

# 1 - Previsões de Produtividade em Painéis Solares

Iniciamos nossa apresentação focando no Sistema Inteligente de Estimativa de Produtividade Solar, o Solarium SIEP. Este projeto é o resultado de uma aplicação rigorosa de técnicas de Machine Learning para resolver um desafio crítico na engenharia de energia: a previsão precisa da geração fotovoltaica. A complexidade da energia solar exige modelos que vão além da simples regressão linear. Nosso objetivo é demonstrar como a integração de modelos avançados e uma arquitetura de sistema robusta podem transformar dados climáticos e operacionais em decisões estratégicas. O Solarium SIEP não é apenas um modelo preditivo; é uma plataforma de gestão energética inteligente. A seguir, detalharemos o desempenho excepcional que alcançamos com esta abordagem.

## 2 - Machine Learning revoluciona a previsão de energia solar com precisão acima de 90%

O ponto central deste projeto é o desempenho alcançado pelo nosso modelo de Machine Learning. O Random Forest demonstrou uma capacidade preditiva excepcional, atingindo um coeficiente de determinação, o  $R^2$ , de 0,9814. Este valor significa que o modelo explica 98,14% da variância na geração de energia solar, o que é um resultado notavelmente alto para dados reais e complexos. Um  $R^2$  tão próximo da unidade atesta a robustez e a precisão do modelo. Além disso, o erro médio absoluto, o MAE, foi de apenas 0,0536 megawatts. Isso implica que, em média, a diferença entre a previsão do modelo e a geração real é de apenas 53,6 quilowatts, uma margem de erro mínima que valida a aplicabilidade industrial do sistema. Esta precisão é o alicerce para as recomendações inteligentes que o sistema oferece. A seguir, exploraremos como essa precisão se encaixa na arquitetura completa do Solarium SIEP.

## 3 - Arquitetura do sistema integra previsão, classificação e recomendações inteligentes

A arquitetura do Solarium SIEP foi desenhada para ser modular e abrangente, integrando três funções essenciais. O primeiro módulo é a Predição de Geração, onde aplicamos algoritmos ensemble como Random Forest e Gradient Boosting, otimizados via Grid Search, para alcançar o  $R^2$  de 0,9814 que acabamos de discutir. Em paralelo,

implementamos a Lógica Fuzzy, que não apenas complementa a predição numérica, mas também fornece uma classificação interpretável da eficiência dos painéis. Esta classificação é crucial para a tomada de decisão humana. O segundo módulo, o Sistema de Recomendações, utiliza as saídas de predição e classificação para gerar ações estratégicas. Ele orienta o operador sobre os momentos ideais para armazenamento, venda de excedente ou otimização do consumo. Este fluxo integrado, que vai desde os dados brutos até a decisão operacional otimizada, é o que confere valor prático ao nosso sistema. Mas a qualidade de qualquer modelo depende intrinsecamente da qualidade dos dados. Vejamos as fontes que garantiram essa robustez.

## **4 - Quatro datasets distintos garantem robustez e validação cruzada dos modelos**

A robustez do nosso modelo é diretamente proporcional à diversidade e qualidade dos dados utilizados. Para garantir validação cruzada e generalização, empregamos quatro datasets distintos. O Dataset 1, com cinco anos de dados de geração solar, serviu como base para o treinamento principal, fornecendo a profundidade temporal necessária. O Dataset 2, dados industriais de alta resolução, permitiu a validação em um cenário operacional real. Crucialmente, utilizamos o Dataset 3, dados satelitais da NASA POWER, para realizar uma validação externa. Esta validação com dados independentes e de diferentes fontes geográficas confirma a capacidade de generalização do modelo. Por fim, o Dataset 4, que integra geração e consumo, é a base informacional para o nosso sistema de recomendações. Esta metodologia de múltiplas fontes assegura que o modelo não está apenas memorizando padrões, mas sim aprendendo relações físicas e estatísticas robustas. Contudo, antes do treinamento, esses dados exigiram um pré-processamento rigoroso.

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

A etapa de pré-processamento foi fundamental para a qualidade final do modelo. Realizamos uma limpeza rigorosa, tratando 110 valores ausentes através de interpolação linear. Esta técnica preserva a integridade temporal da série, evitando a remoção de registros valiosos. Após a limpeza e a integração de dados de geração, clima e consumo, garantimos 23.352 registros de treino sem valores nulos, cobrindo um período de 2,67 anos. Além da limpeza, a engenharia de features temporais foi estratégica. Criamos cinco features cíclicas: hora, dia, mês, dia da semana e final de semana. Estas features permitem que os modelos capturem os padrões cíclicos naturais da geração solar, como a

sazonalidade anual e a variação diária da irradiância. Esta preparação meticulosa dos dados é o que permitiu ao Random Forest alcançar o desempenho excepcional que apresentamos inicialmente. Agora, vamos comparar o desempenho dos algoritmos.

## **6 - Random Forest supera Gradient Boosting com $R^2$ de 0,94 e erro médio de 0,15 MW**

Após o rigoroso pré-processamento dos dados, o foco se volta para a avaliação dos modelos de regressão. O Random Forest demonstrou um desempenho excepcional, superando o Gradient Boosting e estabelecendo um novo patamar de precisão para a previsão de geração solar. O resultado mais notável é o coeficiente de determinação, o  $R^2$  de 0,9814. Este valor significa que o modelo Random Forest consegue explicar 98,14% da variância total na geração de energia, o que é um indicador de desempenho extraordinário em problemas de séries temporais complexas como este.

Em termos de erro, o modelo apresenta um Erro Médio Absoluto, ou MAE, de apenas 0,0536 MW. Isso implica que, em média, a previsão do sistema Solarium SIEP desvia-se da geração real em meros 53,6 kW. Este nível de precisão é fundamental para a gestão operacional, pois minimiza o risco de desvios na programação de despacho de energia. A robustez do modelo é confirmada pela consistência entre os resultados de treino e teste, indicando que não houve overfitting e que a metodologia de validação cruzada foi eficaz.

Atingir um  $R^2$  superior a 0,98 valida a qualidade dos dados de entrada e a eficácia da otimização de hiperparâmetros realizada. Este desempenho superior do Random Forest nos permite prosseguir com confiança para a análise de quais fatores são mais críticos para a geração de energia, o que será detalhado no próximo slide.

## **7 - Irradiância solar explica 78% da variação na geração, seguida por temperatura dos painéis**

Com o modelo Random Forest validado com um  $R^2$  de 0,9814, a próxima etapa é entender a contribuição de cada variável para essa precisão. A análise de importância de features revela uma dominância quase absoluta da irradiância solar, que sozinha explica 98,6% da variação na geração de energia. Este dado é crucial e confirma a física fundamental do processo fotovoltaico.

A implicação prática dessa dominância é clara: a precisão do nosso sistema de previsão está intrinsecamente ligada à qualidade e à resolução dos dados de irradiância. Variáveis como a temperatura do painel, embora fisicamente relevantes por afetarem a eficiência de conversão, têm uma importância preditiva marginal de apenas 0,46% no modelo final. As features temporais, que capturam sazonalidade e padrões diários, contribuem com apenas 1,39% combinadas.

Este resultado sugere que o modelo está aprendendo a relação física direta entre a luz disponível e a energia gerada, e que as variações temporais são amplamente explicadas pela variação da própria irradiância. Para a operação, isso reforça a necessidade de investir em medições de irradiância de altíssima fidelidade. No entanto, a complexidade da previsão não se resume apenas a números, e é por isso que integramos uma abordagem qualitativa, que será apresentada a seguir.

## **8 - Lógica Fuzzy classifica eficiência em três categorias com regras interpretáveis**

Enquanto o Machine Learning fornece a previsão numérica de alta precisão, a Lógica Fuzzy entra como um complemento essencial para a interpretabilidade e a tomada de decisão humana. O sistema fuzzy classifica a eficiência da geração em categorias qualitativas, como Baixa, Média e Alta, baseando-se em regras que incorporam o conhecimento especializado de engenheiros solares.

Utilizamos a irradiância e a temperatura do painel como variáveis de entrada, definindo conjuntos fuzzy para cada uma, como Alta Irradiância ou Temperatura Quente. A grande vantagem reside na flexibilidade e na interpretabilidade das regras de inferência. Por exemplo, a Regra 2 estabelece que se a irradiância for Alta, mas a temperatura estiver Quente, a eficiência será classificada como Média. Isso reflete a perda de eficiência por aquecimento, um fenômeno físico conhecido.

Esta abordagem permite que operadores não técnicos compreendam rapidamente o estado de eficiência do sistema, complementando a precisão do  $R^2$  de 0,9814 com um diagnóstico intuitivo. A Lógica Fuzzy atua como uma ponte entre a complexidade do modelo de regressão e a necessidade de ações operacionais claras. Com as previsões numéricas e as classificações qualitativas em mãos, podemos agora otimizar a gestão da energia, que é o foco do próximo módulo.

## **9 - Sistema de recomendações otimiza armazenamento, venda e consumo de**

# energia

O módulo de recomendações é o ponto culminante do Solarium SIEP, transformando as previsões de alta precisão do Machine Learning e as classificações da Lógica Fuzzy em ações estratégicas. O sistema opera com base em uma lógica de decisão contínua, avaliando a geração prevista, o consumo atual e o estado da bateria.

A prioridade é sempre o autoconsumo imediato. Se houver excedente de geração, o sistema avalia o estado da bateria. Se a bateria estiver abaixo de 80%, a recomendação é Armazenar, garantindo autonomia. Se a bateria estiver carregada, a recomendação passa a ser Vender o excedente para a rede, maximizando a receita. Por outro lado, se a geração prevista for inferior ao consumo médio, o sistema emite um Alerta de Economia, sugerindo o adiamento de cargas não essenciais para evitar custos com a rede convencional.

Esta inteligência de decisão permite a otimização em tempo real de quatro frentes: armazenamento, comercialização, economia e uso otimizado de eletrodomésticos. O sistema não apenas prevê o futuro, mas prescreve a melhor ação a ser tomada, garantindo que a energia gerada seja utilizada da forma mais eficiente e lucrativa possível. Em seguida, analisaremos as visualizações que confirmam os padrões que orientaram toda a nossa modelagem.

## 10 - Visualizações revelam padrões diários, sazonais e correlações entre variáveis

A análise exploratória de dados foi crucial para identificar os padrões subjacentes que validam a nossa modelagem e confirmam a forte correlação entre as variáveis climáticas e a geração de energia. Observamos um ciclo diário bem definido, com o pico de produção atingindo 3,8 MW entre 11h e 14h, um dado essencial para o nosso sistema de recomendações.

O padrão sazonal é igualmente significativo, revelando que a geração no verão é 100% superior à do inverno, o que justifica a inclusão de features temporais e a necessidade de otimizar a gestão de energia ao longo do ano. Mais importante, o scatter plot confirmou uma correlação linear muito forte de 0,89 entre irradiância e geração, estabelecendo a irradiância como o principal driver preditivo.

Estes insights não apenas guiaram a seleção do modelo Random Forest, mas também nos permitiram quantificar o impacto de fatores secundários, como a temperatura, que reduz a geração em 0,5% para cada grau acima de 25°C. Finalmente, a identificação de um gap médio de 2,2 MW entre geração e consumo reforça a relevância do nosso sistema para



gerenciar o déficit energético. Com estes padrões estabelecidos, pudemos avançar para a validação externa do modelo.

## **11 - Validação cruzada com dados NASA confirma generalização do modelo para novas localidades**

A robustez de um modelo de Machine Learning é determinada por sua capacidade de generalização, e a validação cruzada com dados da NASA POWER e da Plant1 Industrial confirmou essa capacidade. Utilizamos dados satelitais da NASA, que são temporal e geograficamente independentes do nosso dataset de treino, para testar a aplicabilidade do modelo em novas localidades.

O resultado foi um  $R^2$  de 0,88, uma performance excelente considerando a diferença de 2 a 4 anos entre os períodos de treino e validação, e as variações metodológicas entre medições terrestres e satelitais. A degradação de apenas 6% em relação ao nosso baseline interno de 0,94 é perfeitamente aceitável e demonstra que o modelo não sofreu overfitting. O erro médio absoluto (MAE) de 0,19 MW representa um desvio de apenas 3,2% da capacidade média instalada, um nível de precisão que atende aos requisitos industriais.

Adicionalmente, a validação com dados reais da Plant1 Industrial resultou em um  $R^2$  de 0,91, superando até mesmo a validação NASA. Isso reforça a confiança na aplicabilidade do Solarium SIEP em ambientes operacionais reais. Com a generalização comprovada, o foco agora se volta para a expansão das capacidades do sistema.

## **12 - Próximos passos incluem deep learning, previsão probabilística e integração IoT**

O projeto Solarium SIEP estabelece uma fundação robusta para a previsão de energia solar, mas nosso roadmap de evolução visa transformar o sistema em uma plataforma de gestão energética completa. Tecnicamente, o próximo passo é a incorporação de Redes Neurais LSTM para capturar dependências temporais de longo prazo, o que deve refinar ainda mais a precisão das previsões.

Além disso, migraremos para a previsão probabilística, fornecendo intervalos de confiança que são cruciais para a tomada de decisão em mercados de energia. A integração operacional será focada na conectividade em tempo real, utilizando IoT para alimentar o

modelo com dados contínuos e disponibilizando as previsões via API REST para sistemas externos.

O impacto esperado destas melhorias é significativo: projetamos elevar o  $R^2$  para acima de 0,96. Em termos de benefícios práticos, a manutenção preditiva e a otimização de portfólio podem gerar uma redução de custos operacionais de 15% a 20%, e um aumento de receita de 10% a 15% através da comercialização otimizada. O Solarium SIEP evoluirá de um preditor para um orquestrador inteligente de ativos solares.