

Victor Gomide Ferraz

**Aplicação de Técnicas de Machine
Learning na Previsão de Geração de
Energia Elétrica na UFG**

Goiânia

2021

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR
VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE
GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC nº 1204/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo dos Trabalhos de Conclusão dos Cursos de Graduação disponibilizado no RI/UFG é de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o(s) autor(a)(es)(as) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG):

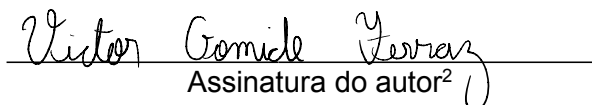
Nome completo do autor: Victor Gomide Ferraz

Título do trabalho: Aplicação de Técnicas de Machine Learning na Previsão de Geração de Energia Elétrica na UFG

2. Informações de acesso ao documento:

Concorda com a liberação total do documento SIM NÃO¹

Independente da concordância com a disponibilização eletrônica, é imprescindível o envio do(s) arquivo(s) em formato digital PDF do TCCG.


Assinatura do autor²

Ciente e de acordo:

Assinatura do(a) orientador(a)²

Data: 09 / 11 / 2021

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante: (a) consulta ao(à)s autor(a)(es)(as) e ao(à) orientador(a); b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo do TCCG. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;

² As assinaturas devem ser originais sendo assinadas no próprio documento, imagens coladas não serão aceitas.

Victor Gomide Ferraz

Aplicação de Técnicas de Machine Learning na Previsão de Geração de Energia Elétrica na UFG

Trabalho de conclusão de curso apresentado na Escola de Engenharia Elétrica, Mecânica e de Computação como requisito para a conclusão do curso de Engenharia de Computação e obtenção do título de Engenheiro de Computação.

Universidade Federal de Goiás – UFG

Escola de Engenharia Elétrica, Mecânica e de Computação (EMC)

Orientador: Prof. Dr. Marcelo Stehling de Castro

Goiânia

2021

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Ferraz, Victor Gomide

Aplicação de Técnicas de Machine Learning na Previsão de Geração de Energia Elétrica na UFG [manuscrito] / Victor Gomide Ferraz. - 2021.

XVII, 17 f.

Orientador: Prof. Dr. Marcelo Stehling de Castro.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de Computação (EMC), Engenharia da Computação, Goiânia, 2021.

1. Energia Solar. 2. Machine Learning. 3. Irradiância. 4. Placas Fotovoltaicas. I. de Castro, Marcelo Stehling, orient. II. Título.

CDU 62+004+005



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

ATA DE AVALIAÇÃO DE PROJETO FINAL

Curso

() Eng. Elétrica () Eng. Mecânica (X) Eng. de Computação
() Projeto Final I (X) Projeto Final II

Título do Trabalho

Aplicação de Técnicas de Machine Learning na Previsão de Geração de Energia Elétrica na UFG

Banca Avaliadora

Membro 1	Marcelo Stehling de Castro
Membro 2	Sanderley Ramos Pires
Membro 3	Gustavo Dias de Oliveira

Discente(s)

Matrícula	Nome
201403470	Victor Gomide Ferraz

NOTAS													
Matrícula	Membro 1				Membro 2				Membro 3				Média
	NPT	NTE	NAA	NF	NPT	NTE	NAA	NF	NPT	NTE	NAA	NF	
201403470	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0	10,0
NPT – Nota plano de trabalho;													
NTE – Nota do trabalho escrito;													
NAA – Nota de apresentação e arguição													
Para Eng. Elétrica, Mecânica e PFC2 da Eng. da Computação: $NF = 0,1 \times NPT + 0,45 \times NTE + 0,45 \times NAA$													
Para PFC1 da Eng. da Computação: $NF = 0,3 \times NPT + 0,7 \times NAA$													

* A APROVAÇÃO DO(S) ALUNO(S) ESTÁ CONDICIONADA À APRESENTAÇÃO DO TRABALHO FINAL AO ORINETADOR COM TODAS AS CORREÇÕES SUGERIDAS PELA BANCA.

OBSERVAÇÕES:

Preencher com modificações solicitadas, caso existam. Em caso de reprovação, informar a justificativa.



Documento assinado eletronicamente por **Marcelo Stehling De Castro, Professor do Magistério Superior**, em 05/11/2021, às 15:35, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sanderley Ramos Pires, Professor do Magistério Superior**, em 05/11/2021, às 15:40, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gustavo Dias De Oliveira, Técnico de Tecnologia da Informação**, em 06/11/2021, às 09:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2458160** e o código CRC **DA4868FD**.

Referência: Processo nº 23070.041130/2021-15

SEI nº 2458160

Aplicação de Técnicas de *Machine Learning* na Previsão de Geração de Energia Elétrica na UFG

Victor Gomide Ferraz¹ — Marcelo Stehling de Castro². Universidade Federal de Goiás (UFG) - Escola de Engenharia Elétrica, Mecânica e de Computação (EMC) - Goiânia, Goiás, Brasil 74601-010, e-mails: victorgomferraz@discente.ufg.br, mcastro@ufg.br²

Resumo — O presente artigo visa analisar a aplicação de técnicas de machine learning na previsão de energia produzida pelas placas fotovoltaicas instaladas no Campus Colemar Natal e Silva, na Universidade Federal de Goiás (UFG). Em um primeiro momento, apresenta-se conceitos e explicações básicas sobre como ocorre a geração de energia fotovoltaica e quais são os principais fatores que a afetam; posteriormente, aborda-se a temática de machine learning e o algoritmo SRV – escolhido para o teste – e os parâmetros que alteram seu comportamento. Em seguida são abordadas as relações entre as variáveis e calculado a relação entre as variáveis meteorológicas com a geração de energia elétrica. Por fim, é apresentado o treinamento com abordagens diversas entre a configuração do algoritmo e a quantidade de variáveis utilizadas para o treino, após os dados são comparados para compreender quais variáveis e configurações afetam sobremaneira a previsão de produção de energia pelas células fotovoltaicas.

Palavras-chaves — energia solar; *machine learning*; irradiância; placas fotovoltaicas.

Abstract — This article aims to analyze the application of machine learning techniques in predicting the energy produced by photovoltaic panel installed on Campus Colemar Natal e Silva, at the Federal University of Goiás (UFG). At first, it presents basic concepts and explanations about how photovoltaic energy is generated and what are the main factor that affect it; later, the thematic of machine learning and the SRV algorithm – chosen for the test – and the parameters that change the behavior are approached. Then, the relationships between the variables are discussed and the relationship between the metrological variables and the generation of electricity is calculated. Finally, training is presented with different approaches between the configuration of the algorithm and the amount of variables used of training, after which the data are compared to understand which variables and settings greatly affect the prediction of energy production by photovoltaic cells.

Keywords — solar energy; *machine learning*; irradiance; photovoltaic cells.

I. INTRODUÇÃO

Desde o final da década de 60 intensificou-se os estudos e discussões sobre a insustentabilidade das atividades humanas e impactos ambientais; entretanto, após mais de 60 anos as fontes de energia majoritariamente usadas no Brasil ocasionam grandes impactos ambientais [1].

O direito a um meio ambiente ecologicamente equilibrado e saudável é assegurado pelo artigo 25 da Declaração Universal dos Direitos Humanos, além de ser um direito positivado no

ordenamento jurídico brasileiro. É nesse contexto que se aponta a relevância do estudo e incentivo à produção de energia solar [2].

Atualmente países ao redor mundo buscam formas alternativas e renováveis de energia, com a intenção de diversificar sua matriz energética e suficiência. Prática que deveria ser adotada pelo Brasil, uma vez que estudos indicam que as dificuldades de abastecimento

de água e geração de energia se intensificarão nos próximos anos [1].

Além disso, o Brasil apresenta uma alta capacidade de gerar energia a partir do sol, com valores de 1550 e 2350 kWh m⁻² ano⁻¹, em específico as regiões Nordeste (5,9 kWh m⁻² ano⁻¹) e Centro-Oeste (5,7 kWh m⁻² ano⁻¹) [1].

Em consonância com essa perspectiva a Universidade Federal de Goiás desenvolveu o Programa de Eficiência Energética, criado em 2018, que obteve investimentos da Agência Nacional de Energia Elétrica (Aneel) revertidos na instalação de placas solares para a produção de energia fotovoltaica [3].

Dessa maneira, este artigo tem como objetivo analisar os dados obtidos pelos módulos solares instalados na Faculdade de Engenharia Mecânica, Elétrica e da Computação, no Campus Colemar Natal e Silva da Universidade Federal de Goiás e com eles aplicar a técnicas de *machine learning*.

Portanto, com o uso de algoritmos aplicados em *machine learning* objetiva-se que o algoritmo seja capaz de correlacionar os dados obtidos pelos módulos fotovoltaicos com os dados meteorológicos e assim, ser capaz de prever a produção de energia das placas solares.

II. GERAÇÃO FOTOVOLTAICA

O Sol é uma das principais fontes de energia para o nosso planeta, a energia proveniente dele possui um grande potencial de utilização, como a energia térmica e elétrica [1].

A energia liberada pelo sol é resultado de reações termonucleares, em que quatro prótons são fundidos em um núcleo de hélio. Estima-se que a taxa de energia gerada pelo sol é constante há bilhões de anos, cujo a potência atual está na casa dos 3,86x10²⁶W [4] [5].

Essa energia chega à Terra por meio de radiação eletromagnética; 81% da energia que chega à Terra está em uma faixa de comprimentos de onda que vai do visível ao

infravermelho próximo. Essa energia alimenta vários processos térmicos, dinâmicos e químicos.

Dentre os processos naturais podemos citar processos como a fotossíntese, ciclo hidrológico e a dinâmica da atmosfera e oceanos. Quanto aos processos tecnológicos podemos citar o aquecimento global, a geração de eletricidade e a climatização do ambiente. Toda essa energia transportada pelo espectro de radiação pode ser captada na forma de calor e luz [6].

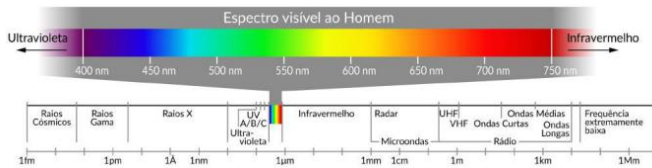


Fig. 1. Espectro da radiação solar incluindo um detalhamento da faixa visível humana [5].

A energia solar recebida pela terra através da radiação solar é captada e convertida em energia elétrica através de células fotovoltaicas; essas células recebem a radiação solar e conseguem gerar uma diferença de potencial, que quando agrupadas em uma placa solar é capaz de gerar um potencial elétrico. Para que isso seja possível essa célula é projetada com base no efeito fotoelétrico. O efeito fotoelétrico foi descoberto pelo físico francês Edmond Becquerel, em 1839, consistente no aparecimento de uma diferença de potencial nos extremos de um material semiconductor, produzido pela absorção da luz [7].

A primeira célula fotovoltaica em funcionamento foi produzida em 1954 e foi muito utilizada pelo programa espacial americano, em razão da sua vantajosa relação entre potência e peso por unidade. Após aperfeiçoada para ser usada no espaço, a tecnologia passou a ser utilizada em aplicações terrestres [7].

As células fotovoltaicas são constituídas por um material semiconductor, composto por duas faces; uma com tendência a doar elétrons ou com a tendência de receber elétrons, porém para que essa passagem de elétrons ocorra é necessária uma energia de ativação que é fornecida pelos fótons oriundos da radiação solar. Com isso teremos um material polarizado, em que a face com acúmulo de elétrons é o polo negativo e a face que perde elétrons passa a ser o polo positivo [7].

O material semiconductor utilizado é composto por uma face de silício dopado com impurezas do tipo *n*, com fósforo, fazendo com que um elétron fique desemparelhado e a outra face com silício do tipo *p*, com boro, criando uma lacuna como pode ser visto na imagem abaixo [7].

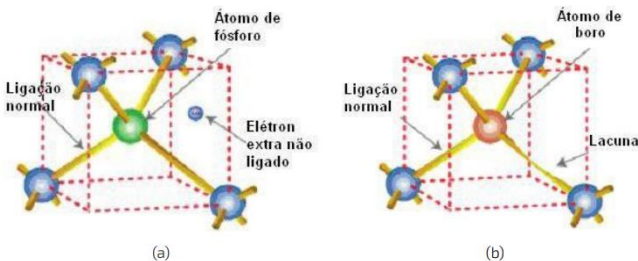


Fig. 2. Silício dopado com impurezas do tipo *n*, com fósforo (a) e silício do tipo *p*, com boro (b) [7].

O silício com impurezas do tipo *n* entra em contato com o silício do tipo *p*, e os elétrons livres do lado *n* preenchem as lacunas do lado *p*. A camada *n*, que perdeu elétrons fica positivamente carregada, enquanto a camada *p*, que recebeu elétrons, fica negativamente carregada. Essas cargas aprisionadas dão origem a um campo elétrico permanente, que dificulta a passagem de mais elétrons do lado *n* para o lado *p*. O equilíbrio é atingido quando o campo elétrico forma uma barreira de potencial capaz de barrar os elétrons livres remanescentes do lado *n*. [7]

O efeito fotoelétrico ocorre nessas condições, pois quando um elétron do lado *p* recebe energia suficiente do fóton da radiação solar e se move para a banda de condução, criando o par elétron-lacuna, o campo elétrico permanente envia o elétron para o lado *n*, não permitindo sua volta, ao mesmo tempo que repele a lacuna para o extremo do lado *p*. Adicionando contatos em suas extremidades do material semiconductor dopado o elétron da camada *n*, passa pela carga elétrica externa e retorna para a lacuna na camada *p*. Como exemplificado na figura abaixo: [7]

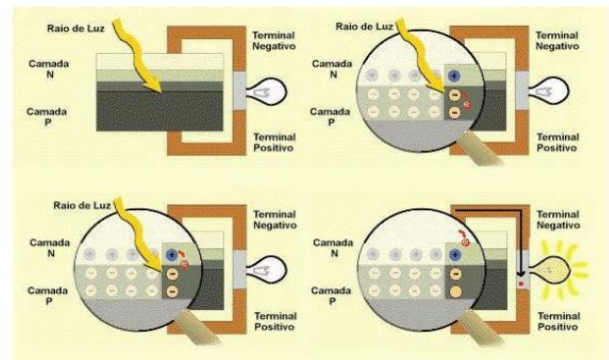


Fig. 3. Efeito fotoelétrico na junção pn [7].

A energia gerada pela célula solar é a que o elétron recebe do fóton para migrar da camada *p* para *n*. Porém nem toda energia é aproveitada. Os fótons fornecidos pela radiação solar devem ter energia suficiente para migrar o elétron da banda de valência para a banda de condução. Quando essa energia não é suficiente, ela passa direto por meio do material semiconductor; quando a energia do fóton é superior ao necessário para migração o excesso é convertido em calor. Essas duas situações representam as perdas de eficiência que podem ocorrer na conversão da energia solar em eletricidade [7].

Existem fatores que diminuem o rendimento de energia. O silício, por ser um material brilhante, pode refletir mais de 30% da luz que incide sobre ele, o que diminui a quantidade de fótons que são absorvidos. Duas técnicas são aplicadas para minimizar esse fator.

A primeira consiste em aplicar uma fina camada de monóxido de silício na superfície da célula, que reduz a reflexão em 10% e uma segunda camada pode reduzir a reflexão para apenas 4%. A segunda técnica consiste em fazer a superfície ter formato de pirâmide ou cone para que a luz refletida incida novamente sobre a célula aumentando a absorção. Como pode ser visto na imagem abaixo: [7]

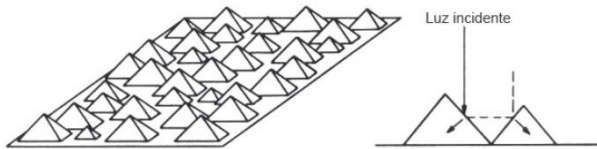


Fig. 4. Espectro da radiação solar incluindo um detalhamento da faixa visível humana [7].

Outro fator relevante relacionado ao efeito fotoelétrico é a irradiância. A irradiância pode ser entendida como uma grandeza utilizada para quantificar a radiação solar, também denominada de irradiação, expressa na unidade de W/m^2 (Watt por metro quadrado), ou seja, considera a potência por área.

Portanto, considerando que a potência é uma grandeza física capaz de expressar a energia que é transportada por um período ou sua variação, pode-se considerar que o aumento da potência da radiação solar, conseqüentemente mais energia será transportada em um intervalo de tempo [6].

Considera-se que a irradiância da luz solar é geralmente em torno de $1000 W/m^2$ e esse é o valor padrão para especificar e avaliar as células fotovoltaicas utilizado pela indústria fotovoltaica, uma vez que será possível analisar sua eficiência com base em uma condição padrão de radiação solar [6].

A irradiância será medida por um sensor e os dados obtidos serão armazenados ao longo do dia – é o modo mais comum de quantificar a energia recebida pelo Sol nos sistemas fotovoltaicos – assim, será possível calcular a quanta energia solar em certa área.

Dessa maneira, a irradiância varia conforme modifica a incidência dos raios solares no decorrer do dia. Ao analisar esses dados em um gráfico, obtém-se o valor da energia recebida durante o dia na unidade de área, esse valor é denominado de insolação.

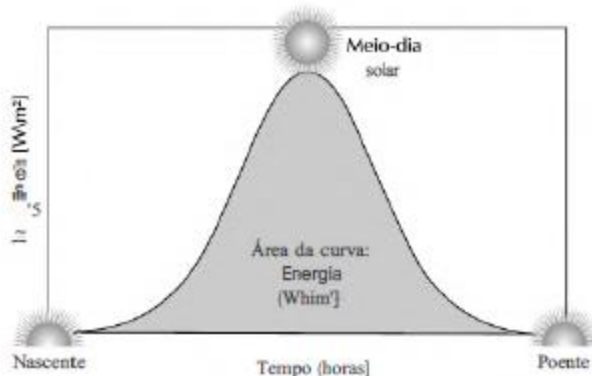


Fig. 5. Perfil de irradiância solar ao longo do dia [6]

Dessa maneira, considera-se a insolação como uma grandeza usada para expressar a energia solar que incide sobre uma determinada área de superfície plana ao longo de determinado intervalo de tempo; sua unidade é Wh/m^2 (watt-hora por metro quadrado). Essa medida é utilizada para realizar o dimensionamento dos sistemas fotovoltaicos e esses dados podem ser usados para compreender a radiação solar em determinada localidade e em consequência sua capacidade de gerar energia solar [6].

Necessário ressaltar que a insolação em determinado local varia ao longo do ano em razão dos diferentes níveis de radiação solar nas estações do ano, da ocorrência de chuvas e da presença ou não de nuvens no céu.

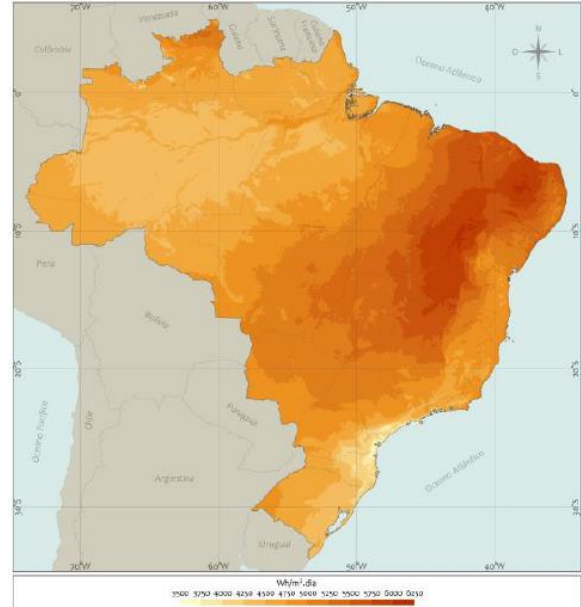


Fig. 6. Total diário de insolação global horizontal, média anual [5].

As grandezas apresentadas anteriormente são fundamentais para compreender como seria um bom funcionamento dos módulos fotovoltaicos, relação demonstrada a seguir.

A célula fotovoltaica – parte integrante dos módulos fotovoltaicos – fornecem uma tensão elétrica média de $0,6 V$, para obter níveis de tensão adequados as células são conectadas em série, com uma tensão resultante igual a soma das tensões individuais das células.

A corrente a ser produzida dependerá da área da célula, pois está intrinsecamente ligada a quantidade de luz recebida, assim, são grandezas proporcionais; quanto maior a área, maior será a quantidade de corrente elétrica gerada. Nos módulos comerciais a corrente elétrica varia entre $8 A$ (módulos cristalinos) e $2 A$ (módulos de filmes finos) [6] [4].

Cabe ressaltar que o módulo fotovoltaico não atua como uma fonte elétrica convencional, uma vez que não apresenta uma tensão de saída constante, pois depende da corrente e vice-versa. A relação entre a corrente elétrica e tensão pode ser percebida na curva do gráfico abaixo:

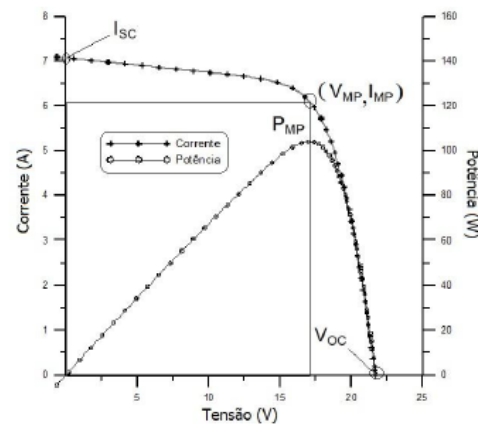


Fig.7 Curva característica $I-V$ e curva de potência $P-V$ para um módulo de potência nominal de 100 Wp [4].

O gráfico demonstra que a cada ponto na curva $I-V$, o produto corrente-tensão representa a potência gerada naquela condição. Também mostra uma curva de potência em função da tensão, chamada de curva $P-V$, que identifica o ponto com o máximo valor da potência [4].

Ademais, o módulo usualmente é identificado pela sua potência elétrica de pico (Wp) que é analisada nas condições-padrão de ensaio (STC – *Standard Test Conditions*), quais são: irradiância solar de 1000 W/m^2 sob uma distribuição espectral padrão para AM 1,5 e temperatura da célula de 25°C [4].

Portanto, conclui-se que a maioria das características de um módulo dependem da irradiância solar e da temperatura.

À medida que a irradiância solar aumenta a corrente elétrica gerada pelo módulo também aumenta. Já a corrente de curto-circuito tem um aumento linear em relação a grandeza mencionada [4].

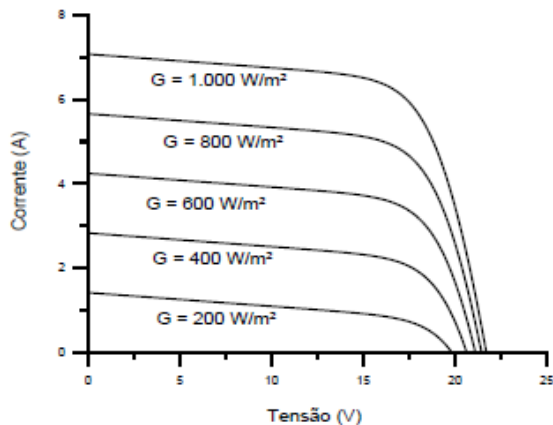


Fig. 8 Efeito causado pela variação da irradiância solar sobre a curva característica $I-V$ para um módulo fotovoltaico de 36 células de silício cristalino (c-Si) a 25°C [4].

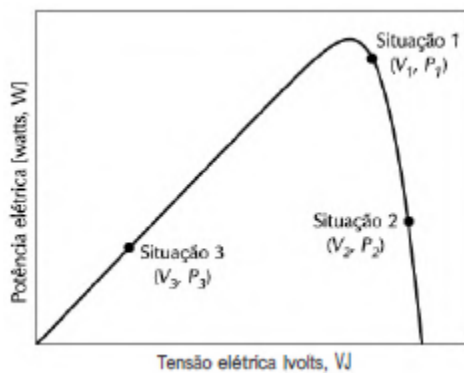


Fig. 9 Pontos de operação do módulo fotovoltaico ao longo da curva $P-V$ para as três situações de operação [6].

Pode-se afirmar que com uma irradiância solar de 1000 W/m^2 o módulo possui a capacidade de fornecer a corrente máxima determinada em seu catálogo, considerando a temperatura de

25°C . Portanto, a corrente está intimamente ligada a quanta luz solar o módulo recebe [6].

Outro fator relevante a ser analisado é a temperatura ambiente, pois ela ocasiona uma variação na temperatura das células que compõe os módulos. O aumento da temperatura da célula ocasiona uma queda relevante na tensão; apesar da corrente elétrica apresentar um aumento, ele é inexpressivo e não é suficiente para suprir a perda causada pela diminuição da tensão [4].

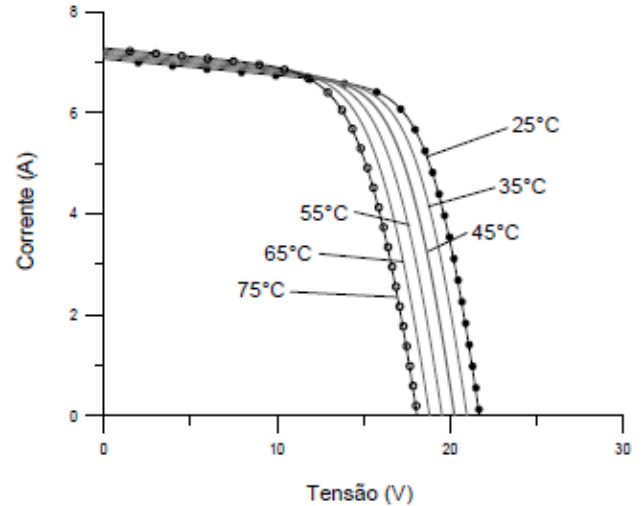


Fig. 10 – Efeito causado pela variação da temperatura das células sobre a curva característica $I-V$ para um módulo fotovoltaico de 36 células de silício cristalino (c-Si) sob irradiância de 1000 W/m^2 [4].

Além disso, em razão da diminuição da tensão devido o aumento da temperatura, a potência fornecida pelo módulo também sofrerá uma queda, uma vez que a potência é o produto da tensão pela corrente do módulo [6].

Devido as condições-padrão de ensaio (STC) não representarem geralmente as condições de operação real, as normas determinam uma temperatura nominal para a operação das células nos módulos, de forma que as características elétricas estejam mais próximas as observadas na prática em campo.

Essa temperatura nominal será obtida com a exposição do módulo a um circuito aberto com irradiância de 800 W/m^2 com uma temperatura de 20°C e com ação do vento com velocidade de 1 m/s . Ela comumente é encontrada nas folhas de dados técnicos dos módulos, identificada pela sigla NOCT (*Nominal Operating Cell Temperatur*) e geralmente está entre a 40 a 50°C .

A NOCT está relacionada às propriedades térmicas e ópticas nos materiais empregados na construção do módulo, assim, se os módulos apresentam os mesmos coeficientes de temperatura (α , β e γ)¹, aquele com a menor NOCT terá melhor desempenho em campo, já que apresentará uma perda menor relacionada à temperatura [4].

Diante o exposto nesse tópico, evidente a capacidade energética proveniente da energia solar. A energia produzida advém do efeito fotoelétrico que ocorre nas células que compõe

¹ Coeficiente α refere-se a variação da corrente de curto-circuito com a temperatura; Coeficiente β refere-se a variação da tensão de circuito aberto com

a temperatura; e Coeficiente γ refere-se a variação da potência máxima do módulo com a temperatura.

os módulos fotoelétricos. A geração de energia elétrica por essas células está diretamente relacionada a irradiância e a temperatura, fatores que serão de suma importância para o estudo proposto neste artigo.

III. MACHINE LEARNING

Na computação escreve-se programas capazes de realizar as mais diversas tarefas por meio de algoritmos; porém, existem tarefas que são simples para seres humanos, mas complexas para se descrever em um algoritmo eficiente, tal como reconhecer as diferenças de um gato e um cachorro através de uma imagem [8].

Com a crescente aumento no volume de dados gerados por diferentes setores, surgiu a necessidade de criar ferramentas computacionais mais sofisticadas, que fossem mais autônomas e que não dependessem de especialistas; o que ocorria até então com o desenvolvimento de sistemas especialistas ou sistemas baseados em conhecimento. Essas ferramentas deveriam ser capazes de criar por si – a partir da experiência passada – uma hipótese, ou função que seja capaz de resolver o problema que se deseja tratar [8].

Em *Machine Learning*, computadores são programados para aprender com base em experiências passadas, aplicando, por exemplo, o princípio de inferência, também denominado indução, no qual se obtém conclusões genéricas a partir de um conjunto particular de exemplos. Assim, algoritmos de *Machine Learning* aprendem a induzir uma função ou hipótese capaz de resolver um problema com base nos dados que representam instâncias do problema a ser resolvido [8].

A. Indução de Hipóteses

Suponha um conjunto de dados composto de informações de pacientes de um hospital. Esse conjunto possui características dos pacientes, como por exemplo: identificação, nome, idade, sexo, estado de origem, sintomas e resultados de exames clínicos. [8]

Um algoritmo de *Machine Learning* é utilizado para aprender uma hipótese – uma regra – capaz de diagnosticar pacientes de acordo com seus atributos de entrada; atributos como: identificação e nome não são considerados entradas relevantes, pois não possuem nenhuma relação com o diagnóstico de uma doença [8].

O desejado é induzir uma hipótese capaz de fazer diagnósticos corretos para novos pacientes, diferentes daqueles que foram utilizados para aprender a regra de decisão. Sendo assim é esperado que a regra seja válida para outros objetos de mesmo domínio ou problema que não fazem parte do conjunto de treinamento; essa capacidade é chamada de generalização da hipótese [8].

Quando uma hipótese apresenta uma baixa capacidade de generalização, pode ser por estar super ajustada aos dados, denominado de *overfitting*; nesse caso a hipótese se especializou nos dados de treinamento. Também pode ocorrer o caso inverso, o *underfitting*, quando a hipótese obtém uma baixa taxa de acerto mesmo ao testar com os dados de treinamento; isto pode ocorrer quando os exemplos de treino

são pouco representativos ou o modelo usado é muito simples e não captura os padrões existentes dos dados [8].

B. Viés Indutivo

Algoritmos de *Machine Learning* tentam aprender a partir dos dados de treinamento, buscam uma hipótese, em um espaço de possíveis hipóteses, capaz de descrever as relações entre os objetos que melhor se ajusta aos dados de treinamentos [8].

Quanto a análise das hipóteses, existem dois tipos de vieses: o viés de representação, que restringe as hipóteses que podem ser geradas; e o viés de busca, que define como as hipóteses são pesquisadas no espaço das hipóteses. Sem o viés não haveria generalização.

Apesar do viés parecer uma limitação, sua ausência em um algoritmo de *Machine Learning* o impossibilitaria de generalizar o conhecimento adquirido durante seu treinamento e aplicá-lo com sucesso a novos dados.

Cada algoritmo possui uma forma de induzir estas hipóteses. As redes neurais artificiais representam uma hipótese por um conjunto de valores reais associados aos pesos das conexões da rede.

Árvores de decisão utilizam uma estrutura de árvore em que cada nó interno é representado por uma pergunta referente ao valor de um atributo e cada nó externo está associado a uma classe. A figura abaixo apresenta três tipos de vieses de representação [8].

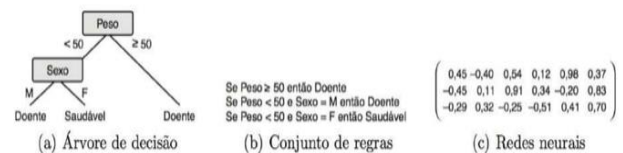


Fig. 9. Diferentes vieses de representação [8].

C. Tarefas de aprendizado

Algoritmos de *Machine Learning* podem ser divididos em dois tipos de tarefas de aprendizado: as **preditivas** e **descritivas** [8].

Tarefas preditivas são aquelas que a meta é encontrar uma hipótese a partir dos dados de treinamento que possa ser utilizada para prever um rótulo ou valor. Esses algoritmos seguem o paradigma de aprendizado supervisionado, ou seja, em seu conjunto de dados possuem a saída que é esperada [8].

Por sua vez, as **tarefas descritivas** são aquelas que buscam descrever um conjunto de dados. Os algoritmos utilizados nessas tarefas não fazem uso do atributo de saída, nesse caso seguem o paradigma de aprendizagem não supervisionada, geralmente são utilizados em uma tarefa de agrupamento de dados, por exemplo [8].



Fig. 10. Hierarquia de aprendizado [8]

IV. CORRELAÇÃO DE PEARSON

O coeficiente de correlação de Pearson é uma forma de medir a relação linear entre duas variáveis que é dada pela fórmula [9]:

$$r = \frac{1}{n-1} \sum \frac{(x_i - \bar{X})(y_i - \bar{Y})}{sx \ sy}$$

Em que:

- n é o número de amostras;
- x_i é a amostra da variável x ;
- \bar{X} é a média da variável x ;
- sx é o desvio padrão da variável x ;
- y_i é a amostra da variável y ;
- \bar{Y} é a média da variável y ;
- sy é o desvio padrão da variável y ;
- r é o coeficiente de correção de Pearson.

Em termos estatísticos duas variáveis se relacionam quando possuem semelhança na distribuição de seus escores, ou seja, elas podem se associar a partir das distribuições ou pelo compartilhamento de variância. A correlação de Pearson é uma medida da variância compartilhada entre duas variáveis [9].

Para um modelo linear se espera que as alterações na variável X também gerem o mesmo impacto na variável Y . Por relação linear, a melhor forma de se ilustrar graficamente um relacionamento de duas variáveis é através de uma linha reta. Portanto para a correlação de Pearson é necessário um compartilhamento de variância e que a variância seja distribuída linearmente [9].

O coeficiente de correção de Pearson (r) varia de -1 a 1. O valor representa a força do relacionamento entre as variáveis. Para esses valores podemos montar uma tabela que defina qual a afinidade entre as duas variáveis [9].

O significado prático varia de autor para autor, abaixo são apresentados dois entendimentos de autores diferentes.

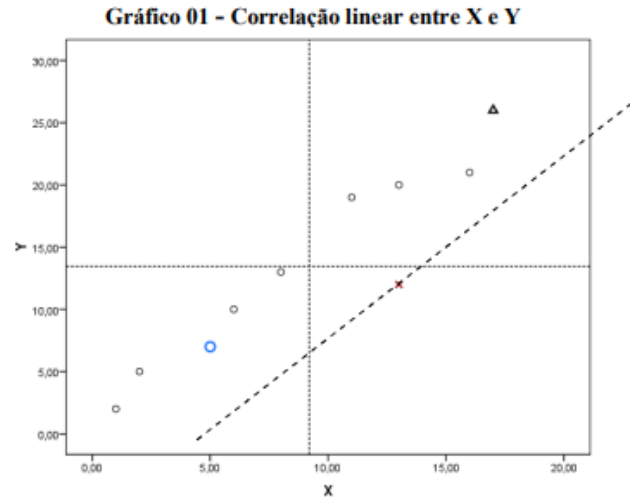
Coefficiente	Relação
0,1-0,29	Pequena
0,39-0,49	Média
0,5-1,0	Grande

(BRITO *apud* Cohen 1998) [9].

Coefficiente	Relação
0,1-0,3	Fraca
0,4-0,6	Moderado
0,7-1	Forte

(BRITO *apud* Dancey e Reidy 2005) [9].

Independente do autor é correto afirmar que quanto mais próximo de um – independente do sinal - maior será a dependência estatística linear entre as variáveis, enquanto para valores mais próximo de zero, mais fraca a relação. Abaixo é apresentado um exemplo de relação linear positiva entre duas variáveis hipotéticas X e Y [9].

Fig. 11. Gráfico relação linear hipotética entre X e Y [9].

V. MÉTODO SVM E SVR

O SVM (*Support Vector Machine*) é um algoritmo que busca uma separação entre classes distintas, analisando um ponto de cada grupo mais próximo da outra classe. O algoritmo escolhe uma reta - ou hiperplano em maiores dimensões - entre dois grupos que se distanciam mais de cada um [10] [11] [12].

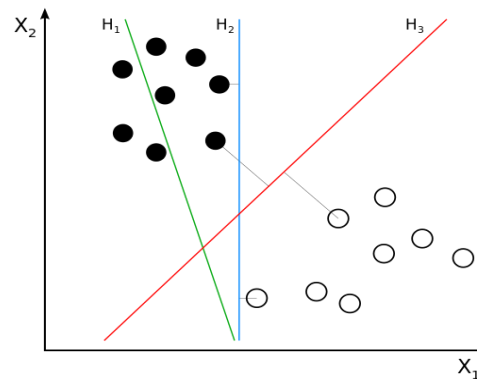


Fig. 12. A melhor reta é a H3, que possui a melhor distância entre os dois grupos [12].

A. SVM não linear

Existem casos em que não é possível separar as classes por uma reta, esses casos se encaixam no método SVM não linear. Neste caso a busca do algoritmo é em traçar mais retas e curvas a fim de separar as classes da melhor forma possível [11] [12].

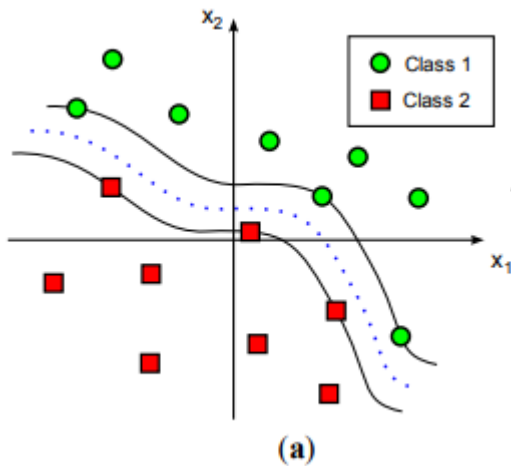


Fig. 13. SVM não linear [12]

SVM's lidam com problemas não lineares, mapeando o conjunto de treinamento de seu espaço original – referenciado como entradas – para um novo espaço de maior dimensão, denominado espaço de características – *feature space*. Problemas que não podem ser resolvidos com SVM linear no espaço original, a depender das entradas de treinamento, podem ser separadas de maneira linear no espaço de características [8].

O uso desse procedimento é motivado pelo teorema de Cover (CARVALHO *apud* Haykin, 1999) [8]. Dado um conjunto de dados não lineares no espaço de entradas X , esse teorema afirma que X pode ser transformado em um espaço de características no qual, com alta probabilidade, os objetos são linearmente separáveis. Duas condições devem ser satisfeitas para isso: a transformação seja não linear e a dimensão do espaço de características seja suficientemente alta [8].

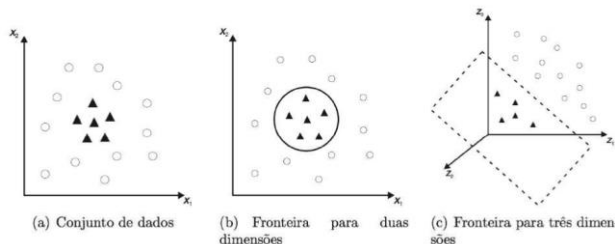


Fig. 14. Exemplo de transformação realizada em conjunto de dados não linear para o espaço de características. [8]

B. Funções Kernel

Como o espaço de características pode ter uma dimensão muito ampla, o custo computacional para realizar o processo pode ser extremamente alto ou inviável. A fim de evitar esse problema são utilizadas as funções *kernel*. [8]

A função *kernel* tem como objetivo receber dois objetos do espaço das entradas e calcular o produto escalar desses objetos no espaço de características. As funções *kernel* mais comuns são: Linear, Polinomial e RBF [8].

C. SVM em Problemas de Regressão (SVR)

O SVM também pode ser utilizado para prever valores contínuos com base nos dados, ao invés de prever a classe que os dados pertencem. [12]

Para regressão o objetivo é encontrar um hiperplano que mais se aproxime dos dados. Diferente da classificação que considera os pontos mais próximos ao hiperplano, na regressão são considerados os pontos mais distantes. Para casos não lineares é utilizado a mesma técnica de transformação do SVM não linear para fazer uma regressão não linear [12].

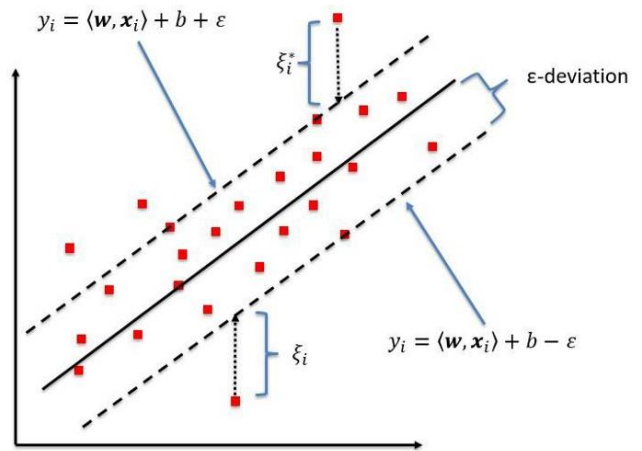


Fig. 15. Regressão por SMV (COUTINHO *apud* Kleynhans) [12]

D. Métricas

Em problemas de previsão, seja ela classificação ou regressão, é importante se ter uma métrica de quanto o seu treinamento está performando, para entender se a acurácia do seu treinamento está boa o suficiente, ou qual é o treinamento que desempenha o menor resultado. Neste sentido serão abordadas algumas métricas para validar treinamentos de regressão.

1) Max Error (ME)

Essa métrica busca encontrar o maior erro possível dentre todas as amostras apresentadas, isto é, serão calculadas as diferenças entre os valores esperados e o valor previsto e será escolhido o maior valor - em módulo - de todo o treinamento [13].

$$MaxError(y, \hat{y}) = \max(|y_i - \hat{y}_i|)$$

Em que:

- y é uma sequência de valores reais;
- \hat{y} é uma sequência de valores previstos;
- \max é uma função que retorna o maior valor de uma sequência de valores;
- y_i é um elemento de y
- \hat{y}_i é um elemento de \hat{y}

2) Mean absolute error (MAE)

Essa métrica realiza a média da soma da diferença absoluta entre o valor esperado e o previsto [13].

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

Em que:

- y é uma sequência valores reais;
- \hat{y} é uma sequência valores previsto;
- y_i é um elemento de y
- \hat{y}_i é um elemento de \hat{y}
- n é número total de amostras.

3) Mean squared error (MSE)

Essa métrica se assemelha a *MEA*, todavia, nessa a diferença entre o valor esperado e o previsto é elevada ao quadrado. [13]

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

Em que:

- y é uma sequência valores reais;
- \hat{y} é uma sequência valores previsto;
- y_i é um elemento de y
- \hat{y}_i é um elemento de \hat{y}
- n é número total de amostras.

4) Root Mean squared error (RMSE)

Essa métrica resolve um problema criado pela métrica anterior, *MSE*, em que a unidade comparada estaria ao quadrado. Para isso é tirado a raiz quadrada do resultado da métrica *MSE*. [13]

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

Em que:

- y é uma sequência valores reais;
- \hat{y} é uma sequência valores previsto;
- y_i é um elemento de y
- \hat{y}_i é um elemento de \hat{y}
- n é número total de amostras.

Apesar da métrica *MAE* apresentar grandes semelhanças com a *RMSE*, o fato desta última apenas tirar a raiz do resultado acentua grandes erros; todavia, a *RMSE* é uma métrica importante para sistemas em que erros grandes são indesejáveis.

VI. DADOS DE GERAÇÃO E CLIMÁTICOS

Os dados utilizados neste estudo foram coletados na Estação Meteorológica da EMC/UFG obtidos através da geração de energia elétrica das placas fotovoltaicas instaladas.

A. Estação Meteorológica da EMC/UFG

A Estação Meteorológica da EMC/UFG está localizada na Escola de Engenharia Elétrica Mecânica e de Computação (EMC) da Universidade Federal de Goiás (UFG) sobre o telhado do Centro de Aulas B. Está em operação desde 27 de julho de 2015, sendo capaz de realizar as seguintes medições:

- Irradiância solar global (W/m²) no plano horizontal, através de um piranômetro da marca Hukseflux, modelo LP02. Possui faixa espectral de 285 nm a 3000 nm, faixa de medição de 0 a 2000 W/m².
- Temperatura do ar (°C) e umidade relativa do ar (%) através de um sensor da marca Campbell, modelo CS2. As medições de temperatura possuem uma resolução de 0,01 (°C), exatidão de ± 0,4°C na faixa de 5 a 40 °C e exatidão de ± 0,9°C na faixa de -40 a 70 °C. Quanto às medições de umidade a resolução é de 0,03 % ± 4% na faixa de 0% a 100%.
- Índice de precipitação pluviométrica (mm), através de um pluviômetro da marca Texas Electronics, modelo TR25I. Possui resolução de 0,2 mm e precisão de 1 % até 50 mm/hr.
- Velocidade do vento (m/s) e direção do vento (°), através de anemômetro da R M Young, modelo 03002-5. A velocidade do vento possui exatidão de ± 0,5 m/s.
- Pressão atmosférica (mBar), através de um sensor da marca Texas Electronics, modelo TB-2012M. Possui exatidão de ± 1,3 mBar. [14]



Fig. 16. Estação meteorológica instalada na EMC/UFG e seus sensores [14].

Os dados são armazenados em um registro central de baixo consumo de energia – *data logger* – que possuía uma capacidade de 8 dias de medição em 2015, tendo sido ampliada,

com a troca do *data logger* ocorrida em abril de 2019, para 90 dias. As medições são realizadas com intervalos de 1 em 1 minuto [14].

B. Sistema Fotovoltaico da EMC/UFG

Como mencionado anteriormente, o sistema fotovoltaico da EMC/UFG está localizado na Escola de Engenharia Elétrica Mecânica e de Computação (EMC) da Universidade Federal de Goiás (UFG) sobre o telhado do Centro de Aulas B. Foi instalado seguindo a orientação do telhado, com 10° de inclinação, parte dos módulos com -15° de desvio azimutal e outra parte com 165° de desvio azimutal, ocupando uma área total de 217 m² [14].

Atualmente o sistema é composto por 133 módulos de silício policristalino da marca SUN EARTH. Cada módulo possui uma potência de 235 Wp e o sistema completo possui uma potência instalada de 31,25 kWp [14].

O telhado onde estão instalados os módulos, são apresentadas duas águas – norte e sul – separadas ao meio por uma plataforma que as divide em leste e oeste. Estão instalados 60 módulos na água nordeste (NE) e 60 módulos na água sudeste (SE) [14].



Fig. 17. Sistema Fotovoltaico EMC/UFG – Lado Leste. [14]

Outros 25 módulos estão instalados na água noroeste (NO), porém não existem módulos instalados na água sudoeste (SO) do telhado [14].



Fig. 18. Sistema Fotovoltaico EMC/UFG – Lado Oeste. [14]

São utilizados sete inversores da marca ELTEK -THEIA – seis com potência de 4,4kW e um com 2,9 kW [14].

Os dados de geração são obtidos, desde o dia 14 de outubro de 2016, através de um monitoramento fornecido pelo fabricante ELTEK. Os dados representam os valores de energia produzida na saída dos inversores [14].

VII. TRATAMENTO DE DADOS

Para esse estudo os dados foram tratados removendo dados considerados como medições erradas e posteriormente foram agrupados a cada uma hora e correlacionados pelo horário.

A. Estação Meteorológica da EMC/UFG

Como apresentado anteriormente, a estação meteorológica é capaz de fornecer cinco tipos de medidas meteorológicas.

Como os dados são obtidos de minuto em minuto, foram agrupados de hora em hora e utilizou-se apenas a média de cada medida para o treinamento.

B. Sistema Fotovoltaico da EMC/UFG

Para os dados de geração utilizou-se apenas a medição de potência ativa acumulada e a temperatura do módulo. Os dados obtidos apresentam uma periodicidade de 15 (quinze) minutos.

Dados que possuíam geração em horários noturnos foram desconsiderados.

Para que a previsão seja feita em kWh, os dados foram somados de hora em hora para serem utilizados no treinamento.

VIII. VARIÁVEIS

Apesar de literaturas apontarem que as variáveis que determinam a geração de energia são apenas a irradiância e a temperatura. Neste trabalho foram analisadas as demais variáveis meteorológicas, como índice pluviométrico e umidade relativa. Ademais, foi analisado como seus valores influenciam na geração de energia.

A. Irradiância

Como apresentado anteriormente, a irradiância é a variável principal na geração de energia, pois através dela que os

módulos solares são excitados, resultando na geração de energia.

Analisou-se, com base nos dados colhidos, o comportamento da geração de energia baseado na irradiância, através de um gráfico de dispersão de energia gerada (kWh) e irradiância (W/m^2)

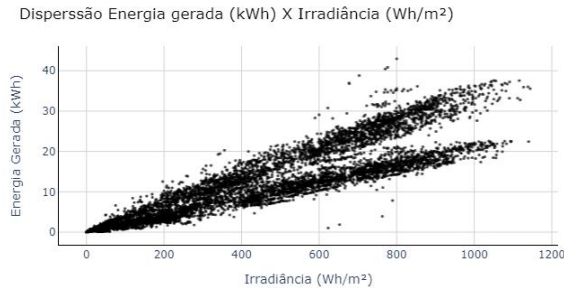


Fig. 19. Gráfico de dispersão entre as variáveis energia gerada e irradiância.

Calculando o valor do coeficiente de Pearson obtemos um valor de 0,95, que é considerada uma relação positiva forte.

B. Temperatura Ambiente

A temperatura do módulo influencia diretamente na geração de energia; quanto maior a temperatura do módulo, menor será a eficiência nessa geração. Contudo, o presente trabalho visa analisar outros dados meteorológicos, pode-se observar a relação entre a temperatura do módulo e a temperatura ambiente.

O gráfico a seguir foi compilado com base na temperatura média por horário, tanto da estação meteorológica como a informado pelo sensor dos módulos.

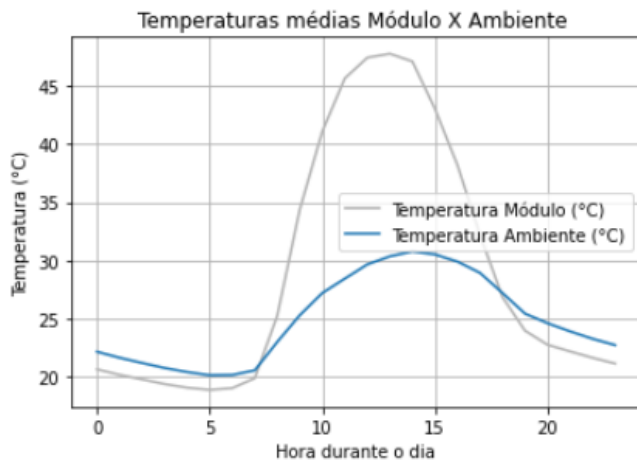


Fig. 20. Gráfico temperaturas média módulo X temperatura ambiente.

Percebe-se que existe uma relação entre os dados obtidos da temperatura do módulo e os dados obtidos da temperatura ambiente.

Foi gerado um gráfico de dispersão entre os dados de temperatura ambiente e os dados de geração de energia.

Dispersão Energia gerada (kWh) X Temperatura Ambiente (°C)

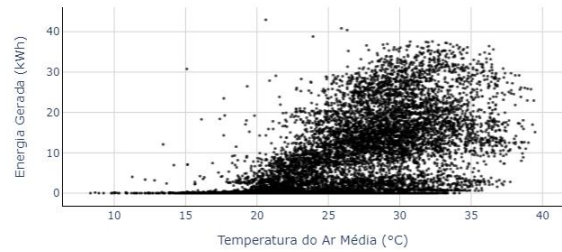


Fig. 21. Gráfico de dispersão entre as variáveis energia gerada e temperatura ambiente.

Como pode-se observar existe um comportamento semelhante ao obtido ao analisar a irradiância, o que demonstra ter uma correlação entre as variáveis com o resultado da geração de energia.

Calculando o valor do coeficiente de Pearson obtém-se um valor de 0,63 que é considerada uma relação positiva moderada.

C. Índice pluviométrico

Para a análise foi plotado novamente o gráfico de dispersão.

Dispersão Energia gerada (kWh) X Índice Pluviométrico (mm)

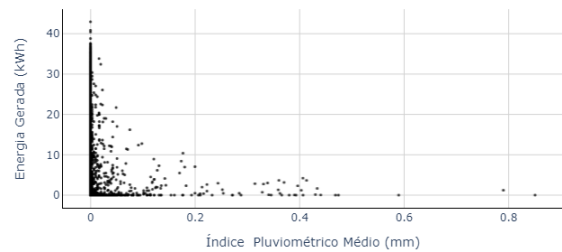


Fig. 22. Gráfico de dispersão entre as variáveis energia gerada e índice pluviométrico médio.

Como pode ser observado, o índice pluviométrico não tem uma relação clara com a geração de energia.

Ao calcular o valor do coeficiente de Pearson obtém-se um valor de -0,05, conclui-se não haver relação entre as variáveis.

D. Umidade relativa do ar

Para a análise foi plotado novamente o gráfico de dispersão.

Dispersão Energia gerada (kWh) X Umidade relativa do ar (%)

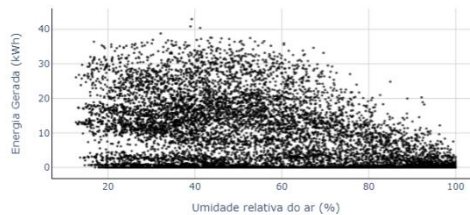


Fig. 23. Gráfico de dispersão entre as variáveis energia gerada e umidade relativa do ar.

Pela análise do gráfico pode-se considerar que a umidade relativa do ar possui certa relação com a geração de energia.

Pode-se notar que a geração de energia é mais baixa para uma umidade relativa entre 80% e 100%.

Ao calcular o valor do coeficiente de Pearson obtém-se um valor de -0,43, que é considerado uma relação negativa moderada, ou seja, é uma variável inversamente proporcional, assim, quanto menor a umidade maior será a geração de energia.

E. Escolha de variáveis

Com base nos dados analisados pode-se montar a seguinte tabela para apresentar a relação entre as quatro variáveis:

-	E	IR	T	IP	U
E	-	0,95	0,63	-0,05	-0,43
IR	0,95	-	0,67	-0,05	-0,45
T	0,63	0,67	-	-0,07	-0,67
IP	-0,05	-0,05	-0,07	-	0,14
U	-0,43	-0,45	-0,67	0,14	-

Em que:

- E – Energia gerada;
- IR – Irradiância;
- T – Temperatura Ambiente;
- IP – Índice pluviométrico;
- U – Umidade relativa do ar;

Para o treinamento do algoritmo, as variáveis escolhidas foram: **irradiância**, **temperatura ambiente** e **umidade relativa do ar**. Sendo que as que tiveram o maior efeito sobre a geração de energia foram a **irradiância** e a **temperatura ambiente**.

IX. TREINAMENTO E RESULTADOS

Para o treinamento foi utilizado a linguagem de programação *Python* com a biblioteca *scikit*; para o treinamento e cálculo de erro, a biblioteca *numpy*; e *pandas* para manipulação de dados. O *dataset* para treinamento possui 16068 registros entre os períodos de 21/05/2019 a 25/09/2021.

Ademais, foi utilizado o método SVR com os diferentes *kernels*: Linear, Polinomial e RBF (*Radial Basis Funcion*).

Os treinamentos foram divididos da seguinte maneira: utilizando uma variável (irradiância), duas variáveis (irradiância e temperatura ambiente) e três variáveis (irradiância, temperatura ambiente e umidade relativa do ar), todavia, não foi possível plotar o gráfico desta última.

Para cada treinamento foi separada 30% das amostras de forma aleatória para a realização do cálculo das métricas. Foram calculadas as seguintes métricas apresentadas no artigo: *max error* (ME), *mean absolute error* (MAE) e *root mean squared error* (RMSE).

Os treinamentos por SVR foram feitos utilizando os parâmetros ϵ , que significa o intervalo de valores que devem ser ignorados durante o treinamento igual a 1 (um), pois desta forma o treinamento tenderá a seguir melhor a distribuição da irradiância, os demais parâmetros como γ e C que define a penalidade do algoritmo, *degree* que define o grau da

função polinomial são mantidos em seus valores padrão do *scikit* para o SRV que são respectivamente 1 e 3.

A. Treinamento com irradiância com função kernel linear

O gráfico abaixo apresenta a dispersão entre as variáveis irradiância e energia gerada, com 3 (três) curvas plotadas, sendo elas a previsão dos dados por SVR com *kernel* linear e as margens ϵ .

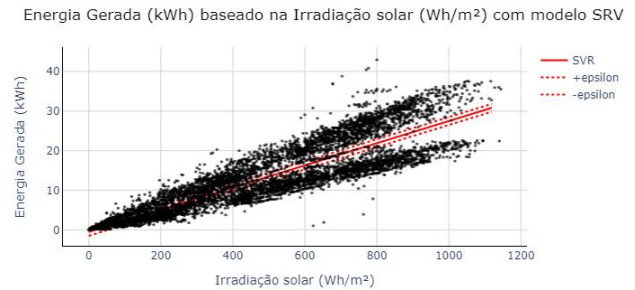


Fig. 24. Gráfico de dispersão entre as variáveis energia gerada e irradiância com curva de previsão SVR com kernel linear treinado com a variável irradiância.

Foram obtidos os seguintes valores para as métricas analisadas:

Métrica	Valor
ME	19,59
MAE	1,80
RMSE	2,82

Ao utilizar esse treinamento pode-se simular um dia arbitrário e comparar com os resultados reais:

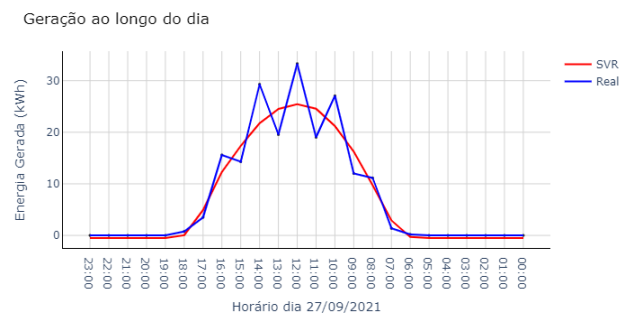


Fig. 25. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel linear treinado com a variável irradiância.

B. Treinamento com irradiância com função kernel polinomial

O gráfico da Fig. 26 apresenta a dispersão entre as variáveis irradiância e energia gerada, com 3 (três) curvas plotadas, em que elas são a previsão dos dados por SVR com *kernel* polinomial e as margens ϵ .

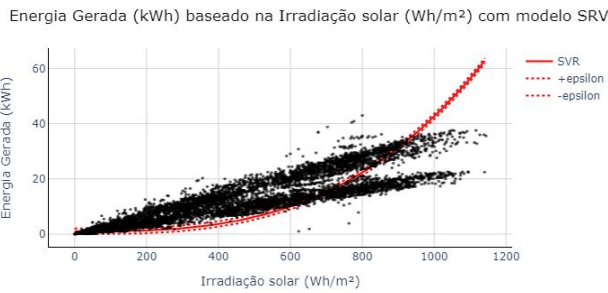


Fig. 26. Gráfico de dispersão entre as variáveis energia gerada e irradiância com curva de previsão SVR com kernel treinado com a variável irradiância.

Foram obtidos os seguintes valores para as métricas analisadas:

Métrica	Valor
ME	33,25
MAE	2,9
RMSE	4,81

Ao utilizar esse treinamento pode-se simular um dia arbitrário e comparar com os resultados reais:

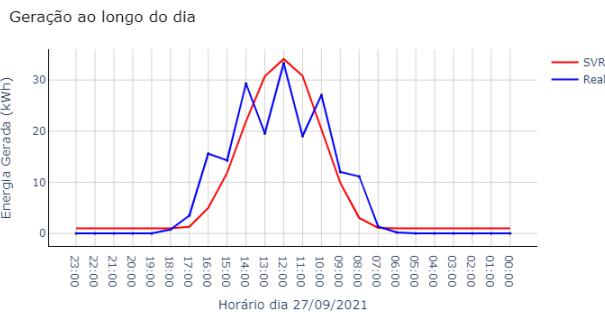


Fig. 27. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel polinomial treinado com a variável irradiância.

C. *Treinamento com irradiância com função kernel RBF*

O gráfico abaixo apresenta a dispersão entre as variáveis irradiância e energia gerada, com 3 (três) curvas plotadas, sendo elas a previsão dos dados por SVR com *kernel RBF -Radial Basis Function* - e as margens épsilons.

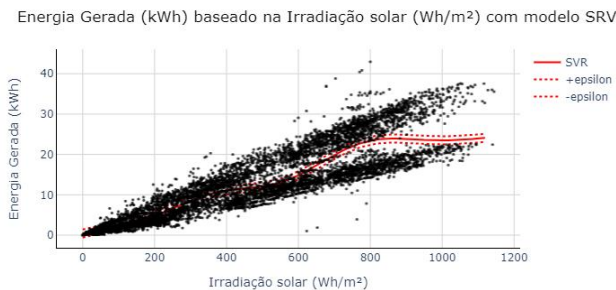


Fig. 28. Gráfico de dispersão entre as variáveis energia gerada e irradiância com curva de previsão SVR com kernel RBF treinado com a variável irradiância.

Foram obtidos os seguintes valores para as métricas analisadas:

Métrica	Valor
ME	19,56
MAE	1,73
RMSE	2,91

Utiliza-se esse treinamento para simular um dia arbitrário e comparar com os resultados reais:



Fig. 29. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel RBF treinado com a variável irradiância.

D. *Treinamento com irradiância e temperatura ambiente com função kernel linear*

O gráfico abaixo apresenta a dispersão entre as variáveis irradiância, temperatura ambiente e energia gerada, com um plano irradiância x temperatura ambiente demonstrando o hiperplano gerado pelo algoritmo SVR com *kernel linear*.

Plano de previsão

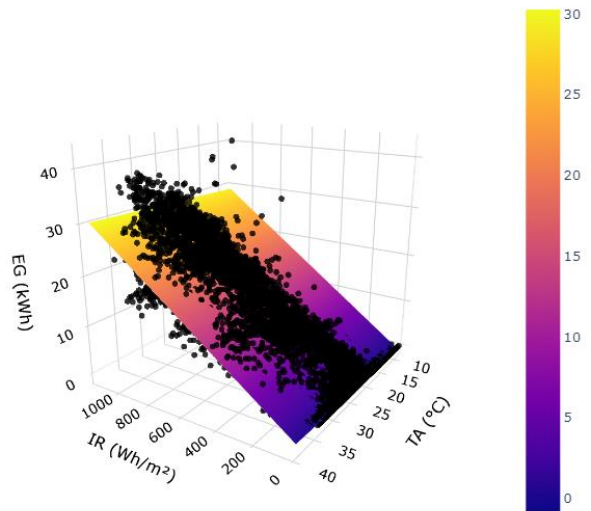


Fig. 30. Gráfico de dispersão entre as variáveis energia gerada (EG), irradiância (IR) e temperatura ambiente (TA) com plano de previsão SVR com kernel linear treinado com as variáveis irradiância e temperatura ambiente.

Foram obtidos os seguintes valores para as métricas analisadas:

Métrica	Valor
ME	20,59
MAE	1,67
RMSE	2,83

Utiliza-se esse treinamento para simular um dia arbitrário e comparar com os resultados reais:

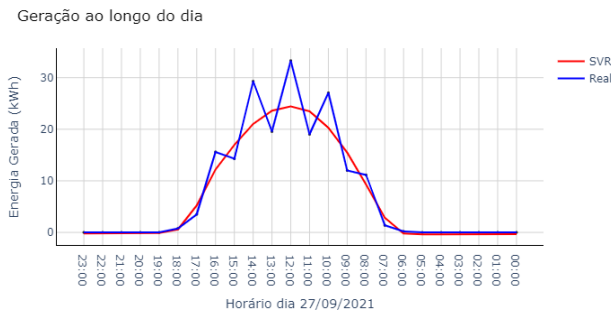


Fig. 31. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel linear treinado com as variáveis irradiância e temperatura ambiente.

E. Treinamento com irradiância e temperatura ambiente com função kernel polinomial

O gráfico da Fig. 32 apresenta a dispersão entre as variáveis irradiância, temperatura ambiente e energia gerada, com um plano irradiância x temperatura ambiente demonstrando o hiperplano gerado pelo algoritmo SVR com *kernel* polinomial. Plano de previsão

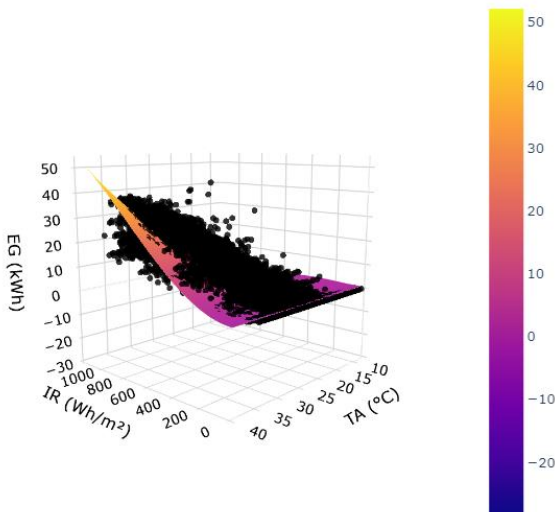


Fig. 32. Gráfico de dispersão entre as variáveis energia gerada (EG), irradiância (IR) e temperatura ambiente (TA) com plano de previsão SVR com kernel polinomial treinado com as variáveis irradiância e temperatura ambiente.

Foram obtidos os seguintes valores para as métricas analisadas:

Métrica	Valor
ME	26,40
MAE	2,31
RMSE	3,68

Utiliza-se esse treinamento para simular um dia arbitrário e comparar com os resultados reais:

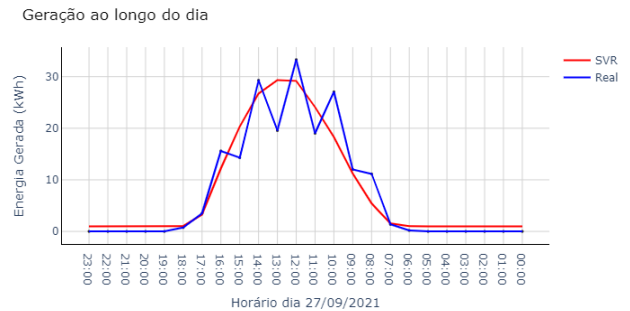


Fig. 33. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel polinomial treinado com as variáveis irradiância e temperatura ambiente.

F. Treinamento com irradiância e temperatura ambiente com função kernel RBF

O gráfico da Fig. 34 apresenta a dispersão entre as variáveis irradiância, temperatura ambiente e energia gerada, com um plano irradiância x temperatura ambiente demonstrando o hiperplano gerado pelo algoritmo SVR com *kernel* RBF. Plano de previsão

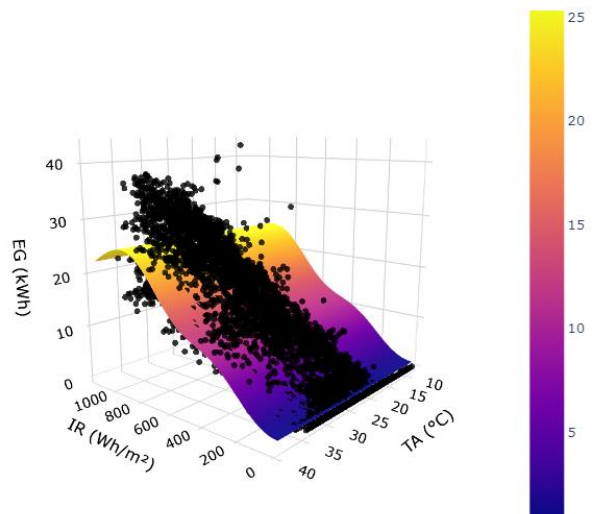


Fig. 34. Gráfico de dispersão entre as variáveis energia gerada (EG), irradiância (IR) e temperatura ambiente (TA) com plano de previsão SVR com kernel RBF treinado com as variáveis irradiância e temperatura ambiente.

Foram obtidos os seguintes valores para as métricas analisadas:

Métrica	Valor
ME	19,93
MAE	2,01
RMSE	2,90

Utiliza-se esse treinamento para simular um dia arbitrário e comparar com os resultados reais:

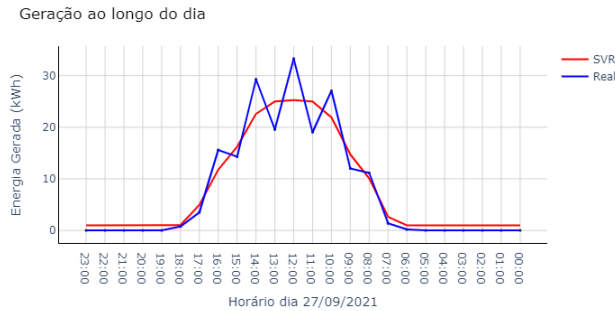


Fig. 35. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel RBF treinado com as variáveis irradiância e temperatura ambiente.

G. Treinamento com irradiância, temperatura ambiente e umidade relativa do ar com função kernel linear

Como não é possível fazer uma representação gráfica com as 3 (três) variáveis será apresentado apenas a tabela das métricas obtidas. As variáveis analisadas foram a irradiância, temperatura e umidade relativa do ar com a função *kernel* linear. Os valores obtidos foram:

Métrica	Valor
ME	22,38
MAE	1,64
RMSE	2,73

Utiliza-se esse treinamento para simular um dia arbitrário e comparar com os resultados reais:

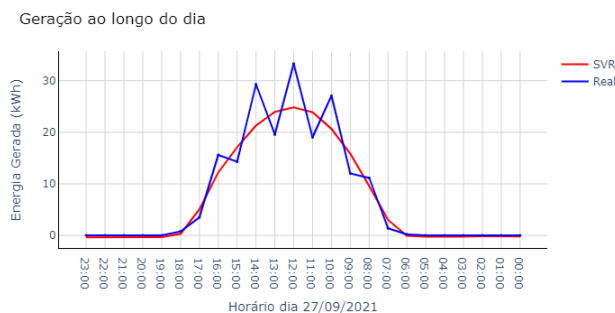


Fig. 36. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel linear treinado com as variáveis irradiância, temperatura ambiente e umidade relativa do ar.

H. Treinamento com irradiância, temperatura ambiente e umidade relativa do ar com função kernel polinomial

Como não é possível fazer uma representação gráfica com 3 (três) variáveis será apresentado apenas a tabela das métricas obtidas. As variáveis analisadas foram a irradiância, temperatura e umidade relativa do ar com a função *kernel* polinomial. Os valores obtidos foram:

Métrica	Valor
ME	22,76
MAE	1,99
RMSE	3,02

Utiliza-se esse treinamento para simular um dia arbitrário e comparar com os resultados reais:

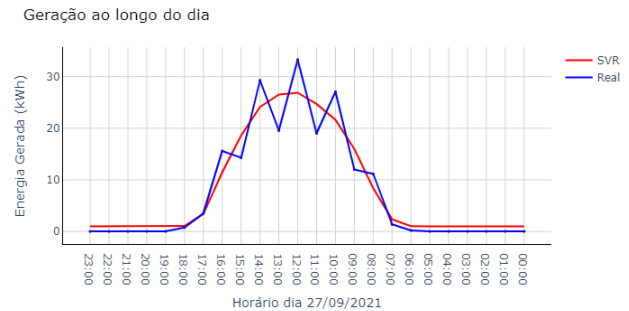


Fig. 37. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel polinomial treinado com as variáveis irradiância, temperatura ambiente e umidade relativa do ar.

I. Treinamento com irradiância, temperatura ambiente e umidade relativa do ar com função kernel RBF

Como não é possível fazer uma representação gráfica com as 3 (três) variáveis será apresentado apenas a tabela das métricas obtidas. As variáveis analisadas foram a irradiância, temperatura e umidade relativa do ar com a função *kernel* RBF. Os valores obtidos foram:

Métrica	Valor
ME	21,27
MAE	1,90
RMSE	2,86

Utiliza-se esse treinamento para simular um dia arbitrário e comparar com os resultados reais:

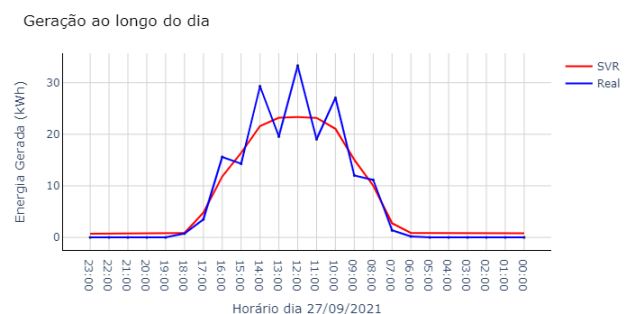


Fig. 38. Gráfico que representa a geração geral dia 27/09/2021 e a previsão feita pelo SVR com kernel RBF treinado com as variáveis irradiância, temperatura ambiente e umidade relativa do ar.

J. Análise das métricas obtidas

Para esta análise serão considerados apenas as métricas MAE e RMSE, pois são as métricas mais relevantes para nosso estudo, já que a ME pode apresentar grandes variações para grandes valores de irradiância. A Tabela a seguir é um compilado dos resultados a coluna “Kernel” representa o kernel utilizado e a coluna “Variáveis” representa o número de variáveis utilizado no treinamento.

Kernel	Variáveis	MAE	RMSE
Linear	1	1,80	2,82
Linear	2	1,67	2,83
Linear	3	1,64	2,73
Polinomial	1	2,90	4,81
Polinomial	2	2,31	3,68
Polinomial	3	1,99	3,02
RBF	1	1,73	2,91
RBF	2	2,01	2,90
RBF	3	1,90	2,86

X. CONCLUSÕES

Conforme explicado anteriormente, algumas variáveis afetam sobremaneira a capacidade de gerar energia das células fotovoltaicas. Dentre elas a literatura destaca a irradiância e a temperatura interna das células.

Ademais, para a análise dos dados obtidos pelas placas fotovoltaicas instaladas no Campus Colemar Natal e Silva, UFG, foi aplicado o algoritmo SVR, em razão da sua facilidade em manipular o treinamento alterando as funções kernel, o que torna possível uma comparação entre seus diferentes tipos. Neste estudo foram analisados os kernels: linear, polinomial e RBF e utilizada a linguagem Python.

As variáveis submetidas foram: irradiância, temperatura ambiente e umidade relativa do ar. Apesar da importância da última variável, essa apresenta menor impacto nos dados obtidos, o que reafirma a relevância da irradiância.

Pode-se observar que independente da variável utilizada no trabalho, o kernel linear foi o que melhor performou nas métricas MAE e RMSE. Também observar-se que para o kernel linear houve um pequeno ganho nas métricas MAE e RMSE ao aumentar o número de variáveis no treinamento.

O treino que apresentou melhor resultado quanto a MAE e RMSE foi ao utilizar o kernel linear com as 3 (três) variáveis – irradiância, temperatura ambiente e umidade relativa do ar.

Ao observar as simulações do dia 27 de setembro de 2021, as realizadas com o kernel linear tiveram pouca variação visualmente. Enquanto as simulações feitas com kernel polinomial tiveram as maiores variações; nota-se que sua curva sofre alteração quando o número de variáveis em seu treinamento é acrescida.

Vale ressaltar que a temperatura da placa está relacionada a sua eficiência, como explicado anteriormente, e apresenta uma

curva no gráfico semelhante à temperatura ambiente. Percebe-se que com os treinamentos do kernel linear a variação foi pequena se comparados com os demais kernels.

Como as variações nas métricas MAE e RMSE ao se adicionarem variáveis no kernel linear são sutis, conclui-se que o algoritmo não conseguiu tirar o melhor proveito desse cenário.

O Treinamento com o kernel polinomial teve a característica de possuir o maior ME uma vez que utiliza o parâmetro de grau da função como 3 (três) e ao aumentar esse grau, a função tenderá a ter um comportamento similar a RBF.

Constata-se que o modelo que obteve as melhores métricas MAE e RMSE é o SVR com kernel linear. Todavia, não se comporta bem com grandes variações, como é perceptível na simulação do dia 27 de setembro de 2021 realizado das 10:00h às 15:00h.

O algoritmo possui uma métrica de erro aceitável para finalidades que não demandem alta precisão, para casos em que ela é necessária, recomenda-se o estudo e treinamento de algoritmos que trabalhem melhor com número de variáveis superiores a 1 (um).

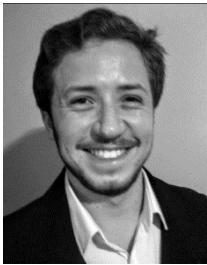
XI. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] L. K. TSURUDA, T. A. MENDES, L. R. VITOR e M. B. SILVEIRA, “A importância da energia solar para o desenvolvimento sustentável e social,” *6th International Workshop Advances in Cleaner Production*, 24 Maio 2017.
- [2] A. G. d. N. Unidas, “Unicef,” 10 Dezembro 1984. [Online]. Available: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. [Acesso em 10 Outubro 2021].
- [3] G. Motta, “Jornal UFG,” Universidade Federal de Goiás, 20 Novembro 2019. [Online]. Available: <https://jornal.ufg.br/n/122121-ufg-instalou-25-mil-lampadas-led-e-comeca-producao-de-energia-solar>. [Acesso em 10 Outubro 2021].
- [4] J. T. Pinho e M. A. Galdino, *Manual de Engenharia para Sistemas Fotovoltaicos*, Rio de Janeiro, 2014, pp. 67-69 144-153.
- [5] E. B. PEREIRA, F. R. MARTINS, A. R. GONÇALVES, R. S. COSTA, F. L. LIMA, R. RÜTHER, S. L. ABREU, G. M. TIEPOLO, S. V. PEREIRA e J. G. SOUZA, em *Atlas brasileiro de energia solar*, São José dos Campos, 2017, pp. 15-19.
- [6] M. G. Villalva e Jonas Rafael Gazoli, *Energia Solar Fotovoltaica*, 1ª ed., São Paulo, São Paulo: Érica Ltda., 2012, pp. 39-81.
- [7] S. D. O. e. a. FROTIN, em *Usina Fotovoltaica JAÍBA SOLAR - Planejamento e Engenharia*, Brasília, Teixeira Gráfica e Editora LTDA, 2017, pp. 62-71.

- [8] A. C. P. L. F. Carvalho, K. Faceli, A. C. Lorena e J. Gama, “Inteligência Artificial: uma abordagem de aprendizado de máquina,” 2011, pp. 1-7, 122-135.
- [9] D. Britto Figueiredo Filho e J. A. da Silva Júnior, “Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r),” *Revista Política Hoje*, pp. 115-120, 2009.
- [10] M. Aamir, S. S. H. Rizvi, M. A. Hashmani, M. Zubair e J. Ahmad, “Machine Learning Classification of Port Scanning and DDoS Attacks: A Comparative Analysis,” 2019.
- [11] R. Ruiz-Gonzalez, J. Gomez-Gil, F. J. Gomez-Gil e V. Martínez, “An SVM-Based Classifier for Estimating the State of Various Rotating Components in Agro-Industrial Machinery with a Vibration Signal Acquired from a Single Point on the Machine Chassis,” 3 11 2014.
- [12] B. Coutinho, “Modelos de Predição | SVM,” 28 Julho 2019. [Online]. Available: <https://medium.com/turing-talks/turing-talks-12-classifica%C3%A7%C3%A3o-por-svm-f4598094a3f1>. [Acesso em 1 10 2021].
- [13] “3.3. Metrics and scoring: quantifying the quality of predictions — scikit-learn 1.0.1 documentation,” [Online]. Available: https://scikit-learn.org/stable/modules/model_evaluation.html. [Acesso em 20 10 2021].
- [14] J. A. A. d. Silva, “Tratamento de dados meteorológicos e análise de desempenho do sistema fotovoltaico da EMC/UFG,” pp. 60-61,94-96.



Marcelo S. de Castro graduou-se em Engenharia Elétrica pela Universidade Federal de Juiz de Fora (1992), com mestrado em Engenharia Elétrica pela Universidade Estadual de Campinas (1995) e doutorado em Engenharia Elétrica pela UnB (2010). Docente Associado da Universidade Federal de Goiás, tendo ingressado em 1996. Possui experiência na área de engenharia de redes, computação paralela e distribuída, comunicações óticas e tecnologias alternativas de última milha (BPL, ZigBee, Wi-Fi). Desenvolve pesquisas em temas que incluem redes de comunicação (5G, Gigabit Wi-Fi), *Smart Grids*, *Smart Cities*, *Smart Campus*, tecnologia da informação e comunicação e gestão aplicadas a projetos de redes de telecomunicações, projetos de automação usando Plataforma Arduino e educação em engenharia.



Victor Gomide Ferraz é graduando em Engenharia de Computação na Universidade Federal de Goiás (2021). Atualmente é desenvolver de software na Oobj Tecnologia em Gestão Fiscal em Goiânia-GO. Tendo experiência no desenvolvimento de soluções tecnológicas que buscam melhorar e otimizar processos.