

# A Transformação do Mercado de Energias Renováveis Utilizando a Análise de Dados

## *Transformation of the Renewable Energy Market Using Data Analysis*

Leonardo Macedo Gonçalves<sup>1</sup>, Vinícius de Melo Puglia<sup>2</sup>, João Carlos Lopes. Fernandes<sup>3</sup>, Renato de Brito Sanchez<sup>4\*</sup>

**Resumo:** O conceito de Big Data, vem revolucionando o mercado na era digital, podendo revolucionar negócios, reunindo grandes dados de informações e auxiliando para tomada de decisões e previsões de mercado, gerando lucro e reduzindo custos. A pesquisa tem como objetivo criar um modelo de análise de dados para processos de energias renováveis, contribuindo com uma análise preditiva de dados de energia solar e eólica, prevendo as condições climáticas com antecedência e permitindo que as plantas de produção aumentem significativamente sua produção, foi usado no projeto a pesquisa quantitativa, no qual contribuiu para compreender quais os principais softwares serão utilizados para conclusão do projeto, gerando um sistema de análise que poderá contribuir para diversos mercados. Com a análise dos dados, será possível obter os modelos de previsão de produção de energia elétrica, em fontes renováveis, nomeadamente sistemas solares fotovoltaicos e turbinas eólicas. Permitindo a previsão da produção de energia elétrica a curto e médio prazo, permitindo assim o balanceamento eficaz do fluxo de energia gerado.

**Palavras-Chave:** Big Data. Data Science. Fotovoltaico. Eólica.

**Abstract:** The Big Data concept has been revolutionizing the market in the digital age and can revolutionize business by gathering big data and helping to make decisions and market forecasts, generating profit and reducing costs. The research aims to create a data analysis model for renewable energy processes, contributing to a predictive analysis of solar and wind energy data, predicting weather conditions in advance and allowing production plants to significantly increase their production. Quantitative research was used in the project, which contributed to understand what the main software will be used to complete the project, generating an analysis system that can contribute to several markets. By analyzing the data, it will be possible to obtain models for forecasting the production of electricity from renewable sources, namely solar photovoltaic systems and wind turbines. Allowing the forecast of short- and medium-term electric power production, thus allowing the efficient balancing of the generated energy flow.

**Keywords:** Big Data. Data Science. Photovoltaic. Wind.

<sup>1</sup>Acadêmico no Curso de Engenharia Mecatrônica, Centro Universitário ENIAC, leomacedo86@hotmail.com.

<sup>2</sup> Pesquisador do NuPE, Engenheiro Eletricista; Engenheiro de Produção. Centro Universitário ENIAC, vinicius\_puglia@hotmail.com.

<sup>3</sup> Professor Doutor e Pesquisador do NuPE, Centro Universitário ENIAC, jlopesf@uol.com.br.

<sup>4</sup> Professor Mestre e Pesquisador do NuPE, Centro Universitário ENIAC, renatobritosanchez@gmail.com. \*Para quem o contato deve ser realizado.

### I. INTRODUÇÃO

Fundamentalmente o progresso significa, pegar a tecnologia e torná-la funcional para a humanidade, sempre com a expectativa de melhorar a experiência de todos ao nosso redor. Com a chegada da indústria 4.0, o termo *Big Data* ganhou muita notoriedade no mercado, trazendo junto dele e fortalecendo diversos seguimentos em *Big Data*, tais como *Data Science* e *Machine Learning*.

Com a revolução tecnológica nunca a humanidade produziu tantos dados como na última década. Cada vez mais são gerados e armazenados dados, que têm um potencial gigantesco de serem transformados em valiosas informações, quanto maior a informações das empresas na década atual, maior o seu poder de decisão e maiores as chances de sucesso.

E como extrair informações a partir de uma infinidade de bits que inundam os bancos de dados em todo o Planeta? É aí que entra área mais conhecida como *Data Science*, com suas técnicas e metodologias.

Desde o start da era da transição energética, estão cada vez mais sendo feitos esforços para eliminar as emissões que contaminam a atmosfera e a busca para reduzir o impacto ambiental passam por querer eliminar o uso do petróleo e do gás natural. Além de reduzir sua utilização, mesmo que continuem sendo explorados. Se isso fosse feito de uma melhor maneira possível? E se buscarmos aprimorar a implementação das energias renováveis? A tecnologia tem avançado para encontrar soluções industriais, baseadas em Inteligência Artificial e em big data, a fim de otimizar a exploração energética e diminuir o gasto excessivo.

Com o mercado energético em alta, a combinação de fontes de energias renováveis, tem impactado uma grande transformação no setor de energia elétrica, sendo uma rede inteligente, sustentável e designada por “*smart grid*”. Neste setor, as redes de energia elétrica deixam de estar presentes em uma estrutura centralizada e passa a se concentrar em uma estrutura aberta. Assim dividida em microproduções de energias renováveis como a solar e fotovoltaica.

De acordo com a última edição de 2017 do *US Energy Information Administration (EIA) Electric Power Monthly*, as fontes de energia renováveis representaram 19,35% da geração elétrica líquida dos EUA durante o primeiro trimestre de 2017.

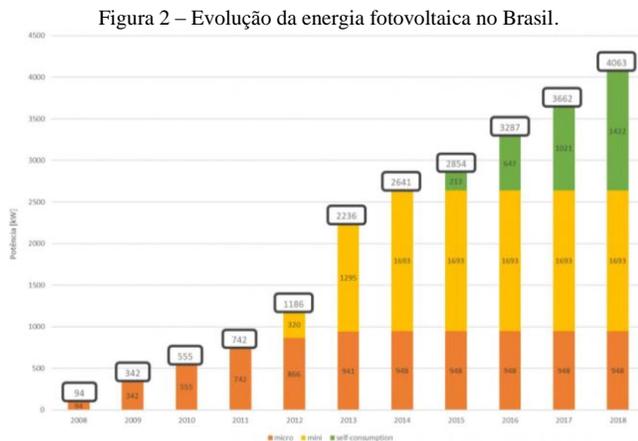
No Brasil, os dados do boletim da Câmara de Comercialização de Energia Elétrica — CCEE indicam que a geração de energia eólica no país cresceu 15,5% no primeiro semestre de 2019 em relação ao mesmo período em 2018. Essa participação provavelmente aumentará à medida que o setor de energia renovável continuar adotando ferramentas e tecnologia de análise dados.

Na Figura 1 é expresso o infográfico do crescimento da geração de energia eólica instalada no Brasil.



Fonte: Revista Forbes, 2019.

Na Figura 2 é expresso o infográfico do crescimento da geração de energia fotovoltaica instalada no Brasil.



Fonte: Centro de Estratégias em Recursos Naturais e Energia, 2019.

Mesmo que a geração de energia renovável seja um dos principais objetivos para os próximos anos, principalmente pela sustentabilidade. Os recursos são intermitentes e imprevisíveis, tais como vento e luz solar, dificultando a produção eficiente. E como resultado, aumenta a dificuldade para as usinas solares e eólicas operarem em seu potencial

máximo, obrigando os ao aumento da quantidade de painéis ou turbinas para alcançar a produção desejada. Assim elevando custos de investimentos e reduzindo lucros, e até mesmo a não alcançar a produção desejada, obrigando a obtenção de energia no mercado, impactando mais ainda os custos de produção e consumo.

O Big Data contribui para esse cenário. As usinas de energia solar e eólica sempre coletaram e geraram dados. Com as novas tecnologias, as análises preditivas e técnicas de *Machine Learning*, esses dados agora podem ser combinados com dados meteorológicos e por satélite. Esta tecnologia de previsão de energia solar e eólica pode mapear as condições climáticas com bastante antecedência, permitindo que as plantas renováveis aumentem significativamente sua produção. O conhecimento do comportamento da produção solar fotovoltaica e eólica tem grande importância nesta gestão inteligente, conduzindo a ganhos econômicos.

Deste forma, este trabalho visa criar um sistema de análise dos dados, tornando possível obter os modelos de previsão de produção de energia elétrica, em fontes renováveis, nomeadamente sistemas solares fotovoltaicos e turbinas eólicas, permitindo a previsão da produção de energia elétrica a curto e médio prazo, assim como o balanço eficiente do fluxo de energia gerada para uma produção futura.

## I. MATERIAIS E MÉTODOS

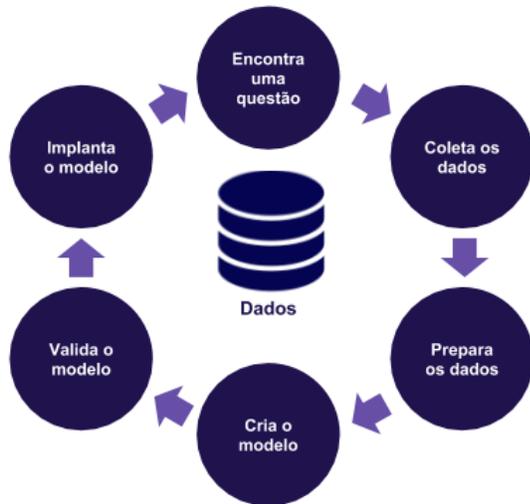
As principais ferramentas utilizadas para desenvolvimento do projeto, são baseadas na linguagem de programação *Python*, que é uma ferramenta muito abrangente para *data science*, com diversas bibliotecas atreladas, no que facilitam o processo de tratamento de dados. Para ambientação e criação do projeto será usado o programa Anaconda, que é um software desenvolvido com todas as bibliotecas e softwares de Python em seu banco de dados. Dentro do programa, será usado Software *Jupyter Notebook*, para processamento e codificação dos dados, a principal biblioteca para tratamento dos dados será utilizada será o Pandas, para tratamento de dados com *Python* e cálculos estatísticos, e para os cálculos matemáticos, em caso de matrizes será utilizada.

Para retiradas dos dados no banco disponível, e integrar esses dados brutos no sistema usarei o sistema da *Microsoft Integration Services*, que usa a linguagem SQL para essa consulta estruturada e também o sistema ETL (extract transform load), e para a plotagem dos gráficos será utilizado a biblioteca *python Matplotlib no Jupyter notebook*. Com toda a pesquisa aplicada, o trabalho será experimental, aprofundando nas bibliotecas, estruturas, módulos e kits de ferramentas para desenvolvimento de data science. Os dados para desenvolvimento da pesquisa serão retirados do INMET – Instituto Nacional de Meteorologias, usando algoritmos de programação e modelagem matemática para estruturação, e comparação dos dados, tendo como foco principal estatística aplicada.

O desenvolvimento da pesquisa para o projeto, se constitui bom base nas tecnologias da Indústria 4.0, e como contribuir com o desenvolvimento do setor, tendo como foco principal a produção de energias renováveis e limpas, para microproduções sustentáveis, podendo contribuir para a descentralização do setor e desenvolvimento econômico.

A criação do projeto de *Data Science* se consiste em algumas etapas principais e cruciais para um projeto robusto. Na Figura 3, descreve um processo de análise de dados, que consiste em seis etapas.

Figura 3 – Processo de Coleta de Dados



Fonte: Data Science Central, 2019.

Na primeira etapa consiste na compreensão do negócio ou, que tem como função entender o que pretende realizar numa perspectiva maior, na segunda etapa está presente a coleta de dados, onde é realizada a aquisição e familiarização dos dados. Na terceira Etapa é executada a preparação dos dados, que é referente a decisão sobre o volume e o tipo de dados a usar no processo de análise. Na quarta etapa consiste na criação do modelo, que é a fase onde é selecionado o algoritmo ou a modelação que vai utilizar na análise de dados. Na quinta etapa, vem a validação do modelo, verificando se alcança adequadamente os objetivos pretendidos. Na sexta fase é perceptível a implantação do modelo, que pode variar de um simples relatório á um complexo processo de prospecção de dados repetível em todo o negócio.

Com as ferramentas definidas, a primeira etapa é a extração dos dados das fontes consultadas e posteriormente realizar os primeiros tratamentos desses dados brutos, por exemplo transformá-los em um arquivo de Excel.

Posteriormente iniciar o processo de análise de dados, importando os dados para um arquivo de Excel e fazendo os primeiros ajustes dos dados obtidos. Com o processo de limpeza e preparação dos dados, com o sistema ETL, é uma ferramenta de software que tem como função auxiliar na estrutura e preparação da análise de dados.

Com o agrupamento dos dados, se inicia o processo de análise dos dados com a linguagem de programação *Python*. Dentro do software Anaconda, na ambientação do software, foi executado todo o processo restante na ambientação do *Jupyter Notebook*, o qual contém uma lista ordenada de células que podem conter código, texto, fórmulas matemáticas, plotagens e imagens, e foi usada somente a linguagem de programação *Python*, no qual incorporou a biblioteca do software Pandas para manipular os dados e analisar os dados, a modelagem matemática para aplicar a estatística e os cálculos, usaremos biblioteca *NumPy*.

Com as bibliotecas usadas, foi trabalhado todas as colunas de dados e as organizei para obter a média dos dados requeridos. Para comparação, foi retirada os dados dos últimos 5 anos de todo o atlas meteorológico de Guarulhos, a fim de exibir os dados históricos obtidos de média de Incidência Solar e Velocidade Média do Vento. E posteriormente com os códigos de *Machine Learning*, realizar a mesma previsão para os 5 primeiros meses de 2020, para a mesma região.

Foi determinado usar no projeto a insolação solar em horas, devido ao objetivo do projeto de fornecer os dados para o setor de produção de energia limpa da região, para posteriormente cada produtor usar a sua razão física e matemática, para conversão dos valores, para compreensão dos resultados podemos usar a fórmula de conversão de insolação solar, para se obter a irradiação solar do local em MJ/m<sup>2</sup>.

Dados em pesquisa a fórmula de conversão:

$$Qg/Qo = a + b n/N \quad (1)$$

onde,

Qo = irradiância solar extraterrestre diária = f (latitude e declinação solar)

Qg = irradiância solar global diária = f (Qo, absorção, difusão, insolação)

n = insolação ou número efetivo de horas de brilho solar f (N e nebulosidade)

N = fotoperíodo = f (latitude e declinação solar)

Da mesma forma a conversão do vento em potencial energético se dá pela potência mecânica disponível no vento irá variar com o cubo da velocidade e pode ser dada pela seguinte equação:

$$Pv = \frac{1}{2} \cdot \rho \cdot A \cdot v1 \quad (2)$$

onde,

Pv = potência média do vento em Watts (W)

ρ = densidade do ar seco = 1,225 kg/m<sup>3</sup>

A = área de varredura do rotor (m<sup>2</sup>)

v1 = velocidade média do vento (m/s)

Como contexto de ordenar o *Dataframe*, que é minimizar os dados e organiza-los, a ferramenta Pandas executa esta atividade para os dados diários a fim de obter a média mensal dos dados requeridos, e com a biblioteca *Numpy*, foi possível elaborar matrizes para obter os cálculos de média com o algoritmo, assim simplificando sua visualização, como mostra a Figura 4:

Figura 4 - Organização do algoritmo

```
In [16]: import pandas as pd
import numpy as np

In [19]: dados = pd.read_csv('C:\PROJETOS\dadosmensais.csv')

In [22]: dados.head()
```

Out[22]:

	Estacao	Data	VelocidadeVentoMaximaMedia	InsolacaoTotal
0	83781	31/01/2015	5.8	196.8
1	83781	28/02/2015	5.6	156.7
2	83781	31/03/2015	6.3	139.3
3	83781	30/04/2015	5.8	169.2
4	83781	31/05/2015	5.6	116.1
5	83781	30/06/2015	5.8	162.1
6	83781	31/07/2015	5.5	134.4
7	83781	31/08/2015	8.3	193.9
8	83781	30/09/2015	6.1	145.2
9	83781	31/10/2015	6.6	125.1
10	83781	30/11/2015	8.3	93.9
11	83781	31/12/2015	6.6	126.9
12	83781	31/01/2016	5.8	139.9

Fonte: Autor.

Com o avanço da programação e dos códigos desenvolvidos, foi obtido por meio das bibliotecas do Pandas. Foi obtido todas as médias históricas propostas para apresentação gráfica dos resultados e assim iniciado os códigos com *Machine Learning*, em que serão utilizados algoritmos de classificação com funções de regressão estatística no software, o qual permite classificar todos os dados obtidos historicamente e irá aprender e avaliar as situações e resultados obtidos em cada dado.

Posteriormente será cruzado com os dados de previsões climáticas futuros, assim ele irá executar os planos de treinos e planos de testes, nos dois eixos X e Y, para posteriormente entregar os resultados e indicando uma acurácia de 99% de chances de acertos. Com as médias propostas para plotagem dos resultados em gráficos, junto da previsão climática para o período proposto.

Assim, podendo ir para a parte final do projeto, usando a biblioteca *Matplotlib*, que responsável pela plotagem gráfica para a linguagem de programação.

## II. RESULTADOS

Com todas as fontes consultadas, e as técnicas, pode ser plotado e apresentado os gráficos dos dados históricos e das previsões climáticas conforme as tabelas abaixo. Na Tabela abaixo traz as médias históricas da velocidade do vento dos últimos 5 anos, para termos base como comparação, na produção eólica.

Tabela 1 – Velocidade Média de Vento nos últimos 5 anos

MÊS	ANO				
	2015	2016	2017	2018	2019
Janeiro	5,8	5,8	4,6	5	5,8
Fevereiro	5,6	5	5,6	5	5
Março	6,3	5,8	5	5,5	5
Abril	5,8	5	5,8	5	4,8
Mai	5,6	5	5	6,6	5
Junho	5,8	5	5,8	4,6	
Julho	5,5	6,6	6,1	5,5	
Agosto	8,3	5,3	5	5,6	
Setembro	6,1	7,1	5	5,8	
Outubro	6,6	5	5,5	9,3	
Novembro	8,3	5,3	5,3	6,3	
Dezembro	6,6	10	6,3	8,3	
Velocidade Média de Vento (m/s)					

Fonte: Autor.

Em comparação com os dados históricos eólicos, pode-se observar um aumento da velocidade máxima média no segundo semestre de todos os anos, considerando um possível aumento de produção nesse período.

Já com os resultados obtidos, pode ser plotado a previsão climática para o primeiro período de 2020, que de acordo com as visualizações históricas e aprendizagem de previsão, pode se acompanhar a tendência de uma menor velocidade no primeiro semestre do ano, conforme mostra o Gráfico 1.

Gráfico 1 - Previsão de vento para 2020.



Fonte: Autor.

Também foi aplicado a mesma relação das médias dos últimos 5 anos, da insolação solar, para termos base de comparação, para produção fotovoltaica, observamos os resultados na Tabela 2.

Tabela 2 – Insolação total em horas nos últimos 5 anos

MÊS	ANO				
	2015	2016	2017	2018	2019
Janeiro	196	139	118	129	256
Fevereiro	156	151	127	147	169
Março	139	179	169	171	188
Abril	169	234	131	190	196
Mai	116	130	136	100	128
Junho	162	163	160	114	
Julho	134	208	200	181	
Agosto	193	164	152	141	
Setembro	145	164	263	133	
Outubro	125	175	157	139	
Novembro	94	150	171	128	
Dezembro	126	206	143	185	
<b>Insolação Total em Horas</b>					

Fonte: Autor.

Também pode-se observar como esperado uma maior insolação no primeiro trimestre dos anos passados, devido a estação do ano.

Com as técnicas também aplicadas de previsão para 2020, o sistema apontou uma porcentagem maior para erro, referente ao mês de maio de 2018, aonde se obteve uma insolação atípica ao período, levando sistema a apontar uma margem de alteração, conforme o Gráfico 2.

Gráfico 2 – Previsão Insolação 2020.



Fonte: Autor.

### III. CONCLUSÃO

O tempo tem um futuro promissor, dado que o termo *Big data*, surgiu em 2010, também com as técnicas de *Data Science*, vindo posteriormente. O projeto aplicado pode-se abranger a todos os tipos de negócios e análises. Ao tema desenvolvido, o projeto teve os resultados que foram propostos alcançados, podendo ser ampliado, aos nichos pertinentes da produção, tais como detecção de falhas técnicas, diminuição de riscos de manutenção, previsão de receita líquida e bruta, com a energia produzida. Tenho certeza de que na próxima década, muito irá se ouvir sobre as novas profissões relacionadas *Big Data*, e sobre o avanço do mercado de energia Limpa.

### IV. REFERÊNCIAS

- [1] Data Science Central. “Programando em Data Science.” Disponível em: [www.datasciencecentral.com](http://www.datasciencecentral.com)
- [2] Foster, P., Fawcett, T. “Data Science para Negócios” Ed, Alta Books. 2016
- [3] Géron, A. “Aprendizado de Máquina com Scikit-Learn e Tensorflow” Ed. Alta Books, 2019.
- [4] IBM, International Business Machines Corporation “IBM Drives the Future of Renewable Energy with New Wind and Solar Forecasting System.” 12-Ago-2013.
- [5] Instituto Nacional de Meteorologia. “Dados em Python.” Disponível em: [http://www.inmet.gov.br/portal/index.php?r=home/page&page=rede\\_estacoes\\_conv\\_graf](http://www.inmet.gov.br/portal/index.php?r=home/page&page=rede_estacoes_conv_graf).
- [6] Jornal da Bioenergia. “Gráfico Energia” Disponível em: <http://www.canalbioenergia.com.br/geracao-eolica-cresce-155-no-primeiro-semester/>
- [7] KNIME. “Plataform Analytics - KNIME.” Disponível em: <https://www.knime.org/knime>.
- [8] Mello, K. “Gráfico Vento” Revista Forbes. Disponível em: <https://forbes.com.br/negocios/2019/07/quais-sao-os-pros-e-contras-do-modelo-energetico-do-brasil/>
- [9] Mouzakis, F. “Energy Big Data Europe” 3rd Workshop for Energy during WindEurope Conference. Disponível em: <http://www.bigdataeurope.eu/energy/>.
- [10] R Project. “The R Project for Statistical Computing.” Disponível em: <https://www.rproject.org/>.
- [11] Rapid Miner. “Open Source Predictive Analytics Platform”. Disponível em: <https://rapidminer.com/>.
- [12] Weather Underground. “Python.” Disponível em: [www.wunderground.com/](http://www.wunderground.com/)
- [13] Weka. “Data Mining with Open Source Machine Learning Software in Java.” Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/index.html>.