

Análise do Impacto das Ações de Manutenção nos Indicadores de Continuidade em Redes de Distribuição utilizando Machine Learning e Regressão com Dados em Painel

Vitor Hugo Ferreira*, Lorena B. Oliveira*, André C. Pinho*, Henrique O. Henriques*, M.Z.Fortes*, Fabricio A. Nunes**, Alejandro C. A. Pose***, Rodrigo B. de Oliveira***

*Universidade Federal Fluminense, PPGEET, Niterói, RJ, Brasil

**Light Serviços de Eletricidade S.A., Rio de Janeiro, RJ, Brasil

***Siglasul Consultoria LTDA, Rio de Janeiro, RJ, Brasil

(e-mails: vhferreira@id.uff.br, lorennaoliveira@id.uff.br, acpinho@id.uff.br, hohenriques@id.uff.br, mzamboti@id.uff.br, fabricio.nunes@light.com.br, apose@siglasul.com.br, rbarbosa@siglasul.com.br)

Abstract: The regulation of the electric energy sector provides for the service of consumers, producers and other agents in the sector in order to guarantee continuity and quality in the supply of electric energy. In the electricity distribution sector, distributors that do not meet the established requirements have their profitability impaired. Thus, meeting the targets for quality indicators is a fundamental objective to be pursued by the distributors. This article presents a methodology based on Machine Learning and Regression with Panel Data to estimate the impact of maintenance investments on the quality indicators of Light's airlines. To this end, annual histories of investments in maintenance actions are used in each line of the company, allowing the identification of actions that have a positive impact on the respective quality indicator (SAIDI, SAIFI and compensations). This paper presents the preliminary results of the P&D project financed by Light and entitled "Management Support Tool for Planning Maintenance Actions for Distribution Networks with Predictable Impact on the indicators DEC, FEC and Financial Compensation" (P&D LIGHT 116/18).

Resumo: A regulamentação do setor de energia elétrica prevê o atendimento dos consumidores, produtores e demais agentes do setor de forma a garantir a continuidade e qualidade no fornecimento de energia elétrica. No setor de distribuição de energia elétrica, as distribuidoras que não atendem aos requisitos estabelecidos possuem sua rentabilidade prejudicada. Assim, o atendimento às metas para os indicadores de qualidade constitui objetivo fundamental a ser perseguido pelas distribuidoras. Este artigo apresenta uma metodologia baseada em *Machine Learning* e Regressão com Dados em Painel para estimação do impacto dos investimentos em manutenção nos indicadores de qualidade das linhas aéreas da Light. Para tanto, são utilizados os históricos anuais de investimentos em ações de manutenção em cada linha da empresa, permitindo identificar as ações que produzem impacto positivo no respectivo indicador de qualidade (DEC, FEC e compensações). Este trabalho apresenta os resultados preliminares do projeto de P&D financiado pela Light e intitulado "Ferramenta de Apoio Gerencial para Planejamento das Ações de Manutenção das Redes de Distribuição com Previsibilidade do Impacto nos Indicadores de DEC, FEC e Compensação Financeira" (P&D LIGHT 116/18).

Keywords: maintenance; power quality index; SAIDI; SAIFI; power distribution; machine learning.

Palavras-chaves: manutenção; indicadores de qualidade; DEC; FEC; distribuição de energia; aprendizado de máquinas.

1. INTRODUÇÃO

Os processos de tomada de decisão e gestão da manutenção quando bem realizados podem proporcionar um aumento na confiabilidade e redução de perdas nos processos (Mamun and Islam, 2016) (Ruschel, 2017). No setor de energia elétrica a eficácia desses processos é essencial para a garantia da

continuidade do serviço e atendimento das regulamentações vigentes (RUDIN et al., 2012).

A ANEEL (Agência Nacional de Energia Elétrica) é a responsável no Brasil por estas regulamentações e busca, por meio de normativas, garantir que a qualidade mínima no atendimento às demandas de energia seja atendida. A partir do PRODIST (Procedimentos de Distribuição de Energia Elétrica

no Sistema Elétrico Nacional) são avaliadas a continuidade do serviço utilizando como comparador os indicadores de continuidade coletivos (DEC e FEC) e individuais (DIC, FIC e DIMIC) (ANEEL, 2020a). Quando os limites destes últimos são transgredidos, a distribuidora deve compensar financeiramente o consumidor em prazo estipulado (ANEEL, 2020b)(Barbosa and Carvalho, 2017).

Tendo em vista as altas taxas pagas anualmente pelas distribuidoras e a necessidade de aumento da confiabilidade, diversos autores têm estudado modelos para a melhoria dos indicadores de continuidade a fim de auxiliar a tomada de decisão gerencial na manutenção (Phoothong *et al.*, 2008), (Teera-achariyakul *et al.*, 2010) e (Mirsaeedi *et al.*, 2018) e expansão das redes (Mussoi and Teive, 2013)(De Campos *et al.*, 2013).

Louback (2018) apresentou em seu trabalho uma metodologia de previsão dos indicadores DEC e FEC utilizando redes neurais artificiais (RNA). Magalhães (2017) propôs os modelos utilizando técnicas de Regressão Linear Múltipla (RLM) e RNAs. Segundo os autores, a previsão dos valores desses indicadores de continuidade pode auxiliar a concessionária na realização de um planejamento mais eficiente da manutenção, aumentando a confiabilidade e oportunizando a identificação de áreas que possuem tendência de piora no tempo.

Garantir a gestão eficiente das despesas é vital para a rentabilidade das distribuidoras de energia elétrica (Brandão *et al.*, 2017). Sendo assim, a busca por soluções que resultem na maximização da eficiência na alocação dos investimentos em manutenção pode contribuir para a melhoria da rentabilidade e qualidade do serviço prestado.

Este trabalho busca então estimar o impacto das ações de manutenção nos indicadores de qualidade, permitindo a alocação de recursos de forma otimizada. Para tanto, o trabalho se divide em duas partes. Na primeira, é utilizado um algoritmo de *Machine Learning* não-supervisionado para o agrupamento das linhas de distribuição segundo suas características geográficas e técnicas. A partir da criação das unidades de análise, uma regressão com dados em painel e efeito fixo, busca estimar, na segunda etapa, o impacto das ações de manutenção nos indicadores das linhas. Esta metodologia foi aplicada ao histórico de ações de manutenção e de indicadores de qualidade por linha da Light S.A para o período de 2016 a 2018, apresentando resultados promissores ao identificar ações que podem melhorar os indicadores em até 97% das linhas aéreas da Light S.A.

2. METODOLOGIA

A metodologia utilizada para a criação de um estimador dos índices de continuidade DEC, FEC e Compensação foi dividida em duas partes tendo em vista a esparsidade dos dados utilizados. A etapa 1 utilizou um algoritmo não-supervisionado de *Machine Learning* para separar as linhas de distribuição de energia elétrica da Light em diferentes agrupamentos a partir da semelhança existente nas características técnicas e geográficas de cada linha.

Após a identificação de agrupamentos de linhas foram construídas regressões com dados em painel para estimação dos impactos de cada ação de manutenção nos indicadores de qualidade em cada agrupamento formado na etapa 1.

2.1 Etapa 1 - Classificação em Famílias

O agrupamento proposto nesta primeira etapa levou em consideração a necessidade de garantir um número mínimo de registros utilizados para cada unidade a ser analisada na etapa 2. Assim, tendo em vista que nem todas as linhas de distribuição possuíam estes históricos de manutenção foi proposto uma clusterização das redes de distribuição a fim de minimizar os erros estatísticos de estimação.

Além desta característica, buscou-se também através da criação de uma nova unidade de análise minorar os efeitos de outros fatores externos, ou seja, aqueles que não podem ser atribuídos exclusivamente à ação investigada.

A base de clusterização utilizada neste trabalho foi obtida por meio de uma extração do *SmallWorld* (referência: outubro de 2018) com as características geográficas, climatológicas, comerciais (clientes) e técnicas (comprimento, equipamentos, tipos de cabo, potência instalada) das 2.372 linhas provenientes dos 2.133 alimentadores da LIGHT. Vale ressaltar que as linhas foram analisadas neste trabalho.

Um estudo de esparsidade dos valores das variáveis originais em relação a quantidade de linhas utilizadas foi realizado previamente a fim de alcançar uma metodologia de uso de dados adequada. Tendo em vista a esparsidade dos dados encontrada, foi proposta uma agregação da base de clusterização a partir dos aspectos construtivos das mesmas. Nesta etapa foi realizada a avaliação do banco de dados por especialistas a fim de segmentar o banco de dados iniciais.

A primeira agregação de linhas foi realizada separando as linhas oriundas de subestações SETD (majoritariamente urbanas) das linhas oriundas de subestações SESD (majoritariamente rurais). Analogamente, as linhas do tipo LDA, LRA, LSA, LSS e LDS também foram agregadas, tendo em vista que esta caracterização da linha contribui para o direcionamento do tipo de ação de manutenção realizada.

2.1.1 Algoritmo de Clusterização Automática

Um problema de clusterização ou agrupamento de dados consiste em, dada uma base de dados X , agrupar (clusterizar) os objetos (elementos) de X de modo que objetos mais similares fiquem no mesmo cluster (grupo ou família) e objetos menos similares sejam alocados para clusters distintos. Existem diversos algoritmos para solução de problemas de clusterização. Um dos mais populares é o algoritmo k-means (Haykin, 1999). Essa técnica busca minimizar a distância dos elementos a um conjunto de k centróides a serem estimados de forma iterativa. Para este cálculo foi utilizada a distância euclidiana, que segundo (Haykin, 1999) é a mais utilizada neste processo. O algoritmo que é iterativo, calcula a distância do ponto ao centro mais próximo dele, devendo minimizar a função apresentada em (1), onde $\chi = \{x_1, x_2, \dots, x_k\}$ são os centróides dos k clusters, e $d(p_i, X)$ é definida como sendo

a distância do ponto ao centro mais próximo dele (Linden, 2009). A Fig. 1 apresenta as iterações do método k-means.

$$d(P, X) = \frac{1}{n} \sum_{i=1}^n d(p_i, X)^2 \quad (1)$$

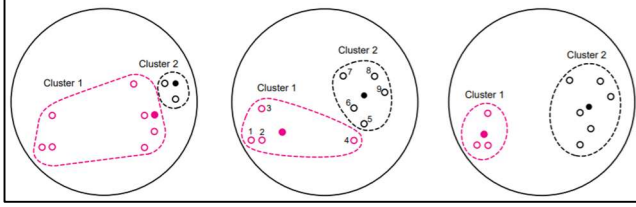


Fig. 1 Representação do método K-means (Faber, 1994).

Tendo em vista que o *k-means* é um modelo que admite como premissa a existência de agrupamentos com distribuições gaussianas centradas nos respectivos centróides (Hamerly, 2004), e que para a utilização do algoritmo é necessário o conhecimento prévio do número *k* (*clusters*), para a estimativa automática deste parâmetro foi utilizado como procedimento o critério de informação bayesiana (*Bayesian Information Criterion - BIC*). Sua utilização permite a construção de uma função de máxima verossimilhança de tal sorte que se obtém valores para o *BIC*. Assim, a partir de testes de formação de agrupamentos com diferentes valores de *k*, o agrupamento mais consistente é aquele com maior valor para a estatística *BIC* (Pelleg, 2000). A Fig. 2 apresenta o gráfico da variação do valor do BIC com o aumento do número de *clusters* e a escolha do número ideal de agrupamentos para os dados apresentados.

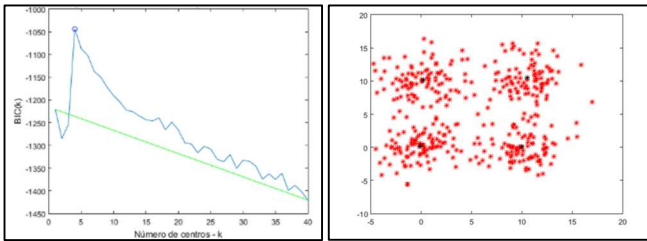


Fig. 2 – Gráfico de evolução do BIC (em azul) obtido a partir de dados artificiais (vermelho).

2.2 Etapa 2 - Regressões com Dados em Painel

A segunda etapa da metodologia busca desenvolver uma formulação matemática que apresente como resultado a estimativa dos indicadores de qualidade DEC e FEC e a compensação financeira a partir dos dados históricos de investimento na manutenção das redes de distribuição.

Conceitualmente o DEC é o intervalo de tempo que, em média, cada consumidor do conjunto considerado ficou privado do fornecimento de energia elétrica, no período de observação. No entanto, segundo (ANEEL, 2020a), apenas as interrupções maiores ou iguais a 3 (três) minutos são consideradas para o cálculo que é realizado conforme apresentado em (2). No caso de violação do limite de continuidade individual do indicador, o cálculo realizado para compensação é dado por (3).

$$DEC = \frac{\sum_{i=1}^{C_c} DIC(i)}{C_c} \quad (2)$$

$$Valor = \left(\frac{DICv}{DICp} - 1 \right) DICp \times \frac{EUSD_m}{730} \times kei \quad (3)$$

- DEC* Duração equivalente de interrupção por unidade consumidora, expressa em horas e centésimos de hora;
- i* Índice de unidades consumidoras atendidas em BT ou MT faturadas do conjunto;
- C_c* Número total de unidades consumidoras faturadas do conjunto no período de apuração, atendidas em BT ou MT;
- DIC(i)* Duração de Interrupção Individual por Unidade Consumidora, excluindo-se as centrais geradoras;
- DICv* duração de interrupção por unidade consumidora ou por ponto de conexão, conforme cada caso, verificada no período considerado, expressa em horas e centésimos de hora;
- DICp* limite de continuidade estabelecido no período considerado para o indicador de duração de interrupção por unidade consumidora ou por ponto de conexão, expresso em horas e centésimos de hora;
- EUSD_m* média aritmética dos encargos de uso do sistema de distribuição correspondentes aos meses do período de apuração do indicador;
- kei* coeficiente de majoração cujo valor deve ser fixado em 15 (BT), 20 (MT) ou 27 (AT).

Da mesma forma, segundo (ANEEL, 2020a), o FEC que representa o número de interrupções que, em média, cada consumidor do conjunto considerado sofreu no período de observação. Neste caso também são considerados para efeito de cálculo apenas as interrupções maiores ou iguais a 3 (três) minutos conforme apresentado em (4). Em casos onde há violação dos limites estipulados é realizada a compensação segundo (5).

$$FEC = \frac{\sum_{i=1}^{C_c} FIC(i)}{C_c} \quad (4)$$

$$Valor = \left(\frac{FICv}{FICp} - 1 \right) FICp \times \frac{EUSD_m}{730} \times kei \quad (5)$$

- FEC* Frequência equivalente de interrupção por unidade consumidora, expressa em número de interrupções e centésimos do número de interrupções;
- FIC(i)* Frequência de Interrupção Individual por Unidade Consumidora, excluindo-se as centrais geradoras;
- FICv* frequência de interrupção por unidade consumidora ou por ponto de conexão, conforme cada caso, verificada no período considerado, expressa em número de interrupções;
- FICp* limite de continuidade estabelecido no período considerado para o indicador de frequência de interrupção por unidade consumidora ou por

ponto de conexão, expresso em número de interrupções e centésimo do número de interrupções.

O algoritmo desenvolvido utilizou regressores multivariados para cada agrupamento formado na etapa 1 utilizando os dados em painel (observações ao longo do tempo).

Os dados utilizados na construção da base para a regressão foram disponibilizados pela Light e contemplaram o horizonte de três anos, de 2016 a 2018. Assim, a base foi composta pelas seguintes informações:

- a) Base de dados de inspeções e de podas;
- b) Base de dados de proteção e de seccionamentos da rede;
- c) Base de dados de aumento de capacidade;
- d) Ações voltadas à robustez da rede; e
- e) Listagem de obras estruturantes.

Três regressores foram construídos nesta etapa, e em cada um deles foi associado um indicador de qualidade como variável dependente, são eles: clientes interrompidos (CI - relacionado como FEC), cliente hora interrompido (CHI - relacionado como DEC) e compensações financeiras. As variáveis ambientais disponibilizadas também foram utilizadas a fim de representar os ofensores não controláveis pelas ações de manutenção, a saber: pluviometria, temperatura, velocidade média e velocidade máxima de vento.

3. REGRESSÃO MÚLTIPLA COM DADOS EM PAINEL

Segundo Gujarati (2011), pode-se afirmar que a regressão é o estudo da dependência de uma variável em relação a uma ou mais variáveis visando estimar e/ou prever o valor médio da primeira em termos dos valores conhecidos ou fixados das segundas. Sua aplicação, neste caso, se dá com a finalidade identificar para cada unidade de análise as ações que impactam na qualidade do serviço prestado pela distribuidora, bem como nas compensações financeiras.

Quando as informações disponíveis relacionam indivíduos (empresa, unidade geográfica, famílias, dentre outras) e tempo, pode-se dizer que estes dados estão em painel e nestes dados tende haver heterogeneidade. As heterogeneidades são efeitos que não podem ser observados por meio do conjunto de dados não podendo, assim, ser estimados em dados *cross-section* ou séries temporais. Desta forma, as utilizações de técnicas de estimação em dados em painel possuem a vantagem de estimar a heterogeneidade contida em cada unidade de análise (Gujarati, 2011).

Gujarati (2009) cita três tipos de métodos de estimativa de modelos de regressão de dados em painel: o modelo empilhado, o modelo de efeitos fixos e o modelo de efeitos aleatórios, sendo a análise do comportamento das heterogeneidades fundamental para a escolha do método mais adequado.

Neste caso, optou-se pela utilização do modelo de efeitos fixos, uma vez que a probabilidade das heterogeneidades das unidades de corte transversal estarem correlacionadas com as

variáveis explicativas é alta. Um exemplo disso pode ser obtido quando se observa a relação entre as ações de manutenção e as influências ambientais que cada linha está sendo submetida tendo em vista sua localização geográfica, ou seja, o maior número de podas se apresenta onde os alimentadores estão sujeitos a grande incidência de vegetação. Em (5) está apresentada a formulação da regressão em painel com efeito fixo.

$$IND_{ij}(k) = \underline{IND}_{ij} + \sum_{l=1}^n b_l^{ij} [x_l^{ij}(k) - \underline{x}_l^{ij}] + e(k) \quad (5)$$

Em que:

$IND_{ij}(k)$ representa o indicador de qualidade (CI, CHI ou compensação) da linha i do conjunto j no ano k ;

\underline{IND}_{ij} o indicador médio (2016 a 2018) de qualidade (CI, CHI ou compensação) da linha i do conjunto j ;

b_l^{ij} o l -ésimo coeficiente do regressor da linha i do conjunto j ;

$x_l^{ij}(k)$ o valor da variável explicativa l na linha i do conjunto j no ano k ;

\underline{x}_l^{ij} o valor médio (2016 a 2018) da variável explicativa l na linha i do conjunto j ; e

$e(k)$ o erro para o ano k .

As variáveis explicativas utilizadas neste modelo levaram em conta as informações ambientais (temperatura, precipitação, velocidade média e velocidade máxima do vento) e investimentos em ações de manutenção.

Vale ressaltar que esta metodologia permite identificar as ações de manutenção que contribuem para melhoria da qualidade do fornecimento em cada linha. Nesse sentido, são analisadas duas hipóteses para a formulação final:

- 1) A hipótese do coeficiente b_l^{ij} ser não nulo, ou seja, avaliar se a ação x_l é relevante para estimativa de $IND_{ij}(k)$;
- 2) Avaliar se a ação x_l é relevante para redução do indicador $IND_{ij}(k)$, ou seja, verificar a hipótese do coeficiente b_l^{ij} ser negativo;

Para tais avaliações o teste de hipótese t é utilizado. E em casos onde a variância é não constante, o erro padrão utilizado no teste t deve ser corrigido utilizando a técnica de correção de White (Gujarati, 2011).

4. RESULTADOS

Diferentes arranjos de entrada de dados foram testados, tanto na etapa 1 na escolha dos dados técnicos a serem considerados para o agrupamento das linhas de distribuição, quanto na etapa 2 na escolha das variáveis do regressor.

Neste trabalho, estão apresentados os resultados das simulações referentes a utilização de 7 variáveis de entrada na

etapa 1, e 6 variáveis de entrada na etapa 2. As tabelas 1, 2 e 3 apresentam os resultados desta análise.

Tabela 1. Percentual de linhas onde o regressor $IND_{ij}(k)$ para o indicador CHI apresentou ao menos um coeficiente b_l^{ij} não-nulo e negativo segundo teste t com correção de White

CHI								
Tipo	Total de Linhas com Ações Explicáveis			Total de Linhas sem Ações Explicáveis			Total de Linhas	
	Q	T	L	Q	T	L	Q	L
SETD LDA	1039	100	65	0	0	0	1039	65,3
SETD LRA	0	0	0	63	100	4	63	4,0
SETD LSA	115	97	7	3	3	0	118	7,4
SESD LDA	360	99	23	3	1	0	363	22,8
SESD LRA	0	0	0	3	100	0	3	0,2
SESD LSA	5	100	0	0	0	0	5	0,3
Total	1519	95	95	72	5	5	1591	100,0

Q- quantidade, T-% tipo, L - % Light

A partir dos dados apresentados, é possível perceber que os percentuais das linhas da Light S.A. que apresentam ao menos uma ação explicável para a unidade de análise na qual está inserida foram elevados, chegando a atingir 97% das linhas de distribuição sob estudo no caso do CI (Tabela 2).

A alta aderência do modelo sugere que em cenários futuros será possível a utilização desta metodologia para auxiliar os tomadores de decisão na gestão da manutenção das redes da distribuidora. O que pode gerar o aumento na confiabilidade, redução de perdas e atendimento das regulamentações vigentes.

Análise dos resultados ainda apresentou as ações que historicamente mais impactaram na confiabilidade do sistema sob estudo, que são: instalação de equipamentos (religadores e seccionadores), manutenção de equipamentos, recondução e aumento de capacidade.

Vale ressaltar que os testes realizados utilizaram também variáveis ambientais, que apesar de auxiliarem o estimador no aumento de precisão tornam-se, em cenários futuros, um complicador para fins de previsão. Além disso, a seleção de ações relevantes por meio da análise somente do sinal do coeficiente do regressor viabiliza a escolha de mais ações explicáveis (99% do total de linhas), porém ainda está sendo analisada visto que a seleção por meio dos testes estatísticos é a mais adequada.

Tabela 2. Percentual de linhas onde o regressor $IND_{ij}(k)$ para o indicador CI apresentou ao menos um coeficiente b_l^{ij} não-nulo e negativo segundo teste t com correção de White

CI								
Tipo	Total de Linhas com Ações Explicáveis			Total de Linhas sem Ações Explicáveis			Total de Linhas	
	Q	T	L	Q	T	L	Q	L
SETD LDA	1039	100	65	0	0	0	1039	65,3
SETD LRA	0	0	0	63	100	4	63	4,0
SETD LSA	111	94	7	7	6	0	118	7,4
SESD LDA	363	100	23	0	0	0	363	22,8
SESD LRA	0	0	0	3	100	0	3	0,2
SESD LSA	0	0	0	5	100	0	5	0,3
Total	1513	95	95	78	5	5	1591	100,0

Compensação Financeira								
Tipo	Total de Linhas com Ações Explicáveis			Total de Linhas sem Ações Explicáveis			Total de Linhas	
	Q	T	L	Q	T	L	Q	L
SETD LDA	1039	100	65	0	0	0	1039	65,3
SETD LRA	0	0	0	63	100	4	63	4,0
SETD LSA	111	94	7	7	6	0	118	7,4
SESD LDA	363	100	23	0	0	0	363	22,8
SESD LRA	0	0	0	3	100	0	3	0,2
SESD LSA	0	0	0	5	100	0	5	0,3
Total	1513	95	95	78	5	5	1591	100,0

Q- quantidade, T-% tipo, L - % Light

Tabela 3. Percentual de linhas onde o regressor $IND_{ij}(k)$ para o indicador Compensação Financeira apresentou ao menos um coeficiente b_l^{ij} não-nulo e negativo segundo teste t com correção de White

Q- quantidade, T-% tipo, L - % Light

5. CONCLUSÕES

Neste artigo, apresentou-se uma metodologia baseada em *Machine Learning* e Regressão com Dados em Painel para estimar o impacto dos investimentos em manutenção nos indicadores de qualidade das linhas aéreas da Light. Na primeira etapa, as características técnicas das redes foram utilizadas para o agrupamento das linhas por semelhança. A partir desses agrupamentos, foi realizada, na segunda etapa, a regressão com dados em painel das ações de manutenção das redes utilizando o modelo de fixo. Os resultados alcançados mostraram-se efetivos tendo em vista os baixos erros de estimativa dos índices DEC, FEC e compensação financeira.

Por meio destes resultados é possível identificar as ações que produzem maior impacto nos indicadores de qualidade, auxiliando para uma eficiente tomada de decisão do planejamento da manutenção. Tal iniciativa pode, além de melhorar a qualidade dos serviços prestados pela distribuidora, proporcionar o atendimento das metas estabelecidas pela ANEEL, e consequentemente a redução dos custos provenientes do não atendimento à regulamentação vigente.

AGRADECIMENTOS

Os autores agradecem à Light e à ANEEL pelo financiamento do P&D "Ferramenta de Apoio Gerencial para Planejamento das Ações de Manutenção das Redes de Distribuição com

Previsibilidade do Impacto nos Indicadores de DEC, FEC e Compensação Financeira" (P&D LIGHT 116/18).

REFERÊNCIAS

- ANEEL^a (2020). Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST, Disponível em: <https://www.aneel.gov.br/prodist>.
- ANEEL^b (2020). Compensação pela Transgressão dos Limites de Continuidade, Disponível em: <https://www.aneel.gov.br/indicadores-de-compensacao-de-continuidade>.
- Barbosa, D. S. A. and Carvalho, P.L. (2017) ‘Impacts of the change in regulation in Brazil: Penalty simulation for violation of collective interruption indicators versus compensation for consumers’, in *CIREN - Open Access Proceedings Journal*. Institution of Engineering and Technology, pp. 729–732. doi: 10.1049/oap-cired.2017.0676.
- Brandão, R. Et al. (2017). Análise de Sustentabilidade Econômica e Financeira das Distribuidoras de Energia Elétrica no Brasil no período de 2009-2015, in *IX CITENEL*, João Pessoa, PB, 2017.
- De Campos, R. et al. (2013) ‘Application of risk analysis in prioritizing investments in terms of value-based management and power quality’, *Journal of Control, Automation and Electrical Systems*, 24(4), pp. 513–521. doi: 10.1007/s40313-013-0032-3.
- Faber, V. (1994) ‘Clustering and the Continuous k-Means Algorithm’, *Los Alamos Science*, (22), pp. 138–144.
- Gujarati, D.N., Porter, D.C. (2011). *Econometria Básica*, Ed. Bookman, 5a. edição.
- Gujarati, D. N. (2019) *ECONOMETRIA: Princípios, teoria e aplicações práticas*. Ed. Saraivauni, 1a edição.
- Hamerly, G. Charles, E. (2004). Learning the k in k-means, in *Advances in neural information processing systems*, 2004.
- Haykin, S., *Neural networks and Learning Machines*. Hamilton: 3rd, 1999.
- LIGHT, Demonstrações Financeiras 2018, Disponível em <http://ri.light.com.br/ptb/7647/674785.pdf>.
- Linden, R. (2009) ‘Técnicas de Agrupamento’, *Revista de Sistemas de Informação da FSMA*, 4, pp. 18–36.
- Louback, F.O. (2018). Previsão dos indicadores de continuidade DEC e FEC usando redes neurais artificiais. Dissertação de Mestrado, UFES.
- Magalhães, E.F.A. (2017). Modelagem e Simulação de Indicadores de Continuidade: Ferramenta auxiliar para a manutenção em redes de distribuição de energia elétrica. Dissertação de Mestrado, UFBA.
- Mamun, K. A. and Islam, F. R. (2016) ‘Reliability evaluation of power network: A case study of Fiji Islands’, in *Proceedings of the 2016 Australasian Universities Power Engineering Conference, AUPEC 2016*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/AUPEC.2016.7749359.
- Mirsaeedi, H. et al. (2018) ‘Electricity Distribution System Maintenance Budgeting: A Reliability-Centered Approach’, *IEEE Transactions on Power Delivery*. IEEE, 33(4), pp. 1599–1610. doi: 10.1109/TPWRD.2017.2772202.
- Mussoi, F. L. R. and Teive, R. C. G. (2013) ‘A multistage value-based model for prioritization of distribution projects using a multiobjective genetic algorithm’, *Journal of Control, Automation and Electrical Systems*, 24(5), pp. 623–637. doi: 10.1007/s40313-013-0060-z.
- Pelleg, D., Moore, A. (2000). X-means: extending k-means with efficient estimation of the number of clusters, in *Proceedings of the Seventeenth International Conference on Machine Learning*, 1:727–734.
- Phoothong, N. et al. (2008) ‘Optimal preventive maintenance budget setting for electric power distribution utilities’, in *5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2008*, pp. 957–960. doi: 10.1109/ECTICON.2008.4600590.
- Rudin, C. et al. (2012). Machine learning for the New York City power grid, *IEEE Trans. Pattern Anal. Mach. Intell.*, 34:328–345. doi:10.1109/TPAMI.2011.108.
- Ruschel, E. et al. (2017). Industrial maintenance decision-making: A systematic literature review, *J. Manuf. Syst.*, 45:180–194. doi:10.1016/j.jmsy.2017.09.003.
- Teera-achariyakul, N. et al. (2010) ‘Optimal allocation of maintenance budgets for reliability target setting’, in *Asia-Pacific Power and Energy Engineering Conference, APPEEC*. doi: 10.1109/APPEEC.2010.5448913.