UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO

Bruno Stinghel Mattedi

PREDIÇÃO DE INDICADORES DE QUALIDADE NO FORNECIMENTO DE ENERGIA ELÉTRICA ATRAVÉS DE REDES NEURAIS

Vitória - ES

Dezembro/2018

Bruno Stinghel Mattedi

PREDIÇÃO DE INDICADORES DE QUALIDADE NO FORNECIMENTO DE ENERGIA ELÉTRICA ATRAVÉS DE REDES NEURAIS

Parte manuscrita do Projeto de Graduação do aluno Bruno Stinghel Mattedi, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito para obtenção do grau de Engenheiro Eletricista.

Vitória - ES

Dezembro/2018

Bruno Stinghel Mattedi

PREDIÇÃO DE INDICADORES DE QUALIDADE NO FORNECIMENTO DE ENERGIA ELÉTRICA ATRAVÉS DE REDES NEURAIS

Parte manuscrita do Projeto de Graduação do aluno Bruno Stinghel Mattedi, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito para obtenção do grau de Engenheiro Eletricista.

Aprovado em 13 de dezembro de 2018.

COMISSÃO EXAMINADORA:

Prof. Dr. Klaus Fabian Côco
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Patrick Marques Ciarelli
Universidade Federal do Espírito Santo
Examinador

Prof. Dr. Oureste Elias Batista
Universidade Federal do Espírito Santo
Examinador

Vitória - ES

Dezembro/2018

AGRADECIMENTOS

Meus agradecimentos,

À Deus por estar sempre de braços abertos para nos acolher.

Aos meus pais, Carmen e Francisco, e minha irmã, Flávia, pelo companheirismo de todas as horas e pela força nos momentos de fraqueza.

Aos meus avós que são meus segundos pais, os quais tenho sempre em meu coração.

Aos meus amigos, quase irmãos, que fiz durante essa jornada, proporcionando descontração perante as dificuldades.

Ao Prof. Dr. Klaus Fabian Coco, por estar sempre disponível para tirar as dúvidas ativamente ajudando com o trabalho.

À banca examinadora pela aceitação do convite e pelo tempo investido para leitura e avaliação desse trabalho.

"Panic is the seed of defeat, so you must maintain your calm and think logically. Look beyond your surroundings, and don't waste your effort on unnecessary details."

(Overlord)

RESUMO

Com a evolução dos processos industriais, o mercado tem se tornado cada vez mais competitivo, onde qualquer vantagem é o suficiente para ser a diferença entre ganhar e perder dinheiro. Apesar de contrário ao senso, o mesmo se aplica ao mercado de energia cativo, onde o consumidor não escolhe a distribuidora de energia. Isso ocorre pelo fato de que a Agência Nacional de Energia Elétrica, com seu poder regulatório, estabeleceu limites para a qualidade do serviço, que ao serem ultrapassados pelas distribuidoras, acarretam em multas à serem pagas. Assim, o controle desses indicadores proporciona uma grande vantagem estratégica, de modo que os investimentos possam ser otimizados para evitar que o limite seja violado. Técnicas simplistas, como modelos autoregressivos integrados de média móvel (*autoregressive integrated moving average* - ARIMA), estão perdendo seu lugar para processos mais sofisticados, como as redes neurais rasas e profundas, a fim de tornar a previsão mais adaptável. Este trabalho descreve o uso de redes neurais para a previsão dos indicadores de continuidade para uma distribuidora de energia elétrica. O foco é criar predições que possuam um bom desempenho, possibilitando a otimização dos gastos com a redução dos valores pagos em multas. O método proposto é utilizar redes neurais rasas do tipo autorregressivo não linear de entradas exógenas (*Nonlinear Autoregressive with exogenous inputs* - NARX) e redes recorrentes do tipo *Long Short-Term Memory* (LSTM). Utiliza-se um banco de dados composto por séries temporais dos indicadores de continuidade, constituídas a partir das ordens de interrupção no fornecimento. Assim, treina-se as redes dentro de determinados intervalos de variação dos seus parâmetros de forma a obter o melhor desempenho e comparar as previsões feitas com os valores reais. A partir das comparações entre os métodos propostos em relação aos métodos clássicos e aos valores observados, foi possível observar que a entrada exógena, por meio da rede NARX, melhorou significativamente o desempenho das predições das séries de alta frequência.

Palavras-chave: *Machine Learning*; *Deep Learning*; Indicadores de Qualidade; Séries Temporais;

ABSTRACT

Together with the evolution of industrial processes, the market has become increasingly competitive, where any advantage is enough to be the difference between winning and losing money. Although contrary to the sense, the same applies to the market of captive energy, where the consumer does not choose the energy distributor. This is due to the fact that the National Electric Energy Agency (ANEEL), with its authority, established limits for the service's quality, which, when overtaken by the companies, entails a fine. Hence, control of these indicators provides a great strategic advantage, because of the oportunit to optimize the investments to prevent the threshold being breached. Simplistic techniques such as integrated autoregressive integrated moving average models (ARIMA) are losing their place to more sophisticated processes such as shallow and deep neural networks, seeking to make prediction more adaptive. This paper describes the use of neural networks to predict the continuity indicators of an power retailer company. The focus is to create predictions that have the best performance in relation to a PPGEE master's research, allowing the optimization of expenses with the reduction of amounts paid in fines. The proposed method is to use a non-linear autoregressive exogenous inputs and recurrent Long Short-Term Memory (LSTM) networks. The data used is composed of time series of the continuity indicators, constituted from the interruption orders in the supply. Thus, the networks are trained within certain ranges of their parameters in order to obtain the best performance and compare the forecasts made with the actual values. From the comparisons between the proposed methods with respect to the classical methods and the observed values, it was possible to observe that the exogenous input, through the NARX network, significantly improved the result of the high frequency series' predictions.

Keywords: Machine Learning; Deep Learning; Quality Indicators; Time Series.

LISTA DE FIGURAS

Figura 1 – Revoluções industriais 13
Figura 2 – Consumo mundial de energia de 1820-2010 17
Figura 3 – Custo, em dólares, devido as falhas no fornecimento de energia para as regiões dos EUA 19
Figura 4 – Comparação entre DEC previsto através da rede NAR 20
Figura 5 – Comparação entre FEC previsto através da rede NAR 20
Figura 6 – Crescimento da População Norte Americana 23
Figura 7 – Variação do valor das ações da Petrobras ao longo do tempo 23
Figura 8 – Janela deslizante no tempo 27
Figura 9 – Tipos de algoritmos de aprendizado de máquina e suas aplicações . . . 29
Figura 10 – Representação de um aprendizado não supervisionado 30
Figura 11 – Modelo de um neurônio 31
Figura 12 – Função de ativação linear 32
Figura 13 – Função de ativação sigmoide 33
Figura 14 – Função de ativação sigmoide e tangente hiperbólica 33
Figura 15 – Rede recorrente com neurônios ocultos 34
Figura 16 – Modelo autorregressivo não linear de entradas exógenas 35
Figura 17 – Modelo autorregressivo não linear 36
Figura 18 – Modelo autorregressivo não linear de entradas exógenas 36
Figura 19 – Gradiente em declive acentuado vs plano 37
Figura 20 – Redes com retropropagação vs dependências antigas 38
Figura 21 – Camada LSTM 38
Figura 22 – Bloco/Célula LSTM 38
Figura 23 – Segregação semanal 42
Figura 24 – Amostra do DEC antes e depois do tratamento de *outliers* 43
Figura 25 – Série temporal de DEC da quinta-feita com e sem *outlier* 43
Figura 26 – Comparação da série após oscilação e raiz quadrada 44
Figura 27 – Subséries após transformada wavelet 45
Figura 28 – Série de Baixa Frequência 46
Figura 29 – Dados Normalizados 46
Figura 30 – Série de alta frequência interpolada 47
Figura 31 – Rede NARX com uma camada oculta 48
Figura 32 – valor observado vs previsão da subsérie de Baixa Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NAR 50
Figura 33 – valor observado vs previsão da subsérie de Média Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NAR 51

Figura 34 – valor observado vs previsão da subsérie de Alta Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NAR 51
Figura 35 – valor observado vs previsão da subsérie de Baixa Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NARX 52
Figura 36 – valor observado vs previsão da subsérie de Média Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NARX 53
Figura 37 – valor observado vs previsão da subsérie de Alta Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NARX 53
Figura 38 – valor observado vs previsão da subsérie de Baixa Frequência para o indicador DEC no dia 05/01/2016 através de uma rede LSTM 54
Figura 39 – valor observado vs previsão da subsérie de Média Frequência para o indicador DEC no dia 05/01/2016 através de uma rede LSTM 55
Figura 40 – valor observado vs previsão da subsérie de Alta Frequência para o indicador DEC no dia 05/01/2016 através de uma rede LSTM 55
Figura 41 – Comparação DEC previsto com o DEC Observado 58
Figura 42 – Comparação FEC previsto com o FEC Observado 58
Figura 43 – DEC e FEC acumulado mensal 59
Figura 44 – Boxplot do erro relativo para subsérie de baixa frequência do indicador DEC 60
Figura 45 – Boxplot do erro relativo para subsérie de média frequência do indicador DEC 60
Figura 46 – Boxplot do erro relativo para subsérie de alta frequência do indicador DEC 61
Figura 47 – Boxplot do erro relativo para subsérie de baixa frequência do indicador FEC 61
Figura 48 – Boxplot do erro relativo para subsérie de média frequência do indicador FEC 62
Figura 49 – Boxplot do erro relativo para subsérie de alta frequência do indicador FEC 62

LISTA DE TABELAS

Tabela 1 – Indicadores Coletivos de Continuidade do Conjunto de São Mateus . . 41
Tabela 2 – Configuração referente ao melhor desempenho obtido para o dia 05/01/2016 referente a rede NAR para o indicador DEC 50
Tabela 3 – Configuração referente ao melhor desempenho obtido para o dia 05/01/2016 referente a rede NARX para o indicador DEC 52
Tabela 4 – Configuração referente ao melhor desempenho obtido para o dia 05/01/2016 referente a rede LSTM para o indicador DEC 54
Tabela 5 – Tabela de erro referente as previsões dos indicadores de continuidade de janeiro/2016 - (Continua) 56
Tabela 6 – Tabela de erro referente as previsões dos indicadores de continuidade de janeiro/2016 - (Continuação) 57
Tabela 7 – MSE das redes propostas em relação aos indicadores DEC e FEC . . . 59

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	*Agência Nacional De Energia Elétrica*
ABRADEE	*Associação Brasileira de Distribuidores de Energia Elétrica*
BC/UFES	*Biblioteca Central / UFES*
CPU	*Unidade Central de Processamento (Central Process Unit)*
DEC	*Duração da Interrupção individual ocorrida em dia crítico por unidade consumidora*
DIC	*Duração de interrupção Individual por unidade consumidora*
DICRI	*Duração da Interrupção individual ocorrida em dia crítico por unidade consumidora*
DMIC	*Duração Máxima de Interrupção Contínua por unidade consumidora*
FEC	*Frequência Equivalente de Interrupção por Unidade Consumidora*
FIC	*Frequência de interrupção Individual por unidade consumidora*
GPU	*Unidade de Processamento Gráfico (Graphics Processing Unit)*
LCEE/UFES	*Laboratório de Computação da Engenharia Elétrica / UFES*
PPGEE	*Programa de Pós Graduação em Engenharia Elétrica / UFES*
UFES	Universidade Federal do Espírito Santo

SUMÁRIO

1 INTRODUÇÃO . . . 13
1.1 Apresentação e Objeto de Pesquisa . . . 13
1.2 Indicadores de Continuidade do Serviço de Distribuição de Energia Elétrica . . . 15
1.2.1 Indicadores Individuais de Continuidade . . . 15
1.2.2 Indicadores Coletivos de Continuidade . . . 16
1.3 Justificativa . . . 17
1.4 Objetivos . . . 19
1.4.1 Objetivo Geral . . . 19
1.4.2 Objetivos Específicos . . . 19
1.5 Trabalho de Referência . . . 20
1.6 Organização do Trabalho . . . 21

2 EMBASAMENTO TEÓRICO . . . 22
2.1 Introdução . . . 22
2.2 Séries Temporais . . . 22
2.3 Decomposição de Séries Temporais . . . 23
2.4 Tratamento de dados . . . 24
2.4.1 Normalização . . . 24
2.4.2 Agregação . . . 24
2.4.3 Limpeza dos dados . . . 25
2.5 Transformada Wavelet de Máxima Sobreposição . . . 26
2.6 Aprendizado de Máquina . . . 28
2.6.1 Redes Neurais Artificiais . . . 30
2.6.2 Aprendizagem Profunda . . . 36
2.6.3 Otimizadores . . . 39

3 METODOLOGIA E ETAPAS DE DESENVOLVIMENTO . . . 41
3.1 Divisão do projeto em etapas . . . 41
3.2 Visão geral . . . 41
3.3 Tratamento de Dados . . . 42
3.4 Treinamento da Rede . . . 47
3.4.1 Rede NAR e NARX . . . 48
3.4.2 Aprendizado profundo com camada LSTM . . . 48
3.5 Previsão da Rede . . . 49
3.5.1 Rede NAR . . . 49

3.5.2 Rede NARX 52
3.5.3 Rede LSTM 53

4 RESULTADOS E CONSIDERAÇÕES FINAIS 56
4.1 Resultados das previsões dos indicadores DEC e FEC para os dias 01/01/2016 à 29/01/2016 56
4.2 Erros para as subséries de frequência 59

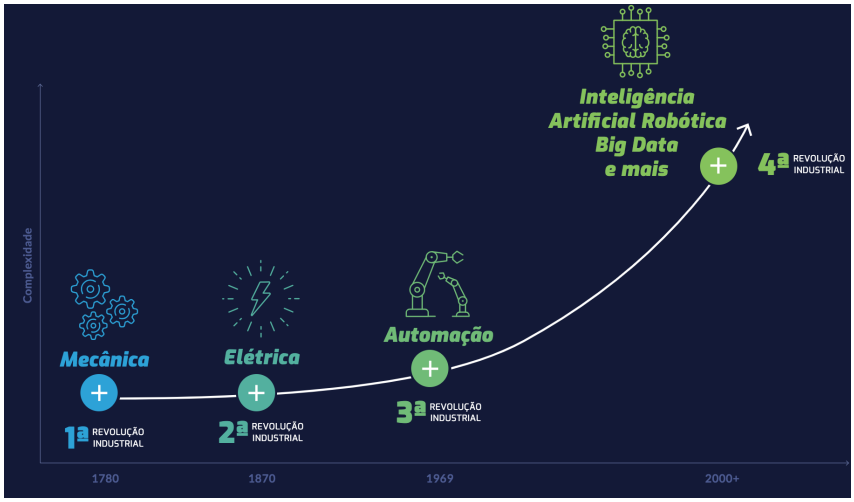
5 CONCLUSÕES E PROJETOS FUTUROS 63
5.1 Trabalhos Futuros 63

REFERÊNCIAS 65

1 INTRODUÇÃO

A indústria é a atividade que tem por finalidade produzir bens materiais comercializáveis a partir da matéria-prima. A partir da primeira revolução industrial em 1780, este setor da economia começou a evoluir rapidamente, sendo que atualmente ocorre a quarta revolução industrial (Figura 1), marcada pelo avanço da inteligência artificial, permitindo que o mundo físico e o virtual estejam cada vez mais próximos.

Figura 1 – Revoluções industriais



Fonte: MME (2017)

Com essa integração, é gerada uma quantidade massiva de dados, sendo necessário o desenvolvimento de técnicas para processar e extrair as informações mais relevantes, bem como utilizá-las de forma a obter uma produção/venda mais eficiente.

Uma das formas de otimizar os processos é por meio de análises preditivas, de modo a reduzir o tempo de parada dos processos e uma melhor gestão dos recursos (LU, 2017).

1.1 Apresentação e Objeto de Pesquisa

Com a indústria 4.0, os mercados têm se tornado cada vez mais competitivos, com mudanças rápidas e significativas, a maioria das empresas é lançada a uma incessante busca por qualidade e produtividade (SANTOS; COLOSIMO; MOTTA, 2007). Neste contexto, se almeja que os recursos sejam utilizados de maneira ótima, ou seja, de forma a maximizar o tempo de operação para garantir a maior taxa de retorno por investimento.

Dessa mesma forma, todavia, com algumas onerações a mais, as empresas brasileiras no

mercado de distribuição de energia elétrica buscam otimizar e ampliar a eficiência de seus serviços. O lucro de uma distribuidora de energia se dá em função da TUSD (Tarifa de Uso do Sistema de Distribuição), que engloba as despesas de capital e os custos de operação e manutenção das redes de distribuição (ABRAADE, 2016), tendo seu rendimento de acordo com a estrutura disponibilizada para os clientes.

Em contrapartida, a ANEEL (Agência Nacional de Energia Elétrica), responsável pela regulação e fiscalização do setor elétrico brasileiro, não permite o repasse automático dos custos operacionais para as tarifas, sendo repassados apenas os custos por ela considerados eficientes. Além disso, define limites para a qualidade do produto e dos serviços prestados, estabelecidos no módulo 8 do PRODIST (ANEEL, 2017c).

Assim, a respeito da qualidade do serviço, a ANEEL exige que as distribuidoras mantenham um padrão de serviços através dos limites dos indicadores de continuidade da rede, seja os coletivos (DEC – Duração Equivalente de interrupção por unidade Consumidora e FEC - Frequência Equivalente de interrupção por unidade Consumidora) ou os individuais (DIC - Duração de interrupção Individual por unidade Consumidora, FIC - Frequência de interrupção Individual por unidade Consumidora, $DMIC$ – Duração Máxima de Interrupção Contínua por unidade consumidora e $DICRI$– Duração da Interrupção individual ocorrida em dia Crítico por unidade consumidora). Quando os indicadores individuais de continuidade são violados, a distribuidora deve pagar uma multa para o consumidor, resultando em devoluções na ordem de milhões de reais diluídos nas contas dos consumidores, todos os anos. Já os indicadores coletivos, não são seguidos diretamente por multas quando violados, sendo cobrado a tomada de medidas e o acompanhamento das ações pela ANEEL. A aplicação de multas ocorre a partir do não comprimento das solicitações.

Conforme dados disponibilizados pela ANEEL, em 2017, foram pagos mais de cem milhões de multas, totalizando 492 milhões de reais gastos com as transgressões dos limites de continuidade pelas distribuidoras. Além disso, é previsto um crescimento médio anual do consumo de eletricidade de 3,7% ao ano até 2026, alcançando o patamar de 741 TWh demandados, contra 516 TWh em 2016 (EPE, 2017). Portanto, com a crescente demanda por energia elétrica e as altas quantias gastas com multas, as distribuidoras buscam um investimento estratégico na rede, de forma a repassar os custos para as tarifas, e o melhor serviço ao usuário, para reduzir os gastos devido às multas. Assim, para obter um melhor resultado com o mesmo investimento, é necessário saber a melhor forma de disponibilização das equipes e pontos críticos de atuação.

1.2 Indicadores de Continuidade do Serviço de Distribuição de Energia Elétrica

Os indicadores de continuidade do serviço são limites impostos pela ANEEL de forma a manter um padrão mínimo de qualidade. Esses podem ser segregados em dois grupos, os coletivos (DEC e FEC) e os individuais (DIC, FIC, $DMIC$ e $DICRI$) e são avaliados através de subdivisões das áreas atendidas pelas distribuidoras, denominadas conjuntos elétricos, onde cada conjunto elétrico pode ser maior ou menor que um município e são compostos por uma ou mais Subestações de Distribuição, conforme definido no módulo 8 do PRODIST.

Para a contabilização dos indicadores, as interrupções devem ser maiores que três minutos, sendo possível o expurgo dos indicadores dada determinadas condições, como:

i Interrupção de dia Crítico (grande quantidade de ocorrências emergenciais - Mod. 1 PRODIST);

ii Decreto de calamidade pública (Mod.8 PRODIST);

iii Interrupção em situação de emergência (Mod.8 PRODIST);

iv Esquema regional de alívio de carga (Mod.1 PRODIST)

1.2.1 Indicadores Individuais de Continuidade

Conforme a ANEEL (2016), os indicadores individuais apuram a qualidade do serviço em cada unidade consumidora e são compostos por:

i DIC: Intervalo de tempo que, no período de apuração, em cada unidade consumidora ou ponto de conexão, ocorreu descontinuidade da distribuição de energia elétrica com duração de $t(i)$ horas, definido por 1.1:

$$DIC = \sum_{i=1}^{n} t(i) \tag{1.1}$$

Sendo n a quantidade de interrupções no período avaliado.

ii FIC: Número n de interrupções ocorridas, no período de apuração, em cada unidade consumidora ou ponto de conexão, definido pela Equação 1.2:

$$FIC = n \tag{1.2}$$

iii $DMIC$: Tempo máximo de duração de uma interrupção contínua de energia elétrica, em uma unidade consumidora ou ponto de conexão, definido pela Equação 1.3:

$$DMIC = t(i)_{max} \tag{1.3}$$

iv $DICRI$: Corresponde à duração de cada interrupção ocorrida em dia crítico, para cada unidade consumidora ou ponto de conexão, definido pela Equação 1.4:

$$DICRI = t_(i)crítico \tag{1.4}$$

Os indicadores DIC e FIC têm seus limites estipulados para períodos mensais, trimestrais e anuais, já o $DMIC$ tem apenas os limites mensais. O $DICRI$ é estabelecido por interrupção em dia crítico, ou seja, independe de período (ANEEL, 2016).

1.2.2 Indicadores Coletivos de Continuidade

Os indicadores coletivos são utilizados pela ANEEL como uma forma de monitorar e avaliar a continuidade do serviço prestado, verificando se estão com um desempenho melhor ou pior que o editado e, caso o desempenho seja abaixo do estabelecido, requisitam, no primeiro momento, que a distribuidora proponha medidas para solucionar o problema.

Os Indicadores coletivos são:

i DEC: Duração Equivalente de Interrupção por Unidade Consumidora, definido pela Equação 1.5:

$$DEC = \frac{\sum_{i=1}^{Cc} DIC(i)}{Cc} \tag{1.5}$$

Sendo Cc a quantidade de clientes na região analisada, por exemplo, os clientes de um conjunto elétrico.

ii FEC: Frequência Equivalente de Interrupção por Unidade Consumidora, definido pela Equação 1.6:

$$FEC = \frac{\sum_{i=1}^{Cc} FIC(i)}{Cc} \tag{1.6}$$

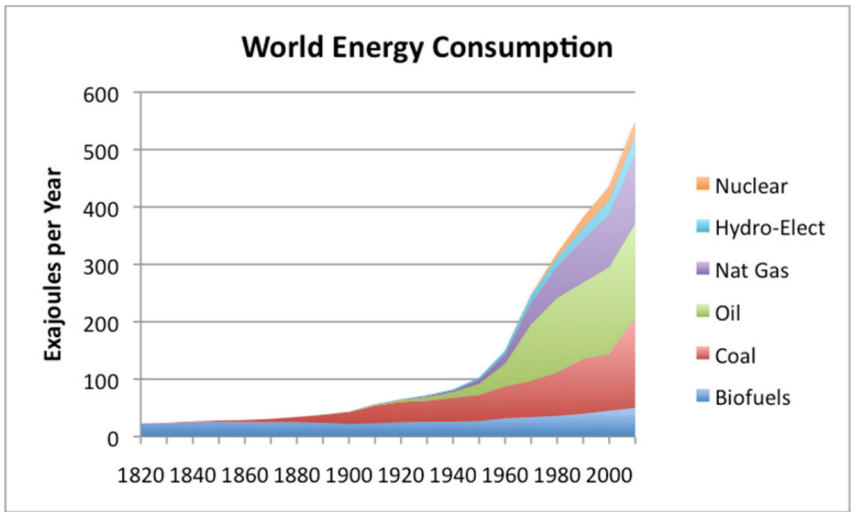
1.3 Justificativa

Desde os primórdios da humanidade criou-se a dependência pela energia, quando o Homo neanderthalensis começou a usar o fogo para a cocção de alimento (CARVALHO, 2014).

Muitos anos depois, a energia do vento passou a ser utilizada para a navegação, e com o passar dos séculos, a madeira passou a ser a principal fonte de combustível. Em seguida, no século XVIII, o carvão tomou a frente devido ao surgimento das máquinas a vapor.

A energia permeia diversos meios, sendo eles sociais, econômicos, climáticos, além de possuir diversas fontes, como o carvão mineral, derivados do petróleo, madeira, energia potencial da água, etc. O consumo de energia cresce ao longo dos anos, conforme o gráfico apresentado por Karjalainen et al. (2014)

Figura 2 – Consumo mundial de energia de 1820-2010



Fonte: Karjalainen et al. (2014)

O estudo da relação entre o consumo de energia e o crescimento da economia é alvo de diversos estudos. Sendo possível observar, para os países integrantes do G-7, que a estagnação do consumo de energia pode prejudicar o crescimento econômico, visto que

existe causalidade no aumento do consumo com o crescimento da economia (SOYTAS; SARI, 2003). Sendo que países com alta renda, em sua maioria, possuem de 95%-100% de sua população com acesso a energia elétrica (RITCHIE HANNA, 2018).

Diante acentuado crescimento do consumo de energia, vem a necessidade da implementação de formas mais fáceis de obtê-la a um baixo custo, com facilidade para transporte e amplas formas de transformação, podendo assim ser utilizada em diversas frentes.

Deste modo, a energia elétrica vem ganhando força sobre os outros tipos de energia intermediária por ser facilmente transformada em outras formas, como a térmica e mecânica, por exemplo, gradualmente substituindo os derivados do petróleo na indústria automotiva. Além da larga aplicação para equipamentos que necessitam de processamento, como em sistemas bancários, automação industrial, máquinas hospitalares, semáforos, etc.

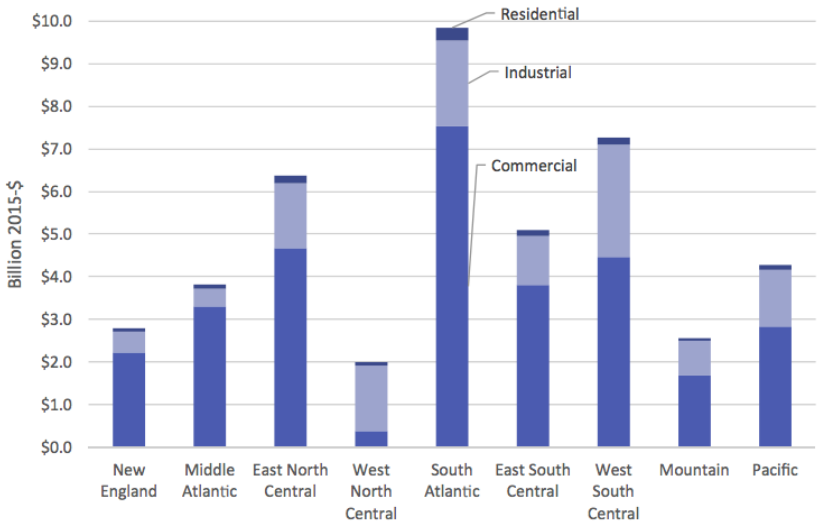
Frente a dependência da energia elétrica, são feitos investimentos para a melhora do fornecimento, entretanto, tais investimentos são de alto custo, principalmente devido às dimensões territoriais do Brasil, onde em 2016 aproximadamente 13,8 Bilhões de reais foram investidos na compra de novos equipamentos, treinamento de pessoas, conscientização da população sobre cuidados com a rede elétrica, expansão da rede, atendimento aos consumidores, combate a furtos e fraudes, dentre outros (ABRAADE, 2017).

Além disso, o estudo realizado por LaCommare et al. (2018), estimou o custo das falhas no fornecimento de energia para os setores comerciais, industriais e residenciais, conforme a Figura 3, sendo possível observar as interrupções geram um custo para os consumidores de bilhões de dólares por ano nas diversas regiões dos Estados Unidos da América (EUA).

O grande crescimento do uso da energia tem demandado altos índices de consumo que também acarretam em altos valores financeiros decorrentes das multas e obrigações referentes aos indicadores de qualidade de energia. Nesse sentido, técnicas de engenharia que busquem modelar e prevenir a ocorrência de falhas e/ou melhorar as estratégias de logística aplicada na alocação de equipes de manutenção preventiva/corretiva são de grande interesse para as concessionárias de distribuição de energia elétrica, bem como para os clientes atendidos.

A elaboração de técnicas de inteligência atreladas à nova modalidade de Indústria 4.0 tem permitido melhorar a qualidade dos serviços reduzindo os indicadores de qualidade ou mesmo prevendo seus valores, de forma a antecipar as ações das equipes envolvidas no processo de tratamento das falhas decorrentes do uso das redes de energia elétrica.

Figura 3 – Custo, em dólares, devido as falhas no fornecimento de energia para as regiões dos EUA



Fonte: LaCommare et al. (2018)

Justifica-se, assim, a implementação de redes neurais para realização da predição dos indicadores DEC/FEC, para que seja possível a realização de investimentos pontuais, de forma que antes que os indicadores coletivos e individuais cresçam, medidas possam ser tomadas para sua prevenção. Aumentando a eficiência dos investimentos, e contribuindo para uma melhoria não só para a empresa, mas também para os clientes, que terão menor número e duração das interrupções.

1.4 Objetivos

1.4.1 Objetivo Geral

O objetivo geral deste trabalho é avaliar diferentes aspectos relacionados à previsão dos indicadores coletivos de continuidade, de uma distribuidora de energia, a partir do estudo de previsão iniciado por uma pesquisa de mestrado do PPGEE.

1.4.2 Objetivos Específicos

- Avaliar o uso de método de aprendizagem profunda aplicada aos mesmos dados da pesquisa para fins de comparação com as redes neurais recorrentes (*Recurrent Neural Network - RNN*) utilizadas;

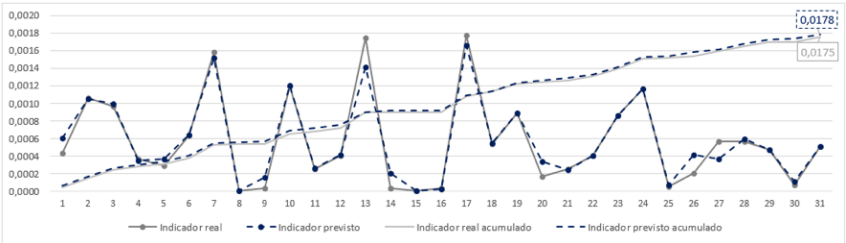
- Avaliar dados exógenos que possam afetar as medidas de DEC/FEC na contribuição de sua previsão nos moldes apresentados na pesquisa.
- Testar modelo de predição baseado em rede de aprendizado profundo com camada LSTM.

1.5 Trabalho de Referência

Esse projeto vista a continuação da pesquisa desenvolvida na dissertação de mestrado feita por Louback (2018), o qual realizou a previsão dos indicadores de continuidade DEC/FEC por meio de uma rede neural não linear autorregressiva, sendo escolhida a melhor dentre as variações dos parâmetros da rede, como numero de neurônios e interpolações. Assim, para obter uma referência adequada para a comparação dos resultados, foram utilizados os mesmos parâmetros para treinar a rede NAR.

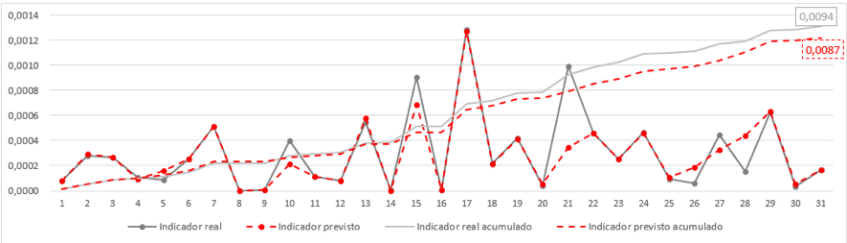
Nesta dissertação, foi também demonstrada a importância do tratamento de *outliers*, bem como a forte correlação entre os dias da semana, obtendo erros na previsão para janeiro de 1,71% para o DEC acumulado e -7,45% para o FEC acumulado, conforme as Figuras 4 e 5.

Figura 4 – Comparação entre DEC previsto através da rede NAR



Fonte: Louback (2018)

Figura 5 – Comparação entre FEC previsto através da rede NAR



Fonte: Louback (2018)

1.6 Organização do Trabalho

Esse trabalho foi divido em 5 capítulos, sendo eles introdução, embasamento teórico, metodologia e etapas de desenvolvimento, resultados e considerações finais, conclusões e projetos futuros e referências bibliográficas.

No primeiro capítulo é introduzido a importância da anergia elétrica no dia-a-dia, os conceitos relacionados indicadores de continuidade do fornecimento de energia, bem como os objetivos deste trabalho.

No segundo capitulo são introduzidos os conceitos relacionados a séries temporais, tratamento de dados e aprendizado de máquina, onde foi feito o estudo de redes do tipo NAR, NARX e de aprendizado profundo com camada LSTM.

No terceiro capitulo são apresentadas as arquiteturas das redes, seus parâmetros de treinamento e exemplificado como são dadas as saídas das redes, para realizar a comparação entre elas.

O quarto capítulo aborda os resultados dos treinamentos para o mês de janeiro de 2016, comparando as redes propostas em questão de indicador acumulado mensal, erro para cada dia e para as previsões das subséries de frequência.

No quinto capítulo são exportas as conclusões obtidas por meio deste trabalho juntamente com propostas para trabalhos futuros.

O sexto capítulo é destinado a apresentação de todas referências bibliográficas utilizadas no trabalho.

2 EMBASAMENTO TEÓRICO

2.1 Introdução

Este capítulo tem por finalidade estabelecer os conceitos teóricos necessários usados no trabalho. Portanto, o capítulo inicia-se com uma explicação conceitual sobre as técnicas de tratamento de dados utilizadas.

Depois passa-se a explicar sobre o que são Redes Neurais, quais os tipos utilizados e como funcionam, limitações em função do hardware utilizado e estado da arte.

2.2 Séries Temporais

Uma série temporal é qualquer conjunto de observações ordenadas no tempo (MORETTIN; TOLOI, 2006), podendo ser contínua ou discreta. As séries temporais possuem diferentes aplicações, como:

i Investigação do mecanismo gerador da série temporal;

ii Previsões de valores futuros da série;

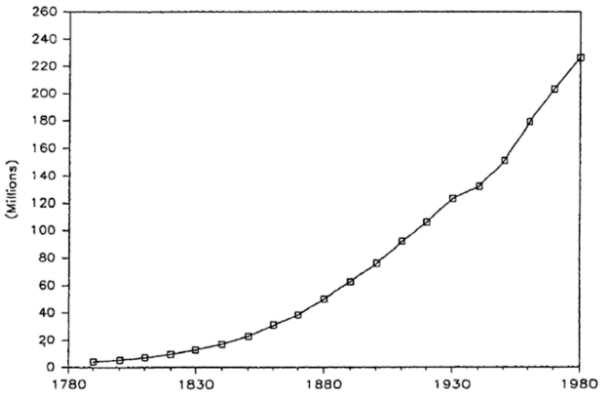
iii Descrição de seu comportamento;

iv Investigação de propriedades relevantes nos dados.

Apesar do nome “temporal”, a série não necessariamente deve variar apenas em função do tempo, podendo ter algum outro parâmetro físico.

Um exemplo de série temporal pode ser formulado pelos dados de crescimento da população norte americana entre o período de 1780 a 1980, indicado pela Figura 6, onde cada ponto é uma amostra da população, formando uma série temporal discreta.

Figura 6 – Crescimento da População Norte Americana

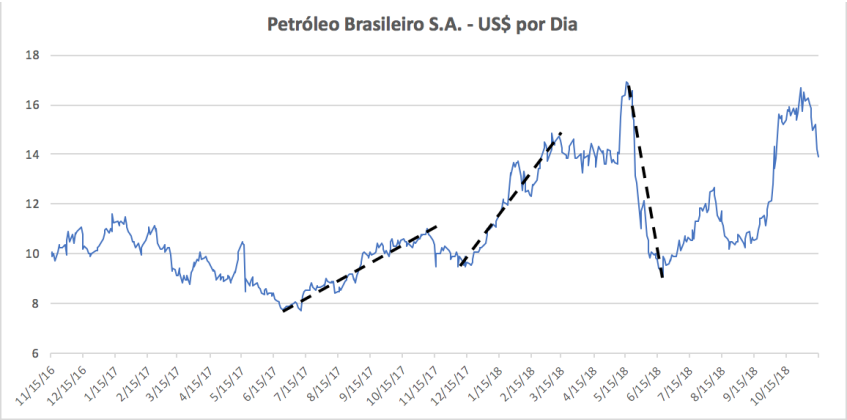


Fonte: (BROCKWELL; DAVIS, 2013)

2.3 Decomposição de Séries Temporais

Em determinados intervalos de tempo, as séries temporais apresentam tendências, seja a de subir ou decrescer, fator comumente observado em dados financeiros, como na Figura 7 onde foram destacadas algumas tendências observadas.

Figura 7 – Variação do valor das ações da Petrobras ao longo do tempo



Fonte: Produção do autor

Assim, para melhor analisar a série, é necessário quebrá-la em segmentos homogêneos, de forma a reduzir descontinuidades, como mudanças súbitas de nível devido a sazonalidades (BROCKWELL; DAVIS, 2013) Para uma série temporal Z_t, um modelo de decomposição consiste em escrever essa série em função de três componentes não observáveis,

$$Z_t = T_t + S_t + a_t \tag{2.1}$$

em que T_t representa a tendência e S_t a sazonalidade, enquanto a_t é uma componente aleatória, de média zero e variância constante (BROCKWELL; DAVIS, 2013). O principal interesse em considerar o modelo acima, é de construir a série livre de sazonalidades.

2.4 Tratamento de dados

Todo grupo de dados tem desafios e oportunidades únicas, de modo que a análise não pode ser reduzida a seguir uma receita (OSBORNE, 2012). O tratamento é um passo primordial, visto que caso não seja realizado, ou mal executado, pode atrapalhar o treinamento e predição do modelo (KUHN; JOHNSON, 2013). Assim, é necessário atentar para que as alterações não insiram tendências no modelo (MALLEY; RAMAZZOTTI; WU, 2016). De maneira genérica, essas etapa refere-se a adição, deleção ou transformação dos dados de treinamento.

O objetivo dessas mudanças é fazer com que a informação que será utilizada para treinamento do modelo esteja de forma adequada, contribuindo para o aprendizado da rede. Nas seções seguintes, foram descritas as transformações estudadas.

2.4.1 Normalização

A normalização é aplicada para que os valores da série sejam escalados para um intervalo específico, geralmente 0-1. Assim, é mais prática a comparação entre os dados que possuem diferentes magnitudes (MALLEY; RAMAZZOTTI; WU, 2016).

Outra forma de normalização é conhecida como Z-Score. Esse método consiste em, a partir da média e do desvio padrão da série, fazer com o que dado seja centralizado de forma a obter média zero e escalado para obter o desvio padrão de um.

2.4.2 Agregação

Quando se possui múltiplos valores do mesmo atributo e o foco do estudo utiliza uma visão macro, esses dados são agregados para formar um único valor (MALLEY; RAMAZZOTTI; WU, 2016). Uma aplicação comum da agregação é para obter informações de grupos ao invés de indivíduos, como a união dos indicadores a nível de ordem de interrupção de energia de forma a obter os indicadores coletivos de um município.

2.4.3 Limpeza dos dados

A limpeza de dados lida com a detecção e remoção de erros e inconsistências, bem como *outliers* e ruídos, de forma a melhorar a qualidade dos dados. Nessa projeto será abordado o tratamento de dados faltantes e de *outliers*.

O primeiro, afeta uma grande quantidade de dados e, para efetuar a realização das análises a manipulação dessas informações faltantes é indispensável. Para manipular esses dados, é possível deletar as informações incompletas ou, através do processo conhecido como *inputation*, substituir os valores faltantes a partir do comportamento anterior ou de outras variáveis, seja de maneira manual ou não (MALLEY; RAMAZZOTTI; WU, 2016).

Por meio do tratamento de *outliers*, ou suavização dos dados, é possível extrair os ruídos e destacar o padrão básico de comportamento. Sendo de grande importância, tendo em vista que a previsão de valores é realizada a partir das observações passadas, tendo em vista que estas informações possuem um padrão de comportamento que pode ser de difícil observação com a presença de *outliers* (MORETTIN; TOLOI, 2006).

Esses, são uma forma de anormalidade/discordância do comportamento usual. A partir disso, um dos métodos mais simples de identificação desses, consiste em uma análise unidimensional, nas quais se o valor é muito grande ou muito pequeno em relação aos demais na sua vizinhança, ele é considerado um *outlier* (SALGADO et al., 2016).

Assim, a partir de um *threshold*, comumente definido por um múltiplo do desvio padrão, os *outliers* são removidos e substituídos por um valor interpolado, como o Polinômio interpolador de hermite cúbica por partes (*Piecewise Cubic Hermite Interpolating Polynomial* - PCHIP) (LOUBACK, 2018).

Entretanto, os *outliers* podem representar valores de ocorrência natural, devido a variações do processo e/ou da população observada, podendo assim, conter informações relevantes sobre comportamentos mais sucintos (SALGADO et al., 2016).

Além da remoção de *outliers*, para facilitar o aprendizado a rede, pode ser realizada a interpolação para tornar as variações mais suaves, como a utilização de *splines*, que retorna a quantidade de pontos desejada entre os dados a partir de uma função.

Dessa forma, é preciso analisar a necessidade, pois de acordo com a realização do tratamento desses valores, considerando mais eventos inesperados, os algoritmos de predição requeridos para aprender esses padrões mais tênues vão se tornando mais complexos.

2.5 Transformada Wavelet de Máxima Sobreposição

A Transformada de Fourier (TF) é uma ferramenta muito utilizada para análise das componentes de frequência de um sinal (LIU, 2010). Ela, ao ser aplicada em um sinal no domínio do tempo, $g(t)$, quebra o sinal em infinitas somas de funções senos e cossenos, sendo definida pela Equação (2.2).

$$L\{g(t)\} = G(s) = \int_{-\infty}^{+\infty} e^{-i\omega t} dt \tag{2.2}$$

Sendo w a frequência angular, dada por $w = 2\pi f$, e $G(s)$ o sinal resultante após a aplicação da TF.

Uma forma mais explicita de observar a soma de senos e cossenos é analisando a Série de Fourier (2.3), que representa um caso da transformada aplicada a sinais periódicos, $x(t)$, onde k representa os múltiplos da frequência natural e a_k é conhecido como coeficiente de Fourier, podendo ser definido pela Equação (2.4), sendo T o período do sinal.

$$R(t) = \sum_{n=-\infty}^{\infty} a_k \exp\left(j\frac{2n\pi t}{T}\right) \tag{2.3}$$

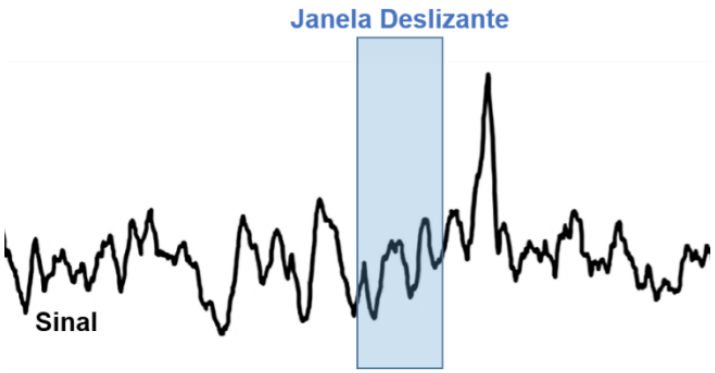
$$a_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t) \exp\left(-j\frac{2n\pi t}{T}\right) dt \tag{2.4}$$

Apesar da TF apresentar uma boa forma de visualização dos dados na frequência, ela não carrega consigo dados do tempo, como o momento que aquela frequência ocorre. Para obter informações do tempo, em conjunto com a frequência, há a Transformada de Fourier de Curto Período (*Short-time Fourier Transform* - STFT) que desloca uma janela no tempo, Figura 8, para obter conhecimento sobre ambos os domínios (LIU, 2010), porém limitado ao tamanho da janela representada por $w(t-\tau)$ na Equação (2.5), centrada em τ.

$$[STFT]\{g(t)\} = G(\tau, s) = \int_{-\infty}^{+\infty} g(t) e^{-i\omega t} w(t-\tau)\, dt \tag{2.5}$$

Assim, conforme Diniz (2018), para auxiliar na análise de sinais, buscando solucionar a limitação da STFT, foi desenvolvida a Transformada Wavelet (TW), definida pela Equação

Figura 8 – Janela deslizante no tempo



Fonte: Diniz (2018)

(2.6), possibilitando a decomposição do sinal em diferentes frequências.

$$[TW](a,b) = \int_{-\infty}^{+\infty} x(t)\psi_{a,b}(t)\,dt \tag{2.6}$$

sendo a o parâmetro de escala, b de translação e $\psi_{a,b}(t)$ a função wavelet base, $\psi(t)$ escalada e deslocada conforme a Equação (2.7):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2.7}$$

Para trabalhar com sinais discretos, é utilizada a Transformada de Wavelet Discreta (DWT - *Discrete Wavelet Transform*), que pode ser expressa por

$$Wf(a,b) = \sum_{n=0}^{N-1} f(n) \cdot \psi^*_{a,b}(n) \tag{2.8}$$

sendo que $Wf(a,b)$ é o coeficiente da DWT, $f(n)$ é o sinal discreto de tamanho N e a base discretizada é

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a_0^j|}}\psi\left(\frac{t - rb_0a_0^j}{a_0^j}\right) \tag{2.9}$$

onde j e r são números inteiros, correspondentes aos parâmetros de dilatação e translação, respectivamente, $a_0 > 1$ é um parâmetro de dilatação fixo, que reflete a variação da dilação,

e b_0 é o fator de translação (passo da translação) o qual depende do fator de dilatação (DINIZ, 2018).

Para utilização no processamento de sinais, o algoritmo da DWT sofre uma sensibilidade à escolha do ponto inicial, ou seja, possui variância no tempo (PERCIVAL; WALDEN, 2000). Devido a essa limitação, foram formulados os algoritmos de Transformadas Wavelets não-decimadas e invariantes no tempo, como o *Maximal Overlap Discrete Wavelet Transform* (MODWT) (DINIZ, 2018).

A MODWT é uma transformada não ortogonal e redundante, ou seja, funcionando como um filtro, capaz de decompor uma série temporal em subséries. Essa transformada possui a capacidade de reter baixas frequências, que outrora seriam descartadas, como no caso da DWT, de modo que a partir da somas dos sinais consegue-se reconstruir o original (CORNISH; BRETHERTON; PERCIVAL, 2006).

Assim, pela característica de decompor uma série mais complexa em outras mais simples, conforme exposto por Diniz (2018), a MODWT é muito utilizada como entradas de modelos de predições.

2.6 Aprendizado de Máquina

A habilidade humana de adquirir conhecimento a partir da experiência e generalizar o aprendizado, aplicando o que foi aprendido em um contexto para uma situação diferente, tem fascinado os cientistas. Afinal, identificar como o conhecimento é adquirido, leva a formas mais eficientes de ensinar (DOUMAS; PUEBLA; MARTIN, 2018).

Com esse fascínio, em 1943, o neurofisiologista McCulloch e o matemático Pitts publicaram um artigo onde modelaram uma rede neural simples utilizando circuitos elétricos. A partir de então se busca criar/reproduzir a maneira como os neurônios funcionam. Dando início a um campo que futuramente, de maneira mais genérica, ficou conhecido como "Aprendizado de Máquinas".

O aprendizado de máquinas busca formas de construir programas que aprendem com a experiência, ou seja, com a habilidade de reconhecer padrões sem depender de codificação diretamente humana para isso (MITCHELL, 1997).

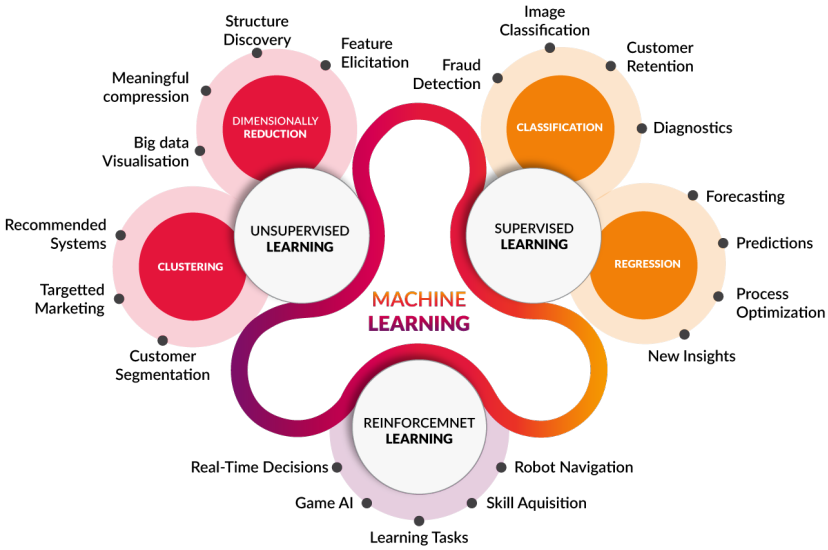
Os algoritmos "aprendem" a partir de tentativas e verificações, assim, com um objetivo definido, são feitos constantes testes de desempenho e, caso a tentativa tenha sido satis-

fatória, é recompensado de forma a buscar naquele entorno. Como definido por Mitchell (1997), "Pode-se dizer que um programa de computador aprende com a experiência **E** a respeito de alguma tarefa **T** e desempenho medido **P**, se ele obtém um resultado em uma tarefa **T** como medido em **P** e evolui com a experiência **E**".

Essa metodologia é empregada em diversas áreas com obtenção de bons resultados, por exemplo, reconhecimento de fala, de imagens, pontos de falha de equipamentos, extração de informações relevantes de dados comerciais, pré-diagnósticos médicos e também para tomadas de decisões, como compra e venda de ações e carros autônomos (MITCHELL, 1997).

Os tipos de algoritmos para aprendizado de máquinas podem ser classificados em três categorias, como na Figura 9.

Figura 9 – Tipos de algoritmos de aprendizado de máquina e suas aplicações



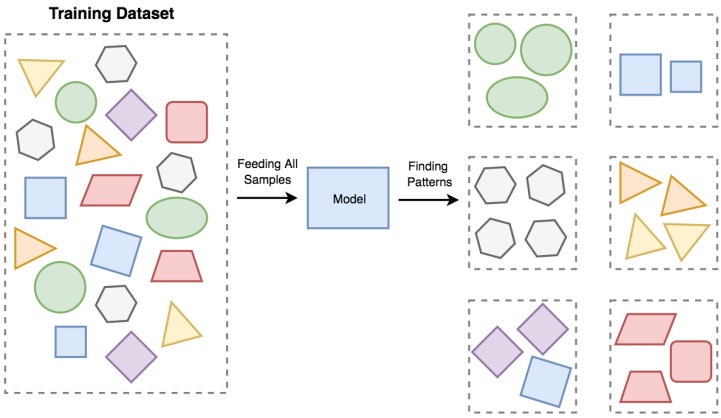
Fonte: Krzyk (2018)

O aprendizado supervisionado, diz respeito a grupo de algoritmos precisa de dados que venham acompanhados de pares entrada-saída, dai a origem da nomenclatura "supervisionado". Por exemplo, para treinar uma rede neural a classificar números em imagens, é necessário que os dados para treinamento venham identificados (GOODFELLOW; BENGIO; COURVILLE, 2016).

Dessa forma, esses algoritmos buscam encontrar parâmetros que permitam um bom desempenho dos dados de treinamento e também nos dados de teste, para garantir que o modelo não está viciado, podendo assim, classificar/estimar uma entrada desconhecida.

O aprendizado não supervisionado, também conhecido por análise de *cluster*, não possui uma resposta certa, ele foi desenvolvido para extrair informações, descobrindo padrões e como os dados estão estruturados de modo a codificá-las na saída, conforme indicado na Figura 10.

Figura 10 – Representação de um aprendizado não supervisionado



Fonte: Krzyk (2018)

Por último, a técnica de aprendizado por reforço busca encontrar a melhor ação a ser tomada, dentre determinadas possibilidades ou regras, a partir de dados do ambiente (BISHOP, 2006).

Assim, por meio de tentativa e erro, cada vez que o algoritmo toma uma decisão ele é recompensado ou punido, guiando-o na direção desejada.

2.6.1 Redes Neurais Artificiais

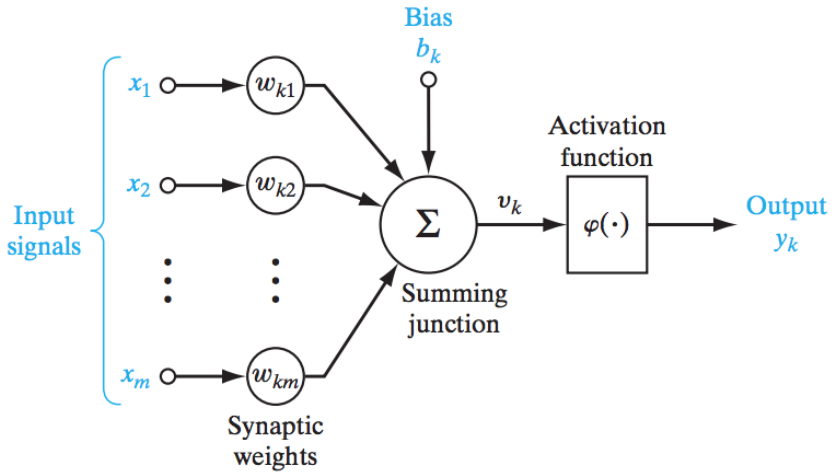
A aplicação de Redes neurais artificiais é uma alternativa para a solução de problemas que não possuem uma solução matemática fechada, tendo como inspiração o comportamento do cérebro humano, ou seja, com alta complexidade, não linearidade e processamento em paralelo (HAYKIN et al., 2009).

Segundo Haykin et al. (2009), da mesma forma que o cérebro, uma rede neural também possui o neurônio como um elemento essencial para o seu funcionamento. Como na Figura 11, um neurônio consiste dos seguinte elementos:

- **Sinapses:** Nas sinapses, as entradas (x_m) são multiplicadas por um peso (w_k), individual, podendo ser um valor tanto positivo quanto negativo.

- **Somador:** O somador, como o próprio nome indica, realiza uma adição das entradas após a etapa anterior. Uma outra entrada dessa etapa é o *bias* (b_k), ou viés, que tem como objetivo aumentar ou diminuir a entrada da função de ativação.
- **Função de ativação:** Tem por finalidade limitar a amplitude do sinal de saída.

Figura 11 – Modelo de um neurônio



Fonte: Haykin et al. (2009)

Os sinais de saída podem ser descritos como na Equação (2.11)

$$y_k = \varphi(u_k + b_k) \tag{2.10}$$

sendo b_k o *bias* e u_k o sinal de entrada representado pela Equação (2.11), para um número m de entradas.

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{2.11}$$

Os neurônios ocultos, possuem o mesmo modelo dado pela Figura 11, se localizam entre os neurônios de entrada e os de saída, podendo formar uma ou varias camadas. São utilizados principalmente devido a sua habilidade de encontrar padrões através de suas características não lineares. Dessa forma, ao acrescentar camadas ocultas a rede, mais padrões são encontrados (HAYKIN et al., 2009).

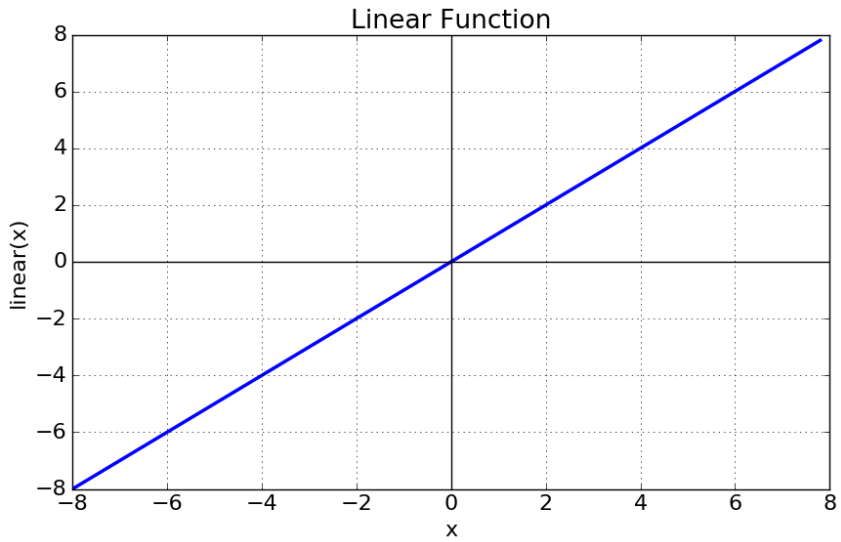
A função de ativação, atribuída anteriormente como φ, possui diversas formas, tendo a escolha de acordo com a necessidade da aplicação. Neste estudo foram utilizadas três funções de ativação, sendo elas a linear, a sigmoide e a tangente hiperbólica.

1. **Função Linear:** Nessa função a saída é diretamente proporcional a entrada, como na Figura 12, podendo ser definida pela Equação (2.12), onde v_k é a saída do somador adicionado ao sinal de *bias*.

$$\varphi_v = cv_k \tag{2.12}$$

sendo c o coeficiente de inclinação da reta.

Figura 12 – Função de ativação linear



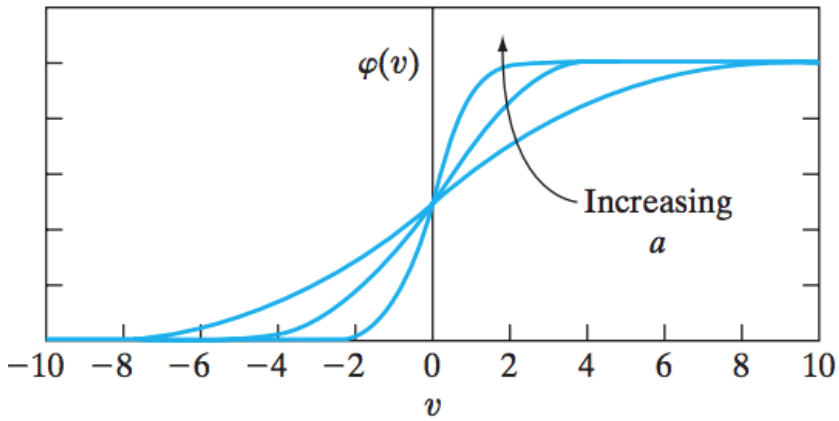
Fonte: Sharma (2017)

2. **Função sigmoide:** Conforme Haykin et al. (2009), a função sigmoide mantém o equilíbrio entre a linearidade e a não-linearidade, esta é uma das funções de ativação mais usadas, com saída limitada ao intervalo [0,1], definida pela Equação (2.13),

$$f(x) = \frac{1}{1 + e^{-av_k}} \tag{2.13}$$

com sinal demonstrado pela Figura 13, onde a é o parâmetro de inclinação.

Figura 13 – Função de ativação sigmoide

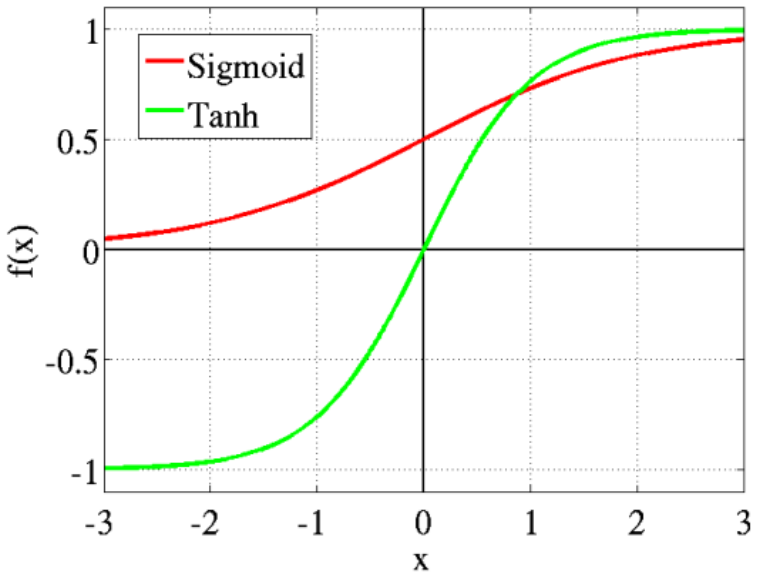


Fonte: (HAYKIN et al., 2009)

3. **Função tangente hiperbólica:** Para possibilitar que o intervalo de saída tenha valores negativos [-1,1] é utilizada a função tangente hiperbólica, que possui as mesmas características, exceto a amplitude, representada pela Equação (2.14), podendo ser observada na Figura 14, pelo sinal em verde, em comparação com a função sigmoide dada em vermelho.

$$f(x) = \frac{1 - e^{-av_k}}{1 + e^{-av_k}} \tag{2.14}$$

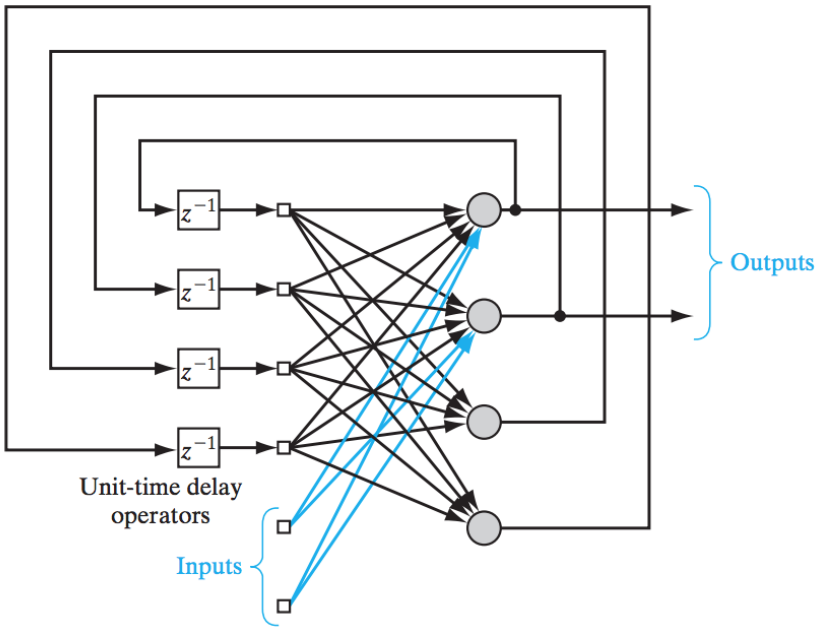
Figura 14 – Função de ativação sigmoide e tangente hiperbólica



Fonte: Sharma (2017)

Dentre as arquiteturas de redes neurais, as redes recorrentes, diferente da estrutura de *feedfoward*, possuem ao menos uma realimentação, ou seja, a saída de um neurônio é realimentada para a entrada de todos os neurônios. Dessa forma, os neurônios não ocultos terão como entrada os dados passados como *input* e as realimentações, enquanto os neurônios ocultos disporão apenas das realimentações (HAYKIN et al., 2009), como exemplificado na Figura 15.

Figura 15 – Rede recorrente com neurônios ocultos

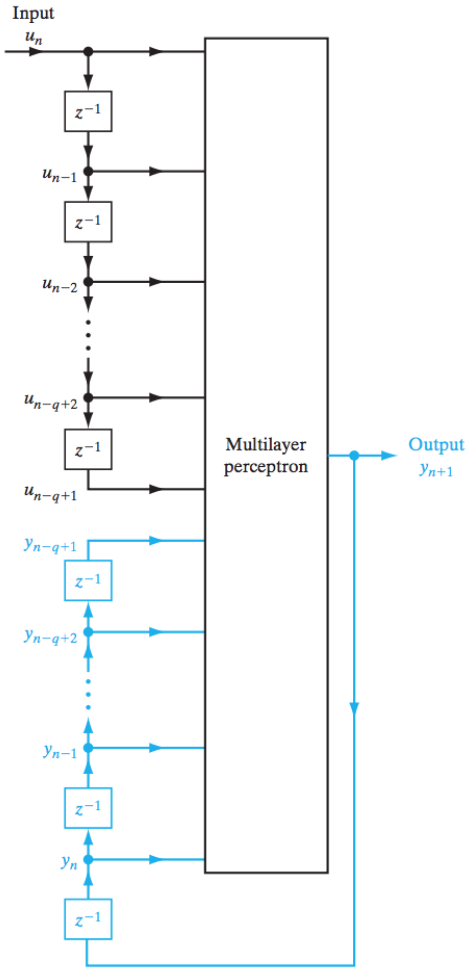


Fonte: Haykin et al. (2009)

Uma variação das redes recorrentes, são os autorregressivo com entradas exógenas. Este modelo possui uma realimentação, como explicado, onde a saída é também é utilizada como entrada. A diferença é que a entrada é passada para uma camada perceptron. Como na Figura 16, a saída está uma unidade de tempo a frente da entrada, ou seja, a camada de entrada da camada perceptron consiste no histórico de valores da entrada $(u_n, u_{n-1}, ..., u_{n-1+q})$ e saída $(y_n, y_{n-1}, ..., y_{n-1+q})$, obtendo a saída y_{n+1} (HAYKIN et al., 2009). Por exemplo, sendo um horizonte de previsão de um, apenas y_{n+1} corresponderá ao valor previsto e o restante aos valores ocorridos, além disso, u, representa uma entrada exógena diferente de y.

Esse modelo é conhecido como autorregressivo não linear de entradas exógenas (NARX - *Nonlinear autoregressive with exogenous inputs*). Exógeno devido ao dados passados na entrada, regressivo em virtude da realimentação e não linear pois a configuração da camada perceptron permite uma não linearidade.

Figura 16 – Modelo autorregressivo não linear de entradas exógenas

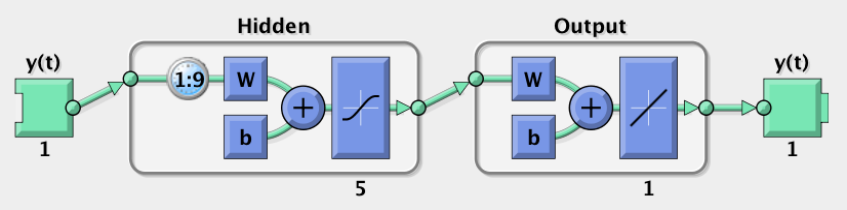


Fonte: Haykin et al. (2009)

Em virtude da NARX possibilitar a utilização do histórico do real valor da série temporal e de dados complementares, ela é muito utilizada para modelagem de séries temporais (BOUSSAADA et al., 2018).

Como uma subdivisão da NARX, existe a rede NAR, que não possui entradas exógenas (Figura 17), sendo uma rede mais simples, logo, desde que o dado de entrada não seja muito complexo, é mais rápida de treinar.

Figura 17 – Modelo autorregressivo não linear



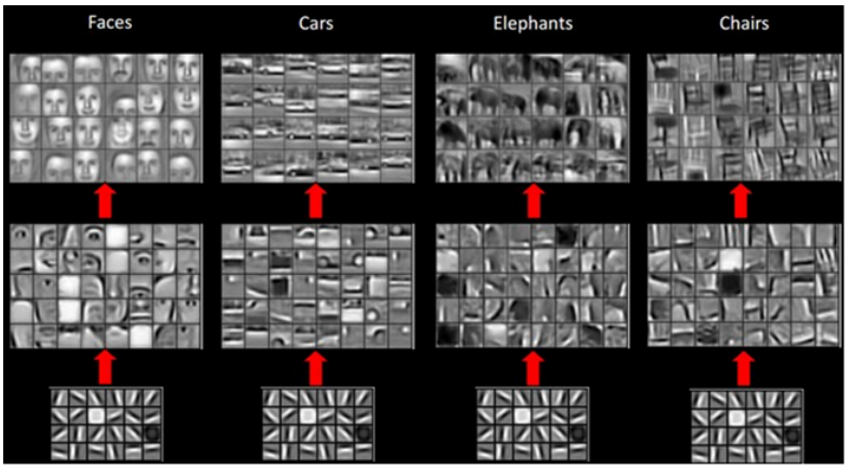
Fonte: Produção do Autor

2.6.2 Aprendizagem Profunda

Um grande problema relacionado ao aprendizado de máquinas é a quantidade de fatores que podem influenciar a entrada, como no exemplo dado por Goodfellow, Bengio e Courville (2016), onde os *pixels* de um cartão mudam o tom conforme a exposição de luz e a forma é alterada de acordo com o ângulo da foto. Logo, é comum realizar um tratamento dos dados para buscar padronizar a entrada.

Entretanto, os algoritmos de aprendizado profundo conseguem resolver este problema a partir de representações mais simples, chegando a conceitos mais complexos, como na Figura 18, onde primeiro são extraídas as bordas, depois as formas e, enfim, a classificação final.

Figura 18 – Modelo autorregressivo não linear de entradas exógenas

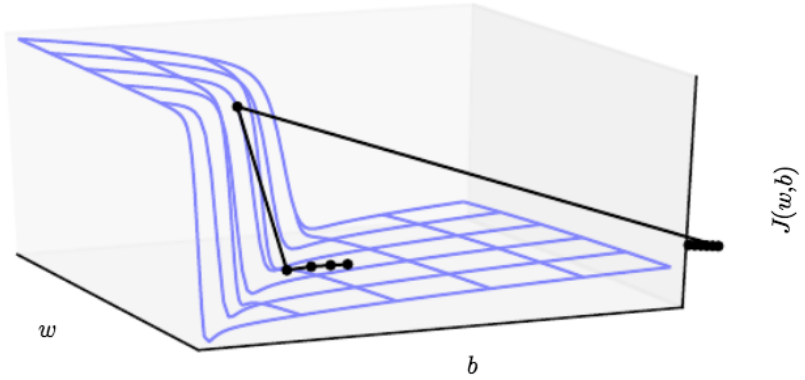


Fonte: Sqalli (2016)

De acordo com a profundidade da rede, informações mais complexas podem ser aferidas, pois as instruções mais a frente podem utilizar os dados extraídos pelas camadas anteriores (GOODFELLOW; BENGIO; COURVILLE, 2016).

Porém, segundo Goodfellow, Bengio e Courville (2016), algumas situações que prejudicam o aprendizado da rede são comuns quando essas possuem muitas camadas. Dentre elas, temos as regiões com inclinação muito elevada, o que pode levar a um gradiente grande, ou como é conhecido, gradiente explosivo. Além disso, também é comum a presença de regiões planas, fazendo com que o gradiente se torne desprezível. Os dois casos podem ser observados na Figura 19.

Figura 19 – Gradiente em declive acentuado vs plano



Fonte: Goodfellow, Bengio e Courville (2016)

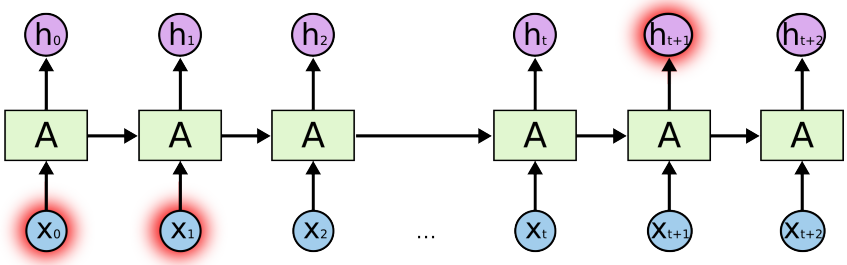
O erro, em redes com retropropagação (*backpropagation*) no tempo, é retroalimentado e tende a explodir ou desaparecer. Isso ocorre pois, conforme demonstrado por Hochreiter e Schmidhuber (1997), a forma como essas redes comumente aprendem faz com que o erro seja escalado de forma a crescer ou decrescer exponencialmente. No primeiro caso, gera uma oscilação nos pesos, fazendo com que a rede tenha um aprendizado instável. Já para o decréscimo, o erro rapidamente desaparece, assim, a rede não aprende em um intervalo de tempo aceitável.

Dessa forma, por apresentar dificuldades em utilizar informações passadas no instante necessário (Figura 20), em 1997 foi apresentada a proposta da camada *Long Short-Term Memory* (LSTM) de modo a proporcionar a passagem do erro sem os problemas anteriores. Nesse estudo, foi introduzido o conceito de célula de memória, permitindo às unidades conter informações fora do fluxo, assim dados mais antigos podem ser utilizados mesmo vários passos depois ou esquecidos, caso não sejam relevantes para o aprendizado, com estrutura exemplificada pela Figura 21.

Dentro de um bloco/célula LSTM, Figura 22, existem três *gates* responsáveis por proporcionar as habilidades de uma camada LSTM, sendo eles o *gates* de entrada (*input gate*), de esquecimento (*forget gate*) e o de saída (*output gate*).

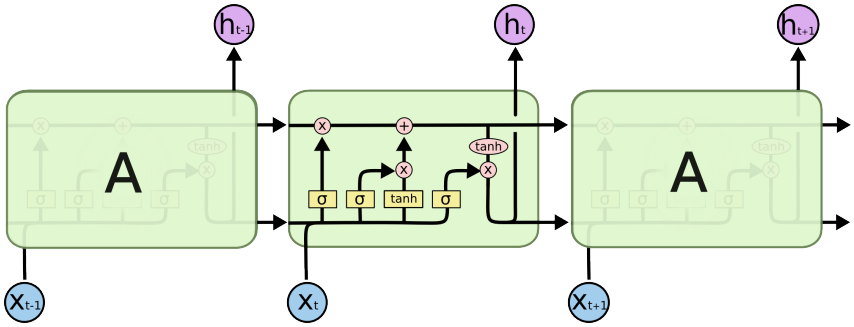
Segundo Cui, Ke e Wang (2018) e tomando a Figura 22 como referência visual, temos os

Figura 20 – Redes com retropropagação vs dependências antigas



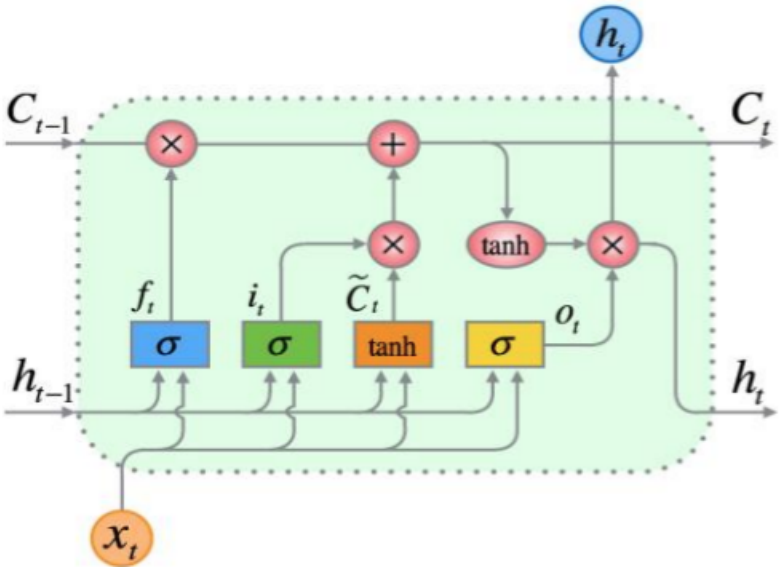
Fonte: Olah (2015)

Figura 21 – Camada LSTM



Fonte: Cui, Ke e Wang (2018)

Figura 22 – Bloco/Célula LSTM



Fonte: Cui, Ke e Wang (2018)

gates como:

1. ***Forget Gate***: Utilizado para decidir que informação deve ser descartada, esquecida,

a partir de h_{t-1} e x_t, resultando em uma saída f_t no intervalo de $[0, 1]$, onde zero representa esquecer totalmente e um indica que C_{t-1} deve ser integralmente armazenada, defino por (2.15).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.15}$$

sendo σ a função de ativação, W e b as matrizes de pesos e *bias*, respectivamente.

2. ***Input Gate***: Nesta etapa é definido o que será atualizado, através de i_t (2.16), e o valor das atualizações, $\tilde{C}_t$ (2.17). Assim, o valor a ser adicionado é dado por $\tilde{C}_t \cdot i_t$, representando o vetor de entrada escalado.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_i] + b_i) \tag{2.16}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_i] + b_C) \tag{2.17}$$

tendo *tanh* como função de ativação.

3. ***Output Gate***: Para composição da saída, primeiro é importante definir C_t. C_t representa o estado da célula, sendo constituído a partir do estado antigo C_{t-1}, onde ao multiplicar por f_t os valores decididos são esquecidos. Em seguida, os valores novos ($\tilde{C}_t \cdot i_t$) são adicionados ao estado antigo, obtendo assim, C_t, como definido em (2.18)

$$C_t = C_{t-1} \cdot f_t + i_t \cdot \tilde{C}_t \tag{2.18}$$

Após aplicar a função de ativação em h_{t-1} e x_t, para definir qual parte no novo estado (C_t) irá ser utilizado na saída, por meio de o_t (2.19), C_t é aplicado a uma função *tanh* para que seu intervalo varia de $[-1, 1]$. Assim, ao multiplicar os dois é obtida a saída h_t, como definido em (2.20)

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.19}$$

$$h_t = o_t \cdot tanh(C_t) \tag{2.20}$$

2.6.3 Otimizadores

Como os pesos sinápticos são responsáveis por parte de como a rede lida com as entradas, é necessário ajustá-los de forma a proporcionar um aprendizado eficiente, assim, os algoritmos

de otimização, ou de aprendizagem, são utilizados como um conjunto de regras responsáveis pelo ajuste destes pesos (ADERALDO, 2017). Esses métodos buscam reduzir a função de custo, geralmente definida pelo erro quadrático médio (*Mean Square Error* - MSE), dado por

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2.21}$$

onde Y_i é o valor observado e $\hat{Y}_i$ o predito, dentro de n predições.

Neste contexto, serão abordados dois métodos de aprendizado, um para redes rasas e outro para redes profundas.

Para redes rasas, devido a lenta convergência do algoritmo de *backpropagation*, um dos métodos criados para superar essa dificuldade é o Levenberg-Marquardt (LM). Para lidar com redes de no máximo algumas centenas de conexões, este método pode ser de 10 a 100 vezes mais rápido que o de backpropagation (FINOCCHIO, 2014).

Tal acréscimo em desempenho, conforme podemos observar em Lourakis (2005), se dá pelo fato de que quanto mais longe a solução atual está longe da desejada, o algoritmo varia os pesos de uma forma lenta, como no método *steepest descent*, buscando garantir a convergência. Ao se aproximar da solução os passos se tornam maiores, como no método de Gauss-Newton.

Para aprendizado profundo, os métodos disponíveis consistem, em sua maioria, de variações do algoritmo de *Stochastic Gradient Descent* (SGD), assim como este em questão. Adam, do inglês *Adaptive Moment Estimation*, foi proposto por Kingma e Ba (2014), como uma forma de compensar a lenta convergência do gradiente, com um bom desempenho em parâmetros não estacionários.

Guan et al. (2017) menciona que o algoritmo Adam calcula as taxas de aprendizado de forma adaptativa com base nas estimativas de primeira e segunda ordem do gradiente, resultando em uma convergência muito mais rápida que o SGD.

Kingma e Ba (2014) acreditam que esse método é promissor para utilização com mercado de ações, por ter implementação direta, computacionalmente eficiente e requerer pouca memória, além de ser eficiente para dados ruidosos e não estacionários.

3 METODOLOGIA E ETAPAS DE DESENVOLVIMENTO

3.1 Divisão do projeto em etapas

O andamento do projeto foi dividido em 3 fases, sendo a primeira de pesquisa bibliográfica (Embasamento Teórico), seguida da execução e a terceira foi a fase de escrita do projeto final.

3.2 Visão geral

Para previsão dos indicadores de continuidade é possível utilizar dados em diversos níveis, como indicadores de toda a distribuidora, dos município, dos conjuntos elétricos, dentre outros. Para realização desse projeto foi utilizado o conjunto de São Mateus, dentre os 58 existentes em 2018, assim como Louback (2018), por apresentar um *DEC* próximo ao limite estabelecido pela ANEEL, observado na Tabela 1, bem como por conter uma estacão meteorológica do Instituto Nacional de Meteorologia (INMET), tornando possível futuros estudos.

Tabela 1 – Indicadores Coletivos de Continuidade do Conjunto de São Mateus

Ano	Clientes	DEC Apurado	DEC Limite	FEC Apurado	FEC Limite
2010	48.930	15,69	14,00	10,04	13,00
2011	35.984	12,00	13,00	6,60	13,00
2012	37.787	13,44	13,00	8,98	12,00
2013	39.653	11,87	13,00	7,71	12,00
2014	41.313	11,56	13,00	10,57	12,00
2015	42.842	12,28	12,00	8,90	11,00
2016	44.315	11,43	11,00	6,46	10,00
2017	44.879	10,59	11,00	7,40	10,00

Fonte: ANEEL (2017a)

Os dados coletados são referentes a janeiro de 2010 à dezembro de 2017, com os indicadores anuais observados na Tabela 1, obtidos com apoio da distribuidora EDP Espirito Santo.

3.3 Tratamento de Dados

Os dados coletados, no período mencionado acima, são referentes aos eventos utilizados para o cálculo dos indicadores, sendo necessário realizar o tratamento de dados para continuar as análises.

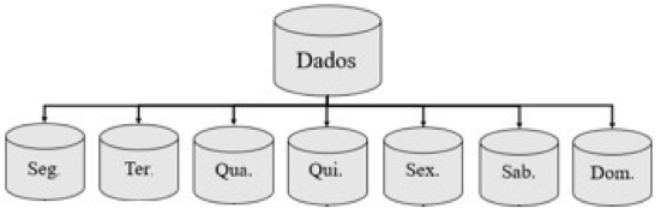
Um evento, segundo a ANEEL (2017b), é um "acontecimento que afete as condições normais de funcionamento de uma rede elétrica, podendo gerar uma ou mais interrupções no fornecimento de energia". Para cada interrupção de longa duração, conforme a legislação, faz necessário armazenar o fator gerador, horário de início e reestabelecimento e número de consumidores afetados.

Os eventos de interrupção são agrupados, formando as ordens, com causa e horário de início da primeira falta de energia notificada. Assim, para cada ordem, é calculado o indicador do conjunto onde ocorreu a interrupção.

Com os indicadores coletivos referentes a cada ordem, é feito o agrupamento dos que pertencem ao conjunto elétrico de São Mateus somando os indicadores coletivos, obtendo assim, uma série temporal do DEC/FEC diário.

Como dito na Seção 2.3, é importante separar as séries temporais em seguimentos com comportamento semelhante. Dessa forma, sabendo que dados de DEC/FEC apresentam uma grande correlação entre os mesmos dias da semana (LOUBACK, 2018), foi realizada a separação dos indicadores de acordo com o dia, criando sete séries temporais, como na Figura 23.

Figura 23 – Segregação semanal

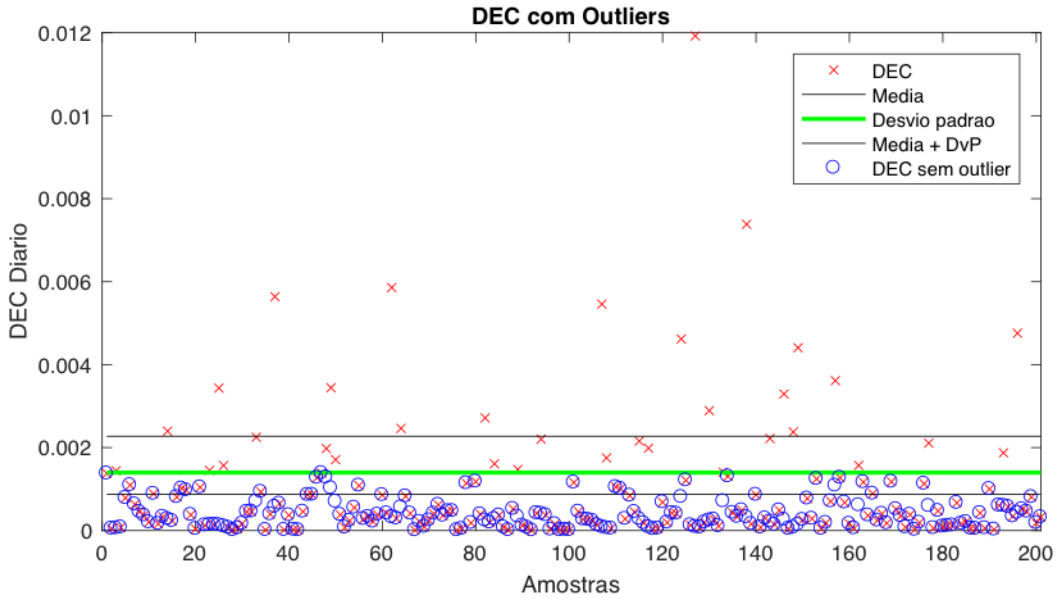


Fonte: Louback (2018), com edições do autor

Após a composição das séries de dia da semana, foi realizado o tratamento de *Outliers*

através do método empregado por (LOUBACK, 2018), onde foram mantidos inalterados os dados no intervalo dado por $[0, \sigma]$, sendo σ o desvio padrão calculado a partir de toda a série temporal. Os valores acima desse intervalo, foram retirados da série e preenchidos com um valor através do algoritmo de PCHIP, conforme a Figura 24, onde é possível observar a média, essa somada ao desvio padrão e a linha de corte realizada, com valor de um desvio padrão.

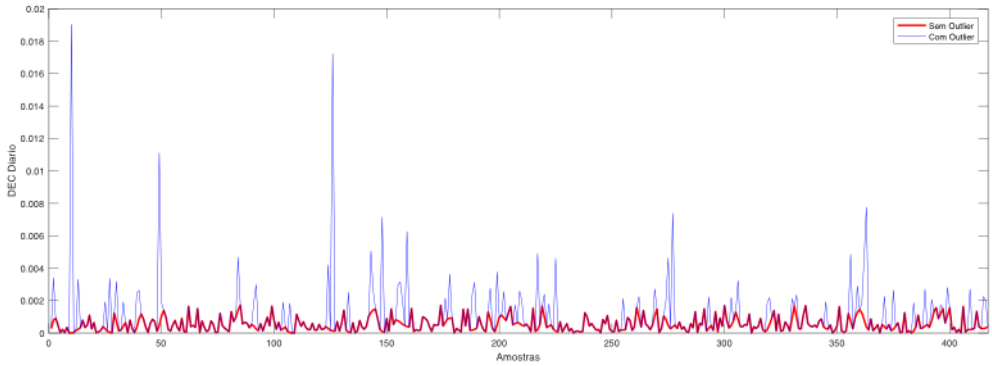
Figura 24 – Amostra do *DEC* antes e depois do tratamento de *outliers*



Fonte: Produção do Autor

A Figura 25 permite a comparação entre a série com e sem *outlier*, permitindo uma considerável suavização do comportamento, vale também observar que os dados acima do limite estabelecidos não são alterados para o valor de um desvio padrão, mas sim aplicados ao PCHIP que leva em consideração os pontos vizinhos.

Figura 25 – Série temporal de *DEC* da quinta-feita com e sem *outlier*

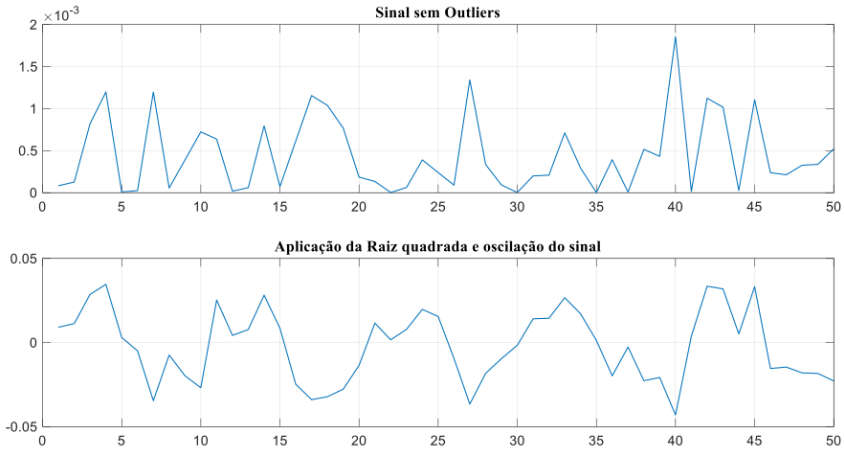


Fonte: Produção do Autor

Dado que os indicadores coletivos de continuidade, DEC/FEC, possuem apenas valores positivos, para melhor explorar as propriedades da função de ativação da tangente hiperbólica que varia no intervalo $[-1, 1]$, Figura 14, foi feita a oscilação em torno do zero, de forma a utilizar a região simétrica, positiva e negativa, da função de ativação (LOUBACK, 2018). A oscilação foi feita em uma janela com cinco amostras, ou seja, formando uma sequência de cinco valor positivos, cinco negativos.

Além disso, para possibilitar retornar a série após oscilação ao valor original, foi realizada a raiz quadrada dos itens, assim, sendo negativos ou positivos ao elevar os valores à potencia de ordem dois, os dados retornaram ao intervalo anterior. Com a aplicação dos dois algoritmos mencionados acima, tem-se a série indicada na Figura 26.

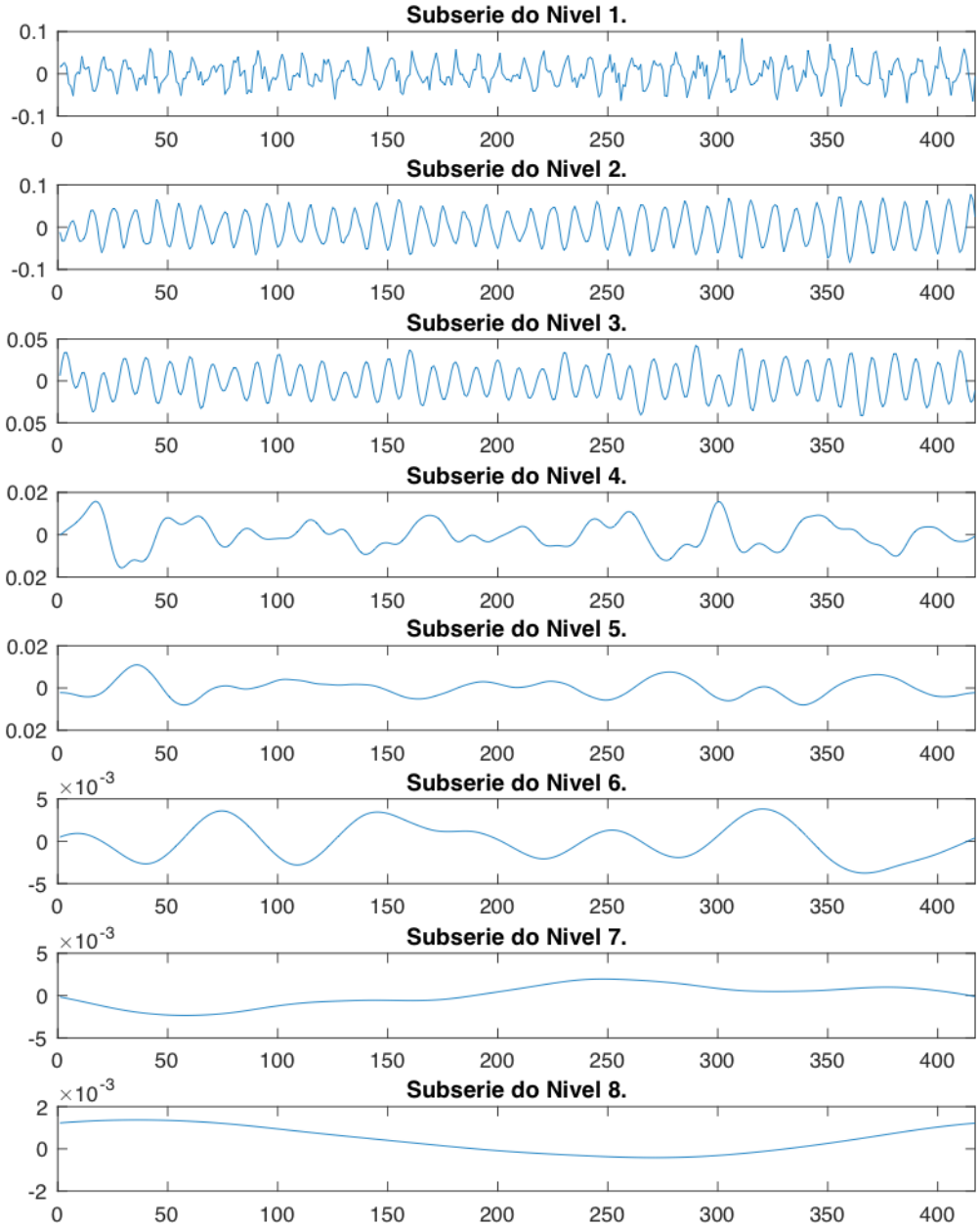
Figura 26 – Comparação da série após oscilação e raiz quadrada



Fonte: Louback (2018)

Através da transformada wavelet, a série foi dividida em 8 subséries, conforme a Figura 27. Pela imagem, é notável a diferença, sendo que quanto maior o nível da subsérie, maiores são as frequências que as compõe, contendo assim, mais informações.

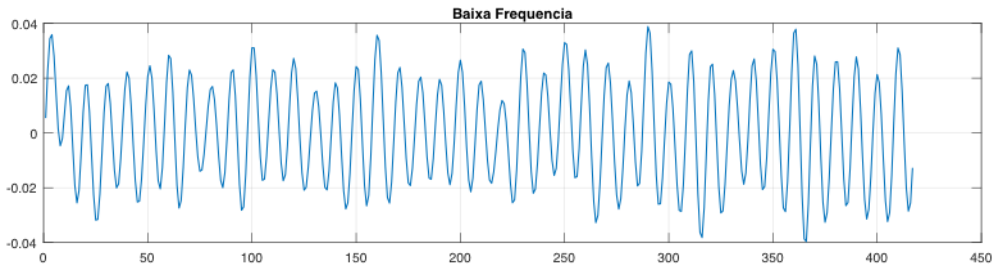
Figura 27 – Subséries após transformada wavelet



Fonte: Produção do Autor

Sabendo que a TW permite a soma das subséries para obtenção do sinal original, foi efetuada a adição dos seis sinais de menor frequência para composição do sinal denominado Baixa Frequência, representada pela Figura 28.

Figura 28 – Série de Baixa Frequência

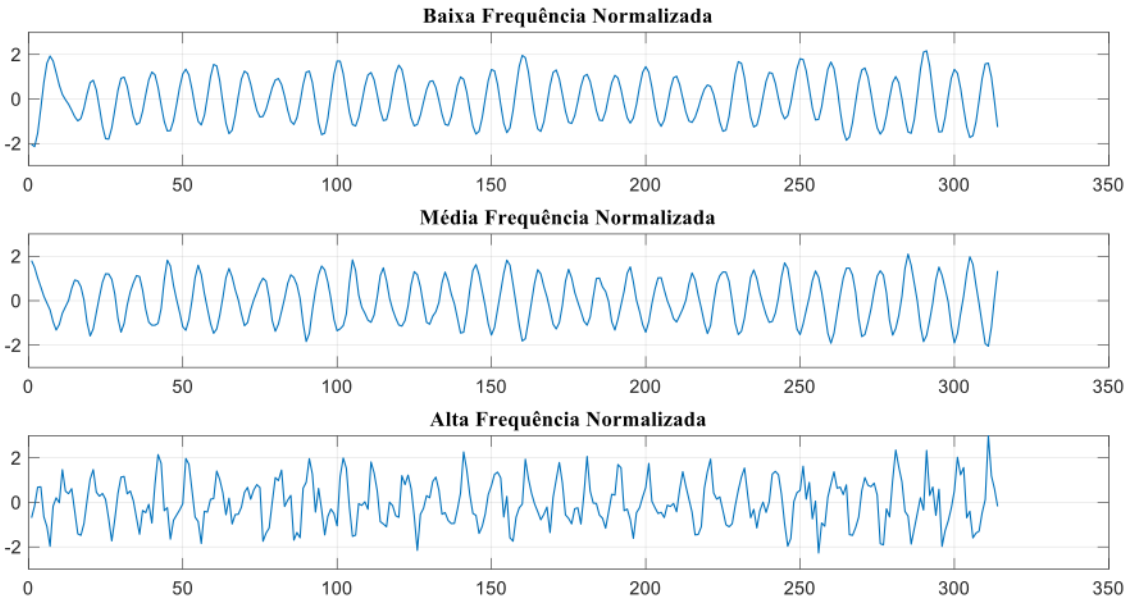


Fonte: Produção do Autor

De modo a evitar que a rede opere na região de saturação da função de ativação e para que a magnitude dos valores não afetem a forma como os pesos e tendências (*bias*) são ajustados, é importante a realização do processo de normalização (GAMBOGI, 2013).

Assim, para melhorar o desempenho da rede foi realizada a normalização dos dados, de forma a obter média zero e desvio padrão um, através do método Z-score, sendo possível observar o resultado através da Figura 29.

Figura 29 – Dados Normalizados

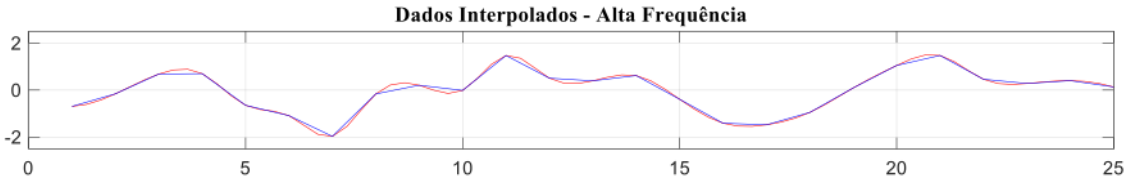


Fonte: Louback (2018)

Por último, com o objetivo de tornar a série de entrada mais suave, para aumentar o desempenho da rede (LOUBACK, 2018), foi empregado o método de interpolação *spline*, onde a partir de uma ordem e quantidade de pontos ao entorno definidos pelo usuário, é

retornado o vetor interpolado, apresentado pela Figura 30 na cor vermelha, enquanto o vetor com os dados originais está em azul.

Figura 30 – Série de alta frequência interpolada



Fonte: Louback (2018), com edições do autor

3.4 Treinamento da Rede

Foram treinados dois tipos de redes neurais para comparação com o trabalho desenvolvido por Louback (2018). A partir da rede por ele utilizada, NAR, foi desenvolvida uma rede NARX e uma rede neural profunda com camada LSTM, ambas possuindo como entrada exógena o dia da semana anterior ao que está sendo previsto.

Para treinamento dos dados foi efetuada a separação de 70% para treinamento e 30% para teste e validação. Sendo que para a comparação foi calculado o erro relativo de cada subsérie de frequência e dia.

A previsão é realizada individualmente para cada frequência, assim, três séries são retornadas com $I.HP$ dados, sendo I o número de pontos interpolados e HP o horizonte de previsão.

As redes foram treinadas diversas vezes, variando alguns de seus parâmetros, como *delay*, interpolação e numero de neurônios. Dessa forma, apenas o treinamento que obteve o melhor desempenho foi salvo.

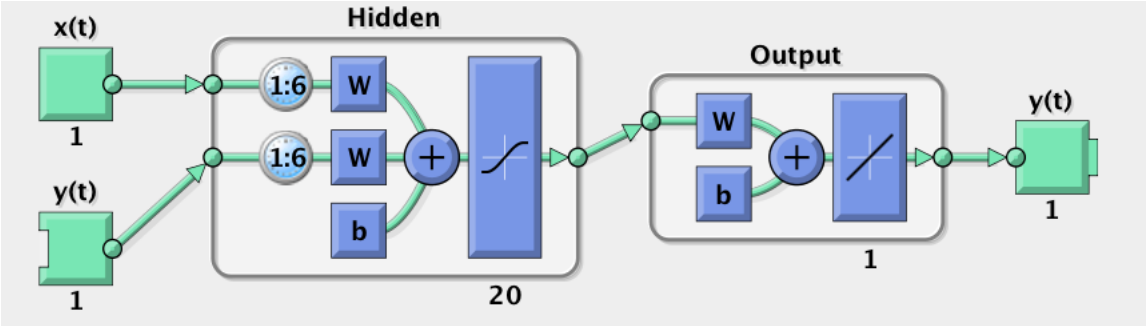
O parâmetro *delay* mencionado define a quantidade de dados passados que a rede utilizará para realizar a predição, ou seja, para um valor de x dias, os dados para treinamento corresponderão as entradas $(y_n, y_{n-1}, ..., y_{n-x})$, obtendo a saída y_{n+1}.

3.4.1 Rede NAR e NARX

Para criar uma base de comparação com as melhorias propostas, realizou-se um treinamento de uma rede NAR, de parâmetros variados por intervalos determinados Louback (2018), sendo eles, $[5, 50]$ neurônios, $[2, 5]$ *delays*, número de pontos para às interpolações foi definido para o intervalo de $[3, 6]$, de forma a manter a rede de comparação. Vale ressaltar que o valor do *delay* é considerado como base, de modo que o utilizado como parâmetro da série depende do número de interpolações realizadas.

A Figura 31 exemplifica a rede NARX utilizada, possuindo uma camada oculta, foi definido um intervalo de variação para os parâmetros da rede, assim como na rede NAR, de modo que os neurônios foram limitados ao intervalo $[20, 65]$, o *delay* a $[2, 5]$ e o número de interpolações foi limitado a $[3, 4]$, pois quanto mais interpolações, maior o *delay* utilizado pela rede.

Figura 31 – Rede NARX com uma camada oculta



Fonte: Produção do Autor

Ambos os algoritmos de treinamento utilizaram o método Levenberg-Marquardt como algoritmo de otimização e a tangente hiperbólica como função de ativação.

3.4.2 Aprendizado profundo com camada LSTM

Para a previsão dos indicadores através de técnicas de aprendizado profundo, foi utilizada uma arquitetura com seis camadas, sendo quatro ocultas, uma de entrada e uma de saída.

A respeito das camadas ocultas, as duas primeiras são do tipo LSTM, com a quantidade de neurônios na primeira camada variando no intervalo de $[400, 600]$, sendo que a segunda camada possui $2/3$ do número de neurônios da primeira. A terceira camada conta com o mesmo número de neurônios da camada anterior, seguida da quarta camada com apenas

um neurônio, visto que o horizonte de previsão é de um dia, sendo que ambas são do tipo totalmente conectadas.

3.5 Previsão da Rede

Buscando exemplificar a previsão da rede para cada frequência, bem como o processo de agrupamento do resultado para constituir o erro final, foi escolhido o dia 5 de Janeiro. A partir desse dia, serão expressos os parâmetros com que a rede obteve o melhor desempenho para o indicador escolhido (DEC), junto com os erros relativos (3.1) e absolutos (3.2) de previsão em cada etapa. A comparação entre os tipos de rede será deixada para a Seção seguinte.

$$E_{rel} = \frac{Ind_{obs} - Ind_{prev}}{Ind_{prev}} \tag{3.1}$$

$$E_{abs} = \frac{Ind_{obs} - Ind_{prev}}{máx(Ind(t)) - min(Ind(t))} \tag{3.2}$$

sendo Ind_{obs} o indicador observado, Ind_{prev} o indicador previsto e $Ind(t)$ a série temporal do indicador a ser previsto.

Vale ressaltar que a escolha do indicador DEC para a apresentação dos resultados abaixo tem como objetivo evitar a repetição, visto que a análise é análoga.

3.5.1 Rede NAR

Os resultados da Tabela 2 indicam a melhor configuração obtida, e seu desempenho, dentre os intervalos proporcionados. Tal configuração obteve um erro final relativo de $E_{rel} = 27,48$ e erro absoluto $E_{abs} = 4,23$.

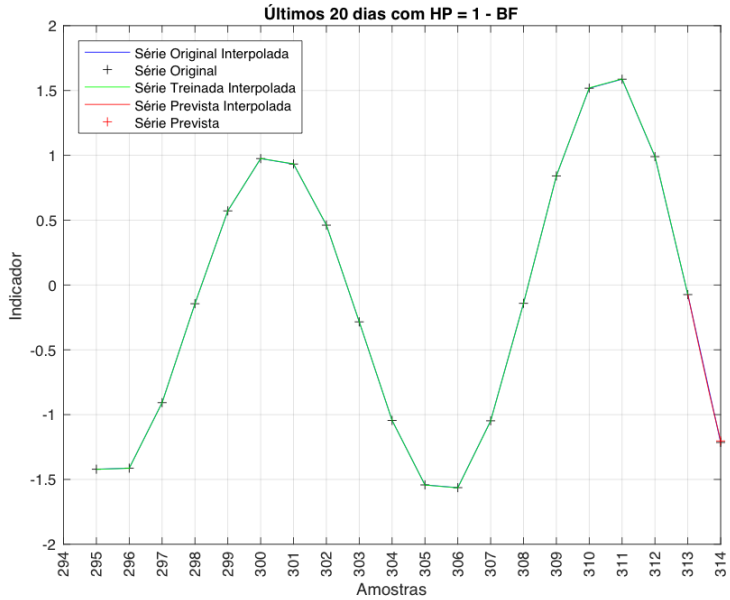
Através das Figuras 32, 33 e 34, é possível observar que os valores previstos, em vermelho, estão bem próximos dos observados, em azul, obtendo os erros indicados na Tabela 2 acima.

Tabela 2 – Configuração referente ao melhor desempenho obtido para o dia 05/01/2016 referente a rede NAR para o indicador *DEC*

Subsérie	Neurônios	Interpolação	Delay	Erro Relativo
BF	5	3	2	-0,0570
MF	35	3	5	-19,1357
AF	35	5	4	-41,8826

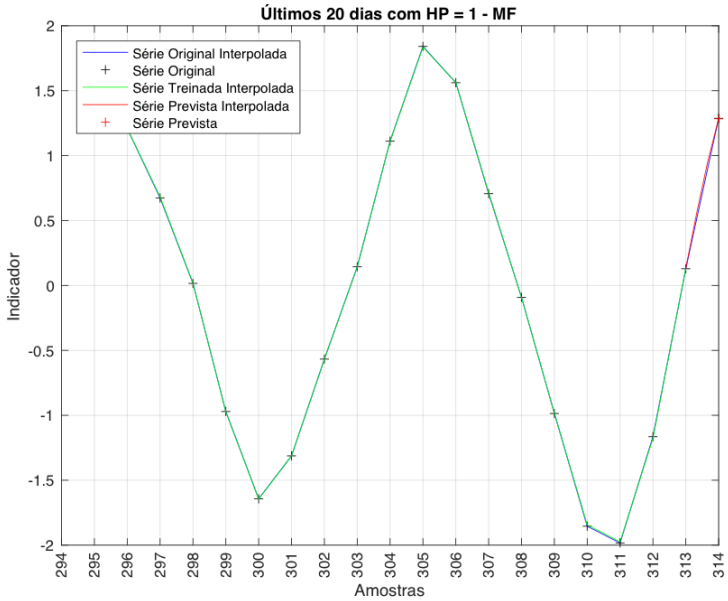
Fonte: Produção do Autor

Figura 32 – valor observado vs previsão da subsérie de Baixa Frequência para o indicador *DEC* no dia 05/01/2016 através de uma rede NAR



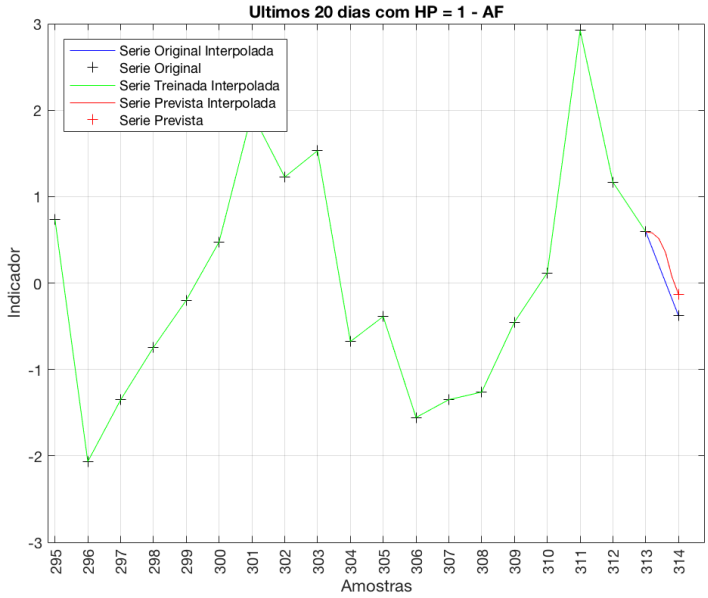
Fonte: Produção do Autor

Figura 33 – valor observado vs previsão da subsérie de Média Frequência para o indicador *DEC* no dia 05/01/2016 através de uma rede NAR



Fonte: Produção do Autor

Figura 34 – valor observado vs previsão da subsérie de Alta Frequência para o indicador *DEC* no dia 05/01/2016 através de uma rede NAR



Fonte: Produção do Autor

3.5.2 Rede NARX

De maneira semelhante ao indicado para a rede NAR, a Tabela 3 indica a melhor configuração obtida, e seu desempenho, dentre os intervalos proporcionados para a configuração em questão. O melhor resultado apresentou um erro total relativo de $E_{rel} = 21,80$ e erro absoluto $E_{abs} = 3,65$.

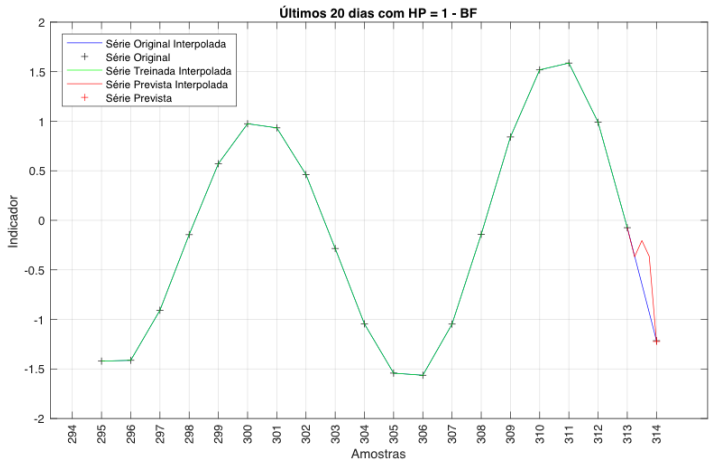
Tabela 3 – Configuração referente ao melhor desempenho obtido para o dia 05/01/2016 referente a rede NARX para o indicador DEC

Subsérie	Neurônios	Interpolação	Delay	Erro Relativo
BF	25	4	3	-0,6560
MF	45	3	4	0,6870
AF	45	3	4	-0,2204

Fonte: Produção do Autor

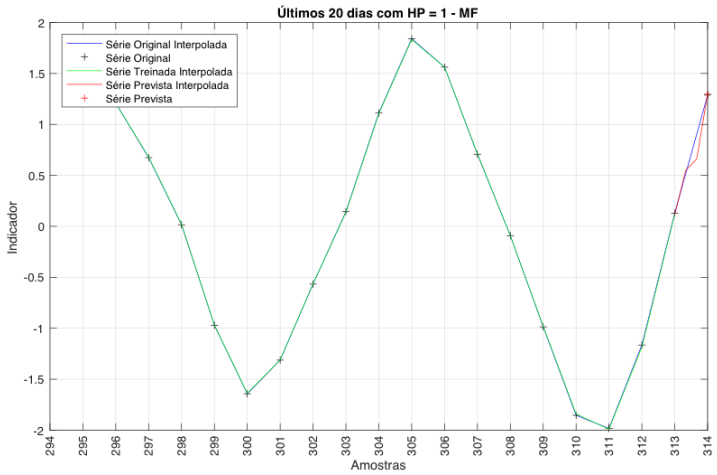
Por meio das Figuras 35, 36 e 37 nota-se que os valores previstos, em vermelho, estão bem próximos dos observados, em azul, obtendo os erros indicados na Tabela 3. As principais diferenças observadas, são referentes aos valores interpolados, de menos importância que o valor final.

Figura 35 – valor observado vs previsão da subsérie de Baixa Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NARX



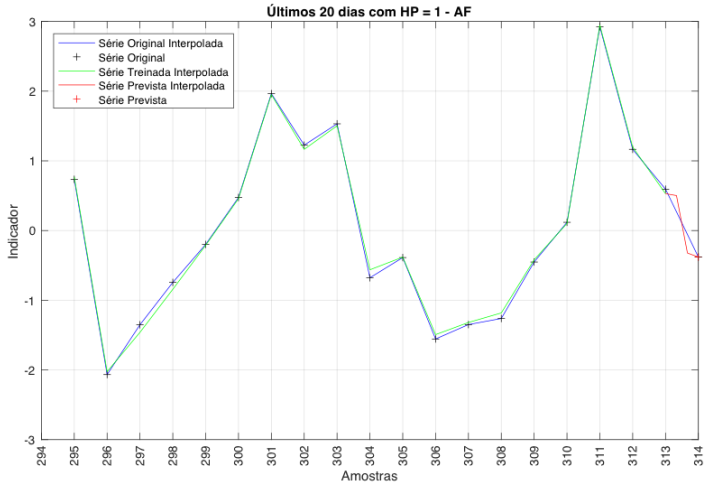
Fonte: Produção do Autor

Figura 36 – valor observado vs previsão da subsérie de Média Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NARX



Fonte: Produção do Autor

Figura 37 – valor observado vs previsão da subsérie de Alta Frequência para o indicador DEC no dia 05/01/2016 através de uma rede NARX



Fonte: Produção do Autor

3.5.3 Rede LSTM

A rede de aprendizado profundo teve como parâmetro de variação os neurônios e a quantidade de interpolações. Dentre as possíveis escolhas, a que obteve o melhor desempenho está apresentada na Tabela 4, obtendo um erro global relativo de $E_{rel} = -8,45$ e erro absoluto $E_{abs} = -1,42$.

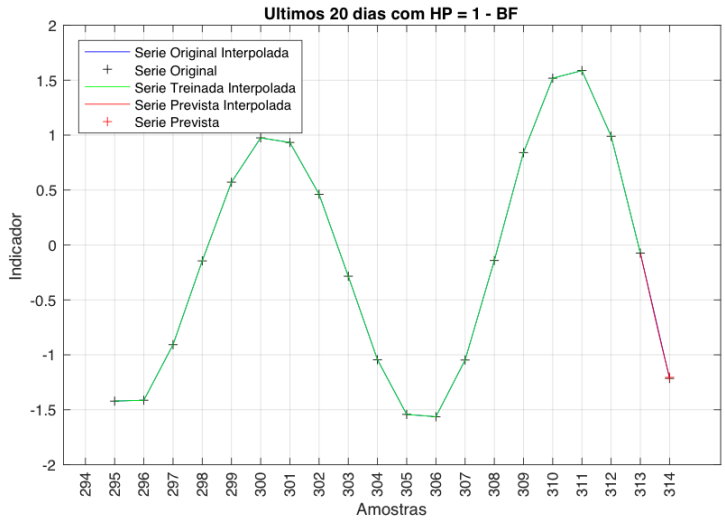
Tabela 4 – Configuração referente ao melhor desempenho obtido para o dia 05/01/2016 referente a rede LSTM para o indicador *DEC*

Subsérie	HL1	HL2	HL3	HL4	Interpolação	Erro Relativo
BF	400	267	267	1	4	-0,7808
MF	400	267	267	1	3	0,3553
AF	400	267	267	1	4	-24,0132

Fonte: Produção do Autor

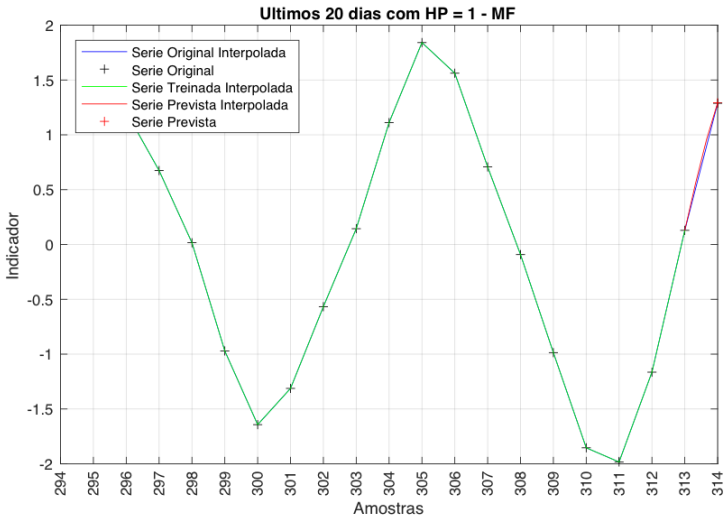
De maneira semelhante, o desempenho da rede pode ser observado por meio das Figuras 38, 39 e 40, onde na Tabela 4 é apresentado um erro maior para a Alta Frequência, esse também pode ser visto no gráfico de AF através da Figura 40.

Figura 38 – valor observado vs previsão da subsérie de Baixa Frequência para o indicador *DEC* no dia 05/01/2016 através de uma rede LSTM



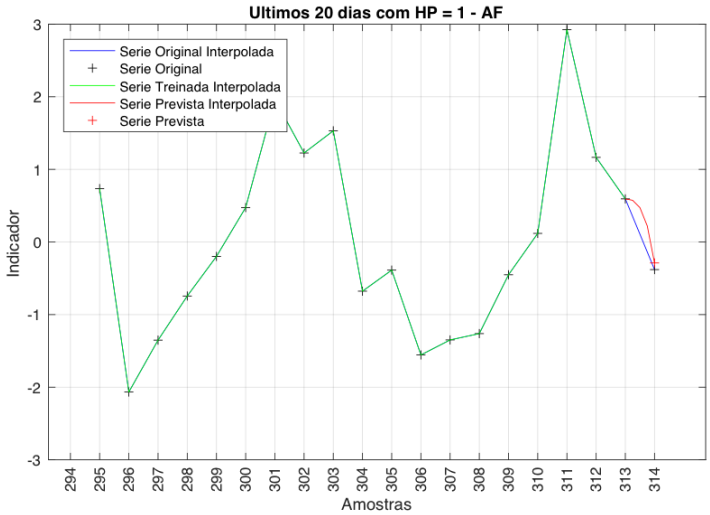
Fonte: Produção do Autor

Figura 39 – valor observado vs previsão da subsérie de Média Frequência para o indicador *DEC* no dia 05/01/2016 através de uma rede LSTM



Fonte: Produção do Autor

Figura 40 – valor observado vs previsão da subsérie de Alta Frequência para o indicador *DEC* no dia 05/01/2016 através de uma rede LSTM



Fonte: Produção do Autor

4 RESULTADOS E CONSIDERAÇÕES FINAIS

Esta seção compara os resultados obtidos nas propostas para previsão dos indicadores de continuidade DEC/FEC por meio de uma rede NARX e uma rede profunda com camada LSTM, em comparação com os realizados por Louback (2018) através de uma rede NAR.

4.1 Resultados das previsões dos indicadores DEC e FEC para os dias 01/01/2016 à 29/01/2016

A partir das previsões obtidas na Seção anterior (3.5), é possível observar os erros absolutos por indicador por rede utilizada, de acordo com cada dia, conforme as Tabelas 5 e 6.

Tabela 5 – Tabela de erro referente as previsões dos indicadores de continuidade de janeiro/2016 - (Continua)

Dia	DEC			FEC		
	NAR	NARX	LSTM	NAR	NARX	LSTM
1	7,32	2,39	5,74	0,05	-3,88	-1,94
2	0,49	10,96	8,25	1,28	10,67	5,09
3	1,06	22,01	20,69	0,04	6,85	1,62
4	1,12	8,42	6,1	0,99	-0,42	-4,78
5	4,23	3,65	-1,42	3,57	0,76	-3,84
6	0,18	22,08	15,91	0,1	4,55	0,8
7	3,82	86,71	81,91	0,06	23,77	19,92
8	0,01	0,14	-0,92	0,01	0,05	-0,15
9	8,55	0,43	0,41	0,02	-0,82	0,27
10	0,29	43,93	43,75	3,9	14,16	15,57
11	0,5	14,86	14,85	0,4	7,59	10,19
12	0,51	23,05	23,61	0,05	0,78	4,32
13	16,72	92,72	96,53	1,09	24,22	26,19
14	9,28	1,76	1,91	0,36	0,46	0,47
15	0,01	0,82	-0,24	8,01	29,6	32,78

Fonte: Produção do Autor

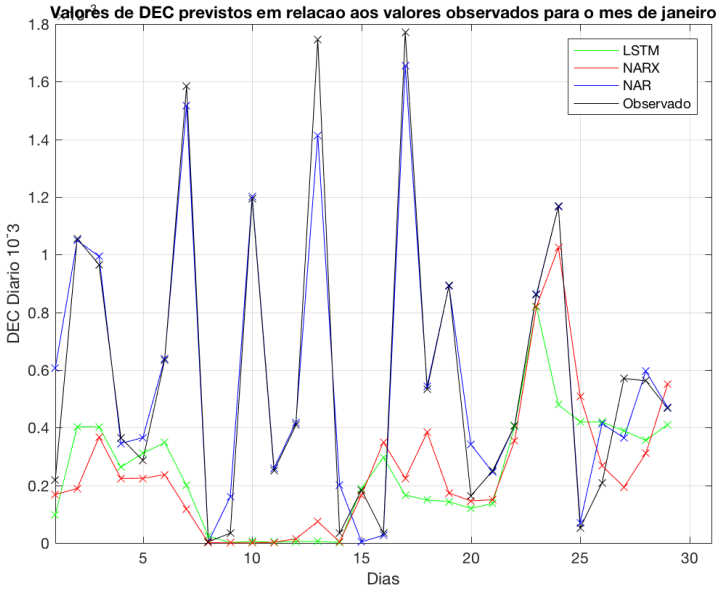
Tabela 6 – Tabela de erro referente as previsões dos indicadores de continuidade de janeiro/2016 - (Continuação)

Dia	DEC			FEC		
	NAR	NARX	LSTM	NAR	NARX	LSTM
16	0,53	35,72	36,39	0,05	-1,21	1,34
17	4,25	56,96	59,11	0,25	43,94	45,2
18	0,57	9,01	23,02	0,11	16,85	12,15
19	0,09	42,11	43,86	0,08	13,87	17,76
20	8,9	0,99	2,41	0,33	-6,92	-3,41
21	0,34	5,95	6,76	31,57	47,45	48,25
22	0,04	2,47	0,16	0,15	8,33	6,81
23	0,28	0,5	0,53	0,06	1,42	-1,44
24	0,03	5,27	25,35	0,06	4,38	5,56
25	1,06	-27,27	-21,99	1,01	-11,69	-11,73
26	11,22	-3,46	-12,34	6,33	-14,36	-11,02
27	10,29	20,88	10,06	4,79	8,72	6,33
28	1,84	14,9	12,22	14,08	-3,2	-5,38
29	0,12	-3,95	2,63	0,18	12,01	15,07

Fonte: Produção do Autor

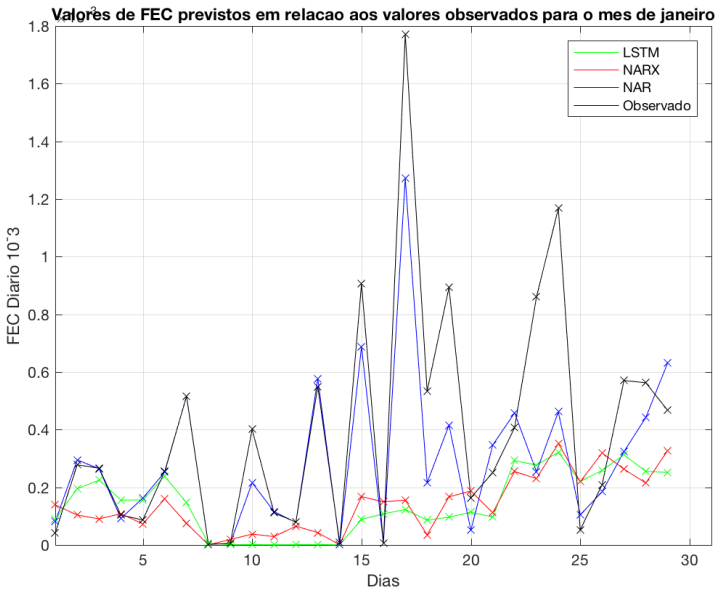
Por meio das Tabelas 5 e 6 é possível observar que apesar de alguns dias as séries NARX e LSTM terem obtido um melhor resultado, isso não representa a maioria. Com as Figuras 41 e 42 nota-se uma grande semelhança entre as séries previstas pela rede NARX e LSTM, entretanto afastados do valor observado.

Figura 41 – Comparação DEC previsto com o DEC Observado



Fonte: Produção do Autor

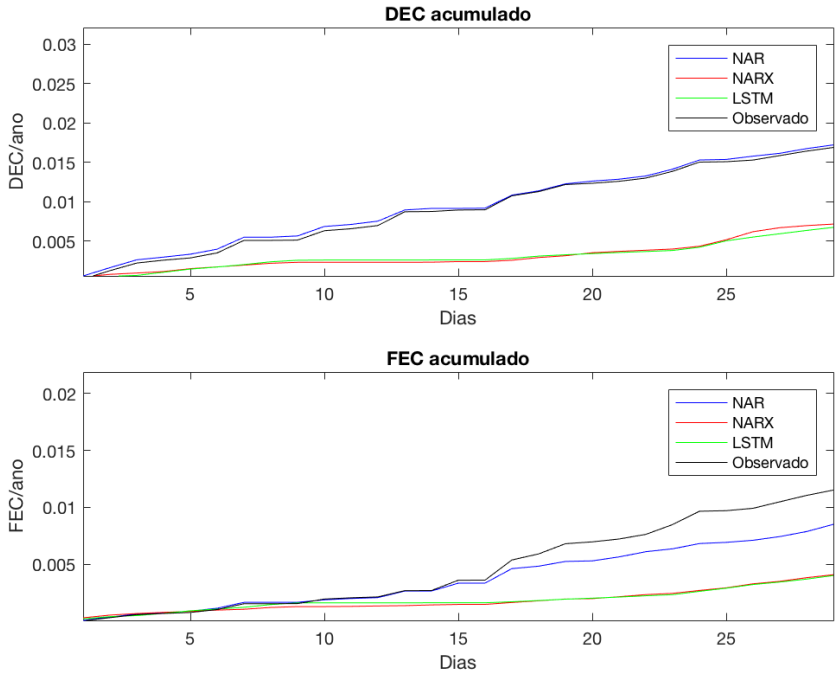
Figura 42 – Comparação FEC previsto com o FEC Observado



Fonte: Produção do Autor

Tal discrepância frente ao valor real para as séries propostas, bem como a proximidade entre elas, fica ainda mais evidentes quando colocadas com o indicador de forma acumulada mensal, como na Figura 43.

Figura 43 – DEC e FEC acumulado mensal



Fonte: Produção do Autor

Analisando o MSE pela Tabela 7, bem como por meio da Figura 43, é possível perceber que a previsão do DEC obteve bons resultados utilizando a rede NAR, ja o FEC teve um desempenho menor. Porém, a rede NAR obteve melhor desempenho para ambos os indicadores quando comparada com a rede NARX e aprendizado profundo com camada LSTM.

Tabela 7 – MSE das redes propostas em relação aos indicadores DEC e FEC

DEC			FEC		
NAR	NARX	LSTM	NAR	NARX	LSTM
1.6652e-08	3.9256e-07	3.9830e-07	5.7409e-08	2.0698e-07	2.1366e-07

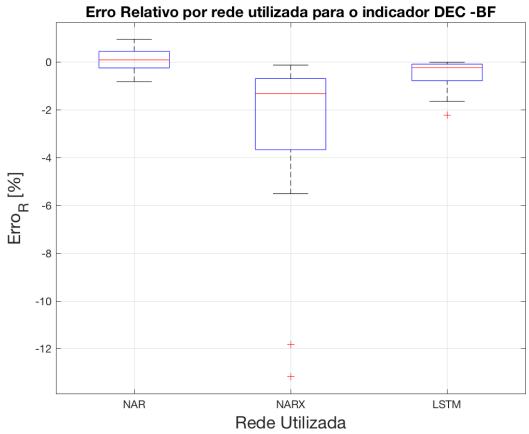
Fonte: Produção do Autor

4.2 Erros para as subséries de frequência

Como cada indicador é dividido em três subséries de frequência (Baixa, Média e Alta) foi feita a análise dessas, de forma a observar o desempenho das redes rasas e da rede

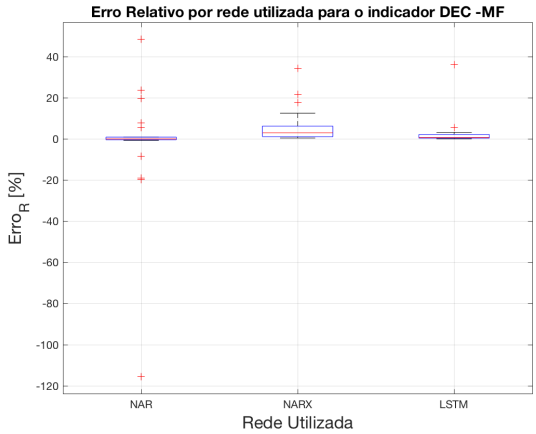
profunda para cada situação. Por meio das Figuras 44, 45, 47 e 48 extrai-se que a rede NARX não obteve um bom desempenho para a baixa e media frequência de ambos os indicadores, com a rede NAR e LSTM foram obtidos erros menos dispersos e próximos a zero. Entretanto, nas Figuras 46 e 49, representando a distribuição do erro para alta frequência, a rede profunda com camada LSTM teve uma grande piora no desempenho e a Rede NARX obteve o melhor resultado das três metodologias, em ambos os indicadores. Assim, acredita-se que a entrada exógena do dia anterior teve uma boa correlação para variações rápidas.

Figura 44 – Boxplot do erro relativo para subsérie de baixa frequência do indicador *DEC*



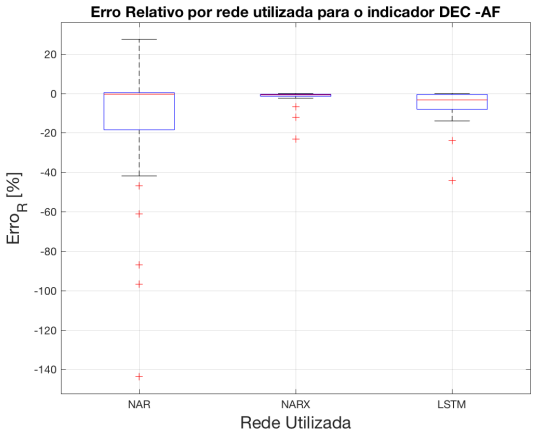
Fonte: Produção do Autor

Figura 45 – Boxplot do erro relativo para subsérie de média frequência do indicador *DEC*



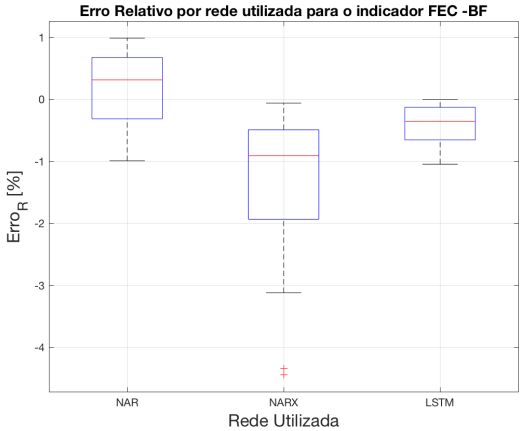
Fonte: Produção do Autor

Figura 46 – Boxplot do erro relativo para subsérie de alta frequência do indicador DEC



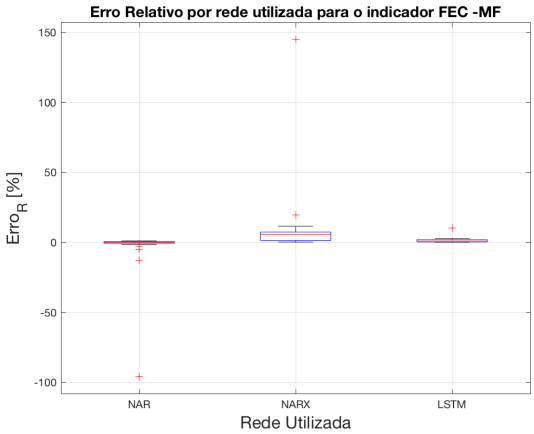
Fonte: Produção do Autor

Figura 47 – Boxplot do erro relativo para subsérie de baixa frequência do indicador FEC



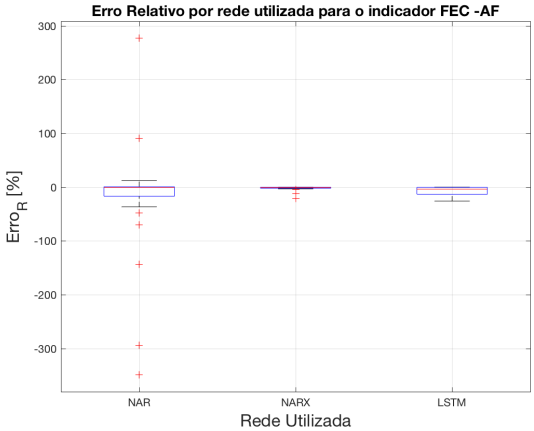
Fonte: Produção do Autor

Figura 48 – Boxplot do erro relativo para subsérie de média frequência do indicador FEC



Fonte: Produção do Autor

Figura 49 – Boxplot do erro relativo para subsérie de alta frequência do indicador FEC



Fonte: Produção do Autor

5 CONCLUSÕES E PROJETOS FUTUROS

Este trabalho teve como objetivo maior a aplicação de técnicas alternativas de predição de forma a comparar com o resultado obtido por Louback (2018) referente a previsão dos indicadores de continuidade coletivos de uma distribuidora de energia elétrica, em que a partir dos resultados obtidos, seja possível uma melhor alocação de recursos da empresa para um horizonte pequeno.

Foi realizada a predição do DEC e FEC por meio de uma rede base para comparação, rede NAR, e duas outras propostas pelo autor, NARX e rede profunda com camada LSTM, possuindo como entradas exógenas o dia da semana anterior ao que está sendo predito.

Para os indicadores DEC e FEC, a rede NAR obteve o melhor desempenho, seguido pela rede NARX. A rede NARX, obteve o melhor desempenho para a alta frequência, possibilitando inferir que o dia anterior tem uma maior influencia para variações rápidas, já a baixa e média que correspondem a variações sazonais, em sua maioria, pouco tem a ver com a serie exógena proposta.

Assim, por meio deste trabalho, foi possível observar que a serie exógena utilizada apresentou um ganho no desempenho da rede de alta frequência, sendo assim, através do treinamento da rede NAR para BF e MF e a rede NARX para AF, obter um resultado ainda mais satisfatório.

Além disso, pelo fato da rede NAR utilizada possuir poucos neurônios, foi possível realizar um "rápido" treinamento dessa. Seguido pela rede profunda LSTM, onde contrário ao senso comum, conseguiu um tempo menor se comparada a rede NARX. Isso se deve a utilização da GPU para a rede profunda, aumentando drasticamente a velocidade de treinamento.

5.1 Trabalhos Futuros

Para um trabalho futuro é possível utilizar todos os outros dias da semana como entrada exógena, não apenas o dia anterior como apresentado neste trabalho. Além disso, as ordens que compõe os indicadores de continuidade possuem algumas causas relacionadas ao clima, como vento forte e descarga atmosférica, podendo ser estudada a separação do DEC/FEC

de acordo com a origem e realizar a soma destes para obter o indicador diário.

Em complemento, outras técnicas de tratamento de dados podem ser empregadas, buscando reduzir a retirada dos *outliers*, por exemplo, a suavização através da média móvel exponencial, que atribui um peso aos valores que decai conforme se afastam do ponto em questão (KALEKAR, 2004).

Devido a gama de variações de parâmetros no treinamento de uma rede neural, tanto rasa como profunda, e que a mudança desses afeta consideravelmente o resultado da rede, foi necessário inserir um limite para que o projeto fosse executado em tempo hábil. Assim, a partir da utilização de computadores com maior capacidade de processamento, como o *Colaboratory* disponibilizado pelo *Google* de forma grátis, é possível aumentar os intervalos de teste dos parâmetros para buscar uma solução mais próxima do ótimo.

REFERÊNCIAS

ABRAADE. Tarifas de Energia. [S.l.]: Associação Brasileira de Distribuidores de Energia Elétrica, 2016. <http://www.abradee.com.br/setor-de-distribuicao/tarifas-de-energia/tarifas-de-energia>. Accessed: 2018-04-14. Citado na página 14.

ABRAADE. Investimentos do Setor. [S.l.]: Associação Brasileira de Distribuidores de Energia Elétrica, 2017. <http://www.abradee.com.br/setor-de-distribuicao/investimentos-do-setor>. Accessed: 2018-10-21. Citado na página 18.

ADERALDO, V. P. Utilização de Redes Neurais Artificiais para Detecção de Anomalia em Redes de Computadores. 2017. Citado na página 40.

ANEEL. Qualidade do Serviço. [S.l.]: Agencia Nacional de Energia Elétrica, 2016. <http://www.aneel.gov.br/qualidade-do-servico2>. Accessed: 2018-04-29. Citado 2 vezes nas páginas 15 e 16.

ANEEL. Indicadores de Continuidade por Conjunto. 2017. Citado na página 41.

ANEEL. Módulo 1. [S.l.]: Agencia Nacional de Energia Elétrica, 2017. <http://www.aneel.gov.br/modulo-1>. Accessed: 2018-04-19. Citado na página 42.

ANEEL. Módulo 8. [S.l.]: Agencia Nacional de Energia Elétrica, 2017. <http://www.aneel.gov.br/modulo-8>. Accessed: 2018-04-19. Citado na página 14.

BISHOP, C. Pattern Recognition and Machine Learning. [S.l.]: Springer-Verlag, New York, 2006. ISBN 978-1-4939-3843-8. Citado na página 30.

BOUSSAADA, Z. et al. A nonlinear autoregressive exogenous (narx) neural network model for the prediction of the daily direct solar radiation. Energies, Multidisciplinary Digital Publishing Institute, v. 11, n. 3, p. 620, 2018. Citado na página 35.

BROCKWELL, P. J.; DAVIS, R. A. Time series: theory and methods. [S.l.]: Springer Science & Business Media, Colorado, 2013. Citado 2 vezes nas páginas 23 e 24.

CARVALHO, J. F. d. Energia e sociedade. Estudos Avançados, SciELO, 2014. Accessed: 2018-10-23. Citado na página 17.

CORNISH, C. R.; BRETHERTON, C. S.; PERCIVAL, D. B. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. Boundary-Layer Meteorology, Springer, v. 119, n. 2, p. 339–374, 2006. Citado na página 28.

CUI, Z.; KE, R.; WANG, Y. Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction. arXiv preprint arXiv:1801.02143, 2018. Citado 2 vezes nas páginas 37 e 38.

DINIZ, A. P. M. Modelos de Previsão do Conteúdo de Silício no Ferro-Gusa Usando Redes Neurais Artificiais. Dissertação (Mestrado) — Universidade Federal do Espirito Santo, Vitória, 2018. Citado 3 vezes nas páginas 26, 27 e 28.

DOUMAS, L. A.; PUEBLA, G.; MARTIN, A. E. Human-like generalization in a machine through predicate learning. arXiv preprint arXiv:1806.01709, 2018. Citado na página 28.

EPE. Plano Decenal de Expansão de Energia 2026. [S.l.]: Empresa de Pesquisa Energética, 2017. 249 p. <http://www.aneel.gov.br/modulo-8>. Accessed: 2018-04-19. Citado na página 14.

FINOCCHIO, M. A. F. Noçõs de Redes Neurais Artificiais. 2014. Citado na página 40.

GAMBOGI, J. A. Aplicação de redes neurais na tomada de decisão no mercado de ações. Universidade de São Paulo, São Paulo, p. 78, 2013. Citado na página 46.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado 3 vezes nas páginas 29, 36 e 37.

GUAN, N. et al. Delay compensated asynchronous adam algorithm for deep neural networks. In: IEEE. Ubiquitous Computing and Communications (ISPA/IUCC), 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on. [S.l.], 2017. p. 852–859. Citado na página 40.

HAYKIN, S. S. et al. Neural networks and learning machines. [S.l.]: Pearson Upper Saddle River, 2009. v. 3. Citado 6 vezes nas páginas 30, 31, 32, 33, 34 e 35.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. Neural computation, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 37.

KALEKAR, P. S. Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi School of Information Technology, v. 4329008, p. 1–13, 2004. Citado na página 64.

KARJALAINEN, J. et al. Energy Models and Scenarios in the Era of Climate Change. [S.l.: s.n.], 2014. ISBN 978-952-249-275-3. Citado na página 17.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014. Citado na página 40.

KRZYK, K. Coding Deep Learning For Beginners. [S.l.]: Towards Data Science, 2018. <https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>. Accessed: 2018-10-25. Citado 2 vezes nas páginas 29 e 30.

KUHN, M.; JOHNSON, K. Applied Predictive Modeling. [S.l.]: Springer, New York, NY, 2013. 27–59 p. ISBN 978-1-4614-6849-3. Citado na página 24.

LACOMMARE, K. H. et al. Improving the estimated cost of sustained power interruptions to electricity customers. Energy, Elsevier, CA-USA, v. 153, p. 1038–1047, 2018. Citado 2 vezes nas páginas 18 e 19.

LIU, C.-L. A tutorial of the wavelet transform. NTUEE, Taiwan, 2010. Citado na página 26.

LOUBACK, F. O. Previsão de Indicadores de Continuidade com a utilização de redes neurais. 88 p. Dissertação (Mestrado) — Universidade Federal do Espirito Santo, Vitória, 2018. Citado 11 vezes nas páginas 20, 25, 41, 42, 43, 44, 46, 47, 48, 56 e 63.

LOURAKIS, M. I. A brief description of the levenberg-marquardt algorithm implemented by levmar. Foundation of Research and Technology, v. 4, n. 1, p. 1–6, 2005. Citado na página 40.

LU, Y. Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration, Elsevier, v. 6, p. 1–10, 2017. Citado na página 13.

MALLEY, B.; RAMAZZOTTI, D.; WU, J. T.-y. Secondary Analysis of Electronic Health Records. [S.l.]: Springer, Cham, 2016. 115–141 p. ISBN 978-3-319-43742-2. Citado 2 vezes nas páginas 24 e 25.

MITCHELL, T. M. Machine Learning. [S.l.]: McGraw-Hill Science/Engineering/Math, 1997. Citado 2 vezes nas páginas 28 e 29.

MME. Inústria 4.0. Ministério da Indústria, Minas e Energias, 2017. Accessed: 2018-10-23. Citado na página 13.

MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. 2. ed. [S.l.]: Egard Blucher, São Paulo, 2006. Citado 2 vezes nas páginas 22 e 25.

OLAH, C. Understanding LSTM Networks. 2015. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Accessed: 2018-11-07. Citado na página 38.

OSBORNE, J. W. Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data. SAGE Publications, 2012. ISBN 9781452289670. Disponível em: <https://books.google.com.br/books?id=bzkXBAAAQBAJ>. Citado na página 24.

PERCIVAL, D. B.; WALDEN, A. T. The maximal overlap discrete wavelet transform. In: _____. Wavelet Methods for Time Series Analysis. [S.l.]: Cambridge University Press, 2000. (Cambridge Series in Statistical and Probabilistic Mathematics), p. 159–205. Citado na página 28.

RITCHIE HANNA, R.-M. Energy Production Changing Energy Sources. [S.l.]: OurWorldInData, 2018. <https://ourworldindata.org/energy-production-and-changing-energy-sources>. Accessed: 2018-10-21. Citado na página 18.

SALGADO, C. M. et al. Secondary Analysis of Electronic Health Records. [S.l.]: Springer, Cham, 2016. 163–183 p. ISBN 978-3-319-43742-2. Citado na página 25.

SANTOS, W. B. dos; COLOSIMO, E. A.; MOTTA, S. B. da. Tempo ótimo entre manutenções preventivas para sistemas sujeitos a mais de um tipo de evento aleatório. SciELO Brasil, 2007. Citado na página 13.

SHARMA, S. Activation Functions: Neural Networks. [S.l.]: Towards Data Science, 2017. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>. Accessed: 2018-10-25. Citado 2 vezes nas páginas 32 e 33.

SOYTAS, U.; SARI, R. Energy consumption and gdp: causality relationship in g-7 countries and emerging markets. Energy Economics, v. 25, n. 1, p. 33 – 37, 2003. ISSN 0140-9883. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0140988302000099>. Citado na página 18.

SQALLI, M. Traffic signs classification with Deep Learning. [S.l.]: Hacker Noon, 2016. <https://hackernoon.com/traffic-signs-classification-with-deep-learning-b0cb03e23efb>. Accessed: 2018-11-07. Citado na página 36.