



Inovação e Regulamentação do Setor Elétrico: O Uso de Algoritmos de Validação de Dados e Web Scraping na Análise de Dados do INDGER

Tema: Novos modelos de negócio

Autores: Mayara Correia Tenorio, Tiburcio Valeriano Dantas Gurgel, Victor Guimarães, Luciano Vianna, Aurélio Miranda

Co-Autores: -

Empresa: Equatorial Pará Distribuidora de Energia S.A

Resumo

A obrigação de envio mensal para a Agência Nacional de Energia Elétrica (ANEEL) das informações dos serviços comerciais solicitados pelos consumidores, dados de faturamento e informações técnicas das redes da distribuidora, imputou uma nova necessidade nas distribuidoras, na geração, validação e controle dessas informações encaminhadas para a agência. A ferramenta de gestão do INDGER do Grupo Equatorial garante a centralização das bases, integridade dos dados e possibilita uma validação prévia e o monitoramento eficiente e contínuo dos indicadores. Com a utilização de *web scraping* para coletar informações abertas da ANEEL e tecnologias de validação, as empresas conseguem realizar benchmarks eficazes, comparando seu desempenho com o cenário nacional. Essa abordagem integrada melhora a transparência na gestão regulatória, antecipa a identificação de problemas e facilita a tomada de decisões, tornando a adoção de tecnologias inovadoras um diferencial estratégico para a modernização dos processos.

1. Introdução

Com o avanço das regulamentações do setor elétrico e as crescentes exigências de envios de relatórios com indicadores sobre as operações das concessionárias para o órgão regulador, torna-se evidente a necessidade das distribuidoras de se prepararem tecnologicamente para essa nova realidade. Essa preparação não se limita apenas à coleta de dados, mas envolve a implementação de ferramentas analíticas avançadas que permitam uma interpretação profunda das informações disponíveis.

Ao utilizar esses dados de forma estratégica, as distribuidoras podem identificar áreas de melhoria, otimizar processos operacionais e, assim, elevar a qualidade dos serviços prestados à comunidade. Essa abordagem proativa não apenas atende às exigências regulatórias, mas também minimiza o risco de penalidades em eventuais processos de fiscalização do órgão regulador, ao antecipar melhorias nos processos. Além disso, essa prática fortalece a relação com os consumidores, que buscam serviços mais eficientes e maior transparência nas operações.

Em 29 de maio de 2020, a ANEEL abriu a Tomada de Subsídio nº 008/2020 (TS008/2020) com o objetivo de coletar opiniões sobre a alteração da estrutura dos Indicadores Gerenciais (INDGER) e a mudança na periodicidade de recebimento desse indicador, de anual para mensal e do aumento do número de campos

de 58 para 187. Em 31 de dezembro de 2020, a TS008/2020 foi encerrada por meio da Nota Técnica nº 130/2020-SFE/ANEEL e na sequência foram aprovadas a Resolução Normativa nº 956/2021, que definiu as informações exigidas dos Agentes de Distribuição, e a Resolução Normativa nº 1.000/2021, que estabeleceu as regras para a prestação do serviço público de distribuição e consolidou uma série de resoluções sobre o serviço de distribuição de energia elétrica, incluindo a Resolução Normativa nº 414/2010.

Assim, o INDGER se tornou um dos relatórios mensais obrigatórios que as distribuidoras de energia elétrica devem enviar à ANEEL, que inclui uma ampla gama de indicadores comerciais, de serviços e técnicos. Atualmente, o INDGER contempla 59 indicadores comerciais abertos por município, 15 indicadores relacionados a 101 serviços também abertos por municípios, ambos reportados por município da concessão, além de 3 entidades técnicas que somam 44 indicadores sendo estes abertos em alimentadores de distribuição, linhas de distribuição e subestações, totalizando 118 campos.

Dessa forma, fica evidente que o INDGER exige a criação de uma solução robusta, ágil e dinâmica para a extração, gestão, validação e envio desses dados à ANEEL. A natureza abrangente e rica em informações dessa base de dados sobre a área de concessão torna essencial a implementação de uma infraestrutura que não apenas atenda às exigências regulatórias, mas também permita uma análise eficiente e estratégica dos dados.

A complexidade é elevada para uma área de concessão, considere então, o cenário da Equatorial Energia, onde em fevereiro de 2023, era responsável pela concessão de sete distribuidoras de energia elétrica no Brasil, que operavam com sistemas técnicos e comerciais distintos, em especial a recém adquirida CELG. Diante do desafio que o processo do INDGER representa para suas operações, o grupo identificou a necessidade de desenvolver um sistema que não apenas colete dados de diversas fontes, mas também realize a validação de cada indicador em suma menor granularidade, gere os indicadores nos formatos exigidos pela ANEEL, forneça relatórios detalhados e apresente painéis gerenciais com uma série de visões dos indicadores analisados pela agência nos planos de resultados.

Em parceria com a empresa Norven, foi criado um validador que utiliza técnicas de mineração de dados para transformar informações dispersas em um fluxo coeso e organizado. Este validador é equipado com centenas de critérios para identificar erros críticos e gerar alertas, permitindo que as áreas responsáveis realizem verificações minuciosas, correções dos erros críticos e validação dos alertas. Essa abordagem não apenas garante a confiabilidade dos dados, mas também promove uma governança eficaz das informações.

Este trabalho visa apresentar a ferramenta desenvolvida em parceria entre o Grupo Equatorial e a Norven, destacando suas aplicações regulatórias e gerenciais, bem como as tecnologias envolvidas no processo. Além disso, busca evidenciar as oportunidades de inovação que existem no setor de regulação da energia, demonstrando como a adoção de soluções tecnológicas, como o uso do Web Scraping, pode não apenas aprimorar a eficiência e a transparência das operações, evitando retrabalho, mas também possibilitar a redução ou até a mitigação de eventuais penalidades.

2. Desenvolvimento

2.1. SISTEMA DE GESTÃO DO INDGER DO GRUPO EQUATORIAL

O desenvolvimento de ferramentas para a automatização de processos é fundamental para otimizar a eficiência e a produtividade nas organizações modernas. De acordo com um estudo de Brynjolfsson e McAfee (2014), a automação pode aumentar significativamente a produtividade, permitindo que tarefas repetitivas sejam realizadas de forma rápida e precisa, o que reduz a probabilidade de erros humanos e libera os

colaboradores para focar em atividades mais estratégicas e criativas. Além disso, a automação possibilita uma análise de dados mais eficaz, permitindo decisões informadas em tempo real.

Ao aumentar a flexibilidade e escalabilidade das operações, as empresas conseguem se adaptar rapidamente às mudanças do mercado, garantindo uma vantagem competitiva sustentável. Nesse contexto de constante evolução tecnológica, investir em automação se torna não apenas uma necessidade, mas uma estratégia crucial para o crescimento e a inovação nas organizações.

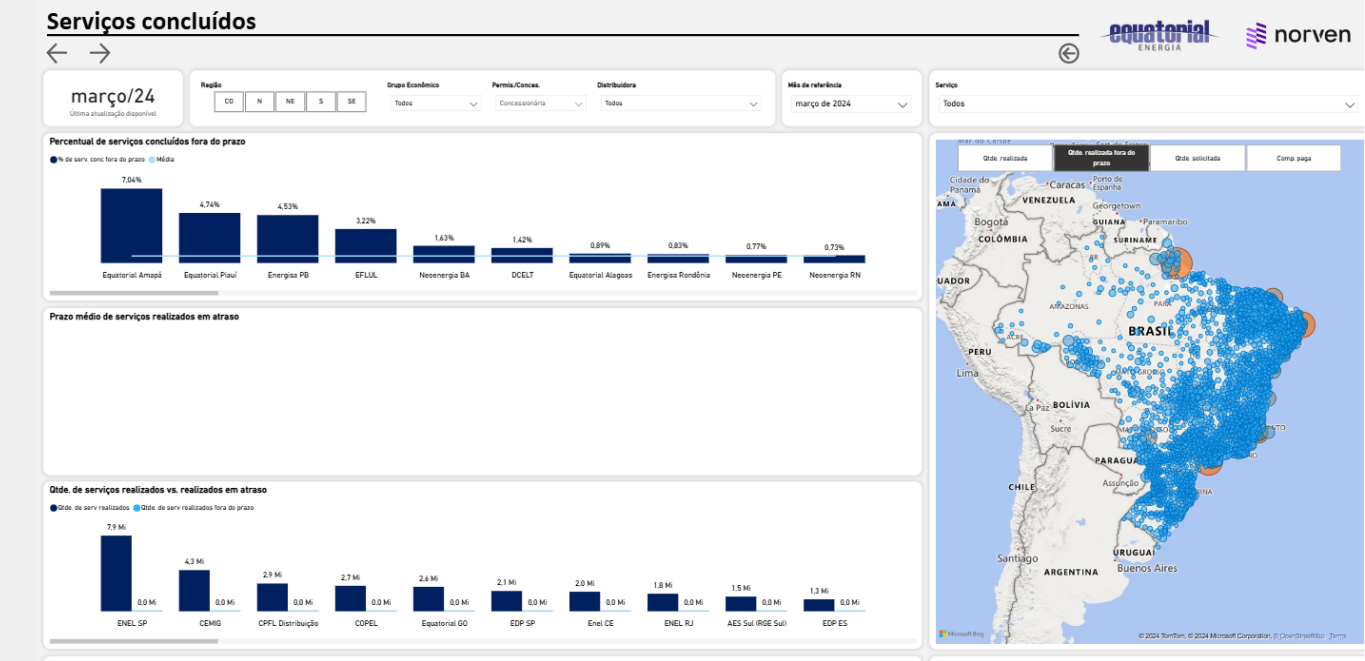
A ferramenta de gestão do INDGER do Grupo Equatorial se baseia em três tipos de dados: os que alimentam o sistema a partir de informações do próprio banco de dados das distribuidoras, os dados que são inseridos manualmente e os dados que são coletados por web scraping dos dados abertos disponibilizados no site da ANEEL. Todos os dados coletados passam por testes de sanidade, que foram implementados para garantir que as informações sejam pré-validadas, eliminando os erros críticos e impedindo não sejam enviados à ANEEL ou apresentados nos painéis gerenciáveis. O sistema foi desenhado para a realização de dupla validação, sendo uma automática dos dados através de regras pré-definidas e outra a etapa de validação do usuário validador que precisa indicar no sistema que o dado está conforme, para assegurar que a informação correta está sendo enviada ao regulador e somente fique disponível para envio para a ANEEL após validação das áreas responsáveis de cada indicador (Figura 1).

Detalhes validação					
COM_033	0	0	102 OK		Validado
COM_034	0	0	102 OK		Validado
COM_035	0	0	102 OK		Validado
COM_036	0	0	102 OK		Validado
COM_037	347	0	<div>33 Maior do que 50% da média dos últimos 12 meses</div> <div>65 OK</div> <div>4 É maior ou igual a que 15 unidades</div>		Em Análise
COM_038	109	0	<div>97 Não está entre 20% e 60% da média dos últimos 12 meses</div> <div>5 OK</div>		Em Análise
COM_039	0	0	102 OK		Em Análise
COM_040	45.549,68	0	102 OK		Em Análise
COM_041	10.278	0	102 OK		Em Análise
COM_042	105	0	102 OK		Em Análise
COM_043	1.836	0	102 OK		Validado
COM_044	0	0	102 OK		Validado

Figura 1 - Imagem do painel de validação do INDGER.

O sistema foi desenvolvido, com foco no processo de validação, dessa forma, ele foi concebido para guardar as bases de dados que deram origem às informações consolidadas definidas pela ANEEL. O sistema permite a realização do *drill down* da informação, que possibilita uma compreensão mais profunda de como esse registro foi gerado produzindo insights adicionais e possibilitando que o validador analise a base que compõe o dado (Ikeda et al., 2012), sem a necessidade de acessar outros sistemas, facilitando assim o processo de validação.

Em paralelo os Painéis de Indicadores auxiliam as áreas nas validações e no acompanhamento dos indicadores. A partir dos dados abertos coletados da ANEEL ainda são gerados painéis que realizam benchmark dos indicadores, disponibilizando o ranqueamento das distribuidoras Brasil por desempenho de cada indicador do INDGER (Figura 2).



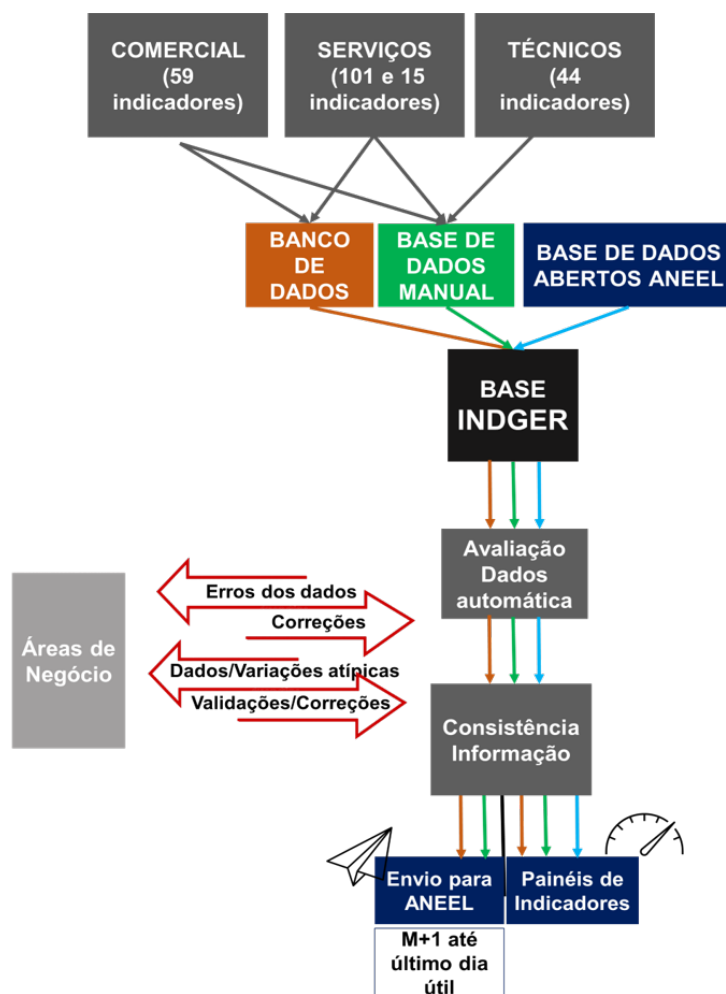


Figura 3 - Fluxo do sistema de gestão do INDGER

2.2. O PAPEL DO WEB SCRAPING NA CONSOLIDAÇÃO E ANÁLISE DOS DADOS NO SISTEMA DE GESTÃO DO INDGER

Na era da informação, a análise de dados tornou-se essencial para a competitividade do mercado, incluindo o mercado energético. O web scraping, uma técnica de coleta automatizada de dados de múltiplas fontes online, emerge como uma ferramenta crucial nesse contexto. Ao consolidar os dados do INDGER, essa técnica transforma informações em insights valiosos que promovem otimizações de processos e inteligência organizacional.

O web scraping permite a extração automatizada de dados de websites, convertendo informações não estruturadas em dados formatados para análise. Quando integrado a um sistema de Business Intelligence (BI), possibilita a realização de benchmarking, análise de tendências e outras funcionalidades de forma eficiente (Zhang et al., 2020). Essa abordagem contínua possibilita o monitoramento de tendências de desempenho ao longo do tempo, facilitando a detecção de padrões e variáveis críticas que podem impactar a performance futura. A consolidação dessas informações no sistema de BI é fundamental para garantir a eficácia das estratégias adotadas, permitindo uma tomada de decisão mais informada e proativa (Chaudhuri et al., 2011).

Na era da Big Data, o volume de informações disponíveis online é imenso, tornando o web scraping ainda mais relevante para a extração eficaz de dados. A integração dessa técnica no processo analítico proporciona diversas vantagens, especialmente na tomada de decisão baseada em dados. Entre os principais benefícios estão:

I) Acesso a Informações em Tempo Rápido: possibilita a coleta de dados atualizados rapidamente, uma característica vital para mercados dinâmicos como o setor de energia. Isso permite que as empresas ajustem suas estratégias de forma imediata, fundamentando decisões em informações precisas (Mishra et al., 2020).

II) Decisões Baseadas em Dados Concretos: a coleta massiva de dados fornece uma base robusta para análises, permitindo decisões fundamentadas em evidências em vez de suposições. Essa abordagem minimiza riscos e maximiza oportunidades (Davenport & Harris, 2007).

III) Automatização e Eficiência na Coleta de Dados: a automação reduz o tempo e o esforço necessários para a obtenção de informações, permitindo que a equipe de análise se concentre no uso estratégico dos dados. Essa eficiência garante que as informações estejam sempre atualizadas, melhorando a agilidade nas decisões (Choudhury et al., 2020).

IV) Vantagem Competitiva: A integração do web scraping com sistemas de BI melhora a eficiência operacional ao eliminar a coleta manual de dados. Essa otimização permite decisões rápidas e baseadas em informações recentes, conferindo à empresa uma vantagem competitiva em um mercado em constante evolução (Chen et al., 2012).

2.3. WEB SCRAPING UTILIZANDO AS BIBLIOTECAS PANDAS E REQUESTS COM MULTITHREADING

Foi implementado um bot de web scraping para extrair dados abertos do INDGER da ANEEL, utilizando as bibliotecas Pandas e Requests, juntamente com técnicas de multithreading. Essa solução foi desenvolvida para o Painel Brasil do INDGER, que compila dados de todas as distribuidoras de energia no Brasil, permitindo a realização de benchmark de todos os indicadores gerenciáveis.

A biblioteca Requests permite enviar solicitações para servidores web e receber respostas, permitindo interação com APIs, páginas da web e outros recursos online. Ela é extremamente útil ao recuperar informações da web ou interagir com serviços online em projetos (McKinney, 2022), possibilitando a extração eficiente de dados diretamente de páginas da web.

Por sua vez, o Pandas oferece funcionalidades avançadas para a manipulação e análise de dados, permitindo a organização das informações coletadas em estruturas de dados chamadas DataFrames, que são tabelas bidimensionais semelhantes às planilhas do Excel (Nelli, 2018). O Pandas permite fácil carregamento, limpeza, transformação e análise de dados. Ele é amplamente usado em tarefas de ciência de dados e análise exploratória, tornando o processo de trabalho com dados mais eficiente e produtivo (McKinney, 2015).

A adoção de multithreading otimiza o processo de scraping ao possibilitar a realização simultânea de múltiplas requisições. Isso resulta em uma redução significativa no tempo total de coleta de dados, especialmente quando se trata de grandes volumes de informações.

A técnica de multithreading é particularmente benéfica para mitigar os impactos da latência de rede, que pode ser um fator limitante em abordagens tradicionais de web scraping, onde as requisições são tratadas de forma sequencial. Essa abordagem não apenas aumentou a eficiência, mas também melhorou a capacidade de resposta do bot, permitindo que ele lide com diferentes fontes de dados simultaneamente, aumentando assim a escalabilidade da operação de scraping.

3. Conclusão

A implementação da ferramenta de gestão do INDGER não apenas assegura a integridade e precisão dos dados enviados à ANEEL, mas também permite um acompanhamento contínuo e eficiente dos indicadores

pelas áreas de negócio. O uso de web scraping para coletar dados abertos da ANEEL, aliado a tecnologias avançadas com a possibilidade do *drill down* para enriquecer a capacidade da validação da informação e a análise dos indicadores, enriquece a capacidade de benchmark, ranqueando cada empresa em relação ao desempenho nacional. Essa abordagem integrada fortalece a transparência e a confiabilidade do processo de gestão regulatória, contribuindo para uma melhor tomada de decisões e aprimoramento contínuo das operações. Assim, a adoção de tecnologias, torna-se um diferencial estratégico para a modernização e a eficiência na gestão dos indicadores regulatórios.

4. Referências bibliográficas

- Agência Nacional de Energia Elétrica - ANEEL. Tomada de Subsídio nº 008/2020, 29 mai. 2020.
- Agência Nacional de Energia Elétrica - ANEEL. Nota Técnica nº 130/2020-SFE/ANEEL, 31 dez. 2020.
- Agência Nacional de Energia Elétrica - ANEEL. Resolução Normativa nº 956/2021, 7 dez. 2021.
- Agência Nacional de Energia Elétrica - ANEEL. Resolução Normativa nº 1.000/2021, 7 dez. 2021.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
- Ikeda, R., Sarma, A. Das, Widom, J., "Logical provenance in data-oriented workflows?", *2012 IEEE 29th International Conference on Data Engineering (ICDE)*, pp.877-888.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An Overview of Business Intelligence Technology. *Communications of the ACM*, 54(8), 88-98.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.
- Choudhury, A., Jha, S., & Ghosh, A. (2020). Role of Automation in Business Intelligence: A Review of Trends and Applications. *Journal of Business Research*, 112, 294-301.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.
- McKinney, W. Python for data analysis. O'Reilly Media, Inc., 2022.
- McKinney, W. & Team, P. D. Pandas—Powerful python data analysis toolkit. *Pandas—Powerful Python Data Analysis Toolkit*, 2015. 162.
- Mishra, S., Singh, A., & Rani, R. (2020). Big Data Analytics in Business: A Review of Applications and Challenges. *International Journal of Business Analytics*, 7(1), 52-67.
- Nelli, F. (2018). The pandas Library—An Introduction. In: Walker, M. A. (Ed.), *Python Data Analytics* (pp. 49-?). Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-3913-1_4 (Acesso em: 08 out. 2024).
- Zhang, L., Chen, L., & Zhou, Y. (2020). Web Scraping for Data Collection: A Review of Techniques and Applications. *Journal of Data and Information Science*, 5(4), 29-50.