

# DSA: Projeto 08

## Modelagem Preditiva em IoT – Previsão de Uso de Energia

---

**Autor:** Rodrigo de Lima Oliveira

Linkedin: <https://www.linkedin.com/in/rodrigolima82/>



# Visão Geral

## Desafio: Consumo de Energia

- Dados: através de IoT (medições através de sensores de temperatura e umidade)
- Objetivo: previsão de consumo de energia de eletrodomésticos
- Resumo dos Dados
- Construção de Variáveis
- Seleção de Atributos
- Metodologia e Resultados do Modelo
- Considerações Finais

# Resumo dos Dados



Total de registros: 19.735  
(dados de treino + teste)



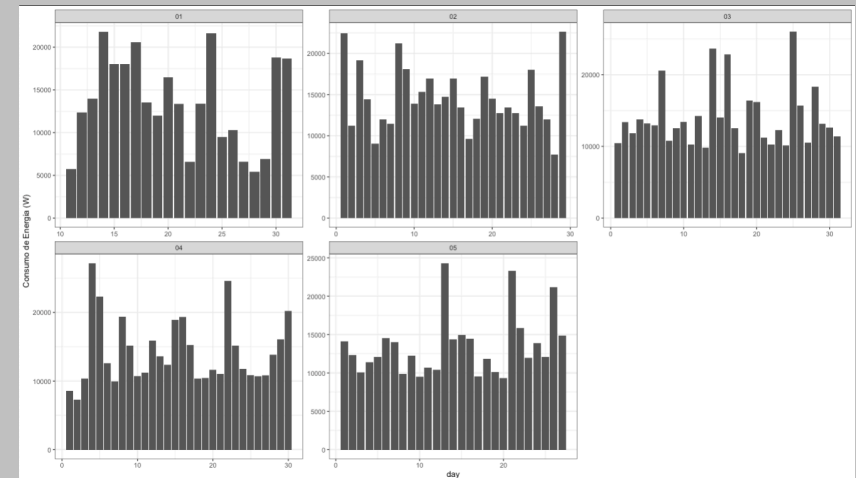
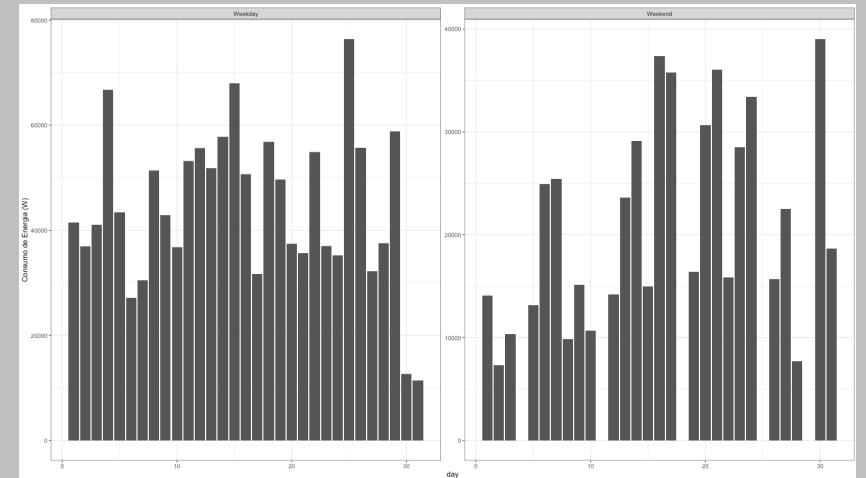
32 atributos



Atributo “Appliances” é o target  
(variável a ser prevista)

# Detalhe dos Dados

- O conjunto de gráficos superior apresenta o consumo de energia em (W) em dias da semana e em finais de semana (o primeiro gráfico de dias da semana e o segundo finais de semana)
- O conjunto de gráficos inferior apresenta o consume de energia de eletrodomésticos distribuídos nos meses de Janeiro à Maio e os dias de cada mês.
- Pelos gráficos é possível observar um pico de consume no mês de janeiro e fevereiro



# Construção de Variáveis

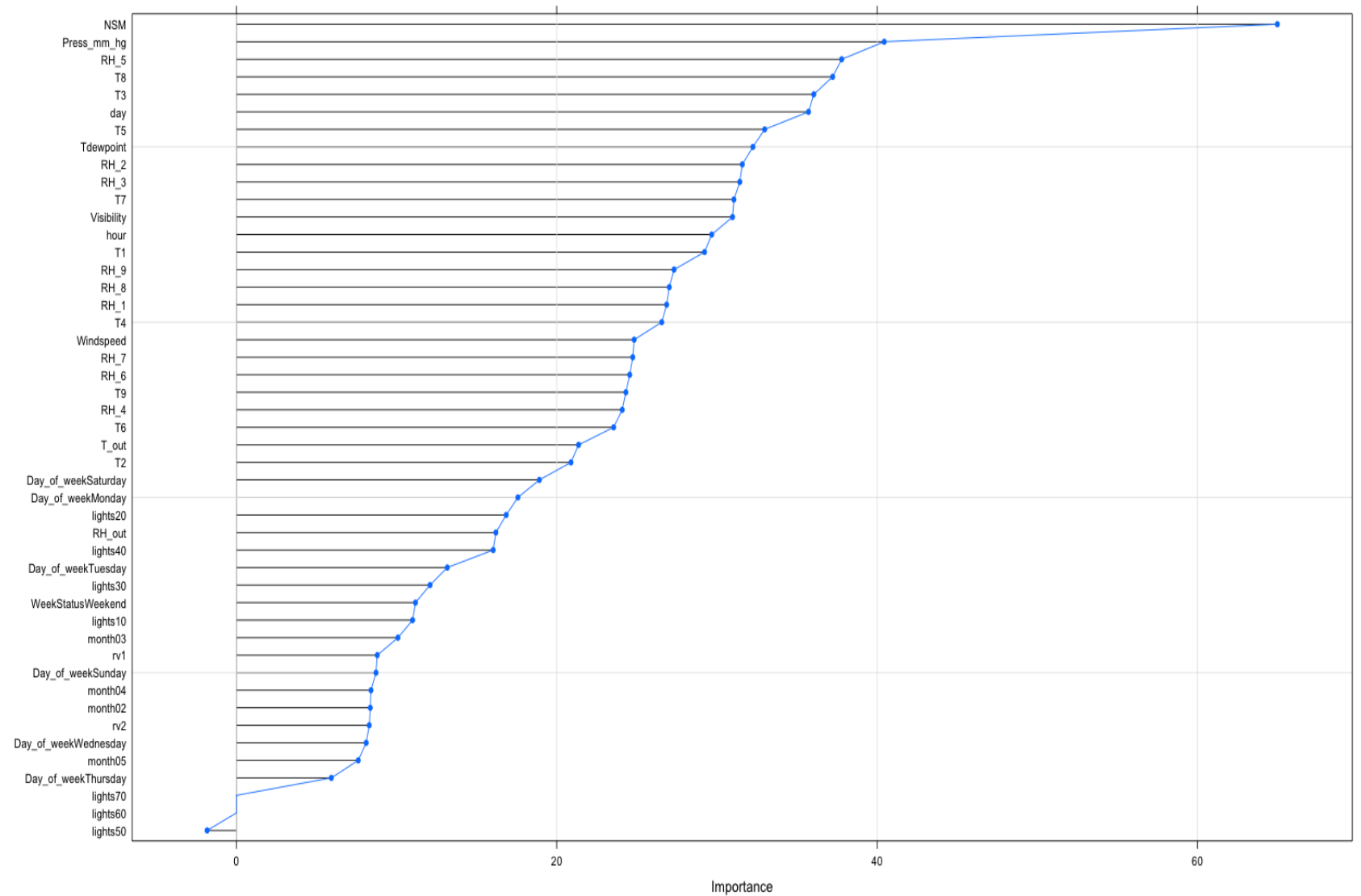
**MONTH:** COLUNA EXTRAÍDA  
DA DATA PARA INDICAR O MÊS  
DO CONSUMO DE ENERGIA

**DAY:** COLUNA EXTRAÍDA DA  
DATA PARA INDICAR O DIA DO  
CONSUMO DE ENERGIA

**HOUR:** COLUNA EXTRAÍDA DA  
DATA PARA INDICAR A HORA  
DO CONSUMO DE ENERGIA

# Seleção de Atributos

- Colunas selecionadas com base na importância
- Aplicado técnica de Feature Selection usando modelo Random Forest
- Para a construção do modelo usei as features com índice de importância acima de 20 (com exceção da RH\_out)



# Métrica

## Métricas para avaliação do modelo

### Root Mean Squared Error (RMSE) e The Coefficient of Determination (R Squared)

- RMSE e  $R^2$  são excelentes métricas para modelos de regressão, além de serem muito fácil de interpretar.
- A **Raiz Quadrada do Erro Quadrático Médio (RMSE)** — nada mais é que a diferença entre o valor que foi previsto pelo modelo e o valor real que foi observado
- Já o **Coeficiente de Determinação ( $R^2$ )** — varia entre 0 e 1 e indica, em percentagem, o quanto o modelo consegue explicar os valores observados. Quanto maior o  $R^2$ , mas explicativo é o modelo, melhor ele se ajusta à amostra.

# Modelos

## **Modelos de Machine Learning avaliados:**

- Multiple Logistic Regression (GLM)
- Generalized Boosted Regression Modeling (GBM)
- eXtreme Gradient Boosting (XGBoost)
- eXtreme Gradient Boosting Otimizado



# Resultado dos modelos

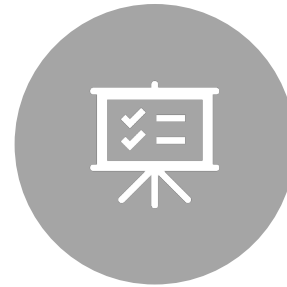
---

Métrica R squared	Treino	Teste
Multiple Logistic Regression (GLM)	14%	-
Generalized Boosted Regression Modeling (GBM)	28%	-
eXtreme Gradient Boosting (XGBoost)	46%	-
eXtreme Gradient Boosting Otimizado	58%	61%

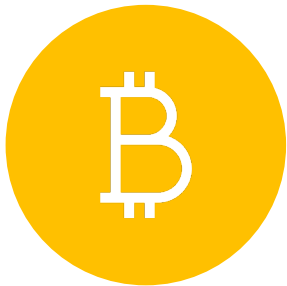
# Considerações Finais



O melhor algoritmo para esse dataset foi o XGBoost



O modelo otimizado foi capaz de explicar 61% da variância nos dados de teste



Realizando a remoção de outliers no dataset, não houve melhora na performance do modelo



O ideal agora seria obter mais dados para aumentar a performance do modelo