

Candidato: Enzo Laragnoit Fernandes

Previsão de Irradiação Solar para Sistemas Fotovoltaicos utilizando *Machine Learning*

**Uma abordagem de Aprendizagem Supervisionada para Previsão de
Irradiação Solar no Estado de São Paulo**

FAPESP

Fundação de Amparo à Pesquisa do Estado de São Paulo

Iniciação Científica

Sorocaba

Julho, 2020

Resumo

Com a crescente demanda por energia, surgem diferentes formas de produção de energia elétrica. Dentre elas tem-se a energia solar fotovoltaica, uma fonte renovável, abundante e de baixo custo, cujas aplicações dependem diretamente da intensidade de irradiação solar. Neste projeto de iniciação científica propõe-se o uso técnicas de Aprendizado de Máquina Supervisionado (*Supervised Learning*) para a implementação de modelos capazes de realizar a previsão de irradiação solar a partir de um conjunto de dados conhecidos *a priori* em localidades específicas dentro do Estado de São Paulo. Uma vez que o conjunto de dados é obtido serão utilizadas técnicas para realizar o pré-processamento e a Engenharia de Atributos a fim de otimizar os algoritmos de aprendizagem supervisionada na base de dados. A capacidade de previsão dos modelos será avaliada em dois instantes: durante a etapa de ajustes, e posteriormente, na etapa de validação. Os resultados serão analisados por meio do tabelamento onde informações como os hiperparâmetros dos modelos, as diferenças entre as previsões e as observações e as métricas de erro para regressão serão registradas. A implementação será realizada na linguagem de programação Python em conjunto com diversas bibliotecas dedicadas para aprendizagem supervisionada, manipulação e visualização do conjunto de dados.

Palavras-chave: aprendizado de máquina, aprendizagem supervisionada, energia renovável, energia solar, irradiação solar.

Abstract

With the increasing demand for energy different forms of electrical energy yield arise. Among them there is the photovoltaic solar energy, a renewable, abundant and low cost energy source whose applications depend directly on the solar irradiation intensity. In this project one proposes the use of Supervised Machine Learning techniques in order to implement models capable of predicting solar irradiation from a known dataset in specific sites within the State of São Paulo, Brazil. Once such dataset is obtained different techniques will be applied aiming to carry the preprocessing and Feature Engineering steps to optimize the Supervised Machine Learning algorithms. Models' capability of prediction will be evaluated in two moments: during the adjustment stage and, subsequently, in the validation stage. Results will be tabulated and analysed regarding the models' hyperparameters, the difference between predicted and observed values and error metrics. Implementation will be carried out using the Python programming language in addition to diverse machine learning, data manipulation and visualization libraries.

Keywords: machine learning, supervised learning, renewable energy, solar energy, solar irradiation.

1 Introdução e Justificativa

Um dos principais desafios para o século XXI é conciliar a busca por soluções energéticas com o desenvolvimento sustentável sem causar maiores prejuízos ao já afetado meio ambiente. Com a crescente demanda por energia elétrica (EPE, 2017) movida pela progressiva informatização da sociedade moderna surgem diferentes formas renováveis de geração de energia elétrica em alternativa aos finitos recursos não renováveis como carvão mineral e petróleo.

Uma das alternativas promissoras é a conversão de energia solar em energia elétrica por meio de painéis solares fotovoltaicos que transformam a irradiação solar em energia elétrica que pode ser prontamente utilizada ou armazenada em baterias sob a forma de energia química. Presente em grande abundância e em praticamente todos os lugares, a quantidade de energia incidente na Terra proveniente do Sol é aproximadamente 10.000 vezes maior que a quantidade de energia consumida pela humanidade (SMETS et al., 2015). A energia solar apresenta vantagens quando comparada a outras formas de energia renovável, como a eólica e a hidrelétrica, pela facilidade de instalação, considerando aspectos de construção, danos ambientais, custos operacionais e manutenção. Além disso, apresenta a possibilidade de ser utilizada sem grandes dificuldades nos ambientes urbanos e também oferece como alternativa um modelo de geração de energia descentralizada no qual consumidores podem gerar sua própria energia. O aumento da eficiência energética e a constante redução dos custos dos sistemas fotovoltaicos justificam sua tendência de crescimento e inserção nas matrizes energéticas (ACHILLES, 2013).

Outra área de grandes avanços e crescentes expectativas é a de Inteligência Artificial, especialmente a subárea de Aprendizado de Máquina (PERRAULT et al., 2019). Possibilitada pelo aprimoramento das técnicas e pelo aumento do poder computacional dos sistemas modernos, sua presença em aplicações utilizadas no cotidiano como reconhecimento de imagens, processamento de linguagem natural, *chatbots* e sistemas de recomendação se tornou praticamente universal em *softwares* nas mais diversas plataformas. Grande esperança é depositada nos *insights* e resultados obtidos da aplicação dessas técnicas para trabalhar com questões no fronteira do conhecimento de tópicos fundamentais como energia, saúde e economia em consonância com a computação científica.

Dentre as aplicações possibilitadas pelo uso das técnicas de Aprendizado de Máquina destaca-se, no contexto deste projeto, o uso de Aprendizado de Máquina Supervisionado (*Supervised Machine Learning*) na competência de aferir previsões a partir de um conjunto de dados já conhecidos *a priori*. Essa tarefa é realizada utilizando diferentes algoritmos como em (HAYKIN, 1994; SUYKENS; VANDEWALLE, 1999), cujo propósito geral é obter uma função $f : X \rightarrow Y$ capaz de realizar o mapeamento entre dados de entrada, chamados de atributos (*features*), e as variáveis alvo (*targets*) cujos valores se objetiva determinar. Para este projeto, as informações relacionadas aos dados meteorológicos serão utilizadas como os dados de entrada X da função f e a radiação solar horizontal global será a variável alvo (*target variable*) a ser determinada.

Além das características intrínsecas à energia solar fotovoltaica já salientadas, há de se

considerar o grande potencial de geração ainda pouco explorado de suas aplicações (PEREIRA et al., 2017). Apesar de dispor de uma matriz energética composta predominantemente por fontes renováveis, a energia solar fotovoltaica representa menos de 1% da capacidade instalada de geração de energia elétrica no Estado de São Paulo (INFRAESTRUTURA E MEIO AMBIENTE, 2019). Outro aspecto relevante é o fato de o Estado de São Paulo consumir mais energia elétrica do que gera, atestando a necessidade de infraestrutura de terceiros para atender à demanda interna por eletricidade. Conforme (INFRAESTRUTURA E MEIO AMBIENTE, 2020), o estado consome aproximadamente 30% da energia elétrica gerada no Brasil (apenas o setor residencial do estado consome mais que as regiões Norte e Nordeste juntas) enquanto gera apenas 12% e por isso tem apresentado uma dependência média de 40% nos últimos anos (INFRAESTRUTURA E MEIO AMBIENTE, 2019).

A principal justificativa para esse projeto reside no fato de os objetivos propostos serem diretamente relacionados ao monitoramento e à análise de viabilidade para implantação de sistemas de geração de energia solar fotovoltaica. O estabelecimento destes sistemas tem o potencial de reduzir a dependência energética do estado de São Paulo ao mesmo tempo que proporciona o aumento da participação de fontes renováveis em sua matriz energética fomentando o desenvolvimento das regiões envolvidas e o estabelecimento de um modelo de geração de energia descentralizado. Neste projeto o uso das técnicas de Aprendizado de Máquina voltado para aplicações de energia solar fotovoltaica representa um dos esforços na busca por soluções para as questões energéticas enfrentadas na atualidade. Ele foi baseado em aplicações similares dessas técnicas em diferentes contextos e localidades como (ALZAHIRANI et al., 2017; LI et al., 2016; ASSOULINE; MOHAJERI; SCARTEZZINI, 2017; LOU et al., 2016; ZENG; QIAO, 2013), na intersecção entre as áreas de Inteligência Artificial e Energia Solar.

Tratando-se de uma Iniciação Científica, o projeto oferece a oportunidade de consolidar e integrar os conhecimentos adquiridos no âmbito da graduação por meio de disciplinas presentes na grade curricular do curso de Ciência da Computação tais como Geometria Analítica, Cálculo Diferencial e Integral, Programação Orientada a Objetos, Estruturas de Dados e Aprendizado de Máquina em aplicações reais.

Por fim, após seu término, o projeto pode ser utilizado como um embasamento para o desenvolvimento de futuros trabalhos tanto no desenvolvimento das técnicas e aplicações de aprendizado de máquina quanto na continuidade do estudo de aplicações de energia renovável, especialmente da energia solar fotovoltaica.

2 Objetivos

O objetivo central desse projeto consiste em utilizar técnicas de Aprendizado Supervisionado a fim de implementar e testar modelos para prever a intensidade de irradiação solar, medida em W/m^2 , em intervalos de antecedência de 30 minutos em pontos específicos dentro do

estado de São Paulo. Este intervalo de antecedência se justifica por procurar atender às especificações do Sistema de Coleta de Dados de Energia em conformidade com as Regras de Comercialização de Energia estabelecidas pela Câmara de Comercialização de Energia Elétrica (COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA, 2020) a fim de corroborar com a redução dos custos de operação e otimização do planejamento das políticas de distribuição de energia. A capacidade de prever a intensidade de irradiação solar, neste contexto, permite não apenas o melhor monitoramento operacional dos pontos de geração já instalados mas também possibilita a análise de viabilidade mais detalhada para os locais candidatos à instalação de infraestrutura voltada à geração de energia elétrica a partir da energia solar fotovoltaica.

Para que o objetivo geral desse projeto seja alcançado foram estabelecidas as seguintes metas que deverão ser executadas conforme o cronograma estipulado na Seção 4:

1. Obtenção do conjunto de dados;
2. Geração do conjunto de dados preprocessados;
3. Seleção dos algoritmos para a implementação dos modelos;
4. Implementação dos modelos de previsão a partir de diferentes algoritmos de aprendizagem supervisionada selecionados;
5. Teste e avaliação da performance dos modelos em relação ao conjunto de dados obtidos *a priori*;
6. Ajuste dos modelos com base nas informações do item 5;
7. Teste e avaliação da performance dos modelos em na previsão de irradiação nos locais específicos durante 1 mês;
8. Elaboração dos Relatórios;

3 Material e Métodos

Os aspectos relativos aos materiais e à metodologia que serão utilizado neste projeto estão divididos em duas etapas: a obtenção dos conjuntos de dados, uma vez determinados os pontos de coleta, e o processamento destes e implementação dos modelos.

3.1 Pontos de Coleta de Dados

A fim de determinar os pontos de coleta de dados dentro do estado de São Paulo dois critérios serão adotados:

- (i) a proximidade dos pontos à linhas de transmissão de energia elétrica;

(ii) a proximidade dos pontos aos centros urbanos mais populosos.

Para o critério (ii) serão utilizadas as informações disponíveis no site do Instituto Brasileiro de Geografia e Estatística (IBGE) referentes ao estado de São Paulo¹. Já no critério (i) serão utilizadas informações obtidas por meio do site da Agência Nacional de Energia Elétrica (ANEEL)², que disponibiliza informações sobre a rede de transmissão de energia em todo o país. Uma vez que esses dados forem obtidos seu escopo será restrito apenas às informações relativas às linhas de transmissão dentro do estado de São Paulo.

Após a obtenção dessas informações, o procedimento a ser realizado para definir os pontos de coleta de dados será a deliberação das coordenadas geográficas dentro do estado que melhor satisfaçam os critérios (i) e (ii) elencados acima.

3.2 Preprocessamento e Implementação

Uma vez que as informações são obtidas haverá a etapa de preprocessamento que é constituída, em linhas gerais, pelo tratamento de inconsistências numéricas nas observações de cada uma das variáveis e normalização dos valores numéricos para a otimização dos algoritmos. Nesta etapa, há também a possibilidade de se utilizar técnicas de *Feature Engineering* para obter combinações de atributos que possam contribuir para a identificação de padrões na variável alvo em relação ao conjunto de dados. Por fim, há separação dos conjunto de dados preprocessado em dois conjuntos de dados, $X = [x_1, x_2, \dots, x_n]^T$ e $Y = [y_1, y_2, \dots, y_n]^T$, de forma que X contenha os dados de entrada do algoritmo e Y contenha as variáveis alvo, também denominadas “rótulos”. Aqui x_i representa o vetor de com m atributos, $x_i = (a_1, a_2, \dots, a_m)$, da i -ésima amostra; e y_i representa o vetor com k rótulos, $y_i = (r_1, r_2, \dots, r_k)$, da i -ésima amostra. Neste cenário, o número de atributos m é relacionado ao conceito de *dimensionalidade* da amostra de dados.

Desta forma, um possível exemplo poderia ser:

$$\begin{aligned}x_i &= (\text{temperatura}, \quad \text{umidade relativa}, \quad \text{pressão atmosférica}, \quad \dots) \\y_i &= (\text{irradiação solar})\end{aligned}$$

Para a implementação dos modelos de previsão, a linguagem de programação a ser utilizada será a linguagem Python por sua versatilidade e praticidade além de dispor de várias bibliotecas com implementações voltadas para Aprendizado de Máquina e Análise de Dados que podem ser utilizadas durante a execução das etapas de implementação desse projeto. Algumas dessas bibliotecas são:

- **Numpy** para operações envolvendo vetores e álgebra linear de forma geral além de ser extensivamente utilizado pelas demais bibliotecas. <<https://numpy.org/>>;

¹ <https://www.ibge.gov.br/cidades-e-estados/sp.html>

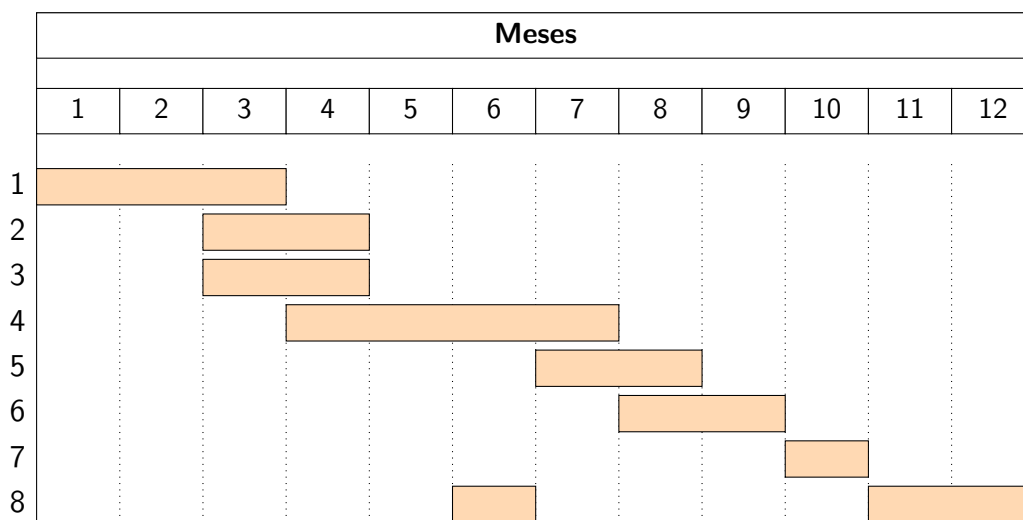
² <https://sigel.aneel.gov.br/portal/home/>

- **Scikit-Learn** para o pré-processamento do conjunto de dados, implementação dos algoritmos aprendizado de máquina e avaliação dos modelos de previsão. <<https://scikit-learn.org/>>;
- **Pandas** para manipulação e pré-processamento do conjunto de dados <<https://pandas.pydata.org/>>;
- **TensorFlow** para implementação de algoritmos de aprendizagem profunda (*Deep Learning*) tais como redes neurais profundas e avaliação dos respectivos modelos <<https://tensorflow.org/>>;
- **Matplotlib** para representação visual dos dados <<https://matplotlib.org/>>.

4 Cronograma

A execução das metas listadas na Seção 2 seguirá o cronograma ilustrado na Tabela 1 com um período de execução total de 12 (doze) meses.

Tabela 1 – Cronograma de desenvolvimento do projeto. Cada linha corresponde a uma meta específica da Seção 2. Cada coluna corresponde a um mês.



Fonte: Elaborada pelo autor.

5 Análise dos Resultados

Os resultados obtidos pelos modelos implementados no decorrer deste projeto serão analisados por meio de tabelamento em dois instantes:

1. Durante a execução do item 5 da Seção 2, na qual os valores obtidos serão comparados com os valores esperados dentro do escopo do conjunto de dados obtidos no item 1. Neste instante os resultados serão utilizados tanto para fins de documentação quanto para realizar os ajustes na implementações dos modelos a fim de realizar o item 7 da Seção 2.
2. Durante a execução do item 7 da Seção 2, na qual os resultados obtidos pelos modelos já ajustados, serão comparados com os valores obtidos de observações nos locais de previsão como forma de validação dos modelos.

Para cada modelo de previsão com o respectivo algoritmo de aprendizagem supervisionada implementado durante o item 4 da Seção 2, haverá o tabelamento dos hiperparâmetros utilizados e das métricas de avaliação para as previsões. Por tratar-se de uma tarefa relacionada à regressão, algumas das métricas que podem ser utilizadas para a avaliação da performance dos modelos durante as duas etapas de análise de resultados apresentadas anteriormente são:

- Correlação (R^2) entre as previsões e as observações;
- MSE – Erro quadrático médio

$$MSE = \frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2 ,$$

- MAPE – Erro percentual absoluto médio

$$MAPE = \frac{1}{n} \sum_{i=0}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% ,$$

- MAE – Erro absoluto médio

$$MAE = \frac{1}{n} \sum_{i=0}^n |y_i - \hat{y}_i| ,$$

onde n indica o número total de amostras, y_i indica o valor numérico da i -ésima observação e \hat{y}_i é o valor numérico da i -ésima previsão realizada pelo modelo.

Referências

- ACHILLES, R. **Energia Solar Paulista: Levantamento do Potencial**. 2013. P. 8.
- ALZAHIRANI, Ahmad et al. Solar Irradiance Forecasting Using Deep Neural Networks. **Procedia Computer Science**, Elsevier, v. 114, p. 304–313, 2017.
- ASSOULINE, Dan; MOHAJERI, Nahid; SCARTEZZINI, Jean-Louis. Quantifying Rooftop Photovoltaic Solar Energy Potential: A Machine Learning Approach. **Solar Energy**, Elsevier, v. 141, p. 278–296, 2017.
- COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA, Câmara de. **Regras De Comercialização de Energia - Módulo: Medição física**. 2020.
- EPE, EDPE. Projeção de Demanda de Energia Elétrica-2017-2026. **Ministério de Minas e Energia**. Rio de Janeiro, p. 95, 2017.
- HAYKIN, Simon. **Neural Networks: a Comprehensive Foundation**. Prentice Hall PTR, 1994.
- INFRAESTRUTURA E MEIO AMBIENTE, Secretaria de. **Balanco Energético do Estado de São Paulo**. 2019.
- _____. **Resumo Executivo: Dados de Produção e Consumo de Energia Elétrica**. 2020.
- LI, Jiaming et al. Machine Learning for Solar Irradiance Forecasting of Photovoltaic System. **Renewable energy**, Elsevier, v. 90, p. 542–553, 2016.
- LOU, Siwei et al. Prediction of Diffuse Solar Irradiance Using Machine Learning and Multivariable Regression. **Applied energy**, Elsevier, v. 181, p. 367–374, 2016.
- PEREIRA, Enio Bueno et al. **Atlas Brasileiro de Energia Solar**. INPE São José dos Campos, 2017. v. 1.
- PERRAULT, Raymond et al. The AI Index 2019 Annual Report. **AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA**, 2019.
- SMETS, Arno HM et al. **Solar Energy: The physics and engineering of photovoltaic conversion, technologies and systems**. UIT Cambridge, 2015.
- SUYKENS, Johan AK; VANDEWALLE, Joos. Least Squares Support Vector Machine Classifiers. **Neural processing letters**, Springer, v. 9, n. 3, p. 293–300, 1999.
- ZENG, Jianwu; QIAO, Wei. Short-term Solar Power Prediction Using a Support Vector Machine. **Renewable Energy**, Elsevier, v. 52, p. 118–127, 2013.