



CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA



Universidade Federal
de Campina Grande

DIOGO PASSOS MENEZES



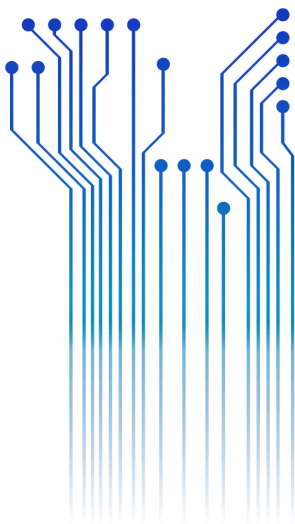
Centro de Engenharia
Elétrica e Informática

TRABALHO DE CONCLUSÃO DE CURSO

**ASSISTENTE BASEADO EM REDES NEURAS ARTIFICIAIS PARA INDICAÇÃO
DE CLIENTES SUSPEITOS DE FRAUDE EM SISTEMA DE DISTRIBUIÇÃO DE
ENERGIA ELÉTRICA DE BAIXA TENSÃO**



Departamento de
Engenharia Elétrica



Campina Grande
2017

DIOGO PASSOS MENEZES

ASSISTENTE BASEADO EM REDES NEURAS ARTIFICIAIS PARA INDICAÇÃO DE CLIENTES
SUSPEITOS DE FRAUDE EM SISTEMA DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DE BAIXA
TENSÃO

*Trabalho de Conclusão de Curso submetido à
Coordenação de Graduação em Engenharia
Elétrica da Universidade Federal de Campina
Grande como parte dos requisitos necessários
para a obtenção do grau de Bacharel em
Ciências no Domínio da Engenharia Elétrica.*

Área de Concentração: Distribuição de Energia

Orientador:

Professor André Dantas Germano, D. Sc.

Campina Grande
2017

DIOGO PASSOS MENEZES

ASSISTENTE BASEADO EM REDES NEURAIIS ARTIFICIAIS PARA INDICAÇÃO DE CLIENTES
SUSPEITOS DE FRAUDE EM SISTEMA DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DE BAIXA
TENSÃO

*Trabalho de Conclusão de Curso submetido à
Coordenação de Graduação em Engenharia
Elétrica da Universidade Federal de Campina
Grande como parte dos requisitos necessários
para a obtenção do grau de Bacharel em
Ciências no Domínio da Engenharia Elétrica.*

Área de Concentração: Distribuição de Energia

Aprovado em 27 / 04 / 2017

Professor George Rossany Soares de Lira, D. Sc.
Universidade Federal de Campina Grande
Avaliador, UFCG

Professor André Dantas Germano, D. Sc.
Universidade Federal de Campina Grande
Orientador, UFCG

Dedico este trabalho à minha esposa, pais e irmã.

AGRADECIMENTOS

Agradeço a Deus, soberano, verdadeiro, justo e misericordioso, que tudo faz como Lhe apraz. A Ele toda a glória, em todo tempo e em todo lugar.

Agradeço à Nathalie Menezes, minha esposa fiel, cuidadosa e dedicada. Por sua companhia e suporte diários, compartilhando alegrias e tristezas.

Agradeço a Welington e Edilene Menezes, meus pais e primeiros educadores. Pelo cuidado sem medidas de sempre, independentemente do lugar do mundo onde eu esteja.

Agradeço a Ísis e Manuel Albuquerque, irmã e cunhado, sempre dispostos a ajudar. Por tudo e mais um pouco que vocês fazem por mim.

Agradeço às famílias Arcenio, Barbosa, Nole e Wood, com as quais me sinto em casa. Pelas portas abertas.

Agradeço a André Oliveira, Áquila Nole, Diogo Costa, Edmilson Oliveira, Eli Dias, Heber Passos, Kelyson Nunes, Nilton Barbosa, Pedro Freitas e Wesley Oliveira, amigos e irmãos com quem posso contar em qualquer que seja o momento. Agradeço, por extensão, a suas noivas e esposas.

Agradeço aos meus companheiros de curso, especialmente Arthur Araújo, Gabriel Rodrigues, Nayara Aguiar, Plateny Ponchet, Thiago Leite e Vítor Silveira, sem os quais esse curso seria infindável. Pelas noites viradas e estudos produtivos (ou não).

Agradeço aos professores e servidores do DEE, em especial Damásio Fernandes, Tarso Vilela, Washington Neves e Tchaikowsky Oliveira. Pelos conhecimentos transmitidos e pelo muito que me ajudaram nessa reta final.

*“In Christ alone, my hope is found
He is my light, my strength, my song
This Cornerstone, this solid ground
[...] Here in the love of Christ I stand”*

Keith Getty e Stuart Townend.

RESUMO

A busca por pontos de medição com irregularidade em um sistema de distribuição de energia se mostra uma tarefa custosa quando a mesma é realizada por meio de um processo não automatizado. Gastam-se tempo e esforço físico-mental dos envolvidos, gastos estes que são intensificados para o caso de clientes atendidos em baixa tensão, onde a coleta do consumo de energia é feita em campo e não há acompanhamento de outras grandezas elétricas por meio de telemetria. Este trabalho, então, implementa um assistente baseado em redes neurais artificiais para auxílio na detecção de irregularidade (potenciais fraudes) na medição de energia mensal. Com base em seus consumos históricos, clientes com e sem fraude foram classificados corretamente a uma taxa de até 83,80% de acerto, por meio de uma rede neural artificial de *perceptrons* de múltiplas camadas treinada pelo método *resilient backpropagation*. Os resultados apresentados nesse trabalho podem ser utilizados para abordar o combate a perdas não-técnicas devido a fraudes, com RNA em sistemas de distribuição de energia elétrica.

Palavras-chave: Fraude, Irregularidade, Perda Não-Técnica, Distribuição, RNA, MLP.

ABSTRACT

The search for measurement points with irregularities in energy distribution systems has shown itself to be a costly task when accomplished through non-automated processes, which require great amounts of time and physico-mental effort of those involved. These costs are increased for low-voltage services, where the collection of energy consumption readings are field-based and there is no monitoring of several other electrical quantities through telemetry. This work, then, implements an assistant based on artificial neural networks theory for detecting irregularities (potential frauds) of monthly energy measurements. With basis in historical consumption, fraudulent and non-fraudulent clients are classified with a success rate of up to 83.80%, utilizing a multilayer perceptron artificial neural network trained with the resilient backpropagation algorithm. The outcomes presented in this work can be further utilized to approach the fighting of non-technical losses caused by fraudulent means, with neural networks in electrical energy distribution systems.

Keywords: Fraud, Irregularity, Non-Technical Losses, Distribution Systems, ANN, MLP.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo de neurônio artificial.....	20
Figura 2 – Descarte de consumo após detecção de fraude	24
Figura 3 – Clientes com consumos considerados regulares	25
Figura 4 – Ilustração das configurações utilizadas.....	27
Figura 5 – Desempenho em função do tamanho da saída	34
Figura 6 – Desempenho em função da configuração de entrada.....	35
Figura 7 – Desempenho em função do número de anos de consumo.....	35
Figura 8 – Desempenho em função da dimensão da camada oculta	35

LISTA DE TABELAS

Tabela 1 – Dez piores desempenhos e suas respectivas configurações.....	31
Tabela 2 – Dez melhores desempenhos e suas respectivas configurações.....	32
Tabela 3 – Desempenho por parâmetro variado.....	33
Tabela 4 – Grau de influência de cada parâmetro variado	36
Tabela 5 – Cinco melhores desempenhos para 2 anos de consumo	37
Tabela 6 – Cinco melhores desempenhos para 3 anos de consumo	37

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
BD	Base de Dados
DHCNOC	Valores de Derivada de Histórico de Consumo Normalizado em Ordem Crescente
HCN	Histórico de Consumo Normalizado
MLP	<i>Multilayer Perceptron</i>
RNA	Redes Neurais Artificiais

LISTA DE SÍMBOLOS

φ_j^k	Função de ativação do neurônio artificial
b_j^k	Polarização do neurônio artificial
i	Índice de sinal de entrada de um neurônio artificial
j	Índice de neurônio artificial em uma camada
k	Índice de camada de uma rede neural artificial
v_j^k	Potencial de ativação do neurônio artificial
w_{ji}^k	Peso sináptico do neurônio artificial
x_i^k	Entrada do neurônio artificial
y_j^k	Saída do neurônio artificial

SUMÁRIO

1	Introdução	14
1.1	Motivação	14
1.2	Objetivo	15
1.3	Estrutura do Texto	15
2	Embasamento Teórico	16
2.1	Perdas Não-Técnicas de Energia	16
2.2	Redes Neurais Artificiais	18
2.2.1	Elementos Neurais.....	19
2.2.2	Aprendizagem de uma RNA tipo MLP por Retropropagação	20
3	Metodologia	22
3.1	Adequação da Base de Dados Empregada.....	23
3.2	Atributos Extraídos	25
3.3	Parâmetros Variados na Análise de Sensibilidade.....	27
4	Resultados.....	30
4.1	Configurações com Menor Desempenho.....	30
4.2	Configurações com Maior Desempenho.....	32
4.3	Análise Individual dos Parâmetros Variados.....	33
4.4	Análise de Assistente para Distribuidora.....	36
5	Conclusão.....	38
	Referências	39
	APÊNDICE A – Tabela de Desempenhos	40

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

Geração, transmissão e distribuição. Essas são as etapas do sistema elétrico pelas quais percorre a energia elétrica comercializada. Nesse processo, uma parcela de energia é perdida e pode ser calculada pela diferença entre a energia que sai do ponto de geração e a energia contabilizada entregue aos consumidores finais. Tal perda é subdividida em perdas na rede básica (oriundas do processo de transmissão) e perdas de distribuição. Estas são ainda segregadas em perdas técnicas e perdas não-técnicas. As perdas técnicas englobam perdas físicas inerentes ao sistema de distribuição, como perdas por aquecimento dos fios condutores, onde há conversão de energia elétrica, que está sendo transmitida, em energia térmica. Já as perdas não-técnicas, também conhecidas como perdas comerciais, estão associadas, por exemplo, a perdas por erro de medição, erro de faturamento da distribuidora de energia e furtos de energia.

Com o propósito de reduzir as perdas não-técnicas, distribuidoras contam com departamentos comerciais de combate a perdas. De forma geral, maiores gastos são destinados à supervisão de clientes grandes e, assim, é comum que várias grandezas elétricas (tensões, correntes, fatores de potência, potência demandada, consumo de energia, etc.) sejam acompanhadas em tempo real por telemetria. Por outro lado, clientes menores – maior parte dos clientes atendidos em baixa tensão – geralmente dispõem de equipamentos de medição mais simples e sem telemetria; assim, auxiliares comerciais são necessários para que a coleta do consumo de energia de clientes seja feita (mensal ou plurimensalmente). Percebe-se, então, que, para a maioria dos clientes atendidos em baixa tensão, a identificação de fraude (especificamente) é dificultada pela limitação de informações disponíveis quando em comparação com os clientes que dispõem de telemetria. Além disso, quanto menos automatizado o processo analítico desses clientes, maiores o tempo e o esforço físico-mental gastos para realizar a análise. Logo, para esse processo, o auxílio de inteligência artificial pode ser tão bem-vindo quanto benéfico. Dependendo da sofisticação dessa inteligência, por exemplo, casos de fraude não-aparentes para um analista podem ser apontados.

1.2 OBJETIVO

Neste trabalho, busca-se implementar um assistente para analistas de distribuidoras responsáveis por indicar suspeitas de fraude em clientes atendidos em baixa tensão. As indicações feitas pelo assistente devem ser (a) aceitas ou (b) revisadas (ainda assim há ganho, uma vez que o assistente reduz o universo de clientes para análise), ficando a critério da distribuidora atendida. Para tanto, algumas configurações de rede neural artificial (RNA) serão testadas e empregadas.

1.3 ESTRUTURA DO TEXTO

No Capítulo 2 é realizado um embasamento teórico acerca de perdas não-técnicas de energia e redes neurais artificiais. Em seguida, no Capítulo 3, a metodologia utilizada é apresentada, em termos material e métodos. Os resultados obtidos são discutidos no Capítulo 4, onde sugestões de aplicação também são feitas. Por fim, as conclusões do trabalho e considerações finais são apresentadas no Capítulo 5.

2 EMBASAMENTO TEÓRICO

No decorrer deste Capítulo, o conceito de perdas não-técnicas de energia, suas formas de combate e benefícios de sua redução são apresentados. Em seguida, redes neurais artificiais são apresentadas e definidas e seus princípios de funcionamento são descritos. Objetiva-se aqui, então, a explanação dos principais temas relacionados ao trabalho desenvolvido.

2.1 PERDAS NÃO-TÉCNICAS DE ENERGIA

Perdas podem ser segregadas quanto à natureza e origem. Quanto à natureza, as perdas podem ser classificadas em perdas de demanda e de energia. Perdas de demanda são a diferença entre a demanda de potência de entrada e de potência de saída de um sistema. Semelhantemente, as perdas de energia são a diferença entre a energia de entrada e de saída de um sistema. Quanto à origem, as perdas podem ser classificadas em perdas técnicas e não-técnicas. Perdas técnicas são a parcela de energia ou demanda perdida no processo de transporte (em linhas de transmissão, alimentadores, etc.) e de transformação (em transformadores e elementos de medição) de demanda ou energia. Por outro lado, perdas não-técnicas são a parcela de energia ou demanda que é efetivamente entregue ao cliente final, mas que não é de fato paga por ele (MÉFFE, 2001).

Segundo ALMEIDA, OLIVEIRA e DANTAS (2006), as cinco grandes causas das perdas-não técnicas de energia, também conhecidas como perdas comerciais, são:

- i. Ligações clandestinas: ligações feitas por clientes que nunca foram regulares, isto é, nunca fizeram parte do cadastro da concessionária;
- ii. Consumidores auto religados: ligações feitas por clientes regulares após ação de suspensão do fornecimento de energia. Certos casos têm a inadimplência como início de um ciclo fraudulento, onde a inadimplência gera a suspensão do serviço, que leva consumidores a adotar medidas fraudulentas, produzindo novos graus de inadimplência (RAMOS, 2014);
- iii. Consumidores cadastrados e regularmente ligados ao sistema: fraudes ou defeitos na medição e erros de faturamento. Os tipos de irregularidades

mais encontrados nos processos analisados pela ANEEL incluem: ligação direta à rede secundária, desvio no ramal de entrada, ligações do medidor invertidas, curto-circuito na entrada ou saída do medidor, terminal de prova aberto, entre outras (RAMOS, 2014);

- iv. Iluminação pública: cadastro desatualizado e erros na estimativa de faturamento;
- v. Consumidores ligados sem medição.

Dadas as diferentes origens de perdas não-técnicas de energia, é natural que variadas formas de combate sejam aplicadas pelas distribuidoras de energia elétrica. Para tal, segundo ARAUJO e SIQUEIRA (2006), é comum que as distribuidoras realizem os seguintes procedimentos:

- i. Identificação de áreas críticas: a identificação clara das áreas com maior incidência de perdas não-técnicas de energia dentro da área de concessão da distribuidora é essencial para a eficácia e a efetividade de suas ações de combate ao problema;
- ii. Balanço energético: o balanço energético aqui referido é o cálculo da diferença entre a energia medida pelos registradores instalados junto aos transformadores de distribuição (medições totalizadoras) e a energia medida dos consumidores conectados aos referidos transformadores;
- iii. Sistemas de faturamento: uma importante ferramenta no combate às perdas não-técnicas de energia no âmbito das distribuidoras é a inserção, nos seus sistemas de faturamento, de ferramentas que possibilitem a obtenção e a gestão de informações precisas referentes a variações acentuadas no consumo de energia de unidades consumidoras;
- iv. Ações de marketing institucional: as distribuidoras vêm lançando mão de ações de marketing institucional, via de regra com o desenvolvimento de campanhas educativas junto a comunidades carentes, em cujo âmbito são prestadas informações sobre a adequada e eficiente utilização da energia elétrica;
- v. Motivação dos colaboradores: para um efetivo combate às perdas não-técnicas, é imprescindível o engajamento de todos os funcionários da empresa;

- vi. Desenvolvimento e utilização de novas tecnologias: várias tecnologias têm sido desenvolvidas e implementadas na busca de soluções mais eficazes para o combate às perdas de energia elétrica, merecendo destaque, entre outras, a utilização de medição externa e de medidores eletrônicos, a blindagem de cabos e o desenvolvimento de novos tipos de medidores e de softwares que empreguem inteligência artificial para aumentar a eficácia das inspeções.

Para os casos em que as medidas de combate às perdas não-técnicas resultam em detecção e regularização de irregularidades na medição, o consumo de energia não contabilizado pode ser recuperado pela distribuidora parcial ou totalmente. São definidos, então, os conceitos de energia recuperada e energia agregada, que possibilitam avaliar o impacto causado por essas medidas na redução das perdas. A energia recuperada se refere à parcela de energia consumida, antes da regularização, que não foi faturada. Já a energia agregada se refere à parcela de energia que passa a ser contabilizada corretamente, após a regularização, em relação ao perfil de consumo anterior à regularização (ALMEIDA, OLIVEIRA e DANTAS, 2006).

Por fim, é válido mencionar que a redução de perdas não-técnicas de energia beneficia tanto às distribuidoras quando seus consumidores. Do ponto de vista das empresas, por exemplo, há redução da parcela de energia não faturada, aumentando sua receita. Já do ponto de vista do consumidor, há uma redução tarifária, já que uma parte das perdas não-técnicas são repassadas para o consumidor em forma de tarifa de energia (ANEEL, 2016).

2.2 REDES NEURAIS ARTIFICIAIS

O cérebro humano, diferentemente do computador digital convencional, opera de uma forma inteiramente singular, funcionando como um sistema de processamento de informações altamente complexo, não-linear e de operação paralela. Ele tem a capacidade de organizar suas estruturas constituintes, conhecidas como neurônios, para realização de certos processamentos (reconhecimento de padrões, percepção e controle motor) de forma rápida e eficiente. Como exemplo, tem-se que o reconhecimento de um rosto familiar, em uma cena não familiar, é feito em cerca de 100-200 ms (HAYKIN, 2008).

Na tentativa de se modelar o processamento de uma rede neural humana, a fim de se extrair suas melhores características, iniciam-se os estudos na área de RNA. Alguns de seus marcos históricos são indicados a seguir. Em 1940, McCulloch e Pitts afirmam que redes de neurônios artificiais são capazes de executar qualquer função aritmética ou lógica. Em 1958, Rosembat desenvolve a rede neural *perceptron* para reconhecimento de padrões. Em 1969, Minsky e Papert demonstram as limitações de um *perceptron* simples, apontando que o mesmo não é capaz de resolver o problema lógico do “ou-exclusivo”. Tal descoberta promove um hiato nas pesquisas relacionadas a RNA, que somente são retomadas na década de 1980, dado o desenvolvimento de ferramentas computacionais e o estudo de diversos autores, como Hopfield e Rumelhart e MacLelland (FERREIRA, 2011).

2.2.1 ELEMENTOS NEURAIIS

Atualmente, define-se uma RNA como sendo um processador distribuído massivamente paralelo composto por unidades elementares de processamento e que tem a propensão a armazenar conhecimento, adquirido por meio de experiência, tornando-o disponível para uso (HAYKIN, 2008). Sua semelhança com o cérebro humano se evidencia pelos seguintes fatores:

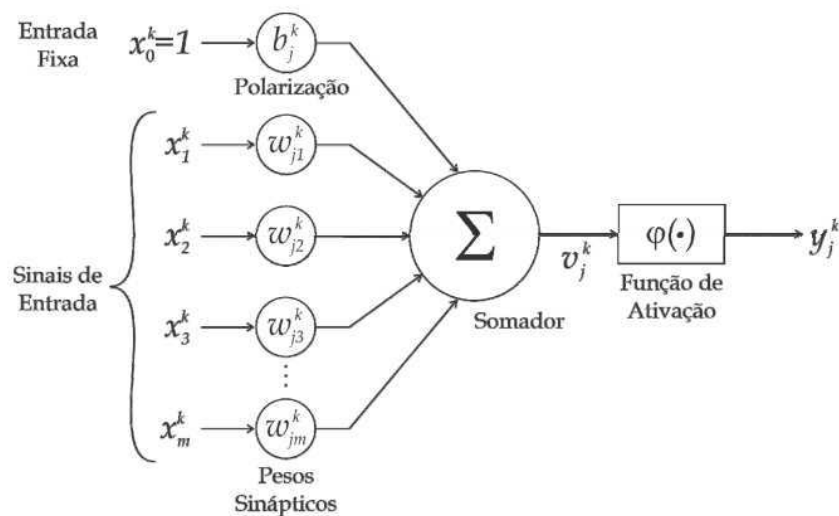
- i. O conhecimento é adquirido a partir do ambiente por meio de um processo de aprendizagem (treinamento);
- ii. O conhecimento adquirido é armazenado nas conexões interneurais, conhecidas como pesos sinápticos.

O modelo da unidade elementar de processamento, o neurônio artificial, é mostrado na Figura 1. Três elementos básicos podem, então, ser identificados (FERREIRA, 2011):

- i. Um conjunto de sinapses ou conexões interneurais, caracterizadas individualmente por um peso sináptico w_{ji}^k , que indica a importância do sinal de entrada i , x_i^k , na determinação da saída do neurônio j da camada k ;

- ii. Um somador, que tem a função de realizar uma combinação linear dos sinais de entrada, ponderados pelos respectivos pesos sinápticos de suas ligações;
- iii. Uma função de ativação, φ_j^k , que limita a amplitude da saída num intervalo normalizado, geralmente $[0,1]$ ou $[-1,1]$. Necessária para que se limite numericamente as informações que se propagam na RNA, buscando evitar a saturação dos neurônios e perda de eficiência da rede.

Figura 1 – Modelo de neurônio artificial



Fonte: (FERREIRA, 2011).

A polarização do neurônio, b_j^k , tem por função deslocar adequadamente o argumento da função de ativação, v_j^k , denominado de potencial de ativação. Com isso, a saída do neurônio, y_j^k , é dada matematicamente por:

$$y_j^k = \varphi_j^k(v_j^k) = \varphi_j^k\left(b_j^k + \sum_i w_{ji}^k x_i^k\right) \quad (1)$$

2.2.2 APRENDIZAGEM DE UMA RNA TIPO MLP POR RETROPROPAGAÇÃO

Após a definição estrutural de uma RNA, segue-se o seu processo de aprendizagem. Primeiramente, para que uma RNA adquira conhecimento, uma base de dados (BD) é necessária. Para o caso de uma RNA destinada à classificação de padrões, uma BD composta por amostras de cada classe do elemento a ser classificado é requerida.

Em seguida (considerando o treinamento por um método supervisionado), tais amostras são dispostas em entradas e saídas desejadas (alvos), onde aquelas são compostas por informações com as quais a RNA será alimentada e estas constituem uma identificação da classe à qual pertence a amostra. Após a definição estrutural da RNA (funções de ativação, número de camadas, tamanho das entradas e dos alvos, etc.), a BD amostral é dividida aleatoriamente em três grupos destinados a etapas distintas, a saber:

- i. Treinamento: etapa de aprendizagem da RNA. Inicialmente, os pesos sinápticos são definidos aleatoriamente, indicando a falta de conhecimento da rede. Posteriormente, as amostras de treinamento são apresentadas sucessivamente à RNA, que, para cada iteração (chamada época), calcula os valores de saída com base na Equação 1. Ao fim de cada época, os alvos são comparados com os valores de saída e os pesos sinápticos são alterados com vistas à minimização da diferença entre alvos e saídas;
- ii. Validação: etapa de supervisão da aprendizagem da RNA. A validação é executada paralelamente ao treinamento. Ao fim de cada época, as amostras de validação são apresentadas à RNA com o objetivo de verificar se o conhecimento adquirido, até então, pode ser generalizado e utilizado para classificar amostras que não foram utilizados na etapa de treinamento;
- iii. Teste: etapa de verificação da aprendizagem. O teste é similar à validação do ponto de vista de que ela serve para avaliar o nível de conhecimento adquirido e de tomada de decisão da RNA. Entretanto, o teste é realizado somente ao final da aprendizagem, quando não mais os pesos sofrem alteração.

As etapas apresentadas caracterizam a aprendizagem de uma RNA supervisionada, já que os alvos são conhecidos e utilizados no treinamento. Caso não se tenha conhecimento das classes às quais pertencem as amostras ou se deseje deixar a criação de classes a critério da RNA, métodos de aprendizagem não-supervisionados podem ser empregados.

3 METODOLOGIA

No decorrer deste Capítulo serão apresentados o material e os métodos escolhidos e utilizados para a implementação do assistente proposto. Em termos de material, coletaram-se amostras de consumo histórico de clientes de uma distribuidora de energia elétrica, as quais definirão a base de dados empregada na pesquisa. Em termos de método, o MATLAB foi utilizado e várias RNA do tipo *Multilayer Perceptron* (MLP) com diferentes configurações foram executadas a fim de se realizar uma análise de sensibilidade dos parâmetros e de se levantar informações sobre desempenho. Em seguida, considerações e escolhas das melhores configurações de RNA foram feitas.

De forma sucinta, a metodologia empregada para atingirem-se os objetivos propostos foram:

- i. Adequação da base de dados e definição do espaço amostral;
- ii. Carregamento da base de dados a partir de uma planilha eletrônica (histórico de consumo e indicação de presença ou não de irregularidade);
- iii. Pré-processamento dos dados de entrada da RNA (normalização do histórico consumo);
- iv. Extração de atributos a partir do histórico de consumo normalizado – visando ressaltar características particulares de cada uma das classes –, agrupamento dos atributos e definição de entradas e saídas;
- v. Procedimento do treinamento, validação e teste de RNA, empregando o Toolbox de RNA do MATLAB;
- vi. Cálculo de parâmetros de desempenho.

Os itens (iii) a (vi) foram executados diversas vezes, pois para cada variação nos parâmetros da extração de atributos e da RNA se fazia necessária nova bateria de testes. Este procedimento caracterizou uma análise de sensibilidade realizada, a qual será mais detalhadamente discutida na Seção 3.3.

Nas Seções 3.1 e 3.2 serão detalhados os itens (i) e (iv), respectivamente. Tais etapas merecem maior atenção por serem muito particulares dentro de cada pesquisa envolvendo técnicas de reconhecimento de padrões. Os demais itens são procedimentos

recorrentes neste tipo de pesquisa, e por isso não serão detalhados neste trabalho. Para maiores informações dentro destes procedimentos, recomenda-se a leitura de (HAYKIN, 2008) e (FAUSETT, 1994).

3.1 ADEQUAÇÃO DA BASE DE DADOS EMPREGADA

Os dados históricos de consumo de clientes foram coletados diretamente em uma distribuidora de energia da região nordeste do Brasil. Como empregaram-se no trabalho RNA treinadas por algoritmos supervisionados, foi necessário que os dados coletados fossem segregados de acordo com as classes a serem consideradas: clientes com e sem suspeita de fraude. Assim, a coleta pode ser dividida em três partes:

- i. Levantamento de clientes com registro de ocorrência de fraude;
- ii. Levantamento de clientes sem registro de ocorrência de fraude, e;
- iii. Consulta do histórico de consumo dos clientes levantados.

É importante perceber que, na etapa (i), a ocorrência de fraude deve ser considerada somente no período em que o consumo é disponibilizado, como explicado e exemplificado nas próximas páginas.

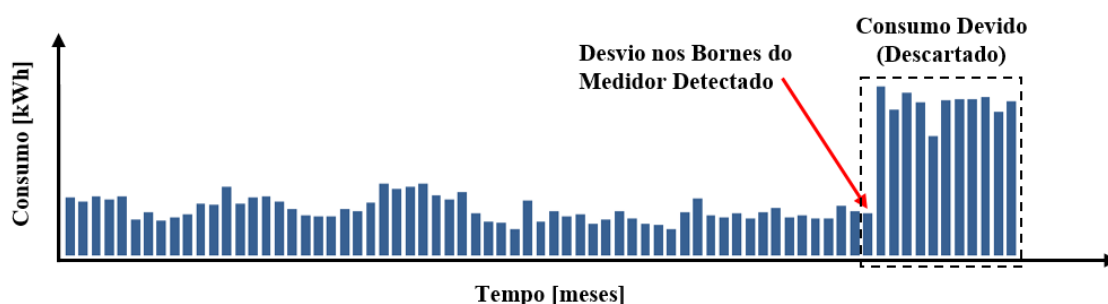
Em se tratando do levantamento de histórico de consumo, o limitante do número de meses disponíveis para o estudo foi definido pelo máximo de meses que a ferramenta de consulta, disponibilizada pela distribuidora, conseguia retornar. Tal limitante se deu em 73 meses, uma vez que a consulta, feita em fevereiro de 2017, retornou consumos de janeiro de 2011 a janeiro de 2017.

No caso dos clientes com fraude, nem todo o consumo histórico deve ser levado em consideração, dado que para a maioria há consumo antes da detecção da fraude (deve ser considerado) e depois da detecção da fraude (deve ser descartado), como representado na Figura 2. Este descarte de meses posteriores à detecção da inconformidade se dá pelo fato de que o assistente será aplicado para indicar a suspeita de fraude antes da descoberta da mesma. Logo, o consumo a partir do mês da descoberta da fraude não pode ser considerado, já que, inclusive, pode representar o consumo devido.

Assim, caso se deseje escolher 72 meses como janelamento para alimentação das RNA, a implicação direta é de que só podem ser escolhidos clientes que apresentaram

fraude em janeiro de 2017, já que fraudes em meses anteriores levam ao descarte de mais de um mês. Isto reduz bastante o número amostral para a classe de clientes fraudadores, tornando a BD muito pequena. Assim, definiu-se limitar o janelamento a no máximo 60 meses, utilizando-se os clientes com irregularidades identificadas entre janeiro e dezembro de 2016 (implicando na consideração de uma janela de consumo de janeiro de 2011 a dezembro de 2015 até de dezembro de 2011 a novembro de 2016). Esta escolha resultou em um grupo de 1666 amostras de clientes fraudadores válidos, dos quais 900 foram aleatoriamente escolhidos para compor a BD empregada.

Figura 2 – Descarte de consumo após detecção de fraude



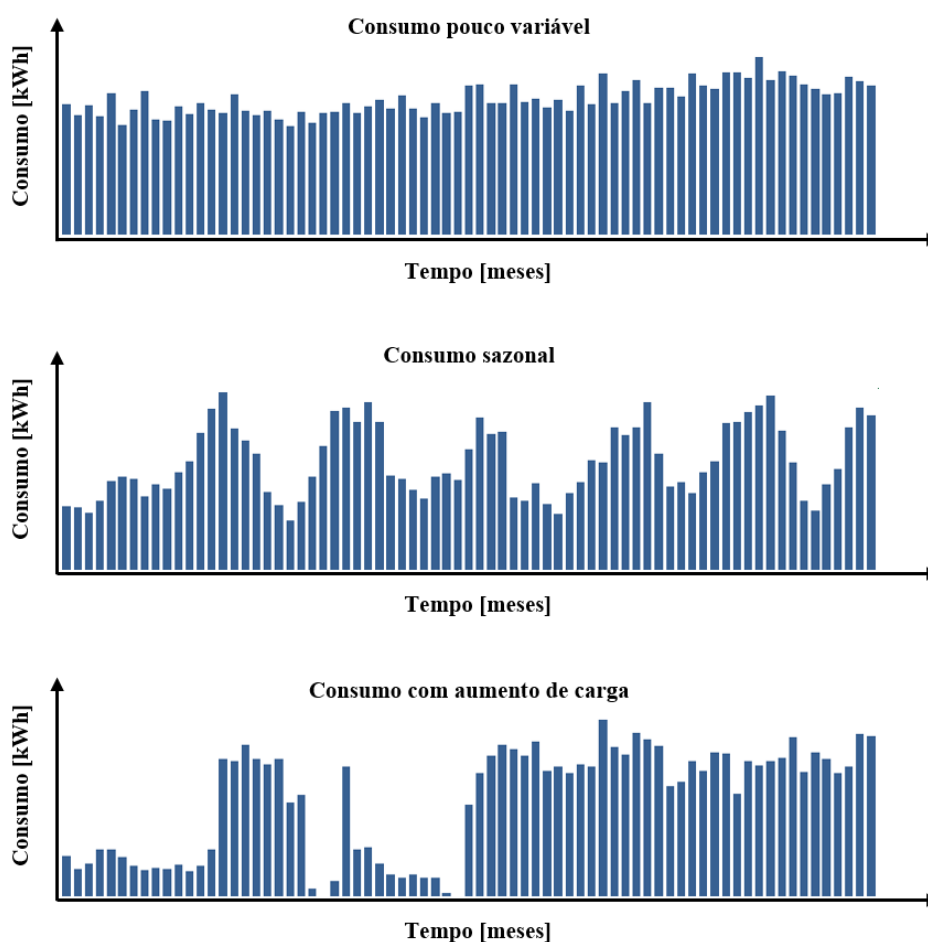
Fonte: o próprio autor.

Em seguida, para o levantamento de clientes regulares, a premissa de “não ocorrência de irregularidade de 2011 a 2017” foi considerada. É importante perceber que essa premissa não garante que os clientes levantados não apresentam fraude; ela apenas garante que esses clientes nunca foram inspecionados ou que, em caso de ocorrência de inspeção, não foi detectada uma fraude. Assim, após levantamento, 5000 clientes foram escolhidos aleatoriamente; desses, 900 foram selecionados visualmente por um especialista a partir de seu consumo histórico. Alguns exemplos são indicados na Figura 3, onde são representados consumos com (a) pouca variação relativa em torno da média, (b) sazonalidade e (c) aumento de carga. Isso permitiu definir para a base de dados que alimenta as RNA o que é considerado consumo histórico regular do ponto de vista de um analista humano. A expectativa é de que os casos que fugissem a este comportamento sensibilizassem a RNA no sentido de auxiliar na tomada de decisão.

Uma vez definida a parte da base de dados constituída de clientes não fraudadores, janelas de consumo de no máximo 60 meses dentro desta BD foram definidas. A localização de cada uma destas janelas dentro dos 72 meses constituintes da BD foi realizada de forma aleatória. Este procedimento foi necessário para garantir que nas duas

metades da BD houvesse um componente aleatório do mês de início do registro dos dados, uma vez que a data de detecção de fraude na classe de clientes com irregularidades é um parâmetro aleatório. Além disso, o janelamento aleatório do grupo amostral de clientes sem irregularidades remove tendências de comportamentos sazonais coletivos dentro do grupo (como por exemplo maior consumo de energia nos meses mais quentes). Caso estas tendências não sejam removidas, é provável que a RNA detecte a sazonalidade (ou não) do consumo em cada classe, em vez de detectar efetivamente os vícios de consumo que denunciam a potencial fraude.

Figura 3 – Clientes com consumos considerados regulares



Fonte: o próprio autor.

3.2 ATRIBUTOS EXTRAÍDOS

Em se tratando de regularidade de consumo, é esperado, de forma geral, que o consumo regular de clientes atendidos em baixa tensão não apresente muitas variações

entre meses. Por outro lado, é esperado que o consumo irregular para clientes de porte semelhante apresente fortes oscilações entre meses (que podem ser causadas, por exemplo, por manipulações da medição do consumo) ou quedas bruscas de consumo (que podem ser causadas, por exemplo, por introdução de uma fraude permanente na medição).

Com essas hipóteses em mente, decidiu-se realizar a extração da primeira derivada do sinal de consumo (já normalizado) para a indicação do seu grau de variação. Uma vez que não existe um momento específico para a ocorrência de irregularidade, resolveu-se ordenar de forma crescente os valores obtidos para a derivada de cada cliente. Assim, nos casos em que esses novos atributos foram utilizados, buscou-se passar para a RNA conhecimento sobre a existência ou não de variações no consumo e suas intensidades.

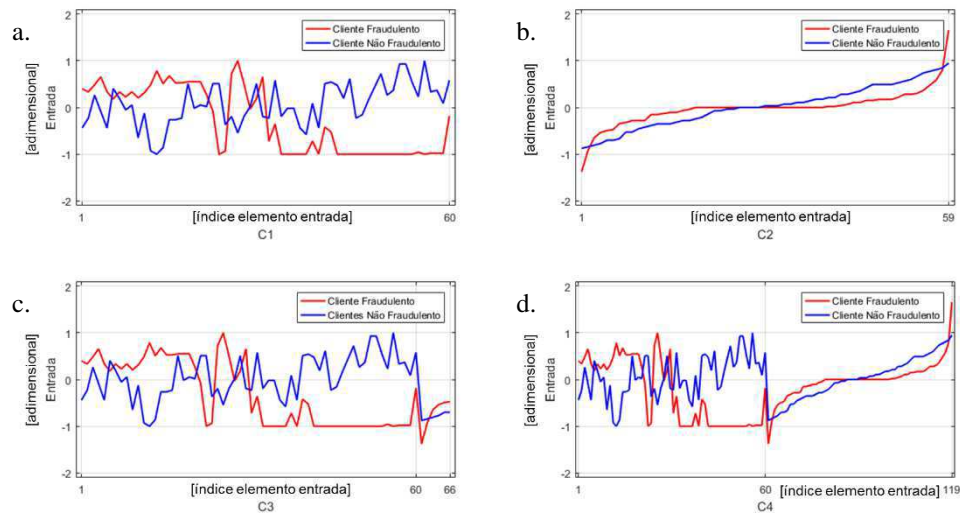
Com o intuito, então, de medir a contribuição (positiva ou negativa) desses novos atributos, quatro configurações de entrada para as RNA (Cn) foram exploradas neste trabalho. A saber:

- C1. **Histórico de consumo normalizado (HCN)**: esta configuração visa propor medidas para analisar o desempenho das RNA quando alimentadas somente com o consumo de cada cliente;
- C2. **Valores da derivada do histórico de consumo normalizado em ordem crescente (DHCNOC)**: esta configuração visa propor medidas para analisar o desempenho das RNA quando alimentadas somente com a distribuição dos valores da derivada do consumo;
- C3. **HCN concatenado com os seis primeiros valores de DHCNOC**: esta configuração visa propor medidas para analisar o desempenho das RNA quando alimentadas com o consumo de cada cliente e 6 representantes de seus valores de derivada de consumo mais negativos;
- C4. **HCN concatenado com DHCNOC**: esta configuração visa propor medidas para analisar o desempenho das RNA quando alimentadas com o consumo de cada cliente e seus valores de derivada do consumo (na íntegra).

Para ilustrar as configurações utilizadas, as entradas de um cliente fraudulento (em vermelho) e de um cliente não fraudulento (em azul) são mostradas na Figura 4. Mais especificamente, as configurações C1, C2, C3 e C4 são representadas, respectivamente, nas Figuras 4a, 4b, 4c e 4d. Percebe-se, então, que, para as configurações C1 e C3, a

queda de consumo é evidenciada para um cliente fraudulento, enquanto um cliente regular apresenta uma variação em torno da média. Para as configurações com DHCNOC, em particular C2 e C4, evidencia-se que o DHCNOC de um cliente regular se apresenta aproximadamente linear, diferentemente do que se mostra para um cliente fraudulento.

Figura 4 – Ilustração das configurações de entrada utilizadas



Fonte: o próprio autor.

3.3 PARÂMETROS VARIADOS NA ANÁLISE DE SENSIBILIDADE

O passo-a-passo metodológico previamente apresentado faz rápida referência à análise de sensibilidade baseada na simulação de várias configurações da RNA. Tais configurações foram obtidas variando certos parâmetros da RNA, enquanto outros foram mantidos fixos. Os parâmetros mantidos fixos durante a análise de sensibilidade foram:

- F1. **Topologia:** MLP de 2 camadas (entrada, camada oculta e saída). A utilização de apenas uma camada oculta tem sido solução suficiente para grande maioria dos problemas resolvidos com RNA (SILVA e OLIVEIRA, 2006);
- F2. **Funções de ativação:** tangente hiperbólica. Essa função apresenta excelentes propriedades de treinamento de RNA, mostradas por vários problemas *benchmark* (KALMAN e KWASNY, 1992);

- F3. **Algoritmo de treinamento:** *resilient backpropagation*. Esse algoritmo foi escolhido por sua alta performance em termos de velocidade e minimização de erro quadrático (ADEOTI e OSANAIYE, 2013);
- F4. **Condições de parada de treinamento:** valores padrão associados ao algoritmo *resilient backpropagation* do *toolbox* de redes neurais artificiais do MATLAB, uma vez que são valores tipicamente válidos para a maioria das aplicações.

Os parâmetros variados durante a execução são elencados abaixo:

- V1. **Número de anos de consumo considerado:** 1 a 5 anos (passo de 1 ano). Uma vez que se dispõe de 60 meses de consumo para cada cliente, para números menores de anos o consumo mais recente é considerado (para 2 anos, por exemplo, apenas os 24 meses mais recentes da base de dados são considerados). Espera-se, entretanto, que melhores resultados sejam obtidos para um maior número de anos, uma vez que mais informações sobre o consumo dos clientes são disponibilizadas;
- V2. **Número de neurônios na camada oculta:** variados entre 0,25 a 2 vezes o número de elementos do vetor de entrada, com passo de 0,25. Neste trabalho, este número de vezes é denominado “dimensão da camada oculta” e o número de neurônios calculado deve ser arredondado para o menor inteiro. A escolha de seus limites foi feita de forma *ad hoc*, a fim de restringir o número de configurações de RNA deste trabalho;
- V3. **Configurações da entrada:** quatro configurações de entrada (C1-C4), conforme mostrado na Seção 3.2;
- V4. **Número de elementos da saída:** saídas com um ou dois elementos são utilizadas. Para saídas com um elemento, o valor -1 indica um cliente sem suspeita de fraude e o valor 1 indica um cliente com suspeita de fraude. Para saídas com dois elementos, o vetor $[-1 \ 1]^T$ indica um cliente sem suspeita de fraude e o vetor $[1 \ -1]^T$ indica um cliente com suspeita de fraude. É importante perceber que a utilização de dois elementos na saída permite, efetivamente, a classificação dos clientes em três classes: clientes com suspeita de fraude (saída $[1 \ -1]^T$),

cliente sem suspeita de fraude (saída $[-1 \ 1]^T$) e clientes destinados a revisão (saídas $[1 \ 1]^T$ e $[-1 \ -1]^T$).

Por fim, cada configuração de RNA foi treinada, validada e testada 20 vezes, para fins de análise de desempenho da configuração em termos estatísticos. O número de vezes escolhido se deu de forma *ad hoc*, visando não restringir o espaço amostral da análise (para cada configuração) e, paralelamente, não incrementar irrestritamente o tempo de simulação.

4 RESULTADOS

Este Capítulo objetiva apresentar e comparar os desempenhos obtidos ao se variar as configurações (combinações dos parâmetros). O desempenho obtido é analisado em termos de:

- i. **Média da Taxa de Acerto:** média aritmética calculada com base nas 20 taxas de acerto de cada configuração. Neste trabalho, considera-se que há acerto por parte da RNA quando os sinais da saída são equivalentes aos sinais da saída esperada. Mede o desempenho;
- ii. **Desvio Padrão da Taxa de Acerto:** desvio padrão amostral calculado com base nas 20 taxas de acerto de cada configuração. Mede a dispersão do desempenho;
- iii. **Máximo da Melhor Época:** máximo do valor de época escolhida como melhor ao fim do treinamento, a partir da coleta das melhores épocas dos 20 treinamentos da configuração. Mede, em épocas, a duração aproximada do treinamento;
- iv. **Mínimo da Melhor Época:** mínimo do valor de época escolhida como melhor ao fim do treinamento, a partir da coleta das melhores épocas dos 20 treinamentos da configuração. Mede, em épocas, a duração aproximada do treinamento.

Ademais, análises são feitas para cada um dos parâmetros variados e seu grau de influência no desempenho da RNA do assistente. Por fim, considerações são feitas do ponto de vista de aplicabilidade do assistente.

4.1 CONFIGURAÇÕES COM MENOR DESEMPENHO

As dez (10) configurações com menor desempenho (das 320 totais, apresentadas no Apêndice A) são apresentadas na Tabela 1. A pior média de taxa de acerto obtida foi de 59,26%.

Tabela 1 – Dez piores desempenhos e suas respectivas configurações

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
2	C2	4	1,75	59,26%	28,76%	61	0
2	C2	5	1,75	61,07%	28,74%	86	0
2	C1	1	0,25	61,09%	5,07%	135	12
2	C1	1	0,5	62,35%	4,44%	45	13
2	C1	1	1,25	62,81%	4,04%	45	12
2	C2	5	2	63,41%	24,76%	61	0
2	C1	1	1,75	63,56%	2,54%	50	17
2	C1	1	0,75	63,57%	3,12%	57	12
2	C1	1	2	64,44%	2,55%	46	14
1	C1	1	0,25	64,78%	3,51%	43	4

Fonte: o próprio autor.

A partir da Tabela 1, evidenciam-se 4 possíveis motivos pelo qual há queda no desempenho da RNA. Percorrendo a tabela da esquerda para a direita, o primeiro motivo aparente está relacionado com o número de saídas, já que 90% dos 10 piores desempenhos apresentam uma saída com dois elementos. Possivelmente, isso se dá pelo fato de a RNA ter de calcular corretamente os dois elementos na saída (ao invés de um) para que o acerto seja considerado. Em outras palavras, não basta acertar um dos elementos, o que reduz a taxa de acerto.

O segundo motivo aparente se relaciona com o tipo de entrada utilizada, onde 70% dos 10 piores desempenhos são de RNA com a configuração de entrada C1. Percebe-se aqui que apenas a normalização do histórico de consumo coletado, sem nenhuma extração adicional de informação no pré-processamento, pode ser a origem de redução do desempenho de uma RNA. Adicionalmente, a concatenação de HCN com apenas os 6 primeiros valores de DHCNOC aparenta promover uma melhoria de desempenho em relação a C1, a ponto de a configuração C3 não fazer parte das 10 configurações com pior desempenho.

Já o terceiro motivo aparente está relacionado com o número de anos considerado. De fato, 70% dos 10 piores desempenhos levam em consideração apenas 1 ano de consumo (apenas 12 valores na entrada, já que a respectiva configuração de entrada é 1).

Por fim, o quarto motivo aparente é proveniente de um encerramento de treinamento precoce, uma vez que há registro de mínimo da melhor época igual a 0. Isso significa que em alguns dos treinamentos, o melhor estado da RNA considerado é o estado

anterior ao treinamento (os valores dos pesos sinápticos finais da RNA são os próprios valores iniciais, aleatoriamente definidos), reduzindo a taxa média de acertos das configurações e aumentando o desvio padrão. Neste caso, seria necessário alterar os parâmetros que definem o critério de parada da RNA no *toolbox* do MATLAB a fim de que mais épocas pudessem ser executadas até que se caracterizasse uma tendência de queda do erro médio quadrático. Este procedimento, contudo, não pode não ser suficiente para que se caracterize um bom desempenho da RNA. A fim de limitar o número de parâmetros variantes neste trabalho, definiu-se que tais os critérios de parada seriam mantidos em seus valores padrão. Tal análise poderá ser abordada em trabalhos futuros.

4.2 CONFIGURAÇÕES COM MAIOR DESEMPENHO

As dez (10) configurações com maior desempenho são apresentadas na Tabela 2. A melhor média de taxa de acerto obtida foi de 83,80%.

Tabela 2 – Dez melhores desempenhos e suas respectivas configurações

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
1	C4	5	1	83,80%	1,54%	44	9
1	C4	5	0,75	83,04%	2,29%	36	8
1	C4	5	0,25	83,04%	1,95%	39	9
1	C4	4	1	83,02%	2,20%	41	14
1	C4	5	0,5	82,98%	2,33%	32	10
1	C4	4	0,25	82,72%	2,15%	33	12
1	C4	4	0,5	82,61%	2,44%	41	13
1	C4	5	1,25	82,52%	4,36%	44	0
1	C4	5	1,75	82,35%	4,05%	43	0
1	C4	4	2	82,28%	2,37%	50	11

Fonte: o próprio autor.

A partir da Tabela 2, evidenciam-se 3 possíveis motivos pelo qual há aumento no desempenho da RNA. Percorrendo a tabela da esquerda para a direita, o primeiro motivo aparente está relacionado com o número de saídas, já que 100% dos 10 melhores desempenhos apresentam uma saída com um elemento.

O segundo motivo aparente se relaciona com o tipo de entrada utilizada, onde 100% dos 10 melhores desempenhos são de RNA que utilizam a configuração de entrada C4. Percebe-se aqui que a extração e utilização adequada de informações adicionais (derivação do consumo) promoveu um ganho no desempenho das RNA.

Já o terceiro motivo aparente está relacionado com o número de anos considerado. De fato, 60% e 40% dos 10 melhores desempenhos levam em consideração 5 e 4 anos, respectivamente. Em termos de entradas, trata-se aqui de vetores de tamanho 119 e 95, respectivamente. Para vetores maiores constituídos de informação não redundante, um resultado melhor no desempenho é esperado, dado que mais informações são disponibilizadas para a RNA.

4.3 ANÁLISE INDIVIDUAL DOS PARÂMETROS VARIADOS

As Seções 4.1 e 4.2 apresentam as 20 configurações com os 10 piores e 10 melhores desempenhos. Com o intuito de aprofundar a análise e confirmar os motivos aparentes apresentados, uma análise individual dos parâmetros variados foi conduzida. Para cada um desses parâmetros, a média de taxa de acerto foi calculada mantendo-se aquele parâmetro constante e variando todo o resto. O resultado é indicado na Tabela 3.

Tabela 3 – Desempenho por parâmetro variado

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto
1	-	-	-	76,54%
2	-	-	-	73,08%
-	C1	-	-	73,15%
-	C2	-	-	73,58%
-	C3	-	-	75,14%
-	C4	-	-	77,38%
-	-	1	-	68,89%
-	-	2	-	74,31%
-	-	3	-	76,34%
-	-	4	-	77,20%
-	-	5	-	77,30%
-	-	-	0,25	75,05%
-	-	-	0,5	75,36%
-	-	-	0,75	75,45%
-	-	-	1	75,40%
-	-	-	1,25	74,82%
-	-	-	1,5	74,44%
-	-	-	1,75	73,82%
-	-	-	2	74,14%

Fonte: o próprio autor.

Com base na Tabela 3, gráficos foram elaborados e são mostrados nas Figuras 5 a 8. Nota-se que os resultados obtidos individualmente corroboram com o que foi apresentado nas Tabelas 1 e 2. Isto é, de fato os piores desempenhos são obtidos nas seguintes condições:

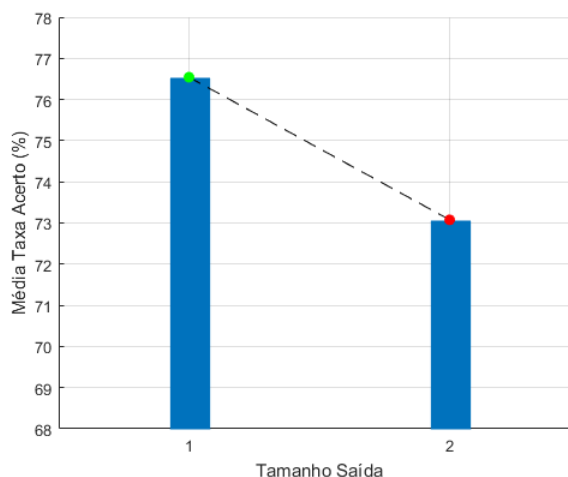
- i. Há camada de saída com 2 elementos;
- ii. Entrada composta apenas pelo HCN, ou;
- iii. Apenas 1 ano de consumo são considerados.

Analogamente, os melhores desempenhos são esperados quando:

- iv. A saída tem apenas de 1 elemento;
- v. A entrada é composta pelo HCN concatenado com DHCNOC, ou;
- vi. São considerados 5 anos de consumo.

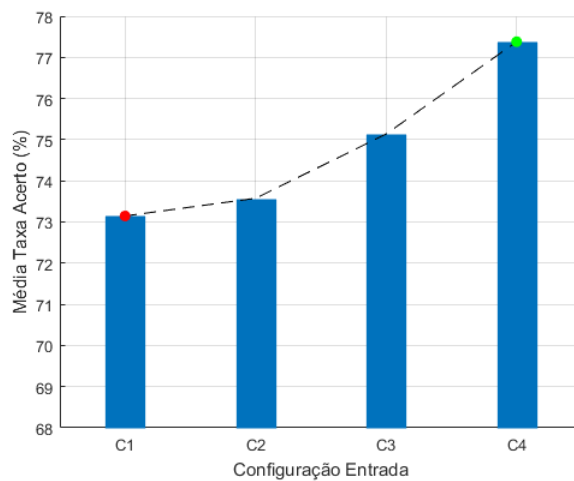
Além disso, é válido destacar que o pior caso esperado (tamanho da saída = 2, configuração da entrada = C1, número de anos = 1 e dimensão da camada oculta = 1,75) e o melhor caso esperado (tamanho da saída = 1, configuração da entrada = C4, número de anos = 5 e dimensão da camada oculta = 0,75) estão, de fato, entre as configurações com piores e melhores desempenhos, conforme apresentado nas Tabelas 1 e 2.

Figura 5 – Desempenho em função do tamanho da saída



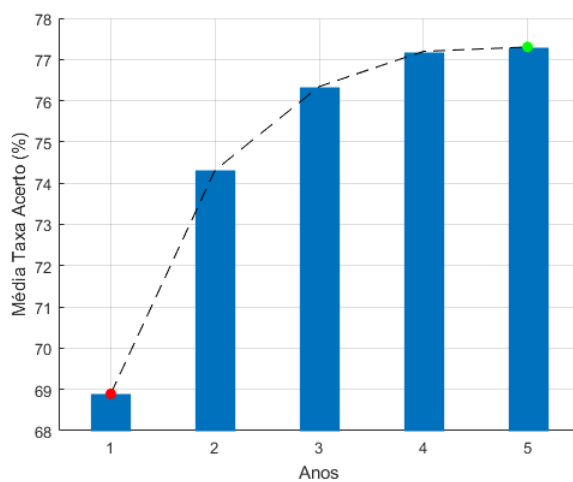
Fonte: o próprio autor.

Figura 6 – Desempenho em função da configuração de entrada



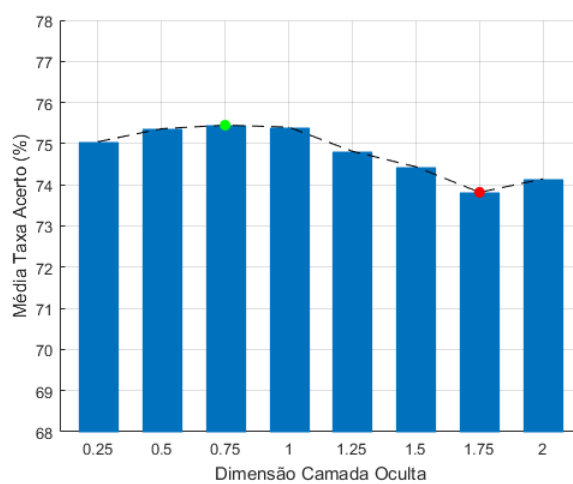
Fonte: o próprio autor.

Figura 7 – Desempenho em função do número de anos de consumo



Fonte: o próprio autor.

Figura 8 – Desempenho em função da dimensão da camada oculta



Fonte: o próprio autor.

Em seguida, para análise do grau de influência de cada parâmetro (dentro de seus respectivos limites neste trabalho), a diferença entre os pontos de máximo e mínimo das Figuras 5 a 8 são apresentados na Tabela 4. É válido destacar que o parâmetro que apresenta menor grau de influência também se encontra em consonância com os resultados encontrados nas Seções 4.1 e 4.2, uma vez que a influência das dimensões não se mostrou evidente nas Tabelas 1 e 2. Para os demais parâmetros, a seguinte ordem de influência é observada: tamanho da saída (variação de 3,46%), configuração da entrada (variação de 4,23%) e número de anos (variação de 8,41%).

Tabela 4 – Grau de influência de cada parâmetro variado

Parâmetro	Variação
Tamanho Saída	3,46%
Configuração Entrada	4,23%
Anos	8,41%
Dimensão Camada Oculta	1,63%

Fonte: o próprio autor.

4.4 ANÁLISE DE ASSISTENTE PARA DISTRIBUIDORA

Com base na Tabela 2, percebe-se que 5 foi o número de anos utilizado na configuração de melhor desempenho. Do ponto de vista prático, para que se exista um histórico de consumo de 5 anos, é necessário que o cliente seja atendido pela distribuidora há 5 anos ou mais. Entretanto, não é preferível que se espere que 5 anos de contrato se completem para que o cliente seja investigado com auxílio do assistente. Além disso, o tempo que se leva para realizar uma consulta de consumo se eleva com o aumento do número de anos a serem consultados. Logo, é desejável que se proponha para a distribuidora a utilização de uma configuração na qual se utiliza um número reduzido de anos.

Caso se proponha um assistente que leva em consideração 2 anos de consumo, a configuração da RNA pode ser escolhida dentre as 5 melhores para 2 anos, apresentadas na Tabela 5. Caso queira se levar em consideração 3 anos, a escolha deve ser feita dentre as 5 melhores para 3 anos, mostradas na Tabela 6.

Tabela 5 – Cinco melhores desempenhos para 2 anos de consumo

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto
1	C4	2	1,5	78,81%	1,82%
1	C4	2	0,75	78,74%	3,37%
1	C4	2	1	78,35%	2,91%
1	C4	2	0,25	78,31%	2,77%
1	C4	2	2	78,22%	2,55%

Fonte: o próprio autor.

Tabela 6 – Cinco melhores desempenhos para 3 anos de consumo

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto
1	C4	3	1	81,26%	1,72%
1	C4	3	1,75	81,24%	2,07%
1	C4	3	0,25	81,00%	2,21%
1	C4	3	2	80,70%	4,54%
1	C4	3	1,25	80,30%	4,14%

Fonte: o próprio autor.

5 CONCLUSÃO

Neste trabalho, perdas não-técnicas de energia, suas causas, seu combate e benefícios da sua redução foram discutidos. Em seguida, o conceito de redes neurais artificiais foi apresentado e discorreu-se sobre seu funcionamento, visando a aplicação de RNA na implementação de um assistente para indicação de clientes fraudulentos de distribuidora de energia elétrica.

Uma metodologia para desenvolvimento do assistente foi proposta, cujo material de trabalho é uma base de dados composta pelo histórico de consumo de (a) clientes conhecidamente fraudulentos e (b) clientes considerados regulares (com base em inspeção visual de seus consumos). Em termos de método, redes do tipo *perceptron* de múltiplas camadas com parâmetros fixos (função de ativação, algoritmo de treinamento, número de camadas ocultas e condições de parada de treinamento) e variados (configuração da entrada, tamanho da saída, número de anos de consumo considerado e dimensão da camada oculta) foram treinadas empregando o ambiente MATLAB. Seus desempenhos foram, então, medidos em termos de taxa média de acerto.

A obtenção de taxas médias de acertos das RNA superiores a 80%, e um máximo de 83,80%, mostra êxito na implementação do assistente proposto. A variação de alguns dos parâmetros das RNA treinadas permitiu tanto a busca pelas melhores e piores configurações de RNA para o assistente, quanto o estudo da influência de cada um desses parâmetros. Desses, o número de ano de consumo se mostrou mais influente, enquanto a dimensão da camada oculta se mostrou a menos relevante. Ademais, sugestões de configuração de RNA foram dadas para a utilização do assistente em distribuidoras de energia elétrica, as quais apresentam sucesso de acerto de até 78,81% (considerando-se 2 anos de histórico de consumo) e 81,26% (considerando-se 3 anos de histórico de consumo).

Para trabalhos futuros, sugere-se uma maior variação da dimensão da camada oculta e outra abordagem de extração de atributos (como utilização de transformadas de Fourier), para fins de comparação com os resultados obtidos neste trabalho. Sugere-se ainda que outros algoritmos de treinamento e topologias sejam testados, em busca de melhores resultados e redução de perdas não-técnicas de energia provenientes de meios fraudulentos.

REFERÊNCIAS

- ADEOTI, O. A., & OSANAIYE, P. A. (2013). Effect of Training Algorithms on the Performance of ANN for Pattern Recognition of Bivariate Process. *International Journal of Computer Applications*, 69.
- ALMEIDA, M. A., OLIVEIRA, W., & DANTAS, P. (2006). Reduções de Perdas de Energia Elétrica na COELBA - Estratégias e Resultados Pós-Privatização. *Revista CIER*, 48.
- ANEEL. (2016). *Agência Nacional de Energia Elétrica*. Acesso em 8 de Dezembro de 2016, disponível em <http://www2.aneel.gov.br/area.cfm?idArea=801&idPerfil=4>
- ARAÚJO, C. M., & SIQUEIRA, C. (2006). Considerações sobre as Perdas na Distribuição de Energia Elétrica no Brasil. *XVII Seminário Nacional de Distribuição de Energia Elétrica*.
- FAUSETT, L. V. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Nova Jersey, EUA: Prentice Hall.
- FERREIRA, T. V. (2011). *Estimação Inteligente da Poluição de Isolamentos Elétricos Baseada nos Vecse do Ruído Ultrassônico*. Tese (Doutorado em Engenharia Elétrica), Universidade Federal de Campina Grande, Departamento de Engenharia Elétrica, Campina Grande.
- HAYKIN, S. S. (2008). *Neural networks and learning machines* (3ª ed.). Hamilton, Ontario, Canadá: Prentice Hall.
- KALMAN, B. L., & KWASNY, S. C. (1992). *Why Tanh: Choosing a Sigmoidal Function*. Artigo, Universidade de Washington, Departamento de Ciências da Computação, Missouri.
- MÉFFE, A. (2001). *Metodologia para Cálculo de Perdas Técnicas por Segmento do Sistema de Distribuição*. Dissertação (Mestrado em Engenharia), Escola Politécnica da Universidade de São Paulo, São Paulo.
- RAMOS, C. C. (2014). *Caracterização de Perdas Comerciais em Sistemas de Energia Através de Técnicas Inteligentes*. Tese (Doutorado em Engenharia Elétrica), Escola Politécnica da Universidade de São Paulo, São Paulo.
- SILVA, E., & OLIVEIRA, A. (2006). *Dicas para a Configuração de Redes Neurais*. Rio de Janeiro: Universidade Federal do Rio de Janeiro.

APÊNDICE A – TABELA DE DESEMPENHOS

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
1	C1	1	0.25	64.78%	3.51%	43	4
1	C1	1	0.5	67.54%	3.30%	73	3
1	C1	1	0.75	68.85%	2.98%	41	3
1	C1	1	1	69.02%	3.18%	67	7
1	C1	1	1.25	68.70%	2.79%	41	6
1	C1	1	1.5	69.02%	3.04%	33	7
1	C1	1	1.75	70.09%	2.00%	43	7
1	C1	1	2	68.87%	2.57%	31	7
1	C1	2	0.25	71.74%	2.50%	44	4
1	C1	2	0.5	73.07%	2.24%	42	6
1	C1	2	0.75	73.39%	2.28%	40	8
1	C1	2	1	73.26%	3.44%	30	7
1	C1	2	1.25	73.56%	3.17%	36	7
1	C1	2	1.5	73.33%	2.31%	33	6
1	C1	2	1.75	74.15%	3.23%	32	7
1	C1	2	2	75.87%	2.61%	38	14
1	C1	3	0.25	76.07%	3.03%	48	13
1	C1	3	0.5	76.37%	2.63%	48	14
1	C1	3	0.75	77.07%	2.23%	40	9
1	C1	3	1	76.81%	2.99%	35	8
1	C1	3	1.25	77.48%	2.79%	32	6
1	C1	3	1.5	76.57%	3.40%	40	12
1	C1	3	1.75	77.22%	2.17%	41	14
1	C1	3	2	77.54%	2.98%	46	7
1	C1	4	0.25	78.41%	3.11%	40	10
1	C1	4	0.5	78.13%	2.26%	37	8
1	C1	4	0.75	79.30%	2.11%	43	10
1	C1	4	1	80.28%	2.54%	50	9
1	C1	4	1.25	77.61%	3.99%	53	0
1	C1	4	1.5	79.46%	2.62%	50	14
1	C1	4	1.75	78.72%	2.39%	59	9
1	C1	4	2	78.87%	2.28%	57	9
1	C1	5	0.25	79.26%	2.39%	28	4
1	C1	5	0.5	79.87%	1.87%	36	8
1	C1	5	0.75	79.61%	2.07%	31	6
1	C1	5	1	78.69%	2.60%	53	9
1	C1	5	1.25	79.39%	2.15%	33	9
1	C1	5	1.5	80.83%	2.23%	46	8

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
1	C1	5	1.75	78.89%	2.84%	53	11
1	C1	5	2	79.28%	2.39%	61	5
1	C2	1	0.25	71.09%	3.65%	61	6
1	C2	1	0.5	72.19%	2.74%	45	9
1	C2	1	0.75	72.24%	2.58%	76	12
1	C2	1	1	72.93%	2.45%	47	9
1	C2	1	1.25	73.39%	2.07%	63	7
1	C2	1	1.5	72.20%	3.40%	41	9
1	C2	1	1.75	73.00%	2.25%	34	8
1	C2	1	2	73.57%	3.32%	76	8
1	C2	2	0.25	75.94%	2.73%	78	3
1	C2	2	0.5	77.26%	1.78%	59	14
1	C2	2	0.75	76.24%	3.09%	50	7
1	C2	2	1	76.13%	2.28%	40	13
1	C2	2	1.25	76.54%	1.92%	46	9
1	C2	2	1.5	76.46%	3.08%	67	9
1	C2	2	1.75	76.02%	4.66%	44	0
1	C2	2	2	76.57%	4.04%	35	0
1	C2	3	0.25	79.48%	2.00%	65	16
1	C2	3	0.5	78.59%	2.26%	63	20
1	C2	3	0.75	79.02%	1.85%	66	17
1	C2	3	1	77.46%	6.75%	55	0
1	C2	3	1.25	78.72%	4.24%	82	0
1	C2	3	1.5	68.87%	10.47%	50	0
1	C2	3	1.75	70.35%	11.83%	78	0
1	C2	3	2	75.74%	9.12%	51	0
1	C2	4	0.25	79.19%	2.39%	91	12
1	C2	4	0.5	79.26%	2.91%	81	3
1	C2	4	0.75	78.50%	8.47%	81	0
1	C2	4	1	74.17%	10.10%	65	0
1	C2	4	1.25	73.15%	11.85%	93	0
1	C2	4	1.5	70.59%	11.68%	70	0
1	C2	4	1.75	68.98%	14.09%	50	0
1	C2	4	2	64.93%	13.89%	70	0
1	C2	5	0.25	79.54%	3.85%	94	3
1	C2	5	0.5	78.91%	6.39%	78	0
1	C2	5	0.75	76.31%	9.62%	111	0
1	C2	5	1	77.61%	8.66%	62	0
1	C2	5	1.25	74.54%	11.88%	80	0
1	C2	5	1.5	67.06%	14.34%	50	0

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
1	C2	5	1.75	67.57%	14.63%	62	0
1	C2	5	2	70.13%	12.59%	79	0
1	C3	1	0.25	69.61%	2.10%	48	5
1	C3	1	0.5	70.78%	2.30%	30	7
1	C3	1	0.75	71.33%	2.33%	36	6
1	C3	1	1	71.81%	2.39%	29	5
1	C3	1	1.25	72.04%	3.01%	33	8
1	C3	1	1.5	71.78%	3.43%	30	5
1	C3	1	1.75	72.24%	3.00%	40	6
1	C3	1	2	71.02%	2.39%	29	6
1	C3	2	0.25	75.20%	2.66%	36	10
1	C3	2	0.5	75.63%	2.31%	34	13
1	C3	2	0.75	76.72%	2.81%	39	7
1	C3	2	1	75.98%	2.33%	35	14
1	C3	2	1.25	76.48%	2.13%	33	10
1	C3	2	1.5	77.94%	1.98%	36	8
1	C3	2	1.75	76.11%	2.57%	33	7
1	C3	2	2	76.80%	2.17%	34	10
1	C3	3	0.25	77.56%	2.31%	29	5
1	C3	3	0.5	77.43%	2.52%	32	5
1	C3	3	0.75	78.87%	2.30%	28	13
1	C3	3	1	78.04%	2.79%	30	10
1	C3	3	1.25	79.04%	1.72%	45	11
1	C3	3	1.5	79.52%	1.97%	39	12
1	C3	3	1.75	77.91%	2.96%	33	11
1	C3	3	2	78.35%	2.68%	46	16
1	C3	4	0.25	80.85%	3.17%	39	11
1	C3	4	0.5	81.07%	2.52%	34	9
1	C3	4	0.75	80.28%	2.41%	28	5
1	C3	4	1	80.52%	2.02%	54	12
1	C3	4	1.25	80.20%	2.27%	46	8
1	C3	4	1.5	80.74%	2.93%	39	8
1	C3	4	1.75	79.78%	2.44%	36	14
1	C3	4	2	78.96%	2.32%	62	9
1	C3	5	0.25	80.59%	2.11%	32	7
1	C3	5	0.5	81.11%	2.35%	36	6
1	C3	5	0.75	80.63%	2.28%	35	9
1	C3	5	1	80.43%	2.64%	51	10
1	C3	5	1.25	80.00%	2.53%	57	10
1	C3	5	1.5	79.70%	2.46%	57	5

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
1	C3	5	1.75	79.20%	3.03%	47	4
1	C3	5	2	79.17%	2.38%	43	10
1	C4	1	0.25	72.44%	2.97%	32	9
1	C4	1	0.5	72.43%	1.92%	43	7
1	C4	1	0.75	72.50%	2.44%	26	7
1	C4	1	1	72.80%	2.82%	38	4
1	C4	1	1.25	72.15%	2.61%	41	11
1	C4	1	1.5	72.15%	2.56%	31	8
1	C4	1	1.75	72.72%	2.52%	33	13
1	C4	1	2	73.63%	2.02%	40	9
1	C4	2	0.25	78.31%	2.77%	44	10
1	C4	2	0.5	77.30%	2.70%	35	11
1	C4	2	0.75	78.74%	3.37%	32	9
1	C4	2	1	78.35%	2.91%	41	11
1	C4	2	1.25	77.80%	1.91%	34	11
1	C4	2	1.5	78.81%	1.82%	31	11
1	C4	2	1.75	77.54%	2.46%	28	11
1	C4	2	2	78.22%	2.55%	53	10
1	C4	3	0.25	81.00%	2.21%	33	8
1	C4	3	0.5	79.87%	4.76%	37	0
1	C4	3	0.75	79.67%	5.87%	37	0
1	C4	3	1	81.26%	1.72%	41	8
1	C4	3	1.25	80.30%	4.14%	49	0
1	C4	3	1.5	79.15%	5.44%	48	0
1	C4	3	1.75	81.24%	2.07%	41	10
1	C4	3	2	80.70%	4.54%	45	0
1	C4	4	0.25	82.72%	2.15%	33	12
1	C4	4	0.5	82.61%	2.44%	41	13
1	C4	4	0.75	82.04%	3.38%	45	0
1	C4	4	1	83.02%	2.20%	41	14
1	C4	4	1.25	80.91%	8.64%	51	0
1	C4	4	1.5	79.98%	8.38%	50	0
1	C4	4	1.75	77.24%	9.35%	42	0
1	C4	4	2	82.28%	2.37%	50	11
1	C4	5	0.25	83.04%	1.95%	39	9
1	C4	5	0.5	82.98%	2.33%	32	10
1	C4	5	0.75	83.04%	2.29%	36	8
1	C4	5	1	83.80%	1.54%	44	9
1	C4	5	1.25	82.52%	4.36%	44	0
1	C4	5	1.5	78.50%	10.25%	47	0

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
1	C4	5	1.75	82.35%	4.05%	43	0
1	C4	5	2	81.78%	8.16%	36	1
2	C1	1	0.25	61.09%	5.07%	135	12
2	C1	1	0.5	62.35%	4.44%	45	13
2	C1	1	0.75	63.57%	3.12%	57	12
2	C1	1	1	64.96%	3.25%	53	14
2	C1	1	1.25	62.81%	4.04%	45	12
2	C1	1	1.5	64.83%	3.06%	47	16
2	C1	1	1.75	63.56%	2.54%	50	17
2	C1	1	2	64.44%	2.55%	46	14
2	C1	2	0.25	68.69%	3.47%	69	9
2	C1	2	0.5	68.93%	3.97%	50	13
2	C1	2	0.75	68.24%	4.29%	47	12
2	C1	2	1	67.70%	2.48%	48	10
2	C1	2	1.25	70.24%	2.57%	44	14
2	C1	2	1.5	68.94%	2.74%	60	10
2	C1	2	1.75	70.17%	2.83%	51	9
2	C1	2	2	68.61%	2.45%	36	11
2	C1	3	0.25	72.37%	3.75%	58	4
2	C1	3	0.5	72.26%	2.62%	36	7
2	C1	3	0.75	71.50%	2.40%	43	10
2	C1	3	1	71.31%	4.26%	50	13
2	C1	3	1.25	73.22%	2.77%	38	10
2	C1	3	1.5	72.33%	3.08%	43	14
2	C1	3	1.75	73.28%	3.07%	47	12
2	C1	3	2	72.87%	2.86%	42	6
2	C1	4	0.25	75.13%	3.81%	42	10
2	C1	4	0.5	75.52%	2.78%	37	15
2	C1	4	0.75	75.69%	2.74%	41	16
2	C1	4	1	74.15%	2.34%	48	12
2	C1	4	1.25	75.81%	2.82%	50	8
2	C1	4	1.5	73.70%	3.05%	59	13
2	C1	4	1.75	74.59%	2.32%	63	8
2	C1	4	2	73.54%	2.92%	77	13
2	C1	5	0.25	76.04%	2.98%	33	7
2	C1	5	0.5	76.07%	2.85%	34	10
2	C1	5	0.75	75.00%	2.29%	42	11
2	C1	5	1	74.96%	2.20%	37	9
2	C1	5	1.25	74.22%	2.84%	38	8
2	C1	5	1.5	74.37%	3.10%	49	7

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
2	C1	5	1.75	74.28%	2.56%	48	12
2	C1	5	2	73.39%	3.95%	60	10
2	C2	1	0.25	68.76%	5.73%	154	17
2	C2	1	0.5	69.00%	4.42%	110	10
2	C2	1	0.75	69.15%	3.42%	79	10
2	C2	1	1	68.57%	2.27%	77	9
2	C2	1	1.25	70.30%	3.11%	121	10
2	C2	1	1.5	69.15%	3.00%	82	8
2	C2	1	1.75	69.41%	2.64%	55	9
2	C2	1	2	69.31%	1.78%	63	8
2	C2	2	0.25	73.72%	2.62%	63	16
2	C2	2	0.5	73.93%	2.46%	57	9
2	C2	2	0.75	74.17%	2.76%	48	11
2	C2	2	1	74.91%	2.65%	55	9
2	C2	2	1.25	73.87%	2.61%	87	12
2	C2	2	1.5	74.28%	2.19%	46	10
2	C2	2	1.75	75.43%	2.43%	62	15
2	C2	2	2	73.83%	2.34%	52	11
2	C2	3	0.25	76.63%	2.51%	71	12
2	C2	3	0.5	76.35%	2.09%	80	12
2	C2	3	0.75	76.44%	3.09%	65	14
2	C2	3	1	76.28%	1.96%	65	18
2	C2	3	1.25	76.30%	3.01%	55	14
2	C2	3	1.5	76.81%	2.93%	73	19
2	C2	3	1.75	76.46%	2.29%	60	19
2	C2	3	2	71.54%	17.00%	69	0
2	C2	4	0.25	77.54%	2.04%	144	23
2	C2	4	0.5	75.04%	12.59%	62	0
2	C2	4	0.75	78.00%	2.42%	82	26
2	C2	4	1	78.19%	1.94%	70	19
2	C2	4	1.25	71.98%	18.72%	94	0
2	C2	4	1.5	76.02%	7.16%	69	0
2	C2	4	1.75	59.26%	28.76%	61	0
2	C2	4	2	68.22%	21.33%	86	0
2	C2	5	0.25	78.13%	3.85%	97	6
2	C2	5	0.5	76.67%	3.74%	91	9
2	C2	5	0.75	75.87%	7.74%	108	0
2	C2	5	1	76.24%	6.76%	97	1
2	C2	5	1.25	66.04%	25.30%	80	0
2	C2	5	1.5	69.35%	17.26%	64	0

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
2	C2	5	1.75	61.07%	28.74%	86	0
2	C2	5	2	63.41%	24.76%	61	0
2	C3	1	0.25	64.98%	4.92%	55	8
2	C3	1	0.5	65.69%	2.86%	35	7
2	C3	1	0.75	66.28%	2.89%	42	11
2	C3	1	1	67.57%	2.40%	38	10
2	C3	1	1.25	65.72%	2.43%	62	9
2	C3	1	1.5	66.78%	2.31%	36	8
2	C3	1	1.75	66.91%	2.89%	40	15
2	C3	1	2	66.83%	3.29%	39	11
2	C3	2	0.25	72.19%	3.49%	73	14
2	C3	2	0.5	73.24%	2.57%	46	16
2	C3	2	0.75	71.57%	2.67%	44	14
2	C3	2	1	71.69%	2.57%	42	12
2	C3	2	1.25	72.50%	2.46%	41	10
2	C3	2	1.5	72.35%	2.66%	30	8
2	C3	2	1.75	73.19%	2.22%	53	12
2	C3	2	2	72.70%	3.12%	38	16
2	C3	3	0.25	73.43%	3.34%	45	10
2	C3	3	0.5	74.74%	2.66%	36	11
2	C3	3	0.75	74.87%	3.41%	32	15
2	C3	3	1	74.96%	2.41%	39	14
2	C3	3	1.25	74.07%	2.15%	41	16
2	C3	3	1.5	73.57%	2.38%	35	11
2	C3	3	1.75	73.70%	2.40%	46	12
2	C3	3	2	74.00%	2.37%	45	20
2	C3	4	0.25	74.76%	2.24%	47	10
2	C3	4	0.5	77.89%	2.30%	34	12
2	C3	4	0.75	75.74%	2.35%	38	12
2	C3	4	1	76.19%	2.61%	38	11
2	C3	4	1.25	76.43%	2.33%	38	10
2	C3	4	1.5	75.30%	2.30%	40	13
2	C3	4	1.75	75.07%	2.37%	59	19
2	C3	4	2	75.65%	2.56%	60	15
2	C3	5	0.25	76.31%	2.52%	30	8
2	C3	5	0.5	76.91%	2.35%	28	9
2	C3	5	0.75	76.70%	2.37%	37	10
2	C3	5	1	76.96%	2.14%	46	11
2	C3	5	1.25	75.33%	1.84%	38	14
2	C3	5	1.5	76.28%	2.71%	59	13

Tamanho Saída	Configuração Entrada	Anos	Dimensão Camada Oculta	Média Taxa Acerto	Desvio Padrão Taxa Acerto	Máximo Melhor Época	Mínimo Melhor Época
2	C3	5	1.75	74.67%	2.13%	79	10
2	C3	5	2	75.80%	3.04%	47	12
2	C4	1	0.25	65.93%	3.94%	41	8
2	C4	1	0.5	67.33%	2.70%	59	13
2	C4	1	0.75	68.06%	2.93%	28	10
2	C4	1	1	67.19%	2.65%	51	10
2	C4	1	1.25	67.17%	2.61%	26	12
2	C4	1	1.5	67.81%	2.74%	33	9
2	C4	1	1.75	67.85%	2.28%	37	12
2	C4	1	2	66.93%	2.87%	40	9
2	C4	2	0.25	73.87%	1.41%	51	12
2	C4	2	0.5	74.43%	2.94%	37	6
2	C4	2	0.75	74.72%	2.04%	35	11
2	C4	2	1	75.19%	1.86%	41	11
2	C4	2	1.25	74.65%	2.84%	34	13
2	C4	2	1.5	74.39%	3.04%	32	14
2	C4	2	1.75	74.46%	2.77%	36	15
2	C4	2	2	73.74%	2.58%	49	16
2	C4	3	0.25	76.09%	2.31%	35	12
2	C4	3	0.5	76.31%	2.33%	46	13
2	C4	3	0.75	77.54%	2.12%	29	13
2	C4	3	1	76.80%	2.43%	47	10
2	C4	3	1.25	76.24%	2.45%	38	14
2	C4	3	1.5	76.59%	1.78%	39	12
2	C4	3	1.75	76.80%	1.48%	62	10
2	C4	3	2	77.09%	2.16%	47	12
2	C4	4	0.25	79.20%	2.29%	33	12
2	C4	4	0.5	78.91%	2.43%	34	15
2	C4	4	0.75	79.56%	2.74%	41	12
2	C4	4	1	79.56%	2.66%	57	12
2	C4	4	1.25	80.09%	1.86%	43	12
2	C4	4	1.5	80.80%	2.06%	50	14
2	C4	4	1.75	79.80%	1.91%	42	13
2	C4	4	2	80.52%	2.59%	54	14
2	C4	5	0.25	80.20%	2.00%	35	10
2	C4	5	0.5	80.59%	2.23%	34	11
2	C4	5	0.75	80.98%	2.31%	49	14
2	C4	5	1	80.35%	2.33%	41	10
2	C4	5	1.25	81.13%	2.29%	61	11
2	C4	5	1.5	81.17%	2.66%	41	11
2	C4	5	1.75	81.44%	3.32%	53	9
2	C4	5	2	81.09%	1.90%	55	8

