



UNIVERSIDADE FEDERAL FLUMINENSE

ESCOLA DE ENGENHARIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
TELECOMUNICAÇÕES

DOUGLAS ARANIL MAGALHÃES BARBOSA

ANALYTICS E MACHINE LEARNING PARA PREVISÃO AUTOMÁTICA
DE CARGA HORÁRIA E SEMANAL

NITERÓI, RJ

2020

UNIVERSIDADE FEDERAL FLUMINENSE
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
TELECOMUNICAÇÕES

DOUGLAS ARANIL MAGALHÃES BARBOSA

ANALYTICS E MACHINE LEARNING PARA PREVISÃO AUTOMÁTICA
DE CARGA HORÁRIA E SEMANAL

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica e de Telecomunicações.

Orientador: Prof. Dsc. VITOR HUGO FERREIRA

Niterói – RJ

2020

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

B238a Barbosa, Douglas Aranil Magalhães
ANALYTICS E MACHINE LEARNING PARA PREVISÃO AUTOMÁTICA DE
CARGA HORÁRIA E SEMANAL / Douglas Aranil Magalhães Barbosa ;
Vitor Hugo Ferreira, orientador. Niterói, 2020.
130 f. : il.

Dissertação (mestrado)-Universidade Federal Fluminense,
Niterói, 2020.

DOI: <http://dx.doi.org/10.22409/PPGEET.2020.m.14729218795>

1. Aprendizado de Máquinas. 2. Analytics. 3. Previsão de
Carga. 4. Operação de sistema elétricos. 5. Produção
intelectual. I. Ferreira, Vitor Hugo, orientador. II.
Universidade Federal Fluminense. Escola de Engenharia. III.
Título.

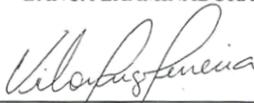
CDD -

DOUGLAS ARANIL MAGALHÃES BARBOSA

ANALYTICS E MACHINE LEARNING PARA PREVISÃO AUTOMÁTICA DE CARGA DIÁRIA E SEMANAL

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Telecomunicações da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica e de Telecomunicações.

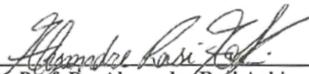
BANCA EXAMINADORA



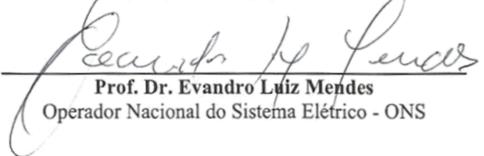
Prof. Dr. Vitor Hugo Ferreira - Orientador
Universidade Federal Fluminense - UFF



Prof. Dr. Bruno Soares Moreira Cesar Borba
Universidade Federal Fluminense - UFF



Prof. Dr. Alexandre Raí Aoki
Universidade Federal do Paraná - UFPR



Prof. Dr. Evandro Luiz Mendes
Operador Nacional do Sistema Elétrico - ONS

Niterói
(outubro/2020)

DEDICATÓRIA

Dedico este trabalho aos meus pais, por toda dedicação, amor, carinho, respeito e confiança durante todos os anos de minha vida.

Aos meus familiares pelas boas energias que me dispensaram ao longo desses últimos anos, em especial para minha avó Dona Izaura, minhas Tias Branca, Iara, Delma e meu tios Jorge e Carlos.

Ao meu avô “Seu Aranil”, que me abençoa todos os dias lá de cima e certamente está fazendo um churrasco para comemorar mais uma vitória de seu neto.

Aos meus avós maternos Dona Maria e Seu Fani que embora não os conheci sei que me iluminam durante a minha caminhada.

Aos meus amigos, por se tornarem a família que escolhi, sempre depositando confiança de que era capaz de alcançar meus objetivos.

AGRADECIMENTOS

A Deus,

A toda a minha família,

Aos meus amigos,

Ao professor Vitor Hugo, por embarcar em minhas ideias loucas e ainda contribuir para o desenvolvimento de mais loucuras ao longo do trabalho. De início meu orientador, agora um amigo para a vida.

A toda equipe de previsão de carga do Operador Nacional do Sistema Elétrico (Fausto, Marcia, Marcela, Maria Helena, Douglas Alexander, Evandro, Fernando, Mirtis, David, Luiz Carlos, Juliana, Olivio, Danilo, Gabriel, Maria Jadnilza, Janaina, Marcio Fera) pela confiança em meu trabalho, companheirismo e aprendizado durante os últimos anos.

RESUMO

Atualmente, o conhecimento a priori do valor de carga para um determinado horizonte futuro representa um dos principais desafios durante o planejamento, operação e expansão de um sistema elétrico de potência. Dessa forma, o presente problema vem sendo abordado na literatura durante um longo período. Nos últimos anos com o avanço do poder computacional os modelos que utilizam redes neurais artificiais ganharam cada vez mais atenção, no entanto, sua utilização não se torna trivial uma vez que tais modelos apresentam diversos parâmetros que precisam ser ajustados/escolhidos para suas aplicações, como: escolha das variáveis de entrada, escolha do número de neurônios e escolha do número de camadas a serem utilizadas. A vista disso, o presente trabalho possui como objetivo o desenvolvimento de um modelo matemático autônomo para previsão automática de carga horária e semanal. Para tanto são utilizadas técnicas de Analytics e Machine Learning, como as redes neurais artificiais com treinamento bayesiano, os modelos de misturas de gaussianas e técnicas de clusterização para segmentação de bases de dados, como os algoritmos de K-médias e X-médias. A escolha da utilização de diferentes modelos em conjunto (redes neurais artificiais e modelos de misturas de gaussianas) permite uma melhor representação do espaço, possibilitando tratar de forma não paramétrica series temporais que apresentam ruídos não gaussianos. Além disso, ao longo do trabalho será apresentada formas autônomas de escolha das variáveis de entrada e controle de complexidade do modelo, possibilitando ainda mais a criação de um modelo autônomo conforme o objetivo descrito. Por fim, como forma de validação do modelo, o mesmo será aplicado a problemas de previsões de carga de energia semanais e horárias, sendo os resultados obtidos avaliados através de métodos apropriados como: a avaliação do erro absoluto médio percentual, erro absoluto, análise do coeficiente de assimetria dos resíduos, Curtose e avaliação dos intervalos de confiança através do índice de avaliação de intervalos de previsão.

Palavras-chave: Aprendizado de Máquinas, Analytics, Previsão de Carga, Redes Neurais Bayesianas, Operação de sistema elétricos.

ABSTRACT

Nowadays, the knowledge of the future load value represents one of the main challenges during the planning, operation, and expansion of an electric power system. Thus, this problem has a lot of works published in the literature for a long time. In recent years, with the advancement of computational power, models that use artificial neural networks have gained more and more attention, however, their use is not trivial, because these models present several parameters that need to be adjust/chosen for their applications, such as: choosing the input variables, choosing the number of neurons, and choosing the number of hidden layers. Therefore, the present work aims to develop an autonomous mathematical model for automatic hourly and weekly load prediction. For this purpose, techniques of Analytics and Machine Learning are used, such as artificial neural networks with Bayesian training, Gaussian mixtures model and clustering techniques for segmentation of databases, such as the algorithms of K-means and X-means. The choice of using different models (artificial neural networks and Gaussian mixtures model) allows a better representation of space, making possible to treat temporal series that present non-Gaussian noises. In addition, throughout the work, autonomous ways of choosing the input variables and controlling the model's complexity will be presented, allow to create an autonomous model according to the objective described. As a form of model validation, it will be applied to weekly and hourly load forecast problems, with the results obtained being evaluated using appropriate methods such as the evaluation of the mean percentage absolute error, absolute error, skewness, kurtosis and assessment of confidence intervals using Prediction Interval Evaluation Indexes.

Keywords: Machine Learning, Analytics, Load Forecasting, Bayesian neural networks, Transmission System Operators.

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Previsões de carga no cenário brasileiro	14
1.2	Pontos de Melhorias	16
1.3	Objetivos do trabalho	17
1.4	Produções científicas oriundas do trabalho	17
1.5	Organização do trabalho.....	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Redes Neurais Artificiais.....	19
2.2	Inferência Bayesiana aplicada ao desenvolvimento de MLP's	23
2.3	Treinamento Bayesiano Aplicado aos MLP's.....	24
2.3.1	Regressão	25
2.3.2	Classificação	34
2.3.3	Determinação automática de relevância.....	37
2.4	Aprendizado não supervisionado	38
2.4.1	Medidas de Similaridade.....	39
2.4.2	K médias (<i>K-means</i>).....	41
2.4.3	X-Médias.....	44
2.5	Misturas de Gaussianas	47
2.6	Resumo e Discussões	49
3	MATERIAIS E MÉTODOS	50
3.1	Metodologia proposta.....	53
3.2	Dados utilizados	57
3.3	Critério de Análise.....	58
3.3.1	Erro Médio Percentual (<i>MAPE</i>).....	58
3.3.2	Erro absoluto	59
3.3.3	Coeficiente de assimetria e Curtose	59
3.3.4	Avaliação dos intervalos de confiança.....	60
4	RESULTADOS	61
4.1	Resultados Semanais	61
4.1.1	Análise de Curtose e do Coeficiente de assimetria	68
4.1.2	Análise do Intervalo de Confiança	71
4.1.3	Benchmark	73
4.2	Resultados Diários.....	74

4.2.1	Análise diária com feriado	75
4.2.2	Análise diária sem feriado.....	79
4.2.3	Análise de Curtose e do Coeficiente de Assimetria	82
4.2.4	Avaliação dos intervalos de Confiança <i>PICP</i>	84
4.2.5	Benchmark	84
5	CONCLUSÃO E TRABALHOS FUTUROS	86
	REFERÊNCIAS BIBLIOGRÁFICAS	89
	APÊNDICES	94

LISTA DE ILUSTRAÇÕES

FIGURA 1 - HORIZONTE DE PREVISÃO DO PMO	16
FIGURA 2 – ESTRUTURA DO NEURÔNIO	19
FIGURA 3 - REDE NEURAL ARTIFICIAL DO TIPO MULTI LAYER PERCEPTRON	21
FIGURA 4 - ESTRUTURA DO MLP UTILIZADO PARA CLASSIFICAÇÃO DE PADRÕES [23], [38]	35
FIGURA 5 - FLUXOGRAMA DO ALGORITMO DO K-MÉDIAS [45].....	43
FIGURA 6 - FLOXOGRAMA DE ORGANIZAÇÃO DO MODELO CONSTRUÍDO.....	51
FIGURA 7 - SUBMERCADO DE ENERGIA [68]	57
FIGURA 8 – VERIFICADO X PREVISTO SIN + ERRO	63
FIGURA 9 – VERIFICADO X PREVISTO SE/CO + ERRO	64
FIGURA 10 – VERIFICADO X PREVISTO SUL + ERRO	64
FIGURA 11- VERIFICADO X PREVISTO NORDESTE + ERRO	65
FIGURA 12- VERIFICADO X PREVISTO NORTE + ERRO	65
FIGURA 13 – ERRO SEMANAL ACUMULADO (SIN)	66
FIGURA 14 - ERRO SEMANAL ACUMULADO (SE/CO)	66
FIGURA 15 - ERRO SEMANAL ACUMULADO (SUL)	67
FIGURA 16 - ERRO SEMANAL ACUMULADO (NORDESTE)	67
FIGURA 17 - ERRO SEMANAL ACUMULADO (NORTE)	68
FIGURA 18 - HISTOGRAMA DOS RESÍDUOS SEMANAIS DO SIN	69
FIGURA 19 - HISTOGRAMA DOS RESÍDUOS SEMANAIS DO SUBSISTEMA SUDESTE/CENTRO- OESTE.....	69
FIGURA 20- HISTOGRAMA DOS RESÍDUOS SEMANAIS DO SUBSISTEMA SUL.....	70
FIGURA 21- HISTOGRAMA DOS RESÍDUOS SEMANAIS DO SUBSISTEMA NORDESTE	70
FIGURA 22- HISTOGRAMA DOS RESÍDUOS SEMANAIS DO SUBSISTEMA NORTE	71
FIGURA 23 - ERRO DIÁRIO SE/CO 168 PASSOS (SE/CO)	77
FIGURA 24- ERRO DIÁRIO SE/CO 48 PASSOS (SE/CO)	77
FIGURA 25 - DESVIO DIÁRIO MÁXIMO 168 (SE/CO)	78
FIGURA 26 - DESVIO DIÁRIO MÁXIMO 48 (SE/CO)	78
FIGURA 27 - DESVIO DIÁRIO MÍNIMO 168 (SE/CO)	78
FIGURA 28 - DESVIO DIÁRIO MÍNIMO 48 (SE/CO)	79
FIGURA 29 - ERROS DIÁRIOS 168 PASSOS SEM FERIADO (SE/CO)	80
FIGURA 30 - ERROS DIÁRIOS 48 PASSOS SEM FERIADO (SE/CO)	81
FIGURA 31 - DESVIO DIÁRIO MÁXIMO 168 S/FERIADO (SE/CO)	81
FIGURA 32 - DESVIO DIÁRIO MÁXIMO 48 S/FERIADO (SE/CO)	82
FIGURA 33 - HISTOGRAMA DOS RESÍDUOS PARA A PREVISÃO 168 HORAS	83
FIGURA 34- HISTOGRAMA DOS RESÍDUOS DAS PREVISÕES 48 HORAS.....	83

LISTA DE TABELAS

TABELA 1 - AEROPORTOS UTILIZADOS E SEUS RESPECTIVOS SUBSISTEMAS [67]	58
TABELA 2 - LISTA DE FERIADOS CONSIDERADOS	62
TABELA 3 - ERRO MÉDIO POR SUBSISTEMA (PREVISÕES SEMANAIS).....	62
TABELA 4 - ANÁLISE PICP PARA PREVISÃO SEMANAL CONSIDERANDO 3 DP	72
TABELA 5 - ANÁLISE PICP PARA PREVISÃO SEMANAL CONSIDERANDO 2 DP	72
TABELA 6 - ANÁLISE PICP PARA PREVISÃO SEMANAL CONSIDERANDO 1 DP	72
TABELA 7 - INTERVALO DE CONFIANÇA MÉDIO.....	73
TABELA 8 - ERROS DIÁRIOS MÉDIOS POR SUBSISTEMA	75
TABELA 9 - MAPE MENSAL.....	76
TABELA 10 - MAPE DIÁRIO	76
TABELA 11 - ERROS MÉDIOS DIÁRIOS SEM FERIADO	79
TABELA 12 - MAPE MENSAL (SEM FERIADOS).....	80
TABELA 13 - MAPE DIÁRIO (SEM FERIADOS)	80
TABELA 14 - ANÁLISE PICP (COM E SEM FERIADOS)	84
TABELA 15 - PREVISÃO DIÁRIA EXEMPLO	94
TABELA 16 - INTERVALO DE CONFIANÇA SEMANAL SE/CO.....	96
TABELA 17 - INTERVALO DE CONFIANÇA SEMANAL SUL	98
TABELA 18 - INTERVALO DE CONFIANÇA SEMANAL NORDESTE.....	101
TABELA 19 - INTERVALO DE CONFIANÇA SEMANAL NORTE.....	103
TABELA 20 - INTERVALO DE CONFIANÇA SEMANAL SIN	106
TABELA 21 - TABELA DE ERROS SEMANAIS	109
TABELA 22 - ERROS DIÁRIOS	112

1 INTRODUÇÃO

O conhecimento a priori do valor da carga que irá se realizar num determinado horizonte futuro de tempo apresenta vital importância para o planejamento e operação segura dos sistemas elétricos de potência. Além disso, devido ao avanço do mercado de energia nos últimos anos, com a desestatização de muitas empresas, os agentes que compõem o setor elétrico devem operar no regime de máxima eficiência, para que seus custos operacionais sejam minimizados e conseqüentemente maximizando a rentabilidade do empreendimento. Dessa maneira, mais uma vez a previsão de carga desponta como principal insumo a ser conhecido para a realização de diversas atividades que fazem parte do dia a dia do setor elétrico.

Na literatura os horizontes de previsão são divididos, basicamente, de três formas: curto, médio e longo prazo. Todavia, as definições desses horizontes não são únicas, fazendo com que estudos de curto ou médio prazo para alguns autores sejam definidos como médio ou longo prazo para outros [1], [2].

Abordando um pouco mais sobre esses horizontes, as previsões de longo prazo são aquelas que se estendem de alguns meses à frente até anos à frente. Geralmente, tais previsões estão relacionadas ao planejamento técnico e financeiro das empresas, no que tange a expansão do parque de geração, compra futura de energia, planejamento das manutenções preventivas anuais e até mesmo para o desenvolvimento de políticas tarifárias [3].

Previsões de médio prazo são aquelas que variam de uma semana à frente até cinco anos, com medições em base diária, semanal ou mensal. Essas previsões embasam a compra de combustíveis para as unidades geradoras, manutenções de equipamentos, otimização do despacho de unidades geradoras entre outras atividades [3].

Por fim, as previsões de curto prazo são aquelas que se estendem de alguns minutos à frente até um mês à frente, tais previsões são discretizadas em minutos, horas ou dias. As previsões de curto prazo servem como insumo para diversos estudos, como os estudos de coordenação hidrotérmica e fluxo de potência ótimo para uma linha de transmissão. Além desses estudos, essas previsões também estão diretamente relacionadas com a operação do sistema elétrico de potência, uma vez que as medidas operativas que deverão ser tomadas na sala de controle, como: ajuste de tap's dos

transformadores, desligamento de circuitos, despacho ou desligamento de unidades geradoras, são embasadas por essas previsões de curto prazo [3].

Dado a importância que esse tema apresenta para os sistemas elétricos de potência, inúmeros trabalhos podem ser encontrados na literatura. Em geral, esses trabalhos podem ser divididos em dois tipos, basicamente, que são: técnicas estatísticas de análise de séries temporais, dentre elas estão os métodos de Suavização Exponencial, Modelos Autorregressivos de média móvel (do inglês *autoregressive moving average* – *ARMA*), Modelos autorregressivo integrado de médias móveis (do inglês *autoregressive integrated moving average* – *ARIMA*), Modelos Autorregressivos Generalizados Orientados por Score (do inglês *Generalized Autoregressive Score* - *GAS*) [4]–[6] e técnicas de inteligência computacional, onde abrangem modelos que utilizam as Redes Neurais Artificiais (RNA's), Lógica Fuzzy, *Deep Learning*, Máquinas de Vetor Suporte [7]–[9]. Além disso, alguns autores propõem técnicas mistas, que interligam um ou mais tipos de modelos como: NeuroFuzzy, Redes neurais com algoritmo genéticos ou otimização por enxame de partículas (do inglês *particle swarm optimization* – *PSO*), entre outros métodos [10]–[12].

Torna-se válido frisar que, ao longo dos últimos anos os modelos de inteligência computacional vêm se despontando, esse fato pode ser justificado em função do avanço do poder computacional, o que representava um grande empecilho para sua utilização e pela sua capacidade de generalização e rápida adaptação às mudanças de comportamento do cenário. Em função disso, inúmeros trabalhos podem ser vistos na literatura utilizando tais modelos, como em [13], [14].

1.1 Previsões de carga no cenário brasileiro

No Brasil, o Operador Nacional do Sistema Elétrico (ONS) é o órgão responsável pela coordenação e controle da operação das instalações de transmissão e geração do país. Além disso, também está sob sua responsabilidade o planejamento da operação dos sistemas isolados, sob a fiscalização e regulação da Agência Nacional de Energia Elétrica [15]. Diante disso, uma das funções do ONS consiste em estabelecer critérios para a operação do sistema interligado nacional (SIN). Sendo esses critérios apresentados em seus procedimentos de redes, que são documentos de caráter normativo elaborados pelo

ONS, com participação dos agentes e crivo da Agência Nacional de Energia Elétrica (ANEEL) [16].

O módulo cinco dos procedimentos de rede abrange as atribuições e responsabilidades dos agentes e do próprio ONS, em relação aos prazos e sistemáticas envolvidas no processo de consolidação de carga. Em função do objetivo do presente trabalho, os submódulos dos procedimentos de rede envolvidos são: Submódulo 5.4 [17] e Submódulo 5.6 [18], que fazem referência as previsões de carga diária para a elaboração da programação diária da operação eletroenergética e para a programação de intervenções em instalações da rede de operação e previsões de carga semanais e mensais para o Programa Mensal da Operação (PMO).

As previsões de carga diária realizadas pelo operador apresentam discretizações semi-horária, conforme definido em [17]. Devido a mudança de paradigma que vem ocorrendo no setor elétrico brasileiro com a entrada do modelo DESSEM [19], modelo esse que visa a realização do despacho hidrotérmico horário, tais previsões são realizadas diariamente inclusive durante os finais de semana. Com isso, torna-se responsabilidade do analista realizar a previsão de carga para o próximo dia ($d+1$), todavia, o modelo DESSEM necessita de previsões de carga para um horizonte de até sete dias à frente. Isto posto, previsões oriundas de modelos são adicionadas a fim de completar o horizonte de previsão necessário. Portanto, o desenvolvimento de modelos que sejam capazes de originar boas previsões num horizonte de até sete dias à frente, com discretização horária ou semi-horária, se torna necessário.

Além das previsões diárias que foram supracitadas, o ONS também necessita de previsões em discretização semanal e mensal, devido ao Programa Mensal da Operação, que possui como objetivo elaborar diretrizes eletroenergéticas para uma operação otimizada do SIN para o próximo mês. Logo, encontra-se definido em procedimentos de rede que os agentes deverão encaminhar ao ONS previsões de carga de energia nas discretizações postas anteriormente. Assim, torna-se de responsabilidade do operador consistir as informações enviadas pelos agentes e realizar as previsões oficiais para os horizontes em questão. A Figura 1 expõe o horizonte de previsão para o PMO, onde atualmente são solicitadas previsões em base mensal para dois meses à frente, sendo o primeiro mês desagregado em semanas operativas, torna-se válido ressaltar que as

semanas operativas se estendem de sábado até sexta. Além disso, destaca-se que essas previsões semanais são revistas todas as semanas.

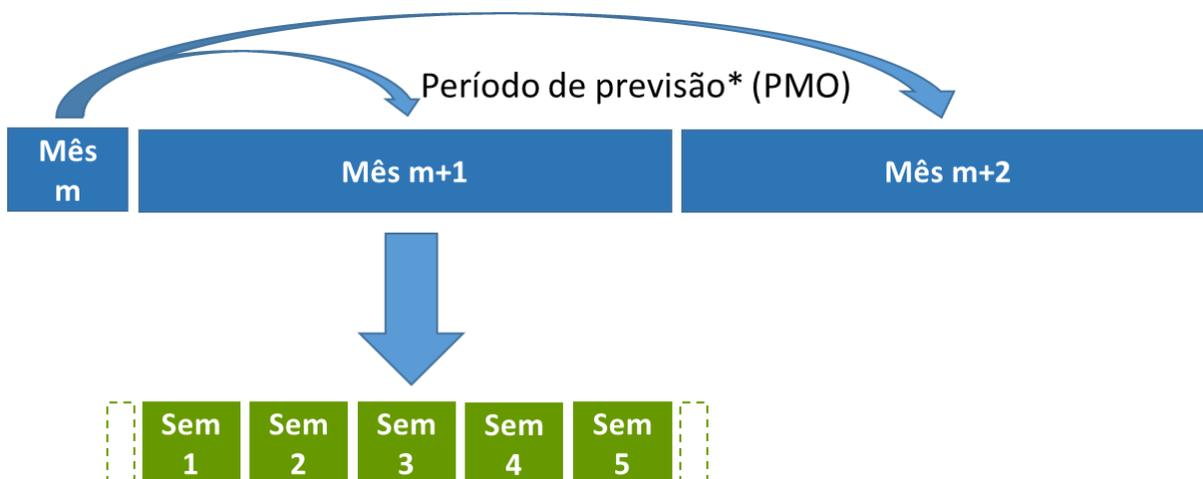


Figura 1 - Horizonte de Previsão do PMO

1.2 Pontos de Melhorias

Como visto, o tema previsão de carga ainda é muito debatido por diversos autores na literatura em função da sua complexidade. Diversos modelos são apresentados e em sua vasta maioria são modelos que podem ser denominados de especialistas, isto é, modelos que são desenvolvidos apenas para uma série temporal específica, o que representa a construção de modelos poucos generalistas e que para sua utilização em um outro tipo de série temporal torna-se necessário uma grande interferência por parte do especialista, sendo essa interferência vista principalmente durante a modelagem dos parâmetros apresentados no modelo e a escolha das variáveis relevantes para o problema [3]. Além disso, outro ponto de melhoria que pode ser apresentado consiste na previsão de intervalos de confiança associados aos valores médios previstos, uma vez que atualmente a maioria dos trabalhos vistos na literatura realizam técnicas a posteriori, como re-amostragem dos erros ou regressões [20], para a construção desses intervalos.

Logo, em função das inúmeras limitações, como por exemplo as limitações supracitadas, torna-se desejado o desenvolvimento de um modelo generalista e robusto para que possa ser utilizado em diferentes tipos de séries temporais, com a mínima interferência por parte do analista. Além disso, dado a elevada incerteza presente durante o processo de previsão, a associação de intervalos de confiança as previsões se torna indicado.

1.3 Objetivos do trabalho

Frente as limitação supracitadas, o presente trabalho possui como objetivo o desenvolvimento de um modelo matemático autônomo para a previsão de diferentes tipos de séries temporais. O modelo proposto surgiu com o objetivo da correção das previsões de precipitação, utilizadas pelo Operador Nacional do Sistema Elétrico, oriundas de modelos físicos. As referências [1], [3] utilizam uma parte do modelo proposta aplicado a problemas de previsão de carga, contudo nessas referências o espaço é modelado através de uma única distribuição gaussiana somente. No presente trabalho, com a utilização do modelo de mistura de gaussianas, o espaço será modelado através de diferentes gaussianas, o que permitirá o ajuste de qualquer função de densidade probabilidade de forma não paramétrica, maiores detalhes teóricos serão apresentados no decorrer do trabalho. Além disso, o modelo proposto irá demandar a mínima interferência do analista durante o processo de construção do modelo, o que irá garantir a sua utilização em diferentes tipos de séries temporais, e intervalos de confiança associados aos valores médios serão previstos.

Para esse desenvolvimento serão utilizadas técnicas de *Machine Learning*, como as redes neurais artificiais com treinamento bayesiano e *Analytics* como os métodos K-Médias, X-Médias e *GMM*, sendo esses métodos detalhados no decorrer do trabalho. Como forma de validar a robustez e a capacidade de generalização do modelo dois horizontes de previsão são adotados, o primeiro deles é a previsão de carga semanal com o horizonte de uma semana à frente, e a segundo é a previsão de carga diária (com discretização horária) com horizonte de até sete dias à frente, o que representa 168 horas.

1.4 Produções científicas oriundas do trabalho

- Barbosa, D.; Ferreira, V.; “Inferência Bayesiana Aplicada a MLPs para previsão probabilística de carga Semanal”, SBSE 2020, UFABC.

Neste trabalho é abordado a previsão de carga de energia (MW_{méd}) uma semana à frente utilizando o modelo neural autônomo sugerido nesta dissertação. As previsões se estenderam ao longo do ano de 2017 e 2018 para os submercado de energia que compõem o Sistema Interligado Nacional (SIN). As previsões obtidas foram validadas e comparadas com métodos sugeridos na literatura, e algumas semanas comparadas com o desempenho

do próprio ONS. Os resultados obtidos se mostraram satisfatórios em função dos baixos desvios médios verificados durante o processo.

1.5 Organização do trabalho

O capítulo 2 apresenta a fundamentação teórica utilizada para o desenvolvimento do modelo, onde é visto uma leve introdução sobre as redes neurais artificiais e em seguida é apresentado a inferência bayesiana aplicada as *MLPs*. Dando continuidade ao capítulo é exposto o método que garante a escolha de forma automática das variáveis de entrada relevantes e por fim são apresentados os métodos de agregação utilizados.

O capítulo 3 apresenta a metodologia proposta de uma forma detalhado, onde é apresentado como os métodos descritos no capítulo dois são “encapsulados” para a construção do modelo proposto. Em seguida será realizada uma breve explanação sobre os dados utilizados e ao final do capítulo é visto os critérios utilizados para a avaliação dos resultados obtidos durante os testes do modelo desenvolvido.

O capítulo 4 aborda os resultados obtidos ao se utilizar a metodologia proposta, onde uma análise crítica e minuciosa é realizada sobre tais previsões.

Por fim, o capítulo 5 apresenta as conclusões do presente trabalho e os desafios futuros que são propostos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Redes Neurais Artificiais

As redes neurais artificiais (RNA's) são modelos matemáticos inspirados no funcionamento do sistema nervoso dos seres humanos, possuindo a capacidade de aquisição, manutenção e generalização do conhecimento [21]. As RNA's podem ser vistas como processadores de sinais paralelamente distribuídos, formadas por unidades de conhecimento que são denominados neurônios [22].

Os neurônios obtêm conhecimentos via um algoritmo de aprendizado, que será apresentado no decorrer do documento. Esses conhecimentos são armazenados sob a forma de pesos sinápticos (ω), que interligam os diversos neurônios como apresentado na figura 2. Nessa figura é retratado um neurônio cuja sua saída c é dada pela equação 2.1 [23].

$$c = \phi\left(\sum_{i=1}^n w_i x_i + b\right) \quad (2.1)$$

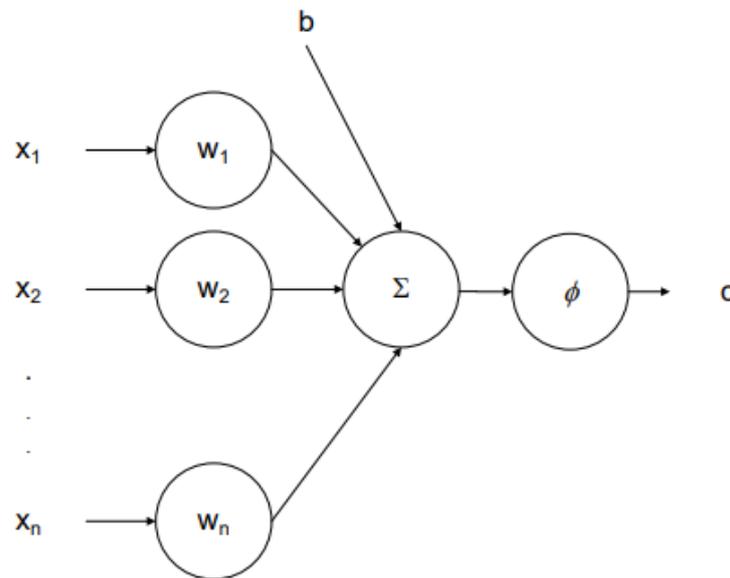


Figura 2 – Estrutura do Neurônio

Na equação 2.1, $\underline{w} \in \mathbb{R}^n$, $\underline{w} = [w_1, w_2, w_3, \dots, w_n]$ que representa o conjunto de pesos que ligam as entradas $\underline{x} \in \mathbb{R}^n$, $\underline{x} = [x_1, x_2, x_3, \dots, x_n]$, ao neurônio em questão.

Essas entradas podem ser oriundas da própria camada de entrada ou até mesmo de outros neurônios, $b \in \mathbb{R}$ representa o bias desse neurônio e por fim $\emptyset(.)$ representa a função de ativação. Em geral, são utilizadas funções de ativação não lineares, o que acarreta numa elevada capacidade de aproximação de diversas funções, por parte do modelo [24]. A seguir, serão apresentadas as funções de ativação mais utilizadas na literatura, que são: sigmoideal ou logística (equação 2.2), tangente hiperbólica (equação 2.3) e linear (equação 2.4).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.3)$$

$$f(x) = x \quad (2.4)$$

Como supracitado, as RNA's visam emular o funcionamento do cérebro humano, para isso os neurônios são dispostos em camadas e de acordo com esse arranjo as RNA's são classificadas. Diversas arquiteturas são apresentadas na literatura, como: *feedforward* (padrão utilizado na rede *Perceptron* de múltiplas camadas), recorrente ou realimentadas (padrão utilizado nas Redes Recorrentes de Hopfield) e reticulada (padrão utilizado nas redes de Kohonen) [21]. Contudo nesse trabalho serão apresentadas somente as redes alimentadas adiante, do inglês *feedforward*. Como seu próprio nome diz, as redes *feedforward* são aquelas em que seus neurônios são conectados de forma consecutiva e adjacente e com isso o fluxo da informação segue sempre numa única direção, conforme apresentado na figura 3.

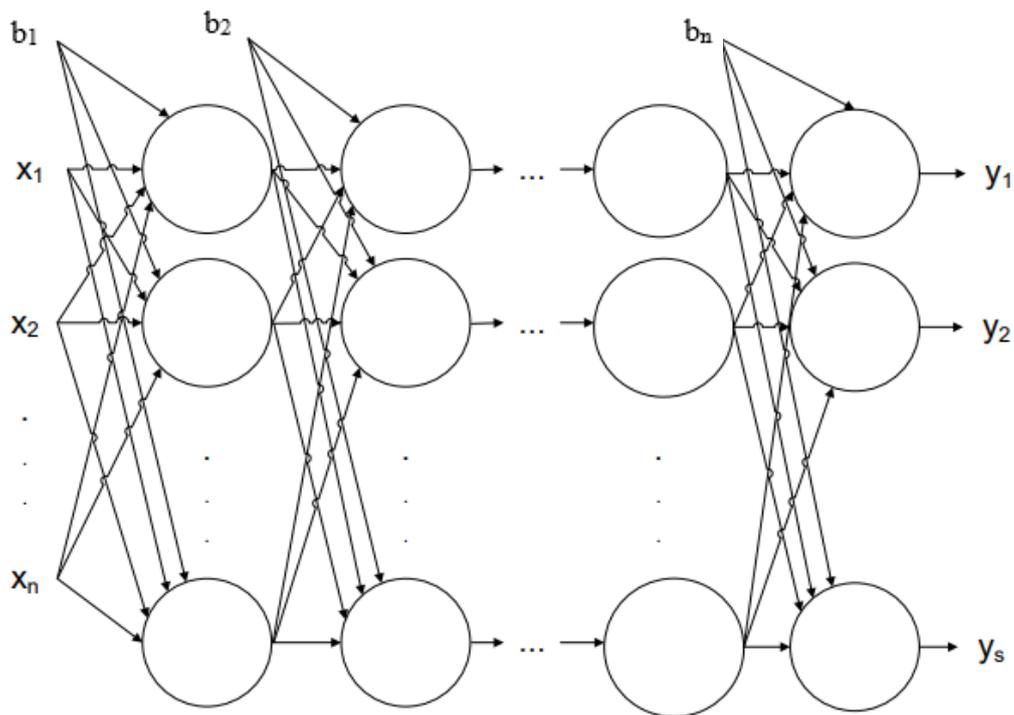


Figura 3 - Rede Neural Artificial do tipo Multi Layer Perceptron

As principais vantagens de utilização dos modelos neurais são [3]:

- Não-linearidade: para o caso mais comumente utilizado, em que a função de ativação dos neurônios da camada oculta é não-linear, o modelo neural resultante da interconexão destas unidades mais simples apresenta considerável grau de não-linearidade. Entretanto, esta vantajosa característica pode ser prejudicial na presença de dados ruidosos, problema que será abordado ao longo deste capítulo.
- Mapeamento entrada-saída: a partir de um conjunto de pares entrada-saída, as redes neurais realizam um mapeamento destes dados, sem a necessidade de desenvolvimento de modelos matemáticos abordando a dinâmica do processo.
- Adaptabilidade: estes modelos apresentam elevada capacidade de adaptação em virtude de mudanças nas condições do ambiente para o qual a rede foi treinada para operar. Para tal, basta treinar a rede novamente, incluindo no conjunto de treinamento os padrões referentes às novas condições operacionais.

- Implementação simples: já que os modelos *feedforward* podem ser vistos como grafos orientados, a implementação destas estruturas é extremamente simples quando comparada com o grau de complexidade dos modelos que podem ser gerados

Como dito, as RNA's são modelos capazes de generalizar inúmeros tipos de funções, para isso o teorema da aproximação universal [25] garante que redes *feedforward* são capazes de aproximar com precisão arbitrária qualquer função contínua $F(x): \mathbb{R}^n \rightarrow \mathbb{R}$. Para tanto, torna-se necessário que a estrutura apresente pelo menos uma camada oculta com número suficiente de neurônios nesta camada e função de ativação limitada, contínua e não constante. Ou seja, redes *feedforward* podem aproximar com precisão arbitrária qualquer função contínua, desde que para isso apresente pelo menos uma camada oculta e um número suficiente de neurônios nessa camada.

Embora a capacidade de generalização dos modelos neurais represente uma grande motivação para sua utilização, a existência de dados ruidosos pode representar um grande empecilho para sua aplicação. Pois, com o elevado grau de não linearidade presente em tais modelos, torna-se possível a representação de traços específicos do conjunto de dados disponível para treinamento, validação e/ou teste do modelo, dando origem ao problema conhecido como *overfitting*. Posto que o objetivo durante a construção do modelo é a representação de forma satisfatória do processo de geração dos dados, representando assim uma boa capacidade de generalização do mesmo, o descuido durante o ajuste dos parâmetros existentes no modelo pode vir a representar um grande empecilho para o alcance desse objetivo.

O processo de ajuste dos parâmetros das redes neurais é realizado durante a etapa de treinamento do modelo. Para isto, inúmeras técnicas são propostas na literatura. No entanto, a grande maioria desses processos visam a minimização do erro médio quadrático entre a resposta do modelo e o valor desejado, não garantindo assim um controle de *overfitting* associado ao processo.

Além do que foi supracitado, outro fator que representa um grande desafio durante a utilização desse tipo de estrutura é a escolha do número ideal de neurônios na camada escondida. Para esse fim, na literatura é sugerida a separação do conjunto de dados disponíveis em três: um para treinamento, outro para validação e outro para teste. Dessa

maneira, inúmeros modelos são desenvolvidos e aquele que apresentar o menor erro para o conjunto de validação é o modelo proposto para ser utilizado. Com isso, surge a técnica de validação cruzada (do inglês *cross-validation*). No entanto, devido a não garantia da escolha do modelo ideal, dado que ao se utilizar diferentes formas de validação (*leave-one-out*, *K-fold*, *bootstrap*) cada uma delas poderá apresentar diferentes resultados [26], muitos autores ainda se encontram bastante céticos em relação à efetividade de tais técnicas [27].

Outro grande desafio durante a utilização dos modelos neurais consiste na escolha das variáveis a serem utilizadas como entradas no modelo. A escolha equivocada desse conjunto de dados, pode ocasionar num aumento do esforço computacional, aumento da complexidade para visualização e entendimento dos dados e até mesmo ocasionar num ruim desempenho do modelo. Contudo, muitos autores ainda acreditam que o conhecimento do analista sobre o problema basta para a escolha desses *inputs*, não implementando assim qualquer método que garanta a escolha dessas variáveis de forma ideal.

Com os fatos mencionados anteriormente, torna-se evidente a importância do desenvolvimento de métodos que visem garantir o controle de complexidade do modelo, evitando assim os problemas de *overfitting*, e métodos que garantam uma boa seleção do número de neurônios e do conjunto de entradas. Para isso, o presente trabalho utiliza a inferência bayesiana aplicado aos *MLPs* como forma de treinamento das RNA's, o que garante um controle de complexidade ao modelo e um método de escolha de forma automática das variáveis de entrada.

2.2 Inferência Bayesiana aplicada ao desenvolvimento de MLP's

No tópico anterior foi abordado que na literatura existem diversos métodos para o treinamento das RNA's, nesse trabalho será utilizada a inferência Bayesiana para esse objetivo. O primeiro a propor essa forma de treinamento foi *David J.C. Mackay* em 1992 [28], sendo suas principais vantagens dadas por [3], [25]:

- O algoritmo de retro propagação do erro tradicional pode ser visto como um caso particular dos resultados obtidos através da aplicação de técnicas de inferência bayesiana.

- Para problemas de regressão, intervalos de confiança podem ser gerados automaticamente.
- Este procedimento permite o desenvolvimento de um algoritmo de determinação automática de relevância das entradas, do inglês *automatic relevance determination (ARD)*, técnica que pode ser utilizada para seleção de variáveis de entrada.
- Através do cálculo da evidência de cada modelo, relacionada com a probabilidade a posteriori de cada estrutura, esta metodologia permite a comparação entre diferentes modelos utilizando somente os dados disponíveis para treinamento.
- Técnicas de inferência bayesiana permitem afirmar em qual região do espaço de entrada devem ser obtidos novos dados com o intuito de aumentar a informação contida no modelo, característica conhecida como aprendizado ativo.

No presente trabalho foi adotado o treinamento bayesiano para dois tipos de objetivos, classificação e regressão. Por conseguinte, serão apresentadas as duas formas de treinamento dos modelos. Aplicações da inferência bayesiana as *MLP's* podem ser encontradas nas referências [29]–[35].

2.3 Treinamento Bayesiano Aplicado aos MLP's

Ao contrário dos treinamentos convencionais que possuem como objetivo a minimização do erro médio quadrático entre a resposta oriunda do modelo e o conjunto de saída desejadas, o treinamento bayesiano apresenta como objetivo a maximização da evidência, o que nada mais é que a maximização da probabilidade a posteriori da realização do conjunto de pesos da rede neural, dados os padrões de entrada e saída. No decorrer do trabalho será visto que alguns hiperparâmetros serão ajustados, além do conjunto de pesos, e com o ajuste desses hiperparâmetros torna-se possível o desenvolvimento de métodos encapsulados para a escolha do conjunto de entradas ideal, conhecido como Determinação Automática de Relevância (do inglês *Automatic Relevance Determination - ARD*).

Visto que, nesta dissertação serão utilizadas as redes neurais bayesianas aplicadas a problemas de classificação e de regressão, os dois tipos de treinamentos serão apresentados. Para isso, será adotado como premissa que a estrutura do modelo já se encontra definida previamente. Subentende-se como estrutura do modelo o número de camadas ocultas, o número de neurônios e o tipo de função de ativação utilizada por cada neurônio.

2.3.1 Regressão

Como supracitado o objetivo do treinamento sobre o ponto de vista bayesiano consiste em encontrar o conjunto de pesos \underline{w} que maximize a probabilidade a posteriori $p(\underline{w} | Y, X)$, apresentada na equação 2.5. Sendo Y o conjunto de saída utilizado para o treinamento e X o conjunto de entradas utilizadas pelo modelo, essa probabilidade pode ser obtida:

$$p(\underline{w} | Y, X) = \frac{p(Y | \underline{w}, X) p(\underline{w} | X)}{p(Y | X)} \quad (2.5)$$

Na equação 2.5, entende-se $p(Y | X)$ como um fator de normalização que não será influenciado em função dos valores de \underline{w} . Além disso, posto que não será necessário modelar a probabilidade de $p(x)$, uma vez que essa probabilidade se encontra associada aos padrões de entradas e as *MLPs* não modelam tal variável, a partir desse momento esse conjunto será desconsiderado da notação.

Dessa forma, para o cálculo da probabilidade a posteriori do vetor \underline{w} , torna-se necessário o conhecimento da distribuição de probabilidade a priori do conjunto de pesos $p(\underline{w})$, bem como sua função de verossimilhança $p(Y | \underline{w})$, que é relacionada com a distribuição de probabilidade do ruído existente na saída.

Portanto, o primeiro desafio consiste em especificar uma distribuição para $p(\underline{w})$ que reflita a falta de conhecimento existente sobre o processo. Dessa forma, a escolha de uma distribuição gaussiana para $p(\underline{w})$ com média nula e matriz de covariância $\alpha^{-1} \underline{I}, \underline{I} \in \mathbb{R}^M \times \mathbb{R}^M$ igual a matriz identidade, como apresentada na equação 2.6, torna-se indicada.

$$p(\underline{w}) = \frac{1}{Z_w(\alpha)} e^{-\frac{\alpha}{2} \|\underline{w}\|^2} \quad (2.6)$$

$$Z_w(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{\frac{M}{2}}$$

Na equação 2.6 foi apresentado o hiperparâmetro $\alpha \in \mathbb{R}^+$, que nesse momento será pressuposto como constante e conhecido. Entretanto, ao longo do trabalho o cálculo desse hiperparâmetro será apresentado.

Com a escolha da distribuição gaussiana para $p(\underline{w})$, surge um algoritmo iterativo para o cálculo de \underline{w} e algumas simplificações podem ser adotadas para as análises subsequentes. Torna-se válido ressaltar que distribuições não informativas também podem ser utilizadas, todavia seu cálculo analítico é impraticável e com isso suas informações são levantadas através de simulações de Markov ou Monte Carlo [36], [37], que além de computacionalmente custosas, inviabilizam o algoritmo de seleção de entradas (*ARD*) que será apresentado [3].

Embora a escolha dessa distribuição simplifique a complexidade durante o desenvolvimento, isso não se torna adequado. Uma vez que as entradas podem apresentar comportamentos distintos, e o conjunto de pesos deverá refletir tais diferenças de comportamento. Ou seja, os pesos poderão apresentar comportamentos distintos entre si fato que até o momento não é permitido. À vista disso, surge a necessidade da escolha de uma distribuição $p(\underline{w})$ segundo agrupamentos específicos, com isso, o método de determinação automática da relevância, supracitado, irá surgir. Visto que o método *ARD* será apresentado posteriormente, para a continuação do desenvolvimento teórico do treinamento bayesiano será admitido uma única distribuição.

Conhecida a distribuição $p(\underline{w})$, torna-se necessário conhecer a função de verossimilhança do modelo. Como dito essa função encontra-se diretamente relacionada com os ruídos presentes nos dados. Suponha que o modelo seja definido pela equação 2.7, sendo ζ o ruído apresentado pelo modelo:

$$d(k) = F(k) + \zeta \quad (2.7)$$

Adotando como premissa que o ruído possui uma distribuição gaussiana com média nula e variância β^{-1} , pode-se encontrar a probabilidade de ocorrência de uma saída específica $d(k)$, dado seu conjunto de entradas (x_k) e o vetor de pesos (\underline{w}) , que é apresentado na equação 2.8. Levando em consideração que todos os dados pertencentes ao conjunto de

treinamento foram obtidos de maneira independente, a função de verossimilhança a ser utilizada será um produtório da probabilidade de ocorrência de cada saída, conforme apresentado e desenvolvida na equação 2.9.

$$p(d_k | \underline{x}_k, \underline{w}) = \frac{e^{-\left\{\frac{\beta}{2} [d_k - f(\underline{x}_k, \underline{w})]^2\right\}}}{\int e^{-\left\{\frac{\beta}{2} [d_k - f(\underline{x}_k, \underline{w})]^2\right\}} d d_k} \quad (2.8)$$

$$p(Y | \underline{w}) = \prod_{k=1}^N p(d_k | \underline{x}_k, \underline{w})$$

$$p(Y | \underline{w}) = \frac{1}{Z_Y(\beta)} e^{-\left\{\frac{\beta}{2} \sum_{k=1}^N [d_k - f(\underline{x}_k, \underline{w})]^2\right\}} \quad (2.9)$$

$$Z_Y(\beta) = \left(\frac{2\pi}{\beta}\right)^{\frac{N}{2}}$$

De maneira análoga a realizada na equação 2.6, β na equação 2.9 é um hiperparâmetro que será adotado como conhecido e fixo, sendo seu cálculo posteriormente apresentado.

De posse das equações 2.6 e 2.9, pode-se aplicar a regra de Bayes e calcular a probabilidade a posteriori de \underline{w} , dado o conjunto de saídas desejadas Y . Dessa forma, o seguinte resultado é encontrado:

$$p(\underline{w} | Y) = \frac{1}{Z_s} e^{-S(\underline{w})}$$

$$Z_s = \int e^{-S(\underline{w})} d \underline{w} \quad (2.10)$$

$$S(\underline{w}) = \frac{\beta}{2} \sum_{k=1}^N [d_k - f(\underline{x}_k, \underline{w})]^2 + \frac{\alpha}{2} \sum_{j=1}^M w_j^2$$

Como mencionado, uma das grandes motivações para a utilização do treinamento bayesiano aplicado as redes neurais é sua capacidade de lidar com o problema de *overfitting*, em função do controle de complexidade inerente ao algoritmo de treinamento.

Esse controle pode ser entendido analisando a equação do funcional $S(\underline{w})$ onde encontram-se presentes dois termos: o primeiro termo diz respeito ao ajuste da resposta do modelo aos dados disponíveis e a segunda parcela diz respeito à inserção do conhecimento prévio através da probabilidade a priori $p(\underline{w})$, o que garante um controle de magnitude do conjunto de pesos e mapeamentos suaves como proposto em sua formulação [25]. Desse modo, com a minimização do funcional $S(\underline{w})$ é encontrado um ponto de equilíbrio entre o ajuste do modelo aos dados disponíveis e o controle de magnitude dos pesos, fazendo com que no modelo não seja visualizado o problema de *overffiting* e garantindo que mapeamentos suaves possam ser gerados.

Anteriormente foi adotado o conhecimento dos hiperparâmetros α e β . Contudo tais hiperparâmetros não são conhecidos e devem ser adicionados ao problema de maximização. Logo, a probabilidade a posteriori a ser maximizada é:

$$p(\underline{\omega}, \alpha, \beta | Y) = p(Y | \underline{w}, \alpha, \beta) p(\underline{w}, \alpha, \beta) \quad (2.11)$$

Existem diversas abordagens para o cálculo desses hiperparâmetros. Na presente dissertação foi utilizado o método de aproximação das evidências, proposto por **Mackay** [28].

Mackay em sua abordagem parte do pressuposto que a probabilidade a posteriori $p(\alpha, \beta | Y)$ irá apresentar pouca dispersão ao redor dos valores mais prováveis para esses hiperparâmetros. Dessa forma, a equação 2.11 poderá ser simplificada dando origem a seguinte expressão:

$$p(\alpha, \beta | Y) = p(Y | \alpha, \beta) p(\alpha, \beta) \quad (2.12)$$

De maneira análoga à que ocorreu para a distribuição de probabilidade $p(\underline{w})$, torna-se necessário a inserção de algum conhecimento prévio sobre a distribuição $p(\alpha, \beta)$. Contudo, o único conhecimento prévio que poderá ser incluído nesta distribuição é a falta de conhecimento sobre ela. Logo, deverá ser adotado uma distribuição de probabilidade que seja insensível aos valores de α e β .

Na equação 2.12 foi omitido $p(Y)$, uma vez que essa probabilidade é independente dos hiperparâmetros em questão. Logo, a maximização da probabilidade a posteriori $p(\alpha, \beta | Y)$ torna-se dependente somente de maximizar a probabilidade dada $p(Y | \alpha, \beta)$, que é conhecida como evidência dos hiperparâmetros [25]. Essa probabilidade poderá ser

obtida através da equação 2.13, que será apresentada somente em sua forma simplificada. Para a simplificação dessa equação foi analisado que o hiperparâmetro α encontra-se relacionado somente com a probabilidade a priori do vetor de pesos \underline{w} e o hiperparâmetro β encontra-se relacionado com a distribuição dos ruídos, com isso a seguinte equação é encontrada:

$$p(Y|\alpha, \beta) = \int p(Y|\underline{w}, \beta) p(\underline{w}, \alpha) d\underline{w} \quad (2.13)$$

Utilizando as equações 2.6 e 2.9 na equação 2.13, pode-se obter a seguinte expressão:

$$p(Y|\alpha, \beta) = \frac{Z_s(\alpha, \beta)}{Z_y(\beta)Z_w(\alpha)} \quad (2.14)$$

$$Z_s(\alpha, \beta) = \int e^{-S(\underline{w})} d\underline{w}$$

Levando em consideração que o funcional $S(\underline{w})$ segue distribuição gaussiana, a probabilidade a posteriori, $p(\underline{w}|Y)$, é encontrada através de uma aproximação quadrática da série de Taylor do funcional $S(\underline{w})$ em torno do vetor \underline{w}^* (conjunto de pesos que minimiza o funcional $S(\underline{w})$), logo Z_s passa a ser dada por:

$$Z_s(\alpha, \beta) = e^{-S(\underline{w}^*)} (2\pi)^{\frac{M}{2}} \{\det [\underline{A}(\underline{w})|_{\underline{w}=\underline{w}^*}]\}^{\frac{1}{2}} \quad (2.15)$$

$$\underline{A}(\underline{w})|_{\underline{w}=\underline{w}^*} = \beta \underline{H}(\underline{w})|_{\underline{w}=\underline{w}^*} + \alpha \underline{I}$$

Sendo:

$\underline{H}(\underline{w})|_{\underline{w}=\underline{w}^*}$ ----- Matriz Hessiana do funcional $E_s(\underline{w})$
calculada utilizando \underline{w}^*

$\underline{I} \in \mathbb{R}^M \times \mathbb{R}^M$ ----- Matriz Identidade

Tomando como base o conjunto a equação 2.15 e utilizando-a em conjunto com as expressões explicitadas anteriormente para $Zw(\alpha)$ e $Zy(\beta)$, nas equações 2.6 e 2.9, pode-se encontrar o logaritmo neperiano da expressão 2.14 (equação 2.16). Logo, será necessário obter o valor que maximize esse logaritmo em relação a α e β , para isto basta derivar tal função em relação a esses parâmetros e as equações 2.17 e 2.18 irão surgir.

$$\ln p(Y|\alpha, \beta) = -\frac{\alpha}{2} \sum_{j=1}^M (w_j^*)^2 + \frac{\beta}{2} \sum_{k=1}^N [d_k - f(x_k, \underline{w})]^2 - \frac{1}{2} \ln \left\{ \det [\underline{A}] \right\} + \frac{M}{2} \ln \alpha + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) \quad (2.16)$$

$$\gamma = \alpha \sum_{j=1}^M (w_j^*)^2 = M - \text{trace} \left\{ \left[\underline{A}(\underline{w}) \Big|_{\underline{w}=\underline{w}^*}^{-1} \right] \right\} = \sum_{j=1}^M \frac{v_i}{v_i + \alpha} \quad (2.17)$$

$$N - \gamma = \beta \sum_{k=1}^N [d_k - f(x_k, \underline{w})]^2 \quad (2.18)$$

Abordando um pouco melhor sobre a equação 2.17, γ é o chamado número efetivo de parâmetros e v_i representa os autovalores da matriz hessiana $\underline{H}(\underline{w})$.

Algumas simplificações foram realizadas durante a apresentação das equações anteriores. Caso o leitor sinta a necessidade de uma maior conhecimento dessas simplificações as referências [3], [23], [38] apresentam maiores detalhes sobre o cálculo.

As equações 2.17 e 2.18 foram calculadas levando em consideração a aproximação quadrática do Funcional $S(\underline{w})$ em torno do seu ponto de mínimo, métodos de segunda ordem de treinamento de *MLP*'s, como o Levenberg-Marquardt, utilizam a cada iteração a aproximação do funcional do risco empírico ao redor do ponto de operação $\underline{w}(l)$. Dessa maneira, uma analogia a este método pode ser realizada, gerando um processo recursivo para a estimativa dos hiperparâmetros α e β :

$$\gamma(l+1) = M - \text{trace} \left\{ \left[\underline{A}(\underline{w}) \Big|_{\underline{w}=\underline{w}^*}^{-1} \right] \right\} = \sum_{j=1}^M \frac{v_i(l)}{v_i(l) + \alpha(l)} \quad (2.19)$$

$$\alpha(l+1) = \frac{\gamma(l+1)}{\sum_{j=1}^M [w_j(l)]^2} \quad (2.20)$$

$$\beta(l+1) = \frac{N - \gamma(l+1)}{\sum_{k=1}^N [d_k - f(\underline{x}_k, \underline{w})]^2} \quad (2.21)$$

Note que a escolha da distribuição de probabilidade a priori $p(\underline{w})$, apresenta algumas inconsistências em relação as propriedades de escalonamento dos mapeamentos a serem realizadas pelas *MLP*'s. Quando transformações lineares são aplicadas aos dados, os funcionais deverão dar origem a modelos similares aos modelos treinados com os dados originais. Consequentemente, a escolha de um funcional regularizador para todo conjunto de pesos \underline{w} , como apresentado na equação 2.6, não satisfaz essa propriedade, sendo necessário a escolha de outra distribuição à priori.

Além do que foi apresentado, é de conhecimento geral que as entradas podem apresentar comportamentos distintos dado as suas diferentes origens, ou seja, as entradas precisam se relacionar de forma explicativa somente com o resultado desejado e não com elas mesmas. À vista disso, intuitivamente pode-se verificar que a escolha de uma distribuição $p(\underline{w})$ para todo conjunto de pesos não se torna razoável. Em razão disso, para a resolução do problema apresentado, inúmeras distribuições gaussianas são adotadas, todas com média nula diferenciando-se somente na matriz de covariância. Logo, uma forma mais fidedigna de inserção do conhecimento prévio é vista.

Tomando como base o que foi supracitado, a probabilidade de ocorrência de um vetor \underline{w}_i pode ser descrita conforme a equação 2.22, apresentada a seguir.

$$p(\underline{w}_i) = \frac{1}{\left(\frac{2\pi}{\alpha_i}\right)^{\frac{M_i}{2}}} e^{-\frac{1}{2}\alpha_i \|\underline{w}_i\|^2} \quad (2.22)$$

Seja g o número de conjuntos nos quais os pesos são agrupados, e $\underline{w}_i \in \mathbb{R}^M$, w_i o vetor de pesos do i -ésimo grupo, contendo M_i elementos e α_i o hiperparâmetro associado a tal conjunto.

Agora considerando-se a independência entre os grupos de pesos, a distribuição de probabilidade a priori $p(\underline{w})$ passa a ser dada pelo produto das distribuições de probabilidade de cada um dos grupos, conforme apresentado na expressão 2.23.

$$p(\underline{w}) = \prod_{i=1}^g p(\underline{w}_i) = \frac{1}{\prod_{i=1}^g \left(\frac{2\pi}{\alpha_i}\right)^{\frac{M_i}{2}}} e^{-\frac{1}{2}\sum_{i=1}^g (\alpha_i \|\underline{w}_i\|^2)} \quad (2.23)$$

Com a nova distribuição de probabilidade a priori de $p(\underline{w})$, torna-se necessário atualizar a expressão 2.10 para a obtenção de um novo funcional $S(\underline{w})$ a ser minimizado. Logo, a maximização da probabilidade a posteriori $p(\underline{w}|Y)$ passa a ser dada pela minimização do novo funcional dado pela equação 2.24.

$$S(\underline{w}) = \frac{\beta}{2} \sum_{k=1}^N [d_k - f(\underline{x}_k, \underline{w})]^2 + \frac{1}{2} \sum_{i=1}^g (\alpha_i \sum_{j=1}^{M_i} w_{ij}^2) \quad (2.24)$$

Mantendo a uniformidade, torna-se necessário a atualização do logaritmo natural da evidência dado que diferentes α 's serão utilizados. Isto posto, as seguintes expressões são encontradas:

$$\gamma_i = \alpha_i \sum_{j=1}^{M_i} w_{ij}^{*2} = M_i - \text{trace}\{\underline{\underline{B}}_i\} \quad (2.25)$$

$$\gamma = \sum_{i=1}^g \gamma_i \quad (2.26)$$

$$\underline{\underline{B}}_i = \left[\underline{\underline{A}}(\underline{w}) \Big|_{\underline{w}=\underline{w}^*}^{-1} \right] \underline{\underline{I}}_i \quad (2.27)$$

Note que, na expressão 2.27, $\underline{\underline{I}}_i \in \mathbb{R}^M \times \mathbb{R}^M$ é uma matriz com elementos nulos, porém com elementos iguais a 1 somente para os componentes do vetor \underline{w} pertencentes ao i -ésimo grupo de pesos, e γ_i o número de parâmetros efetivos associados a este conjunto.

Mantendo a analogia, as expressões para o cálculo do hiperparâmetro α são atualizadas:

$$\underline{\underline{B}}_i(l) = \left[\underline{\underline{A}}(\underline{w}) \Big|_{\underline{w}=\underline{w}(l)}^{-1} \right] \underline{\underline{I}}_i \quad (2.28)$$

$$\gamma_i(l+1) = M_i - \text{trace}\{\underline{\underline{B}}_i(l)\} \quad (2.29)$$

$$\alpha_i(l+1) = \frac{\gamma_i(l+1)}{||\underline{w}_i(l)||^2} \quad (2.30)$$

$$\gamma(l+1) = \sum_{i=1}^g \gamma_i(l+1) \quad (2.31)$$

Portanto, todo o algoritmo de treinamento bayesiano voltado para regressão encontra-se apresentado, realizando uma síntese de tudo que foi visto em forma de fluxograma, tem-se o seguinte [3]:

1. Faça $l = 0$;
2. Inicialize os parâmetros $\underline{w}(l)$ e os hiperparâmetros $\alpha(l)$ e $\beta(l)$;
3. Utilizando alguma técnica de otimização atualize o vetor de pesos $\underline{w}(l+1)$, através da minimização do funcional $S(w)$;
4. Verifique se o critério de parada foi atendido, caso contrário continue o algoritmo;
5. Calcule a matriz hessiana $\underline{\underline{H}}(\underline{w}) \Big|_{\underline{w}=\underline{w}(l)}$;
6. Atualize os hiperparâmetros de cada grupo α_i e β ;
7. Faça $l = l + 1$ e retorne para o passo 3;

Anteriormente, foi exposto que o treinamento bayesiano apresentava algumas vantagens frente aos treinamentos convencionais. Uma delas se tornou evidente ao longo do processo, pois em nem um momento foi apresentado a necessidade de existência de um conjunto de validação além dos valores utilizados para treinamento. Isso ocorre pela estimativa automática de diversos parâmetros de regularização, através do cálculo dos hiperparâmetros $\underline{\alpha}$ e β .

Além disso, através desse processo surge um método capaz de selecionar as entradas relevantes para o processo, denominado *ARD*, que será detalhado a seguir. Outra possibilidade que surge é a de seleção de estruturas, subentenda-se, escolha do número de neurônios na camada oculta. No presente trabalho não será apresentado o processo de

seleção para as estruturas, caso o leitor deseje conhecer melhor esse processo indica-se a leitura da referência [3].

2.3.2 Classificação

Embora o treinamento bayesiano voltado para classificação seja muito semelhante ao treinamento para regressão, algumas ressalvas precisam ser realizadas sobre o processo. Dessa maneira, torna-se necessário o desenvolvimento do presente tópico, para que o leitor possa compreender da melhor forma o algoritmo utilizado.

O processo de classificação de padrões consiste basicamente em agrupar os padrões numa categoria ou classe de acordo com um parâmetro pré determinado. Logo, considerando C classes mutualmente exclusivas entre elas, ou seja, um padrão poderá pertencer somente a uma classe, seja o conjunto $D = \{X, Y\}$ tal que $X = \{\underline{x}_1, \dots, \underline{x}_N\}$, $\underline{x} \in \mathbb{R}^n$, represente o conjunto de entradas e $Y = \{\underline{d}_1, \dots, \underline{d}_N\}$, $\underline{d} \in [0,1]^C$, represente o conjunto de saída, para os padrões pertencentes a mesma classe C_j , $\underline{d}_{ij} = 1$ e $\underline{d}_{jk} = 0$, $k = 1, \dots, C$, $k \neq j$ e $\sum_{k=1}^C \underline{d}_{ik} = 1$. Essa relação garante que os padrões pertencem somente a uma classe, como informado anteriormente.

Neste contexto, realizando uma analogia com o treinamento bayesiano para regressão, o mesmo objetivo de maximizar a probabilidade a posteriori do conjunto de pesos é utilizado. Com isso, o desafio de escolher uma distribuição de probabilidade que reproduza o desconhecimento prévio sobre a solução, também se faz necessário. Assim será adotada uma distribuição gaussiana com média nula e matriz de covariância $\alpha^{-1}I$, onde $\alpha \in \mathbb{R}^+$, como realizado no treinamento previamente apresentado. O cálculo do hiperparâmetro α também será apresentado a seguir, sendo nesse momento adotado o seu conhecimento. Logo, assim como no processo de regressão a equação 2.22 é encontrada, não precisando assim ser reapresentada.

A função de verossimilhança $p(Y|X, \underline{w})$ relaciona-se com a probabilidade de cada padrão pertencer a cada uma das classes. Nesse ponto é identificada a primeira diferença em relação ao treinamento voltado para a regressão. Dado que as saídas do modelo não possuem interpretações estatísticas, ou seja, somente valores numéricos são apresentados na saída do modelo, torna-se necessário um pré processamento desses valores de modo a possibilitar tal interpretação estatística como resposta do processo. Dessa maneira, seja

y_k a k -ésima saída do modelo, representando a probabilidade de um determinado padrão \underline{x}_i pertencer à classe C_k , conforme mostra a expressão 2.32:

$$p\left(\underline{y}_i \in C_k \mid \underline{x}_i\right) = f_k\left(\underline{x}_i, \underline{w}\right) \quad (2.32)$$

Para uma interpretação estatística da saída, seja as classes mutuamente exclusivas $0 \leq y_k \leq 1$ e $\sum_{k=1}^C f_k\left(\underline{x}_i, \underline{w}\right) = 1$ para um dado ponto \underline{x}_i , de modo que as saídas da rede serão obtidas pela expressão 2.33 [25]:

$$y_{ik} = f_k\left(\underline{x}_i, \underline{w}\right) = \frac{\delta_k}{\sum_{i=1}^k \delta_i} \quad (2.33)$$

$$\delta_k = \phi_{saída} \left[\sum_{j=1}^m \varpi_{kj} \phi_{oculta} \left(\sum_{l=1}^n \varpi_{jl} x_l + b_j \right) + b_k \right] \quad (2.34)$$

Onde δ_k representa a saída da função de ativação do k -ésimo neurônio e y_{ik} representa a probabilidade do padrão \underline{x}_i pertencer a classe C_k . Com isso, a estrutura da rede *MLP* utilizada para o módulo de classificação pode ser apresentada como na figura 4.

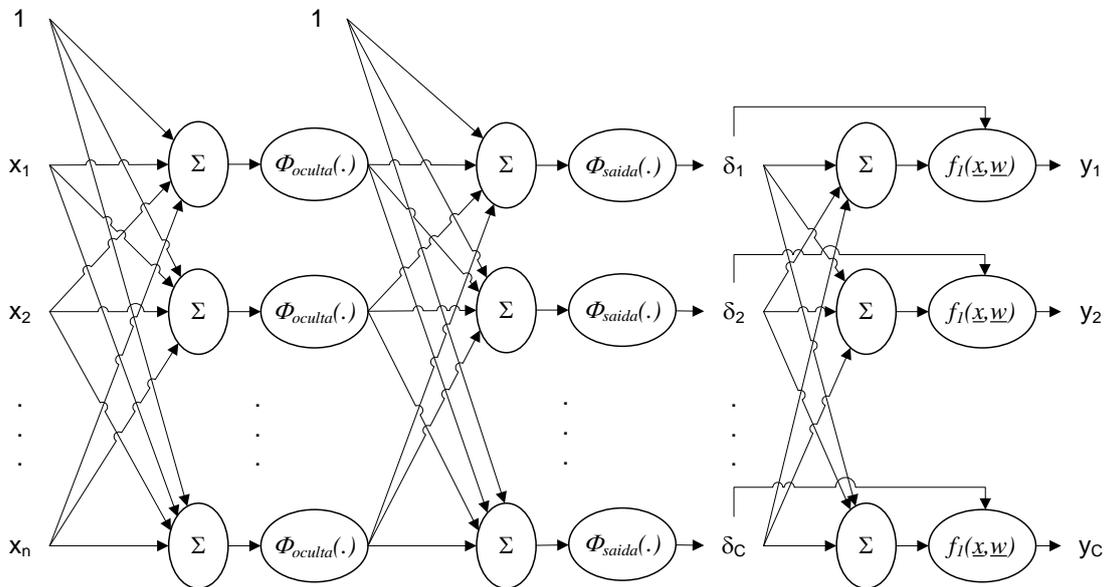


Figura 4 - Estrutura do MLP utilizado para classificação de padrões [23], [38]

Assim sendo, a probabilidade a posteriori de ocorrência de um vetor \underline{d}_i , dados os padrões \underline{x}_i e um vetor de parâmetros \underline{w} , passa a ser dada por:

$$p(\underline{d}_i|\underline{x}_i, \underline{w}) = \prod_{k=1}^C [f_k(\underline{x}_i, \underline{w})]^{d_{ik}} \quad (2.35)$$

Adotando a mesma premissa utilizada no problema de regressão, que os padrões de saída são independentes e identicamente distribuídos, a probabilidade de ocorrência da saída Y dado seu conjunto de entradas X e o vetor de pesos \underline{w} , pode ser descrita da seguinte forma:

$$p(Y|X, \underline{w}) = \prod_{i=1}^N \prod_{k=1}^C [f_k(\underline{x}_i, \underline{w})]^{d_{ik}} \quad (2.36)$$

Com o auxílio da regra de Bayes, pode-se inferir a probabilidade a posteriori de \underline{w} , $p(\underline{w}|X, Y)$, dado seu conjunto de entradas e saídas. A expressão 2.37 apresenta tal função:

$$p(\underline{w}|X, Y) = \frac{1}{Z_S} e^{-S_c(\underline{w})}$$

$$Z_S = \int e^{-S_c(\underline{w})} d\underline{w} \quad (2.37)$$

$$S_c(\underline{w}) = \sum_{i=1}^N \sum_{k=1}^C d_{ik} \ln[f_k(\underline{x}_i, \underline{w})] + \frac{1}{2} \sum_{i=1}^g \alpha_i \sum_{j=1}^M w_{ij}^2 = E_D^C(\underline{w}) + \frac{\alpha}{2} E_C(\underline{w})$$

Note que assim como no processo de regressão, o funcional $S_c(\underline{w})$ é composto por duas parcelas, sendo a primeira responsável por garantir o ajuste do modelo aos dados, evitando assim o problema de *overfitting*, e a segunda parcela responsável pela geração de mapeamentos suaves.

Em vista do que foi supracitado, o algoritmo de treinamento bayesiano aplicado as *MLP's*, voltado para classificação, pode ser resumido da seguinte forma:

1. Faça $l = 0$.
2. Inicialize o vetor de parâmetros $\underline{w}(l)$ e o vetor de hiperparâmetros $\alpha_i(l)$.

3. Utilizando alguma técnica de otimização (por exemplo, um algoritmo de retro propagação do erro), atualize o vetor de parâmetros $\underline{w}(l + 1)$ através da minimização do funcional $S(\underline{w})$ dado pela expressão;
4. Se o critério de parada for atendido, encerre o algoritmo. Do contrário, vá para o passo 5.
5. Atualize os hiperparâmetros $\alpha_i(l + 1)$ utilizando as expressões;
6. Faça $l = l + 1$ e retorne ao passo 3.

2.3.3 Determinação automática de relevância

Com a utilização de funcionais regularizadores durante o treinamento bayesiano, como apresentado nas seções anteriores, as estruturas com parâmetros de pequena magnitude são favorecidas uma vez que produzem mapeamentos suaves. Com isso, é esperado encontrar o equilíbrio entre o ajuste do modelo aos dados e a suavidade do mapeamento, sendo isto de responsabilidade dos hiperparâmetros $\underline{\alpha}$ e β . No presente tópico uma discussão mais detalhada sobre o hiperparâmetro $\underline{\alpha}$ será realizada, uma vez que se encontra sobre sua responsabilidade o controle de magnitude dos conjuntos de pesos do modelo, o que o coloca diretamente relacionado com a importância de determinada variável para a obtenção da resposta desejada.

Ao analisar a equação 2.23 sobre ponto de vista de otimização, torna-se claro que quanto maior for o funcional $\underline{\alpha}_i$ menor será a magnitude dos pesos associados a este funcional. Analisando do ponto de vista probabilístico, dado que o conjunto de pesos é regido por uma distribuição gaussiana com média nula e matriz de covariância regida por $\underline{\alpha}_i$, um elevado valor desse hiperparâmetro pode ser interpretado como uma diminuição da incerteza da informação prévia e com isso tais dados terão uma baixa influência no cálculo da probabilidade a posteriori $p(\underline{w}|Y)$. Ou seja, quanto maior for $\underline{\alpha}_i$, menor a magnitude de \underline{w}_i obtido através da maximização de $p(\underline{w}|Y)$.

Portanto, a análise do valor de cada elemento α_i permitir inferir sobre o impacto dessa entrada no cálculo da saída do modelo e dessa maneira verifica-se a importância da variável em questão. Este procedimento é conhecido como determinação automática da relevância do inglês *ARD* (*automatic relevance determination*). Nesse procedimento às entradas mais relevantes são atribuídos pesos com magnitudes elevadas e às entradas menos relevantes são atribuídos conjuntos de pesos com menor magnitude. Assim,

realizando uma ordenação dos hiperparâmetros α_i é possível verificar qual variável de entrada possui maior relevância para a saída. Contudo, o grande desafio agora consiste em estabelecer um limiar de significância para as variáveis de entradas, uma vez que tal método informa somente a ordem de relevância, não apresentando de forma direta qual variável poderá ser desconsiderada.

Para o levantamento desse limiar de significância, uma variável de prova é adicionada ao conjunto de entradas do modelo. Com essa inserção, torna-se possível verificar quais variáveis estão se correlacionando com a saída e quais poderão ser desconsideradas durante o processo. Logo, o treinamento é realizado com esse espaço de entradas estendido e as variáveis que apresentarem α_i maiores que a variável de prova (α_0), são descartadas ao final do processo. Torna-se necessário ressaltar que esse tratamento deverá ser realizado de forma separada para as variáveis discretas e contínuas. Maiores detalhes podem ser encontrados em [1], [3].

2.4 Aprendizado não supervisionado

Anteriormente, foi apresentada ao leitor as redes neurais artificiais e seu processo de aprendizado, através da aplicação do treinamento bayesiano. Nesse processo torna-se evidente a necessidade da existência de padrões de entradas e saídas, ou seja, para cada conjunto de entradas apresentada ao modelo existe um conjunto de saídas associados a esses valores. Logo, pode-se classificar o processo de treinamento das RNA's como algoritmos de aprendizado supervisionados.

Todavia, em alguns processos a definição de classes através de atributos específicos a ela se torna um tanto quanto inviável, seja pela falta de prévio para classificação dessas informações ou até mesmo pelo elevado número de amostras existentes. Com isso, surgem as técnicas de clusterização, que são algoritmos de aprendizado de máquina não supervisionados capazes de agrupar informações, através de alguma medida de similaridade, sem que nenhuma suposição prévia seja realizada sobre os dados [39]. Ou seja, a tarefa de um algoritmo não supervisionado é identificar instâncias de dados que possam ser agrupadas em conjuntos através de medidas de similaridades, sem que nem uma suposição a priori seja realizada respeito desses grupos. Para isto, são utilizados critérios de informação, que são medidas que visam quantificar as semelhanças ou diferenças presentes no conjunto de dados [40].

Torna-se necessário frisar que algumas propriedades do processo de clusterização deverão ser atendidas, tais como: a não superposição entre os grupos (a), a não existência de clusters vazios (b) e a união de todos os clusters representa o conjunto de dados apresentados para classificação (c). Matematicamente:

$$a) C_i \cap C_j = \emptyset, \forall i \neq j, 1 \leq i \leq k, 1 \leq j \leq k$$

$$b) C_i \neq \emptyset, \forall i, 1 \leq i \leq k$$

$$c) C_1 \cup C_2 \cup C_3 \cup \dots \cup C_k = X$$

Devido à natureza complexa do problema de agrupamento, diversos trabalhos são propostos na literatura. Em geral, alguns pontos importantes são destacados durante a aplicação de tais métodos tais como [41]–[44]:

Passo 1 - Escolha dos dados a serem clusterizados de forma que representem informações efetivamente relevantes para o processo de classificação;

Passo 2 - Selecionar Medidas de similaridade ou dissimilaridade adequadas ao problema em questão. Em geral tais medidas são distâncias entre pares do conjunto de dados a serem clusterizados, definidas de acordo com o conhecimento/experiência do especialista humano;

Passo 3 - Se possível, escolher mais de um algoritmo de agrupamento para realização da tarefa, uma vez que os resultados podem divergir entre os algoritmos. Com isso, fica a carga do especialista a escolha do melhor método;

Passo 4 - Interpretar, avaliar e validar os resultados do processo de agrupamento. Ao final do processo os grupos gerados deverão ser avaliados e validados para ratificar, ou não, sua correteude.

Ressalta-se que no presente trabalho não serão apresentadas formas de validação do processo de clusterização. Contudo, caso o leitor sinta a necessidade de maiores informações, as referências [45], [46] podem suprir tal necessidade.

2.4.1 Medidas de Similaridade

Como supracitado, o relacionamento entre os objetos é descrito através de medidas de similaridade ou dissimilaridade. Visto que o conjunto de dados pode ser

interpretado como um conjunto de pontos em um espaço k-dimensional, medidas de similaridade podem ser utilizadas como métricas de distância entre dois pontos [47].

Seja $\underline{\underline{D}}$ uma matriz de dissimilaridade, definida por uma das equações que será apresentada a seguir, onde cada elemento dessa matriz irá representar uma distância entre pares de objetos:

$$D = \begin{bmatrix} 0 & d(1,2) & d(1,3) & \dots & d(1,p) \\ d(2,1) & 0 & d(2,3) & \dots & d(2,p) \\ d(3,1) & d(3,2) & 0 & \dots & d(3,p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d(n,1) & d(n,2) & d(n,3) & \dots & 0 \end{bmatrix} \quad (2.38)$$

Quanto menor for a distância $d(i,j)$ maior será a similaridade entre os dados. Em contrapartida, quanto maior for a distância $d(i,j)$ menor será essa similaridade.

Posto que a similaridade é informada através da distância entre dois pontos quando trabalhado em espaços multidimensionais, para que seja assegurada a correta quantificação dessa similaridade, torna-se necessária a normalização dos dados uma vez que o atributo que variar numa escala superior poderá sobrepujar os valores dos demais atributos [38].

Na literatura são apresentadas diversas métricas de similaridades, por exemplo: Minkowski, Hamming, Euclidiana. As expressões a seguir irão apresentar as distâncias citadas anteriormente, respectivamente:

$$d(i,j) = \left(\sum_{l=1}^p |x_{il} - x_{jl}|^q \right)^{1/q} \quad (2.39)$$

$$d(i,j) = \sum_{l=1}^p |x_{il} - x_{jl}| \quad (2.40)$$

$$d(i,j) = \sqrt{\sum_{l=1}^p |x_{il} - x_{jl}|^2} \quad (2.41)$$

Para a escolha das funções de distâncias a serem utilizadas, deverão ser obedecidas algumas propriedades, que serão apresentadas nas equações a seguir. A

primeira propriedade equação 2.42 garante que a medição da distância do mesmo ponto seja zero, já a segunda propriedade equação 2.43 assegura a simetria da matriz de similaridade, ou seja, a distância entre os pontos não varia de acordo com o ponto de partida escolhido e a terceira e última propriedade apresentada, equação 2.44, garante que a menor distância entre dois pontos é uma linha reta.

$$D(x, x) = 0 \quad (2.42)$$

$$D(x, y) = D(y, x) \quad (2.43)$$

$$D(x, y) \leq D(x, z) + D(z, y) \quad (2.44)$$

2.4.2 K médias (*K-means*)

A criação de técnicas de agrupamento através de medidas de similaridade é uma tarefa que pode ser realizada trivialmente por qualquer ser humano, porém a criação de um algoritmo confiável e passível de ser implementado computacionalmente torna essa tarefa desafiadora [46] e [48]. MacQueen [49] em 1967 propôs o método denominado K-médias (*K-means*), método que será utilizado nessa dissertação e que provavelmente se tornou um dos métodos mais conhecidos e utilizados para clusterização de dados no mundo [50].

De forma geral, o método K-médias pode ser caracterizado como um algoritmo particional, uma vez que, dado um conjunto de dados (D), o algoritmo irá encontrar K subdivisões de tal conjunto. Para isto, deverá ser definido pelo usuário a quantidade de partições que os dados deverão ser divididos, assim como a medida de similaridade a ser utilizada para esse agrupamento.

A parti das informações definidas pelo usuário, o método irá agrupar os dados de forma a minimizar o somatório da distância dos pontos em relação ao seu centroide, equação 2.45 [51], [52]. Ao final desse processo, os dados semelhantes estarão reunidos no mesmo grupo e os centroides desses K grupos poderão ser interpretados como um valor central (ou médio) do conjunto de pontos pertencentes ao grupo. Torna-se valido frisar que o método K-médias é um algoritmo iterativo, sendo encerrado quando a

distância entre os centroides e suas instâncias não variar ou quando o número máximo de iterações for alcançado.

$$d(p_i, C) = \frac{1}{n} \sum_{i=1}^n d(p_i, C)^2 \quad (2.45)$$

Como supracitado, o método de K-médias se torna um dos métodos de clusterização mais utilizados ao redor do mundo, isto pode ser justificado pela simplicidade de programação e rápida convergência do método. Originalmente, o método é composto por quatro partes, conforme descrito a seguir e ilustrado na figura 5:

1. Primeiramente escolhe-se k -centroides (sementes na figura 5) para que o processo seja inicializado;
2. Calcula-se as distâncias euclidianas de cada agrupamento em relação aos centroides correntes, agrupando o elemento ao grupo que possuir menor distância em relação ao centroide. Nessa parte do algoritmo é observado o maior esforço computacional, uma vez que existindo N pontos e K centroides, são calculadas $N * K$ distâncias;
3. Recalcula-se os centroides para o novo grupo e repete-se o passo 2, considerando os novos centroides;
4. Os passos 2 e 3 devem ser repetidos até que todos os elementos amostrais estejam devidamente alocados em seus grupos. De forma prática, esse processo irá se encerrar após um número determinado de iterações ou devido a não alteração dos centroides, o que indica a separação dos dados.

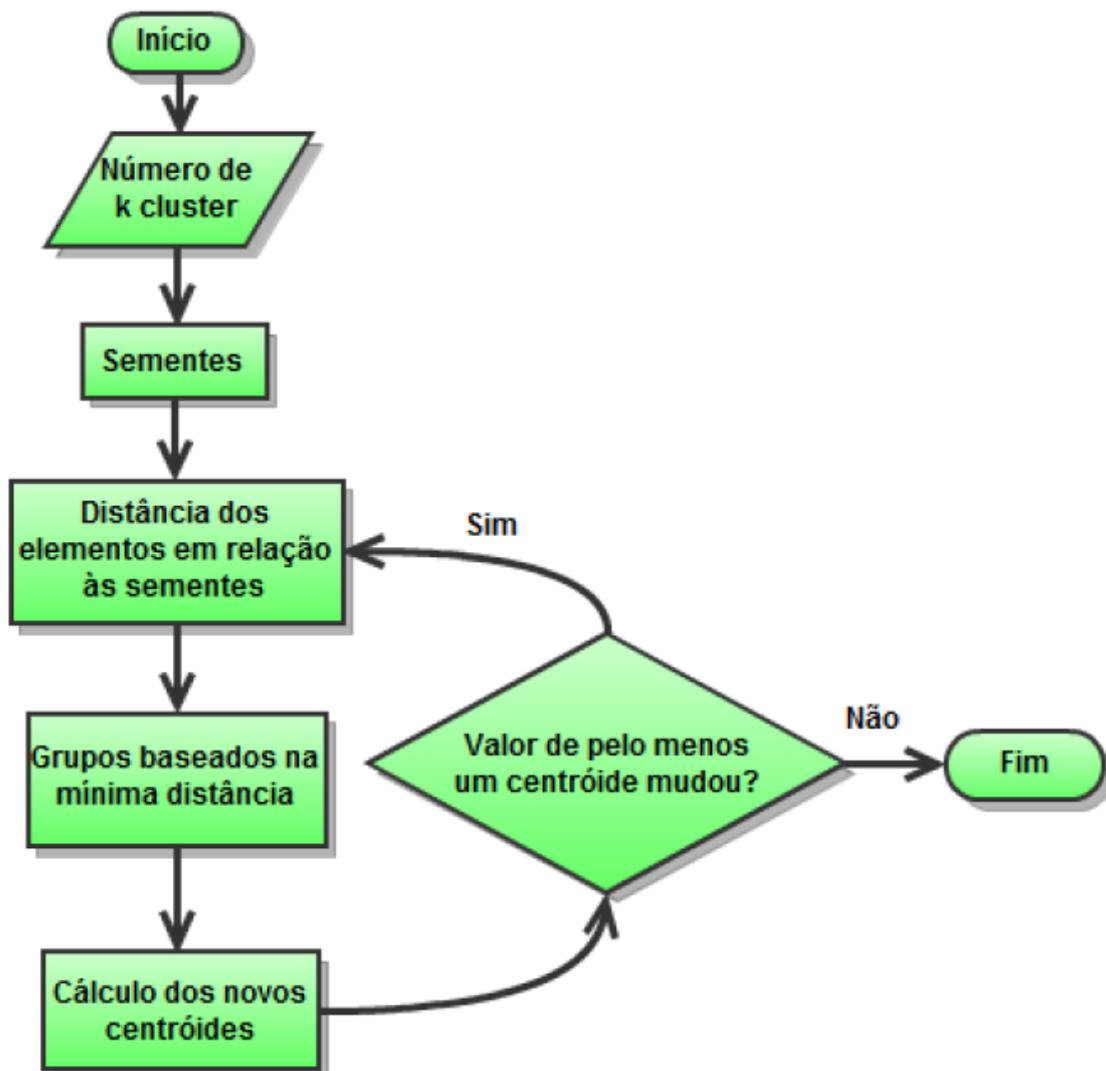


Figura 5 - Fluxograma do Algoritmo do K-médias [45]

Apesar desse método apresentar uma de simples implementação e uma rápida convergência, como dito, algumas limitações podem ser levantadas durante sua utilização. A primeira delas é a definição aleatória do conjunto de centroides iniciais, com isso, diferentes resultados podem ser encontrados ao se utilizar o mesmo conjunto de dados. Para a resolução de tal problema, diversos autores propõem modificações no método original, no presente trabalho não será tratado esse problema e maiores detalhes sobre a resolução dessa limitação podem ser encontradas na referência [53]. A segunda limitação do método é a necessidade de definição pelo usuário da quantidade de grupos que os dados serão divididos. Em alguns casos, nada pode ser inferido previamente sobre os dados, em função do desconhecimento do especialista. Dessa maneira, torna-se necessário a realização de diversos testes para a verificação da quantidade ideal de

agrupamentos, tornando imprescindível a interferência por parte do especialista durante o processo. Nesta dissertação será utilizado o método X-médias, apresentado na próxima seção, para a resolução desse problema.

2.4.3 X-Médias

Em função das limitações supracitadas do método K -médias, principalmente no que tange a escolha da quantidade de grupos a serem utilizados para o particionamento dos dados dado o desconhecimento do analista sobre a melhor forma de agregação dos dados, torna-se válido o desenvolvimento de um método que seja capaz de identificar a quantidade de partições com a mínima interferência por parte do analista. Na literatura são encontradas diferentes metodologias para isto. No presente trabalho será utilizado o método chamado X-médias (do inglês *X-means*), que pode ser entendido como uma variação do K -médias convencional.

Nesta dissertação serão utilizadas duas formas de realização do X-médias, ambas as metodologias utilizam o mesmo critério de informação, o Critério de informação Bayesiano (do inglês *BIC-Bayesian Information Criterion*), para a definição da quantidade ideal de *clusters*. Dessa maneira, será apresentado primeiramente toda a formulação teórica sobre o *BIC*, para em seguida os métodos serem apresentados.

Em analogia à inferência bayesiana aplicada as *MLPs*, o *BIC* é uma métrica que busca equilibrar o ajuste do modelo aos dados com a sua complexidade [54] e por isso foi escolhido como critério de avaliação da clusterização. Para um melhor embasamento do leitor, será apresentado inicialmente a nomenclatura utilizada, onde i é o índice do centroide mais próximo para o i -ésimo ponto de dados, $\underline{\mu}_{(i)}$ é o centroide associado ao i -ésimo ponto durante a iteração, D representa o conjunto de dados de entrada, $D_i \subseteq D$ é o conjunto de pontos que tem $\underline{\mu}_{(i)}$ como centróide, $R = |D|$, $R_i = |D_i|$, M é a dimensão e a matriz de covariância gaussiana dada por $\sum \text{diag}(\sigma^2)$.

Apresentada a nomenclatura utilizada, pode-se expor o cálculo do *BIC*, definido pela equação 2.46.

$$BIC(M_j) = \hat{l}_j(D) - \frac{p_j}{2} \cdot \log R \quad (2.46)$$

Onde:

- M_j representa o modelo utilizado;
- $\hat{l}_j(D)$ é a probabilidade de registro dos dados de acordo com o j -ésimo modelo e tomado no ponto de máxima verossimilhança;
- p_j é o número de parâmetros em M_j e R é o número de pontos.

Adotando como premissa que cada clusters representa uma gaussiana esférica idêntica, pode-se obter a estimativa de máxima verossimilhança para a variância, conforme apresentado na equação 2.47. Além disso, as probabilidades pontuais podem ser obtidas através da equação 2.48 e o log da função de verossimilhança dos registros pode ser verificada pela equação 2.49:

$$\hat{\sigma}^2 = \frac{1}{R - K} \sum_i (x_i - \underline{\mu}_{(i)})^2 \quad (2.47)$$

$$\hat{P}(x_i) = \frac{R_{(i)}}{R} \cdot \frac{1}{\sqrt{2\pi\hat{\sigma}^M}} e^{-\frac{1}{2\hat{\sigma}^2} \|x_i - \underline{\mu}_{(i)}\|^2} \quad (2.48)$$

$$l(D) = \log \prod_i P(x_i) = \sum_i \left(\log \frac{1}{\sqrt{2\pi\hat{\sigma}^M}} - \frac{1}{2\hat{\sigma}^2} \|x_i - \underline{\mu}_{(i)}\|^2 + \log \frac{R_{(i)}}{R} \right) \quad (2.49)$$

Considerando-se apenas o conjunto de pontos D_n que pertencem ao centroide n e associando as equações 2.47 e 2.49, encontra-se a equação 2.50, que descreve o log da função de verossimilhança dos dados de acordo com o j -ésimo modelo.

$$\hat{l}(D_n) = -\frac{R_n}{R} \log(2\pi) - \frac{R_n \cdot M}{2} \log(\hat{\sigma}^2) - \frac{R_n - K}{2} + R_n \log R_n - R_n \log R \quad (2.50)$$

Dessa forma, pode-se calcular através do somatório do log da função de verossimilhança a probabilidade de registros dos dados \hat{l}_j . Por fim, para o cálculo do *BIC*, torna-se necessário a definição da quantidade de parâmetros livres p_j . Esse valor é calculado através do somatório descrito na equação 2.51:

$$p_j = (K - 1) + (M * K) + 1 \quad (2.51)$$

Onde:

- M representa o número de dimensões do conjunto de Dados;
- K representa o número de *clusters* utilizado;

Dessa forma, com a apresentação do critério de informação utilizado para a escolha da quantidade ideal de agrupamentos, pode-se expor os métodos utilizados nessa dissertação para a clusterização. Esta exposição será realizada de forma simplificada, caso o leitor sinta a necessidade maiores detalhes podem ser encontrados nas referências [54]–[56].

2.4.3.1 Método proposto por *Dan Pelleg e Andrew Moore* [56]

A primeira formulação do X-medias que será apresentada foi proposto por Pelleg e Moore [56], nesse método roda-se o K-médias por K_{max} vezes e calcula-se o valor do BIC para cada um dos modelos gerados. Ao final do processo, o modelo que possuir o maior valor de BIC, será apresentado como modelo que melhor representa o conjunto de dados. Ressalta-se que o valor do K_{max} é definido pelo usuário.

Ao se utilizar essa metodologia, torna-se válido frisar o trade-off existente entre a melhor representação do conjunto de dados e o esforço computacional. Caso seja escolhido um valor muito elevado de K_{max} , o esforço computacional será elevado, contudo existe a maior probabilidade de melhor representação do conjunto de dados. Porém, caso seja adotado um valor muito pequeno de K_{max} o método será executado rapidamente, todavia os dados poderão ser agrupados de forma incorreta. Infelizmente, não foi encontrado na literatura trabalhos que apresentem técnicas para definição automática do parâmetro K_{max} .

Para a execução do método proposto, os seguintes passos deverão ser seguidos:

- Passo 1 - Definir o número máximo de centros desejados K_{max} ;
- Passo 2 - Roda-se o método K-médias em loop, variando K de 1 até K_{max} , ou seja, K modelos de agrupamentos são gerados (M_j);
- Passo 3 - Calcula-se o BIC (M_j) para cada um dos modelos M_j ;
- Passo 4 - Ao final do método verifica-se o maior BIC entre os presentes na lista e com isso será descoberto o número de centros que melhor divide a base de dados apresentada ao método.

2.4.3.2 Método proposto por *Tsunenori Ishioka* [41]

Ao contrário do método proposto por Pelleg, Ishioka [41] propõem um método de X-Médias baseado na competição entre pais e filhos, isto é, escolhe-se inicialmente o valor mínimo de grupos para que os dados sejam divididos, em geral $K_0 = 2$, e em seguida, dois agrupamentos filhos são gerados a partir desses pais ($K = 2$). Neste ponto ocorre a competição pelo espaço entre o conjunto pai e os conjuntos filhos e para a escolha do conjunto que melhor representa os dados utiliza-se o *BIC*. Caso seja verificado que o agrupamento pai representa melhor o espaço, seus filhos são descartados. Caso contrário, os agrupamentos filhos se tornam agrupamentos pais e todo processo é novamente realizado. Conseqüentemente, ao final do processo é esperado que os clusters adotados sejam capazes de representar de forma satisfatória todo o conjunto de dados.

Para a execução do método proposto, os seguintes passos deverão ser seguidos:

- Passo 1 - Define-se o número de *clusters* iniciais, em geral $K_0 = 2$ e aplica-se o método de K-médias;
- Passo 2 - Para cada clusters aplica-se novamente o método de K-médias fazendo $K = 2$. Nesse momento os filhos são gerados;
- Passo 3 - Calcula-se o *BIC* para cada um dos modelos, pai (*BIC*) e filhos (*BIC'*), se $BIC < BIC'$ os filhos são “mortos” e o pai não é mais dividido, caso o contrário a divisão continua e os clusters filhos se tornam novos pais. Este processo será encerrado somente quando o *cluster* pai parar de ser dividido.
- Passo 4 - Os passos 2 e 3 são repetidos até a “convergência” do método;
- Passo 5 - Após a convergência do método, o número ideal de clusters será encontrado.

2.5 Misturas de Gaussianas

O presente tópico possui como objetivo realizar uma breve apresentação sobre os principais conceitos relacionados ao modelo de misturas de gaussianas (do inglês *Gaussian Mixture Model (GMM)*) de acordo com sua relevância para essa dissertação. Para um estudo mais criterioso sobre o assunto, sugere-se a leitura dos trabalhos [57], [58].

O modelo de Misturas de Gaussianas pode ser definido como uma família de distribuição formada pela composição de mais de uma distribuição básica [59] e [60]. Esta composição é formada pela ponderação entre as distribuições de probabilidades D_1, D_2, \dots, D_n e suas respectivas probabilidades de ocorrências $\pi_1, \pi_2, \dots, \pi_n$. Ou seja, um *GMM* é definido por uma função de densidade probabilidade $f(\cdot)$, tal que:

$$f(x_n | \psi) = \sum_{k=1}^K \pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k) \quad (2.51)$$

$$\mathcal{N}(x_n | \mu_k, \Sigma_k) = \frac{1}{2\pi^{\frac{M}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right\} \quad (2.52)$$

Sendo π_k a probabilidade a priori de ocorrência de uma determinada distribuição, as seguintes restrições devem ser atendidas:

$$\pi_k \geq 0 \quad (2.53)$$

$$\sum_{k=1}^K \pi_k = 1 \quad (2.54)$$

Sendo K é o número de componentes da mistura, isto é, K descreve o número de distribuições que serão utilizadas para a modelagem de todo o conjunto de dados. Logo, $\psi = \{\{\pi_k\}_{k=1}^K, \{\mu_k\}_{k=1}^K, \{\Sigma_k\}_{k=1}^K\}$ é o conjunto de parâmetros do modelo [58], que nesse caso são as probabilidades de ocorrência e os parâmetros das distribuições gaussianas básicas utilizadas (médias (μ_k) e matrizes de covariância (Σ_k)).

Na equação 2.52 foi utilizado diretamente a distribuição gaussiana, no entanto, existe a possibilidade da implementação de outras distribuições no modelo de mistura. Todavia, a utilização de gaussianas se torna mais interessante, pois de acordo com o ajuste dos parâmetros presentes nos modelos *GMM* inúmeras funções de densidades podem ser aproximadas com precisão arbitrária. À vista disso, diversas aplicações de tais modelos são vistas na literatura, como os trabalhos [61]–[64].

De forma análoga às redes neurais, previamente apresentadas, o grande desafio durante a utilização dos Modelos de Misturas de Gaussianas consiste na estimação de seus parâmetros ψ . Na literatura são encontrados diversos métodos para essa estimação, um exemplo disso consiste na técnica de Máxima Verossimilhança apresentado por [58]. Contudo, tal metodologia não será apresentada nesse trabalho, uma vez que os parâmetros

utilizados nesse modelo (número de componentes da mistura, probabilidade de ocorrência de cada um dos componentes, média e covariância) serão estimados com o auxílio dos métodos previamente apresentados (Redes Neurais, K-Médias e X-Médias). Tal utilização será apresentada com detalhes no próximo capítulo, que contempla a descrição completa da metodologia desenvolvida neste trabalho

2.6 Resumo e Discussões

No presente capítulo foi apresentada toda a fundamentação teórica dos métodos utilizados para a construção do modelo proposto. Na primeira parte foi levantado informações pertinentes as redes neurais artificiais, realizando um leve embasamento sobre as RNA's e em seguida sobre sua forma de treinamento utilizado nesse trabalho. Destaca-se como grande vantagem da utilização do treinamento bayesiano aplicado aos *MLP's* a possibilidade de construção de intervalos de confiança associados aos valores médios previstos quando aplicados a problemas de regressão e quando associados a problemas de classificação a resposta probabilística obtida ao final do processo. Além disso, torna-se válido ressaltar o surgimento da técnica de determinação automática da relevância que será capaz de “escolher” quais variáveis de entradas deverão ser utilizadas.

Em seguida, foi apresentado sobre os métodos de aprendizado de máquina não supervisionado que possuem como objetivo realizar a agregação dos dados utilizando medidas de similaridade, medidas estas que também foram apresentadas. Nesta parte, destaca-se a utilização do método X-médias que garante a escolha da quantidade ideal de grupos (*clusters*). Para essa escolha, o método utiliza o *BIC* como forma de mensurar a quantidade de grupos ideal para a segmentação da base de dados. Apesar de inúmeros trabalhos utilizarem o método K-médias como forma de agregação, dado o desconhecimento do analista a respeito da repartição dos dados e o objetivo desse trabalho, que é a construção de um modelo matemático autônomo, torna-se mais indicado a utilização do método X-médias dado sua maior independência.

Por fim, foi apresentado o modelo de misturas de gaussianas, a utilização desse modelo garante a representação do espaço por inúmeras gaussianas. Dessa forma, series temporais que não seguem um distribuição gaussiana, poderão ser aproximadas de forma não paramétrica.

3 MATERIAIS E MÉTODOS

No segundo capítulo foi apresentada a formulação teórica dos métodos de *Machine Learning* (Redes Neurais com treinamento Bayesiano) e *Analytics* (K-Médias, X-Médias e *GMM*) utilizados nesse trabalho. Contudo, existe a necessidade da exposição de como essas técnicas serão utilizadas em conjunto para a construção do modelo proposto nesse trabalho. Em virtude dessa necessidade, o presente capítulo irá expor ao leitor como tais métodos serão encapsulados durante a construção do modelo, além disso também será apresentado nesse capítulo o conjunto de dados utilizado para o desenvolvimento e validação do modelo construído, assim como os critérios de análise utilizados para verificação dos resultados obtidos.

Para a construção do modelo sugerido nessa dissertação, foram utilizados os ambientes de programação MATLAB e R. A utilização da linguagem R em conjunto com o MATLAB se fez necessária uma vez que os pacotes estatísticos dessa linguagem têm crescido exacerbadamente ao longo do tempo e inúmeras funções foram construídas [65].

A seguir será apresentado o macrofluxo do modelo construído, para elucidar sobre como os métodos apresentados serão “encapsulados”. Esse macrofluxo será debatido de uma forma mais detalhado no decorrer desse capítulo.

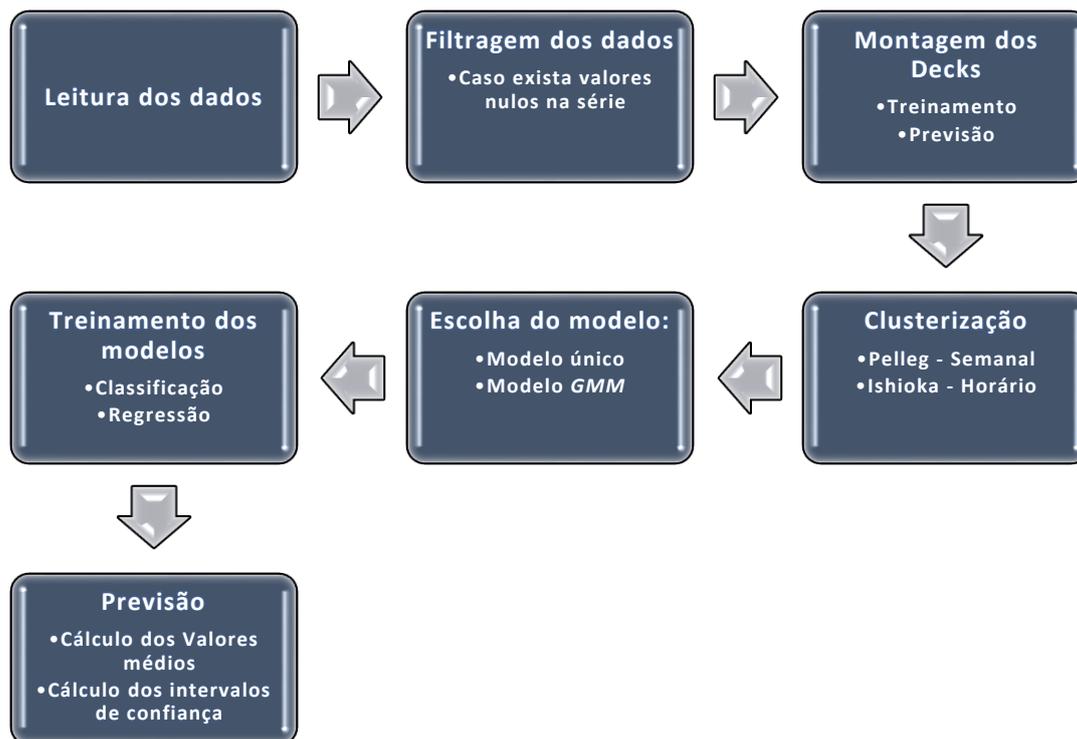


Figura 6 - Fluxograma de organização do modelo construído

De acordo com o macrofluxo apresentado, a seguir será descrito e detalhado cada uma das partes:

Passo 1 - Leitura dos dados de entrada;

Nesse momento é realizada a leitura de todos os dados utilizados, carga e temperatura, para o desenvolvimento e testes do modelo desenvolvido.

Passo 2 - Filtragem dos dados:

Devido a possibilidade da existência de valores nulos na base de dados, nessa etapa é realizada uma rápida verificação, sendo os valores nulos substituídos pela média aritmética entre as amostras da hora anterior e da hora posterior;

Passo 3 - Montagem dos Decks que serão apresentados as **RNA's**:

Nessa parte é realizada a montagem dos padrões de entradas e saída que serão utilizados nos modelos de Redes Neurais. Esses padrões são divididos em dois grupos: um para treinamento e outro para previsão (teste do modelo desenvolvido).

- a. Deck de treinamento;
- b. Deck de previsão:

Passo 4 - Clusterização dos dados de entrada:

Nessa etapa os dados de treinamento são separados em classes utilizando os métodos X-médias para comparação.

- a. Utilização dos métodos **X-médias**:
 - i. Para a previsão de carga semanal de energia foi utilizado o método proposto por *Pelleg* [56];
 - ii. Para previsão de carga horaria foi utilizado o método proposto por *Ishioka* [41];

Passo 5 - Verificação da quantidade de clusters escolhidos para definição de qual modelo será utilizado:

- a. **GMM**;
- b. Modelo Único;

Passo 6 - Treinamento dos modelos:

Nesse ponto os modelos de redes neurais serão treinados, ou seja, todos os parâmetros utilizados nesse modelo serão ajustados, utilizando a metodologia do treinamento bayesiano apresentado no capítulo anterior. Como já descrito, dois módulos baseados em redes neurais são desenvolvidos o primeiro com o intuito de apresentar a probabilidade de ocorrência um determinado grupo (Modelo Classificador) e o segundo modelo com o objetivo de realizar uma regressão onde será previsto um valor médio e uma variância para cada cluster (Modelo Regressor).

- a. Treinamento do modelo de classificação – caso necessário;
- b. Treinamento do modelo de regressão;

Passo 7 - Etapa de previsão

Nesse estágio são realizadas as previsões que serão utilizadas para a avaliação da metodologia proposta. As duas formas de modelagem utilizadas nessa dissertação serão utilizadas nesse momento, a primeira sendo as **RNA's** que irão apresentar como respostas os valores que serão utilizados nos modelos **GMM**.

- a. Execução das **RNA's** como classificador – caso necessário;
- b. Execução das **RNA's** como regressor;
- c. Cálculo do valor médio final de previsão utilizando o modelo **GMM**;

Passo 8 - Cálculo dos intervalos de confiança em função do tipo de modelo utilizado.

- a. Modelo *GMM*;
- b. Modelo único

3.1 Metodologia proposta

De forma geral, a metodologia proposta consiste na integração entre técnicas de *Machine Learning* e *Analytics* para a construção de um modelo autônomo para previsão de séries temporais. No presente trabalho o modelo desenvolvido será aplicado a problemas de previsão de carga, no entanto, devido a sua flexibilidade, este modelo poderá ser aplicado a uma vasta gama de problemas. Com a leitura do capítulo anterior, torna-se claro que diversos desafios são encontrados durante a utilização dos métodos propostos, visto suas peculiaridades. Logo, a presente seção possui como objetivo elucidar ao leitor cada um desses desafios, assim como as soluções propostas para a resolução deles. Dessa forma, também se tornará claro como os métodos foram agregados para a construção do modelo

O primeiro desafio consiste na estimação da quantidade de distribuições a serem utilizadas no modelo *GMM*, uma vez que se torna arriscado assumir que todo conjunto de dados pode ser descrito por uma gaussiana somente. De fato, a capacidade do *GMM* modelar funções de densidade de probabilidade genéricas a partir somente dos dados, constitui o principal motivador para a utilização deste tipo de modelo, viabilizando a abordagem autônoma de desafios de modelagem onde a variável de saída não segue uma distribuição gaussiana. Assim, o modelo *GMM* será utilizado para modelar automaticamente e de forma não-paramétrica a função densidade de probabilidade da variável de saída (série temporal de carga).

Como supracitado, na literatura muitos trabalhos propõem a maximização da verossimilhança para estimação de todos os parâmetros dos modelos *GMM*, sendo um desses parâmetros a quantidade de agrupamentos. As referências [60], [66] informam que a estimação dessa quantidade pode-se resumir no algoritmo K-médias, porém para sua utilização torna-se necessário a definição do número de conjuntos (K) que os dados deverão ser agrupados. Logo, dado o desconhecimento sobre essa quantidade, no presente trabalho foi utilizado o método X-médias que garante a separação automática da base de

dados sem que exista a necessidade de interferência do especialista durante a escolha do número de agrupamentos. Foram utilizadas duas formulações do método de X-médias propostos na literatura e previamente descritos, no método proposto por *Pelleg e Moore* [56] utilizou-se o número máximos de centros igual a 30. Ressalta-se se tornou necessário a que a utilização de dois tipos de algoritmos de clusterização diferentes devido a limitação vista no método proposto por *Ishioka*, pois nesse método o número mínimo de *cluster* parte de dois e em alguns casos isso não se torna realidade. Estes algoritmos de clusterização são aplicados à série temporal de saída visando identificar o número de agrupamentos a ser utilizado no modelo *GMM* para a estimativa da função densidade de probabilidade desta variável.

O segundo desafio durante a utilização dos modelos *GMM* consiste na estimação das probabilidades a priori (π_k) para cada uma das distribuições, além das médias e variâncias (μ_k e Σ_k) utilizados por elas no caso univariado. Para tal fim, utilizou-se as redes neurais artificiais (RNAs) utilizando o treinamento bayesiano, as quais possuem a capacidade de prever o valor esperado e a variância de uma dada variável gaussiana, como também a probabilidade de um dado padrão pertencer a uma dada classe.

Para cumprir esse objetivo, os padrões entrada/saída, utilizados para construção dos modelos de previsão de carga nos diversos horizontes, serão aplicados a dois modelos, um de classificação e um de regressão. Mantidos os padrões de entrada, o modelo de classificação terá como saída a probabilidade deste valor pertencer a cada um dos *clusters* identificados pelo modelo X-médias. Exemplificando, supondo que o X-médias identificou a presença de três *clusters*, o modelo de classificação possuirá 3 saídas, cada uma delas representando a probabilidade da carga pertencer a cada um dos *clusters*. Uma vez treinado, esse modelo tem a capacidade de estimar a probabilidade do referido padrão de entrada estar associado a cada um dos 3 *clusters* identificados, provendo assim uma estimativa para as probabilidades a priori (π_k) das distribuições básicas utilizadas pelo modelo *GMM*.

Paralelamente à etapa de classificação, são desenvolvidos modelos de regressão individuais para cada *cluster* utilizando novamente as RNAs com inferência bayesiana. O motivo para tal uso está relacionado com a capacidade destes modelos em estimar tanto o valor esperado (μ_k) quanto a variância (Σ_k) de uma variável de saída com função densidade de probabilidade gaussiana. Utilizando o exemplo apresentado para o módulo

de classificação (3 clusters), os padrões de entrada e de saída (neste caso o valor da carga) são separados para cada cluster, sendo construídos 3 modelos de regressão independentes para cada um dos *clusters*. Concluído o treinamento, na etapa de previsão o mesmo padrão de entrada é apresentado a cada modelo, sendo fornecidas as estimativas para as médias e variâncias (μ_k e Σ_k) necessárias para uso do modelo *GMM*.

As RNAs utilizadas são do tipo *Multi Layer Perceptron*, contendo apenas uma camada escondida e cinco neurônios em tal camada. Para seu treinamento foi utilizado a inferência bayesiana aplicada as *MLP's* previamente explicitada. Torna-se valido frisar que, em função da modelagem utilizada durante o desenvolvimento dos modelos neurais, é possível a escolha de um número fixo de neurônios na camada escondida, pois caso esse número de neurônios seja superestimado o modelo será capaz de atribuir valores insignificantes de pesos para os neurônios desnecessários.

Por fim, com o auxílio dos resultados obtidos anteriormente, pode-se utilizar a teoria definida nos modelos *GMM* para a estimativa do valor esperado para a carga no instante de tempo de interesse. Aplicando o operador valor esperado em ambos os lados da equação 2.52, obtêm-se:

$$C_{méd} = \underline{C_{previsto}}^t \underline{P_{previsto}} \quad (3.1)$$

Onde: $C_{méd}$ representa o valor esperado da carga previsto pelo modelo ao final do processo, $\underline{C_{previsto}}$ o vetor com os valores esperados previstos para a carga considerando o modelo de regressão construído para cada um dos *clusters* e $\underline{P_{previsto}}$ a probabilidade do ponto a ser previsto pertencer a cada um dos *clusters* construídos, estimada utilizando o modelo de classificação.

Para um melhor entendimento do leitor, será apresentado um exemplo da utilização da equação 3.1. Considerando que após o processo de clusterização (X-médias) a série de carga a ser prevista tenha sido separada em três grupos, sendo desenvolvido um modelo classificador para previsão da probabilidade do padrão de entrada pertencer a cada um dos *clusters* e três modelos de regressão independentes para previsão de carga interna a cada cluster. Concluído o processo de treinamento, um dado padrão de entrada é apresentado ao módulo classificador produzindo as seguintes estimativas para a probabilidade do respectivo padrão pertencer a cada um dos três clusters:

$$\underline{P_{previsto}}^t = [0,5 \ 0,3 \ 0,2]. \quad (3.2)$$

O mesmo padrão de entrada é apresentado aos três modelos de regressão desenvolvidos (um para cada *cluster*) fornecendo as seguintes previsões para os valores esperados da carga:

$$\underline{C_{previsto}}^t = [100 \ 200 \ 300]. \quad (3.3)$$

Assim, o valor esperado da carga calculado pelo método *GMM*, como já apresentado, será a ponderação entre os valores médios previstos e a probabilidade do padrão a ser previsto pertencer a cada um dos clusters, ou seja,

$$C_{méd} = 0,5 * 100 + 0,3 * 200 + 0,2 * 300 = 170. \quad (3.4)$$

Por fim, podem ser calculados os intervalos de confiança associados a tais previsões. Esse cálculo é realizado utilizando a equação (2.51) e a função inversa da distribuição gaussiana a partir dos valores previstos de carga média e variância para cada *clusters*, como também para as probabilidades a priori estimadas pelo módulo de classificação. A partir da definição do usuário acerca da forma desejada para construção dos intervalos de confiança (quartis, percentis), os valores são para cada divisão (quartil ou percentil) são obtidos via simulação da equação (2.51) considerando as estimativas supracitadas. Maiores detalhes sobre a construção dos intervalos de confiança, ao se utilizar os modelos *GMM*, são apresentados no apêndice A.

Torna-se necessário frisar que caso o processo de agrupamento da base de dados indique somente um grupo (*cluster*), é utilizado o chamado modelo único e não mais o *GMM*. Nesse caso não existe a necessidade da execução das RNA's com o objetivo de classificação, uma vez que a probabilidade do padrão pertencer ao grupo será de 100%. Além disso, o cálculo dos intervalos de confiança se torna mais simples, por exemplo, considerando um nível de confiança de 99,73% para o cálculo desse intervalo, basta somente atribuir três desvios padrões ao valor médio previsto, como apresentado na equação (3.5).

$$I = \mu \pm 3\Sigma \quad (3.5)$$

3.2 Dados utilizados

Para o desenvolvimento do presente trabalho foram utilizados os dados de carga dos quatros subsistemas que compõem o SIN (Sistema interligado Nacional), conforme apresentado na figura 6. Além disso, dada a relação da carga com a temperatura, também foram utilizados os dados de temperaturas coletados nos aeroportos de tais regiões como apresentado na tabela 1, caso o leitor sinta a necessidade de conhecer os estados de cada um dos aeroportos a referência [67] poderá auxiliar. Todos os dados utilizados nesse trabalho são dados reais disponibilizados pelo Operador Nacional do Sistema Elétrico (ONS).

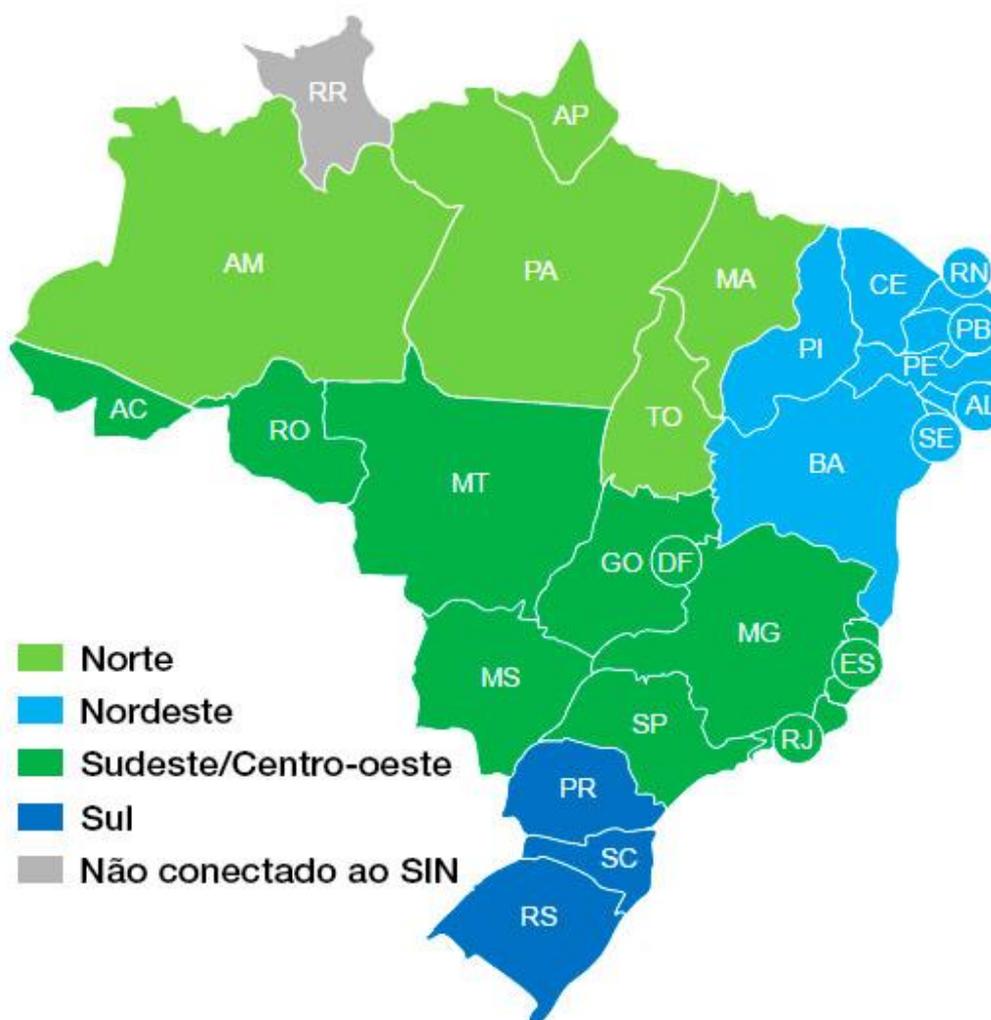


Figura 7 - Submercado de Energia [68]

Tabela 1 - Aeroportos utilizados e seus respectivos subsistemas [67]

Aeroportos	Subsistema
SBEG	N
SBMQ	N
SBSL	N
SBBE	N
SBCT	S
SBLO	S
SBPA	S
SBFL	S
SBMO	NE
SBSV	NE
SBFZ	NE
SBJP	NE
SBRF	NE
SBTE	NE
SBNT	NE
SBVT	SE
SBGO	SE
SBBH	SE
SBGL	SE
SBDN	SE
SBRP	SE

O período de dados disponíveis para desenvolvimento dos modelos se estende do dia 26/12/2009 que é o primeiro dia da primeira semana operativa de 2010 até o dia 03/01/2020 que é o último dia da última semana operativa do ano de 2019. Contudo, devido a mudança no comportamento da série de carga, esse horizonte de dados não será utilizado em sua totalidade.

3.3 Critério de Análise

Visto a necessidade de averiguação da eficácia do modelo proposto, a seguir serão apresentados os métodos utilizados para tal avaliação.

3.3.1 Erro Médio Percentual (*MAPE*)

O erro médio percentual calcula a relação entre a diferença do valor verificado z_f e o valor previsto z_p em função do valor verificado z_f de acordo com a equação 3.6

$$E_{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|z_{f_i} - z_{p_i}|}{z_{f_i}} \times 100 \quad (3.6)$$

3.3.2 Erro absoluto

O erro absoluto verifica a diferença entre o valor verificado e o valor previsto, como apresentado na equação 3.7

$$E_a = z_{f_i} - z_{p_i} \quad (3.7)$$

3.3.3 Coeficiente de assimetria e Curtose

Durante a modelagem do treinamento bayesiano foi assumido como premissa que os resíduos podem ser modelados através de distribuições normais, dessa forma torna-se indicado avaliar o comportamento desses resíduos. Para isto será utilizado o coeficiente de assimetria e Curtose [69]–[73], nesta análise poderá ser visualizado se o modelo apresenta algum tipo de viés, seja ele negativo ou positivo, e se os resíduos seguem uma distribuição normal como adotado nas premissas.

O coeficiente de assimetria informa a respeito da distribuição do resíduo em relação a sua média. Esse coeficiente pode ser calculado através da equação 3.8:

$$Assimetria = \frac{1}{n} \sum_{i=1}^n \left[\frac{(X_i - \bar{X})}{S} \right]^3 \quad (3.8)$$

Já a Curtose visa indicar a respeito do grau de achatamento de uma distribuição de frequência. Dessa forma, de acordo com o valor da curtose pode-se inferir a respeito da forma da distribuição. Existem, basicamente, três classificações que são:

Mesocúrtica – quando apresenta uma medida de curtose **igual** à da distribuição normal, curtose próxima a três.

Platicúrtica – quando apresenta uma medida de curtose **menor** à da distribuição normal, curtose inferior a três.

Leptocúrtica – quando apresenta uma medida de curtose **igual** à da distribuição normal, curtose maior do que três.

Sendo a Curtose definida através da equação 3.9

$$Curtose = \frac{1}{n} \sum_{i=1}^n \left[\frac{(X_i - \bar{X})}{S} \right]^4 \quad (3.9)$$

Onde:

$s =$ desvio padrão

$n =$ Número de amostras

$X_i =$ i – ésimo resíduo da amostra

$\bar{X} =$ média dos resíduos

3.3.4 Avaliação dos intervalos de confiança

Na literatura são encontrados diversos métodos para verificação do intervalo de confiança gerado, no presente trabalho será adotado somente o *PICP – Prediction Interval Evaluation Indexes*. Esse índice possui como objetivo verificar o percentual médio de casos que o valor verificado encontra-se dentro do intervalo de confiança [74]–[76].

$$PICP = \frac{1}{N} \sum_{i=1}^N \varphi_i \quad (3.10)$$

Onde:

$\varphi_i = \begin{cases} 1, & \text{se o valor verificado estiver dentro do intervalo de confiança} \\ 0, & \text{caso o contrário} \end{cases}$

$N =$ Número de amostras

4 RESULTADOS

O presente capítulo possui como objetivo apresentar os resultados obtidos durante a aplicação da metodologia proposta no capítulo anterior. Como benchmark das previsões será utilizado o método *Naive* (ingênuo) como proposto em [35], esse método propõe a repetição do último valor verificado como previsão, a escolha desse método de comparação se torna adequada uma vez que no “dia a dia” de previsão essa forma é utilizada quando as premissas utilizadas encontram-se semelhantes a alguma semana ou dia do histórico. Além disso, também será utilizado como benchmark os indicadores de desempenho do ONS [77], onde previsões com desvios superiores a 3% são consideradas como inadequadas. Torna-se válido ressaltar que este indicador diz respeito às previsões mensais de energia para o sistema interligado nacional, porém o mesmo limiar de 3% será aplicado as previsões geradas nesta dissertação.

4.1 Resultados Semanais

Antes da apresentação dos resultados será mostrado como o modelo foi estruturado no que diz respeito as variáveis de entradas escolhidas e os feriados considerados. Torna-se importante ressaltar que devido ao controle de complexidade presente na formulação do processo de treinamento do modelo, será utilizado um número fixo de neurônios na camada oculta, fato esse mencionado anteriormente.

Os dados utilizados como entrada para o modelo foram: carga e temperaturas (máximas, médias e mínimas) das últimas 4 semanas, temperaturas (máximas, médias e mínimas) previstas para a semana-alvo (semana a ser prevista) e por fim uma variável binária de sete posições para representação de feriados de acordo com seu dia da semana. Foram considerados os 11 feriados nacionais, apresentados na tabela 2, que encontram-se definidos na portaria de nº468 [78].

Tabela 2 - Lista de feriados considerados

Feriado	Data
CONFRAT. UNIVERSAL	01/01/2018
CARNAVAL	13/02/2018
PAIXAO	30/03/2018
TIRADENTES	21/04/2018
DIA DO TRABALHO	01/05/2018
CORPUS CHRISTI	31/05/2018
INDEPENDENCIA DO BRASIL	07/09/2018
NOSSA SRA APARECIDA	12/10/2018
FINADOS	02/11/2018
PROCLAMACAO DA REPÚBLICA	15/11/2018
NATAL	25/12/2018

Para o desenvolvimento do modelo proposto, foi utilizado o período compreendido entre primeira semana operativa do ano de 2014 (28/12/2013 e 03/01/2014) até a penúltima semana operativa do ano de 2016 (24/12/2016 até 30/12/2016) para treinamento do modelo. Devido ao controle de complexidade existente no processo de treinamento, não existe a necessidade de separação de uma parte do conjunto de dados para validação. Em função disso, foi utilizado todo conjunto de dados disponíveis durante o treinamento. Por fim, para teste do modelo foi utilizado o período compreendido entre a primeira semana operativa de 2017 (31/12/2016 até 06/01/2017) até a última semana operativa do ano de 2018 (29/12/2018 até 04/01/2019), representando assim um total de 105 semanas previstas.

Abordando sobre os resultados obtidos durante a utilização do modelo para o período de teste, a tabela 3 apresenta os desvios médios (*MAPEs*) por subsistema. Para o SIN, que é a composição da carga dos quatro subsistemas apresentados, o desvio médio obtido pelo modelo foi de **1,87%**.

Tabela 3 - Erro médio por subsistema (previsões semanais)

Subsistema	Erro (%)
SE/CO	2,00
SUL	3,59
NORDESTE	1,93
NORTE	1,72

As figuras 8, 9, 10, 11 e 12 apresentam as previsões médias em MWméd para as semanas, assim como o *MAPE* associado a cada uma das previsões. Através dessas figuras pode-se observar o comportamento das previsões e dos intervalos de confiança em relação ao valor verificado. Ressalta-se o comportamento das previsões quando existem quedas e/ou acréscimos de energia, pois o modelo consegue reproduzir esse comportamento, mesmo que em níveis diferentes conforme destacado, através de setas nas figuras a seguir. Além disso, torna-se válido frisar a assertividade do intervalo de confiança previsto, dado que a maior parte dos valores verificados se encontram dentro desse intervalo de confiança. Nessas figuras os intervalos de confiança foram construídos utilizando 3 desvios padrões, uma melhor análise em relação aos intervalos de confiança será realizada no próximo tópico, com a utilização do método de avaliação *PICP*.

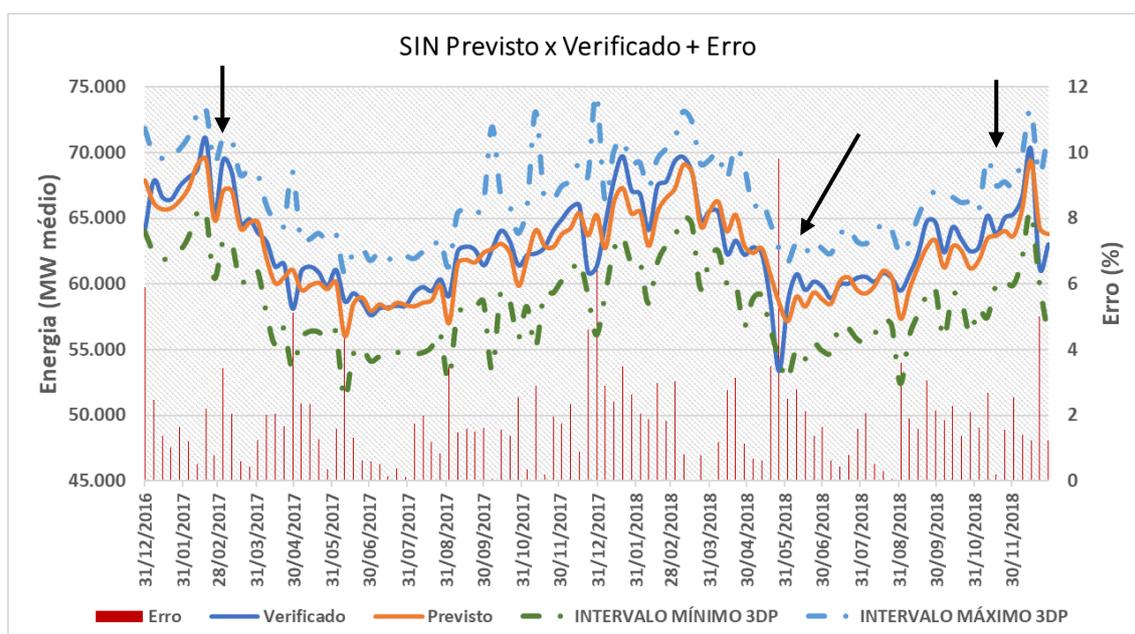


Figura 8 – Verificado x Previsto SIN + Erro

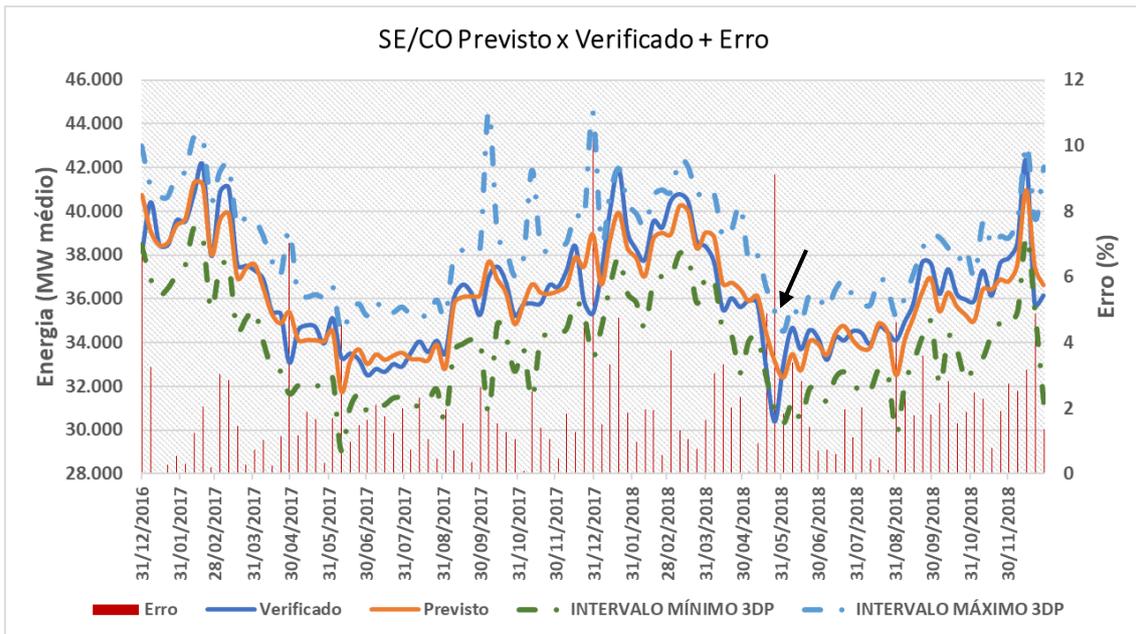


Figura 9 – Verificado x Previsto SE/CO + Erro

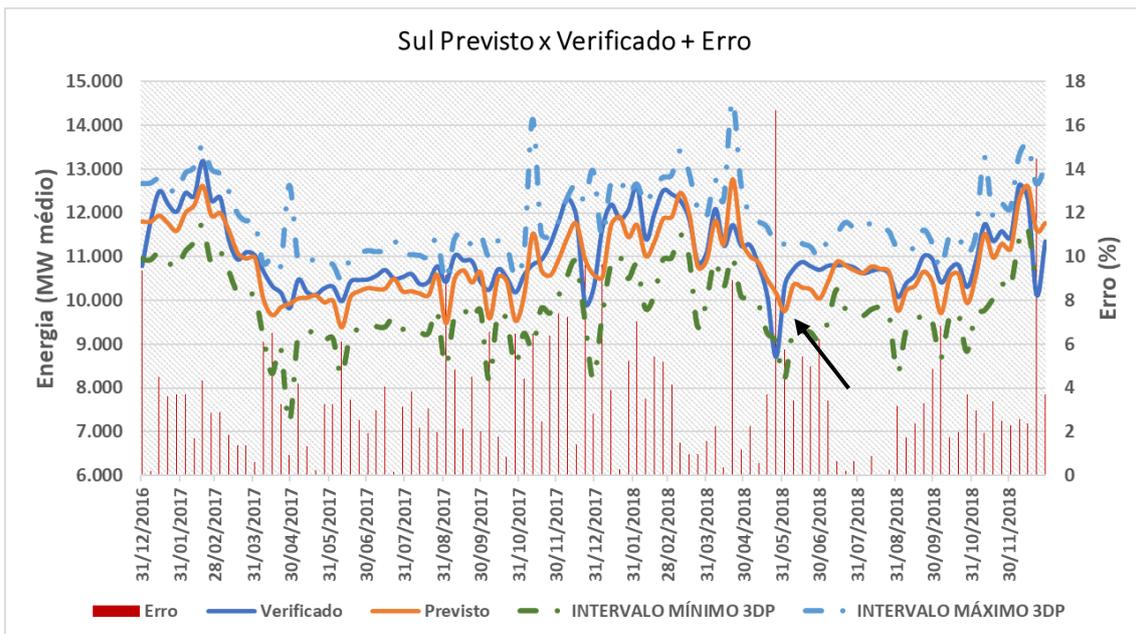


Figura 10 – Verificado x Previsto Sul + Erro

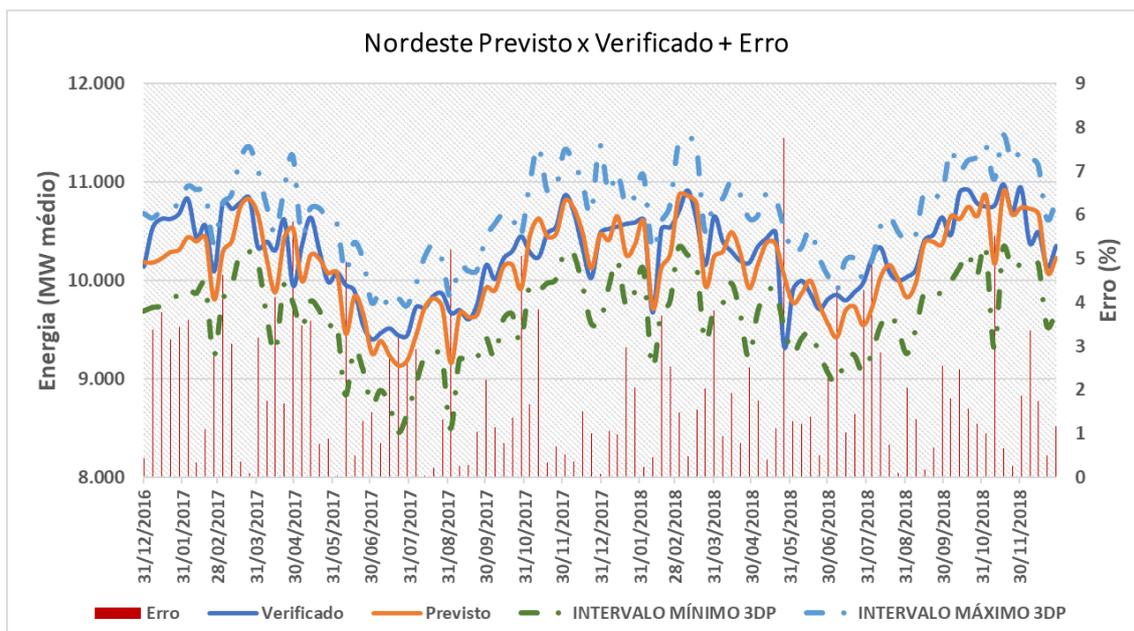


Figura 11- Verificado x Previsto Nordeste + Erro

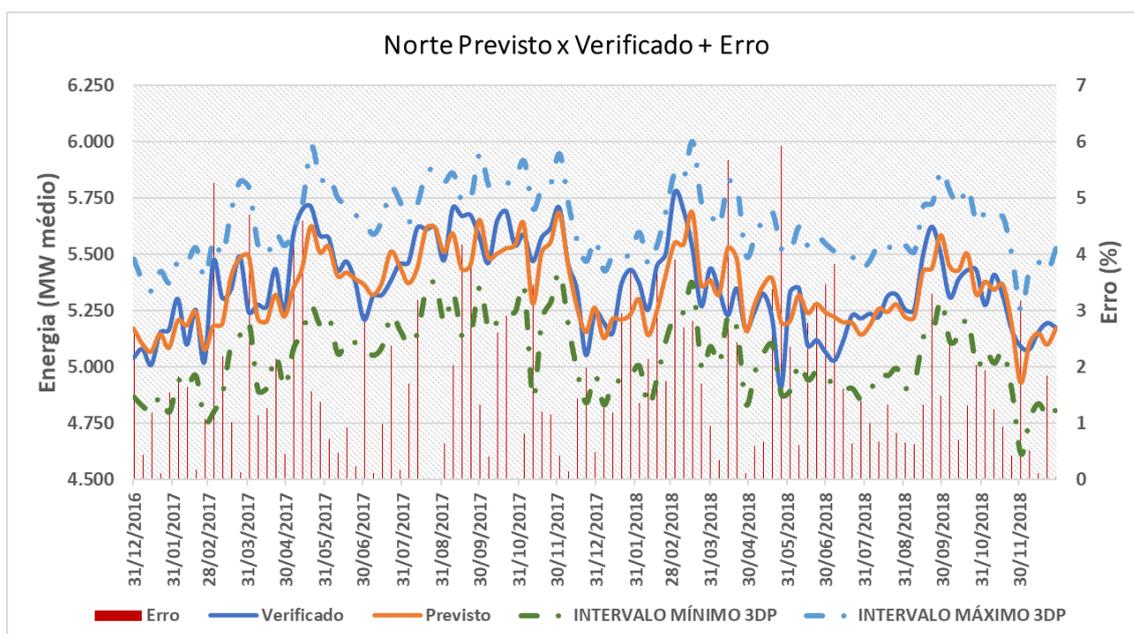


Figura 12- Verificado x Previsto Norte + Erro

Outra análise realizada nessa dissertação diz respeito às distribuições dos erros obtidos. Essa análise se faz necessária para a verificação do comportamento do erro em relação a sua média, ou seja, através dessa avaliação pode-se verificar se o modelo apresenta desvios semelhantes ao desvio médio obtido ou se o modelo apresenta uma elevada dispersão dos desvios em relação ao erro médio. As figuras 13, 14, 15, 16 e 17 apresentam tal consideração, destaca-se que a maioria dos erros se encontram inferiores

a 3% (vide figura 13 e 14 para verificação do SIN e Sudeste/Centro-Oeste, respectivamente). Torna-se válido frisar que o “X” apresentado em tais figuras representa a variação percentual do erro cometido.

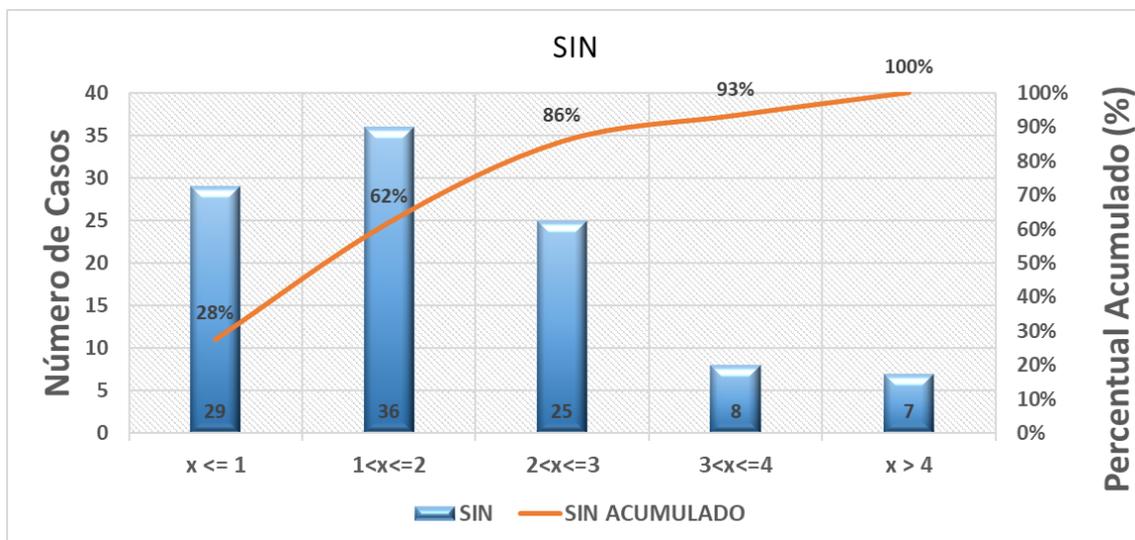


Figura 13 – Erro semanal acumulado (SIN)

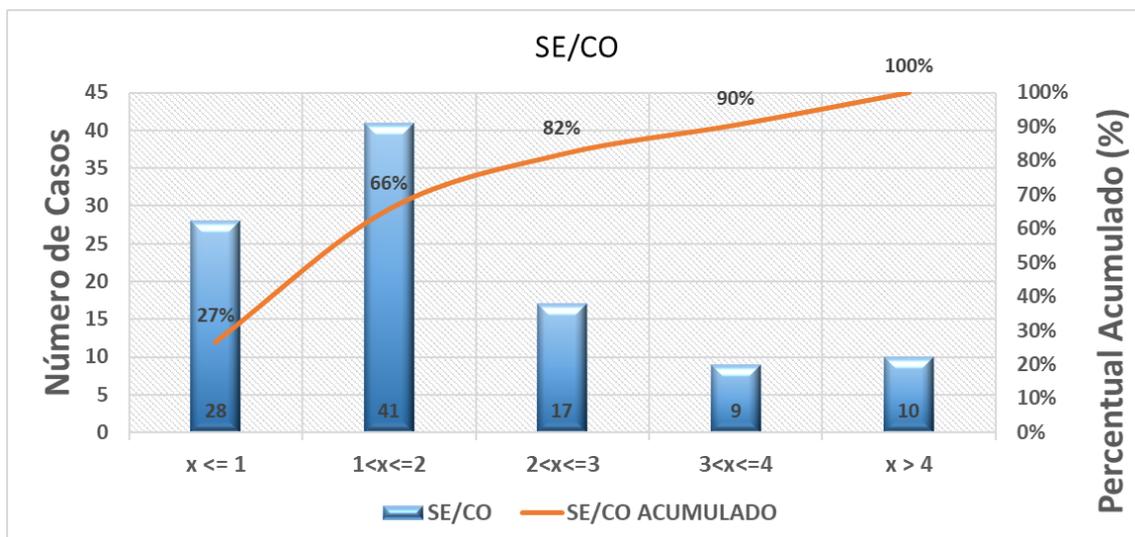


Figura 14 - Erro semanal acumulado (SE/CO)

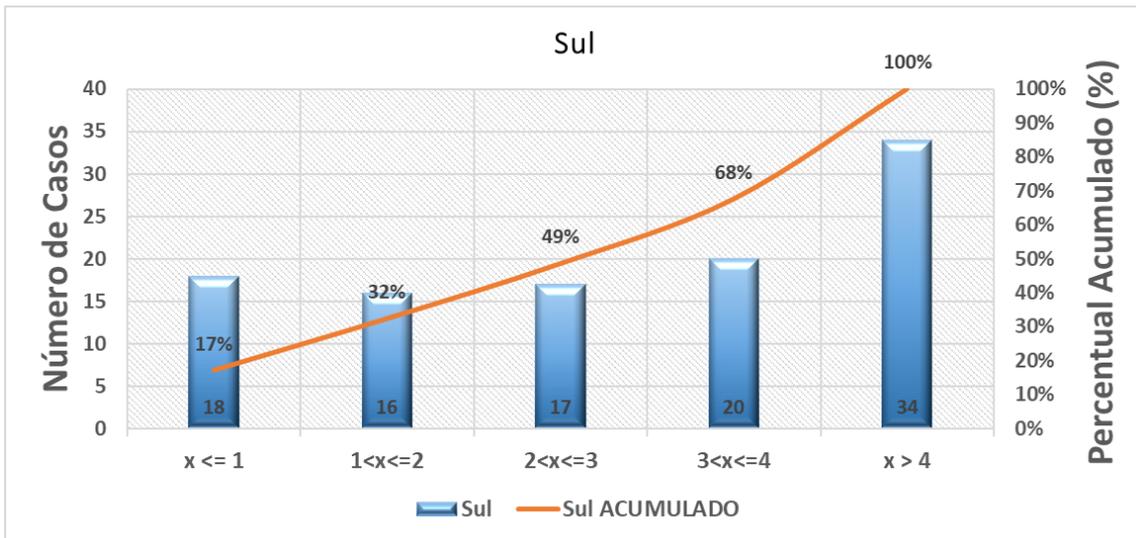


Figura 15 - Erro semanal acumulado (Sul)

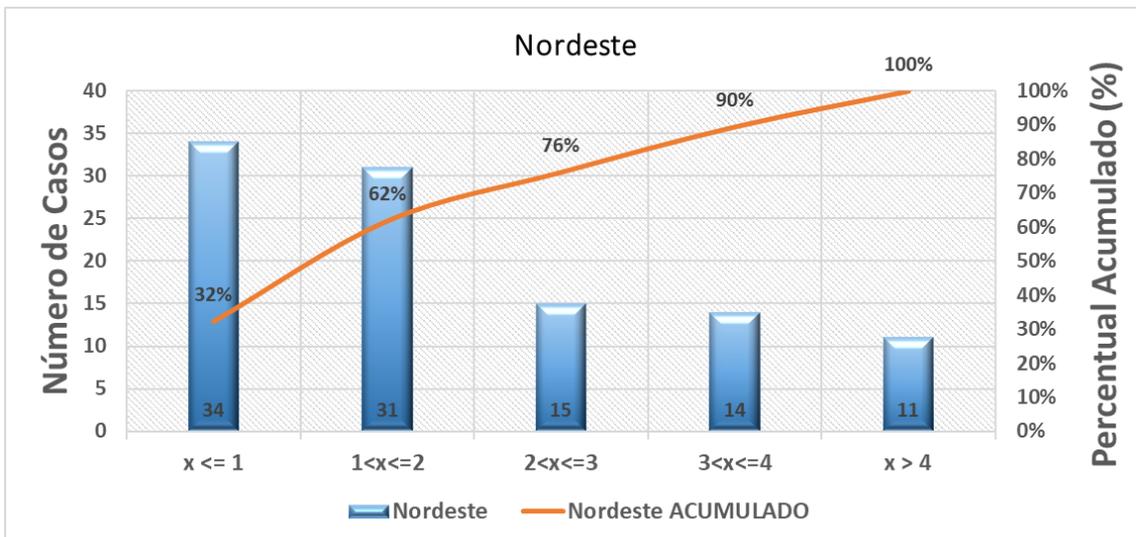


Figura 16 - Erro semanal acumulado (Nordeste)

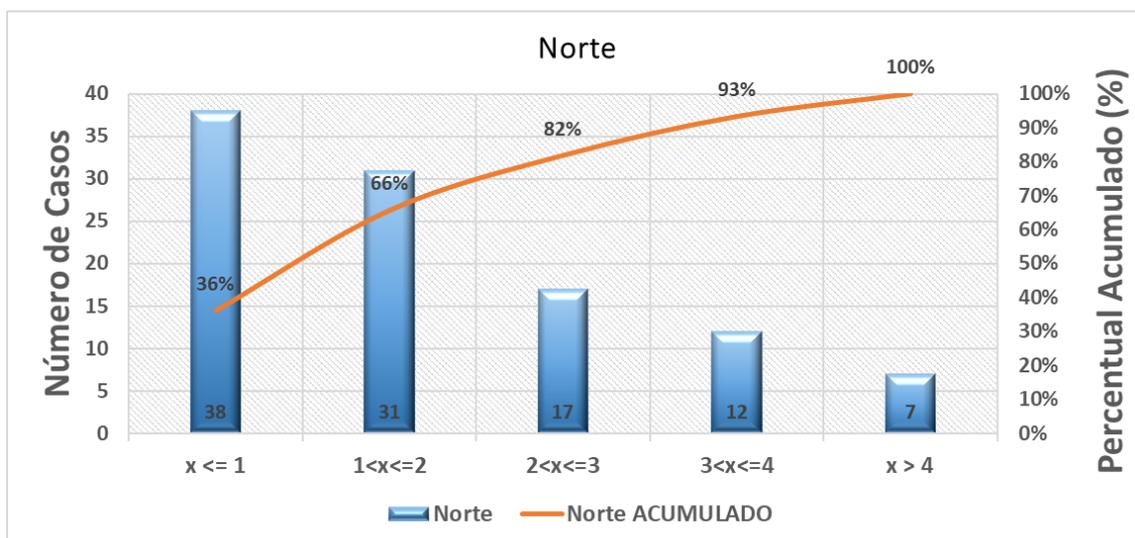


Figura 17 - Erro semanal acumulado (Norte)

4.1.1 Análise de Curtose e do Coeficiente de assimetria

Como supracitado a análise de curtose e do coeficiente de assimetria visa averiguar o comportamento dos resíduos ao se aplicar o modelo proposto nos dados de teste. As figuras 18, 19, 20, 21 e 22 apresentam o histograma dos resíduos. Estes histogramas irão auxiliar no entendimento da Curtose e do coeficiente de assimetria.

Ao se aplicar a equação 3.8, foram obtidos os seguintes coeficientes de assimetria, -1.6, -1.3, -1.4, -0.5 e -0.1 para o SIN, Sudeste/Centro-Oeste, Sul, Nordeste e Norte, respectivamente. Dessa forma, pode-se verificar que exceto nos subsistemas Nordeste e Norte, onde o modelo apresentou um coeficiente de -0.5 e -0.1 o que indica o modelo não apresenta nem um viés, para todos os outros subsistemas o modelo apresentou um viés positivo. O viés apresentado pelo modelo pode ser visualizado nos histogramas a seguir.

Ao se analisar os índices de curtose obtidos, que foram de: 6.4 para o SIN, 5.7 para o subsistema Sudeste/Centro-Oeste, 6.4 para o subsistema Sul, 4.2 para o subsistema Nordeste e 3.0 para o subsistema Norte. Pode-se inferir que os resíduos de todos os subsistemas, exceto o subsistema Norte, não seguem uma distribuição normal.

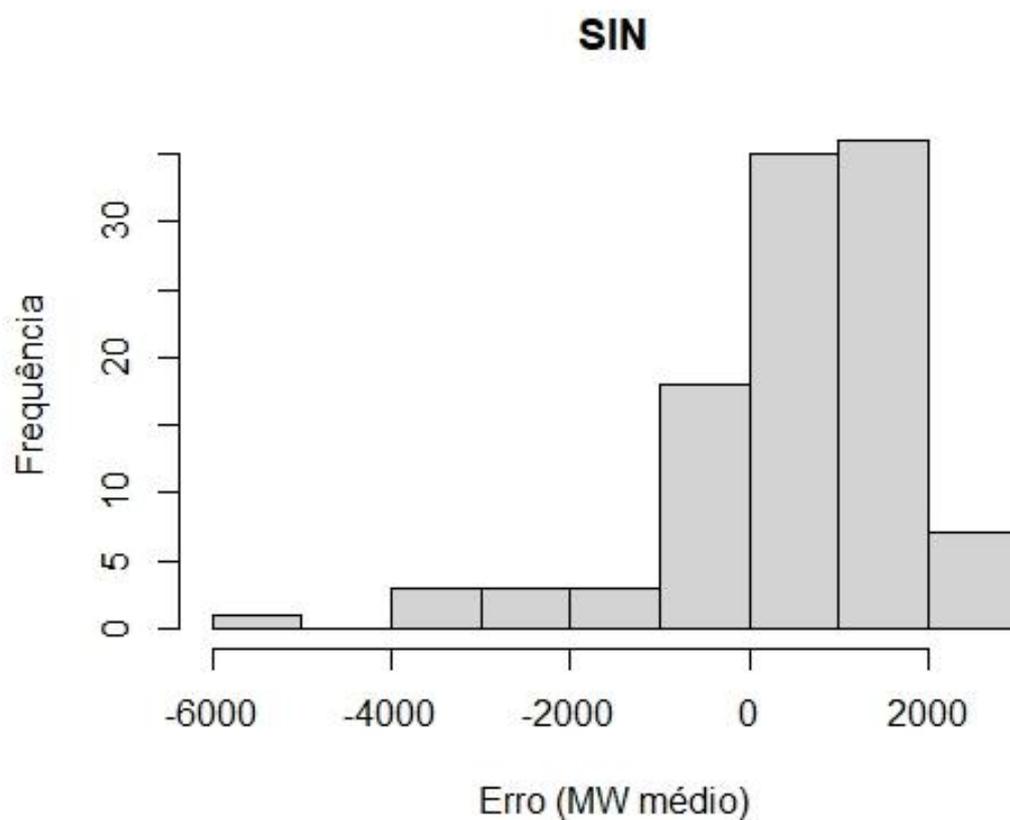


Figura 18 - Histograma dos resíduos semanais do SIN

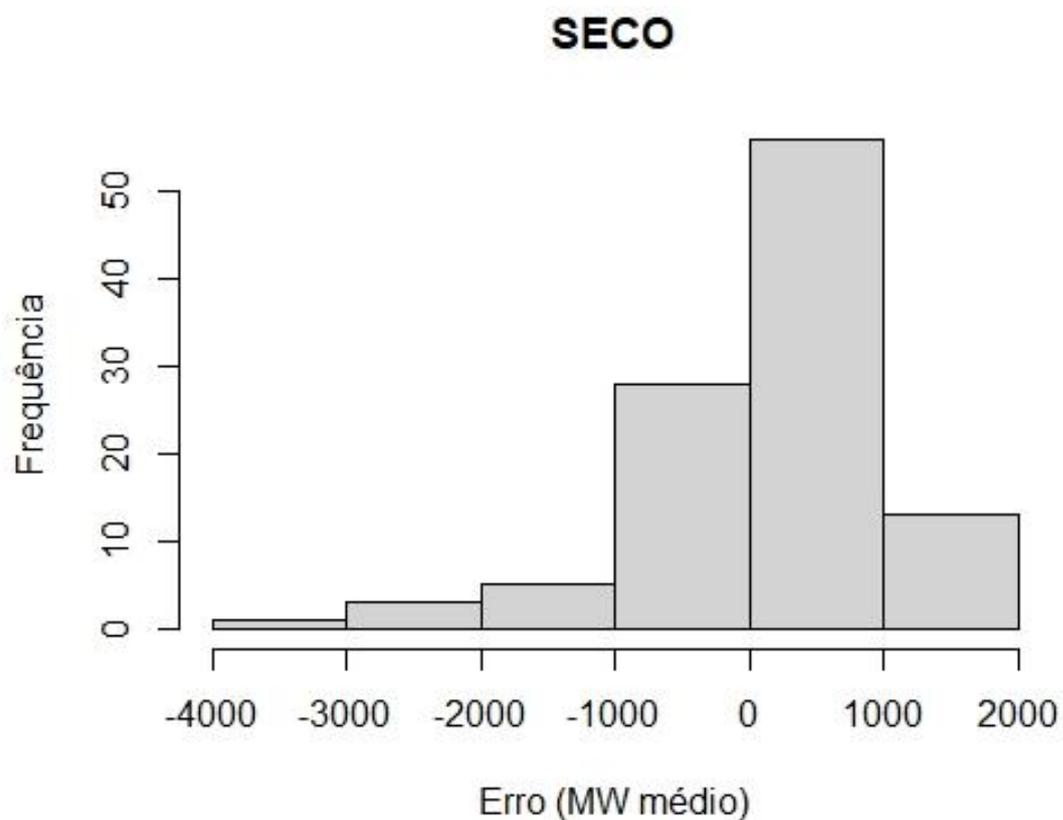


Figura 19 - Histograma dos resíduos semanais do subsistema Sudeste/Centro-Oeste

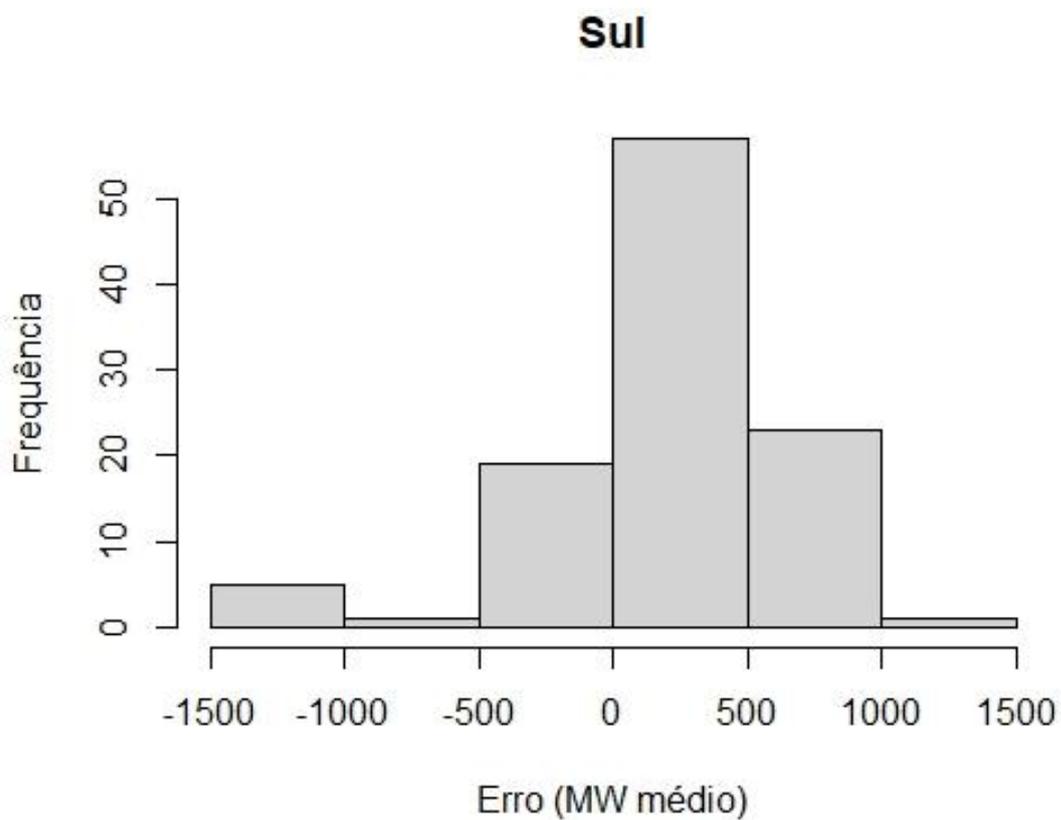


Figura 20- Histograma dos resíduos semanais do subsistema Sul

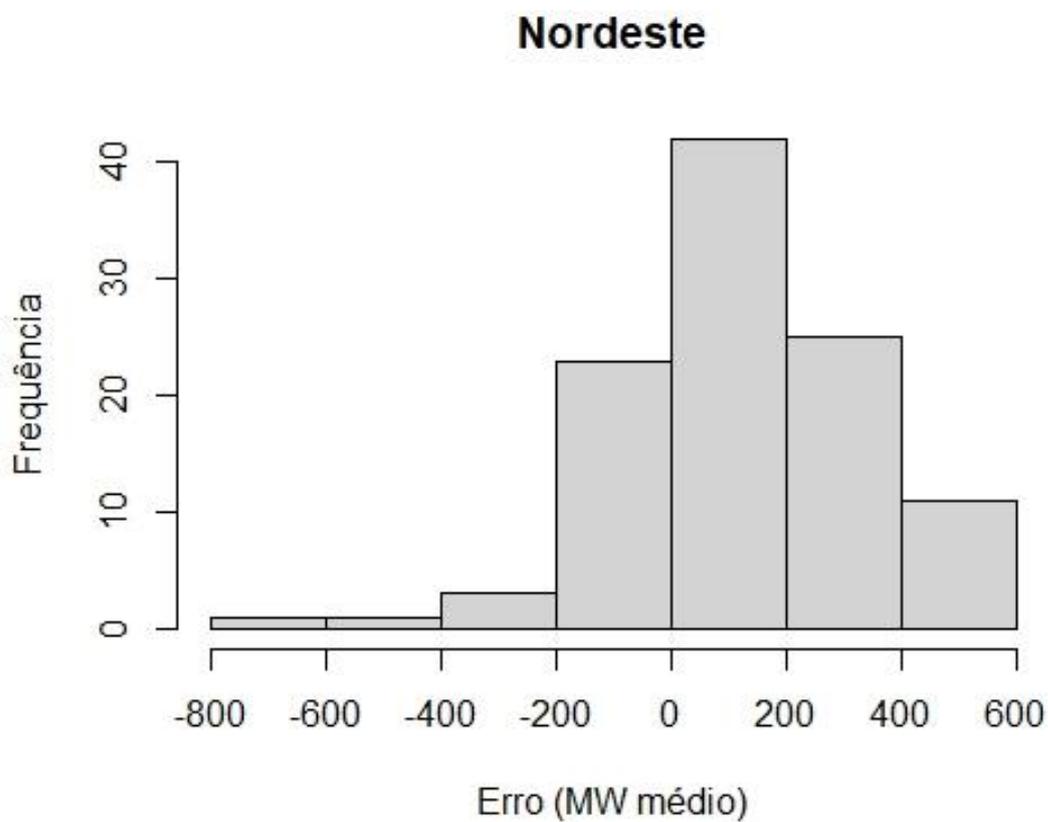


Figura 21- Histograma dos resíduos semanais do subsistema Nordeste

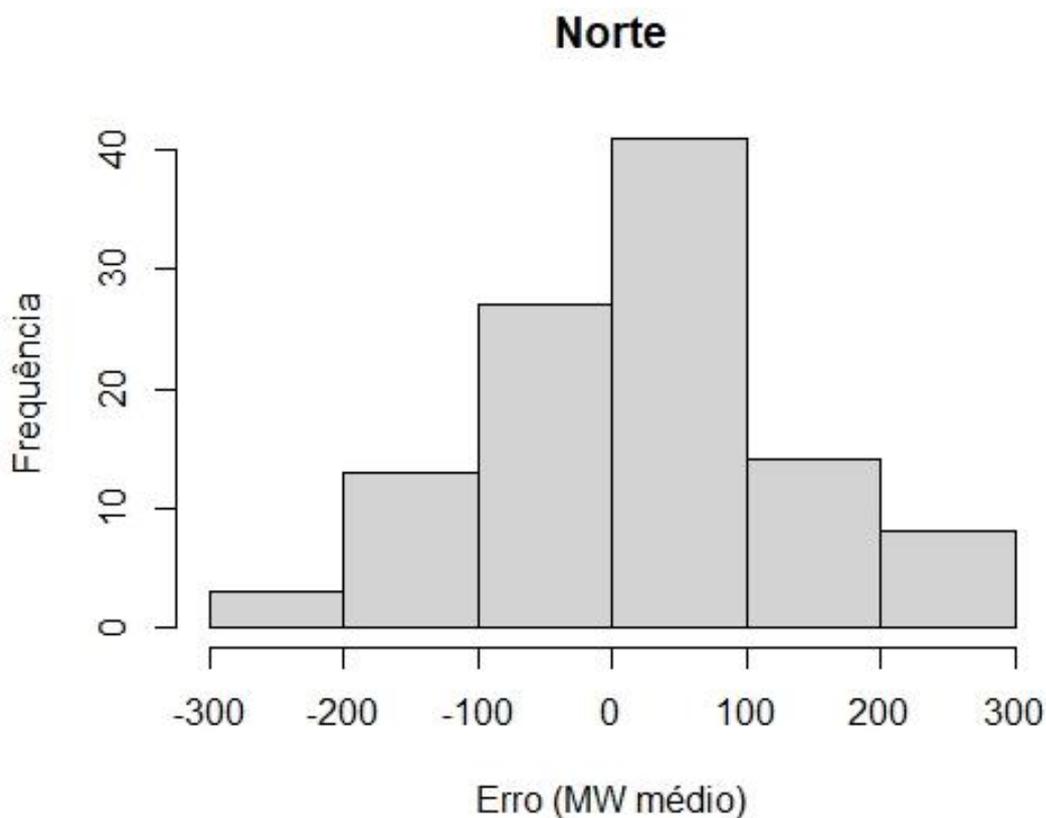


Figura 22- Histograma dos resíduos semanais do subsistema Norte

4.1.2 Análise do Intervalo de Confiança

Como supracitado a utilização do método *PICP*, como forma de avaliar o intervalo de confiança previsto, possui como objetivo verificar o número médio de valores verificados que estão no interior do intervalo de confiança.

No caso das previsões semanais, o processo de clusterização indicou a presença de somente um *cluster*, dessa forma o modelo utilizado foi o denominado “Modelo Único”, nesse tipo de modelo os intervalos de confiança são construídos utilizando basicamente o valor da variância (desvio padrão = raiz quadrada da variância) prevista, como descrito no capítulo anterior.

É de conhecimento geral que quanto maior for o múltiplo de desvio padrão (DP) utilizado, maior a probabilidade do intervalo de confiança abranger o valor verificado (1 DP = 68,26%, 2 DP = 95,44%, 3 DP = 99,73%). Dessa forma, serão realizadas 3 análises variando o número de desvios padrões utilizados em cada uma delas.

Tabela 4 - Análise PICP para previsão semanal considerando 3 DP

INTERVALOS (3DP)		
SUBSISTEMA	N DENTRO	PERCENTUAL
SE	103	98%
S	99	94%
NE	102	97%
N	104	99%
SIN	103	98%

Tabela 5 - Análise PICP para previsão semanal considerando 2 DP

INTERVALOS (2DP)		
SUBSISTEMA	N DENTRO	PERCENTUAL
SE	96	91%
S	87	83%
NE	87	83%
N	95	90%
SIN	100	95%

Tabela 6 - Análise PICP para previsão semanal considerando 1 DP

INTERVALOS (1DP)		
SUBSISTEMA	N DENTRO	PERCENTUAL
SE	68	65%
S	50	48%
NE	59	56%
N	62	59%
SIN	68	65%

Ao verificar os valores obtidos nas tabelas 4, 5 e 6, pode-se concluir que, embora os resultados alcançados estejam divergindo dos valores teóricos esperados e destacados no parágrafo anterior, ainda assim são visualizados excelentes resultados principalmente ao se utilizar três e dois desvios padrões para a construção desses intervalos.

Além disso, torna-se válido a realização de análises em relação a amplitude do intervalo de confiança previsto, pois caso esse intervalo apresente uma elevada amplitude, de nada irá ajudar o especialista no momento de sua decisão. Dessa maneira, a tabela 7 apresenta a amplitude média dos intervalos de confiança previstos, para os diferentes números de desvio padrão utilizados.

Tabela 7 - Intervalo de confiança médio

Intervalo de Confiança Médio (MW médios)			
SUBSISTEMA	3 DP	2 DP	1 DP
SE	4575	3112	1527
S	2039	1348	672
NE	1043	705	353
N	597	404	200
SIN	8253	5569	2752

Com o auxílio da tabela 7, verifica-se que a amplitude dos intervalos de confiança previstos apresenta variações aceitáveis em relação ao valor médio, principalmente ao se utilizar 1 ou 2 desvios padrões.

Por fim, ressalta-se que durante as análises dos intervalos de confiança construídos utilizando três desvios padrões, as semanas que apresentaram valores de carga fora dos intervalos de confiança, foram semanas atípicas devido a existência de feriados em seu período.

Caso exista a necessidade de uma verificação em relação aos intervalos de confiança construídos, maiores detalhes podem ser encontrados no apêndice B.

4.1.3 Benchmark

Abordando um pouco sobre as formas de benchmark utilizadas, temos como primeira forma o método denominado *Naive* (ingênuo). Neste método as previsões adotadas são iguais aos valores verificados na última semana. Logo, ao se utilizar o método *Naive* como forma de previsão obtém-se um desvio médio de 2,27% para o SIN, valor este que se encontra superior ao desvio médio observado ao se utilizar a metodologia proposta (1,87%). Além disso, como mencionado anteriormente, modelos estatísticos apresentam dificuldades de realizar previsões assertivas quando existem mudanças abruptas de cenário, ao contrário do que é esperado ao se utilizar um modelo neural. Isso se torna visível na semana operativa que se encerra no dia 08/06/2018 onde é observado uma variação no comportamento da carga, impulsionada principalmente pelos efeitos decorrentes das variações de temperatura. Para a semana em questão o método *Naive* apresentou um erro médio de 10%, enquanto a metodologia proposta apresentou um desvio de aproximadamente 2,6%.

A segunda forma de benchmark seria utilizando os indicadores de desempenho do ONS. Estes indicadores penalizam previsões, para o SIN, com desvios superiores a 3,0%. Embora essa metodologia seja utilizada fazendo referência a carga mensal, a mesma premissa foi adotada para a carga semanal, à vista disso, conclui-se que o modelo desenvolvido estaria atendendo a meta especificada.

Maiores detalhes sobre os resultados obtidos na previsão podem ser vistos no apêndice C.

4.2 Resultados Diários

De forma semelhante à realizada para as previsões semanais, será apresentado primeiramente a estruturação do modelo e em seguida os resultados alcançados. Durante as análises dos resultados obtidos serão expostos os desvios médios verificado, os dias que apresentaram as melhores e as piores previsões para cada tipo de horizonte estudado e a distribuição dos erros obtidos. Ressalta-se que dois horizontes de previsão serão analisados: o primeiro intitulado como 48 passos diz respeito às previsões para o primeiro dia à frente. O segundo horizonte analisado intitulado como 168 passos diz a respeito às previsões até sete dias à frente. Destaca-se que foram realizadas previsões somente para o subsistema Sudeste/Centro-Oeste, pois diversos resultados serão apresentados e analisados. Com isso, caso fossem apresentados os resultados para os quatro subsistemas não existiriam ganhos para o leitor, além de tornar o documento cansativo para uma boa leitura e entendimento.

Para o desenvolvimento do modelo horário foram utilizados como entrada os valores de temperatura prevista para as 168 horas à frente, valores verificados de carga horaria com os respectivos atrasos: 1, 24, 168, 336, 672, 8712, 8736, 8760, 8784, 8808, esses atrasos foram escolhidos em função da sazonalidade presente na curva de carga horaria, onde em geral os valor de carga apresentam uma elevada relação com o valor verificado da última hora, o valor do dia anterior, o valor da semana anterior e o valor do ano anterior. Além disso, também são apresentados três tipos de variáveis binárias: uma para informar o tipo de dia da semana para a qual a previsão é realizada (segunda, terça, ..., domingo – sete binárias para representação 1 de n), a segunda com o objetivo de qualificar se o período de previsão encontra-se ou não dentro do horário de verão (uma variável – com ou sem efeito) e por fim o último tipo possui como objetivo caracterizar

se o dia em questão é feriado ou não (uma variável – com ou sem efeito). Para treinamento do modelo foi utilizado o período que se estende do dia 01/01/2015 até o dia 31/12/2017 e para teste foi utilizado todo o ano de 2018.

Devido ao pequeno tratamento realizado para os feriados, somente a inserção de uma variável binária como entrada no modelo, duas análises serão realizadas. A primeira delas irá considerar os feriados presentes durante o ano de 2018, feriados esses apresentados na tabela 2. Já a segunda análise irá expurgar todas as previsões que possuem feriados em seus horizontes.

Destaca-se que para as previsões horarias foram geradas utilizando-se o modelo *GMM*, uma vez que durante o processo de clusterização os dados foram divididos em mais de um grupo, conforme detalhado no apêndice D. Em função disso, os intervalos de confiança foram construídos utilizando a função inversa da gaussiana, conforme descrito no capítulo 3.

4.2.1 Análise diária com feriado

Inicialmente, será apresentado através da tabela 8 os erros médios obtidos, para o subsistema Sudeste/Centro-Oeste, ao aplicar o modelo proposto no ano de 2018. Ressalta-se que nesse ano existiram diversos dias atípicos em função da ocorrência da Copa do Mundo no Rio de Janeiro. Contudo, o modelo ainda obteve um erro médio de 3,79%, para as previsões 7 dias à frente e 3,50% para as previsões 2 dias à frente.

Tabela 8 - Erros diários médios por subsistema

Subsistema	Mape 168 (%)	Mape 48 (%)
SE/CO	3,79 %	3,50 %

A tabela 9, apresenta os erros médios obtidos por mês. Nota-se que durante os meses de julho e agosto o modelo proposto apresenta um menor erro médio. O menor valor de erro médio verificados nos meses em questão, pode estar diretamente relacionado com o fato da não existência de feriados nesse horizonte.

Tabela 9 - MAPE Mensal

Meses	Mape 168 (%)	Mape 48 (%)
Janeiro	3,9%	4,0%
Fevereiro	4,5%	4,1%
Março	3,8%	3,7%
Abril	4,2%	3,4%
Mai	4,9%	4,5%
Junho	3,1%	2,9%
Julho	2,9%	2,4%
Agosto	2,7%	2,4%
Setembro	3,0%	2,5%
Outubro	3,4%	3,5%
Novembro	3,2%	3,3%
Dezembro	6,0%	5,3%

Torna-se válido verificar também o comportamento do modelo sugerido em relação aos dias da semana, esse comportamento pode ser visualizado com auxílio da tabela 10. Ressalta-se que quarta-feira e quinta-feira são os dias da semana em que o modelo apresenta os menores erros médios.

Tabela 10 - MAPE diário

Dias	Mape 168 (%)	Mape 48 (%)
Domingo	3,9%	3,5%
Segunda	4,1%	4,6%
Terça	3,9%	4,0%
Quarta	3,6%	3,0%
Quinta	3,5%	2,9%
Sexta	3,7%	3,4%
Sábado	3,7%	3,1%

A seguir, serão apresentadas nas figuras 23, 24 as distribuições dos erros percentuais alcançados por subsistema. Destaca-se que para as previsões 48 passos, os erros, em sua grande maioria, encontram-se inferiores a 3% (vide Figura 24).

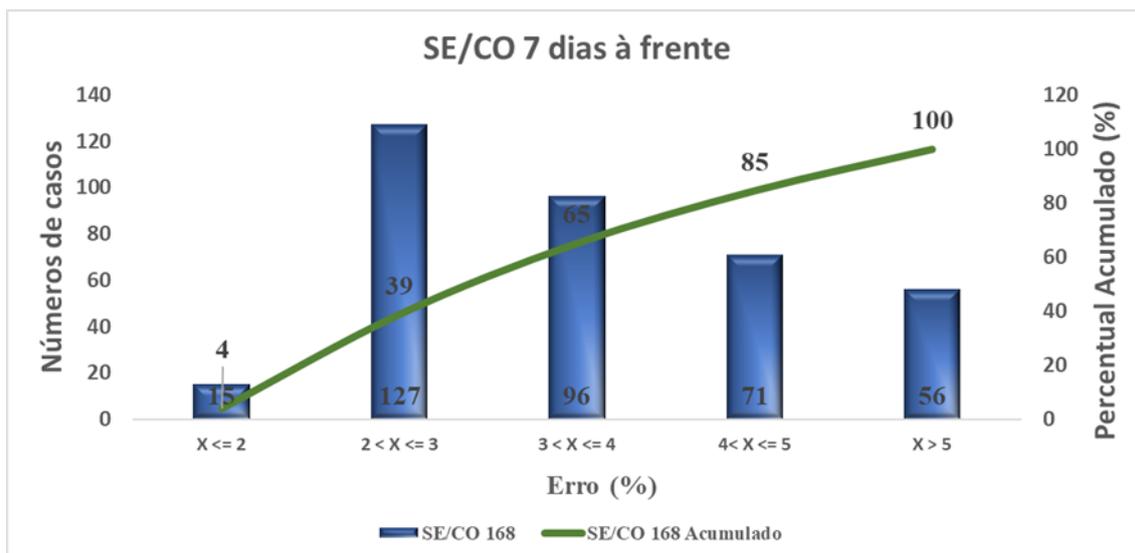


Figura 23 - Erro diário SE/CO 168 passos (SE/CO)

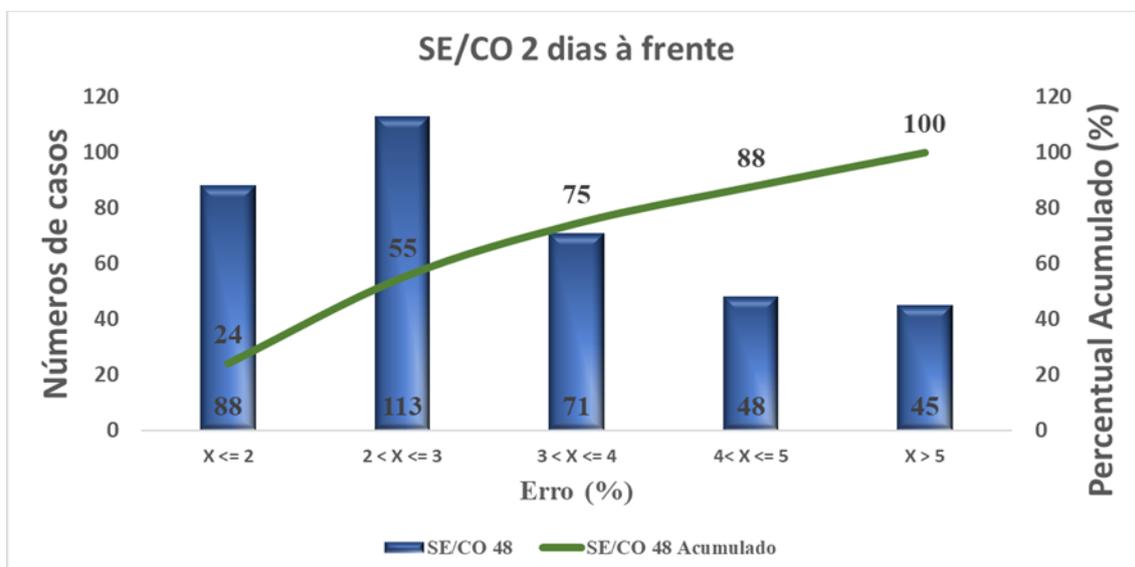


Figura 24- Erro diário SE/CO 48 passos (SE/CO)

Realizada a apresentação dos erros acumulados, as próximas figuras possuem como objetivo expor os dias que tiveram os maiores e menores desvios para cada horizonte de previsão. Com auxílio das figuras 25 e 26, observa-se que os dias que apresentaram maiores desvios são feriados, representando assim mais uma motivação para a realização de uma análise dos resultados expurgando-se tais dias.

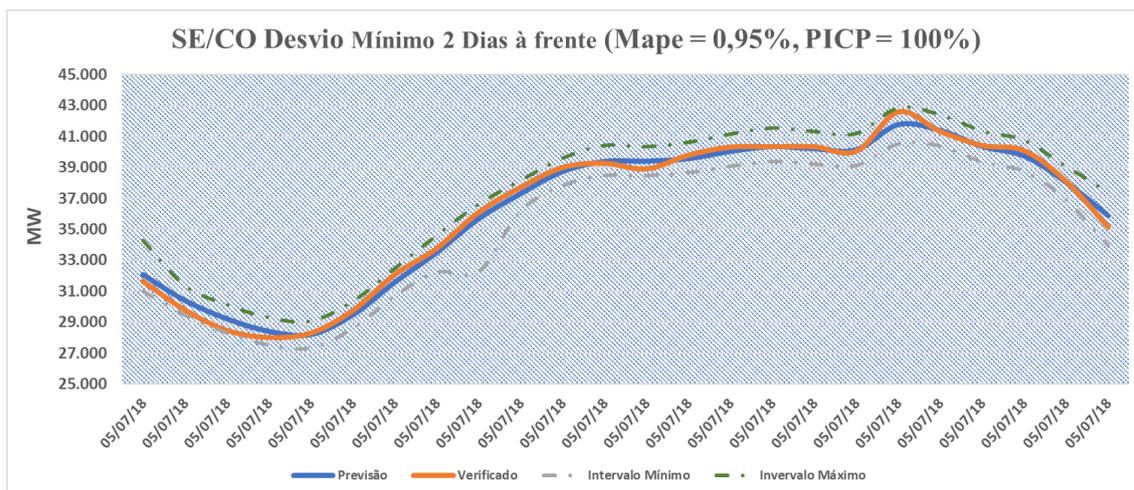


Figura 28 - Desvio Diário Mínimo 48 (SE/CO)

4.2.2 Análise diária sem feriado

Como visto nas figuras anteriores, os dias que apresentaram maiores desvios, foram aqueles que apresentam feriado em seu horizonte de previsão. Logo, torna-se necessário a verificação do desempenho do modelo expurgando-se tais dias.

De maneira semelhante as análises realizadas no tópico anterior, serão apresentadas as distribuições dos desvios por subsistema (vide figuras 29 e 30), os dias com maiores desvios em seus respectivos horizontes de previsão (vide figuras 31 e 32), bem como os erros médios obtidos para o Subsistema Sudeste/Centro-Oeste (vide tabela 11). Nesse tópico serão desconsideradas as análises em relação aos dias que tiverem os menores desvios, visto que esses dias são os mesmos que foram apresentados anteriormente.

Tabela 11 - Erros médios diários sem feriado

Subsistema	Mape 168 (%)	Mape 48 (%)
SE/CO	3,42 %	3,12 %

Com auxílio da tabela 12, pode-se verificar o comportamento mensal do modelo, destaca-se os meses de julho, agosto e setembro, onde o modelo apresentou o menores *MAPEs*. Além disso, com o auxílio da tabela 13 verifica-se o comportamento do modelo em relação aos dias da semana, nesse caso destaca-se que quarta-feira, quinta-feira e sábado foram os dias em que o modelo apresentou os menores erros médios.

Tabela 12 - MAPE Mensal (sem feriados)

Meses	Mape 168 (%)	Mape 48 (%)
Janeiro	3,9%	4,0%
Fevereiro	3,9%	3,6%
Março	3,4%	3,0%
Abril	3,5%	3,0%
Maio	4,4%	3,4%
Junho	3,1%	2,9%
Julho	2,9%	2,4%
Agosto	2,7%	2,4%
Setembro	2,8%	2,4%
Outubro	3,3%	3,6%
Novembro	2,9%	3,0%
Dezembro	4,6%	4,1%

Tabela 13 - MAPE diário (sem feriados)

Dias	Mape 168 (%)	Mape 48 (%)
Domingo	3,6%	3,3%
Segunda	3,5%	4,1%
Terça	3,5%	3,3%
Quarta	3,5%	2,8%
Quinta	3,3%	2,7%
Sexta	3,4%	2,9%
Sábado	3,3%	2,7%

Ao se realizar comparações entre a distribuição dos erros observados com feriados e sem feriados, é verificado que o *MAPE* médio diminui, bem como um aumento no percentual de erro acumulado até 3%.

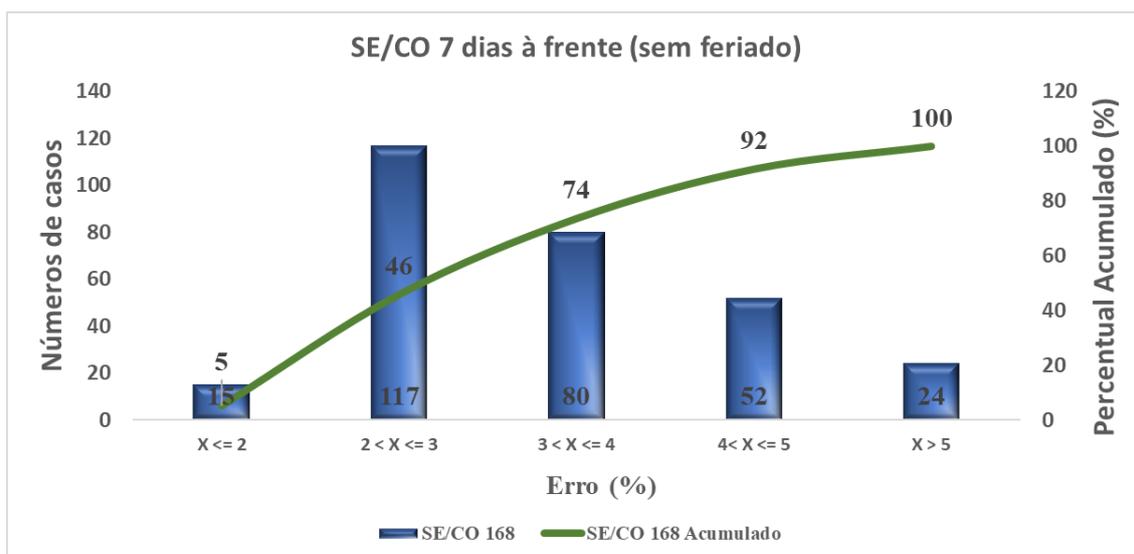


Figura 29 - Erros diários 168 passos sem feriado (SE/CO)

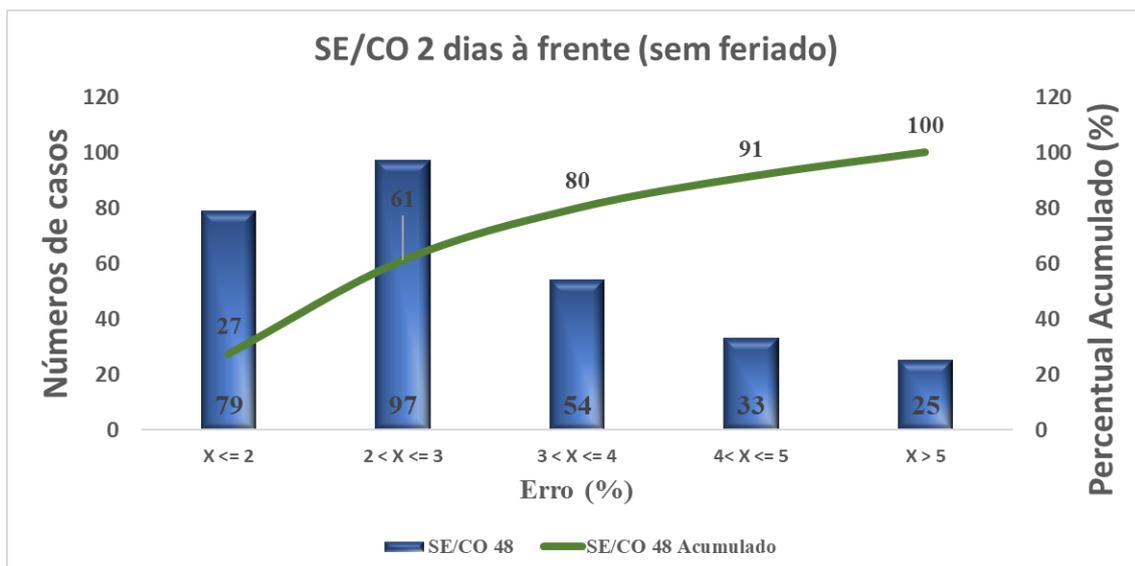


Figura 30 - Erros diários 48 passos sem feriado (SE/CO)

A seguir serão visualizados os dias que apresentaram os maiores desvios em relação ao verificado, observa-se que embora esses valores de erro ainda continuem exacerbados, quando comparados aos valores verificados anteriormente (durante as análises sem expurgar os feriados), existe um ganho nas qualidades das previsões realizadas, conforme esperado.

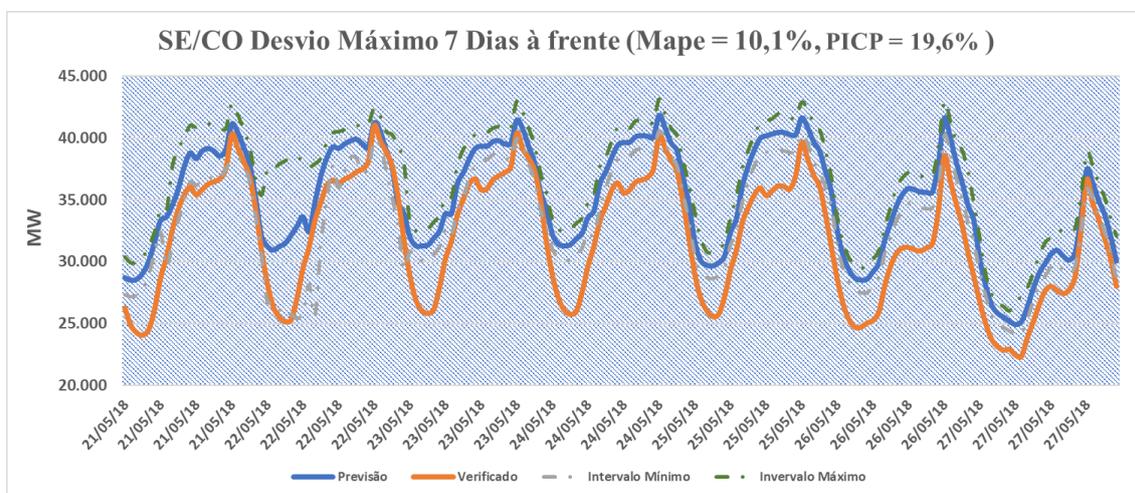


Figura 31 - Desvio Diário Máximo 168 s/feriado (SE/CO)

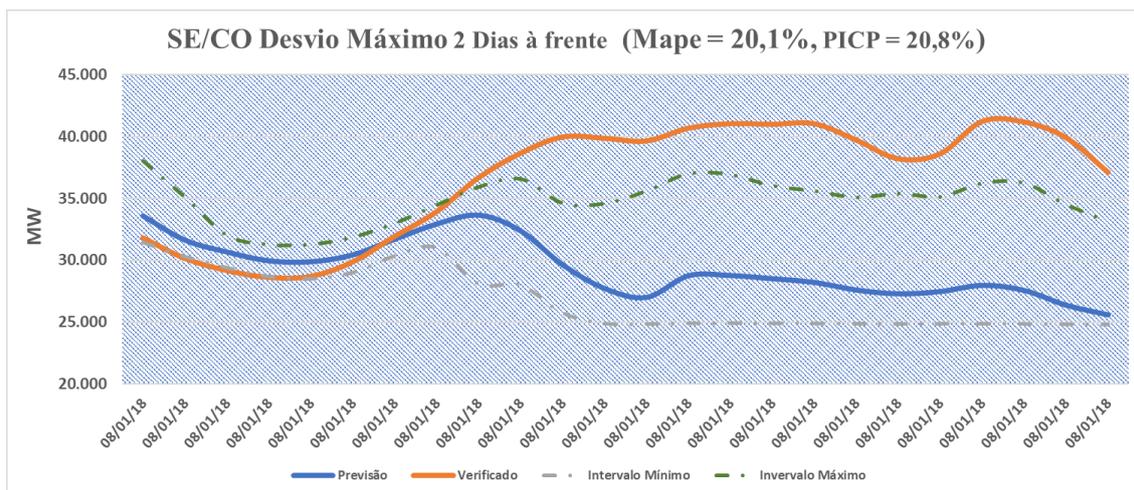


Figura 32 - Desvio Diário Máximo 48 s/feriado (SE/CO)

4.2.3 Análise de Curtose e do Coeficiente de Assimetria

Assim como realizado para as previsões semanais, serão apresentados os resultados obtidos no que tange ao Coeficiente de Assimetria e Curtose. Torna-se válido ressaltar que não será realizado uma análise expurgando os feriados, uma vez que os resultados obtidos foram semelhantes.

Os coeficientes de assimetria alcançados para os casos horários não indicam a presença de nem um viés no resultado do modelo, uma vez que os coeficientes foram de -0.3 para as previsões 168 e 48 horas à frente. Contudo, no que tange em relação a curtose foram obtidos valores de 8.2 e 12.9 para as previsões 168 horas à frente e 48 horas à frente, respectivamente. Dessa forma, pode-se inferir que os resíduos do modelo não seguem o perfil de uma distribuição normal, uma vez que as curtoses obtidas encontram-se superiores ao valor que modela uma distribuição normal.

Através das figuras 33 e 34 pode-se interpretar de uma forma visual os resultados apresentados anteriormente, uma vez que a distribuição dos resíduos se encontra simétricas e com o maior percentual de desvio próximo a zero. E em relação aos resultados da curtose, verifica-se que a distribuição dos resíduos se encontra mais alongada em sua parte superior que a distribuição normal.

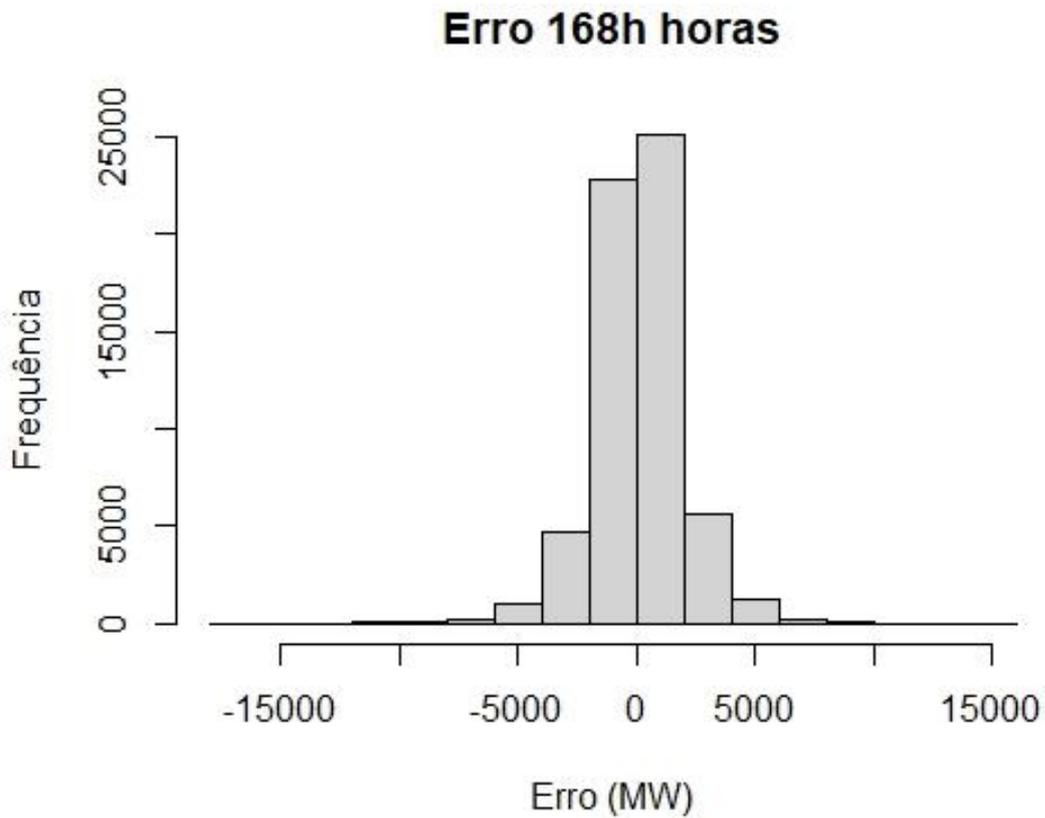


Figura 33 - Histograma dos resíduos para a previsão 168 horas

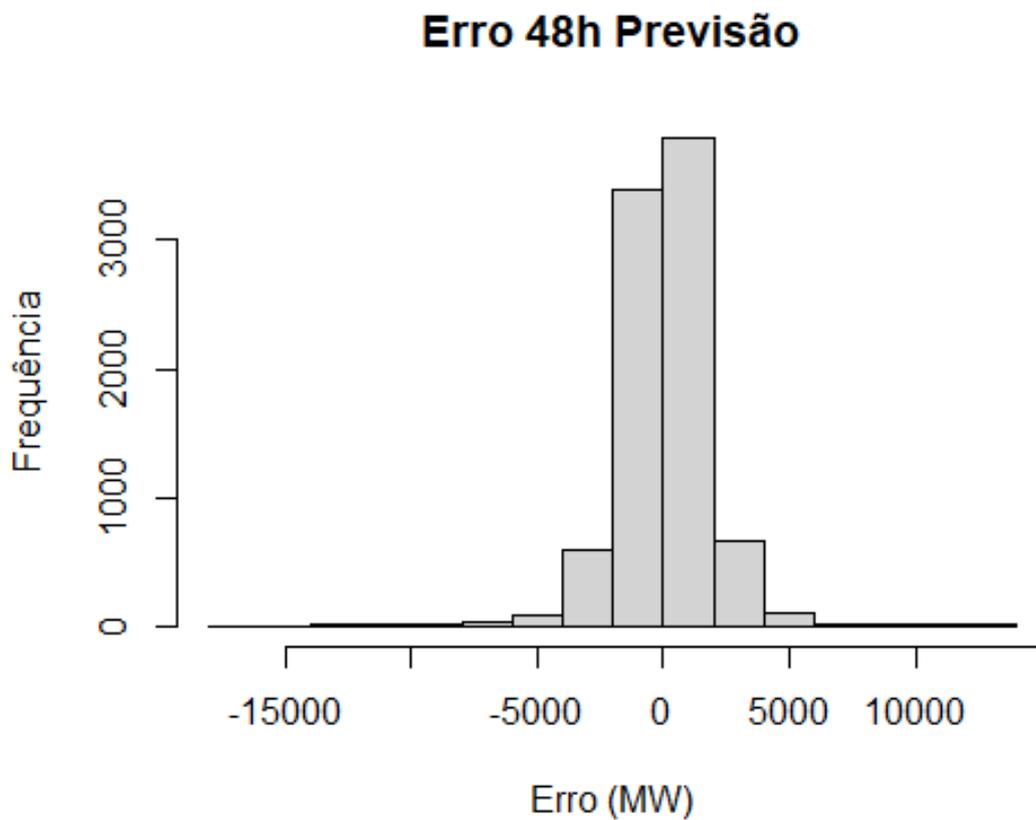


Figura 34- Histograma dos resíduos das previsões 48 horas

4.2.4 Avaliação dos intervalos de Confiança *PICP*

Assim como realizado para as previsões semanais de energia, torna-se interessante verificar a acurácia do modelo em relação aos intervalos de confiança previstos. Para isto, será utilizado novamente o método *PICP*.

Tabela 14 - Análise *PICP* (com e sem feriados)

PICP		
Com Feriado	7 dias à frente	64,3%
	2 dias à frente	67,5%
Sem Feriados	7 dias à frente	65,3%
	2 dias à frente	68,8%

Através da tabela 14 verifica-se que os intervalos de confiança construídos, em ambos os horizontes de previsão, apresentaram resultados satisfatórios, visto que encontram-se superiores a 50% de assertividade. Ressalta-se que embora os desvios médios apresentaram ganhos ao se expurgar os feriados, o mesmo ganho não é visualizado em relação aos intervalos de confiança, uma vez que os resultados se encontram marginalmente próximos. Torna-se válido ressaltar que os intervalos de confiança foram construídos utilizando os percentis 1 e 99, o que teoricamente garantiria 98% de segurança dos valores verificados estarem no interior desse intervalo. Contudo, analisando a tabela anterior esse resultado não é visto, podendo isto ser um dos pontos de melhoria do modelo desenvolvido.

Por fim, torna-se válido informar que os intervalos de confiança previstos apresentaram cerca de 3.284 MW como amplitude média. Dessa forma, ao se analisar tal resultado em conjunto com o *PICP*, pode-se inferir que os intervalos de confiança garantiriam uma maior segurança para o especialista no momento de sua decisão.

4.2.5 Benchmark

Por fim, será apresentado o Benchmark realizado a fim de verificar a qualidade das previsões. A primeira forma de Benchmark que será realizada será utilizando o método *Naive*(ingênuo). Ao se aplicar esse método obtém-se desvios médios horários de 6,5%, para o subsistema SE/CO. Ao comparar com o *MAPE* obtido pelo modelo

desenvolvido, observa-se um ganho de assertividade ao se utilizar a metodologia proposta.

Assim como realizado para as previsões semanais, estendendo-se o raciocínio dos indicadores de confiança do ONS, onde previsões com desvios superiores a 3% são penalizadas, verifica-se que embora as previsões realizadas pelo modelo apresentem um erro médio levemente superior a 3%, este erro se encontra muito próximo do valor desejado.

Maiores detalhes sobre os resultados obtidos, ao se utilizar a metodologia proposta para as previsões horarias, é apresentado no apêndice D.

5 CONCLUSÃO E TRABALHOS FUTUROS

O presente trabalho teve como objetivo o desenvolvimento de um modelo matemático autônomo através da interação de técnicas de *Machine Learning* e *Analytics*. Tais modelos foram desenvolvidos visando a melhoria contínua dos modelos tradicionais de previsão, quais sejam, modelos estatísticos ou modelos baseados em inteligência artificial que não apresentam intervalos de confiança em seus resultados e exigem grande interferência por parte do analista durante o seu desenvolvimento.

Os resultados obtidos, de forma geral, se mostraram muito promissores quando avaliado em relação as previsões médias. Destaca-se que o *MAPE* levemente superior a 3,0% obtido nas previsões horários, pode ser justificado pela ocorrência da Copa do Mundo de Futebol, no ano de 2018, fato esse não apresentado sobre nem uma forma ao modelo.

No que tange os resultados obtidos em relação ao coeficiente de assimetria e curtose, verifica-se que os resíduos obtidos na maioria dos casos, exceto nas previsões semanais para o subsistema norte, não podem ser modelados por uma única distribuição normal, uma vez que a curtose apresentada pelo modelo não segue um resultado típico para uma distribuição normal. Contudo, torna-se importante ressaltar que o resultado alcançado não inviabiliza a utilização do modelo proposto, uma vez que ao se utilizar o modelo *GMM*, pode-se aproximar de forma não paramétricas séries temporais com resíduos não gaussianos. Abordando sobre os coeficientes de assimetria alcançados, verifica-se que nas previsões horarias o modelo construído não apresenta nem um viés em seus resultados, todavia em relação as previsões semanais o modelo construído apresentou um viés positivo na maioria dos subsistemas. Em função disso, maiores análises deverão ser realizadas para a mitigação do viés apresentado.

Em relação aos intervalos de confiança previstos pelos modelos, no que diz a respeito das previsões semanais, tais intervalos demonstraram ser extremamente plausíveis de serem utilizados, uma vez que apresentarem bons resultados na análise *PICP*, assim como uma pequena dispersão em relação ao valor médio previsto. No que tange em relação as previsões horarias, os intervalos de confiança apresentaram uma assertividade superior a 60%, ratificando assim a possibilidade de sua utilização. Dessa

forma, ao se agregar os intervalos de confiança durante o momento de tomada de decisão, isto representará uma maior segurança para os especialistas.

Torna-se válido destacar que a série de carga horaria e semanal apresentam comportamentos distintos, além disso cada um dos subsistemas apresentam um comportamento diferente também. Dessa forma, ao se utilizar o mesmo modelo para ambas as series e no caso semanal o mesmo modelo para todos os subsistemas, torna-se claro a automação obtida no processo, uma vez que não foi preciso o nem uma interferência externa no modelo desenvolvido.

Alguns trabalhos futuros podem dar continuidade às conclusões apresentadas nessa dissertação. O primeiro deles se dá no âmbito do X -médias, onde foram utilizadas duas formulações, porém ambas com o mesmo critério de avaliação para a escolha da quantidade de agrupamentos. Outros critérios de informação poderiam ser testados e verificados para essa escolha como por exemplo o Akaike – AIC [79]. Além disso, a forma proposta por *Ishioka* sugere a separação inicial em dois clusters, contudo ao analisar o processo de divisão e concorrência proposto pelo método, não existiria problemas do processo partir com apenas um cluster. O segundo trabalho futuro seria a criação de um modelo voltado para a previsão de dias “especiais”, como feriados e dias enforcados, uma vez que durante as análises das previsões horários foi demonstrado tal necessidade. Além desses trabalhos citados, torna-se válido verificar a aderência do modelo ao ser desenvolvido utilizando os *GMM* versus o modelo único, dessa forma será ratificado ou não se as premissas utilizadas para justificar o uso dos modelos *GMM* encontra-se correta.

Por fim, ao se falar de carga torna-se claro a correlação que essa série possui com a temperatura, contudo existe uma elevada incerteza associada ao processo de previsão de temperaturas, principalmente com o aumento do horizonte. Dessa forma, primeiramente torna-se indicado avaliar o desempenho do modelo frente aos erros de previsão de temperatura. E dado a necessidade de previsão no PMO de múltiplas semanas à frente, até seis semanas, a inserção de cenários de temperaturas se torna indicado. Com a criação desses cenários associados aos intervalos de confiança previstos, pode-se analisar o nível de incerteza das previsões em função das variações meteorológicas.

Realizando uma síntese de tudo que foi apresentado, o desenvolvimento de um modelo autônomo encontra-se cada vez mais próximo, contudo, a interferência por parte do especialista em algumas situações, como a escolha das variáveis a serem informadas para o modelo, ainda se torna necessário. A inferência bayesiana aplicada as *MLP's* torna o processo de previsão mais robusto, essencialmente pelos intervalos de confiança que podem ser gerados a partir desse método, todavia algumas limitações podem ser verificadas durante o uso do modelo, principalmente no que tange a construção desses intervalos de confiança uma vez que, embora o resultado represente um ganho para o especialista, em alguns casos o resultado encontra-se divergindo do que se é esperado teoricamente. Logo, o presente trabalho apresentou um modelo matemático autônomo, onde a mínima interferência por parte do especialista é vista. Além disso, ao se aplicar o modelo a problemas de previsão de carga, bons resultados podem ser visualizados e ao mesmo tempo uma enorme gama de possibilidade para trabalhos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] V. H. Ferreira e A. P. A. da Silva, “Toward Estimating Autonomous Neural Network-Based Electric Load Forecasters”, *IEEE Trans. Power Syst.*, vol. 22, n° 4, p. 1554–1562, Nov. 2007, doi: 10.1109/TPWRS.2007.908438.
- [2] T. Hong e S. Fan, “Probabilistic electric load forecasting: A tutorial review”, *International Journal of Forecasting*, vol. 32, n° 3, p. 914–938, Jul. 2016, doi: 10.1016/j.ijforecast.2015.11.011.
- [3] V. H. Ferreira, “Desenvolvimento de Modelos Neurais Autônomos para Previsão de Carga Elétrica”, Tese (Doutorado), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- [4] Shyh-Jier Huang e Kuang-Rong Shih, “Short-term load forecasting via ARMA model identification including non-gaussian process considerations”, *IEEE Trans. Power Syst.*, vol. 18, n° 2, p. 673–679, maio 2003, doi: 10.1109/TPWRS.2003.811010.
- [5] L. C. M. de Andrade e I. N. da Silva, “Very Short-Term Load Forecasting Based on ARIMA Model and Intelligent Systems”, in *2009 15th International Conference on Intelligent System Applications to Power Systems*, Curitiba, nov. 2009, p. 1–6, doi: 10.1109/ISAP.2009.5352829.
- [6] L. Wei e Z. Zhen-gang, “Based on Time Sequence of ARIMA Model in the Application of Short-Term Electricity Load Forecasting”, in *2009 International Conference on Research Challenges in Computer Science*, Shanghai, China, dez. 2009, p. 11–14, doi: 10.1109/ICRCCS.2009.12.
- [7] L. Clark, D. Lou, D. Michelle, G. T., e G. C., “Day-Ahead Load Forecasting using Support Vector Regression Machines”, *ijacsa*, vol. 9, n° 3, 2018, doi: 10.14569/IJACSA.2018.090305.
- [8] M. A. A. Amin e Md. A. Hoque, “Comparison of ARIMA and SVM for Short-term Load Forecasting”, in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*, Jaipur, India, mar. 2019, p. 1–6, doi: 10.1109/IEMECONX.2019.8877077.
- [9] A. Almalag e G. Edwards, “A Review of Deep Learning Methods Applied on Load Forecasting”, Cancun, Mexico, 2017, p. 511–516, doi: 10.1109/ICMLA.2017.0-110.
- [10] M. Q. Raza e Z. Baharudin, “A review on short term load forecasting using hybrid neural network techniques”, in *2012 IEEE International Conference on Power and Energy (PECon)*, Kota Kinabalu, Malaysia, dez. 2012, p. 846–851, doi: 10.1109/PECon.2012.6450336.
- [11] H. Quan, D. Srinivasan, e A. Khosravi, “Uncertainty handling using neural network-based prediction intervals for electrical load forecasting”, *Energy*, vol. 73, p. 916–925, ago. 2014, doi: 10.1016/j.energy.2014.06.104.
- [12] M. Ghofrani, K. West, e M. Ghayekhloo, “Hybrid time series-bayesian neural network short-term load forecasting with a new input selection method”, in *2015 IEEE Power & Energy Society General Meeting*, Denver, CO, USA, jul. 2015, p. 1–5, doi: 10.1109/PESGM.2015.7286140.
- [13] H. S. H. Carlos E. Pedreira e Reinaldo Castro Souza, “Neural Networks for Short-Term Load Forecasting: A Review and Evaluation”.
- [14] I. Dimoulkas, P. Mazidi, e L. Herre, “Neural networks for GEFCom2017 probabilistic load forecasting”, *International Journal of Forecasting*, vol. 35, n° 4, p. 1409–1423, out. 2019, doi: 10.1016/j.ijforecast.2018.09.007.
- [15] “ONS - Operador Nacional do Sistema Elétrico”, *ONS - Operador Nacional do Sistema Elétrico*. <http://ons.org.br:80/paginas/sobre-o-ons/o-que-e-ons> (acessado abr. 05, 2020).

- [16] “ONS - Operador Nacional do Sistema Elétrico”, *ONS - Operador Nacional do Sistema Elétrico*. <http://ons.org.br:80/paginas/sobre-o-ons/procedimentos-de-rede/o-que-sao> (acessado abr. 05, 2020).
- [17] Operador Nacional do Sistema Elétrico, “Submódulo 5.4 Consolidação da previsão de carga para a programação diária da operação eletroenergética e para a programação de intervenções em instalações da Rede de Operação”, Operador Nacional do Sistema Elétrico, 862/2019, jan. 2020. Acessado: abr. 05, 2020. [Online]. Disponível em: <http://www.ons.org.br/%2FProcedimentosDeRede%2FM%C3%B3dulo%205%2FSubm%C3%B3dulo%205.4%2FSubm%C3%B3dulo%205.4%202020.01.pdf>.
- [18] Operador Nacional do Sistema Elétrico, “Submódulo 5.6 Consolidação da previsão de carga para a elaboração do Programa Mensal da Operação Energética”, Operador Nacional do Sistema Elétrico, 839/18. Acessado: abr. 05, 2020. [Online]. Disponível em: <http://www.ons.org.br/%2FProcedimentosDeRede%2FM%C3%B3dulo%205%2FSubm%C3%B3dulo%205.6%2FSubm%C3%B3dulo%205.6%202018.12.pdf>.
- [19] E. Cepel, “DESSEM - Modelo de Despacho Hidrotérmico de Curto Prazo”. http://www.cepel.br/pt_br/produtos/dessem-modelo-de-despacho-hidrotermico-de-curto-prazo.htm (acessado abr. 05, 2020).
- [20] A. P. A. da Silva e L. S. Moulin, “Confidence intervals for neural network based short-term load forecasting”, *IEEE Trans. Power Syst.*, vol. 15, n° 4, p. 1191–1196, nov. 2000, doi: 10.1109/59.898089.
- [21] I. N. da Silva, D. H. Spatti, e R. A. Flauzino, *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas*, 2° ed. São Paulo, Brasil: Artliber, 2016.
- [22] H. L. M. do Amaral, “Desenvolvimento de uma nova metodologia para previsão do consumo de energia elétrica de curto prazo utilizando redes neurais artificiais e decomposição de séries temporais.”, Tese (Doutorado), Universidade de São Paulo, São Paulo, Brasil, 2019.
- [23] D. D. S. Lopes, “Modelos Neurais Autônomos para Classificação e Localização de Defeitos em Linhas de Transmissão”, Dissertação (Mestrado), Universidade Federal Fluminense, Niterói, Rio de Janeiro, 2017.
- [24] J. Brownlee, *Deep Learning with Python*, vol. 1.15. Vermont-Australia: Machine Learning Martery, 2018.
- [25] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, New York: Oxford University Press, 1995.
- [26] B. Efron e R. J. Tibshirani, *An Introduction to the Bootstrap*, vol. 57. CRC Press; Edição: Softcover reprint of the original 1st ed. 1993 (15 de maio de 1994), 1994.
- [27] Z. Cataltepe, Y. S. Abu-Mostafa, e M. Magdon-Ismail, “No Free Lunch for Early Stopping”, *Neural Computation*, vol. 11, n° 4, p. 995–1009, maio 1999, doi: 10.1162/089976699300016557.
- [28] D. J. C. Mackey, “Bayesian Methods for Adaptive Models”, Dissertação (Mestrado), California Institute of Technology, Pasadena, California, USA, 1992.
- [29] Yuan Ning, Yufeng Liu, e Qiang Ji, “Bayesian - BP Neural Network based Short-term Load Forecasting for power system”, in *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, Chengdu, China, ago. 2010, p. V2-89-V2-93, doi: 10.1109/ICACTE.2010.5579151.
- [30] Hui-Feng Shi e Yan-Xia Lu, “Bayesian Neural networks for short term load forecasting”, in *2009 International Conference on Wavelet Analysis and Pattern Recognition*, Baoding, China, jul. 2009, p. 160–165, doi: 10.1109/ICWAPR.2009.5207407.

- [31] X. Zhang e S. Mahadevan, “Bayesian neural networks for flight trajectory prediction and safety assessment”, *Decision Support Systems*, vol. 131, p. 113246, abr. 2020, doi: 10.1016/j.dss.2020.113246.
- [32] F. Vivarelli e C. K. I. Williams, “Comparing Bayesian neural network algorithms for classifying segmented outdoor images”, *Neural Networks*, vol. 14, n° 4, p. 427–437, maio 2001, doi: 10.1016/S0893-6080(01)00024-7.
- [33] U. v. Toussaint, S. Gori, e V. Dose, “Invariance priors for Bayesian feed-forward neural networks”, *Neural Networks*, vol. 19, n° 10, p. 1550–1557, dez. 2006, doi: 10.1016/j.neunet.2006.01.017.
- [34] M. Ghofrani, D. Carson, e M. Ghayekhloo, “Hybrid clustering-time series-bayesian neural network short-term load forecasting method”, in *2016 North American Power Symposium (NAPS)*, Denver, CO, USA, set. 2016, p. 1–5, doi: 10.1109/NAPS.2016.7747865.
- [35] H. S. Hippert e J. W. Taylor, “An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting”, *Neural Networks*, vol. 23, n° 3, p. 386–395, abr. 2010, doi: 10.1016/j.neunet.2009.11.016.
- [36] D. Niu, H. Shi, e D. D. Wu, “Short-term load forecasting using bayesian neural networks learned by Hybrid Monte Carlo algorithm”, *Applied Soft Computing*, vol. 12, n° 6, p. 1822–1827, jun. 2012, doi: 10.1016/j.asoc.2011.07.001.
- [37] Y. Kwon, J.-H. Won, B. J. Kim, e M. C. Paik, “Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation”, *Computational Statistics & Data Analysis*, vol. 142, p. 106816, fev. 2020, doi: 10.1016/j.csda.2019.106816.
- [38] R. P. Martins, “Previsão Probabilística de Energia Solar Utilizando Redes Neurais com Inferência Bayesiana e Misturas de Gaussianas”, Dissertação (Mestrado), Universidade Federal Fluminense, Niterói, Rio de Janeiro, 2018.
- [39] K. M. CASSIANO, “Análise de Séries Temporais Usando Análise Espectral Singular (SSA) e Clusterização de Suas Componentes Baseada em Densidade”, Tese (Doutorado), Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio, Rio de Janeiro, 2014.
- [40] L. Kaufman e P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, 1° ed. Wiley-Blackwell, 2005.
- [41] A. K. J. Richard C. Dubes, *Algorithms for Clustering Data*. New Jersey: Prentice-Hall, 1988.
- [42] A. K. Jain, “Data clustering: 50 years beyond K-means”, *Pattern Recognition Letters*, vol. 31, n° 8, p. 651–666, jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
- [43] R. L. Lachi e H. V. da Rocha, “Aspectos básicos de clustering: conceitos e técnicas. Universidade Estadual de Campinas”, Instituto de Computação Universidade Estadual de Campinas, Campinas, 05–003, 2005.
- [44] S. Theodoridis e K. Koutroumbas, *Pattern Recognition*, 4° ed. USA: Academic Press, 2009.
- [45] J. A. Sacramento, “Variabilidade de Índices de Confiança de Sistemas de Potência”, Dissertação (Mestrado), Universidade Federal Fluminense, Niteroi, Rio de Janeiro, 2012.
- [46] A. F. de Oliveira, “Favorecendo o Desempenho do k-Means via Métodos de Inicialização de Centroides”, Dissertação (Mestrado), Centro Universitário Campo Limpo Paulista, Campo Limpo Paulista, 2018.
- [47] R. Goldschmidt e E. Passo, *Data mining: um guia Prático*. Rio de Janeiro: Campos Elsevier, 2005.

- [48] C. C. Aggarwal e K. R. Chandan, *Data Clustering: Algorithms and Applications*. Chapman and Hall/CRC, 2013.
- [49] J. Macqueen, “SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS”, 1967, p. 281–297.
- [50] L. C. P. Velasco, N. R. Estoperez, R. J. R. Jayson, C. J. T. Sabijon, e V. C. Sayles, “Day-ahead Base, Intermediate, and Peak Load Forecasting using K-Means and Artificial Neural Networks”, (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 9, n^o 2, p. 62–67, 2018, doi: DOI: 10.14569/IJACSA.2018.090210.
- [51] A. Al-Wakeel, J. Wu, e N. Jenkins, “k -means based load estimation of domestic smart meter measurements”, *Applied Energy*, vol. 194, p. 333–342, maio 2017, doi: 10.1016/j.apenergy.2016.06.046.
- [52] S. A. Azad, A. B. M. S. Ali, e P. Wolfs, “Identification of typical load profiles using K-means clustering algorithm”, in *Asia-Pacific World Congress on Computer Science and Engineering*, Nadi, Fiji, nov. 2014, p. 1–6, doi: 10.1109/APWCCSE.2014.7053855.
- [53] M. E. Celebi, Org., *Partitional Clustering Algorithms*. Springer International Publishing, 2015.
- [54] T. Ishioka, “An Expansion of X-means for Automatically Determining the Optimal Number of Clusters - Progressive Iterations of K-means and Merging of the Clusters”, Calgary, Alberta, Canada, jul. 2005, p. 91–96.
- [55] P. Humas e S. K. Wasan, “Comparative Analysis of k-mean Based Algorithms”, abr. 2010, vol. 10, p. 314–318.
- [56] D. Pelleg e A. Moore, “X-means: Extending K-means with Efficient Estimation of the Number of Cluster”, 2000, p. 727–734.
- [57] G. J. McLachalan e T. Krishnan, *The EM Algorithm and Extensions*, 2^o ed. New York: Wiley Series in Probability and Statistics, 2008.
- [58] G. J. McLachalan e D. Peel, *Finite Mixture Models*, 1^o ed. New York: Wiley Series in Probability and Statistics, 2000.
- [59] N. M. Portela, “Modelo de Mistura de Gaussianas Fuzzy Contextual”, Tese (Doutorado), Universidade Federal de Pernambuco, Pernambuco, 2015.
- [60] C. M. Bishop, *Pattern Recognition and Machine Learning*, 2^o ed. New York: Springer, 2011.
- [61] J. Sun e Z. Ji, “Bounded asymmetric Gaussian mixture model with spatial constraint for image segmentation”, in *2016 International Conference on Progress in Informatics and Computing (PIC)*, Shanghai, China, dez. 2016, p. 369–373, doi: 10.1109/PIC.2016.7949528.
- [62] S. T. Ali, K. Goyal, e J. Singhai, “Moving object detection using self adaptive Gaussian Mixture Model for real time applications”, in *2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)*, Bhopal, out. 2017, p. 153–156, doi: 10.1109/RISE.2017.8378144.
- [63] N. Ding, H. Ma, H. Gao, Y. Ma, e G. Tan, “Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model”, *Computers & Electrical Engineering*, vol. 79, p. 106458, out. 2019, doi: 10.1016/j.compeleceng.2019.106458.
- [64] R. H. Sun e R. J. Chol, “Subspace Gaussian mixture based language modeling for large vocabulary continuous speech recognition”, *Speech Communication*, vol. 117, p. 21–27, fev. 2020, doi: 10.1016/j.specom.2020.01.001.
- [65] P. 06 S. 2019 | 16:30 GMT, “The Top Programming Languages 2019 - IEEE Spectrum”, *IEEE Spectrum: Technology, Engineering, and Science News*.

- <https://spectrum.ieee.org/computing/software/the-top-programming-languages-2019> (acessado dez. 27, 2020).
- [66] T. F. Covões, “Algoritmos evolutivos para modelos de mistura de gaussianas em problemas com e sem restrições”, Doutorado em Ciências de Computação e Matemática Computacional, Universidade de São Paulo, São Carlos, 2014.
- [67] decoleseufuturo, “Siglas dos aeroportos do Brasil”, *Decole seu Futuro*, abr. 05, 2017. <https://www.aerotd.com.br/decoleseufuturo/siglas-dos-aeroportos-do-brasil/> (acessado dez. 27, 2020).
- [68] P. R. A. C. em Intermediário e M. L. de Energia, “O que é o PLD – Preço de Liquidação das Diferenças? – Mercado Livre de Energia – Grugreen”. <http://grugreen.eng.br/o-que-e-o-pld-preco-de-liquidacao-das-diferencas/> (acessado abr. 03, 2020).
- [69] “2.6 - Curtose - Estatística Básica | Portal Action”. <http://www.portalaction.com.br/estatistica-basica/26-curtose> (acessado dez. 28, 2020).
- [70] “2.5 - Coeficiente de Assimetria - Estatística Básica | Portal Action”. <http://www.portalaction.com.br/estatistica-basica/25-coeficiente-de-assimetria> (acessado dez. 28, 2020).
- [71] D. N. Joanes e C. A. Gill, “Comparing measures of sample skewness and kurtosis”, *J Royal Statistical Soc D*, vol. 47, n° 1, p. 183–189, mar. 1998, doi: 10.1111/1467-9884.00122.
- [72] W. Liao, X. Liu, G. Li, e T. Chen, “Evaluation of thermal-hydraulic performance of dimpled tube from the perspective of skewness and kurtosis”, *International Journal of Thermal Sciences*, vol. 156, p. 106469, out. 2020, doi: 10.1016/j.ijthermalsci.2020.106469.
- [73] “1.3.5.11. Measures of Skewness and Kurtosis”. <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm> (acessado dez. 28, 2020).
- [74] A. Mashlakov, L. Lensu, A. Kaarna, V. Tikka, e S. Honkapuro, “Probabilistic Forecasting of Battery Energy Storage State-of-Charge under Primary Frequency Control”, *IEEE J. Select. Areas Commun.*, vol. 38, n° 1, p. 96–109, jan. 2020, doi: 10.1109/JSAC.2019.2952195.
- [75] A. Kavousi-Fard, A. Khosravi, e S. Nahavandi, “A New Fuzzy-Based Combined Prediction Interval for Wind Power Forecasting”, *IEEE Trans. Power Syst.*, vol. 31, n° 1, p. 18–26, jan. 2016, doi: 10.1109/TPWRS.2015.2393880.
- [76] A. Khosravi, S. Nahavandi, D. Creighton, e A. F. Atiya, “Lower Upper Bound Estimation Method for Construction of Neural Network-Based Prediction Intervals”, *IEEE Trans. Neural Netw.*, vol. 22, n° 3, p. 337–346, mar. 2011, doi: 10.1109/TNN.2010.2096824.
- [77] “ONS - Operador Nacional do Sistema Elétrico”, *ONS - Operador Nacional do Sistema Elétrico*. http://ons.org.br:80/Paginas/sobre-ons/indicadores-ons/indicador_ipc.aspx (acessado abr. 08, 2020).
- [78] “Portaria estabelece feriados nacionais e pontos facultativos em 2018”, *Ministério da Economia*. <http://www.planejamento.gov.br/noticias/portaria-estabelece-feriados-nacionais-e-pontos-facultativos-em-2018> (acessado abr. 08, 2020).
- [79] P. C. Emiliano, “FUNDAMENTOS E APLICAÇÕES DOS CRITÉRIOS DE INFORMAÇÃO: AKAIKE E BAYESIANO”, Dissertação (Mestrado), Universidade Federal de Lavras, Lavras, Minas Gerais, 2009.

APÊNDICES

A. Construção dos intervalos de confiança em modelos GMM

A seguir será detalhado sobre a forma de construção dos intervalos de confiança ao se utilizar os modelos *GMM*. Supondo que o processo de X-médias indique a presença de 3 *clusters*, serão treinados 1 modelo de classificação e 3 modelos de regressão, como o objetivo de realizar as previsões das probabilidades do padrão de entrada pertencer a cada um dos *clusters* dos valores médios e das variâncias associadas aquele padrão, sendo tais previsões detalhadas na tabela 15.

Tabela 15 - Previsão diária exemplo

Clusters	1	2	3
Saídas estimadas	33.800	31.900	33.300
Variância estimadas	8,18E+06	1,33E+05	2,15E+05
Probabilidade	0,25	0,05	0.70

Após esse processo, o valor médio previsto é obtido aplicando-se a equação 3.1. Observa-se que o cluster 3 possui a maior probabilidade de ocorrência, logo, o valor médio previsto (33.355 MW) será mais próximo da saída estimada desse *cluster*. Durante a construção dos intervalos de confiança, o mesmo raciocínio se torna válido, uma vez que as probabilidades testes serão ponderadas pela probabilidade de ocorrência do respectivo *cluster*.

Em seguida, será calculado os valores de cada um dos percentis (1 até 99) com o objetivo de construir os intervalos de confiança. Nesse momento utiliza-se a função inversa da gaussiana para determinar o valor de “x_teste” em que sua probabilidade seja igual ao percentil desejado. Ressalta-se que o modelo *GMM* representa o espaço através de “N” distribuições gaussianas, e, em função disso o cálculo para encontrar o valor de “x_teste” leva em consideração essas distribuições modeladas (média e variância prevista para cada um dos *clusters*), assim como suas respectivas probabilidades de ocorrência.

Dado que inúmeras distribuições são utilizadas para modelar o espaço, encontrar o valor de “x_teste” que apresente probabilidade igual à do percentil desejado, não se

torna uma tarefa simples. Com isso, o cálculo para encontrar o valor de “x_teste” desejado será iterativo, conforme apresentado no pseudo código a seguir:

Passo 1 - Define-se x1 e x2 iguais aos limites de busca de x_teste (no presente trabalho esses limites foram definidos de acordo com os limites de normalização da variável prevista).

Passo 2 - As probabilidades para as variáveis x1 e x2 são calculadas. Destaca-se que as probabilidades de finais serão a ponderação entre as probabilidades das gaussianas modeladas e as probabilidades de ocorrência de cada um dos *clusters*.

Passo 3 - Compara-se prob(x1) e prob(x2) com o percentil desejado, $dif(x) = prob(x) - \text{percentil}$, caso o valor absoluto de uma dessas diferenças ($dif(x1)$ ou $dif(x2)$) esteja dentro da tolerância, o processo é encerrado e x_teste é ajustado igual a variável da respectiva probabilidade em que gerou a diferença dentro da tolerância. Caso contrário, x1 e x2 são atualizados.

Passo 4 - Atualização de x1 e x2: Existem três possíveis formas de atualização dessas variáveis, essas atualizações irão ocorrer da seguinte forma:

a. Caso $dif(x1) < 0$ e $dif(x2) > 0$:

i. $x2 = (x1 + x2) / 2$

b. Caso $dif(x1) < 0$ e $dif(x2) < 0$:

i. $x1 = x2$ e $x2 = \text{limite superior}$

c. Caso $dif(x1) > 0$ e $dif(x2) > 0$

i. $x2 = x1$ e $x1 = \text{limite inferior}$

Passo 5 - Retorne ao passo 1.

Esse processo é repetido para todos os percentis sendo o intervalo de confiança construído ao final do processo. Nessa dissertação foram utilizados como intervalo de confiança os percentis 1 e 99.

B. INTERVALOS DE CONFIANÇA SEMANAIS**Tabela 16 - Intervalo de Confiança Semanal SE/CO**

FSO	INTERVALO MÍNIMO 3DP	INTERVALO MÁXIMO 3DP	INTERVALO MÍNIMO 2DP	INTERVALO MÁXIMO 2DP	INTERVALO MÍNIMO 1DP	INTERVALO MÁXIMO 1DP
31/12/2016	38478	43006	39233	42251	39988	41497
07/01/2017	36972	41208	37678	40502	38384	39796
14/01/2017	36197	40728	36952	39973	37707	39218
21/01/2017	36478	40675	37178	39976	37877	39276
28/01/2017	37093	41641	37851	40883	38609	40125
04/02/2017	37501	41817	38221	41098	38940	40379
11/02/2017	39235	43385	39926	42693	40618	42001
18/02/2017	38967	43477	39718	42725	40470	41973
25/02/2017	35557	40502	36381	39678	37205	38854
04/03/2017	37444	41810	38172	41082	38899	40355
11/03/2017	37640	42145	38390	41394	39141	40643
18/03/2017	34558	39380	35362	38576	36165	37773
25/03/2017	35071	39732	35847	38955	36624	38178
01/04/2017	35478	39653	36174	38957	36870	38261
08/04/2017	34076	38890	34878	38088	35681	37285
15/04/2017	32857	37756	33673	36939	34490	36123
22/04/2017	32696	37086	33427	36354	34159	35623
29/04/2017	31660	39136	32906	37890	34152	36644
06/05/2017	32027	36267	32734	35560	33440	34854
13/05/2017	32169	36079	32821	35427	33473	34776
20/05/2017	32051	36197	32742	35506	33433	34815
27/05/2017	32028	36105	32708	35425	33387	34746
03/06/2017	32423	36591	33117	35897	33812	35202
10/06/2017	29048	34493	29955	33586	30863	32678
17/06/2017	30984	35332	31709	34607	32433	33883
24/06/2017	31635	35771	32324	35082	33014	34393
01/07/2017	30846	35228	31576	34498	32307	33767
08/07/2017	31071	35868	31871	35069	32670	34269
15/07/2017	31139	35309	31834	34614	32529	33919
22/07/2017	31460	35369	32112	34718	32763	34066
29/07/2017	31472	35623	32164	34931	32856	34239
05/08/2017	31233	35303	31911	34625	32589	33946
12/08/2017	31080	35424	31804	34700	32528	33976
19/08/2017	31258	35146	31906	34498	32554	33850
26/08/2017	31877	35947	32555	35269	33233	34590
02/09/2017	30480	35222	31270	34432	32061	33641
09/09/2017	33630	37941	34348	37222	35067	36504
16/09/2017	33932	38229	34648	37513	35364	36797
23/09/2017	34149	38100	34808	37442	35466	36783

30/09/2017	34105	38262	34798	37569	35491	36876
07/10/2017	30830	44581	33122	42289	35413	39997
14/10/2017	34836	38930	35518	38248	36201	37565
21/10/2017	34263	38255	34928	37590	35593	36924
28/10/2017	32687	37033	33412	36309	34136	35585
04/11/2017	33717	37814	34400	37131	35083	36448
11/11/2017	31446	41892	33187	40151	34928	38410
18/11/2017	34282	38291	34950	37623	35618	36954
25/11/2017	34266	38205	34922	37549	35579	36892
02/12/2017	34375	38353	35038	37690	35701	37027
09/12/2017	34653	38595	35310	37938	35967	37281
16/12/2017	35884	39922	36557	39249	37230	38576
23/12/2017	35272	39675	36006	38941	36740	38207
30/12/2017	33363	44542	35226	42679	37089	40816
06/01/2018	34692	38637	35349	37979	36007	37322
13/01/2018	36597	40806	37299	40104	38000	39403
20/01/2018	37887	41976	38569	41295	39250	40613
27/01/2018	36268	40393	36956	39705	37643	39018
03/02/2018	35845	39914	36523	39236	37201	38558
10/02/2018	34731	39337	35499	38570	36267	37802
17/02/2018	36789	40728	37445	40072	38102	39415
24/02/2018	37038	40990	37697	40331	38355	39672
03/03/2018	37020	40878	37663	40235	38306	39592
10/03/2018	38089	42380	38804	41665	39519	40950
17/03/2018	37979	42036	38656	41360	39332	40684
24/03/2018	35823	40802	36653	39972	37482	39142
31/03/2018	36991	41058	37669	40380	38347	39702
07/04/2018	36843	40782	37500	40126	38156	39469
14/04/2018	34546	38854	35264	38136	35982	37418
21/04/2018	34324	39171	35132	38363	35940	37556
28/04/2018	32621	40275	33896	39000	35172	37724
05/05/2018	34084	37715	34689	37110	35294	36504
12/05/2018	34238	37994	34864	37368	35490	36742
19/05/2018	32410	36509	33093	35826	33776	35142
26/05/2018	30893	35413	31647	34660	32400	33906
02/06/2018	30310	34521	31012	33819	31714	33117
09/06/2018	31250	35710	31993	34967	32737	34224
16/06/2018	30521	34959	31261	34219	32000	33479
23/06/2018	31898	36224	32619	35503	33340	34782
30/06/2018	31789	36114	32510	35393	33231	34672
07/07/2018	31329	35550	32032	34846	32736	34143
14/07/2018	32409	36448	33082	35775	33755	35102
21/07/2018	32661	36869	33362	36168	34064	35466
28/07/2018	32152	36117	32813	35456	33473	34795
04/08/2018	31689	35793	32373	35109	33057	34425

11/08/2018	31816	35732	32469	35080	33122	34427
18/08/2018	32819	36946	33507	36258	34195	35570
25/08/2018	32591	36439	33232	35798	33874	35156
01/09/2018	29864	35179	30749	34293	31635	33407
08/09/2018	32159	36117	32819	35458	33479	34798
15/09/2018	33145	37183	33818	36510	34491	35837
22/09/2018	34278	38305	34949	37634	35620	36962
29/09/2018	34946	38939	35612	38274	36277	37608
06/10/2018	32077	38798	33197	37677	34317	36557
13/10/2018	34278	38308	34950	37636	35621	36965
20/10/2018	33627	37697	34305	37018	34983	36340
27/10/2018	32280	38280	33280	37280	34280	36280
03/11/2018	32971	37036	33648	36358	34326	35681
10/11/2018	33318	39538	34354	38501	35391	37465
17/11/2018	34401	38441	35074	37767	35747	37094
24/11/2018	34905	38858	35564	38199	36223	37540
01/12/2018	34826	38808	35490	38144	36153	37481
08/12/2018	35498	39688	36196	38989	36895	38291
15/12/2018	38858	43077	39561	42374	40264	41671
22/12/2018	35327	39641	36046	38922	36765	38203
29/12/2018	31223	42036	33025	40234	34828	38432

Tabela 17 - Intervalo de Confiança Semanal Sul

FSO	INTERVALO MÍNIMO 3DP	INTERVALO MÁXIMO 3DP	INTERVALO MÍNIMO 2DP	INTERVALO MÁXIMO 2DP	INTERVALO MÍNIMO 1DP	INTERVALO MÁXIMO 1DP
31/12/2016	10951	12664	11236	12378	11522	12093
07/01/2017	10927	12678	11219	12386	11511	12094
14/01/2017	11099	12782	11379	12501	11660	12221
21/01/2017	10906	12625	11193	12339	11479	12052
28/01/2017	10736	12451	11021	12165	11307	11879
04/02/2017	11116	12894	11412	12598	11709	12301
11/02/2017	11326	13033	11611	12748	11895	12464
18/02/2017	11718	13530	12020	13228	12322	12926
25/02/2017	10878	13000	11231	12646	11585	12293
04/03/2017	11120	12887	11414	12592	11709	12298
11/03/2017	10694	12503	10996	12201	11297	11900
18/03/2017	10191	11996	10491	11695	10792	11394
25/03/2017	10101	11813	10386	11528	10672	11242
01/04/2017	10119	11818	10402	11535	10685	11251
08/04/2017	9177	10842	9454	10564	9732	10287
15/04/2017	8329	10988	8772	10545	9216	10102
22/04/2017	8916	10771	9225	10462	9535	10153
29/04/2017	7214	12627	8117	11725	9019	10823

06/05/2017	9185	10902	9471	10615	9757	10329
13/05/2017	9250	10869	9520	10599	9789	10329
20/05/2017	9295	10930	9568	10658	9840	10385
27/05/2017	9090	10815	9377	10528	9665	10240
03/06/2017	9144	10821	9423	10542	9703	10262
10/06/2017	8285	10473	8649	10108	9014	9743
17/06/2017	9208	10943	9497	10654	9787	10365
24/06/2017	9345	11072	9633	10785	9921	10497
01/07/2017	9430	11133	9713	10849	9997	10565
08/07/2017	9407	11111	9691	10827	9975	10543
15/07/2017	9393	11144	9685	10852	9977	10560
22/07/2017	9611	11372	9904	11079	10198	10785
29/07/2017	9353	11064	9638	10778	9923	10493
05/08/2017	9376	11041	9654	10764	9931	10486
12/08/2017	9286	11023	9575	10733	9865	10444
19/08/2017	9249	10989	9539	10699	9829	10409
26/08/2017	9751	11403	10026	11127	10301	10852
02/09/2017	8510	10439	8831	10117	9153	9796
09/09/2017	9612	11390	9908	11094	10205	10797
16/09/2017	9810	11566	10103	11274	10396	10981
23/09/2017	9554	11265	9839	10980	10124	10695
30/09/2017	9779	11482	10063	11198	10346	10914
07/10/2017	8184	10987	8652	10520	9119	10053
14/10/2017	9704	11355	9979	11080	10254	10805
21/10/2017	9581	11280	9864	10997	10147	10714
28/10/2017	8535	10526	8867	10194	9199	9862
04/11/2017	9250	10994	9540	10704	9831	10413
11/11/2017	8926	14132	9794	13264	10661	12397
18/11/2017	9817	11527	10102	11242	10387	10957
25/11/2017	9713	11429	9999	11143	10285	10857
02/12/2017	10107	11802	10389	11519	10672	11237
09/12/2017	10564	12337	10860	12042	11155	11746
16/12/2017	10874	12644	11169	12349	11464	12054
23/12/2017	9603	12292	10051	11844	10499	11396
30/12/2017	8214	12957	9004	12167	9795	11376
06/01/2018	9607	11456	9915	11148	10223	10840
13/01/2018	10797	12646	11105	12338	11413	12030
20/01/2018	11036	12789	11328	12497	11620	12205
27/01/2018	10506	12387	10820	12074	11133	11760
03/02/2018	10803	12657	11112	12348	11421	12039
10/02/2018	9820	12203	10217	11806	10614	11409
17/02/2018	10110	12580	10522	12169	10933	11757
24/02/2018	10927	12807	11240	12494	11554	12180
03/03/2018	10976	12858	11290	12545	11604	12231
10/03/2018	11517	13417	11833	13100	12150	12783

17/03/2018	11027	12936	11345	12618	11664	12300
24/03/2018	9415	12111	9864	11662	10314	11213
31/03/2018	9947	11888	10271	11564	10594	11241
07/04/2018	10905	12750	11213	12443	11520	12135
14/04/2018	10305	12294	10636	11963	10968	11631
21/04/2018	10997	14557	11591	13964	12184	13371
28/04/2018	10107	12665	10533	12239	10960	11812
05/05/2018	10076	11961	10390	11647	10704	11333
12/05/2018	9919	11803	10233	11489	10547	11175
19/05/2018	9270	11672	9670	11271	10070	10871
26/05/2018	9059	11274	9429	10905	9798	10536
02/06/2018	8213	11288	8725	10775	9238	10263
09/06/2018	9304	11404	9654	11054	10004	10704
16/06/2018	9292	11297	9626	10962	9960	10628
23/06/2018	9279	11220	9602	10896	9926	10573
30/06/2018	9092	10989	9408	10673	9724	10356
07/07/2018	9505	11360	9814	11050	10123	10741
14/07/2018	10248	11513	10458	11302	10669	11091
21/07/2018	9816	11776	10143	11450	10470	11123
28/07/2018	9712	11661	10037	11337	10362	11012
04/08/2018	9648	11609	9975	11282	10302	10956
11/08/2018	9794	11759	10122	11432	10449	11104
18/08/2018	9848	11598	10139	11306	10431	11014
25/08/2018	9730	11577	10037	11269	10345	10961
01/09/2018	8354	11176	8825	10706	9295	10235
08/09/2018	9301	11144	9608	10837	9915	10529
15/09/2018	9451	11220	9745	10925	10040	10630
22/09/2018	9711	11604	10027	11289	10342	10973
29/09/2018	9525	11314	9823	11016	10121	10718
06/10/2018	8681	10730	9023	10389	9364	10047
13/10/2018	9606	11459	9914	11150	10223	10841
20/10/2018	9722	11484	10016	11190	10309	10897
27/10/2018	8837	11034	9203	10668	9569	10302
03/11/2018	9684	11439	9976	11146	10269	10854
10/11/2018	9769	13268	10352	12685	10936	12102
17/11/2018	10034	11929	10350	11613	10665	11297
24/11/2018	10388	12214	10692	11910	10996	11605
01/12/2018	10278	12086	10579	11785	10880	11483
08/12/2018	11356	13284	11677	12962	11998	12641
15/12/2018	11652	13568	11971	13249	12291	12929
22/12/2018	10532	12670	10888	12314	11245	11957
29/12/2018	10502	13045	10926	12621	11349	12197

Tabela 18 - Intervalo de Confiança Semanal Nordeste

FSO	INTERVALO MÍNIMO 3DP	INTERVALO MÁXIMO 3DP	INTERVALO MÍNIMO 2DP	INTERVALO MÁXIMO 2DP	INTERVALO MÍNIMO 1DP	INTERVALO MÁXIMO 1DP
31/12/2016	9689	10683	9855	10517	10021	10352
07/01/2017	9728	10638	9879	10486	10031	10335
14/01/2017	9742	10709	9903	10548	10064	10386
21/01/2017	9830	10747	9983	10594	10136	10442
28/01/2017	9847	10778	10002	10623	10157	10468
04/02/2017	9922	10958	10094	10786	10267	10613
11/02/2017	9870	10925	10046	10749	10221	10573
18/02/2017	9973	10907	10129	10752	10285	10596
25/02/2017	9239	10378	9429	10188	9619	9999
04/03/2017	9802	10786	9966	10622	10130	10458
11/03/2017	9929	10869	10086	10712	10242	10556
18/03/2017	10282	11227	10440	11069	10597	10912
25/03/2017	10298	11360	10475	11183	10652	11006
01/04/2017	10196	11134	10352	10978	10509	10822
08/04/2017	9659	10766	9844	10582	10028	10397
15/04/2017	9278	10490	9480	10288	9682	10086
22/04/2017	9956	10916	10116	10756	10276	10596
29/04/2017	9783	11269	10031	11022	10279	10774
06/05/2017	9520	10470	9678	10311	9837	10153
13/05/2017	9788	10732	9945	10575	10103	10417
20/05/2017	9702	10740	9875	10567	10048	10394
27/05/2017	9519	10627	9704	10442	9888	10258
03/06/2017	9571	10593	9742	10423	9912	10253
10/06/2017	8839	10087	9047	9879	9255	9671
17/06/2017	9295	10390	9478	10207	9660	10025
24/06/2017	9097	10223	9285	10035	9472	9848
01/07/2017	8738	9781	8912	9607	9086	9433
08/07/2017	8884	9894	9052	9726	9221	9557
15/07/2017	8748	9731	8912	9567	9076	9404
22/07/2017	8456	9823	8684	9595	8911	9367
29/07/2017	8629	9748	8816	9561	9002	9375
05/08/2017	8927	9951	9098	9780	9268	9610
12/08/2017	9216	10233	9386	10063	9555	9894
19/08/2017	9276	10366	9458	10184	9639	10003
26/08/2017	9241	10222	9405	10058	9568	9895
02/09/2017	8496	9832	8718	9609	8941	9387
09/09/2017	9197	10147	9356	9989	9514	9831
16/09/2017	9161	10100	9317	9944	9474	9787
23/09/2017	9208	10106	9358	9956	9507	9807
30/09/2017	9419	10436	9588	10266	9758	10097
07/10/2017	9226	10573	9451	10349	9675	10124

14/10/2017	9602	10683	9782	10503	9962	10323
21/10/2017	9662	10651	9827	10486	9992	10322
28/10/2017	9380	10459	9560	10279	9740	10099
04/11/2017	9944	10973	10116	10802	10287	10630
11/11/2017	9898	11362	10142	11118	10386	10874
18/11/2017	9974	10916	10131	10759	10288	10602
25/11/2017	10012	10938	10166	10784	10320	10629
02/12/2017	10293	11326	10465	11154	10637	10982
09/12/2017	10276	11195	10429	11041	10582	10888
16/12/2017	9969	11007	10142	10834	10315	10661
23/12/2017	9571	10679	9756	10495	9941	10310
30/12/2017	9558	11376	9861	11073	10164	10770
06/01/2018	9924	10900	10087	10738	10249	10575
13/01/2018	10206	11096	10354	10948	10503	10799
20/01/2018	9751	10770	9920	10600	10090	10430
27/01/2018	9886	10856	10048	10694	10210	10533
03/02/2018	10109	11067	10269	10907	10429	10748
10/02/2018	9117	10322	9318	10121	9519	9920
17/02/2018	9667	10601	9823	10445	9978	10290
24/02/2018	9783	10757	9945	10595	10108	10432
03/03/2018	10336	11396	10513	11220	10690	11043
10/03/2018	10262	11461	10462	11262	10662	11062
17/03/2018	10181	11379	10380	11179	10580	10979
24/03/2018	9388	10508	9575	10321	9762	10135
31/03/2018	9695	10788	9878	10606	10060	10424
07/04/2018	9761	10832	9939	10653	10118	10475
14/04/2018	9975	11008	10147	10836	10319	10664
21/04/2018	9677	10879	9878	10679	10078	10479
28/04/2018	9210	10637	9448	10399	9686	10161
05/05/2018	9674	10657	9838	10493	10002	10329
12/05/2018	9886	10884	10052	10718	10219	10551
19/05/2018	9904	10845	10061	10688	10217	10531
26/05/2018	9543	10550	9710	10382	9878	10214
02/06/2018	9257	10291	9430	10119	9602	9947
09/06/2018	9408	10338	9563	10183	9718	10028
16/06/2018	9476	10522	9650	10348	9825	10173
23/06/2018	9271	10242	9433	10080	9595	9918
30/06/2018	9069	10054	9233	9890	9397	9726
07/07/2018	8954	9899	9111	9741	9269	9584
14/07/2018	9189	10204	9358	10035	9527	9866
21/07/2018	9245	10229	9409	10065	9573	9901
28/07/2018	9069	10028	9228	9868	9388	9708
04/08/2018	9250	10196	9407	10038	9565	9881
11/08/2018	9535	10544	9703	10376	9871	10208
18/08/2018	9657	10659	9824	10492	9991	10325

25/08/2018	9520	10501	9683	10338	9847	10174
01/09/2018	9257	10402	9448	10212	9639	10021
08/09/2018	9493	10455	9654	10295	9814	10134
15/09/2018	9876	10899	10047	10728	10217	10558
22/09/2018	9900	10883	10064	10719	10228	10556
29/09/2018	9871	10866	10037	10700	10203	10534
06/10/2018	9987	11320	10209	11098	10431	10876
13/10/2018	10128	11121	10293	10955	10459	10790
20/10/2018	10278	11219	10435	11062	10592	10906
27/10/2018	10051	11254	10251	11053	10452	10853
03/11/2018	10343	11376	10515	11204	10687	11032
10/11/2018	9312	11029	9598	10743	9884	10457
17/11/2018	10329	11478	10520	11287	10712	11095
24/11/2018	10122	11217	10305	11035	10487	10852
01/12/2018	10168	11310	10358	11120	10549	10929
08/12/2018	10206	11251	10380	11077	10554	10903
15/12/2018	10174	11172	10340	11006	10507	10839
22/12/2018	9539	10629	9721	10447	9902	10266
29/12/2018	9668	10790	9855	10603	10042	10416

Tabela 19 - Intervalo de Confiança Semanal Norte

FSO	INTERVALO MÍNIMO 3DP	INTERVALO MÁXIMO 3DP	INTERVALO MÍNIMO 2DP	INTERVALO MÁXIMO 2DP	INTERVALO MÍNIMO 1DP	INTERVALO MÁXIMO 1DP
31/12/2016	4868	5478	4970	5376	5071	5275
07/01/2017	4826	5381	4919	5289	5011	5196
14/01/2017	4803	5337	4892	5248	4981	5159
21/01/2017	4872	5424	4964	5332	5056	5240
28/01/2017	4801	5372	4896	5277	4991	5181
04/02/2017	4940	5473	5029	5384	5118	5295
11/02/2017	4908	5463	5001	5370	5093	5278
18/02/2017	4962	5530	5056	5435	5151	5340
25/02/2017	4761	5396	4867	5290	4973	5184
04/03/2017	4802	5564	4929	5437	5056	5310
11/03/2017	4887	5498	4988	5396	5090	5294
18/03/2017	5103	5699	5202	5599	5302	5500
25/03/2017	5144	5825	5258	5712	5371	5598
01/04/2017	5194	5793	5294	5693	5394	5593
08/04/2017	4896	5540	5003	5433	5111	5325
15/04/2017	4909	5496	5007	5398	5105	5300
22/04/2017	5051	5590	5141	5500	5231	5410
29/04/2017	4911	5538	5016	5434	5120	5329
06/05/2017	5083	5610	5171	5522	5258	5434
13/05/2017	5152	5722	5247	5627	5342	5532

20/05/2017	5270	5978	5388	5860	5506	5742
27/05/2017	5188	5830	5295	5723	5402	5616
03/06/2017	5220	5849	5325	5744	5430	5639
10/06/2017	5059	5750	5174	5635	5290	5520
17/06/2017	5120	5718	5220	5618	5320	5519
24/06/2017	5108	5675	5203	5580	5297	5486
01/07/2017	5089	5643	5181	5550	5273	5458
08/07/2017	5053	5588	5142	5499	5231	5410
15/07/2017	5092	5656	5186	5562	5280	5468
22/07/2017	5230	5792	5324	5698	5417	5605
29/07/2017	5165	5734	5260	5640	5355	5545
05/08/2017	5101	5644	5191	5553	5282	5463
12/08/2017	5155	5726	5250	5631	5345	5536
19/08/2017	5357	5870	5443	5785	5528	5699
26/08/2017	5373	5866	5455	5784	5538	5702
02/09/2017	5199	5822	5303	5718	5407	5614
09/09/2017	5324	5860	5414	5771	5503	5681
16/09/2017	5137	5734	5237	5634	5336	5535
23/09/2017	5154	5752	5254	5653	5353	5553
30/09/2017	5356	5949	5455	5850	5554	5751
07/10/2017	5166	5805	5272	5699	5379	5592
14/10/2017	5204	5803	5304	5703	5404	5603
21/10/2017	5238	5815	5334	5719	5431	5623
28/10/2017	5247	5834	5345	5736	5443	5638
04/11/2017	5354	5915	5447	5821	5541	5728
11/11/2017	4863	5702	5003	5562	5143	5422
18/11/2017	5215	5799	5313	5702	5410	5605
25/11/2017	5290	5815	5378	5728	5465	5640
02/12/2017	5421	5946	5509	5859	5596	5771
09/12/2017	5182	5721	5272	5631	5361	5541
16/12/2017	4978	5555	5074	5459	5170	5362
23/12/2017	4842	5469	4947	5364	5051	5260
30/12/2017	4964	5565	5064	5465	5164	5365
06/01/2018	4833	5426	4932	5327	5031	5229
13/01/2018	4934	5494	5028	5401	5121	5307
20/01/2018	4938	5484	5029	5393	5120	5302
27/01/2018	4960	5506	5051	5415	5142	5324
03/02/2018	5004	5597	5103	5498	5202	5399
10/02/2018	4828	5457	4933	5353	5038	5248
17/02/2018	4957	5543	5054	5445	5152	5348
24/02/2018	5122	5696	5218	5601	5314	5505
03/03/2018	5233	5866	5339	5761	5444	5655
10/03/2018	5247	5842	5346	5743	5445	5644
17/03/2018	5371	6001	5476	5896	5581	5791
24/03/2018	4998	5726	5120	5605	5241	5483

31/03/2018	5093	5680	5191	5582	5288	5484
07/04/2018	5018	5629	5120	5527	5222	5426
14/04/2018	5233	5824	5331	5725	5430	5627
21/04/2018	5168	5794	5272	5690	5377	5585
28/04/2018	4836	5494	4946	5385	5056	5275
05/05/2018	4980	5574	5079	5475	5178	5376
12/05/2018	5065	5661	5164	5562	5264	5462
19/05/2018	5100	5684	5197	5587	5295	5490
26/05/2018	4879	5522	4986	5415	5093	5308
02/06/2018	4897	5512	5000	5410	5102	5307
09/06/2018	5015	5623	5117	5521	5218	5420
16/06/2018	4943	5536	5042	5437	5141	5338
23/06/2018	4983	5578	5082	5479	5182	5380
30/06/2018	4943	5545	5043	5444	5143	5344
07/07/2018	4934	5512	5030	5416	5127	5319
14/07/2018	4903	5496	5002	5397	5101	5299
21/07/2018	4906	5488	5003	5391	5100	5294
28/07/2018	4855	5434	4951	5338	5048	5241
04/08/2018	4895	5475	4992	5378	5089	5282
11/08/2018	4971	5546	5067	5451	5163	5355
18/08/2018	4963	5527	5057	5433	5151	5339
25/08/2018	4995	5563	5089	5468	5184	5373
01/09/2018	4906	5537	5011	5432	5116	5327
08/09/2018	4933	5513	5029	5416	5126	5319
15/09/2018	5136	5714	5232	5618	5329	5521
22/09/2018	5149	5727	5245	5631	5342	5534
29/09/2018	5297	5872	5393	5777	5489	5681
06/10/2018	5113	5778	5224	5667	5335	5556
13/10/2018	5140	5718	5236	5622	5333	5525
20/10/2018	5211	5795	5309	5697	5406	5600
27/10/2018	5005	5640	5110	5534	5216	5428
03/11/2018	5070	5688	5173	5585	5276	5482
10/11/2018	5018	5669	5126	5560	5235	5452
17/11/2018	5068	5664	5168	5564	5267	5465
24/11/2018	4889	5488	4989	5388	5089	5288
01/12/2018	4615	5249	4721	5143	4826	5038
08/12/2018	4771	5442	4883	5330	4994	5218
15/12/2018	4839	5464	4943	5360	5047	5255
22/12/2018	4776	5426	4884	5317	4992	5209
29/12/2018	4815	5533	4934	5413	5054	5293

Tabela 20 - Intervalo de Confiança Semanal SIN

FSO	INTERVALO MÍNIMO 3DP	INTERVALO MÁXIMO 3DP	INTERVALO MÍNIMO 2DP	INTERVALO MÁXIMO 2DP	INTERVALO MÍNIMO 1DP	INTERVALO MÁXIMO 1DP
31/12/2016	63986	71831	65294	70523	66601	69216
07/01/2017	62454	69904	63695	68662	64937	67421
14/01/2017	61841	69556	63127	68270	64413	66984
21/01/2017	62086	69472	63317	68241	64548	67010
28/01/2017	62476	70242	63771	68948	65065	67654
04/02/2017	63479	71143	64756	69866	66034	68588
11/02/2017	65339	72805	66583	71561	67828	70316
18/02/2017	65619	73444	66923	72140	68228	70836
25/02/2017	60435	69277	61908	67803	63382	66329
04/03/2017	63168	71046	64481	69733	65794	68420
11/03/2017	63149	71014	64460	69703	65771	68393
18/03/2017	60134	68301	61495	66940	62856	65579
25/03/2017	60614	68730	61966	67377	63319	66024
01/04/2017	60986	68398	62222	67163	63457	65928
08/04/2017	57808	66038	59180	64667	60552	63295
15/04/2017	55373	64730	56932	63171	58492	61611
22/04/2017	56620	64362	57910	63072	59201	61781
29/04/2017	53569	68570	56069	66070	58569	63570
06/05/2017	55815	63248	57054	62009	58293	60770
13/05/2017	56359	63402	57532	62228	58706	61054
20/05/2017	56318	63846	57573	62591	58827	61336
27/05/2017	55825	63378	57084	62119	58342	60860
03/06/2017	56358	63855	57608	62605	58857	61356
10/06/2017	51230	60804	52826	59208	54421	57612
17/06/2017	54608	62383	55904	61087	57200	59791
24/06/2017	55185	62741	56445	61482	57704	60223
01/07/2017	54102	61785	55382	60505	56663	59224
08/07/2017	54415	62461	55756	61120	57097	59779
15/07/2017	54373	61840	55618	60595	56862	59351
22/07/2017	54757	62356	56023	61090	57290	59823
29/07/2017	54619	62169	55878	60910	57136	59652
05/08/2017	54637	61939	55854	60722	57071	59505
12/08/2017	54736	62406	56015	61127	57293	59849
19/08/2017	55141	62371	56346	61166	57551	59961
26/08/2017	56241	63438	57441	62238	58640	61039
02/09/2017	52685	61315	54123	59876	55561	58438
09/09/2017	57763	65338	59026	64076	60288	62813
16/09/2017	58041	65629	59305	64365	60570	63100
23/09/2017	58065	65223	59258	64030	60451	62837
30/09/2017	58659	66128	59904	64883	61149	63638
07/10/2017	53407	71947	56497	68857	59587	65767

14/10/2017	59346	66771	60584	65534	61821	64296
21/10/2017	58744	66002	59953	64792	61163	63583
28/10/2017	55850	63852	57184	62519	58517	61185
04/11/2017	58265	65696	59503	64458	60742	63219
11/11/2017	55134	73088	58126	70095	61118	67103
18/11/2017	59288	66533	60495	65325	61703	64118
25/11/2017	59281	66387	60465	65203	61649	64018
02/12/2017	60197	67426	61402	66221	62607	65017
09/12/2017	60675	67849	61870	66653	63066	65457
16/12/2017	61704	69128	62941	67891	64179	66653
23/12/2017	59289	68115	60760	66644	62231	65173
30/12/2017	56099	74440	59156	71383	62213	68326
06/01/2018	59055	66419	60283	65192	61510	63965
13/01/2018	62534	70042	63786	68791	65037	67539
20/01/2018	63612	71019	64847	69785	66081	68550
27/01/2018	61621	69142	62874	67888	64128	66635
03/02/2018	61761	69234	63007	67989	64252	66743
10/02/2018	58497	67319	59967	65849	61438	64379
17/02/2018	61523	69452	62844	68131	64166	66809
24/02/2018	62870	70250	64100	69020	65330	67790
03/03/2018	63566	71000	64805	69761	66044	68522
10/03/2018	65115	73100	66446	71769	67776	70438
17/03/2018	64558	72352	65857	71053	67156	69754
24/03/2018	59624	69147	61211	67560	62798	65973
31/03/2018	61726	69414	63008	68133	64289	66851
07/04/2018	62527	69993	63771	68749	65016	67505
14/04/2018	60058	67981	61378	66660	62699	65340
21/04/2018	60166	70402	61872	68696	63578	66990
28/04/2018	56774	69071	58824	67022	60873	64972
05/05/2018	58814	65907	59996	64725	61179	63543
12/05/2018	59108	66341	60313	65136	61519	63930
19/05/2018	56683	64710	58021	63372	59359	62034
26/05/2018	54375	62759	55772	61362	57169	59964
02/06/2018	52677	61612	54166	60123	55655	58634
09/06/2018	54977	63075	56326	61725	57676	60375
16/06/2018	54232	62313	55579	60966	56926	59619
23/06/2018	55431	63264	56737	61959	58042	60653
30/06/2018	54892	62702	56194	61400	57496	60099
07/07/2018	54721	62320	55988	61054	57254	59787
14/07/2018	56749	63661	57901	62509	59053	61357
21/07/2018	56628	64362	57917	63073	59206	61784
28/07/2018	55788	63241	57030	61999	58272	60756
04/08/2018	55482	63073	56747	61808	58012	60542
11/08/2018	56117	63582	57361	62338	58605	61094
18/08/2018	57287	64730	58527	63489	59768	62249

25/08/2018	56835	64081	58042	62873	59250	61665
01/09/2018	52381	62294	54033	60642	55685	58990
08/09/2018	55886	63229	57110	62005	58334	60781
15/09/2018	57608	65016	58842	63781	60077	62547
22/09/2018	59038	66519	60285	65273	61532	64026
29/09/2018	59639	66992	60864	65767	62090	64541
06/10/2018	55857	66626	57652	64831	59447	63036
13/10/2018	59152	66606	60394	65363	61636	64121
20/10/2018	58839	66194	60065	64969	61291	63743
27/10/2018	56173	66208	57846	64536	59518	62863
03/11/2018	58067	65539	59313	64294	60558	63049
10/11/2018	57417	69503	59432	67489	61446	65475
17/11/2018	59831	67512	61111	66232	62391	64952
24/11/2018	60304	67777	61550	66532	62795	65286
01/12/2018	59887	67453	61148	66192	62409	64931
08/12/2018	61830	69664	63136	68359	64442	67053
15/12/2018	65522	73281	66815	71988	68109	70695
22/12/2018	60174	68365	61540	67000	62905	65635
29/12/2018	56207	71404	58740	68871	61273	66339

C. TABELA DE ERROS SEMANAIS**Tabela 21 - Tabela de Erros Semanais**

Semana Operativa	Mape SIN	Mape SECO	Mape S	Mape NE	Mape N
31/12/2016	6,43	10,30	2,82	0,08	0,49
07/01/2017	9,82	9,12	16,70	7,75	5,92
14/01/2017	5,12	7,04	0,91	5,84	0,46
21/01/2017	5,91	6,81	9,37	0,44	2,62
28/01/2017	3,48	4,90	3,72	1,12	3,38
04/02/2017	5,01	4,89	14,48	0,51	1,85
11/02/2017	3,48	4,75	0,29	2,99	2,95
18/02/2017	3,59	4,62	3,18	2,05	0,66
25/02/2017	4,51	4,57	6,10	4,93	0,47
04/03/2017	4,60	4,46	10,07	1,00	1,99
11/03/2017	3,03	3,75	4,16	1,50	3,92
18/03/2017	3,05	3,59	3,29	0,69	3,31
25/03/2017	2,80	3,39	3,44	1,23	0,61
01/04/2017	2,77	3,34	0,37	1,94	5,68
08/04/2017	2,42	3,33	3,92	0,98	1,19
15/04/2017	2,47	3,25	0,20	3,39	0,44
22/04/2017	1,24	3,16	2,38	1,74	0,11
29/04/2017	1,17	3,07	2,26	0,95	0,35
06/05/2017	3,43	3,03	2,88	4,62	5,27
13/05/2017	2,05	2,85	1,85	3,06	2,20
20/05/2017	2,28	2,82	1,74	2,48	0,71
27/05/2017	2,11	2,81	5,42	1,39	2,78
03/06/2017	2,56	2,75	2,29	1,87	3,17
10/06/2017	1,60	2,63	2,00	2,22	1,33
17/06/2017	1,39	2,52	2,59	3,37	0,52
24/06/2017	2,89	2,50	6,46	3,84	3,45
01/07/2017	1,60	2,49	2,98	1,00	1,94
08/07/2017	1,13	2,35	1,18	2,52	0,11
15/07/2017	1,99	2,31	2,16	0,04	3,19
22/07/2017	1,90	2,31	1,74	1,34	0,62
29/07/2017	2,67	2,29	1,95	5,52	1,25
05/08/2017	1,86	2,14	6,83	1,82	2,52
12/08/2017	0,52	2,10	2,98	0,80	0,10
19/08/2017	2,21	2,06	4,36	1,11	0,18
26/08/2017	2,07	2,03	0,00	4,92	1,01
02/09/2017	3,13	2,02	8,91	0,79	2,44
09/09/2017	0,10	1,99	3,15	2,67	0,18
16/09/2017	0,79	1,97	0,19	1,45	0,64
23/09/2017	3,53	1,97	9,19	5,22	0,65

30/09/2017	1,88	1,97	3,48	0,47	2,14
07/10/2017	0,07	1,95	6,56	1,14	0,41
14/10/2017	2,98	1,95	5,43	3,69	3,67
21/10/2017	1,57	1,92	2,50	0,26	0,43
28/10/2017	2,08	1,89	3,69	1,22	2,03
04/11/2017	2,33	1,87	1,31	3,57	4,60
11/11/2017	2,64	1,85	5,22	2,06	3,69
18/11/2017	2,49	1,83	5,77	1,29	2,37
25/11/2017	2,34	1,82	7,25	0,37	0,15
02/12/2017	2,16	1,79	4,86	2,56	1,48
09/12/2017	1,57	1,77	2,37	0,19	1,33
16/12/2017	0,14	1,75	4,06	2,86	0,99
23/12/2017	1,58	1,69	3,25	0,04	0,72
30/12/2017	1,26	1,66	0,24	0,77	1,56
06/01/2018	0,60	1,65	1,94	1,49	2,95
13/01/2018	0,01	1,65	1,59	3,82	0,96
20/01/2018	1,55	1,54	1,78	0,78	2,61
27/01/2018	1,37	1,53	1,98	1,56	1,32
03/02/2018	1,60	1,52	2,14	0,28	4,17
10/02/2018	2,90	1,52	9,66	1,06	1,36
17/02/2018	0,62	1,49	2,53	1,28	0,23
24/02/2018	0,59	1,46	1,36	0,36	1,02
03/03/2018	1,37	1,42	4,99	0,51	3,11
10/03/2018	0,20	1,40	2,47	0,34	1,21
17/03/2018	1,24	1,35	3,69	1,17	0,04
24/03/2018	0,81	1,32	1,50	0,49	2,71
31/03/2018	0,88	1,27	1,42	1,52	1,44
07/04/2018	1,37	1,27	0,86	1,38	2,92
14/04/2018	0,51	1,25	1,71	0,34	1,65
21/04/2018	0,38	1,25	0,15	3,20	2,38
28/04/2018	2,35	1,16	4,19	3,47	4,23
05/05/2018	1,67	1,13	3,27	1,69	2,14
12/05/2018	1,57	1,10	0,67	4,28	1,39
19/05/2018	1,17	1,07	3,05	0,22	0,03
26/05/2018	0,01	1,06	0,98	1,55	2,81
02/06/2018	1,95	1,05	6,42	0,71	1,16
09/06/2018	2,54	1,05	6,47	5,07	0,03
16/06/2018	2,02	1,04	6,13	1,75	1,15
23/06/2018	1,33	0,98	3,44	0,50	0,93
30/06/2018	2,03	0,98	7,03	0,24	1,36
07/07/2018	0,62	0,93	0,58	0,41	0,67
14/07/2018	0,19	0,79	3,36	0,67	0,94
21/07/2018	0,79	0,76	0,97	2,04	1,71
28/07/2018	1,23	0,74	0,60	3,20	4,71
04/08/2018	1,74	0,74	3,81	2,95	1,71

11/08/2018	0,62	0,72	3,42	4,34	3,83
18/08/2018	1,64	0,72	6,22	2,53	3,47
25/08/2018	1,48	0,70	4,82	0,26	2,04
01/09/2018	0,43	0,60	0,66	1,02	1,61
08/09/2018	1,84	0,59	5,20	2,54	1,75
15/09/2018	1,64	0,54	3,69	3,44	1,55
22/09/2018	0,30	0,50	0,04	0,75	1,33
29/09/2018	1,74	0,47	7,42	0,52	0,43
06/10/2018	0,83	0,47	1,97	1,33	0,00
13/10/2018	0,52	0,43	0,89	2,85	0,68
20/10/2018	1,51	0,35	4,51	1,06	3,91
27/10/2018	0,35	0,33	3,27	0,89	1,38
03/11/2018	1,21	0,31	3,71	3,59	1,84
10/11/2018	1,01	0,29	3,62	3,16	0,11
17/11/2018	0,42	0,28	1,37	0,10	0,13
24/11/2018	2,05	0,25	6,53	4,12	1,27
01/12/2018	0,77	0,20	2,87	2,81	1,08
08/12/2018	0,06	0,12	0,27	0,11	0,83
15/12/2018	0,35	0,09	4,43	1,68	0,82
22/12/2018	0,68	0,06	2,26	1,74	0,60
29/12/2018	1,38	0,03	4,52	3,77	1,18
Média	1,87	2,00	3,59	1,93	1,72

D. TABELA DE ERROS E AMPLITUDE DOS INTERVALOS DE CONFIANÇA
DIÁRIOS

Tabela 22 - Erros Diários

Dia	Mape 168 SE	Mape 48 SE	PICP 168 SE	PICP 48 SE	Número de Clusters	Intervalo de confiança médio
01/01/2018	3,2	2,9	76%	96%	9	4370
02/01/2018	4,2	2,1	63%	96%	8	6696
03/01/2018	2,6	1,4	82%	96%	10	4037
04/01/2018	2,5	2,3	79%	71%	9	3779
05/01/2018	3,8	2,6	64%	67%	11	2812
06/01/2018	3,2	4,0	71%	58%	12	3012
07/01/2018	5,7	20,1	66%	21%	10	2854
08/01/2018	2,2	1,2	84%	100%	8	4323
09/01/2018	2,1	2,6	82%	71%	9	3771
10/01/2018	2,3	2,4	80%	79%	11	4029
11/01/2018	3,1	3,2	75%	67%	11	3657
12/01/2018	2,0	1,6	88%	96%	11	3798
13/01/2018	2,8	2,4	69%	79%	10	3712
14/01/2018	3,2	2,6	64%	88%	9	4463
15/01/2018	4,1	2,3	49%	92%	10	3071
16/01/2018	4,0	2,4	46%	71%	11	3486

17/01/2018	5,5	3,8	45%	38%	12	3653
18/01/2018	6,3	5,3	33%	33%	10	3225
19/01/2018	4,7	5,0	38%	38%	9	3172
20/01/2018	4,1	3,6	51%	83%	9	3875
21/01/2018	4,6	7,5	49%	25%	11	3304
22/01/2018	4,4	6,9	51%	25%	9	4784
23/01/2018	3,6	2,2	66%	71%	10	2902
24/01/2018	4,1	1,7	67%	100%	9	2620
25/01/2018	5,3	3,0	46%	79%	10	2583
26/01/2018	4,7	4,1	47%	42%	10	4258
27/01/2018	6,1	6,2	46%	50%	10	3611
28/01/2018	3,8	1,6	64%	100%	11	3818
29/01/2018	4,7	4,3	52%	46%	9	2977
30/01/2018	4,7	8,6	52%	0%	9	3570
31/01/2018	4,1	4,5	57%	54%	10	3460
01/02/2018	3,3	2,4	72%	88%	9	4244
02/02/2018	3,3	2,5	68%	71%	9	3325
03/02/2018	3,2	2,9	71%	79%	9	2992
04/02/2018	3,8	4,9	60%	58%	10	3385
05/02/2018	3,7	3,5	60%	71%	9	3255

06/02/2018	4,4	1,6	67%	96%	9	2906
07/02/2018	4,4	2,0	67%	83%	9	4187
08/02/2018	5,2	3,9	58%	54%	10	3303
09/02/2018	6,4	3,0	45%	50%	12	4377
10/02/2018	8,4	4,2	38%	54%	10	4785
11/02/2018	7,4	13,9	51%	13%	10	3524
12/02/2018	6,2	10,1	39%	8%	10	3607
13/02/2018	5,3	1,9	59%	75%	11	7991
14/02/2018	4,4	2,1	61%	88%	10	4470
15/02/2018	4,8	1,8	62%	100%	11	3336
16/02/2018	3,9	2,7	64%	75%	9	3524
17/02/2018	4,2	4,0	63%	58%	9	4348
18/02/2018	4,3	8,0	57%	29%	11	3912
19/02/2018	2,9	3,0	71%	71%	12	4075
20/02/2018	3,2	1,4	71%	88%	9	2961
21/02/2018	4,4	2,4	54%	71%	9	2817
22/02/2018	3,7	2,2	67%	92%	11	3528
23/02/2018	4,8	4,7	60%	88%	9	3783
24/02/2018	4,2	4,4	53%	38%	9	2902
25/02/2018	3,5	4,3	68%	79%	10	3056

26/02/2018	4,5	6,7	53%	38%	9	3157
27/02/2018	5,0	5,4	44%	25%	11	3191
28/02/2018	3,1	4,3	65%	21%	9	3175
01/03/2018	3,2	1,9	72%	83%	9	2942
02/03/2018	4,5	3,4	56%	54%	10	2654
03/03/2018	4,6	2,5	52%	88%	10	2901
04/03/2018	3,2	2,7	67%	75%	9	5066
05/03/2018	3,3	3,7	60%	54%	9	3154
06/03/2018	3,8	1,7	63%	83%	12	3143
07/03/2018	2,8	2,0	73%	83%	10	3563
08/03/2018	3,5	2,3	64%	79%	10	2854
09/03/2018	4,7	2,1	56%	79%	9	2373
10/03/2018	2,8	1,8	74%	96%	10	4643
11/03/2018	2,5	2,9	77%	83%	10	2616
12/03/2018	3,6	4,1	64%	58%	11	2856
13/03/2018	3,5	4,1	64%	42%	9	3647
14/03/2018	3,3	1,7	64%	83%	9	3384
15/03/2018	3,6	2,2	71%	67%	11	5388
16/03/2018	2,4	2,6	79%	83%	8	5738
17/03/2018	3,8	2,9	63%	71%	10	3269

18/03/2018	3,6	9,1	69%	33%	10	2998
19/03/2018	4,1	3,8	61%	58%	12	3513
20/03/2018	3,4	3,0	65%	58%	12	3571
21/03/2018	3,4	4,4	67%	29%	9	3216
22/03/2018	3,0	2,1	76%	79%	9	6921
23/03/2018	2,8	2,3	77%	79%	9	3489
24/03/2018	4,0	3,7	71%	63%	9	2411
25/03/2018	4,2	2,7	67%	88%	10	3213
26/03/2018	5,3	4,3	58%	42%	9	3073
27/03/2018	5,0	2,5	67%	71%	10	2981
28/03/2018	5,3	1,6	57%	92%	10	4770
29/03/2018	7,8	23,3	43%	8%	13	4088
30/03/2018	4,0	4,0	64%	63%	10	2877
31/03/2018	3,2	3,7	70%	83%	9	3240
01/04/2018	3,3	1,5	61%	96%	9	3229
02/04/2018	2,5	2,6	79%	71%	8	2379
03/04/2018	2,9	2,3	77%	75%	10	5492
04/04/2018	3,2	3,4	59%	46%	10	3148
05/04/2018	2,8	3,8	72%	54%	9	2757
06/04/2018	2,0	1,8	86%	83%	10	3087

07/04/2018	2,5	2,0	74%	71%	11	2952
08/04/2018	3,4	1,3	76%	96%	9	3146
09/04/2018	4,2	3,3	63%	50%	10	4355
10/04/2018	4,5	2,6	60%	71%	10	3606
11/04/2018	4,2	2,1	73%	92%	11	4107
12/04/2018	4,2	2,7	58%	83%	10	3082
13/04/2018	5,0	2,6	57%	88%	10	3964
14/04/2018	5,4	10,5	51%	21%	10	2483
15/04/2018	6,1	8,2	61%	67%	10	3144
16/04/2018	4,4	4,0	67%	75%	11	2823
17/04/2018	5,6	4,8	43%	17%	9	2904
18/04/2018	5,5	3,0	59%	67%	10	3622
19/04/2018	5,0	2,5	58%	88%	11	3200
20/04/2018	3,8	3,2	67%	100%	11	2347
21/04/2018	3,2	3,9	83%	46%	12	2476
22/04/2018	3,5	4,6	54%	33%	9	4532
23/04/2018	2,9	2,2	71%	96%	12	2906
24/04/2018	3,2	2,1	69%	79%	13	2694
25/04/2018	4,7	2,3	83%	83%	9	3488
26/04/2018	4,1	1,5	77%	100%	10	3629

27/04/2018	4,8	3,4	73%	83%	9	2923
28/04/2018	5,2	2,6	57%	75%	9	7392
29/04/2018	6,5	5,0	50%	63%	10	4258
30/04/2018	6,8	4,7	50%	79%	9	2789
01/05/2018	5,6	6,6	51%	38%	9	3955
02/05/2018	5,6	10,4	60%	21%	11	5276
03/05/2018	2,7	3,0	85%	79%	12	2803
04/05/2018	1,8	1,2	93%	100%	9	7468
05/05/2018	2,3	1,3	83%	96%	9	2980
06/05/2018	2,9	2,0	79%	92%	9	2423
07/05/2018	2,1	1,6	83%	96%	12	3790
08/05/2018	2,2	1,8	82%	79%	13	3703
09/05/2018	2,0	1,4	88%	100%	10	2750
10/05/2018	2,8	2,0	79%	83%	13	5005
11/05/2018	2,1	2,7	90%	88%	10	2845
12/05/2018	2,2	3,1	84%	79%	10	2842
13/05/2018	2,3	1,1	79%	92%	11	3675
14/05/2018	2,8	1,8	74%	88%	10	3806
15/05/2018	4,0	1,6	75%	96%	9	4223
16/05/2018	3,5	1,3	68%	100%	10	3507

17/05/2018	4,2	1,6	65%	92%	9	3179
18/05/2018	5,2	2,8	60%	96%	11	2803
19/05/2018	7,1	5,0	32%	38%	10	3479
20/05/2018	9,0	5,9	29%	63%	8	3232
21/05/2018	10,1	9,0	20%	67%	9	2835
22/05/2018	8,2	6,3	25%	38%	9	4957
23/05/2018	7,4	3,6	24%	46%	9	2802
24/05/2018	8,5	7,7	21%	4%	8	2743
25/05/2018	9,3	11,4	25%	4%	8	3338
26/05/2018	8,6	7,2	24%	17%	10	4157
27/05/2018	9,7	8,0	33%	29%	9	3730
28/05/2018	6,1	9,1	60%	50%	11	3656
29/05/2018	3,7	6,5	74%	21%	9	3259
30/05/2018	3,9	6,3	76%	79%	9	2788
31/05/2018	5,0	6,6	61%	58%	9	3096
01/06/2018	3,4	3,9	67%	79%	10	2910
02/06/2018	3,4	3,4	69%	63%	4	2610
03/06/2018	2,5	2,8	86%	71%	11	2733
04/06/2018	3,1	2,7	76%	83%	10	3494
05/06/2018	4,3	4,0	65%	58%	10	2828

06/06/2018	3,1	4,2	77%	79%	8	3041
07/06/2018	2,7	2,2	78%	92%	9	3739
08/06/2018	4,0	2,9	70%	100%	10	3606
09/06/2018	3,6	2,6	80%	88%	10	3273
10/06/2018	3,3	1,9	76%	92%	9	5969
11/06/2018	4,7	2,5	59%	92%	10	2994
12/06/2018	3,5	2,4	61%	88%	9	2905
13/06/2018	3,1	4,2	73%	54%	10	2847
14/06/2018	2,6	3,8	77%	63%	9	2596
15/06/2018	2,7	3,2	74%	71%	8	3116
16/06/2018	2,3	2,1	80%	88%	10	5297
17/06/2018	2,9	1,6	73%	100%	10	3719
18/06/2018	2,9	1,7	81%	96%	9	3152
19/06/2018	3,2	1,0	66%	96%	11	3434
20/06/2018	4,0	2,9	55%	63%	11	3399
21/06/2018	3,8	3,0	60%	92%	8	3242
22/06/2018	3,0	3,8	64%	67%	10	3833
23/06/2018	4,5	4,2	43%	67%	9	2660
24/06/2018	3,3	4,2	63%	33%	11	2552
25/06/2018	2,6	3,9	77%	58%	8	3113

26/06/2018	2,1	3,1	88%	50%	11	2536
27/06/2018	2,4	1,9	80%	83%	13	3560
28/06/2018	2,4	3,0	81%	75%	9	4884
29/06/2018	2,2	3,1	83%	71%	9	3826
30/06/2018	2,4	2,3	78%	88%	12	2834
01/07/2018	2,5	1,7	71%	92%	12	2827
02/07/2018	1,8	2,5	85%	71%	10	3080
03/07/2018	2,4	1,5	88%	92%	11	3740
04/07/2018	2,5	1,0	82%	100%	12	6383
05/07/2018	2,7	1,2	76%	100%	10	4437
06/07/2018	2,7	1,0	75%	100%	4	3742
07/07/2018	3,2	1,1	71%	96%	12	3159
08/07/2018	3,5	7,5	70%	17%	12	5304
09/07/2018	3,2	2,5	66%	83%	10	2934
10/07/2018	4,2	3,1	45%	46%	13	4315
11/07/2018	3,6	1,9	67%	92%	11	3201
12/07/2018	3,4	2,6	57%	79%	12	2582
13/07/2018	3,8	1,7	49%	92%	10	3103
14/07/2018	4,6	1,9	43%	71%	9	3418
15/07/2018	4,5	3,4	36%	42%	9	3084

16/07/2018	4,4	3,0	34%	71%	8	5383
17/07/2018	2,3	2,7	67%	42%	12	3117
18/07/2018	2,5	2,6	67%	50%	12	2718
19/07/2018	2,9	3,5	64%	33%	11	2978
20/07/2018	2,3	1,9	70%	75%	11	3622
21/07/2018	2,2	1,6	73%	88%	12	2661
22/07/2018	2,3	1,6	65%	92%	8	3652
23/07/2018	1,9	1,6	79%	92%	12	2945
24/07/2018	2,7	2,7	65%	63%	10	3664
25/07/2018	2,9	2,0	57%	71%	14	3129
26/07/2018	3,2	1,3	61%	83%	9	3401
27/07/2018	3,4	2,8	55%	88%	9	2944
28/07/2018	2,3	2,7	72%	63%	10	2872
29/07/2018	2,2	4,3	65%	29%	10	4469
30/07/2018	3,1	3,0	51%	46%	10	2838
31/07/2018	1,8	1,7	83%	88%	8	3321
01/08/2018	2,3	1,0	74%	100%	9	3740
02/08/2018	2,4	2,4	70%	71%	10	3305
03/08/2018	2,1	1,9	71%	67%	13	6790
04/08/2018	2,6	1,7	67%	79%	10	4085

05/08/2018	2,8	2,9	67%	83%	12	2627
06/08/2018	3,2	3,6	60%	46%	10	3851
07/08/2018	2,6	3,0	65%	42%	10	3519
08/08/2018	2,2	1,2	75%	92%	13	2488
09/08/2018	2,7	3,0	70%	67%	12	4322
10/08/2018	3,1	3,5	71%	71%	11	4390
11/08/2018	2,6	3,3	67%	71%	10	3167
12/08/2018	3,0	2,5	67%	75%	12	2952
13/08/2018	3,1	3,7	54%	50%	11	3167
14/08/2018	2,7	2,3	66%	54%	10	2426
15/08/2018	2,7	1,7	78%	83%	11	3714
16/08/2018	2,5	2,1	70%	79%	11	4119
17/08/2018	2,4	1,3	70%	88%	8	2889
18/08/2018	4,2	2,4	48%	71%	9	2895
19/08/2018	4,3	3,0	45%	42%	11	3006
20/08/2018	4,4	3,7	44%	54%	12	2636
21/08/2018	2,8	2,3	66%	58%	8	2782
22/08/2018	2,4	2,5	67%	50%	12	5336
23/08/2018	2,8	4,3	60%	38%	10	3657
24/08/2018	2,0	2,4	73%	71%	9	3551

25/08/2018	3,2	4,2	59%	54%	12	3349
26/08/2018	2,1	1,8	76%	83%	10	3206
27/08/2018	1,8	1,3	79%	100%	12	4087
28/08/2018	1,7	1,2	87%	92%	11	4438
29/08/2018	1,7	1,2	82%	88%	11	3530
30/08/2018	2,1	1,5	78%	83%	12	2723
31/08/2018	2,3	2,5	73%	79%	9	2752
01/09/2018	2,8	2,2	68%	71%	12	2744
02/09/2018	3,0	2,4	73%	71%	12	2479
03/09/2018	4,3	2,5	55%	67%	12	5720
04/09/2018	3,3	1,8	65%	75%	9	3337
05/09/2018	2,9	2,5	80%	75%	11	3268
06/09/2018	2,8	4,9	71%	50%	12	2998
07/09/2018	4,6	4,3	59%	46%	11	2802
08/09/2018	2,4	3,9	69%	58%	12	3589
09/09/2018	2,2	1,9	79%	83%	9	2934
10/09/2018	2,6	2,6	72%	75%	12	2943
11/09/2018	2,3	1,5	83%	100%	12	2848
12/09/2018	2,0	1,7	83%	75%	13	3194
13/09/2018	2,4	2,7	79%	88%	12	2758

14/09/2018	2,6	1,3	73%	100%	9	2682
15/09/2018	2,3	1,1	79%	96%	9	3145
16/09/2018	2,0	2,0	76%	83%	10	4544
17/09/2018	2,3	1,9	70%	75%	11	3014
18/09/2018	3,2	3,7	55%	38%	9	2768
19/09/2018	4,0	2,3	55%	71%	9	3268
20/09/2018	3,5	1,4	48%	71%	8	2772
21/09/2018	4,5	3,0	46%	54%	9	2871
22/09/2018	4,3	1,2	52%	83%	12	4934
23/09/2018	4,6	6,3	45%	46%	12	5115
24/09/2018	4,2	5,2	39%	33%	11	2732
25/09/2018	2,6	2,8	70%	54%	10	2658
26/09/2018	1,8	1,8	83%	75%	9	2538
27/09/2018	2,0	2,2	81%	100%	10	3169
28/09/2018	2,4	1,3	74%	100%	11	4893
29/09/2018	3,1	1,5	65%	88%	11	6155
30/09/2018	2,1	2,1	79%	67%	9	3558
01/10/2018	2,7	4,3	64%	29%	12	3688
02/10/2018	3,0	2,0	65%	63%	13	2424
03/10/2018	3,2	3,4	63%	50%	9	2821

04/10/2018	2,6	4,9	70%	29%	11	5269
05/10/2018	2,9	2,4	68%	83%	8	4780
06/10/2018	3,7	6,8	70%	67%	11	3045
07/10/2018	2,3	1,3	77%	100%	10	2624
08/10/2018	3,5	3,5	61%	33%	10	3856
09/10/2018	3,5	2,2	63%	67%	9	2588
10/10/2018	3,4	3,0	66%	63%	10	6602
11/10/2018	4,4	4,2	63%	96%	10	13041
12/10/2018	4,7	4,4	47%	58%	9	2907
13/10/2018	4,9	3,9	48%	46%	10	2822
14/10/2018	5,8	6,5	39%	33%	11	2832
15/10/2018	4,2	4,2	46%	38%	9	3398
16/10/2018	3,6	3,3	54%	33%	12	5750
17/10/2018	2,9	3,3	61%	50%	12	8701
18/10/2018	3,7	2,1	65%	96%	10	3482
19/10/2018	3,8	4,6	61%	33%	12	3323
20/10/2018	3,3	5,3	64%	33%	10	3514
21/10/2018	3,0	3,1	64%	75%	12	2964
22/10/2018	2,5	2,9	75%	63%	10	3300
23/10/2018	3,0	3,8	67%	46%	9	5500

24/10/2018	2,6	3,4	77%	54%	10	3492
25/10/2018	3,2	3,0	63%	71%	11	3069
26/10/2018	2,5	2,0	76%	79%	9	2732
27/10/2018	2,9	3,2	74%	54%	10	2183
28/10/2018	3,6	1,9	78%	92%	13	3043
29/10/2018	2,7	1,9	77%	88%	11	8535
30/10/2018	2,8	3,3	74%	63%	10	3643
31/10/2018	3,5	3,7	64%	46%	8	4388
01/11/2018	2,9	3,3	79%	79%	9	4107
02/11/2018	2,9	2,9	77%	75%	11	4487
03/11/2018	2,6	3,0	70%	54%	9	2673
04/11/2018	2,4	3,2	76%	54%	9	4353
05/11/2018	2,9	2,5	73%	75%	10	3641
06/11/2018	2,5	1,3	72%	88%	12	2976
07/11/2018	2,6	2,3	70%	75%	9	3050
08/11/2018	3,6	2,0	58%	79%	9	3279
09/11/2018	3,9	3,0	55%	71%	11	3336
10/11/2018	3,9	4,1	52%	33%	11	4413
11/11/2018	4,9	3,9	45%	38%	10	4246
12/11/2018	4,2	4,2	52%	29%	10	3013

13/11/2018	4,1	3,3	64%	33%	10	3955
14/11/2018	4,4	8,3	65%	63%	12	5602
15/11/2018	3,6	2,1	73%	83%	10	3241
16/11/2018	3,3	3,5	64%	54%	11	4258
17/11/2018	3,6	3,7	61%	71%	9	3072
18/11/2018	3,5	4,4	66%	42%	9	3227
19/11/2018	3,7	4,1	56%	42%	9	3125
20/11/2018	3,7	4,6	55%	46%	12	2556
21/11/2018	2,7	3,2	69%	54%	9	3857
22/11/2018	2,4	3,1	77%	71%	11	5116
23/11/2018	2,6	3,7	77%	58%	10	3247
24/11/2018	3,0	2,3	63%	71%	9	3345
25/11/2018	2,6	2,7	73%	75%	13	3215
26/11/2018	2,5	1,8	80%	83%	12	2909
27/11/2018	2,0	1,3	82%	100%	8	2583
28/11/2018	2,5	2,6	77%	71%	15	4491
29/11/2018	2,6	5,5	73%	25%	11	3206
30/11/2018	2,4	2,6	82%	67%	11	3083
01/12/2018	2,1	1,5	86%	100%	10	3672
02/12/2018	2,0	2,2	89%	100%	10	3786

03/12/2018	2,1	3,1	86%	79%	9	3210
04/12/2018	2,4	2,0	76%	75%	13	4148
05/12/2018	2,6	2,9	80%	71%	9	3963
06/12/2018	2,9	2,5	64%	71%	11	3488
07/12/2018	3,0	2,0	63%	75%	9	3544
08/12/2018	3,6	3,5	59%	75%	12	3629
09/12/2018	4,1	2,2	54%	79%	10	5283
10/12/2018	4,7	1,9	50%	92%	11	5123
11/12/2018	6,9	3,7	39%	38%	13	4016
12/12/2018	5,5	4,0	33%	38%	10	3002
13/12/2018	6,6	5,0	35%	25%	10	3789
14/12/2018	6,4	4,9	32%	42%	10	3534
15/12/2018	7,8	9,5	14%	0%	11	3918
16/12/2018	7,9	13,7	28%	17%	11	3386
17/12/2018	5,8	5,8	42%	33%	11	3112
18/12/2018	7,2	3,5	46%	54%	10	3368
19/12/2018	7,6	1,3	65%	100%	10	3724
20/12/2018	11,0	2,2	52%	83%	10	2974
21/12/2018	10,1	3,8	48%	96%	11	7587
22/12/2018	9,6	7,0	50%	42%	9	2704

23/12/2018	13,9	15,9	15%	13%	11	4343
24/12/2018	11,5	24,9	32%	42%	11	2792
25/12/2018	4,6	4,3	69%	63%	9	2868
26/12/2018	5,2	5,4	80%	83%	10	3562
27/12/2018	4,1	3,2	78%	71%	9	2759
28/12/2018	6,2	7,6	64%	83%	12	5185
29/12/2018	6,0	5,1	59%	75%	10	3894
30/12/2018	5,9	4,1	55%	83%	11	2904
31/12/2018	6,1	4,9	60%	92%	11	3557