

Redes Inteligentes: Mineração de Dados como Ferramenta de Apoio à Análise de Grandes Volumes de Dados em Subestações Subterrâneas de Energia *

Jonas F. Schreiber * Ivan E. M. Kühne *
Luciano A. Destefani ** Airam T. Z. R. Sausen *
Maurício de Campos * Paulo S. Sausen *

* *Departamento de Ciências Exatas e Engenharias, Universidade Regional do Noroeste do Estado do Rio Grande do Sul, RS, (e-mails: jonasfs@gmail.com, oitentaetres@gmail.com, airam@unijui.edu.br, campos@unijui.edu.br, sausen@unijui.edu.br).*

** *Prefeitura Municipal de Ijuí, RS, (e-mail: luciano@ijui.rs.gov.br).*

Abstract: The Smart Grids can be defined as the superposition of a unified communication and control system to the existing electricity distribution system, in order to overcome the vulnerabilities caused by the fact that Electric Power Systems use old technologies despite being increasingly interconnected. Several authors point out the need and the potential of using advanced analysis techniques, covered under paradigms such as Data Mining, so that the data generated by this unified system can be effectively exploited, capable of supporting the management process of the systems energy distribution. This article presents a case study on the application of this type of technique to the data generated by the monitoring system of underground substations installed and operated by the State Electricity Distribution Company in the city of Porto Alegre. In the presented case study, a mathematical model was developed to predict the performance of the transformers of these substations using data collected by the monitoring system and the Multiple Linear Regression statistical technique. Based on the validation of this model, future works capable of improving the presented process are proposed.

Resumo: As Redes Inteligentes podem ser definidas como a superposição de um sistema unificado de comunicação e controle ao sistema existente de distribuição de energia elétrica, de forma a superar as vulnerabilidades causadas pelo fato dos Sistemas Elétricos de Potência utilizarem tecnologias antigas apesar de estarem cada vez mais interconectados. Diversos autores apontam a necessidade e o potencial do uso de técnicas avançadas de análise, abrangidas sob paradigmas como a Mineração de Dados, para que os dados gerados por esse sistema unificado sejam explorados de forma efetiva, capaz de apoiar o processo de gestão dos sistemas de distribuição de energia. Nesse artigo é apresentado um estudo de caso sobre a aplicação desse tipo de técnica aos dados gerados pelo sistema de monitoramento das subestações subterrâneas instalado e operado pela Companhia Estadual de Distribuição de Energia Elétrica na cidade de Porto Alegre. No estudo de caso apresentado, foi desenvolvido um modelo matemático capaz de prever o rendimento dos transformadores dessas subestações a partir de grandezas coletadas pelo sistema de monitoramento e da técnica estatística de Regressão Linear Múltipla. A partir da validação desse modelo, são propostos trabalhos futuros capazes de aprimorar o processo apresentado.

Keywords: Big Data; Data Mining; Knowledge Discovery in Databases; Mathematical Modeling; Multiple Linear Regression; Smart Grid.

Palavras-chaves: Big Data; Descoberta de Conhecimento em Bancos de Dados; Mineração de Dados; Modelagem Matemática; Rede Inteligente; Regressão Linear Múltipla.

1. INTRODUÇÃO

Conforme o Electric Power Research Institute (2008), o conceito de Rede Inteligente pode ser entendido como a superposição de um sistema unificado de comunicação e controle ao sistema existente de distribuição de energia elétrica. Essa superposição permite que informações sobre o sistema de distribuição possam ser disponibilizadas às entidades interessadas, como os responsáveis pela sua operação e os consumidores, dentro de um espaço de tempo adequado para que decisões corretas possam ser tomadas.

Uma das primeiras menções a esse conceito é encontrada no trabalho de Amin and Wollenberg (2005), onde são analisadas as vulnerabilidades dos Sistemas Elétricos de Potência (SEP) responsáveis pelo suprimento de energia elétrica aos consumidores. Os autores apresentam que a infraestrutura dos SEP está cada vez mais interconectada, o que a torna vulnerável a determinados tipos de ameaças que podem se propagar sob a forma de falhas de grande impacto, capazes de afetar as operações de diversos setores da economia que dependem diretamente de um suprimento energético seguro e confiável, como os transportes, as comunicações e as finanças.

Entretanto, ainda segundo Amin and Wollenberg (2005), os princípios e elementos principais da interconexão dos SEP foram estabelecidos antes da adoção massiva de computadores e de redes de comunicações. Dessa forma, os autores apontam a necessidade de que o sistema de distribuição seja convertido em uma grande plataforma de computação distribuída, integrada por componentes, como subestações e usinas, capazes de funcionar como agentes independentes aptos à comunicação e à cooperação com os demais componentes.

Para que isso possa acontecer, cada componente do sistema de distribuição deve estar conectado a uma série de sensores, de forma que possa avaliar as suas próprias condições de funcionamento através da análise dos dados colhidos por esses sensores, bem como comunicar essas condições aos componentes vizinhos (Amin and Wollenberg, 2005). A adoção desses sensores leva à geração em grande velocidade de um elevado volume de dados heterogêneos, que podem ser armazenados em bancos de dados. Por outro lado, as características de funcionamento das Redes Inteligentes podem gerar inconsistências nos bancos de dados utilizados devido a eventuais falhas de leitura e de comunicação.

As características dos dados gerados pelas Redes Inteligentes produzem a necessidade de aplicação de técnicas de Mineração de Dados (Data Mining), conforme apontado no relatório técnico do IEEE Smart Grid (2018) e nos trabalhos de autores como Stimmel (2014), Sagiroglu et al. (2016), Zhou et al. (2016) e He et al. (2017). Isso ocorre porque as técnicas tradicionais de análise não são capazes de lidar adequadamente com esses conjuntos de dados, que possuem atributos relacionados ao conceito de *Big Data*.

O termo Big Data vem sendo aplicado desde trabalhos como o de Cox and Ellsworth (1997), que o utilizaram para descrever os conjuntos de dados utilizados na Visualização Científica, cujos volumes extrapolavam a capacidade da

memória local e dos sistemas de armazenamento locais e remotos disponíveis na época. Atualmente, conforme a definição apresentada por Manyika et al. (2011), o conceito é associado a conjuntos de dados cujo tamanho impede que sejam coletados, armazenados, geridos e analisados por ferramentas tradicionais de banco de dados.

O conceito de *Big Data* está associado aos atributos de volume, velocidade e variedade, chamados de três "vês", que, conforme Zhou et al. (2016), representam ao mesmo tempo um desafio e uma oportunidade. Em trabalhos como o de Narayanan et al. (2017) e do IEEE Smart Grid (2018) são apresentados mais dois atributos associados ao conceito de *Big Data*: o valor e a veracidade.

Atualmente o processo de coleta, armazenamento e análise de quantidades massivas de dados é utilizado como método de pesquisa em diversas áreas do conhecimento, abrangendo tanto aplicações puramente acadêmicas quanto aquelas voltadas para a resolução de problemas na área da indústria e do setor de serviços. Devido a essa diversidade de aplicações, é difícil estabelecer um consenso sobre o termo que deve ser utilizado para definir esse processo.

Devido a essas divergências, nesse artigo será utilizado o termo Mineração de Dado (*Data Mining*), encontrado com frequência na literatura, de forma a englobar as diversas terminologias propostas, como Análise de Dados (*Data Analysis* ou *Data Analytics*), Aprendizado de Máquina (*Machine Learning*), *Big Data Analytics*, Ciência de Dados (*Data Science*) e Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases - KDD*).

No trabalho de Yu et al. (2015) são apresentadas algumas possibilidades de aplicação da Mineração de Dados às Redes Inteligentes. Em curto prazo, são citadas a detecção do roubo de energia, a detecção da integração de painéis solares, o desenvolvimento de uma projeção de demanda de carga mais granular, a visualização dos sistemas de distribuição de energia, a estimativa do estado da rede e o diagnóstico dos equipamentos. Já em longo prazo são citadas a modelagem do comportamento do consumidor em resposta a várias estruturas de preços e incentivos e o processo de planejamento da transformação do sistema de distribuição.

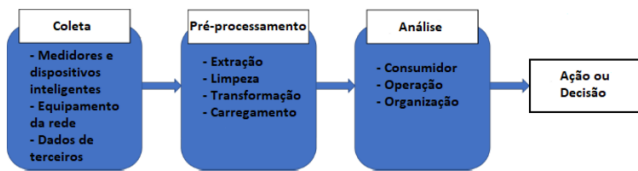
Conforme Zhou et al. (2016), a aplicação adequada da Mineração de Dados pode proporcionar diversas funcionalidades relacionadas ao apoio efetivo às decisões de todos os produtores, operadores e consumidores envolvidos com as Redes Inteligentes, bem como aos responsáveis pela sua regulamentação. De acordo com essa fonte e com a IBM Corporation (2012), essas funcionalidades incluem a otimização da geração e operação de energia elétrica em tempo real, a predição acurada da demanda de energia e resposta rápida a essa demanda, a detecção de falhas e a recuperação após a sua ocorrência, a descoberta de padrões de consumo e o suprimento de energia de forma confiável e econômica.

Na Fig. 1 é exemplificada a utilização da Mineração de Dados como ferramenta de apoio à decisão, conforme descrito por Stimmel (2014). Segundo a autora, os dados de fontes como medidores e dispositivos inteligentes são coletados, pré-processados e analisados, gerando informações que subsidiam ações e tomadas de decisões. Dessa forma, é

* Pesquisa realizada com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

resumido o processo que começa com a adoção massiva de sensores defendida por Amin and Wollenberg (2005) e leva às funcionalidades relacionadas pela IBM Corporation (2012) e por Zhou et al. (2016).

Figura 1. Exemplo de extração de valor através da Mineração de Dados (adaptado de Stimmel (2014)).



As análises realizadas nos dados provenientes das Redes Inteligentes através da Mineração de Dados podem revelar padrões implícitos, que não seriam descobertos através da aplicação de técnicas de análise tradicionais. Segundo o IEEE Smart Grid (2018), essas análises podem ser classificadas em quatro categorias: descritiva, diagnóstica, preditiva e prescritiva. A seguir são explicadas cada uma dessas categorias de análise, de acordo com essa fonte e com Dang-Ha et al. (2015).

- (1) Análise descritiva: fornece informação sobre o que já ocorreu e se constitui do primeiro passo na tentativa de identificação de dados e informações úteis para processamento adicional;
- (2) Análise diagnóstica: procura entender as causas de eventos e o comportamento do sistemas, de forma a identificar desafios e oportunidades;
- (3) Análise preditiva: utilizada para fazer previsões probabilísticas de forma a identificar tendências, com objetivo de determinar o que pode ocorrer no futuro;
- (4) Análise prescritiva: aplicada na identificação do melhor resultado possível de eventos, dados os parâmetros do sistema, e na elaboração de estratégias para a gestão de eventos similares no futuro. Utiliza ferramentas como técnicas de simulação e suporte de decisões para explorar estratégias ótimas para que se possa aproveitar uma oportunidade futura ou mitigar um risco futuro.

No trabalho de Chaganti and Chowdhury (2004) é abordada a utilização do algoritmo C4.5 para a geração de árvores e regras de decisão. A partir da aplicação do algoritmo citado ao conjunto de dados gerado, foram classificadas as falhas e foi determinado o número de instâncias que uma falha é classificada corretamente, dadas as correntes e tensões das medições de sistemas de potência. Segundo os autores, essas técnicas permitem um aumento no desempenho do sistema, o que tende a melhorar o relacionamento com os consumidores.

No trabalho de Hor and Crossley (2005) é discutida a aplicação da teoria dos conjuntos aproximados como uma forma de classificação não-supervisionada dentro de uma subestação de distribuição de energia. A classificação não-supervisionada permite o agrupamento automático dos dados, sem a necessidade de classes pré-definidas ou de conhecimento a priori. A teoria dos conjuntos aproximados também é utilizada no trabalho de Crossley and Liu (2009), que trabalham com a eliminação dos atributos que, embora

presentes no conjunto bruto de dados, não são necessários para caracterizar completamente o conhecimento.

No trabalho de Dobaev et al. (2016) é proposto o desenvolvimento de um sistema modular de apoio à decisão que pode ser utilizado pelas companhias de distribuição de energia elétrica na gestão centralizada dos seus sistemas. No trabalho de Bhuiyan et al. (2017) é abordado o processamento de grandes volumes de dados oriundos das unidades de medição fasorial (*phasor measurement unit* - PMU) para medir a correspondência entre eles durante determinados eventos e, com isso, detectar a ocorrência de anomalias no sistema de distribuição e a sua localização.

Nesse artigo é demonstrada a aplicação de uma técnica estatística de regressão linear múltipla ao banco de dados proveniente do monitoramento das subestações subterrâneas da Companhia Estadual de Distribuição de Energia Elétrica (CEEE-D) na cidade de Porto Alegre. O objetivo da aplicação dessa técnica é desenvolver um modelo matemático, denominado Modelo de Estimativa de Rendimento de Transformadores, capaz de prever o rendimento dos transformadores das subestações a partir do banco de dados disponível.

Cabe ressaltar que não é criado nenhum banco de dados novo e nenhuma informação adicional é requerida, sendo utilizado o conjunto de dados já existente, criado a partir das leituras dos sensores instalados junto às subestações. Além do desenvolvimento e disponibilização do modelo, o mesmo é validado com dados reais. Com isso, é possível obter subsídios para o desenvolvimento de trabalhos futuros envolvendo o mesmo estudo de caso ou sistemas semelhantes, de forma a explorar o potencial da aplicação da Mineração de Dados aos registros gerados pelas Redes Inteligentes.

Muitas vezes os dados das Redes Inteligentes são coletados e usados apenas para alarmar determinadas ocorrências ou eventos, sem que haja uma exploração plena do seu potencial para a descoberta de padrões e conhecimento relevantes para os responsáveis pela gestão desses sistemas de distribuição. Uma possibilidade de utilização do modelo proposto é no auxílio à manutenção preventiva e à detecção de falhas nos transformadores das subestações. Dessa forma, podem ser detectadas as subestações que estejam apresentando problemas mais graves e, por isso, devam ter uma maior prioridade na execução da manutenção preventiva, principalmente quando existem recursos limitados para a realização dessa tarefa.

2. DESENVOLVIMENTO

A partir da pesquisa apresentada na Seção 1, foi selecionado um estudo de caso para a aplicação das técnicas disponíveis. Como diferencial de trabalhos semelhantes, o estudo de caso escolhido é um problema real, referente ao sistema de monitoramento de subestações subterrâneas instalado e operado pela CEEE-D na cidade de Porto Alegre. A partir dos dados disponíveis, foi desenvolvido e validado um modelo de regressão linear múltipla aplicável aos transformadores das subestações subterrâneas monitoradas, denominado Modelo de Estimativa de Rendimento de Transformadores.

O modelo desenvolvido está relacionados às categorias de análise descritiva e diagnóstica, conforme definido por Dang-Ha et al. (2015) e o IEEE Smart Grid (2018). Isso ocorre porque o modelo fornece informações sobre o que já ocorreu no sistema, subsidiando análises adicionais, e possibilita entender o seu comportamento e a causa de eventos, identificado desafios e oportunidades. Dessa forma, as informações encontradas através da aplicação do modelo podem subsidiar ações e a tomada de decisões, conforme proposto por Stimmel (2014).

O sistema reticulado em estudo é instalado na região central de Porto Alegre, sendo alimentado com tensões primárias de 13,8 kV e tensões secundárias de 127/220V. É composto por transformadores submersíveis de 500 kVA alojados em câmaras subterrâneas sob o leito das ruas e a capacidade nominal de corrente do sistema é de 1.600 amperes. São monitoradas diversas grandezas analógicas e digitais desse sistema de forma constante, com intervalo de coleta de dez segundos.

Após realizadas as medições, os dados coletados nas subestações são transmitidos, através de uma rede de comunicação, até um servidor remoto que é encarregado de tratá-los e armazená-los em um banco de dados. Devido às características de funcionamento do sistema de monitoramento, podem haver falhas de leitura e de comunicação que impedem o registro de determinadas leituras. A estrutura do banco de dados original disponibilizado é composta por dois grandes grupos de grandezas: analógicas e digitais. Para o desenvolvimento do modelo, foram selecionadas as grandezas analógicas diretamente relacionadas ao rendimento dos transformadores, que são apresentadas na Tabela 1.

Tabela 1. Grandezas utilizadas.

Grandeza	Símbolo
Tensão do Secundário A	VAs
Tensão do Secundário B	VBs
Tensão do Secundário C	VCs
Corrente Primário A	IA
Corrente Primário B	IB
Corrente Primário C	IC
Corrente Secundário A	IAs
Corrente Secundário B	IBs
Corrente Secundário C	ICs
Temperatura Ambiente	TempAmb
Temperatura do Transformador	TempTrafo

O estudo foi realizado sobre as amostras coletadas pelo sistema de monitoramento da CEEE-D no período de janeiro a março de 2017, época em que esse sistema abrangia cinquenta e seis subestações. Uma vez que as características do sistema fazem com que o registro dos dados seja sujeito a falhas diversas, foi determinado que seriam consideradas apenas as subestações que apresentavam mais de cem mil amostras no período de janeiro a março de 2017, o que representa cerca de 15% do número teórico de amostras disponíveis. Com base nesse critério, foram escolhidas vinte e sete subestações.

Em seguida foi aplicado um filtro sobre as amostras das vinte e sete subestações selecionadas, de forma que foram removidos os *outliers* presentes nos dados. A partir da seleção das subestações e da aplicação desse filtro, evita-se que as subestações com uma quantidade elevada de falhas nos registros comprometam a acurácia do modelo proposto.

Como critério de seleção dos conjunto de treinamento e de validação, foi determinado que as subestações seriam divididas em 75% para treinamento e 25% para validação. Assim, foram selecionadas vinte subestações para o conjunto de treinamento e sete para o conjunto de validação. As subestações que apresentam o maior número de amostras válidas foram selecionadas como conjunto de treinamento, como forma de aumentar a acurácia do modelo, e são apresentadas na Tabela 2 em ordem decrescente do número de amostras válidas. As demais subestações, que foram selecionadas como conjunto de validação, são apresentadas na Tabela 3, também em ordem decrescente do número de amostras válidas.

Tabela 2. Conjunto de treinamento.

Subestação	Quantidade de Amostras
94/15D	743.719
119/5	724.087
106/9A	717.139
83/9	701.915
95/11B	669.865
31/13	664.136
121/9	654.809
87/17	650.115
86/1	649.793
87/13	625.337
97/5	615.704
71/3D	613.206
101/9	611.459
102/11B	602.757
100/13	577.483
96/13	567.972
85/15B	566.182
87/9	558.917
42/3A	551.902
34/3A	547.736

Tabela 3. Conjunto de validação.

Subestação	Quantidade de Amostras
74/11B	527.206
48/5	493.419
67/13	435.769
68/1	218.811
55/5	157.720
34/5	95.941
112/13	94.771

O modelo foi composto pelas variáveis independentes relativas às grandezas utilizadas e é apresentado na Equação 1. Os coeficientes desse modelo foram ajustados ao conjunto de treinamento com auxílio da ferramenta computacional MATLAB MathWorks (2020a) e seu pacote *Statistics and Machine Learning Toolbox* MathWorks (2020b). O resultado desse ajuste é apresentado na Tabela 4.

$$\eta = \beta_0 + \beta_1 VAs + \beta_2 VBs + \beta_3 VCs + \beta_4 IA + \beta_5 IB + \beta_6 IC + \beta_7 IAs + \beta_8 IBs + \beta_9 ICs + \beta_{10} TempAmb + \beta_{11} TempTrafo \quad (1)$$

Após a obtenção dos valores dos coeficientes do modelo, foi calculado o rendimento estimado para as amostras do conjunto de validação e foi obtido o valor médio para cada uma das subestações. Através desse processo, pode ser avaliada a acurácia do modelo e podem ser detectadas eventuais diferenças significativas entre as médias dos

Tabela 4. Valores de β para o Modelo de Estimativa de Rendimento de Transformadores.

Coefficiente	Grandeza	Estimativa
β_0	Intercepto	-1,4905
β_1	VAs	$-4,4043 \times 10^{-4}$
β_2	VBs	$2,4507 \times 10^{-2}$
β_3	VCs	$-6,9359 \times 10^{-3}$
β_4	IA	$-9,1113 \times 10^{-2}$
β_5	IB	$-1,6122 \times 10^{-2}$
β_6	IC	$-4,1973 \times 10^{-2}$
β_7	IAs	$6,8892 \times 10^{-4}$
β_8	IBs	$3,1716 \times 10^{-4}$
β_9	ICs	$5,1259 \times 10^{-4}$
β_{10}	TempAmb	$-4,2191 \times 10^{-4}$
β_{11}	TempTrafo	$7,3808 \times 10^{-4}$

valores reais e estimados. Na Tabela 5 são apresentadas as médias do rendimento real e do rendimento estimado para cada uma das subestações do conjunto de validação, bem como os erros relativos entre esses valores e o valor médio do erro relativo.

Tabela 5. Resultado agregado das simulações.

Subestação	Média do Rendimento (%)		Erro Relativo (%)
	Real	Estimado	
34/5	75,68	74,75	1,23
48/5	58,48	65,98	12,82
55/5	69,66	71,72	2,96
67/13	72,89	72,39	0,69
68/1	76,86	73,47	4,41
74/11B	77,19	74,39	3,63
112/13	89,63	81,52	9,05
Média do Erro Relativo (%)			4,97

Foram criados gráficos que apresentam visualmente as diferenças entre os dados reais, obtidos pelos sensores, com os dados estimados pelo modelo, como um elemento de apoio para a análise dessas diferenças. Alguns desses gráficos são apresentados a seguir, com os dados coletados representados na cor azul e os dados estimados representados na cor laranja. Foram selecionados dois casos em que o rendimento estimado foi próximo do rendimento real e dois casos em que houve um erro significativo entre esses valores. Nesses dois últimos casos, são apresentadas as hipóteses relacionadas à ocorrência desses erros.

O rendimento da Subestação 67/13 é apresentado na Fig. 2, onde pode ser observado que não existe uma diferença significativa entre os valores. Conforme a Tabela 5, a média do rendimento real é 72,89% e a média do rendimento estimado é 72,39%. Isso representa um erro de 0,69%, o que é um valor menor do que a média do erro relativo de 4,97%. Essa diferença não é suficiente para apontar a necessidade de uma análise mais aprofundada no rendimento do transformador desta subestação. No contexto da utilização do modelo desenvolvido para auxiliar na manutenção ou mesmo detecção de falhas nos transformadores, essa subestação não teria necessidade de ser melhor avaliada, principalmente quando existem outras subestações com maior prioridade para a execução da manutenção preventiva.

De forma semelhante, pode ser observado que existe pouca diferença entre os dados reais e estimados para a Subestação 34/5, conforme é apresentado na Fig. 3. Segundo a Tabela 5, a média do rendimento real é 75,68%, enquanto a média

do rendimento simulado é 74,75%. Isso representa um erro de 1,23%, o que é um valor menor do que a média do erro relativo de 4,97%. Essa subestação também não teria que ser priorizada em relação à execução da manutenção preventiva.

Figura 2. Rendimento da Subestação 67/13.

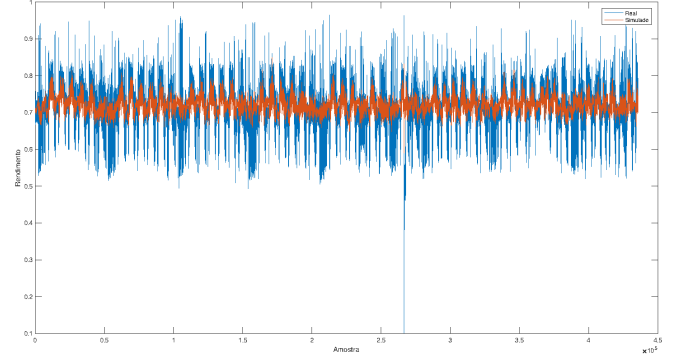
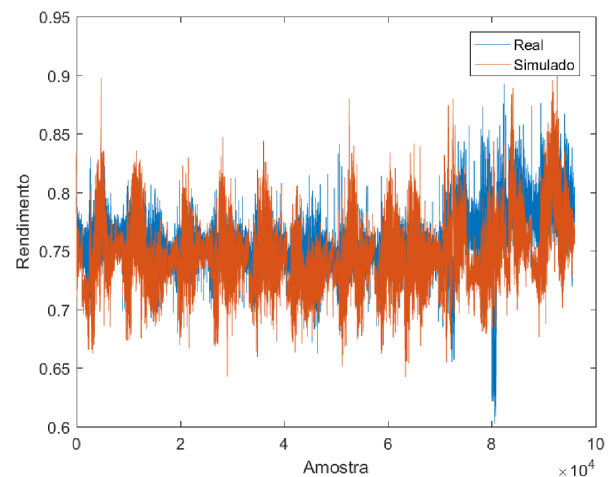


Figura 3. Rendimento da Subestação 34/5.



O rendimento da Subestação 48/5 é apresentado na Fig. 4. É possível visualizar que o rendimento real está significativamente abaixo do rendimento estimado. Conforme a Tabela 5, essa foi a subestação que apresentou a maior diferença entre esses valores. A média de rendimento real é 58,48%, enquanto a média estimada é 65,98%. Isso representa um erro de 12,82%, o que é um valor maior do que a média do erro relativo de 4,97%. Dessa forma, o rendimento real dessa subestação ficou abaixo do valor previsto pelo modelo.

O resultado apresentado na Fig. 4 sugere que a Subestação 48/5 poderia estar apresentando algum problema de mau funcionamento, o que mereceria uma maior investigação por parte da concessionária. Entretanto, como pode ser observado na Fig. 5, o carregamento do transformador está muito abaixo da sua capacidade nominal de 1.600 amperes. Nestes casos, o custo total de operação do transformador é alto e interfere diretamente no seu rendimento.

Em contrapartida, conforme apresentado na Fig. 6, a Subestação 112/13 apresentou um rendimento real acima daquele que foi estimado pelo modelo. Essa diferença

Figura 4. Rendimento da Subestação 48/5.

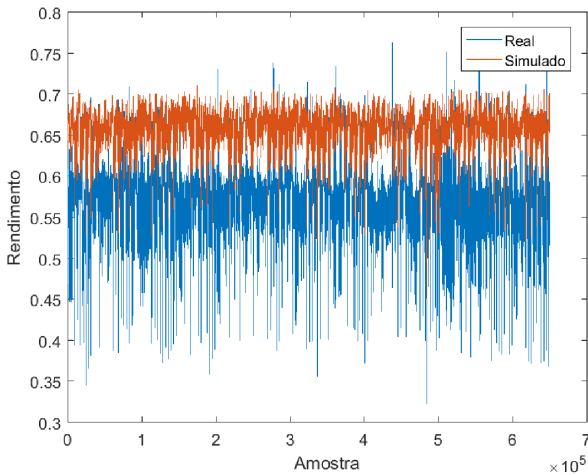
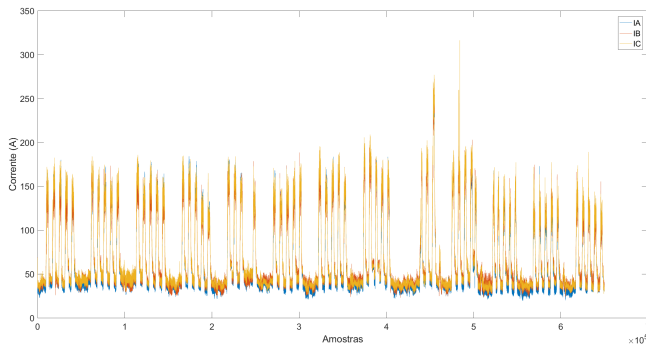


Figura 5. Corrente do secundário da Subestação 48/5.



também pode ser percebida na Tabela 5, em que é apresentada uma média de rendimento real de 89,63% e uma média de rendimento estimado de 81,52%. Isso representa um erro de 9,05%, o que é um valor maior do que a média do erro relativo de 4,97%.

Pela posição geográfica, a Subestação 112/13 está localizada em uma região predominantemente comercial. Em hipótese, a carga aplicada nessa região não é tão agressiva para o transformador em relação àquele aplicada em uma região industrial, que possui características que interferem diretamente no rendimento dos transformadores. Na Fig. 7 é apresentado o comportamento da corrente do secundário da Subestação 112/13, sendo possível observar a divisão semanal da utilização deste transformador.

3. CONCLUSÕES

As Redes Inteligentes podem ser caracterizadas como a superposição de um sistema unificado de comunicação e controle ao sistema existente de distribuição de energia elétrica, de forma a agregar diversas funcionalidades proporcionadas pelas tecnologias atuais. Os sistemas de monitoramento associados a essa superposição geram um volume massivo de dados, cujo potencial para a descoberta de conhecimento relevante para a gestão dos sistemas de distribuição elétrica nem sempre é plenamente explorado. Esses dados apresentam as características de volume, velocidade, variedade,

Figura 6. Rendimento da Subestação 112/13.

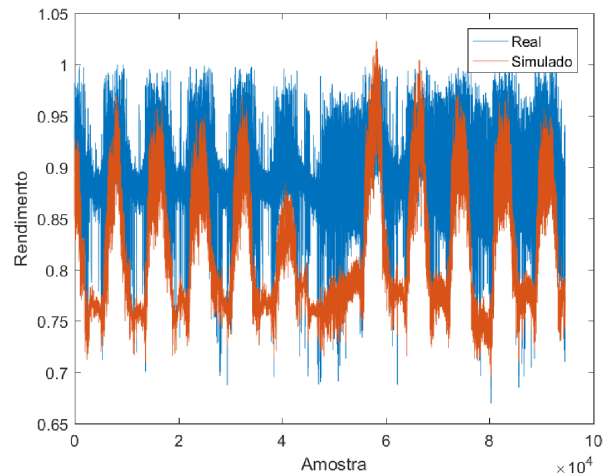
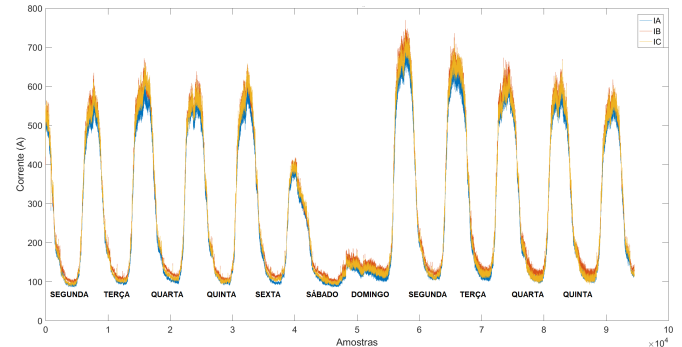


Figura 7. Corrente do secundário da Subestação 112/13.



valor e veracidade associadas ao conceito de *Big Data*, o que representa, ao mesmo tempo, uma oportunidade e um desafio.

Assim, diversos trabalhos apontam que a aplicação da Mineração de Dados se mostra vantajosa na análise dos dados provenientes dos sistemas de monitoramento das Redes Inteligentes. A partir dessa aplicação, podem ser descobertos padrões que estão implícitos nos dados analisados e que poderiam permanecer ocultos caso os dados fossem analisados através de análise tradicionais. Esses padrões, quando contextualizados em relação aos sistemas estudados, podem se transformar em conhecimento relevante para o seu entendimento e gestão.

Dentro desse contexto de análise dos dados provenientes das Redes Inteligentes, esse artigo apresentou o desenvolvimento do Modelo de Estimativa de Rendimento de Transformadores, criado a partir dos dados reais coletados e registrados pelo sistema de monitoramento dos transformadores das subestações subterrâneas da CEEE-D localizadas na cidade de Porto Alegre - RS. Esse modelo foi validado a partir da sua aplicação aos dados do conjunto de validação, de forma que foi constatada uma média de erro relativo de 4,97% entre as médias do rendimento real e do rendimento estimado.

Para as duas subestações que apresentaram um erro relativo maior do que essa média, foi constatado que elas apresentavam condições de operação que justificavam essa diferença, sendo as conclusões validadas por um especialista no domínio. Dessa forma, caracterizou-se a análise descritiva e diagnóstica proposta nesse trabalho, uma vez que foi possível entender o comportamento do sistema em estudo e identificar as causas possíveis para os eventos fora de seu comportamento esperado, o que serve de subsídio para ações e tomada de decisões por parte dos profissionais responsáveis por sua gestão.

O modelo apresentado nesse trabalho foi capaz de obter conhecimento a partir de um banco de dados que não era utilizada para esse fim, pois o monitoramento se restringia a alarmar a ocorrência de eventos específicos e pré-definidos. A partir do desenvolvimento e disponibilização desse modelo é possível ampliar significativamente o conhecimento sobre eventos até então desconhecidos. A ocorrência desses eventos, que estava implícita nos dados armazenados, só foi tornada explícita a partir da aplicação da técnica estatística de Regressão Linear Múltipla.

Como continuidade do trabalho, propõe-se um maior aprofundamento na análise dos dados e na avaliação das conclusões obtidas a partir desse processo. Por exemplo, podem ser exploradas técnicas estatísticas que diminuem o erro relativo encontrado, como forma de aumentar a acurácia dos modelos. Dessa forma, acredita-se que seja possível apresentar uma contribuição mais efetiva para os responsáveis pela gestão dos sistemas de distribuição de energia elétrica, o que pode incluir a descoberta de padrões implícitos nos dados que não são de conhecimento para especialistas no domínio. Nesses trabalhos futuros, pretende-se também uma documentação mais explícita das etapas do processo de Mineração de Dados, de forma que seja proposto um *framework* para análise dos dados provenientes das Redes Inteligentes.

AGRADECIMENTOS

Os autores agradecem o apoio da Companhia Estadual de Distribuição de Energia Elétrica (CEEE-D) do Estado do Rio Grande do Sul, através do Programa de Pesquisa e Desenvolvimento da Agência Nacional de Energia Elétrica (ANEEL), pelo apoio técnico e financeiro.

REFERÊNCIAS

- Amin, S.M. and Wollenberg, B.F. (2005). Toward a Smart Grid: Power Delivery for the 21st Century. *IEEE power and energy magazine*, 3(5), 34–41.
- Bhuiyan, S.M., Khan, J.F., and Murphy, G.V. (2017). Big Data Analysis of the Electric Power PMU Data from Smart Grid. In *SoutheastCon 2017*, 1–5. IEEE.
- Chaganti, M. and Chowdhury, B.H. (2004). Data Mining of Substation Relay Data. In *IEEE Power Engineering Society General Meeting, 2004.*, 369–375. IEEE.
- Cox, M. and Ellsworth, D. (1997). Application-Controlled Demand Paging for Out-of-Core Visualization. In *Proceedings of Visualization'97*, 235–244. IEEE.
- Crossley, P. and Liu, Y. (2009). Data Mining of the Substation Data in Distribution Network Using Rough Set and Genetic Algorithms. In *2009 44th International Universities Power Engineering Conference (UPEC)*, 1–5. IEEE.
- Dang-Ha, T.H., Olsson, R., and Wang, H. (2015). The Role of Big Data on Smart Grid Transition. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, 33–39. IEEE.
- Dobaev, A., Maslakov, M., and Dedegkaeva, A. (2016). Development of Decision Support System for Data Analysis of Electric Power Systems. In *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 1–4. IEEE.
- Electric Power Research Institute (2008). The Green Grid - Energy Savings and Carbon Emissions Reductions Enabled by a Smart Grid.
- He, X., Ai, Q., Qiu, R.C., Huang, W., Piao, L., and Liu, H. (2017). A Big Data Architecture Design for Smart Grids Based on Random Matrix Theory. *IEEE transactions on smart Grid*, 8(2), 674–686.
- Hor, C.L. and Crossley, P. (2005). Analysis of Substation Data for Knowledge Extraction. In *IEEE Power Engineering Society General Meeting, 2005*, 656–662. IEEE.
- IBM Corporation (2012). Managing Big Data for Smart Grids and Smart Meters. URL <ftp://public.dhe.ibm.com/software/pdf/industry/IMW14628USEN.pdf>.
- IEEE Smart Grid (2018). Big Data Analytics in the Smart Grid. URL <http://resourcecenter.smartgrid.ieee.org/sg/product/publications/SGWP0003>.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A.H. (2011). Big Data: The Next Frontier for Innovation, Competition and Productivity.
- MathWorks (2020a). MATLAB. URL <https://www.mathworks.com/products/matlab.html>.
- MathWorks (2020b). Statistics and Machine Learning Toolbox. URL <https://www.mathworks.com/products/statistics.html>.
- Narayanan, U., Paul, V., and Joseph, S. (2017). Different Analytical Techniques for Big Data Analysis: a Review. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 372–382. IEEE.
- Sagiroglu, S., Terzi, R., Canbay, Y., and Colak, I. (2016). Big Data Issues in Smart Grid Systems. In *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, 1007–1012. IEEE.
- Stimmel, C.L. (2014). *Big Data Analytics Strategies for the Smart Grid*. Auerbach Publications.
- Yu, N., Shah, S., Johnson, R., Sherick, R., Hong, M., and Loparo, K. (2015). Big Data Analytics in Power Distribution Systems. In *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 1–5. IEEE.
- Zhou, K., Fu, C., and Yang, S. (2016). Big Data Driven Smart Energy Management: from Big Data to Big Insights. *Renewable and Sustainable Energy Reviews*, 56, 215–225.