

GLOBAL SOLUTION


2TDPSC – 2024



FIAP **GLOBAL SOLUTION 2024**



2TDS – Turmas de Agosto



FIAP

Projeto: Modelo de Classificação para Dados de Consumo Energético

Descrição Geral: Este projeto utiliza técnicas de Machine Learning para criar um modelo de classificação baseado em dados históricos de consumo energético. O objetivo principal é categorizar os registros de consumo diário em três classes: Baixo, Médio e Alto. O modelo foi desenvolvido para explorar padrões de consumo energético e fornecer insights que podem ser aplicados em sistemas de monitoramento e otimização de energia.

Etapas do Desenvolvimento:

1. Coleta e Preparação dos Dados:
 - Um conjunto de dados foi gerado com informações de consumo energético por hora, dia e mês. Os dados foram organizados em um arquivo CSV com 120 registros, abrangendo um período de 5 dias.
 - Os registros foram enriquecidos com uma nova coluna de classificação baseada em limites predefinidos de consumo diário.
2. Pré-processamento dos Dados:
 - Transformação e normalização das variáveis de entrada para garantir escalas consistentes.
 - Divisão dos dados em conjuntos de treino (70%) e teste (30%) para avaliar a performance do modelo.
 - Codificação da variável de saída em categorias numéricas para facilitar o treinamento do modelo.
3. Treinamento dos Modelos:
 - Dois algoritmos de Machine Learning foram aplicados:
 - Random Forest: Modelo baseado em múltiplas árvores de decisão para capturar relações não lineares.
 - K-Nearest Neighbors (KNN): Algoritmo que classifica com base na proximidade entre os pontos de dados.
 - Os modelos foram treinados usando as variáveis de entrada: consumo por hora e consumo mensal.
4. Avaliação dos Modelos:
 - As performances dos modelos foram avaliadas utilizando métricas como acurácia, matriz de confusão e relatório de classificação (precision, recall e f1-score).

- Ambos os modelos apresentaram uma acurácia inicial de aproximadamente 30.56%, indicando desafios em capturar padrões claros devido à limitada quantidade de dados e simplicidade das features.

5. Visualizações:

- Uma matriz de confusão foi gerada para cada modelo, destacando os acertos e erros nas classificações.
- Foi criado um gráfico de importância das variáveis para o modelo Random Forest, evidenciando as contribuições das variáveis de entrada no processo de decisão.

Conclusões: Embora o desempenho inicial dos modelos não seja ideal, o projeto demonstra um pipeline completo para classificação de dados de consumo energético. As áreas de melhoria incluem:

- Refinamento dos thresholds de classificação.
- Aumento da base de dados para maior generalização.
- Inclusão de novas variáveis (como condições climáticas ou tipo de dispositivo).
- Otimização de hiperparâmetros para melhorar a performance dos algoritmos.

Tecnologias e Ferramentas Utilizadas:

- Python: Linguagem principal para manipulação de dados e desenvolvimento do modelo.
- Pandas e NumPy: Para manipulação e análise de dados.
- Scikit-learn: Biblioteca para construção e avaliação dos modelos de Machine Learning.
- Matplotlib e Seaborn: Para visualização gráfica dos resultados.

Aplicações Finais: Este modelo pode ser integrado a sistemas de monitoramento energético para fornecer insights sobre o comportamento de consumo, permitindo otimizações, recomendações e ações direcionadas para economizar energia.

EQUIPE

GABRIEL TORRES FERNANDES – RM553635