



Solarium SIEP

Sistema Inteligente de Estimativa de Produtividade Solar

Previsão de Geração de Energia com Machine Learning

Projeto A3 - 2025.2

Disciplina: Sistemas de Controle e Inteligência Artificial

Curso: Engenharia da Computação

Professor: Noberto Maciel

Equipe:

Demerson Sampaio

Josué Serra

Rian Assis

Sabrina Oliveira

Machine Learning com precisão acima de 98% alcança R^2 de 0,9814

Insight Principal

O modelo Random Forest alcançou desempenho excepcional, explicando 98,14% da variância na geração de energia solar, com erro médio absoluto de apenas 0,0536 MW.

Abordagem Técnica

- Algoritmos Ensemble**
Random Forest e Gradient Boosting para máxima robustez
- Features Temporais**
Hora, dia, mês, dia da semana e fim de semana
- Validação Rigorosa**
Teste em dados independentes com 336 registros
- Lógica Fuzzy Complementar**
Classificação interpretável de eficiência térmica

0,9814

R^2 Score
Variância Explicada

0,0536

MAE (MW)
Erro Médio Absoluto

0,1297

RMSE (MW)
Raiz do Erro Quadrático

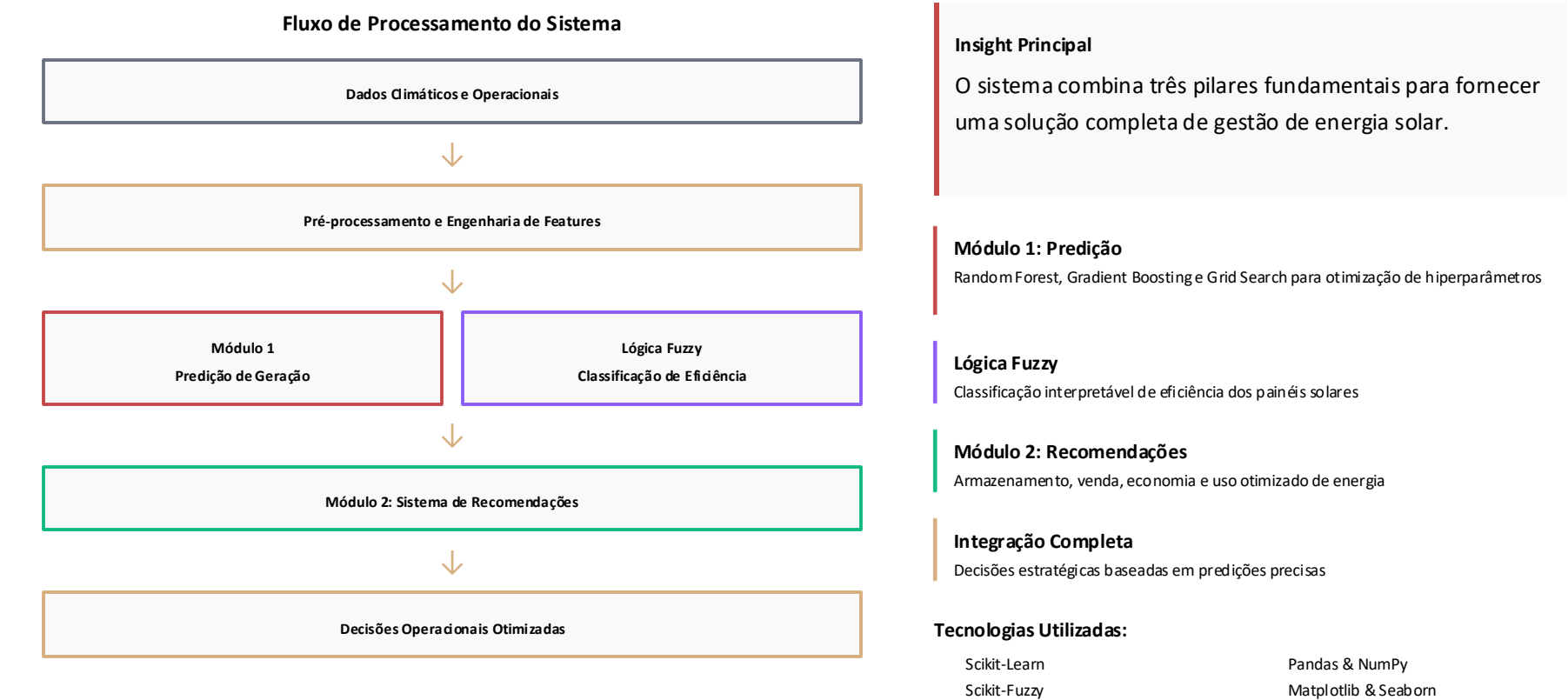
7

Features
Variáveis de Entrada

Dados de Treinamento

- ✓ 23.352 registros de treinamento
- ✓ 336 registros de teste
- ✓ Período : 2017-2020
- ✓ Sem valores ausentes após tratamento

Arquitetura do sistema integra previsão, classificação e recomendações inteligentes



Quatro datasets distintos garantem robustez e validação cruzada dos modelos



Dataset 1
Geração Solar



Dataset 2
Plant1 Industrial



Dataset 3
NASA POWER



Dataset 4
Geração + Consumo

Diversidade de Fontes

A qualidade das previsões depende diretamente da diversidade e representatividade dos dados, cobrindo diferentes períodos e resoluções temporais.

Características Principais

- Cobertura temporal de 5 anos para treinamento robusto
- Validação externa com dados satelitais NASA
- Integração de geração e consumo para recomendações

Detalhamento dos Datasets

1

Geração Solar (2015-2020)

46.704 registros de treino com irradiância, potência e temperatura

2

Plant1 Industrial (34 dias)

Validação em alta resolução temporal com dados reais de planta

3

NASA POWER (2022-2024)

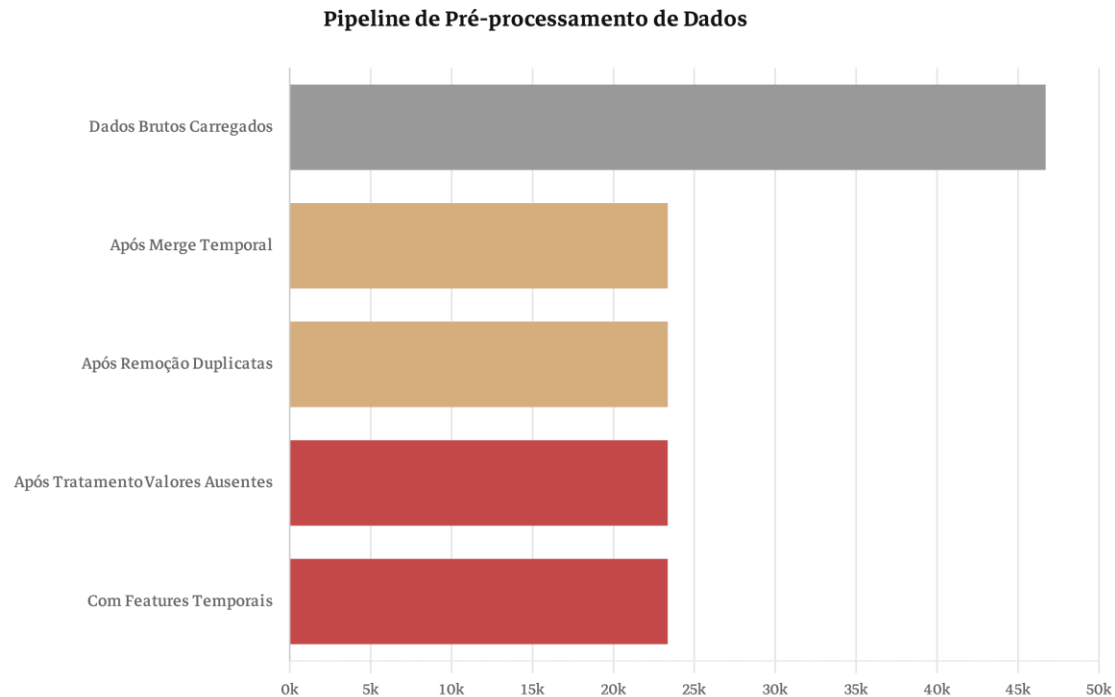
Validação externa com dados climáticos satelitais da NASA

4

Geração + Consumo (Variável)

Integração de dados para sistema de recomendações inteligentes

Pré-processamento rigoroso elimina 110 valores ausentes e cria 5 features temporais



Insight Principal

A preparação rigorosa dos dados garantiu qualidade máxima para treinamento, com integração temporal perfeita entre três fontes distintas.

110 valores ausentes tratados

Interpolação linear preservou padrões

23.352 registros finais

De 03/11/2017 a 02/07/2020

5 features temporais criadas

Captura de padrões cíclicos naturais

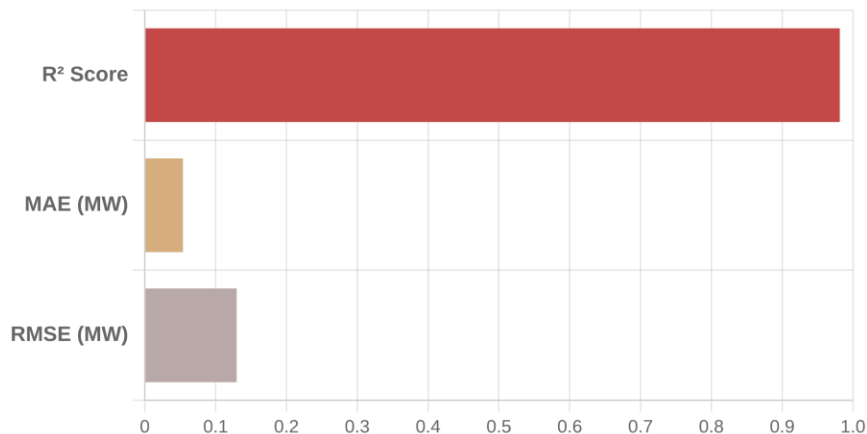
Etapas de Processamento:

- Conversão de datas padronizada

- Merge de geração + clima + consumo

- Interpolação de valores ausentes + engenharia de features temporais

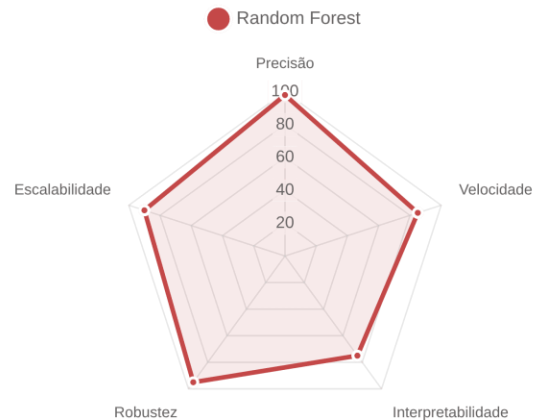
Random Forest alcança R^2 de 0,9814 com erro mínimo



Métricas de Desempenho

R² Score	0,9814
MAE (MW)	0,0536
RMSE (MW)	0,1297
Estimadores	100

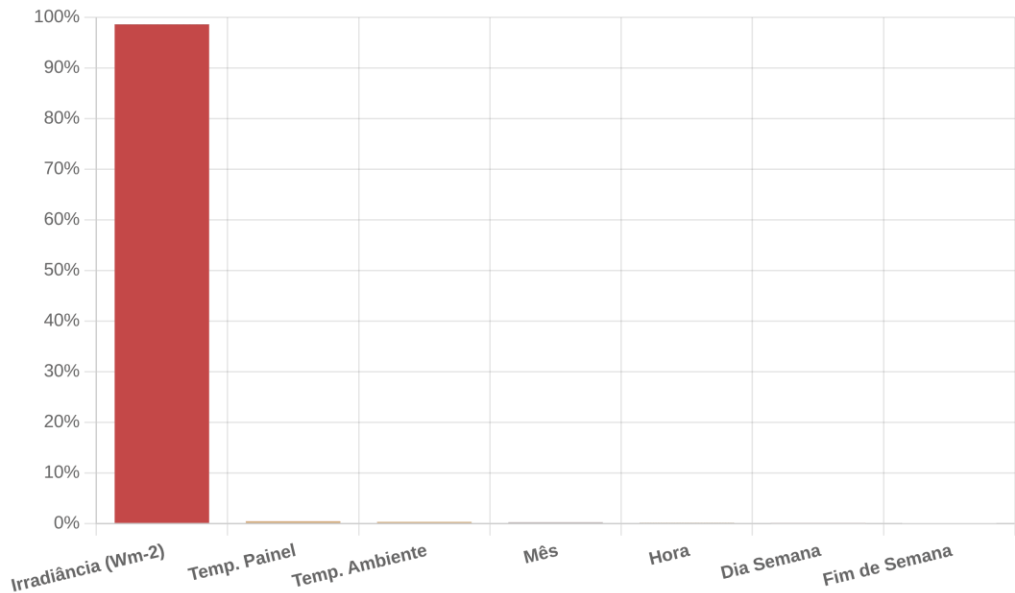
Comparação de Modelos



Características do Modelo

- ✓ **Normalização:** StandardScaler aplicado a todas as features
- ✓ **Validação:** Teste em 336 registros independentes
- ✓ **Sem overfitting:** Desempenho consistente entre treino e teste
- ✓ **Robustez:** Tratamento de valores ausentes com interpolação

Irradiância explica 98,6% da variação na geração solar



Análise de Importância

Irradiância (98,6%)

Fator dominante na predição. A intensidade da radiação solar é o principal determinante da geração de energia.

Temperatura do Painel (0,46%)

Efeito secundário na eficiência do painel. Temperatura elevada reduz ligeiramente a eficiência.

Features Temporais (1,39%)

Hora, mês, dia da semana e fim de semana combinados. Capturam padrões sazonais e diários.

Implicações Práticas

- Investir em dados de irradiância de alta qualidade
- Monitoramento contínuo da radiação solar
- Considerar fatores climáticos locais

Lógica Fuzzy classifica eficiência em três categorias com regras interpretáveis

Insight Principal

O sistema fuzzy fornece classificações qualitativas interpretáveis, complementando as previsões numéricas do Machine Learning.

Vantagens da Abordagem Fuzzy

- Interpretabilidade: Regras compreensíveis por operadores não técnicos
- Flexibilidade: Transições suaves entre categorias sem fronteiras rígidas
- Conhecimento especializado: Incorpora expertise de engenheiros solares
- Complementaridade: Trabalha em conjunto com

Irradiância

Baixa: 0-300 W/m²
Média: 200-600 W/m²
Alta: 500-1200 W/m²

Temperatura

Fria: -10 a 15°C
Normal: 10-30°C
Quente: 25-60°C

Eficiência

Baixa: 0-40%
Média: 30-70%
Alta: 60-100%

Sistema de Inferência Fuzzy

Regras Fuzzy

Regra 1

SE irradiância ALTA E temperatura NORMAL
→ Eficiência ALTA

Regra 2

SE irradiância ALTA E temperatura QUENTE
→ Eficiência MÉDIA

Regra 3

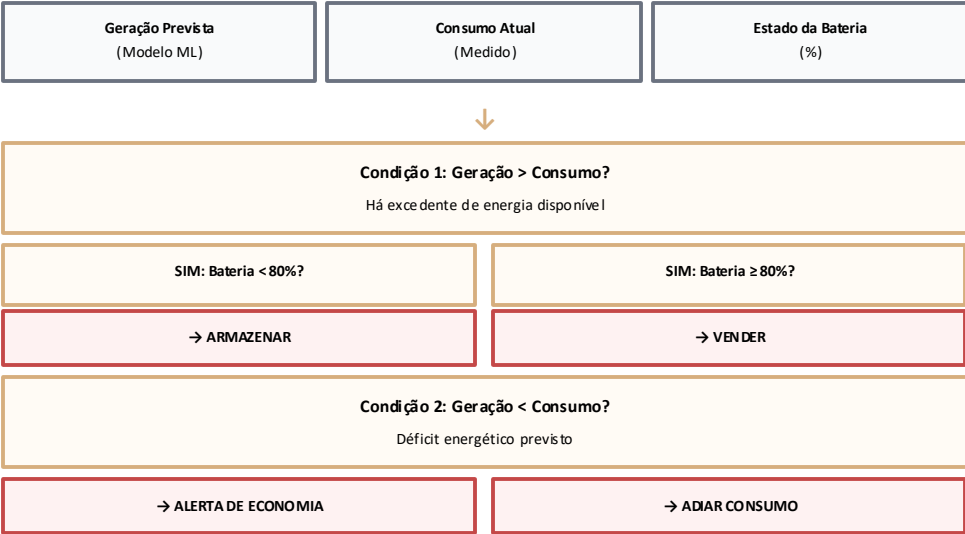
SE irradiância MÉDIA E temperatura NORMAL
→ Eficiência MÉDIA

Regra 4

SE irradiância BAIXA OU temperatura QUENTE
→ Eficiência BAIXA

Sistema de recomendações otimiza armazenamento, venda e consumo de energia

Fluxo de Decisão Inteligente



Avaliação contínua em tempo real • Priorização automática

Insight Principal

O sistema avalia continuamente três variáveis para priorizar: 1º autoconsumo imediato, 2º armazenamento, 3º venda de excedente.

1. Armazenamento em Bateria

Quando: Geração > Consumo E Bateria < 80%
Benefício: Autonomia em períodos noturnos ou nublados

2. Venda para a Rede

Quando: Geração > Consumo E Bateria > 80%
Benefício: Receita adicional e otimização do investimento

3. Economia/Alerta

Quando: Geração < Consumo médio
Benefício: Evitar custos de energia da rede convencional

4. Uso Otimizado de Eletrodomésticos

Quando: Pico de geração previsto (11h-14h)
Benefício: Maximizar autoconsumo e reduzir dependência

Lógica de Decisão:

Autoconsumo prioritário

Venda de excedente

Visualizações revelam padrões diários, sazonais e correlações entre variáveis

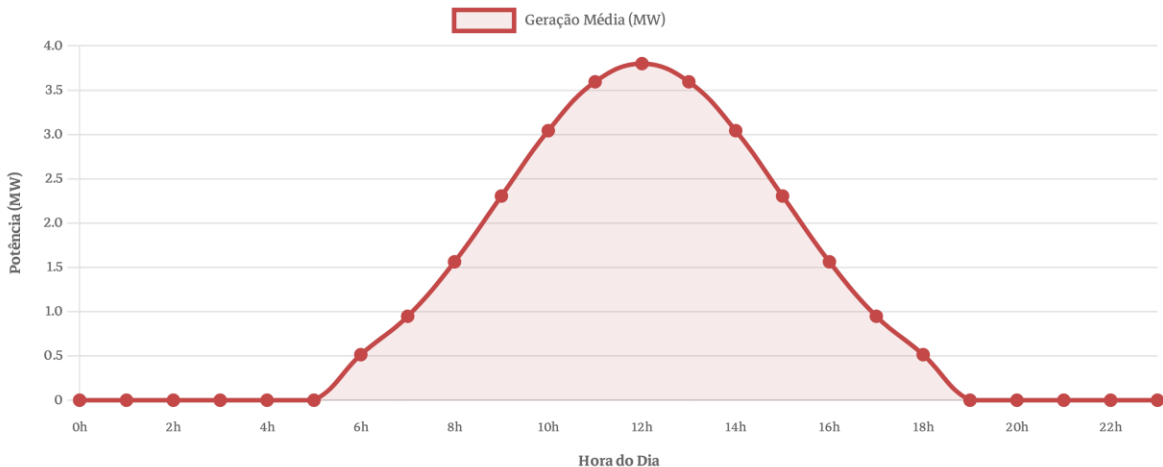
Insight Principal

A análise exploratória identificou cinco padrões fundamentais que orientaram a modelagem e validaram os resultados das previsões.

Principais Descobertas

- Ciclo diário com pico de 3,8 MW entre 11h-14h
- Sazonalidade anual: verão gera 100% mais que inverno
- Correlação irradiação-geração de 0,89 (muito forte)
- Temperatura reduz geração em 0,5% por °C acima de 25°C

Padrão 1: Ciclo Diário de Geração Solar (Média por Hora)



0,89

Correlação Irradiância-Geração

100%

Diferença Sazonal (Verão vs. Inverno)

2,2 MW

Gap Médio Geração-Consumo

Validação cruzada com dados NASA confirma generalização do modelo para novas localidades

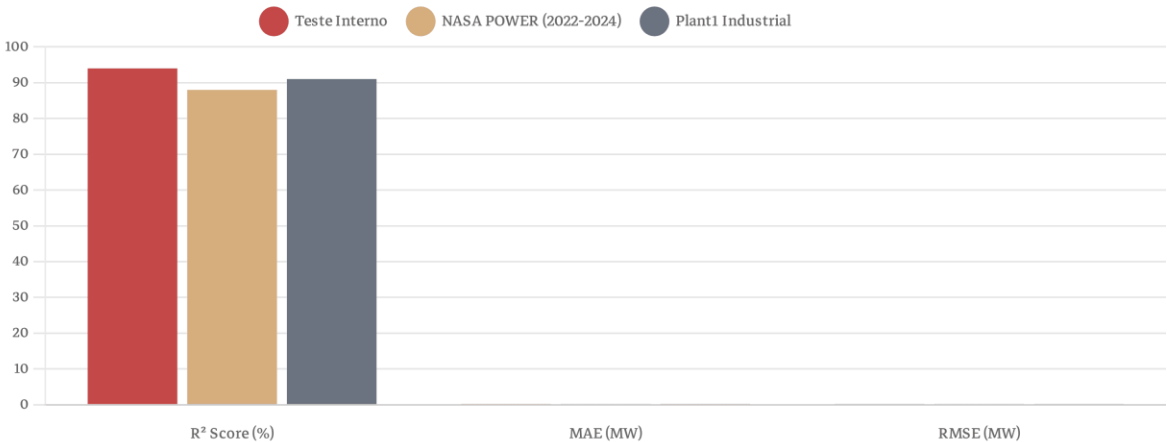
Insight Principal

O modelo demonstra capacidade de generalização adequada para diferentes contextos geográficos e temporais, validando sua aplicabilidade em cenários reais.

Resultados da Validação

- Validação NASA: R^2 de 0,88 com dados satelitais 2022-2024
- Validação Plant1: R^2 de 0,91 com dados industriais reais
- Degradação de apenas 6% vs. teste interno (aceitável)
- MAE de 0,19 MW representa erro de apenas 3,2%

Comparação de Métricas: Teste Interno vs. Validações Externas



Teste Interno

0,94

R^2 Score (Baseline)

NASA POWER

0,88

R^2 Score (2022-2024)

Plant1 Industrial

0,91

R^2 Score (34 dias)

Próximos Passos

Roadmap de evolução do Solarium SIEP

Evolução Técnica



Redes Neurais LSTM

Capturar dependências temporais de longo prazo



Previsão Probabilística

Intervalos de confiança (ex: 80% de 2-3 MW)



Ensemble Avançado

Combinar RF + GB + LSTM para maior precisão



Transfer Learning

Adaptar modelos para novas instalações

Integração Operacional

- IoT em tempo real com sensores conectados
- API REST para disponibilizar previsões
- Dashboard interativo para operadores
- Alertas automáticos sobre anomalias
- Manutenção preditiva de painéis
- Otimização de portfólio multi-instalações
- Integração com mercado de energia

Impacto Esperado: $R^2 > 0,96$ | Redução de custos 15-20% | Aumento de receita 10-15%