

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

LUIZA BUTZGE

**Aplicação de algoritmos inteligentes para
previsão de carga no curto prazo em sistemas
de distribuição**

Porto Alegre

2020

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

LUIZA BUTZGE

**Aplicação de algoritmos inteligentes para previsão de
carga no curto prazo em sistemas de distribuição**

Projeto de Diplomação como requisito parcial à obtenção do título de graduanda em Engenharia Elétrica na Universidade Federal do Rio Grande do Sul

Orientador: Prof. Dr. Maicon Jaderson Silveira Ramos

Porto Alegre

2020

FOLHA DE APROVAÇÃO

Luiza Butzge

Aplicação de algoritmos inteligentes para previsão de carga no curto prazo em sistemas de distribuição

Projeto de Diplomação como requisito parcial à obtenção do título de graduanda em Engenharia Elétrica na Universidade Federal do Rio Grande do Sul
Orientador: Prof. Dr. Maicon J. S. Ramos

Aprovada em: Porto Alegre, 26 de novembro de 2020.

BANCA EXAMINADORA:

Prof. Dr. Maicon Jaderson Silveira Ramos
Universidade Federal do Rio Grande do Sul

Prof. Dr. Flávio Antônio Becon Lemos
Universidade Federal do Rio Grande do Sul

Ms. Carlos Eduardo Cauduro Figueiredo
Rio Grande Energia - RGE (CPFL Energia)

CIP - Catalogação na Publicação

Butzge, Luiza
Aplicação de algoritmos inteligentes para previsões
de carga no curto prazo em sistemas de distribuição /
Luiza Butzge. -- 2020.
73 f.
Orientador: Maicon Jaderson Silveira Ramos.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Escola de
Engenharia, Curso de Engenharia Elétrica, Porto
Alegre, BR-RS, 2020.

1. Previsão de carga. I. Jaderson Silveira Ramos,
Maicon, orient. II. Título.

Agradecimentos

Agradeço a meus pais, Carlos e Nair, e a minha irmã Julia, pela confiança no meu desenvolvimento acadêmico e pelo apoio emocional demonstrado por todos esses anos de graduação.

Sou grata também a meu orientador Professor Doutor Maicon Jaderson Silveira Ramos pelas valiosas informações, correções e sugestões que contribuíram para a qualidade do presente trabalho.

Agradeço a meus colegas de curso, pela oportunidade do convívio e pela cooperação mútua durante estes anos, e a universidade, por me proporcionar um ensino de qualidade.

Um especial agradecimento a meu amigo Renan e a minha prima Vanessa pelo apoio emocional e afetivo incondicional durante a realização do curso.

Resumo

Este trabalho tem como objetivo obter modelos para previsão da curva de carga no horizonte de curto prazo, considerando medições de potência em transformadores de força instalados nas subestações de distribuição. Para tanto, foram examinadas variáveis ditas relevantes, como potência ativa e fatores climáticos das regiões em que esses transformadores se encontram, mais especificamente Santa Maria e São Leopoldo. Para isso, foram analisados modelos da área afim de ressaltar vantagens e desvantagens de cada trabalho e, em seguida, construir a metodologia de modelos em Aprendizado de Máquina. Utilizou-se então quatro modelos computacionais, Regressão de Vetores de Suporte, Árvore de Decisão, Florestas de Decisão Aleatória e Memória de Longo e Curto Prazo, construídos em Python. Os resultados encontrados no trabalho foram obtidos segundo comparações entre dados reais e dados previstos de potência ativa levando em conta métodos estatísticos de cálculos de erros e correlação entre as variáveis com a intenção de tornar os modelos mais exatos, diminuindo assim os erros encontrados. Por fim, é discutida a viabilidade de tais modelos para a previsão de carga, fontes de erros e possíveis melhoras na implementação dos modelos para trabalhos futuros.

Palavras-chave: Previsão de curvas de carga; Aprendizado de Máquina; Redes Neurais

Abstract

The present paper aims to develop prediction models to forecast the load of power transformers for a short period of time located in the substations in question with the help of computational tools. In this matter, it's analysed important variables, active power and weather factors, where those transformers are located, more precisely Santa Maria and São Leopoldo. The models created by important authors are analysed to raise the advantages and disadvantages of each work so it's possible to develop the methodology of the models in Machine Learning. This paper uses four models, Support Vector Regression, Decision Tree, Random Forest and Long Short Term Memory, made in Python language. The results found are obtained comparing the real active power data with the predicted active power data taking into account statistical methods to calculate the errors and the correlations between the variables. Lastly, it's discussed the viability of those models to predict power loads, the common founts of errors and possible improvements in the models for future works.

Keywords: Load forecasting; Machine Learning; Neural Networks

Lista de Figuras

Figura 1 – Curva de carga típica industrial para cargas > 10MWh	18
Figura 2 – Tipos de Aprendizado de Máquina	21
Figura 3 – Árvore de Decisão exemplo: ‘Sair de casa segundo a previsão do tempo’	22
Figura 4 – Resultados da previsão de carga	25
Figura 5 – Fluxograma dos modelos de previsão	27
Figura 6 – Divisão teste e treino dos dados	30
Figura 7 – Diagrama da criação do modelo	32
Figura 8 – Diagrama das variáveis	33
Figura 9 – Comparação curva de carga com e sem fatores climáticos para a Árvore de Decisão	39
Figura 10 – Comparação curva de carga com e sem fatores climáticos para a Florestas de Decisão Aleatória	39
Figura 11 – Comparação curva de carga com e sem fatores climáticos para a Memória de Longo e Curto Prazo	40
Figura 12 – Segunda-feira: Modelo Árvore de Decisão	41
Figura 13 – Segunda-feira: Modelo Florestas de Decisão Aleatória	41
Figura 14 – Segunda-feira: Modelo Memória de Longo e Curto Prazo	42
Figura 15 – Quarta-feira: Modelo Árvore de Decisão	43
Figura 16 – Quarta-feira: Modelo Florestas de Decisão Aleatória	43
Figura 17 – Quarta-feira: Modelo Memória de Longo e Curto Prazo	44
Figura 18 – Sexta-feira: Modelo Árvore de Decisão	45
Figura 19 – Sexta-feira: Modelo Florestas de Decisão Aleatória	45
Figura 20 – Sexta-feira: Modelo Memória de Longo e Curto Prazo	46
Figura 21 – Sábado: Modelo Árvore de Decisão	47
Figura 22 – Sábado: Modelo Florestas de Decisão Aleatória	47
Figura 23 – Sábado: Modelo Memória de Longo e Curto Prazo	48
Figura 24 – Domingo: Modelo Árvore de Decisão	48
Figura 25 – Domingo: Modelo Florestas de Decisão Aleatória	49
Figura 26 – Domingo: Modelo Memória de Longo e Curto Prazo	49
Figura 27 – Natal: Modelo Árvore de Decisão	50
Figura 28 – Natal: Modelo Florestas de Decisão Aleatória	50
Figura 29 – Natal: Modelo Memória de Longo e Curto Prazo	51
Figura 30 – Finados: Modelo Árvore de Decisão	52
Figura 31 – Finados: Modelo Florestas de Decisão Aleatória	52
Figura 32 – Finados: Modelo Memória de Longo e Curto Prazo	53

Figura 33 – Comparação curva de carga com e sem fatores climáticos para a Árvore de Decisão	55
Figura 34 – Comparação curva de carga com e sem fatores climáticos para a Florestas de Decisão Aleatória	55
Figura 35 – Comparação curva de carga com e sem fatores climáticos para a Memória de Longo e Curto Prazo	56
Figura 36 – Segunda-feira: Modelo Árvore de Decisão	57
Figura 37 – Segunda-feira: Modelo Florestas de Decisão Aleatória	57
Figura 38 – Segunda-feira: Modelo Memória de Longo e Curto Prazo	58
Figura 39 – Quarta-feira: Modelo Árvore de Decisão	58
Figura 40 – Quarta-feira: Modelo Florestas de Decisão Aleatória	59
Figura 41 – Quarta-feira: Modelo Memória de Longo e Curto Prazo	59
Figura 42 – Sexta-feira: Modelo Árvore de Decisão	60
Figura 43 – Sexta-feira: Modelo Florestas de Decisão Aleatória	60
Figura 44 – Sexta-feira: Modelo Memória de Longo e Curto Prazo	61
Figura 45 – Sábado: Modelo Árvore de Decisão	62
Figura 46 – Sábado: Modelo Florestas de Decisão Aleatória	62
Figura 47 – Sábado: Modelo Memória de Longo e Curto Prazo	63
Figura 48 – Domingo: Modelo Árvore de Decisão	64
Figura 49 – Domingo: Modelo Florestas de Decisão Aleatória	64
Figura 50 – Domingo: Modelo Memória de Longo e Curto Prazo	65
Figura 51 – Natal: Modelo Árvore de Decisão	66
Figura 52 – Natal: Modelo Florestas de Decisão Aleatória	66
Figura 53 – Natal: Modelo Memória de Longo e Curto Prazo	67
Figura 54 – Finados: Modelo Árvore de Decisão	68
Figura 55 – Finados: Modelo Florestas de Decisão Aleatória	68
Figura 56 – Finados: Modelo Memória de Longo e Curto Prazo	69

Lista de Tabelas

Tabela 1 – Comparação entre os resultados dos algoritmos RVS e MLCP	24
Tabela 2 – Comparação dos modelos RN e Árvore de Decisão	24
Tabela 3 – Comparação dos resultados encontrados pelos algoritmos	25
Tabela 4 – Dias de análise das curvas de carga	35
Tabela 5 – Dias de análise das curvas de carga	36
Tabela 6 – Resultados das métrica de avaliação dos modelos para SE São Leopoldo	37
Tabela 7 – Resultados dos modelos sem fatores climáticos	38
Tabela 8 – Resultados das métrica de avaliação dos modelos para SE Santa Maria	53
Tabela 9 – Resultados dos modelos sem fatores climáticos	54

Lista de Abreviaturas e Siglas

AD	<i>Árvore de Decisão</i>
RVS	<i>Regressão de Vetor de Suporte</i>
FDA	<i>Floresta de Decisão Aleatória</i>
MLCP	<i>Memória de Longo e Curto Prazo</i>
R ²	<i>R quadrado</i>
MQE	<i>Erro quadrado médio</i>
EAM	<i>Erro absoluto médio</i>
UTC	<i>Tempo Universal Coordenado</i>

Sumário

1	INTRODUÇÃO	13
1.1	Justificativa	14
1.2	Estruturação do trabalho	15
1.3	Estrutura do trabalho	15
2	REVISÃO BIBLIOGRÁFICA	17
2.1	Consumo de energia no Brasil	17
2.2	Introdução a previsão de curvas de carga	18
2.3	Modelos de Aprendizado de Máquina e Redes Neurais	21
2.4	Análise de estudos relacionados	23
3	METODOLOGIA	26
3.1	Etapas de construção do modelo	26
3.2	Organização da base de dados	27
3.3	Definição das variáveis	27
3.4	Processamento dos dados	29
3.5	Divisão Treino e Teste	30
3.6	Aplicação do algoritmo inteligente	30
3.7	Aplicação das métricas de comparação	31
3.8	Funcionamento geral dos modelos	32
4	ESTUDO DE CASO	34
4.1	São Leopoldo	34
4.2	Santa Maria	35
5	RESULTADOS E DISCUSSÕES	37
5.1	Resultados encontrados para São Leopoldo	37
5.1.1	Análise da relevância das variáveis climáticas no modelo	38
5.1.2	Análise dos dias da semana	40
5.1.3	Análise do final de semana	46
5.1.4	Análise de feriados	50
5.2	Resultados encontrados para Santa Maria	53
5.2.1	Análise da relevância das variáveis climáticas no modelo	54
5.2.2	Análise dos dias da semana	57
5.2.3	Análise do final de semana	61
5.2.4	Análise de feriados	65

6	CONCLUSÕES	70
6.1	Propostas de Trabalhos Futuros	71
	REFERÊNCIAS BIBLIOGRÁFICAS	72

1 Introdução

Um dos principais fatores de desenvolvimento e sustentabilidade econômica de uma nação é a capacidade de prover energia elétrica para seus consumidores. Sabendo disso e analisando o crescente aumento no consumo de energia, visto nas últimas décadas, sabe-se a importância que o bom planejamento no setor elétrico tem para o progresso de um país.

O caso brasileiro, por exemplo, prevê para os próximos 9 anos, até 2029, um crescimento no Produto Interno Bruto (PIB) de, aproximadamente, 1,9% ao ano segundo o Banco Mundial (Cavalcanti; Júnior, 2018), ao passo que, para o setor elétrico brasileiro, a previsão está na ordem de 28%. Considerando essa diferença, órgãos como o Conselho Nacional de Políticas Energéticas (CNPE), o Comitê de Monitoramento do Setor Elétrico (CMSE) e a Agência Nacional de Energia Elétrica (ANEEL), responsáveis por promover políticas e diretrizes, apresentaram um planejamento extenso para 2029.

Dentre os planos do governo está um investimento de cerca de R\$ 2,34 trilhões distribuídos entre geração centralizada e descentralizada, com R\$ 353 bilhões, sistemas de transmissão, R\$ 104 bilhões, e produtos como o petróleo, a biomassa e o gás natural, com o restante R\$ 1,9 trilhão (ENERGIA, 2020).

No quesito geração, estima-se um significativo crescimento nas energias renováveis eólica e solar, com o aumento e a criação de plantas de grande porte e da microgeração/minigeração, chegando a representar 11% de toda a produção energética do país. Usinas térmicas também apresentaram alta de 8,7% a 10,4% do total, uma diferença que chega a quase 4MW.

Para o sistema de transmissão, o planejamento adota medidas de expansão das linhas de transmissão, com cerca de 40 mil quilômetros a mais do que o apresentado em 2019, e melhora na capacidade de transformação, quase a duplicando.

O principal objetivo desse plano é, portanto, acompanhar o consumo juntamente com uma melhora na qualidade da energia distribuída e na confiabilidade do sistema.

Essas mudanças na estrutura física do sistema elétrico brasileiro podem abrir espaço para discussões de emprego de tecnologias de inteligência artificial em diversos pontos da rede. As redes elétricas inteligentes são arquiteturas de distribuição mais modernas e seguras com o objetivo de integrar todos os usuários da rede de forma digital. Por exemplo, em pontos específicos é possível instalar sensores que enviam dados relativos ao consumo diretamente da unidade consumidora.

A partir desses dados, concessionárias de energia podem planejar intervenções ou reparos na rede de forma mais eficiente como a troca de transformadores antes que apresentem falhas por estarem operando em sobrecarga dentre outras inúmeras vantagens.

1.1 Justificativa

As distribuidoras brasileiras de energia elétrica vêm investindo anualmente valores cada vez maiores em automação, medição e telecomunicação. É fato que elas ainda precisam melhorar bastante no que tange à confiabilidade do sistema, quando comparadas aos sistemas de distribuição existentes em países desenvolvidos. Cabe destacar que, nesses países, os índices de qualidade são significativamente superiores aos brasileiros, uma vez que as redes de distribuição nos centros urbanos são normalmente subterrâneas e não aéreas, como é o caso brasileiro. De qualquer forma, os investimentos feitos nas áreas de automação, medição e telecomunicação permitem melhorar aspectos de operação e controle do sistema elétrico.

Considerando o constante crescimento da carga e a importância cada vez maior de se prever o comportamento futuro dos consumos de energia e potência, este trabalho propõe a aplicação de algoritmos inteligentes para previsão da curva de carga de transformadores de potência instalado em subestações de distribuição juntamente com métodos estatísticos de análise de dados. Dessa forma, variáveis climáticas, consumo de potência ativa, faixa horária do consumo e dias da semana serão analisados.

Sabe-se que atualmente existem vários algoritmos inteligentes para tratamento de dados, entende-se ser oportuno aplicar tais algoritmos na estimativa da curva de carga em transformadores de potência visando o curto prazo, considerando medições das grandezas elétricas de interesse e das variáveis climáticas.

Assim sendo, o presente trabalho analisa os dados de dois transformadores instalados em duas subestações da região sul do Brasil, nas cidades de Santa Maria e São Leopoldo. A opção dessas subestações se deu devido a disponibilidade das medições de carga e das variáveis climáticas.

Este tipo de estudo pode auxiliar as concessionárias de energia no planejamento das operações uma vez que essas empresas devem garantir que as alterações no padrão de consumo de seus clientes não altere a operação do sistema elétrico como um todo.

Com as informações do histórico de potência registrado junto aos transformadores e a integralização de variáveis climáticas em um modelo de previsão de carga, espera-se melhor previsibilidade de como o equipamento irá se comportar em determinadas condições. Atualmente, o comportamento do transformador só é verificado pelos dados obtidos através dos sensores de medições no momento em que essas condições ocorrem. Segundo o modelo

criado portanto, esse comportamento será conhecido de antemão, logo será possível antever possíveis falhas e sobrecarregamento de subestações, por exemplo.

O foco principal será, portanto, responder a perguntas acerca da expressividade das variáveis escolhidas na alteração dos padrões de consumo de energia elétrica, da precisão do modelo de previsão, da utilização dele para prever curvas de cargas futuras de curto prazo e da extrapolação para subestações.

1.2 Estruturação do trabalho

Tendo estabelecido os objetivos e as perguntas a serem respondidas, a construção deste trabalho foi feita de forma gradual. Partiu-se da pesquisa acadêmica ao desenvolvimento de códigos de programação para, por fim, apresentar os resultados obtidos e conclusões. De modo breve, cada etapa realizada será discutida abaixo.

Inicialmente, a atenção estava voltada à aquisição dos dados meteorológicos - concedidos pela Empresa Inmet de meteorologia- e dos dados de potências ativas registradas junto aos transformadores de força instalados em duas subestações, assentido por uma concessionária de energia local.

Posteriormente, usando algoritmos computacionais em Python e técnicas de Aprendizado de Máquina, realizaram-se previsões de carga no curto prazo, considerando os dados obtidos de variáveis meteorológicas e de potência, prevendo-se, assim, a demanda esperada para o respectivo transformador. Consequentemente foram estudadas as variáveis escolhidas e suas correlações, identificando aquelas mais significativas, para garantir que a previsão esteja mais próxima do real possível.

Por fim, foram comparados dados reais, disponibilizados pela concessionária, com os resultados obtidos para o mesmo período de análise. Em relação às comparações feitas, foram também expostas diferentes regressões utilizadas tendo como base parâmetros estatísticos para o melhor resultado possível.

A partir dos dados de 2019, posteriormente, analisou-se a possibilidade de expandir essas previsões para consumos futuros de curtíssimo prazo, dentre uma a algumas horas de 2020. E por fim, foram obtidas conclusões dos resultados obtidos e discutidos meios de aprimorar os mesmos.

1.3 Estrutura do trabalho

Posteriormente ao panorama geral, são apresentadas a divisão em cinco capítulos deste trabalho, mais a bibliografia, nos seguintes tópicos:

- Capítulo 1, Introdução:

Breve apresentação do panorama do setor elétrico brasileiro e importância da previsão de consumo no planejamento dos sistemas de distribuição. A partir disso, foram construídos a justificativa e os respectivos objetivos que serão abordados nesta dissertação.

- Capítulo 2, Revisão Bibliográfica:

Descreve o encaminhamento dos temas abordados juntamente com o embasamento teórico citando, portanto, importantes trabalhos publicados acerca do assunto.

- Capítulo 3, Metodologia:

Fornece a descrição detalhada do método de pesquisa tanto na escolha de dados, variáveis e informações quanto aos algoritmos utilizados. Além disso, serão consideradas as limitações provindas dessas escolhas.

- Capítulo 4, Estudo de Caso:

Apresenta detalhes da metodologia aplicadas a cidades de Santa Maria e São Leopoldo.

- Capítulo 5, Resultados e Discussões:

Apresenta os resultados obtidos pelos métodos utilizados a partir de soluções estatísticas, gráficos e tabelas relevantes.

- Capítulo 6, Conclusões:

Analisa os aspectos importantes da dissertação e as melhorias que poderiam ser implementadas.

2 Revisão Bibliográfica

Este capítulo tem como objetivo demonstrar o estado do conhecimento acerca da previsão de carga, em um curto espaço de tempo, tomando como referência diversas metodologias existentes. Dessa forma, foram verificados que autores empregam e sugerem estruturas de processos distintos manifestando, em suas pesquisas, vantagens e desvantagens da utilização de seus modelos.

A revisão abaixo foi, portanto, construída para observar o panorama geral de consumo do Brasil, o processo de obtenção da curva de carga e suas aplicações, detalhando cada etapa e cada contribuição dos autores citados.

2.1 Consumo de energia no Brasil

O consumo nacional de energia elétrica somou aproximadamente 481 TWh em 2019, crescimento de cerca de 1,4% na comparação com o ano anterior, de acordo com os dados divulgados pela Empresa de Pesquisa Energética (EPE).

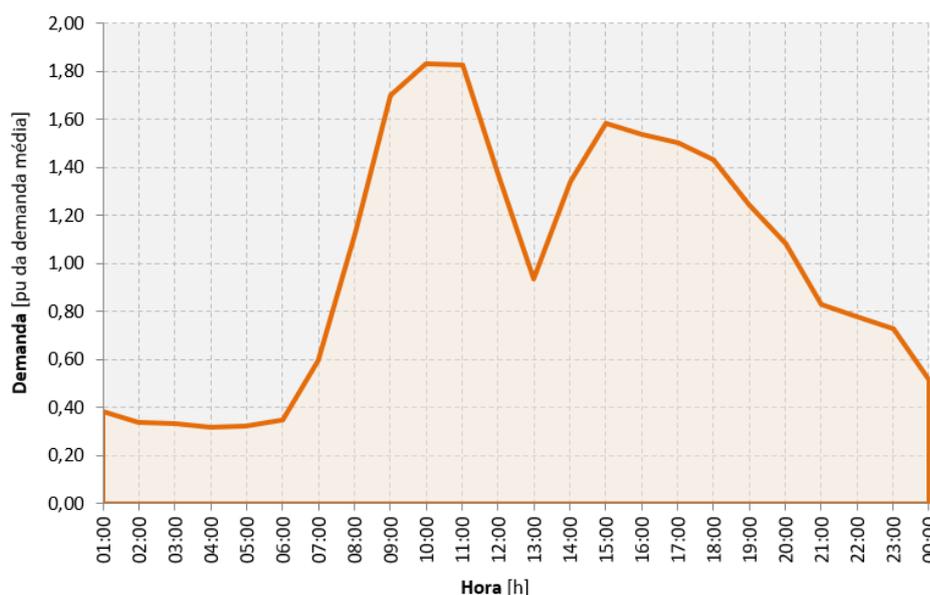
No entanto, as maiores taxas de crescimento em relação a 2018 ocorreram em classes pouco expressivas em quantidade de consumidores como Iluminação Pública, Serviço Público e Consumo Próprio. Dentre as classes residencial, comercial, rural, industrial e outros, a única que sofreu redução de consumo foi a classe industrial, com queda de 8,9%.

Em termos numéricos, consumidores residenciais e industriais são responsáveis por grande parte do consumo do país com 64,4% da energia total, representando 142,4 TWh e 167,6 TWh respectivamente. As classes comerciais e outras por sua vez contiveram 19,1% e 16,6%, respectivamente.

Essas informações tornam-se relevantes à medida que as curvas de carga são extremamente influenciadas pela classe de consumidor assim como por fatores climáticos, socioeconômicos, sazonalidades dentre outros.

A Figura 1, por exemplo, ilustra a curva de carga típica mensal para um consumidor industrial com consumo acima de 10 MWh. Esta curva foi obtida junto a uma distribuidora de energia elétrica, localizada no estado do Rio Grande do Sul. O objetivo é meramente ilustrativo, uma vez que as curvas para os clientes podem apresentar algumas diferenças conforme características da região e da atividade.

Figura 1 – Curva de carga típica industrial para cargas >10MWh



Fonte – Elaborado pela autora

Com base na Figura 1, percebe-se um padrão no consumo de energia na qual ocorrerem dois pontos de pico de consumo um às 10h e outro às 15 horas, de modo que esse comportamento representa a dinâmica de operação de grande parte dos clientes industriais. Esses padrões, entretanto, são bastante diferentes para consumidores residenciais, rurais e comerciais.

As próximas seções se encarregarão de analisar as especificidades das curvas de carga adentrando as classificações utilizadas e os métodos mais comuns de previsão realizados pelos autores da área.

2.2 Introdução a previsão de curvas de carga

Previsões adequadas de curva de carga são questões críticas para operações do dia a dia de concessionárias de energia, indústrias e outros serviços. Decisões como compra ou venda de energia e mudanças de planejamentos são baseadas em tais previsões, ou seja, são informações de vital importância para *Stakeholders*¹ nos mercados de geração, transmissão e distribuição de energia elétrica, (Khan; Noor, 2020).

Tendo em vista que cada utilização tem sua especificidade, é necessário escolher o tipo de modelo mais adequado para cada caso. Logo, diversos autores separam os modelos em categorias a partir do espaço de tempo que desejam utilizar.

Modelos de previsão de carga podem ser agrupados em quatro grupos: previsões de curtíssimo prazo, curto prazo, médio prazo e longo prazo englobando um dia, duas semanas, um ano e três anos, respectivamente (Hong; Fan, 2016).

¹ *Stakeholders*: são as pessoas e as organizações que podem ser afetadas por um projeto ou empresa, de forma direta ou indireta, positiva ou negativamente

No curtíssimo prazo, que tem como horizonte de previsão de algumas horas à frente da previsão, são utilizadas como entradas apenas os valores de consumo de instantes anteriores à previsão.

Na previsão de curto prazo, o interesse é de alguns dias adiante, logo são mais utilizados como dados de entrada as informações de carga e temperatura da área de estudo.

Nas previsões de demanda de médio prazo, é analisado um horizonte de meses. Esse tipo de previsão é muito utilizada para o planejamento e manutenções das redes elétricas, planejamento de compra de combustíveis e negociações de contratos.

No longo prazo, trabalha-se com previsões anuais. Esse tipo de previsão normalmente fornecem informações para produtores e distribuidores sobre a evolução da demanda, o que possibilita o estudo de estratégias de expansão das linhas de transmissão, das usinas geradoras ou dos vários componentes da rede elétrica.

Para Khan e Noor (2020), os modelos são divididos apenas em três categorias: curto prazo, incluindo poucas horas até uma semana, médio prazo, abrangendo poucas semanas até poucos meses, e longo prazo, válidos de um ano a vários anos.

Para este trabalho, a separação sugerida por Khan e Noor (2020) torna-se mais adequada, uma vez que objetivam-se previsões da curva entre uma hora a um dia, encaixando o modelo de previsão na categoria de curto prazo.

O processo de previsão de curto prazo, parte do princípio de que o mesmo mecanismo que gerou a demanda no passado irá gerar a demanda futura. Portanto, sabendo o mecanismo gerador, este é usado para gerar as demandas futuras.

A previsão de demanda para curto prazo é então produzida por meio da extrapolação de relações pré-determinadas entre carga e variáveis correlacionadas, tais como faixa horária e fatores climáticos. Determinar essa relação corresponde à sequência de dois passos a serem realizados: primeiro, identificar a relação entre a carga e as variáveis relacionadas e segundo, quantificar essa relação a partir de parâmetros de técnicas de estimação (Soliman, 2010).

As várias metodologias criadas para resolver os problemas de estimação no curto prazo têm incorporado ideias tradicionais como métodos estatísticos, análise de regressões, modelo auto-regressivo integrado de médias móveis (ARIMA), técnicas de suavização exponencial e filtro de Kalman. Atualmente, algoritmos baseados em inteligência artificial têm gerado grande progresso em sistemas não-lineares de previsão como métodos de lógica fuzzy, de agrupamento, redes neurais, máquinas de vetores de suporte (SVM), técnicas híbridas, entre outros, (Rafi; Nahid-Al-Masood, 2020).

Métodos de previsões de carga podem ser classificados em termos do grau de complexidade aplicados na análise matemática, existindo dois tipos de métodos: o quan-

titativo e o qualitativo. Métodos quantitativos são utilizados, em geral, para obterem uma antecipação precisa da carga para planejamentos. Os métodos quantitativos podem ser divididos em três categorias: técnicas tradicionais, técnicas tradicionais modificadas e técnicas computacionais (Lekshmi; Subramanya, 2019).

A previsão tradicional é utilizada para determinar as futuras demandas de carga para o desenvolvimento da infraestrutura e tendências de desenvolvimento de um país. No período inicial, essas estimativas foram feitas usando técnicas matemáticas convencionais. Com o desenvolvimento de ferramentas matemáticas, essas técnicas foram aprimoradas para vários campos de estudos de previsão dentre eles os métodos de Regressão, Regressão Múltipla e Métodos de suavização exponencial.

As técnicas tradicionais modificadas são aprimoradas para que haja atualização automática em modelos de previsão segundo variações nas condições ambientais. Algumas das formas aprimoradas das técnicas tradicionais são previsão de carga adaptativa, regressões e técnicas baseadas em máquinas de suporte.

Técnicas computacionais surgiram para aumentar a eficiência dos modelos. O *Soft-computing*, como é chamado, é uma abordagem que tenta imitar a capacidade excepcional da mente humana de aprender e raciocinar em um ambientes imprevisíveis para replicá-la em sistemas computacionais inteligentes. As técnicas computacionais formam uma variedade que inclui lógica difusa, redes neurais, algoritmos evolutivos etc.

Este trabalho preocupa-se em analisar duas das diferentes técnicas usando ferramentas computacionais para implementar métodos de regressões e de redes neurais (técnica computacional).

A abordagem da regressão comumente utiliza uma função linear como método de previsão, ou seja, recorre à relação de causa e efeito entre as variáveis climáticas e a variável de potência elétrica para usá-la durante a previsão. Com uma combinação linear, o método da regressão assume sem justificativas uma relação linear e, portanto, não possui a versatilidade de identificar os elementos espaciais e temporais não-lineares presente na relação entre essas variáveis, (Patel; Pandya; Aware, 2015).

A abordagem da série temporal utilizada na rede neural, por sua vez, é definida como pares de dados gerados sequencialmente no tempo. Esse modelo assume que, na ausência de grandes interrupções em fatores críticos de eventos recorrentes, os dados deste evento no futuro estarão correlacionados com os eventos do passado e este poderá ser expresso pelo modelo desenvolvido segundo os eventos do passado, (Deng; Jirutitjaroen, 2010).

Vale ressaltar ainda que as diferentes técnicas recorrerão à previsões de carga utilizando dados de fatores climáticos como temperatura, velocidade dos ventos, umidade e pressão atmosférica. O uso de parâmetros climáticos em previsões de curto prazo tem

sido adotado em modelos estatísticos, de regressões e redes neurais apresentando bons resultados quando há uma relação significativa entre demanda e condições climáticas (Silva et al., 2019).

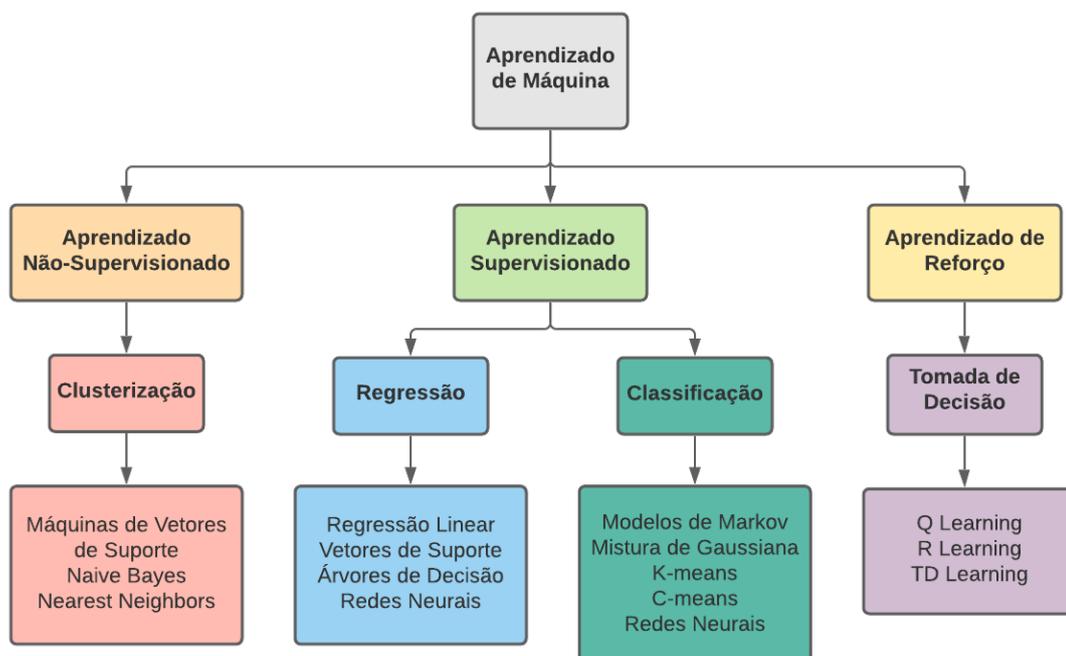
A previsão de carga de curto prazo com informações do histórico de demanda podem ser vinculadas à variáveis não relacionadas ao consumo uma vez que se analisado o horizonte de previsão, pode-se inferir um impacto significativo das variáveis meteorológicas sobre o padrões de consumo unitário (Silva et al., 2017).

2.3 Modelos de Aprendizado de Máquina e Redes Neurais

O aprendizado de máquina explora a construção de algoritmos que podem aprender com seus erros e fazer previsões sobre dados a partir de três abordagens de aprendizagem: supervisionada, não supervisionada e por reforço. Isso permite produzir decisões e resultados confiáveis e repetíveis.

O presente trabalho analisa tipos de regressões em Aprendizado de Máquina para a previsão de carga no curto prazo: Árvore de Decisão (AD), Florestas de Decisão Aleatória (FDA) e Regressão de Vetores de Suporte (RVS) e um modelo de rede neural recorrente chamado Memória de Longo e Curto Prazo (MLCP). Neste sentido, este trabalho fica dentro da categoria de aprendizado de máquina supervisionado segundo o Fluxograma da Figura 2.

Figura 2 – Tipos de Aprendizado de Máquina

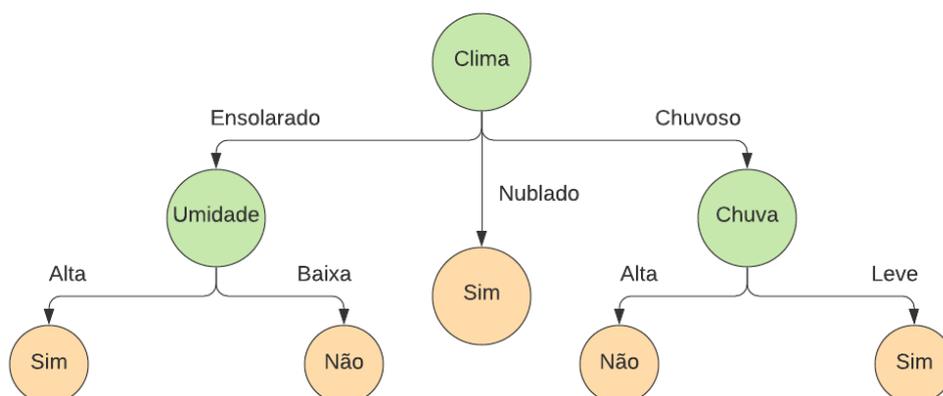


Fonte – Elaborado pela autora

O conceito básico das árvores de decisão é computar tempos instantes à frente de uma resposta ou de uma classe segundo conjuntos conhecidos de entradas. Ao crescer uma árvore binária, cada teste de nó interno é aplicado a uma entrada e com base no resultado, o processo de tomada de decisão hierárquica flui em direção a ramos esquerdo ou direito ou subárvores. Quando uma eventual folha nó aparece, a previsão é fundamentada e realiza a média ou agrega todos os pontos de dados de treinamento que chegam a esse terminal (Chowdhury et al., 2018).

O principal objetivo deste classificador, portanto, é realizar a previsão do alvo baseado nos diversos atributos de entrada, mais especificamente, cada nó da árvore de decisão corresponde a uma variável de entrada e esse nó é ramificado para todos os possíveis valores dessa variável de entrada. Em contrapartida, os nós-folha representam a classe alvo de acordo com os valores das variáveis de entrada, desenhando um caminho da raiz até as folhas. A Figura 3 demonstra uma Árvore de Decisão simbólica.

Figura 3 – Árvore de Decisão exemplo: ‘Sair de casa segundo a previsão do tempo’



Fonte – Elaborado pela autora

A árvore de decisão é amplamente utilizada para construir modelos de previsão em virtude de sua simplicidade, do poder de compreensão do modelo e da fácil implementação computacional, (Dong et al., 2020).

A Florestas de Decisão Aleatória é um algoritmo de máquina flexível derivado da Árvore de Decisão (Lahouar; Ben Hadj Slama, 2015). Esse método cria diversas árvores de decisão de forma aleatória ao mesmo tempo, nesse sentido, alcança melhores resultados, em geral, do que o método de Árvore de Decisão (Ali et al., 2012). O método Florestas de Decisão Aleatória é um algoritmo que constrói uma coleção de árvores não correlacionadas e, em seguida, calcula a média delas (Breiman, 2001). O método de Florestas de Decisão Aleatória é utilizado usualmente para construir modelos de previsões uma vez que parâmetros são facilmente manipuláveis e o modelo é capaz de convergir para uma solução global ótima (Xiaoyu Wu et al., 2016).

A Regressão de Vetores de Suporte, por sua vez, é um algoritmo de aprendizado supervisionado que toma como entrada um conjunto de dados e prediz, segundo cada entrada, qual classes a entrada faz parte. Dessa maneira, o método RVS é considerado uma regressão linear binária já que cada dado pertence a uma de duas categorias. A Regressão de Vetores de Suporte é especialmente apropriado para pequenas amostras de entrada e tem se tornado uma técnica atraente para vários problemas de previsões (Ye; Yang, 2019).

No que diz respeito as redes neurais recorrentes, sabe-se que muitos pesquisadores utilizam esse método para trabalhos com séries temporais uma vez que são especialmente úteis no processamento dos dados já que incluem um *loop*² de *feedback*³ pelo qual a saída do passo n-1 é retroalimentada à rede para afetar o resultado do passo n. Desse modo, as redes recorrentes produzem modelos dinâmicos, que mudam ao longo do tempo, de forma a produzir classificações mais precisas dependentes do contexto em que se encontram inseridos.

Um dos métodos mais comuns levemente adaptados das redes neurais recorrentes usuais é a chamada Memória de Longo e Curto Prazo, ou MLCP, proposto por Hochreiter e Schmidhuber em 1997. Este método elimina o problema do alongamento no processamento gerado pelo *Backpropagation*⁴ que cria uma insuficiência no decaimento do erro (Cui et al., 2020). Os MLCPs ajudam a preservar o erro que pode ser copiado por tempo e camada, ou seja, mantêm um erro mais constante para que a rede possa aprender durante vários passos seguidos. Ainda segundo Cui, He, Di, Lu, Dai e Lv (2020), MLCPs tem sido aplicadas em previsão de carga uma vez que geram erros menores comparados a outros métodos e aumentam significativamente a precisão do modelo criado.

2.4 Análise de estudos relacionados

A utilização das MLCPs para a previsão da carga da Estônia no curto prazo, segundo variáveis de potência em MW para cada hora entre 1 de julho de 2017 a 31 de julho de 2017, se fez eficiente (Jiang et al., 2018). O trabalho levou em consideração dados históricos de uma hora anterior e comparou-os, segundo parâmetros de Raiz Quadrada do Erro Médio (RQEM) e perdas, com um modelo de Aprendizado de Máquina utilizando RVS. Os seguintes resultados foram encontrados:

² *Loop*: conjunto de instruções que um programa de computador percorre e repete um significativo número de vezes até que sejam alcançadas as condições desejadas.

³ *Feedback*: informação que o emissor obtém da reação do receptor à sua mensagem, e que serve para avaliar os resultados da transmissão.

⁴ *Backpropagation*: algoritmo de aprendizado de máquina que recalcula o valor dos pesos atribuídos de camadas de neurônios.

Tabela 1 – Comparação entre os resultados dos algoritmos RVS e MLCP

Modelos de previsão	RQEM	PERDAS
MLCP	38,74	0,0439
RVS	43,16	0,0872

Fonte – Adaptado a partir de Jiang (2018)

As diminuições nas medições da raiz quadrada do erro médio e do parâmetro de perda dos algoritmos confirmam maior precisão do modelo MLCP. Segundo o estudo, o modelo Memória de Longo e Curto Prazo consegue acompanhar melhor a suavidade da curva da carga, enquanto o método RVS apresenta picos destoantes da realidade.

Um segundo estudo comparando diferentes metodologias de previsões de carga de curto prazo é realizado a partir da implementação de uma rede neural densa (RND), utilizando *Backpropagation* e de um modelo de Árvore de Decisão (AD) (Dehalwar et al., 2016). Os dados de variáveis climáticas foram adicionados a ambos os modelos bem como a hora do dia, o dia da semana, a carga das 24 horas prévias, a carga média das últimas 24h e a carga das 168h prévias. Os parâmetros de comparações utilizados foram o Erro Absoluto Médio (EAM), a Média Percentual Absoluta do Erro (MPAE) e a Raiz Quadrada do Erro Médio (RQEM).

Os resultados obtidos pelo estudo encontram-se na Tabela 2.

Tabela 2 – Comparação dos modelos RN e Árvore de Decisão

Modelos	MPAE	EAM [MW]
RND	1,9%	167,91
AD	1,54%	136,39

Fonte – Adaptado a partir de Dehalwar (2016)

A comparação dos modelos, segundo a porcentagem do erro, define que o modelo de Árvore de Decisão apresentou melhor precisão em comparação com o modelo de rede neural.

Outro estudo realizado, com ênfase na área de previsão de curva de carga de curto prazo, compara a implementação de uma Árvore de Decisão (AD) com vários outros métodos de previsão de carga (Liu et al., 2018). Os dados de carga foram obtidos em 96 pontos de uma área da China em um intervalo de 15 minutos de 22 a 31 de outubro de 2014 para prever o dia 01 de novembro. As variáveis de entrada utilizadas foram a potência no mesmo horário de três dias anteriores, a temperatura média de três dias anteriores, a umidade média de três dias anteriores, a temperatura do dia e a umidade do dia.

Os algoritmos de comparação RVS, ARIMA, rede neural (RN) pelo método de *Backpropagation* e os valores reais foram utilizados juntamente com as métricas Erro

Absoluto Médio (EAM), Média Percentual Absoluta do Erro (MPAE) e Raiz Quadrada do Erro Médio (RQEM). Os resultados do estudo encontram-se na Tabela 3.

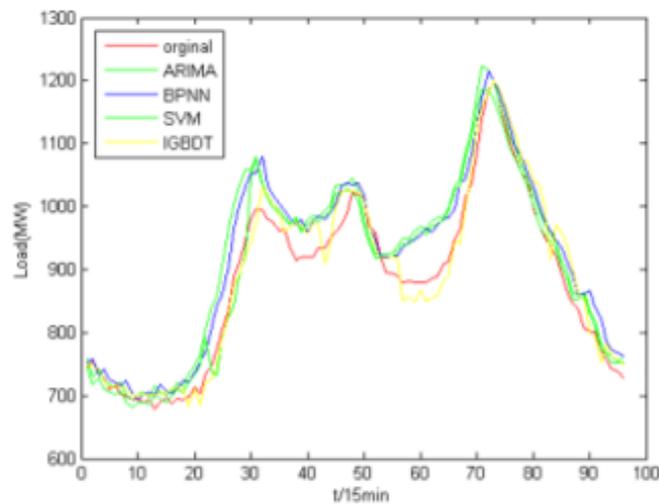
Tabela 3 – Comparação dos resultados encontrados pelos algoritmos

Modelos	EAM [MW]	MPAE [%]	RQEM [MW]
ARIMA	33,79	3,78	42,46
RN	30,29	3,18	37,13
RVS	29,67	2,59	40,57
AD	19,36	2,17	22,67

Fonte – Adaptado a partir de Liu (2018)

Em comparação aos outros modelos, a Árvore de Decisão apresentou os menores valores no erro absoluto médio, na raiz do erro quadrado e no erro percentual absoluto médio, logo, pode-se concluir que o modelo é o mais preciso e mais robusto entre os analisados. A curva de carga gerada pelos algoritmos foi apresentada na Figura 4 abaixo.

Figura 4 – Resultados da previsão de carga



Fonte – Adaptado a partir de Liu (2018)

Através da Figura 4 pode-se perceber que o modelo de Árvore de Decisão é aquele que melhor acompanha os valores originais de potência analisados no período.

A análise dos três trabalhos descritos acima serve de referência para os modelos de regressão e de redes neurais que serão desenvolvidos a seguir. As métricas de precisão e os modelos utilizados pelos autores Jiang, Dehalwar e Liu foram replicados e alterados de acordo com a disponibilidade dos dados e de recursos computacionais.

3 Metodologia

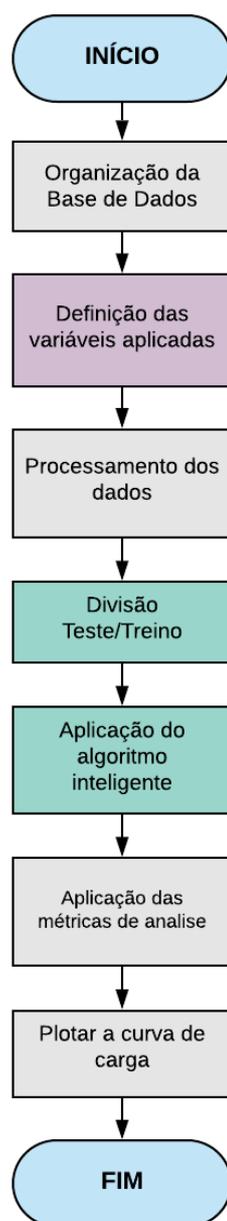
A metodologia aplicada neste trabalho baseia-se na análise de dados para estimativa da curva de carga em transformadores de potência. Os transformadores de potência instalados nas subestações das distribuidoras de energia são equipamentos de suma relevância na operação do sistema de distribuição. Esses equipamentos possuem monitoramento de carga (potências ativa e reativa, corrente e tensão), além do monitoramento das temperaturas do óleo e do enrolamento. Esses equipamentos podem apresentar uma sensibilidade significativa no aumento das temperaturas internas quando combinados os fatores carga e as temperatura ambiente. Assim, prever os níveis de carregamento de um transformador de potência, projetando sua curva de carga pode ser uma ferramenta bastante útil para evitar problemas na operação do sistema em dias com alto carregamento e altas temperaturas. Buscando realizar a projeção da curva de carga do transformador de forma mais assertiva possível, este trabalho implementa uma metodologia capaz de analisar os dados históricos de carga medidos junto à subestação e os dados climáticos registrados na cidade onde a subestação se encontra. Desse modo, tem-se um modelo capaz de estimar curvas de carga de curto prazo para o transformador de força utilizando previsões climáticas (24h) e dados de carga registrados junto ao equipamento até a hora atual.

3.1 Etapas de construção do modelo

Os modelos de Aprendizado de Máquina (Árvore de Decisão (AD), Florestas de Decisão Aleatória (FDA) e Regressão de Vetores de Suporte (RVS) e Rede Neural (Memória de Longo e Curto Prazo (MLCP) utilizados no presente trabalho apresentam estrutura idêntica. Nesse contexto, sete etapas são construídas para obter o modelo de previsão de potência ativa para o curto prazo dos transformadores ligados ao barramento das respectivas subestações: seleção da base de dados, definição das variáveis aplicadas, processamento dos dados, divisão dos dados em teste e treino, aplicação dos algoritmos inteligentes, aplicação das métricas escolhidas e por fim a obtenção das curvas de carga.

As etapas são ilustradas através do Fluxograma da Figura 5 e serão explicadas a seguir.

Figura 5 – Fluxograma dos modelos de previsão



Fonte – Elaborado pela autora

3.2 Organização da base de dados

Nesta fase, a coleta da base de dados é realizada a partir de dados disponibilizados pela concessionária de energia local e pela empresa de meteorologia. Os dados de potência ativa, reativa e fatores climáticos do ano de 2019 são coletados e inseridos no algoritmo.

3.3 Definição das variáveis

Em seguida, com o objetivo de modelar o problema, foram selecionadas variáveis dependentes listadas abaixo, consideradas relevantes, que alterem ou impactem de al-

guma maneira o comportamento da variável de saída em questão, a potência ativa dos transformadores conectados ao barramento da subestação.

- Histórico de Potência Ativa da hora anterior (MW)
- Temperatura máxima, mínima e instantânea ($^{\circ}\text{C}$)
- Umidade máxima, mínima e instantânea (%)
- Pressão (mmHg)
- Velocidade do vento (m/s)
- Dias da semana
- Hora do dia (h)

Essas variáveis foram escolhidas porque, na prática, representam as principais alterações nos hábitos de consumo horário/diário de unidades consumidoras de energia elétrica. A flutuação na carga deve-se principalmente a dois fatores: a regularidade sazonal diária/semanal atribuída aos hábitos de produção e consumo de energia elétrica e a fatores mais imediatos provindos de mudanças nas condições climáticas da região em questão (Qiuyu et al., 2017).

As variações de temperaturas (máximas e mínimas) podem afetar o comportamento de consumo de energia elétrica por parte dos usuários, tais como: ajuste de chuveiros elétricos, acionamentos de aparelhos condicionadores de ar e maior uso de refrigeradores. Considerando os pontos citados, entende-se que estes podem representar significativos incrementos no consumo de energia elétrica. Para Hong e Kim (2015), a temperatura atmosférica é o principal fator causador de variações na demanda e pode ser descrita na forma de uma equação quadrática com a eletricidade.

Os valores de umidade podem alterar a sensação térmica da região, por exemplo, quando o tempo está quente e demasiadamente úmido a sensação será maior do que a marcada pelos termômetros. Caso a temperatura seja baixa e a umidade grande, a sensação térmica será de que está mais frio do que realmente as medições confirmam. Logo, pode-se esperar que valores de umidades relativas do ar são capazes de modificar também o comportamento do consumo de energia elétrica de consumidores.

A velocidade dos ventos é outra variável a ser considerada uma vez que, juntamente com a umidade relativa do ar, acarreta uma alteração da sensação térmica e, como visto, também contribui para mudanças no comportamento dos consumidores.

Ainda segundo Hong e Kim (2015), o impacto de variáveis como umidade e velocidade do vento na variação de carga pode não ser tão facilmente calculado como no caso da temperatura, entretanto, essas variáveis apresentaram uma significativa correlação com a demanda quando analisadas em um curto período de tempo de 6 a 24 horas.

A variável dias da semana é também contabilizada posto que, se tomado como exemplo consumidores residenciais, em geral a curva de carga residencial depende de diversos fatores, dentre eles socioeconômicos, preferências do cliente, clima, tipo de aplicação, (Elma; Selamoğullar, 2017). Isso implica na criação de diferentes modelos para a demanda de energia elétrica para os dias da semana e para o final de semana, para o verão e para o inverno, por exemplo. Essas diferenças, portanto, foram ponderadas através da variável dias da semana.

3.4 Processamento dos dados

Tendo em vista que o período de análise dos dados meteorológicos é relativo ao período de 2019 e registrado de hora em hora, a base de dados utilizada para cada transformador terá um total de 8760 amostras para cada fator climático incluído no estudo. Sabe-se ainda que a base de dados para a variável de potência ativa coletada no mesmo período possui periodicidade de quinze minutos, logo vê-se a necessidade de ajustar o tempo de aquisição desses dados para enquadrar-se ao tempo de aquisição das variáveis climáticas.

Outro ponto importante de divergência a ser considerado entre os bancos de dados de potência e de variáveis climáticas é o fuso horário. As empresas de meteorologia e concessionárias de energia podem utilizar dados em fuso horários distintos, sendo necessária subtração ou adição de uma a algumas horas para convertê-los ao horário local ou ao fuso horário padrão UTC. Essas alterações são importantes uma vez que os dados climáticos e de potência precisam estar alinhados visto que são variáveis extremamente dependentes do tempo de ocorrência.

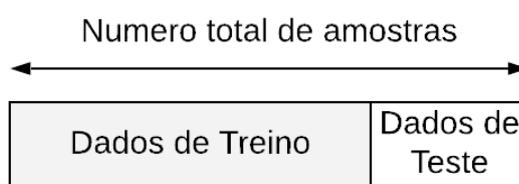
O enquadramento dos dados climáticos e de potência no tempo permite também realizar uma fusão entre os dois conjuntos de dados, com o objetivo de excluir linhas de dados presentes em um conjunto de dados e ausentes em outro e, assim, diminuir fontes de erros na previsão da carga. Essas assimetrias são normalmente encontradas quando ocorrem erros na coleta de dados, tanto por mau funcionamento do software quanto do equipamento. Neste caso, o algoritmo filtra duas colunas, Dia e Hora, dos dois conjuntos de dados, a partir da função Merge, e retira a linha caso um dos conjuntos não apresente o dia ou a hora em questão, segundo a função Dropna. Essas funções são pré-programadas nos ambientes de programação Python e muito utilizadas no pré-processamento de dados.

Por fim, a variável categórica “Dias da Semana” é transformada em uma variável quantitativa de modo a permitir que o software reconheça essa variável e utilize-a durante seu aprendizado.

3.5 Divisão Treino e Teste

Após a etapa de pré processamento dos dados na qual os dados são ajustados, realiza-se a separação em conjuntos de dados de treino e de teste. Esse método permite dividir os dados entre o conjunto que aprende as correlações entre as variáveis dependentes, representando 80% do total, e o conjunto que testa o algoritmo, 20% dos dados restantes. A Figura 6, demonstra a divisão da base de dados em teste e treino.

Figura 6 – Divisão teste e treino dos dados



Fonte – Elaborado pela autora

Esta etapa é realizada de maneira não aleatória, ou seja, são escolhidos os primeiros 80% dos dados para treino e os últimos 20% para testar a aderência do modelo. Isso permite a construção da curva de carga de vinte e quatro horas diárias sem falhas nas previsões.

3.6 Aplicação do algoritmo inteligente

Nesta etapa, seleciona-se o modelo de Aprendizado de Máquina a ser utilizado. Esse processo é realizado manualmente diretamente nas linhas de código pela IDE Spyder na versão 3.9.0 do Python .

Para os casos de aplicação da Árvore de Decisão (AD), Florestas de Decisão Aleatória (FDA) e Regressão de Vetores de Suporte (RVS) é utilizado a biblioteca *Sklearn* do Python que possui esses modelos já programados. A biblioteca *Sklearn*, ou *Scikit-learn*, é um módulo Python para Aprendizado de Máquina baseado em *NumPy*, *SciPy* e *Matplotlib*. Ele fornece ferramentas simples e eficientes para mineração e análise de dados acessível a todos e reutilizável em diversos contextos. A implementação desses métodos, portanto, requer apenas a inserção dos conjuntos de dados de treino/teste e de alguns parâmetros específicos para funcionamento dessas regressões.

Para o modelo de série temporal criado pela utilização da rede neural Memória de Longo e Curto Prazo (MLCP), utiliza-se a biblioteca *Keras* do Python, capaz de criar redes neurais segundo parâmetros de quantidade de neurônios, tamanho do lote, épocas e funções de ativação. Esta biblioteca, foi desenvolvida para facilitar experimentações rápidas, isto é, sem que tenha que dominar cada um dos *backgrounds* ¹ (*Tensorflow* ou

¹ *Backgrounds*: conjunto das condições, circunstâncias ou antecedentes de uma situação, acontecimento ou fenômeno.

Theano), de maneira rápida e eficiente. O modelo apresenta uma camada de entrada, camadas ocultas utilizando o modelo MLCP e uma última camada de saída. É necessário também, como os modelos de Aprendizado de Máquina, inserir os conjuntos de dados de treino/teste e parâmetros específicos para o correto funcionamento do modelo.

Algumas outras bibliotecas do Python são utilizadas: *Pandas*, *Numpy* e *Matplotlib*. *Pandas* é uma biblioteca para manipulação e análise de dados perfeita para iniciar análises exploratórias de dados, pois permite ler, manipular, agregar e plotar os dados em poucos passos. A biblioteca *Numpy*, por sua vez, é uma ferramenta excelente para realizar inúmeras operações com matrizes. A biblioteca *Matplotlib* permite implementar ótimos gráficos e, através deles, reproduzir inúmeros dados de forma mais clara.

3.7 Aplicação das métricas de comparação

Parâmetros estatísticos que calculam o erro e a proporção da variância são utilizados como meio de comparação entre os modelos e de aderência de cada modelo aos valores reais de potência ativa do transformador em questão. Dentre uma gama variada de parâmetros estatísticos são colocados em execução três: R quadrado, média dos erros quadráticos (MEQ) e o erro absoluto médio (EAM).

A métrica R quadrado, ou R², corresponde a proporção da variância entre os valores previstos pelo modelo e valores reais de potência do transformador em um modelo de regressão. Desse modo, quanto maior o valor de R² mais adequado é modelo para o problema.

O MEQ por sua vez, calcula a média dos erros quadrados entre os valores observados (reais) e previsões (valores gerados pelos modelos criados). Esse erro é calculado através da Equação 3.1.

$$MEQ = \frac{1}{n} \sum_{j=1}^n |y_j - \bar{y}_j|^2 \quad (3.1)$$

onde y_j é a medida real de potência ativa, \bar{y}_j é o valor previsto de potência ativa e n o número de amostras.

Por último, a EAM corresponde ao erro absoluto médio entre dois pares de observação de mesma natureza, ou seja, entre os valores reais e os valores previstos. Sendo assim, tanto para EAM quanto para MEQ quanto menor seus valores, mais adequado o modelo.

$$EAM = \frac{1}{n} \sum_{j=1}^n |y_j - \bar{y}_j| \quad (3.2)$$

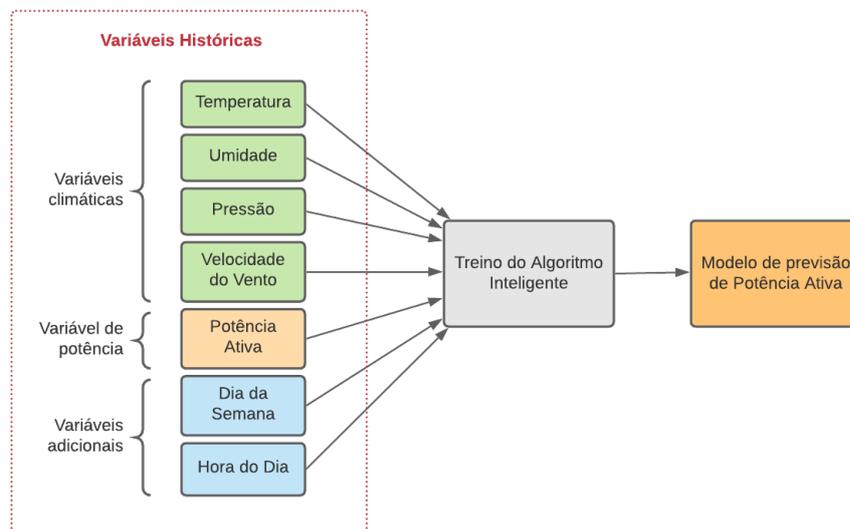
onde y_j é a medida real de potência ativa, \bar{y}_j é o valor previsto de potência ativa e n o número de amostras.

3.8 Funcionamento geral dos modelos

De modo geral, pode-se dividir em dois blocos o processo de previsão de curva de carga. Primeiro, cria-se um modelo de previsão através de dados históricos de potência ativa e de variáveis climáticas, representado pelo conjunto de dados de treino. Para em seguida, utilizar esse modelo inserindo novas variáveis climáticas previstas e de potência ativa da hora anterior para de fato realizar a previsão de carga horária e a aderência do modelo.

O diagrama da Figura 7 apresenta as variáveis históricas de entrada sendo inseridas para criação do modelo de previsão de carga.

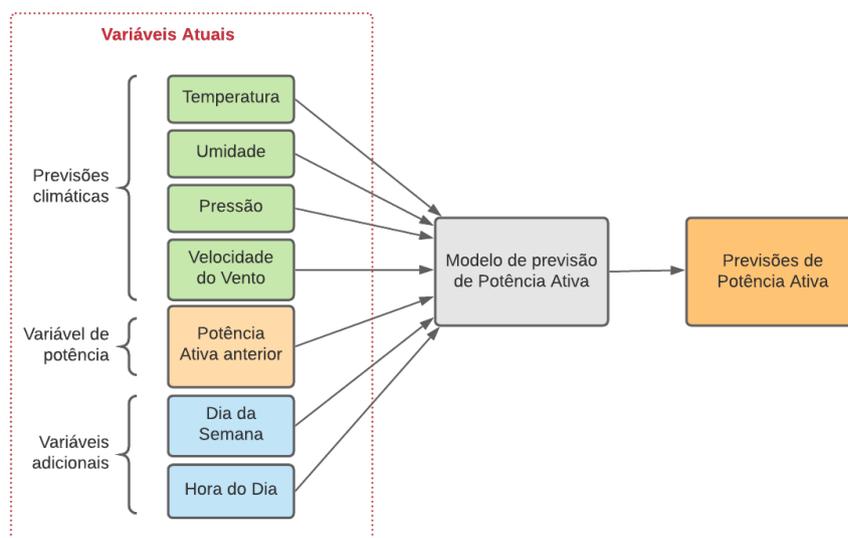
Figura 7 – Diagrama da criação do modelo



Fonte – Elaborado pela autora

A Figura 8 demonstra o processo final para a previsão de carga com as variáveis atualizadas incluídas no modelo criado.

Figura 8 – Diagrama das variáveis



Fonte – Elaborado pela autora

O modelo de previsão descrito é então empregado em um estudo de caso para dois transformadores de força, um em Santa Maria e outro na região de São Leopoldo, e será detalhado a seguir.

4 Estudo de Caso

Este capítulo traz informações relevantes da aplicação da metodologia utilizada para as duas cidades em questão: Santa Maria e São Leopoldo. Informação sobre as bases de dados, características de cada região e outras informações pertinentes são aqui apresentadas.

4.1 São Leopoldo

O município de São Leopoldo localiza-se na região metropolitana de Porto Alegre. Essa região é muito conhecida por verões longos, quente e abafado e invernos curtos e amenos. Ao longo do ano, em geral, a temperatura varia de 10°C a 31°C e raramente é inferior a 4°C ou superior a 36°C.

Os dados de temperatura, umidade, pressão e velocidade dos ventos foram coletados de hora em hora de todos os dias do ano de 2019 com a intenção de caracterizar melhor as condições climáticas no período de funcionamento dos transformadores.

Uma base de dados é então criada, contendo dados meteorológicos provindos da cidade de Campo Bom, cidade suficientemente próxima a São Leopoldo, cerca de 20km, para considerar as previsões do tempo adequados para a cidade em questão, e dados de potência do transformador conectado ao barramento da subestação da região.

Ao utilizar a etapa de processamento de dados, explicado na seção 3.4 do presente trabalho, tem-se que a base de dados para a cidade de São Leopoldo é constituída por 8734 linhas para as variáveis de temperatura, pressão, umidade, velocidade dos ventos, dia da semana, hora e potência da hora anterior. Neste contexto, sabe-se que o tamanho dos conjuntos de dados de treino e teste são de 80% e 20% o que representa, respectivamente, 6987 e 1747 linhas.

Os resultados das previsões de carga são então extraídos para todos os modelos analisados, Florestas de Decisão Aleatória, Regressão de Vetores de Suporte, Árvore de Decisão e Memória de Longo e Curto Prazo, de modo a contemplar os parâmetros estabelecidos R2, EAM e MEQ (Tabela 6).

Em seguida, é analisado o comportamento de cada modelo ao excluir as variáveis climáticas para verificar a importância das mesmas na qualidade do resultado das previsões. Logo, são extraídos novamente os parâmetros de comparação e erro de cada modelo, Tabela 7, e analisados os resultados encontrados.

Uma comparação entre as curvas de carga com variáveis climáticas e sem de um dia aleatório é feita para exemplificar a diferença encontrada nos resultados de modo a expor a importância das variáveis climáticas na resolução das previsões de carga, Figuras 9, 10 e 11.

Após esse processo, uma análise dos dias da semana é realizada com três dias escolhidos aleatoriamente: segunda-feira, quarta-feira e sexta-feira, com o objetivo de demonstrar se existe um padrão seguido pelas curvas de carga durante esses dias. O mesmo é realizado para o final de semana, sábado e domingo, e feriados nacionais, como o Natal e Finados. Vale ressaltar que esta análise é realizada com a presença das variáveis climáticas em todos os modelos. A Tabela 4 apresenta os dias aleatórios utilizados para construir as curvas de carga mencionadas.

Tabela 4 – Dias de análise das curvas de carga

	Dia Analisado
Segunda-Feira	02/12/2019
Quarta-Feira	11/12/2019
Sexta-Feira	08/11/2019
Sábado	30/11/2019
Domingo	27/10/2019
Natal	25/12/2019
Finados	02/11/2019

Fonte – Elaborado pela autora

A previsão da carga aos finais de semana (sábados e domingos) são realizadas uma vez que o comportamento da carga pode não apresentar padrões tão definidos como nos dias de semana, devido ao comportamento dos consumidores e devido a possibilidade de eventos que de alguma forma podem alterar os hábitos de consumo de energia elétrica.

Assim como no final de semana, os feriados ou dias com eventos importantes (eleições, jogos de futebol, capítulos finais de novelas, entre outros), podem também apresentar comportamento incomum da carga. Desse modo uma análise de tais dias se vê necessária.

Os detalhes dos resultados obtidos nesse estudo de caso são apresentados no capítulo de Resultados e Discussões.

4.2 Santa Maria

A cidade de Santa Maria localiza-se na região central do Estado do Rio Grande do Sul. Semelhante à cidade de São Leopoldo, essa região também é conhecida por verões longos, quente e abafado e invernos curtos e amenos. Ao longo do ano, em geral, a temperatura varia de 10 °C a 31 °C e raramente é inferior a 3 °C ou superior a 35 °C.

De maneira semelhante a São Leopoldo, a base de dados da cidade de Santa Maria é constituída por dados meteorológicos da estação climática da cidade extraídos hora a hora de todos os dias de 2019 e de valores de potência de um transformador de força da subestação presente na região.

O processamento dos dados para a cidade de Santa Maria é aplicado a base de dados contendo 8747 linhas para todas as variáveis inclusas nos modelos: temperatura, pressão, umidade, velocidade dos ventos, dia da semana, hora e potência da hora anterior. Os conjuntos de dados de treino e teste são divididos em 6997 e 1750, respectivamente. É necessário salientar que os número de dados totais para São Leopoldo e para Santa Maria são diferentes devido ao processo de limpeza nos dados, explicados na seção 3.4 da Metodologia, uma vez que falhas no processamento dos dados ou nos equipamentos podem ocorrer de maneira independente em cada cidade. Isso resulta na exclusão de linhas pelo algoritmo de maneira a impedir que condições extremas não realistas sejam inclusas no modelo de previsão.

Os resultados dos parâmetros estabelecidos são então extraídos dos quatro modelos analisados para modelos com variáveis climáticas, Tabela 8, e sem as mesmas, Tabela 9. E, assim como em São Leopoldo, é retirado um dia aleatório para verificar a diferença entre as curvas de carga sem e com fatores climáticos para os modelos em questão, Figuras 33, 34 e 35.

Por fim, é realizada a análise dos dias da semana, final de semana e feriado para expor o comportamento da curva de carga nesses casos. Vale ressaltar que essas análises são realizadas com as informações climáticas presentes nos modelos.

A Tabela 5 apresenta os dias aleatórios escolhidos para construção das curvas de carga.

Tabela 5 – Dias de análise das curvas de carga

	Dia Analisado
Segunda-Feira	09/12/2019
Quarta-Feira	04/12/2019
Sexta-Feira	15/11/2019
Sábado	02/11/2019
Domingo	26/10/2019
Natal	25/12/2019
Finados	02/11/2019

Fonte – Elaborado pela autora

Os resultados encontrados para Santa Maria são exibidos e explicados na seção 5.2 do Capítulo de Resultados e Discussões.

5 Resultados e Discussões

Este capítulo apresenta os resultados obtidos com a aplicação dos algoritmos inteligentes para previsão da demanda dos transformadores de força, localizados na subestação da cidade de São Leopoldo e Santa Maria. Os parâmetros de comparação, descritos anteriormente, R quadrado, Erro Absoluto Médio e Erro Médio Quadrado são então utilizados para verificar a aderência do modelo a base de dados criada.

Em seguida, analisa-se o comportamento dos modelos sem variáveis climáticas, com a intenção de observar a sensibilidade dos transformadores para fatores climáticos importantes, como a temperatura. E por fim, é observado o comportamento dos modelos utilizados no presente trabalho, Árvore de Decisão (AD), Florestas de Decisão Aleatória (FDA) e Memória de Longo e Curto Prazo (MLCP) para dias da semana, finais de semana e feriados. Esse estudo permite concluir se os modelos são adequados tanto para cargas mais comportadas visualizadas durante a semana quanto para eventos incomuns onde a carga não se comporta como esperado.

5.1 Resultados encontrados para São Leopoldo

Utilizando as métricas descritas na seção 3.7, avaliou-se a aderência dos modelos de Aprendizado de Máquina e de Redes Neurais de previsão de carga para o transformador instalado na subestação de São Leopoldo. Os seguintes resultados são então extraídos:

Tabela 6 – Resultados das métrica de avaliação dos modelos para SE São Leopoldo

Modelos	R2	MSE	MAE
Regressão de Vetores de Suporte	0,1083	49,207	5,640
Árvore de Decisão	0,9258	4,091	1,312
Florestas de Decisão Aleatória	0,9627	2,058	1,016
Memória de Longo e Curto Prazo	0,9350	3,580	1,450

Fonte – Elaborado pela autora

O método RVS apresenta os piores resultados, atingindo apenas 10,83% de precisão e erros significativos quando comparados aos outros modelos. A explicação desses resultados está na metodologia que a regressão de Vetor de Suporte utiliza, na qual a variável de saída é classificada em apenas dois grupos, semelhante a um sistema binário. Considerando que o padrão de potência ativa real utilizado como resposta da saída do sistema é um número entre 0 e 53,45 (com duas casas decimais após a vírgula), é de se esperar um resultado pobre ao classificar a saída em apenas dois grupos. Por esse motivo, esse tipo de regressão não é indicado para previsões de carga e não será considerado para as análises a seguir.

O modelo *Árvore de Decisão* apresentou resultados bastante satisfatórios com 92,58% de exatidão, ou seja, a árvore criada é capaz de seguir um fluxo específico selecionando os ramos corretos para chegar em uma resposta de saída bastante aderente a medição real de potência ativa.

O modelo *Florestas de Decisão Aleatória*, por sua vez, apresenta melhor desempenho em relação ao modelo *Árvore de Decisão* já que é capaz de criar diversas árvores de decisão ao mesmo tempo. O crescimento de cerca de 4% no parâmetro R2 e o decréscimo nos valores de erro, se comparado aos demais modelos, garantem a *Florestas de Decisão Aleatória* a melhor aderência adquirida na previsão de carga do presente trabalho.

O modelo de rede neural recorrente apresentou 93,5% de precisão, o que deve-se em parte a falta de habilidade do método em considerar eventos futuros nas previsões do momento atual.

5.1.1 Análise da relevância das variáveis climáticas no modelo

A Tabela 7 apresenta os resultados do comportamento dos modelos ao retirar as variáveis climáticas dos mesmos.

Tabela 7 – Resultados dos modelos sem fatores climáticos

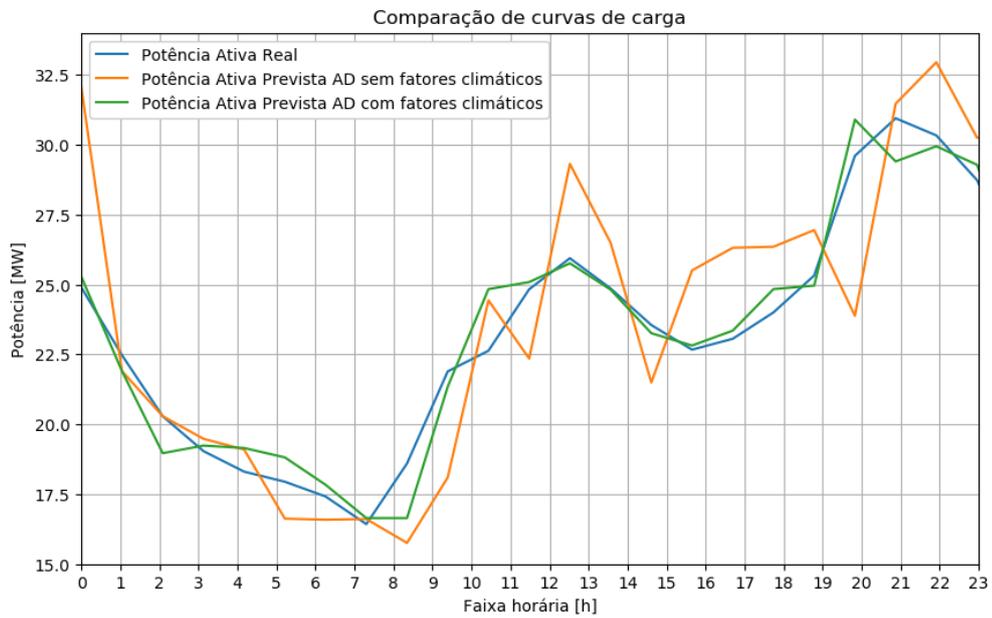
Modelo	R2	MSE	MAE
Árvore de Decisão	0,8027	10,885	2,525
Florestas de Decisão Aleatória	0,8041	10,806	2,564
Memória de Longo e Curto Prazo	0,9017	5,415	1,859

Fonte – Elaborado pela autora

A partir da Tabela 7 percebe-se que para todos os modelos há redução no parâmetro R2 e crescimento de quase 100% nos erros quadrático médio e absoluto médio.

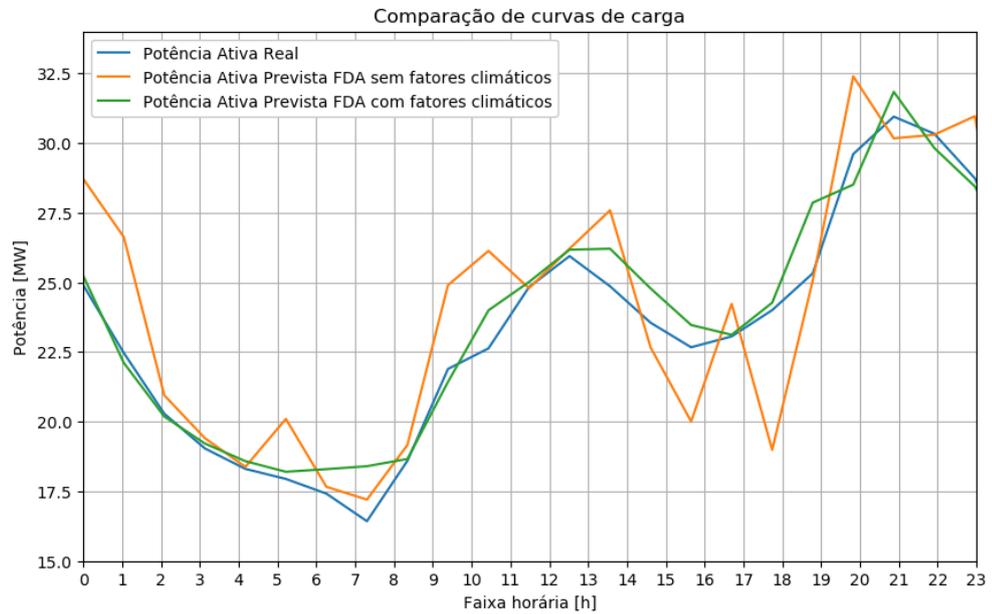
Com o objetivo de melhorar a visualização desta diferença, escolhe-se um dia aleatório para construção da curva de carga com e sem os fatores climáticos. As Figuras 9, 10 e 11 apresentam os resultados obtidos pelos modelos *Árvore de Decisão*, *Florestas de Decisão Aleatória* e *Memória de Longo e Curto Prazo*, respectivamente.

Figura 9 – Comparação curva de carga com e sem fatores climáticos para a Árvore de Decisão



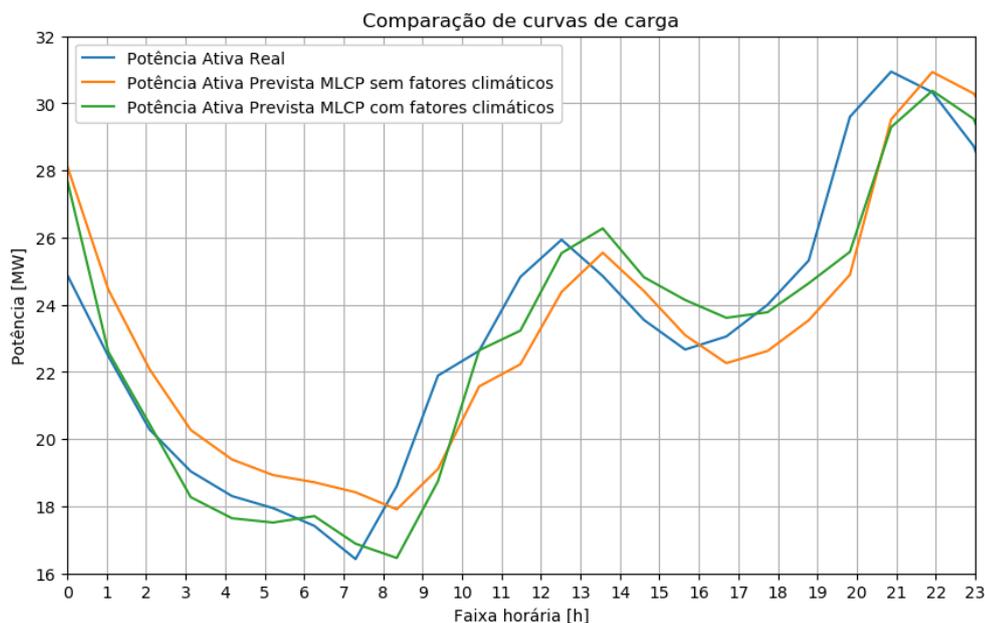
Fonte – Elaborado pela autora

Figura 10 – Comparação curva de carga com e sem fatores climáticos para a Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 11 – Comparação curva de carga com e sem fatores climáticos para a Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

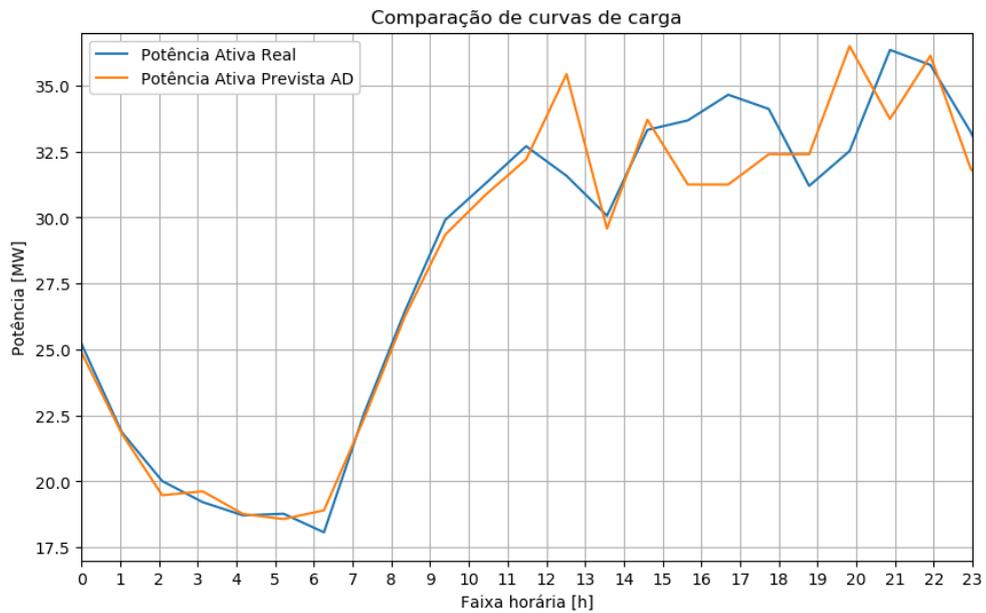
Para os métodos de Árvore de Decisão e Florestas de Decisão Aleatória fica claro que as variáveis climáticas são fatores de grande importância na previsão do modelo. A falta de tais fatores influenciam na capacidade dos valores previstos seguirem a curva de carga real medida no transformador, ou seja, apresentam picos e vales bastante destoantes em regiões onde há suavidades na curva. Logo, a diminuição dos fatores R quadrado e o aumento nos valores de erro são condizentes com as curvas de carga.

O modelo de rede neural, por outro lado, por tratar o problema como uma série temporal está mais preocupado em manter a periodicidade e a suavidade da curva real do que com a representatividade das variáveis nas previsões realizadas. Por esse motivo, percebe-se que a curva sem fatores climáticos é capaz de seguir a curva de carga real medida, contudo, apresenta erros maiores. Esse aumento no erro é causado justamente pela falta dessas variáveis climáticas no modelo uma vez que elas interferem comportamento das cargas conectadas ao transformador da subestações.

5.1.2 Análise dos dias da semana

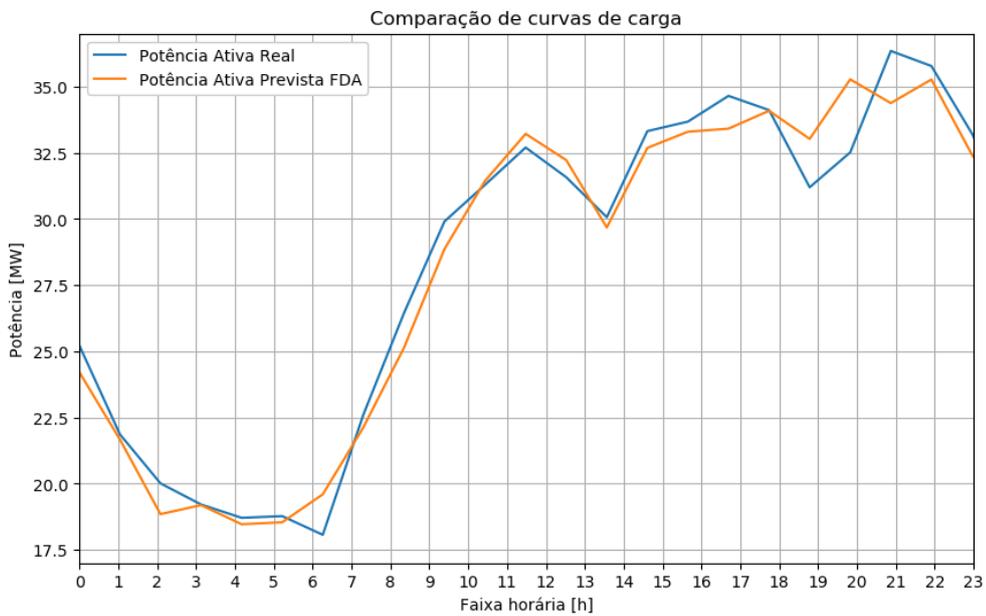
As Figuras 12,13, 14, 15,16, 17, 18,19 e 20 apresentam as curvas de carga dos modelos Árvore de Decisão, Florestas de Decisão Aleatória e Memória de Longo e Curto Prazo para a segunda-feira, quarta-feira e sexta-feira, respectivamente. Destaca-se aqui que as estimativas apresentadas levam em consideração as variáveis climáticas inseridas no modelo.

Figura 12 – Segunda-feira: Modelo Árvore de Decisão



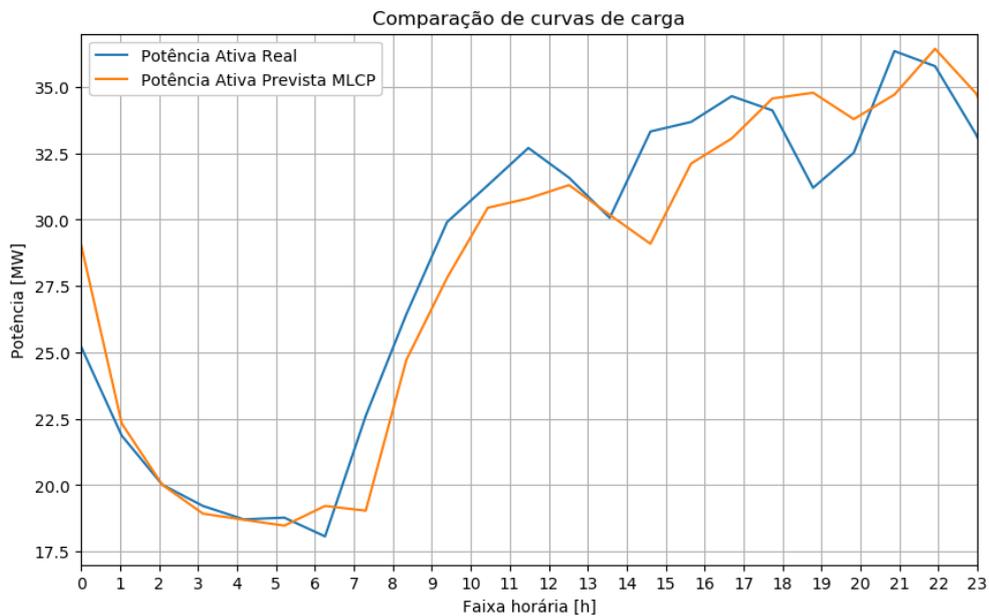
Fonte – Elaborado pela autora

Figura 13 – Segunda-feira: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 14 – Segunda-feira: Modelo Memória de Longo e Curto Prazo

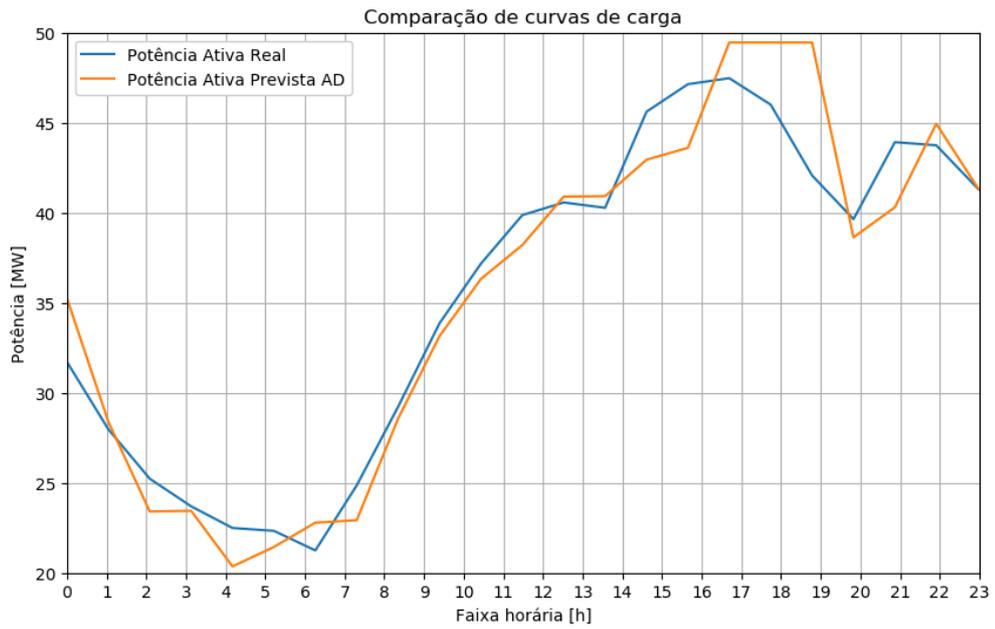


Fonte – Elaborado pela autora

Ao analisar os gráficos das Figuras 12,13 e 14, percebe-se que a curva que melhor acompanha a carga real medida no transformador é aquela produzida pelo modelo Florestas de Decisão Aleatória, resultado coerente se comparado a Tabela 6 na qual o modelo apresenta melhor aderência e menores valores de erros.

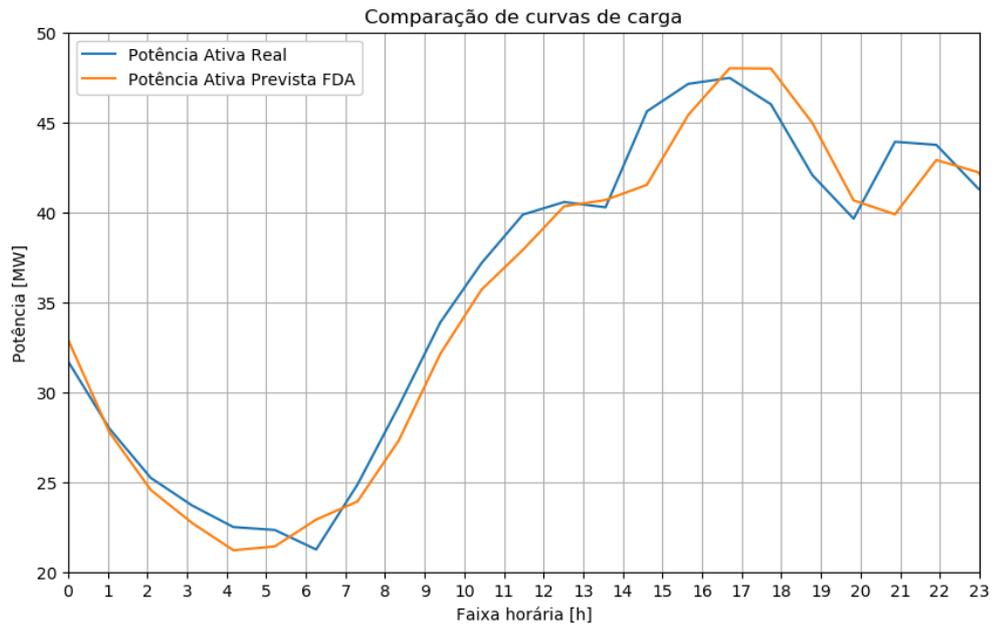
A curva FDA apresenta resultados bastante semelhantes ao real entre a 00h e as 17h, após as 17 horas o modelo torna-se mais inconsistente devido, em partes, a maior variação da carga. De modo semelhante, os modelos de AD e MLCP apresentam melhores respostas durante a manhã, onde há um crescimento mais suave da carga, e resultados menos satisfatórios durante a tarde e a noite.

Figura 15 – Quarta-feira: Modelo Árvore de Decisão



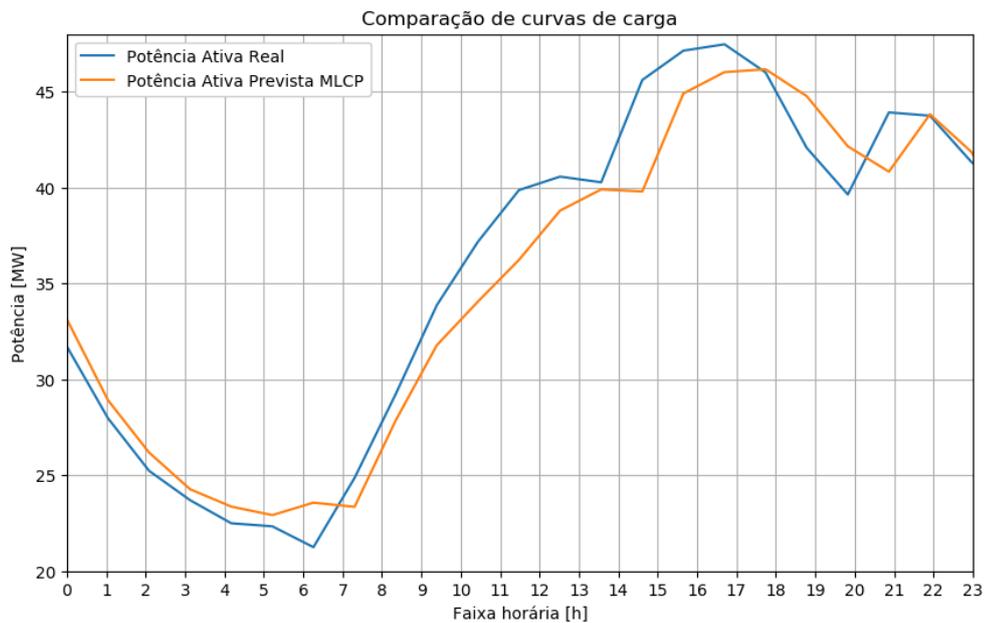
Fonte – Elaborado pela autora

Figura 16 – Quarta-feira: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 17 – Quarta-feira: Modelo Memória de Longo e Curto Prazo

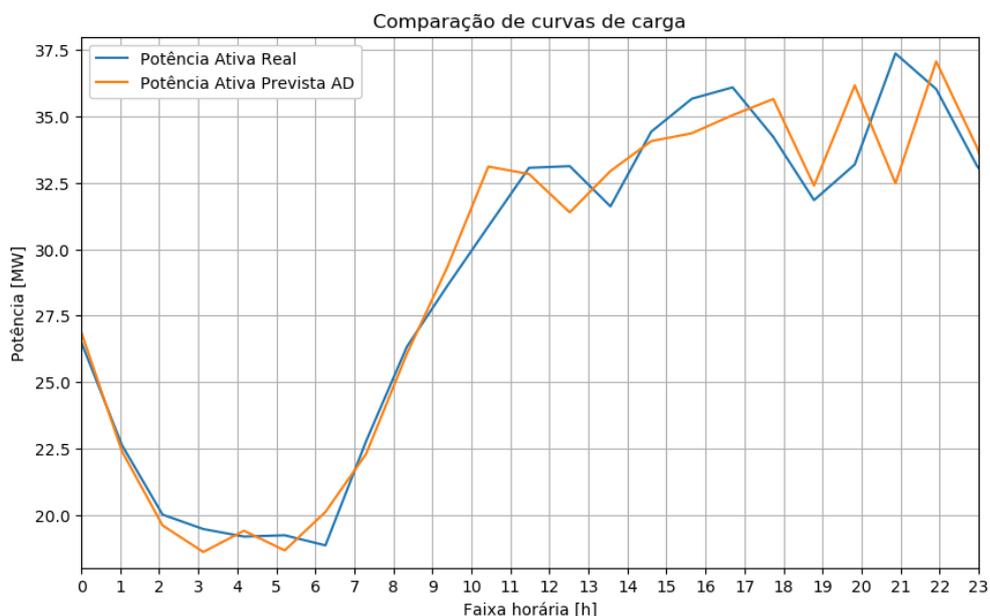


Fonte – Elaborado pela autora

A quarta-feira analisada possui comportamento semelhante a segunda-feira com decaída nas primeiras horas da manhã, aumento constante até o início da tarde e picos e vales de potência durante a noite.

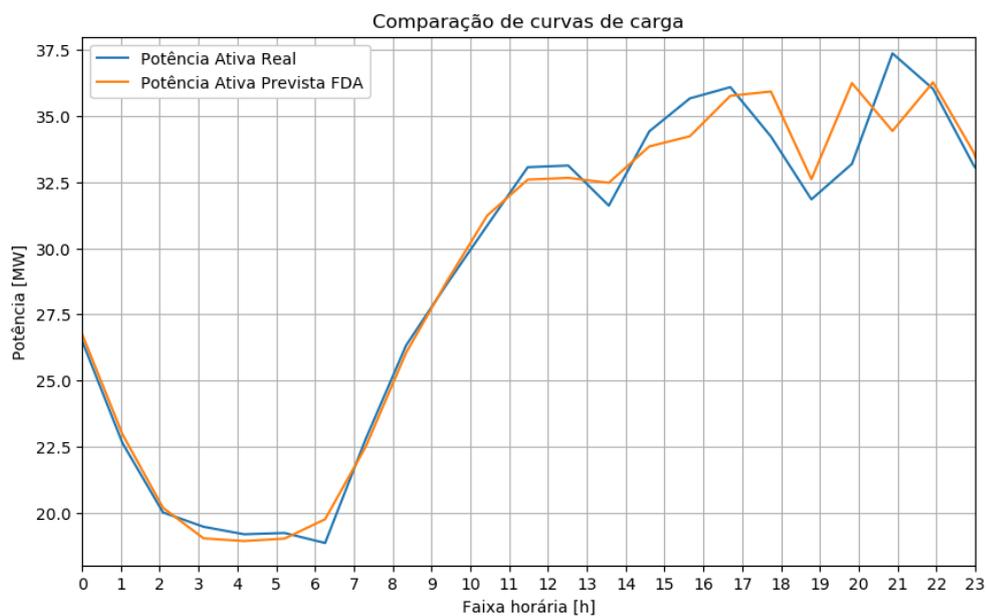
O modelo Árvore de Decisão apresentou os maiores erros durante a tarde/noite, momento em que há maior variação de carga, uma vez que não consegue acompanhar pequenas alterações nos valores de potência ativa real. O modelo FDA, por sua vez, possui boa aderência ao longo do dia assim como o modelo MLCP. Esse último, apresenta atraso esperado de uma hora provinda da variável de potência da hora anterior na qual garante boa aderência porém erros mais expressivos do que o modelo FDA.

Figura 18 – Sexta-feira: Modelo Árvore de Decisão



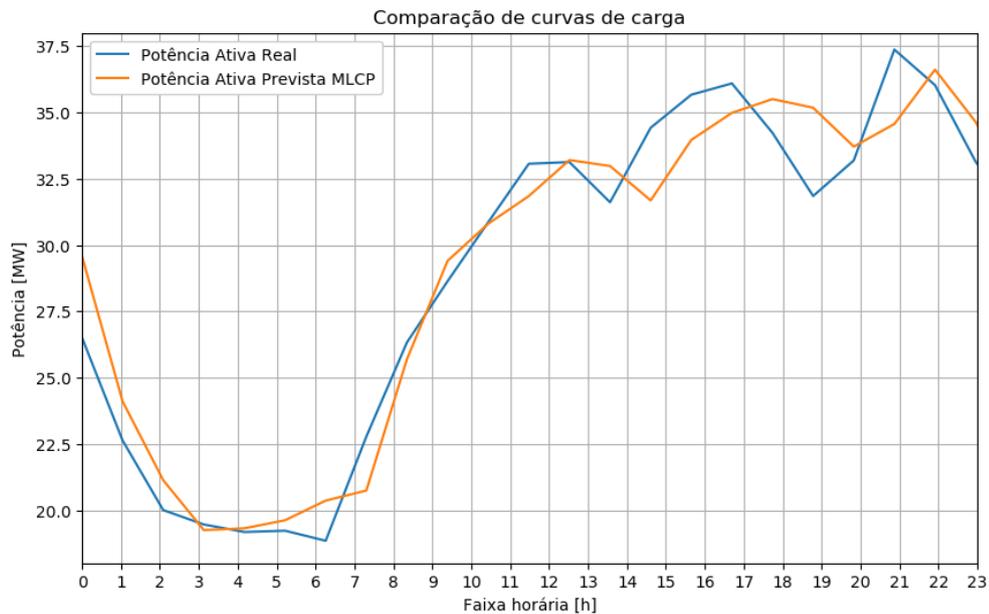
Fonte – Elaborado pela autora

Figura 19 – Sexta-feira: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 20 – Sexta-feira: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

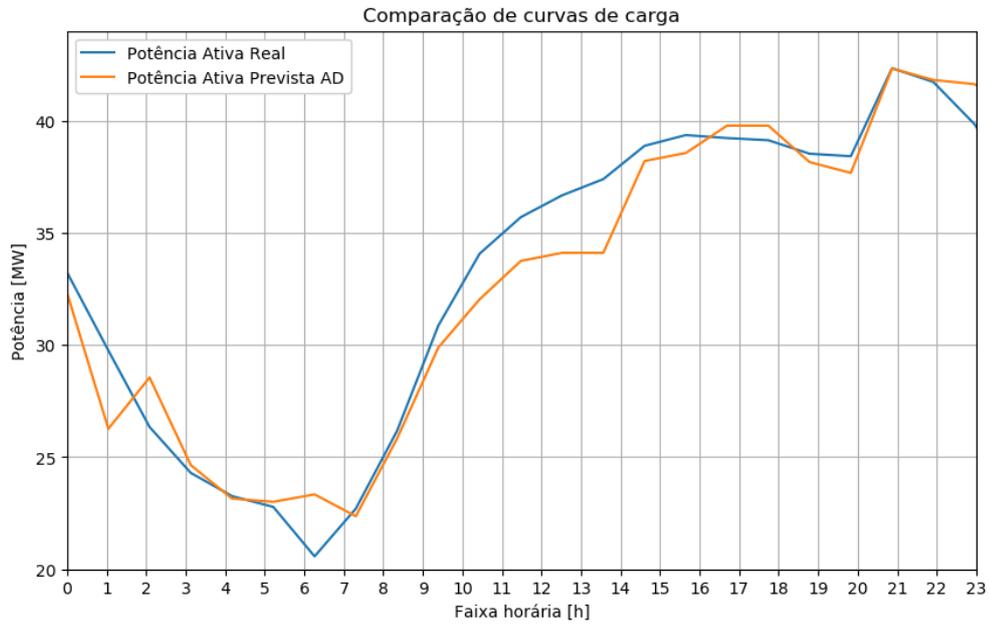
A sexta-feira 18,19 e 20, com comportamento análogo à segunda e quarta-feira apresenta erros relativos maiores para a regressão Árvore de Decisão, atraso de uma hora para a MLCP e melhores resultados para o modelo Florestas de Decisão Aleatória.

Outros dias da semana foram também testados e apresentaram o mesmo comportamento de carga: leve decréscimo durante a madrugada seguido de aumento significativo na potência ativa até o início da tarde, a tarde e a noite os valores de potência são bastante diversos, motivo pelo qual os modelos possuem maiores erros se comparado aos valores reais.

5.1.3 Análise do final de semana

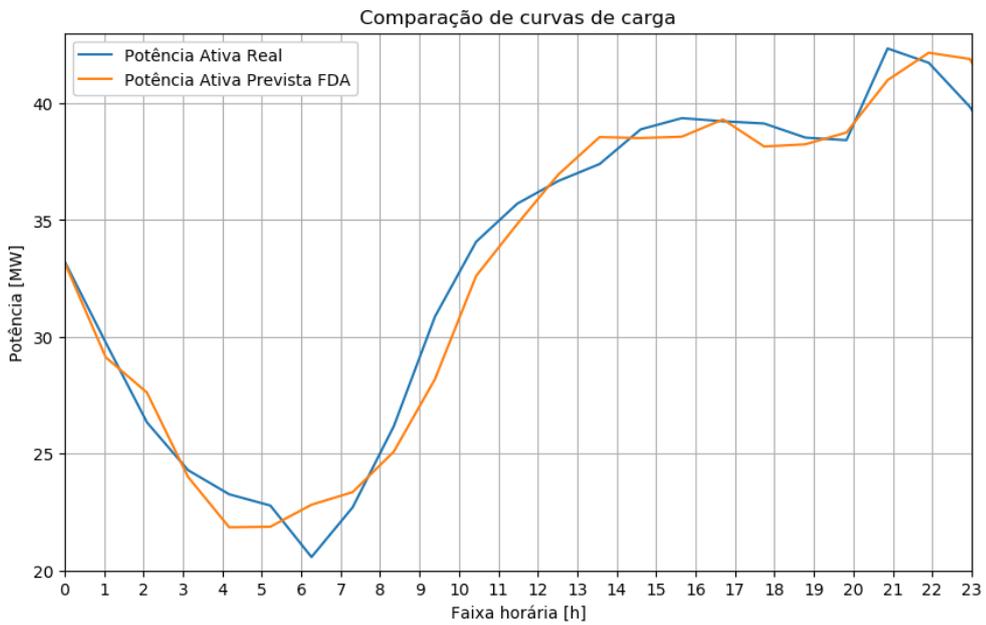
As Figuras 21, 22, 23, 24, 25 e 26 apresentam os resultados dos modelos analisados para sábado e domingo, respectivamente.

Figura 21 – Sábado: Modelo Árvore de Decisão



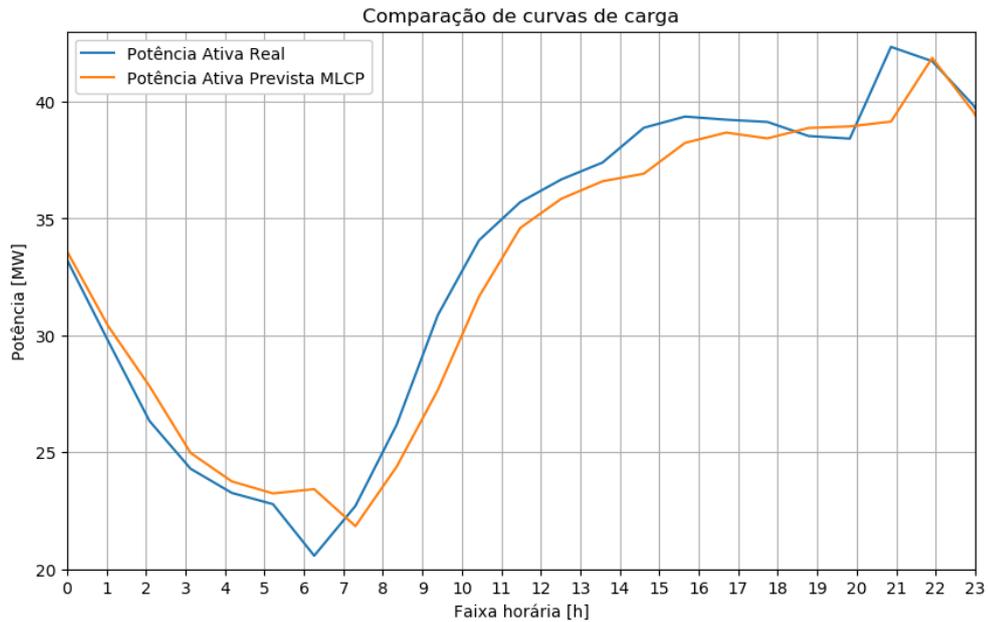
Fonte – Elaborado pela autora

Figura 22 – Sábado: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

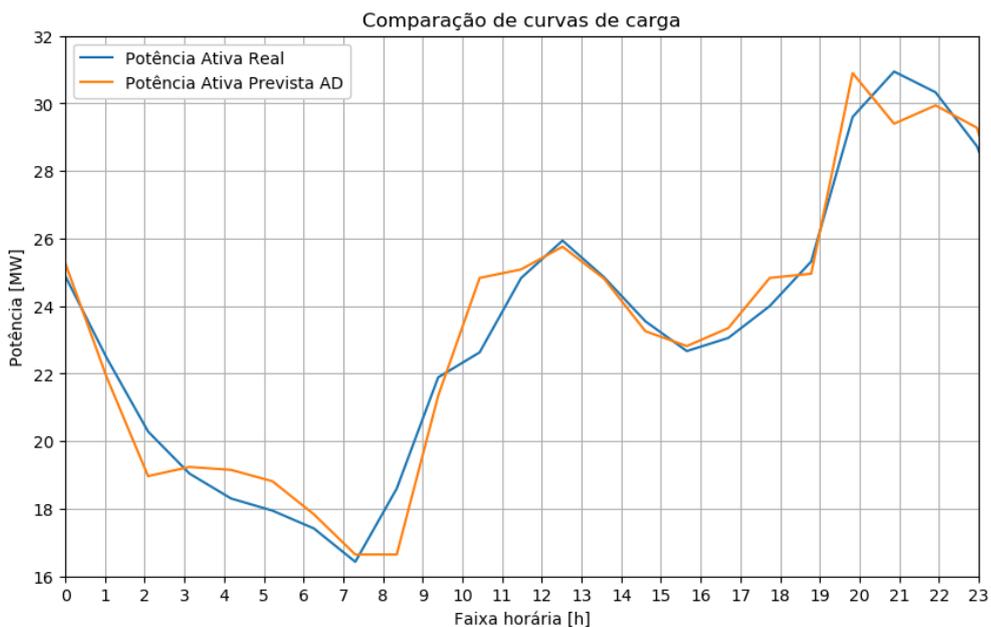
Figura 23 – Sábado: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

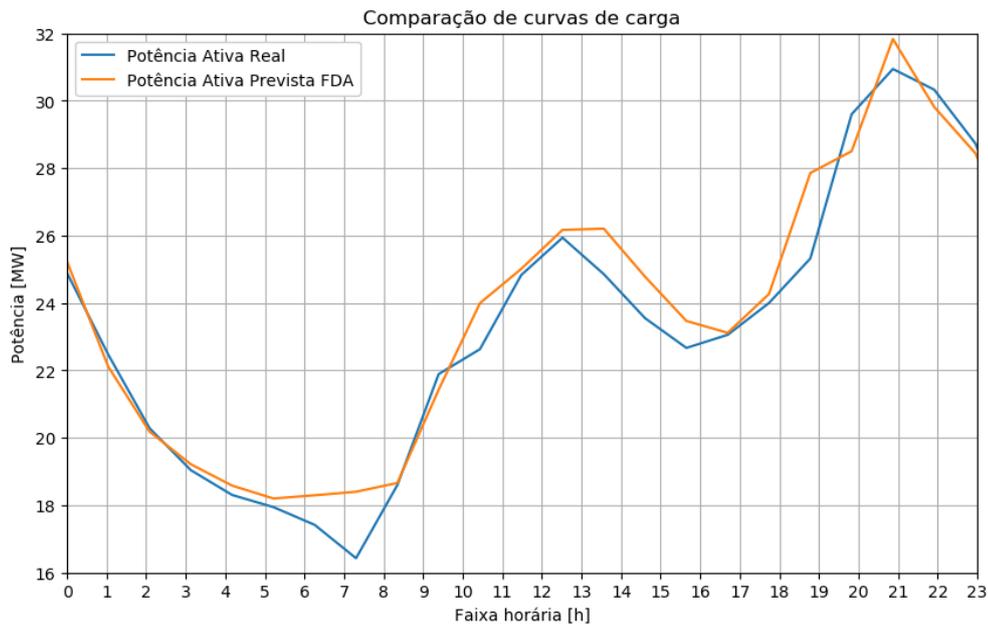
Ao analisar os gráficos produzidos em um sábado é possível perceber que os três modelos apresentam previsões semelhantes aos valores reais de potência ativa e ao comportamento dos dias durante a semana. A carga apresenta formato de U durante a manhã até início da tarde, segue em crescimento até a noite, período onde a potência apresenta pequenas flutuações.

Figura 24 – Domingo: Modelo Árvore de Decisão



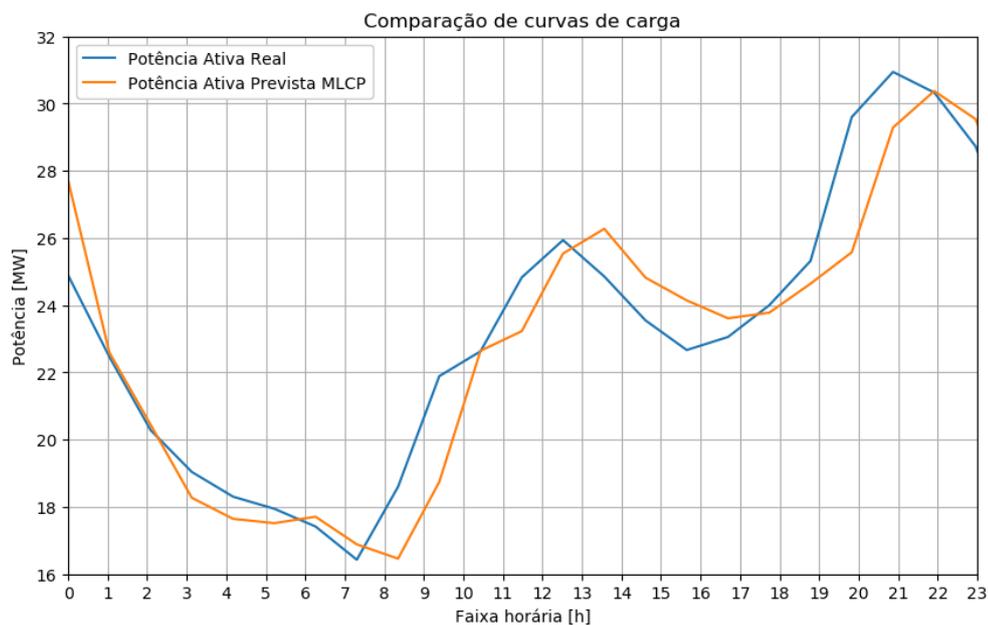
Fonte – Elaborado pela autora

Figura 25 – Domingo: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 26 – Domingo: Modelo Memória de Longo e Curto Prazo



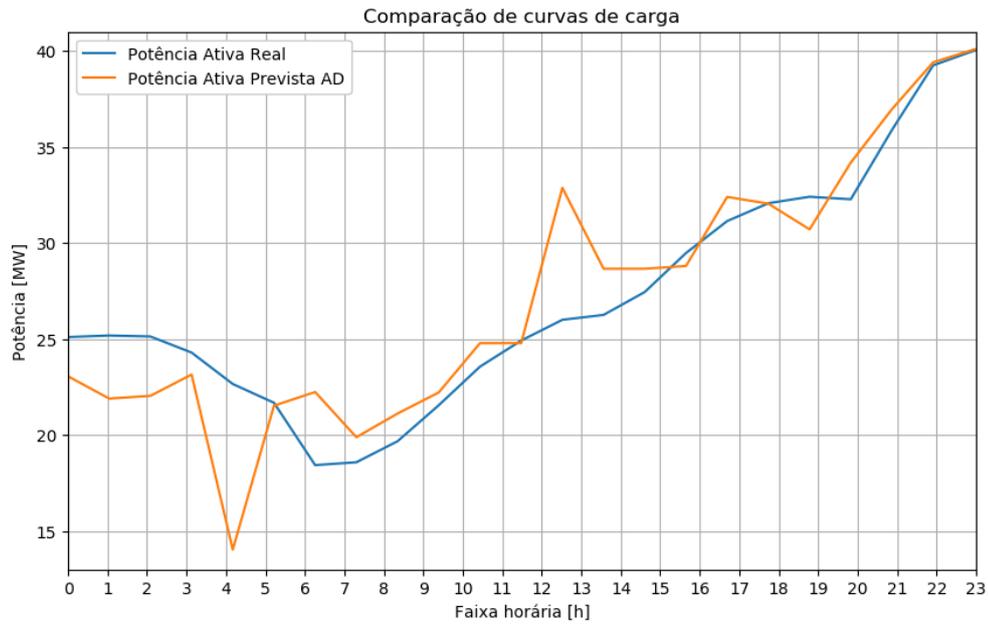
Fonte – Elaborado pela autora

O domingo, diferentemente do sábado, apresenta curva de carga com mudanças mais bruscas, apresentando um vale, durante a manhã, e dois picos, um no início da tarde e outro durante a noite. Esse comportamento é sentido pelos três modelos em questão com algumas diferenças: o modelo AD apresenta menor capacidade de seguir a suavidade da curva, o modelo MLCP apresenta um atraso de cerca de uma hora na maioria das previsões e o modelo FDA, aquele com maior aderência, em geral consegue acompanhar os valores reais de potência ativa.

5.1.4 Análise de feriados

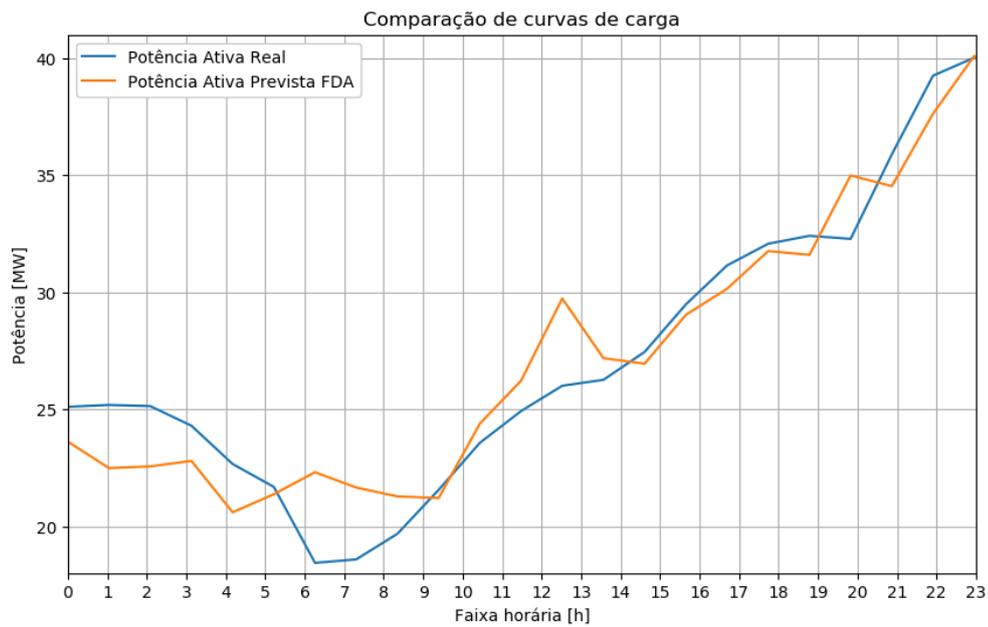
As Figuras 27, 28, 29, 30, 31 e 32 apresentam os resultados encontrados pelos modelos no Natal e no feriado de Finados, respectivamente.

Figura 27 – Natal: Modelo Árvore de Decisão



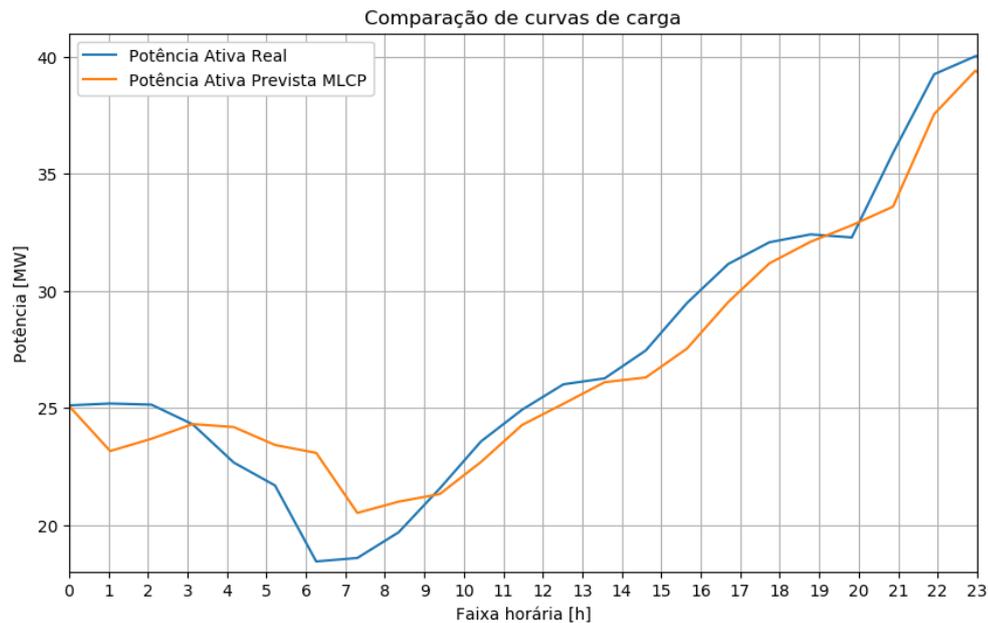
Fonte – Elaborado pela autora

Figura 28 – Natal: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 29 – Natal: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

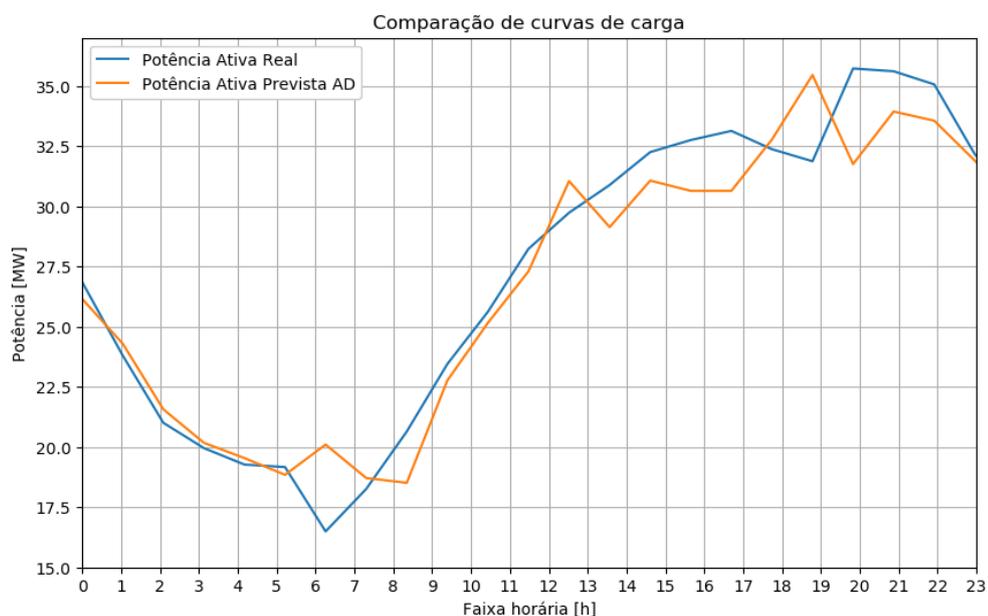
A análise dos gráficos permite observar que a curva de carga durante o Natal é bastante distinta de qualquer outro dia da semana ou final de semana. A carga, para o ano de 2019, apresenta valores intermediários durante a madrugada seguida de leve queda nas primeiras horas da manhã e aumento constante durante a tarde e a noite.

O modelo de previsão AD possui bastante dificuldade de acompanhar a curva real de potência ativa, chegando a apresentar picos e vales 'aleatórios' de carga durante o dia. O modelo é capaz apenas de acompanhar a tendência de aumento ou queda na potência ativa diária.

O modelo FDA, assim como o AD, apresenta certa dificuldade de acompanhamento, entretanto, apresenta erros menores e picos menos expressivos durante o dia.

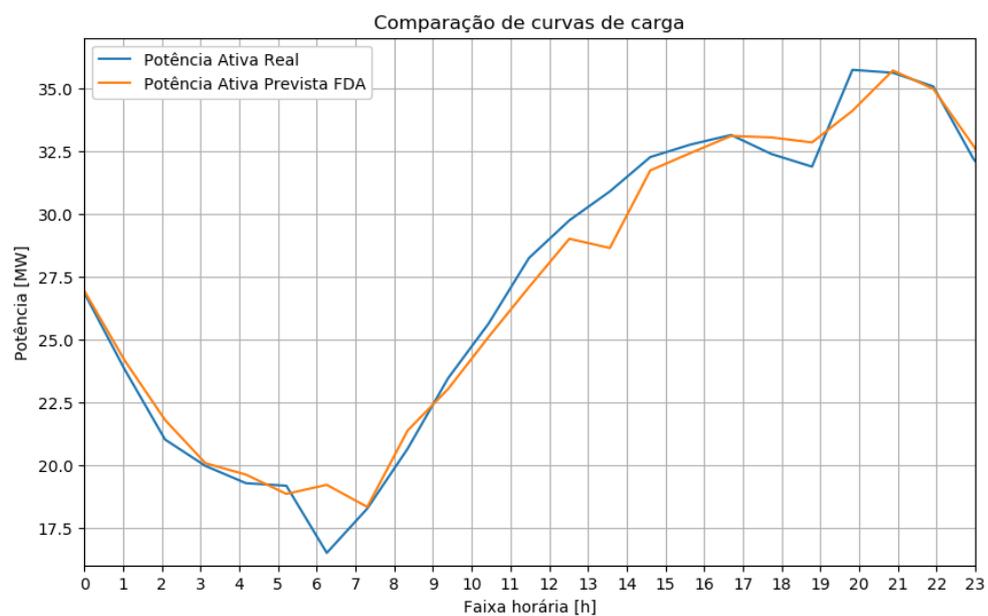
Por sua vez, o modelo MLCP apresenta melhores previsões, uma vez que está mais preocupado com a série temporal potência ativa da hora anterior do que com comportamentos em semelhantes condições em dias passados ou futuros utilizados pelos métodos de Árvore de Decisão e Florestas de Decisão Aleatória.

Figura 30 – Finados: Modelo Árvore de Decisão



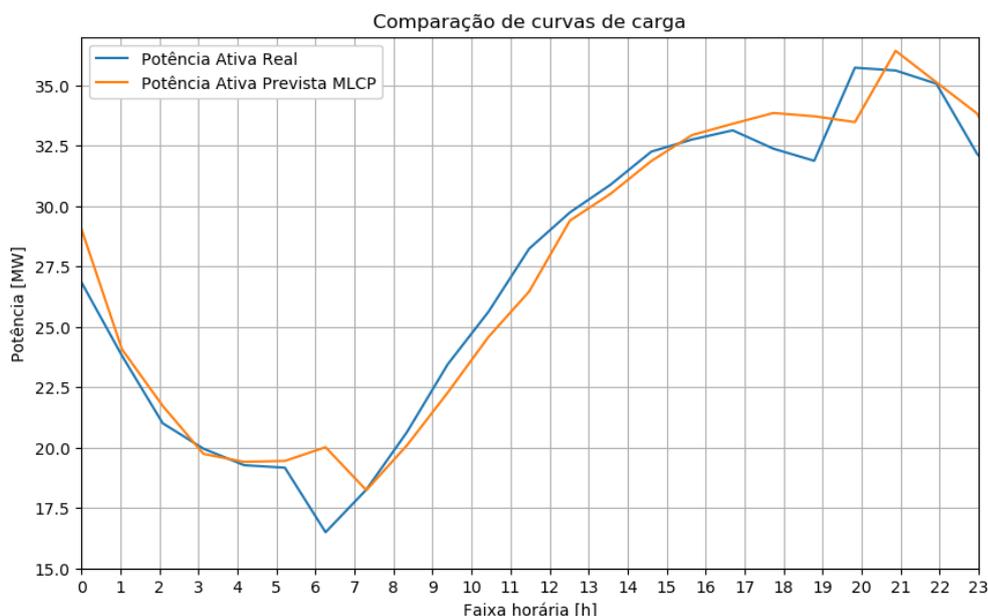
Fonte – Elaborado pela autora

Figura 31 – Finados: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 32 – Finados: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

Analisando as Figuras 30, 31 e 32 conclui-se que o feriado de Finados apresenta comportamento semelhante ao apresentado pelos dias da semana. O modelo Árvore de Decisão, como de hábito, apresenta o pior desempenho, com previsões bastante 'aleatórias' durante a tarde/noite enquanto os modelos de Florestas de Decisão Aleatória e Memória de Longo e Curto Prazo são capazes de seguir a tendência e as nuances da curva de carga real do transformador.

5.2 Resultados encontrados para Santa Maria

As métricas são novamente utilizadas para medir a aderência dos modelos de Aprendizado de Máquina e de Redes Neurais na previsão de carga do transformador da subestação de Santa Maria. Os seguintes resultados são então extraídos:

Tabela 8 – Resultados das métrica de avaliação dos modelos para SE Santa Maria

Modelo	R2	MSE	MAE
Regressão de Vetores de Suporte	-0,0608	33,514	4,357
Árvore de Decisão	0,8509	4,708	1,421
Florestas de Decisão Aleatória	0,9394	1,912	0,987
Memória de Longo e Curto Prazo	0,9586	1,3055	0,824

Fonte – Elaborado pela autora

O resultado de aderência para o modelo RVS apresentou valores negativos, ou seja, o modelo não é apto a prever a carga do transformador com as variáveis escolhidas. Por esse motivo, o modelo será excluído das análises realizadas a seguir.

Assim como no caso de São Leopoldo, pode-se perceber que o modelo Árvore de Decisão apresenta resultados bastante satisfatório com relação ao parâmetro R quadrado, atingindo 85% de aderência, e valores de erros considerados baixos.

O modelo Florestas de Decisão Aleatória, aprimorado do modelo de Árvore de Decisão, apresentou melhora significativa nos valores de R quadrado, com 93% de aderência, e valores ainda menores nos erros, principalmente no erro quadrático médio.

Diferentemente do caso de São Leopoldo, o modelo MLCP apresentou os melhores resultados de aderência, 95%, e os menores valores de erro quadrático e absoluto médios. Essa diferença entre as cidades de São Leopoldo e Santa Maria é criada principalmente pela diferença nos dados climáticos e de potência de cada região, na qual cada variável possui diferentes relevâncias para o modelo em questão. Em geral, por exemplo, Santa Maria possui maiores variações de temperatura atmosférica em relação a variação da carga se comparado a São Leopoldo, o que é melhor sentido pelo modelo MLCP já que insere maior grau de relevância ao acontecimento anterior do que a essa relação entre carga e temperatura.

5.2.1 Análise da relevância das variáveis climáticas no modelo

Os resultados encontrados para os três modelos Árvore de Decisão, Florestas de Decisão Aleatória e Memória de Longo e Curto Prazo sem a presença de variáveis climáticas são demonstrados pela Tabela 9.

Tabela 9 – Resultados dos modelos sem fatores climáticos

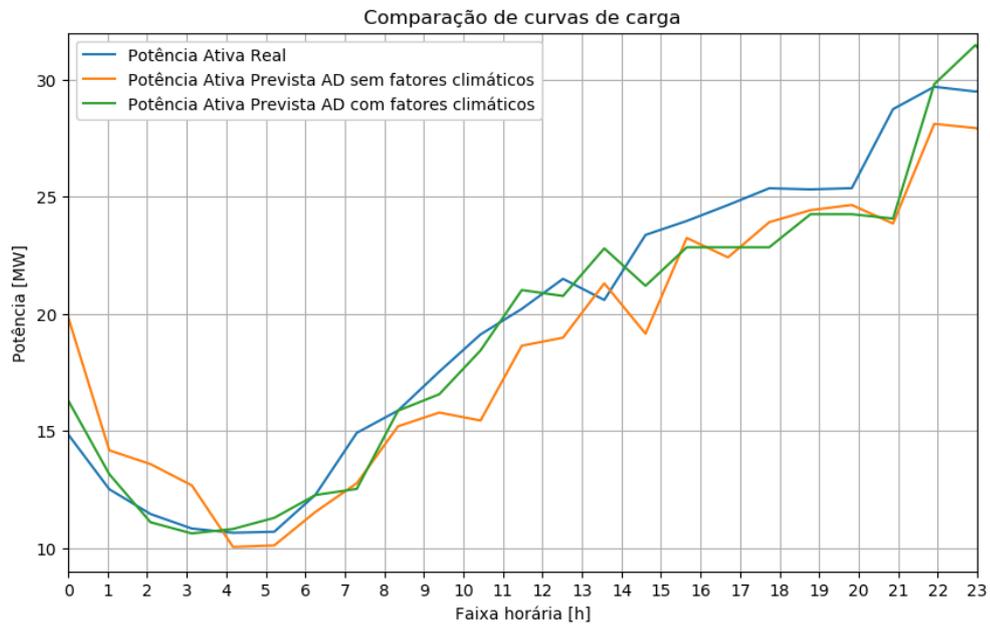
Modelo	R2	MSE	MAE
Árvore de Decisão	0,8154	5,831	1,782
Florestas de Decisão Aleatória	0,8192	5,711	1,794
Memória de Longo e Curto Prazo	0,9250	2,369	1,207

Fonte – Elaborado pela autora

Segundo a Tabela 9 os modelos apresentam redução no parâmetro R2 e crescimento significativos nos erros quadrático médio e absoluto médio.

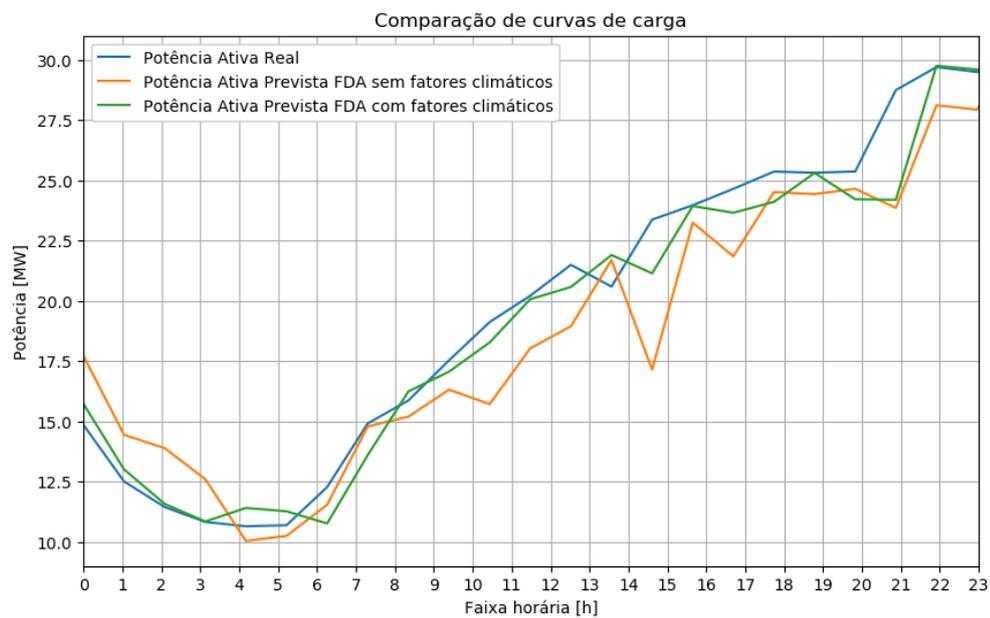
As Figuras 33, 34 e 35 apresentam os resultados da comparação entre curvas de carga com e sem variáveis climáticas pelos modelos Árvore de Decisão, Florestas de Decisão Aleatória e Memória de Longo e Curto Prazo, respectivamente.

Figura 33 – Comparação curva de carga com e sem fatores climáticos para a Árvore de Decisão



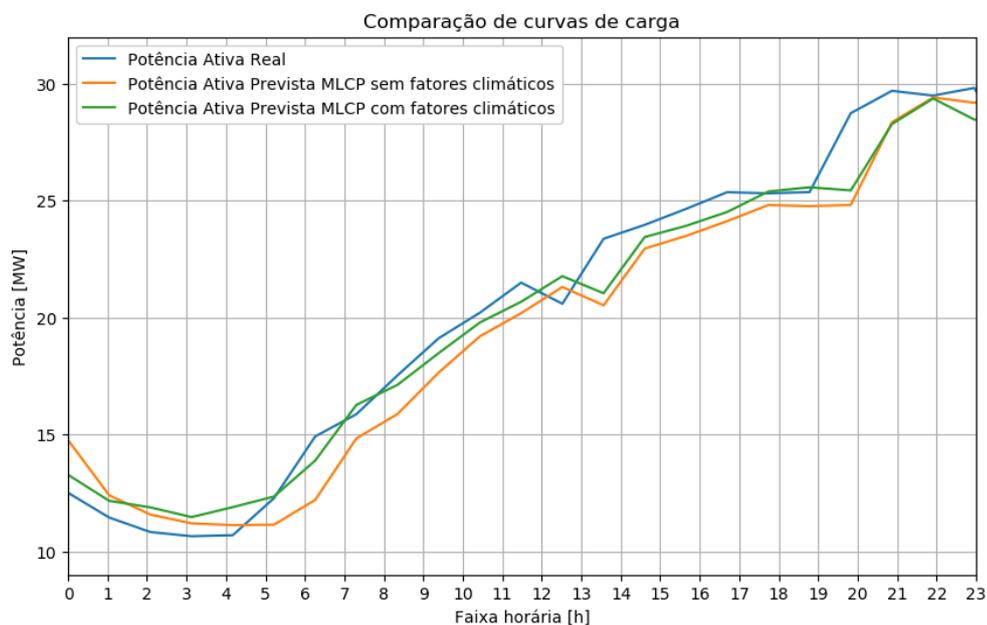
Fonte – Elaborado pela autora

Figura 34 – Comparação curva de carga com e sem fatores climáticos para a Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 35 – Comparação curva de carga com e sem fatores climáticos para a Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

Para o modelo Árvore de Decisão, as variáveis climáticas são fatores com pouca importância na previsão de carga uma vez que o parâmetro R^2 , utilizado para quantificar a aderência dos modelos, apresenta uma redução de apenas 4%, valor considerado pouco significativo. Isso é uma indicação de que o modelo não é tão capaz de construir árvores de decisões mais genéricas capazes de prever o quanto a alteração nas condições climáticas afetam o valor da potência do transformador.

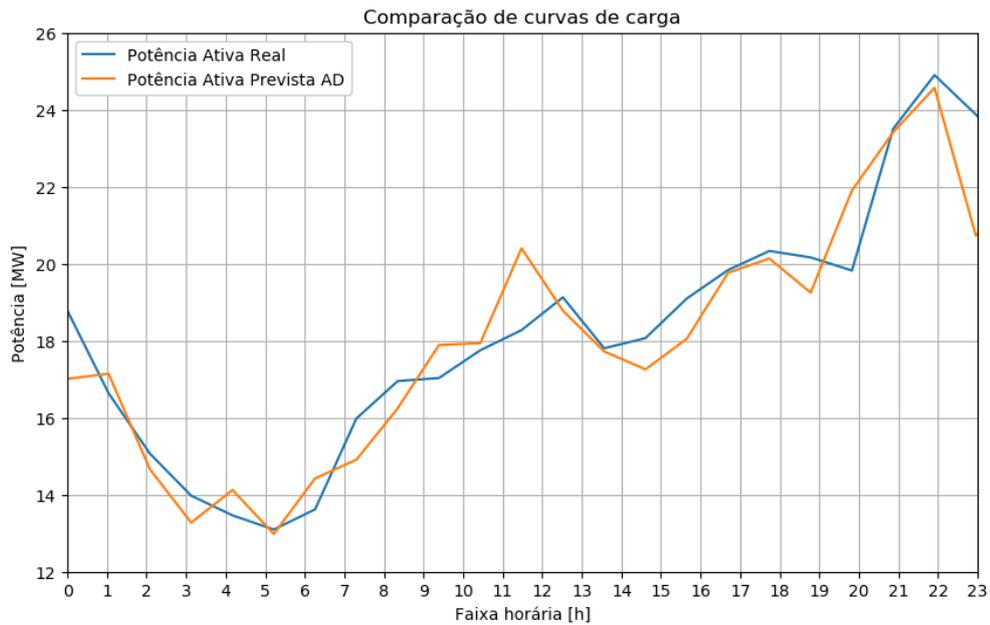
Por outro lado, o modelo Florestas de Decisão Aleatória, derivado do Árvore de Decisão com aumento de potência computacional ao criar diversas árvores de decisão simultâneas, apresenta maior dependência das variáveis climáticas na previsão da carga. A falta de tais fatores acarretam na diminuição de cerca de 12% no parâmetro de aderência e aumentos nas métricas de erro. No caso da Figura 34, há a criação de um vale entre as 14h e 15 horas com valor muito anormal de potência, o que corrobora na piora das métricas de comparação entre os modelos.

O modelo de rede neural, assim como para a cidade de São Leopoldo, trata o problema como uma série temporal, ou seja, está menos preocupado com as variáveis climáticas e mais interessado em manter a linha do tempo. Por esse motivo, a curva sem fatores climáticos é quase tão capaz de seguir a curva de carga real de potência ativa quanto as previsões mantendo as variáveis climáticas. O único porém está no aumento dos erros quadrático e absoluto médios verificados na Figura 35.

5.2.2 Análise dos dias da semana

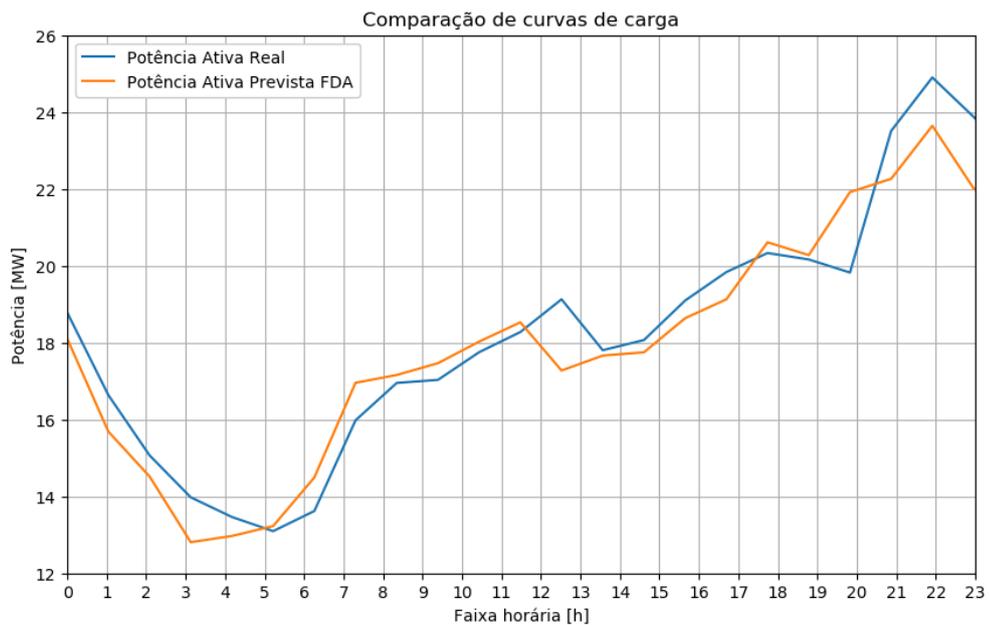
As Figuras 36, 37, 38, 39, 40, 41, 42, 43, 44 apresentam as curvas de carga da segunda, quarta e sexta-feira, respectivamente. Novamente, as curvas de carga analisadas a seguir possuem fatores climáticos em seus modelos de previsão.

Figura 36 – Segunda-feira: Modelo Árvore de Decisão



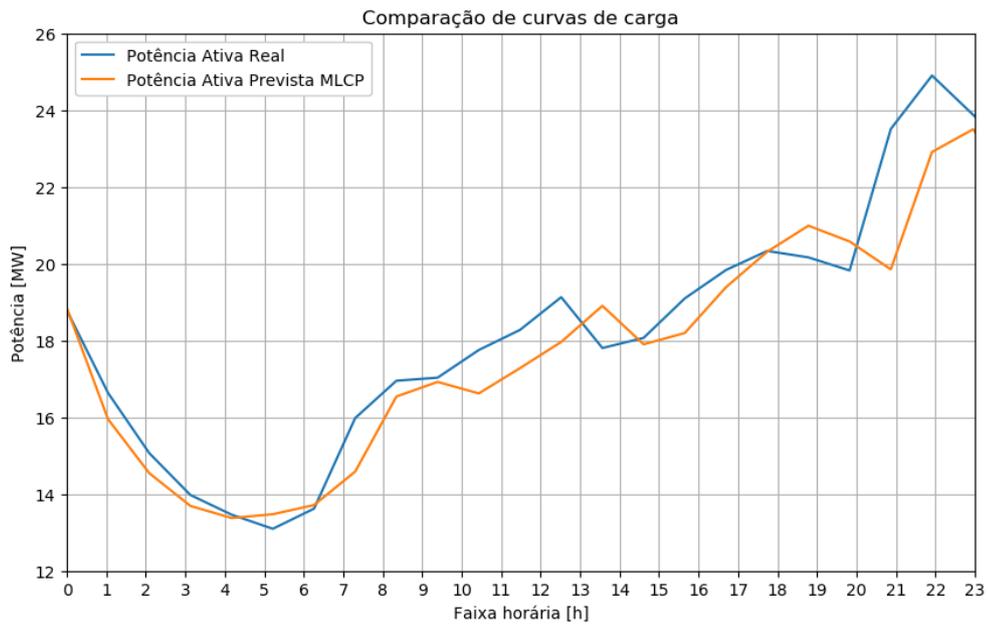
Fonte – Elaborado pela autora

Figura 37 – Segunda-feira: Modelo Florestas de Decisão Aleatória



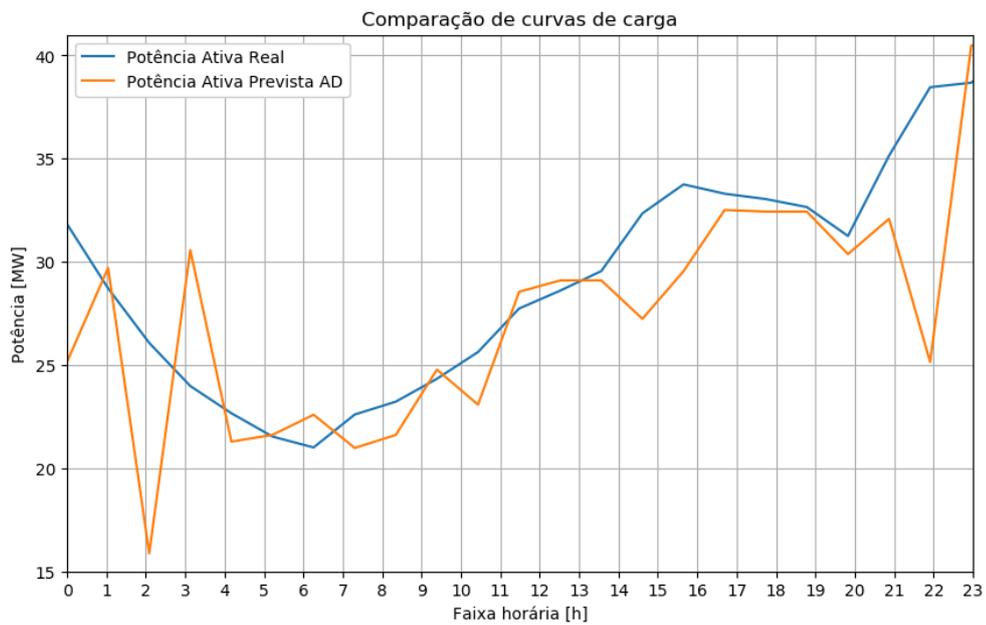
Fonte – Elaborado pela autora

Figura 38 – Segunda-feira: Modelo Memória de Longo e Curto Prazo



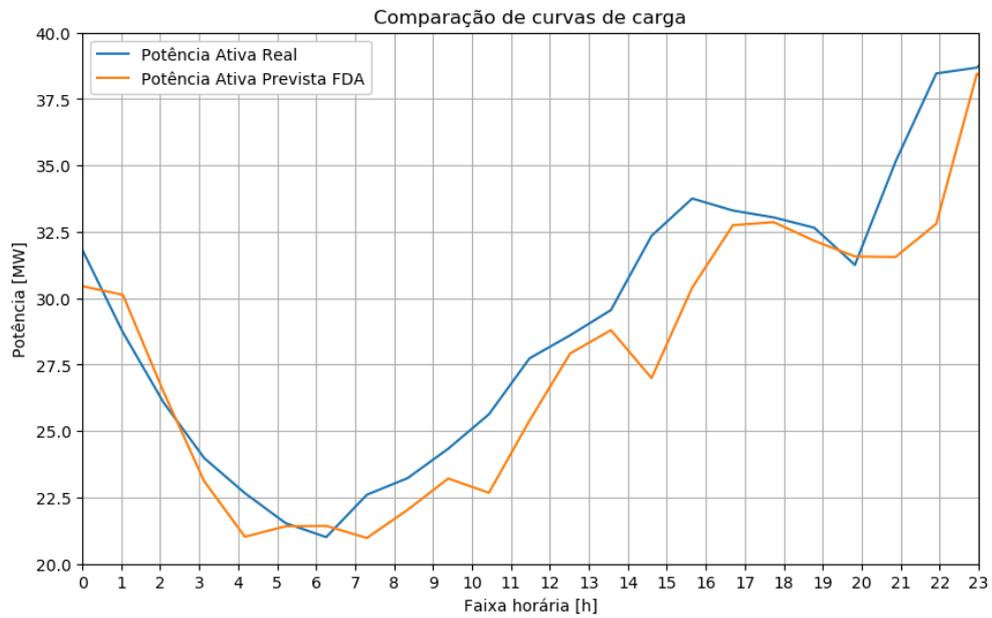
Fonte – Elaborado pela autora

Figura 39 – Quarta-feira: Modelo Árvore de Decisão



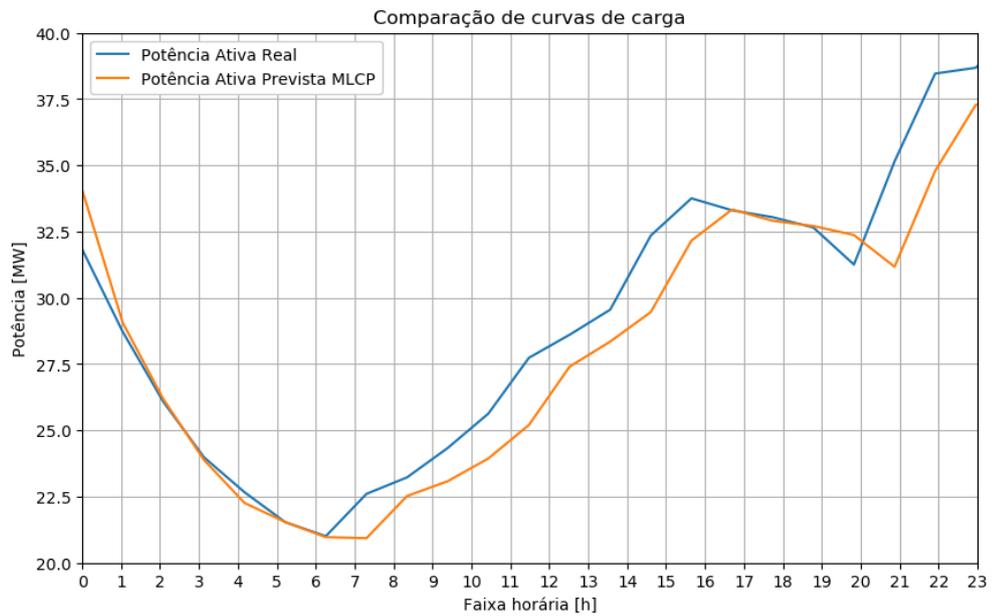
Fonte – Elaborado pela autora

Figura 40 – Quarta-feira: Modelo Florestas de Decisão Aleatória



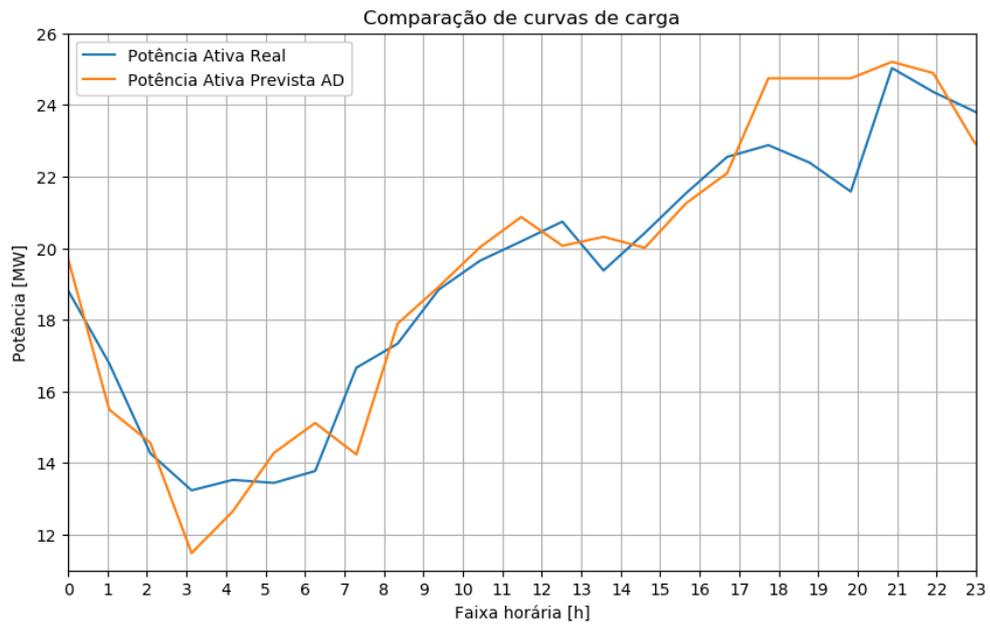
Fonte – Elaborado pela autora

Figura 41 – Quarta-feira: Modelo Memória de Longo e Curto Prazo



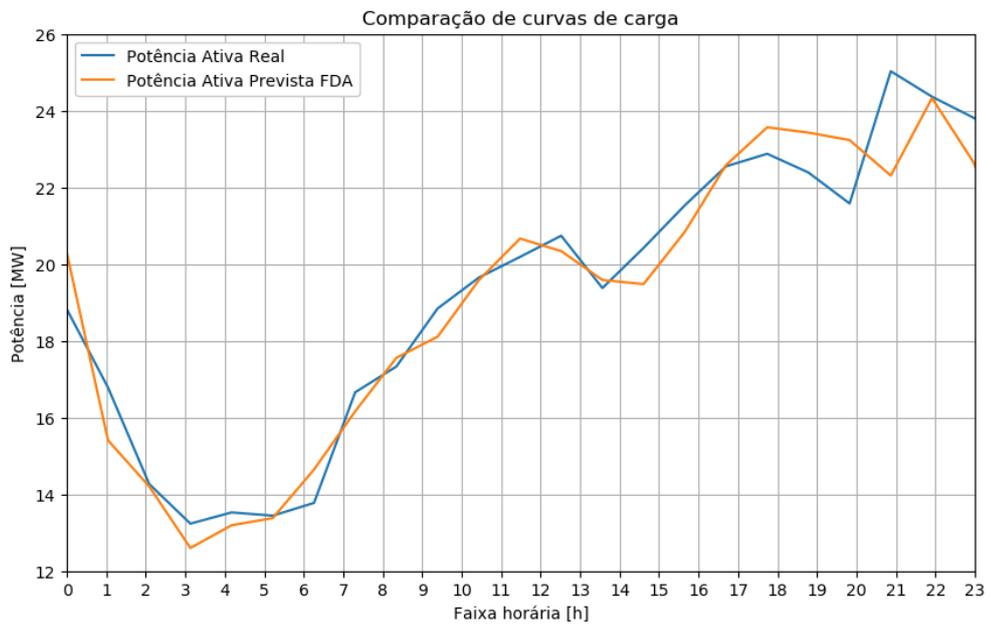
Fonte – Elaborado pela autora

Figura 42 – Sexta-feira: Modelo Árvore de Decisão



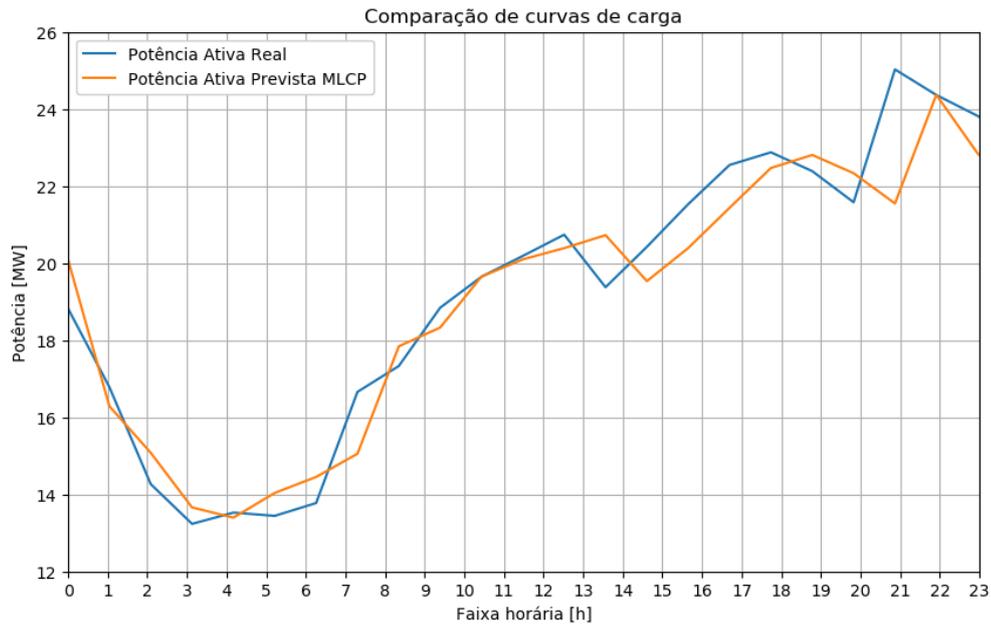
Fonte – Elaborado pela autora

Figura 43 – Sexta-feira: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 44 – Sexta-feira: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

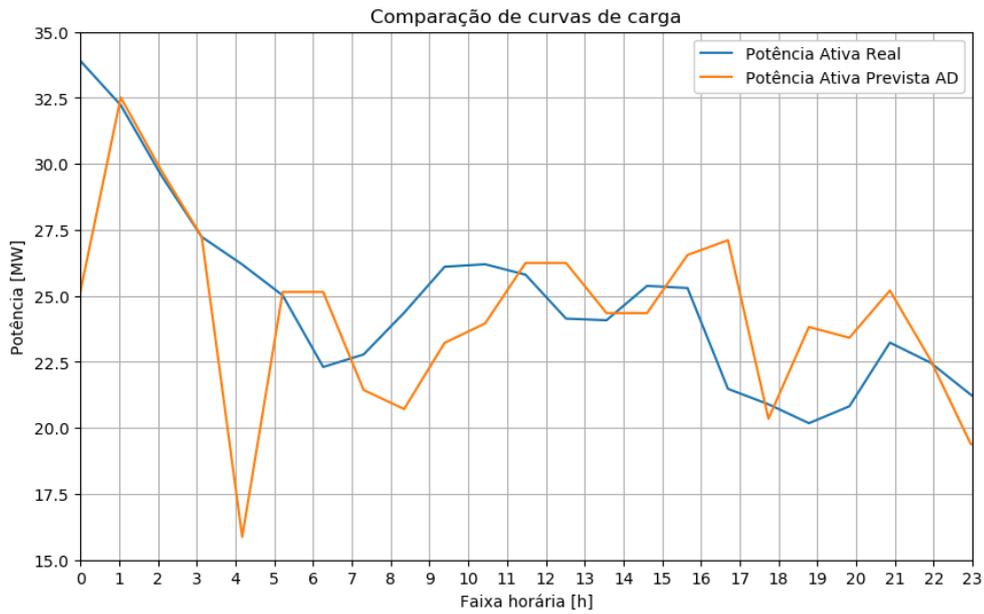
Segundo as Figuras 36, 37, 38, 39, 40, 41, 42, 43, 44, pode-se perceber que os dias da semana para a cidade apresentam curvas de carga bastante similares, com queda durante as primeiras horas da manhã e ascensão na potência até o fim da noite. Esse aumento no consumo lentamente e chega ao pico as 22h.

De modo geral, o modelo AD apresenta picos de potência, especialmente na quarta-feira, pouco condizentes com a curva real de potência ativa, enquanto os modelos FDA e MLCP são capazes de manter a linearidade e tendência dos valores reais.

5.2.3 Análise do final de semana

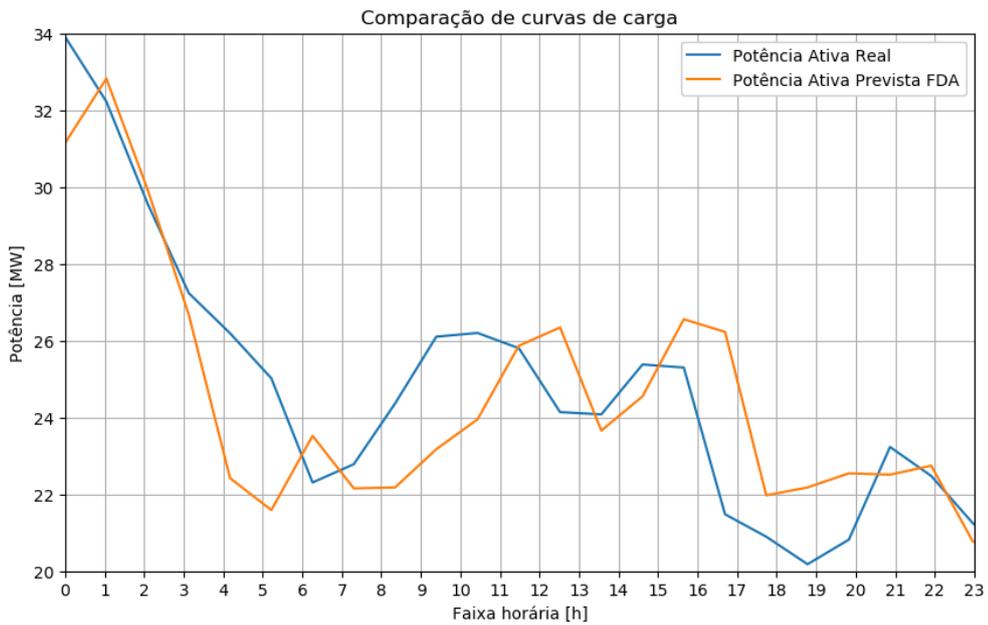
As curvas de carga das Figuras 45, 46 e 47, para o sábado, e 48, 49 e 50 para o domingo, são apresentadas a seguir.

Figura 45 – Sábado: Modelo Árvore de Decisão



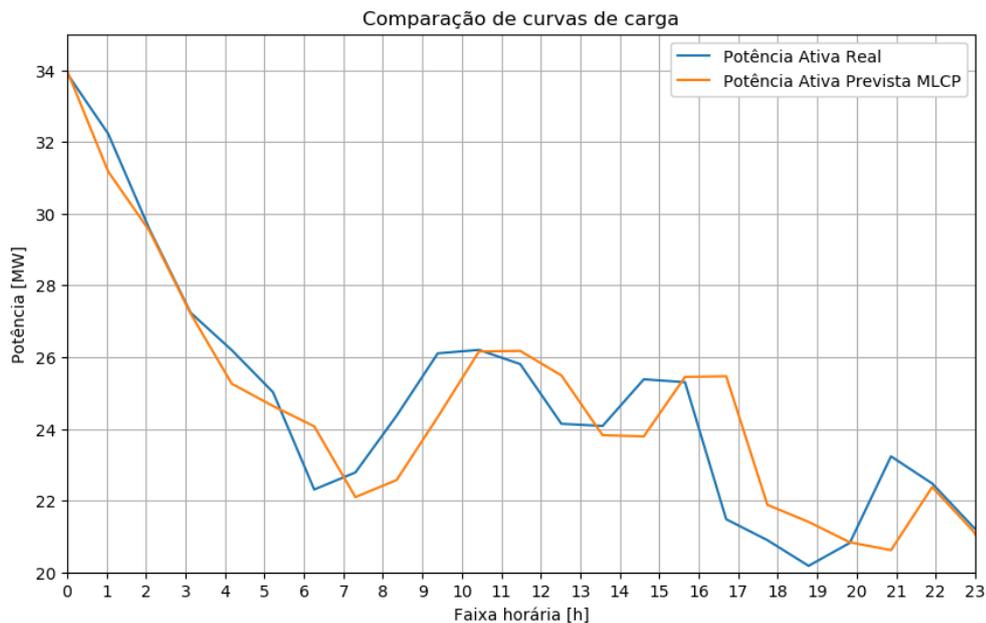
Fonte – Elaborado pela autora

Figura 46 – Sábado: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 47 – Sábado: Modelo Memória de Longo e Curto Prazo



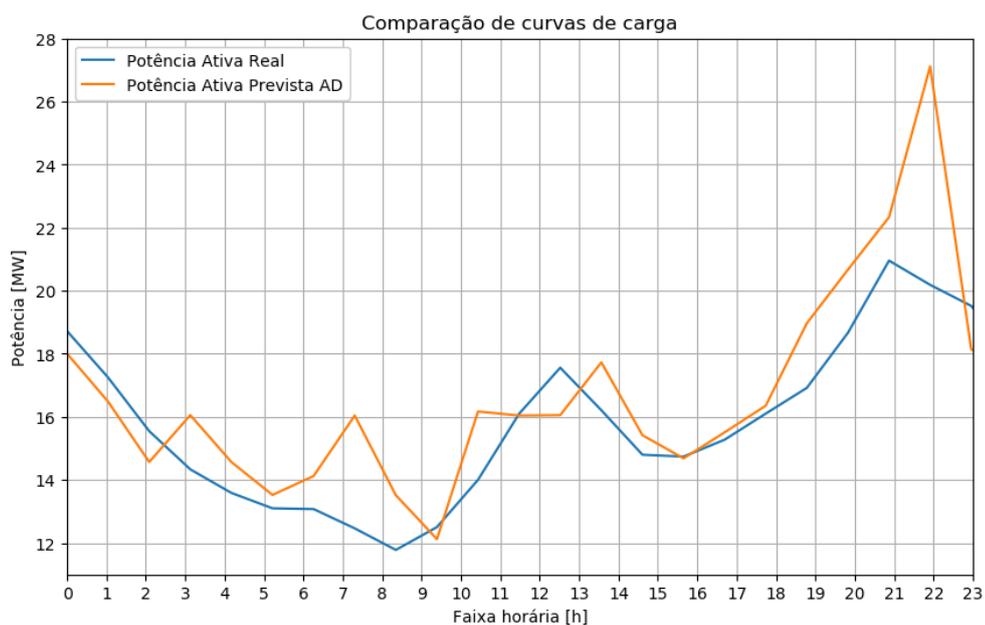
Fonte – Elaborado pela autora

O sábado apresenta uma curva de carga bastante diferente se comparados aos dias da semana. A curva apresenta uma tendência de queda de potência durante o dia, com a maior queda registrada entre a 00h e 06h, e alternância de picos e vales durante a tarde e a noite.

A análise dos gráficos permite concluir facilmente que o modelo MLCP possui a maior vantagem nas previsões de potência quando comparados aos modelos FDA e AD. Os erros, encontrados no gráfico da MLCP, provêm fortemente da variável de potência da hora anterior já que o modelo apresenta um atraso de uma hora entre os valores reais de potência ativa e aqueles previstos pelo algoritmo.

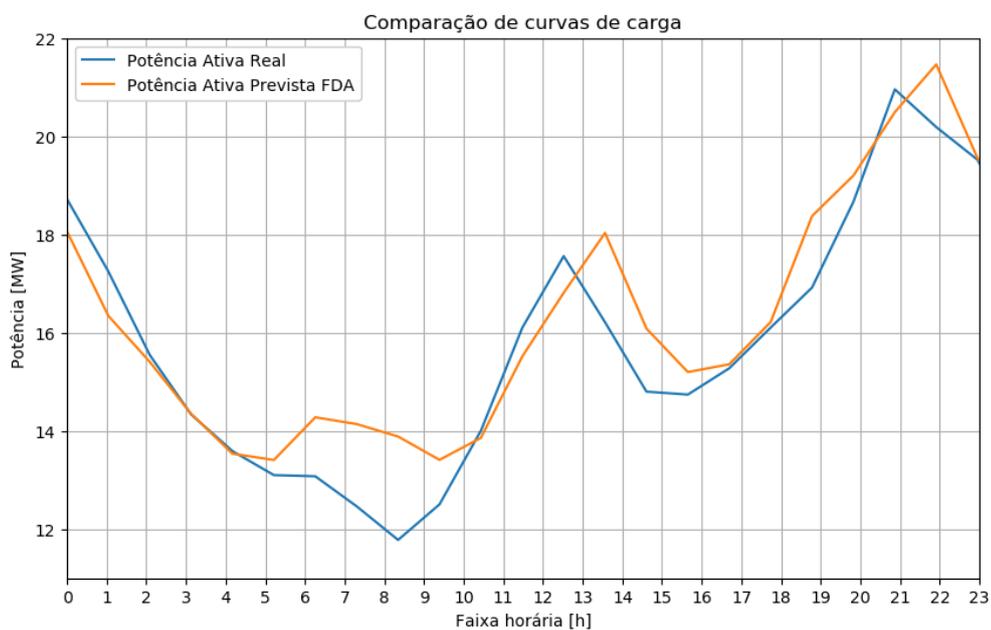
Os modelos Florestas de Decisão Aleatória e Árvore de Decisão, por sua vez, apresentam grande dificuldade em seguir a curva original.

Figura 48 – Domingo: Modelo Árvore de Decisão



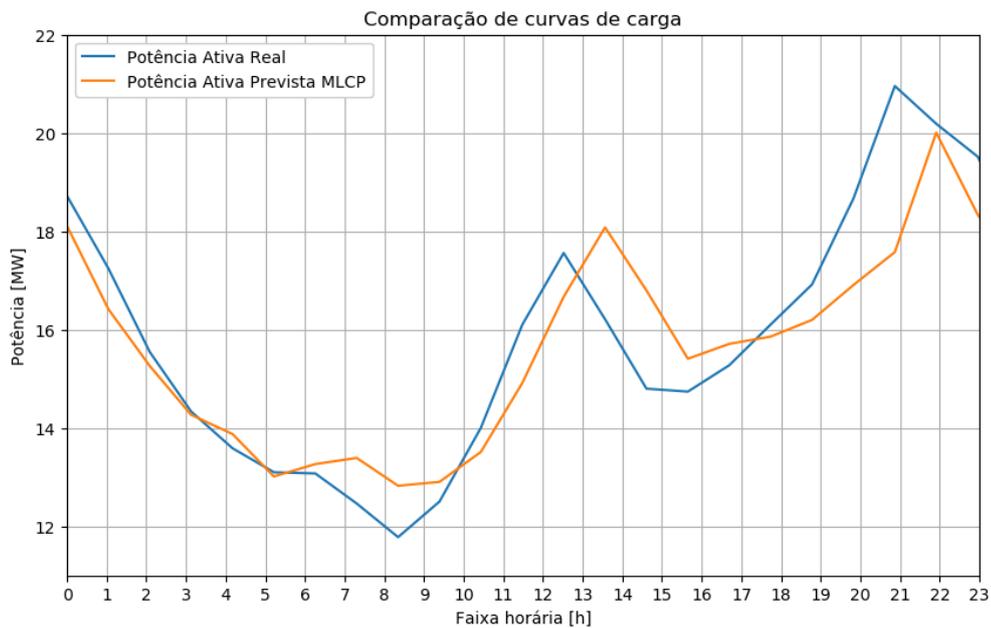
Fonte – Elaborado pela autora

Figura 49 – Domingo: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 50 – Domingo: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

O domingo, por sua vez, apresenta uma curva de carga caracterizada por grandes mudanças na potência ao longo do dia com queda expressiva no período da manhã e dois picos de potência no início da tarde e durante a noite.

Essa mudança brusca de potência acarreta em picos 'aleatórios' nas previsões do modelo Árvore de Decisão, demonstrado pela Figura 48 entre as 02h e as 09h e outro as 22h.

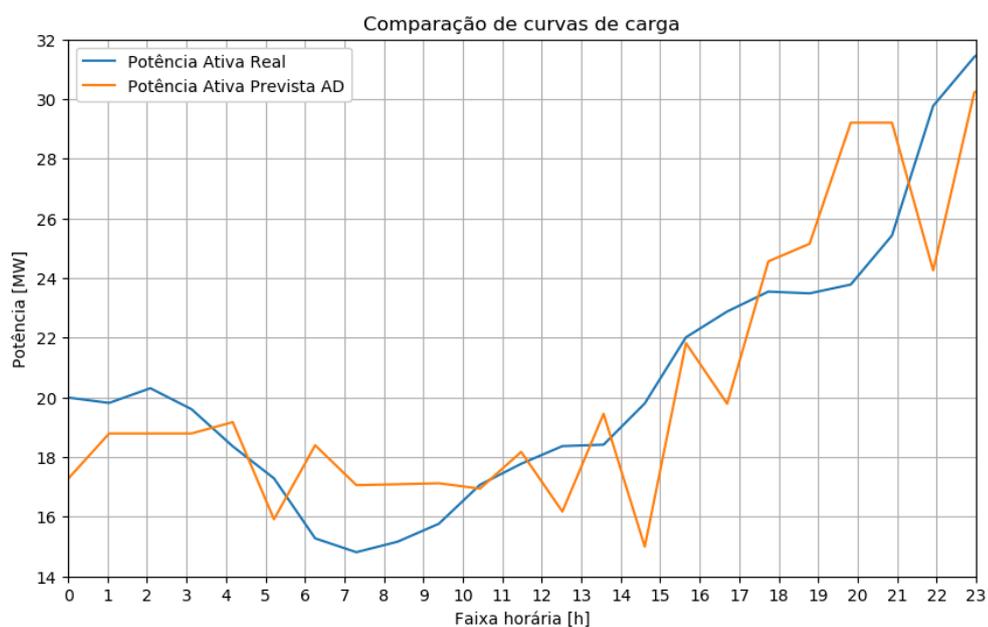
O modelo Florestas de Decisão Aleatória, por sua vez, apresenta previsões condizentes com a realidade com a exceção do vale criado as 08h na qual o modelo não consegue detectar a queda no consumo.

Novamente, os valores previstos pelo modelo MLCP alcançam os valores reais de potência do transformador com um leve atraso provindos da variável de potência da hora anterior.

5.2.4 Análise de feriados

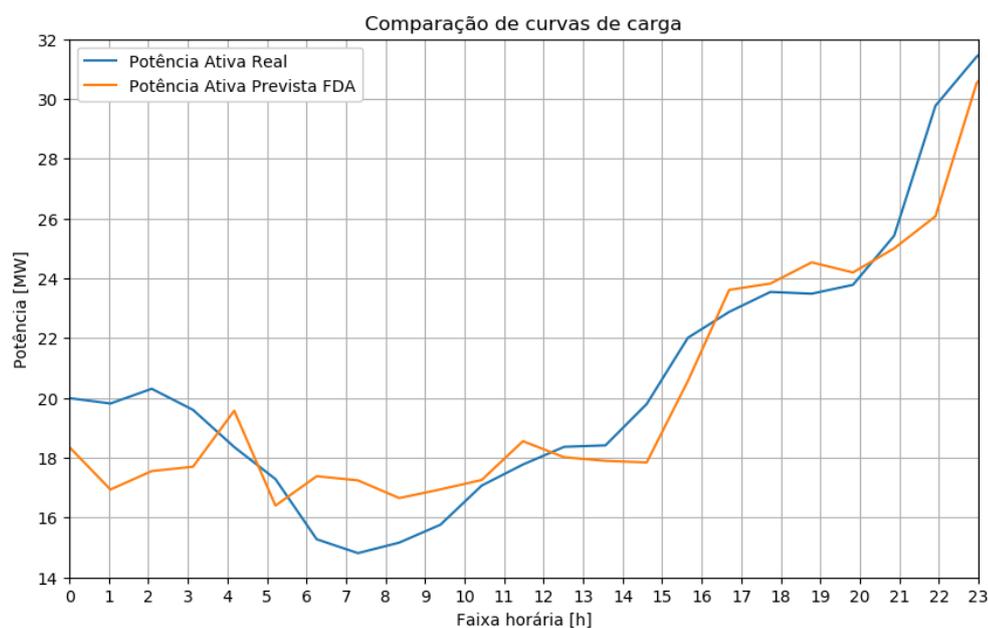
As Figuras 51, 52, 53, 54, 55 e 56 apresentam os resultados encontrados pelos modelos no Natal e no feriado de Finados, respectivamente.

Figura 51 – Natal: Modelo Árvore de Decisão



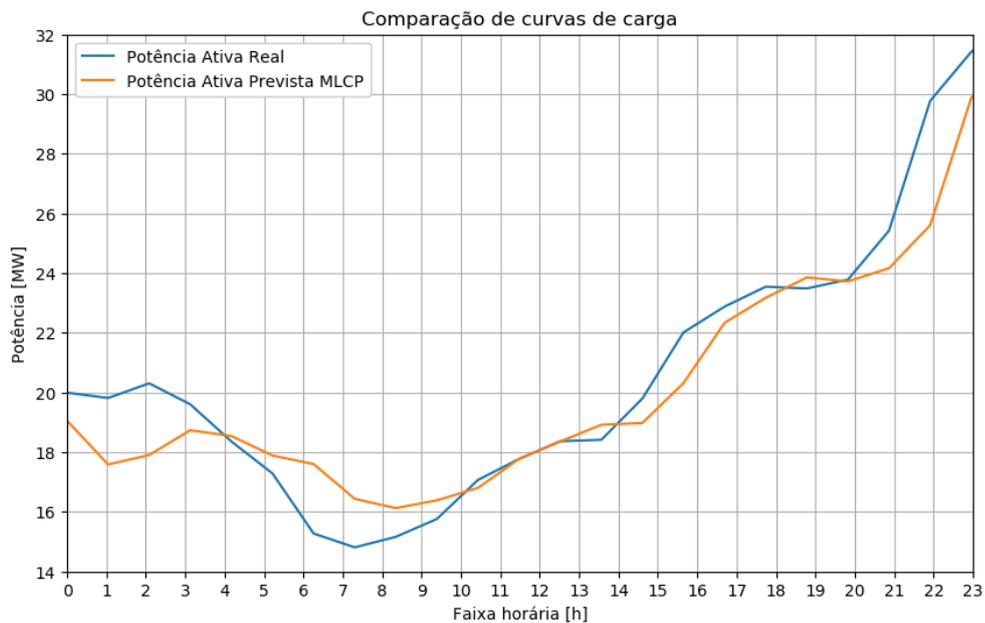
Fonte – Elaborado pela autora

Figura 52 – Natal: Modelo Florestas de Decis o Aleat ria



Fonte – Elaborado pela autora

Figura 53 – Natal: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

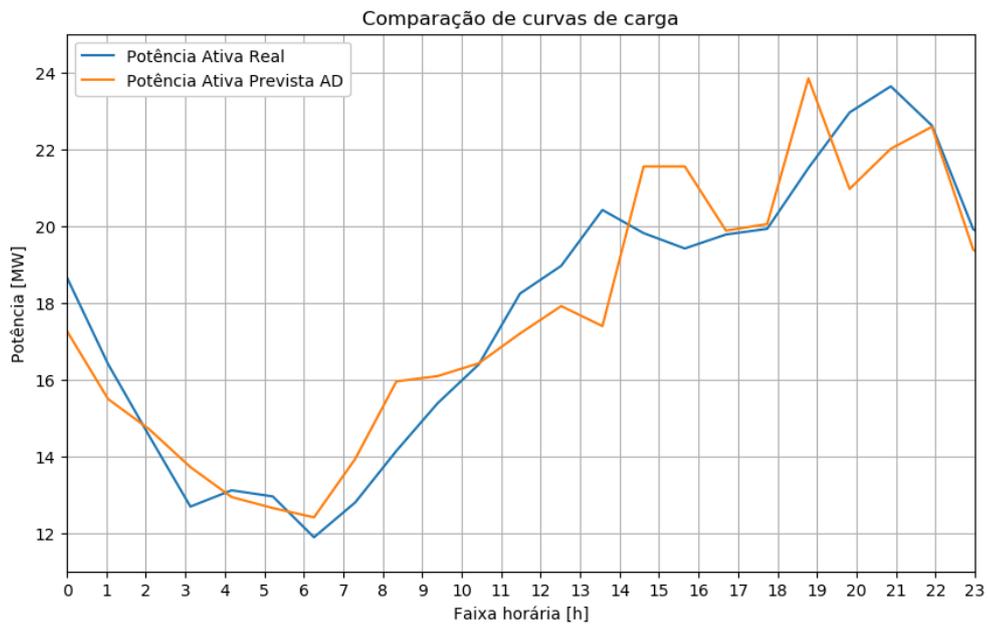
Assim como para São Leopoldo, a análise dos gráficos permite observar que a curva de carga durante o Natal é bastante distinta de qualquer outro dia da semana ou final de semana. A carga do Natal para o ano de 2019 apresenta valores intermediários durante a madrugada, provavelmente providas das ceias de Natal realizadas na véspera, seguida de leve queda nas primeiras horas da manhã e aumento constante durante a tarde e a noite.

O modelo de previsão AD possui dificuldades de acompanhamento da curva real, apresentando picos e vales 'aleatórios' de carga durante o dia. O modelo é capaz apenas de acompanhar a tendência de aumento ou queda na potência do dia.

O modelo FDA, assim como o AD, apresenta certa dificuldade de acompanhamento, entretanto, apresenta erros menores e picos menos expressivos durante o dia.

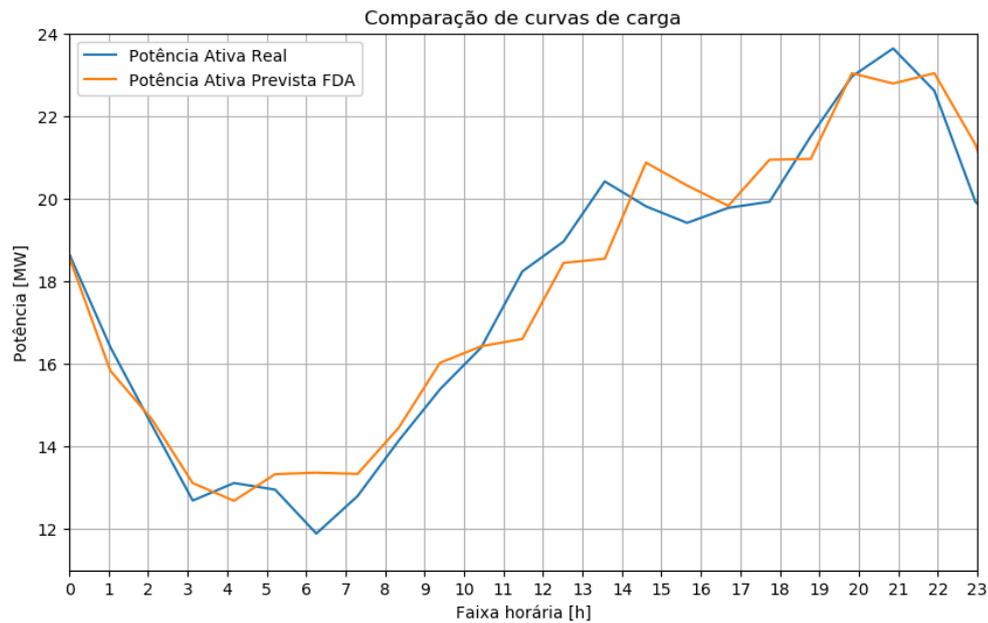
Por sua vez, o modelo MLCP apresenta melhores previsões uma vez que está mais preocupado com a série temporal potência da hora anterior do que com comportamentos em semelhantes condições em dias passados ou futuros utilizados pelos métodos de Árvore de Decisão e Florestas de Decisão Aleatória.

Figura 54 – Finados: Modelo Árvore de Decisão



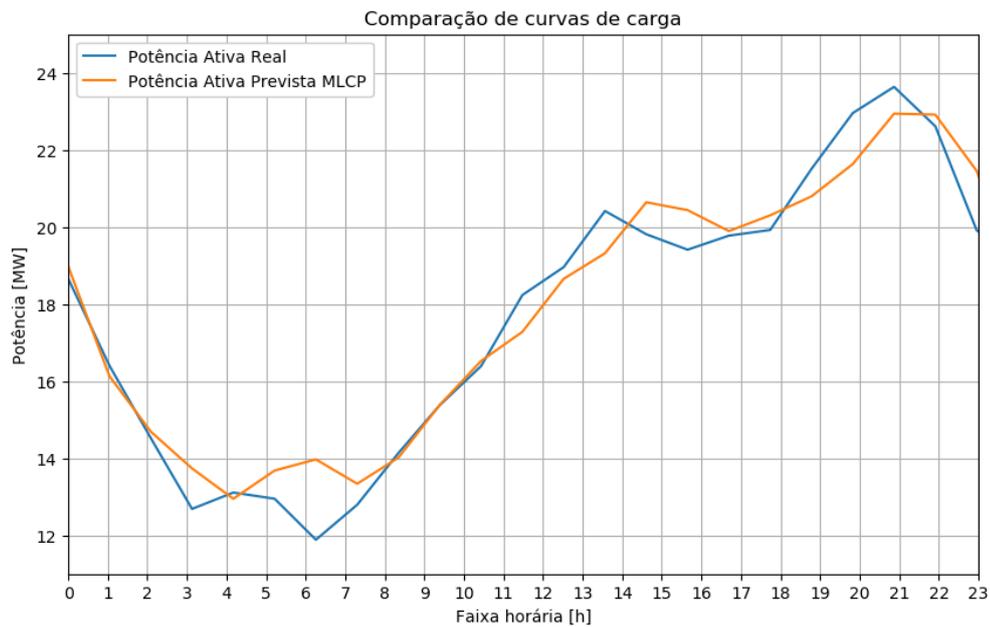
Fonte – Elaborado pela autora

Figura 55 – Finados: Modelo Florestas de Decisão Aleatória



Fonte – Elaborado pela autora

Figura 56 – Finados: Modelo Memória de Longo e Curto Prazo



Fonte – Elaborado pela autora

Semelhante ao comportamento de dias da semana, o feriado de Finados apresenta queda nas primeiras horas da manhã, aumento durante a tarde e pico pela noite.

O modelo de previsão AD, assim como no natal, possui dificuldades de acompanhamento da curva real de potência ativa, apresentando picos e vales 'aleatórios' de carga especialmente durante a noite.

Os modelos FDA e MLCP apresentam bons resultados em comparação com a curva de carga real.

6 Conclusões

A metodologia desenvolvida neste trabalho foi aplicada para estimar as curvas de carga de transformadores de potência conectados em subestações de distribuição através de ferramentas computacionais de Aprendizado de Máquina e Redes Neurais, considerando variáveis elétricas medidas junto ao barramento da subestação e fatores climáticos.

A estimativa de carga feita considerando a medição de potência ativa da hora anterior nos modelos analisados mostrou-se capaz de prever a demanda de potência da hora atual com excelente precisão. No caso do modelo de rede neural, por exemplo, essa variável recebeu maior peso se comparado aos outros elementos, já que é possível ver um atraso de uma hora em grande parte das curvas de carga analisadas.

Importante destacar que a consideração de variáveis climáticas nos modelos mostrou-se bastante importante para aprimorar as estimativas, uma vez que essas variáveis ajudaram a melhorar os resultados, aproximando a curva de carga estimada da curva de carga real medida. O comportamento não linear da temperatura, umidade, pressão e velocidade dos ventos em relação a potência ativa conseguiu ser quantificado pelos modelos, especialmente o Florestas de Decisão Aleatória e o Memória de Longo e Curto Prazo, resultando na melhoria dos valores encontrados.

Dentre os modelos implementados e analisados, os melhores resultados foram obtidos pela Florestas de Decisão Aleatória e pela rede neural Memória de Longo e Curto Prazo. Em ambas as cidades, Santa Maria e São Leopoldo, o modelo de rede neural apresentou notável performance alcançando 93,50% e 95,86% nos parâmetros de R quadrado, respectivamente. O modelo Florestas de Decisão Aleatória por sua vez, atingiu também valores de R2 significativos, 96,27% para São Leopoldo e 93,94% para Santa Maria. Esses modelos foram capazes de seguir a curva de carga real para os dias da semana, finais de semana e até mesmo feriados em um único modelo de previsão. Essa característica torna-se relevante se considerado que diversos autores utilizam modelos separados para finais de semana e feriados devido a especificidade da curva de carga desses dias.

Nesse aspecto, o modelo de regressão Árvore de Decisão apresentou resultados insatisfatórios, uma vez que este foi capaz de prever curvas de carga bastante realistas para os dias da semana, tendo apresentado dificuldades em acompanhar o consumo de feriados e finais de semana no qual não há um padrão visível. Nesse caso, a combinação de mais de um modelo é recomendada para que se obtenha estimativas mais assertivas.

Por fim, cabe destacar que os modelos desenvolvidos apresentam resultados satisfatórios para previsão da carga no curto prazo (período de 24h). Ainda, estes podem apresentar resultados inferiores se aplicados para períodos de tempo maiores.

6.1 Propostas de Trabalhos Futuros

Os modelos de projeção de carga criados a partir de algoritmos inteligentes agrupando informações de potência e de fatores climáticos da região tem como vantagem, para transformadores de força, a possibilidade de inferência da temperatura interna desse transformador (temperatura do enrolamento), fator relevante que interfere na capacidade do mesmo em suportar potências maiores ou menores em condições de sobrecarga. Neste sentido, algumas melhorias no presente trabalho podem ser realizadas:

- A realimentação em tempo real dos dados de potência e dos fatores climáticos em um sistema integrado com o próprio transformador pode facilitar o controle por parte de concessionárias de energia, já que esses dados podem auxiliar no planejamento e remanejamento de carga de transformadores e de subestações inteiras.
- Averiguar o comportamento da temperatura e da carga em diferentes condições. No presente trabalho são contempladas duas cidades; São Leopoldo e Santa Maria, que apresentam diferentes relações entre a temperatura do local e do consumo de energia elétrica. Uma análise dos bancos de dados segundo um Projeto de Experimentos completo pode determinar como cada variável influencia no modelo de previsão, possibilitando uma comparação entre as duas cidades em questão. Essa comparação poderá evidenciar quais as condições são mais favoráveis para o funcionamento dos transformadores, quais condições apresentam um risco para a rede elétrica, e como cada região deve obter um ajuste na carga segundo os componentes climáticos do local.

Referências Bibliográficas

- Ali, J. et al. Random forests and decision trees. In: *2012 IJCSI International Journal of Computer Science Issues*. [S.l.: s.n.], 2012. v. 9, n. 5.
- Breiman, L. Random forests. In: . [S.l.: s.n.], 2001. v. 45, n. 1, p. 5–32.
- Cavalcanti, M. A. F. H.; Júnior, J. R. C. S. *Cenários macroeconômicos para o período de 2020-2031*. 2018. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/9274/1/cc41_nt_cen\%C3\%A1rios_macroeconomicos.pdf>.
- Chowdhury, D. et al. Zone wise hourly load prediction using regression decision tree model. In: *2018 International Conference on Innovation in Engineering and Technology (ICIET)*. [S.l.: s.n.], 2018. p. 1–6.
- Cui, C. et al. Research on power load forecasting method based on lstm model. In: *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*. [S.l.: s.n.], 2020. p. 1657–1660.
- Dehalwar, V. et al. Electricity load forecasting for urban area using weather forecast information. In: *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*. [S.l.: s.n.], 2016. p. 355–359.
- Deng, J.; Jirutitijaroen, P. Short-term load forecasting using time series analysis: A case study for singapore. In: *2010 IEEE Conference on Cybernetics and Intelligent Systems*. [S.l.: s.n.], 2010. p. 231–236.
- Dong, X. et al. Application of adaboost algorithm based on decision tree in forecasting net power of circulating power plants. In: *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. [S.l.: s.n.], 2020. v. 1, p. 747–750.
- Elma, O.; Selamoğullar, U. S. A survey of a residential load profile for demand side management systems. In: *2017 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*. [S.l.: s.n.], 2017. p. 85–89.
- ENERGIA, M. de Minas e. *Plano Decenal de Expansão de Energia*. 2020. Disponível em: <<https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/Documents/PDE\%202029.pdf>>.
- Hong, T.; Fan, S. Probabilistic electric load forecasting: a tutorial review. In: . [S.l.: s.n.], 2016.
- Jiang, Q. et al. Electricity power load forecast via long short-term memory recurrent neural networks. In: *2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC)*. [S.l.: s.n.], 2018. p. 265–268.
- Khan, S. R.; Noor, S. Short term load forecasting using svm based puk kernel. In: *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. [S.l.: s.n.], 2020. p. 1–9.

- Lahouar, A.; Ben Hadj Slama, J. Random forests model for one day ahead load forecasting. In: *IREC2015 The Sixth International Renewable Energy Congress*. [S.l.: s.n.], 2015. p. 1–6.
- Lekshmi, M.; Subramanya, K. N. A. Short-term load forecasting of 400kv grid substation using r-tool and study of influence of ambient temperature on the forecasted load. In: *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*. [S.l.: s.n.], 2019. p. 1–5.
- Liu, S. et al. Short-term load forecasting based on gbdt combinatorial optimization. In: *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*. [S.l.: s.n.], 2018. p. 1–5.
- Patel, H.; Pandya, M.; Aware, M. Short term load forecasting of indian system using linear regression and artificial neural network. In: *2015 5th Nirma University International Conference on Engineering (NUiCONE)*. [S.l.: s.n.], 2015. p. 1–5.
- Qiuyu, L. et al. Short-term load forecasting based on load decomposition and numerical weather forecast. In: *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*. [S.l.: s.n.], 2017. p. 1–5.
- Rafi, S. H.; Nahid-Al-Masood, N. Highly efficient short term load forecasting scheme using long short term memory network. In: *2020 8th International Electrical Engineering Congress (iEECON)*. [S.l.: s.n.], 2020. p. 1–4.
- Silva, L. N. et al. Development of brazilian multi region short-term load forecasting model considering climate variables weighting in ann model. In: *2017 52nd International Universities Power Engineering Conference (UPEC)*. [S.l.: s.n.], 2017. p. 1–6.
- Silva, L. N. et al. Diagnostic and input selection tool applied on weather variables for studies of short-term load forecasting. In: *2019 8th International Conference on Modern Power Systems (MPS)*. [S.l.: s.n.], 2019. p. 1–6.
- Soliman, S. Electrical load forecasting: Modeling and model construction. In: . [S.l.: s.n.], 2010.
- Xiaoyu Wu et al. A two-stage random forest method for short-term load forecasting. In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. [S.l.: s.n.], 2016. p. 1–5.
- Ye, J.; Yang, L. Short-term load forecasting using ensemble empirical mode decomposition and harmony search optimized support vector regression. In: *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. [S.l.: s.n.], 2019. p. 851–855.