



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
CURSO DE ENGENHARIA ELÉTRICA**

JAN LUCAS DE OLIVEIRA GOMES

**ESTUDO DE PREVISÃO DE IRRADIÂNCIA SOLAR USANDO
REGRESSÃO POR VETORES DE SUPORTE**

FORTALEZA

2018

JAN LUCCAS DE OLIVEIRA GOMES

ESTUDO DE PREVISÃO DE IRRADIÂNCIA SOLAR USANDO
REGRESSÃO POR VETORES DE SUPORTE

Trabalho de conclusão de curso apresentado ao Curso de Engenharia Elétrica do Departamento de Engenharia Elétrica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Paulo Cesar Marques de Carvalho.

Coorientador: Prof. Me. Marcello Anderson Ferreira Batista Lima.

FORTALEZA

2018

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- G614e Gomes, Jan Luccas de Oliveira.
Estudo de previsão de irradiância solar usando regressão por vetores de suporte / Jan Luccas de Oliveira Gomes. – 2018.
57 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Elétrica, Fortaleza, 2018.
Orientação: Prof. Dr. Paulo Cesar Marques de Carvalho.
Coorientação: Prof. Me. Marcello Anderson Ferreira Batista Lima.
1. Energia solar. 2. Previsão de irradiância solar. 3. Regressão por vetores de suporte. I. Título.
CDD 621.3
-

JAN LUCCAS DE OLIVEIRA GOMES

ESTUDO DE PREVISÃO DE IRRADIÂNCIA SOLAR USANDO
REGRESSÃO POR VETORES DE SUPORTE

Trabalho de conclusão de curso apresentado ao
Curso de Engenharia Elétrica do Departamento
de Engenharia Elétrica da Universidade Federal
do Ceará, como requisito parcial à obtenção do
título de Bacharel em Engenharia Elétrica.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Prof. Dr. Paulo Cesar Marques de Carvalho (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Arthur Plínio de Souza Braga
Universidade Federal do Ceará (UFC)

Prof. Dr. Raimundo Furtado Sampaio
Universidade Federal do Ceará (UFC)

A Deus.

Aos meus pais, Nadia e Mezaque
e a todos os meus familiares e amigos.

AGRADECIMENTOS

À Universidade Federal do Ceará pela oportunidade de ingressar numa instituição de ensino superior e pela oportunidade de adquirir conhecimento.

Ao Prof. Dr. Paulo Cesar Marques de Carvalho e ao Prof. Me. Marcello Anderson Ferreira Batista Lima, pela excelente orientação.

Ao Prof. Dr. Arthur Plínio de Souza Braga e ao Prof. Dr. Raimundo Furtado Sampaio pela disponibilidade de estar na banca examinadora e pelas ótimas sugestões.

A todos os meus familiares que me auxiliaram nos momentos difíceis.

Aos colegas do curso que me incentivaram a chegar até aqui.

“A persistência é o menor caminho para o êxito.” (Charles Chaplin)

RESUMO

A intermitência do recurso solar é uma das barreiras à utilização de parques fotovoltaicos em um sistema elétrico. O crescimento da geração fotovoltaica demanda estudos de previsão solar para permitir o balanceamento entre consumo e produção de eletricidade. Este trabalho tem como objetivo apresentar um estudo de previsão de irradiação solar global horizontal realizado a partir de uma série de dados históricos de irradiância solar global horizontal coletada em Fortaleza-CE entre 2003 e 2005. A série histórica coletada foi dividida em 70% e 30% para treinamento e para teste respectivamente. O estudo de previsão foi baseado no algoritmo de Regressão por Vetores de Suporte (SVR – *Support Vector Regression*). O algoritmo foi implementado na linguagem de programação Python utilizando a biblioteca de aprendizado de máquina Scikit-learn. O horizonte de previsão é de 10 minutos. Em seguida comparou-se os dados previstos com os dados medidos para o período de teste e avaliações de erros foram feitas. De modo geral foi observado que os erros tendem a aumentar nos períodos do nascer e do pôr do sol. A partir da análise dos resultados obtidos nesse estudo, verificou-se que até 56,4% das previsões geraram erros menores de até 15%, o que representa um resultado satisfatório em comparação com estudos já realizados na área.

Palavras-chave: Energia Solar. Previsão de Irradiância Solar. Regressão por Vetores de Suporte.

ABSTRACT

The intermittence of solar resource is one of the barriers for the use of photovoltaics plants in a electric power system. The growth of photovoltaic generation demands solar forecasting studies to allow the balance between consumption and electricity production. This work has as objective to present a study of solar global horizontal irradiance forecasting made from historical series of data on solar global horizontal irradiance collected in Fortaleza-CE between 2003 and 2005. The data series were divided in 70% and 30% for training and testing set respectively. The forecasting study was based on Support Vector Regression (SVR) algorithm. The algorithm was implemented in Python programming language through Scikit-learn machine learning library. The forecast horizon is 10 minutes. Afterwards predicted and measured data during testing period were compared and evaluation of errors were made. In general, it was observed that errors tend to increase during sunrise and sunset periods. From the analysis of the obtained results in this study, it was verified that 56,4% of forecasts got errors up to 15%, which represents a satisfactory result

Keywords: Solar Energy. Solar Irradiance Forecasting. Support Vector Regression.

LISTA DE FIGURAS

Figura 1	– Irradiação global horizontal diária no território nacional	19
Figura 2	– Sazonalidade do potencial de geração fotovoltaica no Brasil	20
Figura 3	– Componentes de irradiância solar para a atmosfera	23
Figura 4	– Conjunto de dados linearmente separável	27
Figura 5	– Possíveis hiperplanos para a separação das classes	28
Figura 6	– Hiperplano ótimo para a separação das classes	28
Figura 7	– Cálculo da distância d entre os hiperplanos H_1 e H_2	30
Figura 8	– Expressão gráfica da função de perda ε -insensível	32
Figura 9	– Exemplo de uma série temporal com janela de previsão com 5 elementos	40
Figura 10	– Fluxograma do algoritmo de previsão	41

LISTA DE GRÁFICOS

Gráfico 1	– Evolução da potência instalada FV mundial	16
Gráfico 2	– Comparação das médias diárias de irradiação global horizontal entre as regiões brasileiras e alguns países europeus	18
Gráfico 3	– Evolução da potência instalada acumulada da fonte fotovoltaica no Brasil .	21
Gráfico 4	– Evolução da potência solar contratada e acumulada dos últimos leilões de energia	21
Gráfico 5	– Irradiância solar nos 5 primeiros dias de medição	38
Gráfico 6	– Irradiação solar diária em Fortaleza durante um ano	39
Gráfico 7	– Comparação entre medições e previsões para 5 dias de outubro de 2004	42
Gráfico 8	– Erro percentual das previsões para 5 dias de outubro de 2005.....	43
Gráfico 9	– Comparação entre medições e previsões para 5 dias de fevereiro de 2005..	44
Gráfico 10	– Erro percentual das previsões para 5 dias de fevereiro de 2005	44
Gráfico 11	– Dispersão dos valores medidos e previstos	45

LISTA DE TABELAS

Tabela 1 – Alguns trabalhos de previsão de irradiância desenvolvidos anteriormente	25
Tabela 2 – Funções kernel mais comuns	36
Tabela 3 – Resultado da avaliação das previsões no período de teste	46

LISTA DE ABREVIATURAS E SIGLAS

ABSOLAR	Associação Brasileira de Energia Solar Fotovoltaica
ANEEL	Agência Nacional de Energia Elétrica
AR	<i>Auto-Regressive</i>
ARIMA	<i>Auto-Regressive Integrated Moving Average</i>
ANN	<i>Artificial Neural Network</i>
BEN	Balanco Energético Nacional
BP	<i>British Petroleum</i>
CE	Ceará
EPE	Empresa de Pesquisa Energética
FV	Fotovoltaico
IRENA	<i>International Renewable Energy Agency</i>
KKT	Karush-Kuhn-Tucker
k-NN	<i>k-Nearest Neighbors</i>
MME	Ministério de Minas e Energia
NREL	<i>National Renewable Energy Laboratory</i>
nRMSE	<i>Normalized Root Mean Square Error</i>
ONS	Operador Nacional do Sistema
PDE	Plano Decenal de Energia
ProGD	Programa de Desenvolvimento da Geração Distribuída de Energia Elétrica
REN	Resolução Normativa
REN21	<i>Renewable Energy Policy Network for the 21st Century</i>
RMSE	<i>Root Mean Square Error</i>
R ²	Coefficiente de determinação
RF	<i>Random Forest</i>
SIN	Sistema Interligado Nacional
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
UN	<i>United Nations</i>
UNFCCC	<i>United Nations Framework Convention on Climate Change</i>

LISTA DE SÍMBOLOS

G	Irradiação global horizontal
G_{dif}	Irradiação difusa horizontal
G_{dir}	Irradiação direta horizontal
G_n	Irradiação direta normal
θ_z	Ângulo zenital
m	Margem de separação
$f(x)$	Função de separação
\mathbf{w}	Vetor de pesos ajustável
\mathbf{x}	Vetor de dados de entrada
b	Bias
ε	Precisão
ξ	Variável de folga
C	Ponderação dos Erros
α	Multiplicador de Lagrange
η	Multiplicador de Lagrange
$\Phi(\mathbf{x})$	Função mapeadora
$K(\mathbf{x}_1, \mathbf{x}_2)$	Função kernel
Y_{prev}	Valor previsto para a amostra
Y_{real}	Valor real da amostra
Y_{med}	Valor médio das amostras

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Motivação	14
1.2	Objetivos	15
1.3	Estrutura do trabalho	15
2	ESTADO ATUAL DA GERAÇÃO FOTOVOLTAICA	16
2.1	Estado atual da geração FV no mundo	16
2.2	Estado atual da geração FV no Brasil	17
3	PREVISÃO SOLAR	23
3.1	Irradiância solar global	23
3.2	Modelos de previsão solar	24
4	REGRESSÃO POR VETORES DE SUPORTE	27
4.1	Máquinas de Vetores de Suporte	27
4.1.1	<i>Problema de otimização para as SVM</i>	29
4.2	Regressão por vetores de suporte	31
4.2.1	<i>Funções Kernel</i>	35
5	METODOLOGIA E RESULTADOS	38
5.1	Base de Dados	38
5.2	Implementação do código em Python para a previsão de irradiância solar ...	39
5.3	Avaliação da precisão das previsões	41
5.4	Resultados	42
6	CONCLUSÃO E TRABALHOS FUTUROS	47
6.1	Conclusão	47
6.2	Trabalhos Futuros	48
	REFERÊNCIAS	49
	APÊNDICE A – CÓDIGO COMPUTACIONAL	52

1 INTRODUÇÃO

À medida que a população mundial cresce e os países buscam oferecer maior qualidade de vida e desenvolvimento a suas populações, aumenta a demanda por eletricidade. Segundo previsões das Nações Unidas (UN – *United Nations*), a população mundial crescerá de 7,6 bilhões de habitantes em 2017 para 9,8 bilhões em 2050, representando um aumento de cerca de 29% (UN, 2017).

Simultaneamente a este processo, cresce a preocupação com as mudanças climáticas causadas pela liberação de gases do efeito estufa, a maioria proveniente da queima de combustíveis fósseis. Como mostra desta crescente preocupação, em dezembro de 2015, 195 países assinaram o Acordo de Paris durante a Conferência das Nações Unidas sobre as Mudanças Climáticas. Entre os seus objetivos, está o de limitar o aumento da temperatura média global em até 1,5 °C acima dos níveis pré-industriais (UNFCCC, 2015).

Para alcançar estes objetivos, os países têm buscado diversificar sua matriz energética no intuito de reduzir a dependência de combustíveis fósseis e simultaneamente favorecer uma maior utilização de fontes renováveis, dentre as quais destaca-se a fonte fotovoltaica (FV).

Ao contrário da geração a partir de fontes tradicionais, como as termelétricas, em que a energia é produzida de forma contínua, a geração FV apresenta característica intermitente causada por fatores climáticos. A intermitência é um dos entraves para a integração de parques FV no sistema elétrico, na medida que a intermitência dificulta o controle da geração provocando excesso ou escassez de energia no sistema (LIMA, 2016).

1.1 Motivação

A intermitência da geração FV traz como consequência a necessidade de armazenamento de energia ou de integração da fonte intermitente com outras fontes, de modo a atender a demanda. Estes processos devem vir acompanhados de estudos de previsão de consumo e de geração, de forma a estimar a quantidade de energia a ser adicionada no sistema elétrica para atender a demanda.

Para minimizar os impactos da intermitência na geração FV, foi proposto por este trabalho um estudo de previsão de irradiância solar com horizonte de 10 minutos, de forma a aumentar a eficiência econômica dos mecanismos de compensação de energia e proporcionar estabilidade da rede para os operadores do sistema elétrico.

1.2 Objetivos

Este trabalho tem como objetivo apresentar um estudo de previsão de irradiância global horizontal para a localidade de Fortaleza – CE com horizonte de previsão de 10 minutos, utilizando o algoritmo de aprendizado de máquina de regressão por vetores de suporte (SVR – *Support Vector Regression*). Para esta função foi implementado um código computacional utilizando a linguagem Python por meio da biblioteca Scikit-learn.

A previsão solar desempenha um importante papel no controle de sistemas de geração FV e permite que os operadores do sistema elétrico fazer o controle entre geração e consumo e alocar as cargas para serem supridas por outras fontes quando a geração FV não estiver disponível. A irradiância global horizontal é mais requerida em estudos de previsão por impactar diretamente na geração FV (CHATUVERDI; ISHA, 2016).

Os objetivos específicos deste trabalho são:

- Preparar os dados coletados de irradiância global horizontal dividindo em conjuntos de treinamento e testes;
- Implementar um algoritmo computacional baseado em SVR a fim de fornecer previsões a partir dos dados de treinamento;
- Realizar as previsões a partir do algoritmo implementado;
- Analisar comparativamente as previsões obtidas com os dados medidos e realizar uma quantificação dos erros;

1.3 Estrutura do trabalho

O trabalho está organizado em 6 capítulos. No capítulo 1 são mostradas as motivações e a apresentação geral do trabalho. No capítulo 2 é mostrado o panorama atual da geração FV no Brasil e no resto do mundo. No capítulo 3 é mostrado uma descrição das técnicas de previsão solar. No capítulo 4 é feita a abordagem teórica da SVR. No capítulo 5 é apresentada a metodologia do trabalho e a análise dos resultados. Por fim, no capítulo 6 são mostradas as conclusões finais e propostas para trabalhos futuros.

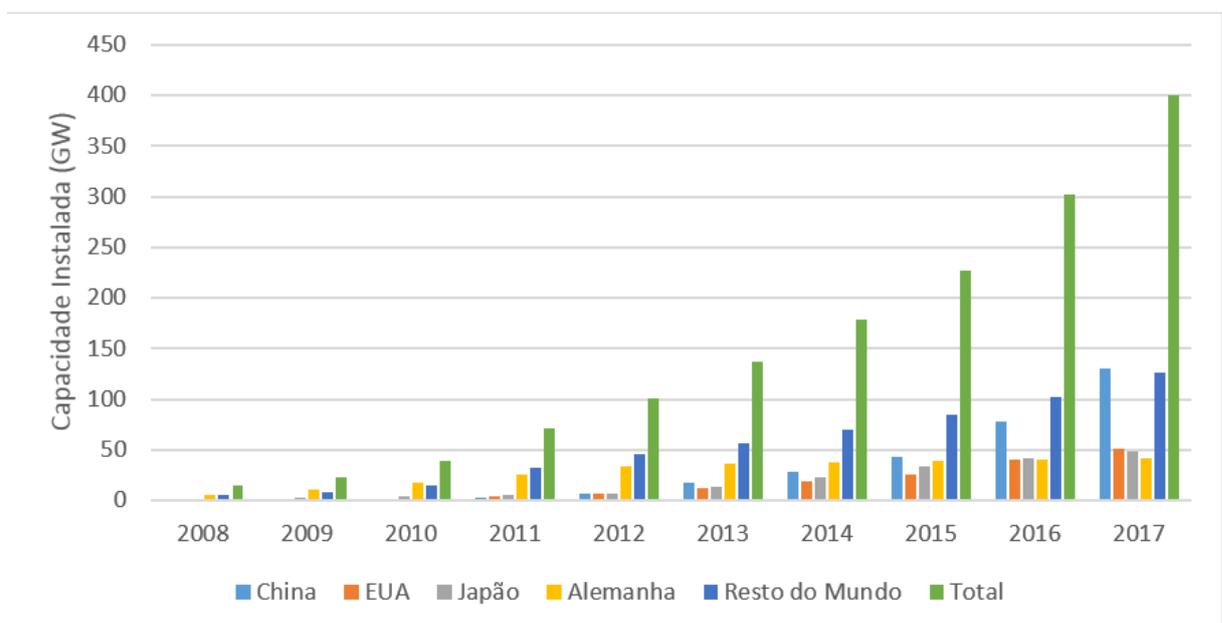
2 ESTADO ATUAL DA GERAÇÃO FOTOVOLTAICA

A geração de eletricidade por fonte fotovoltaica está em franco crescimento ao redor do mundo. Em 2017, mais potência instalada foi adicionada para a geração FV do que em qualquer outra tecnologia de geração de energia. Neste ano foram adicionados cerca de 98 GW em relação ao ano de 2016, totalizando cerca de 402 GW de potência instalada acumulada em 2017. Estima-se que neste mesmo ano, a energia fotovoltaica seja responsável por aproximadamente 3,36 milhões de empregos diretos e indiretos, sendo aproximadamente 2/3 destes na China. No Brasil, a energia solar é responsável por cerca de 10 mil postos de trabalho (REN21, 2018).

1.1 Estado atual da geração FV no mundo

Em 2017, China, EUA, Índia, Japão e Turquia foram os países que mais contribuíram com o acréscimo de potência instalada. O Brasil também tem despontado no mercado global ao ser o 10º nesta lista. Na lista dos países com mais capacidade instalada, a China lidera com cerca de 131 GW, mais que o dobro dos EUA, o segundo país na lista com 51 GW, Japão e Alemanha seguem na sequência (REN21, 2018). O Gráfico 1 mostra o crescimento histórico da potência instalada FV no mundo nos últimos anos 10 anos.

Gráfico 1 – Evolução da potência instalada FV mundial



Fonte: Adaptado de BP (2018)

Em 2017, a geração de eletricidade através da fonte fotovoltaica no mundo foi de 442,6 TWh, correspondendo a 1,73% da geração elétrica mundial. A geração fotovoltaica também tem apresentado um crescimento vertiginoso, tendo crescido 35,2% em relação a 2016 e 49,7% por ano entre 2006 e 2016. A maior parte desta geração está concentrada da China, EUA, Japão e Alemanha, os quais contribuem com 24,4%, 17,6%, 14,1% e 9,1% da geração elétrica a partir da fonte fotovoltaica respectivamente (BP, 2018).

Fatores econômicos têm contribuído para atrair mais interesse pela geração FV. Entre 2009 e 2017, o custo de produção elétrica para a geração FV foi reduzido em 72%, passando de 304 para 86 US\$/MWh, o que faz com que a fonte fotovoltaica já comece a ter competitividade com outras fontes de energias alternativas. Como comparação, em 2017, os custos de produção para a produção elétrica a partir de fonte eólica *on-shore* e *off-shore* foram de US\$ 67/MWh e US\$ 124/MWh respectivamente. Nos EUA, os custos de produção a partir da fonte fotovoltaica já são competitivos até com fontes tradicionais como o carvão e termelétricas a gás. Enquanto os custos da geração fotovoltaica foram de US\$ 54/MWh, a geração a carvão e a termelétrica a gás custou US\$ 66/MWh e US\$ 49/MWh respectivamente em 2017 (FRANKFURT, 2018).

A demanda maior por fontes fotovoltaicas também tem sido incentivada pelo crescente aumento da eficiência dos módulos fotovoltaicos. Nos últimos 10 anos, a eficiência dos módulos FV comerciais baseados em silício, como os mono-cristalinos e os poli-cristalinos, cresceram de 12% para 17%. No mesmo período, a eficiência dos módulos baseados em telureto de cádmio, como os de filme fino, tiveram sua eficiência incrementada de 9% para 16% (FRAUNHOFER, 2018).

Com todos esses incentivos, está prevista que a fonte FV alcance 7.122 GW de capacidade instalada em 2050, o que representaria 36,64% da capacidade instalada para geração elétrica mundial no mesmo ano (IRENA, 2018).

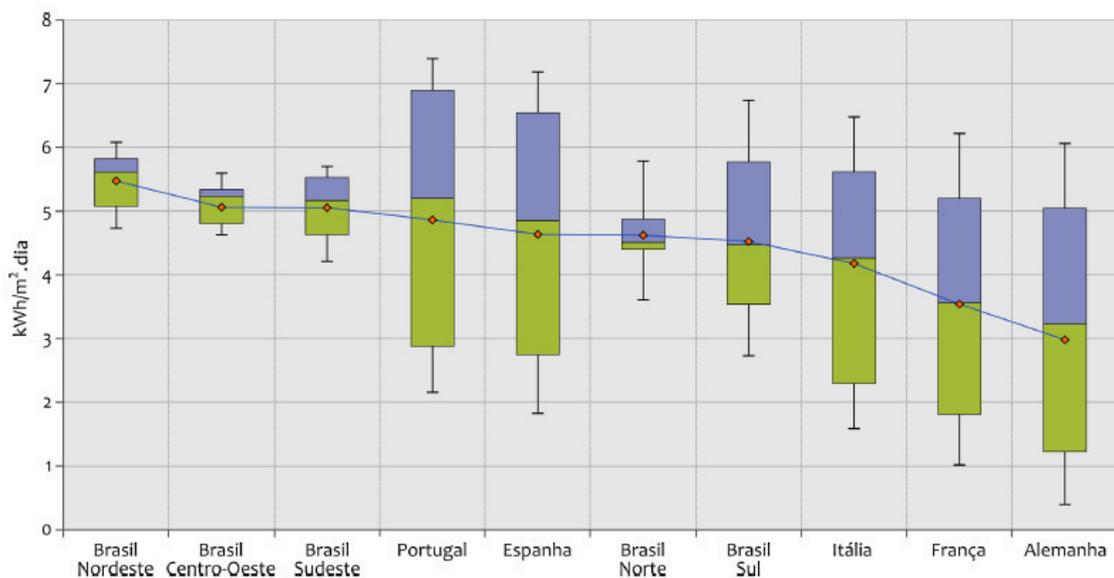
2.2 Estado atual da geração FV no Brasil

Diferentemente do que ocorre na maioria dos países do mundo, a utilização de fontes renováveis para a produção de eletricidade já é dominante no Brasil, mais precisamente com a utilização de fonte hidroelétrica. Segundo o mais recente balanço energético nacional (BEN 2017), as hidroelétricas responderam por 68,1% da geração elétrica brasileira em 2016. Desde a crise de abastecimento que afetou o sistema elétrico brasileiro em 2001, ocasionada em parte pela escassez de chuvas, o país tem buscado diversificar sua matriz elétrica, para

depender menos do regime de chuvas, inicialmente com a geração termelétrica e mais recentemente com a geração eólica e solar.

Por ser um país de extensão continental e ter a maior parte do seu território nos trópicos, o recurso solar é abundante por todo o território brasileiro. O Gráfico 2 mostra uma comparação das médias diárias de irradiação global horizontal entre as regiões brasileiras e alguns países europeus.

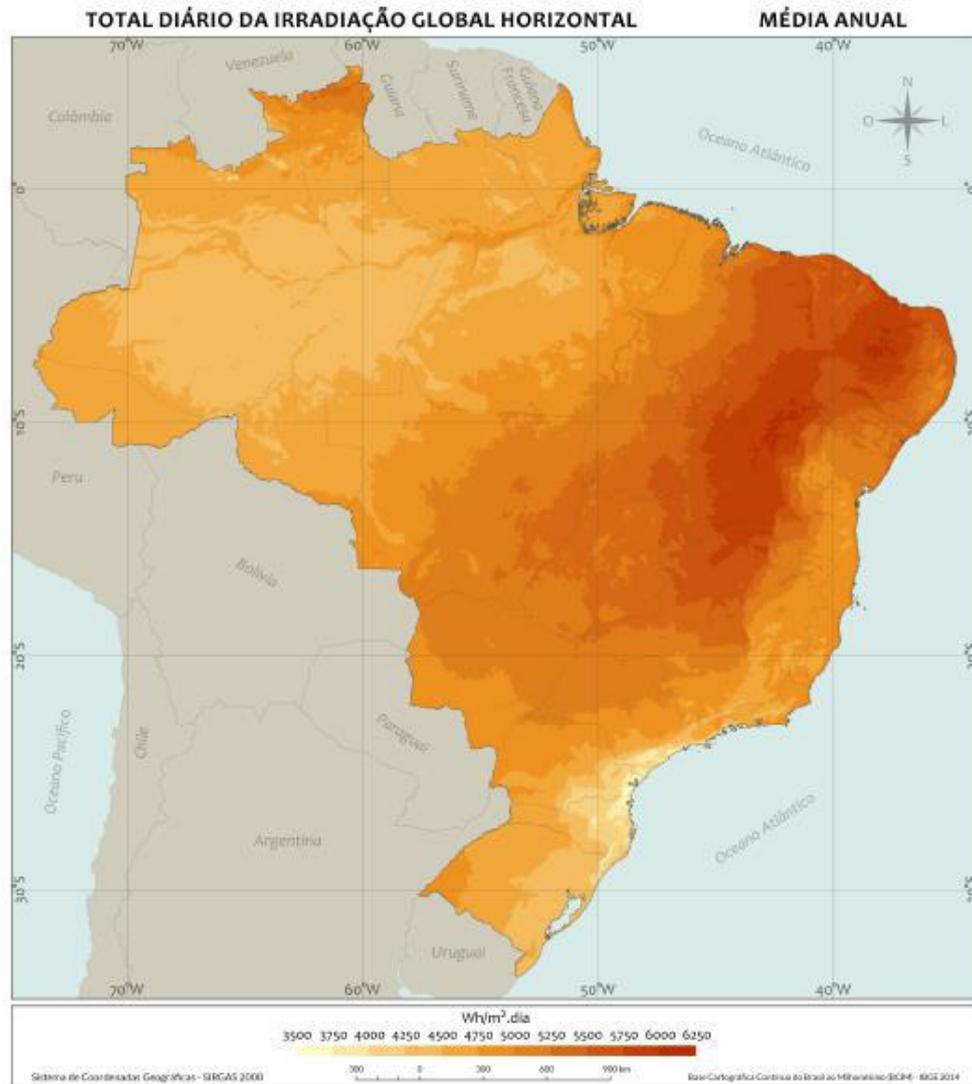
Gráfico 2 – Comparação das médias diárias de irradiação global horizontal entre as regiões brasileiras e alguns países europeus



Fonte: Pereira et al. (2017)

Verifica-se a partir do Gráfico 2 que todas as regiões brasileiras apresentam um alto nível de irradiação solar média, variando de cerca de 5,5 kWh/m².dia no Nordeste a 4,5 kWh/m².dia no Sul do país. Enquanto isso, a Alemanha, 4º país com maior geração FV, apresenta nível de irradiação solar média de cerca de 3,0 kWh/m².dia. Mesmo com níveis menores de irradiação, a Alemanha gerou 39,9 TWh de eletricidade através da fonte fotovoltaica em 2017 enquanto foram gerados 0,7 TWh no Brasil no mesmo ano, um valor 57 vezes maior (BP, 2018). Isto mostra o quanto subaproveitado é o recurso solar no Brasil. Nota-se também que a variabilidade da irradiação média em todas as regiões brasileiras é baixa, a exceção da região Sul, por ser uma região localizada em altas latitudes. A Figura 1 mostra a disponibilidade do recurso solar no território nacional através das médias anuais de irradiação solar horizontal global diária, onde destaca-se o Nordeste como a região com os maiores índices de irradiação, seguido do Centro-oeste e Sudeste.

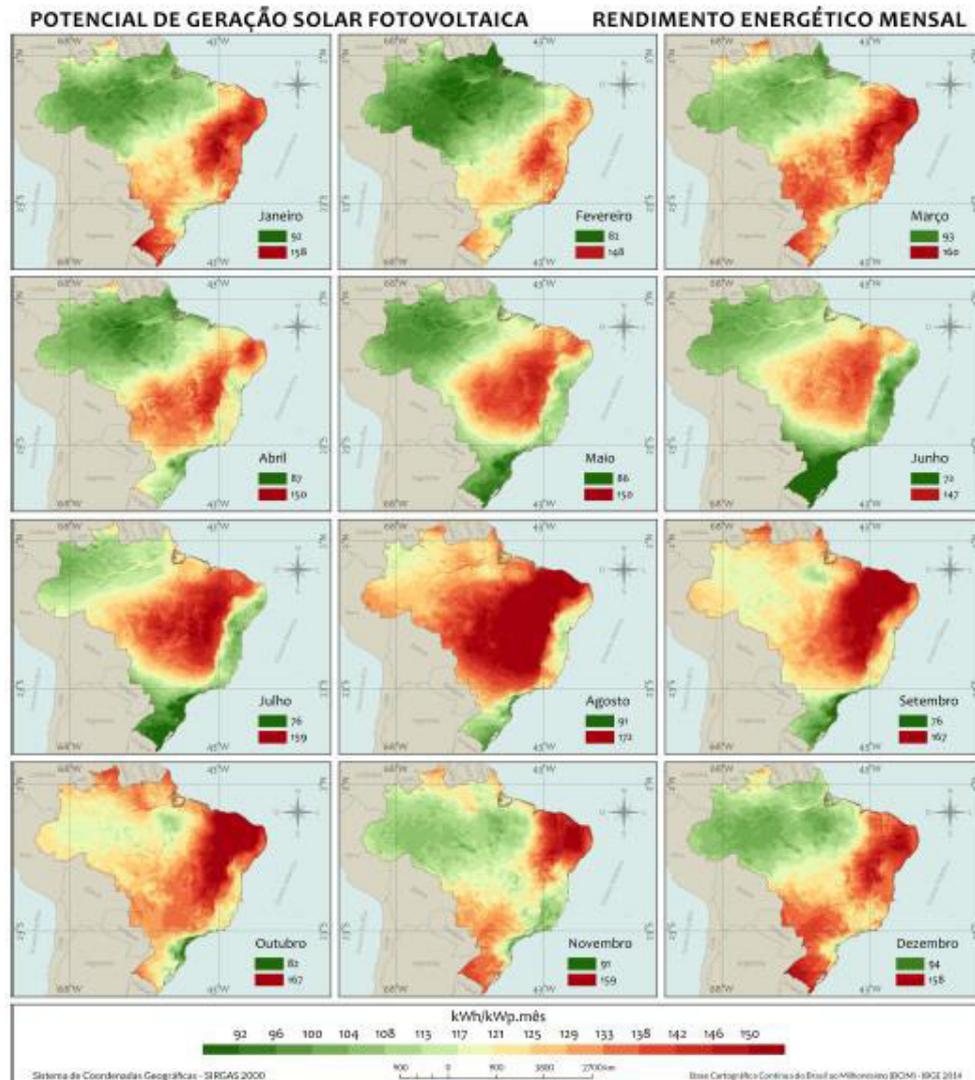
Figura 1 – Irradiação global horizontal diária no território nacional



Fonte: Pereira et al. (2017)

O potencial de geração fotovoltaica no Brasil é mostrado de forma sazonal na Figura 2, onde mostra-se o comportamento do potencial de geração mês a mês ao longo do ano no território nacional. Devido a variação climática sazonal, nos meses do verão, o potencial de geração é máximo nos estados do Sul e Sudeste, o que coincide com os meses de demanda máxima registrados pelo Operador Nacional do Sistema (ONS). Por ter natureza distribuída, a geração fotovoltaica pode agir de forma complementar a outras fontes de geração do sistema desafogando os sistemas de transmissão do Sistema Interligado Nacional (SIN). No Nordeste, os meses de maior potencial de geração fotovoltaica coincidem com os meses de seca, portanto a utilização da energia fotovoltaica serviria de forma que as represas do sistema hidroelétrico poupem água ou liberem os recursos hídricos para outras atividades, como o consumo humano.

Figura 2 – Sazonalidade do potencial de geração fotovoltaica no Brasil

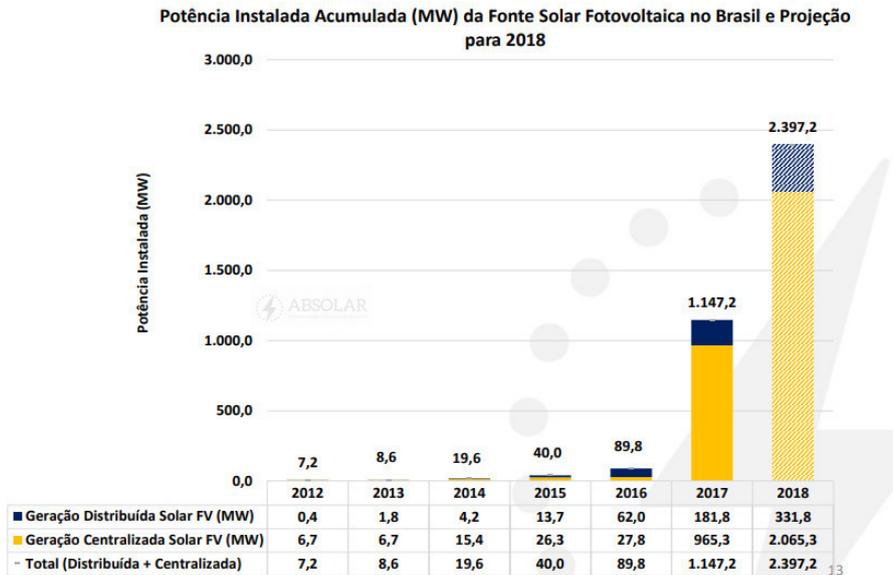


Fonte: Pereira et al. (2017)

Atualmente no Brasil, a capacidade instalada para a geração FV centralizada é de 1.322 MW, equivalente a 0,83% de toda a potência instalada nacional. Outros 817 MW estão em processo de implantação enquanto outros 1.285 MW ainda não tiveram sua implantação iniciada (ANEEL, 2018). Em setembro de 2018, os estados com maior capacidade instalada eram Bahia, Minas Gerais, Piauí, São Paulo e Rio Grande do Norte respectivamente com 413 MW, 321 MW, 210 MW, 150 MW e 106 MW respectivamente (ONS, 2018).

Em 2016, 85 GWh de eletricidade foram gerados através da fonte fotovoltaica em, contribuindo com 0,01% da geração elétrica nacional, tendo crescido 44,7% em relação a 2015 (EPE, 2017). O Gráfico 3 mostra o crescimento da potência instalada da fonte fotovoltaica no país nos últimos anos.

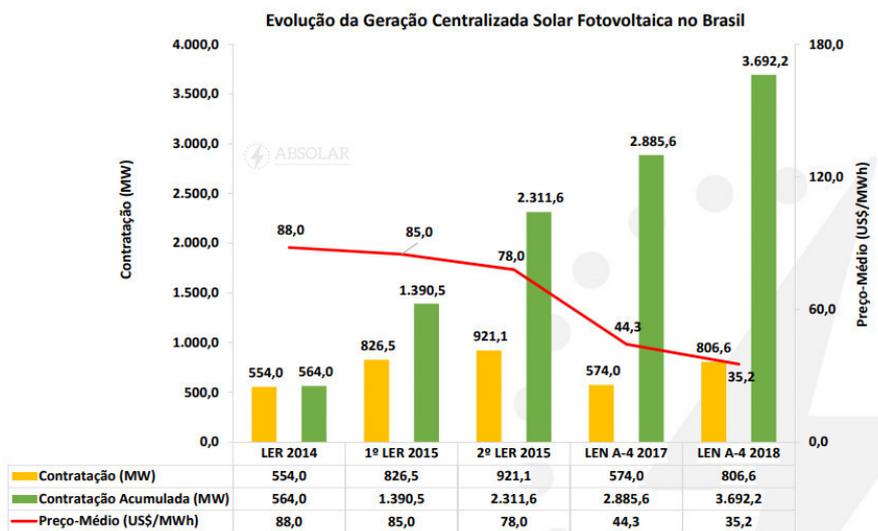
Gráfico 3 – Evolução da potência instalada acumulada da fonte fotovoltaica no Brasil



Fonte: Sauaia (2018)

Segundo o plano decenal de energia (PDE 2026), estão previstos acréscimos de 1.000 GW/ano de capacidade a partir de 2021 e o país poderá alcançar 9.660 MW ao final do período decenal, representando 5,8% da potência instalada nacional. Este cenário parece factível tendo em vista o atual ritmo de evolução da potência de energia solar contratada nos últimos leilões de energia, como se pode observar no Gráfico 4.

Gráfico 4 – Evolução da potência solar contratada e acumulada nos últimos leilões de energia



Fonte: Sauaia (2018)

Apesar do crescimento rápido nos últimos anos, a energia solar ainda é subaproveitada no país. O potencial solar brasileiro é maior que o potencial de todas os outros tipos de fonte renovável juntos. Enquanto o potencial hídrico e eólico são de 280 GW e 300 GW respectivamente, o potencial solar é de 10.000 GW (ABSOLAR, 2016).

Em 2015, o Ministério de Minas e Energia (MME) lançou o Programa de Desenvolvimento da Geração Distribuída de Energia Elétrica (ProGD) com o objetivo de ampliar a geração distribuída no país, sobretudo solar. O programa prevê que 2,7 milhões de unidades consumidoras, entre elas residências, comércios e indústrias, gerarão 48 TWh de energia em 2030, o que equivale à metade da geração anual da usina hidroelétrica de Itaipu. Outra meta do programa é que as energias renováveis, além da hídrica, tenham de 28% a 33% de participação na matriz energética em 2030 (MME, 2015).

A partir de 2012, a ANEEL estabeleceu através da resolução 482/2012, a regulamentação da micro e da minigeração e do mecanismo de compensação de energia, onde uma unidade consumidora pode injetar excedente gerado por telhados solares na rede elétrica, acumulando créditos que podem ser compensados em kWh. Em 2016, a ANEEL flexibilizou as regras através da REN 687/2015, estendendo o conceito de telhados solares para condomínios, consórcios, cooperativas e autoconsumo remoto. Com esta flexibilização, a ANEEL estima que até 2024, 1,2 milhão de unidades consumidoras passarão a produzir sua própria energia, totalizando 4,5 GW de potência instalada (ANEEL, 2015).

Com os incentivos a micro e minigeração de energia através do mecanismo de compensação de energia, cresce a necessidade de modelos de previsão confiáveis para proporcionar aos consumidores um bom controle desse mecanismo de compensação de modo a aumentar sua eficiência econômica (SAMANTA et al, 2014).

Para os operadores do sistema elétrico, modelos de previsão confiáveis com horizonte de curto prazo são necessários para assegurar a estabilidade da rede elétrica com sistemas FV conectados (SILVA, 2018).

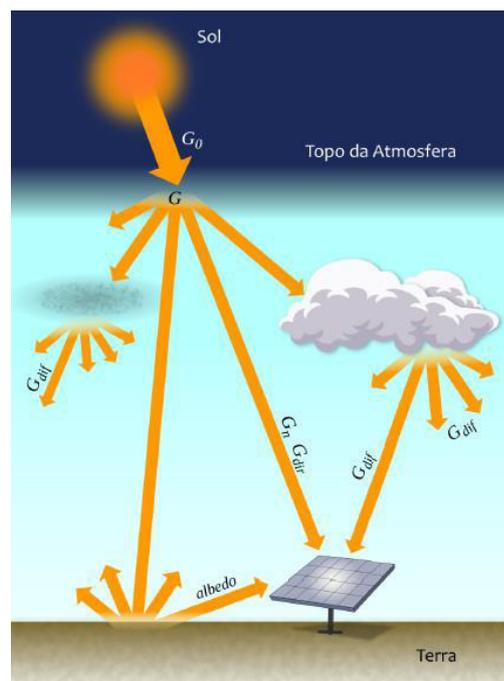
3 PREVISÃO SOLAR

O sol é a maior fonte de energia nas proximidades da terra e é responsável por todos os fenômenos climáticos e pela existência de vida na terra. O sol é uma estrela que emite radiação em todas as direções do espaço. Existem duas medidas associadas a radiação solar, a irradiância e a irradiação. A Irradiância é a potência da radiação solar por unidade de área e é medida em W/m^2 . Nas camadas mais altas da atmosfera terrestre, o nível de irradiância solar é de cerca de $1366 W/m^2$. Já a irradiação é quantidade de energia solar que incide numa superfície durante um determinado período de tempo e é medida em Wh/m^2 (NREL, 2018).

3.1 Irradiância solar global

Ao entrar na atmosfera, parte da irradiância solar sofre espalhamento ou é absorvida pelas nuvens, moléculas e partículas suspensas na atmosfera. A parcela de radiação solar que atinge o solo e não sofre este processo de espalhamento é denominada irradiância solar direta. A parcela de radiação que atinge o solo proveniente de todas as direções da atmosfera após passar pelo processo de espalhamento é denominada de irradiância solar difusa. A Figura 3 mostra as componentes da irradiância solar após entrar na atmosfera terrestre.

Figura 3 – Componentes da irradiância solar na atmosfera



Fonte: Pereira et al. (2017)

A soma das parcelas direta e difusa da irradiância é denominada irradiância global horizontal e é dada pelas equações (3.1) e (3.2) (PEREIRA et al, 2017).

$$G = G_{dif} + G_{dir} \quad (3.1)$$

$$G = G_{dif} + G_n \cos \theta_z \quad (3.2)$$

Em que G é a irradiância global horizontal, G_{dif} é a irradiância difusa horizontal, G_{dir} é a irradiância direta horizontal, G_n é a irradiância direta normal e θ_z é o ângulo zenital, que é o ângulo formado entre o raio solar e a vertical local (PEREIRA et al, 2017).

A irradiância solar global é a variável mais importante nos estudos de previsão solar, por impactar diretamente no valor da potência de saída de um sistema FV. Por isso faz-se necessário adotar boas técnicas de previsão afim de prever com precisão tal parâmetro.

3.2 Modelos de previsão solar

Os modelos de previsão solar são importantes para diminuir riscos técnicos e custos relacionados a incerteza na geração e para viabilizar a programação de despacho de energia de fontes convencionais. As previsões podem considerar um único sistema ou vários sistemas FV. Os modelos de previsão podem ser classificados em modelos paramétricos e não-paramétricos (ALMEIDA et al, 2017).

Os modelos paramétricos consideram um conjunto de equações e parâmetros para descrever o comportamento de cada elemento do sistema fotovoltaico. Entre os modelos paramétricos, há vários submodelos (ALMEIDA, 2017):

- Modelos de decomposição para estimar as componentes direta e difusa da radiação global no plano horizontal;
- Modelos de transposição para estimar a irradiância eficaz no plano inclinado do painel fotovoltaico através dos valores de irradiância direta e difusa;
- Modelos de painel fotovoltaico para estimar a potência CC a partir da irradiância no plano inclinado e da temperatura ambiente;
- Modelos de inversor para estimar a potência CA a partir da potência CC;
- Modelos de cabeamento e demais componentes elétricos da instalação;

A incerteza nos modelos paramétricos depende do desempenho dos submodelos utilizados, da disponibilidade de informações acerca dos componentes do sistema e da incerteza

das medições. A principal vantagem destes modelos é poder prever a potência injetada na rede antes da instalação do sistema fotovoltaico.

Os modelos não-paramétricos não requerem qualquer conhecimento das características internas do sistema fotovoltaico. Tais modelos se baseiam no uso de séries temporais de variáveis meteorológicas e de potência gerada pelos painéis. Neste caso, a qualidade dos dados de entrada impacta diretamente no nível de incerteza destes modelos. Estas características revelam a principal desvantagem dos modelos não-paramétricos que é a necessidade de o sistema fotovoltaico estar em operação por tempo suficiente para a coleta de dados a serem utilizados na previsão. Dentre as vantagens dos modelos não-paramétricos destacam-se a compensação de erros associados as entradas, já que o modelo aprende a associar as entradas de valor incorretos com valores corretos na saída. Quanto apresentado ao mesmo erro na entrada na fase de previsão, a saída é automaticamente corrigida.

Dentre os modelos de previsão não-paramétricos, destacam-se os modelos estatísticos. Estes são baseados em técnicas de aprendizado de máquina que utilizam séries históricas e buscam minimizar os erros entre os valores previstos e medidos durante a fase de treinamento, e a partir disto se constrói a capacidade do modelo de generalizar a partir dos exemplos aprendidos para previsão. Dentre os modelos estatísticos, destacam-se métodos de regressão como o Auto Regressivo (AR – *Auto-Regressive*) e o Auto Regressivo Integrado de Médias Moveis (ARIMA – *Auto-Regressive Integreated Moving Average*), k-Vizinhos mais Próximos (k-NN – *k-Nearest Neighbors*), Floresta Aleatória (RF – *Random Forest*), Redes Neurais Artificiais (ANN – *Artificial Neural Network*) e Máquinas de Vetores de Suporte (SVM – *Support Vector Machine*) (ALMEIDA, 2017). Na Tabela 1 são mostrados alguns trabalhos desenvolvidos para a previsão de irradiância ultimamente técnicas de aprendizado de máquina.

Tabela 1 – Alguns trabalhos de previsão de irradiância desenvolvidos anteriormente

Autor	Técnica	Parâmetro	Local	Resultado
Bae et al., 2016	SVM	Irradiância	Coréia do Sul	$R^2 = 0,9182$ a $R^2 = 0,9685$
Huang et al., 2018	SVM	Irradiância	EUA	$R^2 = 0,9219$
Zhou et al.; 2017	SVM	Irradiância	China	RMSE = 77,64 W/m ² a RMSE = 136,58 W/m ²
Mukherjee et al.; 2018	ANN	Irradiância	Índia	RMSE = 57,25 W/m ² a RMSE = 75,77 W/m ²
Hussain e Alili, 2016	ARIMA	Irradiância	Emirados Árabes	$R^2 = 0,8573$ a $R^2 = 0,8863$

Fonte: Elaborada pelo autor

Este estudo propõe a utilização de Máquinas de Vetores de Suporte, mais precisamente sua versão adaptada para regressão, a Regressão por Vetores de Suporte (SVR – *Support Vector Rgression*) para a implementação do modelo de previsão de irradiância solar.

A SVR foi escolhida por apresentar boa capacidade de generalização e ser eficiente para tratar problemas não-lineares. Outra vantagem é por ter problema de otimização convexa, produzindo uma única solução global para o problema, apresentando uma vantagem em comparação com as redes neurais artificiais, que possuem problemas com mínimos locais (LORENA; CARVALHO, 2007).

4 REGRESSÃO POR VETORES DE SUPORTE

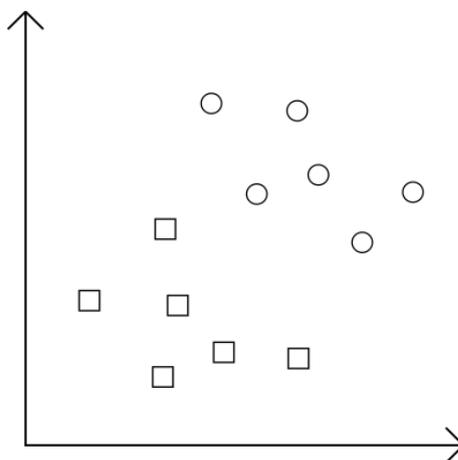
A regressão por vetores de suporte (SVR – *Support Vector Regression*) é um algoritmo adaptado do algoritmo das máquinas de vetores de suporte (SVM – *Support Vector Machine*) para propósitos de regressão. A SVM é um modelo de aprendizado de máquina desenvolvido por Vapnik em 1992 e é baseado na teoria do aprendizado estatístico desenvolvida por Vapnik e Chervonenkis em 1974 (SMOLA; SCHÖLKOPF, 2004).

Técnicas de aprendizado de máquinas podem ser supervisionadas ou não-supervisionadas. No aprendizado supervisionado, o algoritmo é exposto previamente às entradas e saídas desejadas e a partir disso é feita a extração de conhecimento capaz de produzir saídas corretas para novas entradas não introduzidas previamente. No aprendizado não-supervisionado não há exemplos a serem aprendidos, o algoritmo de aprendizado de máquina tende a representar ou agrupar as entradas segundo uma medida de qualidade (LORENA; CARVALHO, 2007). Neste trabalho foi adotado o aprendizado supervisionado em que a SVR é exposta a um conjunto de dados de teste $\{x_i, y_i\}$, e a partir disto tem que gerar uma função regressora capaz de prever corretamente saídas futuras.

4.1 Máquinas de Vetores de Suporte

Dado um conjunto de dados de i pontos que pertencem a duas classes diferentes, este conjunto de dados será linearmente separável se for possível traçar um hiperplano que separe completamente as duas classes, como ilustrado na Figura 4.

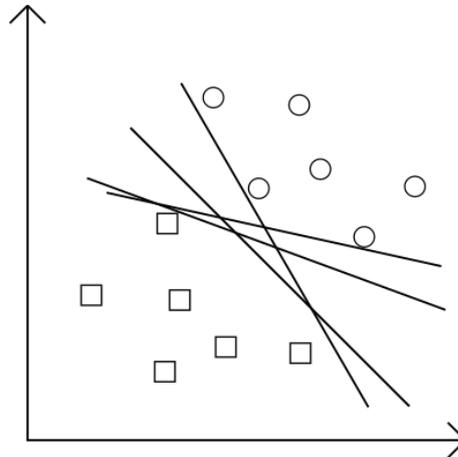
Figura 4 – Conjunto de dados linearmente separável



Fonte: elaborada pelo autor

Um número infinito de hiperplanos pode ser utilizado para separar o conjunto de dados da Figura 4, como mostra a Figura 5.

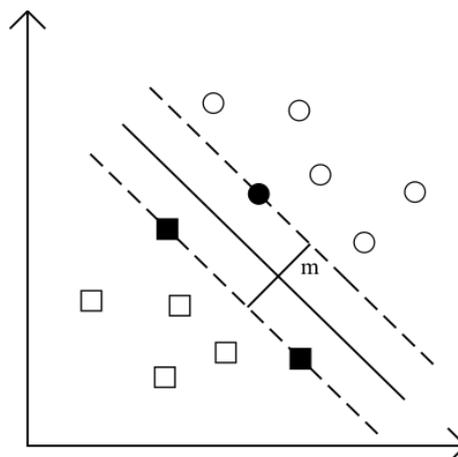
Figura 5 – Possíveis hiperplanos para a separação das classes



Fonte: elaborada pelo autor

Cada hiperplano da Figura 5 possui uma distância em relação às amostras mais próximas, esta distância é denominada margem de separação. A SVM tem como finalidade determinar o hiperplano ótimo que produza uma margem de separação m máxima, como é o caso mostrado na Figura 6.

Figura 6 – Hiperplano ótimo para a separação das classes



Fonte: elaborada pelo autor

O hiperplano ótimo (linha contínua) é obtido através dos pontos (em preto). Estes pontos são denominados vetores de suporte, daí vem o nome “Máquinas de Vetores de Suporte” para esta técnica de aprendizado de máquina.

4.1.1 Problema de otimização para as SVM

Considere um problema de classificação entre duas classes linearmente separáveis. Seja uma amostra de treinamento T do tipo $\{x_i, y_i\}$, em que x_i representa o i -ésimo valor de entrada e y_i a sua respectiva classe, em que y_i só pode assumir dois valores, nesse caso $+1$ ou -1 . Necessita-se determinar o hiperplano ótimo que separe as duas classes com a maior margem possível e sem erros. A equação de um hiperplano separador é definida como:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = 0 \quad (4.1)$$

Em que \mathbf{w} é um vetor normal ao hiperplano descrito, \mathbf{x} é um vetor de amostras e b é um bias. A notação $\mathbf{w} \cdot \mathbf{x}$ representa o produto escalar entre \mathbf{w} e \mathbf{x} . Seleciona-se \mathbf{w} e b de modo que os pontos que se situam nos extremos da margem, ou seja, os vetores de suporte, satisfaçam a seguinte equação:

$$|\mathbf{w} \cdot \mathbf{x} + b| = 1 \quad (4.2)$$

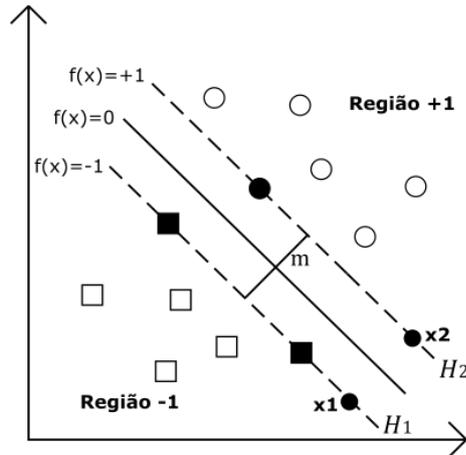
A partir de (4.2) pode-se definir uma função que classifique os elementos do conjunto T de dados:

$$f(\mathbf{x}) = \begin{cases} \mathbf{w} \cdot \mathbf{x} + b \geq +1 & \text{se } y_i = +1 \\ \mathbf{w} \cdot \mathbf{x} + b \leq -1 & \text{se } y_i = -1 \end{cases} \quad (4.3)$$

A equação (4.3) pode ser reescrita como:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall (\mathbf{x}_i, y_i) \in T \quad (4.4)$$

Considere um ponto x_1 no hiperplano $H_1: \mathbf{w} \cdot \mathbf{x} + b = -1$ e um ponto x_2 no hiperplano $H_2: \mathbf{w} \cdot \mathbf{x} + b = +1$, separados por uma distância perpendicular, conforme mostrado na Figura 7.

Figura 7 – Cálculo da distância d entre os hiperplanos H_1 e H_2 

Fonte: Elaborada pelo autor

Como \mathbf{x}_1 e \mathbf{x}_2 são vetores, $\mathbf{x}_1 - \mathbf{x}_2$ também será um vetor. Este vetor terá magnitude m e mesma direção do vetor \mathbf{w} que é perpendicular ao hiperplano de separação. Logo:

$$(\mathbf{x}_1 - \mathbf{x}_2) = m \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (4.5)$$

$$\mathbf{x}_2 = \mathbf{x}_1 + m \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (4.6)$$

Sabendo-se que $\mathbf{w} \cdot \mathbf{x}_1 + b = -1$ e $\mathbf{w} \cdot \mathbf{x}_2 + b = +1$, substitui-se (4.6) nestas equações. Desta forma, obtêm-se:

$$\mathbf{w} \cdot \left(\mathbf{x}_1 + m \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) + b = +1 \quad (4.7)$$

$$m \frac{\|\mathbf{w}\|^2}{\|\mathbf{w}\|} + \mathbf{w} \cdot \mathbf{x}_1 + b = +1 \quad (4.8)$$

Que resulta em:

$$m = \frac{2}{\|\mathbf{w}\|} \quad (4.9)$$

A partir da equação (4.9), verifica-se que a minimização da norma de \mathbf{w} produz uma margem de separação maior. Desta forma, o problema de otimização da SVM consiste em:

$$\text{Minimizar: } \frac{1}{2} \|\mathbf{w}\|^2 \quad (4.10)$$

$$\text{Sujeito a: } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall (\mathbf{x}_i, y_i) \in T \quad (4.11)$$

Este problema também é chamado de problema primal. As restrições em (4.11) asseguram que não haja dados de treinamento entre as margens de separação das classes. Por esse motivo este tipo de SVM é conhecida como SVM com margens rígidas.

4.2 Regressão por vetores de suporte

O problema de otimização descrito em (4.10) e (4.11) foi desenvolvido para casos de classificação de dados linearmente separáveis e não admite erros. Contudo, problemas de regressão podem ser resolvidos pelo método das SVM com algumas modificações.

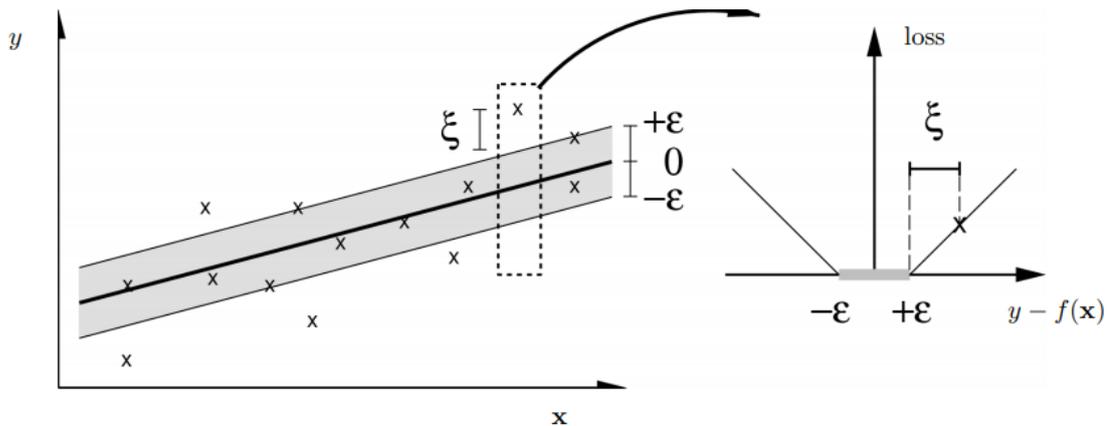
A partir de um certo conjunto de treinamento $\{(x_1, y_1), \dots, (x_m, y_m)\}$, a SVR determina uma função $f(x)$ que possua uma precisão ε em relação aos dados de treinamento, formando um tubo em torno da função. O problema de otimização em (4.10) e (4.11) passa a ser:

$$\text{Minimizar: } \frac{1}{2} \|\mathbf{w}\|^2 \quad (4.12)$$

$$\text{Sujeito a: } \begin{cases} y_i - \mathbf{w} \cdot \mathbf{x}_i + b \leq \varepsilon \\ \mathbf{w} \cdot \mathbf{x}_i + b - y_i \leq \varepsilon \end{cases} \quad (4.13)$$

Na maioria das aplicações, entretanto, alguns erros necessitam serem tolerados. Desta forma introduz-se variáveis de folga ξ_i, ξ_i^* no problema de otimização em (4.12) e (4.13), o que permite lidar com ruídos e *outliers* no conjunto de dados. Esta modificação permite que alguns pontos fiquem fora da região de tubo em torno da função de regressão. Tais pontos recebem uma penalidade associada através de uma função de perda ε -insensível que é apresentada em (4.14) e graficamente na Figura 8:

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{se } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{caso contrário} \end{cases} \quad (4.14)$$

Figura 8 – Expressão gráfica da função de perda ε -insensível

Fonte: Smola e Schölkopf (2001)

Com a introdução das variáveis de folga, onde ξ_i está associada com os dados acima da margem superior e ξ_i^* com os dados abaixo da margem inferior, o problema de otimização primal para a SVR em (4.12) e (4.13) passa a ser:

$$\text{Minimizar: } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (4.15)$$

$$\text{Sujeito a: } \begin{cases} y_i - \mathbf{w} \cdot \mathbf{x}_i - b \leq \varepsilon + \xi_i \\ \mathbf{w} \cdot \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (4.16)$$

A constante $C > 0$ em (4.15) pondera o nível de complexidade do modelo e o quão toleráveis serão os erros maiores que ε (SMOLA; SCHÖLKOPF, 2004). Somente pontos fora do tubo contribuem para o cálculo da equação (4.15).

O problema primal em (4.15) e (4.16) pode ser resolvido através da introdução de uma ferramenta matemática chamada função lagrangiana, que permite encontrar os máximos ou os mínimos de uma função sujeita a restrições. Tal função engloba as restrições do problema primal em (3.15) associadas a parâmetros chamados de multiplicadores de Lagrange.

A partir do problema de otimização (4.15) e (4.16), constrói-se a função lagrangiana:

$$\begin{aligned}
L = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) - \sum_{i=1}^m (\eta_i \xi_i + \eta_i^* \xi_i^*) \\
- \sum_{i=1}^m \alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w} \cdot \mathbf{x}_i + b) \\
- \sum_{i=1}^m \alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w} \cdot \mathbf{x}_i - b)
\end{aligned} \tag{4.17}$$

Onde $\alpha_i, \alpha_i^*, \eta_i$ e $\eta_i^* \geq 0$ são multiplicadores lagrangianos. A solução do problema primal é dada pelos pontos de sela da função lagrangiana em (4.17), onde necessita-se minimizar \mathbf{w}, b, ξ_i e ξ_i^* e ao mesmo tempo maximizar $\alpha_i, \alpha_i^*, \eta_i$ e η_i^* . Os mínimos da função L são obtidos a partir de derivadas parciais em relação a \mathbf{w}, b, ξ_i e ξ_i^* . Logo:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^m (\alpha_i^* - \alpha_i) \mathbf{x}_i = 0 \tag{4.18}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0 \tag{4.19}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \tag{4.20}$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \tag{4.21}$$

Substituindo (4.18), (4.19), (4.20) e (4.21) em (4.17), obtém-se:

$$\begin{aligned}
L: -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathbf{x}_i \cdot \mathbf{x}_j - \varepsilon \sum_{i=1}^m (\alpha_i - \alpha_i^*) + \\
\sum_{i=1}^m y_i (\alpha_i - \alpha_i^*)
\end{aligned} \tag{4.22}$$

A partir de (4.19) e (4.20) verifica-se que $\alpha_i = C - \eta_i$, como $\eta_i \geq 0$, os multiplicadores lagrangianos α_i, α_i^* são limitados por C . Constrói-se então um novo problema de otimização, que é chamado de dual.

$$\begin{aligned} \text{Maximizar: } & -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathbf{x}_i \cdot \mathbf{x}_j \\ & - \varepsilon \sum_{i=1}^m (\alpha_i - \alpha_i^*) + \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (4.23)$$

$$\text{Sujeito a: } \begin{cases} \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (4.24)$$

A partir da derivada (4.18) determina-se o valor de \mathbf{w} :

$$\mathbf{w} = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \mathbf{x}_i \quad (4.25)$$

Substituindo o valor de \mathbf{w} em (4.1), determina-se a função de regressão $f(\mathbf{x})$:

$$f(\mathbf{x}) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \mathbf{x}_i \cdot \mathbf{x}_j + b \quad (4.26)$$

A partir de (4.26) verifica-se que \mathbf{w} é resultado de uma combinação linear dos pontos de treinamento \mathbf{x}_i , ponderados por α_i e α_i^* agem como pesos. Nota-se que $f(\mathbf{x})$ não depende da dimensão dos dados de entrada, mas do número de α_i e α_i^* não-nulos, ou seja, depende apenas dos vetores de suporte.

Pode-se estabelecer uma relação entre as soluções do problema primal e dual através das condições de Karush-Kuhn-Tucker (KKT) para problemas de otimização. As condições KKT.

$$\alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w} \cdot \mathbf{x}_i + b) = 0 \quad (4.27)$$

$$\alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w} \cdot \mathbf{x}_i - b) = 0 \quad (4.28)$$

$$(C - \alpha_i) \xi_i = 0 \quad (4.29)$$

$$(C - \alpha_i^*) \xi_i^* = 0 \quad (4.30)$$

$$\xi_i \xi_i^* = 0 \quad (4.31)$$

$$\alpha_i \alpha_i^* = 0 \quad (4.32)$$

Da condição (4.32), conclui-se que nunca poderão existir variáveis duais α_i e α_i^* que sejam simultaneamente não-nulas. Verifica-se pelas condições (4.29) e (4.30) que só amostras (x_i, y_i) com correspondente $\alpha_i = C$ e $\alpha_i^* = C$ terão valores ξ_i e ξ_i^* não-nulos, portanto tais amostras estarão fora do tubo. Das condições (4.27) e (4.28), verifica-se que as variáveis duais α_i e α_i^* só serão não-nulas quando $|f(x_i) - y_i| \geq \varepsilon$, portanto quando as amostras estiveram fora do tubo. Para $|f(x_i) - y_i| < \varepsilon$, ξ_i e ξ_i^* devem ser não-nulos, portanto α_i e α_i^* devem ser nulos para que condição KKT seja satisfeita. Logo, percebe-se pelas conclusões obtidas que o valor de \mathbf{w} não dependerá de todos os pontos, mas somente daqueles com $\alpha_i, \alpha_i^* > 0$. Tais pontos são chamados de vetores de suporte.

Enquanto o valor de \mathbf{w} é obtido explicitamente através dos dados de treinamento, o valor de b é obtido através das condições KKT. De acordo com (4.29) e (4.30), $\xi_i = \xi_i^* = 0$ se $\alpha_i, \alpha_i^* < C$. A partir de (4.27) e (4.28), obtém-se o valor de b por:

$$b = y_i - \mathbf{w} \cdot \mathbf{x}_i - \varepsilon \quad (4.33)$$

$$b = y_i - \mathbf{w} \cdot \mathbf{x}_i + \varepsilon \quad (4.34)$$

Onde o valor final de b é tomado através da média de todos os valores de b calculados.

4.2.1 Funções Kernel

As SVR consideradas anteriormente tem desempenho satisfatório em aplicações com conjunto de dados com distribuição aproximadamente linear, já que as SVR descritas podem tolerar alguns ruídos e *outliers*.

Porém há muitas aplicações onde os dados de treinamento não possuem distribuição minimamente linear e o modelo da SVR não pode estimar uma função linear que aproxime os dados de entrada, mesmo após a introdução de variáveis de folga. As SVR lidam com isso através de uma função $\Phi(\mathbf{x})$ para mapear o conjunto de treinamento no espaço original ou de entrada X para um espaço de alta dimensão chamado de espaço de características F . Isto é feito através de funções kernel. Uma função kernel K é definida como:

$$K(\mathbf{x}_1, \mathbf{x}_2) = \Phi(\mathbf{x}_1) \cdot \Phi(\mathbf{x}_2), \forall (\mathbf{x}_1, \mathbf{x}_2) \in X \quad (4.35)$$

Algumas dos tipos de kernel mais utilizados em SVM incluem os polinomiais, sigmoidais e os RBF (do inglês, *radial basis function*). Um esboço destas funções é mostrado na Tabela 2.

Tabela 2 – Funções kernel mais comuns

Tipo de Kernel	Função $K(\mathbf{x}_i, \mathbf{x}_j)$	Parâmetros
Polinomial	$((\mathbf{x}_i \cdot \mathbf{x}_j) + 1)^d$	d
Sigmoidal	$\tanh(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + k)$	δ, k
RBF	$\exp(-\sigma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	σ

Fonte: Yu e Kim (2012).

Em SVR não-lineares, os produtos escalares entre vetores no problema de otimização são substituídos por produtos escalares entre funções de mapeamento $\Phi(\mathbf{x})$ no espaço de característica, que por sua vez são substituídos por funções kernel no espaço de entrada, por terem cálculo menos complexo e requererem menos capacidade computacional para o cálculo. Com a função kernel, o problema de otimização dual em (4.23) e (4.24) para SVR não-lineares passa a ser:

$$\begin{aligned} \text{Maximizar: } & -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) \\ & - \varepsilon \sum_{i=1}^m (\alpha_i - \alpha_i^*) + \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (4.36)$$

$$\text{Sujeito a: } \begin{cases} \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (4.37)$$

Outra vantagem da utilização de funções kernel é que não há mais a necessidade de determinar a função mapeadora $\Phi(\mathbf{x})$. A função de regressão $f(\mathbf{x})$ passa a depender da função kernel no espaço de entrada.

$$f(\mathbf{x}) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) K(\mathbf{x}_i, \mathbf{x}_j) + b \quad (4.39)$$

A partir da abordagem teórica da SVR, pode-se determinar os parâmetros a serem utilizados durante a calibração da SVR no programa implementado. A SVR foi implementada através da biblioteca Scikit-Learn da linguagem Python da seguinte forma:

$$reg = SVR(kernel = 'rbf', C = 1.0, epsilon = 0.01)$$

Em que:

- *kernel*: é o tipo de função kernel a ser utilizada. Após testes, o kernel RBF foi adotado por oferecer melhor desempenho.
- *epsilon*: é o diâmetro ϵ do tubo em torno da função de regressão. Os pontos dentro do tubo em torno da função de regressão não serão penalizados através da função de perda ϵ -insensível da equação (4.14). Após várias tentativas, adotou-se o valor de 0,01 para este parâmetro.
- *C*: é o parâmetro de regularização C da equação (4.15) que pondera a penalização dos erros e o nível de complexidade do modelo. Após várias tentativas, adotou-se o valor de 1,0 para este parâmetro.

Após a calibração, a SVR foi implementada através das seguintes funções da biblioteca Python: *reg.fit()* e *reg.predict()*. Em que *reg.fit()* treina a SVR com o conjunto de treinamento e *reg.predict()* gera as previsões a serem comparadas com o conjunto de teste. O programa implementado é apresentado detalhadamente no apêndice A.

5 METODOLOGIA E RESULTADOS

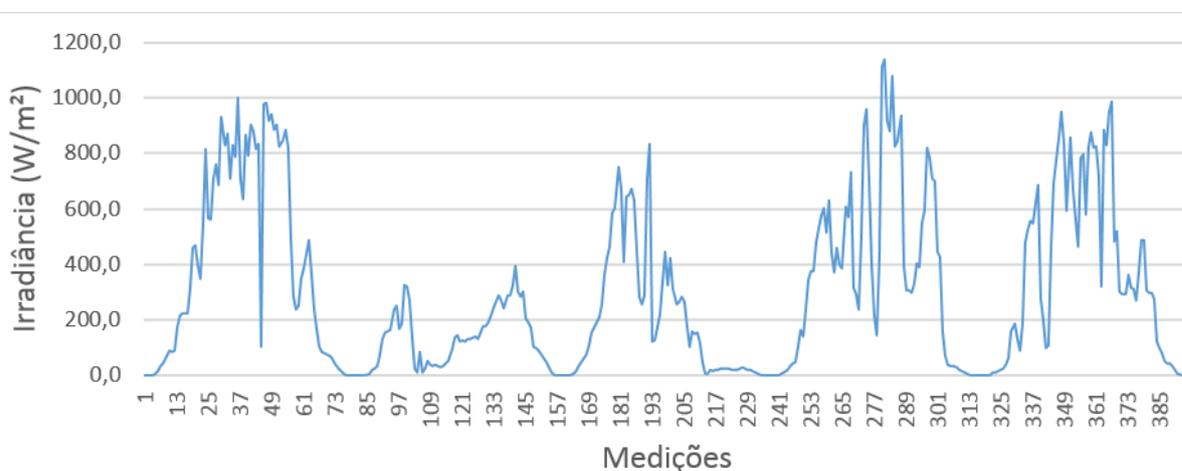
5.1 Base de dados

A base de dados utilizada neste estudo de previsão de irradiação solar foi coletada em Fortaleza-CE (latitude: $-0,3^{\circ} 43'$, longitude: $-38^{\circ} 32'$) por meio de um piranômetro (NRG Systems, LI-200SZ). Foram coletados dados de temperatura, irradiação solar diária e irradiância solar, dos quais somente os últimos foram utilizados no presente estudo. Os dados coletados para cada mês foram arranjados em planilhas em intervalos de 10 minutos, obtidos por meio de média aritmética dos dados processados pelo piranômetro a cada 2 segundos. A coleta de dados ocorreu entre abril de 2003 até abril de 2005 (exceto agosto, setembro e uma parte de junho de 2004), das 05:00 as 18:00, totalizando 53.744 medições distribuídas em cerca de 680 dias. Destas medições, 70% foram utilizadas para treinamento da SVR, enquanto os outros 30% foram utilizadas para teste e comparação com os valores previstos pela SVR.

Os dados das planilhas foram reorganizados em uma planilha única em arquivo .xlsx com 4 colunas e 53.744 linhas. As colunas representam respectivamente da esquerda para a direita o índice, o horário, o período horário e a irradiância global horizontal.

O Gráfico 5 mostra a variação da irradiância solar nos 5 primeiros dias de medições (entre 4 e 8 de abril de 2003). Oscilações abruptas durante as medições foram causadas pela movimentação das nuvens durante o dia. Em meses da estação chuvosa, há dias em que a irradiância solar atinge níveis muito baixos em comparação com a média anual.

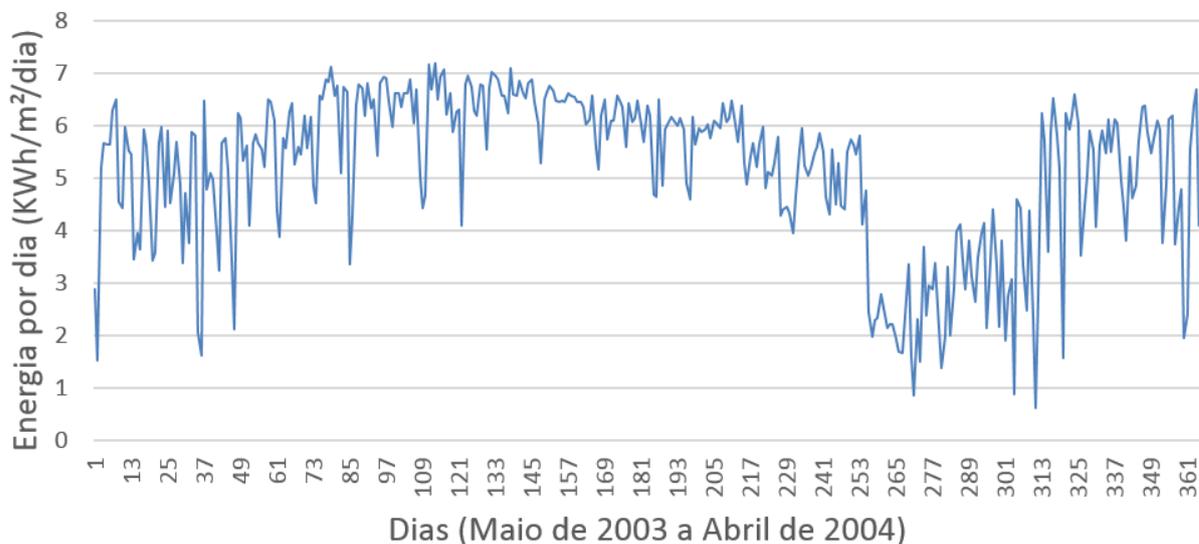
Gráfico 5 – Irradiância solar nos 5 primeiros dias de medições



Fonte: Elaborada pelo autor

O clima em Fortaleza é tropical semiúmido e se caracteriza por apresentar uma estação seca e uma estação chuvosa durante o ano. O período seco ocorre na segunda metade do ano e, por consequência, apresenta altos níveis de irradiação solar na superfície devido à baixa nebulosidade. O período chuvoso ocorre na primeira metade do ano e é marcado pelas chuvas, consequentemente apresentando baixos níveis de irradiação solar. Esta característica climática pode ser vista através dos dados medidos de irradiação solar diária nos painéis mostrados no Gráfico 6.

Gráfico 6 – Irradiação solar diária em Fortaleza durante um ano



Fonte: Elaborada pelo autor

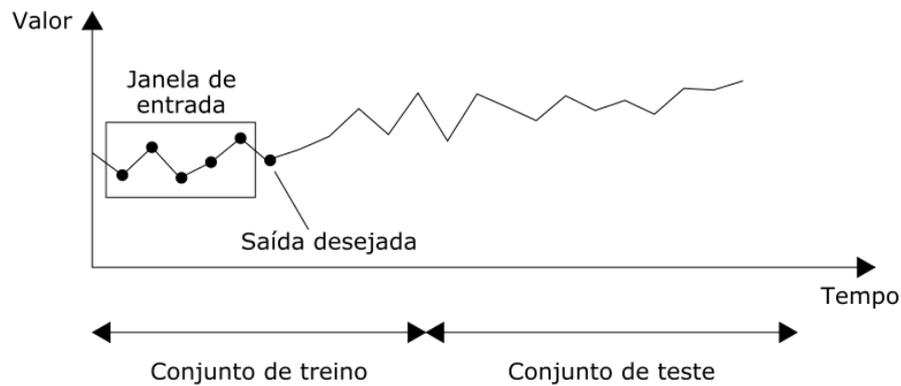
5.2 Implementação do código em Python para a previsão de irradiância solar

Para o estudo de previsão foi implementado na linguagem Python um código computacional utilizando a biblioteca de aprendizado de máquina do Python denominada Scikit-learn. A biblioteca Scikit-learn possui diversas funções para algoritmos de aprendizado de máquina como SVM, SVR e redes neurais. No Apêndice A é apresentado o código implementado para a previsão de irradiância solar

Os dados coletados de irradiância são considerados como uma série temporal, e, portanto, apresenta uma característica cíclica. Devido a isto, cada saída desejada (y) é obtida através de uma janela de previsão formada pelos N dados de entrada (X) obtidos em observações passadas. A janela fornece na entrada padrões ao algoritmo que a partir disto consegue realizar o aprendizado para prever saídas futuras. Um exemplo de uma série temporal

com janela de previsão de 5 elementos para prever um elemento na saída desejada é mostrado na Figura 9.

Figura 9 – Exemplo de uma série temporal com janela de previsão com 5 elementos



Fonte: Elaborada pelo autor

A janela de entrada deve ser suficiente grande para que o modelo aprenda a partir de todas as variações detectadas. Após várias tentativas, foi adotada uma janela de previsão igual a 60, a qual se considerou suficiente larga para oferecer uma boa capacidade de previsão para a SVR. Dessa forma cada saída a ser prevista é uma função das 60 medições anteriores, de forma que:

$$\begin{aligned}
 x_n &= f(x_{n-1}, x_{n-2}, x_{n-3} \dots x_{n-58}, x_{n-59}, x_{n-60}) \\
 x_{n+1} &= f(x_n, x_{n-1}, x_{n-2} \dots x_{n-57}, x_{n-58}, x_{n-59}) \\
 x_{n+2} &= f(x_{n+1}, x_n, x_{n-1} \dots x_{n-56}, x_{n-57}, x_{n-58}) \\
 &\vdots
 \end{aligned}$$

O programa implementado realiza a leitura e prepara a base de dados de forma a transformar a série temporal de dados medidos em uma sequência de janelas de entradas passadas e saídas de dados futuros a serem previstas. Após a preparação, os dados são divididos e os primeiros 70% são utilizados para treinamento e os 30% restantes para teste. Em seguida é feita a configuração da SVR. Após isso, o algoritmo executa as funções contidas na biblioteca Scikit-learn para realizar as previsões. Ao final da execução, erros são calculados e gráficos são traçados para comparar os valores medidos e previstos. Na Figura 10 é apresentado o fluxograma do algoritmo implementado.

Figura 10 – Fluxograma do algoritmo de previsão



Fonte: Elaborada pelo autor

5.3 Avaliação da precisão das previsões

Índices de erro médio quadrático (RMSE – *Root Mean Square Error*) e sua versão normalizada percentual (nRMSE – *Normalized Root Mean Square Error*) foram adotados para avaliar o desempenho do modelo de previsão e são calculados por meio das equações (5.1) e (5.2):

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (Y_{prev} - Y_{real})^2} \quad (5.1)$$

$$nRMSE(\%) = 100 \cdot \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N \left(\frac{Y_{prev} - Y_{real}}{Y_{med}} \right)^2} \quad (5.2)$$

Em que Y_{prev} representa o valor previsto pelo modelo, Y_{real} é o valor das medições, Y_{med} é o valor médio das amostras de teste no período considerado e N é o número de previsões. O RMSE foi adotado por dar maior peso a erros grandes, pois é preferível evitar poucos erros abruptos a muitos erros pequenos.

Também foi adotado o coeficiente de determinação R^2 que mede o quanto a variância de uma das variáveis pode ser explicada pela variância de outra. O coeficiente R^2 pode ser calculado pela expressão 4.3.

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_{real} - Y_{prev})^2}{\sum_{i=1}^N (Y_{real} - Y_{med})^2} \quad (5.3)$$

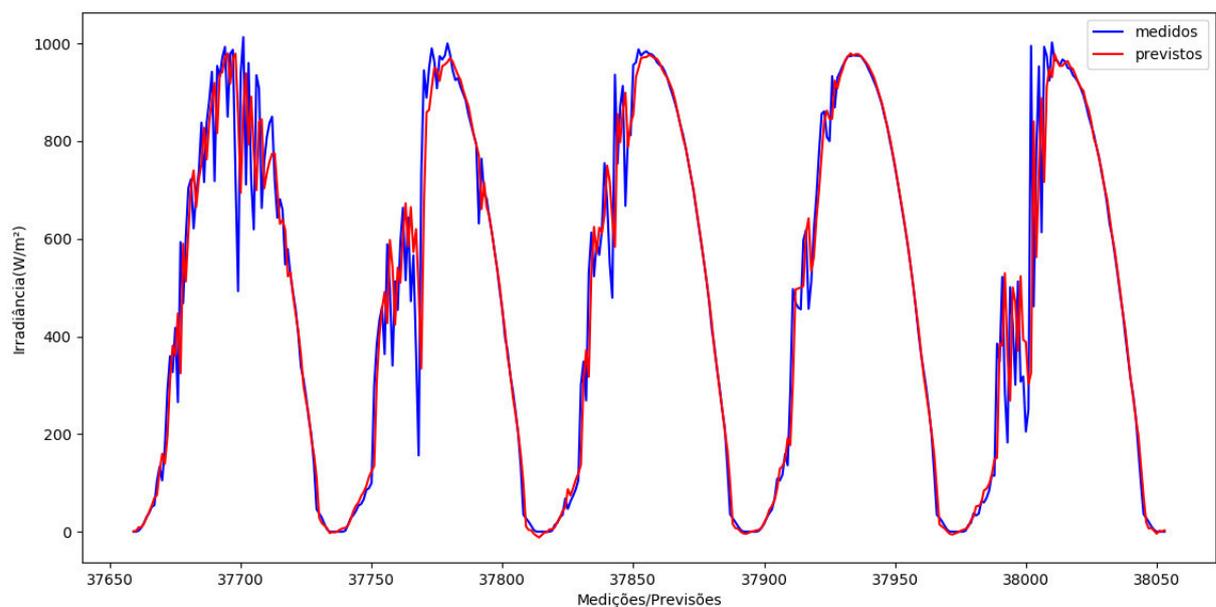
Em que R^2 pode assumir valores entre 0 e 1, em que 0 indica que não existe relação entre as variáveis e 1 indica uma relação linear perfeita.

Também foi realizada comparações com o método da persistência. Este método assume que a previsão no período $t + \lambda$ é igual ao valor atual no período t . Apesar de não parecer ser muito eficiente, este método é particularmente bom para horizontes de previsão de curto prazo.

5.4 Resultados

O Gráfico 7 ilustra o resultado do teste comparativo para um primeiro conjunto de dados medidos e previstos pelo programa para um intervalo de 5 dias entre 9 e 13 de outubro de 2004.

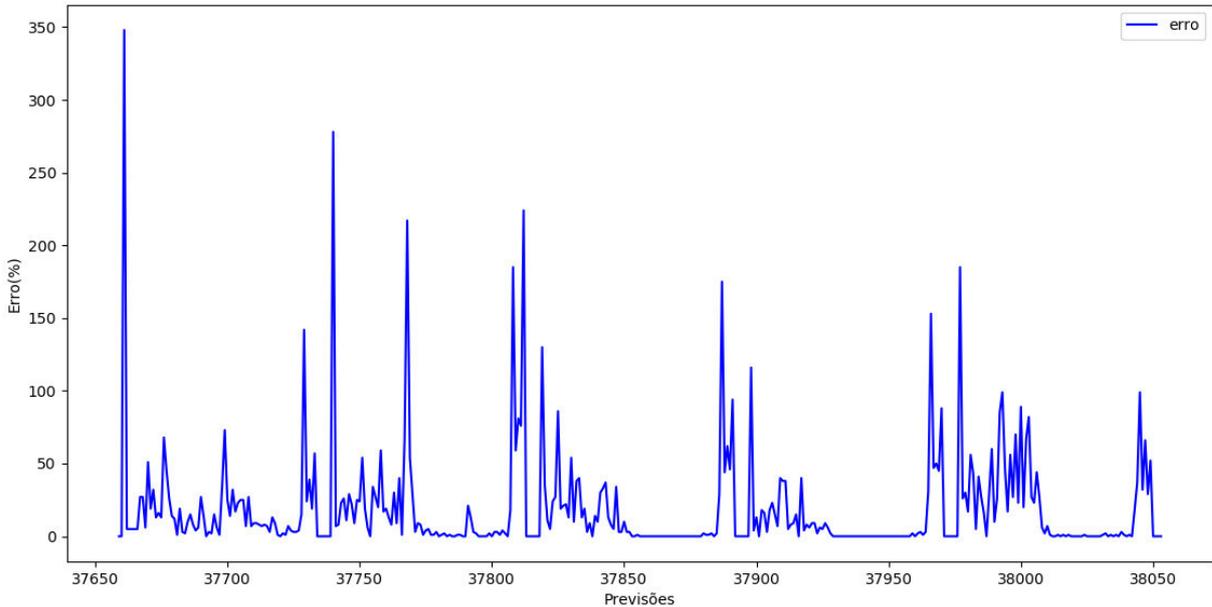
Gráfico 7 – Comparação entre medições e previsões para 5 dias de outubro de 2004



Fonte: Elaborada pelo autor

O Gráfico 8 mostra os erros percentuais de previsão para cada amostra durante o período entre 9 a 13 de outubro de 2004.

Gráfico 8 – Erro percentual das previsões para 5 dias de outubro de 2004



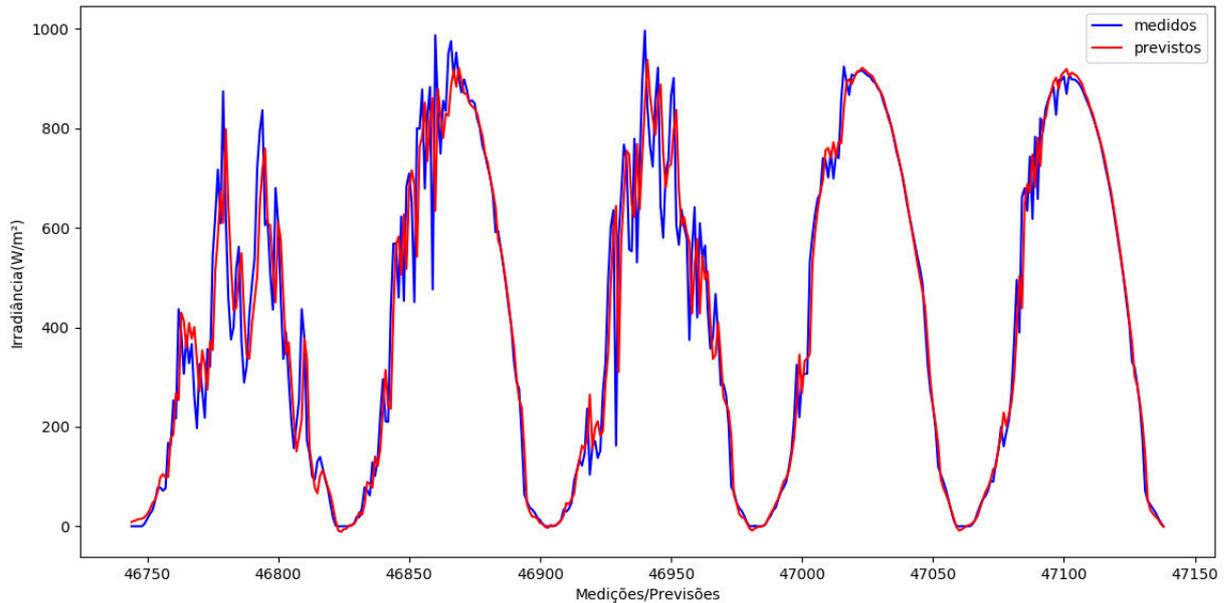
Fonte: Elaborada pelo autor

Para o primeiro conjunto de teste, a SVR foi capaz de produzir valores previstos dentro de uma margem tolerável de erros, especialmente durante os períodos ensolarados, à exceção de alguns picos abruptos no período do nascer e do pôr do sol, que podem ser explicados pelos valores baixíssimos de irradiância nestes períodos que acabam maximizando quaisquer erros obtidos. Tais erros podem ter sido causados pela falta de dados em alguns períodos de junho, agosto e setembro de 2004, o que pode ter comprometido em parte a capacidade de generalização da SVR. De qualquer forma, por se tratar de um horário de baixa demanda de energia, tais erros tendem a ter impactos mínimos no planejamento da operação dos sistemas de geração e distribuição de energia.

No primeiro teste, 52,40% das previsões tiveram um erro de até 5%, 62,78% das previsões obtiveram um erro de até 10% e 68,10% das previsões erraram por menos de 15%. Ao considerarmos o período de 10 às 16 horas, em que a irradiância é maior, 76,67% das previsões tiveram erro menor que 5%, em 87,22% das previsões o erro foi menor que 10% e em 90% das previsões produziram um erro menor que 15%.

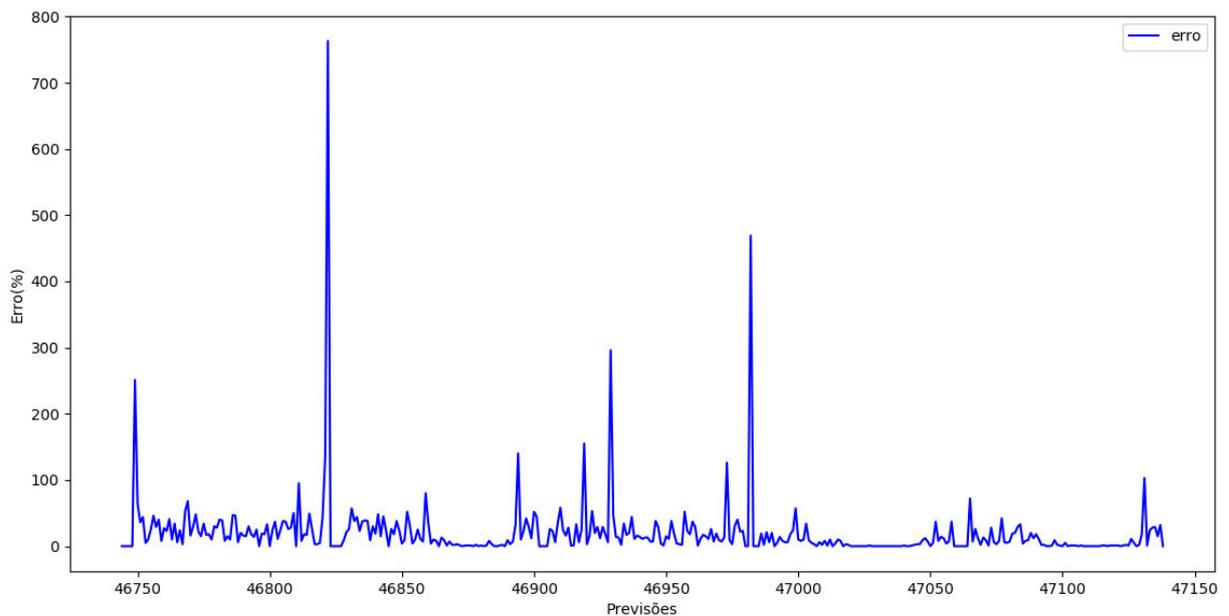
Os Gráficos 9 e 10 mostram respectivamente um segundo teste para um segundo conjunto de dados previstos e medidos para um intervalo de 5 dias entre 1 e 5 de fevereiro de 2005 e os erros percentuais de previsão para cada amostra durante o mesmo período de tempo.

Gráfico 9 – Comparação entre medições e previsões para 5 dias de fevereiro de 2005



Fonte: Elaborada pelo autor

Gráfico 10 – Erro percentual das previsões para 5 dias de fevereiro de 2005



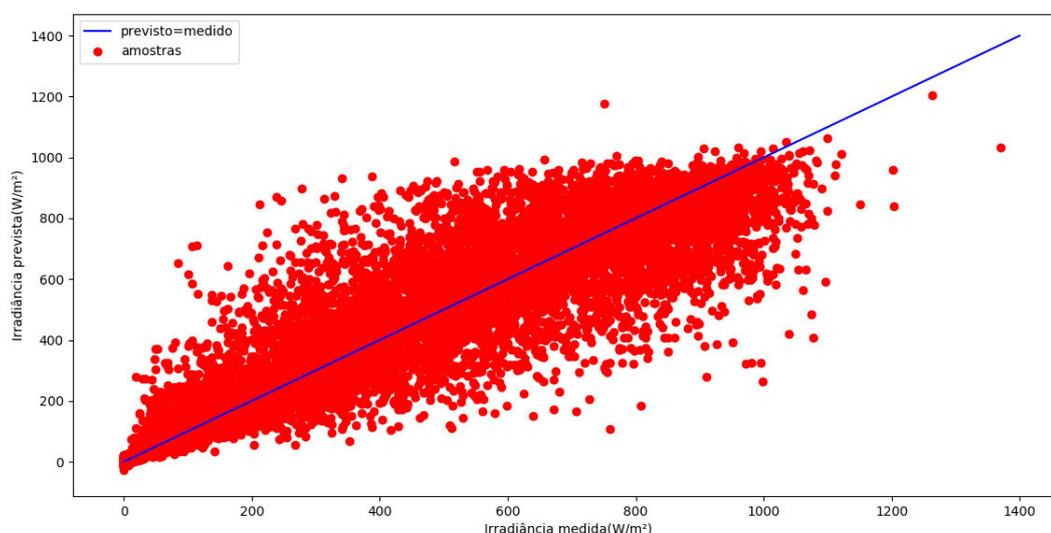
Fonte: Elaborada pelo autor

Para o segundo conjunto de teste, o comportamento da previsão permaneceu o mesmo, com picos abruptos durante o nascer e no pôr do sol. Percebe-se também que o erro percentual diminui à medida que há menos nebulosidade.

No segundo teste, 39,74% das previsões tiveram um erro de até 5%, 54,17% das previsões obtiveram um erro de até 10% e 62,78% das previsões erraram por menos de 15%. Entre 10 e 16 horas, o percentual de previsões com erros menores que 5%, 10% e 15% foi de 54,44%, 63,88% e 74,44% respectivamente.

Para o conjunto total de dados de treinamento, 35,17% das previsões tiveram um erro de até 5%, 47,09% das previsões tiveram erros de até 10% e 56,42% das previsões produziram erros de até 15%. Entre 10 e 16 horas, as previsões com erros menores que 5%, 10% e 15% representaram respectivamente 44,40%, 58,66% e 68,90% do total de previsões. O Gráfico 11 mostra a dispersão dos valores de irradiância medidos e previstos.

Gráfico 11 – Dispersão dos valores medidos e previstos de irradiância



Fonte: Elaborada pelo autor

O Gráfico 11 mostra que a maior parte dos valores previstos se aproximam dos valores medidos, já que a maior parte dos pontos se aproximam da linha diagonal. Os resultados da avaliação das previsões para cada mês do período de teste são sumarizados na Tabela 3.

Tabela 3 – Resultado da avaliação das previsões no período de teste

Mês	RMSE (W/m ²)	nRMSE (%)	R ²
Out/04	104,64	5,22	0,9120
Nov/04	113,07	8,47	0,8916
Dez/04	114,07	9,90	0,8736
Jan/05	100,24	8,57	0,8988
Fev/05	103,89	8,77	0,8969
Mar/05	111,10	13,17	0,8826
Abr/05	103,17	9,85	0,8921
Média	107,47	9,68	0,8932
Persistência	120,99	13,40	0,8646

Fonte: Elaborada pelo autor

A partir da Tabela 3, percebe-se que o nRMSE vai aumentando nos meses chuvosos, entre dezembro e abril, o que indica uma quantidade de erros abruptos maiores devido a variação brusca da irradiação causada pela nebulosidade. No mês de outubro, um mês da estação seca, o nRMSE foi menos da metade do nRMSE para março, um mês tipicamente chuvoso.

O valor do RMSE obtido para o período de teste esteve dentro da faixa dos valores de RMSE obtidos nos trabalhos anteriores apresentados na Tabela 1. Enquanto o RMSE obtido neste estudo foi de 107,47 W/m², os valores de RMSE dos trabalhos anteriores selecionados estiveram na faixa entre 57,25 W/m² e 136,58 W/m².

Comparar valores de RMSE não é o mais conveniente, já que este aumenta em locais de baixa latitude como é o caso de Fortaleza. O R² por sua vez, não depende da localização. Enquanto o valor de R² obtido neste estudo foi de 0,8932 os valores de R² nos trabalhos anteriores selecionados variaram entre 0,8573 e 0,9685, o que demonstra que as previsões deste estudo tiveram um desempenho satisfatório.

Verifica-se que o desempenho da SVR foi superior ao método da persistência em todas as ferramentas de análise, mesmo se tratando de um horizonte de previsão de apenas 10 minutos em que o método da persistência costuma ser mais eficiente.

6 CONCLUSÃO E TRABALHOS FUTUROS

6.1 Conclusão

Neste trabalho foi apresentado um programa computacional para previsão da irradiância solar com horizonte de previsão de 10 minutos, de modo a aumentar eficiência econômica dos mecanismos de compensação de energia e auxiliar os operadores do sistema elétrico a assegurar a estabilidade da rede.

O programa foi implementado na linguagem Python, através da biblioteca Scikit-learn, utilizando SVR. A SVR foi adotada por ser uma técnica de aprendizado de máquina que apresenta uma boa capacidade de generalização e ter bom desempenho para a resolução de problemas não-lineares. A partir dos valores previstos pelo programa comparou-se com os valores medidos.

A previsão obteve resultados satisfatórios em comparação com os resultados obtidos em trabalhos anteriores. O modelo obtido conseguiu se aproximar dos valores reais. Entretanto, observaram-se erros percentuais muito altos no período do nascer do sol e do pôr do sol, o que pode ser explicado pelos valores baixíssimos medidos para estes horários dos valores que acabam elevando os erros percentuais, mas como a geração nestes períodos é muito baixa os impactos destes erros agudos acabam sendo mínimos. Em 56% das previsões, foram obtidos erros menores que 15%. No período entre 10:00 e 16:00, no qual a irradiância e consequentemente a geração é maior, cerca de 70% das previsões obtiveram um erro menor que 15%.

Verificou-se que os erros tendem a variar sazonalmente ao longo do ano, com o erro tendendo a crescer em meses chuvosos e, portanto, de maior nebulosidade e diminuir em meses de seca. Enquanto em outubro, um mês de seca, o nRMSE observado foi de 5,22%, no mês de março, um mês chuvoso, o nRMSE foi de 13,17%.

Os erros podem ter sido causados pela falta de alguns dados em alguns períodos no intervalo de coleta de dados, o que pode ter comprometido em parte a capacidade de generalização da SVR. Recomenda-se uma base de dados maior e com dados disponíveis de forma contínua que consigam captar toda a variação anual de irradiância para reduzir os erros em projetos futuros.

6.2 Trabalhos Futuros

Ao fim deste estudo, nota-se algumas melhorias que podem ser feitas neste trabalho, das quais destacam-se:

- Adotar uma base de dados maior para realizar uma análise sazonal mais completa.
- Adicionar mais informações como entrada para a SVR, como temperatura e humidade.
- Realizar análises separadas para dias chuvosos, nublados e ensolarados para avaliar a variação do desempenho das previsões.
- Implementar uma janela de previsão maior para melhorar a capacidade de generalização do modelo e conseqüentemente diminuir os erros.

REFERÊNCIAS

ABSOLAR. **Potencial da energia solar ainda é subaproveitado no país**. Associação Brasileira de Energia Solar Fotovoltaica, 2016. Disponível em: <<http://www.absolar.org.br/noticia/noticias-externas/potencial-da-energia-solar-ainda-e-subutilizado-no-pais.html>>. Acesso em: 9 de outubro de 2018.

ALMEIDA, M. P. **Implicações técnicas da inserção em grande escala da geração fotovoltaica na matriz elétrica**. Tese de Doutorado em Ciências. Universidade de São Paulo. São Paulo, Maio de 2017.

ALMEIDA, M. P; MUÑOZ, M; PARRA, I. D. L; PERPIÑAN, O. **Comparative Study of PV Power Forecast Using Parametric and Nonparametric PV Models**. Solar Energy. v. 155, p. 854-866, 2017.

ANEEL. **Banco de Informações de Geração**. Disponível em: <<http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>>. Acesso em: 8 de setembro de 2018. 2017.

ANEEL. **ANEEL amplia possibilidades para micro e minigeração distribuída**. Agência Nacional de Energia Elétrica, 2015. Disponível em: <http://www2.aneel.gov.br/aplicacoes/noticias/Output_Noticias.cfm?Identidade=8955&id_area=90>. Acesso em: 20 de outubro de 2018.

BAE, K. Y; JANG, S. J; DAN, K. S. **Hourly Solar Irradiance Prediction Based on Support Vector Machine and Its Error Analysis**. IEEE Transactions on Power Systems. v. 32, n. 2. 2016.

BP. **Renewable Energy - BP Statistical Review of World Energy**, 67. ed, 2018. Disponível em: <<https://www.bp.com/content/dam/bp/en/corporate/pdf/energy-economics/statistical-review/bp-stats-review-2018-renewable-energy.pdf>>. Acesso em: 29 de setembro de 2018.

CHATURVEDI, D. K; ISHA. **Solar Power Forecasting: A Review**. International Journal of Computer Applications. v. 145, n. 6, p. 28-50, 2016.

DIAS, M. S. **O uso de máquina de suporte vetorial para regressão (SVR) na estimação e estrutura a termo da taxa de juros do Brasil**. Dissertação de Mestrado – Departamento de Matemática. Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2007.

EPE. **Balanco Energético Nacional 2017**. Relatório Síntese – Ano Base 2016. Rio de Janeiro. 2017.

FRANKFURT SCHOOL. **Global Trends in Renewable Energy Investment 2018**. Frankfurt. 2018.

FRAUNHOFER INSTITUTE FOR SOLAR ENERGY SYSTEMS. **Photovoltaics Report**. Freiburg, 2018.

HUANG, C; WANG, L; LAI, L.L. **Data-Driven Short-term Solar Irradiance Forecasting Based on Information of Neighboring Sites**. IEEE Transactions on Industrial Electronics. 2018.

HUSSAIN, S; ALILI, A, A. **Day Ahead Hourly Forecast of Solar Irradiance for Abu Dhabi, UAE**. 2016 the 4th IEEE International Conference on Smart Energy Grid Engineering. 2016.

IRENA. **Global Energy Transformation – A roadmap to 2050**, 2018.

LIMA, M. A. B. F. **Teoria do portfólio aplicada à previsão de recursos solar e eólico**. Dissertação de Mestrado em Engenharia Elétrica. Universidade Federal do Ceará. Fortaleza, 2016.

LORENA, A. C; CARVALHO, A. C. P. L. F. **Uma introdução às Support Vector Machines**. Revista de Informática Teórica e Aplicada. Porto Alegre: Universidade Federal do Rio Grande do Sul, v. 14, n. 2, p. 43-67, 2007.

MME. **Programa de Desenvolvimento de Geração Distribuída de Energia Elétrica. Ações de estímulo à geração distribuída com base em fontes renováveis**. Brasília, 2015.

Disponível

em: <<http://www.mme.gov.br/documents/10584/3013891/15.12.2015+Apresenta%C3%A7%C3%A3o+ProGD/bee12bc8-e635-42f2-b66c-fa5cb507fd06?version=1.0>>. Acesso em: 20 de outubro de 2018.

MME. **Plano Decenal de Expansão de Energia 2026**. Brasília, 2017.

MUKHERJEE, A; AIN, A; DUSGUPTA, P. **Solar Irradiance Prediction from Historic Trends Using Deep Neural Networks**. 2018 IEEE International Conference on Smart Energy Grid Engineering. 2018.

NREL. **Solar Resource Glossary**. National Renewable Energy Laboratory. 2018. Disponível em: <<https://www.nrel.gov/grid/solar-resource/solar-glossary.html>>. Acesso em: 20 de outubro de 2018.

ONS. **Boletim Mensal de Geração Solar Fotovoltaica Setembro/2018**. Brasília. 2018.

PEREIRA, E. B; MARTINS, F. R; GONÇALVES, A. R; COSTA, R. S; LIMA, F. J. L. D; RÜTHER, R; ABREU, S. L. D; TIEPOLO, G. M; PEREIRA, S. V; SOUZA, J. G. D. **Atlas Brasileiro de Energia Solar**, 2^a ed, São José dos Campos: INPE, 2017.

REN 21. **Renewables 2018: Global Status Report**, Paris, p. 19-67, 2018.

SAMANTA, M; SRIKANTH, B. K; YERRAPRAGADA, J, B; **Short-Term Power Forecasting of Solar PV Systems Using Machine Learning Techniques**. 2014.

SAUAIA, R. L. **Energia Solar Fotovoltaica: Panorama, Oportunidades e Desafios.**

Associação Brasileira de Energia Solar. 1º Worskhop de Energia Solar Fotovoltaica do Operador Nacional do Sistema Elétrico – ONS, 2018. Disponível em:

<<http://ons.org.br/AcervoDigitalDocumentosEPublicacoes/02-Sector-Dr.RodrigoLopesSuaia-Absolar.pdf>>.

SILVA, L. C. C. T. **Decentralized Solar PV Generation Forecating Based on Peer-to-peer Approach.** Dissertação de Mestrado Integrado em Engenharia de Energia e Meio Ambiente. Universidade de Lisboa. Lisboa, 2018.

SMOLA, A. J; SCHÖLKOPF, B. **Learning with Kernels.** Cambridge: MIT Press, p. 17, 2001.

SMOLA, A. J; SCHÖLKOPF, B. **A tutorial on Support Vector Regression.** Statistics and Computing. Norwell: Kluwer Academic Publishers, v. 14, n. 3, p. 199-226, 2004.

ZHOU, L; WU, H; XU, T; MEI, F; LI, Y; YUAN, X; LIU, H. **Ultra-Short Term Hybrid Power Forecasting Model for Photovoltaic Power Station with Metereological Monitoring Data.** 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation. 2017.

UN – DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS. **World Population Prospects The 2017 Revision.** New York, 2017.

UNFCCC. **Historic Paris Agreement on Climate Change: 195 nations set path to keep temperature rise well below 2 degree Celsius.** United Nations Framework Convention on Climate Change. New York, 2015. Disponível em:<<https://unfccc.int/news/finale-cop21>>. Acesso em: 07 de Outubro de 2018.

YU, H; KIM, S. **SVM Tutorial – Classification, Regression and Ranking.** Handbook of Natural Computing. Berlin: Springer, p. 479-506, 2012.

APÊNDICE A – CÓDIGO COMPUTACIONAL

```

#-----
#importa pacotes e bibliotecas
#-----
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from math import sqrt
#-----
#extrai os dados da planilha do excel para “data”
#em que:
# index = 1º coluna de “data” – índice das medições
# hora = 3º coluna de “data” - horas
# rad = 4º coluna de “data” - valores de irradiância
#-----
data = pd.read_excel(r'C:\...\Dados.xlsx')
index = data.iloc[:, 0]
hora = data.iloc[:, 2]
rad = data.iloc[:, 3]
#-----
#transforma os dados de irradiância em “rad” numa série temporal do tipo “DataFrame”
#-----
df = pd.DataFrame(rad)
#-----
#função que cria a janela de previsão com N elementos
#-----
def create_lags(df, N):
    for i in range(N):
        df['Lag' + str(i+1)] = df.temp.shift(i+1)
    return df
#-----

```

```

#cria uma janela de previsão com 60 elementos
#-----
df = create_lags(df,60)
df = df.dropna()
#-----
#define e normaliza as entradas X e saídas Y e divide em 70% para treino e 30% para teste
#-----
y = df.temp.values/1000
X = df.iloc[:, 1:].values/1000
train_idx = int(len(df) * .7)
X_train, y_train, X_test, y_test = X[:train_idx], y[:train_idx], X[train_idx:], y[train_idx:]
#-----
#cria uma SVR, treina e prediz os valores
#-----
reg = SVR(kernel = 'rbf', C=1.0, epsilon=0.01)
reg.fit(X_train, y_train)
y_pred = reg.predict(X_test)
#-----
#divide os valores de hora e índice em 70% treino e 30% teste
#-----
index = index.iloc[60:]
hora = hora.iloc[60:]
y_test = y_test*1000
y_pred = y_pred*1000
index_train, index_test = index[:train_idx], index[train_idx:]
hora_train, hora_test = hora[:train_idx], hora[train_idx:]
#-----
#plota o gráfico das medições e previsões entre 9 e 13 outubro de 2004
#-----
plt.plot(index_test[20:415], y_test[20:415], color = 'blue', label = 'medidos')
plt.plot(index_test[20:415], y_pred[20:415], color = 'red', label = 'previstos')
plt.xlabel('Medições/Previsões')
plt.ylabel('Irradiância(W/m²)')
plt.legend(loc = 'upper right')

```

```

plt.show()
#-----
#plota o gráfico das medições e previsões entre 1 e 5 de fevereiro de 2005
#-----
plt.plot(index_test[9105:9500], y_test[9105:9500], color='blue', label='medidos')
plt.plot(index_test[9105:9500], y_pred[9105:9500], color='red', label='previstos')
plt.xlabel('Medições/Previsões')
plt.ylabel('Irradiância(W/m²)')
plt.legend(loc='upper right')
plt.show()
#-----
#plota o gráfico de dispersão dos valores medidos e previstos
#-----
p = np.array([0, 1400])
plt.plot(p, p, color='blue', label='previsto=medido')
plt.scatter(y_test, y_pred, color='red', label='amostras')
plt.xlabel('Irradiância medida(W/m²)')
plt.ylabel('Irradiância prevista(W/m²)')
plt.legend(loc='upper left')
plt.show()
#-----
#função que calcula o erro percentual para cada previsão
#em que:
# erro = vetor de erros percentuais para cada amostra
#-----
def erro(test, pred):
    erro = np.repeat(0, len(test))
    for i in range(len(test)):
        if test[i]==0:
            erro[i] = 0
        else:
            erro[i] = 100 *(np.abs(test[i]-pred[i]))/test[i]
    return erro
erro = erro(y_test, y_pred)

```

```

#-----
#imprime o gráfico dos erros percentuais para 9 e 13 de outubro de 2004
#-----
plt.plot(index_test[20:415], erro[20:415], color='blue', label ='erro')
plt.xlabel('Previsões')
plt.ylabel('Erro(%)')
plt.legend(loc = 'upper right')
plt.show()
#-----
#imprime o gráfico dos erros percentuais para 1 e 5 de fevereiro de 2005
#-----
plt.plot(index_test[9105:9500], erro[9105:9500], color='blue', label ='erro')
plt.xlabel('Previsões')
plt.ylabel('Erro(%)')
plt.legend(loc = 'upper right')
plt.show()
#-----
#função que calcula os erros maiores que n%
#em que:
# erro = vetor de erros percentuais
#-----
def percent_erro(erro, n):
    cont = 0
    for i in range(len(erro)):
        if erro[i]<=n:
            cont = cont + 1
    return 100*cont/(len(erro))
#-----
#função que calcula os erros menores que %n entre um período de horas
#em que:
# hor = vetor de horas
# h_ini = hora inicial
# h_fin = hora final
#-----

```

```

def percent_erro_hora(erro, hor, h_ini, h_fin, n):
    cont_erro = 0
    cont_hora = 0
    for i in range(len(erro)):
        if hor.iloc[i]>=h_ini and hor.iloc[i]<h_fin:
            cont_hora = cont_hora + 1
            if erro[i]<=n:
                cont_erro = cont_erro + 1
    return 100*cont_erro/cont_hora
percent_erro_hora(erro, hora_test, 10, 16, 5)
#-----
#cálculo do erro médio quadrático
#-----
rmse = sqrt(mean_squared_error(y_test, y_pred))
rmse
#-----
#cálculo do coeficiente de determinação R2
#-----
r2_score(y_test, y_pred)
#-----
#função para cálculo do erro médio quadrático normalizado
#em que:
# test = dados medidos do período de teste
# pred = dados previstos para o período de teste
#-----
def n_rmse(test, pred):
    erro = np.repeat(0, len(test))
    media = (sum(test)/len(test))
    soma = 0
    for i in range(len(test)):
        erro[i] = ((pred[i] - test[i])/(media))**2
        soma = soma + erro[i]
    return 100*(soma/(len(test)))*(1/2)
n_rmse(y_test, y_pred)

```

```
#-----  
#Método da persistência  
#-----  
w = df.iloc[:, 1].values  
w_train, w_test = w[:train_idx], w[train_idx:]  
rmse = sqrt(mean_squared_error(y_test, w_test))  
rmse  
r2_score(y_test, w_test)  
n_rmse(y_test, w_test)
```