

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET
DEPARTAMENTO DE COMPUTAÇÃO– DC
ENGENHARIA DE COMPUTAÇÃO– ENC

Leticia Amaral da Cunha

**Predição de Carga de Energia Elétrica
no Brasil utilizando técnicas de Deep
Learning**

São Carlos
2022

Leticia Amaral da Cunha

**Predição de Carga de Energia Elétrica
no Brasil utilizando técnicas de Deep
Learning**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Roberto Santos Inoue

Coorientador: Marcelo Suetake

São Carlos

2022

Agradecimentos

Agradeço à minha família, em especial a minha mãe, pelo constante incentivo e encorajamento. Agradeço aos amigos que fiz nessa jornada, pelo apoio e companhia em todos os momentos.

Agradeço também aos professores Marcelo e Roberto pela oportunidade, tempo e paciência durante todo o processo de construção deste trabalho.

Resumo

O mercado livre de energia tem se destacado por ser um forma competitiva e dinâmica de negociação de um ativo antes restrito a um consumo passivo. A liberdade de negociação proporcionada aos integrantes desse mercado tornam a predição de seu preço, assim como os fatores que o afetam de suma importância. Este estudo tem como propósito o desenvolvimento de três modelagens apoiadas em técnicas distintas de redes neurais artificiais recorrentes e a sua análise comparativa. Para tal, devem ser utilizadas as arquiteturas do tipo LSTM *Feed Forward*, LSTM bidirecional e GRU para cada uma das séries temporais exploradas. Dados de carga de energia elétrica média do mercado livre de energia brasileiro entre os anos de 2011 a 2021 são aplicados como base de dados para treinamento, validação e teste. Para a avaliação dos resultados foram utilizadas métricas de ampla aplicação como erro médio quadrático e erro percentual absoluto médio. A análise dos modelos verifica o desempenho das técnicas aplicadas, bem como possíveis causas para a performance das métricas estabelecidas. Os resultados apontaram os tipos LSTM bidirecional e GRU como resultados viáveis da modelagem.

Palavras-chave: Mercado Livre de Energia. Carga de Energia Elétrica. Sistema Integrado Nacional. Séries temporais. Redes Neurais. Redes Neurais Recorrentes. LSTM. GRU.

Abstract

The open electricity market has stood out for being a competitive and dynamic way of trading an asset that was previously restricted to passive consumption. The freedom of negotiation provided to the participants of this market makes the prediction of its price, as well as the factors that affect it, extremely important. The purpose of this study is to develop three models based on different techniques of recurrent artificial neural networks and their comparative analysis. For this, the LSTM Feed Forward, LSTM bidirectional and GRU architectures must be applied to each of the time series explored. Average electric energy load data from the Brazilian free energy market between the years 2011 to 2021 are applied as an input for training, validation and testing processes. For the evaluation of the results, metrics of wide application were used, such as mean square error and mean absolute percentage error. The analysis of the models verifies the performance of the applied techniques, as well as possible causes for the performance of the established metrics. The results showed the bidirectional LSTM and GRU types as viable modeling results.

Keywords: Open Electricity Market. Electrical Load. National Integrated System. Time Series. Neural Networks. Recurrent Neural Networks. LSTM. GRU.

Lista de ilustrações

Figura 1 – Mapa dos diferentes subsistemas e Linhas de transmissão do SIN . . .	24
Figura 2 – Gráfico do PLD por região	25
Figura 3 – Hierarquia de conceitos, Deep Learning	27
Figura 4 – Componentes e estrutura primordial de uma rede neural	29
Figura 5 – Função Sigmóide	30
Figura 6 – Função ReLU	30
Figura 7 – Estrutura Muitos-para-Um	31
Figura 8 – Estrutura Muitos-para-Muitos	32
Figura 9 – Arquitetura LSTM	33
Figura 10 – Arquitetura GRU	33
Figura 11 – Valores de carga para cada subsistema entre 2012 e 2017	43
Figura 12 – Valores de carga para cada subsistema entre 2018 e 2021	44
Figura 13 – Diagrama das etapas de desenvolvimento da modelagem	45
Figura 14 – Gráficos com valores faltantes, anos de 2016 e 2018	48
Figura 15 – Prévia do conjunto de dados para região Norte	49
Figura 16 – Prévia do conjunto de dados para região Nordeste	49
Figura 17 – Prévia do conjunto de dados para região Sudeste	49
Figura 18 – Prévia do conjunto de dados para região Sul	49
Figura 19 – Autocorrelação, região Norte em duas escalas temporais	50
Figura 20 – Autocorrelação, região Nordeste em duas escalas temporais	51
Figura 21 – Autocorrelação, região Sul em duas escalas temporais	52
Figura 22 – Autocorrelação, região Sudeste em duas escalas temporais	53
Figura 23 – Correlação, região Norte	55
Figura 24 – Correlação, região Nordeste	55
Figura 25 – Correlação, região Sul	55
Figura 26 – Correlação, região Sudeste	55
Figura 27 – Valores previstos e reais, região Norte	56

Figura 28 – Valores previstos e reais, região Nordeste	57
Figura 29 – Valores previstos e reais, região Sul	59
Figura 30 – Valores previstos e reais, região Sudeste	60

Lista de tabelas

Tabela 1 – Tipos de consumidores no mercado de Energia Elétrica brasileiro . . .	23
Tabela 2 – Variáveis e suas definições	42
Tabela 3 – Resultados, região Norte	57
Tabela 4 – Resultados, região Nordeste	58
Tabela 5 – Resultados, região Sul	58
Tabela 6 – Resultados, região Sudeste	60
Tabela 7 – Comparativo de resultados entre regiões	61

Lista de siglas

ABRACEEL Associação Brasileira dos Comercializadores de Energia

ABRADEE Associação Brasileira de Distribuidores de Energia Elétrica

AM Aprendizado de Máquina

ANEEL Agência Nacional de Energia Elétrica

CCEE Câmara de Comercialização de Energia Elétrica

DL *Deep Learning*

EMA Erro Médio Absoluto

EMPA Erro Médio Percentual Absoluto

EPE Empresa de Pesquisa Energética

EQM Erro Quadrático Médio

GRU *Gated Recurrent Units*

IA Inteligência Artificial

LSTM *Long Short Term Memory*

MLE Mercado Livre de Energia

ONS Operador Nacional do Sistema Elétrico

OSELM *Online Sequential Extreme Learning Machine*

PLD Preço da Liquidação das Diferenças

ReLU *Rectified Linear Unit*

REQM Raiz do Erro Quadrático Médio

RESEB Reestruturação do Setor Elétrico Brasileiro

RNR Redes Neurais Recorrentes

SIN Sistema Integrado de Energia

Sumário

1	INTRODUÇÃO	17
1.1	Motivação	19
1.2	Objetivos	19
1.3	Organização	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	O Setor Elétrico Brasileiro e o Mercado Livre de Energia	21
2.2	Ambiente de Contratação Livre	22
2.3	Sistema Interligado Nacional	23
2.4	Preço da Liquidação das diferenças	24
2.5	Series Temporais	25
2.6	Inteligência Artificial	26
2.7	Deep Learning	26
2.8	Aprendizagem Supervisionada, Não Supervisionado e Semi-Supervisionado	28
2.9	Redes Neurais Artificiais	28
2.10	Função de Ativação	29
2.11	Redes Neurais Recorrentes	30
2.11.1	Many-to-One e Many-to-Many	31
2.11.2	Long Short Term Memory	32
2.11.3	Gated Recurrent Units	33
2.12	Treinamento	34
2.13	Normalização	34
2.14	Métodos de Avaliação de Resultados	34
2.14.1	Erro Quadrático Médio	35
2.14.2	Raiz do Erro Quadrático Médio	35
2.14.3	Erro Médio Percentual Absoluto	35

2.14.4	Erro Médio Absoluto	35
3	TRABALHOS RELACIONADOS	37
4	METODOLOGIA	41
4.1	Ferramentas	41
4.2	Dados	42
4.3	Etapas do estudo proposto	44
4.3.1	Aquisição dos dados históricos de carga	44
4.3.2	Limpeza dos dados	45
4.3.3	Análise exploratória dos dados	45
4.3.4	Pré processamento dos dados	45
4.3.5	Definição da modelagem a ser utilizada	46
4.3.6	Treinamento das Redes Neurais	46
4.3.7	Avaliação dos resultados	46
5	DESENVOLVIMENTO E ANÁLISE DOS RESULTADOS . .	47
5.1	Desenvolvimento	47
5.2	Avaliação dos Resultados	55
	Conclusão	63
5.2.1	Trabalhos futuros	64
	REFERÊNCIAS	65

Capítulo 1

Introdução

Pode-se tomar como fato a indispensabilidade da energia elétrica como insumo, seja para a manutenção de indústrias ou para telecomunicações, serviços e até o desenrolar das atividades diárias em um centro urbano. Segundo dados da Empresa de Pesquisa Energética (EPE), mesmo em um ano abalado pela Covid-19 houve um crescimento no consumo em relação a 2019 igual ou acima de 3% nos setores Industriais, Residenciais e Agropecuários (EPE, 2021).

No Brasil a distribuição de energia é responsabilidade de um sistema que engloba 26 unidades federativas - estando apenas o estado de Roraima desconectado do grupo - dada a proporção continental, sua complexidade é proporcional (SILVA, 2020). Dessa forma, é necessário que o território nacional seja coberto por diversas redes de transmissão menores e que comportem diferentes voltagens. Além disso, para a administração da operação o país ainda é subdividido entre quatro regiões distintas, estando todas interligadas e conectando geradores a distribuidores.

Do ponto de vista de um consumidor residencial, cujo único papel é o acesso restrito ao serviço do provedor elétrico de sua região, pode ser difícil entender as particularidades e escala do setor elétrico brasileiro. Contudo, esse não é o cenário completo, a um grupo de consumidores é permitida a negociação de contratos de compra e venda de energia elétrica em um modelo de mercado aberto, eles se subdividem entre dois grupos, os chamados consumidores livres e especiais. Para esses, o preço pago pelo ativo pode ser negociado, assim a compreensão do sistema e fatores que influenciam o valor do mesmo são essenciais.

Como em grande parte dos mercados consumidores, o preço de um ativo tem influência direta da demanda pelo mesmo (VARIAN, 2015). Posto isso, a predição de carga de energia elétrica pode ser usada pelos integrantes do mercado livre de energia para otimizarem

suas estratégias de compra, garantindo um melhor preço.

Tendo em vista as matrizes principais utilizadas para a geração energética brasileira, seu valor final não é determinado apenas por sua demanda, mas também por uma série de outros fatores que influenciam esse setor. Alguns deles são as condições de temperatura, geração de energia atual, possível entrada de novas fontes de produção, entre outros.

É importante notar que para esse grupo de agentes que podem fazer a comercialização livre são definidas três escalas temporais para a definição de preços, a de longo, médio e curto prazo. Elas delimitam, respectivamente, períodos de anos, meses e dias. Tomando em conta os últimos dois intervalos de tempo, pode-se observar um mercado dinâmico, marcado por flutuações de valor e certa volatilidade, tornando sua predição uma proposta relevante.

Em virtude do grande volume de dados produzidos através dos anos no mercado de energia, a construção de modelos matemáticos que se mantenham relevantes sem a utilização de meios computacionais passa a ser uma tarefa bastante árdua. Em contrapartida, o rápido desenvolvimento computacional nas últimas décadas, tanto no que é relativo a software, como de hardware, tornaram a aplicação de métodos de inteligência artificial cada vez mais acessíveis (CHEN, 2016).

Para a predição de modelos não lineares o uso de ferramentas de inteligência artificial como redes neurais tem se mostrado de grande utilidade. Levando em conta o caso de estudo de modelagem de uma série temporal, isto é, a predição de carga de energia elétrica para uma determinada escala temporal, o uso de redes neurais recorrentes é ainda mais adequado (HEWAMALAGE; BERGMEIR; BANDARA, 2021).

A respeito de redes neurais é importante contextualizar que elas são parte da área de *Deep Learning* (DL) e definem uma estrutura que tem como objetivo emular o funcionamento do cérebro humano para realizar o reconhecimento de padrões e solução de problemas. Dado o escopo de aplicação da técnica seu estudo foi subdividido em diferentes aplicações, e por consequência, arquiteturas que tem como finalidade a solução de problemas distintos.

Redes neurais recorrentes em específico, tem como seu principal diferencial frente a outros tipos de redes neurais a capacidade de persistir informações. *Long Short Term Memory* (LSTM) bidirecional, *LSTM Feed Forward* e *Gated Recurrent Units* (GRU) são algumas arquiteturas dessa aplicação, e ainda que tenham componentes diversos em sua definição elas viabilizam a retenção de informações anteriores por meio da existência de *loops* em sua estrutura. Posto isso, a manutenção de características passadas é o que torna adequada a aplicação da técnica em problemas que envolvem séries temporais ou análise de textos.

Por último, a possibilidade de criação de modelos e métodos computacionais altamente complexos como redes neurais cujas definições são feitas a partir de uma linguagem computacional de alto nível como Python fornecem uma solução bastante útil para a modelagem

de dados em cenários com grande volume de informações.

1.1 Motivação

Como apontado anteriormente, o Brasil apresenta um cenário de expansão não apenas no seu fornecimento de energia elétrica, como também do volume de participantes no mercado livre de energia. Tendo isso em vista, a predição de carga de energia elétrica e por consequência seu preço se mostra fundamental para o desenvolvimento de estratégias que privilegiem o lucro nas operações de compra e venda de energia elétrica.

A disponibilidade de dados confiáveis de carga energética para os últimos 12 anos permite o desenvolvimento de um trabalho a partir de um cenário real. A evolução das linguagens de programação, bibliotecas e até poder de processamento computacional também propiciam a exploração de técnicas e modelos robustos a partir de ferramentas gratuitas executadas em hardwares de computadores pessoais. Por último, como importante motivação para a realização do estudo está o fato de que a partir da construção de modelos com predições satisfatórias é possível aplicar essas estruturas em operações reais.

1.2 Objetivos

Esse trabalho tem como objetivo propor e implementar modelos para a predição de carga de energia elétrica no Brasil. Para tal vão ser utilizadas técnicas de inteligência artificial, mais especificamente DL, na criação de modelos baseados nas arquiteturas de redes neurais recorrentes, LSTM de estruturas sequenciais e bidirecionais e GRU.

Para a realização das etapas de treinamento e teste devem ser utilizados dados reais de carga de energia no Brasil entre os anos de 2012 e 2021 fornecidos pelo Operador Nacional do Sistema Elétrico Operador Nacional do Sistema Elétrico (ONS). Quanto à avaliação dos dados do modelo, devem ser utilizadas as métricas de Erro Quadrático Médio (EQM), Raiz do Erro Quadrático Médio (REQM), Erro Médio Percentual Absoluto (EMPA) e Erro Médio Absoluto (EMA).

1.3 Organização

O trabalho se organiza em cinco seções: fundamentação teórica, trabalhos relacionados, metodologia, análise de resultados e conclusão.

Na primeira parte são contextualizados os principais temas que envolvem o trabalho. Entre eles, o mercado de energia elétrica brasileiro, inteligência artificial, redes neurais recorrentes e métodos de avaliação de resultados. Em seguida, são expostos alguns dos trabalhos de referência que compõem o estado da arte para predição de carga e preço de

energia elétrica e em diversos países e que utilizam exclusivamente de métodos computacionais para atingir seus objetivos. A terceira seção fundamenta a metodologia aplicada no desenvolvimento do trabalho, isto é, processos, softwares, dados e métodos de avaliação dos resultados. Seu principal objetivo é esclarecer os pormenores das etapas, suas atividades e motivações.

Tendo isso exposto, são relatados os resultados da aplicação de cada um dos processos pontuados na seção anterior. A partir disso é feita a análise dos resultados obtidos da aplicação da metodologia, além disso são feitas uma proposições para trabalhos futuros nesta mesma temática. Por último, está presente uma conclusão do trabalho desenvolvido.

Capítulo 2

Fundamentação Teórica

2.1 O Setor Elétrico Brasileiro e o Mercado Livre de Energia

O setor elétrico brasileiro tem sua origem em 1879, com o processo de iluminação elétrica da atual estação Central do Brasil (SILVA, 2011). Desde então o segmento vem se desenvolvendo em variadas frentes, não apenas na sua expansão para a complexa malha atual, mas também na sua regulamentação e ampliação de seus atores participantes. A evolução dessa infraestrutura serviu como pilar do processo de industrialização e urbanização brasileira ao longo do último século.

Segundo a Associação Brasileira de Distribuidores de Energia Elétrica (ABRADEE) é possível descrever a indústria da energia elétrica a partir de dois vieses, o técnico e o regulatório. O primeiro define que os atores da indústria são os geradores espalhados pelo território e as linhas de transmissão e distribuição que os interligam. É papel desses a coordenação do equilíbrio dinâmico entre o que é produzido e consumido, além da segurança de manutenção do serviço. Já o aspecto regulatório é formado pelos agentes que podem produzir, transportar ou comercializar a energia elétrica, garantindo que os processos de comercialização aconteçam de forma legal (ABRAADE, 2019).

Apesar de seu longo processo de evolução, é possível apontar o plano de Reestruturação do Setor Elétrico Brasileiro (RESEB) na década de 1990 como a última grande reforma do sistema brasileiro, e a mais relevante no contexto desse trabalho. Nesse período ficou definido o interesse nacional em reformar o setor para seguir uma política de estado regulamentador, isso é, assegurando a regulamentação e fiscalização do setor, mas não a execução de todas as suas atividades (LORENZO, 2001). Nesse sentido, o setor passa a

ser composto tanto de companhias privadas, quanto de autarquias e agências públicas, de forma a otimizar o serviços e segurança energética do país.

Apesar dos estudos realizados para o projeto do novo modelo, após uma grande crise energética são propostas novas alterações no setor em 2004. Dentre as mudanças está a criação de dois ambientes de comercialização, o regulado e o livre.

Segundo o art. 57 do Decreto nº 5.163/2004 (BRASIL, 2004), fica definido o ambiente de contratação regulado como aquele onde as operações comerciais são feitas entre os agentes vendedores e os de distribuição, nesse formato os consumidores - também chamados de cativos - estão limitados a consumir energia do fornecedor pertencente a sua rede. Já no ambiente livre, é legal que os consumidores (podendo ser livres ou especiais) possam escolher e negociar contratos de energia entre diferentes produtores de forma a decidir por quem o proporciona melhores condições.

A respeito dos consumidores livres é importante pontuar que apenas compradores que necessitem de uma carga igual ou maior que 10MW, com tensão também maior ou igual a 69kV podem se enquadrar nessa categoria. Para esse segmento também é vetada a compra de energia no mercado cativo, sendo obrigatória a comercialização com participantes autorizados dessa esfera.

Ainda relacionado ao Mercado Livre de Energia (MLE), os consumidores especiais são definidos como compradores de carga maior ou igual a 500kW a qual são permitidas as negociações com geradores de energia em regime de produção independente e autoprodução, ou de fontes solar, eólica e biomassa cuja potência seja menor ou igual a 30000kW, definido em § 5º, do art. 26 da Lei nº 9.648/1998 (BRASIL, 1998).

2.2 Ambiente de Contratação Livre

Conforme apresentado na seção 2.1, atualmente o sistema elétrico Brasileiro se divide em três tipos de consumidores, os cativos, especiais e livres. Enquanto o primeiro têm diversas restrições relacionadas aos seus possíveis fornecedores, aos dois últimos é permitida a negociação ou compra diretamente no mercado livre de energia. Características de cada um dos perfis podem ser encontradas na Tabela 1.

Segundo dados da Associação Brasileira dos Comercializadores de Energia (ABRACEEL) disponibilizados em 2022 (ABRACEEL, 2022) e relativos ao ano anterior, o MLE correspondeu a 34% de toda a energia consumida no país, além disso foi observado um aumento de 25% em relação ao ano anterior no ingresso de novos consumidores desse mercado. É possível entender a diversificação dos consumidores de energia no país no último ano, dado que esse era um setor antes bastante restrito.

Tratando dos atores que envolvem o conjunto, o Ambiente de Contratação Livre Brasileiro engloba os consumidores finais - ou seus representantes- e vendedores de energia que podem realizar negociações diretamente entre si. Estão disponíveis para compra a

energia convencional e incentivada (proveniente de fontes energéticas limpas, como eólica, por exemplo).

Quanto aos valores contratados, eles podem ser definidos a partir da negociação prévia entre as partes ou a partir do preço atual, o Preço da Liquidação das Diferenças (PLD). Esse último fornece a cada hora, dia ou ano um valor tido como ideal para a energia em cada uma das escalas temporais comercializadas no mercado livre. O preço então pode ser aplicado tanto na definição de contratos de compra e venda, quanto na cobrança do recurso utilizado sem negociação anterior.

Tratando especificamente dos contratos, eles são firmados em função da carga de energia contratada, a duração dos mesmos podem variar entre o longo e curto prazo, isto é, de anos a dias. Essa característica é de grande interesse uma vez que possibilita que os custos da energia sejam previstos no curto, médio ou longo prazo. Assim, é observada uma redução no risco de operação para o contratante (ABRACEEL, 2020).

Tabela 1 – Tipos de consumidores no mercado de Energia Elétrica brasileiro

Tipo	Descrição	Demanda mínima
Consumidor Cativo	Deve consumir energia elétrica exclusivamente da concessionária de sua região. Não tem poder de escolha ou negociação.	Não possui demanda mínima
Consumidor Especial	Deve consumir obrigatoriamente energia incentivada, isso é de fontes renováveis. Tem como vantagem descontos no consumo da demanda.	Maior que 500 KW e menor que 1.500 KW
Consumidor Livre	Pode adquirir energia convencional ou incentivada. Tem a possibilidade de negociar contratos segundo demanda ou conveniência	Maior que 1.000KW

Fonte: Tabela produzida pela autora

2.3 Sistema Interligado Nacional

O Sistema Integrado de Energia (SIN), criado a partir da Resolução Nº 351, de 11 de novembro de 1998 (BRASIL, 1998) e atualmente definido pela Resolução normativa ANEEL Nº 1.017, de 19 de abril de 2022 (BRASIL, 2022) tem como papel a produção e transmissão das matrizes energéticas hidrelétricas, térmicas e eólicas, estando atualmente o foco na primeira (ONS, 2022b). O controle das operações e estruturas do SIN pertence a ONS, já a fiscalização é responsabilidade da Agência Nacional de Energia Elétrica (ANEEL).

Além dos papéis mencionados anteriormente, o SIN também realiza a tarefa de gerir os recursos energéticos de modo a garantir a segurança e disponibilidade de acesso contínuo às cargas elétricas. Uma vez que sua atuação é presente em praticamente toda a extensão do

país (apenas o estado de Roraima está desconectado do sistema), podem ser estabelecidas conexões entre variadas regiões. Dessa forma, ao balancear as características climáticas e hidrográficas o SIN promove não apenas a otimização das cargas, como também a priorização da produção baseada em matrizes energéticas limpas.

Levando em conta seu papel gestor e aglutinador, bem como as proporções do país, o SIN divide o território entre quatro grandes sub-regiões - Sudeste/Centro-Oeste, Sul, Norte e Nordeste-. Além disso, existem também os Sistemas Isolados, eles não estão conectados na infraestrutura mencionada anteriormente dadas complicações geográficas ou inacessibilidade. Um mapa dos subsistemas e linhas de transmissão pode ser observado na Figura 1.

Figura 1 – Mapa dos diferentes subsistemas e Linhas de transmissão do SIN



Fonte: ONS (2022a, Mapa dinâmico do SIN)

2.4 Preço da Liquidação das diferenças

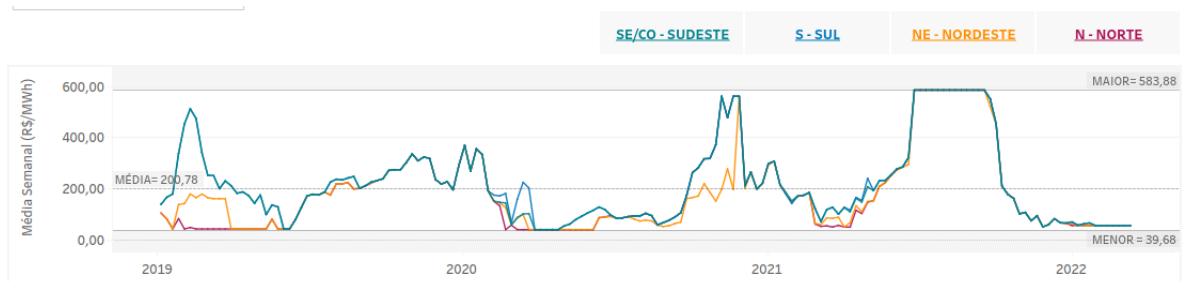
O PLD é obtido pela Câmara de Comercialização de Energia Elétrica (CCEE) a cada hora e para cada um dos subsistemas, o valor final, dado em reais por megawatt hora, é determinado por um modelo matemático que possui um limite máximo e mínimo determinado pela instituição e aplicado para cada cálculo.

O PLD é de grande relevância para as negociações de contratos no mercado livre de energia uma vez que tem como objetivo o equilíbrio entre a demanda, a produção e o

potencial de geração hidrelétrica, posto isso ele fornece o mecanismo a ser utilizado para encontrar a precificação ideal do cenário atual (GANIM, 2009).

É importante salientar que a modelagem aplicada para obter o PLD leva em conta como variáveis de entrada a geração hidrelétrica para todos os sub mercados, as condições climáticas, as demandas de energia, os preços dos combustíveis, o custo do deficit de energia e a relação de possibilidade de entrada de novos empreendimentos geradores de energia. Assim sendo, a volatilidade das condições de entrada e por consequência o preço do PLD (ver Figura 2) tornam a sua predição uma estratégia interessante para os agentes do mercado livre de energia.

Figura 2 – Gráfico do PLD por região



Fonte: CCEE (2022, Painel de preços CCEE)

2.5 Series Temporais

Uma série temporal nada mais é que uma sequência de anotações realizadas de acordo com uma escala temporal, a taxa de nascimentos ou carga de energia média consumida em um país são alguns exemplos desse tipo de série. Elas podem ser lineares ou não, discretas ou contínuas, além disso, são representadas por quatro parâmetros, tendência, ciclo, sazonalidade e aleatoriedade.

A tendência de uma série temporal indica a partir do curto prazo qual é a movimentação que ocorre de forma gradual no longo prazo do conjunto dos dados, como ascendente e decrescente. Já a sazonalidade representa os padrões de dados que se mostram com regularidade nos dados, um exemplo dessa característica seria o efeito da mudança de estações em observações pluviométricas em uma determinada área. Por sua vez, os ciclos em uma série se apresentam como variações periódicas, mas que não seguem os mesmos formatos ou intervalos das sazonais. Por último, a aleatoriedade assinala componentes que não podem ser claramente reconhecidos ou comportamentos sem padrão algum (DODGE, 2008).

Tendo as características das séries temporais em vista, o estudo e processo de análise das mesmas permite encontrar resultados bastante relevantes, como a predição de com-

portamento de um fenômeno, separação de ruídos de sinais, teste de hipóteses a partir de dados anteriores, simulações, entre outras aplicações (BROCKWELL; DAVIS, 2016).

2.6 Inteligência Artificial

Segundo Russell e Norvig (2020), é possível definir Inteligência Artificial (IA) por meio de oito perspectivas distintas, elas são, o esforço para fazer com que máquinas tenham mentes, a automação de atividades relacionadas ao pensamento humano, o uso de modelos computacionais para estudo da mente, a pesquisa da computação que torna possível o pensar, a arte de desenvolver máquinas capazes de realizar esforço humano, o estudo de como fazer a máquina superar o humano, o desenvolvimento de agentes inteligentes e por último, a ligação com comportamentos inteligentes. É interessante notar que todas as descrições citadas anteriormente têm sido trabalhadas por pesquisadores distintos através das últimas décadas. Assim, a tentativa de estabelecer uma definição única torna-se uma proposta reducionista para um campo vasto como o de IA.

Levando em conta a diversidade de campos do tópico, o Teste de Turing, proposto em 1950 por Alan Turing, define uma delimitação de aspectos necessários para que uma máquina opere como uma IA viável, (SAYGIN; CICEKLI; AKMAN, 2000). Para tal, é preciso o desenvolvimento de diferentes áreas correlatas às atividades que envolvem inteligência humana, elas são, processamento de linguagem natural (equivalente a fala), visão computacional (visão), robótica (movimentação), representação de conhecimento, raciocínio automatizado e aprendizado de máquina (aprendizagem).

Dada a ampla abrangência de áreas a serem exploradas no campo da IA é importante salientar que além da computação, a filosofia, matemática, economia, psicologia, linguística, entre outros são os domínios de estudo que servem como fundamentação do tema.

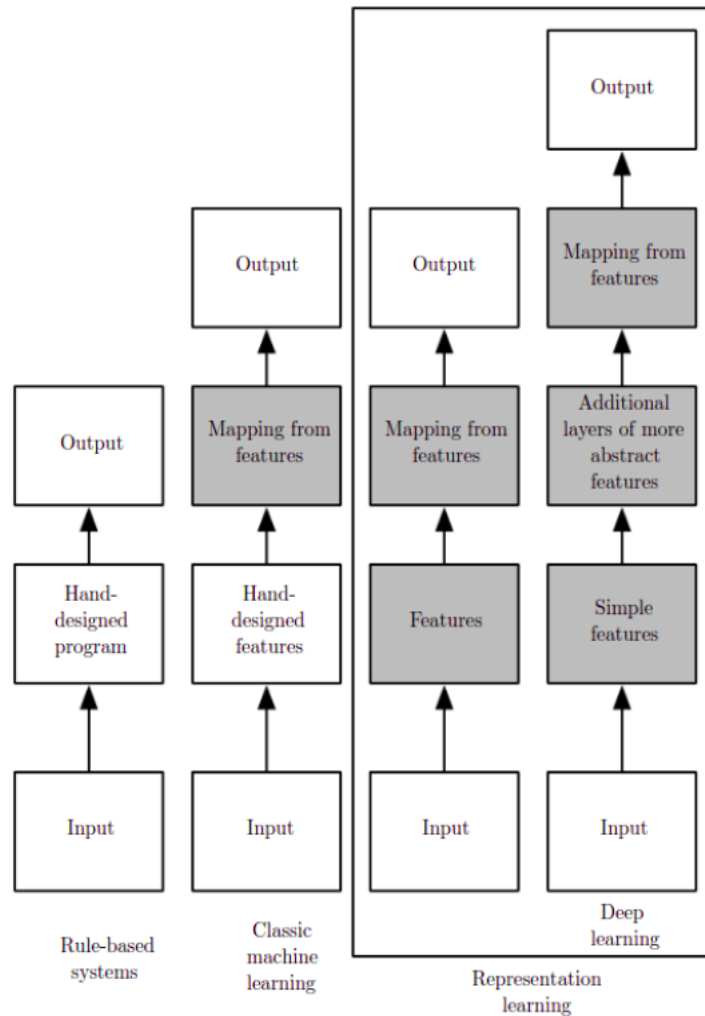
2.7 Deep Learning

Observando o início da história da IA nota-se que de forma inaugural a área esteve focada na solução de problemas tidos como difíceis para a cognição humana, mas com descrições matemáticas claras, e portanto, apropriadas para máquinas inteligentes (ZHANG et al., 2018). Contudo, em meio ao avanço do campo foi atestada a complexidade de descrição de atividades tidas como naturais para um humano, como a tradução de uma frase entre idiomas, ou a diferenciação entre duas espécies distintas de um animal.

Assim, Deep Learning se estabelece como uma sub-área da Inteligência Artificial, especificamente, Machine Learning, na qual se tem como fito o desenvolvimento de técnicas voltadas para problemas complexos nos quais não é possível estabelecer regras determinísticas, para tal, é necessário que a máquina aprenda uma hierarquia de conceitos a partir

de versões mais simples (DENG; YU et al., 2014). Além disso, um agente humano não deve precisar definir formalmente e extensivamente regras, isto é, deve ser possível inferir a partir da experiência, como explicitado na Figura 3. É possível apontar como dois grandes grupos de problemas que Deep Learning trabalha os de classificação e previsão.

Figura 3 – Hierarquia de conceitos, Deep Learning



Fonte: Russell e Norvig (2020)

A classificação resolve situações na qual a partir de um grande conjunto de dados o objetivo é separá-los em grupos de similaridades. O número e característica de interesse no agrupamento podem ser pré definidos ou não, um exemplo é a classificação automática de emails como spam.

Já o segundo segmento propõe a previsão de características ou comportamentos futuros com determinado nível de confiabilidade a partir da criação de modelos ou estruturas fundamentadas em dados de treinamento ou amostras de informações do passado. A predição de valores de carga de energia em um país é um exemplo de problema desse tipo.

2.8 Aprendizagem Supervisionada, Não Supervisionado e Semi-Supervisionado

Está definido que as técnicas de DL exigem que a máquina realize processos de aprendizagem, para tal é possível a aplicação de uma das três técnicas, aprendizado supervisionado, não supervisionado ou semi-supervisionado. Cada uma das três oferece um nível de liberdade para o formato dos dados de entrada, bem como sua aplicação é adequada para um conjunto de problemas diferentes.

Segundo Rezende, Monard e Carvalho (1999), “No aprendizado supervisionado é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido”. Isso é, além das variáveis de entrada, os resultados esperados no processo de aprendizagem de início devem ser fornecidos. Esse formato é apropriado para aplicações em problemas que envolvem a classificação de entradas entre categorias pré definidas, ou realizar um processo de regressão quando os resultados são uma função contínua.

Já o aprendizado não supervisionado envolve o treinamento apenas com as variáveis de entrada, nesse cenário a finalidade é a classificação dos valores em grupos somente a partir das características analisadas pelo algoritmo de DL, em outras palavras, a criação de uma representação dos dados (GHAHRAMANI, 2004). Os processos de agrupamentos e redução de dimensão exigem o uso da técnica.

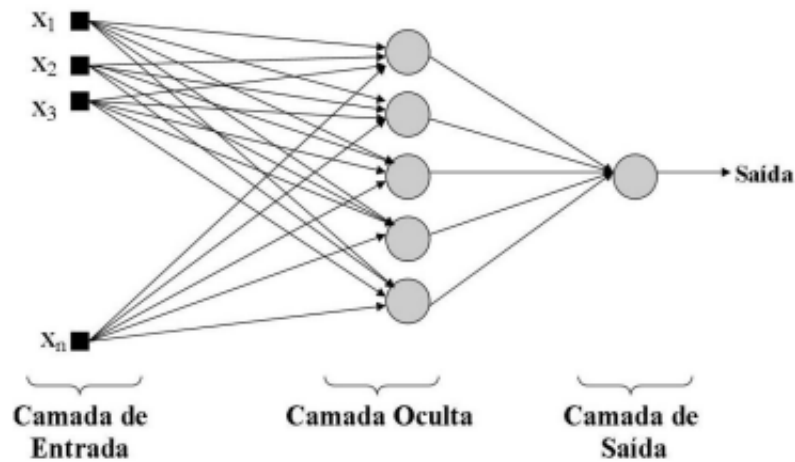
Por último, o processo semi-supervisionado engloba os dois anteriores ao permitir a entrada de dados de treinamento com ou sem a presença dos resultados esperados. Por exemplo, no caso de uma categorização um conjunto proporcionalmente menor ao total fornece categorias de saída, todo o restante dos dados devem ter suas classificações inferidas. O uso dessa metodologia é interessante, pois oferece um melhoramento para o processo de aprendizado não supervisionado (SANCHES, 2003)

2.9 Redes Neurais Artificiais

Redes Neurais Artificiais são estruturas de deep learning que mimetizam, dadas as devidas proporções, o comportamento de organizações biológicas, como o cérebro humano e seus neurônios. Os componentes primordiais da técnica são quatro variando apenas em arranjo e valores, eles são o neurônio, a camada, os pesos e a função de ativação (como representado na Figura 4).

Enquanto a quantidade de camadas, neurônios e função de ativação presentes na rede neural são definidos com antecedência, os pesos finais da mesma resultam inteiramente do processo de treinamento. É importante notar que o treinamento ocorre de forma iterativa, isto é, a cada execução - chamada de época - são usados os valores das variáveis

Figura 4 – Componentes e estrutura primordial de uma rede neural



Fonte: (GONÇALVES, 2003)

de entrada para definir o conjunto de pesos presentes em todas as camadas que mais se mostra adequado para obter os valores das variáveis dependentes do treinamento.

Tendo em vista as características mencionadas anteriormente, pode-se apontar a dificuldade de compreensão de como foram obtidos os valores dos pesos em cada época como um fator de complexidade da técnica. Além disso, como em qualquer modelo de aprendizagem supervisionada, o resultado final pode ter se moldado de forma excessiva aos dados fornecidos na etapa de treinamento, também chamado de *overfitting*.

Ainda assim, a possibilidade da construção de um modelo sem qualquer definição de características previamente a etapa de treinamento, e também o processamento de dados em diferentes formatos são grandes vantagens do uso de redes neurais (LIVINGSTONE; MANALLACK; TETKO, 1997).

Aprofundando a compreensão das estruturas que formam a rede neural, elas são dispostas entre três grupos, a de entrada, as intermediárias ocultas e a de saída, cada uma delas é composta por uma quantidade pré determinada de nós. Um nó, por sua vez, é composto por n entradas e uma saída, cada uma com um peso atribuído, à soma desses valores é então utilizada como entrada para a função de ativação do nó como mostrado na Imagem. O resultado é dado como valor de entrada para nós da próxima camada e assim sucessivamente até alcançar a camada de saída com o resultado final.

2.10 Função de Ativação

A escolha da função de ativação de uma rede neural é uma etapa primordial para alcançar resultados adequados na fase de testes. Uma vez que ela determina se um

neurônio vai ser ativado ou não, ela afeta diretamente a estrutura dos dados de saída (SHARMA; SHARMA; ATHAIYA, 2017). Tendo em vista a importância desse parâmetro, para otimizar a saída é relevante selecionar uma função similar ao comportamento desejado. Tomando como exemplo uma rede neural que pretende inferir um resultado binário, o uso de uma função de etapa binária é uma escolha viável para a modelagem.

Duas funções comumente utilizadas em problemas de deep learning são a Sigmóide - também chamada de logística- e a Unidade Linear Retificada *Rectified Linear Unit* (ReLU), cada uma pode ser conferida na Figura 5 e Figura 6 respectivamente.

Figura 5 – Função Sigmóide

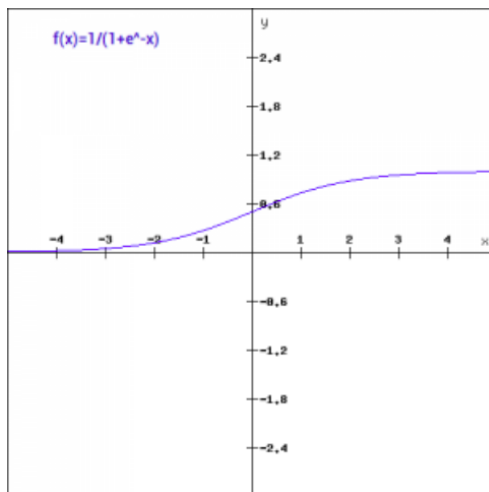
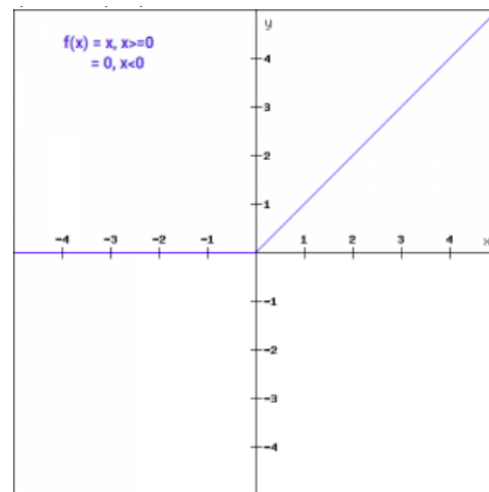


Figura 6 – Função ReLU



Fonte: (SHARMA; SHARMA; ATHAIYA, 2017)

A função logística é não linear e faz com que os valores de entrada sejam levados a pontos extremos da função, dessa forma, seu resultado é sempre entre um intervalo estabelecido. Também não linear, a ReLU possibilita que caso o valor de entrada seja negativo o nó simplesmente não se ative. Isso é interessante, pois permite que ligações entre menos nós sejam levados em conta (DING; QIAN; ZHOU, 2018), e portanto sua aplicação em contextos adequados tende a exigir menor computação (especialmente em aplicações de redes de múltiplas camadas intermediárias). Por último é necessário pontuar que ambas as aplicações podem apresentar o problema da dissipação do gradiente.

2.11 Redes Neurais Recorrentes

Como mencionado anteriormente, redes neurais são compostas de elementos básicos, variando em sua composição e parâmetros, esse é o caso da família de redes neurais recorrentes. Nesse tipo de arquitetura o processamento de dados em uma camada pode ter duas formas de execução, apenas a passagem para outra camada (seja ela a próxima ou a anterior) ou a realimentação da mesma, criando assim um comportamento similar a memória (MEDSKER; JAIN, 1999). A realimentação dos dados no formato de um

loop permite que além do treinamento convencional, a rede também seja capaz de reter informações de iterações anteriores, esse aspecto é ideal para aplicação em problemas de séries temporais, processamento de linguagem natural, entre outros.

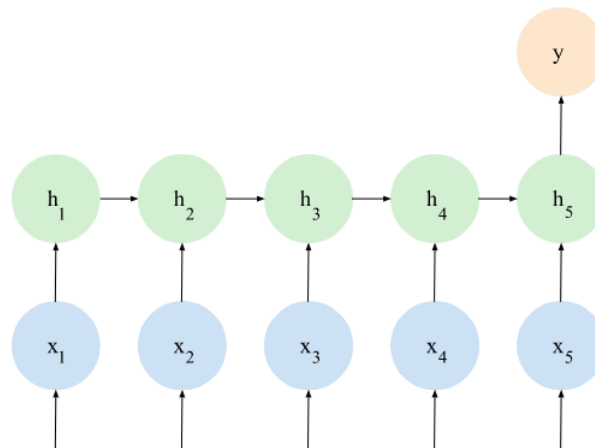
Nas próximas seções vão ser descritas três arquiteturas desse tipo, Many to One, LSTM e GRU.

2.11.1 Many-to-One e Many-to-Many

Na estrutura do tipo Muitos-para-Um (Many-to-One) temos uma rede neural com um número finito de entradas e apenas uma variável de saída, como representado na Figura 7. Essa estrutura é importante no contexto de redes neurais recursivas uma vez que permite o uso de variáveis dependentes que formam um contexto.

Um exemplo de aplicação do formato muitos para um são Redes Neurais Recorrentes (RNR) para a classificação de sentimentos a partir de textos, nesse cenário cada palavra possui uma qualidade atrelada a ele, contudo juntas as palavras geram um contexto que pode ser classificado de forma mais completa. Alguns usos para essa arquitetura são classificações e predição de séries (AMIDI; AMIDI, 2022).

Figura 7 – Estrutura Muitos-para-Um



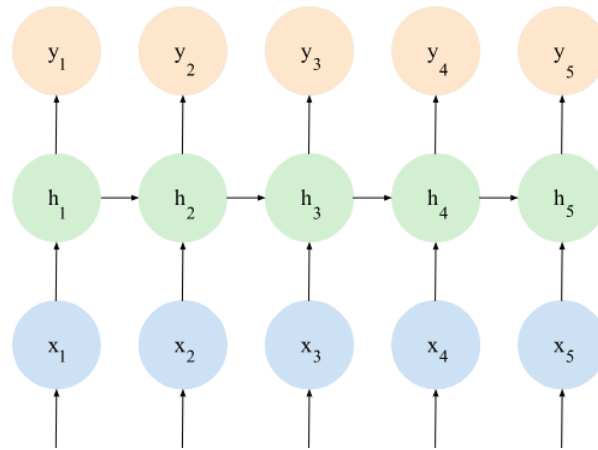
Fonte: Imagem produzida pela autora

Já na arquitetura Muitos-para-Muitos (Many-to-Many) temos um número finito de entradas, bem como outro quantidade de saídas. Nesse formato, além de proporcionar um contexto na estrada ele também é capaz de gerar um conjunto de informações na saída Figura 8.

Para esse caso, um exemplo de aplicação apropriada seria a construção de uma RNR com o propósito de prever um comportamento dado por uma série temporal. A modelagem nesse caso tomaria como entrada um conjunto de informações sequenciais referentes aos

último dias da série. Como saída geraria outro grupo de dados, múltiplas previsões dos próximos pontos.

Figura 8 – Estrutura Muitos-para-Muitos



Fonte: Imagem produzida pela autora

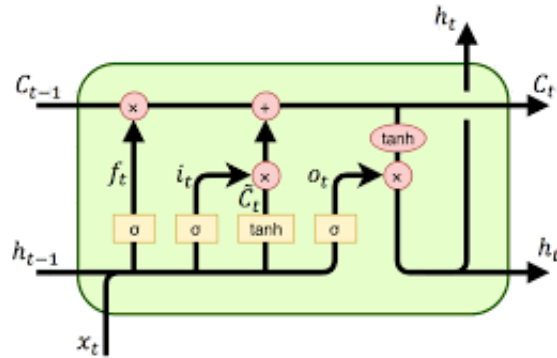
2.11.2 Long Short Term Memory

As redes neurais do tipo LSTM foram apresentadas pela primeira vez em 1997 (HOCHREITER; SCHMIDHUBER, 1997) se diferenciam em relação a outras do mesmo gênero por sua capacidade de reter informações a longo prazo. Ela foi projetada tendo em vista o problema de explosão do gradiente e é adequada para a aplicação em cenários de classificação e predição de séries temporais.

Seu diferencial está na presença de três gates (chaves) além da célula tradicional, a de entrada, saída e “esquecimento”, eles definem quais informações devem ser retidas na rede, e por assim dizer, na memória (Figura 9). Respectivamente, a entrada usa a função de ativação para avaliar a relevância dos dados e produzir um vetor de elementos que devem ser utilizados na rede, já a próxima aplica o resultado da entrada seguido da saída da última iteração para ponderar a importância dos dados recebidos, o resultado é então aplicado a uma função de ativação binária. Por último, o portão de saída, no qual é aplicada a função sigmoide para selecionar as saída da célula.

É interessante notar que arquitetura LSTM pode ainda se apresentar de diversas formas, entre elas as do tipo Feed-Foward e Bidirecional. Enquanto a primeira permite a propagação em apenas uma direção, a segunda propõe a aplicação da informação em dois sentidos. Isso significa que a rede tem como entrada tanto a sequência ordenada, quanto o seu inverso (SCHUSTER; PALIWAL, 1997). Essa característica é bem aproveitada em contextos nos qual uma informação anterior pode afetar a futura.

Figura 9 – Arquitetura LSTM

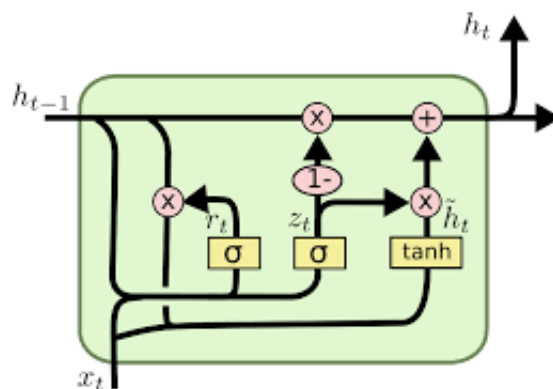


Fonte: (OLAH, 2015)

2.11.3 Gated Recurrent Units

A arquitetura das GRU também é proposta como uma RNR que pretende mitigar o problema do gradiente. Similar a LSTM no que diz respeito a aplicações, sua estrutura se distingue em seus gates, sendo implementados portões de atualização e reset da célula (Figura 10). Enquanto o primeiro permite que seja selecionada a quantidade de informações que devem ser passadas adiante o segundo filtra os dados a serem mantidos ou negligenciados. Essa técnica se mostra bastante importante ao acrescentar o gate de atualização, e assim mitigar o efeito do do gradiente reduzido (DEY; SALEM, 2017).

Figura 10 – Arquitetura GRU



Fonte: (OLAH, 2015)

2.12 Treinamento

Para obter um modelo válido a partir de uma rede neural é preciso antes realizar uma etapa de treinamento, são necessários dados de entrada adequados para a aplicação na rede, bem como dos seus resultados esperados equivalentes. Tendo em mente a estrutura e propósito de aplicação das RNRs é interessante aplicar diferentes formatos para os dados de entrada em vez de apenas aplicá-los todos de forma sequencial.

Tratando especificamente de problemas de séries temporais, a definição de um período de tempo A de dados de entrada a serem aplicados para uma outra quantidade de dias B de saída do modelo é uma técnica bastante utilizada que é chamada de Janelas Deslizantes (CHU, 1995). A técnica define dois tamanhos padrão, um para a janela de observações de entrada e outro para o intervalo de saída. Em seguida são utilizados os intervalos estabelecidos para o treinamento, sempre utilizando um recorte sequencial nos valores decididos. Para a aplicação otimizada da técnica é necessário ter em mente as características dos dados utilizados no problema e sua autocorrelação de forma a definir períodos coerentes.

2.13 Normalização

Para a aplicação em redes neurais é necessário que os dados das variáveis independentes tenham certo grau de uniformidade. Isso é, é preciso que eles sigam um formato único, como decimal e preferencialmente estejam dentro de um intervalo de números inteiros pré definidos (BHANJA; DAS, 2018). Assim, para que não exista uma grande variação na escala dos dados é aplicado o processo de normalização.

A função *MinMaxScaler* é uma das técnicas amplamente utilizadas no processo de normalização. Para encontrar o equivalente nessa nova escala deve ser utilizada a Equação 1. Nela é subtraído o valor de máximo do mínimo desejado na nova escala, em seguida ele é multiplicado pelo valor de resultado da divisão nos quais os parâmetros x , $x_{mínimo}$ e $x_{máximo}$ significam, respectivamente, valor de entrada, valor mínimo e valor máximo antes da normalização. Por último, o valor encontrado é somado ao valor de mínimo dos novos valores.

$$x_{norm} = \frac{(máximo - mínimo)(x - x_{mínimo})}{x_{máximo} - x_{mínimo}} + x_{mínimo} \quad (1)$$

2.14 Métodos de Avaliação de Resultados

Para avaliar a qualidade de um modelo é preciso encontrar métricas que comparem os valores obtidos aos reais, isso é quantifiquem erros. A partir dessa variável é possível validar ou não as predições, além de fornecer um diagnóstico para o desenvolvimento

do modelo. Nesta sessão vão ser tratados quatro métricas diferentes de avaliação de resultados que são comumente utilizadas para modelos de deep learning.

2.14.1 Erro Quadrático Médio

O EQM é uma fórmula estatística para a avaliação da diferença média entre um valor de referência e um segundo. Ele é comumente utilizado para a quantificação de erros em modelos matemáticos e medição da acurácia do mesmo. Seu resultado pode ser encontrado através da aplicação da Equação 2, sendo x o valor de referência, \hat{x} valor a ser avaliado e n o número de observações.

$$EQM = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|^2 \quad (2)$$

2.14.2 Raiz do Erro Quadrático Médio

O REQM é uma fórmula dada pela raiz do EQM. Por ser aplicada a raiz quadrada, seu resultado é mais sensível a erros, além disso, a técnica retorna a mesma escala dos valores esperados o que torna sua aplicação bastante interessante. Seu calculo está explicito na Equação 3, da mesma forma x , \hat{x} e n representam a o valor esperado, o previsto a quantidade de pontos avaliados.

$$REQM = \sqrt{\frac{\sum_{i=1}^n |x_i - \hat{x}_i|^2}{n}} \quad (3)$$

2.14.3 Erro Médio Percentual Absoluto

O EMPA é um método estatístico para avaliar a acurácia do modelo a partir do erro de predição sobre um conjunto de observações. Seu resultado é definido pela Equação 4, ele é encontrado por meio da soma da diferença absoluta de cada par de observações, seguida da divisão pelo número total de apontamentos. Quanto menor o valor maior é a acurácia do modelo.

É importante ressaltar que sua aplicação não é adequada para todos conjuntos de dados. Uma vez que o método se utiliza do valor esperado como divisor, caso esse tenha valores pequenos e próximos a zero o resultado da métrica cresce podendo tender a um número infinito.

$$EMPA = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - \hat{x}_i|^2}{x_i} 100\% \quad (4)$$

2.14.4 Erro Médio Absoluto

O EMA calcula a média de erros a partir do modulo do erro entre o valor real e o esperado para cada uma das observações. Desse modo, erros negativos são mitigados e é

possível ter uma avaliação mais clara do erro médio do modelo. Seu calculo é dado pela Equação 5.

$$EMA = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (5)$$

Capítulo 3

Trabalhos Relacionados

Nesta seção vão ser explorados alguns estudos de diferentes fontes relacionados à temática desenvolvida no trabalho. Os autores mencionados no decorrer do capítulo servem como aprendizado e referência do estado da arte no contexto de previsão de séries temporais e também, especificamente, carga no setor de energia elétrica. Dentre os pesquisadores do tema é possível observar a exploração de técnicas variadas na criação dos modelos. É notável na literatura o uso de métodos de inteligência artificial e estatística, além disso se fazem bastante presentes as etapas de pré-processamento dos dados, como seleção de atributos e escalas temporais.

No trabalho intitulado “*A Study on Neural Networks Approach to Time Series Analysis*” Katarya e Rastogi (2018) os autores têm como objetivo a exploração de algumas técnicas comumente aplicadas na análise de séries temporais. O método ARIMA é o primeiro a ser citado, já que é também o precursor deste campo. Ainda que o mesmo seja relevante, é apontado que a generalização do resultado como um modelo linear pode gerar previsões divergentes das esperadas.

Ao longo do artigo são apontadas três técnicas de redes neurais, *Feed Forward*, recorrentes e LSTM, mesmo que essas tenham uma convergência lenta elas são avaliadas como aproximadores universais para predição de séries temporais. Tratando da estrutura LSTM, sua capacidade de retenção de informações é evidenciada como importante característica.

Por fim é interessante apontar a discussão a respeito de Máquinas de Aprendizado Extremas, isto é, a aplicação de diferentes algoritmos de treinamento em Aprendizado de Máquina (AM). São abordadas as variações da técnica *Online Sequential Extreme Learning Machine* (OSELM), seu desenvolvimento foi motivado pela convergência lenta dos métodos *backpropagation*. Ao introduzir um sistema *fuzzy* de inferência, a técnica

produz resultados com acurácia.

Tendo analisado um trabalho que explora a predição de séries temporais de modo geral é possível agora contextualizar trabalhos que contribuam de modo direto à pesquisa de modelos de energia elétrica. No artigo “*Multidimensional Feeding of LSTM Networks for Multivariate Prediction of Energy Time Series*” (SUCCETTI et al., 2020) é aprofundada a comparação entre redes LSTM de apenas uma e múltiplas variáveis de entrada, assim como duas arquiteturas diferentes para a rede.

Quanto às estruturas LSTM construídas, sua única diferença está na quantidade de camadas recorrentes existentes na rede, enquanto o primeiro modelo parte de uma única faixa o segundo aplica duas *layers* do tipo LSTM de forma sequencial. O desempenho de predição será avaliado a partir do treinamento e aplicação de dados de produção energética de uma planta fotovoltaica nos Estados Unidos, a carga de saída, a temperatura e a velocidade do vento são colocados como variáveis independentes.

Feita a normalização dos dados é estabelecida uma janela de entrada de dois meses para horizontes de predição de 1, 3 e 7 dias. Ainda que os resultados não indiquem um consenso quanto a quantidade de dias a serem previstos ou um grupo de variáveis de entrada que devem ser aplicados para otimizar o modelo, é observado que o uso de duas camadas sequenciais LSTM apresenta uma melhora em 70% dos casos estudados.

Enquanto a pesquisa anterior (SUCCETTI et al., 2020) explora de forma mais restrita as técnicas de predição do modelo os trabalhos *Efficient Short-Term Electricity Load Forecasting for Effective Energy Management* (KHAN et al., 2022) e *Energy Load Forecasting Using a Dual-Stage Attention-Based Recurrent Neural Network* (OZCAN; CATAL; KASIF, 2021) propõem a construção de múltiplas estruturas para a otimização do modelo. Ambos os estudos têm como objetivo de predição os dados de consumo de energia em uma única residência na França a cada minuto entre os anos de 2006 e 2010 por meio de arquiteturas híbridas.

A arquitetura desenvolvida pelo primeiro trabalho (KHAN et al., 2022) consiste de duas etapas, a limpeza dos dados e a aplicação dos mesmos em uma rede composta por uma camada convolucional residual seguida de uma rede do tipo LSTM. Como janelas temporais são utilizados intervalos de 120 minutos de entrada para 60 de saída e 48 horas para a predição de um dia completo.

O uso de uma rede convolucional na entrada se destaca nesse trabalho, essa aplicação é justificada ao pontuar que a técnica tem seu uso apropriado nas atividades de classificação e processamento de linguagem natural. Isso posto, o uso dessa arquitetura como entrada oferece à próxima etapa dados cujas características mais relevantes já foram extraídas. Foram obtidas métricas de resultados melhores que qualquer outra estrutura simples como regressão linear, máquina de suporte vetorial, GRU, entre outros.

Já no estudo proposto por Ozcan, Catal e Kasif (2021) no lugar da camada convolucional residual é aplicado o mecanismo de *Attention*. Essa estrutura usa os dados de

entrada da rede para estabelecer pesos de relevância, em outras palavras, ela permite que características importantes da rede sejam mantidas com facilidade, criando assim um contexto. Além do uso de *Attention* na entrada da arquitetura, ela também é presente entre duas camadas LSTM.

Em conformidade com os trabalhos apresentados anteriormente nessa seção, a etapa de pré-processamento dos dados também é realizada, é feita a imputação dos dados faltantes e normalização dos mesmos. Para a modelagem são aproveitadas todas as variáveis do *dataset*. Novamente, em comparação com modelos de arquitetura consolidada, como LSTM e GRU os resultados obtidos foram de modo geral melhores.

Avaliando ambos os trabalhos Khan et al. (2022) e Ozcan, Catal e Kasif (2021), é possível atestar uma adequação maior de modelos híbridos na predição do conjunto de dados explorado. Ainda assim seria interessante avaliar comparativamente as arquiteturas partindo dos mesmos parâmetros de entrada e saída de modo a entender com mais profundidade os pontos que mais contribuem na otimização da predição.

Seguindo uma abordagem distinta as anteriores, em “*An effective deep learning neural network model for short-term load forecasting*” (LI et al., 2020) os autores não apenas realizam uma comparação entre modelos diferentes de redes neurais mas também propõem uma *loss function* nova. Partindo do objetivo de predição de carga horária para um dia completo em um sistema, são trabalhadas propostas de redes perceptron, convolutivas e recorrentes usando tanto funções de perda já estabelecidas quanto o novo algoritmo desenvolvido.

Tendo isso em vista, são utilizados dois conjuntos de dados diferentes para a construção dos modelos. Em ambos são selecionadas diversas variáveis de entrada, entre elas, data, dia da semana, temperatura, feriados, entre outros. Quanto a nova função, ela é relevante no contexto desse artigo uma vez que utiliza os valores horários em um mesmo dia no cálculo da perda.

A partir da etapa de validação e testes foram obtidos resultados para a nova *loss function* entre 7.3% e 2.82% no que diz respeito a métrica EMPA. É interessante apontar que dentre as modelagens de redes utilizadas a que sobressaiu foi a do tipo convolucional.

A proposta de predição de carga de energia em smart grids por meio de redes neurais profundas será explorada como último estudo da seção. Em “*Energy load forecasting model based on deep neural networks for smart grids*”, (MOHAMMAD; KIM, 2020), os autores constroem duas arquiteturas distintas *Deep Feed-Forward Neural Network* e *Deep Recursive Neural Network* que modelam dados coletados do *New York Independent System Operator* e fazem uma análise comparativa dos resultados encontrados.

O pré-processamento dos dados é apresentado como primeira etapa da modelagem descrita no artigo. A limpeza, normalização e reestruturação da base de dados estudada são as atividades fundamentais nesse momento. Dados incorretos ou vazios são substituídos pela média das observações, o processo de suavização também é aplicado de forma a

minimizar pontos extremos. Posto isso, são selecionados atributos tidos como relevantes para a modelagem a partir da experiência e conhecimento dos autores.

Ao construir as redes do artigo são exploradas algumas variações de números de neurônios nas camadas ocultas, camadas e funções de ativação como métrica para seleção do melhor modelo dentre as variações propostas deve ser utilizada a função de erro médio percentual absoluto. Dados os parâmetros de seleção das redes elas foram comparadas com outras variações mais simples como de regressão linear e redes neurais simples. A projeto de rede *Deep RNN* com função de ativação tangente, de duas camadas com 10 neurônios em cada, superou todas as outras propostas com EMPA de 0.61% contra valores entre 0.68% a 2.88%.

Tendo discriminados alguns dos trabalhos que integram o estado da arte no contexto de predição de séries temporais e carga de energia é possível confirmar a importância da etapa de pré-processamento em todos os trabalhos. O processo de seleção das variáveis independentes e tamanhos de janelas de entrada e saída dos modelos estão presentes também em muitos dos trabalhos mencionados. Por fim, foi notada a exploração de diferentes arquiteturas de redes neurais bem como o uso de redes neurais profundas como artifício principal para a otimização dos resultados.

Capítulo 4

Metodologia

Este capítulo descreve os métodos, processos e ferramentas utilizadas no desenvolvimento do experimento, tendo sempre como objetivo a predição de valores de carga elétrica média na escala diária para o mercado livre de energia. Além disso, a seção também pormenoriza a base de dados sobre a qual o trabalho é aplicado. O estudo proposto consiste das seguintes etapas:

1. Aquisição dos dados históricos de carga
2. Limpeza dos dados
3. Análise exploratória da série temporal
4. Pré processamento dos dados
5. Definição das modelagens a serem aplicadas
6. Treinamento das Redes Neurais
7. Avaliação dos resultados

4.1 Ferramentas

Para a execução das redes neurais será utilizada a linguagem de programação Python na versão 3.7. Para a realização do pré-processamento dos dados e análise dos resultados será utilizada a biblioteca Pandas, posto que ela fornece funções flexíveis para aplicação em DataFrames. A implementação de gráficos e imagens será conduzida com o uso das

bibliotecas Matplotlib e Seaborn. Quanto às técnicas de *Deep Learning*, devem ser aplicadas as funcionalidades da biblioteca TensorFlow, aliada a API Keras com a finalidade de criar e treinar as redes neurais, seu uso é interessante uma vez que ela é gratuita e de alto nível. Todas as tecnologias citadas acima são tanto gratuitas quanto *open-source*, além disso possuem grandes comunidades ativas e ampla documentação.

No que diz respeito à infraestrutura de desenvolvimento, foi utilizada unicamente a IDE Pycharm na sua distribuição profissional e um repositório de versionamento criado pela autora e disponibilizada de forma online por meio da plataforma GitHub.

4.2 Dados

Os dados utilizados para a realização do estudo foram obtidos no Portal de Dados Abertos do ONS, eles são divulgados e mantidos de forma pública pelo Operador Nacional do Sistema Elétrico, sendo divulgados mensalmente. O serviço disponibiliza dados em diferentes granularidades, de acordo com as escalas de precificação aplicadas no mercado livre de energia. É relevante apontar que a plataforma fornece informações agrupadas por ano desde 2000, tanto em formato csv, quanto xlsx.

Tendo em vista o objetivo de predição de carga no curto prazo, a escala adequada para a aplicação dos estudos é a diária, isto é, uma linha de anotação por data. Para a realização da pesquisa deve ser utilizado o conjunto de dados agrupados entre os anos de 2012 até dezembro de 2021.

No que diz respeito às variáveis contidas na base de dados, todos os anos publicados contém a mesma seleção de atributos, eles são, carga elétrica (definida por megawatt médio), data de referência, subsistema e seu respectivo código. Dada a estrutura geográfica estabelecida pelo ONS os dados são categorizados a partir do subsistemas do SIN (Norte, Nordeste, Sul e Sudeste). Quanto aos tipos das variáveis, elas se encontram descritas na Tabela 6:

Tabela 2 – Variáveis e suas definições.

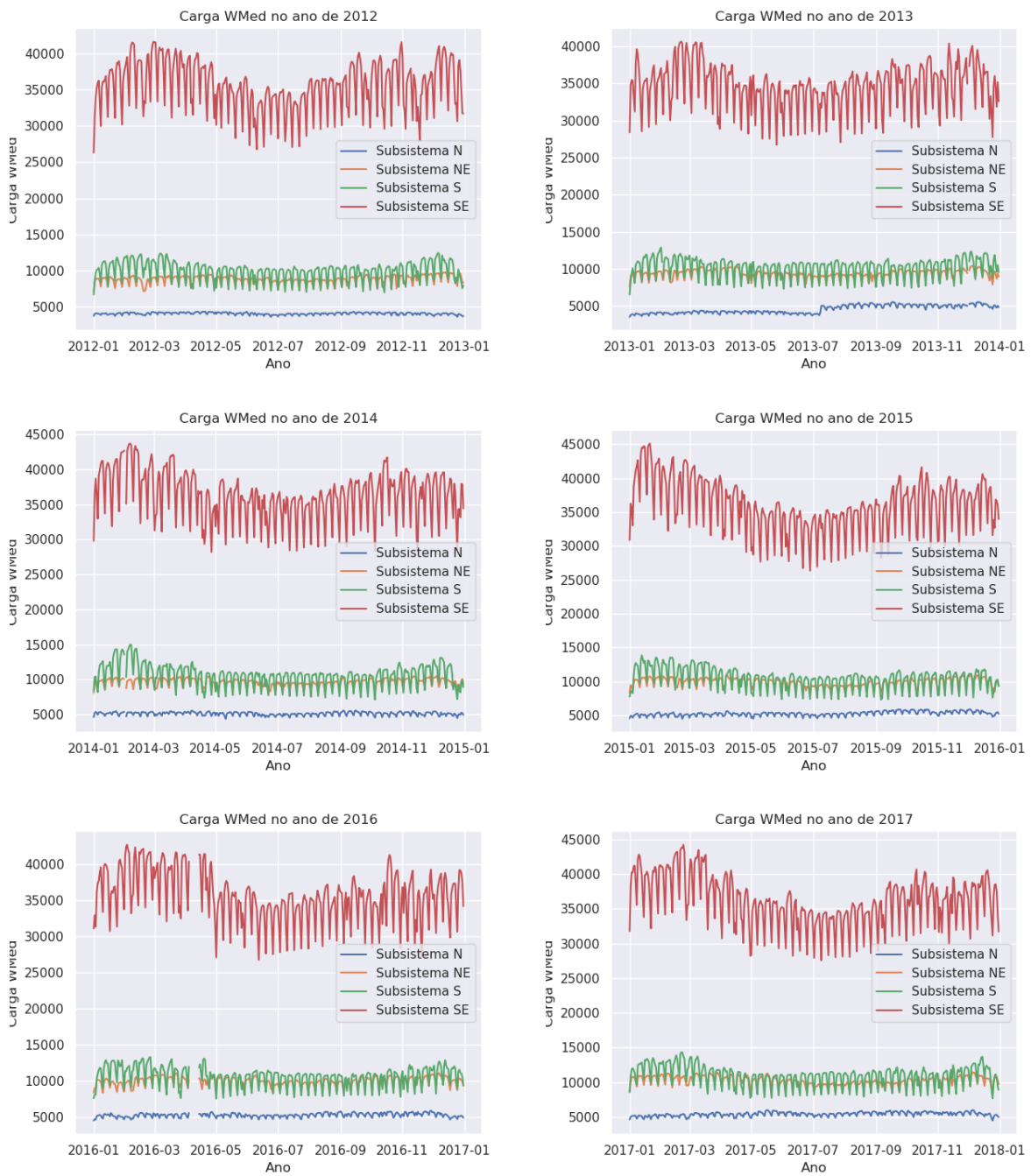
Variável	Tipo	Significado	Exemplo
Carga MWmed	Decimal	Carga em MWat médio para o subsistema indicado na data de referência	12000.00
Subsistema	Literal	Subsistema do SIN para qual a anotação se refere	NORTE
ID	Literal	Sigla do subsistema	NO
Data	DateTime	Data na qual a carga foi observada	08-01-2022

Fonte: Tabela produzida pela autora

Tratando das anotações de carga definida pelo dataset, como pode ser observado na Figura 11 e Figura 12, os valores diferem amplamente entre os quatro subsistemas previstos, além disso existe uma certa manutenção do alcance de carga entre os subsistemas

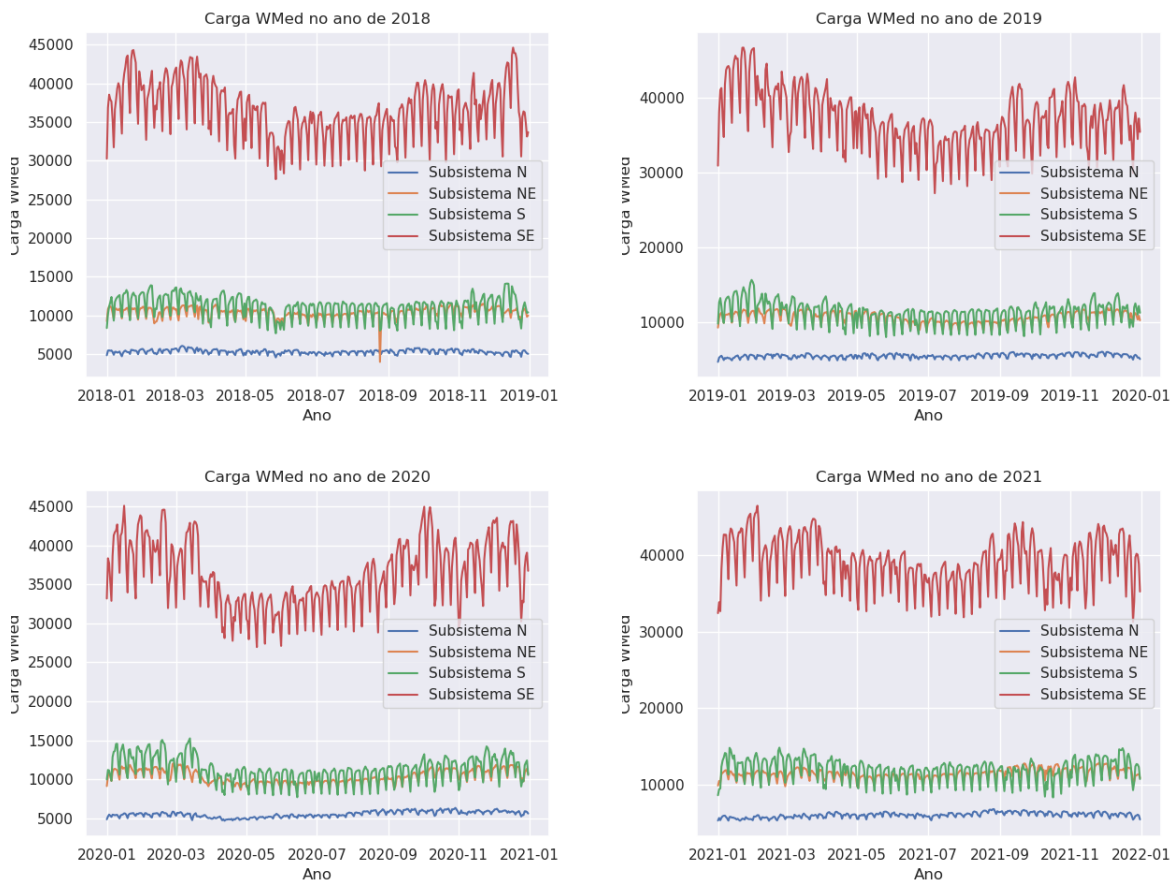
através dos anos. Dessa forma, o conjunto total dos dados se divide em 4 séries distintas que devem ter todos os seus processos realizados separadamente.

Figura 11 – Valores de carga para cada subsistema entre 2012 e 2017



Fonte: Imagens produzidas pela autora

Figura 12 – Valores de carga para cada subsistema entre 2018 e 2021



Fonte: Imagens produzidas pela autora

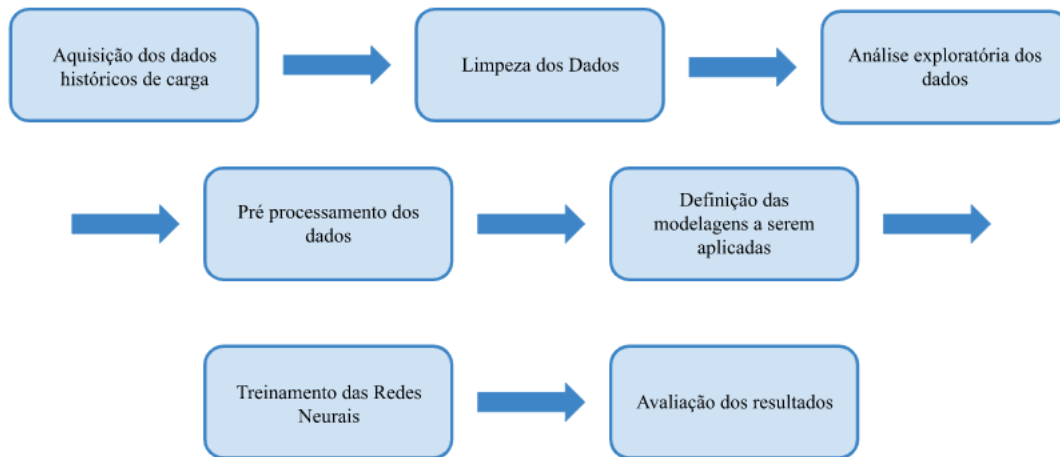
4.3 Etapas do estudo proposto

De forma a adequar os dados obtidos para o treinamento das redes neurais é necessário realizar uma série de pré-processamento nos dados. Esses processos vão desde a avaliação de informações faltantes, até a adição de novas variáveis e avaliação de correlação entre as propriedades, um diagrama da sequência de processos desenvolvidos pode ser observada na Figura 13.

4.3.1 Aquisição dos dados históricos de carga

Como primeira etapa a ser realizada está a adição de outras quatro variáveis independentes no modelo, o dia da semana, dia, mês e ano da variável a ser prevista. Assim, ficam estabelecidas como variáveis independentes as anteriores e as adicionadas, totalizando oito.

Figura 13 – Diagrama das etapas de desenvolvimento da modelagem



Fonte: Imagem produzida pela autora

4.3.2 Limpeza dos dados

Como próximo passo foi realizada a limpeza dos dados, isto é, verificação de anotações nulas ou incorretas, caso existam ocorrências desse tipo o procedimento realizado será a imputação dos dados.

4.3.3 Análise exploratória dos dados

A análise exploratória dos dados fornecerá como resultados o aprofundamento dos comportamentos e características de cada um dos subsistemas vistos como séries temporais individuais. Devem ser implementadas visualizações como a correlação entre as variáveis independentes e a ser prevista, além da correlação entre pontos da variável dependente de forma a entender se existem comportamentos cíclicos ou sazonais.

4.3.4 Pré processamento dos dados

A etapa de pré-processamento tem como principais atividades a normalização dos dados, bem como a seleção dos atributos relevantes para a predição e seleção do tamanho da janela de observações de entrada e saída. Essa etapa é realizada com o objetivo de aumentar a acurácia do modelo final. Para tal, todas as variáveis são normalizadas para valores entre zero e um, no caso de datas e dias da semana, os valores foram convertidos para valores numéricos crescentes e em seguida normalizados. Já a seleção de variáveis teve os resultados da análise exploratória como fontes do embasamento para a tomada de decisão. Atributos de alta correlação absoluta foram selecionados, e aspectos cíclicos das séries serviram como base para a definição dos intervalos de *input* e *output* do modelo.

4.3.5 Definição da modelagem a ser utilizada

Foram escolhidas três técnicas de DL, especificamente Redes Neurais Recorrentes para a realização da predição. Podendo ser consideradas de complexidade crescente elas são, LSTM do tipo *Feed Forward*, LSTM bidirecional e GRU. Todos os parâmetros aplicados foram encontrados de forma empírica.

Como função de ativação das redes neurais será utilizada a tangente hiperbólica. Já a respeito da estrutura das redes elas são definidas pela mesma arquitetura, duas camadas do tipo LSTM ou GRU sequenciais com 200 neurônios cada, em seguida duas camadas intermediárias com 40 e 20 neurônios cada, por último está presente uma camada de saída com 7 neurônios. Também foi aplicado um índice de *Dropout* de 25%.

O otimizador aplicado nas modelagens será o Adam. Já o tamanho do lote de entrada será de 28 dias, enquanto a janela de saída de 7.

4.3.6 Treinamento das Redes Neurais

Para o processo de treinamento, validação e teste o *dataset* inicial deve ser dividido em 70% treinamento, 20% validação e 10% teste. O modelo deve ser treinado por 150 épocas, podendo ter sua parada antecipada pelo delta de erro menor que 0.1 por ao menos 10 vezes seguidas resultante da aplicação do conjunto de validação no treinamento.

O mesmo método de treinamento foi aplicado para os três tipos de redes.

4.3.7 Avaliação dos resultados

Para a avaliação dos resultados são definidas quatro medidas de avaliação dos resultados, erro quadrático médio, raiz quadrada do erro quadrático médio, erro absoluto médio e MAPE.

Capítulo 5

Desenvolvimento e Análise dos Resultados

Conforme descrito na seção de Metodologia, o desenvolvimento e análise de resultados descritos nesta seção tem como propósito a construção de um modelo de predição de valor de Carga elétrica média. O cenário da pesquisa é o mercado livre de energia com base nos dados diários de quatro subsistemas elétricos brasileiros. Para tal, as seguintes etapas foram aplicadas de forma sequencial: aquisição dos dados históricos de preço, limpeza dos dados, análise exploratória da série temporal, pré-processamento dos dados, definição das modelagens a serem aplicadas, treinamento das redes neurais e finalmente a análise dos resultados.

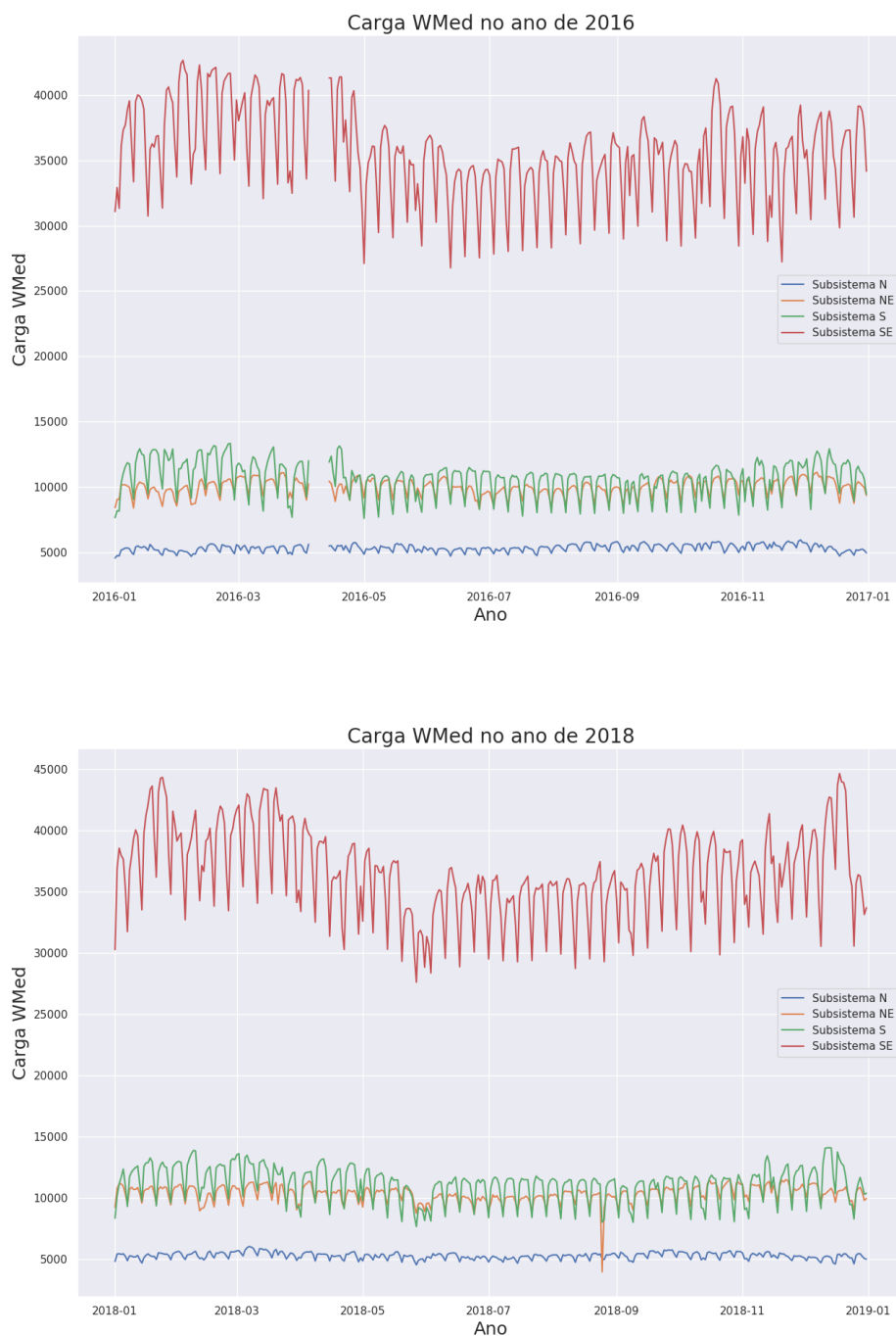
5.1 Desenvolvimento

Tendo importados os dados foi então preciso realizar a avaliação e limpeza dos mesmos. Foi possível observar a existência de valores incorretos ou não numéricos tanto por meio de funções da linguagem Python quanto por métodos visuais como uso dos gráficos como mostrado na Figura 14.

Constatada a necessidade da limpeza dos dados foi feita a imputação dos mesmos. Para tal, foi aplicada a duplicação do valor anterior à anotação faltante por meio de uma função. É importante mencionar que dadas as características do conjunto foi preciso usar o valor da data anterior e do mesmo subsistema para o qual o dado era problemático.

Em seguida foram renomeadas as colunas do *dataset* de forma a melhorar a legibilidade dos nomes. Além disso, a partir da coluna de data foram acrescentadas as seguintes colunas aos dados importados, dia da semana, dia mês e ano. Para a aplicação das

Figura 14 – Gráficos com valores faltantes, anos de 2016 e 2018



Fonte: Imagens produzidas pela autora

informações nas redes neurais foi imprescindível a transformação dos dados literais em numéricos. Por fim, o conjunto total de informações foi dividido entre os subsistemas, criando assim quatro séries temporais distintas a serem modeladas como pode ser visto em Figura 15, Figura 16, Figura 18 e Figura 17.

Figura 15 – Prévia do conjunto de dados para região Norte

ID	Subsistema	Data	Carga_WMed	Dia_Semana	Dia	Mes	Ano
N	NORTE	2019-01-01	4688.341208	1	1	1	2019
N	NORTE	2019-01-02	5251.927167	2	2	1	2019
N	NORTE	2019-01-03	5409.962875	3	3	1	2019
N	NORTE	2019-01-04	5504.575375	4	4	1	2019
N	NORTE	2019-01-05	5247.999458	5	5	1	2019

Figura 16 – Prévia do conjunto de dados para região Nordeste

ID	Subsistema	Data	Carga_WMed	Dia_Semana	Dia	Mes	Ano
NE	NORDESTE	2019-01-01	9273.089917	1	1	1	2019
NE	NORDESTE	2019-01-02	10591.019708	2	2	1	2019
NE	NORDESTE	2019-01-03	11079.841750	3	3	1	2019
NE	NORDESTE	2019-01-04	11260.702417	4	4	1	2019
NE	NORDESTE	2019-01-05	10752.104750	5	5	1	2019

Figura 17 – Prévia do conjunto de dados para região Sudeste

ID	Subsistema	Data	Carga_WMed	Dia_Semana	Dia	Mes	Ano
SE	SUDESTE	2021-01-01	32419.959958	4	1	1	2021
SE	SUDESTE	2021-01-02	33849.020750	5	2	1	2021
SE	SUDESTE	2021-01-03	32674.559833	6	3	1	2021
SE	SUDESTE	2021-01-04	38192.798833	0	4	1	2021
SE	SUDESTE	2021-01-05	40956.483042	1	5	1	2021

Figura 18 – Prévia do conjunto de dados para região Sul

ID	Subsistema	Data	Carga_WMed	Dia_Semana	Dia	Mes	Ano
S	SUL	2021-01-01	8564.132540	4	1	1	2021
S	SUL	2021-01-02	9352.434667	5	2	1	2021
S	SUL	2021-01-03	9449.490167	6	3	1	2021
S	SUL	2021-01-04	12307.582125	0	4	1	2021
S	SUL	2021-01-05	13113.233084	1	5	1	2021

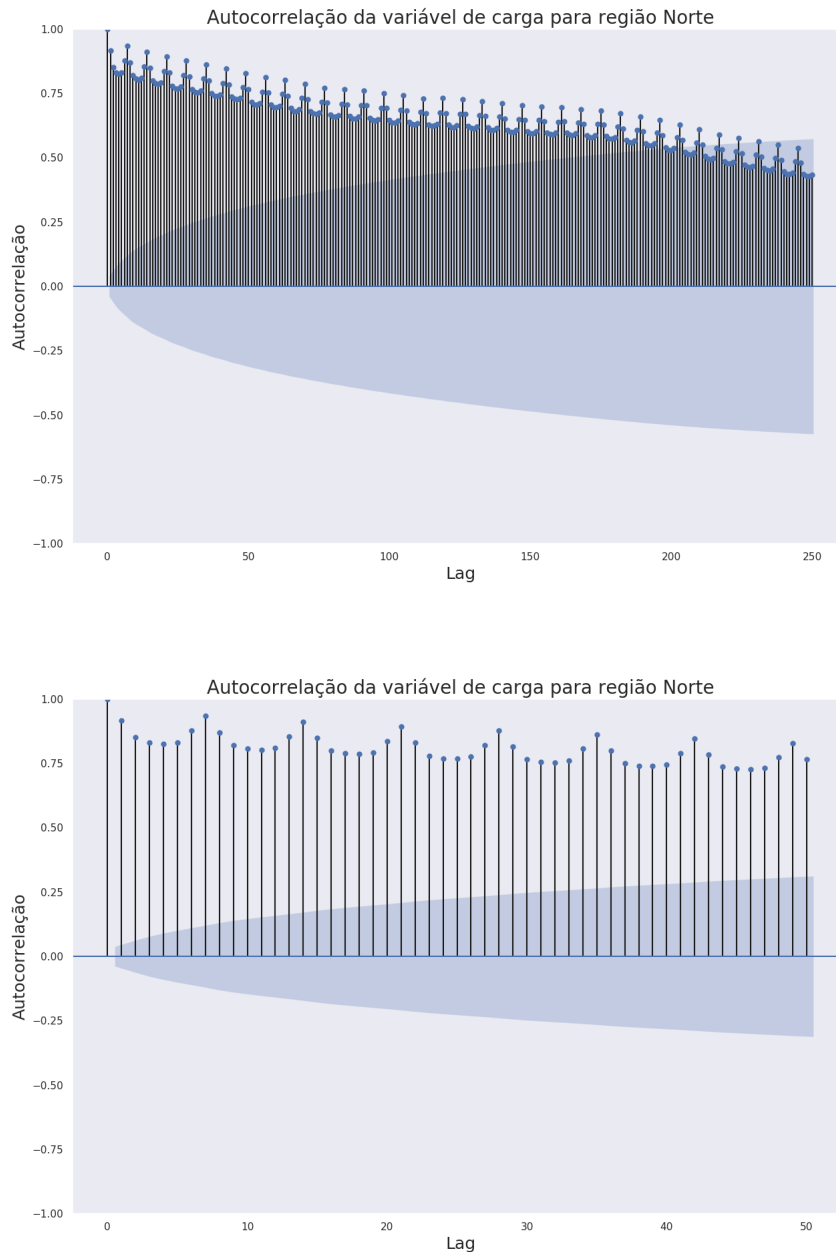
Fonte: Produzido pela autora

O processo de análise exploratória foi desenvolvido tendo em vista a seleção dos atributos, bem como a definição dos intervalos em dias de entrada e saída no processo de modelagem das redes neurais. Para cada uma das regiões foi construída uma matriz de correlação entre as suas variáveis e outros dois gráficos de autocorrelação em escalas diferentes.

Como apoio para a definição das janelas temporais de entrada e saída foram construídos gráficos de autocorrelação da variável de carga média em duas escalas para auxiliar a

visualização sobre diferentes perspectivas. A leitura da imagem é feita por meio da análise dos pontos de autocorrelação e da identificação de um padrão a partir de seus máximos locais. Os gráficos para cada uma das séries temporais estão definidas na Figura 19, Figura 20, Figura 21 e Figura 22.

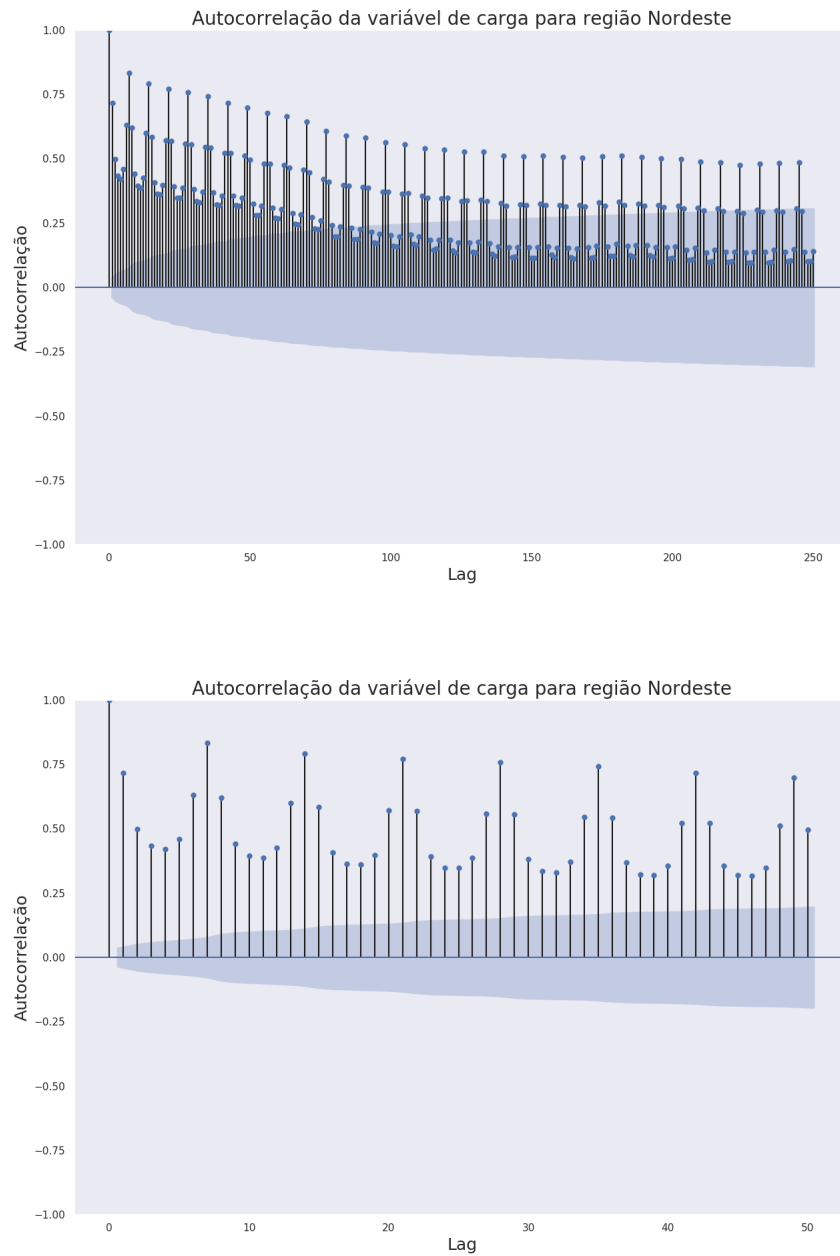
Figura 19 – Autocorrelação, região Norte em duas escalas temporais



Fonte: Produzido pela autora

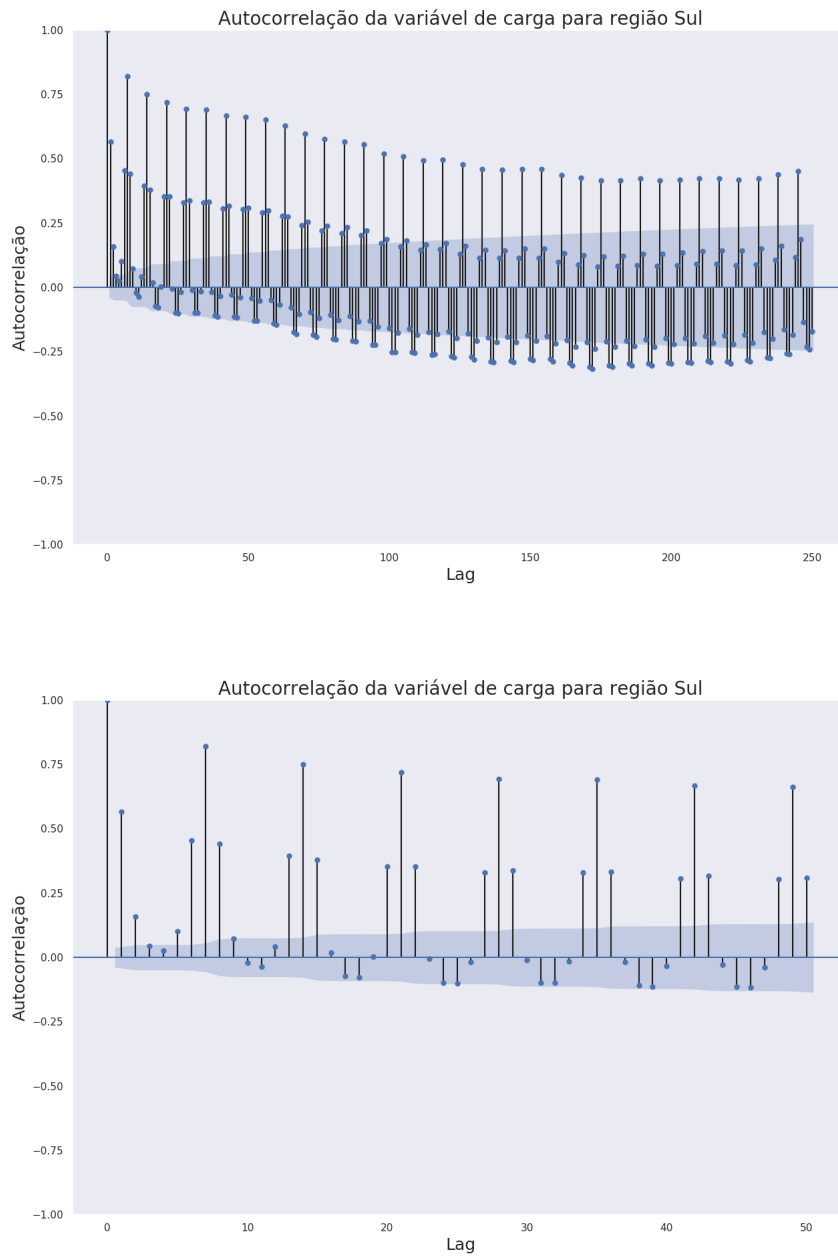
O uso dessa técnica é relevante pois a partir de cada um dos pontos de observação da variável estudada é quantificado a correlação (representada no eixo y), a distância entre o ponto inicial e o n-ésimo da série temporal é dada pelo *lag* (representado no eixo x). A

Figura 20 – Autocorrelação, região Nordeste em duas escalas temporais



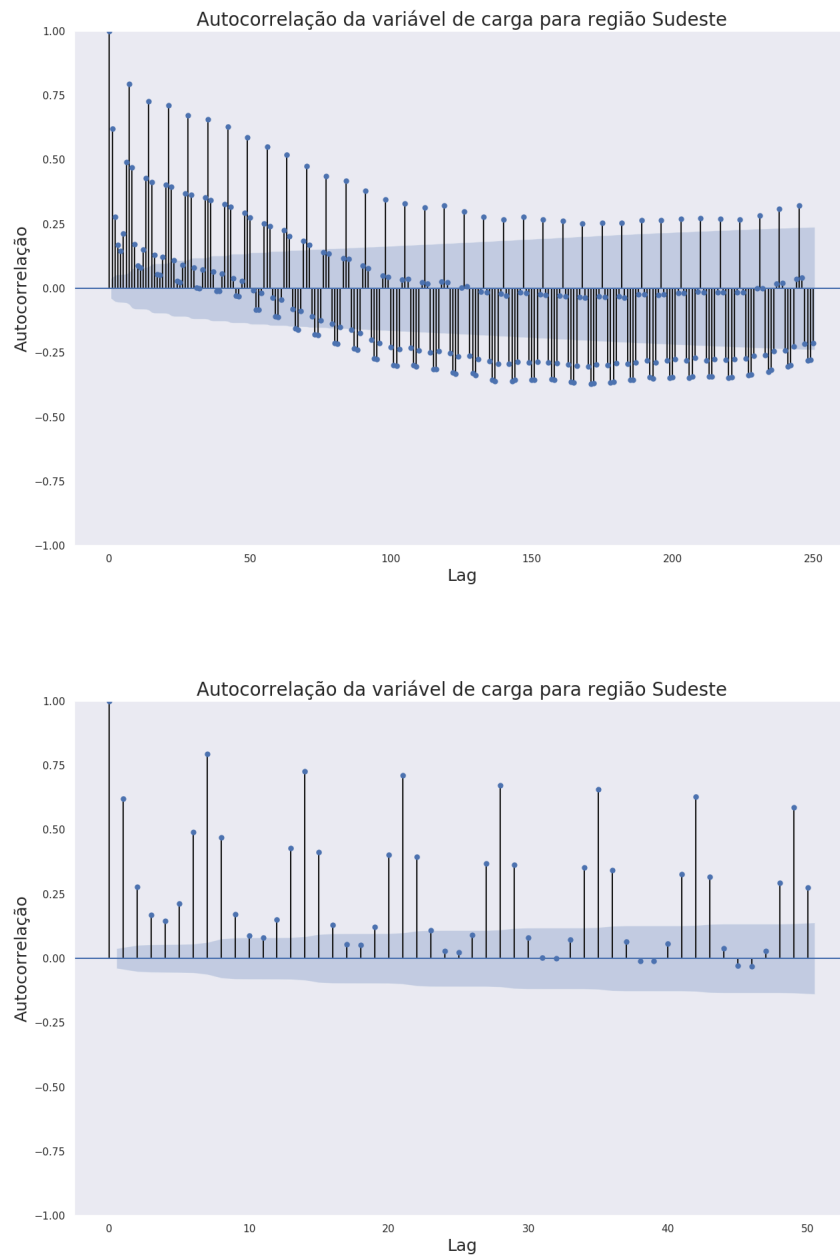
Fonte: Produzido pela autora

Figura 21 – Autocorrelação, região Sul em duas escalas temporais



Fonte: Produzido pela autora

Figura 22 – Autocorrelação, região Sudeste em duas escalas temporais



Fonte: Produzido pela autora

região azul que envolve o eixo horizontal do gráfico representa o intervalo de confiança das avaliações, caso os pontos de autocorrelação estejam fora dessa área isso significa que ele é estatisticamente significativo. Como padrão para o desenvolvimento da modelagem o intervalo aplicado é obtido através da fórmula de Bartlett (FRANCO; ZAKOIAN, 2009)

A partir das figuras foi possível constatar dentre todas as séries a presença de um maior nível de autocorrelação no intervalo de sete dias, ou uma semana completa. Assim sendo, é adequado que tanto a entrada quanto a saída seja um múltiplo de sete.

Com o objetivo de otimizar os resultados e prever um intervalo em curto prazo fica definida uma janela de vinte e oito dias de entrada e sete para a saída

Os gráficos de correlação entre as variáveis independentes de cada um das séries podem ser observadas na Figura 23, Figura 24, Figura 25 e Figura 26 . Por meio das imagens obtidas, ficou clara a presença de correlação inversa entre as informações de Carga X Dia da Semana e também certo nível entre Carga X Mês e Carga X Ano.

Ainda que as informações de ano possuam um nível de correlação entre 0.5 e 0.2 para todos os sistemas, ela não deve ser aplicada a modelagem. Uma vez que as janelas de entrada e saída de informações têm uma escala consideravelmente menor, a informação de ano pode ser tomada como a tendência das séries temporais e assim sua aplicação não é de grande relevância. Isso posto, são selecionadas três variáveis de entrada do modelo, carga média, dia da semana e mês.

No que diz respeito à definição das redes neurais, para encontrar a arquitetura utilizada neste desenvolvimento foram realizadas diversas experimentações com ampla variação nos parâmetros tanto de entrada e saída do modelo quanto para a própria arquitetura. As métricas de avaliação REQM e EMPA foram utilizadas para a comparação entre os diferentes testes. É importante notar que o estudo tem como metodologia a construção de três tipos de redes neurais LSTM, elas são, LSTM *feed forward*, LSTM bidirecional e GRU. As validações foram conduzidas a partir da rede neural do tipo bidirecional e seus parâmetros devem ser estendidos na construção de todos os modelos.

Feito isso, ficou definido como arquitetura das redes duas camadas do tipo que define a rede - cada uma com 200 neurônios cada -, seguidas de duas camadas densas - com 40 e 20 neurônios, respectivamente - por último estão presentes uma camada de *dropout* de 25% e a camada de saída com 7 neurônios.

A janela de entrada dos dados é de 28 dias, a de saída 7. Os modelos devem ser treinados por 150 épocas, além do erro médio dos dados de treinamento também essa métrica também é levada em conta tendo aplicado os dados do conjunto de validação. É de grande importância apontar que essa última métrica é avaliada como critério de parada antecipada do modelo.

Figura 23 – Correlação, região Norte

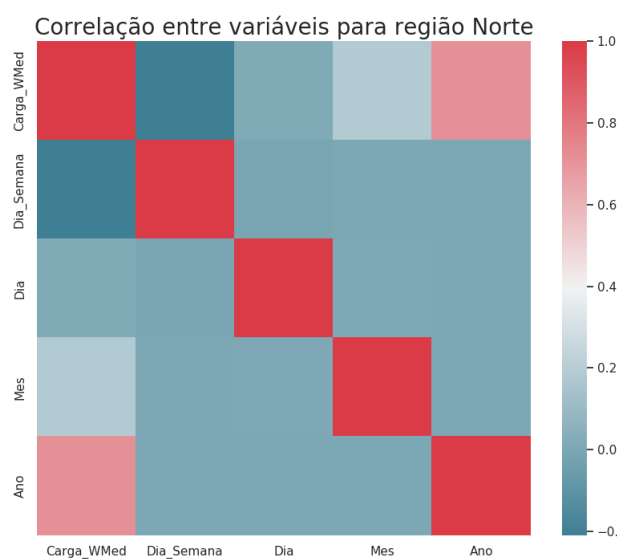


Figura 24 – Correlação, região Nordeste

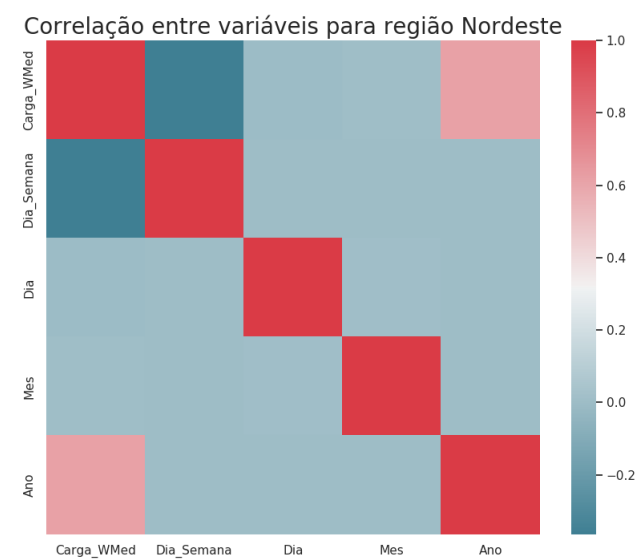


Figura 25 – Correlação, região Sul

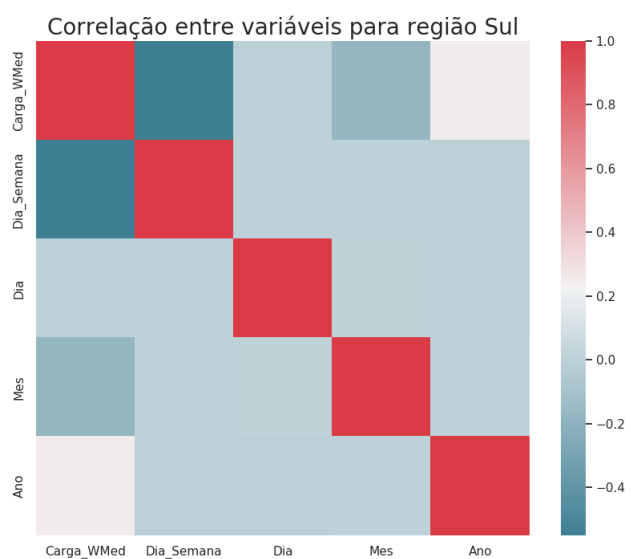
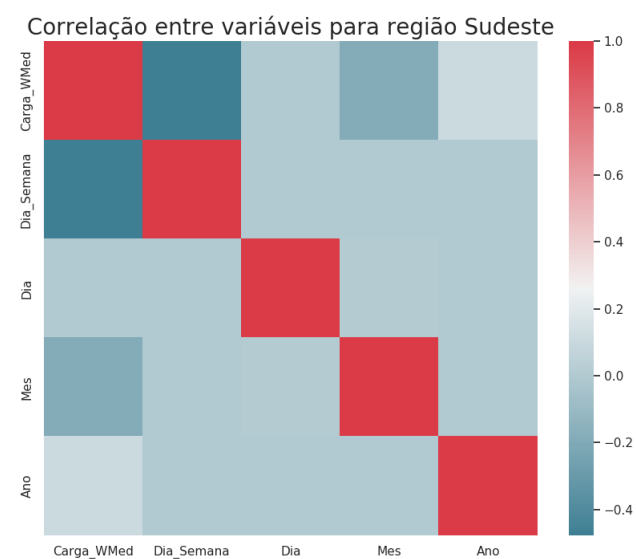


Figura 26 – Correlação, região Sudeste



Fonte: Produzido pela autora

5.2 Avaliação dos Resultados

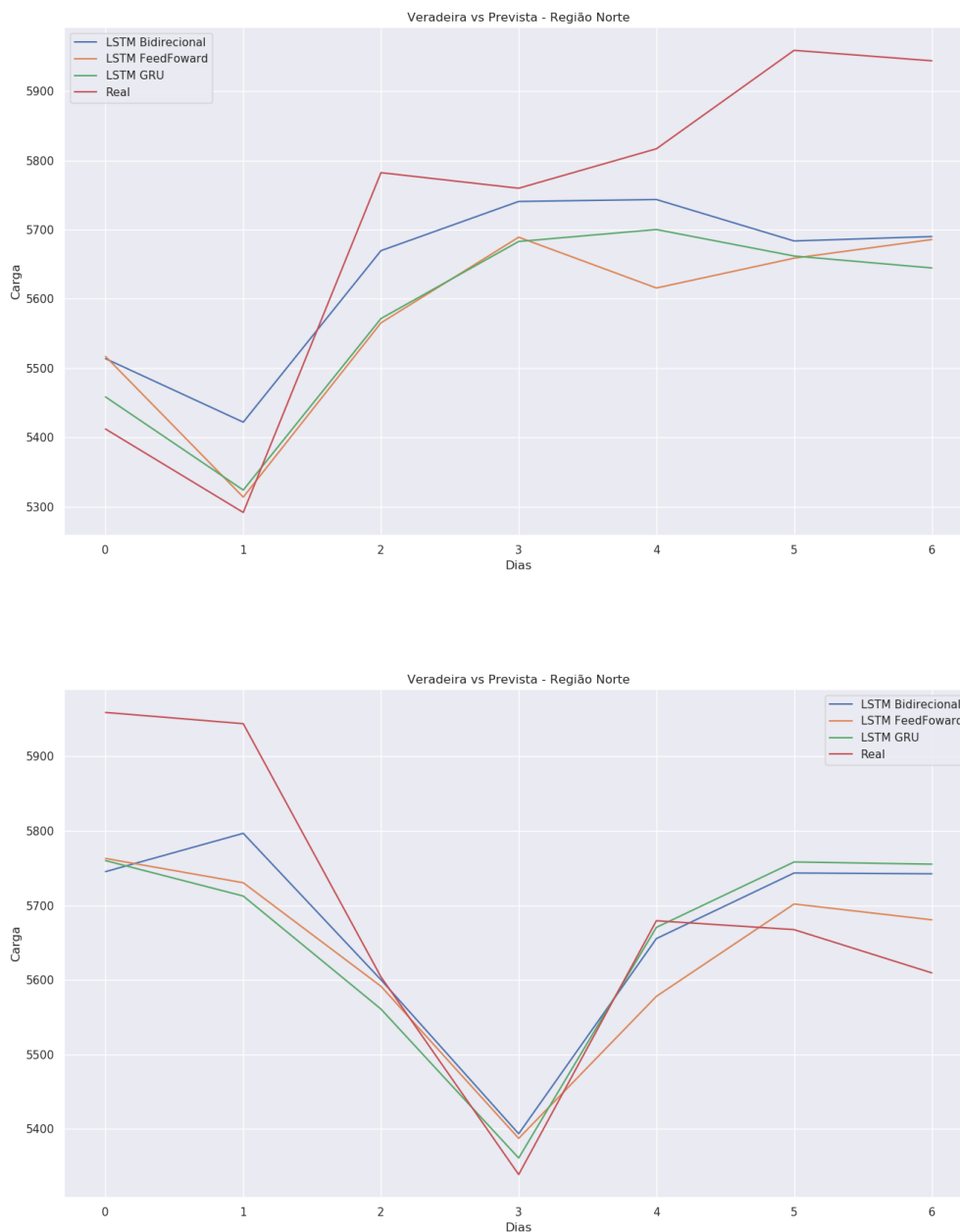
Tendo os processos de desenvolvimento realizados e o modelo construído, são utilizadas as seguintes métricas de avaliação no conjunto de teste, EQM, REQM, EMPA e EMA. Nesta seção os resultados vão ser analisados de forma independente para cada uma das regiões, ao final será feita uma análise das predições como um todo.

Dado que cada uma das séries temporais utilizadas no desenvolvimento da predição possui uma escala distinta para a variável dependente, os valores aplicados no processo

de avaliação são os normalizados para os cálculos de EQM, REQM e EMA. Já o EMPA é construído com base nos valores de erro em sua escala real, isso posto que essa não é uma métrica adequada para a aplicação em dados de valores próximos a zero.

Iniciando com a região Norte é possível observar que todas as modelagens obtiveram níveis próximos de desempenho (Tabela 3). Ainda assim, a rede LSTM bidirecional ofereceu uma melhor predição. Do gráfico (ver Figura 27) é perceptível que todas as redes conseguiram prever pontos de pico de mínimos da série, contudo ainda são encontradas lacunas em seqüências estáveis ou de máximo.

Figura 27 – Valores previstos e reais, região Norte



Fonte: Produzido pela autora

Com um erro médio absoluto de 0,1027 para a modelagem bidirecional contra 0,1084 e 0,1057 das estruturas *Feed Forward* e GRU, respectivamente, é notável que mesmo a primeira obtendo os melhores resultados, todas forneceram boas estimativas para os dados. Isso também pode ser apontado ao constatar que o EMPA de todas permaneceu em aproximadamente entre 2,6% e 2,7%.

Tabela 3 – Resultados, região Norte

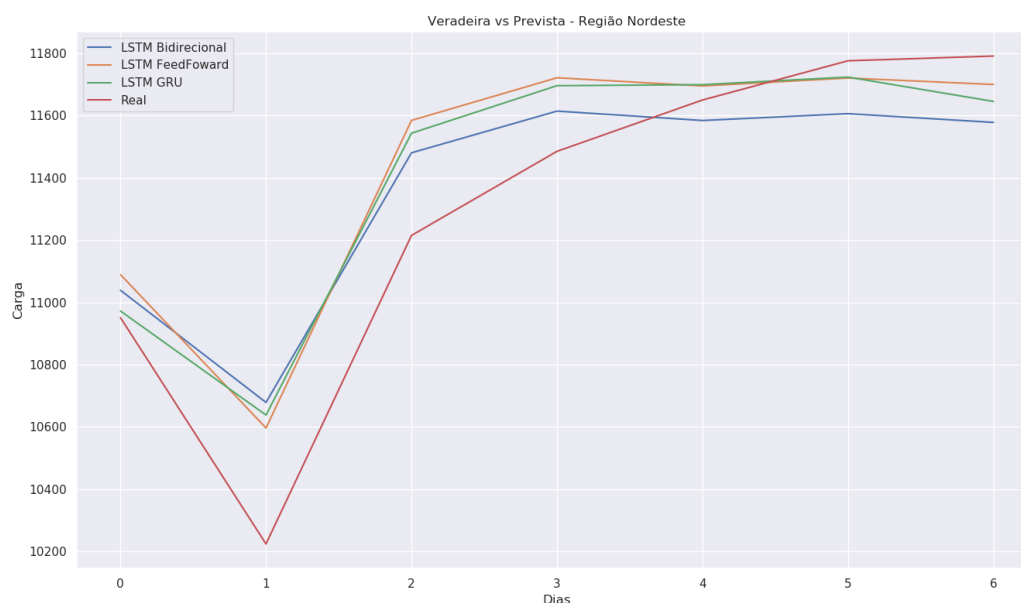
Rede Neural	EQM	REQM	MAPE	Erro médio
LSTM FeedForward	0.0184142	0.128303	2.683324	0.108370
LSTM Bidirecional	0.016676	0.122109	2.607511	0.102744
GRU	0.017616	0.125554	2.736923	0.105670

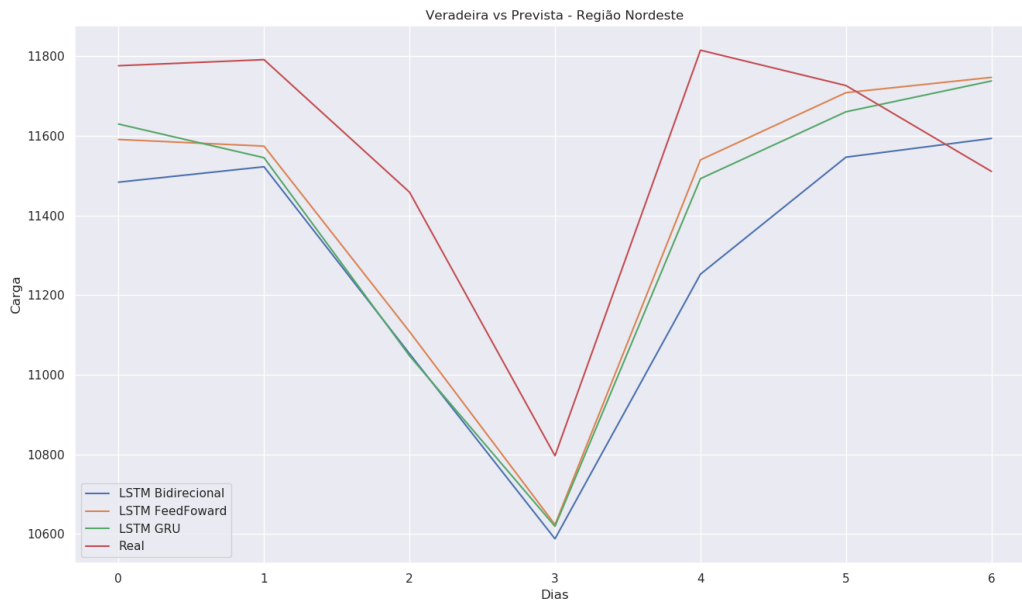
Fonte: Tabela produzida pela autora

Tomando a série representante da região Nordeste do país, a modelagem que teve como melhores resultados das métricas de avaliação foi também a LSTM do tipo bidirecional. Observando a Figura 28 é percebido que enquanto os modelos em grande parte consegue acompanhar as variações da série, ainda existe um intervalo quase de forma contínua entre os valores previstos e o esperado.

Com EQM todos entre 0,018 e 0,021 e REQM entre 0,1264 e 0,1366 - como pode ser visto na Tabela 4 - fica clara a proximidade dos resultados. Ainda que a rede LSTM bidirecional tenha obtido a predição mais próxima à real segundo três métricas os valores do seu EMPA foram os piores. Uma justificativa para tal resultado é que esse avaliador penaliza mais erros negativos.

Figura 28 – Valores previstos e reais, região Nordeste





Fonte: Produzido pela autora

Tabela 4 – Resultados, região Nordeste

Rede Neural	EQM	REQM	MAPE	Erro médio
LSTM FeedForward	0.021330	0.136665	3.459778	0.109516
LSTM Bidirecional	0.018078	0.126446	3.385324	0.106773
GRU	0.020969	0.136684	3.346564	0.113342

Fonte: Tabela produzida pela autora

Analisando os resultados da região Sul, temos a técnica GRU como detentora dos melhores resultados. A partir da Tabela 5 podemos notar a unanimidade dos valores para as métricas do GRU, também é observado um intervalo maior no EMPA 5,2863% versus aproximadamente 5,7499% para os outros dois critérios.

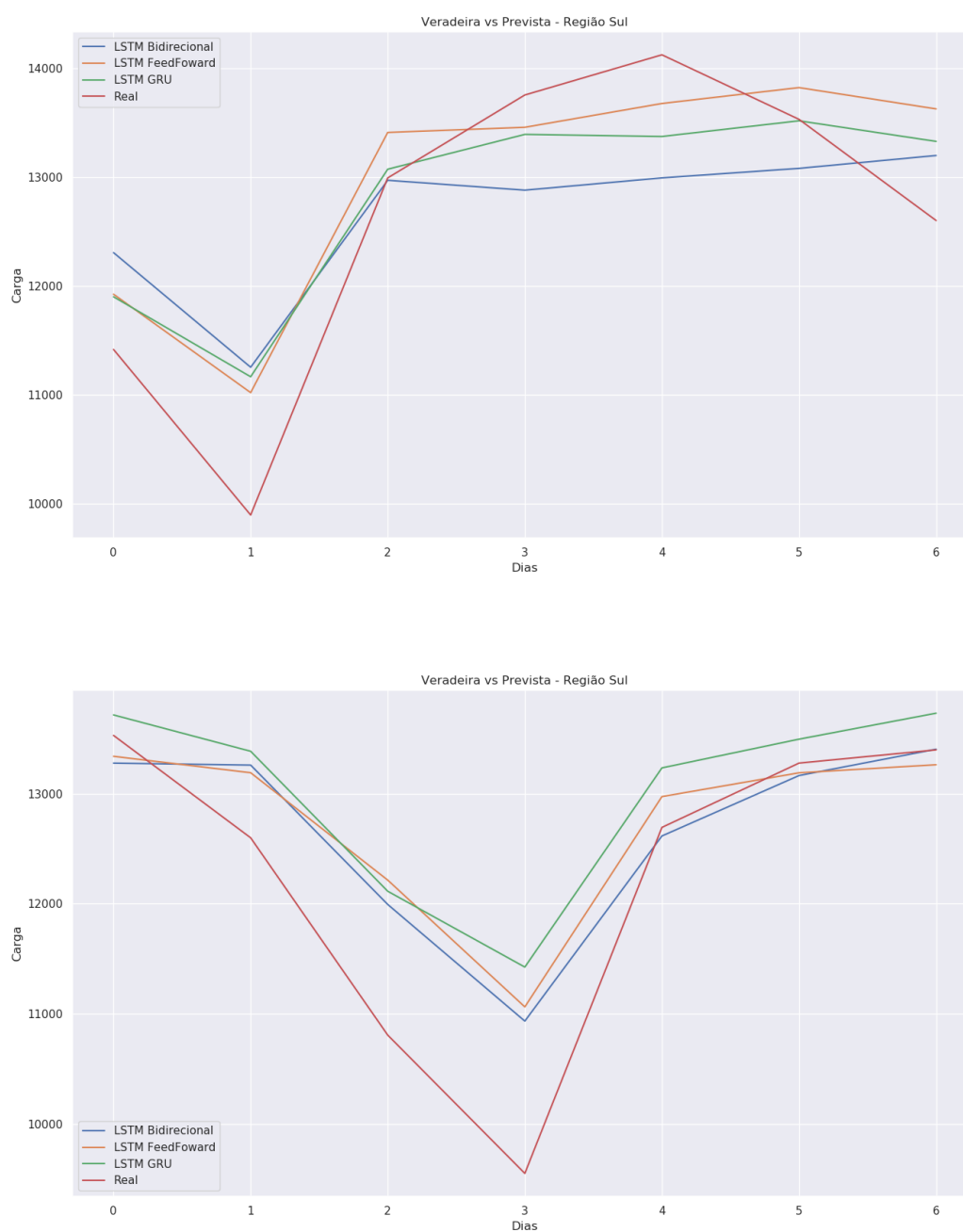
Tabela 5 – Resultados, região Sul

Rede Neural	EQM	REQM	MAPE	Erro médio
LSTM FeedForward	0.019395	0.122031	5.780112	0.102901
LSTM Bidirecional	0.022481	0.142277	5.719734	0.120768
GRU	0.016271	0.112818	5.286275	0.095054

Fonte: Tabela produzida pela autora

Graficamente - Figura 29 - é notado que todos os modelos conseguiram obter resultados relevantes, além disso eles acompanham as variações das séries de forma adequada, tendo suas principais divergências nos picos mínimos para os valores esperados. Ainda que esse resultado apresente um leve achatamento dos valores preditos, seus valores de máximos e mínimos são os mais próximos aos valores reais.

Figura 29 – Valores previstos e reais, região Sul



Fonte: Produzido pela autora

Para a região Sudeste foi tido novamente como unanimidade os resultados da rede GRU na apuração de todas as métricas de avaliação, Figura 30. Além de expressivos por seus resultados, os números possuem um intervalo maior entre os outros métodos, como 0,0163 no EQM contra 0,0194 e 0,0225 para as outras e 0,0950 versus 0,1029 e 0,1208 (LSTM e LSTM Bidirecional, respectivamente).

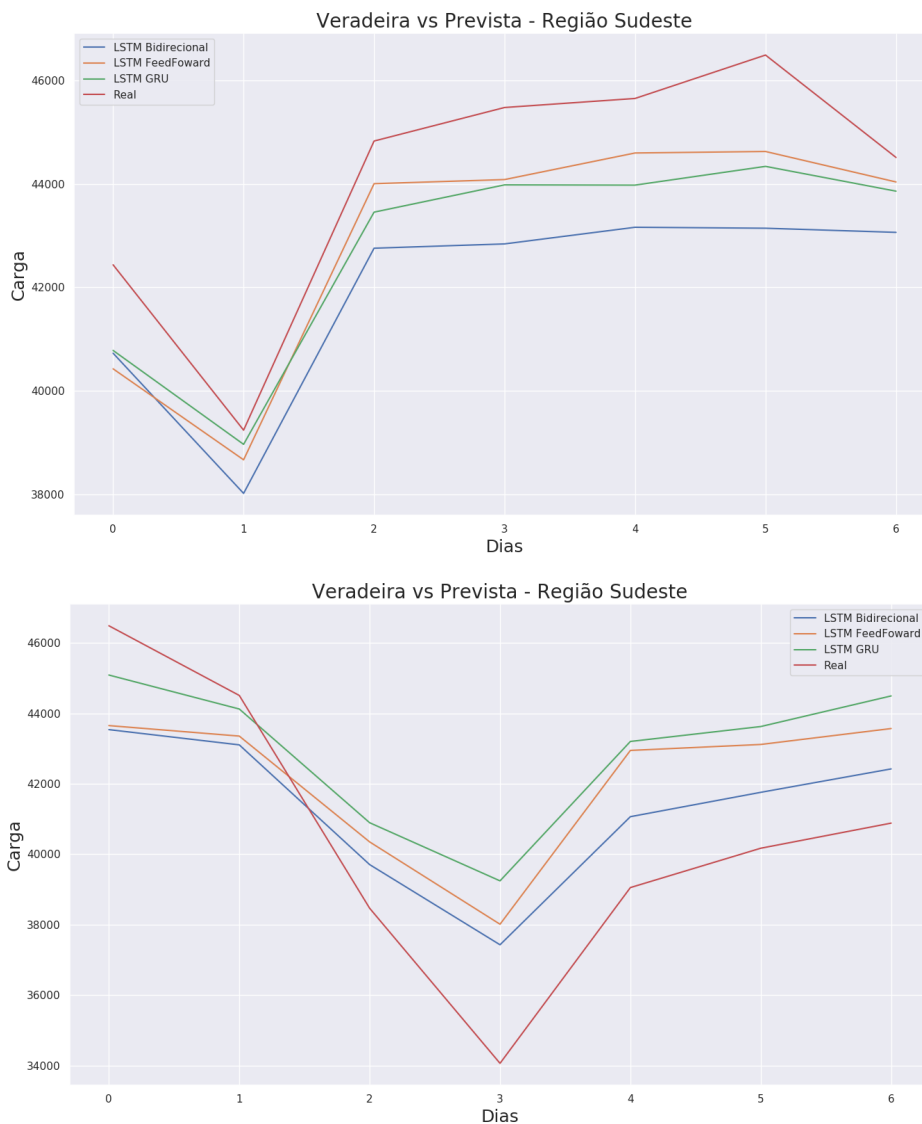
Na avaliação gráfica foi percebido o acompanhamento para as oscilações de todos os

Tabela 6 – Resultados, região Sudeste

Rede Neural	EQM	REQM	MAPE	Erro médio
LSTM FeedForward	0.013532	0.103851	3.553912	0.088934
LSTM Bidirecional	0.018221	0.120107	3.423790	0.100372
GRU	0.012642	0.100309	3.297136	0.085470

Fonte: Tabela produzida pela autora

Figura 30 – Valores previstos e reais, região Sudeste



Fonte: Produzido pela autora

modelos, mas uma proximidade maior nos comportamentos e variações para o modelo GRU.

Por fim, tendo todas as análises realizadas é possível perceber a prevalência de resultados positivos em modelos tidos como mais complexos, no caso desse estudo, redes neurais LSTM Bidirecionais e GRU. Conforme explícito na Tabela 7, o melhor resultado de todas

Tabela 7 – Comparativo de resultados entre regiões

Região	Rede Neural	Erro médio	MAPE	Erro médio em escala
Norte	LSTF Bidirecional	0.102744	2.607511	155.821213
Nordeste	LSTF Bidirecional	0.106773	3.385324	379.793884
Sul	GRU	0.095054	5.286275	616.834106
Sudeste	GRU	0.085470	3.297136	1281.387451

Fonte: Tabela produzida pela autora

as modelagens obteve EMPA inferior a 5,29%, além disso os erros médios normalizados não passaram de 0,11 para nenhum dos modelos.

É interessante notar que cada uma das duas redes performou melhor em séries com características em comum. Isto é, enquanto nos testes realizados a técnica de LSTM Bidirecional obteve melhores métricas em séries que possuíssem amplitudes de valores de máximo e mínimo menores, a arquitetura GRU retornou valores superiores para conjuntos cujas diferenças entre seus valores fossem maiores. Assim, a aplicação de LSTM Bidirecional se adequou às regiões Norte e Nordeste, na qual a amplitude entre os valores de mínimo e máximo foram respectivamente de 129% e 133%. Já a estrutura GRU foi conveniente ao uso nas regiões Sul e Sudeste, nas quais as variações são de 179% e 149% para o conjunto de teste.

Conclusão

Ao desenvolver este trabalho foi possível aprofundar o entendimento em conceitos de inteligência artificial, bem como pré processamento de dados, treinamento de redes neurais e modelagem de séries temporais. Com o uso da linguagem de programação Python e apoiada na biblioteca de código aberto de aprendizado de máquina Tensorflow foram construídas três redes neurais recorrentes (LSTM feed-forward, bidirecional e GRU), tendo como contexto a predição de carga de energia no mercado livre de energia brasileiro.

Foi notável a importância da divisão do conjunto de dados entre as sub-regiões assim como a seleção dos atributos para a obtenção dos resultados. A partir da análise da correlação e autocorrelação das variáveis dos dados foi possível definir respectivamente, os atributos relevantes e os intervalos de janela de entrada e saída da modelagem.

Foi possível observar as técnicas de LSTM Bidirecional e GRU com melhor desempenho em cenários distintos, com EMPA médio de 3,6441%. Conforme demonstrado nos cenários de teste entanto a primeira se sobressaiu em cenários para os quais os valores de máximo e mínimo tinham maior diferença do segundo foi notado o oposto.

Tomando como base os trabalhos referenciados no Capítulo 3, foi notada uma conformidade dos resultados obtidos com outros projetos mencionados. Os valores para a métrica de EMPA em particular superaram parte dos casos da modelagem proposta por (LI et al., 2020), nos quais os valores EMPA da modelagem proposta variam em média entre 7.3% e 2.82%.

É relevante pontuar ainda a diversidade de técnicas e arquiteturas possíveis de construção dentro do tema de redes neurais recorrentes. Assim, a exploração de novas redes, ainda que com os mesmos parâmetros de entrada podem trazer otimizações à predição.

Finalmente, foi possível constatar o desenvolvimento de modelagens de redes neurais recorrentes como uma técnica viável na modelagem de séries temporais, em especial de carga elétrica.

5.2.1 Trabalhos futuros

Com o fito de expandir o trabalho realizado neste documento são propostas algumas alterações que poderiam ser aplicadas de forma a buscar melhorias nos resultados:

- ❑ Construção e aplicação das mesmas estruturas de modelos em um conjunto de dados em uma escala de horas;
- ❑ Aquisição de mais variáveis independentes, como fatores climáticos ou econômicos, para aplicação no modelo
- ❑ Aquisição dos dados em agrupamentos estaduais em lugar das sub regiões do SIN;
- ❑ Aplicação de algoritmos genéticos no processo de treinamento e validação das modelagens;

Referências

ABRAADE. **Entenda a indústria da energia elétrica. Modulo I, A indústria da energia elétrica.** 2. ed. [S.l.], 2019.

ABRACEEL. **Cartilha do Consumidor Livre de Energia.** [S.l.], 2020. 8-23 p. Disponível em: <<https://abraceel.com.br/biblioteca/cartilhas/2020/10/cartilha-do-consumidor-livre-de-energia/>>. Acesso em: 2022-07-10.

_____. **Boletim ABRACELL da energia livre.** [S.l.], 2022. 1-3 p. Disponível em: <<https://abraceel.com.br/biblioteca/boletim/2022/04/boletim-anual-de-mercado-2021/>>. Acesso em: 2022-07-10.

AMIDI, A.; AMIDI, S. Recurrent neural networks cheatsheet. 2022. Disponível em: <<https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks>>. Acesso em: 2022-07-10.

BHANJA, S.; DAS, A. **Impact of Data Normalization on Deep Neural Network for Time Series Forecasting.** arXiv, 2018. Disponível em: <<https://arxiv.org/abs/1812.05519>>.

BRASIL. Lei nº 9.648 de 27 de maio de 1998. **Diário Oficial**, 1998.

_____. Resolução nº 351, de 11 de novembro de 1998. **Diário Oficial da União**, 1998.

_____. Decreto nº 5.163 de 30 de julho de 2004. **Diário Oficial da União**, 2004.

_____. Resolução normativa aneel nº 1.015, de 12 de abril de 2022. **Comunicado CCEE**, 2022.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting.** 3. ed. [S.l.]: Springer - New York, 2016.

CCEE. Painel de preços ccee. **Site Câmara de Comercialização de Energia Elétrica**, 2022. Disponível em: <<http://www.ons.org.br/paginas/sobre-o-sin/mapas>>. Acesso em: 2022-07-10.

CHEN, J. X. The evolution of computing: Alphago. **Computing in Science Engineering**, v. 18, n. 4, p. 4-7, 2016.

- CHU, C.-S. J. Time series segmentation: A sliding window approach. **Information Sciences**, v. 85, n. 1, p. 147–173, 1995. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/002002559500021G>>.
- DENG, L.; YU, D. et al. Deep learning: methods and applications. **Foundations and trends® in signal processing**, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014.
- DEY, R.; SALEM, F. M. Gate-variants of gated recurrent unit (gru) neural networks. In: **2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)**. [S.l.: s.n.], 2017. p. 1597–1600.
- DING, B.; QIAN, H.; ZHOU, J. Activation functions and their characteristics in deep neural networks. In: **2018 Chinese Control And Decision Conference (CCDC)**. [S.l.: s.n.], 2018. p. 1836–1841.
- DODGE, Y. **The Concise Encyclopedia of Statistics**. 1. ed. [S.l.]: Springer - New York, 2008. 536-539 p.
- EPE. **BEN 2021 Relatório Síntese 2021 (Ano Base 2020)**. [S.l.], 2021. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-601/topico-588/BEN_S\C3\%ADntese_2021_PT.pdf>. Acesso em: 2022-07-10.
- FRANCO, C.; ZAKOIAN, J.-M. Bartlett’s formula for a general class of non linear processes. **Munich Personal RePEc Archive Paper**, 2009. Disponível em: <<https://mpra.ub.uni-muenchen.de/id/eprint/13224>>.
- GANIM, A. **SETOR ELÉTRICO BRASILEIRO Aspectos Regulamentares, Tributários e Contábeis**. 1. ed. [S.l.]: Rio de Janeiro - Editora Synergia, 2009. 136-139 p.
- GHAHRAMANI, Z. **Unsupervised Learning**. [S.l.]: Springer - Berlin Heidelberg, 2004. 72-112 p.
- GONÇALVES, A. R. **Redes Neurais Artificiais**. 1. ed. [S.l.]: Springer - New York, 2003.
- HEWAMALAGE, H.; BERGMEIR, C.; BANDARA, K. Recurrent neural networks for time series forecasting: Current status and future directions. **International Journal of Forecasting**, v. 37, n. 1, p. 388–427, 2021. ISSN 0169-2070. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169207020300996>>.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.
- KATARYA, R.; RASTOGI, S. A study on neural networks approach to time-series analysis. In: **2018 2nd International Conference on Inventive Systems and Control (ICISC)**. [S.l.: s.n.], 2018. p. 116–119.
- KHAN, Z. A. et al. Efficient short-term electricity load forecasting for effective energy management. **Sustainable Energy Technologies and Assessments**, v. 53, p. 102337, 2022. ISSN 2213-1388. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2213138822003897>>.

- LI, N. et al. An effective deep learning neural network model for short-term load forecasting. **Concurrency and Computation: Practice and Experience**, v. 32, n. 7, p. e5595, 2020. E5595 cpe.5595. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5595>>.
- LIVINGSTONE, D.; MANALLACK, D.; TETKO, I. Data modelling with neural networks: Advantages and limitations. p. 135–142, 1997.
- LORENZO, H. C. O setor elétrico brasileiro: passado e futuro. **Perspectivas: Revista de Ciências Sociais**, v. 24-25, p. 162–165, 2001.
- MEDSKER, L.; JAIN, L. C. **Recurrent neural networks: design and applications**. [S.l.]: CRC press, 1999.
- MOHAMMAD, F.; KIM, Y.-C. Energy load forecasting model based on deep neural networks for smart grids. **International Journal of System Assurance Engineering and Management**, v. 11, 2020. ISSN 0976-4348.
- OLAH, C. **Understanding LSTM Networks**. [S.l.], 2015. Disponível em: <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>.
- ONS. Mapa dinâmico do sin. **Site Operador Nacional do Sitem Elétrico**, 2022. Disponível em: <<http://www.ons.org.br/paginas/sobre-o-sin/mapas>>. Acesso em: 2022-07-10.
- _____. O sistema interligado nacional. **Site Operador Nacional do Sitem Elétrico**, 2022. Disponível em: <<http://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>>. Acesso em: 2022-07-10.
- OZCAN, A.; CATAL, C.; KASIF, A. Energy load forecasting using a dual-stage attention-based recurrent neural network. **Sensors**, v. 21, n. 21, 2021. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/21/21/7115>>.
- REZENDE, S. O.; MONARD, M. C.; CARVALHO, A. C. P. L. Sistemas inteligentes para engenharia: Pesquisa e desenvolvimento. In: **Anais III Workshop de Sistemas Inteligentes para Engenharia**. Belo Horizonte: Editora UFMG, 1999.
- RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4. ed. [S.l.]: Pearson - London, 2020.
- SANCHES, M. K. **Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados**. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo), 2003.
- SAYGIN, A. P.; CICEKLI, I.; AKMAN, V. Turing test: 50 years later. **Minds and machines**, Springer, v. 10, n. 4, p. 463–518, 2000.
- SCHUSTER, M.; PALIWAL, K. Bidirectional recurrent neural networks. **IEEE Transactions on Signal Processing**, v. 45, n. 11, p. 2673–2681, 1997.
- SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. **towards data science**, v. 6, n. 12, p. 310–316, 2017.

SILVA, B. G. d. **Evolução do setor elétrico brasileiro no contexto econômico nacional: uma análise histórica e econométrica de longo prazo**. 21 p. Dissertação (Dissertação (Mestrado em Energia) - Energia) — Uiversidade of São Paulo, 2011.

SILVA, D. R. **Contextualização do Setor Elétrico Brasileiro**. [S.l.], 2020. Disponível em: <<http://repositorio.ipea.gov.br/handle/11058/10098>>.

SUC CETTI, F. et al. Multidimensional feeding of lstm networks for multivariate prediction of energy time series. In: **2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / ICPS Europe)**. [S.l.: s.n.], 2020. p. 1–5.

VARIAN, H. R. **Microeconomia - Uma Abordagem Moderna**. [S.l.]: Rio de Janeiro - Elsevier, 2015.

ZHANG, W. et al. On definition of deep learning. In: **2018 World Automation Congress (WAC)**. [S.l.: s.n.], 2018. p. 1–5.