, explicando 88,3% da variância dos dados e com o menor erro médio quadrático.

4 Discussão

Os resultados destacam que os modelos baseados em *Gradient Boosting* (XGBoost e LightGBM) são mais adequados para este problema, graças à sua capacidade de capturar relações não lineares e interações complexas entre as variáveis. Modelos mais simples, como Regressão Linear e Árvore de Decisão, apresentaram desempenho inferior devido à sua limitação em lidar com a complexidade dos dados.

5 Conclusão

O sistema desenvolvido demonstrou eficácia na previsão de valores de reembolso, com destaque para o modelo XGBoost. Essa abordagem pode ser aplicada por órgãos públicos e empresas para otimizar processos de reembolso e planejamento financeiro.

5.1 Próximos Passos

- Coletar mais dados para aprimorar as previsões.
- Ajustar os hiperparâmetros dos modelos para melhorar ainda mais o desempenho.
- Incorporar variáveis adicionais, como dados climáticos ou regionais.

Links Importantes

- **Vídeo explicativo:** <https://youtu.be/4JXDqonKUgk>
- **Repositório do projeto:** <https://github.com/Goncalvs98/gs2semia.git>

Referências

- Documentação oficial do XGBoost: <https://xgboost.readthedocs.io/>
- Documentação do LightGBM: <https://lightgbm.readthedocs.io/>
- Artigos sobre Machine Learning em Python: <https://scikit-learn.org/>