

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**UMA ABORDAGEM BASEADA EM REDES
NEURAIS ARTIFICIAIS E CLUSTERIZAÇÃO
PARA PREVISÃO DE CURTO PRAZO DA
DEMANDA DE ENERGIA ELÉTRICA**

ERIC REINOLDO SCREMIN

ORIENTADOR: PROF. DR. RICARDO A. S. FERNANDES

São Carlos – SP
Agosto/2018

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**UMA ABORDAGEM BASEADA EM REDES
NEURAIS ARTIFICIAIS E CLUSTERIZAÇÃO
PARA PREVISÃO DE CURTO PRAZO DA
DEMANDA DE ENERGIA ELÉTRICA**

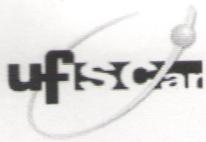
ERIC REINOLDO SCREMIN

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Ricardo A. S. Fernandes

São Carlos – SP

Agosto/2018



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Eric Reinoldo Scremin, realizada em 17/08/2018:

Prof. Dr. Ricardo Augusto Souza Fernandes
UFSCar

Prof. Dr. Fábio Anderson Silva Borges
UFPI

Prof. Dr. Samuel Lourenço Nogueira
UFSCar

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Fábio Anderson Silva Borges e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Ricardo Augusto Souza Fernandes

AGRADECIMENTOS

Agradeço primeiramente a minha mãe Elizabeth a qual sempre me deu apoio, me incentivou e que não apenas nesses anos durante toda minha vida me ensinou a levantar a cada queda. Também agradeço meus irmãos por toda sabedoria repassada que a experiência deles fizeram a caminhada mais segura. Obrigado a todos meus amigos pelos momentos de descontração, pelas novas amizades desenvolvidas em todo o processo deste mestrado.

Ao meu orientador Professor Dr. Ricardo por toda a orientação, incentivo, a oportunidade de trabalho e principalmente a paciência demonstrada durante o desenvolvimento do projeto. A todos os professores envolvidos durante o processo de formação assim como a universidade e o departamento a qual fornece esse curso de pós-graduação. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Por último agradeço a presença, até quase o fim deste trabalho, de meu amigo de quatro patas Chico que pereceu recentemente, este sem proferir uma palavra foi responsável por momentos felizes mesmo quando estava perdido, iluminando meus dias com alegria.

"A dog is the only thing on earth that loves you more than he loves himself."

Josh Billings (1818 - 1885)

RESUMO

A energia elétrica possui um papel crucial no desenvolvimento de um país, pois, a mesma influencia diretamente muitos setores da sociedade. Nesse sentido, a previsão da demanda de energia elétrica é de suma importância para a manutenção e crescimento dos sistemas elétricos de potência. Atualmente, existem diferentes abordagens utilizadas para tal previsão. Além disso, existem muitas variáveis que podem influenciar o desempenho desse processo. Dentre estas variáveis, pode-se destacar aquelas derivadas do clima. Portanto, este trabalho propõe a aplicação de uma Rede Neural Artificial do tipo *Multilayer Perceptron* com algoritmo de treinamento de Levenberg-Marquardt, utilizando-se de variáveis referentes a temperatura do ar e a demanda de energia elétrica. Conjuntamente, o trabalho objetiva averiguar a relação entre as variáveis, fazendo uso do método de agrupamento *k-means* nos dados de entrada. Os resultados mostram que essa abordagem baseada em agrupamentos pode contribuir fornecendo previsões com baixas taxas de erro. Entretanto, obteve-se resultados ligeiramente melhores quando os dados não são agrupados.

Palavras-chave: Demanda de energia elétrica, Previsão de séries temporais, Redes neurais artificiais, Clima.

ABSTRACT

Electricity plays a crucial role in the development of a country because it directly influences many sectors of the society. In this sense, the demand forecasting is of paramount importance for the maintenance and growth of the electric power systems. Currently, there are different approaches used for such forecasting. In addition, there are many variables that can influence the performance of this process. Among these variables, one can highlight those derived from the climate. Therefore, this work proposes the application of an Artificial Neural Network of Multilayer Perceptron type with Levenberg-Marquardt training algorithm, using temperature and demand as input variables. Moreover, another objective of this work is to investigate the relationship between variables, making use of the k-means clustering method on the input data. The results show that this clustering-based approach obtains predictions with low error rates. However, slightly better results were obtained when the data were not clusterized.

Keywords: Electric load forecasting, Time-series forecasting, Artificial neural networks, Climate.

LISTA DE FIGURAS

2.1	Caminho da Previsão de demanda (Adaptado de (GUPTA; PAL, 2017)).	21
2.2	Mapa de densidade de ocorrências das palavras-chave.	30
2.3	Mapa de densidade de ocorrências das palavras-chave, divididas em dois grupos.	31
2.4	Número de Publicações por Ano.	32
2.5	Número de Publicações por País.	33
3.1	Visão geral da metodologia proposta.	35
3.2	Áreas de atuação da PJM. Fonte:(PJM, 2018)	36
3.3	Áreas de atuação da EKPC. Fonte: Kentucky Energy Profile 2015.	37
3.4	Perfil da geração de energia no estado do Kentucky. Fonte: Kentucky Energy Profile (2015).	38
3.5	Temperatura média por dia da região do estudo referente aos anos de 2014 a 2016.	40
3.6	Demanda de energia elétrica da EKPC referente aos anos de 2014 a 2016.	41
3.7	Entrada dos dados no modelo neural.	41
3.8	Entrada dos dados no modelo neural com a utilização de <i>clusters</i>	42
3.9	Modelo da rede neural proposta.	43
4.1	Agrupamentos de temperatura e demanda por hora.	45
4.2	Resultados obtidos pela MLP com topologia de 25 neurônios na camada oculta.	47
4.3	Erro absoluto percentual apresentado após 25 repetições de treinamento e teste ao longo dos meses de 2016 (RNA sem entradas clusterizadas).	48
4.4	Diagrama da configuração de <i>clusters</i>	49
4.5	Resultados obtidos pela MLP com a utilização de topologia composta.	50

4.6	Percentual do erro absoluto apresentado após 25 repetições de treinamento e teste ao longo dos meses de 2016 (RNAs com entradas clusterizadas).	50
4.7	Histograma dos erros apresentados pelos resultados obtidos sem a separação de dados.	51
4.8	Histograma dos erros apresentados pelos resultados obtidos com a separação de dados.	51

LISTA DE TABELAS

2.1	Indicadores da Análise Bibliométrica.	28
2.2	Relação das Palavras-Chave por Periódicos.	29
2.3	Relação das Palavras-Chave por Autores.	29
2.4	Quantidade de Citações por País.	33
3.1	Correlação de Pearson obtida entre os valores de temperatura e de demanda da energia elétrica (de 4 horas anteriores) em relação à variável a ser predita (demanda).	39
3.2	Correlação de Spearman obtida entre os valores de temperatura e de demanda da energia elétrica (de 4 horas anteriores) em relação à variável a ser predita (demanda).	39
4.1	Tabela de Distribuição nos Agrupamentos	46
4.2	Resultados obtidos referentes à aplicação do modelo neural sem clusterização das entradas e com variação topológica.	47
4.3	Resultados obtidos referentes à aplicação do modelo neural com clusterização das entradas e com variação topológica.	49
4.4	Sumarização de resultados obtidos pela literatura correlata.	52

SUMÁRIO

CAPÍTULO 1 – INTRODUÇÃO	17
1.1 Motivação e Justificativa	17
1.2 Objetivos	18
1.3 Organização	19
CAPÍTULO 2 – REFERENCIAL TEÓRICO E ESTUDO BIBLIOMÉTRICO	21
2.1 Previsão de Demanda em Horizonte de Longo Prazo	22
2.2 Previsão de Demanda em Horizonte de Médio Prazo	23
2.3 Previsão de Demanda em Horizonte de Curto Prazo	23
2.4 Previsão de Demanda em Horizonte de Curtíssimo Prazo	26
2.5 Estudo Bibliométrico	27
2.6 Definição da Base de Dados e Buscas	27
2.7 Análise Sumarizada	28
2.8 Análise de Palavras-Chave	28
2.9 Análise do Número de Publicações	32
CAPÍTULO 3 – PROCEDIMENTOS E METODOLOGIA PROPOSTA	35
3.1 Dados Coletados para o Experimento	36
3.2 Clusterização	38
3.3 Organização dos Dados Processados	40
3.4 Modelo Proposto Baseado em Redes Neurais Artificiais	42

3.5	Métricas de Avaliação de Desempenho	43
CAPÍTULO 4 – RESULTADOS E DISCUSSÕES		45
4.1	<i>Clusters</i> obtidos entre demanda e temperatura	45
4.2	Resultados obtidos pelos modelos neurais	46
4.2.1	Modelo Neural com entradas não clusterizadas	46
4.2.2	Modelo Neural com entradas clusterizadas	48
4.3	Discussões	51
CAPÍTULO 5 – CONCLUSÕES		55
REFERÊNCIAS		57

Capítulo 1

INTRODUÇÃO

A energia elétrica possui um papel crucial no desenvolvimento de um país, pois a mesma influencia diretamente muitos setores da sociedade. Atualmente, a energia elétrica é comercializada como qualquer outra *commodity*, porém tem suas peculiaridades. Neste sentido, pode-se destacar sua alta volatilidade, a qual está atrelada à sua dependência a diversos outros fatores, a saber: preço do combustível, variações na geração hidroelétrica, crescimento econômico do país, histórico de preços e de demanda da energia elétrica (AMJADY; KEYNIA, 2009).

1.1 Motivação e Justificativa

Além dos fatores supracitados, esta dissertação ainda é motivada pelo fato de que a demanda da energia elétrica também pode ser afetada por fatores externos que ocorrem durante os dias de uma semana, bem como pelas estações do ano (MANDAL; SRIVASTAVA; PARK, 2009). É importante destacar que, diferente de outras *commodities*, ela não pode ser armazenada em grandes quantidades. Portanto, a oferta deve satisfazer a demanda de energia elétrica dos consumidores no momento em que a mesma for exigida (AMJADY; KEYNIA, 2009; MANDAL; SRIVASTAVA; PARK, 2009; GARCIA-MARTOS; RODRIGUEZ; SANCHEZ, 2012; MOHSENIAN-RAD; LEON-GARCIA, 2010; AMJADY; HEMMATI, 2006). Assim, a estimação da demanda da energia elétrica passou a ser essencial para o planejamento estratégico de concessionárias de energia elétrica ou de indústrias, pois, possibilita a minimização dos custos relacionados à energia elétrica. Contudo, essa não é uma tarefa trivial, visto que existe uma infinidade de fatores (atributos) que podem influenciar o processo de previsão. Em (GUPTA; PAL, 2017), os autores listam grande parte dos atributos utilizados para estas previsões, tais como, a umidade do ar, taxa de desenvolvimento econômico, temperatura do ar, além de atividades humanas consideradas cíclicas. Em função desse comportamento cíclico, (BOROOJENI et al., 2017) buscam dividir tais padrões por dia, se-

mana e/ou mês. Outros autores como, (ÇEVİK; ÇUNKAŞ, 2016) fazem essa divisão em função de feriados/finais de semana e dias úteis.

Ademais, outra motivação se dá pelas investigações científicas que se estendem às ferramentas de previsão propriamente ditas. Em (SUGANTHI; SAMUEL, 2012), os autores ressaltam que, nas últimas décadas, surgiram ferramentas que parte de métodos estritamente estatísticos e vão até métodos de *soft computing*. Dentre as ferramentas estatísticas é importante destacar os modelos autorregressivos como ARMA (*Autoregressive Moving Average*) e ARIMA (*Autoregressive Integrated Moving Average*). Em relação às ferramentas de *soft computing*, destacam-se as Redes Neurais Artificiais (RNA) e os Sistemas de Inferência *Fuzzy*.

De acordo com o contexto previamente mencionado e também em função da Revisão Bibliográfica que será apresentada no Capítulo 2, foi possível notar que diversos países têm investido em ferramentas para a previsão de demanda da energia elétrica, as quais visam aprimorar o sistema, tornando-o mais confiável, seguro, inteligente e a um custo mais baixo, conforme citado em (AMJADY; HEMMATI, 2006). Assim, nota-se que essas são características alvo das *Smart Grids*.

Dessa forma, essa dissertação se apoia em uma abordagem que visa empregar RNA e técnicas de separação e agrupamento de dados referentes aos dados públicos de demanda de energia elétrica e climatológicos do mercado PJM (*Pensylvannia - New Jersey - Maryland*). Tal mercado representa a interconexão energética do leste dos Estados Unidos. Os dados climatológicos são provenientes de leituras realizadas pela *University of Kentucky*. Cabe informar que foram selecionados tais dados devido à completude dos mesmos, pois não foi possível obter dados brasileiros com a mesma qualidade.

1.2 Objetivos

Os objetivos gerais desta dissertação consistem na investigação, desenvolvimento, projeto, implementação e validação de RNA para a previsão da demanda de energia elétrica no âmbito das *Smart Grids*, ou seja, predispondo de uma grande quantidade de dados que deverão ser tratados ou agrupados para que se tornem úteis aos sistemas de informação de concessionárias e indústrias que façam parte do Mercado Livre de Energia. Dentro desse contexto, alguns objetivos específicos podem ser pautados em quatro itens principais:

- Gerar uma base de dados homogênea que possa armazenar os dados históricos referentes à série temporal da demanda, bem como das séries de dados coletados

referentes à temperatura. Essa homogeneidade é necessária para que os dados sejam disponibilizados a outras pesquisas do Laboratório de Inteligência Artificial Aplicada (LIAA);

- Permitir que novas abordagens, desenvolvidas por meio da aplicação de sistemas inteligentes possam ser testadas e validadas;
- Verificar o comportamento de RNA em conjunto com o método *k-means* no processo de previsão da demanda da energia elétrica;
- Realizar comparações de resultados e comportamentos da abordagem proposta em relação a outros métodos encontrados na literatura.

1.3 Organização

Esta dissertação está organizada da seguinte forma:

- Capítulo 2 - apresenta a revisão da literatura correlata, destacando-se os métodos e abordagens aplicados para longo, médio, curto e curtíssimo prazos, enfatizando a exposição das técnicas de previsão empregadas, origem e época dos dados, bem como as taxas de erro obtidas. Ainda apresenta o levantamento bibliométrico realizado, onde mostra-se a tendência de publicações na área e os pesquisadores mais relevantes (em termos de citações);
- Capítulo 3 - introduz a metodologia proposta para essa dissertação, demonstra os aspectos intrínsecos à base de dados obtida, detalhando a origem e sua organização. Ainda são apresentados os detalhes da abordagem proposta, ou seja, sua configuração e técnicas de validação e avaliação de desempenho;
- Capítulo 4 - são apresentados os resultados obtidos pela abordagem proposta, os quais são discutidos em função da literatura correlata;
- Capítulo 5 - destina-se a apresentar as conclusões obtidas pela realização deste trabalho, bem como a proposta de continuidade de pesquisa no âmbito do laboratório.

Capítulo 2

REFERENCIAL TEÓRICO E ESTUDO BIBLIOMÉTRICO

Este capítulo tem como objetivo enriquecer o embasamento teórico sobre o tema central dessa dissertação, ou seja, a previsão de demanda da energia elétrica. Portanto, na sequência, as subseções irão apresentar os horizontes de previsão comumente encontrados na literatura, bem como os tipos de atributos e as ferramentas de previsão empregadas.

Nos últimos 50 anos, com a importância da previsão da demanda da energia elétrica já reconhecida, foram desenvolvidos inúmeros métodos. Neste sentido, as previsões puderam ser divididas em: longo, médio, curto e curtíssimo prazos.

Antes de apresentar a literatura que representa cada horizonte de previsão, é importante comentar que independente do tipo de modelo e horizonte de previsão utilizado para a demanda da energia elétrica, o processo de previsão, via de regra, segue algumas etapas fundamentais, conforme pode ser notado no fluxograma da Figura 2.1.

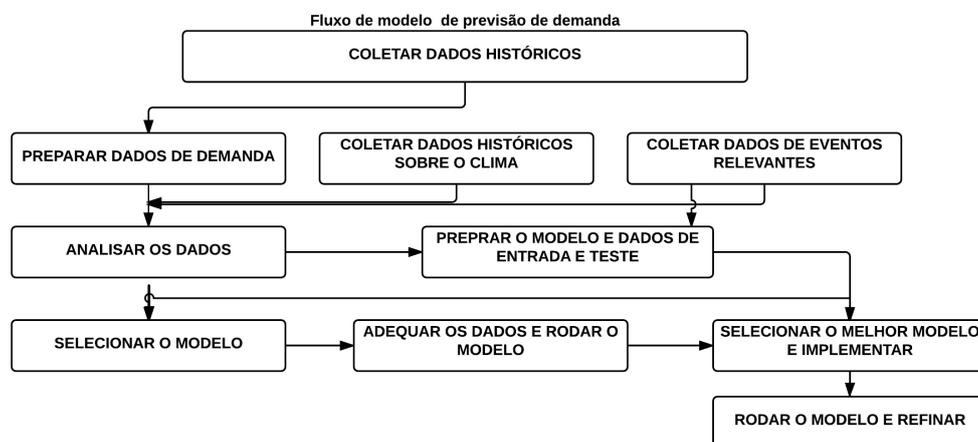


Figura 2.1: Caminho da Previsão de demanda (Adaptado de (GUPTA; PAL, 2017)).

O fluxograma expresso pela Figura 2.1 destaca as principais etapas em que as abordagens para previsões de demanda de energia elétrica são submetidas. A princípio, realiza-se a coleta de dados históricos, a qual é fundamental para a previsão de demanda. Em seguida, se necessário, faz-se a coleta de dados históricos correlatos, que podem afetar a série temporal (como por exemplo, dados climatológicos). Assim, os dados podem ser visualizados e preparados para o modelo escolhido para a predição, conforme explanado nas seções subsequentes.

2.1 Previsão de Demanda em Horizonte de Longo Prazo

A previsão de longo prazo é caracterizada por espaços de tempo entre um e vários anos, podendo chegar a décadas. Neste conceito, (AL-SABA; EL-AMIN, 1999) aplicam RNA comparadas a métodos auto-regressivos em dados de demanda provenientes da Arábia Saudita, onde considerou-se um horizonte de 30 anos.

Aplicações de modelos matemáticos também podem trazer bons resultados para este tipo de previsão, conforme proposto por (HYNDMAN; FAN, 2010), onde os autores fazem uso de um modelo aditivo semi paramétrico, relacionando os dados de demanda de energia elétrica de um estado da Austrália. São utilizados atributos a respeito da temperatura da região. No entanto, essa é uma prática incomum para horizontes de longo prazo, pois o clima é uma característica fundamentalmente de curto e médio prazo. Contudo, os autores demonstram, a partir de um estudo das estações do ano, que o consumo de energia tende a ser maior em situações extremas de temperatura devido ao uso de aquecedores ou de equipamentos de ar condicionado.

Em (OH et al., 2016), os autores realizam uma previsão para os anos de 2014 até 2030 utilizando de um método de múltiplas regressões lineares. Assim, os autores notaram que, em Singapura, cerca de 31% do consumo de energia elétrica (com variância de 2%) é devido à utilização de sistemas de ar condicionado.

Na abordagem proposta por (LI; YAN; FENG, 2016), os autores fazem uso de variáveis da urbanização, variação econômica e expansão territorial da cidade. Neste sentido, quanto maior o crescimento econômico e demográfico da uma região da China, maior será o consumo de energia elétrica. Tal abordagem foi empregada para médio e longo prazos.

Por fim, em (GUPTA; PAL, 2017), os autores fazem uma revisão onde citam que as previsões de longo prazo podem empregar atributos demográficos e econômicos da região a ser estudada, bem como atributos que dizem respeito às mudanças de tecnologia.

2.2 Previsão de Demanda em Horizonte de Médio Prazo

Diferente da previsão supracitada, em médio prazo o objetivo é analisar o espaço de tempo de semanas a meses, porém, em alguns casos esse intervalo pode chegar a um ano.

O artigo publicado por (LEE; HONG, 2015) faz uso de um modelo dinâmico híbrido que analisa a dependência que a temperatura do ar tem em relação à carga. Esse modelo é aplicado em dados da área metropolitana de Seul. Mostra resultados significativamente melhores que os modelos ARIMA e Koyck.

A utilização de RNA para a previsão de médio prazo também é comum. Em (El Desouky; El Kateb, 2000), os autores relacionam um método ARIMA com RNA, incluindo como entradas os dados referentes à cidade de Jeddah na Arábia Saudita. Os atributos de entrada foram: temperatura máxima do mês atual; temperatura máxima do mês anterior e dois meses anteriores; pico de utilização de energia elétrica do mês anterior e de dois meses anteriores; valores máximos da temperatura do mês a ser previsto; maior temperatura média do mês anterior; componente indicador do tempo que representa a posição dos dados, 0 para o início dos dados e 1 para o final. Os resultados indicaram valores de MAPE (*Mean Absolute Percentage Error*) e RMSE (*Root Mean Square Error*) inferiores aos obtidos com métodos autoregressivos.

Destaca-se também o trabalho de (NAGI et al., 2011), onde o objetivo é prever picos de utilização diária. Para tanto, foi empregada uma abordagem baseada em Redes Neurais do tipo SOM (*Self Organizing Maps*) e SVM (*Support Vector Machine*). Assim, as redes SOM foram utilizadas na clusterização de dados, dividindo-os em dois *clusters*. Após esse passo, aplicou-se uma regressão de suporte vetorial (SVR, do inglês *Support Vector Regression*) que é responsável por distribuir os dados entre os dois *clusters*. Este modelo foi aplicado em dados coletados do mercado de energia PJM e também do mercado da península da Malásia.

Ainda em relação ao uso de RNA, em (NEZZAR et al., 2016) os autores propõe diversas topologias para uma Rede Neural de Auto Regressão Não Linear com um pré-processamento de dados mensais referente à produção de energia elétrica da Argélia entre os anos de 2000 e 2012.

2.3 Previsão de Demanda em Horizonte de Curto Prazo

A maior produção científica situa-se na previsão de demanda a curto prazo, ou seja, em um horizonte que abrange de horas até semanas. Por este motivo e também por este ser o foco da presente dissertação, foram consideradas apenas as publicações relevantes que envolvam Redes

Neurais Artificiais.

Em (KIM; YOUN; KANG, 2000), os autores citam que RNA pode apresentar um comportamento insatisfatório para previsões de feriados, pois, segundo os autores, são dias atípicos. Dessa forma, a previsão por RNA poderia retornar picos (variâncias) indesejáveis. Assim, eles propõem uma abordagem híbrida de RNA com Sistemas *Fuzzy*, onde a inferência é realizada para prever máximos e mínimos valores para os dias atípicos. Os dados referentes aos anos de 1990 a 1995 foram fornecidos pela KEPCO (*Korea Electric Power Corporation*). O MAPE total obtido foi de 1,78%.

Na proposta de (PETRIDIS et al., 2001), os autores fazem o uso de um preditor Bayesiano. Além disso, tal preditor é expandido para incluir além da formulação probabilística, generalizações não probabilísticas ao contexto de classificação de séries temporais e estimação de parâmetros. Esse modelo foi aplicado na previsão de demanda de energia em dados referentes à Grécia no intervalo dos anos de 1989 até 1994, obtendo resultados com MAPE de 2,3%.

Um modelo baseado em RNA combinado à Transformada Wavelet é proposto por (ZHANG; DONG, 2001), com o intuito de prever a demanda da energia em alguns estados da Austrália referentes ao ano de 1999. Apresentando resultados com APE (*Absolute Percentage Error*) para cada modelo de predição proposto, de um passo adiante até sete passos posteriores, onde para um passo o APE médio foi de 0,01% com picos de 0,05%. Entretanto, para a escala de sete passos, onde é aplicado o modelo híbrido proposto, o APE apresentado se situa entre 0,05% e 0,1%.

No artigo de (KALAITZAKIS; STAVRAKAKIS; ANAGNOSTAKIS, 2002), é feita uma análise comparativa de abordagens baseadas em redes neurais. As seguintes redes são comparadas, *Multilayer Perceptron* (MLP) com *Backpropagation* e taxa de aprendizado adaptativa, outra com *Backpropagation* e codificação Gaussiana, rede neural com ativação de pesos aleatória com e sem o algoritmo de treinamento *Moving Window Regression*, redes com função de base radial (RBF, do inglês *Radial Basis Function*), redes neurais recorrentes em tempo real e redes neurais recorrentes de autoregressão. Todos os dados são referentes à ilha de Creta, na Grécia. Os resultados obtidos que as redes neurais recorrentes de autoregressão apresentam o RMSE mínimo de 1,71.

Redes Neurais Wavelet foram utilizadas na previsão de dias semelhantes, por (CHEN et al., 2010), onde empregou-se dados do estado de New England referentes aos anos de 2006 a 2007. Nesse caso, foram analisados, por exemplo, dias em uma mesma estação do ano, apenas terças-feiras, apenas feriados em distintos anos, etc. Os autores dividem o erro de acordo com o mês

do ano, apresentando um MAPE mínimo de 0,03%.

Já em (SHEIKHAN; MOHAMMADI, 2012), propõe-se o uso combinado de meta-heurísticas, a saber: Algoritmos Genéticos e Colônias de Formigas, denominado GA-ACO (do inglês, *Genetic Algorithm - Ant Colony Optimization*). Ademais, também faz uso de outro modelo denominado PCA-MLP (*Principal Component Analysis - Multilayer Perceptron*) que combina a análise de componentes principais à RNA do tipo MLP. Ambos os modelos propostos apresentam bons resultados, obtendo um MAPE de 2,49% médio referente ao PCA-MLP e de 1,51% para o GA-ACO-MLP para previsões da semana posterior. Os testes foram realizados de acordo com as previsões de demanda da *West-Tehran Province Power Distribution Company*, entre os anos de 2009 e 2010.

As redes neurais recorrentes também são bastante utilizadas nessa área. Em (LI et al., 2014), os autores fazem uso de um modelo híbrido da rede neural de Elman. Os testes do modelo foram realizados com dados de 2008 da região de Chongqing, na China. Assim, obteve-se precisão considerada elevada pelos autores, com APE que oscila entre 0,24% e 4,6%.

Neste mesmo ano, (KOUHI; KEYNIA; Najafi Ravadanegh, 2014) realizaram um pré-processamento dos dados baseado na seleção de atributos e previsão com uma RNA do tipo MLP, onde o algoritmo de treinamento aplicado foi o de Levenberg-Marquardt. Os dados utilizados são do mercado PJM, do estado de New England, referentes ao ano de 2006. Obteve-se um MAPE de 2,62% e, após otimizações do aprendizado da RNA por meio dos algoritmos de enxame de partículas e diferencial evolucionário, foi possível obter um MAPE ainda menor, 1,72%.

Um novo modelo híbrido foi proposto por (FARD; AKBARI-ZADEH, 2014), o qual combina o modelo ARIMA com SVR e o algoritmo de busca Cuckoo. Como metodologia, os autores aplicam as técnicas de previsão separadamente e também em conjunto. Os testes são realizados com dados da demanda de energia do Irã, no ano de 2009. Obtendo com este modelo híbrido um MAPE de 0,4004%.

Conforme pôde ser notado, o mercado de New England é bastante utilizado como um *benchmark* para a análise de novos modelos de previsão. Neste sentido, em (GHAYEKHLOO; ME-NHAJ; GHOFRANI, 2015) é empregada uma rede neural Bayesiana que é caracterizada por ser menos dependente a variações em quantidades massiva de dados. Neste modelo, a atualização dos pesos da rede é realizado com um Algoritmo Genético a fim de minimizar o erro. Os dados de entrada da rede passam por um *Framework* de pré-processamento, onde aplica-se discretização e também a Transformada Wavelet. O modelo proposto foi testado em dados referentes a 5 anos de demanda, do intervalo de 2008 até 2013. Tais dados foram separados, sendo 80% deles utilizados para treinamento e o restante para validação do modelo. Os resultados

com a aplicação do modelo proposto alcançaram um MAPE de 0,3932%. Em contrapartida, o mesmo modelo sem otimização obteve um MAPE de 1,0037%.

Dada a popularidade das RNA, (DUDEK, 2016) lista e compara distintos modelos, como: MLP, RBF, GRNN (*Generalized Regression Neural Network*), FCN (*Fuzzy Counterpropagation Networks*) e as redes SOM. Todas as RNA foram aplicadas em dados coletados para previsões de dia posterior. Tais dados se referem à demanda de energia em diversos locais, saber: Polônia para um período entre 2002 e 2004; França para o período de 2007 a 2009; Grã-Bretanha no período de 2007 a 2009; e região de Victoria na Austrália no período de 2006 a 2008. Todos os dias atípicos foram removidos da base de dados (por exemplo, feriados). Dentre os resultados obtidos, a GRNN apresentou melhor precisão. Contudo, a diferença do MAPE para o segundo melhor método foi de aproximadamente 0,07%.

Utilizando-se também de dados fornecidos pela PJM, (SAHAY; TRIPATHI, 2014) fazem uso de RNA, onde são consideradas variáveis climatológicas. A base de dados abrange os anos de 2007 a 2011. Assim, fizeram parte dessa base de dados, os seguintes atributos: hora do dia, dia da semana, indicador de feriado (1 para sim, 0 para não), média da demanda das 24 horas antecedentes, a demanda registrada das 24 horas anteriores, ponto de orvalho e temperatura do ar. Os resultados foram obtidos para o ano de 2012 e o MAPE foi de 3,14%. Contudo, existem variações e erros inferiores dependendo do período observado.

Esta dissertação tem sua abordagem semelhante à de (SAHAY; TRIPATHI, 2014), porém faz a previsão apenas de uma concessionária de energia do conglomerado PJM, em vez de usar os dados que representam todo o mercado de energia representado pelo PJM.

2.4 Previsão de Demanda em Horizonte de Curtíssimo Prazo

No horizonte de curtíssimo prazo, o processo de previsão é mais sensível aos erros. Portanto, a variedade de métodos aplicáveis é menor quando comparada ao de curto prazo.

Um modelo de caixa cinza (*Gray-model*) foi proposto por (YAO; CHI; CHEN, 2003) devido à sua menor dependência de quantidades massivas de dados. Os resultados mostram dois dilemas, a dissipação dos dados gerados pelo modelo e a extrapolação de tentativas, onde foram adaptadas modificações para a minimização do problema. Sem as modificações, os resultados foram totalmente inadequados para a previsão das séries temporais. Ao adicionar as modificações, pode-se obter resultados mais confiáveis para este tipo de previsão, a qual fora testada em dados de Taiwan.

A observação e a previsão de demanda de curtíssimo prazo é de interesse particular a aplicações em *Smart Grids*, onde os programas de *Demand Response* são favorecidos. Em (HSIAO, 2015), os autores buscam realizar uma previsão de demanda para residências de Taiwan. Assim, são analisadas as agendas diárias de uso de equipamentos eletrodomésticos. Esse contexto combina estações do ano, temperatura do ar, dia da semana e tipo de dia. Foram identificados 17 atributos relevantes à categorização de tipos de consumo. Todos os dados foram, a princípio, classificados em *clusters*. Esta abordagem obteve resultados superiores aos métodos de autoregressão como o ARIMA.

Por fim, em (NUNES; CARLI, 2016), os autores fazem uso do modelo ANFIS (*Adaptive Network-based Inference System*) que é aplicado a dados referentes ao consumo de energia em quatro cidades brasileiras, a saber: Mogi-Guaçu, Ubatuba, Votuporanga e Cordeirópolis. As medições são realizadas em intervalos de 5 minutos durante uma semana. Os resultados obtidos, em termos de MAPE, foram: para Cordeirópolis 0,83%, para Mogi-Guaçu 1,45%, para Ubatuba 0,46% e para Votuporanga 1,88%.

2.5 Estudo Bibliométrico

A Bibliometria é um ramo da Ciência da Informação onde são aplicados métodos estatísticos e matemáticos a fim de definir indicadores que venham a demonstrar a evolução da ciência em determinadas áreas do conhecimento. Neste sentido, podem ser obtidos indicadores em função da autoria, país de origem, citações, palavras-chave mais relevantes, dentre outros. Nessa dissertação, o estudo bibliométrico teve o objetivo de comprovar a relevância científica do tema (previsão de demanda de energia elétrica). Para tal finalidade, empregou-se o software R junto ao pacote bibliometrix, sendo este último desenvolvido por (ARIA M.; CUCCURULLO, 2016).

2.6 Definição da Base de Dados e Buscas

A princípio, faz-se necessária a busca e obtenção de significativo número de publicações para que as análises estatísticas sejam devidamente realizadas. Portanto, foram realizadas buscas na base de dados Scopus, onde definiu-se as seguintes palavras-chave: *load forecast*, *load prediction* e *electricity forecast*. Ademais, foram adicionadas às cláusulas de busca, termos como *time-series*. Cabe ainda informar que todas as publicações encontradas foram delimitadas para os anos de 2000 a 2016 e somente produções com tipologia *Article* ou *Surveys* foram

consideradas. Deste modo, artigos publicados em anais de eventos foram desconsiderados.

Baseado nos filtros previamente descritos, a busca retornou 523 publicações. Assim, foi importado um arquivo .bib que é dado como entrada ao pacote bibliometrix.

2.7 Análise Sumarizada

Partindo-se dos 523 artigos, pôde-se notar que estes são provenientes de 222 distintas fontes que apresentaram uma média de 27,8 citações por publicação. As informações dos indicadores gerais podem ser visualizadas por meio da Tabela 2.1.

Tabela 2.1: Indicadores da Análise Bibliométrica.

Artigos	523
Fontes	222
Palavras Chave (Periódicos)	3849
Palavras Chave (Autores)	1405
Período	2000-2016
Média de Citações por publicação	27,8
Autores	1248
Aparecimento de Autores	1618
Artigos de Único Autor	23
Artigos com múltiplos Autores	1225
Artigos por Autor	0,419
Autores por Artigo	2,39
Coautores por artigo	3,09
Índice de Colaboração	2,6

2.8 Análise de Palavras-Chave

O bibliometrix proporciona a realização de um tipo de análise referente às palavras-chave encontradas na busca. Este tipo de análise fornece informações das principais palavras-chave disponíveis e possibilita a expansão ou minimização do escopo de pesquisa. As Tabelas 2.2 e 2.3 são organizadas dividindo as palavras-chave e suas respectivas quantidades de ocorrências, respectivamente por periódicos e por autores. Assim, a Tabela 2.2 trata-se da quantidade de todas em relação às aparições das palavras-chave. Por outro lado, a Tabela 2.3 apresenta a

relação entre as palavras-chave das publicações com os autores.

A observação dos dados mostra uma relevância dos métodos baseados em RNA, surgindo entre as dez palavras-chave mais frequentes, seja por periódicos ou por autores. Esse fato demonstra, inclusive, a popularidade das RNA em publicações na área de previsão de demanda de energia elétrica.

Outra análise relacionada às palavras-chave pode ser feita em função do agrupamento de determinadas palavras, dividindo tais grupos por meio de mapas de densidade, conforme pode ser visualizado nas Figuras 2.2 e 2.3.

Tabela 2.2: Relação das Palavras-Chave por Periódicos.

Time Series	78
Load Forecasting	62
Forecasting	49
Short-Term Load Forecasting	47
Neural Networks	27
Neural Network	23
Artificial Neural Network	21
Time Series Analysis	19
Phase Space Construction	16
Power System	14

Tabela 2.3: Relação das Palavras-Chave por Autores.

Forecasting	166
Electric Load Forecasting	142
Time Series	109
Neural Networks	96
Regression Analysis	62
Time Series Analysis	54
Load Forecasting	53
Mathematical Models	53
Electricity	48
Electric Utilities	38

O mapa de densidade da Figura 2.2 demonstra que as palavras-chave dos maiores agru-

grau de relacionamento entre os dois *clusters*. Assim, é possível afirmar que haveria um forte impacto se apenas um dos *clusters* de palavras-chave fosse escolhido durante as buscas.

Neste caso, notou-se que se fossem excluídos termos de busca referentes à previsão do preço da energia elétrica, isso poderia comprometer os resultados obtidos, pois há publicações que tratam de ambos os temas, ou seja, previsão de demanda e preço da energia elétrica.

2.9 Análise do Número de Publicações

Ao analisar o histórico de publicações no período delimitado entre os anos de 2000 e 2016, notou-se um crescimento significativo do número de publicações nesta área. Em 2000, apenas 11 artigos foram publicados em periódicos e um crescimento real se deu apenas no ano de 2008, com 39 publicações. Entretanto, no ano de 2009, a produção científica nessa área atingiu seu pico (com 62 publicações). Portanto, entre 2008 e 2016, há uma produção científica anual média de 45 artigos, sendo este um número bastante elevado. Visto que a média de publicações anual entre os anos de 2000 e 2007 era de 14 artigos, observa-se que o número de publicações triplicou, conforme mostrado na Figura 2.4.



Figura 2.4: Número de Publicações por Ano.

Baseado no conjunto de dados, também foi possível verificar a produtividade acadêmica por país. A Figura 2.5 ilustra um *ranking* dos países mais produtivos.

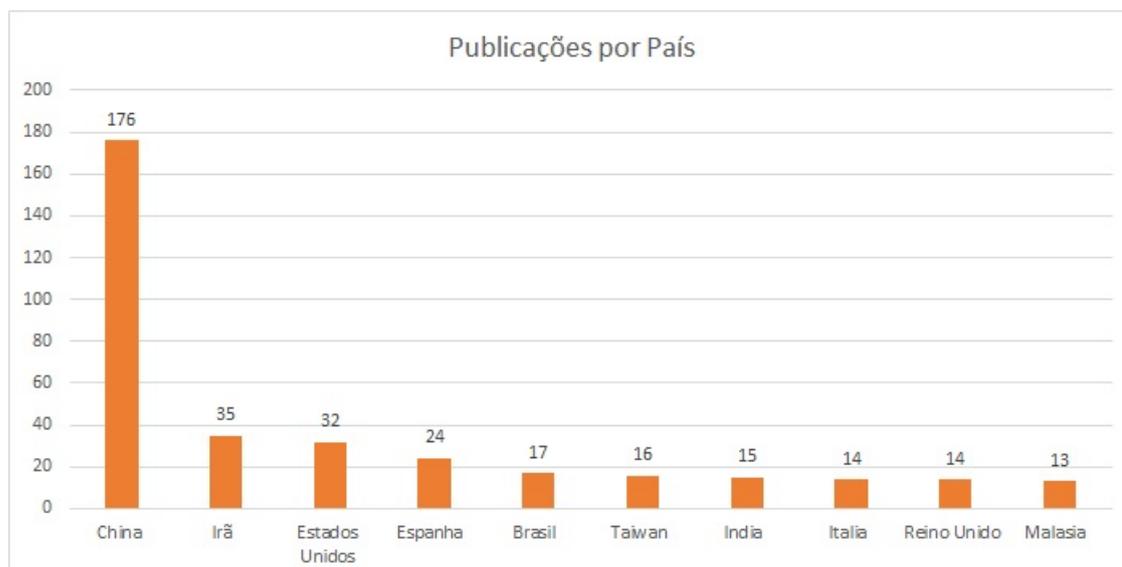


Figura 2.5: Número de Publicações por País.

Conforme o gráfico da Figura 2.5, observa-se que a China tem sido o país mais produtivo, seguida de Irã e Estados Unidos.

Entretanto, ao analisar a quantidade de citações por país, nota-se que Taiwan, seguido de Espanha e China são os três países com produções científicas mais relevantes nessa área, conforme apresentado na Tabela 2.4.

Tabela 2.4: Quantidade de Citações por País.

País	Total de Citações	Média de Citações
Taiwan	2052	128,25
Espanha	2041	85,04
China	1902	10,81
Irã	1246	35,60
Estados Unidos	1221	38,16
Reino Unido	886	63,29
Arábia Saudita	413	68,83
Austrália	390	32,50
Índia	366	24,40
Dinamarca	362	45,25

Ao analisar a Tabela 2.4, nota-se que a China (maior produtor de publicações) tem a pior contribuição científica, visto que sua média de citações é muito baixa. Taiwan, por sua vez,

além de possuir a maior quantidade de citações, também possui a maior média. O Brasil seria o décimo primeiro colocado em termos de número de citações (330 citações), porém, teria uma média de citações de 19,41, ou seja, maior do que a da China.

Por meio desse estudo bibliométrico verificou-se que a constante produção científica em relação à previsão de demanda de energia elétrica, desde 2000, mostra a preocupação que os países têm em obter informação estratégica que venha de encontro com uma gestão eficiente das matrizes energéticas.

Capítulo 3

PROCEDIMENTOS E METODOLOGIA PROPOSTA

Para este trabalho, empregou-se uma metodologia baseada em *clustering* junto a RNAs. Os procedimentos necessários para a execução dessa metodologia encontram-se ilustrado por meio da Figura 3.1.

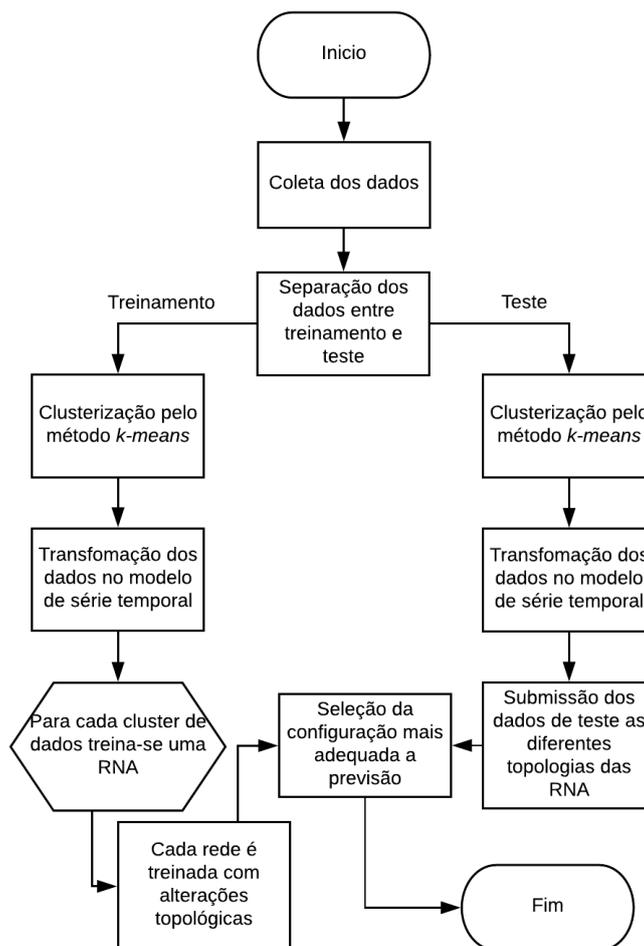


Figura 3.1: Visão geral da metodologia proposta.

Com intuito de melhor apresentar os detalhes de cada um dos blocos da metodologia proposta, estes são demonstrados nas subseções que seguem.

3.1 Dados Coletados para o Experimento

Para o estudo proposto foram coletados dados referentes ao mercado PJM entre os anos de 2014 e 2016. O PJM trata-se de um mercado composto pela transmissão de energia da região leste dos Estados Unidos. Até os anos 2000, o PJM era responsável pela maior quantidade de venda de energia elétrica no mundo, sendo ultrapassada pela *European Integrated Energy Markets*.

A Figura 3.2 mostra toda a área de atuação e de fornecimento de energia do PJM. Sua região de atuação abarca os seguintes estados: Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, e o Distrito de Columbia.

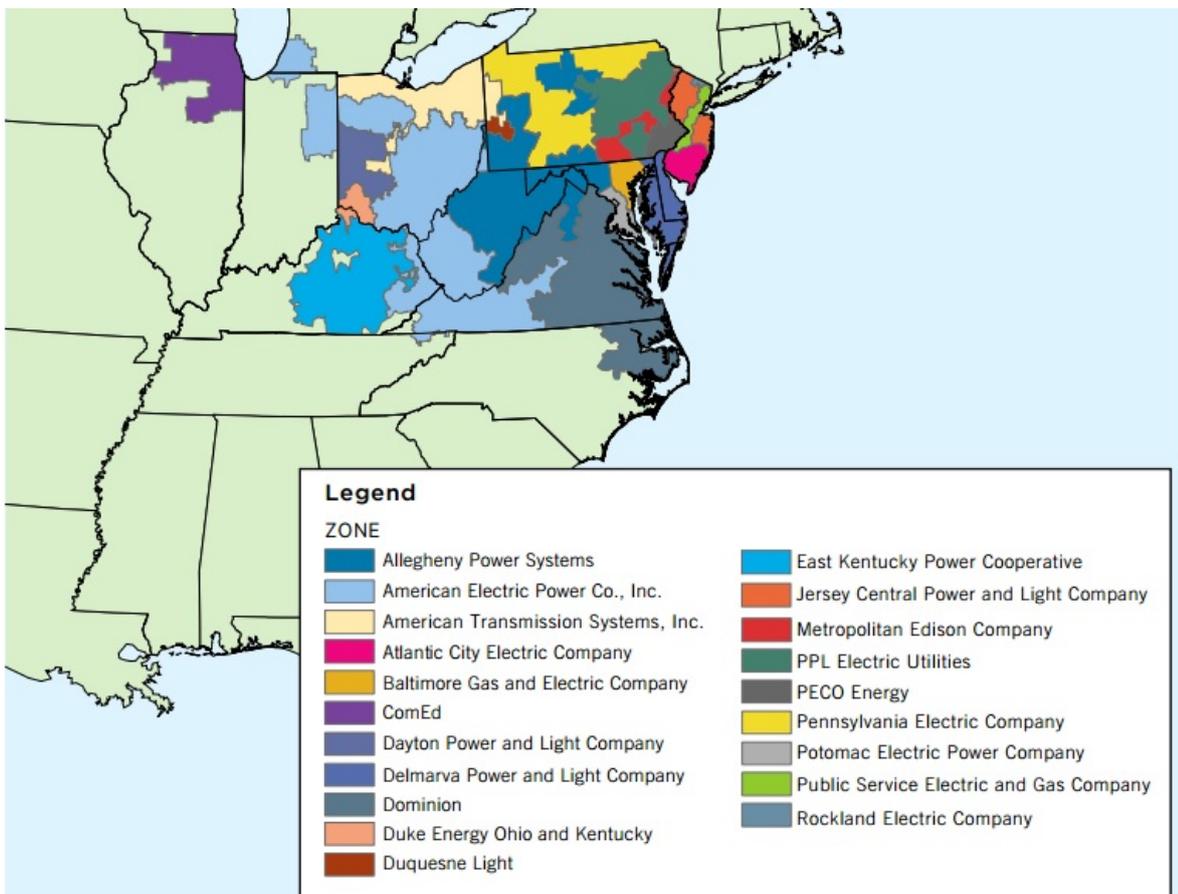


Figura 3.2: Áreas de atuação da PJM. Fonte:(PJM, 2018)

Como supracitado, este estudo tem por objetivo limitar o cálculo da demanda ao estado

americano do Kentucky. Portanto, os dados são separados para conter exclusivamente informações referentes ao mesmo. Neste caso, os dados são fornecidos pela empresa *East Kentucky Power Cooperative* (EKPC). De acordo com a proposta de trabalho, que tem como objetivo relacionar a demanda de energia com as informações climatológicas, a obtenção dos dados climatológicos referentes ao estado do Kentucky foi feita junto à *University of Kentucky*.

Em termos climatológicos, o estado do Kentucky é caracterizado como subtropical. A variância de temperatura apresentada durante o ano registra mínimas de 23 °F (-5 °C) e máximas de 87 °F (31 °C) em sua área total de 104.659 km², sendo pouco maior que o estado de Pernambuco. Devido a seu clima subtropical, o estado tem as quatro estações do ano bem definidas.

Mais especificamente, a área de atuação da EKPC encontra-se ilustrada na Figura 3.3

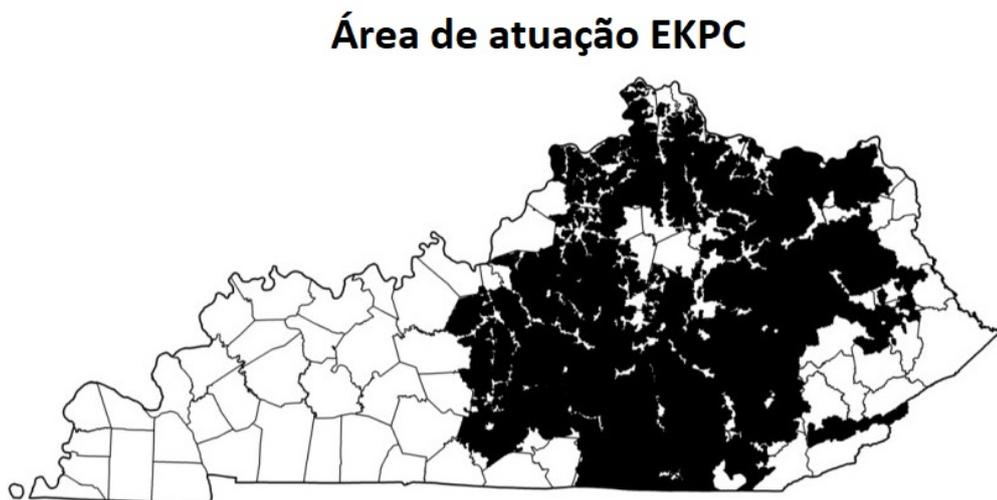


Figura 3.3: Áreas de atuação da EKPC. Fonte: Kentucky Energy Profile 2015.

Os Estados Unidos têm uma matriz energética onde cerca de 39% de toda energia gerada é proveniente de termoelétricas, que utilizam carvão, como principais provedoras de energia elétrica para o país. O estado do Kentucky tem uma porção ainda mais significativa de geração por meio de termoelétricas, conforme mostrado no gráfico da Figura 3.4.

A porção de utilização de geração de energia pelo uso de carvão é de cerca de 92%. Tratando-se de um recurso não renovável, a previsão de demanda por energia elétrica é ainda mais vantajosa, pois uma precisão de alta confiabilidade resulta na economia por recursos e, conseqüentemente, em preços menores para a energia elétrica. O perfil da matriz energética do estado em questão também fornece informações que indicam que a maioria dos consumidores residenciais do estado contêm aquecimento por meio de energia elétrica.

Geração de Energia no Kentucky por Combustível, 2014

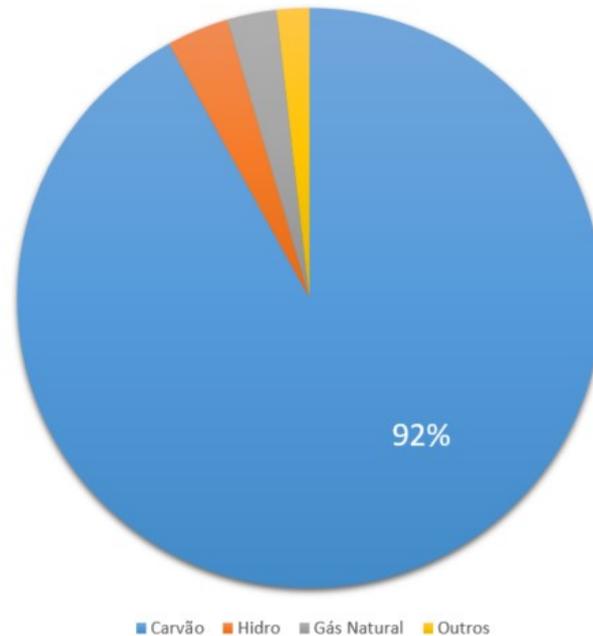


Figura 3.4: Perfil da geração de energia no estado do Kentucky. Fonte: Kentucky Energy Profile (2015).

3.2 Clusterização

A clusterização (ROKACH; MAIMON, 2005) (*clustering*) trata-se de uma técnica para agrupamento automático de dados, que se baseia em graus de semelhança ou diferença, onde este grau faz parte da definição do problema. Partindo deste ponto, os agrupamentos dos dados coletados de demanda de energia elétrica e a respectiva temperatura são gerados pela técnica denominada *k-Means* (MACQUEEN, 1967). Assim, empregou-se a distância euclidiana para avaliar as distâncias de pontos formados pela temperatura e demanda em relação aos centróides (ou seja, os centros dos *clusters* obtidos pelo *k-Means*). Este cálculo é demonstrado por meio da Equação 3.1:

$$d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}. \quad (3.1)$$

Esta técnica faz a separação de grupos baseada na observação de todas as n amostras dos dados. Assim, é estabelecido um valor k que representa o número de *clusters*. Na sequência, deve inicializar a posição espacial dos centros desses *clusters*, chamados de centróides. Na sequência, calcula-se as distâncias entre cada uma das amostras e os centróides, de forma que a amostra é incluída no *cluster* de menor distância. Após atribuir todas as amostras aos *clusters*,

deve-se recalcular a posição espacial dos centróides com base na média das amostras pertencentes ao respectivo *cluster*. Esse processo é repetido de forma iterativa, onde a convergência ocorre ao estabilizar as posições dos centróides. Ressalta-se que os métodos de clusterização podem abranger até n dimensões possíveis. No escopo deste trabalho, a utilização de *clusters* foi realizada em função das dimensões expressas pela temperatura do ar e da demanda registrada em determinado horário.

A utilização de agrupamentos se dá pelo fato de a temperatura e a demanda de energia elétrica apresentarem elevados índices de correlação linear. Portanto, os agrupamentos podem indicar um incremento nas incidências de pares de dados que possuam maior correlação com a variável a ser predita. Além disso, o agrupamento dos dados poderá contribuir para o aprendizado do modelo neural.

Dessa forma, calculou-se a correlação de Pearson (Equação 3.2) para que a análise fosse realizada. De acordo com a Tabela 3.1 é possível verificar uma correlação positiva entre as demandas de 4 horas anteriores à demanda a ser prevista. Em contrapartida, nota-se uma correlação negativa para todas as temperaturas de 4 horas prévias à demanda a ser prevista. Isto implica que, de acordo com as definições de correlação linear, quanto mais próximo o valor da correlação de 1, mais forte será a correlação entre as variáveis.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (3.2)$$

Tabela 3.1: Correlação de Pearson obtida entre os valores de temperatura e de demanda da energia elétrica (de 4 horas anteriores) em relação à variável a ser predita (demanda).

D-4	T-4	D-3	T-3	D-2	T-2	D-1	T-1
0.8030	-0.4585	0.8733	-0.45279	0.9353	-0.4509	0.9805	-0.4505

A Tabela 3.1 ainda demonstra que os elementos mais próximos do valor alvo da previsão têm maior correlação, justificando assim a abordagem de agrupamentos proposta.

Ainda é importante mencionar que a correlação de Spearman também pode ser aplicada. Dessa forma, a mesma é apresentada na Tabela 3.2.

Tabela 3.2: Correlação de Spearman obtida entre os valores de temperatura e de demanda da energia elétrica (de 4 horas anteriores) em relação à variável a ser predita (demanda).

D-4	T-4	D-3	T-3	D-2	T-2	D-1	T-1
0.73696	-0.26690	0.82908	-0.25208	0.91139	-0.24376	0.97278	-0.24056

Os resultados da Tabela 3.2 confirmam a análise já realizada por meio da correlação de Pearson.

3.3 Organização dos Dados Processados

Conforme previamente mencionado, os dados são referentes aos anos de 2014 a 2016. Assim, foi realizada a divisão de 66% desses dados para treinamento da rede neural, enquanto o restante fora separado para a realização dos testes. Esta divisão de dados é a mesma para todos os métodos discutidos neste trabalho. Tratando-se do escopo dos anos de 2014 até 2016, esta divisão proporciona testar a metodologia proposta com o ano de 2016.

Ao analisar graficamente os dados de demanda e temperatura, nota-se uma tendência cíclica, conforme pode ser visualizado nas Figuras 3.5 e 3.6 que correspondem, respectivamente, à temperatura e à demanda dos anos de 2014 a 2016.

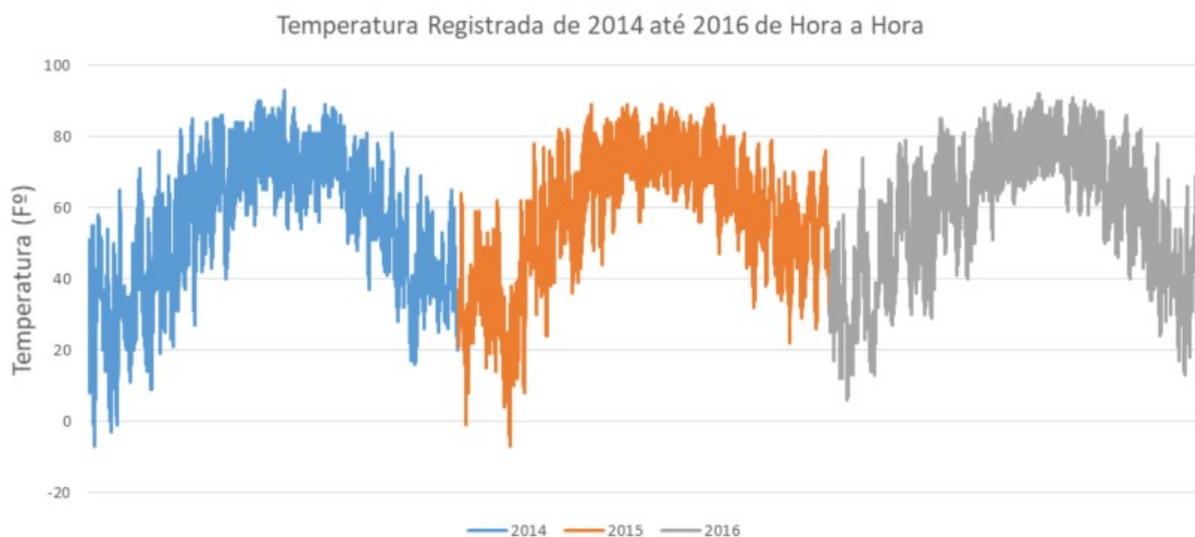


Figura 3.5: Temperatura média por dia da região do estudo referente aos anos de 2014 a 2016.

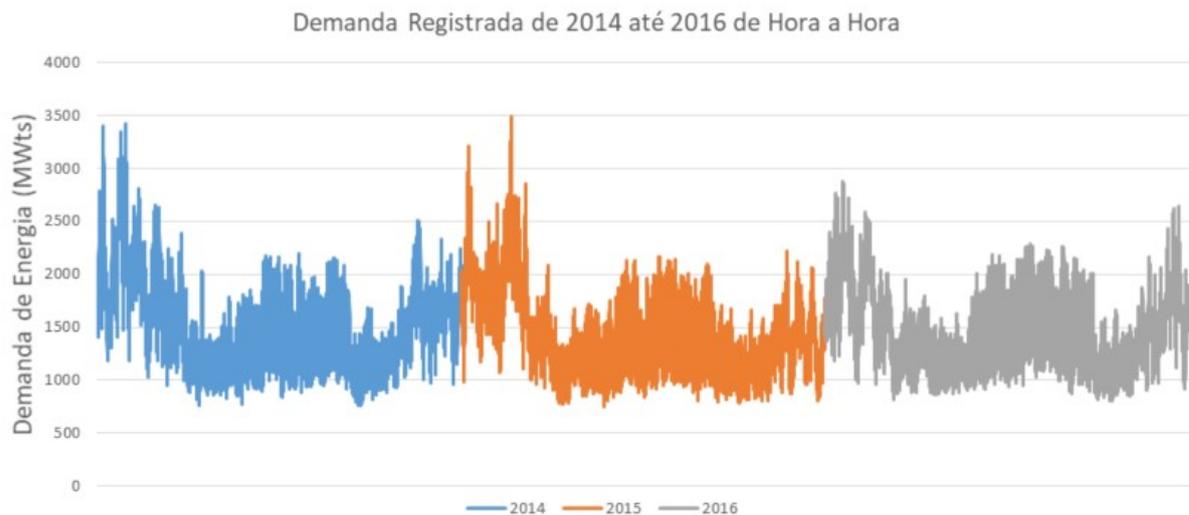


Figura 3.6: Demanda de energia elétrica da EKPC referente aos anos de 2014 a 2016.

Ainda com relação aos gráficos de temperatura e demanda, nota-se que há certos picos de utilização de energia em conjunto com extremos de temperatura. Contudo, como supracitado, a temperatura não é exclusivamente responsável pela demanda de energia elétrica.

Mediante o fornecimento dos dados de demanda pela EPKC e de temperatura pela *University of Kentucky*, faz-se necessário agrupá-los hora a hora. As estações climatológicas foram escolhidas com base nos seguintes critérios: (i) menor quantidade de dados faltantes; e (ii) posição geográfica da mesma.

A organização dos dados para ambas as escalas se dá de forma semelhante, onde os dados são agrupados em um modelo de série temporal para servirem de treinamento e teste da RNA. Basicamente, pode-se dizer que a entrada de dados corresponde a uma janela ($d-n$ para a demanda e $t-n$ para a temperatura) das séries de demanda e de temperatura, onde a saída do modelo durante o processo de treinamento é representada pela demanda d . Neste estudo, considerou-se as últimas dez horas, $n = 10$, registradas para a temperatura do ar e da demanda da região analisada. A Figura 3.7 demonstra as entradas do modelo neural.

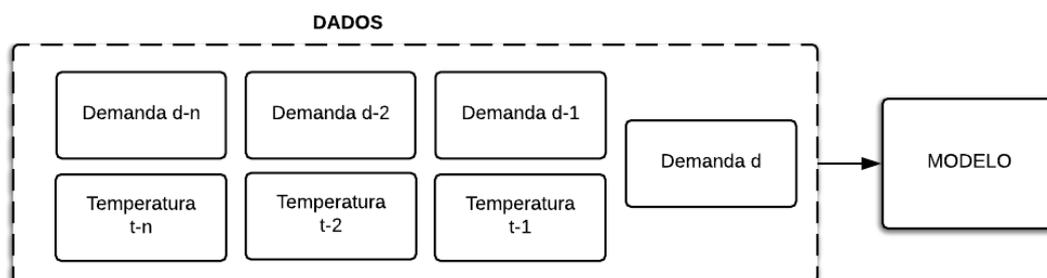


Figura 3.7: Entrada dos dados no modelo neural.

Para a utilização dos mesmos dados referentes à separação em *clusters*, as entradas seguem a mesma proporção em quantidade de atributos (ou seja, as últimas 10 horas), onde a diferença é marcada pelos atributos serem todos condizentes aos *clusters* das últimas horas. Essa formatação de entradas baseada em *clusters* pode ser visualizada na Figura 3.8.

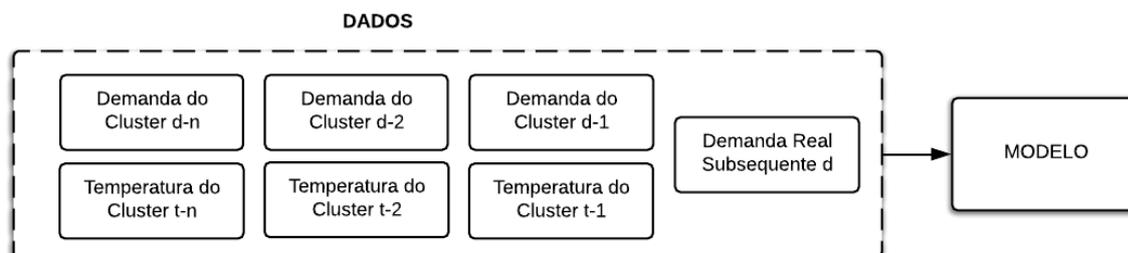


Figura 3.8: Entrada dos dados no modelo neural com a utilização de *clusters*

Dessa forma, optou-se por gerar 4 *clusters* que confrontam temperatura e demanda, os quais serão responsáveis por agrupar as amostras nesses *clusters* que, por consequência, corresponderão a novas séries temporais, porém com descontinuidade nos dados. Sabe-se que gerar descontinuidade nos dados de séries temporais não é uma prática comum, visto que os métodos clássicos baseados em ARMA, ARIMA e GARCH evidentemente irão apresentar erros elevados. Sendo este o motivo pelo qual optou-se por escolher como modelo de previsão as Redes Neurais Artificiais.

3.4 Modelo Proposto Baseado em Redes Neurais Artificiais

As Redes Neurais Artificiais têm sido corriqueiramente empregadas na previsão de séries temporais. Conforme (HAYKIN, 1998), uma rede neural representa um massivo processador paralelo composto de unidades de processamento simples, que lembra o cérebro em dois aspectos:

1. O conhecimento é obtido pela rede por um processo de aprendizado;
2. Conexões entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

Entretanto, em uma RNA, o aprendizado é realizado por meio de um algoritmo encarregado de ajustar os pesos sinápticos de modo que a rede atinja o objetivo para a qual foi designada. Nessa dissertação, empregou-se uma RNA do tipo *Multilayer Perceptron*, com algoritmo de treinamento de Levenberg-Marquardt (HAGAN; MENHAJ, 1994).

Com o intuito de realizar observações sobre os resultados obtidos pelas RNAs propostas, os experimentos foram realizados por meio de distintas topologias. Neste sentido, variou-se a quantidade de neurônios entre 5 e 50. Esse procedimento fora realizado para a abordagem que emprega os *clusters* como entradas e também para a abordagem sem *clusters*.

A arquitetura das RNAs utilizadas está ilustrada na Figura 3.9, que ainda indica as funções de ativação específicas de cada uma das camadas neurais. A camada neural escondida utiliza a função de ativação Tangente Sigmoidal (que atua em um intervalo de valores entre -1 e 1), enquanto a camada neural de saída utiliza uma função linear, a qual é bastante comum em redes com arquitetura *feedforward*.

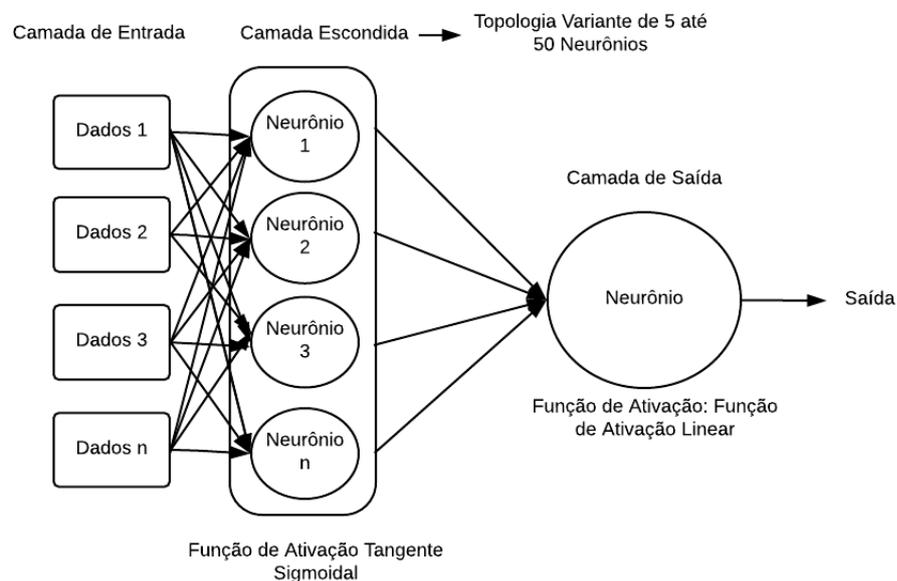


Figura 3.9: Modelo da rede neural proposta.

3.5 Métricas de Avaliação de Desempenho

Objetivando a avaliação dos resultados obtidos nos experimentos, fora utilizado os seguintes métodos de cálculo de erro: RMSE (*Root Mean Squared Error*); MAE (*Mean Absolute Error*); e MAPE (*Mean Absolute Percentage Error*). Estes cálculos são apresentados, respectivamente, por meio das Equações 3.3, 3.4 e 3.5:

$$RMSE = \sqrt{\frac{1}{|T|} \sum_{(u,i)} (\hat{r}_{ui} - r_{ui})^2}, \quad (3.3)$$

$$MAE = \frac{1}{|T|} \sum_{(u,i)} |\hat{r}_{ui} - r_{ui}|, \quad (3.4)$$

$$MAPE = \frac{1}{|T|} \sum_{(u,i)} \left| \frac{\hat{r}_{ui} - r_{ui}}{r_{ui}} \right|. \quad (3.5)$$

onde \hat{r}_{ui} representa o valor previsto e r_{ui} é o valor real e conhecido para os dados de validação.

O RMSE representa a raiz quadrada do Erro Quadrático Médio (MSE, do inglês *Mean Square Error*). Esse confere uma maior sensibilidade do modelo a grandes erros. Além disso, a raiz garante que o RMSE se mantenha na mesma dimensão da variável a ser prevista. O MAE mostra a variação média da diferença entre os valores previstos e reais da série temporal, enquanto o MAPE reflete diretamente a precisão da previsão obtida, mostrando o erro médio em proporção aos valores reais.

Capítulo 4

RESULTADOS E DISCUSSÕES

Este capítulo visa apresentar os resultados obtidos com a metodologia proposta. Assim, a princípio, são mostrados e discutidos os *clusters* gerados. Em seguida, são apresentados os resultados obtidos pelos modelos de predição baseados em Redes Neurais Artificiais. Por fim, realiza-se uma comparação entre resultados obtidos com e sem a utilização do método de clusterização das entradas. Cabe ressaltar que esta comparação se faz apenas em relação ao horizonte de curto prazo (demanda de energia elétrica diária).

4.1 *Clusters* obtidos entre demanda e temperatura

Conforme previamente mencionado, optou-se por estabelecer 4 *clusters* de dados, os quais foram gerados pelo algoritmo *k-Means*. Neste sentido, foi possível observar o comportamento apresentado na Figura 4.1 em relação à clusterização.

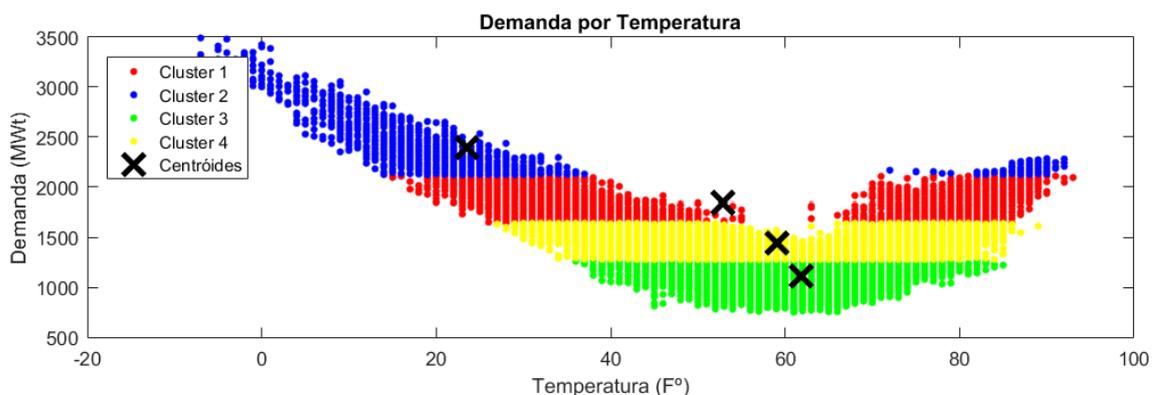


Figura 4.1: Agrupamentos de temperatura e demanda por hora.

Nota-se uma maior demanda de energia, demonstrada principalmente pelos agrupamentos de cor azul e vermelho, que são referentes aos extremos de temperatura, estes representam

também maior consumo de energia, isto possivelmente decorrente da utilização de equipamentos para conforto térmico (aquecedores e condicionadores de ar). Portanto, ambos os *clusters* mostram-se condizentes com a observação feita junto aos gráficos das Figuras 3.5 e 3.6.

Cabe observar a demanda mínima para a região sob estudo, onde os agrupamentos de cor verde e amarelo, na Figura 4.1, indicam o comportamento habitual da região em fatores de utilização de energia elétrica. Estes agrupamentos representam aproximadamente 71% de todos os dados utilizados, conforme mostrado na Tabela 4.1.

Tabela 4.1: Tabela de Distribuição nos Agrupamentos

Cluster	Quantidade
1	23,06%
2	6,1%
3	39,96%
4	30,78%

Ainda é possível notar que o *cluster 2* possui uma maior presença de *outliers*. Verifica-se também a que os *clusters 3 e 4* podem representar a necessidade básica de energia elétrica da região sem a interferência da temperatura, corroborando com a correlação negativa apresentada na Seção 3.2.

4.2 Resultados obtidos pelos modelos neurais

Após a clusterização dos dados, os modelos neurais foram treinados e testados. Durante a etapa de treinamento, a quantidade de neurônios da camada oculta foi variado a fim de obter um quadro comparativo. Portanto, o número de neurônios foi variado de forma incremental, de 5 em 5 neurônios até atingir 50.

4.2.1 Modelo Neural com entradas não clusterizadas

Os resultados obtidos para o modelo neural sem as entradas de dados clusterizadas são apresentados na Tabela 4.2, em termos de MAE, MAPE e RMSE, onde considera-se o ano de 2016 como dados de teste.

Com base nesses resultados, nota-se que o MAPE mostrou-se bastante estável, independente da topologia, visto que sua variação ficou entre 2,48% e 2,25%. Ademais, a melhor

configuração obtida, foi denotada por uma topologia com 25 neurônios na camada oculta. Dessa forma, analisou-se graficamente (Figura 4.2) uma janela menor dentro do conjunto de dados de teste, segunda semana do mês de janeiro de 2016.

Tabela 4.2: Resultados obtidos referentes à aplicação do modelo neural sem clusterização das entradas e com variação topológica.

Neurônios	MAE	MAPE(%)	RMSE
5	35,02	2,48	45,08
10	32,83	2,32	42,45
15	32,83	2,32	42,3
20	32,32	2,28	41,78
25	31,92	2,25	41,45
30	32,12	2,27	41,70
35	32,14	2,27	41,58
40	32,40	2,28	41,96
45	32,52	2,30	41,97
50	32,58	2,29	41,96

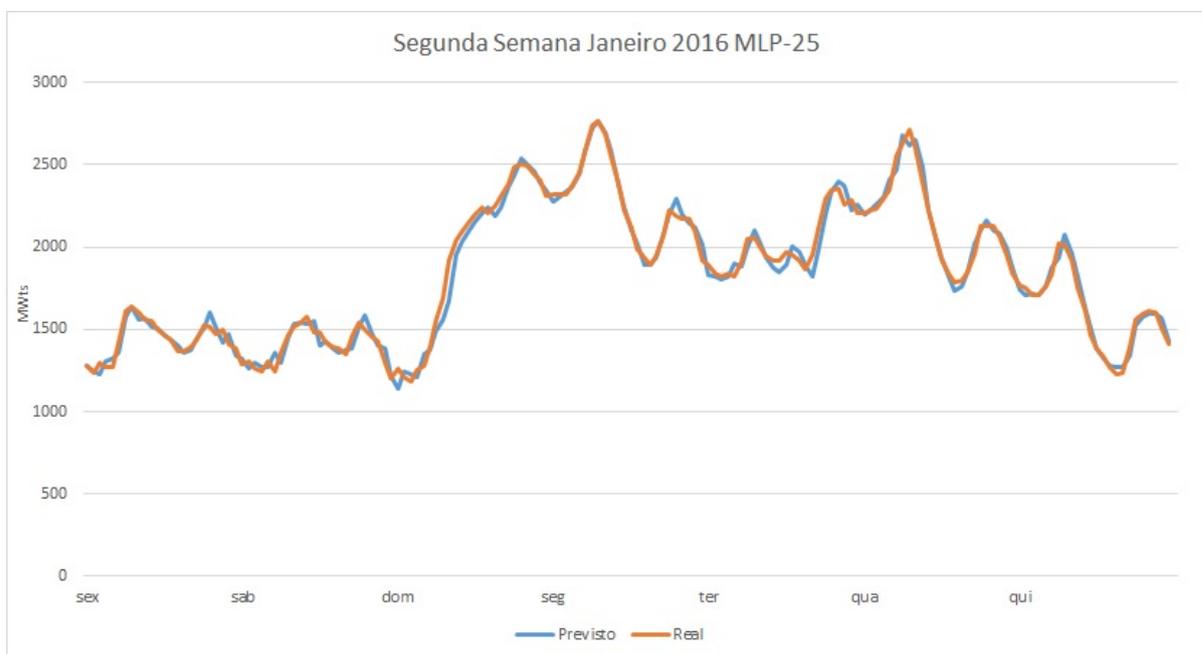


Figura 4.2: Resultados obtidos pela MLP com topologia de 25 neurônios na camada oculta.

O isolamento de um período de tempo se faz necessário para obter a devida visualização do resultado alcançado, onde o eixo vertical representa a demanda de energia elétrica necessária

para suprir a região de concessão da EKPC (em MW) e o eixo horizontal indica os dias da semana.

Visando analisar a robustez da melhor topologia entrada para a MLP (com 25 neurônios), os treinamentos e testes foram repetidos 25 vezes, com o intuito de averiguar os erros absolutos encontrados. No entanto, para aprofundar estas análises, os erros absolutos foram verificados mês a mês, conforme mostrado na Figura 4.3.

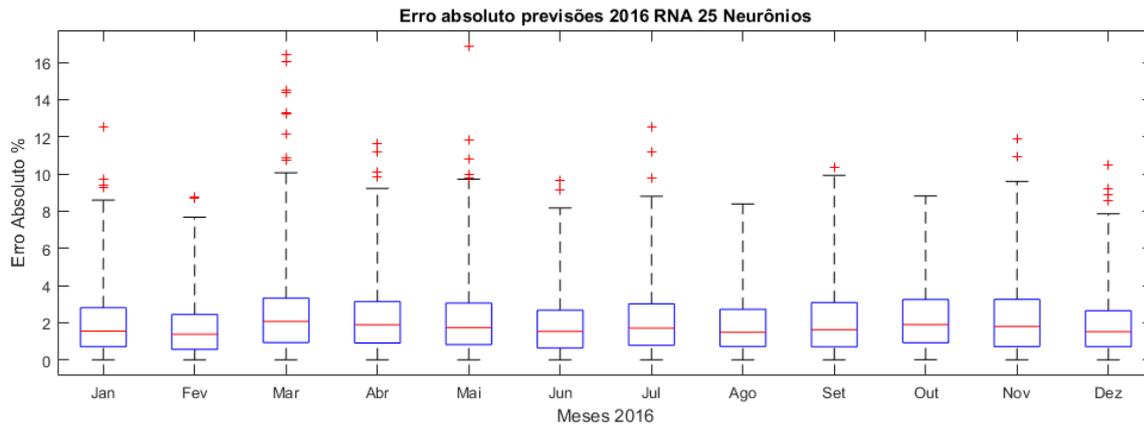


Figura 4.3: Erro absoluto percentual apresentado após 25 repetições de treinamento e teste ao longo dos meses de 2016 (RNA sem entradas clusterizadas).

A Figura 4.3 mostra a presença de *outliers* em algumas previsões realizadas pelo método, onde o mês de Maio apresentou o erro máximo de previsão, com 16,88% de erro absoluto percentual. Em contrapartida, o erro mínimo obtido foi de 0,00019%.

4.2.2 Modelo Neural com entradas clusterizadas

O mesmo procedimento previamente realizado, fora repetido utilizando-se do modelo neural com entradas clusterizadas. Os resultados obtidos podem ser verificados por meio da Tabela 4.3.

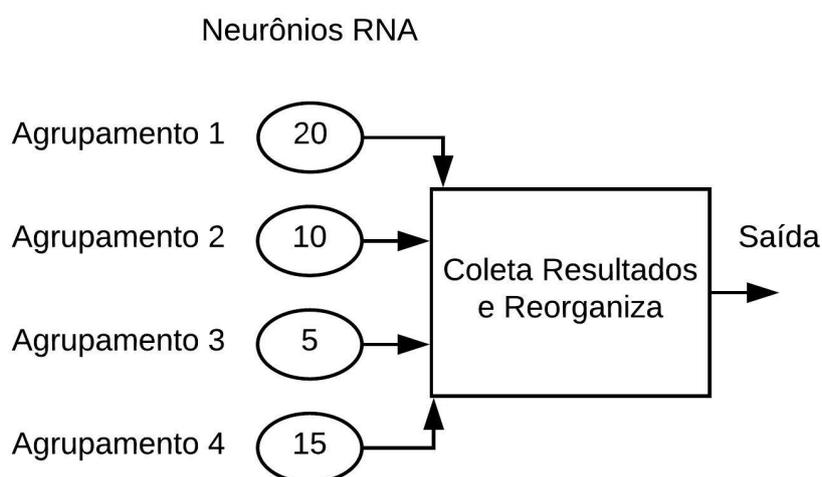
Ao observar a Tabela 4.3, nota-se que o MAPE do *cluster 1* variou entre 2,62% e 2,89%; para o *cluster 2* essa variação ficou entre 2,24% e 4,12%; o *cluster 3* apresentou uma variação entre 2,99% e 3,21%; e o *cluster 4* apresentou variação entre 3,28% e 3,47%.

Assim, adotou-se como melhor configuração: para o *cluster 1* a MLP com 20 neurônios; para o *cluster 2* a MLP com 10 neurônios; para o *cluster 3* a MLP com 5 neurônios; e para o *cluster 4* a MLP com 15 neurônios. Dessa forma, o MAPE médio entre as configurações escolhidas para cada *cluster* é de 2,77%.

Tabela 4.3: Resultados obtidos referentes à aplicação do modelo neural com clusterização das entradas e com variação topológica.

Neurônios	MAPE C1	MAPE C2	MAPE C3	MAPE C4
5	2,68	2,47	2,99	3,29
10	2,65	2,24	3,00	3,33
15	2,73	2,93	3,07	3,24
20	2,62	2,85	3,00	3,35
25	2,70	3,85	3,06	3,29
30	2,74	2,85	3,04	3,28
35	2,76	2,55	3,13	3,46
40	2,87	4,12	3,12	3,41
45	2,82	3,28	3,01	3,47
50	2,89	3,44	3,21	3,44

Mediante os resultados obtidos, foram selecionadas as melhores RNAs, ou seja, aquelas que representam cada *cluster*. Essa configuração pode ser visualizadas por meio da Figura 4.4.

**Figura 4.4: Diagrama da configuração de *clusters*.**

Assim, os mesmos procedimentos apresentados para o modelo neural sem clusterização das entradas foram replicados à avaliação dessa melhor configuração de RNAs para entradas clusterizadas. Neste sentido, na sequência, é mostrada a janela que representa a segunda semana do mês de janeiro de 2016.

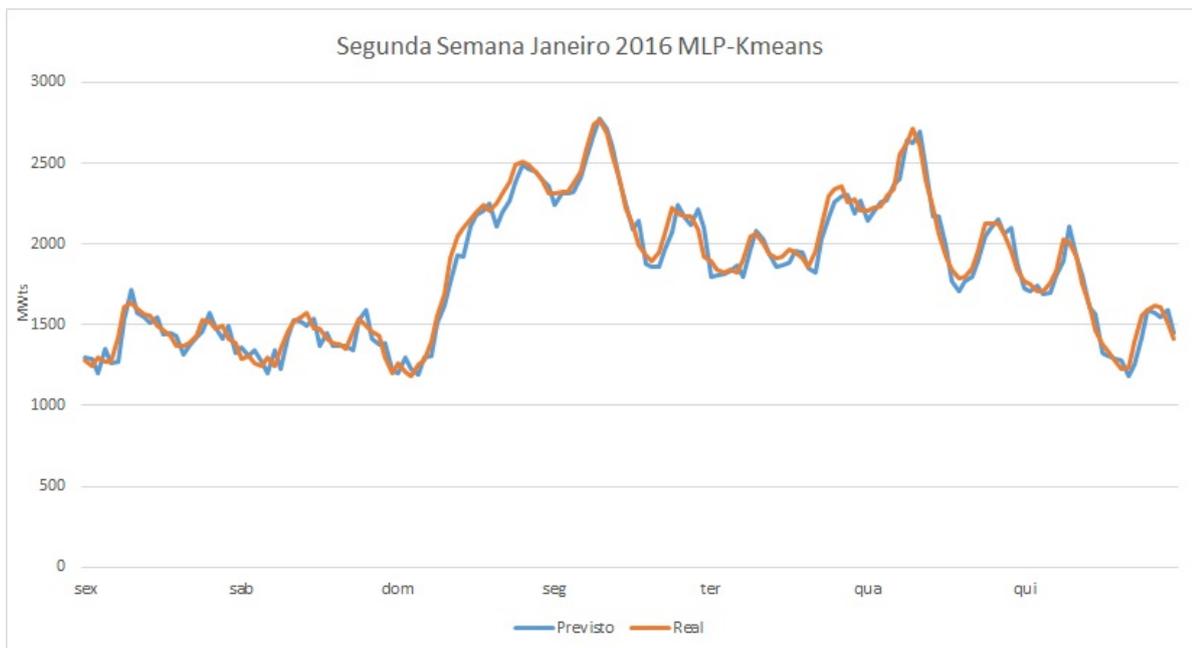


Figura 4.5: Resultados obtidos pela MLP com a utilização de topologia composta.

Após a análise gráfica, ainda fora verificada a robustez da melhor configuração do sistema composto por quatro RNAs que fazem a previsão em função dos *clusters*. Novamente, essa análise fora realizada mediante a execução de 25 treinamentos e testes. Os resultados obtidos podem ser visualizados por meio de Box Plots apresentados na Figura 4.5.

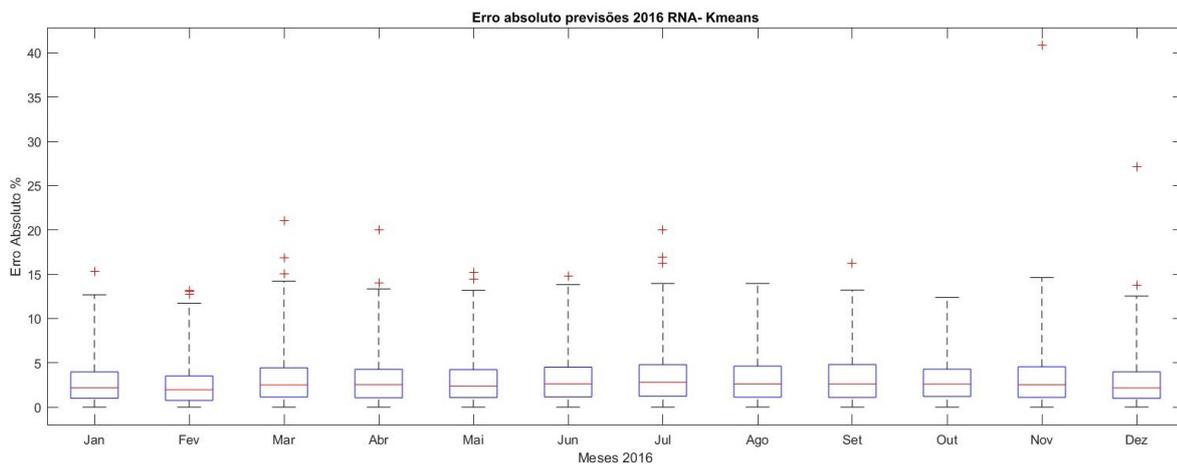


Figura 4.6: Percentual do erro absoluto apresentado após 25 repetições de treinamento e teste ao longo dos meses de 2016 (RNAs com entradas clusterizadas).

Ao contrastar os gráficos das Figuras 4.6 e 4.3 é possível observar a ocorrência de picos significativos nos meses de novembro e dezembro. O máximo erro absoluto percentual de previsão fora de 40.82%, enquanto o mínimo foi de 0.000026%, inclusive, menor do que aquele obtido para a RNA sem clusterização das entradas.

A fim de enriquecer a demonstração dos resultados, são apresentados histogramas de erro obtidos para ambos os métodos desenvolvidos nesta dissertação. A Figura 4.7 representa o histograma do erro obtida pela utilização da RNA sem entradas clusterizadas e a Figura 4.8 demonstra os erros referentes à utilização de clusterização nos dados de entrada.

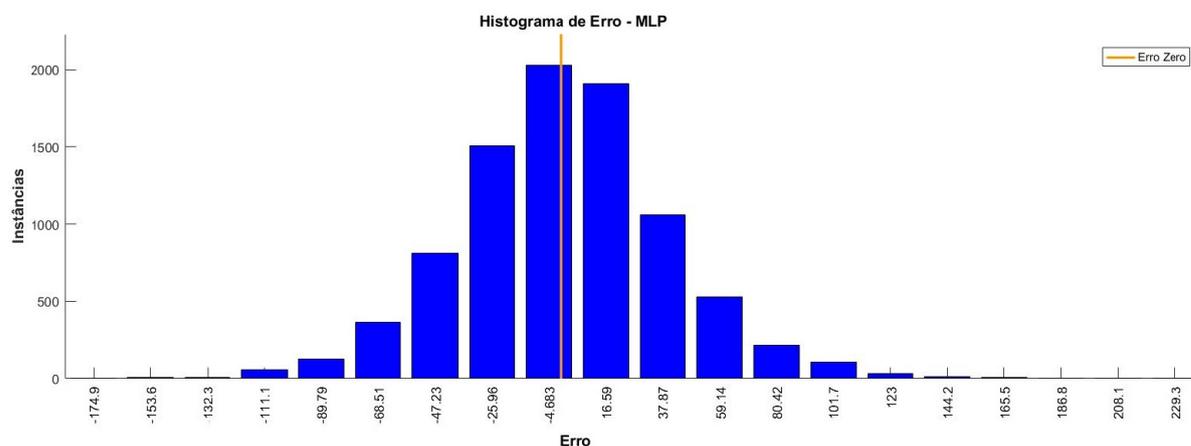


Figura 4.7: Histograma dos erros apresentados pelos resultados obtidos sem a separação de dados.

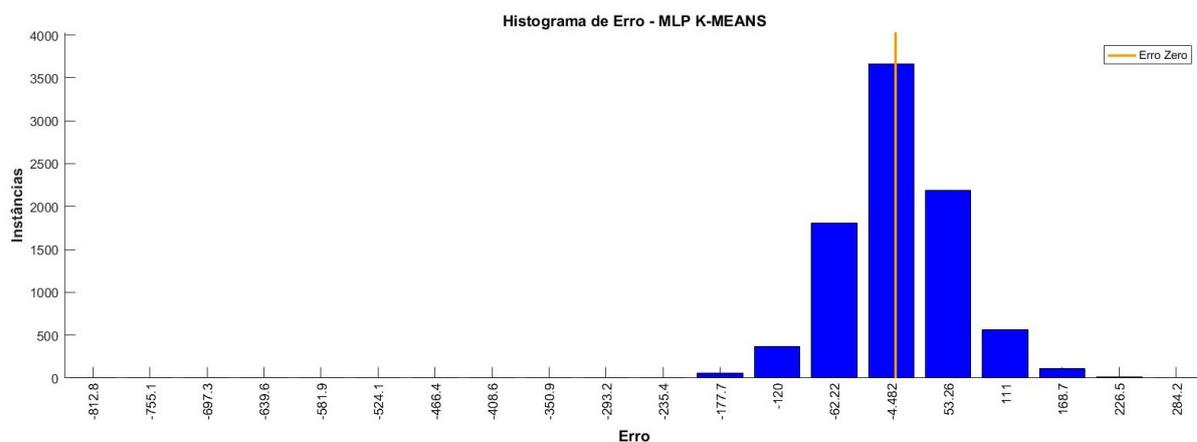


Figura 4.8: Histograma dos erros apresentados pelos resultados obtidos com a separação de dados.

Analisando ambos os gráficos, comparativamente, nota-se uma incidência maior de instâncias próximas ao erro 0 (zero) para a metodologia proposta baseada em clusterização dos dados.

4.3 Discussões

Esta sessão apresenta a comparação entre os resultados obtidos com alguns trabalhos relacionados na literatura correlata. Ressalta-se a utilização de diferentes bases de dados, assim como a diferença entre os períodos de tempo utilizados por cada autor. Conforme previamente

mencionado, este trabalho fez uso de dados referentes ao conglomerado PJM, especificamente da concessionária EKPC no estado americano do Kentucky. Essa análise é apresentada por meio da Tabela 4.4.

Tabela 4.4: Sumarização de resultados obtidos pela literatura correlata.

Publicação	Método e Resultado	Dados
(KOUHI; KEYNIA; Najafi Ravanagh, 2014)	Usa um método de seleção caótica de atributos relevantes junto com uma MLP e otimização por enxame de partículas. Obteve MAPE entre 2,62 e 1,72 para os dias separados do teste.	Utiliza dados de temperatura para a região de New England. Porém, apesar de utilizar dados da PJM, para tal mercado, não emprega a temperatura. Todos os dados são referentes ao ano de 2006.
(SAHAY; TRIPATHI, 2014)	Utiliza uma rede neural do tipo MLP. Obteve MAPE de 3,14% para o PJM e 1,59% para dados referentes à região de New England	Utiliza dados de temperatura para a região de New England, porém não usa dados de temperatura para PJM. Os dados são referentes a 2012.
Metodologia Proposta	Esta dissertação utiliza RNAs do tipo MLP com o algoritmo de treinamento de Levenberg-Marquadt, com e sem a clusterização de dados pelo método de <i>k-Means</i> . Obteve MAPES de 2,25% (sem a utilização de <i>clusters</i>) e 3,06% (com a utilização de <i>clusters</i>)	Faz uso de dados referentes à EKPC que está contida no conglomerado do PJM. Os dados são referentes aos anos de 2014 a 2016.

Apesar de apresentar resultados ligeiramente inferiores, a técnica baseada na utilização de *clusters* pode apresentar uma vantagem estratégica para o objetivo da previsão de demanda de energia elétrica, demonstrada pelo histograma dos erros da Figura 4.8, visto que grande parte das instâncias estão concentradas na redondeza do erro 0 (zero).

A análise comparativa é dificultada pela utilização de uma concessionária específica do conglomerado PJM. Entretanto, para garantir um melhor critério de comparação, foram seleti-

onados apenas artigos que utilizaram dados do PJM e fizeram previsões de demanda de energia elétrica de curto prazo, especificamente no escopo de uma hora adiante.

Capítulo 5

CONCLUSÕES

A proposta dessa dissertação proporcionou o estudo de conceitos referentes a métodos de aprendizado de máquina aplicados à tarefa de previsão de séries temporais, mais especificamente no escopo da demanda de energia elétrica. Além disso, um dos objetivos foi coleta e limpeza da base de dados, bem como o pré-processamento por meio de clusterização.

Assim, esta pesquisa contribui com uma metodologia que agrega a clusterização dos dados a modelos de regressão baseados em Redes Neurais Artificiais com arquitetura *feedforward*. A partir da clusterização dos dados, utilizando o algoritmo *k-Means*, mostrou-se eficaz principalmente na análise de comportamento da série temporal, corroborando inclusive com as análises de correlação. Além disso, ao empregar Redes Neurais do tipo MLP com algoritmo de aprendizado de Levenberg-Marquardt, notou-se que tal modelo é eficiente e competitivo em relação a outros modelos de regressão propostos na literatura correlata. Apesar desses resultados, ainda acredita-se que outras técnicas de pré-processamento e extração de características possam contribuir para aprimorar o modelo. Tais melhorias devem ser constantemente buscadas, visto a previsão de demanda de energia elétrica é estratégica e essencial ao desenvolvimento de qualquer país, a qual irá garantir a economia de recursos e, conseqüentemente, uma redução de custos de geração.

Outra contribuição significativa foi disponibilização de uma nova base de dados e de scripts para as pesquisas realizadas no Laboratório de Inteligência Artificial Aplicada (LIAA).

Este trabalho também proporcionou perspectivas de continuidade e elaboração de trabalhos futuros. Como comprovado por meio da análise bibliométrica, a pesquisa ainda é relevante cientificamente. A proposta da identificação de distúrbios nas séries temporais, visando a redução de *outliers* nas previsões, bem como uma análise detalhada de distintas variáveis de entrada poderão trazer avanços para a área.

REFERÊNCIAS

- AL-SABA, T.; EL-AMIN, I. Artificial neural networks as applied to long-term demand forecasting. v. 13, n. October 1998, p. 189–197, 1999.
- AMJADY, N.; HEMMATI, M. Energy price forecasting: Problems and proposals for such predictions. *IEEE Power and Energy Magazine*, v. 4, n. 2, p. 20–29, 2006. ISSN 15407977.
- AMJADY, N.; KEYNIA, F. Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. *Energy Conversion and Management*, v. 50, n. 1, p. 2976–2982, 2009. ISSN 0885-8950.
- ARIA M.; CUCCURULLO, C. *bibliometrix: A R tool for comprehensive bibliometric analysis of scientific literature*. 2016. Disponível em: <<http://www.bibliometrix.org>>.
- BOROOJENI, K. G. et al. A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon. *Electric Power Systems Research*, Elsevier Ltd, v. 142, p. 58–73, 2017. ISSN 03787796.
- ÇEVIK, H. H.; ÇUNKAŞ, M. A Fuzzy Logic Based Short Term Load Forecast for the Holidays. v. 6, n. 1, p. 57–61, 2016.
- CHEN, Y. et al. Short-Term Load Forecasting : Similar Day-Based Wavelet Neural Networks. v. 25, n. 1, p. 322–330, 2010.
- DUDEK, G. Neural networks for pattern-based short-term load forecasting: A comparative study. *Neurocomputing*, Elsevier, v. 205, p. 64–74, 2016. ISSN 09252312.
- El Desouky, A. A.; El Kateb, M. M. Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA. In: *IEE Proceedings: Generation, Transmission and Distribution*. [S.l.: s.n.], 2000. v. 147, n. 4, p. 213–217. ISSN 13502360.
- FARD, A. K.; AKBARI-ZADEH, M.-R. A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting. *Journal of Experimental and Theoretical Artificial Intelligence*, Taylor and Francis Ltd., v. 26, n. 2, p. 167–182, 2014. ISSN 0952813X.
- GARCIA-MARTOS, C.; RODRIGUEZ, J.; SANCHEZ, M. J. Forecasting electricity prices by extracting dynamic common factors: application to the Iberian Market. *IET Generation, Transmission & Distribution*, v. 6, n. 1, p. 11–20, 2012. ISSN 17518687.
- GHAYEKHLOO, M.; MENHAJ, M. B.; GHOFRANI, M. A hybrid short-term load forecasting with a new data preprocessing framework. *Electric Power Systems Research*, Elsevier Ltd, v. 119, p. 138–148, 2015. ISSN 03787796.

- GUPTA, V.; PAL, S. An Overview of Different Types of Load Forecasting Methods and the Factors Affecting the. v. 5, n. Iv, p. 729–733, 2017.
- HAGAN, M. T.; MENHAJ, M. B. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks*, v. 5, n. 6, p. 989–993, 1994. ISSN 19410093.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2nd. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. ISBN 0132733501.
- HSIAO, Y.-H. Household electricity demand forecast based on context information and user daily schedule analysis from meter data. *IEEE Transactions on Industrial Informatics*, IEEE Computer Society, v. 11, n. 1, p. 33–43, 2015. ISSN 15513203.
- HYNDMAN, R. J.; FAN, S. Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, v. 25, n. 2, p. 1142–1153, 2010. ISSN 08858950.
- KALAITZAKIS, K.; STAVRAKAKIS, G. S.; ANAGNOSTAKIS, E. M. Short-term load forecasting based on artificial neural networks parallel implementation. *Electric Power Systems Research*, v. 63, n. 3, p. 185–196, 2002. ISSN 03787796.
- KIM, K.-H.; YOUN, H.-S.; KANG, Y.-C. Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method. *IEEE Transactions on Power Systems*, v. 15, n. 2, p. 559–565, 2000. ISSN 08858950.
- KOUHI, S. b.; KEYNIA, F.; Najafi Ravadanegh, S. A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection. *International Journal of Electrical Power and Energy Systems*, Elsevier Ltd, v. 62, p. 862–867, 2014. ISSN 01420615.
- LEE, W.-J.; HONG, J. A hybrid dynamic and fuzzy time series model for mid-term power load forecasting. *International Journal of Electrical Power and Energy Systems*, Elsevier Ltd, v. 64, p. 1057–1062, 2015. ISSN 01420615.
- LI, P. et al. Application of a hybrid quantized Elman neural network in short-term load forecasting. *International Journal of Electrical Power and Energy Systems*, v. 55, p. 749–759, 2014. ISSN 01420615.
- LI, Y.; YAN, Z.; FENG, D. Mid/long-term load forecasting model considering urbanization characteristics. *Dianli Zidonghua Shebei/Electric Power Automation Equipment*, Electric Power Automation Equipment Press, v. 36, n. 4, p. 54–61, 2016. ISSN 10066047.
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 1967. p. 281–297. Disponível em: <<https://projecteuclid.org/euclid.bsm/1200512992>>.
- MANDAL, P.; SRIVASTAVA, A. K.; PARK, J. W. An effort to optimize similar days parameters for ANN-based electricity price forecasting. *IEEE Transactions on Industry Applications*, v. 45, n. 5, p. 1888–1896, 2009. ISSN 00939994.
- MOHSENIAN-RAD, A. H.; LEON-GARCIA, A. Optimal residential load control with price prediction in real-time electricity pricing environments. *IEEE Transactions on Smart Grid*, v. 1, n. 2, p. 120–133, 2010. ISSN 19493053.

- NAGI, J. et al. A computational intelligence scheme for the prediction of the daily peak load. *Applied Soft Computing Journal*, v. 11, n. 8, p. 4773–4788, 2011. ISSN 15684946.
- NEZZAR, R. M. et al. Mid-long term load forecasting using multi-model artificial neural networks. *International Journal on Electrical Engineering and Informatics*, School of Electrical Engineering and Informatics, v. 8, n. 2, p. 389–401, 2016. ISSN 20856830.
- NUNES, I.; CARLI, L. Efficient Neurofuzzy Model to Very Short-Term Load Forecasting. *IEEE Latin America Transactions*, IEEE Computer Society, v. 14, n. 2, p. 721–728, 2016. ISSN 15480992.
- OH, S. J. et al. Forecasting long-term electricity demand for cooling of Singapore’s buildings incorporating an innovative air-conditioning technology. *Energy and Buildings*, Elsevier Ltd, v. 127, p. 183–193, 2016. ISSN 03787788.
- PETRIDIS, V. et al. A Bayesian multiple models combination method for time series prediction. *Journal of Intelligent and Robotic Systems: Theory and Applications*, v. 31, n. 1-3, p. 69–89, 2001. ISSN 09210296.
- PJM. *Pennsylvania-New Jersey-Maryland Interconnection*. 2018. Disponível em: <<http://www.pjm.com>>.
- ROKACH, L.; MAIMON, O. Clustering methods. In: *The Data Mining and Knowledge Discovery Handbook*. [S.l.: s.n.], 2005.
- SAHAY, K. B.; TRIPATHI, M. M. Day ahead hourly load forecast of PJM electricity market and iso new england market by using artificial neural network. *2014 IEEE PES Innovative Smart Grid Technologies Conference, ISGT 2014*, p. 1–5, 2014.
- SHEIKHAN, M.; MOHAMMADI, N. b. Neural-based electricity load forecasting using hybrid of GA and ACO for feature selection. *Neural Computing and Applications*, v. 21, n. 8, p. 1961–1970, 2012. ISSN 09410643.
- SUGANTHI, L.; SAMUEL, A. A. Energy models for demand forecasting — A review. *Renewable and Sustainable Energy Reviews*, Elsevier Ltd, v. 16, n. 2, p. 1223–1240, 2012. ISSN 1364-0321.
- YAO, A. W. L.; CHI, S. C.; CHEN, J. H. An improved Grey-based approach for electricity demand forecasting. v. 67, p. 217–224, 2003.
- ZHANG, B.-L.; DONG, Z.-Y. An adaptive neural-wavelet model for short term load forecasting. *Electric Power Systems Research*, v. 59, n. 2, p. 121–129, 2001. ISSN 03787796.