



Douglas Diniz Landim

**ANÁLISE DE DEMANDA DO RESTAURANTE  
UNIVERSITÁRIO DO ICT UNIFESP VIA REDES  
NEURAIIS**

São José dos Campos, SP



Douglas Diniz Landim

# **ANÁLISE DE DEMANDA DO RESTAURANTE UNIVERSITÁRIO DO ICT UNIFESP VIA REDES NEURAS**

Trabalho de conclusão de curso apresentado ao Instituto de Ciência e Tecnologia – UNIFESP, como parte das atividades para obtenção do título de Bacharel em Ciência da Computação.

Universidade Federal de São Paulo – UNIFESP

Instituto de Ciência e Tecnologia

Bacharelado em Ciência da Computação

Orientador: Prof. Dr. Marcos Gonçalves Quiles

São José dos Campos, SP

Outubro de 2020

Douglas Diniz Landim

# **ANÁLISE DE DEMANDA DO RESTAURANTE UNIVERSITÁRIO DO ICT UNIFESP VIA REDES NEURAIIS**

Trabalho de conclusão de curso apresentado ao Instituto de Ciência e Tecnologia – UNIFESP, como parte das atividades para obtenção do título de Bacharel em Ciência da Computação.

Trabalho para apresentar em Outubro/2020:

---

**Prof. Dr. Marcos Gonçalves Quiles**  
Orientador

---

**Professora**  
Dra. Daniela Leal Musa

---

**Professora**  
Dra. Regina Célia Coelho

São José dos Campos, SP

Outubro de 2020

*Este trabalho é dedicado aos meus pais que apoiaram e sacrificaram esforços para me manter ativo nessa jornada, a todos os professores que me somaram conhecimentos, oportunidades e esperanças indo além de suas rotinas e agendas em prol do ensino, e principalmente à todos que me motivaram me oferecendo desafios para que eu pudesse enfrentá-los superando meus próprios limites*



# Agradecimentos

**Minha Jornada** Minha jornada pela graduação foi marcada por muita persistência, dificuldades e fracassos. Agradeço primeiramente a Deus por me dar fé e alimentar minha persistência e esperança. Apesar de todo o conteúdo técnico das mais de 40 disciplinas do meu curso, o que mais me agregou aprendizado foi o ambiente desafiador desta universidade; que somado à muitas dificuldades pessoais, acidentes, contra-tempos de saúde, profissão e família; constituiu o conjunto perfeito de desafios que me transformou em uma pessoa forte e destemida para enfrentar as cobranças do mercado, mais convicto e perseverante a cada nova tentativa de conquistar meus objetivos. Agradeço à minha família por sempre me apoiar dando tudo de si, aos meus professores que me orientaram e me motivaram, nas reuniões e chats online até nos finais de semana, aos meus amigos universitários, e a todos os colegas e colaboradores que conheci durante a graduação na Unifesp.

**Em especial, agradeço:** À professora Daniela Musa por ser minha primeira coordenadora de curso e orientadora quando ingressei na Unifesp.

Aos professores das disciplinas que formaram minha base de conhecimento da ciência da computação e que me deram grande preparo para as minhas atividades acadêmicas, profissionais e nos diversos processos seletivos que participei no mercado. Aos professores Reginaldo Kuroshu, Bruno Kimura, Álvaro Fazenda, Regina Coelho, Arlindo Flavio, Antonio Chaves, Ana Luiza e Otavio Lemos.

À professora Camila Bertini que na disciplina de simulação de sistemas me motivou nos primeiros estudos no tema deste trabalho, com uma realização de correlação com reamostragem de consumo x temperatura em 2016.

À professora Flavia Martins de estatística, que me orientou algumas vezes em sua sala sobre a introdução teórica de predições com redes neurais, sua tese de doutorado foi bem complementar e enriquecedora na fundamentação teórica deste trabalho.

Aos professores Vinícius Veloso e Fabio Faria, por me apresentarem a disciplina de inteligencia artificial. E ao Vinícius Veloso por me orientar na primeira parte deste trabalho e em todo o desenvolvimento de fundamentação teórica da primeira parte.

Ao professor Marcos Quiles por me orientar nesta segunda parte do trabalho, me apresentando desde o início o trabalho com python, tensorflow, scikit learning, sobre as orientações mais específicas do aprendizado da rede neural. Em me orientar nas métricas de avaliação dos

modelos e sobre a toda a metodologia experimental, além das revisões no texto final.

À equipe de Data Science e Machine Learning Engineering da empresa 2RP-NET, especialista em análise de fraudes, dados, e trabalhos com aprendizado de máquina em Big-data, da qual recentemente fui integrado em setembro/2020 graças ao aprendizado adquirido neste trabalho acadêmico. E por ter disponibilizado um período extenso da reunião de equipe durante o nosso expediente para discutirmos os resultados e métricas deste trabalho, equipe da qual é composta por profissionais mestres e doutores em data science e machine learning.

À todos os outros professores, colegas, alunos e colaboradores do instituto de ciência e tecnologia da UNIFESP.

À minha família por ter me apoiado nesse longo período na UNIFESP.

*Mesmo desacreditado e ignorado por todos, não posso desistir, pois para mim, vencer é nunca desistir.*  
*(Albert Einstein)*



# Resumo

O presente trabalho tem como objetivo o estudo de métodos para a previsão de vendas do restaurante universitário da Unifesp para evitar super-projeção de demanda com consequência de desperdício de alimentos, ou subprojeção com consequência de docentes ou discentes sem refeições. Em uma investigação anterior, realizada como trabalho de disciplina, o autor empregou métodos estatísticos para a análise do comportamento de consumo de refeições. Neste trabalho, foram desenvolvidos modelos de aprendizado de máquina, mais especificamente, redes neurais perceptron de múltiplas camadas e redes *gated recurrent units*, com métodos de análises de dados coletados, preparação e pré-processamento das informações, seleção e avaliação dos melhores modelos e conclusões finais.

**Palavras-chave:** Redes Neurais Artificiais, Previsão de demanda, Aprendizado de Máquina, Inteligência Artificial, Perceptron Múltiplas camadas.



# Abstract

This current work aims to study methods for forecasting meals of Unifesp university restaurant to avoid over-projection of demand resulting from food waste, or under projection with the consequence of teachers or students without meals. In a previous investigation, carried out as discipline work, the author used statistical methods to analyze meal consumption behavior. Machine learning models were developed in this work, more specifically, multilayer perceptron neural networks and gated recurrent units networks, with data analysis methods, preparation and pre-processing of information, selection, and evaluation of the best models and final conclusions.

**Keywords:** Artificial Neural Networks, Demand Prediction, Machine Learning, Artificial intelligence, Perceptron Multiple layers.



# Lista de ilustrações

Figura 1 – Métodos de previsão de demanda. . . . .	25
Figura 2 – Neurônio Artificial. . . . .	28
Figura 3 – Neurônio Artificial Perceptron. . . . .	29
Figura 4 – Problema linearmente separável (a) e não separável (b). . . . .	29
Figura 5 – Exemplos de funções de Ativação. . . . .	30
Figura 6 – Rede de perceptrons com múltiplas camadas. . . . .	31
Figura 7 – Ponto ótimo de parada da validação cruzada. . . . .	32
Figura 8 – Fases de treino da <i>MLP-Back-Propagation</i> . . . . .	33
Figura 9 – Comportamento de otimizadores para MLP treinadas com <i>Backpropagation</i> . . . . .	34
Figura 10 – Arquitetura do modelo GRU. . . . .	35
Figura 11 – Rede Neural Perceptron de Múltiplas Camadas. . . . .	38
Figura 12 – Rede Neural Perceptron de Múltiplas Camadas . . . . .	39
Figura 13 – Restaurante Universitário do ICT-Unifesp . . . . .	41
Figura 14 – Domínio temporal da fase 1. . . . .	49
Figura 15 – Domínio temporal da fase 2 . . . . .	50
Figura 16 – Resultado do modelo RNN_ENDO_2 obtido sobre o conjunto de dados aleatoriamente ordenado sobre o tempo. . . . .	50
Figura 17 – Resultado do modelo RNN_ENDO_2 obtido sobre o conjunto de dados com ordenação corrigida . . . . .	50
Figura 18 – Estimativa do restaurante para o ano de 2019. . . . .	51
Figura 19 – Gráfico de dispersão da estimativa de consumo do restaurante para o ano de 2019. . . . .	52
Figura 20 – Correlação entre consumo e vendas de almoço. . . . .	52
Figura 21 – Gráfico de dispersão entre consumo e vendas de almoço. . . . .	54
Figura 22 – Vendas de <i>tickets</i> normalizados com teto de 3x o desvio padrão. . . . .	54
Figura 23 – Vendas de <i>tickets</i> escalada entre 0 a 1. . . . .	54
Figura 24 – Correlação de consumo de almoço e jantar de 1 dia anterior. . . . .	55
Figura 25 – Gráfico de dispersão entre consumo e jantar de 1 dia anterior. . . . .	55
Figura 26 – Gráfico de violino da distribuição do consumo na segunda feira. . . . .	56
Figura 27 – Gráfico de violino da distribuição do consumo na terça feira. . . . .	56
Figura 28 – Gráfico de violino da distribuição do consumo na quarta feira. . . . .	56
Figura 29 – Gráfico de violino da distribuição do consumo na quinta feira. . . . .	56
Figura 30 – Gráfico de violino da distribuição do consumo na sexta feira. . . . .	57

Figura 31 – 1a Fase : Relação da distribuição do consumo com o avanço do semestre, Correlação (r) = -0.35. . . . .	57
Figura 32 – Gráfico de dispersão da distribuição do consumo com o avanço do semestre.	58
Figura 33 – Topologia do modelo MLP1, Ferramenta NETRON. . . . .	59
Figura 34 – Gráfico de treino do modelo MLP1, RMSE = 130,62 . . . . .	60
Figura 35 – Topologia do modelo MLP2. . . . .	61
Figura 36 – Gráfico de treino do modelo MLP2. RMSE = 107,97. . . . .	62
Figura 37 – Topologia do modelo RNN_EXO_1. . . . .	63
Figura 38 – Topologia do modelo RNN_ENDO_2. . . . .	64
Figura 39 – Análise de anomalias preditivas do RNN_ENDO_2 . . . . .	65
Figura 40 – Treino do modelo RNN_EXO_1 na 1a fase, RMSE = 132.94 . . . . .	67
Figura 41 – Gráfico de treino do modelo RNN_EXO_1 na 2a fase, RMSE = 109.97 . . . . .	67
Figura 42 – Gráfico de dispersão de teste do modelo RNN_EXO_1, 1a fase . . . . .	69
Figura 43 – Gráfico de dispersão de teste do primeiro semestre, RNN_EXO_1 treinado na segunda fase. . . . .	69
Figura 44 – Teste do modelo RNN_EXO_1, 1a fase. . . . .	70
Figura 45 – Teste do primeiro semestre do RNN_EXO_1 treinado na segunda fase. . . . .	70
Figura 46 – Gráfico final de teste do modelo RNN_EXO_1. . . . .	71
Figura 47 – Gráfico de dispersão do modelo RNN_EXO_1. . . . .	72
Figura 48 – Topologia do modelo MLP_ENDO_1 . . . . .	78
Figura 49 – Topologia do modelo RNN_ENDO_1 . . . . .	79
Figura 50 – Topologia do modelo RNN_EXO_2 . . . . .	80
Figura 51 – Topologia do modelo RNN_EXO_3 . . . . .	81

# Lista de tabelas

Tabela 1 – Formato original dos dados originais obtidos pelo restaurante universitário . . . . .	43
Tabela 2 – Transformação dos registros do restaurante em uma série temporal. . . . .	43
Tabela 3 – Estrutura final do conjunto de dados indexados por data . . . . .	45
Tabela 4 – Métricas da estima de consumo do restaurante para o ano de 2019 . . . . .	52
Tabela 5 – Comparação de consumo com um dia anterior . . . . .	53
Tabela 6 – Erros de predições do modelo RNN_ENDO_2 na 1a fase . . . . .	65
Tabela 7 – Anomalias de predições do modelo RNN_ENDO_2 na 1a fase . . . . .	65
Tabela 8 – Métricas do melhor modelo: RNN_ENDO_2 . . . . .	66
Tabela 9 – RNN_EXO_1 TREINADO NA 1A FASE,TESTE 1o SEMESTRE 2019 . . . . .	68
Tabela 10 – RNN_EXO_1 TREINADO NA 2A FASE, TESTE 1o SEMESTRE 2019 . . . . .	68
Tabela 11 – RNN_EXO_1 TREINADO NA 2A FASE, TESTE ANO DE 2019 . . . . .	71
Tabela 12 – Previsões e erros de todos os modelos . . . . .	84
Tabela 13 – Erros quantitativos de todos os modelos . . . . .	84
Tabela 14 – Métricas estatísticas e de treino de todos os modelos . . . . .	85
Tabela 15 – Métricas gráficas de todos os modelos . . . . .	85



# Lista de abreviaturas e siglas

ICT	Instituto de Ciência e Tecnologia
R.U.	Restaurante Universitário
UNIFESP	Universidade Federal de São Paulo
UFV	Universidade Federal de Viçosa
UNESP	Universidade Estadual Paulista Júlio de Mesquita Filho
BDMEP	Banco de Dados Meteorológicos para Ensino e Pesquisa
RNA	Rede Neural Artificial
MLP	Multi Layer Perceptron
GRU	Gated Recurrent Unit
RMSE	Root Mean Squared Error



# Sumário

<b>1</b>	<b>Introdução</b>	<b>21</b>
<b>2</b>	<b>Fundamentação teórica</b>	<b>25</b>
2.1	Métodos de Previsão de Demanda	25
2.1.1	Séries Temporais	26
2.2	Redes Neurais Artificiais	27
2.2.1	Perceptron	28
2.2.2	Rede MultiLayer Perceptron (MLP)	30
2.2.3	Rede perceptron Múltiplas Camadas com <i>Backpropagation</i> .	32
2.2.4	Redes Recorrentes: O modelo GRU	34
2.2.4.1	Dropout	35
<b>3</b>	<b>Trabalhos relacionados</b>	<b>37</b>
3.1	Trabalho de comparações de métodos de previsão de demanda.	37
3.2	Previsão de demanda em restaurantes universitários	37
3.3	Previsão de demanda em outros ambientes	39
<b>4</b>	<b>Metodologia</b>	<b>41</b>
4.1	Área de estudo	41
4.2	Descrição dos dados	42
4.3	Obtenção e tratamento dos dados	42
4.3.1	Dados endógenos	42
4.3.2	Dados exógenos	43
4.3.3	Pré-processamento	44
4.3.4	Tratamento dos dados para entrada nos modelos	45
4.3.5	Fases Experimentais	46
4.4	Definição e treino dos modelos	47
4.4.1	Hiper parâmetros : Função de ativação e otimizador	47
4.5	Teste e Métricas de avaliação	48
<b>5</b>	<b>Resultados</b>	<b>49</b>
5.1	Organização do conjunto de dados	49
5.1.1	Manipulação e pré-processamento do conjunto de dados	50
5.2	Avaliação das Variáveis	51
5.2.1	Estimativas de consumo do restaurante	51
5.2.2	Análise das variáveis endógenas	51

5.2.2.1	Consumo do dia vigente em relação às vendas de tickets do dia anterior . . . . .	52
5.2.2.2	Normalização e escala de <i>features</i> . . . . .	53
5.2.2.3	Consumo atual em relação ao consumo do jantar de 1 dia anterior. . . . .	55
5.2.3	Análise da sazonalidade semanal . . . . .	56
5.2.4	Análise das variáveis exógenas . . . . .	57
5.2.4.1	Consumo atual em relação ao avanço do semestre . . . . .	57
5.3	Protocolo Experimental . . . . .	58
5.3.1	Avaliação do aprendizado do problema da predição de refeições por meio de redes neurais MLP . . . . .	58
5.3.1.1	Ajuste empírico de topologia do primeiro modelo perceptron . . . . .	58
5.3.2	Topologias dos melhores modelos . . . . .	62
5.3.2.1	Modelo misto RNN_EXO_1, melhor resultado na segunda fase e em todo o trabalho . . . . .	62
5.3.2.2	Modelo endógeno GRU RNN_ENDO_2, melhor resultado na primeira fase . . . . .	63
5.3.3	Diferenças principais dos resultados entre as fases experimentais . . . . .	65
5.4	Resultados com o melhor modelo, RNN_EXO_1 . . . . .	66
5.4.1	Comparativo do treino entre as duas fases . . . . .	66
5.4.2	Comparativo de teste do modelo no primeiro semestre, entre as duas fases . . . . .	67
5.4.3	Teste final do modelo RNN_EXO_1 para as predições no RU do ICT-Unifesp . . . . .	70
<b>6</b>	<b>Conclusão . . . . .</b>	<b>73</b>
6.1	Conclusões gerais . . . . .	74
 <b>Anexos . . . . .</b>		<b>75</b>
<b>ANEXO A Topologias de outros modelos de redes testados . . . . .</b>		<b>77</b>
<b>ANEXO B Métodos e códigos experimentais . . . . .</b>		<b>83</b>
B.1	Primeira Fase Experimental . . . . .	83
B.2	Segunda Fase Experimental . . . . .	83
B.3	Tabela completa das métricas de todos os modelos . . . . .	84
<b>Referências . . . . .</b>		<b>87</b>

# 1 Introdução

Dentre as definições mais aceitas sobre segurança alimentar está a versão cunhada durante a Cúpula Mundial da Alimentação de 1996 (SHAW, 2007), que a enuncia como uma situação em que todas as pessoas, a todo momento, tenha condições física, social e econômica de acesso a alimentação segura, saudável e nutritiva para uma vida saudável e ativa. Porém, Segundo a Organização Mundial da Saúde (ORGANIZATION et al., 2009) mais de 1 bilhão de pessoas no mundo possuem uma dieta nutritivamente insuficiente e mais que o dobro de pessoas tem carência de micronutrientes.

Segundo Webb et al. (2006) a segurança alimentar pode ser decomposta em 3 pilares essenciais, disponibilidade, possibilidade de acesso e utilização racional. Estes conceitos são intrinsecamente hierárquicos uma vez que a disponibilidade não garante acesso, que por sua vez não garante sua utilização racional. Com os avanços na produção agrícola e a abertura dos mercados econômicos mundiais, ainda que não tenha solucionado o problema, possibilitou passos largos nos dois primeiros pilares da segurança alimentar supracitados. Assim, nas primeiras décadas do século XXI o terceiro pilar, da utilização racional dos recursos, tem se mostrado um importante fator a ser considerado neste novo mundo globalizado (BARRETT, 2010).

No final do ano de 2019 uma doença altamente transmissível foi identificada na província de Wuhan na china, comumente chamada de o novo Coronavírus (COVID-19), a qual foi declarada estar em estado de pandemia em março de 2020, e até o presente momento infectou mais de 40 milhões de pessoas e ocasionou 1 milhão de mortes<sup>1</sup>. Ainda que com esforços globais para contenção e remediação da pandemia de COVID-19, não existem vacinas ou curas cientificamente comprovadas e a necessidade de *lockdown* ainda é pertinente para evitar novas ondas de contaminação. Os efeitos da pandemia foram devastadores em praticamente todas as áreas econômicas, principalmente o setor de alimentação que é essencial para a manutenção da vida e não tem espaço para *lockdown* (GALANAKIS, 2020).

No dia 22 de julho de 2020, a Associação Brasileira da Indústria e Consumo de alimentos, realizou uma conferência sobre os impactos da pandemia do COVID-19 na cadeia produtiva brasileira de alimentos<sup>2</sup>. Nesta conferência foram reunidos especialistas e executivos do setor agroindustrial para analisar mudanças de comportamento de consumo nos setores terciários que lidam diretamente com o fornecimento de alimentos ao consumidor final. A partir desta análise

---

<sup>1</sup> <https://www.worldometers.info/coronavirus/>

<sup>2</sup> <https://www.abia.org.br/noticias/abia-participa-de-live-para-discutir-os-impactos-da-pandemia-na-cadeia-produtiva-de-alimentos>

da nova rotina do consumidor, estão sendo buscados novos métodos de se prever o formato atual de consumo e a nova demanda para estes setores. Por fim, foram discutidos também métodos de produção e abastecimento do agronegócio para o setor terciário.

Assim, neste período caótico, em que a mão de obra global reduziu aproximadamente 25% durante os primeiros meses de pandemia (HUFF et al., 2015), vem atona a importância do desenvolvimento de processos que utilizem os recursos alimentícios da forma mais racional possível. Neste contexto, uma área que chama a atenção é a de previsão de demandas indústrias ou de estabelecimentos que lidam com alimentos perecíveis em grande quantidade, uma vez que o alimento tem prazo de validade e a produção por demanda pode evitar armazenamento por longos períodos ou descartes indesejado de alimentos prontos para consumo.

Neste sentido, tornam-se viáveis a previsão de demandas para refeições em restaurantes industriais ou de grande porte, como os restaurantes universitários (RU) em que são ofertadas refeições a preços acessíveis para a comunidade estudantil da instituição, através de subsídios no valor repassado aos alunos. Geralmente os RU's são terceirizados por meio de licitações em que o governo do estado uma porcentagem de cada refeição produzida para a empresa gestora do restaurante. Devido a regulamentações sanitárias, refeições preparadas mas não consumidas até o final do expediente devem ser descartadas para evitar contaminações e garantir a segurança alimentar do consumidor <sup>3</sup>. Dessa forma, a produção em excedente (não consumida) de refeições por parte dos RU's gera não somente um gasto exagerado de recursos financeiros públicos mas também uma utilização pouco racional dos recursos alimentícios.

Neste contexto, o restaurante universitário da Unifesp em São José dos Campos é um bom estudo de caso uma vez que vende aproximadamente 90 mil refeições anualmente, em que cada refeição tem o valor fixo de R\$2,50 para alunos e um subsidio médio de aproximadamente R\$9,00 por refeição na ultima década (2011-2019). Com isto, foram subsidiados mais de 6 milhões de reais durante os anos de 2011 a 2019.

Uma forma de se reduzir esse desperdício está na desenvolvimento de métodos de predição de consumo (LOPES, 2008; ROCHA; MATOS; FREI, 2011). Normalmente, as abordagens para previsão da demanda de consumo de um restaurante universitário se resumem a análise exploratória de dados coletados, como por exemplo, as vendas computadas na semana ou mês anterior. Além disso, informações externa, denominadas dados exógenos, também podem ser considerados no processo de predição, como por exemplo dados climáticos, dados do calendário anual, feriados, dentre outras informações que podem ser relevantes para a estimativa de consumo.

A predição de consumo de redes restaurantes universitários já foi abordada em outros

<sup>3</sup> <https://super.abril.com.br/mundo-estranho/o-que-acontece-com-a-comida-que-sobra-dos-restaurantes/>

trabalhos. Por exemplo, no trabalho de Lopes (2008) realizado na UFV, é utilizado um modelo de rede neural perceptron de 1 camada oculta e 1 neurônio de saída, utilizando os 5 dias anteriores como parâmetros de predição com erro de 3% sobre o total consumido, e no trabalho de Rocha, Matos e Frei (2011) realizado na UNESP é utilizado também um modelo de rede neural com 1 camada oculta e 1 neurônio de saída, utilizando apenas 1 parâmetro que informa o número de refeições do dia anterior, e outros parâmetros informando médias de dias anteriores e informações relativas à data de consumo, este modelo obteve erro de 9,5% calculado sobre o total consumido.

Assim, este trabalho consiste na aplicação de modelos de redes neurais para a previsão de demanda das refeições fornecidas no restaurante universitário da Unifesp no campus de São José dos Campos. Para isso, será utilizado um conjunto de dados históricos disponíveis no sistema da Unifesp e outras informações externas (exógenas) que podem impactar o comportamento de consumo e seu processo de predição.

Os objetivos gerais deste trabalho compõem a construção e comparação de modelos de Redes Neurais Artificiais para a previsão da demanda de refeições do restaurante universitário do ICT-UNIFESP. Especificamente, tem-se como objetivos:

- a) Obter e preprocessar os dados de consumo e venda de refeições do ambiente do restaurante e dados do ambiente externo ao consumo, como por exemplo dados climáticos, visando seu uso como entrada de modelos de aprendizado de máquina para a obtenção das predições de consumo na saída destes modelos;
- b) Realizar análises exploratórias e descritivas de todos os dados;
- c) Construir e avaliar modelos preditivos de redes neurais;
- d) Realizar análises sobre as métricas de predições dos modelos apontando suas características viáveis para uma predição de consumo.

O restante deste documento está organizado da seguinte forma. O capítulo 2 introduz a fundamentação necessária a compreensão deste trabalho; A literatura relacionada ao tema é descrita no capítulo 3; O capítulo 4 apresenta a metodologia empregada nesse estudo; Os resultados são sumarizados no capítulo 5; As principais conclusões desse trabalho bem como as indicações de investigações futuras são apresentadas no capítulo. 6; Por fim, o capítulo A apresenta alguns resultados adicionais gerados durante os experimentos e os códigos implementados.



## 2 Fundamentação teórica

Neste capítulo será feita uma breve revisão das características principais das series temporais, assim como das redes neurais, especialmente das que serão aplicadas no problema de interesse, chamadas de redes Multilayer Perceptron (MLP) e redes Gated Recurrent Unit (GRU).

### 2.1 Métodos de Previsão de Demanda

A previsão de demanda pode ser definida como um processo de busca de informações que geram vendas de um produto ou um conjunto de produtos.

Em Moreira (1998) são definidos os métodos de previsão de demanda como processos que utilizam tais informações para estimar uma demanda futura, podendo se encaixar nesta definição desde métodos subjetivos e intuitivos até métodos matemáticos e computacionais. No mesmo trabalho, são classificados os modelos de previsão em dois grandes grupos: métodos qualitativos e métodos quantitativos.

Em Junior (2007) foi feita a avaliação de diversos métodos de previsão de demanda, os quais são esquematizados na Figura 1.

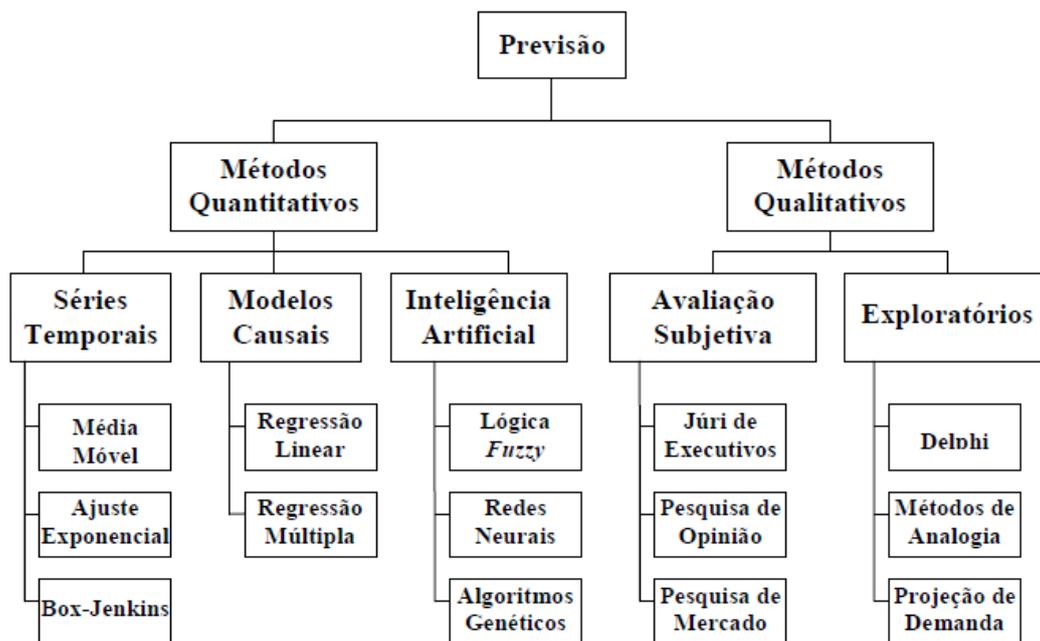


Figura 1: Métodos de previsão de demanda.

Fonte: (JUNIOR, 2007).

Os métodos qualitativos são definidos em (JUNIOR, 2007) como um julgamento dos dados expostos sem processamento analítico, tais como agrupamento e classificação de dados. Esses métodos não fornecem novas informações numéricas nem modelos preditivos.

Já os métodos quantitativos são definidos, no mesmo trabalho, como sendo analíticos e baseados em modelos matemáticos para realizar previsões. Esses métodos analisam padrões de comportamento de um histórico de dados, visando prever comportamento futuro.

Em Junior (2007) também é mencionado que os dados coletados para modelos de previsão, ao ser projetados graficamente, evidenciam comportamentos que podem ser generalizados de forma subjetiva pelos gestores dos dados. Em todos os casos, a análise de dados é necessária para selecionar parâmetros de demanda objetivos para fazer previsões. Porém, apenas a análise dos dados pode ser insuficiente, e se realizada com critérios incorretos pode comprometer as conclusões dos estudos.

A análise de dados é um processo amplo que visa tratar os dados desde sua aquisição e pré-processamento até a sua interpretação a partir de ferramentas complexas de mineração. Nas diferentes etapas, são abrangidas diversas técnicas matemáticas, probabilísticas, estatísticas, computacionais e heurísticas. Na Seção 2.1.1 serão comentadas as principais características das séries temporais, o qual será o método de previsão aplicado ao problema prático de interesse.

### 2.1.1 Séries Temporais

De acordo com Morettin (1987), uma série temporal é um conjunto de observações ordenadas em função do tempo, comumente iguais, apresentando uma dependência serial entre ela. Também pode ser definida como uma realização de um conjunto de variáveis aleatórias  $X = \{x_1, x_2, \dots, x_T\}$ , ordenadas no tempo, onde  $T$  representa o comprimento da série, como é feito em (DAVILA, 201-). A relação entre as variáveis é comumente descrita pela função de distribuição conjunta delas, ou, em outros casos, pela média e covariâncias.

Entre os objetivos do análise de séries temporais, podem ser destacados os seguintes: i) descrever o comportamento das séries identificando tendências e variações, ii) analisar e modelar a dependência entre as observações, e iii) fazer previsões de valores futuros da série.

As series temporais tem uma componente determinística e uma componente aleatória, e podem ser contínuas ou discretas, dependendo do tipo de observação das variáveis. São chamadas de estacionárias quando as propriedades de média, variância e covariância são mantidas no tempo. A tendência faz referência à taxa de crescimento ou de decréscimo, podendo ser linear, exponencial ou amortecido. Já a oscilação da tendência é conhecida como ciclo. Além disso, podem apresentar sazonalidade, isto é, exibir comportamento que tende a se repetir em certo numero de períodos de tempo.

Exemplos de cada uma dessas propriedades, assim como técnicas aplicadas para estudo e correção de cada uma delas podem ser vistas em (EHLERS, 2009). Essas características, junto com o estudo da componente aleatória, fornecem informações de interesse para as aplicações práticas. Devido ao objetivo deste trabalho, serão aplicadas series temporais para eventos com sazonalidade, a fim de estimar previsões. As técnicas para fazer isto usam métodos de regressão, e algumas aplicações praticas podem ser vistas incluem previsão de consumo de energia elétrica realizados , como feito em Almeida (2013), Ruas (2012) e Silva (2010), e da previsão de demanda de produtos cosméticos, como pode ser visto em Junior (2007).

## 2.2 Redes Neurais Artificiais

As redes neurais artificiais são ferramentas com raiz multidisciplinar, pois são nutridas por conhecimentos de neurociência, matemática, estatística, física, ciência da computação e engenharia (HAYKIN, 2007), e fazem parte da grande área de conhecimento denominada Inteligência Artificial (IA), termo cunhado em (KAPLAN; HAENLEIN, 2019).

De forma resumida, poderia ser definido que "a inteligência artificial é o ramo da ciência da computação que se ocupa do comportamento inteligente."(LUGER, 2004). Os sistemas de inteligência artificial buscam resolver funções e problemas inspirados em duas características humanas: capacidade de abstração e aprendizagem com o erro.

No contexto de IA um neurônio é uma unidade considerada fundamental para o processamento de informação, e as redes neurais são conjuntos de neurônios artificiais interconectados através de relações, funções lógicas e matemáticas (HAYKIN, 2007). Os neurônios de uma rede são capazes de processar múltiplos valores de entradas e reagir rapidamente produzindo uma resposta relacionada à essas entradas, simulando o comportamento do cérebro humano.

Inspirado por uma busca de um modelo computacional do neurônio biológico, o primeiro modelo de neurônio artificial, denominado MCP, foi proposto no artigo *A Logical Calculus of the Ideas Immanent in Nervous Activity*. (MCCULLOCH; PITTS, 1943), uma ilustração adaptada por (LEMOS, 2003) deste modelo pode ser visualizada na figura 2.

McCulloch era psiquiatra e neuroanatomista e passou cerca de 20 anos refletindo e estudando sobre a representação do sistema nervoso, em 1942 ele convidou Pitts, que era matemático, para fazer parte das suas pesquisas.

A estrutura do neurônio artificial reage a um vetor de entradas e as sinapses são representadas por pesos numéricos. Uma função de transferência, também chamada função de ativação, avalia uma combinação linear dos valores da entrada e os pesos das sinapses, determinando se o neurônio é ativado ou não dependendo do valor obtido. Se o neurônio é ativado é emitido um

valor de saída 1, caso contrário se emite um valor de saída 0.

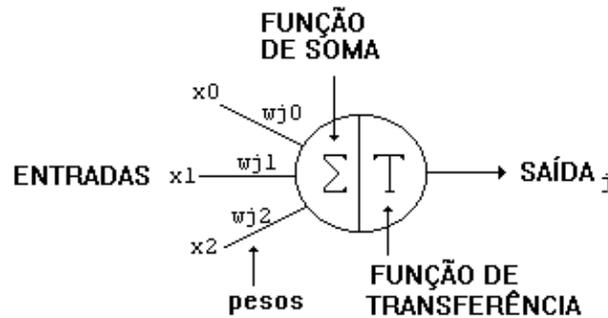


Figura 2: Neurônio Artificial.

Fonte:(LEMOS, 2003)

Todo o funcionamento deste modelo então é reduzido a responder se a soma ponderada recebida é maior que um valor numérico estabelecido. Contudo, associado a este neurônio não foi proposto uma forma automática para ajuste dos pesos, ou seja, não foi dado um algoritmo de aprendizagem para treinar o neurônio. Esse problema foi contornado posteriormente, pela formulação do Perceptron.

Em Haykin (2007) no capítulo 1.9 denominado *Notas Históricas*, é apresentada com mais detalhes a fascinante história de desenvolvimento das redes neurais desde a concepção inicial do estudos do neurônio biológico até redes complexas de aprendizagem supervisionada chamadas de *Máquinas de Vetor Suporte*.

### 2.2.1 Perceptron

O modelo de neurônio proposto em Mcculloch e Pitts (1943), apesar de simular um neurônio biológico e resolver algumas tarefas lógicas e matemáticas, não atendia o objetivo principal da Inteligência Artificial: A capacidade de aprendizado. Para poder utilizar esta estrutura, era necessário conhecer o ajuste dos pesos das entradas, o qual não era um problema trivial em muitos casos.

O primeiro neurônio com um algoritmo de aprendizado foi proposto em Rosenblatt (1958) e foi nomeado de Perceptron. Nesse trabalho, os pesos das conexões são ajustados de forma autônoma com a introdução de pesos associados e um valor bias, a fim de buscar um reconhecimento autônomo de padrões. Na Figura 3 é apresentado um esquema do funcionamento da estrutura.

Porém, em Minsky e Papert (1969) foi provado que, devido ao modelo aprendizado limitado a uma combinação linear, o perceptron poderia resolver apenas problemas linearmente

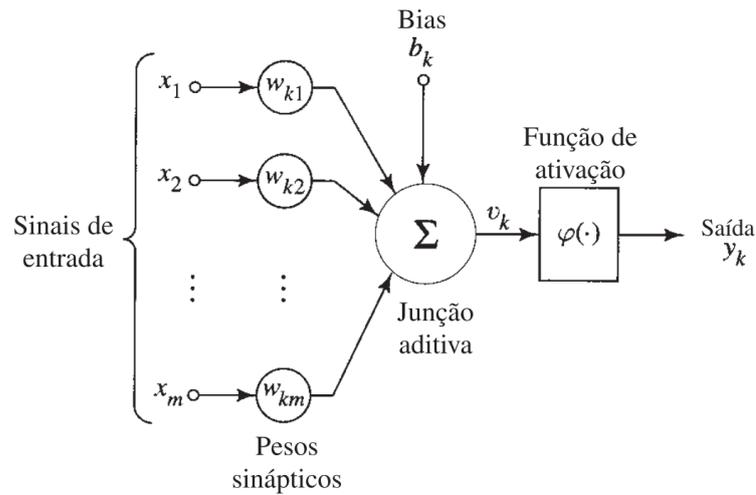


Figura 3: Neurônio Artificial Perceptron.

Fonte: Haykin (2007)

separáveis. Na Figura 4 são apresentados dois problemas simples, um que o perceptron pode resolver (a), e outro que não (b).

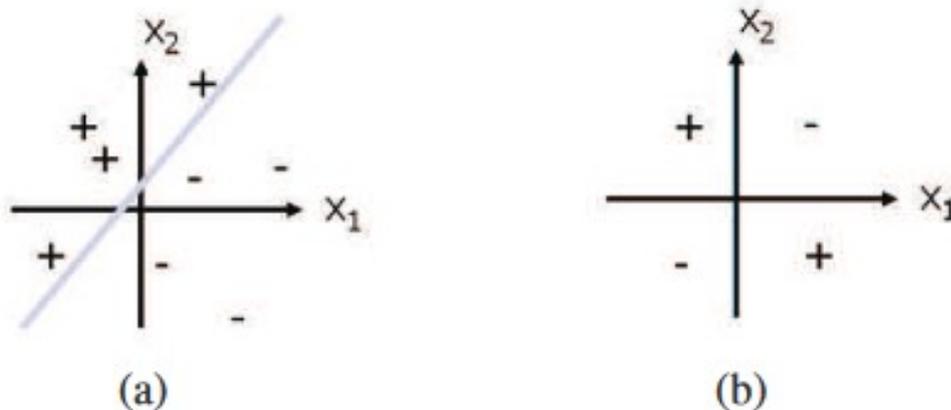


Figura 4: Problema linearmente separável (a) e não separável (b).

Fonte:(MARIANO, 2014).

Quase duas décadas depois, foi apresentado em (RUMELHART; HINTON; MCCLELLAND, 1986) o primeiro modelo de rede neural, chamada de rede perceptron, aplicando o treinamento por combinações lineares à um conjunto de perceptrons interligados. Essa abordagem permitia resolver problemas mais complexos por meio de uma combinação de soluções.

A rede possuía apenas uma camada de entrada, uma única saída, e uma função de ativação  $\varphi$  (HAYKIN, 2007). A função de ativação da rede perceptron ainda poderia ser linear ou não linear. Na Figura 5 são ilustradas algumas funções de ativação comumente utilizadas.

Em Almeida (2013) é analisado o processo de aprendizado da rede perceptron, de forma

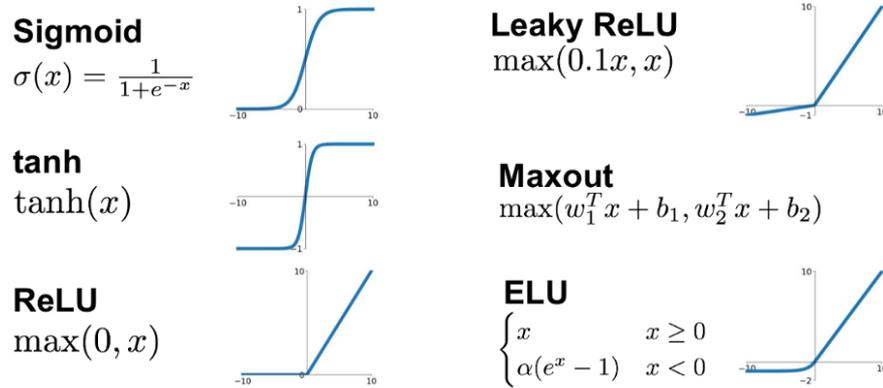


Figura 5: Exemplos de funções de Ativação.

Fonte: (MC AI, 2020)

supervisionada. Neste processo, a estrutura aprende a relacionar um conjunto observado de variáveis de entrada na rede com um ou mais valores de saída esperados denominado como valores reais ou valores verdades. Depois são avaliados os resultados do aprendizado, comparando estes valores com os valores gerados pela rede perceptron sobre o mesmo conjunto de dados, e partir desta comparação é calculada a medida de erro do treinamento.

Um critério de parada do algoritmo de treino é verificar se o erro é aceitável ou não. Caso afirmativo, a rede neural mantém os valores dos pesos das sinapses obtidos no momento. Caso contrário, é feita uma nova época de treino tentando ajustar os pesos para obter um erro menor. O outro critério de parada é atingir um número máximo permitido de épocas de treinamento. O reajuste de pesos é denominado taxa de aprendizagem.

O seguinte passo no desenvolvimento de modelos de redes neurais está relacionado com a topologia que determina a quantidade dos perceptrons na rede e a forma como eles se conectam, gerando redes de múltiplas camadas.

### 2.2.2 Rede MultiLayer Perceptron (MLP)

A possibilidade de combinar duas ou mais camadas de perceptrons foi dada pela utilização de um perceptron combinador de sinal de saída. Com ele as redes neurais são ampliadas para várias colunas de perceptrons interconectados. Cada coluna é denominada uma camada oculta da rede neural. A última camada deve ter o número de perceptrons correspondente ao número de saídas desejadas. Na Figura 6 é apresentada uma rede neural com 2 camadas ocultas, e cuja camada de saída possui 3 neurônios.

Em Braga A. de P.; Carvalho (2000) é postulado que através de uma camada intermediária é possível aproximar qualquer função contínua e, ainda mais, que duas camadas intermediárias são suficientes para aproximar qualquer função matemática. Se a utilização de duas ou

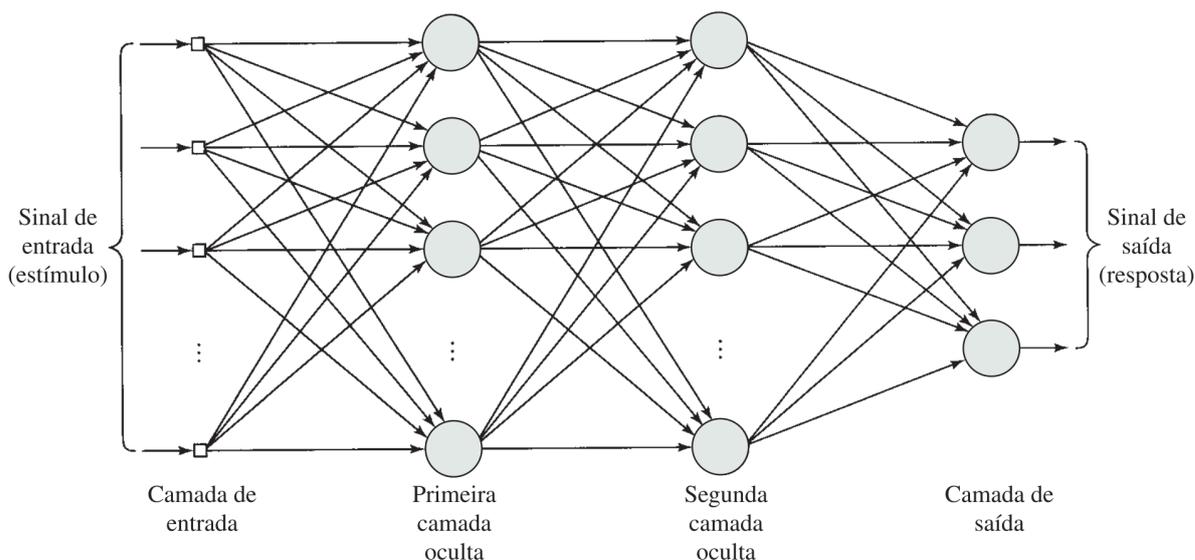


Figura 6: Rede de perceptrons com múltiplas camadas.

Fonte: (HAYKIN, 2007)

mais camadas pode facilitar o treinamento da rede, torna-se inviável a utilização de um número grande destas, pois em cada camada oculta o erro é estimado a partir do erro na camada anterior, o que gera perda de precisão.

Nas aplicações práticas, tem-se visto que em alguns casos, que a capacidade de abstração e reconhecimento de padrões das redes neurais sobrepassa as capacidades humanas. Em outros casos, uma rede pode não produzir uma resposta esperada resolvendo erroneamente um problema, assim como o cérebro humano, devido à limitações de aprendizado ou falhas nos treinos.

Para treinar de modo supervisionado uma rede MLP, o conjunto de dados de entrada deve ser dividido em em dois subconjuntos, um de *treino* e um de *validação*. Esses subconjuntos podem ser separados com diversas técnicas. Assim, por exemplo em Data Science Academy (2020) é apresentada uma forma heurística de separação de conjuntos em ordem aleatória, com 70% dos dados para treino e 30% dos dados para validação.

Em relação ao número máximo de tentativas de treino permitidas, tem-se que treinos muito prolongados tendem a memorizar pesos dos valores observados nos dados de treino. Isto se traduz em perda capacidade de generalização da rede, implicando em uma dificuldade para avaliar entradas fora dos dados de treino. Esse fenômeno é conhecido como *overfitting*.

Existe uma outra forma de treinar uma rede MLP, chamada de validação cruzada, que foi apresentada em (KOHAVI, 1995). Essa técnica consiste em intercambiar os conjuntos de treino e de validação em diferentes épocas de treino. Nesse caso, a medida do erro de validação

passa por um processo de avaliação levando em consideração o número de épocas. Avaliando o erro quadrático médio de ambos os conjuntos é possível detectar o início do *overfitting*. Por isso, o ponto ótimo de parada do treinamento está associado ao limite inferior deste erro quadrático médio no conjunto de validação ilustrado na figura 7 como ponto de parada antecipada ao *overfitting*.

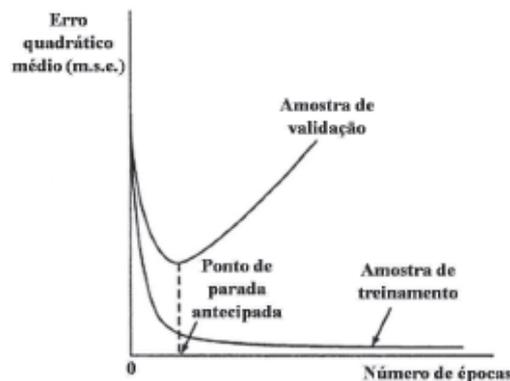


Figura 7: Ponto ótimo de parada da validação cruzada.

Fonte: (HAYKIN, 2007)

### 2.2.3 Rede perceptron Múltiplas Camadas com *Backpropagation*.

O método de aprendizado para redes neurais de múltiplas camadas denominado de *Backpropagation* foi apresentado em (RUMELHART; HINTON; WILLIAMS, 1986), como abreviação de *backward propagation of errors*, em português, retro-propagação de erros. Nesse método, o treinamento é feito em duas fases:

***Feed-forward*** É apresentado um vetor de entrada com vetor de saída conhecido aos neurônios da primeira camada, e calculado um vetor de saída seguindo o fluxo natural das operações na rede.

***Feed-backward*** É calculado o gradiente do erro, para obter informação que induz decréscimo na função, dada pela direção oposta ao gradiente. Com isso, são atualizados os pesos de todas as camadas, começando pela última e seguindo o fluxo inverso da rede.

Essas duas fases são esquematizadas na Figura 8.

Neste método, a medida que o gradiente do erro é calculado das camadas inferiores para as superiores, sua norma decresce com velocidade exponencial. Isto faz que nas camadas mais próximas à entrada os ajustes de pesos sejam pequenos, tornando o aprendizado nelas mais lento. Este problema é conhecido na literatura como *vanishing gradient problem*, o problema

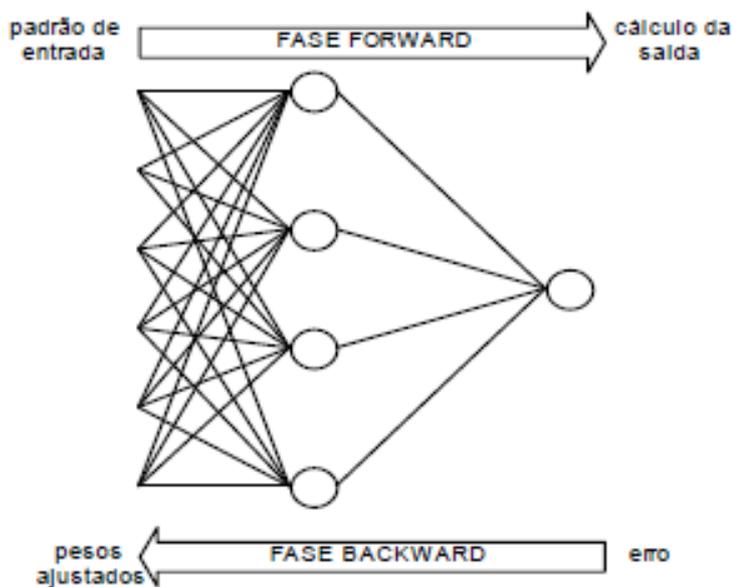


Figura 8: Fases de treino da *MLP-Back-Propagation*.

Fonte: (ALMEIDA, 2013)

de dissipação do gradiente. Usualmente, os valores das taxas de aprendizagem ficam entre 0,2 e 0,8.

Nos treinamentos de redes neurais MLP com *backpropagation*, a validação ocorre só com a fase *feedforward*, obtendo-se os erros quadráticos da camada de saída com o dado de validação observado.

Como o ponto ótimo de parada é um limite inferior, o mesmo somente é descoberto quando superado após algumas épocas de treino, dado que em procedimentos práticos a obtenção de erro é oscilatória e pelo qual é necessário manter salvos os parâmetros obtidos durante estas épocas.

**Otimizador de reajuste dos pesos** São conhecidos vários algoritmos que otimizam a convergência do reajuste dos pesos no treino de *backpropagation*, como os que serão mencionados a seguir. O otimizador Momentum acelera o reajuste dos pesos em busca dos erros globais mínimos, e o RMSProp impede a busca na direção das oscilações. Um terceiro otimizador, denominado Adam pela abreviação de Adaptive Moment optimization, combina essas duas características. Para o algoritmo ADAM, a taxa de aprendizado pode ser arbitrada mas em Brownlee (2020) é mencionado que uma constante com valor 0.001 tem produzido resultados positivos em problemas de predições.

Na Figura 9 é comparado o comportamento de alguns otimizadores. Nesta figura é possível ver que quanto menor o custo de treinamento, maior a velocidade de convergência ao

reajuste ideal dos pesos. Além disso, é visível a vantagem computacional do ADAM frente à outros otimizadores, quando se aumenta o número de iterações.

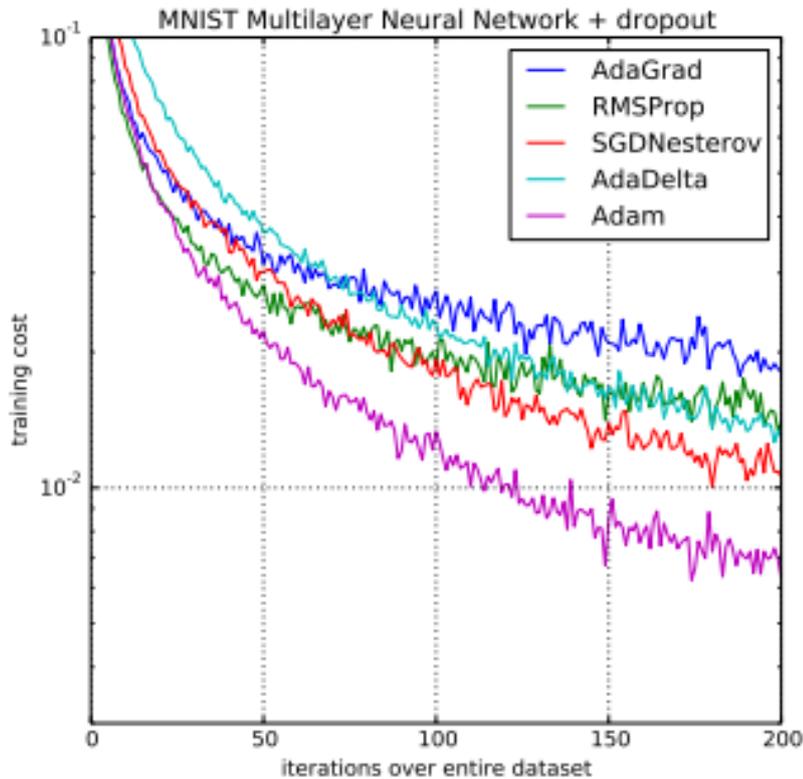


Figura 9: Comportamento de otimizadores para MLP treinadas com *Backpropagation*.

Fonte : (BROWNLEE, 2020)

#### 2.2.4 Redes Recorrentes: O modelo GRU

As redes GRU, abreviação de *Gated Recurrent Unit*, foram apresentadas por primeira vez em (KYUNGHYUN et al., 2014), sendo uma adaptação das redes LSTM (Long Short-Term Memory).

As redes LSTM foram apresentadas em (HOCHREITER; SCHMIDHUBER, 1997), e utilizam blocos de memória chamados células, que permitem que certas informações sejam mantidas na rede. A manipulação da informação é feita por portões (*gates*, pelo qual o procedimento comumente é chamado *gating*). Para estas redes, existem três tipos de portões: i) portão de esquecimento, para remover informações que já não são úteis, ii) portão de entrada, para adicionar de informações úteis ao estado da célula, e iii) portão de saída, para extrair informações úteis do estado da célula.

As redes LSTM permitiram a resolução de problemas mais complexos, porém, apresentavam ainda o problema de dissipação do gradiente, pelo qual a memória não conseguia manter informações de sequências longas, pelo qual era usado o termo de memória de curto prazo.

As redes recorrentes GRU resolveram este problema trocando o uso do estado das células pelo uso de um estado oculto com dois novos portões. Esses portões, chamados de portão de atualização (*update*) e de redefinição (*reset*) decidem quais informações devem ser passadas para a saída e podem ser treinados para manter informações de sequencias longas, sem sofrer dissipação dos valores.

Em (Data Science Academy, 2020), esses portões são citados como as estruturas úteis para solucionar problemas de predições. Na Figura 10 são apresentados dois modelos de redes, um LSTM e um GRU, indicando os portões em cada um delas.

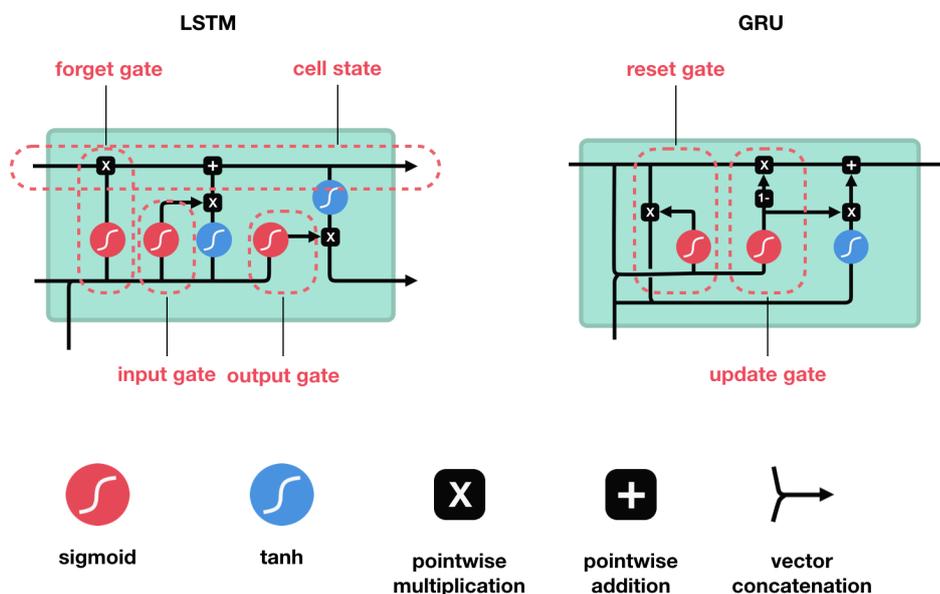


Figura 10: Arquitetura do modelo GRU.

Fonte: (Data Science Academy, 2020)

As redes GRU permitiram então a resolução de problemas com sequencias longas de dados, resolvendo o problema de memória a curto prazo. Então a maior dificuldade que poderia surgir ainda nos treinamentos supervisionados teria a ver com o overfitting, pelo qual foi proposta mais uma ferramenta para evitar que a rede memoriza-se além do desejado: o método *dropout*.

#### 2.2.4.1 Dropout

O método de *Dropout* foi apresentado em (HINTON et al., 2012), termo traduzido ao português na literatura como abandono, e propõe a remoção temporária de alguma célula da rede.

Em (SRIVASTAVA et al., 2014) foi aplicado *dropout* no treino de uma rede neural, no qual, em cada época de treinamento são desligadas aleatoriamente algumas células da camada

de entrada e algumas outras das camadas ocultas, sendo todas ligadas novamente no final da época. Com isso, em cada época do treino apenas uma amostra dos dados é processada por um subconjunto das células ocultas.

Esta técnica procura que a aleatoriedade da escolha de células em cada época induza uma redução na dependência entre elas as células no processo de ajuste, fazendo que cada unidade gere padrões que não dependam dos aprendidos pelas outras. No momento do teste, todos os pesos são multiplicados pela probabilidade da sua célula ter sido desligada.

Em (SRIVASTAVA et al., 2014) são obtidos bons resultados no treino da rede com *dropout* quando, em cada época, são desligadas 50% das células em camadas ocultas e 20% das células na camada de entrada.

## 3 Trabalhos relacionados

Este capítulo descreve os principais trabalhos relacionados encontrados na literatura referenciada neste trabalho. A primeira seção cita a primeira referência literária encontrada na execução deste trabalho que realiza comparações de métodos de previsão de demanda. A segunda seção cita pesquisas relacionadas ao tema deste trabalho, a predição de demanda em restaurantes universitários, contendo estudos relacionados com uma parte dos métodos executados neste trabalho que são os modelos de redes neurais MLP, e o diferencial deste trabalho em relação aos outros é a inclusão de redes neurais recorrentes modernas, denominadas GRU, fundamentadas no tópico 2.2.4. A última seção cita o maior volume de referências encontradas durante as pesquisas de predição de demanda em geral, e que não corresponderam ao tema deste trabalho.

### 3.1 Trabalho de comparações de métodos de previsão de demanda.

Junior (2007) Realiza um trabalho de comparação entre os métodos estocásticos (Método de suavização exponencial, modelos de Box-Jenkins) e modelos de aprendizado de máquina (Redes Neurais), ilustrados na Figuras 1, os quais são usados para a previsão da demanda de produtos cosméticos distribuídos em séries temporais. Entre as Redes Neurais, encontramos redes do tipo *feedforward* com o algoritmo de treino por *backpropagation* que foi o principal foco no trabalho de previsão do R.U na Universidade Federal de Viçosa e na Universidade Estadual Paulista Júlio de Mesquita Filho, e que também fundamentou parte do desenvolvimento deste trabalho de predição no ICT Unifesp. Neste trabalho do autor, também são analisadas diversas medidas de performance preditivas e é feito uma análise comparativa final destas medidas entre os métodos citados.

### 3.2 Previsão de demanda em restaurantes universitários

No estudo estatístico feito por Landim (2016), foi analisada a correlação entre a temperatura e o consumo de refeições nos dias de vendas do restaurante universitário do campus ICT-Unifesp, sendo que os dados continham apenas uma pequena amostra das vendas do segundo semestre de 2016. Devido ao baixo volume de ocorrências, os dados foram submetidos à reamostragem via *bootstrap*. De acordo com os gráficos das amostras, identificou-se que a correlação mostrada nos gráficos da primeira metade do semestre e do período total do semestre formaram distribuições bimodais. Porém, na segunda metade do semestre formou-se uma

distribuição unimodal. Portanto, concluiu-se que outras variáveis e outros modelos de análises deveriam ser utilizados para esta previsão de demanda.

Lopes (2008) faz o mesmo estudo deste cenário do ICT-Unifesp aplicado na Universidade Federal de Viçosa (UFV). Neste estudo, os dados utilizados foram somente o histórico de vendas do restaurante universitário, e nenhuma variável de ambiente foi coletada como temperatura, precipitação, número de alunos matriculados, etc. O algoritmo utilizado foi o *Traincgp* (*Conjugate gradient backpropagation with Polak-Ribiere updates*) no software Matlab. Este algoritmo não envolve o cálculo das derivadas segundas das variáveis e converge ao mínimo da função quadrática em um número finito de iterações como cita o autor. Foram então considerados para cada nó da rede neural, o dia da semana (como segunda, terça, quarta, quinta e sexta) e cada camada dessa rede utilizando os 5 dias anteriores para cada nó (as 5 segundas anteriores, 5 terças anteriores e assim sucessivamente) e, por fim, obtido um modelo pela rede que apresentou erro máximo de 3. A rede neural aplicada neste trabalho é representada na Figura 11.

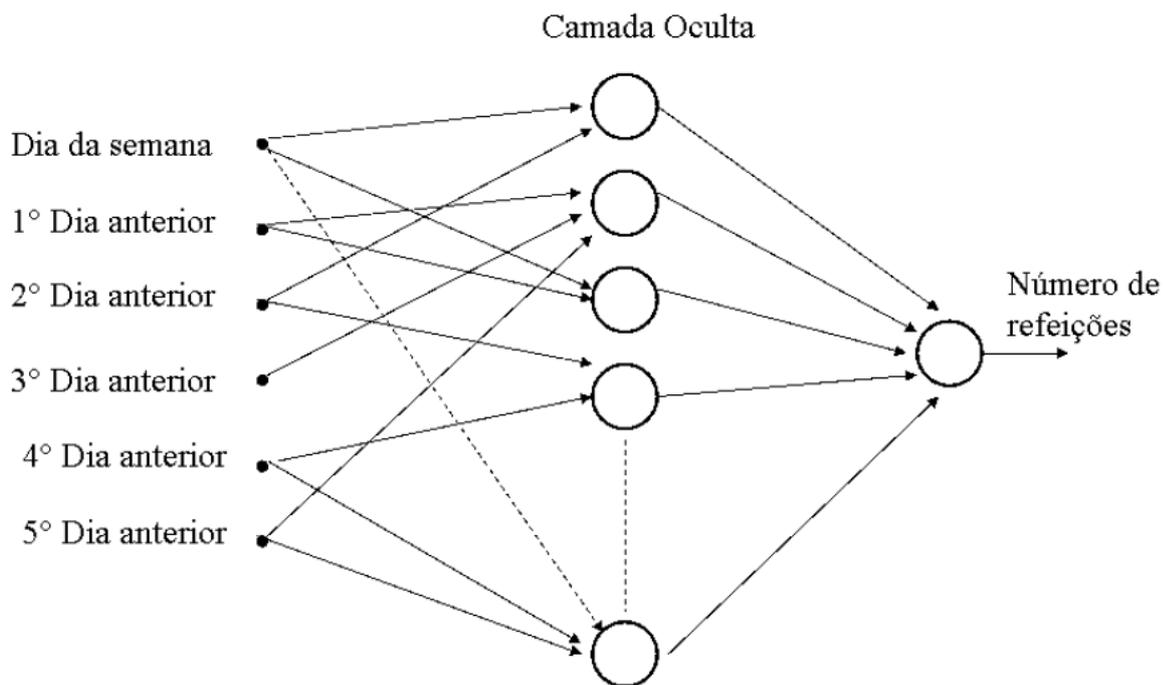


Figura 11: Rede Neural Perceptron de Múltiplas Camadas.

Fonte:(LOPES, 2008).

Rocha, Matos e Frei (2011) também realiza o estudo de demanda no restaurante universitário da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), novamente com os métodos de redes neurais artificiais com backpropagation e utilizando apenas como fonte de dados (o histórico numérico das vendas realizadas), e outras variáveis intermediárias obtidas a partir deste, como médias de subconjunto de observações (médias de segundas-feiras). A única variável de ambiente coletada foi o número de feriados próximos à observação de venda. No

estudo do total de dias analisados, verifica-se que em 73% (187 dias), o método de média simples propiciou um maior erro em relação à RNA, que por sua vez ocasionou um erro maior nos 23% (69 dias) restantes. Em se tratando de menor desperdício, observa-se que a RNA apresenta erros maiores que 50 refeições em 13 dias, enquanto o método da média simples apresenta erros maiores que 50 refeições em 58 dias, concluindo-se então que o método de RNA foi bem mais eficiente do que o cálculo de média simples utilizado pela administração do restaurante universitário. A Figura 12 apresenta um esquema da rede neural aplicada nesse trabalho.

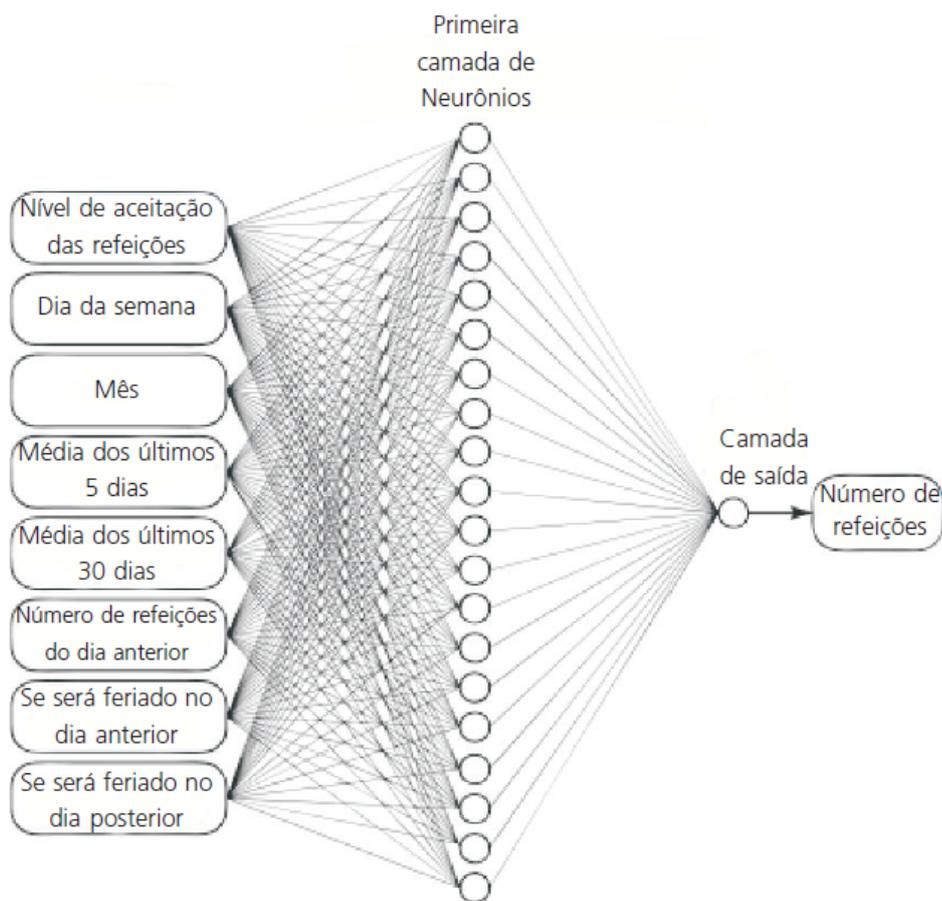


Figura 12: Rede Neural Perceptron de Múltiplas Camadas

Fonte: (ROCHA; MATOS; FREI, 2011)

Tanto o modelo apresentado em (ROCHA; MATOS; FREI, 2011) quando em (LOPES, 2008) possuem uma única camada oculta.

### 3.3 Previsão de demanda em outros ambientes

Ruas (2012) faz uma análise de previsão de demanda de energia elétrica no estado do Paraná, entre os anos de 2004 e 2006, utilizando redes neurais artificiais e máquinas de veto-

res de suporte. Apesar de não ser o mesmo exemplo do cenário do restaurante universitário do ICT-Unifesp, temos a distribuição dos dados de consumo coletados como uma série temporal. Nesta pesquisa de previsão de demanda de energia elétrica foi utilizada uma rede parcialmente recorrente de Elman, que permite a previsão de um passo de tempo à frente. Para que seja possível realizar a previsão para vários pontos à frente, é necessário utilizar os valores já previstos, ou seja, a saída da rede, como entradas da mesma.

Almeida (2013) analisa um cenário semelhante de demanda de energia elétrica, porém utilizando-se técnicas de previsão de demanda com Rede Neural Artificial do tipo Multilayer Perceptron combinado com lógica fuzzy que permite colocar variáveis de temperatura (entre outras) em um conjunto de regras que impactam no problema.

Silva (2010) também aplica técnicas de redes neurais para previsão de demanda de energia elétrica, com o estudo de variáveis climáticas, porém através de um modelo de MAPA SOM - (Self-Organizing Map) que é um tipo de rede neural desenvolvido para reconhecimento de padrões. Apesar de ser um modelo não supervisionado, o modelo é ideal para organizar as principais variáveis impactantes e descartáveis na previsão. O mapa som utilizado pelo autor apresenta os dados associados aos seus neurônios de forma que padrões similares encontram-se em neurônios contíguos, tendo uma organização topológica. Deste modo é possível se extrair relações abstratas entre as variáveis do vetor de dados através da sua posição nos mapas componentes, que por meio de uma escala de cores mostram a quantidade de uma variável específica em cada neurônio do mapa.

## 4 Metodologia

Neste Capítulo é descrita a metodologia experimental deste trabalho, a qual consiste nos seguintes passos:

- Coleta de dados endógenos e exógenos.
- Transformação de cada registro de dado endógeno (os dados de consumo e vendas), em uma série temporal com intervalo de cinco dias anteriores.
- Análises exploratórias dos conjuntos de dados endógenos e exógenos com o conjunto de dados a serem previstos.
- Construção e treino dos modelos exclusivamente endógenos e dos modelos mistos, duplicados em duas fases experimentais com diferentes domínios temporais.
- Análises comparativas dos resultados dos modelos.

### 4.1 Área de estudo

A área de estudo deste trabalho é o restaurante universitário do Instituto de Ciência e Tecnologia da Unifesp (ICT-Unifesp) em São José dos Campos. O campus do ICT foi inaugurado no ano de 2007 visando suprir as demandas científicas e tecnológicas da região do Vale do Paraíba. A Figura 13 apresenta um dia comum de utilização do espaço físico do restaurante universitário do ICT.



Figura 13: Restaurante Universitário do ICT-Unifesp

## 4.2 Descrição dos dados

Este trabalho fará uso de dois tipos de dados, endógenos e exógenos, os quais são descritos a seguir.

Dados endógenos: são os dados do domínio de predição neste caso. Para este problema, são as quantidades diárias de refeições consumidas no almoço e no jantar do Restaurante Universitário (RU) do ICT-Unifesp, quantificada diariamente pela número de alunos que passam no ponto de acesso, a catraca. Também são considerados dados endógenos a quantidade diária de *tickets* de refeição vendidos pelo Restaurante. Em ambos casos, as informações são tidas dos dias letivos. Esses dados são transformados em entrada de redes neurais, em formato de série temporais.

Dados exógenos: são todos os outros dados fora do domínio de predição. Para este problema, são parâmetros derivados das datas das observações, como por exemplo o dado categórico que representa o dia da semana (segunda-feira à sexta-feira), e os dados climáticos.

## 4.3 Obtenção e tratamento dos dados

A obtenção dos dados é realizada por meio de duas fontes distintas, sendo os dados endógenos inteiramente fornecidos através do setor de tecnologia da informação do ICT Unifesp, parte dos dados exógenos da data de registro dos dados endógenos coletados, e a parte restante dos dados exógenos, os dados climáticos, são obtidos através de uma estação meteorológica mais próxima ao ICT Unifesp, localizada na cidade de Taubaté-SP.

### 4.3.1 Dados endógenos

Os dados históricos de consumo no restaurante foram retirados do atual sistema banco de dados de refeições subsidiadas do Hospital São Paulo, que gerencia os dados dos refeitórios de todas as unidades da Unifesp. Apenas alguns funcionários autorizados têm acesso ao banco de dados da instituição, portanto para coletar tais dados o presente trabalho obteve uma autorização com a direção do campus ICT-Unifesp. Foram solicitados os dados de consumo apenas para os discentes de graduação, visto que o banco de dados ainda possui as informações de consumo de docentes, alunos de pós-graduação e visitantes, porém claro com menor relevância em termos quantitativo. Além disso, o padrão de consumo destes outros estratos do meio acadêmico pode influenciar o processo de predição das demandas trazendo tendências diferentes. Na tabel 1.

Após a coleta, os dados de consumo do restaurante foram transformados em um processo de aproximação por uma série temporal, para um intervalo de cinco dias, e em cada regis-

Tabela 1: Formato original dos dados originais obtidos pelo restaurante universitário

DATA	(19/12/2017)	(18/12/2017)
VENDAS CAFÉ	0	0
VENDAS ALMOÇO	24	71
VENDAS JANTAR	0	0
VENDAS REFEIÇÃO	24	71
TOTAL VENDAS	24	71
ENTR. CAFÉ	0	0
ENTR. ALMOÇO	42	70
ENTR. JANTAR	3	24
TOTAL ENTR. REFEIÇÃO	45	94
TOTAL ENTRADA	45	94

tro de venda acrescentados cinco novos atributos contendo os valores passados, deste mesmo atributo, em um intervalo de cinco dias anteriores. Este processo adapta o conjunto de dados para o processo de memorização das entradas, estruturando o formato compatível de leitura de dados nos modelos de redes neurais aplicados.

A Tabela 2 exemplifica a nova estrutura de um registro de dado do restaurante, com um intervalo temporal de cinco dias anteriores. Nota-se que o valor de consumo da data 20/04/2017 foi removido do conjunto de dados, por se tratar o valor supervisionado a ser previsto, dado que o processo de aprendizado das redes neurais utilizam apenas dados no passado, a partir de um dia anterior.

DATA	(19/12/2017)
1 DIA ANTERIOR	500
2 DIAS ANTERIORES	00
3 DIAS ANTERIORES	300
4 DIAS ANTERIORES	200
5 DIAS ANTERIORES	100

Tabela 2: Transformação dos registros do restaurante em uma série temporal.

### 4.3.2 Dados exógenos

Os dados exógenos correlacionados com o consumo se dividem em dois tipos principais, dados climáticos coletados de estações meteorológicas próximas ao ICT-Unifesp, e dados derivados das datas dos registros de consumo.

Em relação as variáveis climáticas utilizadas como dados exógenos, foram considerados parâmetros que possam afetar o consumo de refeições de forma indireta, como temperatura média ambiente, pressão atmosférica, umidade e velocidade do vento. Tais parâmetros podem ser obtidos de forma gratuita pelo BDMEP - Banco de Dados Meteorológicos para Ensino e Pesquisa, pertencente à instituição pública INMET - Instituto Nacional de Meteorologia, per-

tendente ao Ministério da agricultura, pecuária e abastecimento. É necessário um cadastro na plataforma do INMET<sup>1</sup> para a obtenção dos dados. A instituição contém dados registrados de forma digital desde 1961 no país inteiro, os dados históricos referentes a períodos anteriores a 1961 ainda não estão em forma digital e, portanto, estão indisponíveis no BDMEP. Importante ressaltar que o BDMEP leva 90 dias para registrar cada nova data.

Além dos dados ambientais, também foram gerados dados exógenos a partir dos dados de consumo coletados. A informação de data contida nos índices dos registros dos dados endógenos, foi derivada em diversas informações que representam o comportamento de consumo em relação à sazonalidade da frequência dos alunos influenciada pelas agendas de atividades acadêmicas.

Os seguintes parâmetros foram definidos:

- Semestre 1 ou 2 em formato categórico e binário;
- Dia da semana em formato categórico e binário;
- Distância em dias até o registro anterior e posterior;
- Avanço do semestre em escala percentual;
- Avanço do mês em escala percentual.

O consumo distribuído em uma janela de cinco dias para entrada nas redes MLP seguiu padrão semelhante com os trabalhos de previsão de demanda em R.U realizados por Lopes (2008) e Rocha, Matos e Frei (2011), ilustrado na Figura 11. Por fim, a Tabela 3 representa o conjunto de dados estruturados e preparados para o processo de divisão em domínios de treino, validação e teste para o treino dos modelos.

Por fim, a Tabela 3 representa o conjunto de dados estruturados e preparados para o processo de divisão em domínios de treino, validação e teste para o treino dos modelos.

### 4.3.3 Pré-processamento

Na etapa de pré-processamento é realizada a transformação dos dados endógenos em séries temporais, de comprimento de cinco dias, normalização com remoção de *outliers*, e aplicação de escala 0 a 1, para que todos os dados correspondam à um mesmo domínio de aprendizado. Após a conclusão destas etapas, o conjunto de dados foi preparado para as fases experimentais 1 e 2, que realizaram uma divisão do conjunto final de dados em intervalos temporais distintos. Apesar da distribuição dos dados se apresentarem em datas em função do

<sup>1</sup> <http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>

Estrutura final do conjunto de dados indexados por data:		
identificador	nome da variável	tipo de variável
0	SEMESTRE_1	int64
1	SEMESTRE_2	int64
2	SEGUNDA	int64
3	TERCA	int64
4	QUARTA	int64
5	QUINTA	int64
6	SEXTA	int64
7	DISTANCIA_DIA_ANTERIOR	int64
8	DISTANCIA_DIA_POSTERIOR	int64
9	PERC_CONCLUSAO_SEM	float64
10	PERC_CONCLUSAO_MES	float64
11	PRESSAO_ATMOSFERICA	float64
12	TEMPERATURA	float64
13	UMIDADE	int64
14	VENTO	float64
15	VENDAS_ALMOCO	int64
16	VENDAS_ALMOCO_1	int64
17	VENDAS_ALMOCO_2	int64
18	VENDAS_ALMOCO_3	int64
19	VENDAS_ALMOCO_4	int64
20	VENDAS_ALMOCO_5	int64
21	ENTR_ALMOCO	int64
22	ENTR_ALMOCO_1	int64
23	ENTR_ALMOCO_2	int64
24	ENTR_ALMOCO_3	int64
25	ENTR_ALMOCO_4	int64
26	ENTR_ALMOCO_5	int64
27	ENTR_JANTAR	int64
28	ENTR_JANTAR_1	int64
29	ENTR_JANTAR_2	int64
30	ENTR_JANTAR_3	int64
31	ENTR_JANTAR_4	int64
32	ENTR_JANTAR_5	int64

Tabela 3: Estrutura final do conjunto de dados indexados por data

tempo se classificando em um modelo de série temporal, assume-se a hipótese que o comportamento dos mesmos também é impactado por relações causais com outros variáveis exógenas, como recesso acadêmico, feriados, eventos, precipitações intensas que causam trânsito local e impactam na logística e frequência do público, entre outras variáveis de causas menos aparentes

#### 4.3.4 Tratamento dos dados para entrada nos modelos

Os dados endógenos, após estruturados na tabela final do conjuntos de dados, passam pelas seguintes transformações:

- Cálculo do desvio padrão de cada vetor de atributos, e normalização dos valores máximos para o teto de 3x o desvio padrão, e mínimo de 0;
- Transformação dos dados em escala de 0 e 1.

Os dados exógenos não passam pela transformação em série temporal, portanto os mesmos foram tratados de acordo com os passos:

- Transformação dos dados em escala de 0 e 1;
- Os parâmetros categóricos binários (dias da semana e semestre) já estão escalados por serem categorias binárias.

#### 4.3.5 Fases Experimentais

O processo experimental foi realizado em dois roteiros distintos de divisão do domínio temporal do conjunto de dados, e os resultados obtidos entre as duas fases foram comparados.

O conjunto de dados contemplando o período de 2017 a 2019, foi dividido em conjunto de treino, validação e teste da seguinte maneira:

**1º Fase com validação no 1º semestre de 2018 e teste no 1º semestre de 2019** Neste roteiro, o semestre de validação que compõe o conjunto de dados para o treino *backpropagation* das redes neurais contempla o primeiro semestre de 2018 e o conjunto de teste contempla o primeiro semestre de 2019. Os dados de 2017 contemplando o 1º e 2º semestre, e 2018 contemplando o 2º semestre, foram usados para treino. Os resultados obtidos nesta divisão foram usados para validar a hipótese de que os modelos aprendem especificamente a sazonalidade de consumo no primeiro semestre, se saindo melhor nos testes realizados no primeiro semestre de 2019, em comparação aos outros modelos treinados com validação no ano todo de 2018. Portanto, o conjunto de dados da primeira fase contempla o seguinte domínio:

- Conjunto de treino dos modelos, contemplando o primeiro e segundo semestre de 2017, e segundo semestre de 2018;
- Conjunto de validação dos modelos, contemplando o primeiro semestre de 2018;
- Conjunto de teste dos modelos, contemplando o primeiro semestre de 2019.

**2º Fase com treino em 2017, validação em 2018 e teste em 2019** Nesta fase, os conjuntos foram divididos conforme sua descrição e o melhor modelo encontrado passa por uma última etapa de teste no domínio da primeira fase (teste somente no primeiro semestre de 2019). As métricas obtidas neste teste foram comparadas com o melhor modelo da primeira fase.

## 4.4 Definição e treino dos modelos

No conjunto de dados deste trabalho, os dados obtidos se dividem em dados temporais e endógenos (tal que cada registro de consumo e venda trás a informação de seu domínio em um intervalo temporal de cinco dias anteriores) e dados discretos e exógenos, sendo variáveis categóricas de data para cada registro, e variáveis climáticas.

Portanto foi necessária a implementação de modelos específicos para entradas temporais e modelos específicos para as entradas discretas. Para a saída final foi implementado um comitê de redes neurais endógenas e exógenas, com um neurônio perceptron na saída, recebendo os dois valores dos modelos endógenos e exógenos para a regressão das saídas das duas redes ao valor que será a predição do consumo.

### Modelos endógenos

- Desenvolvimento das redes perceptron de baixa profundidade para avaliar o aprendizado da rede;
- Aumento da profundidade da rede e avaliar as mudanças da função de perda RMSE;
- Implementação e avaliação dos modelos com redes recorrentes GRU, conforme a Figura. 10 que são especialmente desenvolvidos para o aprendizado com memorização de dados, e no caso deste trabalho, podem memorizar as sazonalidades semanais de consumo (em um intervalo de cinco dias).

### Modelos Mistos : Endógenos e Exógenos

- Para os dados temporais (consumo e venda) utilizou-se os melhores modelos endógenos dos experimentos anteriores para as entradas endógenas.
- Para os dados discretos e categóricos adaptou-se a entrada destes dados para rede perceptron
- Concatenou-se a saída das duas redes neurais em um perceptron criando um comitê de redes neurais para obter a saída final prevista.

#### 4.4.1 Hiper parâmetros : Função de ativação e otimizador

Conforme fundamentado no Capítulo 2 na Seção 2.2.1, a função de ativação dá a capacidade do perceptron, quando conectado em rede, de resolver problemas lineares e não lineares, agregando adaptação e improviso ao resolver programas que não estão contidos em seus dados de alimentação. Portanto para as camadas ocultas das redes neurais MLP desenvolvidas será

aplicada a função ReLu e para o neurônio de saída será aplicada a função linear, e o otimizador de treino realiza a função de otimizar o tempo de convergência do reajuste dos pesos à valores ideais, sendo escolhido o otimizador ADAM com taxa de aprendizado definido em 0,001.

## 4.5 Teste e Métricas de avaliação

As principais métricas de avaliação dos modelos são a Raiz do Erro Quadrático Médio ou em inglês *Root Mean Squared Error* (RMSE), O coeficiente de correlação de Pearson (R), e o coeficiente "chi-quadrado" definido como  $R^2$ . Estas métricas estatísticas foram utilizado nas etapas de teste para avaliar a proximidade das predições do modelo com o comportamento real de consumo. As equações 4.1 e 4.2 apresentam a formulação para o RMSE e o coeficiente de Pearson (R), respectivamente.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^{est} - x_i^{obs})^2} \quad (4.1)$$

$$R = \frac{\frac{1}{n} \sum_{i=1}^n (x_{est} - \bar{x}_{est}) * (x_{obs} - \bar{x}_{obs})}{\sigma_{est} * \sigma_{obs}} \quad (4.2)$$

Onde “est” são os valores estimados; “obs” são os valores reais; n é o numero de amstras ;  $\sigma$  é o desvio padrão; R é a correlação linear;  $\bar{x}$  é a média de x. Avaliou-se também os erros positivos e negativos entre os valores previstos e reais, para representar quantas refeições seriam descartadas e quantas estariam em falta se a produção de refeições fosse de acordo com as predições do modelo.

## 5 Resultados

Este Capítulo descreve os principais resultados experimentais obtidos durante esta pesquisa. Como descrito no Capítulo 4, os experimentos foram conduzidos em duas fases. Contudo, optou-se por apresentar neste Capítulo, apenas os principais resultados obtidos. Os demais resultados estão disponíveis nos Anexos deste documento. As primeiras seções introduzem a organização dos dados, uma breve análise das variáveis e o protocolo experimental, respectivamente.

### 5.1 Organização do conjunto de dados

O procedimento de coleta de dados foi realizado de acordo com a metodologia da Seção 4.3.1 para os dados endógenos, e de acordo com os passos da Seção 4.3.2 foram obtidos os dados exógenos. Ambos os conjuntos de dados coletados foram estruturados conforme a Tabela 3, contendo um intervalo temporal de registros, desde o 12 de abril de 2017 (2017-04-12) para o primeiro registro até o 16 de dezembro de 2019 (2019-12-16) para o último registro, totalizando 514 registros de consumo de refeições em dias letivos.

Conforme a metodologia definida na Seção 4.3.5 este conjunto de dados com o total de 514 registros, foi duplicado para 2 fases experimentais distintas, cada fase com uma organização específica do conjunto de dados. O conjunto de dados da primeira fase experimental foi organizado de acordo com a Figura 14. Esta fase tem o conjunto de validação contemplando exclusivamente o primeiro semestre de 2018, indicando que o primeiro semestre de 2018 pudesse apresentar um movimento de consumo e vendas semelhante ao primeiro semestre de 2019, sendo a ideal para testes envolvendo somente o primeiro semestre do conjunto de testes.

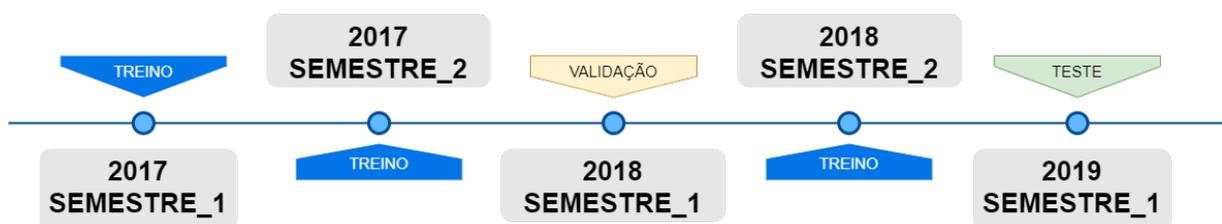


Figura 14: Domínio temporal da fase 1.

Para a segunda fase o conjunto de dados foi organizado de acordo com a Figura 15. Para esta fase o conjunto de validação selecionado contemplou o ano letivo de 2018 por completo, já para os experimentos de teste dos modelos os dados selecionados contemplaram o todo o ano letivo de 2019.

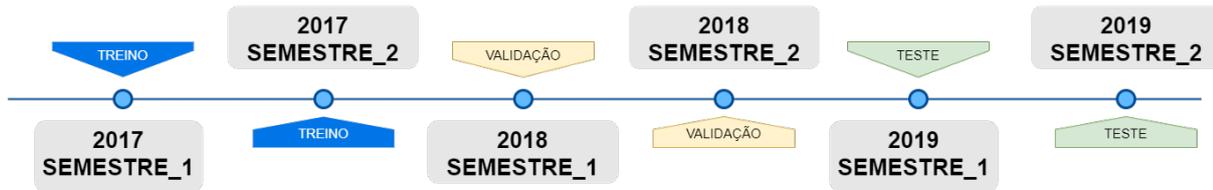


Figura 15: Domínio temporal da fase 2

### 5.1.1 Manipulação e pré-processamento do conjunto de dados

Buscando organizar os dados brutos obtidos para a sua posterior aplicação nos modelos foram encontrados algumas peculiaridades. A primeira dificuldade encontrada nos experimentos foi um comportamento anômalo do resultados de previsão para o modelo RNN\_ENDO\_2. A linha azul na Figura 16 representa uma previsão refeições do modelo RNN\_ENDO\_2, e a linha vermelha valores reais de consumo do primeiro semestre de 2019. Com isto, foi possível observar que em ambos os conjuntos (real e predito) este comportamento estava presente.

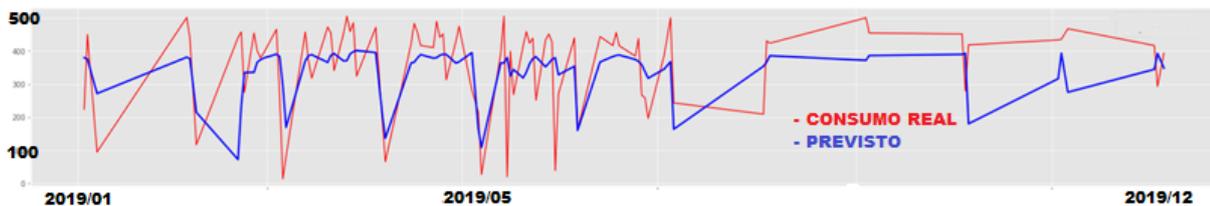


Figura 16: Resultado do modelo RNN\_ENDO\_2 obtido sobre o conjunto de dados aleatoriamente ordenado sobre o tempo.

Após uma análise exploratória foi descoberto que os registros continham um erro na indexação por data, onde a estampa de datas se apresentava trocada, ou seja os dias por meses e vice-versa. Após a correção deste problema indexação os dados de consumo real e previsão produziram resultados realísticos e dentro do formato esperado, conforme apresentado na Figura 17, que apresenta um exemplo de predição pelo modelo RNN\_ENDO\_2 e os dados reais para o período do almoço.

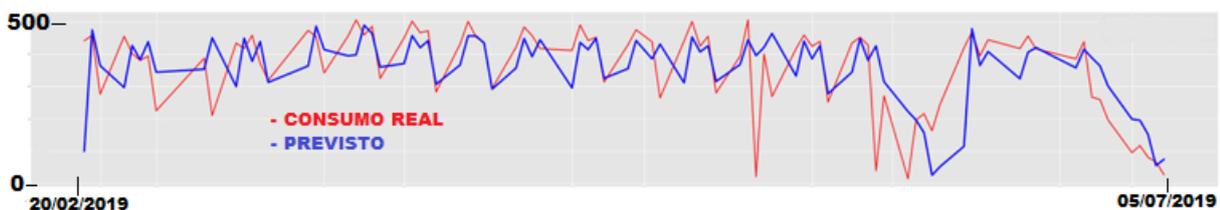


Figura 17: Resultado do modelo RNN\_ENDO\_2 obtido sobre o conjunto de dados com ordenação corrigida

## 5.2 Avaliação das Variáveis

Nessa Seção serão feitos alguns comentários em relação as características das variáveis que foram mais importantes para o problema, e serão apresentados alguns gráficos dos estudo estatístico feito para avaliar as relações entre elas.

### 5.2.1 Estimativas de consumo do restaurante

A análise da técnica de estimação de consumo, realizada de forma subjetiva em relação ao consumo da semana anterior, utiliza-se do cálculo de 30% de produção acima do consumo do quinto dia anterior. Este método de estimativa é foi adotado para tolerar descartes devido à existência de multa contratual por falta de refeições. Ainda, é possível observar que este modelo de 30% a mais produz um comportamento linear, representado pela linha azul na Figura 18, estando distante do comportamento real de consumo, indicado pela linha vermelha.

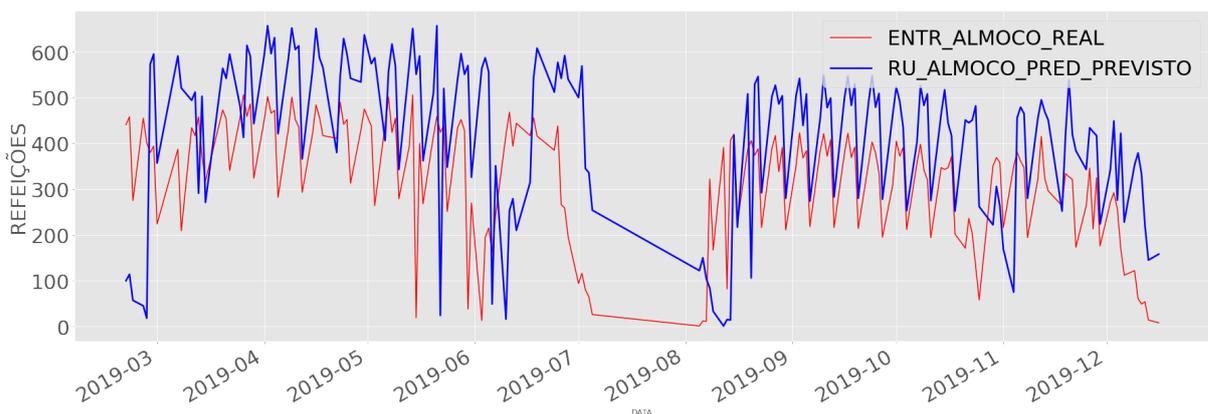


Figura 18: Estimativa do restaurante para o ano de 2019.

Apesar da estimativa seguir as tendências de quedas e aumento de consumo, a Figura 19 apresenta a dispersão gerada entre a estimativa do R.U e o consumo real no ano de 2019, demonstrando que a regressão linear (linha vermelha) tem o eixo totalmente descentralizado com a função identidade da estimativa ideal (representada pela diagonal imaginária formada entre a origem do gráfico e o vértice superior direito).

Assim esse formato de predição gera também um erro maior do que 30% no somatório total de refeições descartadas no semestre, ocasionado pelo comportamento oscilatório do consumo, conforme a Tabela 4.

### 5.2.2 Análise das variáveis endógenas

As variáveis endógenas são os parâmetros temporais de entrada nos modelos MLP e GRU, correspondentes ao domínio de consumo no restaurante.

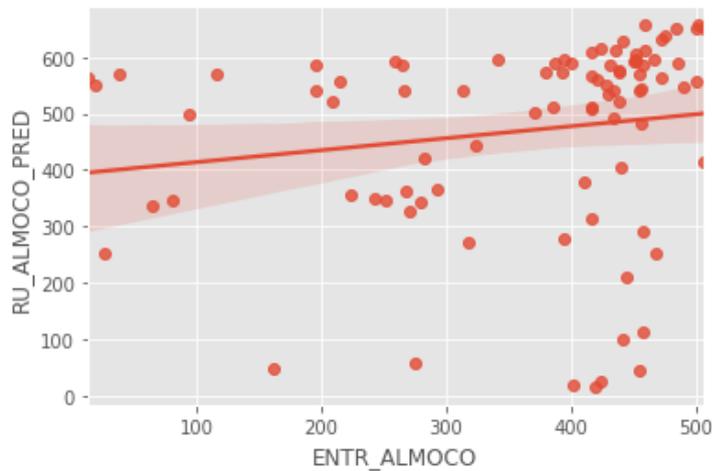


Figura 19: Gráfico de dispersão da estimativa de consumo do restaurante para o ano de 2019.

Consumo com margem 30% acima do 5o dia anterior	
TOTAL DE REFEIÇÕES CONSUMIDAS	58653
TOTAL DE REFEIÇÕES ESTIMADAS	76262
CORRELAÇÃO (r)	0.4006
P-value	2.0845e-08
RMSE	191.7620
SOMA DOS ERROS POSITIVOS	23412
SOMA DOS ERROS NEGATIVOS	-5803
ERRO ABSOLUTO MEDIANO	133.0
ERRO ABSOLUTO PERCENTUAL MEDIO	205.6113%

Tabela 4: Métricas da estima de consumo do restaurante para o ano de 2019

### 5.2.2.1 Consumo do dia vigente em relação às vendas de tickets do dia anterior

É possível notar na Figura 20 que as vendas de *tickets* no período de almoço apresentaram comportamento diferente no ano de 2017 em comparação aos anos seguintes devido à uma limitação imposta pelo restaurante, a partir de 2018 os alunos poderiam adquirir apenas 2 tickets por dia. Possivelmente, esta limitação foi dada para aproximar o comportamento de consumo de 1 ou 2 dias seguintes à venda do ticket. Esta limitação pode ser interpretada como método de auxílio à gestão para a produção de refeições e para o tratamento de desperdício.

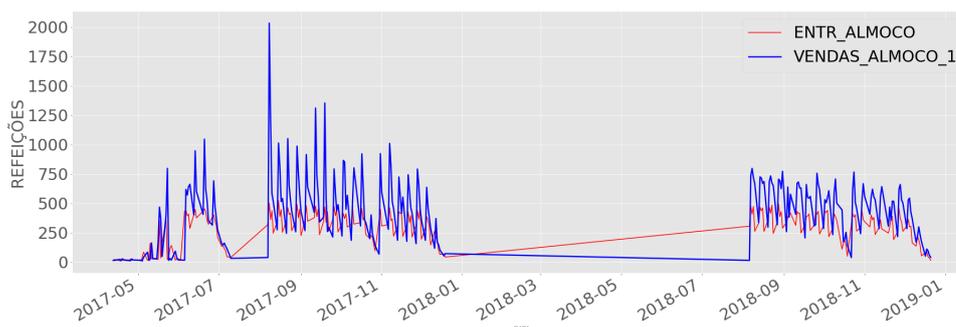


Figura 20: Correlação entre consumo e vendas de almoço.

Mesmo com o valor *outlier* de 2000 vendas em um único dia, e com a nova limitação de compras de *tickets* a partir de 2018, o consumo no horário de almoço está fortemente relacionado com as vendas de *tickets* no período do almoço de um dia anterior. Nota-se também que os alunos se adaptaram à limitação imposta para utilização dos *tickets* com prazo de validade de apenas dois dias, conforme o valor do coeficiente de correlação é de aproximadamente em 72%, como apresentado na Tabela 5.

Tabela 5: Comparação de consumo com um dia anterior

CONSUMO EM RELAÇÃO ÀS VENDAS DE 1 DIA ANTERIOR	
CORRELAÇÃO (r)	0.7255528038157009
P-value	5.399561176138223e-41
RMSE	260.5399426736619
ERRO ABSOLUTO MEDIANO	139.0
ERRO ABSOLUTO PERCENTUAL MEDIO	90.18

Há outros fatores não previstos envolvidos, como possíveis falha de registros de vendas no sistema, bem como o valor *outlier* de 2000 vendas pode ser interpretado com a migração de sistema e banco de dados de refeições que ocorreu em 2017 da unidade Talim do ICT-Unifesp para o banco de dados do Hospital São Paulo. Possivelmente também foram importadas vendas do sistema antigo sem a diferenciação de datas.

A soma das vendas de refeições no horário do almoço, obteve um total de 242282 para todos os 514 registros no conjunto de dados. Neste mesmo período o valor real de consumo, ou seja alunos que efetivamente ingressaram no restaurante (passaram pela catraca) totalizou 163752 refeições. Apesar de notória, a diferença de 78.530 refeições vendidas acima do consumo real não foi passível de investigar neste trabalho. Ressaltando que estes valores foram obtidos no conjunto original de dados fornecidos pelo fiscal de contrato do R.U do ICT-Unifesp por meio de solicitação via e-mail para este fiscal.

### 5.2.2.2 Normalização e escala de *features*

O processo de normalização e escala é demonstrado nesta Seção com a *feature* de vendas de *tickets* de 1 dia anterior, pois entre todas esta é a que produziu *outliers* de maior destaque. A normalização dos dados é feita com o teto de 3x o desvio padrão médio, logo o pico de 2000 vendas foi normalizado para o valor arredondado de 1356 refeições. Mesmo com a normalização, o comportamento linear desta *feature* se manteve, conforme apresentado na Figura 22.

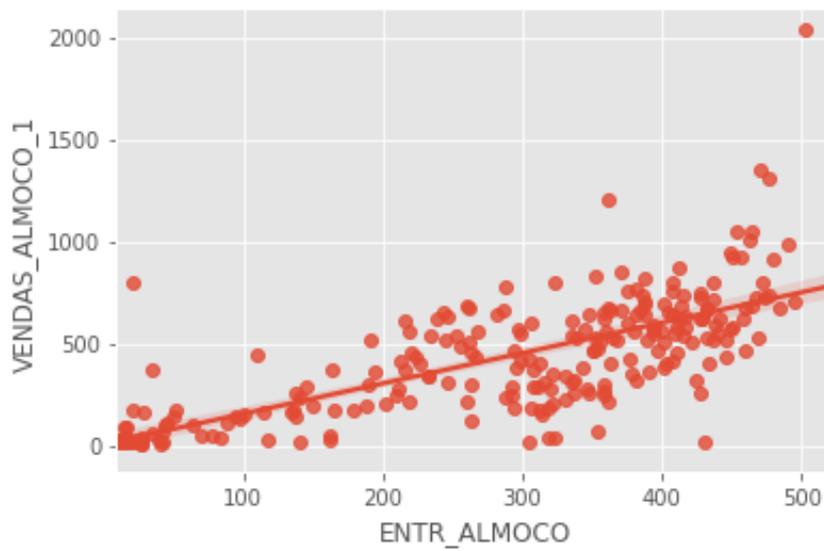


Figura 21: Gráfico de dispersão entre consumo e vendas de almoço.

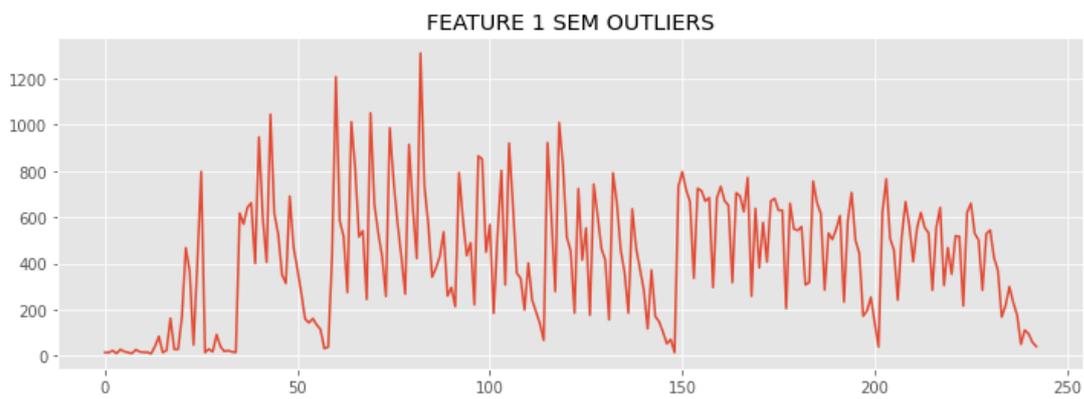


Figura 22: Vendas de *tickets* normalizados com teto de 3x o desvio padrão.

Após a normalização foi realizada a padronização da escala de 0 a 1 nesta e conforme é observado na Figura 23, o comportamento linear foi novamente mantido.

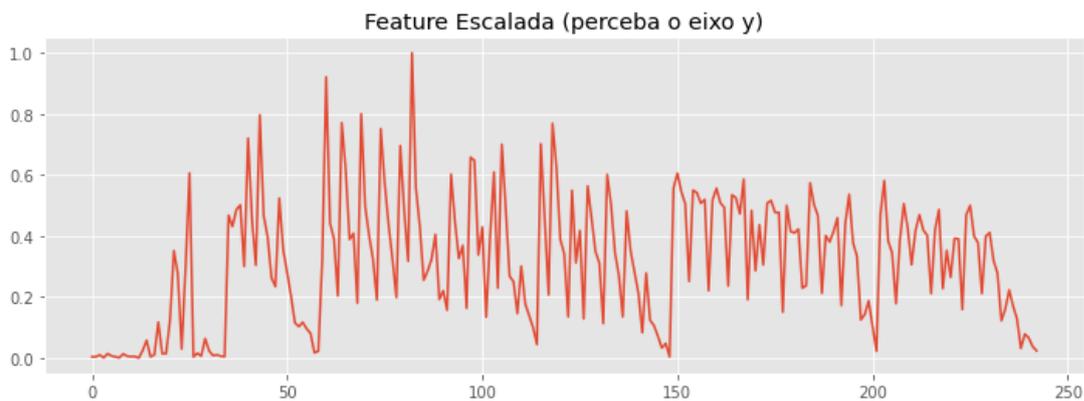


Figura 23: Vendas de *tickets* escalada entre 0 a 1.

Este processo de normalização e escala foi realizado para todas as métricas endógenas e também para as métricas climáticas.

### 5.2.2.3 Consumo atual em relação ao consumo do jantar de 1 dia anterior.

Buscando encontrar e avaliar os possíveis relacionamentos entre as diversas métricas utilizadas, esta análise ganhou destaque como um efeito anômalo e provavelmente casual encontrando entre os dados. Apesar das grades curriculares e horários dos alunos que consomem refeições no almoço serem, geralmente, dispares aos alunos que consomem o jantar na noite do dia anterior, nota-se uma relação evidente entre estas 2 variáveis. Esse comportamento pode ser evidenciado pela congruência entre os parâmetros, como demonstrado na Figura 24 e pela elevada correlação obtida na regressão linear ( $R = 0,7655$ ) entre esses 2 consumos, apresentada na Figura 25. Ainda que apresente uma relevante correlação, não foi possível determinar uma causa evidente para este efeito anômalo encontrado.

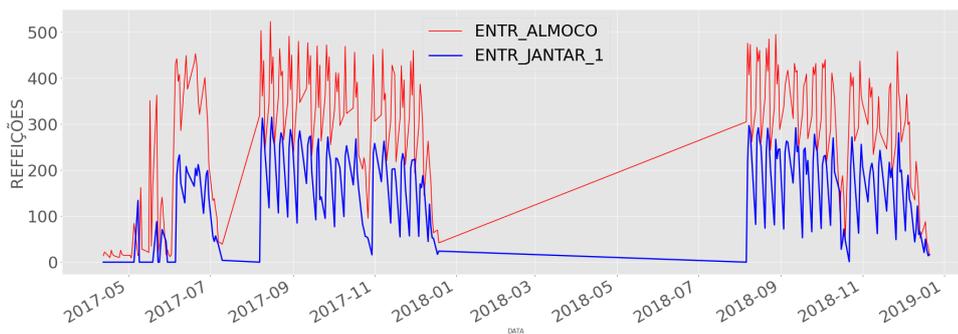


Figura 24: Correlação de consumo de almoço e jantar de 1 dia anterior.

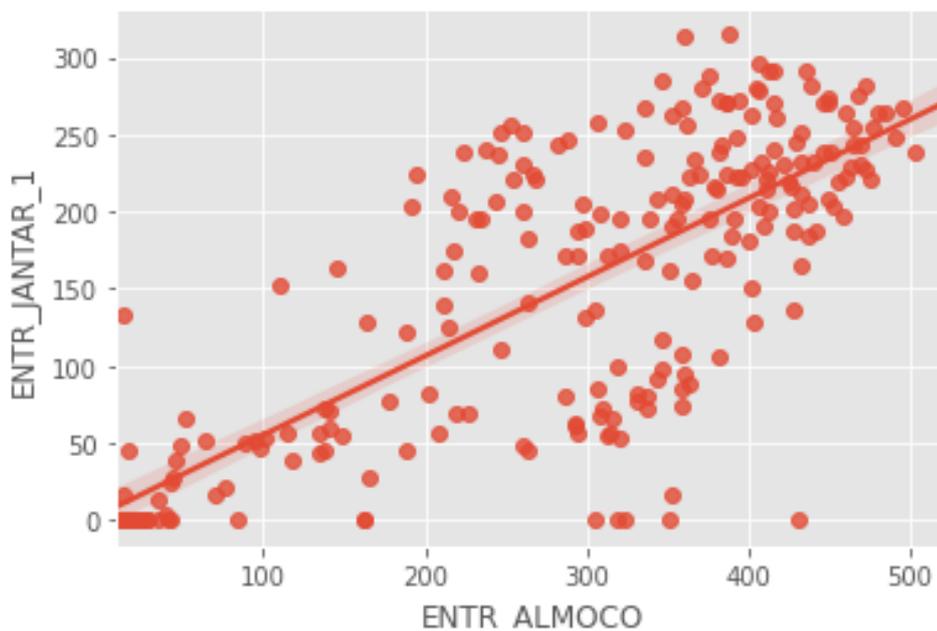


Figura 25: Gráfico de dispersão entre consumo e jantar de 1 dia anterior.

### 5.2.3 Análise da sazonalidade semanal

Os gráficos de consumo a seguir da Figura 26, representando a segunda-feira, até a Figura 30, representando a sexta-feira, são gerados para as *features* categóricas binárias, com a funcionalidade *violin-plot* da biblioteca *seaborn*, própria para distribuição de variáveis categóricas binárias em um conjunto de dados.

O violino azul com o valor 1 representa a distribuição do consumo ao longo do conjunto total de dados. O violino com valor zero pode ser ignorado e é um retorno padrão no gráfico da ferramenta, representando o complemento do consumo para o dia da semana considerado. Nas sextas feiras, o consumo teve escala de distribuição menor para todo o conjunto 2019. Foi notório que apesar da alternância de grades horárias durante a troca de semestres no ano de 2019, os dias de terça e quinta feira concentraram o maior movimento de consumo.

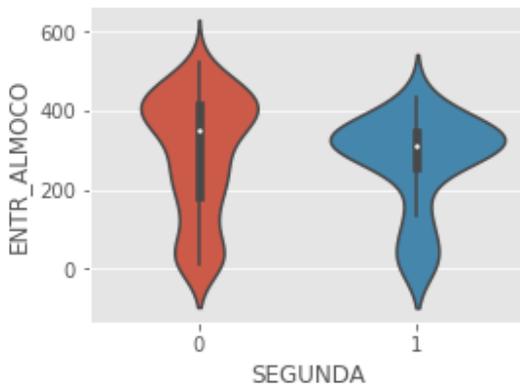


Figura 26: Gráfico de violino da distribuição do consumo na segunda feira.

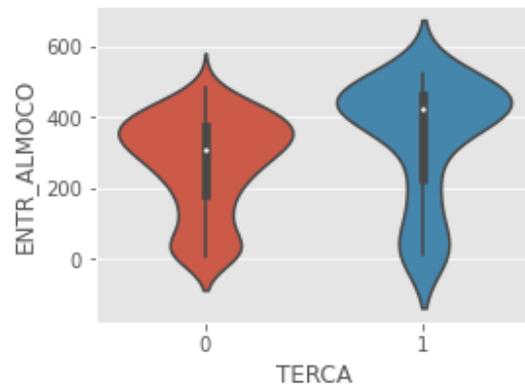


Figura 27: Gráfico de violino da distribuição do consumo na terça feira.

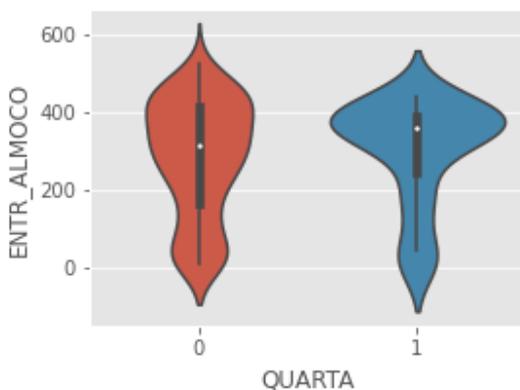


Figura 28: Gráfico de violino da distribuição do consumo na quarta feira.

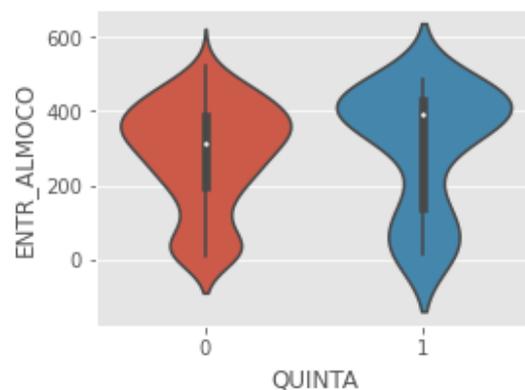


Figura 29: Gráfico de violino da distribuição do consumo na quinta feira.

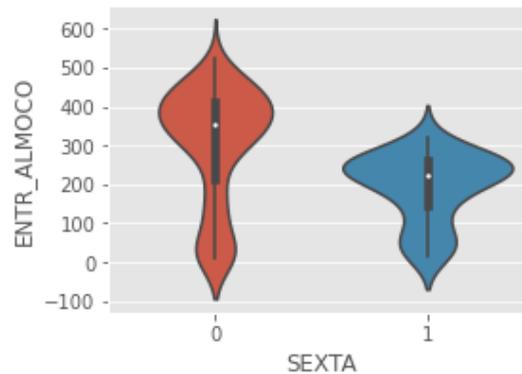


Figura 30: Gráfico de violino da distribuição do consumo na sexta feira.

## 5.2.4 Análise das variáveis exógenas

As variáveis exógenas correspondem aos parâmetros, de domínio discreto, que são utilizados exclusivamente nos modelos de redes neurais mistos, e são lidos pelas camadas MLP destes modelos.

### 5.2.4.1 Consumo atual em relação ao avanço do semestre

Para esta análise foi necessário restringir o domínio de análise para 1 semestre, o consumo em relação ao avanço do semestre teve queda abrupta nos últimos dias do semestre, portanto a correlação dos conjuntos de dados das Figuras 31 e 32 obteve valor negativo.

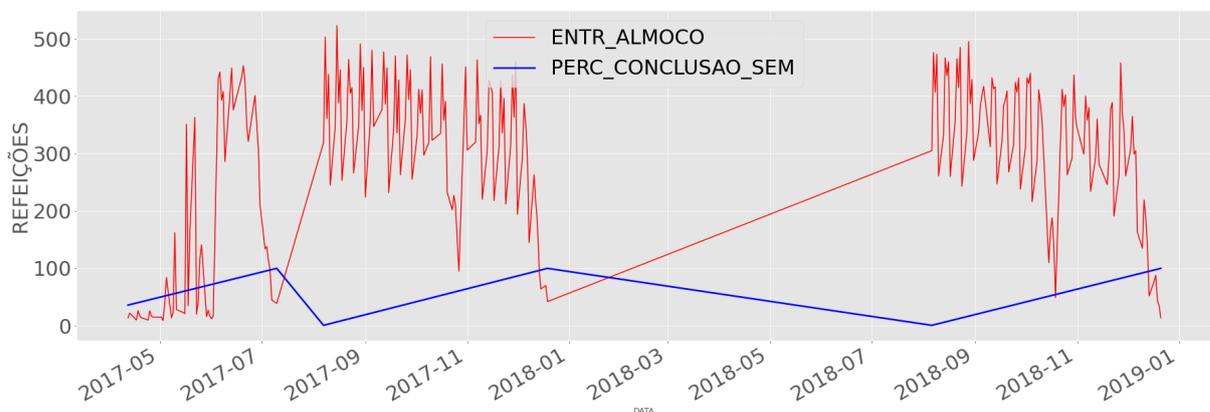


Figura 31: 1a Fase : Relação da distribuição do consumo com o avanço do semestre, Correlação ( $r$ ) = -0.35.

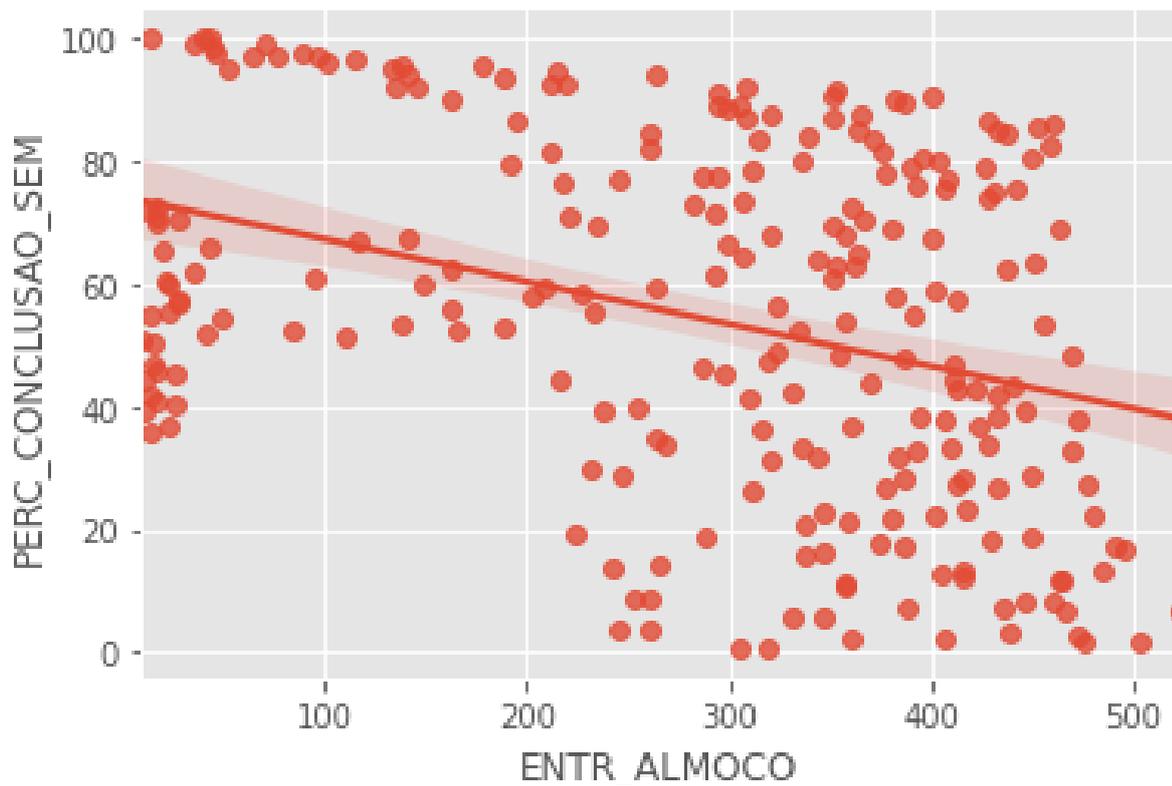


Figura 32: Gráfico de dispersão da distribuição do consumo com o avanço do semestre.

## 5.3 Protocolo Experimental

Nesta Seção se iniciam os experimentos com os modelos de redes neurais. O primeiro experimento avalia o potencial de aprendizado dos modelos em prever o consumo do RU, por meio de um modelo básico de rede neural, ao concluir que as topologias mais básicas de redes neurais tem potencial de aprendizado sobre o problema foram desenvolvidas novas topologias. Em seguida, foi selecionado a topologia e a fase experimental que trouxe os melhores resultados e por fim a mesma é analisada e indicada como a solução final do trabalho.

### 5.3.1 Avaliação do aprendizado do problema da predição de refeições por meio de redes neurais MLP

#### 5.3.1.1 Ajuste empírico de topologia do primeiro modelo perceptron

O primeiro experimento com redes neurais realizado na primeira fase experimental, avaliou a capacidade de aprendizado do modelo perceptron sobre a sazonalidade dos dados endógenos, referentes ao domínio de consumo de refeições no R.U, verificando se o comportamento de consumo no restaurante pôde ser aprendido por este tipo de rede neural, portanto foi definida 1 rede neural inicial perceptron com apenas 1 camada oculta contendo 1 neurônio

para 15 parâmetros de entrada (mesmo número de parâmetros endógenos) e com 1 neurônio de saída, denominada de MLP1.

Os parâmetros endógenos correspondem à uma série temporal de intervalo de 5 dias anteriores para consumo de refeições no período do almoço, jantar e de vendas de *tickets* no período do almoço. O modelo foi denominado MLP1, sua ilustração pode ser visualizada na Figura 33 obtida pela ferramenta NETRON.

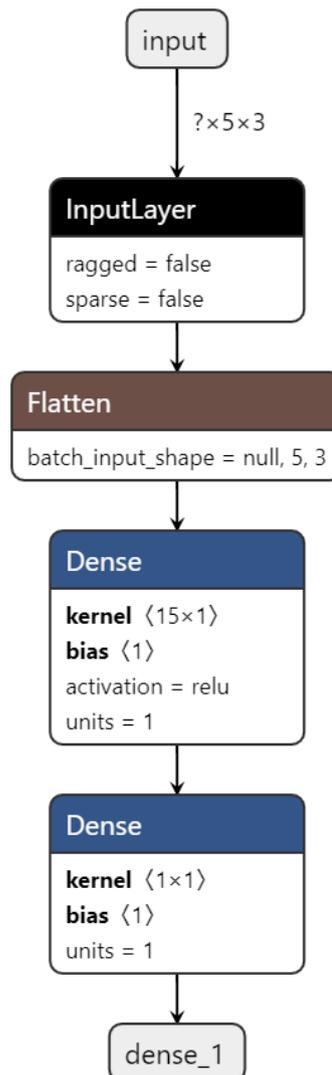


Figura 33: Topologia do modelo MLP1, Ferramenta NETRON.

Cada camada do modelo MLP1 corresponde à um bloco com título **Dense** nesta Figura, a primeira aresta da Figura entre o bloco *input* e *InputLayer* demonstra as 3 séries temporais dos parâmetros de entrada, com intervalo de 5 dias passados cada. **ENTR\_ALMOCO**, **ENTR\_JANTAR** e **VENDAS\_ALMOCO**. O bloco *Flatten* converte cada dia de entrada das séries temporais em um parâmetro de entrada da rede neural MLP, conforme modelo conceitual do trabalho de Lopes (2008) ilustrado na Figura 11 que utiliza apenas 1 parâmetro endógeno

com intervalo temporal também de 5 dias. A primeira camada oculta desta rede pode ser visualizada no primeiro bloco denso da Figura que demonstra a função de ativação ReLu deste neurônio, e o número de unidades desta camada sendo 1. A camada de saída é o último bloco da Figura, também com 1 unidade, e como a função de ativação é linear, ela não é exibida na descrição do bloco.

O treino deste modelo foi executado, obtendo RMSE com o valor 130,62 sobre o conjunto de validação, neste caso da primeira fase sendo os dados do primeiro semestre de 2018, e é possível notar na Figura 34 que as linhas da função de perda e treino demonstraram um treino adequado até a última época de treino, e como não houve uma aplicação de *Validation Loss* simultânea com *Train Loss*, o treino não produziu um *overfitting*.

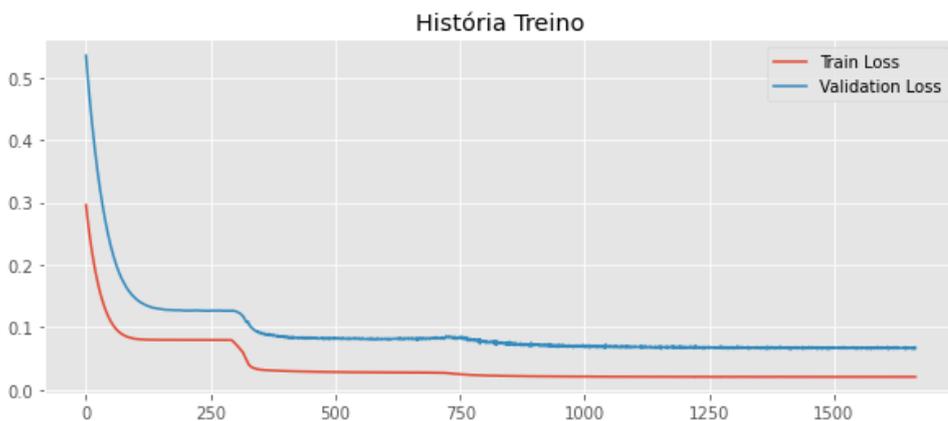


Figura 34: Gráfico de treino do modelo MLP1, RMSE = 130,62

Portanto foi aumentada a profundidade do modelo MLP1 obtendo-se o modelo MLP2 com topologia ilustrada na Figura 35, e após o treino deste modelo foi possível notar a diminuição do RMSE para o valor de 107.97, como observado na Figura 36. É possível notar que a partir da época 300 de treino, a linha *Train Loss* começou a sofrer um decréscimo de erro a medique o *Validation Loss* iniciou um ganho de erro a partir da época 400. Como a reconfiguração de topologia da rede produziu melhora de resultados até um *overfitting* saturar esta melhoria para esta topologia, se validou a hipótese de que a predição do consumo no restaurante, pode ser aprendida por modelos simples de redes neurais, e portanto a pesquisa seguiu com a definição de novos modelos demonstrados na próxima Seção.

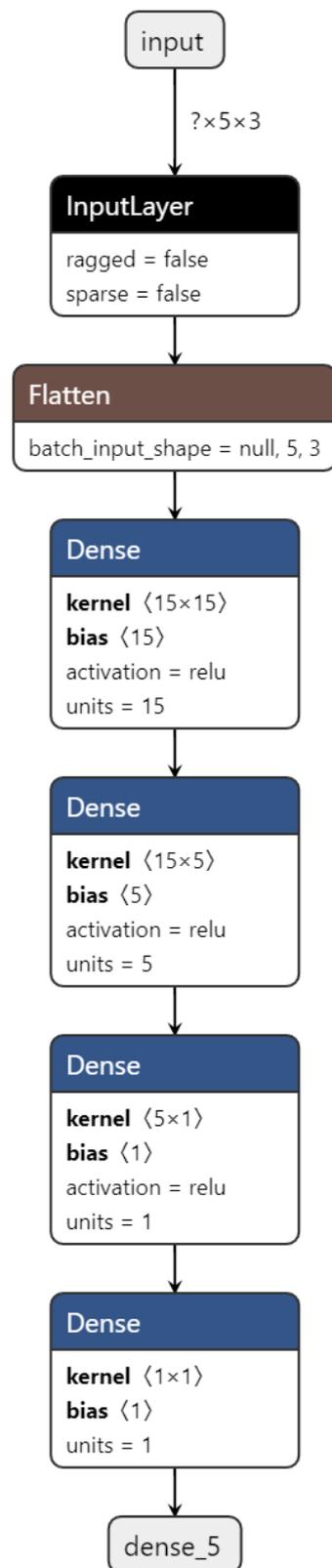


Figura 35: Topologia do modelo MLP2.

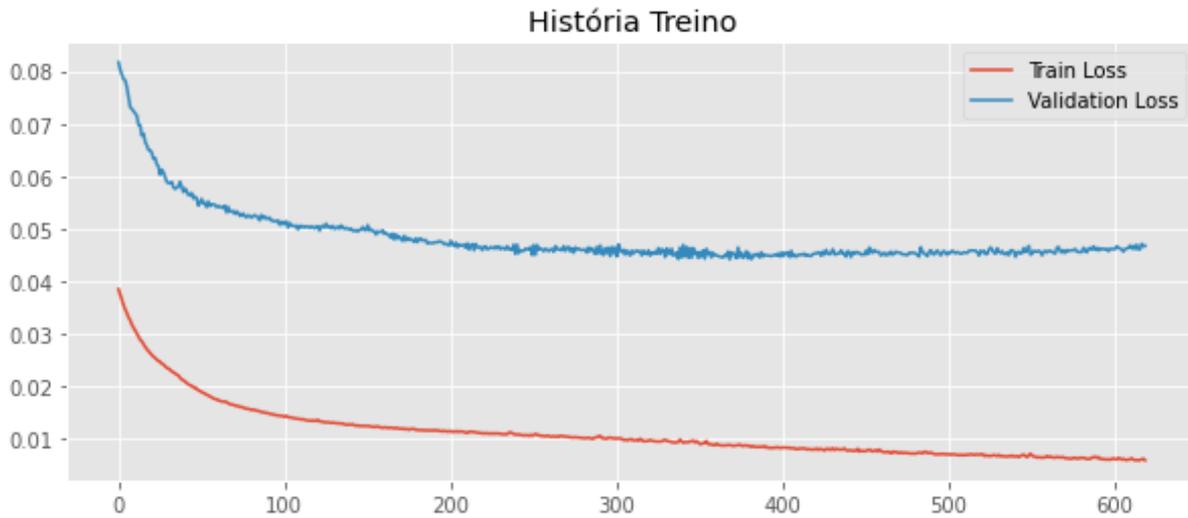


Figura 36: Gráfico de treino do modelo MLP2. RMSE = 107,97.

### 5.3.2 Topologias dos melhores modelos

Nesta seção são feitos alguns comentários sobre as topologias dos modelos que obtiveram resultados, apresentados com figuras esquemáticas da rede. Para o resto dos modelos treinados, o mesmo procedimento é feito no anexo A.

#### 5.3.2.1 Modelo misto RNN\_EXO\_1, melhor resultado na segunda fase e em todo o trabalho

Interpretando o digrama do primeiro modelo misto, RNN\_EXO\_1, na Figura 37 o bloco com título GRU à esquerda na Figura de topologia do modelo, trata as entradas endógenas (temporais), assim como exemplificado nos modelos GRU endógenos. O bloco com título **Dense** é uma rede MLP que recebe um *input* com 10 parâmetros de 1 dimensão, portanto todos discretos, correspondendo aos 4 parâmetros climáticos (temperatura, umidade, pressão e vento), 1 parâmetro para o dia da semana vigente, 1 para o semestre vigente e 4 parâmetros de controle do calendário (distância da data anterior, posterior, avanço do semestre, avanço do mês). A saída dos blocos GRU e MLP são concatenadas e tratadas pelo bloco MLP de saída.

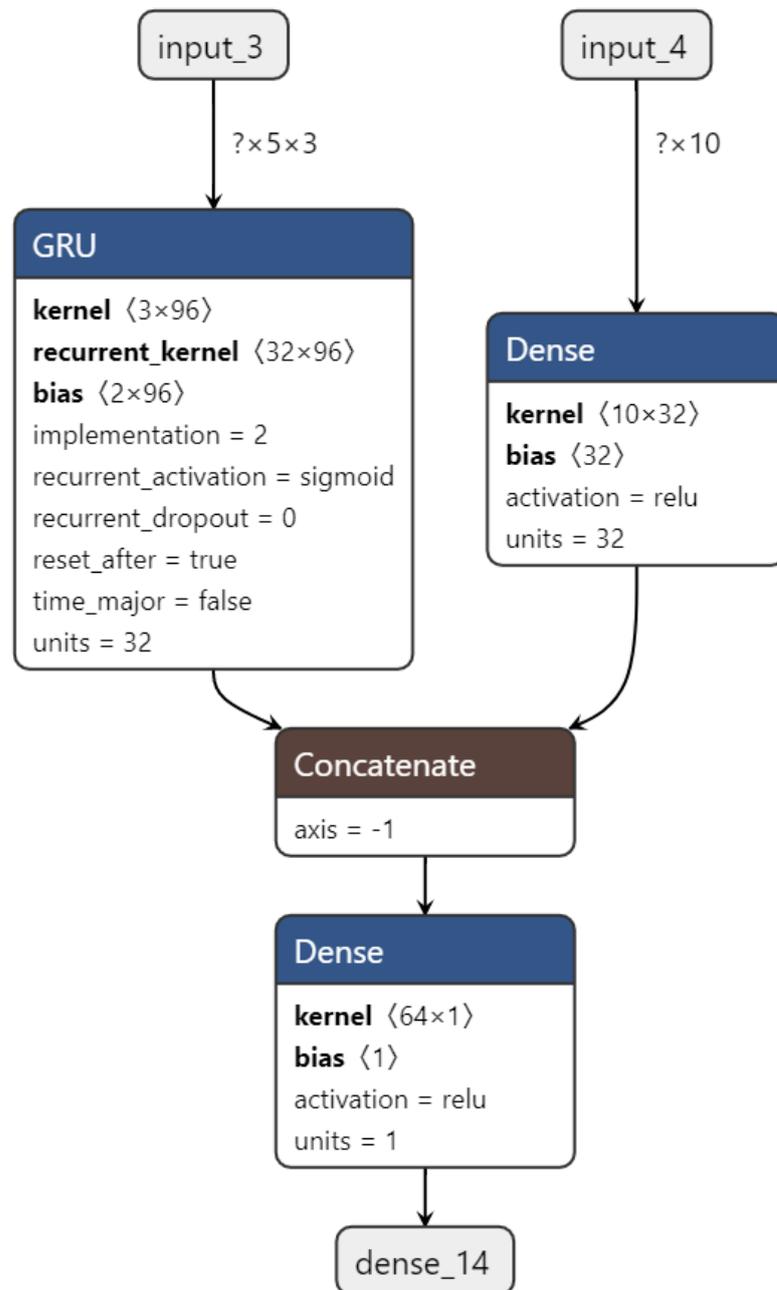


Figura 37: Topologia do modelo RNN\_EXO\_1.

### 5.3.2.2 Modelo endógeno GRU RNN\_ENDO\_2, melhor resultado na primeira fase

Este modelo foi obtido através de uma segunda reconfiguração do primeiro modelo GRU, RNN\_ENDO\_1 detalhado na Figura 49 com o aumento da profundidade de unidades deste modelo anterior, em formato regressivo de 16 unidades na primeira camada, 8 unidades na segunda e 4 na terceira, e com a inclusão do recurso *dropout* fundamentado no tópico 2.2.4.1, sua topologia final é observada na Figura 38. Este modelo produziu o melhores resultados na primeira fase experimental, detalhados na Figura 39 na próxima Seção.

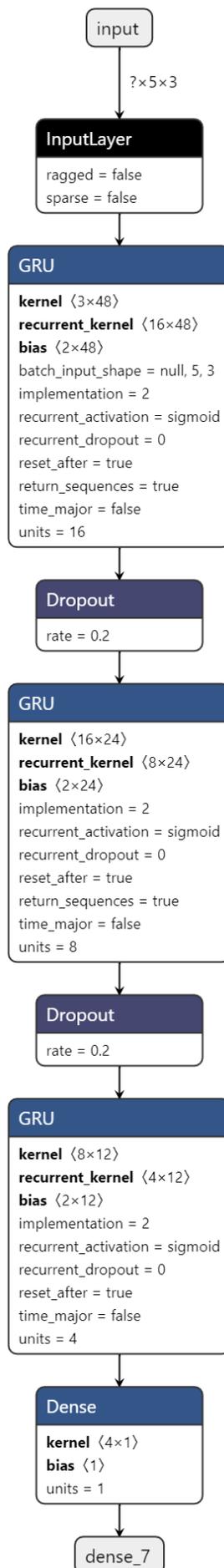


Figura 38: Topologia do modelo RNN\_ENDO\_2.

### 5.3.3 Diferenças principais dos resultados entre as fases experimentais

**Diferenças entre os melhores modelos** Para os experimentos da 1a fase, o modelo que produziu o menor RMSE no conjunto de testes com vantagem em todas as outras métricas foi o modelo endógeno, RNN\_ENDO\_2, com algumas anomalias de predição, como observado na Figura 39.

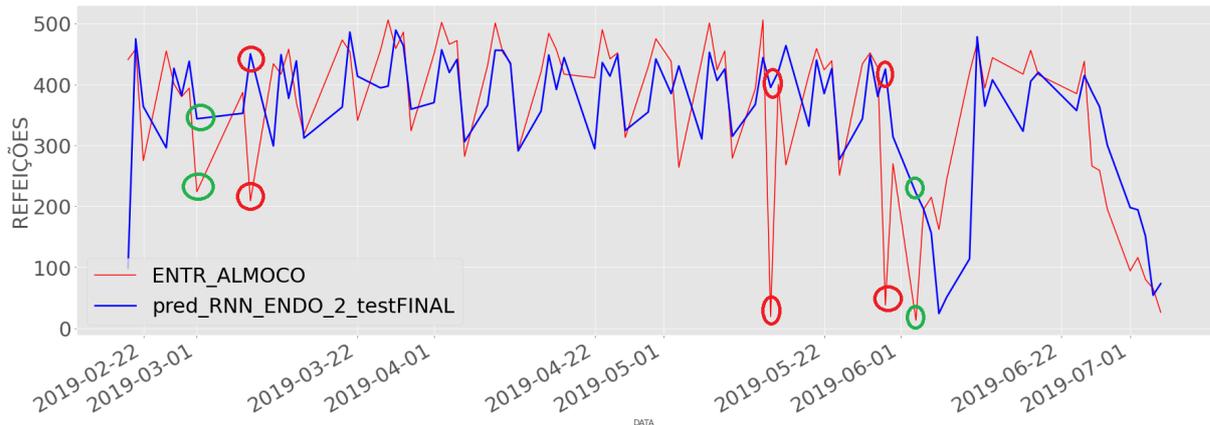


Figura 39: Análise de anomalias preditivas do RNN\_ENDO\_2

Os pontos verdes *outliers* representam predições que corresponderam a tendência de alta ou baixa de consumo mas com erros discrepante, e os pontos vermelhos representam predições com tendência inversa ao consumo. As justificativas para a predição dentro da tendência, se encontram na Tabela 6, denotando datas especiais que não poderiam corresponder ao processo de aprendizado do modelo.

Tabela 6: Erros de predições do modelo RNN\_ENDO\_2 na 1a fase

Data	Consumo	Justificativa
01/03/2019 (sexta feira)	224	Sexta Feira pré - carnaval
03/06/2019 (segunda feira)	13	Segunda Feira pós paralisação estudantil

As justificativas para previsões onde o modelo seguiu tendência oposta ao consumo também corresponderam à datas especiais, conferidas na Tabela 7.

Tabela 7: Anomalias de predições do modelo RNN\_ENDO\_2 na 1a fase

Data	Consumo	Justificativa
08/03/2019 (sexta feira)	209	Sexta Feira pós - carnaval
15/05/2019 (quarta feira)	19	Paralisação estudantil na praça Afonso Pena
30/05/2019 (quinta feira)	38	Paralisação estudantil na Praça Afonso pena

Nas métricas deste modelo, é observado na Tabela 8 que a soma dos erros positivos, correspondeu à um descarte de aproximadamente 3479 refeições, e o erro quadrático médio de previsão foi de aproximadamente 108 refeições.

Tabela 8: Métricas do melhor modelo: RNN\_ENDO\_2

Melhor modelo:	RNN_ENDO_2:
Total_Consumidas	31962
Total_Previstas	31465,61133
Erro_Total_Previsao	-496,3886719
Percentual_Erro_Total	-1,5530%
Correlação	0,595439895
P-value	9,42215E-10
RMSE	108,0663015
Soma dos erros Negativos	2982,567947
Soma dos erros Positivos	3478,957266
ERRO_ABS_MEDIANO	46,70721436
ERRO_ABSOLUTO_PERCENTUAL_MEDIO	74,93539002

Já na segunda fase, todos os modelos obtiveram melhoras no erro de treino sobre o conjunto de validação, e foi obtido o modelo com as melhores previsões do trabalho, o modelo misto **RNN\_EXO\_1** detalhado em sua própria Seção a seguir. É importante notar que as 2 fases produziram melhores modelos de classes distintas, a primeira com um modelo que utiliza apenas dados endógenos, e que contempla um conjunto de validação e teste com amplitude de apenas 1 semestre, e a segunda com um modelo que utiliza dados temporais e discretos, e que utiliza conjunto de validação e teste com amplitude de 1 ano. Isso denota que resultados melhores foram conquistados sem nenhuma alteração de parâmetros e hiper-parâmetros nos modelos, alterando-se apenas a organização temporal dos conjuntos de dados.

Durante o teste de todos os modelos, apenas o primeiro semestre contemplou datas especiais onde estes modelos produziram anomalias de previsões, ilustrado como exemplo na Figura 39.

## 5.4 Resultados com o melhor modelo, RNN\_EXO\_1

Este modelo, representado na Figura 37 obteve os melhores resultados de todo este trabalho, quando foi treinando na segunda fase experimental. É notório sua melhoria de resultados com uma única mudança da organização dos conjuntos de dados entre as fases experimentais.

### 5.4.1 Comparativo do treino entre as duas fases

Nota-se que este modelo produziu convergiu à um *overfitting* de forma mais lenta na primeira fase, ilustrado na Figura 40 em comparação ao treino da segunda fase demonstrando

uma convergência à um *overfitting* rápida e acentuada a partir da época 100, trazendo melhores resultados com a parada antecipada de treino, ilustrado na Figura 41. A diferença dos valores da métrica RMSE entre estes 2 treinos, produzindo um resultado melhor na segunda fase, de  $RMSE = 109,97$  em comparação ao resultado da primeira fase de  $RMSE = 132,94$ .

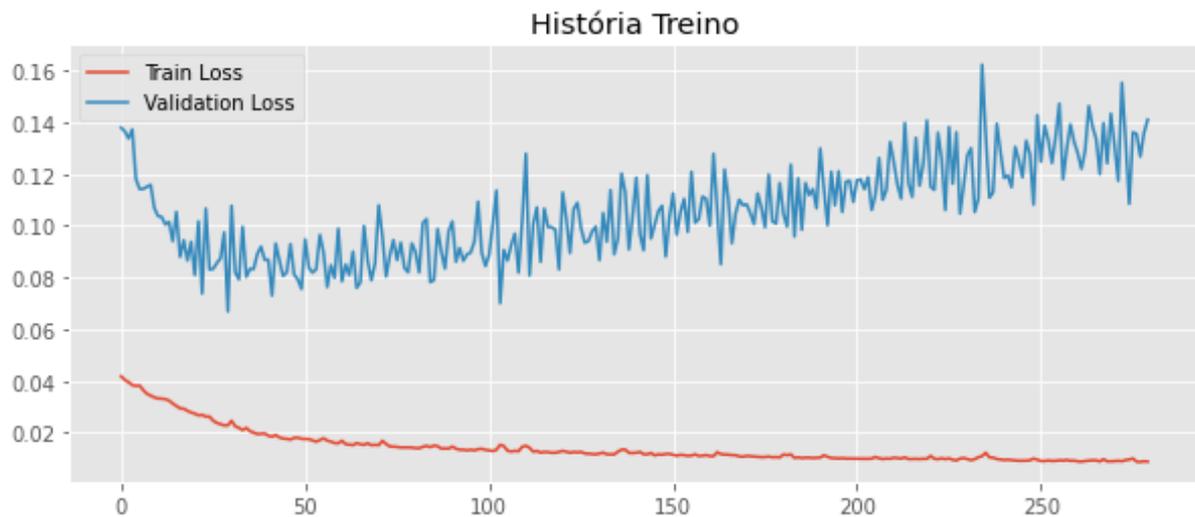


Figura 40: Treino do modelo RNN\_EXO\_1 na 1a fase,  $RMSE = 132,94$

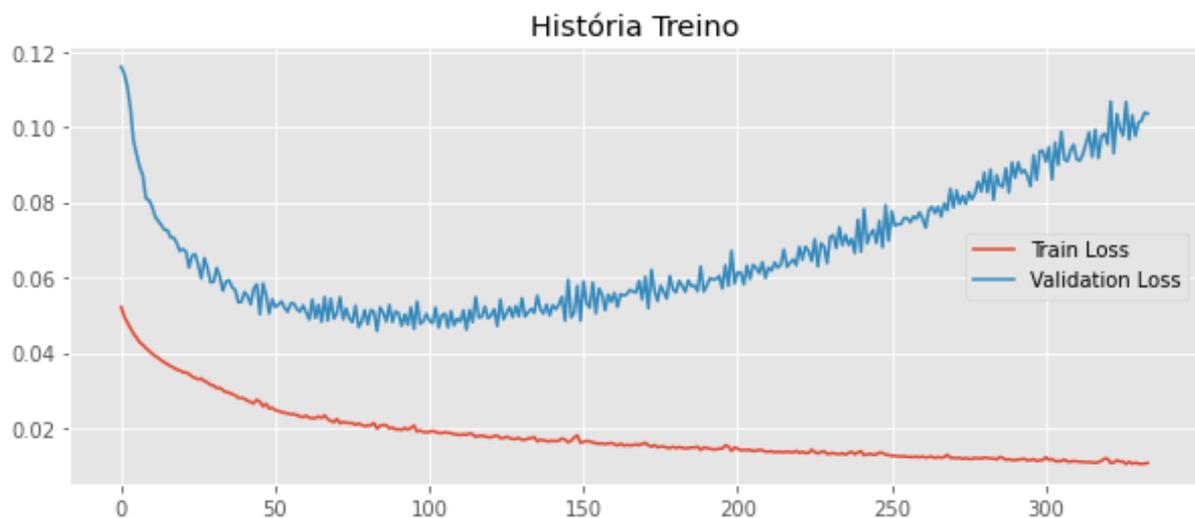


Figura 41: Gráfico de treino do modelo RNN\_EXO\_1 na 2a fase,  $RMSE = 109,97$

#### 5.4.2 Comparativo de teste do modelo no primeiro semestre, entre as duas fases

Como este modelo treinado na segunda fase obteve os melhores resultados do trabalho, foram recalculadas as métricas para o teste de modelo dentro do domínio do primeiro semestre de 2019 para realizar uma comparação equivalente com sua versão treinada e testada também no primeiro semestre de 2019 na primeira fase.

A Tabela 9 demonstra que o RMSE de teste do modelo treinado na primeira fase obteve um desempenho menos satisfatório do que o RMSE treinado na segunda fase de acordo com a Tabela 10. O RMSE sendo menor no treino deste modelo na segunda fase trás uma melhoria em todas as outras métricas em comparação com o seu treino na primeira fase. A soma dos erros positivos de predições também foi menor na segunda fase, o que impacta menor em descarte de refeições. O RMSE deste modelo testado na primeira fase e alcançando RMSE = 106,208 também se saiu melhor do que o melhor modelo da primeira fase, o RNN\_ENDO\_2 que alcançou RMSE = 108,06.

Tabela 9: RNN\_EXO\_1 TREINADO NA 1A FASE,TESTE 1o SEMESTRE 2019

RNN_EXO_1 TREINADO NA 1A FASE,TESTE 1o SEMESTRE 2019	
RMSE	124.49
TOTAL DE REFEIÇÕES CONSUMIDAS	31962
TOTAL DE REFEIÇÕES PROJETADAS	28728.816
ERRO DE PREVISÃO	-3233.1839
PERCENTAGEM DE ERRO	-10.11%
CORRELAÇÃO (r)	0.41
P-value	6.59e-05
R2	0.16
SOMA DOS ERROS POSITIVOS	2709.17
SOMA DOS ERROS NEGATIVOS	5942.35
ERRO ABSOLUTO MEDIANO	85.59
ERRO ABSOLUTO PERCENTUAL MÉDIO	90.98%

Tabela 10: RNN\_EXO\_1 TREINADO NA 2A FASE, TESTE 1o SEMESTRE 2019

RNN_EXO_1 TREINADO NA 2A FASE, TESTE 1o SEMESTRE 2019	
RMSE	106.2080
TOTAL DE REFEIÇÕES CONSUMIDAS	31962
TOTAL DE REFEIÇÕES PROJETADAS	32170.24
ERRO DE PREVISÃO	208.2460
PERCENTAGEM DE ERRO	0.6515%
CORRELAÇÃO (r)	0.59
P-value (p)	1.4143e-09
R2	0.3485
SOMA DOS ERROS POSITIVOS	3454.8698
SOMA DOS ERROS NEGATIVOS	3246.6228
ERRO ABSOLUTO MEDIANO	59.5414
ERRO ABSOLUTO PERCENTUAL MÉDIO	83.2671%

O gráfico de dispersão do modelo treinado na primeira fase, conforme Figura 42 também se saiu pior, mais distante da borda superior direita do gráfico, em relação a dispersão do modelo treinado na segunda fase conforme a Figura 43.

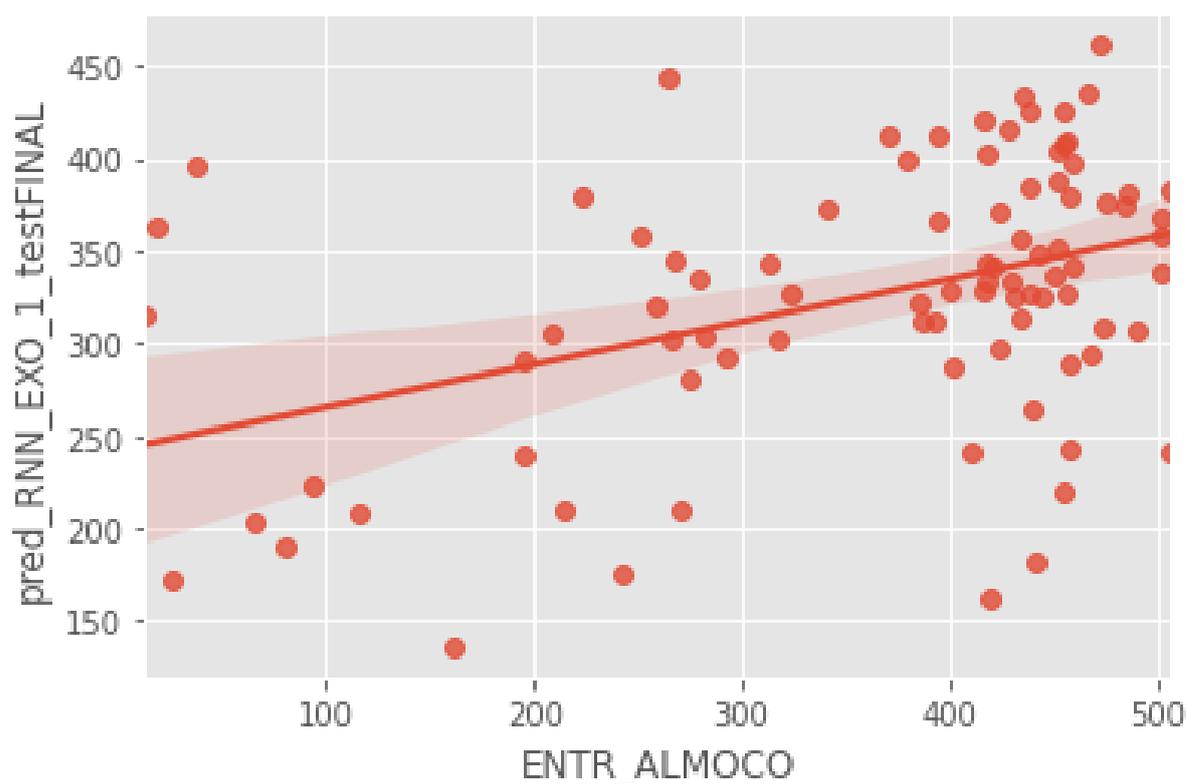


Figura 42: Gráfico de dispersão de teste do modelo RNN\_EXO\_1, 1a fase

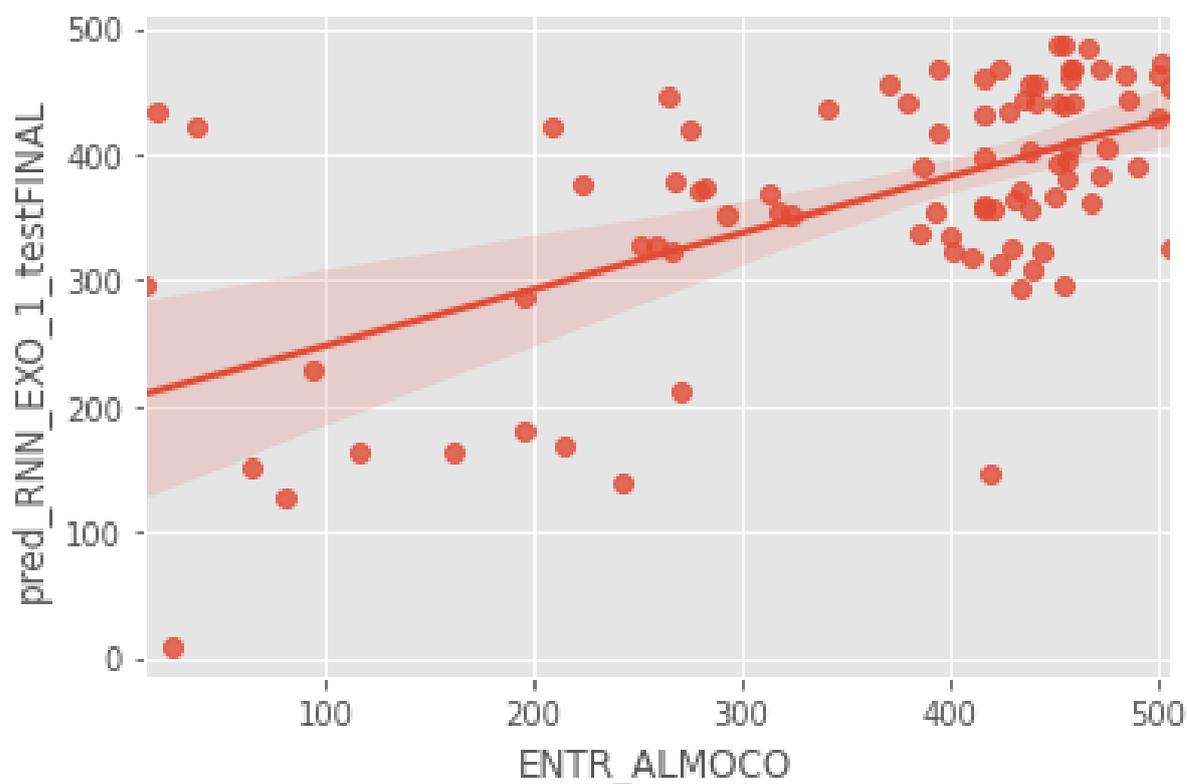


Figura 43: Gráfico de dispersão de teste do primeiro semestre, RNN\_EXO\_1 treinado na segunda fase.

Por fim na comparativa entre os gráficos de predição, o modelo RNN\_EXO\_1 treinado na primeira fase produziu predições piores, e não aprendeu a sazonalidade semanal do consumo, como pode ser observado na Figura 44, é possível notar também, na Tabela 9, que a correlação entre os valores previstos e o consumo real, bem como o valor  $R^2$  foi inferior em comparação às métricas do modelo treinado na segunda fase, e que este modelo treinado na segunda fase aprendeu melhor a sazonalidade semanal e mensal do consumo como pode ser observado na Figura 45.

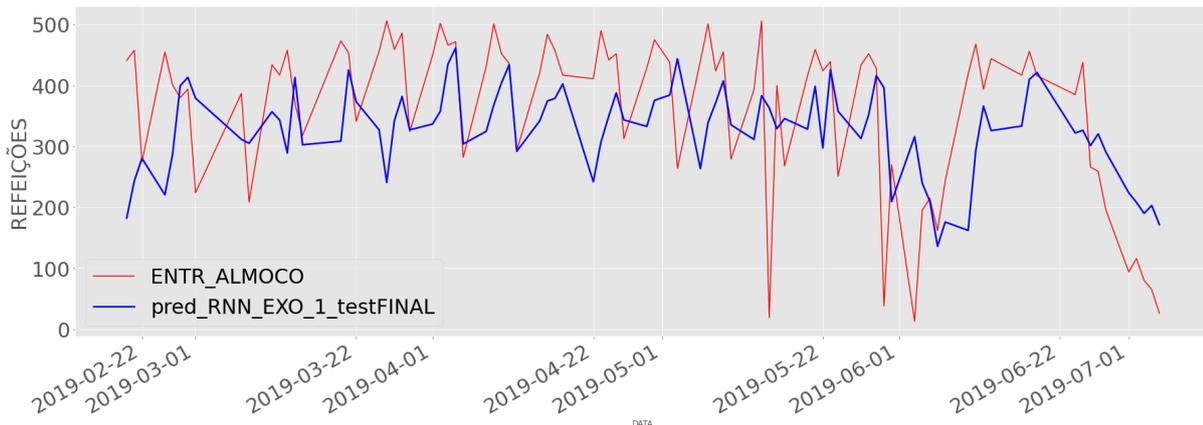


Figura 44: Teste do modelo RNN\_EXO\_1, 1a fase.

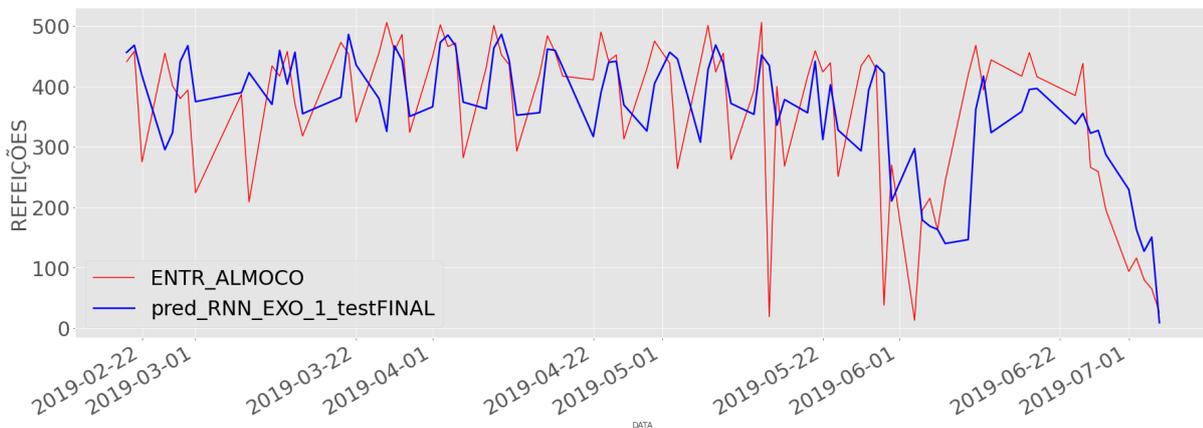


Figura 45: Teste do primeiro semestre do RNN\_EXO\_1 treinado na segunda fase.

### 5.4.3 Teste final do modelo RNN\_EXO\_1 para as predições no RU do ICT-Unifesp

O teste final do modelo RNN\_EXO\_1 por fim produziu os melhores resultados com seu treino na segunda fase e sendo testado para o ano inteiro de 2019. O RMSE observado na Tabela 11 foi notoriamente inferior à todos os modelos treinados e testados em todo o trabalho.

Tabela 11: RNN\_EXO\_1 TREINADO NA 2A FASE, TESTE ANO DE 2019

RNN_EXO_1 TREINADO NA 2A FASE, TESTE ANO DE 2019	
RMSE	99.36
TOTAL DE REFEIÇÕES CONSUMIDAS	58653
TOTAL DE REFEIÇÕES PROJETADAS	62048.04
ERRO DE PREVISÃO	3395.04
PERCENTAGEM DE ERRO	5.78%
CORRELAÇÃO (r)	0.67
P-value (p)	3.29e-25
R2	0.45
SOMA DOS ERROS POSITIVOS	8163.18
SOMA DOS ERROS NEGATIVOS	4768.13
ERRO ABSOLUTO MEDIANO	55.23
ERRO ABSOLUTO PERCENTUAL MÉDIO	83.2671%

O descarte de refeições foi obtido pela soma dos erros positivos atingiu o valor de 4768 refeições. É possível notar na Figura 46 que o modelo aprendeu bem a sazonalidade mensal e semanal do consumo, mas obteve erro discrepante para o primeiro valor previsto do segundo semestre, o erro foi justificável pois seu conjunto de treino contempla apenas 1 ano com 1 alternância de semestre impossibilitando um aprendizado melhor sobre este comportamento. O gráfico de dispersão ilustrado na Figura 47 também demonstra uma boa regressão linear sobre os valores previstos pelo modelo e o valor real de consumo, se aproximando da função identidade de uma previsão ideal.

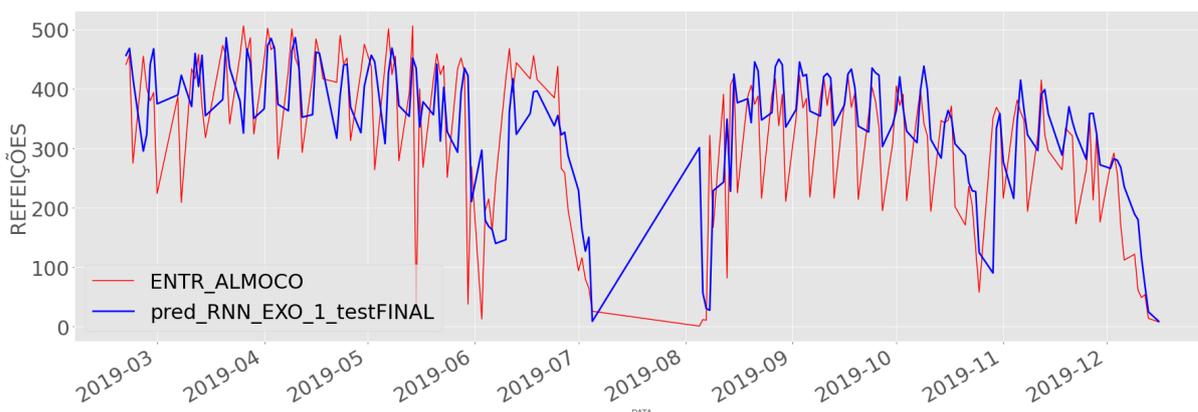


Figura 46: Gráfico final de teste do modelo RNN\_EXO\_1.

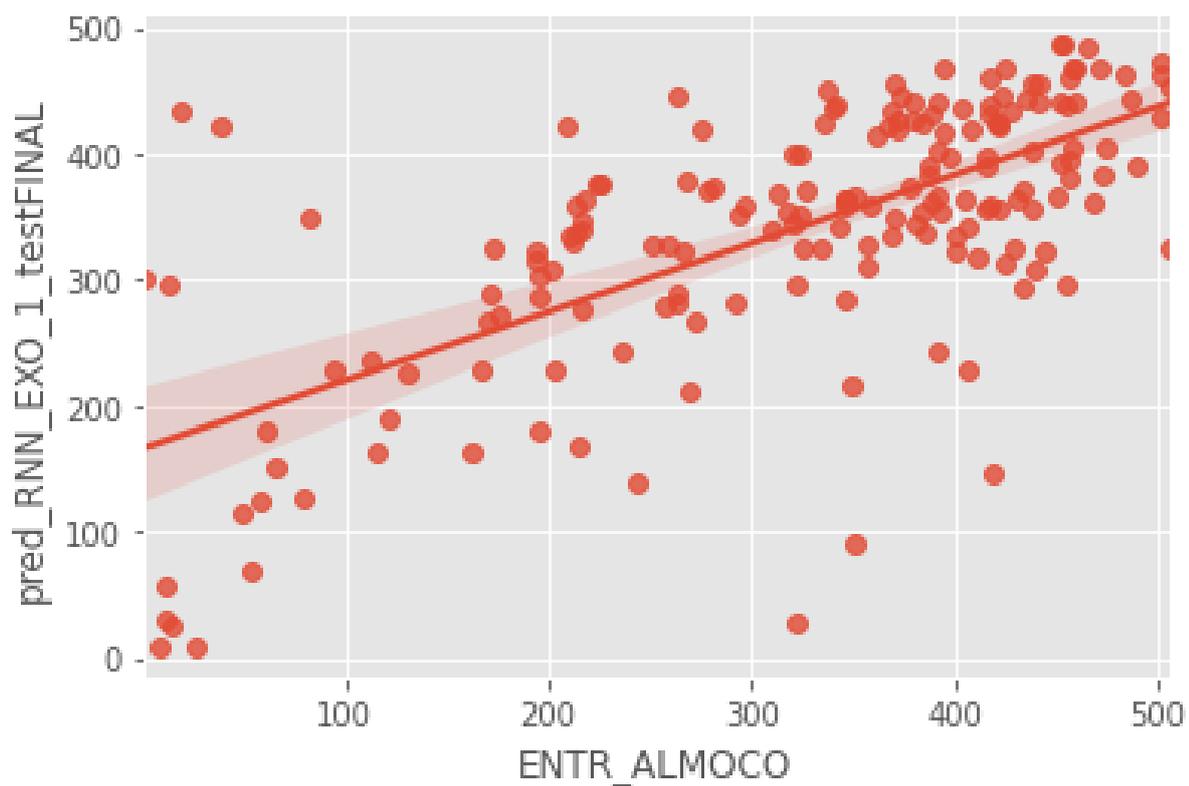


Figura 47: Gráfico de dispersão do modelo RNN\_EXO\_1.

Resultados obtidos pelos outros métodos testados estão disponíveis no B.

## 6 Conclusão

Primeiramente, neste trabalho foi possível avaliar importância da metodologia de divisão do conjunto de dados em séries temporais. Visto que com um conjunto de dados de sazonalidade temporal, é notório que a ordenação dos dados na separação dos conjuntos de treino, teste e validação devem seguir uma ordem cronológica para os modelos aprenderem com o passado e realizarem previsões para o futuro. O comportamento anômalo da previsão identificada no capítulo de divisão do conjunto de dados da fase 1, demonstrado na figura 16, e sua eliminação após a organização correta da série temporal, conforme a figura 17, reforça a importância da sequência temporal dos dados para o correto aprendizado dos modelos testados.

Com relação ao método de produção de refeições com margem de erro e análise da semana anterior foi possível observar que mesmo com a produção 30% acima do consumo na semana anterior, no fim de cada semestre, o restaurante do ICT-Unifesp mais do que 30% são descartados. Este comportamento oscilatório do consumo e o acréscimo de *outliers* acaba ampliando o erro dos modelos em prever as refeições. No ano de 2019, seguindo este método, 23 mil refeições foram descartadas. Neste trabalho O Modelo RNN\_EXO\_3, que apresentou o maior descarte entre todos os 12 modelos testados, obteve um valor máximo de 8914 descartes. Isso evidencia a necessidade de se implementar métodos eficientes para a produção e planejamento de refeições no restaurante universitário da Unifesp.

Em relação aos ajustes empíricos da topologia dos modelos durante a etapa de validação dos primeiros modelos desenvolvidos foi possível notar a redução do RMSE ao longo do aprofundamento da rede Perceptron para treino e avaliação sob o conjunto de validação, validando a hipótese de que os modelos tem capacidade de aprendizado do problema em relação ao ajuste da topologia dos mesmos.

Apesar do conjunto de dados conter 2 características que informam a distância em dias para o próximo registro e o registro anterior para os modelos identificarem feriados e recessos prolongados, alguns eventos no calendário, como paralisações, não são muito bem representados, indicando a necessidade de uma exploração mais aprofundada destas características que possam representar melhor este comportamento.

Avaliando o modelo de melhor desempenho na primeira fase, com validação restrita ao primeiro semestre de 2018, os modelos endógenos se saíram melhor do que os modelos mistos. Isso pode significar que os atributos exógenos foram ruidosos durante o aprendizado. Estes atributos são a maioria compostos de sazonalidade anual tais como as climáticas, limitadas às estações do ano. O modelo RNN\_EXO\_1 da 2a fase, obteve o melhor desempenho

entre todos os modelos avaliados neste trabalho, porém algumas melhorias são indicadas:

- Aumentar o conjunto de dados para o modelo se ajustar às sazonalidades semestrais e à troca de semestres. Os atributos categóricos que indicam os semestres, dia da semana, bem como os que quantificam recessos (distância registro anterior e posterior) têm potencial de agregar aprendizado nessa questão. Ainda é necessário uma diversificação maior do conjunto de dados, dado que este modelo foi treinado apenas com 1 período de sazonalidade anual (1 ano para treino, outro ano para validação e um terceiro ano para teste).
- Acrescentar atributos de eventos importantes para identificar eventos e paralisações.
- Um atributo de cardápio tem potencial de aumentar a qualidade da predição.
- Um atributo representando o número de alunos matriculados em cada período de cada dia da semana tem grande potencial de aumentar a predição.
- Pesquisas podem ser feitas para uma melhor transformação dos dados de entrada no modelo perceptron, pois são dados discretos, enquanto os dados que entram na camada GRU são temporais (com intervalo de 5 dias).

## 6.1 Conclusões gerais

O fenômeno mais evidente neste trabalho foi a melhoria significativa em todos os modelos de redes neurais artificiais avaliados, apenas alterando-se a organização do conjunto de dados entre a primeira e segunda fase, sem interferência em qualquer parâmetro ou hiperparâmetro destes modelos. Assim, denota-se a necessidade de mais estudos e experimentos relacionados com organização de conjunto de dados para predições de séries temporais de consumo. As análises diversas de previsão de demanda para o tema abordado requerem extensos métodos de implementação e estruturação de dados.

A aplicação de métodos de treino com retro-propagação, visto a diversidade de parâmetros no aprendizado de máquina dentro de apenas uma análise, onde pode-se montar infinitas topologias diferentes com base na estrutura dos dados coletados, tem potencial para aumentar a performance dos modelos na previsão das demandas par ao RU.

As heurísticas sobre a definição de topologia apesar de diversas, não são determinísticas, e o processo requer análise exploratória, subjetiva e empírica sobre o tema ou problema a ser abordado. Todavia, foi notória a eficiência dos modelos de aprendizado de máquina em trabalhos relacionados à restaurantes universitários. Como no caso do RU do ICT-Unifesp não há qualquer modelo atual de previsão e a falta de um modelo causa desperdício de alimentos e prejuízo ao restaurante, a abordagem desta pesquisa e sua continuação com novos métodos torna-se viável e importante para auxiliar em uma gestão assertiva dos recursos.

# Anexos



# ANEXO A – Topologias de outros modelos de redes testados

**Modelo endógeno MLP\_ENDO\_1** O primeiro modelo experimental MLP para obtenção das predições tem topologia definida na figura 48, denominado MLP\_ENDO\_1, este modelo tem 64 neurônios na primeira camada oculta e 32 neurônios na segunda camada oculta, e assim como todos os outros modelos, 1 neurônio na camada de saída.

**Modelo endógeno GRU RNN\_ENDO\_1** Este primeiro modelo experimental GRU foi definido com 16 unidades GRU, com topologia conceitual definida de acordo com a figura 10 no capítulo de fundamentação teórica MLP\_ENDO\_1 e tem um neurônio percpetron para emissão do sinal de saída, representado pelo bloco **Dense** na figura 49

**Modelo misto RNN\_EXO\_2** O Segundo modelo misto foi definido com o aumento da profundidade das camadas GRU e MLP do modelo anterior, conforme observado na figura 50 foram acrescentadas 1 camada GRU e 1 camada MLP (Dense).

**Modelo misto RNN\_EXO\_3** Para o terceiro e último modelo misto, representado na figura 51, foi feita uma reconfiguração do modelo misto anterior, RNN\_EXO\_2, com a utilização do recurso dropout nas saídas das camadas GRU.

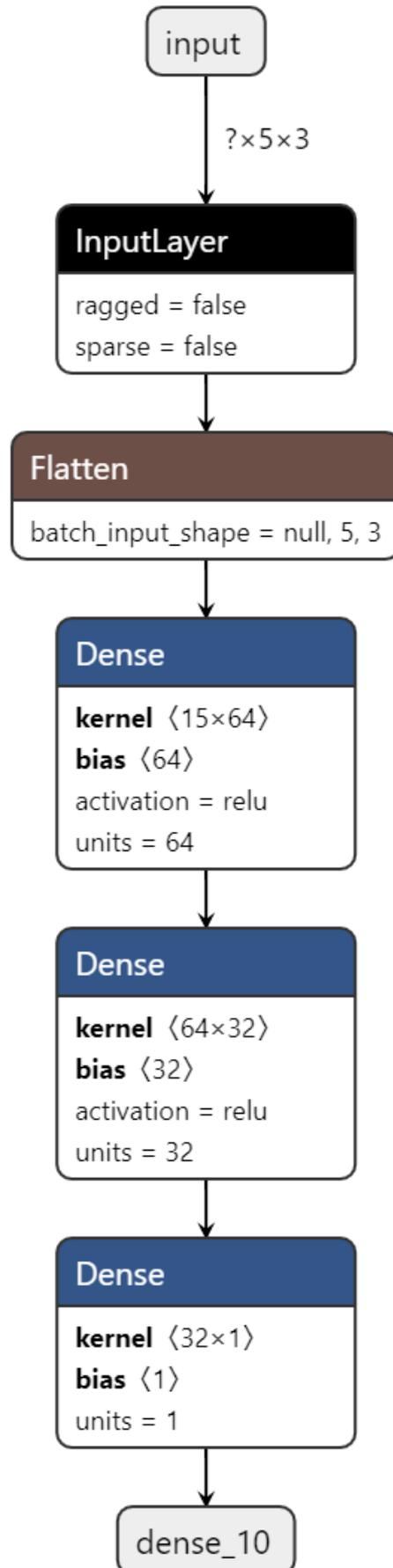


Figura 48: Topologia do modelo MLP\_ENDO\_1

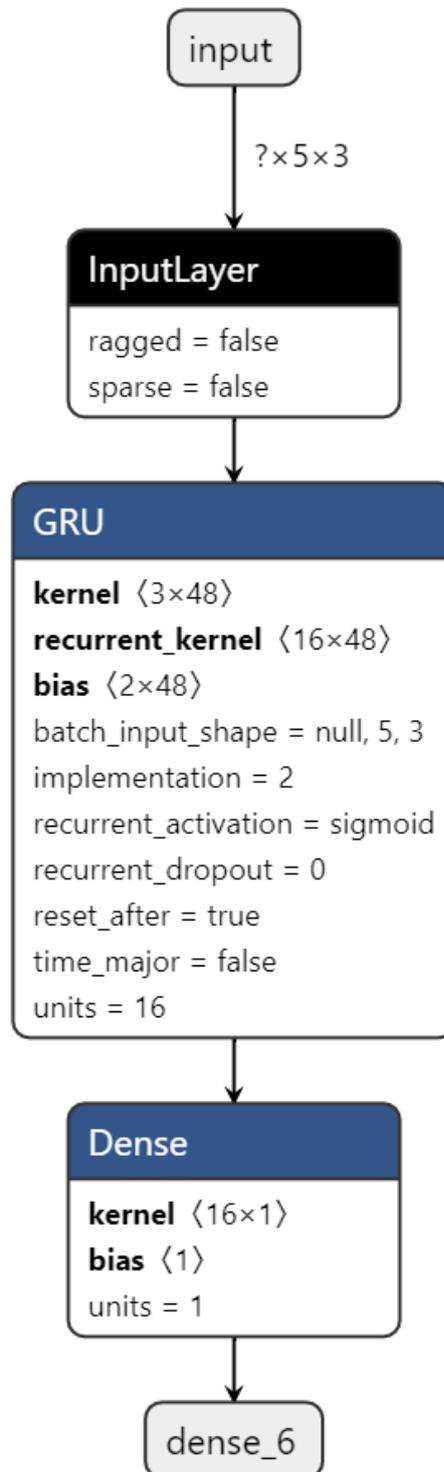


Figura 49: Topologia do modelo RNN\_ENDO\_1

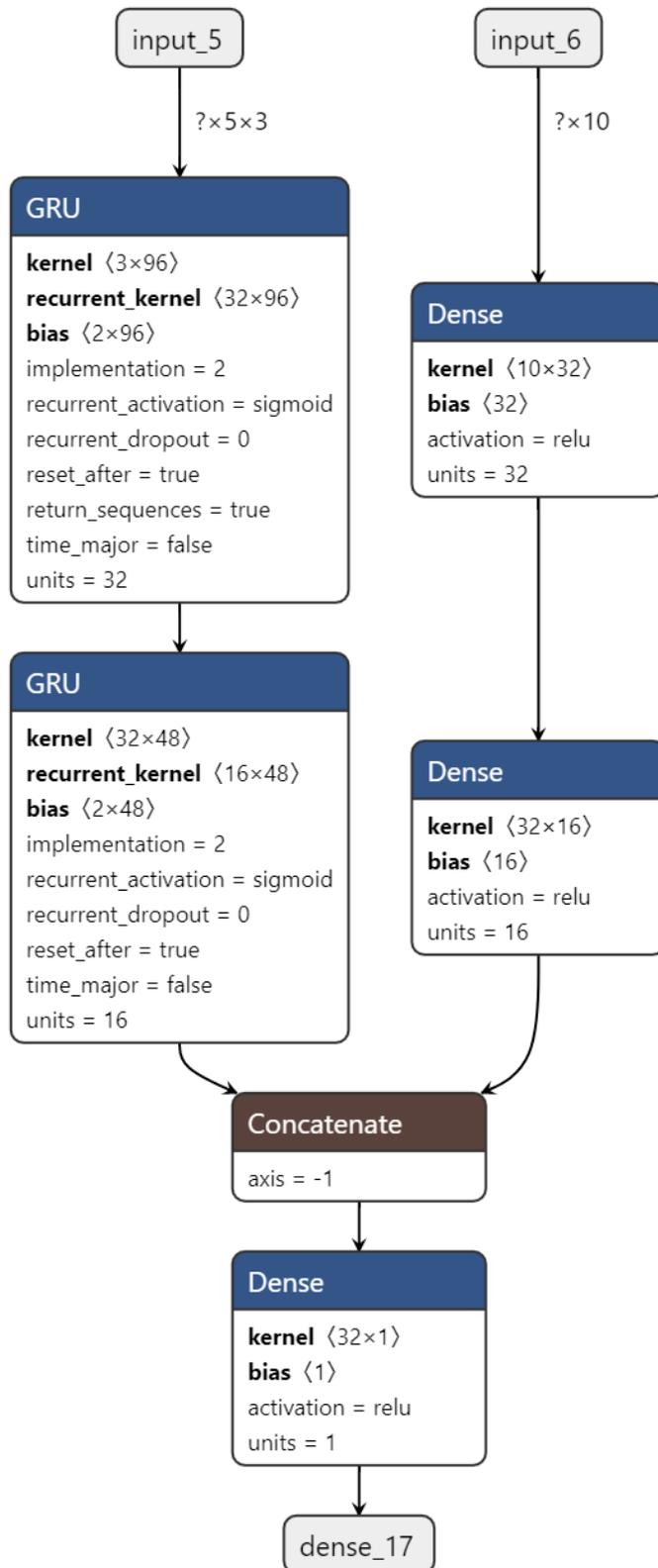


Figura 50: Topologia do modelo RNN\_EXO\_2

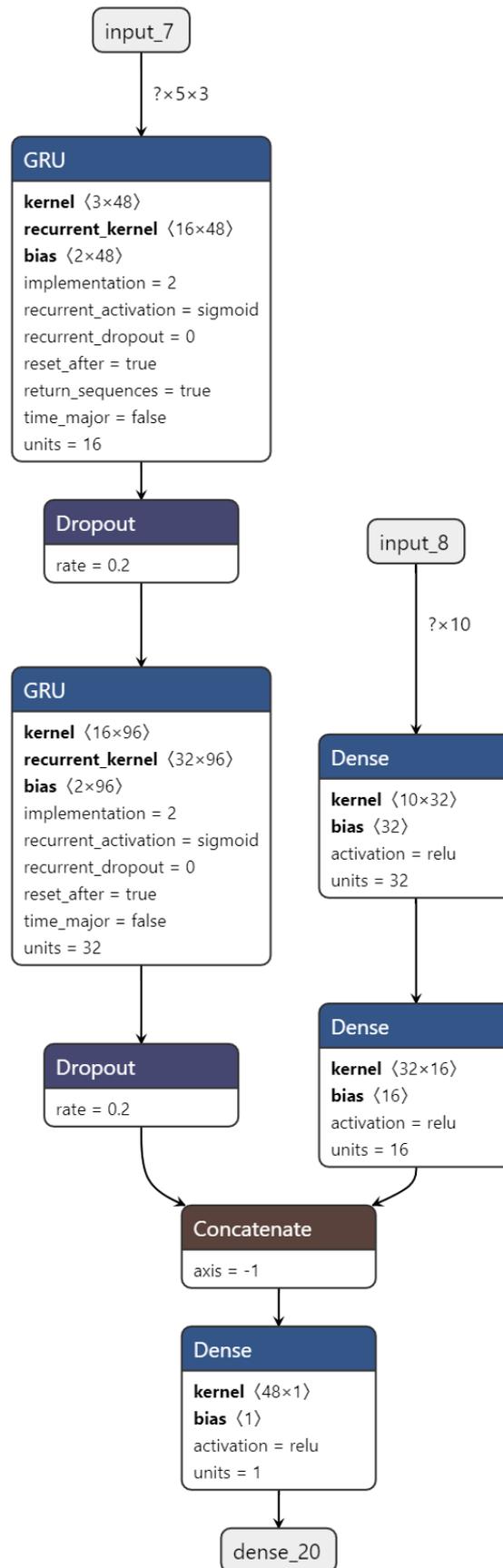


Figura 51: Topologia do modelo RNN\_EXO\_3



# ANEXO B – Métodos e códigos experimentais

## B.1 Primeira Fase Experimental

**Pré-Processamento e treino dos modelos** No repositório abaixo, encontram-se os gráficos, métricas e tabelas de todos os modelos treinados na primeira fase experimental, bem como a análise exploratória detalhada de todas as variáveis e parâmetros de entrada nos modelos, e também o detalhamento do pré-processamento dos conjuntos de dados.

Os experimentos foram realizados na plataforma Google Colab, em linguagem Python. Nenhuma configuração de ambiente é necessária para reproduzir os experimentos, bastando acessar o link para a plataforma google colab no interior do documento, através de um navegador de internet e executá-los caso seja de interesse.

Os experimentos se encontram acessíveis em formato de documento Jupyter Notebook, com recursos de indexação e documentação com blocos de texto e imagens para uma fácil compreensão e acesso instantâneo das informações. [https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case1\\_\\_Experimentos\\_2609.ipynb](https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case1__Experimentos_2609.ipynb)

**Importação e aplicação de métricas dos modelos** No link abaixo encontra-se uma página para a importação dos modelos contidos no repositório deste trabalho, já treinados, e encontram-se também o conjuntos de dados pré-processados para o calculo das métricas finais da primeira fase. [https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case1\\_\\_ModelsTests\\_DriverCode.ipynb](https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case1__ModelsTests_DriverCode.ipynb)

## B.2 Segunda Fase Experimental

**Pré-Processamento e treino dos modelos** Repositório para os gráficos, métricas e tabelas de todos os modelos treinados na segunda fase experimental. [https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case2\\_\\_Experimentos\\_2709.ipynb](https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case2__Experimentos_2709.ipynb)

**Importação e aplicação de métricas dos modelos** Repositório para a importação dos modelos já treinados e conjuntos de dados já pré-processados para o calculo das métricas finais da segunda fase. [https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case2\\_\\_ModelsTests\\_DriverCode.ipynb](https://github.com/ddlandim/monografy-ann-demand-prediction/blob/master/case2__ModelsTests_DriverCode.ipynb)

### B.3 Tabela completa das métricas de todos os modelos

A seguir são listadas as tabelas com todos resultados experimentais.

Tabela 12: Previsões e erros de todos os modelos

MODELO	TOTAL CONSUMIDAS	TOTAL PREVISTAS	ERRO TOTAL PREVISAO	ERRO TOTAL PERC PREVISAO
<b>MODELOS ENDÓGENOS</b>				
RNN_ENDO_1 1A FASE	31962	30927	-1035	-3.2394
RNN_ENDO_1 2A FASE	58653	60412	1759	2.9991
RNN_ENDO_2 1A FASE	31962	31466	-496	-1.5530
RNN_ENDO_2 2A FASE	58653	61855	3202	5.4594
MLP_ENDO_1 1A FASE	31962	32370	408	1.2754
MLP_ENDO_1 2A FASE	58653	60039	1385	2.3611
<b>MODELOS EXÓGENOS</b>				
RNN_EXO_1 1A FASE	31962	28729	-3233.	-10.1157
RNN_EXO_1 2A FASE	58653	62048	3395	5.7883
RNN_EXO_2 1A FASE	31962	30823	-1139	-3.5631
RNN_EXO_2 2A FASE	58653	63161	4507	7.6849
RNN_EXO_1 1A FASE	31962	29426	-2536	-7.9350
RNN_EXO_3 2A FASE	58653	58348	-305	-0.5196
MLP_ENDO_1 (**)	31962	31678	-285	-0.8901
RNN_EXO_1 (**)	31962	32170	208	0.6515

(\*\*) **MODELOS TREINADOS NA 2A FASE E TESTADOS NO DOMÍNIO DA 1A FASE**

Tabela 13: Erros quantitativos de todos os modelos

MODELO	TOTAL SOMA ERROS POSITIVOS	TOTAL SOMA ERROS NEGATIVOS	ERRO ABS MEDIANO	ERRO ABS PERC MEDIO
<b>MODELOS ENDÓGENOS</b>				
RNN_ENDO_1 1A FASE	-3281	4316	74.0192	87.4987
RNN_ENDO_1 2A FASE	-7709	5950	58.4248	107.8793
RNN_ENDO_2 1A FASE	-2983	3479	46.7072	74.9353
RNN_ENDO_2 2A FASE	-8335	5133.	55.7799	101.2846
MLP_ENDO_1 1A FASE	-4306	3898	71.8950	92.5815
MLP_ENDO_1 2A FASE	-7097	5712	53.8804	98.5516
<b>MODELOS EXÓGENOS</b>				
RNN_EXO_1 1A FASE	-2709	5942	85.5910	90.9869
RNN_EXO_1 2A FASE	-8163	4768	55.2355	224.9068
RNN_EXO_2 1A FASE	-3045	4184	63.5989	88.2687
RNN_EXO_2 2A FASE	-9677	5170	64.6863	230.9423
RNN_EXO_1 1A FASE	-3418	5954	100.4442	98.7906
RNN_EXO_3 2A FASE	-8608	8913	85.1866	236.5670
MLP_ENDO_1(**)	-351	3797	65.6641	84.9684
RNN_EXO_1 (**)	-3455	3247	59.5414	83.2671

(\*\*) **MODELOS TREINADOS NA 2A FASE E TESTADOS NO DOMÍNIO DA 1A FASE**

Tabela 14: Métricas estatísticas e de treino de todos os modelos

MODELO	r_2 value	std_err	RMSE
<b>MODELOS ENDÓGENOS</b>			
RNN_ENDO_1 1A FASE	0,2277	0,0600	115,5925
RNN_ENDO_1 2A FASE	0,4034	0,0417	101,1817
RNN_ENDO_2 1A FASE	0,3545	0,0721	108,0663
RNN_ENDO_2 2A FASE	0,3933	0,0474	105,3284
MLP_ENDO_1 1A FASE	0,2716	0,0947	128,0541
MLP_ENDO_1 2A FASE	0,4354	0,0484	101,1515
<b>MODELOS EXÓGENOS</b>			
RNN_EXO_1 1A FASE	0,1699	0,0551	124,4907
RNN_EXO_1 2A FASE	0,4508	0,0445	99,3650
RNN_EXO_2 1A FASE	0,2710	0,0639	112,9921
RNN_EXO_2 2A FASE	0,3539	0,0417	107,8493
RNN_EXO_1 1A FASE	0,1153	0,0306	124,6581
RNN_EXO_3 2A FASE	0,1953	0,0366	117,0316
MLP_ENDO_1 (**)	0,2896	0,0788	116,6204
RNN_EXO_1 (**)	0,3485	0,0658	106,2080

(\*\*) MODELOS TREINADOS NA 2A FASE E TESTADOS NO DOMÍNIO DA 1A FASE

Tabela 15: Métricas gráficas de todos os modelos

MODELO	CORRELAÇÃO	p-value	slope	intercept
<b>MODELOS ENDÓGENOS</b>				
RNN_ENDO_1 1A FASE	0,4772	2,5815E-06	0,3022	241,6487
RNN_ENDO_1 2A FASE	0,6351	5,9900E-22	0,4601	183,6360
RNN_ENDO_2 1A FASE	0,5954	9,4221E-10	0,4956	177,5496
RNN_ENDO_2 2A FASE	0,6271	2,7285E-21	0,5125	174,6988
MLP_ENDO_1 1A FASE	0,5212	1,9211E-07	0,5365	172,9566
MLP_ENDO_1 2A FASE	0,6599	3,9898E-24	0,5704	146,0429
<b>MODELOS EXÓGENOS</b>				
RNN_EXO_1 1A FASE	0,4122	6,5907E-05	0,2313	242,4375
RNN_EXO_1 2A FASE	0,6714	3,2984E-25	0,5421	166,2111
RNN_EXO_2 1A FASE	0,5206	1,9952E-07	0,3613	219,0125
RNN_EXO_2 2A FASE	0,5948	8,3516E-19	0,4149	213,3027
RNN_EXO_1 1A FASE	0,3396	0,00126742	0,1025	297,1444
RNN_EXO_3 2A FASE	0,4419	4,2231E-10	0,2424	242,4514
MLP_ENDO_1 (**)	0,5382	6,36258E-08	0,4668	190,4162
RNN_EXO_1 (**)	0,5903	1,41433E-09	0,4468	203,2738

(\*\*) MODELOS TREINADOS NA 2A FASE E TESTADOS NO DOMÍNIO DA 1A FASE



# Referências

- ALMEIDA, V. A. d. *Previsão De Carga Através De Modelos Neuro-fuzzy*. 91 p. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2013. Disponível em: <<http://monografias.poli.ufrj.br/monografias/monopoli10007620.pdf>>. Citado 4 vezes nas páginas 27, 29, 33 e 40.
- BARRETT, C. B. Measuring food insecurity. *Science*, American Association for the Advancement of Science, v. 327, n. 5967, p. 825–828, 2010. Citado na página 21.
- BRAGA A. DE P.; CARVALHO, A. P. L. F. L. *Redes neurais artificiais: teoria e aplicações*. [S.l.]: Livros técnicos e científicos, 2000. Citado na página 30.
- BROWNLIE, J. Ebook machine learning mastery. 2020. Access date: 3 oct. 2020. Disponível em: <<https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>>. Citado 2 vezes nas páginas 33 e 34.
- Data Science Academy. Deep learning book. 2020. Access date: 3 oct. 2020. Disponível em: <<http://deeplearningbook.com.br/arquitetura-de-redes-neurais-gated-recurrent-unit-gru/>>. Citado 2 vezes nas páginas 31 e 35.
- DAVILA, V. H. *Introdução às Séries Temporais*. 201–. Diapositivas. Disponível em: <<https://www.ime.unicamp.br/~hlaachos/MaterialSeries.pdf>>. Citado na página 26.
- EHLERS, R. *Análise de Séries Temporais*. 2009. Disponível em: <<https://sites.icmc.usp.br/~ehlers/stemp/cap2.pdf>>. Citado na página 27.
- GALANAKIS, C. M. The food systems in the era of the coronavirus (covid-19) pandemic crisis. *Foods*, Multidisciplinary Digital Publishing Institute, v. 9, n. 4, p. 523, 2020. Citado na página 21.
- HAYKIN, S. *Redes Neurais, Princípios e Práticas, 2nd Ed.* [S.l.]: Bookman, 2007. Citado 5 vezes nas páginas 27, 28, 29, 31 e 32.
- HINTON, G. E. et al. *Improving neural networks by preventing co-adaptation of feature detectors*. 2012. Citado na página 35.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 34.
- HUFF, A. G. et al. How resilient is the united states' food system to pandemics? *Journal of environmental studies and sciences*, Springer, v. 5, n. 3, p. 337–347, 2015. Citado na página 22.
- JUNIOR, A. M. *Análise de métodos de previsão de demanda baseados em séries temporais em uma empresa do setor de perfumes e cosméticos*. Dissertação (Mestrado) — Pontifícia Universidade do Paraná. Curitiba, 2007. Citado 4 vezes nas páginas 25, 26, 27 e 37.
- KAPLAN, A.; HAENLEIN, M. Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, Elsevier, v. 62, n. 1, p. 15–25, 2019. Citado na página 27.

- KOHAVI, R. Study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. [S.l.: s.n.], 1995. p. 1137–1143. Citado na página 31.
- KYUNGHYUN, C. et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. [S.l.: s.n.], 2014. Citado na página 34.
- LANDIM, D. D. Análise de demanda no restaurante universitário do instituto de ciência e tecnologia da unifesp - universidade federal de são paulo. *Projeto da Disciplina de Simulação de Sistemas*, Universidade Federal de São Paulo, v. 1, n. 1, p. 6, 2016. Disponível em: <<https://github.com/ddlandim/monografy-ann-demand-prediction/tree/master/referencias>>. Citado na página 37.
- LEMOS, E. *Análise de crédito bancário com uso de data mining: Redes Neurais e Árvores de Decisão*. Dissertação (Mestrado) — Universidade Federal do Paraná, 2003. Citado 2 vezes nas páginas 27 e 28.
- LOPES, L. Modelos para previsão de demanda no restaurante universitário utilizando técnicas de redes neurais. *Congresso Nacional De Excelência Em Gestão. Universidade Federal De Viçosa.*, Universidade Federal De Viçosa, IV, n. 1, p. 14, 2008. Disponível em: <<http://www.novos cursos.ufv.br/projetos/ufv/gesep/www/wp-content/uploads/5>> Citado 6 vezes nas páginas 22, 23, 38, 39, 44 e 59.
- LUGER, G. F. *Inteligência Artificial. Estruturas e Estratégias para a Solução de Problemas Complexos, 4a ed.* [S.l.]: Porto Alegre: Bookman, 2004. Citado na página 27.
- MARIANO, F. C. M. Q. *Redes Neurais na predição de valores energéticos de alimentos para aves*. 130 p. Tese (Doutorado) — Universidade Federal de Lavras, 2014. Citado na página 29.
- MC AI. Overview. 2020. Access date: 3 oct. 2020. Disponível em: <<https://mc.ai/complete-guide-of-activation-functions/>>. Citado na página 30.
- MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 127–147, 1943. Citado 2 vezes nas páginas 27 e 28.
- MINSKY, M.; PAPERT, S. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press, 1969. Citado na página 28.
- MOREIRA, D. A. Produção e operações. *São Paulo: Pioneira*, 1998. Citado na página 25.
- MORETTIN, P. A. *Análise de séries temporais*. [S.l.]: Editora Blucher, 1987. Citado na página 26.
- ORGANIZATION, W. H. et al. Global prevalence of vitamin a deficiency in populations at risk 1995-2005: Who global database on vitamin a deficiency. World Health Organization, 2009. Citado na página 21.
- ROCHA, J. C.; MATOS, F. D.; FREI, F. Utilização de redes neurais artificiais para a determinação do número de refeições diárias de um restaurante universitário. *Rev. Nutr., Campinas*, Universidade Estadual Paulista Júlio de Mesquita Filho., n. 1, p. 8, 2011. Citado 5 vezes nas páginas 22, 23, 38, 39 e 44.

- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, p. 386–408, 1958. Citado na página 28.
- RUAS, G. I. e. S. Previsão de demanda de energia elétrica utilizando redes neurais artificiais e support vector regression. *Copel Distribuição S/A (COPEL)*, Universidade Federal do Paraná(UFPR), v. 1, n. 1, p. 10, 2012. Disponível em: <[http://www.dcc.fc.up.pt/~ines-enia07\\_html/pdf/27927.pdf](http://www.dcc.fc.up.pt/~ines-enia07_html/pdf/27927.pdf)>. Citado 2 vezes nas páginas 27 e 39.
- RUMELHART, D.; HINTON, G.; MCCLELLAND, J. A general framework for parallel distributed processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, v. 1, 1986. Citado na página 29.
- RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. *Nature*, v. 323, p. 533–536, 1986. Citado na página 32.
- SHAW, D. J. World food summit, 1996. In: *World Food Security*. [S.l.]: Springer, 2007. p. 347–360. Citado na página 21.
- SILVA, C. e. a. Metodologia para previsão de demanda de energia elétrica em curtíssimo prazo baseada em um modelo probabilístico discreto. *Desenvolvimento de Algoritmos e Software para Previsão de Carga na Área de Concessão da MuxEnergia - Horizonte de Curtíssimo Prazo*, Universidade Federal do Pampa - UNIPAMPA, v. 1, n. 1, p. 4, 2010. Citado 2 vezes nas páginas 27 e 40.
- SRIVASTAVA, N. et al. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, v. 15, p. 1929–1958, 06 2014. Citado 2 vezes nas páginas 35 e 36.
- WEBB, P. et al. Measuring household food insecurity: why it's so important and yet so difficult to do. *The Journal of nutrition*, Oxford University Press, v. 136, n. 5, p. 1404S–1408S, 2006. Citado na página 21.