

ARTIGO REF: 6758

## **ANÁLISE DO COMPORTAMENTO DE VARIÁVEIS CLIMÁTICAS SOB A ÓTICA DA GERAÇÃO ENERGÉTICA RENOVÁVEL**

**Cássio Rangel Paulista<sup>(\*)</sup>, Roberta Alvarenga dos Santos, Henrique Rego Monteiro da Hora,  
Milton Erthal Júnior**

Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFFluminense),  
Campos dos Goytacazes, Brasil

<sup>(\*)</sup>*Email:* cassio.paulista@iff.edu.br

### **RESUMO**

O monitoramento e/ou previsão de eventos climáticos é essencial para diversas atividades humanas. A mineração de dados se apresenta como uma das técnicas que auxilia a descoberta de conhecimento em grandes bases de dados. O objetivo deste trabalho é extrair conhecimento de dados climáticos utilizando técnicas de agrupamento e regressão, de forma a se obter curvas que melhor representem o comportamento da velocidade do vento e radiação solar, contidas no grupo trimestral que apresentar maior potencial de geração energética. A base de dados consultada foi fornecida pelo Instituto Nacional de Meteorologia (INMET). O Weka foi utilizado para a aplicação do agrupamento, e o Curve Expert para a obtenção das curvas. Percebeu-se que os meses que obtiveram a maior ocorrência de velocidades de vento acima de 4 m/s foram setembro, outubro e novembro. A curva referente ao dia de maior incidência solar apresentou coeficiente de correlação (0,96), num formato sinusoidal.

### **INTRODUÇÃO**

Num cenário em que a velocidade das mudanças no setor tecnológico mundial e a necessidade de adaptação é cada vez maior, a análise de informações em grandes bases de dados se torna uma atividade pertinente, demandando o uso de técnicas e ferramentas que tornem o processo de coleta, análise e utilização das informações mais ágil e confiável. Assim se configura a extração de conhecimento das massas de dados meteorológicos, que são coletados por estações espalhadas pelo território brasileiro.

Segundo Wagner, Zabadal e Ribeiro (2016), o monitoramento e/ou previsão de eventos climáticos é essencial para diversas atividades humanas. Determinadas áreas de conhecimento, como por exemplo, na agricultura, ou em outros ramos de atividades como na indústria ou transporte, há a necessidade de se ter informações confiáveis para seus planejamentos. Outra área, como a energética, é beneficiada com a possibilidade de se prever, ainda que com certo grau de erro, a incidência de irradiação solar ou a velocidade do vento, variáveis essas que estão intimamente ligadas com a geração fotovoltaica e eólica, respectivamente. Sempre que elas estão disponíveis, há o potencial de geração de energia elétrica, se incorporadas em redes inteligentes, também conhecidas por smart grids (JARRAH, 2016). A existência desse tipo de rede em território brasileiro foi estudada por Di Santo *et al.* (2015), onde se verificou os detalhes de sete smart grids no país.

Baba, Vaz e Costa (2014) ressaltaram que a análise de dados meteorológicos apresenta suporte para a previsão de fenômenos, para a avaliação de dados, para a tomada de decisões, e também auxilia na escolha e localização para implantação de sistemas utilizando fontes

renováveis de energia. Mas, analisar esses dados gerados por estações de monitoramento ambiental não é uma tarefa tão simples a qual conseguimos realizar sem uso de ferramentas específicas. A mineração de dados (Data Mining - DM) é uma dessas técnicas, na qual auxilia a descoberta de conhecimento em grandes bases de dados, transformando dados em conhecimento útil de forma inteligente e automática (REIS, 2015).

O objetivo deste trabalho é extrair conhecimento de dados climáticos utilizando técnicas de DM, especificamente, agrupamento e regressão, de forma a se obter curvas que melhor representem o comportamento da velocidade do vento e radiação solar contida no grupo trimestral que apresente maior potencial de geração de energia. As informações são oriundas da estação meteorológica de superfície automática do Instituto Nacional de Meteorologia (INMET), denominada Campos-A607, que está situada na latitude  $-21.714767^\circ$  e longitude  $-41.344017^\circ$ , no Parque Aldeia, na cidade de Campos dos Goytacazes/ RJ. A técnica de agrupamento é aplicada com a utilização do Weka. As regressões são realizadas com o auxílio do Curve Expert (HYAMS, 2009).

## REVISÃO DA LITERATURA

A fundamentação teórica desse trabalho foi iniciada com uma contextualização do cenário energético no Brasil entre 1990 e 2014, em comparação com cinco países da Organização de Cooperação e de Desenvolvimento Econômico (OCDE) que se destacam na produção de energia fotovoltaica, de acordo com Machado e Miranda (2015). Em seguida, foi realizada uma explanação do conceito de DM e a descoberta do conhecimento (Knowledge Discovery in Databases - KDD).

### Cenário energético brasileiro em comparação com países da OCDE

Realizando um corte temporal de 1990 a 2014, pôde-se perceber que o Brasil não se destacou positivamente quanto à intensidade de emissão de  $\text{CO}_2$  (IEA, 2016). O valor absoluto de dióxido de carbono liberado aumentou consideravelmente na faixa de tempo considerada. A Itália e Alemanha apresentaram redução da quantidade de  $\text{CO}_2$  emitido. Os outros países praticamente mantiveram a quantidade de gás emitido por queima de combustíveis fósseis. Tais valores foram representados pela extensão da área das bolhas demonstradas na Figura 1.

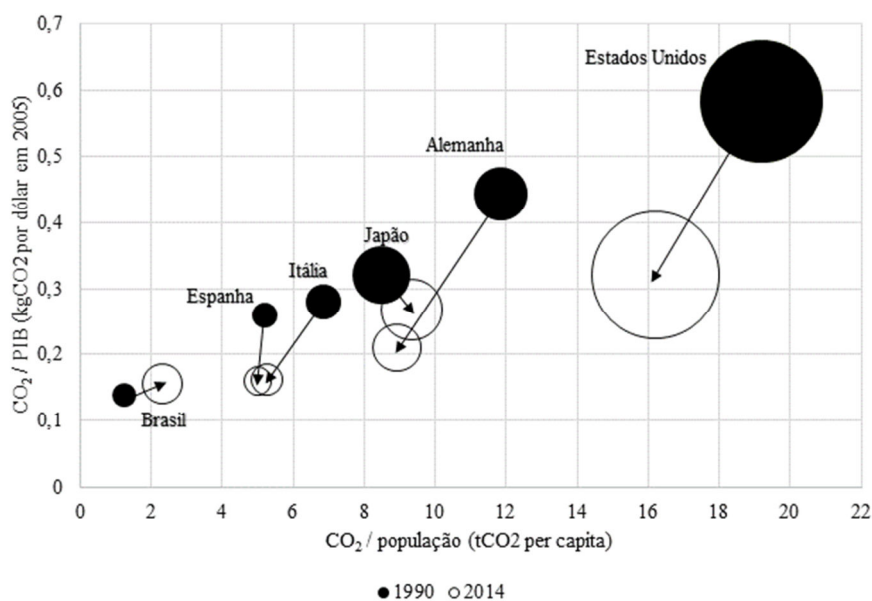


Fig. 1 - Comparação de intensidade de emissão de  $\text{CO}_2$  entre o Brasil e países membros da OCDE.

O deslocamento gráfico em direção à origem do gráfico significa redução da intensidade de emissão de CO<sub>2</sub>. A representação da questão brasileira demonstrou que o país continuou emitindo mais gases do efeito estufa em relação ao seu crescimento econômico (PIB), e ao crescimento populacional. Ainda que 70,6% da energia brasileira seja oriunda das usinas hidrelétricas (GOLDEMBERG, 2015), percebeu-se que o investimento em avanço tecnológico em relação às soluções energéticas de menor impacto ambiental é relevante, e a análise de dados históricos de monitoramento ambiental pode se apresentar como uma ferramenta valiosa para o estudo de viabilidade técnico-econômica de instalação de geradores FV, eólicos, ou até mesmo geração híbrida.

### Mineração de dados e extração do conhecimento

As informações disponíveis vêm crescendo nos últimos anos e inúmeros fatores contribuíram para esse significativo aumento. Em consequência disso, bancos de dados passaram a possuir verdadeiros tesouros de informação e, devido ao seu volume, ultrapassaram a habilidade técnica e a capacidade humana em sua captação e interpretação, tornando-se necessária a aplicação de técnicas e ferramentas automáticas que agilizassem o processo de extração de informações relevantes de grandes volumes de dados (CARDOSO; MACHADO, 2008).

Segundo Santos, Menezes e Hora (2014), o KDD surgiu para atender a este novo cenário. Fayyad, Piatetsky-Shapiro e Smyth (1996) definiram essa técnica como sendo um processo de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados. Tal processo foi mostrado na Figura 2.

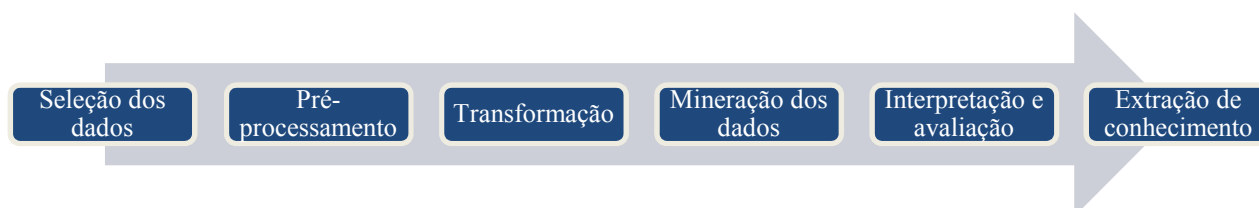


Fig. 2 - Visão geral das etapas que compõe o processo KDD.

Steiner *et al.* (2006) apontou que até 1995, muitos autores consideravam os termos KDD e DM como sinônimos, porém o primeiro refere-se a todo processo de descoberta de conhecimento útil nos dados, enquanto o segundo trata-se de uma etapa desse processo. Wagner, Zabadal e Ribeiro (2016) definiram DM como um processo de extração de informações de uma grande base de dados para tomada de decisões, que é aplicada em diversas áreas como empresas, pesquisas e indústrias que utilizam os resultados, como exemplo, para melhoria de processos ou analisar tendências.

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), existem diversos métodos de DM para encontrar respostas ou extrair conhecimento em repositórios de dados, sendo os mais importantes para o KDD: classificação, modelos de relacionamento entre variáveis, análise de agrupamento, sumarização, modelo de dependência, regras de associação e análise de séries temporais. Nesse trabalho foi utilizada a técnica de agrupamento.

### METODOLOGIA

Os procedimentos que foram executados nesse artigo corresponderam aos passos descritos na Figura 3. Os dados climáticos foram obtidos da base de dados brasileira do Instituto Nacional de Meteorologia (INMET), criado em 1909, com o objetivo de fornecer dados meteorológicos

para a sociedade do Brasil, de forma a auxiliar os processos decisórios e promover o desenvolvimento sustentável do país. O órgão faz parte do Ministério de Agricultura, Pecuária e Abastecimento, possuindo sede em Brasília e dez distritos de meteorologia distribuídos estrategicamente nas capitais (INMET, 2016).



Fig. 3 - Metodologia do trabalho.

O corte temporal considerado no presente artigo foi realizado para o ano de 2016, de janeiro a dezembro. Foram criados dez grupos trimestrais, de forma a se promover uma superposição de meses em cada grupo. A estrutura dos grupos trimestrais foi modelada conforme a Figura 4. No total, foram mineradas 8708 instâncias.

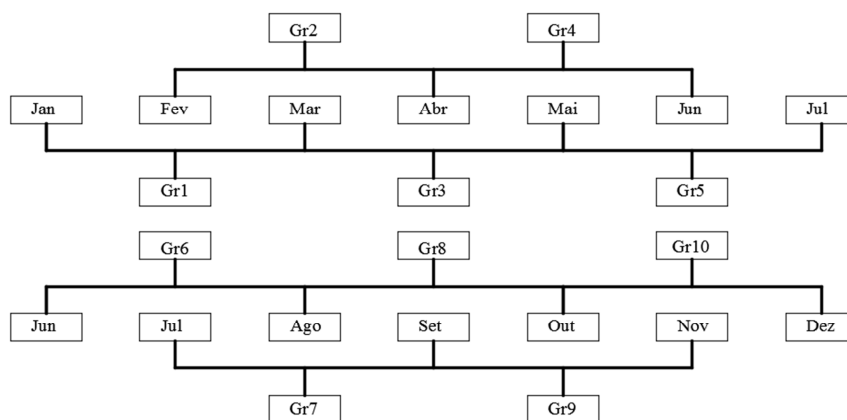


Fig. 4 - Estrutura dos grupos trimestrais analisados.

### Clusterização

O algoritmo de agrupamento utilizado foi o *Expectation Maximization* (EM). Essa técnica atribuiu uma distribuição de probabilidade para cada instância, indicando a probabilidade de ela pertencer a um núcleo, ou cluster. O código que representou a configuração completa do algoritmo foi “EM -I 1000 -N -1 -X 10 -max -1 -ll-cv 1.0 E-6 -ll-iter 1.0 E-6 -M 1.0 E-6 -K 10 -num-slots 1 -S 100”.

A Tabela 1 apresentou os resultados gerados após a aplicação do algoritmo. As iniciais “V” e “R” foram definidas para designar a velocidade do vento, em m/s, e a radiação solar, em KJ/m<sup>2</sup>, respectivamente. Os valores apresentados representaram as médias calculadas pelo método de mineração. O grupo trimestral destacado foi selecionado a partir dos parâmetros de viabilidade de geração de energia eólica, de acordo com Amarante *et al.* (2001), e que obteve o maior número de instâncias contidas nele.

### Análise de regressão

O software Curve Expert foi utilizado como ferramenta para se detectar a curva que melhor representa o comportamento dos dados de velocidade do vento e radiação solar, para o período relativo ao grupo trimestral “Gr9”. O programa foi configurado para realizar até 100 iterações, tendo como critério de parada o alcance de uma precisão de 1 micron (10-6).

Tabela 1 - Resultados da aplicação do algoritmo EM, por grupo trimestral.

j	Gr1		Gr2		Gr3		Gr4		Gr5	
	V	R	V	R	V	R	V	R	V	R
0	2,3	0,0	2,1	0,0	3,1	259,1	3,1	1245,6	2,8	368,7
1	3,6	1617,2	3,6	1655,9	3,4	2662,8	2,2	0,0	2,3	0,0
2	-	-	-	-	3,4	1377,1	-	-	3,2	1633,2
3	-	-	-	-	2,1	0,0	-	-	-	-

j	Gr6		Gr7		Gr8		Gr9		Gr10	
	V	R	V	R	V	R	V	R	V	R
0	3,1	920,3	4,6	2400,3	4,1	1143,7	4,2	1374,7	4,8	2941,5
1	3,0	162,7	2,5	0,0	2,7	0,0	2,9	0,0	2,7	0,0
2	2,3	0,0	3,0	147,7	4,9	2626,4	-	-	3,4	296,4
3	4,2	2209,7	3,7	978,5	3,2	206,6	-	-	4,2	1328,9

Para a velocidade do vento, buscou-se obter três gráficos de demonstrassem a velocidade em função do tempo ao longo dos 90 dias, com cada curva correspondendo a um mês do grupo trimestral. Como os dados são obtidos de forma horária, foi necessário calcular a mediana diária para cada mês. Os resultados com as curvas obtidas, coeficientes de correlação e equações, foram mostrados na Figura 5.

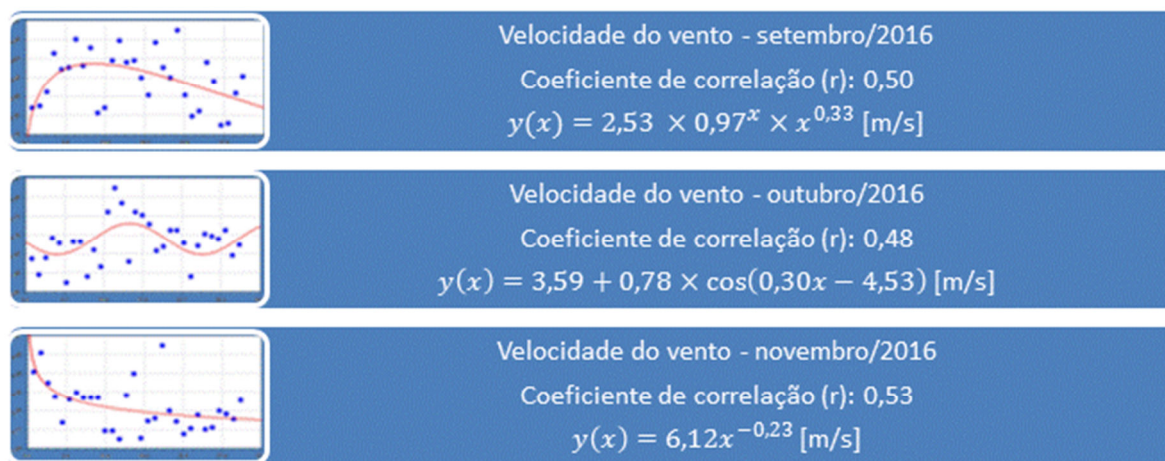


Fig. 5 - Velocidade do vento x dia do mês.

A curva relacionada à radiação solar foi obtida tendo como referência o dia contido no grupo trimestral “Gr9” que apresentou a maior radiação solar. Após análise dos dados do INMET, verificou-se que a ocorrência dessa radiação foi em 19 de novembro de 2016, às 15h, com um valor de 4020 kJ/m<sup>2</sup>. A equação que definiu o comportamento da radiação solar durante o dia especificado foi uma curva senoidal, conforme descrito pela Equação 1.

$$y(x) = 1471,36 + 2103,13 \times \cos(30,39x - 4,41) \text{ [KJ/m}^2\text{]} \quad (1)$$

O coeficiente de correlação obtido para a curva sinusoidal foi de 0,96, o que demonstrou alta correlação entre a curva e os dados obtidos na base do INMET imputados no Curve Expert. A curva obtida foi demonstrada na Figura 6.

## CONSIDERAÇÕES FINAIS

A utilização de técnicas de mineração de dados, especificamente, agrupamento e regressão, foram demonstradas e, nesse sentido, verificou-se que as mesmas são relevantes no estudo do comportamento de variáveis climáticas, possibilitando a extração de conhecimento de uma grande quantidade de dados coletados por estações automáticas. Vale notificar que as informações da base utilizada, INMET, são brutas e não sofreram nenhum tratamento prévio pelo instituto para serem divulgadas ao público.

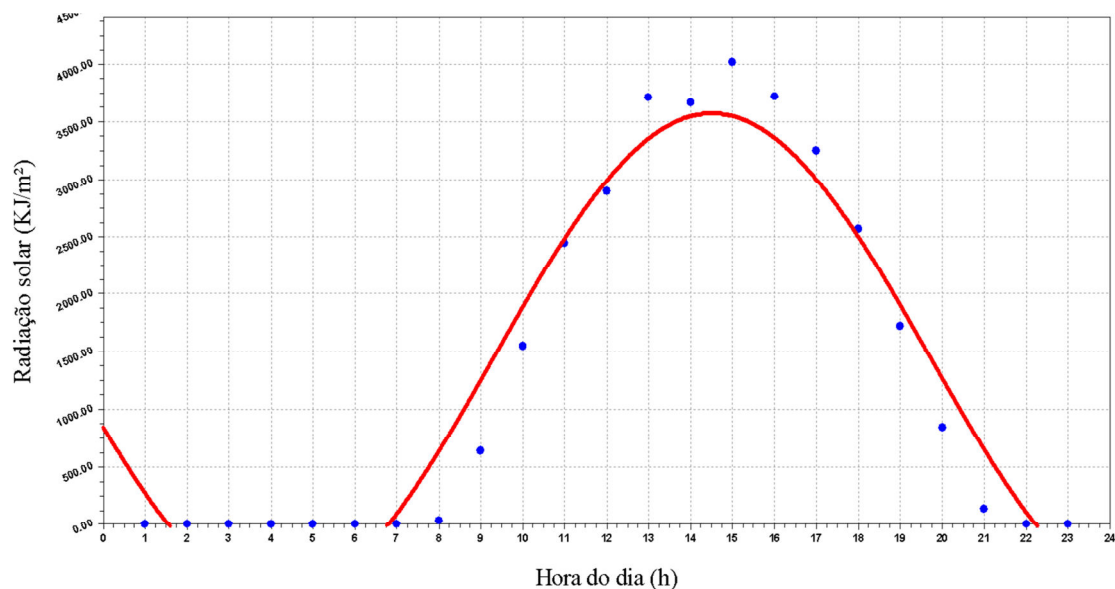


Fig. 6 - Curva de radiação solar durante o dia 19 de novembro de 2016.

A aplicação da técnica de agrupamento por grupos trimestrais demandou esforço braçal para preparação dos arquivos antes de serem carregados no *Weka*. Ainda assim, o recurso foi considerado eficaz, com a utilização do algoritmo EM, para se verificar quais grupos continham as instâncias que detinham as maiores velocidades de vento, com suas respectivas radiações solares. Desse modo, observou-se que, em 2016, os meses que obtiveram a maior ocorrência de velocidades de vento acima de 4 m/s foram setembro, outubro e novembro. A modelagem das curvas de velocidade do vento para esses meses, via regressão, revelou coeficientes de correlação abaixo de 0,53, coeficiente para novembro de 2016. Concluiu-se que os dados dos ventos levantados pela estação automática não seguem um padrão facilmente modelado matematicamente.

A curva referente ao dia de maior incidência solar para novembro 2016 foi a que apresentou o maior coeficiente de correlação (0,96) através de um gráfico sinusoidal. Nesse aspecto, constatou-se que a radiação solar possuiu um comportamento melhor definido em relação à velocidade do vento. Os valores de radiação foram considerados coerentes, sendo uma das evidências o pico que ocorreu às 15h, fato comumente comprovado empiricamente.

Para trabalhos futuros, sugere-se a utilização de outros bancos de dados, como o SWERA, para poder se comparar as curvas obtidas, e se verificar se haveria maior aderência dos gráficos às informações disponíveis na base. Sugere-se também a definição de uma equação geral para a radiação solar, na faixa de 1 ano, para se poder utilizá-la em modelos de simulação a eventos discretos que tratem de geração de energia fotovoltaica.

## REFERÊNCIAS

- [1]-AMARANTE, O. A. C. DO et al. Atlas do potencial eólico brasileiro Brasília, 2001.
- [2]-BABA, R. K.; VAZ, M. S. M. G.; COSTA, J. D. Correção de dados agrometeorológicos utilizando métodos estatísticos. Revista Brasileira de Meteorologia, v. 29, n. 4, 9 jun. 2014.
- [3]-CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. Revista de administração pública, v. 42, n. 3, p. 495-528, jun. 2008.
- [4]-DI SANTO, K. G. et al. A review on smart grids and experiences in Brazil. Renewable and Sustainable Energy Reviews, v. 52, p. 1072-1082, dez. 2015.
- [5]-FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. AI magazine, v. 17, n. 3, p. 37, 1996.
- [6]-GOLDEMBERG, J. O estado atual do setor elétrico brasileiro. Revista USP, n. 104, p. 37-44, 2015.
- [7]-HYAMS, D. G. CurveExpert Copyright © 1995-2007. Hixson, Tennessee: Daniel Hyams, 2009.
- [8]-IEA. International Energy Agency. Disponível em: <<http://www.iea.org/>>. Acesso em: 19 jun. 2016.
- [9]-INMET. Instituto Nacional de Meteorologia. Disponível em: <[http://www.inmet.gov.br/portal/index.php?r=home/page&page=sobre\\_inmet](http://www.inmet.gov.br/portal/index.php?r=home/page&page=sobre_inmet)>. Acesso em: 22 dez. 2016.
- [10]-JARRAH, M. Modeling and Simulation of Renewable Energy Sources in Smart Grid Using DEVS Formalism. Procedia Computer Science, v. 83, p. 642-647, 2016.
- [11]-MACHADO, C. T.; MIRANDA, F. S. Photovoltaic Solar Energy: A Briefly Review. Revista Virtual de Química, v. 7, n. 1, 2015.



[12]-REIS, R. X. DOS. A utilização do fenômeno Big Data na Administração Pública - a experiência do PENSA na Prefeitura Municipal do Rio de Janeiro. Trabalho de conclusão de curso—Rio de Janeiro: Escola de Contas e Gestão, 2015.

[13]-SANTOS, A. C. DE S. G. DOS; MENEZES, T. DE P.; DA HORA, H. R. M. Análise do perfil de aluno e egresso de cursos técnicos por meio de data mining: estudo de caso no Instituto Federal Fluminense. # Tear: Revista de Educação, Ciência e Tecnologia, v. 3, n. 1, 2014.

[14]-STEINER, M. T. A. et al. Study of a medical problem using KDD, with emphasis on exploratory data analysis. Gestão & Produção, v. 13, n. 2, p. 325-337, maio 2006.

[15]-WAGNER, M. M.; RIBEIRO, V.; ZABADAL, J. Mineração de Dados Meteorológicos Empregando Dados de Temperatura: o caso de uma cidade gaúcha. Interfaces Científicas - Exatas e Tecnológicas, v. 2, n. 2, p. 57-70, 11 out. 2016.