

ESTUDO DE CONDIÇÕES CLIMÁTICAS UTILIZANDO MINERAÇÃO DE DADOS: ESTUDO DE CASO EM UMA PLANTA GERADORA DE ENERGIA EÓLICA

Elias Rocha Gonçalves Júnior (Professor Instrutor I na Universidade Candido Mendes - Campos) E-mail: eliasrgjunior1@gmail.com

Virgínia Siqueira Gonçalves (Professora Instrutora I na Universidade Candido Mendes - Campos)

Geórgia Regina Rodrigues Gomes (Professora Adjunto da Universidade Federal Fluminense)

Claudio Luiz Melo de Souza (Professor Adjunto na Universidade Candido Mendes - Campos)

Resumo: Devido às condições climáticas atuais e preocupações quanto a geração de energia por fontes renováveis, a energia eólica torna-se uma alternativa interessante do ponto de vista ambiental. Com isso, este artigo tem como objetivo analisar um conjunto de dados climatológicos de uma estação meteorológica automática, situada no município de Natal/RN, com o intuito de verificar parâmetros como velocidade média dos ventos e temperatura ambiente. Para tal, utilizaram-se técnicas de mineração de dados (KDD) e a ferramenta WEKA 3.8.1 para determinar o período mais adequado do dia para utilização da planta eólica em questão, garantindo melhor eficiência, aumentando a vida útil e reduzindo intervenções de manutenção. A partir dos resultados, pode-se afirmar que o período da Tarde é mais adequado a utilização dos equipamentos de geração de energia, além de demonstrar que temperaturas altas garantem que a velocidade média dos ventos seja adequada, assegurando eficiência na geração de energia elétrica.

Palavras-chave: energia eólica, mineração de dados, WEKA.

STUDY OF CLIMATE CONDITIONS UTILIZING DATA MINING: A CASE STUDY IN A WIND ENERGY GENERATION PLANT

Abstract: Due to the current climatic conditions and concerns about the generation of energy from renewable sources, wind energy becomes an interesting alternative from environmental point of view. The aim of this article is to analyze a set of climatological data from an automatic meteorological station, located in the city of Natal / RN, in order to verify parameters such as average wind speed and ambient temperature. For this, data mining techniques (KDD) and WEKA tool 3.8.1 were used to determine the most appropriate period of time for the use of the wind power plant in question, ensuring better efficiency, increasing the service life and reducing maintenance interventions. From the results, it can be stated that the afternoon period is more adequate to use the energy generation equipment, in addition to demonstrating that high temperatures ensure that average wind speed is adequate, ensuring efficiency in generation of electric energy.

Keywords: wind power, data mining, WEKA.

1. INTRODUÇÃO

A humanidade enfrenta o desafio de suprir a demanda de energia evitando agressões ao meio ambiente. Segundo Pereira et al. (2012), o mundo está consumindo principalmente energia proveniente de combustível fóssil e, tratando-se de consumo final de energia, esta corresponde a 79% da matriz energética mundial. Nesse contexto, os processos convencionais de geração de energia elétrica estão entre os principais responsáveis pelas emissões de Gases de Efeito Estufa (PAZHERI et al., 2014).

As energias renováveis estão se tornando em algumas nações a principal estratégia no sentido de atenuar mudanças climáticas, em decorrência de poluentes atmosféricos oriundos de atividades industriais (AKELLA et al., 2009). Apesar de demonstrar inúmeros avanços,

provavelmente ocorrerão modificações nas bases climáticas, ecológicas e sociais com efeitos catastróficos para aos seres vivos (ALLEN; VARGA, 2013).

Para mensurar a dimensão da importância desses incentivos para a energia eólica no caso brasileiro, Dalmaz et al. (2008) aponta o Programa de Incentivo às Fontes Alternativas de Energia Elétrica (PROINFA), com o objetivo de incentivar a participação de fontes renováveis. Para tal, há realização de leilões e, atualmente, é financiador de cerca de 95% dos projetos (Silva et al., 2016).

Segundo o Atlas do Potencial Eólico Brasileiro (CRESESB, 2001), o Brasil tem um potencial de geração eólico estimado em 143,47 GW, considerando apenas os locais com velocidade média anual dos ventos acima de 7 m/s. Algumas regiões que se destacam são: o litoral do nordeste, principalmente do Ceará e do Rio Grande do Norte, o litoral do Rio Grande do Sul, as Serras Gaúcha e Catarinense, alguns locais do litoral Catarinense e a região entre Paraná e Santa Catarina.

Diante deste panorama, este trabalho tem por objetivo avaliar, utilizando técnicas de Mineração de Dados, a variação da velocidade dos ventos e da temperatura ambiente identificando a viabilidade de utilização da planta geradora de Natal/RN e o período do dia mais adequado para utilização dos equipamentos, com base em aspectos técnicos.

2. REFERENCIAL TEÓRICO

2.1. VIABILIDADE DE GERAÇÃO ELÉTRICA

Dentre estas fontes energéticas chamadas de “limpas”, isto é, que não envolvem qualquer processo de combustão ou etapa de transformação térmica, a energia eólica é apontada como a fonte de energia renovável mais promissora. Este papel de destaque é devido ao potencial que tem demonstrado no que concerne aos custos de produção, segurança, fornecimento da energia e, principalmente, sustentabilidade ambiental (Martins, Guarniere & Pereira, 2008).

Segundo Dutra (2004), estas informações de velocidade do vento são importantes, porém, para determinação do potencial eólico para geração elétrica, é necessária a análise desses dados durante vários anos.

A fim de distinguir quais regiões possuem mais atratividade para instalação de parques eólicos, Reinhard (2009) afirma que diversos aspectos determinantes, sendo estes econômicos, técnicos, sociais e ambientais, devem ser levados em consideração nos modelos de apoio à decisão. Desse modo, a decisão não pode ser baseada unicamente em um único eixo-dimensional de avaliação, como custo ou benefício, que, em alguns casos, são conflitantes de natureza distintas e inerentes à decisão.

2.2. MINERAÇÃO DE DADOS

Dentre estas fontes energéticas chamadas de “limpas”, isto é, que não envolvem qualquer processo de combustão ou etapa de transformação térmica, a energia eólica é apontada como a fonte de energia renovável mais promissora. Este papel de destaque é devido ao potencial que tem demonstrado no que concerne aos custos de produção, segurança, fornecimento da energia e, principalmente, sustentabilidade ambiental (Martins, Guarniere & Pereira, 2008).

Segundo Dutra (2004), estas informações de velocidade do vento são importantes, porém, para determinação do potencial eólico para geração elétrica, é necessária a análise desses dados durante vários anos.

As técnicas relacionadas à extração de conhecimento e padrões tendo como base um banco de dados é uma área da inteligência computacional denominada como Knowledge Discovery in Databases (KDD).

Fayyad (1996) expõe a Mineração de Dados como um processo para identificação, nos dados utilizados como base, padrões válidos, úteis e compreensíveis. O processo de KDD é composto por seleção, pré-processamento e limpeza, transformação, mineração de dados e interpretação/avaliação. O processo compreende todo o ciclo que o dado percorre até tornar-se conhecimento ou informação (Figura 1).

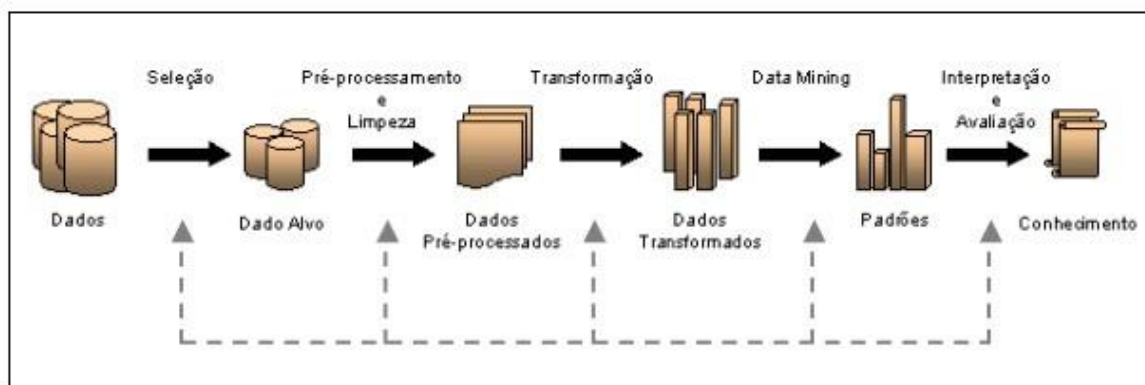


Figura 1 – O ciclo do processo de KDD.

Ao citar uma tabela, formate-a no mesmo padrão das figuras, porém, situe a legenda acima da tabela. Um exemplo pode ser observado na Tabela 1.

Fayyad et al. (1996) relata as etapas deste processo, que inclui:

- Limpeza dos dados: etapa na qual são eliminados ruídos e dados inconsistentes;
- Integração dos dados: etapa na qual diferentes fontes de dados podem ser combinadas produzindo um único repositório de dados;
- Seleção: etapa na qual são selecionados os atributos que interessam ao usuário. Por exemplo, o usuário pode decidir que informações como endereço e telefone não são de relevantes para decidir se um cliente é um bom comprador ou não;
- Transformação dos dados: etapa onde os dados são transformados num formato apropriado para aplicação de algoritmos de mineração (por exemplo, através de operações de agregação);
- Mineração: etapa essencial do processo consistindo na aplicação de técnicas inteligentes a fim de se extrair os padrões de interesse;
- Avaliação ou Pós-processamento: etapa na qual são identificados os padrões interessantes de acordo com algum critério do usuário;
- Visualização dos Resultados: etapa na qual são utilizadas técnicas de representação de conhecimento a fim de apresentar ao usuário o conhecimento minerado.

Para Salvador et al. (2009), a clusterização, uma das técnicas da mineração de dados, busca agrupar os dados de tal maneira que seja capaz de potencializar a similaridade dos objetos de um mesmo grupo e/ou diferença entre grupos distintos.

Já para Kant e Ansari (2016), clusterização é uma técnica de aprendizado não supervisionada para encontrar grupos homogêneos de um conjunto de objetos de dados ou pontos de dados. O objetivo do agrupamento é dividir objetos de dados em grupos de forma que os objetos pertencentes ao mesmo cluster sejam semelhantes entre si e se diferenciam dos outros grupos (AGGARWAL; REDDY, 2014).

Segundo Fayyad et al. (1996), árvore de decisão pode ser definida modelo preditivo que pode ser visualizado na forma de uma árvore, daí seu nome. Cada ramo da árvore é uma questão de classificação e cada folha é uma partição do conjunto de dados com sua classificação.

A tarefa de associação tem como objetivo identificar associações entre registros de dados que, de alguma maneira, estão ou devem estar relacionados. Sua premissa básica é encontrar elementos que implicam na presença de outros em uma mesma transação (SCHUNEIDER, 2002). Alguns algoritmos que utilizam os conceitos desta tarefa são as regras de associação e os padrões sequenciais.

3. METODOLOGIA

O presente artigo trata da aplicação de KDD sobre um conjunto de dados climatológicos que são essenciais para determinação do período do dia mais adequado a utilização da planta eólica de geração de energia. Para tal, os dados levantados foram velocidade do vento a uma altura de 10 metros e humidade do ar, ambos medidos a cada minuto do dia, ao longo de um determinado período. Deve-se salientar que este artigo não utiliza fatores econômicos e tecnológicos na avaliação do sistema renovável para geração de energia elétrica.

A ferramenta de data mining utilizada foi o WEKA versão 3.8.1 e foi desenvolvida pela Universidade de Waikato na Nova Zelândia. Ela pode ser definida como uma coleção de algoritmos do tipo machine learning para executar tarefas de data mining.

O WEKA vem sendo cada vez mais aplicado e algumas características interessantes ajudam a explicar seu sucesso (MARKOV e RUSSELL, 2006):

- Contém diversos algoritmos para data mining, web mining e machine learning;
- Tem código aberto e está disponível na Web gratuitamente;
- É relativamente fácil de usar, inclusive por pessoas que não sejam especialistas;
- Proporciona recursos flexíveis para experimentos;
- É mantido atualizado, pois novos algoritmos são adicionados logo que surgem na literatura.

O processo de mineração de dados compreende os seguintes (MARKOV e RUSSELL, 2006):

- Levantar a fontes de dados (bancos de dados, relatórios, etc);
- Efetuar uma limpeza dos dados para “carregar” para o WEKA;
- “Carregar” para o WEKA o arquivo pós-limpeza dos dados;

- Buscar padrões relevantes.

A descoberta do conhecimento sobre as condições climatológicas foi realizada com o levantamento de 23.041 dados gerados nos primeiros quinze dias do mês de Março (período entre 01/03/2016 a 15/03/2016) oriundos da estação meteorológica de superfície automática do Instituto Nacional de Meteorologia (INMET). A estação está situada em uma área costeira na cidade de Natal, capital do estado do Rio de Grande do Norte.

Para identificar padrões e propriedades comuns, foram utilizados três variáveis, sendo período do dia (Quadro 1), velocidade média dos ventos (Quadro 2), e temperatura ambiente (Quadro 3). Foram utilizadas as técnicas de clusterização, associação e árvore de decisão, para observar a relação dos resultados encontrados.

Quadro 1 – Critérios de classificação do período do dia.

Período do dia	Entre 00h00min e 05h59min	Entre 06h00min e 11h59min	Entre 12h00min e 17h59min
Classificação	Madrugada	Manhã	Tarde

Quadro 2 – Critérios de classificação da velocidade média dos ventos.

Velocidade média dos ventos	Abaixo de 4 m/s	Acima de 4 m/s
Classificação	Inadequada	Adequada

Quadro 3 – Critérios de classificação da temperatura ambiente.

Temperatura ambiente	Valores entre 18,0°C e 23,9°C	Valores entre 24,0°C e 28,9°C	Valores acima de 29,0°C
Classificação	Amena	Média	Alta

4. RESULTADOS E DISCUSSÃO

O estudo teve como objeto de análise dois atributos: velocidade média dos ventos e a radiação temperatura ambiente ao longo de um período de quinze dias, medindo-se as variáveis a cada minuto do dia

4.1. CRIAÇÃO DA ÁRVORE DE DECISÃO

Na criação de árvore de decisão, utilizou-se o algoritmo de J48, que demonstrou que o período da tarde é o mais adequado, pois a velocidade média dos ventos é sempre classificada como “Adequada”, evitando usos desnecessários e aumentando os períodos entre as manutenções. Além disso, foi possível identificar que no período da manhã, a velocidade dos ventos é adequada quando a temperatura ambiente for alta e também que, no período da noite, esta só é considerada adequada com temperaturas altas e médias. Já para o período da madrugada, apenas com temperaturas amenas, que recomenda-se a utilização da planta geradora. A referida árvore possui confiança de aproximadamente 72% (Figura 2).

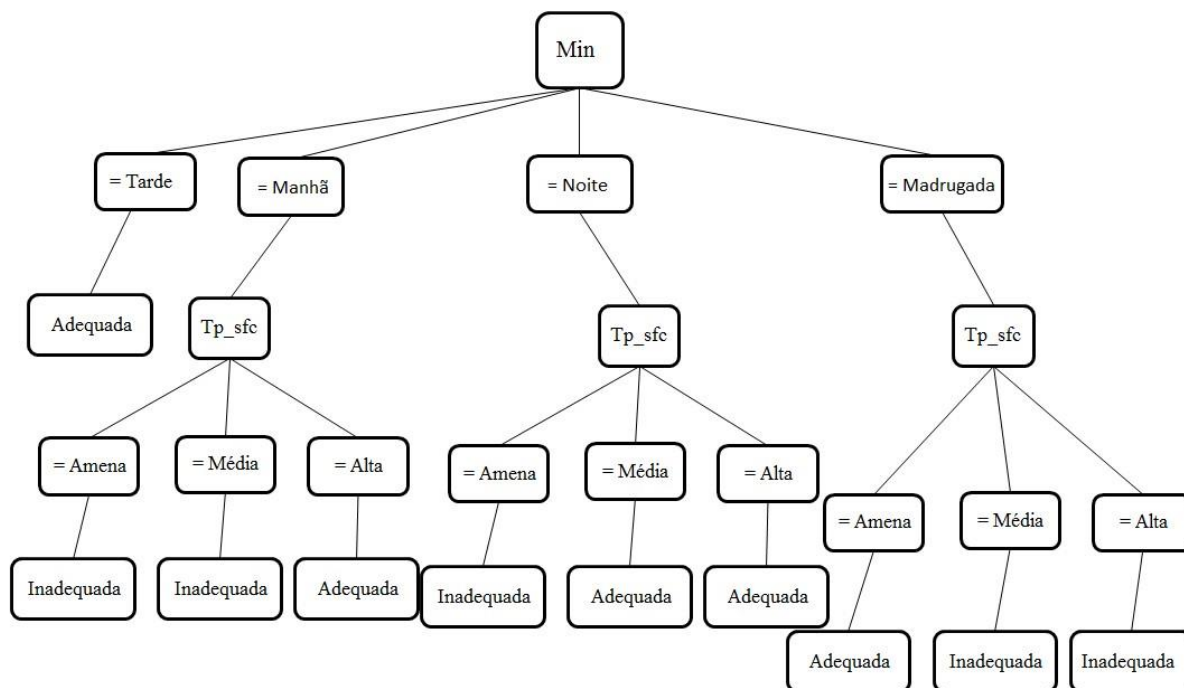


Figura 2 – Árvore de decisão gerado pelo algoritmo J48 - WEKA

4.2. ETAPA DE CLUSTERIZAÇÃO

Na etapa de clusterização, utilizou-se o algoritmo de SimpleKMeans com a formação de cinco clusters (Figura 3). O SimpleKMeans é um algoritmo de agrupamento particional simples desenvolvido por MacQueen em 1967 (MACQUEEN, 1967), o qual é baseado em um processo iterativo de duas fases no qual todo o conjunto de dados é dividido em K clusters desconexos entre si (KUSHWAHA et al., 2017).

```
Final cluster centroids:
Attribute      Full Data      Cluster#
                (23040.0)      0          1          2          3          4
-----
min            Tarde Madrugada Tarde Manhã Noite Madrugada
tp_sfc         Média Média Alta Média Média Média
ws_10m        Adequada Adequada Adequada Inadequada Adequada Inadequada

Time taken to build model (full training data) : 0.09 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      4904 ( 21%)
1      5897 ( 26%)
2      5237 ( 23%)
3      3797 ( 16%)
4      3205 ( 14%)
```

Figura 3 – Clusters gerado pelo algoritmo SimpleKMeans - WEKA

O cluster #1 identifica que se estiver no período da tarde e com temperatura alta, está será adequada a utilização da planta geradora, com as melhores condições possíveis (Figura 3). Este apresenta 26% dos dados utilizados, o que, para este caso, é satisfatório, além de complementar a informação obtida na árvore de decisão, a qual indica que no período da tarde, a velocidade média dos ventos é sempre adequada.

O cluster #2 corrobora o fato apresentado na árvore de decisão que indica o período da manhã, apresentando temperatura amena, como inadequado a geração de energia, bem como o cluster #4, que refere-se ao período da madrugada, com temperatura média, com velocidade média dos ventos inadequada.

No caso do cluster #3, há a indicação de que quando a temperatura ambiente for média e no período da noite, indica-se a utilização dos equipamentos para gerar energia, pois as condições são ideais para tal, assim como é explanado na árvore de decisão, apresentada anteriormente.

4.3. OBTENÇÃO DAS REGRAS DE ASSOCIAÇÃO

Para obtenção de regras de associação, foi utilizado o aplicativo Apriori, que gerou um total de dez regras, com confiança variando de 90% a 81%, ou seja, deveras relevante. Dentre as regras obtidas, destacam-se as seguintes:

1. min=Tarde tp_sfc=Alta 4692 ==> ws_10m=Adequada 4222 <conf:(0.9)>
lift:(1.47) lev:(0.06) [1353] conv:(3.87)
3. min=Tarde 5760 ==> ws_10m=Adequada 5113 <conf:(0.89)> lift:(1.45)
lev:(0.07) [1591] conv:(3.45)
4. tp_sfc=Alta 5764 ==> ws_10m=Adequada 5071 <conf:(0.88)> lift:(1.44)
lev:(0.07) [1547] conv:(3.23)
8. min=Tarde ws_10m=Adequada 5113 ==> tp_sfc=Alta 4222 <conf:(0.83)>
lift:(3.3) lev:(0.13) [2942] conv:(4.3)

Há a indicação do período da tarde como aquele que apresenta condições de velocidade média dos ventos adequada, além de relacioná-lo a temperaturas altas, com confiança elevada e significativa. Também foi indicado que, sempre que a temperatura for classificada como “Alta”, a velocidade dos ventos será propício a geração de eletricidade.

Estas regras corroboram as informações obtidas a partir da utilização das outras duas técnicas de mineração de dados utilizados na realização deste trabalho.

5. CONCLUSÃO

A aplicação do KDD utilizando o software WEKA 3.8.1 sobre a base de dados meteorológicos mostrou-se uma ferramenta útil para determinar o período mais adequado do dia para utilização da planta eólica em questão, garantindo melhor eficiência na geração de energia e reduzindo as manutenções.

Segundo as análises realizadas, pode-se afirmar que o período da Tarde é mais adequado a utilização dos equipamentos de geração eólica de energia, além de demonstrar que temperaturas altas garantem que a velocidade média dos ventos seja adequada, ou seja, passível de geração de energia elétrica.

Como proposta para estudos futuros, sugere-se aumentar o número de dados meteorológicos para que os especialistas tenham um conhecimento mais real das condições climáticas da região nos outros meses do ano, para determinar se há sazonalidade dos ventos ou se há um padrão médio anual.

AGRADECIMENTOS

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo.

REFERÊNCIAS

AGGARWAL, C.C.; REDDY, C.K. *Data clustering: Algorithms and Applications*. Boca Raton: CRC Press, 2014.

AKELLA, A.K.; SAINI, R.P.; SHARMA, M.P. *Social, economical and environmental impacts of renewable energy systems*. *Renewable Energy*, v. 34, n. 2, p. 390-396, 2009.

ALLEN, P.; VARGA, L. *Exploring possible energy futures for the UK: evolving power generation*. *Emergence: Complexity & Organization*, v. 15, n. 2, p. 24-47, 2013.

CRESESB. *Atlas do Potencial Eólico brasileiro de 2001*. Disponível em: <http://www.cresesb.cepel.br>. Acessado em: 03 de julho de 2017.

DALMAZ, A.; PASSOS, J.C.; COLLE, S. *Energia eólica para geração de eletricidade e a importância da previsão*. *Revista ABCM –Engenharia*, v. 13, n. 1, p. 27-32, 2008.

DUTRA, R. M. *Energia Eólica*. In: *Alternativas Energéticas Sustentáveis no Brasil*. Ed. Relume Dumará. Rio de Janeiro, 2004.

EMPRESA DE PESQUISA ENERGÉTICA – EPE. *Anuário estatístico de energia elétrica 2012*. Rio de Janeiro: EPE, 2012. Disponível em: http://www.epe.gov.br/AnuarioEstatisticodeEnergiaEletrica/20120914_1.pdf. Acessado em: 01 de junho de 2017.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *From Data Mining to Knowledge Discovery in Databases*. *AI magazine*, v. 17, n.3, p. 37-54, 1996.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. *Advances in Knowledge Discovery and Data mining*. Califórnia: AAAI Press, 1996.

KANT, S.; ANSARI, I.A. *An improved K means clustering with Atkinson index to classify liver patient dataset*. *International Journal of System Assurance Engineering and Management*, v. 7, n. S1, p. 222-228, 2016.

KUSHWAHA, N.; PANT, M.; KANT, S.; JAIN, V.K. *Magnetic optimization algorithm for data clustering*. *Pattern Recognition Letters*, 2017 (in press).

INMET. *Monitoramento das Estações Automáticas*. Disponível em: <http://www.inmet.gov.br/SONABRA/maps/automaticas.php>. Acessado em: 01 de junho de 2017.

MACQUEEN, J. *Some methods for classification and analysis of multivariate observations.* In: *5^o Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: Universidade da California, 1967, p. 281-297.

MARKOV, Z.; RUSSELL, I. *An Introduction to the WEKA Data Mining System.* In: *11th annual SIGCSE conference on Innovation and technology in computer science education.* P. 367 - 368. Bologna, Itália, 2006.

MARTINS, F.R.; GUARNIERI, R.A.; PEREIRA, E.B. *O Aproveitamento da Energia Eólica.* *Revista Brasileira de Ensino de Física*, v. 30, n. 1, p. 1-13, 2008.

PEREIRA, M. G.; CAMACHO, C. F.; FREITAS, M. A. V.; SILVA, N. F. *The renewable energy market in Brazil: Current status and potential.* *Renewable & Sustainable Energy Reviews*, v. 16, n. 6, p. 3786-3802, 2012.

PAZHERI, F.R.; OTHMAN, M.F.; MALIK, N.H. *A review on global renewable electricity scenarior.* *Renewable and Sustainable Energy Reviews*, v. 31, p. 835-845, 2014.

REINHARD, R.; ANTUNES, C.H.; DIAS, L.C. *Assessing the performance of biogas plants with multi-criteria and data envelopment analysis.* *European Journal of Operational Research*, v. 197, n. 3, p. 1084–1094, 2009.

SALVADOR, H.G.; CUNHA, A.M.; CORRÊA, C.S. *Vedalogic: um método de Verificação de Dados Climatológicos Apoiado em Modelos Minerados.* *Revista brasileira de Meteorologia*, São Paulo, v. 24, n. 4, 2009.

SCHUNEIDER, L.F. *Mineração de Dados - Conceitos.* Universidade Federal do Rio Grande do Sul, UFRS, 2002.

SILVA, R.C.; MARCHI NETO, I.; SEIFERT, S.S. *Electricity supply security and the future role of renewable energy sources in Brazil.* *Renewable and Sustainable Energy Reviews*, v. 59, p. 328-341, 2016.