

## Análise de desempenho de um modelo de geração de alvos de fiscalização de perdas não técnicas

Wesley Biffe\*. Poleana Gehrke\*\*. Nelson K. Neto\*\*\*. Fernando G. K. Guarda\*\*\*\*. Alzenira da Rosa Abaide\*\*\*\*\*.

\*Universidade Federal de Santa Maria, Cachoeira do Sul, RS, Brasil  
(Tel: 303-555-5555; e-mail: wesley.biffe@acad.ufsm.br).

\*\*Universidade Federal de Santa Maria, Cachoeira do Sul, RS, Brasil  
(Tel: 303-555-5555; e-mail: poleana.gehrke@acad.ufsm.br).

\*\*\*Universidade Federal de Santa Maria, Cachoeira do Sul, RS, Brasil  
( e-mail: nelson.knak@ufsm.br).

\*\*\*\*Departamento de Eletromecânica e Sistemas de Potência, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil ( e-mail: fernando.guarda@ufsm.br).

\*\*\*\*\* Departamento de Eletromecânica e Sistemas de Potência, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil ( e-mail: alzenira@ufsm.br)

---

**Abstract:** The electrical system is made up of the generation, transmission and distribution sectors, these sectors having losses that can be divided into technical losses and non-technical losses. The objective of the present study is to analyze the performance of a prediction model and identification of outliers referring to non-technical losses. The methodology is based on the use of SPSS Software that, through the data mining process, performs a search for patterns, enabling the appointment of consumers in a situation of probable irregularity. Through the analysis of the assertiveness of the current model, it is sought to improve it and create a new predictive model in order to achieve more accurate results.

**Resumo:** O sistema elétrico é constituído pelos setores de geração, transmissão e distribuição, sendo estes setores detentores de perdas que podem ser divididas em perdas técnicas e perdas não técnicas. O objetivo do presente estudo é analisar o desempenho de um modelo de previsão e identificação de pontos discrepantes referentes às perdas não técnicas. A metodologia tem como base a utilização do Software SPSS que, através do processo de mineração de dados, realiza uma busca por padrões possibilitando o apontamento de consumidores em situação de provável irregularidade. Através da análise de assertividade do atual modelo busca-se o aprimoramento do mesmo e a criação de um novo modelo preditivo na intenção de alcançar resultados mais precisos.

**Keywords:** non-technical losses; SPSS; data mining; database; inspection; electricity.

**Palavras-chaves:** perdas não técnicas; SPSS; mineração de dados; banco de dados; fiscalização; energia elétrica.

---

### 1. INTRODUÇÃO

O sistema elétrico de potência é constituído principalmente pelos setores de geração, transmissão e distribuição. Estes setores possuem perdas, que são divididas em perdas técnicas - PT e perdas não técnicas - PNT ( Piotrowski et al. 2021). As perdas técnicas (PT) mais significativas em um sistema de distribuição ocorrem nos condutores primários, sendo causadas principalmente pela resistência dos condutores e o consequente aquecimento (Efeito Joule). Já as perdas não-técnicas representam a parcela de energia elétrica que é consumida mas não é faturada pela concessionária. Essas perdas impactam negativamente o desempenho das concessionárias de energia elétrica, causando grandes

prejuízos financeiros, além de afetar na qualidade da energia que é entregue aos consumidores.

Furto de energia é um tipo de PNT causada devido à energia consumida não faturada. Já a fraude, popularmente conhecida como “rato”, é uma adulteração por parte do consumidor, ocasionando o pagamento de uma menor parcela da quantidade de energia utilizada (Penin, 2008).

No gráfico 1 é possível observar o montante relativo a perdas técnicas e perdas não-técnicas no ano de 2019 (ANEEL, 2020). É possível notar que aproximadamente 13,77% da energia injetada no sistema é perdida.

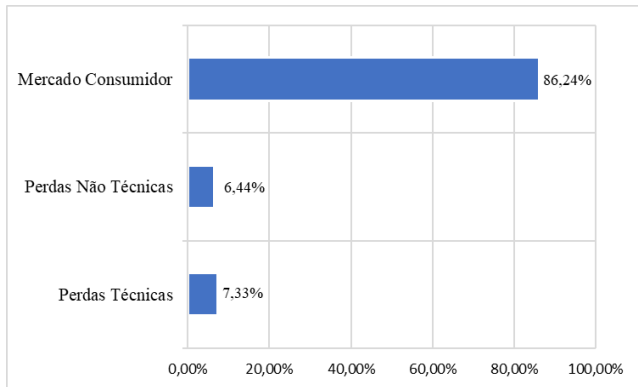


Fig. 1. Perdas referentes a energia injetada.

A redução das perdas no sistema elétrico de potência é um dos grandes objetivos tanto para os consumidores quanto para a concessionária, sendo benéfica para ambos. Isso porque os consumidores regulares terão redução dos valores regulatórios, ou seja, redução na tarifa a ser paga. Já a concessionária será capaz de realizar maiores investimentos nos seus sistemas elétricos (ANEEL, 2020).

O presente estudo tem como objetivo realizar análise de desempenho de um modelo de previsão e identificação de *outliers*, ou seja, dos pontos discrepantes, referente às perdas não-técnicas. Para isso será utilizado o *Software* SPSS, no qual é realizada uma busca de padrões através do processo de mineração de dados, possibilitando a identificação de consumidores irregulares que estejam furtando ou fraudando energia elétrica.

## 2. METODOLOGIA

Nesta seção é apresentado a metodologia para identificação das perdas não-técnicas, analisando o desempenho do modelo de previsão realizado através do *software* SPSS.

A metodologia consiste na caracterização do modelo, onde é apresentado o fluxograma de uma distribuidora de energia, juntamente com as variáveis utilizadas, processos e o modelo preditivo. Além disto é composta pela caracterização do *software* SPSS, a exposição dos parâmetros de análises realizadas demonstrando a assertividade do modelo e o diagnóstico de desempenho do fluxograma. Na figura 2 pode-se observar o esquemático utilizado na para a estruturação da metodologia.



Fig. 2. Modelo esquemático da metodologia

### 2.1 Caracterização do modelo

O diagrama da figura 3 mostra o modelo esquemático traçado, apresentando as bases de dados adotadas, com a adoção de dados exógenos, como informações climáticas e geográficas,

extração de conteúdo e pré-processamento dos mesmos, base de dados PNT, processamento de dados, emissão de indicadores e atualização da base de dados.

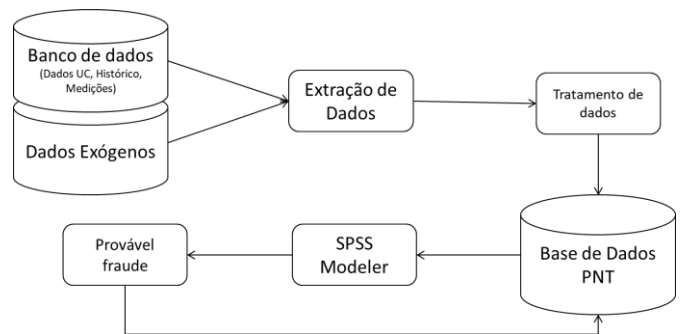


Fig. 3. Caracterização do fluxograma.

O banco de dados da empresa analisada disponibiliza as seguintes informações:

- Fiscalizações;
- Dados comerciais das unidades consumidoras;
- Dados de consumo;
- Dados de telemediação;

A base de dados “fiscalizações” é referente às fiscalizações realizadas para possíveis furtos que foram detectados em 2008. Além disso, essa base de dados fornece:

- Número referente ao cliente;
- Código da unidade consumidora;
- Data em que a fiscalização foi realizada;
- Parecer da fiscalização (Regular, Fraude)

Na base de dados “Dados Comerciais de UCs”, são fornecidos os dados sobre:

- Classe de consumo;
- Subgrupo de faturamento;
- Tipo de conexão;
- Dados do corte de energia.

Na base de dados “Dados de Consumo”, são armazenadas as leituras das UCs atendidas pela empresa, sendo uma base importante para identificação de padrões e anomalias. Nesta base de dados, são obtidos:

- Médias de consumo;
- Data de leitura;
- Tipo de leitura;

Já a base de dados “telemediação” é referente àquelas unidades que possuem medição remota, onde são obtidos:

- Código da unidade consumidora;
- Tipo de telemediação;

A distribuidora fornece também um banco de dados exógenos formado por:

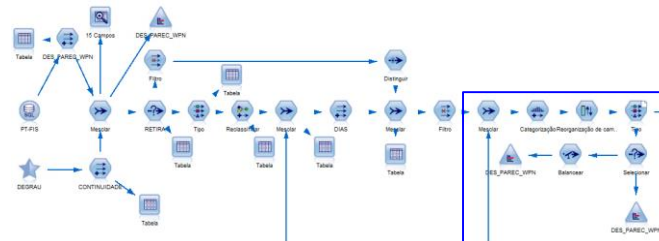
- Dados climáticos;
- Informações geográficas;
- Histórico de pagamentos;
- Dados socioeconômicos.

O processamento corresponde à etapa de mineração de dados. A partir dos dados provenientes dos bancos de dados, são aplicadas funções de filtro de dados, descartando aqueles dados desnecessários, mescla dos bancos de dados, criação de variáveis e categorização de variáveis contínuas. Essa criação e/ou categorização de variáveis possui como critério o aumento da assertividade do modelo, não seguindo necessariamente uma técnica ou referência definida. Por fim é aplicada a função de *Sampling* e balanceamento com o intuito de obter amostras de dados de igual grandeza. Ao final, são listadas 12 variáveis que servirão como base para o modelo preditor. Conforme listadas na tabela 1. Nesta tabela são descritas as doze entradas de acordo com as entradas estabelecidas pelo Grupo Equatorial Energia e uma saída, no intuito de obter o parecer da fiscalização, verificando se esse consumidor é um possível fraudador de energia elétrica.

**Tabela 1. Descrição das variáveis**

Entradas	Saída
Média de consumo anterior à data referência	Descrição do parecer da fiscalização
Continuidade	
Descrição classe consumidora segundo a ANEEL	
Indica se é casa ou apartamento	
Quantidade de vezes onde o consumidor teve alteração de 80% no consumo de energia	
Código localização	
Código tipo fase	
Data da leitura - data de conclusão da fiscalização	
Média de consumo após a data de referência	
Categorização da variável "dia"	
Comparação do consumo atual com o anterior e verifica a ocorrência de um degrau	
Categorização da variável "variação"	

Na figura 4 é possível visualizar a função de *sampling*, sendo realizada no fluxograma, com o objetivo de balancear os dados e obter quantidades iguais de amostras como 50% de dados classificados em irregularidade e 50% como consumidor normal. Esse processo é realizado para a criação de um novo modelo preditivo.



**Fig. 4. Demonstração da função *sampling*.**

Após o processo de balanceamento dos dados, os dados são treinados de forma a aplicar em modelos de árvores de decisão, floresta aleatória e redes neurais artificiais.

A árvore de decisão é um algoritmo não-paramétrico que pode ser utilizado para classificação. Ela tem como objetivo determinar um atributo alvo a partir de atributos previsoires. Esse método é baseado em procura, em que quando um espaço de possíveis hipóteses é percorrido, são sistemas orientados a criar estruturas simbólicas que sejam compreensíveis por humanos.

Segundo Oliveira (2018) a árvore de decisão utiliza a estratégia de dividir para conquistar, onde um problema complexo é decomposto em subproblemas mais simples e recursivamente a mesma estratégia é aplicada a cada subproblema.

As árvores de decisão são compostas por nós, onde ela apresentará um nó raiz como sendo o ponto de partida. Esse nó raiz terá nós filhos, que também gerarão filhos, sucessivamente até chegar ao nó folha. Com isso a árvore de decisão é capaz de armazenar as regras de classificação através de seus nodos, e o nodo final apresenta a decisão a ser tomada (LENZ, 2020).

Ainda, segundo Lenz (2020), sua finalidade é determinar quais campos de informação na base de dados são importantes e se relacionam com o problema.

A floresta aleatória é um algoritmo de aprendizado de máquina que se utiliza de uma série de árvores de decisão sequenciais, ou seja, ao invés de empregar o resultado de somente uma árvore, é aplicado um número maior de árvores de decisão, e de maneira randômica, são selecionadas para um resultado final. A principal ideia é que a combinação de diversas árvores de decisão geradas aleatoriamente deve aumentar a assertividade do modelo como um todo (MARIANO 2021).

A rede neural artificial (RNA) é conceituada como uma das mais modernas e poderosas ferramentas de IA e de processamento de dados da atualidade. É inspirada no contexto do sistema nervoso humano, e dentre todas as características das RNAs, uma das mais importantes é a habilidade de aprender e com isso melhorar o seu desempenho. No ambiente de RNA, aprendizagem ou treinamento corresponde ao processo de ajuste dos parâmetros livres da rede através de um mecanismo de apresentação de estímulos ambientais, conhecidos como padrões.

Para treinamento de rede, este trabalho utilizou o algoritmo de retropropagação de erro, baseado na regra de aprendizagem por correção de erro. Basicamente o algoritmo de aprendizagem por retropropagação de erro consiste de dois passos através de diferentes camadas da rede: um passo para frente, durante o qual todos os pesos são mantidos fixos, e no passo para trás, os pesos sinápticos são todos ajustados de acordo com a regra de correção de erro. (CASTRO, 2016)

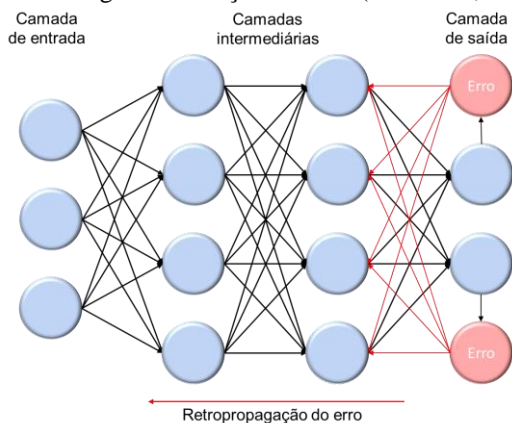


Fig. 5. Exemplificação do funcionamento da rede neural artificial com algoritmo de retropropagação.

### 2.3 Caracterização da ferramenta

Para a implementação da metodologia é utilizado o Software SPSS Modeler para encontrar os furtos de energia elétrica. Este *Software* realiza a mineração de dados e averiguação de texto para a construção de modelos e agrupamentos. Possuindo como base o banco de dados da empresa, são realizados modelos de previsão com a capacidade de diferenciar dados que são muito diferentes do esperado, encontrando assim possíveis furtos.

A análise dos dados é realizada através dos padrões já existentes no banco de dados da respectiva empresa através da mineração de dados.

Essa mineração realiza a combinação de dados armazenados no banco de dados com interações efetuadas com esses dados, sendo realizados por meio de inteligência artificial. O uso desta ferramenta permite que sejam obtidas informações essenciais como: padrões, anomalias e estruturas dentre os dados (Penin, 2008).

No presente trabalho o Software é utilizado com o objetivo realizar a integração e tratamento de diversas bases de dados para a detecção dos pontos fora da curva, ou seja, os possíveis furtos. Desta forma, são criados modelos de previsão, para o caso de detecção de furtos de energia que identificam dados que são muito diferentes dos demais a ponto de gerar suspeitas. Para se obter um modelo ainda mais assertivo na detecção das fraudes de energia, são inseridos dados exógenos como: temperatura, dados socioeconômicos e informações de redes sociais.

### 2.4 Exposição dos parâmetros de análises

O fluxograma do SPSS é formado por quatro banco de dados distintos: PT-FIS, DEGRAU, DADOS\_UC e TELEMED. A

figura 6 demonstra o fluxograma geral composto por este banco de dados.

O propósito do estudo de caso é analisar o nó de mescla e o nó retira no intuito de verificar o comportamento destes nós no fluxograma.



Fig. 6. composição do fluxograma

A primeira análise realizada foi no supernó (Degrau - estrela), onde nesta parte do fluxo já existem diversos filtros. Desta forma a mescla funciona por meio de uma chave escolhida, COD\_UN\_CONS\_WHC, que foi analisada através de uma tabela.

Nessa análise foram gerados 753.492.392 registros que foram lidos, no entanto não foram apresentados na saída pelo grande volume de dados.

A análise seguinte é realizada no nó de derivação denominado de DES\_PAREC\_WPN. Este nó foi analisado através da auditoria de dados e por meio do gráfico de distribuição. Este nó apresenta 100% de campos completos e 100% de registros completos.

Campos completos (%)	Registros completos (%)
100%	100%

Campo	Medição	Valores	Existem	Ação	Imputação omes.	Método	% Cond.	Registros válidos	Valor nulo	Sequência de ca.	Espaço em bran.	Valor em branco
COD_UN_CONS	Contínuo	0	25	Nanum	Nanica	Fixa	100	15197	0	0	0	0
DTA_CONCL_SE	Contínuo	0	0	Nanum	Nanica	Fixa	100	15197	0	0	0	0
COD_PAREC	Categórico	--	--	Nanica	Fixa	100	15197	0	0	0	0	0
DES_PAREC_WPN	Categórico	--	--	Nanica	Fixa	100	15197	0	0	0	0	0

(a)

Campo	Gráfico de Amostra	Medição	Min.	Máx.	Média	Erro Dev.	Assimétrico	Exclusivo	Válido
COD_UN_CONS	[Gráfico]	Contínuo	15952886	1005215719.000	54331921.550	42958595.941	18.070	--	15197
DTA_CONCL_SE	[Gráfico]	Contínuo	04-01-200	15-12-2020 09:00:00	--	--	--	--	15197
COD_PAREC_WPN	[Gráfico]	Categórico	--	--	--	--	--	7	15197
DES_PAREC_WPN	[Gráfico]	Categórico	--	--	--	--	--	7	15197

(b)

Valor	Proporção	%	Contagem
001	9.1	9.1	1383
002	3.24	3.24	493
003	13.31	13.31	2022
005	73.5	73.5	11170
010	0.24	0.24	36
015	0.17	0.17	26
017	0.44	0.44	67

(c)

Fig. 7. (a) Análise por meio de tabela. (b) Análise por meio da auditoria de dados. (c) Análise por meio do gráfico de distribuição.

A análise realizada por meio do gráfico de distribuição, demonstra que no valor de 005 encontra-se o maior percentual armazenado de 73,5%, sendo este um consumidor indicado como regular. Esta informação é de extrema importância para a criação de um modelo com uma base de dados balanceada, para a projeção de um novo modelo preditivo.

Ademais segue a classificação dos demais percentuais apresentados na tabela 2.

Tabela 2. Descrição dos percentuais

003	Avaria
010	Avaria
015	Avaria
005	Normal
017	Irregularidade
001	Irregularidade
002	Irregularidade

Após as análises realizadas no nó de derivação foi analisado o nó de mescla, sendo a primeira mescla presente no fluxograma. Este nó foi analisado por meio da auditoria de dados.

Fig. 8. Análise da mescla por meio da auditoria de dados

Nesta análise foi obtido 6,67% dos campos completos e 25,18% de registros completos. Estes resultados refletem os dados sem registro de fiscalização, demonstrando que apenas um quarto das unidades consumidoras já foi fiscalizada. O próximo nó analisado foi o nó RETIRA, onde foram consideradas duas análises, uma com o  $DTA\_CONCL\_SERV\_WOD > DTA\_REF\_WHC$  e outra considerando o  $@NULL(DTA\_CONCL\_SERV\_WOD)$ .



Fig. 9. Nó retira com as expressões

A primeira análise realizada considerou o  $@NULL(DTA\_CONCL\_SERV\_WOD)$ . Nesta análise os campos  $DTA\_CONCL\_SERV\_WOD$ ,  $COD\_PAREC\_WPN$  e  $DES\_PAREC\_WPN$  estão listados com \$null\$, o que é esperado pois a expressão determina exatamente essa função. Como esta expressão demonstra, um total de 60% de campos completos e 0% de registros completos. Este resultado indica que os filtros aplicados no fluxograma estão corretos, permitindo a passagem apenas dos registros sem fiscalização.

Fig. 10. Análise do nó retira por meio da expressão  $@NULL(DTA\_CONCL\_SERV\_WOD)$ .

Além disso foram realizadas análises com a expressão  $DTA\_CONCL\_SERV\_WOD > DTA\_REF\_WHC$ . Com essa expressão onde havia o \$null\$ nos campos, agora estão preenchidos, demonstrando um total de 80% dos campos completos e 98,71% dos registros completos.

Fig. 11. Análise do nó retira por meio da expressão  $DTA\_CONCL\_SERV\_WOD > DTA\_REF\_WHC$ .

### 2.5 Assertividade do modelo

Para a análise de assertividade do modelo, adotou-se uma amostra de 41250 unidades consumidoras, sendo destas, 39990 inspecionadas sem fraude e 1260 inspecionadas e constatadas como fraude. Dentre todos os modelos estudados, o que apresentou o melhor desempenho global foi a rede neural

artificial (RNA) com 77% de precisão, sendo esta, também, a responsável por demonstrar o melhor resultado na identificação de unidades consumidoras sem fraude de energia, atingindo um percentual de acerto de 78%. Já para a identificação de unidades com fraude, a floresta aleatória obteve um resultado pouco melhor em relação a RNA e mais significativo se comparado à árvore de decisão.

**Tabela 3. Descrição da assertividade**

	Árvore de Decisão		Floresta Aleatória		Rede Neural Artificial	
Precisão (41250 UCs)	64%		75%		77%	
UCs sem fraude (39990 UCs)	64%	25504	75%	30074	78%	31055
UCs com fraude (1260 UCs)	63%	795	69%	870	69%	865

### 3 RESULTADOS E ANÁLISES

Neste artigo, analisou-se as perdas não técnicas por meio do desempenho de um modelo de previsão, no intuito de realizar o reconhecimento de consumidores irregulares, por meio do *software* SPSS.

A análise realizada demonstra o percentual armazenado de 73,5% realizado por meio do gráfico de distribuição, esse valor indica o percentual de consumidores regulares presentes no fluxo analisado. Além disso, por meio deste fluxograma obteve-se o percentual de consumidores considerando diversos métodos em busca de uma boa assertividade. O método com o melhor desempenho é por meio da rede neural artificial, alcançando uma precisão de 77%, constituindo de 69% de possíveis unidades consumidoras com fraude, e 78% de possíveis unidades consumidoras sem fraude de energia elétrica.

No entanto, a assertividade deste modelo ainda pode ser melhorada, no intuito de obter-se uma maior assertividade ao fluxograma. No intuito de melhorar os resultados obtidos, pretende-se realizar o aprimoramento do fluxo, por meio dos resultados obtidos, a criação de um novo modelo preditivo, utilizando dos dados balanceados por meio da função *sampling* e a criação de modelos com árvores de decisão, floresta aleatória e redes neurais artificiais, no objetivo de alcançar resultados ainda melhores.

### 4 CONCLUSÃO

Neste trabalho foi apresentado o método para detecção de perdas não-técnicas em sistemas de distribuição de energia, utilizando za o *software* SPSS Modeler para criar um modelo preditor que tem como entrada dados de fiscalizações, comerciais, consumo e telemedição. A inserção de dados

exógenos ao modelo permite um aumento significativo da assertividade na detecção de furtos de energia elétrica.

### 5 REFERÊNCIAS

ANEEL, Relatório de Perdas. Perdas de Energia Elétrica na Distribuição. Brasília, DF. Disponível em: <<https://www.aneel.gov.br/documents/654800/18766993/Relat%C3%B3rio+Perdas+de+Energia+2019.pdf/6cb0bf36-4074-bbc3-d15d-ed370f44b34b>>. Acesso em: 11 fev. 2022.

Castro, D.G.F.L.N. (2016). *Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações*. Porto Alegre: Editora Saraiva.

Lenz, M., Neumann, F., Santarelli, R. e Salvador, D. (2020). *Fundamentos de Aprendizagem de Máquina*. Porto Alegre: Grupo A.

Mariano, D.C., Marques, L., Silva, M. e al., E. (2021). *Data Mining*. Porto Alegre: Grupo A.

Oliveira, Edson Farias de. (2018). Avaliação da distorção harmônica total de tensão no ponto de acoplamento comum industrial usando o processo KDD baseado em medição. Tese de Doutorado. Universidade Federal do Pará, Belém, 2018.

Penin, Carlos Alexandre de Sousa (2008). *Combate, Prevenção e Otimização das Perdas Comerciais de Energia Elétrica*. Tese (Doutorado), Universidade de São Paulo, São Paulo, 2008.

Piotrowski, Leonardo J., et. al., (2021). Análise das Perdas de Energia no Sistema Elétrico de Distribuição Brasileiro. 13<sup>th</sup> Seminar on Power Electronics and Control (SEPOC 2021).