



Douglas Diniz Landim

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

São José dos Campos, SP

Douglas Diniz Landim

ANÁLISE DE DEMANDA DO RESTAURANTE UNIVERSITÁRIO DO ICT UNIFESP VIA REDES NEURAIS

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

Universidade Federal de São Paulo – UNIFESP

Instituto de Ciência de 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 NEURAIS

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

Trabalho para apresentar em Outubro/2020:

Prof. Dr. Marcos Gonçalves Quiles
Orientador

Professor
Convidado 1

Professor
Convidado 2

Professor
Convidado 3

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 Otávio 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 previsões 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 apre-

sentando 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 underprojection 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 the behavior meals consumption. In this work, machine learning models were developed, more specifically, multilayer perceptron neural networks and gated recurrent units networks, with methods of analysis of collected data, 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.

Listas de ilustrações

Figura 1 – Tipos de métodos quantitativos.	33
Figura 3 – Métodos de previsão de demanda.	33
Figura 5 – Dados de demanda de sazonalidade diária, em função do horário.	35
Figura 7 – Neuronio Biológico	36
Figura 9 – Neuronio Artificial.	37
Figura 11 – Problemas Linearmente Separáveis e Não Separáveis.	38
Figura 13 – Problemas Linearmente Não Separáveis. Múltiplas Soluções.	39
Figura 14 – Neurônio Artificial Perceptron.	39
Figura 16 – Funções de Ativação.	40
Figura 18 – Rede de perceptrons com múltiplas camadas.	41
Figura 20 – Ponto ótimo de parada da validação cruzada.	43
Figura 22 – Fases de treino da MLP-Back-Propagation.	44
Figura 24 – Rede Neural Perceptron de Múltiplas Camadas.	44
Figura 26 – Rede Neural Perceptron de Múltiplas Camadas.	45
Figura 28 – Arquitetura do modelo GRU	49
Figura 30 – Entradas de dia da semana em cofatores.	57
Figura 32 – Funções de Ativação	61
Figura 34 – Comparison of Adam to Other Optimization Algorithms Training a Multi-layer Perceptron	62
Figura 36 – Domínio temporal da 1a fase	66
Figura 37 – Resultado do modelo RNN_ENDO_2 obtido sobre o conjunto de dados aleatoriamente ordenado sobre o tempo	66
Figura 38 – Resultado do modelo RNN_ENDO_2 obtido sobre o conjunto de dados com ordenação corrigida	67
Figura 39 – 1a fase, Conjunto verdade de treino	67
Figura 41 – 1a fase, Conjunto verdade de validação	67
Figura 43 – 1a fase, Conjunto verdade de teste	68
Figura 45 – Correlação entre consumo e vendas de almoço.	68
Figura 46 – Gráfico Scatter entre consumo e vendas de almoço.	68
Figura 47 – Vendas de tickets normalizados com teto de 3x o desvio padrão.	70
Figura 48 – Vendas de tickets escalada entre 0 a 1.	70
Figura 49 – Correlação entre consumo e produção com 30% acima do consumo da semana anterior.	70
Figura 50 – Gráfico Scatter entre consumo e produção com margem de 30%. acima do consumo da semana anterior	71

Figura 51 – Correlação de consumo de almoço e jantar de 1 dia anterior.	72
Figura 52 – Gráfico Scatter entre consumo e jantar de 1 dia anterior.	72
Figura 53 – Gráfico violino da distribuição do consumo na segunda feira.	73
Figura 54 – Gráfico violino da distribuição do consumo na terça feira.	73
Figura 55 – Gráfico violino da distribuição do consumo na quarta feira.	73
Figura 56 – Gráfico violino da distribuição do consumo na quinta feira.	73
Figura 57 – Gráfico violino da distribuição do consumo na sexta feira.	73
Figura 58 – Relação da distribuição do consumo com o avanço do semestre.	74
Figura 60 – Gráfico scatter da distribuição do consumo com o avanço do semestre.	74
Figura 61 – Relação da distribuição do consumo com o avanço do mês.	74
Figura 63 – Gráfico scatter da distribuição do consumo com o avanço do mes.	75
Figura 64 – Correlação da distribuição do consumo com a temperatura.	75
Figura 65 – Gráfico scatter da distribuição do consumo com a temperatura.	75
Figura 66 – Correlação da distribuição do consumo com a umidade.	76
Figura 67 – Gráfico scatter da distribuição do consumo com a umidade.	76
Figura 68 – Correlação da distribuição do consumo com a velocidade do vento em m/s. .	76
Figura 69 – Gráfico scatter da distribuição do consumo com a velocidade do vento. . . .	76
Figura 70 – Correlação da distribuição do consumo com a pressão atmosférica.	76
Figura 71 – Gráfico scatter da distribuição do consumo com a pressão atmosférica. . . .	76
Figura 72 – Topologia do modelo MLP1	77
Figura 73 – Gráfico de treino do modelo MLP1.	77
Figura 75 – Topologia do modelo MLP2	78
Figura 76 – Gráfico de treino do modelo MLP2.	78
Figura 78 – Topologia do modelo MLP_ENDO_1	79
Figura 79 – Treino do modelo MLP_ENDO_1	79
Figura 80 – Avaliação do modelo MLP_ENDO_1	80
Figura 81 – Topologia do modelo RNN_ENDO_1	80
Figura 82 – Treino do modelo RNN_ENDO_1	80
Figura 83 – Avaliação do modelo RNN_ENDO_1	81
Figura 84 – Topologia do modelo RNN_ENDO_2	82
Figura 85 – Treino do modelo RNN_ENDO_2	82
Figura 86 – Avaliação do modelo RNN_ENDO_2	82
Figura 87 – Comparação da etapa de validação entre MLP_ENDO_1 e RNN_ENDO_1 .	83
Figura 88 – Topologia do modelo RNN_EXO_1	84
Figura 89 – Treino do modelo RNN_EXO_1	84
Figura 90 – Avaliação do modelo RNN_EXO_1	84
Figura 91 – Grafico scatter de avaliação do modelo RNN_EXO_1	85
Figura 92 – Topologia do modelo RNN_EXO_2	85
Figura 93 – Treino do modelo RNN_EXO_2	86

Figura 94 – Avaliação do modelo RNN_EXO_2	86
Figura 95 – Grafico scatter de avaliação do modelo RNN_EXO_2	86
Figura 96 – Topologia do modelo RNN_EXO_3	87
Figura 97 – Treino do modelo RNN_EXO_3	88
Figura 98 – Avaliação do modelo RNN_EXO_3	88
Figura 99 – Grafico scatter de avaliação do modelo RNN_EXO_3	88
Figura 100 –Teste MLP_ENDO_1	89
Figura 101 –Teste RNN_ENDO_2	89
Figura 102 –Analise de anomalias preditivas do RNN_ENDO_2	90
Figura 103 –Grafico scatter de teste do RNN_ENDO_2	91
Figura 104 –Teste do modelo RNN_EXO_1	91
Figura 105 –Grafico scatter de teste do modelo RNN_EXO_1	91
Figura 106 –Teste do modelo RNN_EXO_2	92
Figura 107 –Grafico scatter de teste do modelo RNN_EXO_2	92
Figura 108 –Teste do modelo RNN_EXO_3	93
Figura 109 –Grafico scatter de teste do modelo RNN_EXO_3	93
Figura 110 –Domínio temporal da 2a fase	94
Figura 111 –Conjunto de dados de treino, ano de 2017 completo	95
Figura 112 –Conjunto de dados de validação, ano de 2018 completo	95
Figura 113 –Conjunto de dados de teste, ano de 2019 completo	95
Figura 114 –Tecnica do restaurante para a 2a fase.	96
Figura 115 –Gráfico scatter da tecnica do restaurante para a 2a fase	96
Figura 116 –Gráfico de treino do modelo MLP_ENDO_1 na 2a fase.	97
Figura 117 –Gráfico de validação do modelo MLP_ENDO_1 na 2a fase.	97
Figura 118 –Gráfico de treino do modelo RNN_ENDO_1 na 2a fase.	97
Figura 119 –Gráfico de validação do modelo RNN_ENDO_1 na 2a fase.	97
Figura 120 –Gráfico de treino do modelo RNN_ENDO_2 na 2a fase.	98
Figura 121 –Gráfico de validação do modelo RNN_ENDO_2 na 2a fase.	98
Figura 122 –Gráfico de treino do modelo RNN_EXO_1 na 2a fase.	98
Figura 123 –Gráfico de validação do modelo RNN_EXO_1 na 2a fase.	98
Figura 124 –Gráfico de treino do modelo RNN_EXO_2 na 2a fase.	99
Figura 125 –Gráfico de validação do modelo RNN_EXO_2 na 2a fase.	99
Figura 126 –Gráfico de treino do modelo RNN_EXO_3 na 2a fase.	99
Figura 127 –Gráfico de validação do modelo RNN_EXO_3 na 2a fase.	99
Figura 128 –Gráfico de teste do modelo MLP_ENDO_1.	100
Figura 129 –Gráfico de scatter do modelo MLP_ENDO_1.	100
Figura 130 –Gráfico de teste do modelo RNN_ENDO_1.	101
Figura 131 –Gráfico de scatter do modelo RNN_ENDO_1.	101
Figura 132 –Gráfico de teste do modelo RNN_ENDO_2.	101

Figura 133 –Gráfico de scatter do modelo RNN_ENDO_2.	101
Figura 134 –Gráfico de teste do modelo RNN_EXO_1.	102
Figura 135 –Gráfico de scatter do modelo RNN_EXO_1.	102
Figura 136 –Gráfico de teste do modelo RNN_EXO_2.	102
Figura 137 –Gráfico de scatter do modelo RNN_EXO_2.	102
Figura 138 –Gráfico de teste do modelo RNN_EXO_3.	103
Figura 139 –Gráfico de scatter do modelo RNN_EXO_3.	103
Figura 140 –Gráfico de teste do RNN_EXO_1 segunda fase no domínio da primeira fase	104
Figura 141 –Gráfico scatter Gráfico de teste do RNN_EXO_1 segunda fase no domínio da primeira fase	104

Listas de tabelas

Tabela 1 – Formato dos dados originais obtidos pelo restaurante universitário	56
Tabela 2 – Transformação dos registros do restaurante em uma série temporal	56
Tabela 3 – Estrutura final do conjunto de dados indexados por data	58
Tabela 4 – Comparação de consumo com um dia anterior	69
Tabela 5 – Estimações do restaurante para o 1o Semestre de 2019	71
Tabela 6 – Métricas de avaliação do modelo MLP_ENDO_1.	80
Tabela 7 – Métricas de avaliação do modelo RNN_ENDO_1.	81
Tabela 8 – Métricas de avaliação do modelo RNN_ENDO_2	83
Tabela 9 – Métricas do modelo RNN_EXO_1	85
Tabela 10 – Métricas do modelo RNN_EXO_2	86
Tabela 11 – Métricas do modelo RNN_EXO_3	88
Tabela 12 – Métricas dos melhores modelos RNN_ENDO_2 x MLP_ENDO_1	89
Tabela 13 – Métricas do melhor modelo: RNN_ENDO_2	90
Tabela 14 – Erros do modelo RNN_EXO_1	92
Tabela 15 – Erros do modelo RNN_EXO_2	93
Tabela 16 – Erros do modelo RNN_EXO_3	93
Tabela 17 – Métricas da técnica do restaurante para a 2a fase.	96
Tabela 18 – Tabela de teste do RNN_EXO_1 segunda fase no domínio da primeira fase .	104
Tabela 19 – Previsões e erros de todos os modelos	105
Tabela 20 – Erros quantitativos de todos os modelos	105
Tabela 21 – Métricas estatísticas e de treino de todos os modelos	106
Tabela 22 – Métricas gráficas de todos os modelos	106

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

1	Introdução	25
1.1	Contextualização e Motivação	25
1.2	Definição do problema	27
1.3	Justificativas	27
1.4	Objetivos	27
1.5	Metodologia e Experimentos	27
2	Fundamentação Teórica	31
2.1	Análises dos Dados	31
2.1.1	Tipos de dados	31
2.1.2	Análise Exploratória dos Dados	31
2.1.3	Métodos de Previsão	32
2.1.4	Métodos de previsão de Demanda	33
2.1.5	Séries Temporais	34
2.1.6	Componentes Temporais	34
2.2	Inteligencia Artificial	35
2.2.1	Neurônio Artificial	35
2.2.2	Redes Neurais Artificiais	38
2.2.3	Rede Perceptrons Múltiplas Camadas - MLP	41
2.2.4	Perceptrons Múltiplas Camadas com Retro Propagação de Erro	43
2.2.5	Treino da MLP - BackPropagation para 1 sinal de saída	46
2.2.6	Redes Recorrentes GRU	48
3	Trabalhos relacionados	51
3.1	Previsão de demanda em restaurantes universitários	51
3.2	Previsão de demanda em outros ambientes	52
3.3	Revisão literária para previsão de demanda	53
4	Metodologia	55
4.1	Pré-processamento	55
4.1.1	Dados endógenos	55
4.1.2	Dados exógenos	56
4.1.3	Formato final do conjunto de dados	58
4.1.4	Tratamento dos dados para entrada nos modelos	58
4.1.5	Fases Experimentais	59
4.2	Definição e treino dos modelos	60

4.2.1	Topologia	60
4.2.1.1	Modelos endógenos	60
4.2.1.2	Modelos Mistos : Endógenos e Exógenos	60
4.2.2	Função de ativação	61
4.2.2.1	Para a entrada e camadas ocultas	61
4.2.2.2	Para a saída	61
4.2.3	Otimizador	62
4.3	Teste e Métricas	63
5	Resultados	65
5.1	Fase Experimental I	65
5.1.1	Pré-Processamento	66
5.1.1.1	Divisão dos conjuntos	66
5.1.1.2	Análise das variáveis endógenas	68
5.1.1.3	Análise da sazonalidade semanal	72
5.1.1.4	Análise das variáveis exógenas	73
5.1.2	Definição e treino dos modelos	76
5.1.2.1	Ajuste empírico de topologia com os modelos perceptron	77
5.1.2.2	Modelos Endógenos	79
5.1.2.3	Modelos mistos endógenos e exógenos	84
5.1.3	Teste e métricas	88
5.1.3.1	Endógenos : GRU x MLP	89
5.1.3.2	Melhor modelo Endógeno	89
5.1.3.3	Modelos mistos	91
5.1.4	Melhor modelo da 1a fase	94
5.2	2a Fase Experimental	94
5.2.1	Pré-Processamento	94
5.2.2	Treino e Validação dos modelos	96
5.2.3	Teste	100
5.2.4	Melhor Modelo	103
5.2.5	Tabela completa de todos os modelos	104
6	Conclusão	107
6.1	A importância da metodologia de divisão do conjunto de dados em séries temporais	107
6.2	Sobre o método de produção de refeições com margem de erro e análise da semana anterior	107
6.3	Sobre a sazonalidade semanal	107
6.4	Sobre o ajuste empírico da topologia dos modelos	107
6.5	Sobre os resultados preliminares da etapa de validação	108

6.6	Sobre os erros anômalos de predição	108
6.7	Sobre o modelo com melhor resultado	108
6.8	Conclusões gerais	109
Referências	111
7 Anexos	113
7.1	1a Fase Experimental	113
7.1.1	Pré-Processamento e treino dos modelos	113
7.1.2	Importação e aplicação de métricas dos modelos	113
7.2	2a Fase Experimental	113
7.2.1	Pré-Processamento e treino dos modelos	113
7.2.2	Importação e aplicação de métricas dos modelos	113

1 Introdução

1.1 Contextualização e Motivação

A previsão da demanda é o fator principal da eficiência de qualquer modelo de negócio produtivo. Sendo um ponto de extrema importância para qualquer empresa ou indústria, permitindo o ajuste de todo o mecanismo de operações para atendê-la com a melhor eficiência possível, maximizando lucros, minimizando perdas e, principalmente, atendendo todas as necessidades do cliente.

Todo restaurante universitário enfrenta problemas de previsão de demanda de refeições e prejuízos com a falta de vendas e ou o descarte de refeições não vendidas. Um dos grandes problemas enfrentados hoje no mundo é a elevação dos preços dos alimentos. O valor, além de monetário, é moral pois a alimentação é o recurso primitivo de base da humanidade que hoje ainda enfrenta um difícil acesso a este em muitas regiões carentes. O descarte de alimentos, provocado por suas limitações e durabilidade, também afeta o meio-ambiente. Isso tem causado preocupações para a população em geral e também para empresas como restaurantes que sofrem diretamente os reflexos da variação no preço dos alimentos e na demanda. Atualmente, o restaurante universitário do ICT - UNIFESP não possui um sistema que ajude na gestão de compras dos alimentos e sofre dos problemas acima relatados.

No restaurante universitário do ICT - UNIFESP, as refeições são fornecidas de segunda a sexta-feira. O caso particular de restaurantes universitários envolve um fluxo de demanda influenciado pelo dia da semana, visto que a demanda é influenciada pela quantia de alunos presentes na universidade, que por sua vez é influenciada pela grade de aulas determinada semestralmente por dia da semana.

Devido às condições burocráticas no ambiente do restaurante que compreendem fidelidade de contrato e exclusividade, pois o restaurante se encontra em localização que o faz ser o único provedor de refeições ao público do campus. os parâmetros das predições não contemplarão fatores externos como por exemplo restaurantes concorrentes.

Outro ponto importante é o tratamento dos valores obtidos pelo banco de dados do ICT Unifesp, que retorna valores de vendas de tickets no ponto de venda do restaurante, e valores da passagem dos alunos no ponto de acesso ao refeitório, portanto este trabalho prevê a quantidade de alunos que acessam o restaurante e utiliza demais parâmetros como suporte à predição.

O estudo da relação de vendas, temperatura, outras variáveis climáticas e do ambiente, já é comum em outros cenários; dentre eles, o de maior destaque é na demanda de energia elétrica. Os cenários de vendas de alimentos perecíveis ganha também destaque apesar de se encontrar investimentos maiores de previsão de demanda na indústria de produção de energia

elétrica. O objetivo, tanto no cenário deste trabalho - o restaurante universitário, como em outros cenários é o mesmo: atender toda a demanda de consumo e evitar transtorno a qualquer consumidor pela falta do produto desejado e evitar prejuízos de produção não consumida. Tais prejuízos impactam não só o fornecedor, mas o consumidor, visto que um fornecimento de produto e serviço com um bom planejamento de demanda poupa recursos ao produtor. Tais recursos podem ser investidos em melhor qualidade de produto e menor preço ao consumidor caso seja obtido um modelo de previsão viável para evitar sobreestimação de demanda com consequência de desperdício de alimentos, ou subestimação com consequência de docentes ou discentes sem refeições.

Tal problema tem sido impactante e frequente para o restaurante que informa que em alguns dias no mês passa por sobreestimação e desperdício superior a 30% do total produzido.

O modelo analisado neste trabalho é o comportamento dos consumidores de um restaurante universitário (R.U), o qual projeta sua compra de insumos e alocação de recursos na entrada de seu modelo de negócio, e projeta sua saída, que é a produção de refeições em quantidade numérica e inteira distribuída em função do tempo em dias. O consumo feito por alunos não se comporta de maneira determinística, já que este consumo é facultativo. Portanto, nem todos os alunos que estão na universidade se alimentarão no R.U. todos os dias.

De acordo com o contrato presente entre o R.U e a universidade, o mesmo deve atender totalmente à demanda do público, sendo multado se caso algum consumidor fique sem alimentação. Porém, este mesmo contrato não trata refeições que não são consumidas; logo, o R.U. deve lidar integralmente o prejuízo de refeições produzidas acima da demanda de consumo. O que o R.U. faz com essas sobras está fora do escopo deste trabalho.

Tais refeições fornecidas aos alunos são, em parte, subsidiadas pela universidade. O período de dados obtido para este trabalho compreende até Agosto de 2018 e segue um modelo de contrato antigo no qual o restaurante recebe R\$2,50 do aluno e R\$9,14 da universidade totalizando R\$11,64 por refeição.

De agosto em diante, iniciando no segundo semestre de 2018, a universidade subsidia R\$5,44 da refeição do aluno e o mesmo R\$2,50, o R.U. recebe o total de R\$7,94. Logo, esta previsão de demanda corresponde também aos interesses da administração do campus local, que periodicamente deve realizar uma alocação de recursos financeiros para subsidiar todas estas refeições consumidas.

É necessário então entender e descobrir quais elementos influenciam este consumo, de que forma estes elementos exercem tal influência e em qual intensidade ela ocorre.

Neste capítulo o principal objetivo é definir todos os conceitos que são necessários para o entendimento total deste trabalho, analisando a forma que os dados são entendidos, e as ferramentas com os quais possam ser trabalhados.

1.2 Definição do problema

O problema definido neste trabalho é encontrar um modelo de previsão de demanda por meio de algoritmos de regressão via aprendizado de máquina.

1.3 Justificativas

As atuais abordagens de previsão do restaurante universitário, que envolvem análise exploratória semanal e dedução subjetiva dos dados de consumo de refeições, chamados de dados endógenos, executadas pelos responsáveis pelo restaurante são falhas por não serem calculadas considerando-se uma visão ampla de todo um histórico de dados de grande amostragem. Além disso, não são consideradas outras informações, chamadas de dados exógenos, como dados climáticos, dados de calendário anual, feriados próximos, dentre outros. Essas falhas causam desperdício de alimentos por super-estimação de demanda e prejuízo de refeições descartadas, e falta de atendimento por sub-estimação de demanda de consumo.

1.4 Objetivos

Objetivo geral Construir e comparar modelos de Redes Neurais Artificiais para a previsão da demanda de refeições do restaurante universitário do ICT-UNIFESP.

Objetivos específicos

- Estruturar o conjunto de dados, e realizar análises exploratórias sobre os dados;
- Construir, executar previsões, e realizar análises comparativas da qualidade dos modelos de redes neurais construídos;
- Apontar o melhor ou os melhores modelos para a solução do objetivo do trabalho;
- Indicar possíveis melhorias e trabalhos futuros com o modelo encontrado;

1.5 Metodologia e Experimentos

Análises dos dados Por se tratar de um trabalho de previsão de demanda, este trabalho realizou a coleta e tratamento dos dados de consumo do restaurante universitário da Unifesp, a coleta de dados climáticos que influenciam em tal demanda e o cálculo dos dados de calendário derivados das datas de coleta, por exemplo, o dia da semana, a estação do ano, o número do semestre (se par ou ímpar), horário, etc.

Após a estruturação dos dados, realizou-se uma análise exploratória fundamentada pelo trabalho de ([JUNIOR, 2007](#)), para previsões de demanda em geral. O modelo de dados de venda

estruturados recebeu o acréscimo de dados exógenos de variáveis climáticas, como possíveis fatores de influencia no consumo, conforme ocorre no trabalho de previsão de demanda de energia elétrica fundamentado pelo trabalho de (ALMEIDA, 2013; RUAS, 2012; SILVA, 2010).

Os trabalhos de (JUNIOR, 2007) e (SILVA, 2010) fundamentam também uma classificação de análises dos dados estruturados indicando os métodos de previsão de demanda para este modelo quantitativo que são a Regressão Linear Múltipla e RNAs.

Assim, a primeira investigação usa a análises estatísticas de regressão linear múltipla fundamentada pela pesquisa de (CLARICE G.B. DEMÉTRIO; ZOCCHI, 2011) por meio de métricas aplicadas pelo framework `scipy` e plot de gráficos scatter por meio da biblioteca `seaborn`, que além de realizar nos uma construção dos gráficos relacionando os dados exógenos considerados e o comportamento de consumo, obtendo uma regressão linear simples entre estes 2 valores supervisionados, também são retornadas outras métricas como medidas de correlação, *p-value*, R^2 , e outras métricas que serão documentadas para trabalhos futuros.

Aprendizado de máquina As análises de inteligência artificial, fundamentadas por trabalhos de previsão de demanda em R.U. da Universidade Federal de Viçosa, de acordo (LOPES, 2008) e na Universidade Estadual Paulista Júlio de Mesquita por (FERNANDO., 2011), concluem a segunda aplicação da técnica de redes neurais artificiais com o modelo de Multilayer Perceptron (MLP) fundamentado em (HAYKIN, 1994), incrementando-se na técnica a inclusão de variáveis climáticas utilizadas nos trabalhos de previsão de (ALMEIDA, 2013; RUAS, 2012; SILVA, 2010), modificando a topologia como em (LOPES, 2008) para receber tais dados.

E também são considerados modelos modernos, como as redes recorrentes GRU (Gated Recurrent Unit), fundamentados por (Data Science Academy, 2020).

Após a modificação da topologia, os modelos são duplicados e executados em 2 fases experimentas, sendo a primeira com os experimentos focados para validação e teste no primeiro semestre de 2019, e na segunda fase os experimentos contemplaram o ano de 2018 inteiro para validação e o ano de 2019 inteiro para teste.

Portanto na primeira fase experimental, considerou-se 2 modelos MLP para a exploração e aprendizado do problema. 3 modelos que tem parâmetros de entrada endógenos, e 3 modelos mistos que tem parâmetros endógenos e exógenos de entrada.

Na segunda fase experimental, alterou-se os domínios de treino, validação e teste dos modelos da primeira fase e coletou-se novos resultados.

Estes novos resultados foram comparados com a primeira fase experimental para decidir se o melhor modelo para cada semestre é treinado e validado com um semestre ou com o ano todo.

Resultados O capítulo de resultados demonstrou graficamente as análises de parâmetros de predição com a passagem de alunos no ponto de acesso do restaurante, bem como as previsões dos modelos sobre a real passagem. Através destas análises gráficas e de valores estatísticos retornados, selecionou-se os melhores modelos para o objetivo final do trabalho que é determinar um modelo de aprendizado de máquina para realizar previsões de consumo.

Análise dos resultados De acordo com o trabalho de ([MARIANO, 2014](#)), a comparação e a discussão dos resultados de uma predição com redes neurais artificiais deve ser feita com base em medidas de avaliação de erro absoluto médio (EAM), e raiz do erro quadrado médio (REQM). Além destas análises, para o tema deste trabalho, considerou-se os valores de refeições descartadas e em falta (quando o valor previsto é menor que o consumo real), bem como documentou-se outras métricas obtidas pela plataforma experimental (TensorFlow e bibliotecas seaborn, scipy, scikit-learning) para a realização de análises e trabalhos futuros.

Conclusões Por fim, conclui-se que a predição via modelos de redes neurais possibilita um tratamento de produção e predição de consumo mais eficiente do que uma produção baseada na simples análise do consumo da semana anterior. Também se concluiu os efeitos notórios dos parâmetros de previsões considerados, bem como o alto contraste dos resultados entre os modelos da primeira e segunda fase experimental. E se concluiu entre todos os modelos experimentados, que o modelo misto treinado, validado e executado no domínio temporal da segunda fase experimental obteve melhores resultados e que pode ser utilizado e melhorado em trabalhos futuros, para a realização da predição do consumo no restaurante universitário do ICT Unifesp.

2 Fundamentação Teórica

Na seção de Análises dos Dados, vemos que para realizar os experimentos com as redes neurais artificiais com o objetivo de solucionar o problema de predição de refeições no restaurante universitário do ICT Unifesp, foram necessárias algumas fundamentações teóricas para coletar e preparar os dados de entrada nas redes, realizar análises exploratórias nestes dados, estipular os métodos de previsão, e considerar o tratamento dos dados como uma série temporal.

Em seguida fundamentou-se a teoria para a construção das redes neurais artificiais, que neste trabalho construiu-se por meio de 2 modelos, as redes neurais MLP (Multilayer Perceptron) e as redes recorrentes GRU (Gated Recurrent Unit).

2.1 Análises dos Dados

2.1.1 Tipos de dados

Para este trabalho considerou-se 2 tipos de dados: Endógenos e Exógenos.

Endógenos Dados endógenos são os dados do domínio de predição, neste caso, são os valores de consumo de refeições representado pela quantificação da passagem dos alunos no ponto de acesso, a catraca, do restaurante universitário do ICT Unifesp. São considerados endógenos também os dados de consumo de jantar, e os dados de vendas de tickets de refeições, todos discretizados e somados por dias letivos. Além disso, neste trabalho, os dados endógenos são transformados em entrada das redes neurais, em formato de série temporais, fundamentada nas seções adiante.

Exógenos Os dados exógenos são todos os outros dados fora do domínio de predição, incluso parâmetros derivados das datas das observações como por exemplo o dado categórico que representa o dia da semana (segunda-feira à sexta feira), outros dados que derivam da data dos registros dos dados endógenos e por fim os dados climáticos.

2.1.2 Análise Exploratória dos Dados

(JUNIOR, 2007) cita, em seu trabalho, que dados coletados em modelos de previsão possuem informações que quando são projetadas graficamente evidenciam comportamentos que em alguns casos podem ser visualizados e generalizados de forma subjetiva pelos gestores dos dados. Em todos os casos, a análise exploratória é necessária para selecionar o melhor método de análise que enquadrou-se neste comportamento.

Somente a análise exploratória não é o suficiente e realizada nos intervalos ou critérios incorretos pode comprometer seriamente as conclusões do comportamento dos dados, e que por sua vez pode comprometer seriamente a decisão dos gestores responsáveis por estes dados, no cenário de uma previsão de demanda. Isto ocorre atualmente no cenário de previsão da demanda de refeições do ICT-UNIFESP, onde a universidade e o estabelecimento que fornece as refeições não tem nenhum modelo de previsão de demanda.

De acordo com o gestor da atual empresa que fornece refeições no ICT, a análise utilizada para se prever as refeições é observar dentro da semana o dia anterior de consumo. Em variações de 300 para 450 refeições aproximadamente, isso tem provocado um desperdício médio de 150 refeições diárias. Em geral, de acordo com o restaurante, todos os dias o mesmo trabalha com um erro e um descarte de 30% das refeições que são trazidas e consumidas ao campus. Estima-se então que no período de 2011 - 01/08/2018 os estabelecimentos tenham tido um prejuízo de R\$1.885.938,40, e de 30% de R\$78.386,85 no atual período de 01/08/2018 - 31/10/2018 totalizando o montante R\$1.964.325,25. Aproximadamente 2 milhões de reais em prejuízo acumulado desde 2011.

2.1.3 Métodos de Previsão

(JUNIOR, 2007) realiza, em seu trabalho, uma revisão bibliográfica extensa abordando principais métodos de previsão de consumo sazonais, no cenário de uma indústria cosmética. Tais métodos estatísticos de previsão se dividem em 2 ramificações, sendo quantitativos ou qualitativos. Métodos qualitativos fazem um julgamento dos dados expostos sem um sistema de processamento analítico para se produzir novos modelos ou dados, eles são úteis para sistemas de agrupamento, clusterização ou classificação de dados, sem fornecer novas informações numéricas ou modelos preditivos.

Métodos quantitativos que é o foco deste trabalho, são analíticos e se baseiam em um modelo matemático para realizar previsões.

Para realizar tais previsões os métodos quantitativos necessitam de um histórico de dados, para analisar padrões em seu comportamento e predizer o futuro que irá agir dentro deste padrão.

Estes métodos se ramificam em 2 tipos, as séries temporais e os modelos causais.

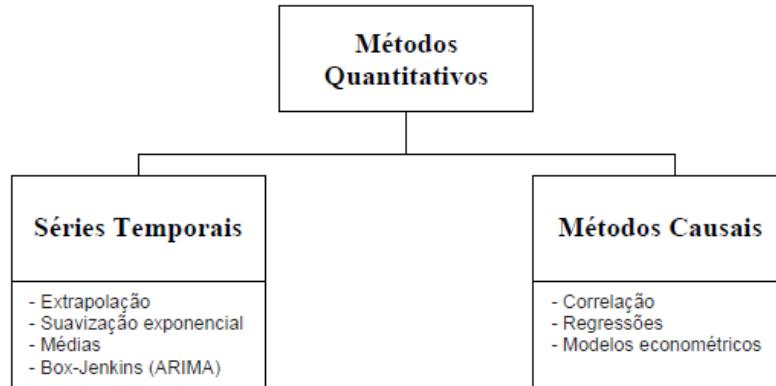


Figura 1: Tipos de métodos quantitativos.

Fonte: ([JUNIOR, 2007](#))

2.1.4 Métodos de previsão de Demanda

O autor supracitado também referencia métodos estatísticos especialmente selecionados para uma previsão de demanda, com a atenção de que alguns métodos qualitativos foram criteriosamente selecionados para prever uma demanda industrial, onde geralmente são previstas pelos métodos quantitativos.

O comportamento dos dados deste trabalho, apesar de ter uma distribuição de datas em função do tempo se classificando em um modelo de série temporal, assume-se a hipótese que tem tal comportamento impactado por relações causais com outras variáveis 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.

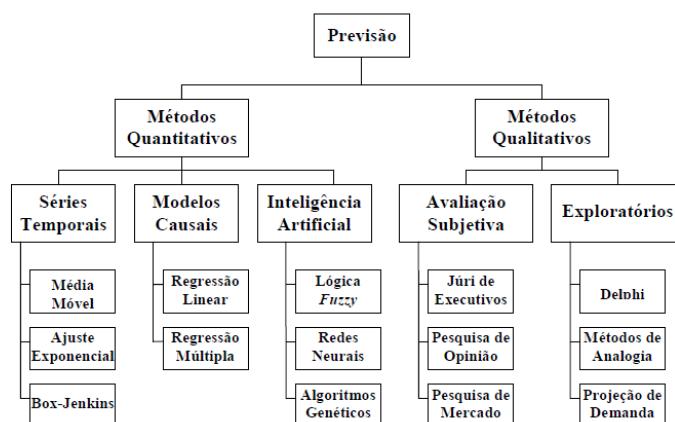


Figura 3: Métodos de previsão de demanda.

Fonte: ([JUNIOR, 2007](#)).

2.1.5 Séries Temporais

De acordo com ([MORETTIN, 1987](#)) em seus trabalhos, uma série temporal é um conjunto de observações ordenadas em função do tempo, comumente iguais, apresentando uma dependência serial entre elas, sendo um dos objetivos do estudo de séries temporais, analisar e modelar essa dependência. Além disso, séries temporais são analisadas pelos seus principais movimentos, como tendência, sazonalidade e a componente aleatória, sendo a tendência e sazonalidade as propriedades que mais ganham destaque em pesquisas de previsão de demanda de carga elétrica. Os meios mais comuns de se analisar a componente sazonal são o Método de Regressão (paramétrico), Método de Médias Móveis (não paramétrico), e Método de Diferença Sazonal (sazonalidade determinística). Neste trabalho, a distribuição dos dados do Restaurante se dá de forma paramétrica, ou seja, as vendas do restaurante se distribuem em função do tempo e com influências de parâmetros como o dia da semana por exemplo.

([ALMEIDA, 2013](#)) também cita em sua análise que para realizar a previsão de uma determinada série temporal é possível utilizar diferentes métodos. Pode-se classificá-los basicamente entre métodos estatísticos e baseados em inteligência artificial. Dentre os métodos estatísticos não abordados neste trabalho, destacam-se:

- Regressão Linear Múltipla;
- Alisamento Exponencial;
- Média Móvel Integrada Auto-Regressiva (ARIMA)

Entre os métodos baseados em inteligência artificial tem-se:

- Redes Neurais Artificiais;
- Lógica Fuzzy.

Dos quais abordamos as Redes Neurais Artificiais.

2.1.6 Componentes Temporais

É importante observar que os métodos de previsão em séries temporais buscam uma redução da série temporal à um modelo estacionário e a decomposição da série em componentes de tendência, ciclo, sazonalidade (próxima figura) e termo aleatório. A tendência entende-se pelo movimento persistente dos dados em uma direção, o ciclo indica o movimento oscilatório desta tendência, sazonalidade indica comportamento regular assumido de forma repetitiva e o termo aleatório se dá por movimentos irregulares presentes na série.

As técnicas de previsão para evidências sazonais, usam métodos de regressão, que podem ser observados nos trabalhos de previsão de consumo de energia elétrica realizados pelas

pesquisas de ([ALMEIDA, 2013](#)), ([RUAS, 2012](#)) e de ([SILVA, 2010](#)), além , da previsão de demanda de produtos cosméticos pelo trabalho de ([JUNIOR, 2007](#))

O cenário deste estudo analisa a sazonalidade da frequência de alunos do ICT - UNI-FESP, inferida pelas grades semestrais de suas disciplinas, que consequentemente inferem na sazonalidade de consumo de refeições dentro do restaurante universitário.

Na figura 5 temos um exemplo de sazonalidade diária de dados.

Através da Figura percebe-se a sazonalidade diária, com picos de consumo próximo às 09:00, 14:00 e 20:00 horas, e baixo consumo durante a madrugada, entre 00:00 e 06:00 horas.

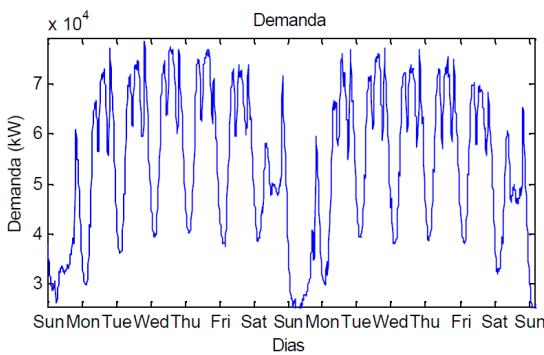


Figura 5: Dados de demanda de sazonalidade diária, em função do horário.

Fonte: ([RUAS, 2012](#))

2.2 Inteligencia Artificial

A Inteligência artificial "A inteligência artificial é o ramo da ciência da computação que se ocupa do comportamento inteligente."([LUGER, 2004](#)). O termo "Inteligencia Artificial" surgiu em meados de 1956 em uma conferência, chamada Dartmouth Summer Research Project on Artificial Intelligence, sediada nos Estados Unidos em Dartmouth College, Hanover, New Hampshire, a conferência foi organizada por John McCarthy e teve como proposta reunir matemáticos, cientistas da computação e pesquisadores buscando meios de como fazer com que as máquinas usem linguagem, abstrações de formulários e conceitos, para resolver problemas que eram reservados aos humanos, e que consigam melhorar o seu próprio desempenho.

Sistemas de inteligência artificial buscam então resolver funções e problemas que seres humanos conseguem resolver melhor do que máquinas convencionais, usando sua capacidade de abstração e aprendizagem com o erro.

2.2.1 Neurônio Artificial

História e inspiração da Inteligência Artificial Redes Neurais Artificiais são elementos de inteligência artificial da ciência da computação, inspirados no funcionamento do cérebro hu-

mano. São formadas por neurônios artificiais interconectados que são capazes de processar múltiplos valores de entradas, e reagir produzindo uma resposta relacionada à essas entradas. Como qualquer outro método de aprendizado de máquina, este modelo busca obter um aprendizado a partir dos dados de entrada recebidos, criando uma capacidade de generalização de problemas e assim buscam o objetivo principal de resolver novos problemas com este aprendizado.

Uma rede neural pode não produzir uma resposta esperada, resolvendo erroneamente um problema, assim como o cérebro humano tem limitações de aprendizado, que levam o homem a cometer falhas de decisões ou ações por um aprendizado mal treinado. Por isso as características fundamentais que tornam uma rede neural artificial em uma boa solução, é um bom planejamento de sua topologia, e método de treinamento, que serão explicados a seguir.

(BRYAN KOLB; ARIF, 2014) demonstra em sua pesquisa que a propriedade fundamental do cérebro é a sua capacidade de mudar com uma grande variedade de experiências, inclusive lesões que provocam perdas de neurônios, abordando princípios de plasticidade cerebral, mostra também que a capacidade de aprendizado de mamíferos que sofreram uma redução de neurônios causada por lesões, pode ser restaurada não somente pela recuperação desses neurônios perdidos, mas sim por novas conexões e sinapses entre outros neurônios. Ou seja, a rede sofre uma mudança de topologia.

(ZORZETTO, 2012) Mostra em um experimento recente que o cérebro humano possui atualmente 86 bilhões de neurônios interconectados, e que sua capacidade de aprendizado e habilidades evolutivas também vêm aumentando em conjunto com seu número de neurônios e topologia de rede neural, que há 30 milhões de anos atrás tinha apenas 2,5 bilhões de neurônios e era apenas um animal arborícola quadrúpede.

Neurônio Biológico O modelo de neurônio artificial surge então, com a busca da inteligência artificial de reproduzir o comportamento de aprendizado humano, reproduzindo computacionalmente seu elemento biológico principal de aprendizagem: O neurônio biológico.

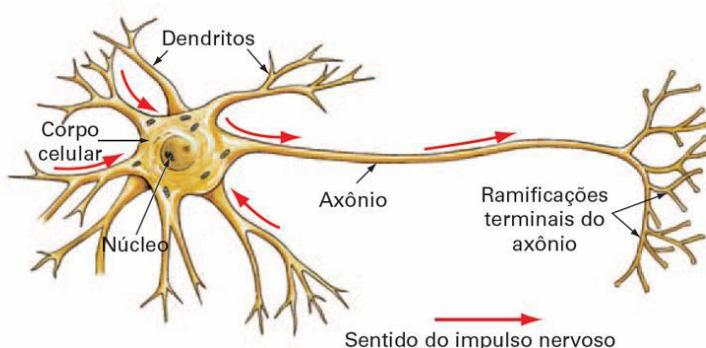


Figura 7: Neuronio Biológico

Fonte: <https://pt.khanacademy.org/science/biology/human-biology/neuron-nervous-system/v/anatomy-of-a-neuron>.

Um neurônio biológico figura 7 trata as informações recebidas por meio de seus dendritos e processa-as em seu corpo celular, tal reação à esses estímulos recebidos gera um sinal de saída como resposta aos estímulos, enviado através do axônio. Esse estímulo é repassado como sinal de entrada, através de outros neurônios por meio de seus dendritos, e o ciclo se repete em uma vasta rede. Controlando essas conexões, pontos de contato entre a resposta de um neurônio e a entrada de outro, existem as sinapses. Elas funcionam como agentes que permitem a interação acontecer ou inibem, e são acionadas por um conjunto somatório de estímulos. Se tal somatório de estímulos for satisfatório, elas permite a transmissão de sinal elétrico pelo axônio até o dendrito de um neurônio vizinho, formando um ciclo de aprendizado em uma rede neural biológica.

Neuronio Artifical - MCP Warren McCulloch e Walter Pitts em 1942 observando o neurônio biológico iniciam a busca de um modelo computacional do mesmo. 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. Em 1943 lançaram o artigo "A Logical Calculus of the Ideas Immanent in Nervous Activity." chegando em um modelo matemático do neurônio artificial:

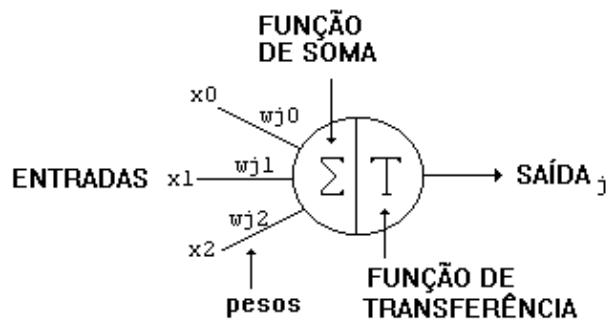


Figura 9: Neuronio Artificial.

Fonte:

<http://redesneuraisartificiais.blogspot.com/2010/10/o-primeiro-modelo-de-um-neuronio-criado.html>

Assim como no neurônio biológico, o modelo de neurônio artificial da figura 9 reage à um vetor de entradas x_0 à x_2 onde tem suas sinapses representadas por pesos numéricos w_{j0} à w_{j2} , com uma soma ponderada dessas entradas que é controlada por uma função de transferência ou função de ativação, determinando se essa soma é maior que um valor numérico. Se essa soma for satisfatória o neurônio é ativado emitindo um valor de saída 1, caso contrário se emite um valor de saída 0. Todo o funcionamento deste modelo então é reduzido a responder se a soma recebida é maior que um valor numérico esperado. Contudo, associado a este neurônio não foi proposto uma forma automática para ajuste dos pesos, ou seja, não há um algoritmo de aprendizagem (HAYKIN, 1994).

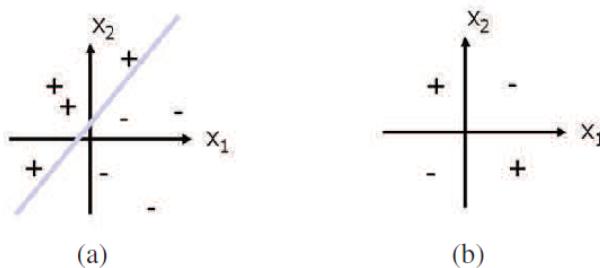
2.2.2 Redes Neurais Artificiais

Perceptron O modelo de Warren McCulloch e Walter Pitts, apesar de conseguir simular o modelo de neurônio biológico e resolver algumas tarefas lógicas e matemáticas não atendia o objetivo principal da Inteligência Artificial: A capacidade de aprendizado. Para utilizar um modelo de neurônio artificial a fim da solução de um problema era necessário conhecer o ajuste dos pesos das entradas, e em uma tarefa de um cenário complexo e com muitas variáveis, ou pesos não perceptíveis ao valor esperado de saída, o ajuste não se torna trivial.

Em 1958, o cientista da computação Frank Rosenblatt desenvolveu o primeiro modelo de rede neural artificial com um algoritmo de aprendizado para o neurônio MCP, denominada de Perceptron, para solucionar este problema de ajuste de pesos. Basicamente neste novo modelo, os pesos das conexões são ajustados de forma autônoma com a introdução de pessoas associados e um valor bias, a fim de buscar um reconhecimento autônomo de padrões. Tarefas de reconhecimento de padrões simples com separações lineares os seres humanos conseguem realizar de forma trivial mas ainda era um desafio para tal problema se resolvido por uma máquina.

Então em 1969, Marvin Minsky e Seymour Papert realizaram uma publicação comprovando essa limitação de aprendizado à uma combinação linear, e provaram que o perceptron é limitado à resolução de problemas que são resolvidos ou classificados por apenas 1 linha, ou seja, problemas linearmente separáveis que consiste na mesma limitação do processo de regressão linear múltipla.

Na figura 11 o perceptron não pode solucionar o problema (b). E essa limitação desmotivou e parou os estudos com perceptrons.



Exemplos linearmente separável (a) e não separável linearmente (b)

Figura 11: Problemas Linearmente Separáveis e Não Separáveis.

Fonte: ([MARIANO, 2014](#)).

Em 1986 James McClelland e David Rumelhart proporam o método de desenvolvimento de uma rede neural de perceptrons, tangendo mais ainda a inspiração da solução no cérebro humano. O processo resume o treinamento do perceptron simples aplicado à um conjunto de perceptrons interligados, e assim solucionando problemas complexos que podem ser

resolvidos com uma combinação de soluções.

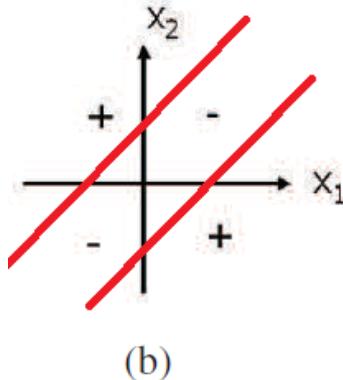


Figura 13: Problemas Linearmente Não Separáveis. Múltiplas Soluções.

A rede perceptron , conforme demonstrado no trabalho de([MARIANO, 2014](#)), possui apenas uma camada de entrada e saída, a saída utiliza como função de ativação a função degrau $\delta(\cdot)$ que define quando o neurônio emitirá o sinal lógico 1 ou quando emitirá o sinal lógico 0.

O sinal de saída do perceptron entende-se então por: $y_j = \delta(\sum_{i=1}^n x_i W_{ji} + b_j)$

- x_i - sinais de entrada do neurônio;
- w_{ji} - pesos sinápticos do neurônio;
- b_j - bias ou limiar de ativação;
- $\delta(\cdot)$ - função de ativação;
- y_j - sinal de saída do neurônio.

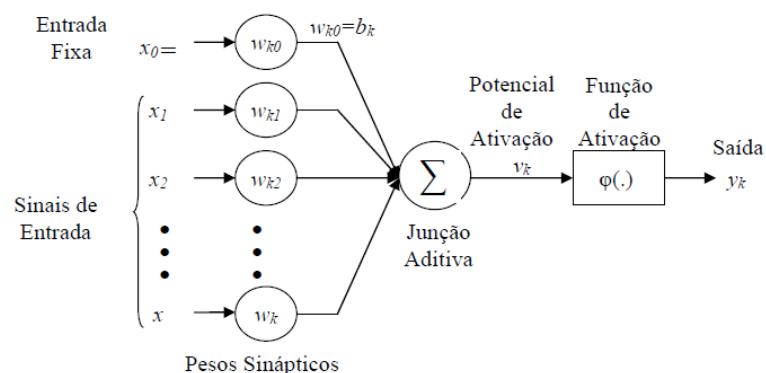
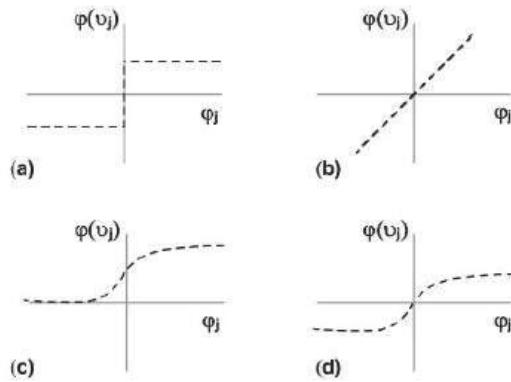


Figura 14: Neurônio Artificial Perceptron.

retirado de ([JUNIOR, 2007](#))

Em algumas literaturas o bias pode ser reduzido a um peso W_0 com entrada X_0 fixo em 1 no neurônio, a representação gráfica da topologia pode mudar, mas no somatório de saída, o cálculo continua o mesmo.

A função de ativação δ se apresenta de forma linear ou não linear, determinando a saída de um neurônio a partir do seu potencial de ativação.



Tipos de função de ativação: (a) função degrau, (b) função linear, (c) função sigmoide, (d) função tangente hiperbólica

Figura 16: Funções de Ativação.

Fonte: ([MARIANO, 2014](#)):

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.

Perceptron - Treino Supervisionado ([ALMEIDA, 2013](#)) cita em sua análise que o processo de aprendizado do perceptron pode ocorrer de forma supervisionada quando o neurônio deve aprender a relacionar um conjunto observado de variáveis à um valor observado de saída deste mesmo conjunto. O neurônio recebe os sinais de entrada X_i e produz uma saída Y_i através do combinador linear e função δ de ativação; compara essa saída Y_i com a observação i obtida do conjunto de dados (ponto chave do processo de treino supervisionado) e por fim essa comparação irá gerar um erro e .

Critério de parada De acordo com algum critério adotado em cada contexto de aplicação do perceptron, este erro pode ser aceitado e o neurônio mantém o valor de seus pesos W_i que impactam na saída Y_i desejada. A aprendizagem do neurônio também pode atingir um critério de parada após N épocas de treinamento, diversos critérios de paradas podem ser combinados.

Época de treinamento Se o critério de parada não for aprovado, uma nova época de treinamento (ou repetição de interação) é iniciada mesmos valores X_i e Y_i passados, os valores de pesos W_i são reajustados com a taxa de aprendizagem, buscando o objetivo de se obter um erro menor.

Taxa de aprendizagem Esse reajuste de pesos W_i denomina-se taxa de aprendizagem, α que pode ter valores de escolha livre ao contexto de aplicação do perceptron, para reajustar estes pesos W_i . Assim que a taxa de aprendizagem ajusta os pesos, uma nova época $N+1$ de treinamento está se iniciando buscando novamente um erro menor. Ressaltando então que o critério de parada pode ser acionado e interromper o reinicio do processo, se for estipulado um limite para o valor de N combinado ou não com um limite para o erro.

2.2.3 Rede Perceptrons Múltiplas Camadas - MLP

A solução base para se combinar 2 ou mais camadas de perceptrons a fim de se resolver um problema com a combinação de 2 ou mais soluções lineares, é a utilização de um perceptron combinador de sinal de saída, já que cada perceptron pode ter múltiplas entradas e somente uma saída. Dessa forma as redes neurais vão formando colunas de perceptrons interconectados. Cada coluna é denominada uma camada da rede neural. A ultima camada deve ter o número de perceptrons correspondente ao número de saídas desejadas. Na figura abaixo, encontra-se uma rede neural com duas camadas intermediárias e 3 saídas.

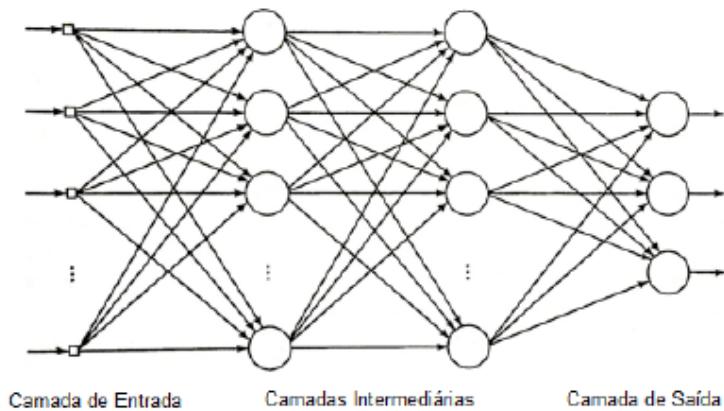


Figura 18: Rede de perceptrons com múltiplas camadas.

Fonte: ([ALMEIDA, 2013](#))

Esta rede denominada MLP (Multilayer Perceptron) possui uma camada de entrada onde cada nó representa uma variável a ser considerada ao problema a ser analisado, e pelo menos uma camada intermediária.

Nesta camada intermediária os neurônios possuem geralmente uma função de ativação sigmoidal logística ou tangente hiperbólica, e conceitualmente no mínimo 1 neurônio desta primeira camada oculta deve receber no mínimo 2 conexões de entrada. E uma camada de saída, na borda à direita, com o número de neurônios correspondente ao número de soluções buscadas.

Treino e Validação da MLP O conjunto de dados de entrada na rede MLP supervisionada, deve ser dividido em 2 partes principais, Treinamento e Validação. É importante salientar que as observações de ambos os conjuntos devem originar do mesmo conjunto de dados para representar o mesmo problema, as observações Y_i relacionadas aos vetores i de variáveis $X_{i1}, X_{i2}, \dots, X_{ij}$ do conjunto de treino, devem ter a mesma estrutura, mesmo número j de variáveis X por vetor de entrada e devem representar o mesmo problema que as observações Y_i do conjunto de validação.

O conjunto de dados da validação e treino somados devem formar exatamente o conjunto original, sem informações excedentes ou em falta.

Os dados de treino e dados de validação, podem ser separados de forma aleatória, sendo os dados de treino os responsáveis pelos reajustes de pesos e capacidade de generalização da rede, e os dados de validação responsáveis pelo processo de validação pós-treino.

É importante adotar um bom critério de parada de treino, pois um treino prolongado tende a convergir em ajustes de pesos memorizados dos valores observados nos dados de treino e isso causa o fenômeno de overfitting na rede neural, que é a perda da capacidade de generalização. Pesos sinápticos que sofrem processo prolongado de reajustes acaba "viciando" a rede neural para reconhecer apenas os dados de treino.

Parada do erro mínimo: O critério de parada do erro mínimo encerra o treinamento da RNA quando a mesma obtém um erro menor que o mínimo estipulado para o valor observado, este é o critério mais simples, adotado nos casos onde existe um limiar de erro já determinado pelo problema, entre o valor observado e o valor a se predizer.

Parada por número de épocas: Um critério de parada pode ser limitado também ao número de épocas de treino. A determinação deste número de épocas pode ocorrer por tentativa e erro, visto que haverá a convergência de um número pequeno para uma baixa capacidade de aprendizado, e a convergência de um número grande para o processo de overfitting. Logo é necessário realizar experimentos que validem número de épocas fora desses intervalos de convergência.

Parada por validação cruzada: Por fim a validação cruzada é a técnica que se utiliza os dados dos conjuntos de validação e treino de forma cruzada, neste processo então, os dados de treino são utilizados no processo interativo de aprendizagem, e no fim deste processo o conjunto é validado com os dados de validação, obtendo-se um novo erro de validação. A medição do erro de validação passa por um processo de avaliação em função do número de épocas, a fim de se detectar um ponto de número de épocas onde o erro quadrático médio da amostra de validação sofre uma curva de crescimento após um decréscimo.

É notório observar que o erro quadrático médio das amostras de treinamento sofrerá

um decréscimo em função do aumento do número de épocas de treinamento, convergindo ao processo de overfitting, ou memorização da rede.

O ponto de parada no limite inferior do erro quadrático médio da amostra de validação, será o ponto ótimo de parada de treinamento.

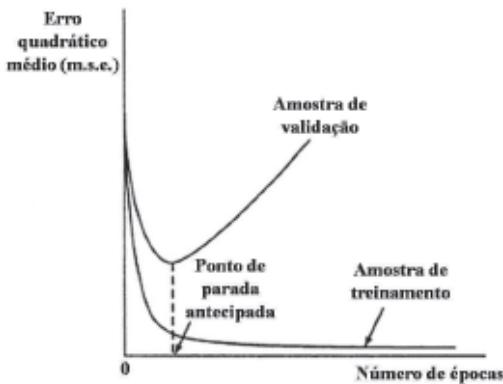


Figura 20: Ponto ótimo de parada da validação cruzada.

Retirado de ([MARIANO, 2014](#))

O treinamento de uma rede MLP para previsões de demanda de restaurantes universitários que obteve sucesso em ([LOPES, 2008](#)) e ([FERNANDO., 2011](#)) é a retro propagação de erro.

2.2.4 Perceptrons Múltiplas Camadas com Retro Propagação de Erro

A rede Perceptron Múltiplas Camadas com Retro Propagação de Erro (do inglês, M.L.P Backpropagation), faz o reajuste dos pesos sinápticos dos neurônios através de duas fases:

Feed-forward Nesta primeira fase de treino os sinais $X_{i0}, X_{i1}, \dots, X_{in}$ com sua respectiva saída Y_i do conjunto de dados são apresentados à todos os neurônios da primeira camada. O processo de propagação do sinal de saída de cada neurônio segue o princípio do neurônio artificial apresentado na seção anterior, que envia o sinal de saída como um sinal de entrada ao neurônio seguinte.

Feed-backward Nesta fase é obtido um valor de erro da camada de saída. Este erro é utilizado na equação de reajuste do peso sináptico das conexões dos neurônios da camada de saída com o sinal de saída dos neurônios da última camada oculta. E depois este erro é propagado realizando outra equação de reajuste de peso sináptico das conexões dos neurônios nas camadas anteriores, no sentido contrário em direção à camada de entrada. Isto permite que os pesos sináptico de todas as camadas de intermediárias neste processo tenham seus pesos ajustados.

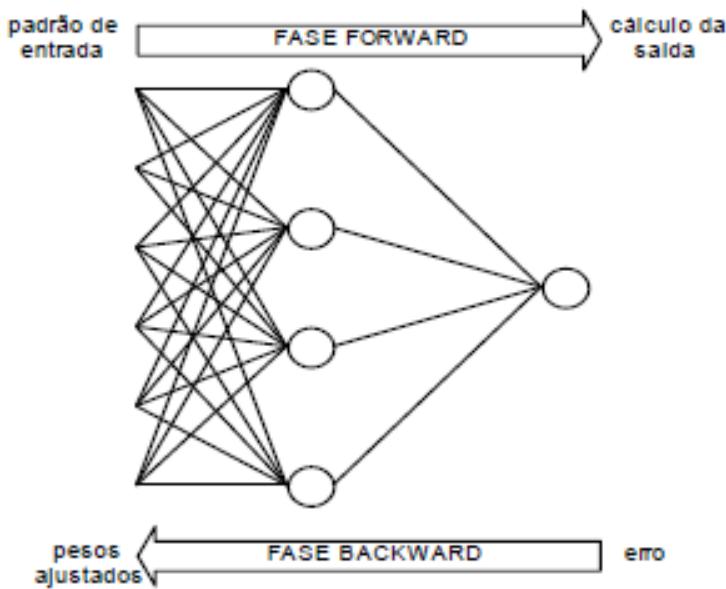


Figura 22: Fases de treino da MLP-Back-Propagation.

Fonte: ([ALMEIDA, 2013](#))

MLP Backpropagation em R.U em trabalhos relacionados De acordo com o trabalho de ([LOPES, 2008](#)), a rede neural perceptron de múltiplas camadas é utilizada para tratar a previsão de demanda do R.U da UFV, utilizando apenas como variáveis quantitativas as 5 últimas observações anteriores ao dia a se analisar, e como variáveis qualitativas o dia da semana variando de segunda a sexta, em valores binários.

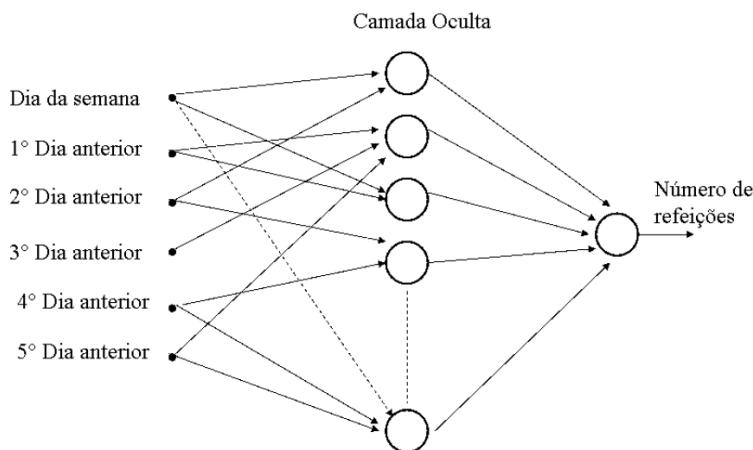


Figura 24: Rede Neural Perceptron de Múltiplas Camadas.

Fonte: ([LOPES, 2008](#)).

Conforme o trabalho realizado por ([FERNANDO., 2011](#)) com neurônios artificiais para prever a demanda do R.U da UNESP, envolve apenas uma única camada de entrada e uma segunda camada para saída, porém com uma diversidade maior de variáveis de entrada correlacionadas com o consumo do restaurante.

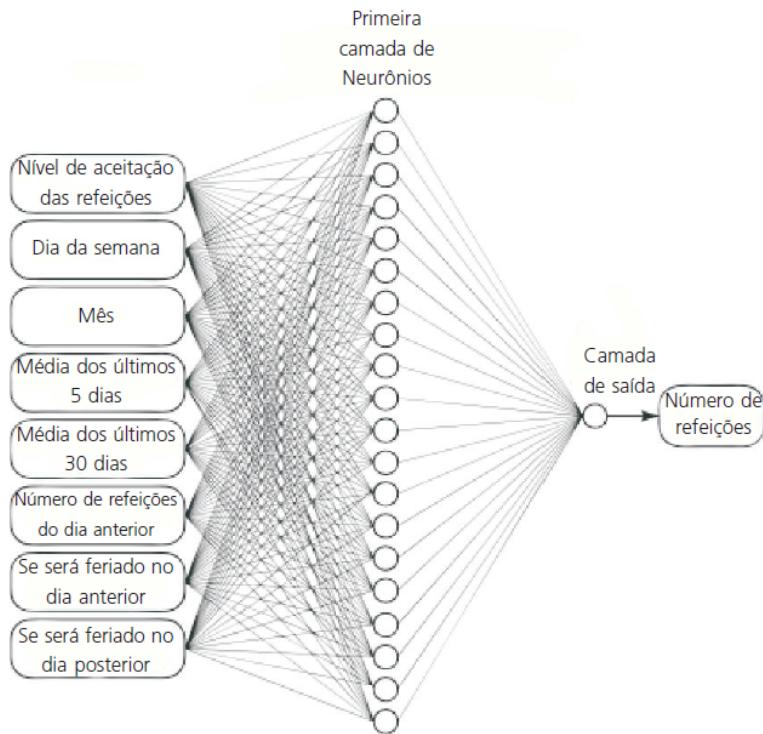


Figura 26: Rede Neural Perceptron de Múltiplas Camadas.

Fonte: ([FERNANDO., 2011](#))

Ambos os modelos possuem topologia de uma camada oculta para entrada dos dados, e uma camada de saída. que de acordo com a análise de([RAGA A. DE P.; CARVALHO, 2000](#)) é citado que através de uma análise de Cybenko, uma camada intermediária é o suficiente para aproximar qualquer função contínua e 2 camadas intermediárias são suficientes para aproximar qualquer função matemática, e devendo ser observado o fato de que em alguns casos, a utilização de 2 ou mais camadas pode facilitar o treinamento da rede, porém a utilização de um grande número de camadas intermediárias ou ocultas, é inviável, pois em cada uma delas a estimativa do erro se trata de uma estimativa da estimativa do erro da camada anterior, e este cascamenteamento de estimativas pode se tornar menos preciso à medida que cresce.

Em conformidade com o trabalho de ([MARIANO, 2014](#)), o número de neurônios na primeira camada oculta é proporcional à dimensão do espaço de observação. Logo em modelos preditivos de demanda, supracitados, observa-se que no mínimo, na primeira camada oculta, é utilizado um número de neurônios igual ao número de variáveis que influenciam no dado preditivo.

Parâmetros de treino da MLP Backpropagation A definição da função δ de ativação interfere na linearidade do modelo a ser analisado, sendo a função sigmoide a mais popular na literatura.

A taxa de aprendizagem, define a velocidade de reajuste dos pesos, podendo variar de 0 a 1. Ressalta-se que uma taxa próxima de 1 provoca picos oscilatórios na taxa de aprendizado, e taxa próxima de 0 provoca lentidão da convergência de aprendizagem. Valores comuns de utilização ficam entre 0,2 e 0,8.

2.2.5 Treino da MLP - BackPropagation para 1 sinal de saída

No caso do problema de predição de demanda onde se busca apenas um valor de saída, que é a previsão de vendas em relação às variáveis de entrada, o cálculo de reajuste dos pesos da camada de saída é reduzido à apenas 1 neurônio na camada de saída.

O treino da rede perceptron com backpropagation obtém o sinal de saída no último neurônio s da rede, através da propagação dos sinais de saída dos neurônios anteriores da rede, feito pela aplicação da combinação linear dos sinais de entrada com os pesos sinápticos em uma função de ativação. É calculado o erro quadrático deste sinal de saída e a regra do Gradiente Descendente com base neste erro é utilizada para o reajuste dos pesos sinápticos em todas as conexões de todos os neurônios. O processo é denominado regra delta generalizada, Δ

O principal objetivo de todo o método backpropagation conforme o avanço de épocas, é o treino em uma determinada época reduzir a média de erros quadráticos do conjunto de validação ($m.s.e$) em n iterações de validação desta época:

$m.s.e = \frac{1}{n} \sum_{k=1}^n e(k)$ É importante observar que 1 época consiste no par treino e validação. Logo após o treino, os erros $e(k)$ serão obtidos novamente através das entradas Y_i do conjunto de validação, e respostas propagadas Y_μ na camada de saída da rede já treinada, para então o $m.s.e$ ser calculado.

O valor $m.s.e$ será o critério de parada do treinamento de toda a rede e deve ser observado a cada época de treinamento, assim que este valor atingir um ponto ótimo, conforme figura 20 deste ponto ótimo o critério de parada será verdadeiro e o treinamento deve parar, pois a partir desse ponto a rede converge à um overfitting (divergindo de uma capacidade de generalização, e memorizando os dados de treino, podendo ter sua estimativa eficiente somente no conjunto de dados de treino).

Ressaltando que toda saída de todo neurônio j da rede, com n entradas, é calculada através da equação:

$$Y_\mu(j) = F(\mu)$$

Onde $\mu = (\sum_{i=1}^n (X_{i_n}(j) * W_{i_n}(j) + b)$ é o potencial de ativação do neurônio j e F é a função de ativação do neurônio.

Fase Feedforward Nesta fase, em determinada época (k), em um vetor de n variáveis X_i correspondentes à um total de venda Y_i , todos os neurônios j propagam o sinal de saída através do cálculo de $Y_{i_\mu}(j) = F(\mu(j))$. No cálculo de μ os pesos sinápticos de cada neurônio j

conectado à uma entrada $Xi_n(j)$ são $Wi_n(j)$. A saída estimada pela rede será $Yi_\mu(s)$.

Se a iteração i da época k for a primeira, todos os pesos Wi_nj em todos os neurônios, são inicializados com valores aleatórios pequenos.

O valor de $Yi_\mu(j)$, dos neurônios da primeira camada oculta, são obtidos de forma que $Xi_n(j)$ são os sinais de entrada das variáveis de 1 à n .

O valor de $Yi_\mu(j)$, dos neurônios das próximas camadas, inclusive a de saída, utiliza o sinal de saída Yi_μ dos neurônios da camada anterior conectados à j como um sinal de entrada Xi_n .

Fase Feedbackward O Processo de reajuste dos pesos sinápticos com o objetivo de atingir a minimalização do erro quadrático da validação, é feito pelo método do Gradiente Descendente, reajustando os pesos sinápticos $Wi_n(j)$ correspondente à cada neurônio j da rede, para um valor $\Delta Wi_n(j)$, da seguinte forma:

$$\Delta Wi_n(j) = \eta * \delta_j * Xi_n(j)$$

Regra δ para neurônio de saída Exclusivamente para a camada de saída, δ é calculado por

$$\delta_s = (Yi - Yi_\mu(s)) * F'(\mu(s)).$$

O valor dos n pesos $Wi_n(s)$ do neurônio de saída s com n entradas vindas das saídas de neurônios j das camadas anteriores, é obtido por $\Delta Wi_n(s) = \eta * \delta_s * Yi_\mu(j)$

Regra δ para neurônios de camadas ocultas Nesta etapa, os neurônios n da camada oculta obterão seu δ através da seguinte equação usando o δ dos m neurônios da camada posterior que se conectam à ele:

$$\delta_n = F'(\mu(n)) * (\sum \delta_m * Wi_nm)$$

Na rede neural de 1 camada oculta e 1 neurônio de saída, a soma será reduzida à apenas 1 neurônio, ficando então a regra δ da seguinte forma:

$$\delta_n = F'(\mu(n)) * (\delta_s * Wi_ns)$$

Por fim os pesos das n entradas $Wi_n(n)$ conectadas ao neurônio n terão seus pesos reajustados para a próxima iteração da seguinte forma:

$$\Delta W(i+1)_n(n) = \eta * \delta_n * X i_n(n)$$

Após o reajuste chegar na primeira camada oculta, um novo vetor de observações i é apresentado à rede repetindo, iniciando novamente a fase feed-forward e repetindo o ciclo até o fim das entradas.

Validação de treino Após se encerrar uma época, que percorreu todas as i -ésimas entradas do conjunto de treino, um conjunto de dados de validação, previamente separado, é apresentado à rede. A validação ocorre só com a fase feedforward, obtendo-se os erros quadráticos da camada de saída com o dado de validação observado. Então é calculado o $m.s.e$, obtendo-se a média de todos os i -ésimos erros quadráticos do conjunto de validação. O valor $m.s.e$ deve ser acompanhado até que se atinja um ponto ótimo, ou seja, um limite inferior após k épocas. Quando este limite for atingido em uma época k_o , todos os parâmetros da rede dentro da época k_o serão os parâmetros desejados do modelo final da rede.

Quando a curva do $m.s.e$ estiver sendo realizada, não se saberá o ponto ótimo de parada, até que ele comece a ser superado, logo é importante arbitrar um número l para salvar os parâmetros das l épocas anteriores, em busca do ponto ótimo de parada, conforme figura 20

2.2.6 Redes Recorrentes GRU

De acordo com o livro Deep Learning Book desenvolvido pela ([Data Science Academy, 2020](#)), no capítulo de arquitetura de redes neurais gated recurrent unit gru, as redes recorrentes GRU são uma melhoria das redes perceptron recorrentes clássicas encontradas na literatura, pois utilizam elementos que decidem o valores de saída das unidades, e que podem memorizar informações de entrada dadas por um grande intervalo de tempo, sem sofrer dissipação destes valores, com o aproveitamento de recursos de memória computacionais. Portanto serão indicadas na aplicação deste trabalho para a memorização da sazonalidade semestral e anual dos dados do restaurante universitário.

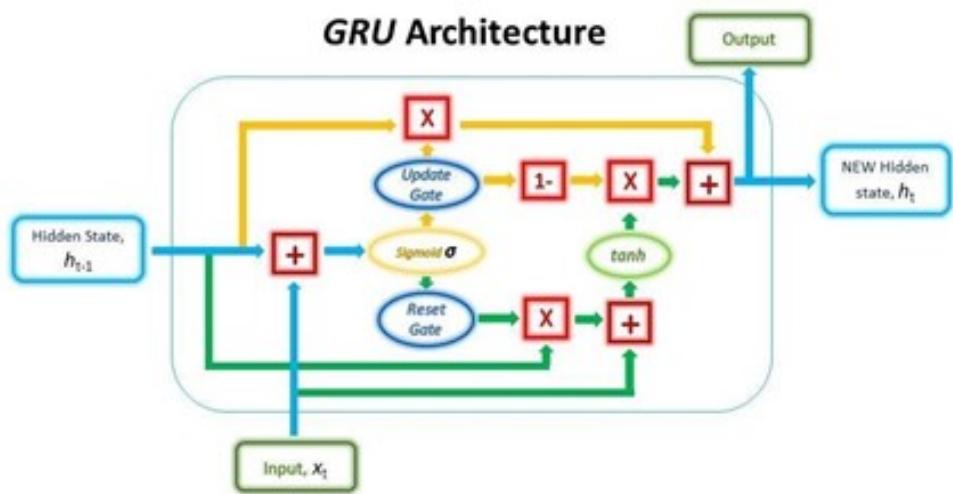


Figura 28: Arquitetura do modelo GRU

Fonte: <http://deeplearningbook.com.br/arquitetura-de-redes-neurais-gated-recurrent-unit-gru/>

3 Trabalhos relacionados

Este capítulo descreve os principais trabalhos relacionados encontrados na literatura referenciada neste trabalho. A seção de Revisão literária para previsão de demanda, cita a primeira referência literária encontrada na execução deste trabalho. A seção de previsão de demanda em outros ambientes cita o maior volume de referencias encontradas durante as pesquisas de predição de demanda, e a seção de Previsão de demanda em restaurantes universitários cita especificamente os estudos relacionados com uma parte dos métodos executados neste trabalho que são os modelos de redes neurais MLP, já o diferencial deste trabalho em relação à todos os trabalhos referenciados é a inclusão de redes neurais recorrentes modernas, denominadas GRU, fundamentadas no capítulo anterior, na seção de Inteligência Artificial, subseção Redes Recorrentes GRU.

3.1 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 da 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.

([FERNANDO., 2011](#)) também realiza o estudo de demanda no restaurante universitário da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), novamente com os méto-

dos 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), e 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.

3.2 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 votores 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.

3.3 Revisão literária para previsão de demanda

(JUNIOR, 2007) realiza um trabalho de comparação entre diversos métodos estocásticos e modelos de aprendizado de máquina, 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.

4 Metodologia

Em resumo a metodologia experimental deste trabalho 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 5 dias anteriores.
- Analises 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 2 fases experimentais com diferentes domínios temporais.
- Analises comparativas dos resultados dos modelos.
- Conclusões sobre os resultados e seleção do melhor modelo.

4.1 Pré-processamento

Na etapa de pre-processamento dos dados, além da obtenção dos dados, também é realizada a etapa de transformação dos dados em séries temporais, normalização com a remoção de outliers, e aplicação da 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.

4.1.1 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 todos as unidades da Unifesp.

Apenas alguns funcionários autorizados tem acesso ao banco de dados do sistema de refeições da instituição, entre eles o fiscal de contrato do restaurante universitário. Para obter tais dados neste trabalho, foi necessário obter uma autorização com a direção do campus ICT - UNIFESP e em seguida solicitar a exportação dos dados ao fiscal. Para este trabalho foram solicitados os dados de consumo exclusivamente de alunos, pois podem ser obtidos também os dados de professores, alunos de pós-graduação e visitantes.

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

Tabela 1: Formato dos dados originais obtidos pelo restaurante universitário

Os dados exportados pelo sistema do restaurante da Unifesp foram obtidos de acordo com o formato da tabela 1

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

Após a coleta, os dados de consumo do restaurante foram transformados em um processo de aproximação por uma série temporal, para um intervalo de 5 dias, e em cada registro de venda são acrescentados 5 novos atributos contendo os valores passados, deste mesmo atributo, em um intervalo de 5 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 desenvolvidos. A tabela 2 representa a nova estrutura de um registro de dado do restaurante, com um intervalo temporal de 5 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 1 dia anterior.

4.1.2 Dados exógenos

Os dados exógenos correlacionados com o consumo se dividem em 2 tipos principais, os dados climáticos coletados de estações meteorológicas próximas ao ICT Unifesp, e dados derivados das datas dos registros de consumo.

Dados Climáticos Também obteve-se variáveis climáticas como dados exógenos, para a influência de fatores externos (temperatura média ambiente, pressão atmosférica, umidade e velocidade do vento). Tais dados 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, pertencente ao MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO do Governo Brasileiro.

É necessário um cadastro no site <http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep> 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.

Dados de Calendário 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 dia da semana seguiu o modelo binário de acordo com os trabalho de previsão de demanda em R.U realizados por ([LOPES, 2008](#)) e ([FERNANDO., 2011](#)).

Dia da semana	Valor				
Segunda-feira	1	0	0	0	0
Terça-feira	0	1	0	0	0
Quarta-feira	0	0	1	0	0
Quinta-feira	0	0	0	1	0
Sexta-feira	0	0	0	0	1

Figura 30: Entradas de dia da semana em cofatores.

Retirado de ([LOPES, 2008](#)).

4.1.3 Formato final do conjunto de dados

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

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.

4.1.4 Tratamento dos dados para entrada nos modelos

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

- Calculo 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 são 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.

4.1.5 Fases Experimentais

O processo experimental é realizado em 2 roteiros distintos de divisão do domínio temporal do conjunto de dados, e os resultados obtidos entre as duas fases serão comparados.

O conjunto de dados contemplando o período de 2017 a 2019, é 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, são usados para treino. Os resultados obtidos nesta divisão são 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 são 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 são comparadas com o melhor modelo da primeira fase.

4.2 Definição e treino dos modelos

4.2.1 Topologia

Sobre a necessidade de se implementar modelos mistos No conjunto de dados deste trabalho, os dados obtidos se dividem em dados temporais (onde cada registro de consumo e venda trás a informação de seu domínio em um intervalo de 5 dias anteriores) e dados discretos onde temos variáveis categóricas de data para cada registro, e variáveis climáticas, porém sem a aplicação de um intervalo temporal, ou seja, os dados exógenos são discretos, e os endógenos são temporais. Portanto é 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 perceptron na saída, recebendo os 2 valores dos modelos endógenos e exógenos para a regressão das saídas das 2 redes ao valor que será a predição do consumo.

4.2.1.1 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 28 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 5 dias).

4.2.1.2 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 2 redes neurais em um perceptron criando um comitê de redes neurais para obter a saída final prevista.

4.2.2 Função de ativação

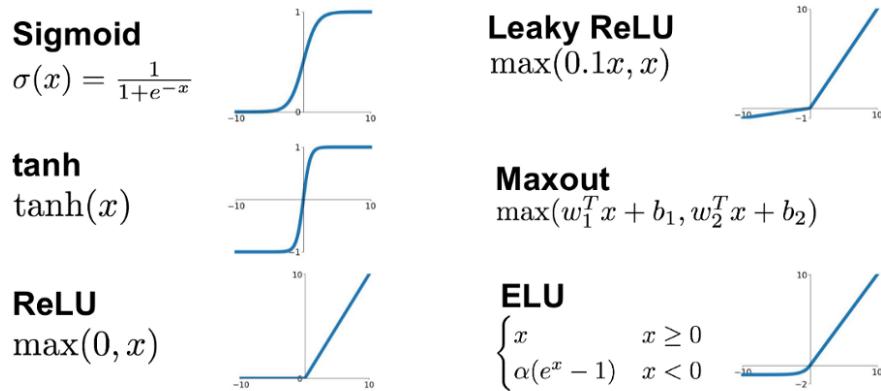


Figura 32: Funções de Ativação

Fonte: ([MC AI, 2020](https://mc.ai/complete-guide-of-activation-functions/)), <https://mc.ai/complete-guide-of-activation-functions/>

4.2.2.1 Para a entrada e camadas ocultas

A função ReLu é simples e eficiente para a aplicação nos experimentos, visto que na fase feed-forward tem efeito parecido com a função identidade, e na fase feed-backward durante o reajuste dos pesos pelo otimizador, produz efeito degrau zerando valores negativos e sendo adequada na aplicação do domínio de entrada deste trabalho, visto que todas as variáveis endógenas e exógenas não possuem valores negativos.

4.2.2.2 Para a saída

Para os valores previstos a função escolhida é a linear, pois na fase feed-backward do reajuste dos pesos pelo algoritmo de treino backpropagation, a derivada da função linear se torna zero, mantendo a execução do algoritmo de treino transformando apenas os valores de entrada e valores das camadas ocultas das redes, sem interferir no valor de saída.

4.2.3 Otimizador

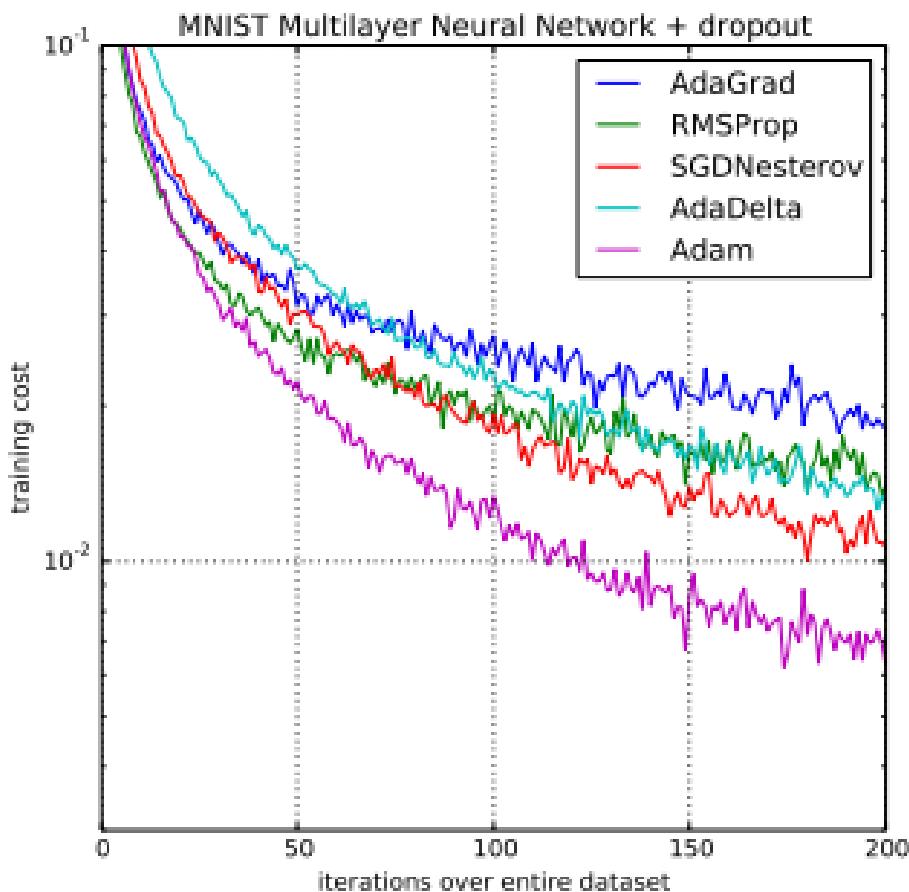


Figura 34: Comparison of Adam to Other Optimization Algorithms Training a Multilayer Perceptron

Fonte : ([Jason Brownlee, 2020](#))

<https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>

Para este trabalho foi escolhido o otimizador ADAM no treino de backpropagation. A vantagem deste otimizador é a fusão das melhores características de 2 otimizadores : Momenum e RMSProp.

O Momentum acelera o reajuste dos pesos em busca dos erros globais mínimos.

O RMSProp impede a busca na direção das oscilações.

Adam ou Adaptive Moment otimization combina estas 2 heurísticas. O coeficiente de aprendizado escolhido para o otimizador, ou conhecido como "learning rate", foi definido para a constante alpha, com valor 0.001. Esta constante tem produzido resultados positivos em problemas de forecast (previsões) de acordo com o artigo fundamentado em ([Jason Brownlee, 2020](#)).

4.3 Teste e Métricas

A principal métrica de avaliação dos modelos, é a Raiz do Erro Quadrático Médio, obtido pela chamada de função "mean_squared_error" do framework Tensorflow.Keras utilizado para a modelagem,treino e predição dos modelos de redes neurais.

O coeficiente de correlação de Pearson, obtido pela biblioteca scipy na chamada de função `scipy.stats.pearsonr(true,pred)`" e o coeficiente "chi-quadrado" definido como R^2 obtido na chamada de função `scipy.stats.linregress`" foi utilizado nas etapas de teste para avaliar a proximidade das previsões do modelo com o comportamento real de consumo (Se acompanha as quedas e subidas de consumo ao longo do tempo).

Documentou-se, para trabalhos futuros, outras métricas estatísticas obtidas por esta chamada de função como os valores de slope, intercept, p_value e std_err.

A chamada de função `sns.regplot(x=arr_true,y=arr_pred,data=df)` da biblioteca seaborn retorna um gráfico scatter dos valores preditos e reais para a avaliação dos erros, médias e tendência de previsão.

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 previsões do modelo.

5 Resultados

Este capítulo reúne os resultados das 2 fases experimentais deste trabalho e seleciona por fim o melhor resultado entre as fases.

A primeira fase experimental na etapa de pré-processamento realiza antes do desenvolvimento dos modelos, o procedimento de análise exploratória dos dados fundamentado no trabalho de ([JUNIOR, 2007](#)).

O parâmetro “ENTR_ALMOCO” da tabela [3](#) é o atributo alvo da predição, pois se trata do número de entradas no ponto de acesso do restaurante, portanto esse parâmetro é frequentemente utilizado nos gráficos comparativos em formato de linha, sendo geralmente comparado com algum outro parâmetro de entrada nos modelos, denominado de “Feature” neste trabalho, nomenclatura utilizada por reconhecidas referências na área de Data Science, conforme o artigo de ([Rebecca Vickery](#),).

Nesta primeira fase também são desenvolvidos 2 modelos de redes MLP para suporte na análise exploratória do problema da predição do restaurante e é verificado se esses modelos tem capacidade de aprendizado sobre o problema.

Depois de verificada a capacidade de aprendizado do problema foram desenvolvidos 3 modelos endógenos, sendo uma rede MLP e duas redes GRU. Os resultados de predição foram comparados visando a seleção do melhor modelo endógeno.

Após a seleção do melhor modelo, foram desenvolvidos três modelos mistos. Ao contrário dos anteriores, estes modelos recebem como entrada tanto features endógenas quanto exógenas. Da mesma forma, o melhor modelo com ambas as features também é selecionado.

Na segunda fase, é alterado o domínio temporal experimental para a divisão dos conjuntos e repetidos os passos da primeira fase a partir do desenvolvimento dos modelos endógenos.

Após selecionado o melhor modelo da segunda fase, foram comparados os melhores modelos das duas fases. Para uma comparação justa, visto que o melhor modelo da 1a fase tem domínio de teste restrito ao primeiro semestre de 2019, o modelo obtido na segunda fase é novamente testado apenas no domínio do primeiro semestre obtendo-se novas métricas de teste.

Ao final das duas fases experimentais, o modelo com melhor predição dos dados é apresentado como um possível preditor para a demanda do RU.

5.1 Fase Experimental I

O domínio experimental desta fase segue de acordo com a figura [36](#). Nesta fase temos 1 modelo selecionado na comparação de resultados entre 3 modelos endógenos e 3 modelos

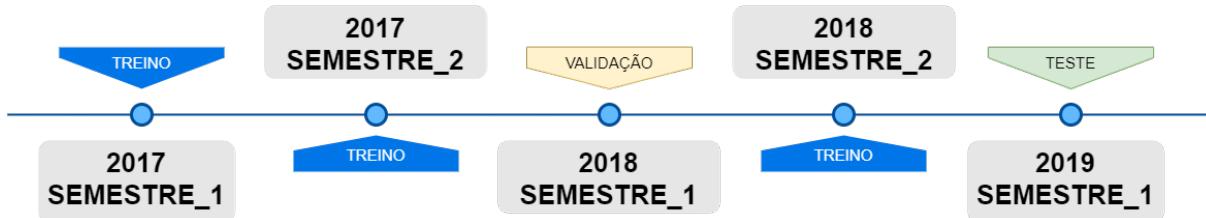


Figura 36: Domínio temporal da 1a fase

exógenos desenvolvidos, treinados e testados. A hipótese sobre este modelo segue sobre o seu conjunto de validação que contempla apenas o primeiro semestre de 2018, portanto é o indicado para semestres com sazonalidade, grade horária de disciplinas e comportamento de consumo semelhante ao 1º semestre de 2018.

5.1.1 Pré-Processamento

Nesta subseção é realizada, no conjunto de treino, a análise exploratória das features e do comportamento de consumo, e também executada a técnica atual de previsão do restaurante no conjunto de teste, composta pela produção de refeições, para cada dia, com margem de 30% acima do consumo do mesmo dia da semana anterior.

5.1.1.1 Divisão dos conjuntos

Dificuldades encontradas e resolvidas A primeira dificuldade encontrada foi um comportamento anômalo do gráfico de previsão e valores reais. O conjunto de dados foi lido trocando datas (dia por mês e mês por dia). A indexação por data na divisão de conjuntos produzia gráficos fora do formato esperado, de acordo com a figura 37.

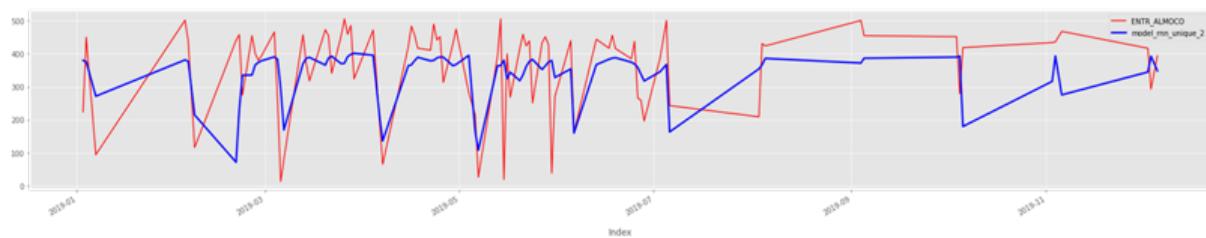


Figura 37: Resultado do modelo RNN_ENDO_2 obtido sobre o conjunto de dados aleatoriamente ordenado sobre o tempo

Corrigindo a importação dos registros para o padrão brasileiro na ferramenta pandas os resultados foram melhores e condizentes com o comportamento real de consumo, além disso nota-se que a primeira fase filtra dados obtidos apenas para o 1º semestre de 2019 e no caso da indexação incorreta na figura 37 a ordenação ultrapassou o mês de julho, este foi o comportamento que indicou a investigação no código de importação.

A figura 38 demonstra os resultados obtidos após a correção da leitura dos dados, onde todos os modelos obtiveram resultados melhores. Isto também demonstra que ordenações alea-



Figura 38: Resultado do modelo RNN_ENDO_2 obtido sobre o conjunto de dados com ordenação corrigida

tórias na divisão dos conjuntos de treino, teste e validação, não seriam indicadas para a execução deste trabalho. Nas figuras 40a , 42a e 44a são exibidos os valores reais de consumo nos conjuntos de treino (2017-1,2017-2,2018-2), validação (2018-1) e teste (2019-1).

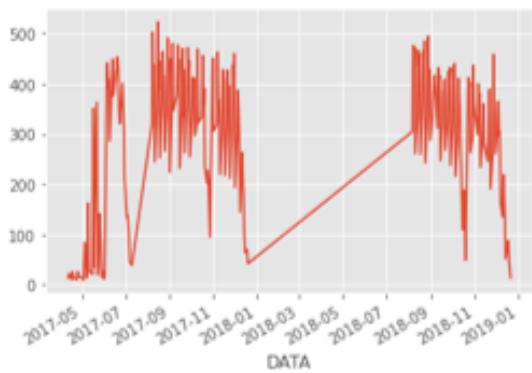


Figura 39: 1a fase, Conjunto verdade de treino

Dados do 1o e 2o semestre de 2017
e 2o semestre de 2018.

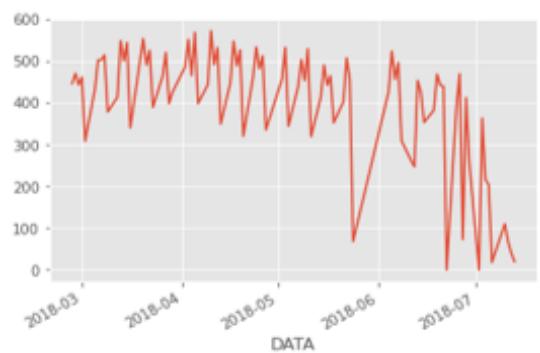


Figura 41: 1a fase, Conjunto verdade de validação

Dados do 1o semestre de 2018.

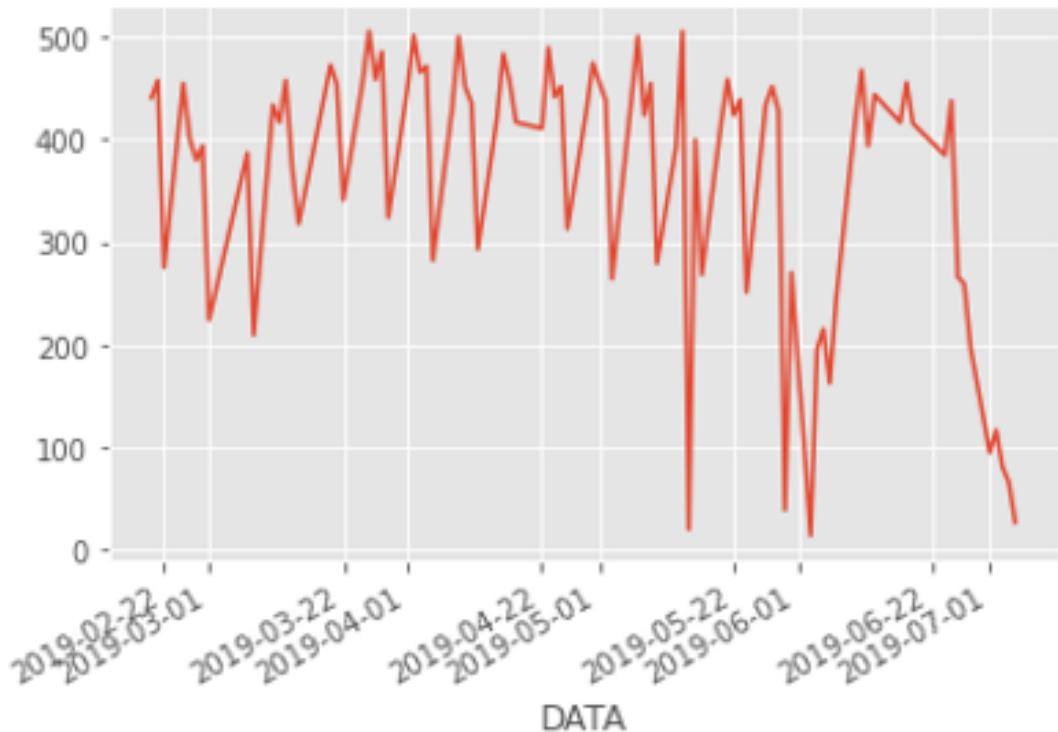


Figura 43: 1a fase, Conjunto verdade de teste

Dados do 1o semestre de 2019.

5.1.1.2 Análise das variáveis endógenas

Consumo em relação às vendas de 1 dia anterior

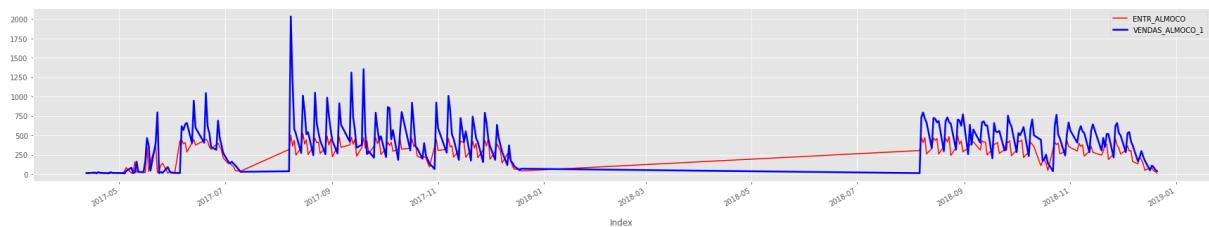


Figura 45: Correlação entre consumo e vendas de almoço.

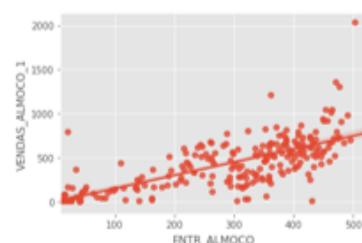


Figura 46: Gráfico Scatter entre consumo e vendas de almoço.

Tabela 4: 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
TOTAL DE REFEIÇÕES PROJETADAS	104694
TOTAL DE REFEIÇÕES CONSUMIDAS	69544
TOTAL DE REFEIÇÕES SUB PROJETADAS	-4703
TOTAL DE REFEIÇÕES SUPER PROJETADAS	39853
ERRO ABSOLUTO MEDIANO	139.0
ERRO ABSOLUTO PERCENTUAL MÉDIO	90.18

É possível notar na figura 45 que as vendas de ticket 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, que os alunos comprassem apenas 2 tickets por dia. Possivelmente esta limitação foi dada para aproximar o comportamento de consumo de 1 até 2 dias seguintes à vendas de tickets no dia vigente, esta limitação pode ser interpretada como método de auxílio à gestão para a produção de refeições e para o tratamento de desperdício.

Mesmo com o outlier de 2000 vendas em um dia, e com a nova limitação de compras de tickets a partir de 2018, o consumo está fortemente relacionado com as vendas de tickets de 1 dia anterior e que os alunos se adaptaram à utilização em curto prazo dos tickets, conforme o valor de correlação aproximado em 72% que pode ser conferido na tabela 4.

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

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, da seção anterior, pois entre todas as features esta é a que produziu outliers com 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 e mesmo com a normalização, o comportamento linear desta feature, conforme figura 47, se manteve. E após a normalização foi realizada a aplicação da escala de 0 a 1 na feature e conforme é observado na figura 48, o comportamento linear da feature também se manteve.

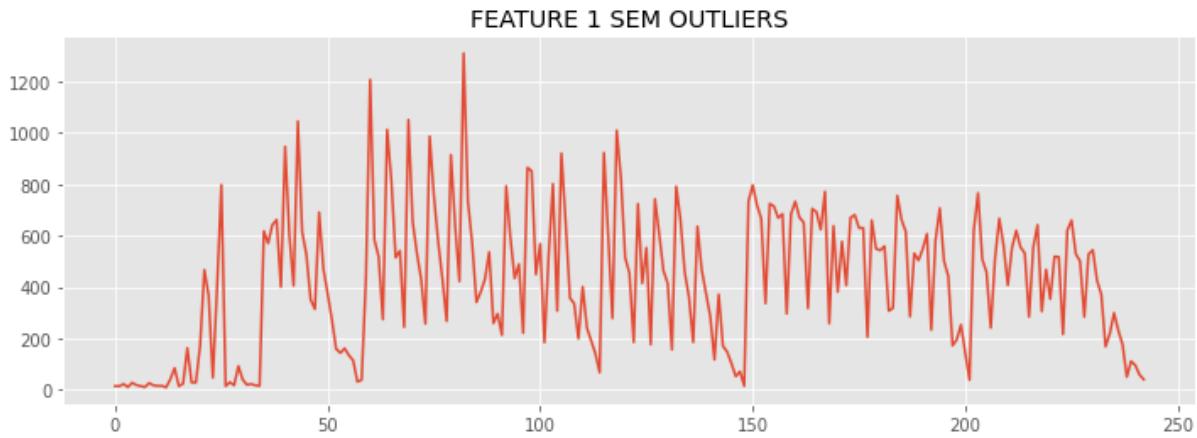


Figura 47: Vendas de tickets normalizados com teto de 3x o desvio padrão.

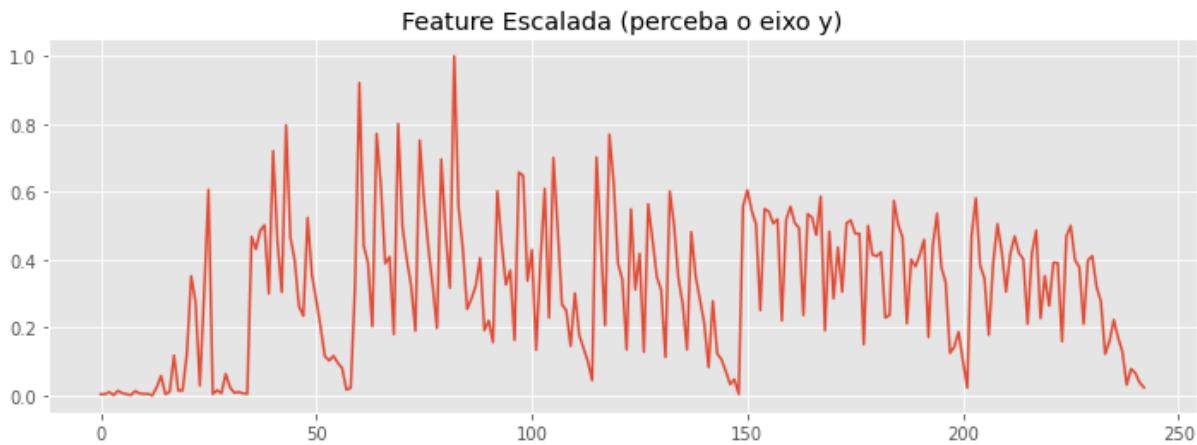


Figura 48: Vendas de tickets escalada entre 0 a 1.

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

Análise da técnica do restaurante

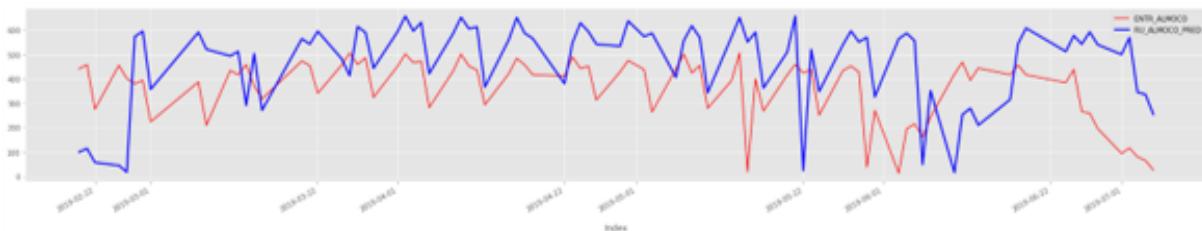


Figura 49: Correlação entre consumo e produção com 30% acima do consumo da semana anterior.

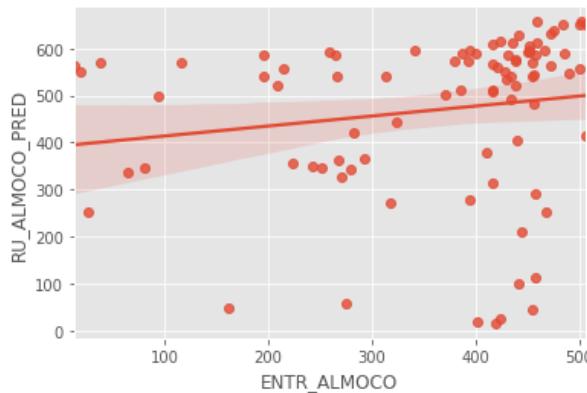


Figura 50: Gráfico Scatter entre consumo e produção com margem de 30%. acima do consumo da semana anterior

Consumo vigente em relação ao 5º dia anterior	
CORRELAÇÃO (r):	0.1593
Pi (p) :	0.1380
RMSE =	224.4733
TOTAL DE REFEIÇÕES PROJETADAS =	41351
TOTAL DE REFEIÇÕES CONSUMIDAS =	31962
TOTAL DE REFEIÇÕES SUB PROJETADAS =	-3650
TOTAL DE REFEIÇÕES SUPER PROJETADAS =	13039
ERRO ABSOLUTO MEDIANO =	152.0
ERRO ABSOLUTO PERCENTUAL MÉDIO =	167.2659%

Tabela 5: Estimações do restaurante para o 1º Semestre de 2019

A análise da técnica de estimativa de consumo, do restaurante, foi feita com o cálculo de 30% de produção acima do consumo do 5º dia anterior. É possível notar que o modelo do R.U é feito para tolerar descartes devido à multa contratual para falta de refeições e o que a produção de 30% de refeições acima do consumo da semana anterior, conforme a figura 49 produz um comportamento linear distante do comportamento real de consumo, apesar de seguir as tendências de quedas e aumento de consumo, na figura 50 do gráfico scatter gerada pela biblioteca seaborn também demonstra que a regressão linear (representada pela linha vermelha no gráfico) tem o eixo totalmente descentralizado com a função identidade entre o consumo e a predição ideal (representada pela diagonal formada entre a origem do gráfico e o vértice superior direito), gerando também um erro muito maior do que 30% no somatório total de refeições descartadas no semestre, devido ao comportamento oscilatório do consumo, conforme a tabela 5.

Consumo atual em relação ao consumo do jantar de 1 dia anterior. Apesar de que os alunos que consomem refeições no almoço geralmente são de período e de grade horária diferente dos alunos que consomem o jantar no período da noite, nota-se uma relação evidente entre os 2 consumos, evidenciada pelo comportamento linear na figura 51 com correlação ($r = 0.7655$), e

pela regressão linear entre esses 2 consumos na figura 52.

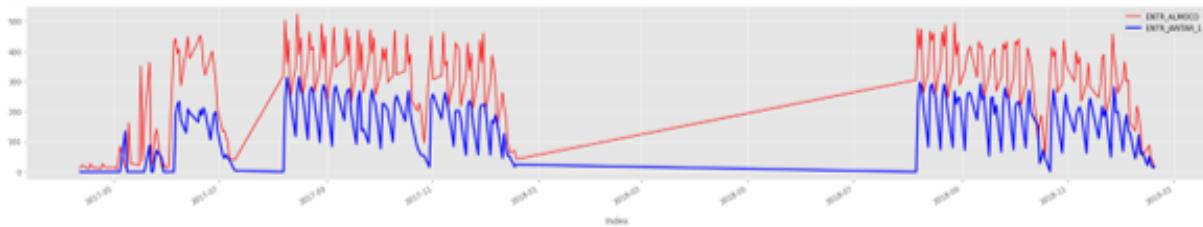


Figura 51: Correlação de consumo de almoço e jantar de 1 dia anterior.

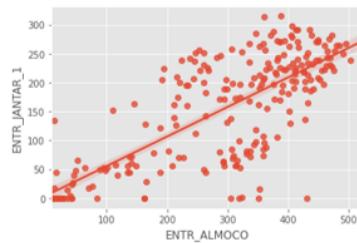


Figura 52: Gráfico Scatter entre consumo e jantar de 1 dia anterior.

5.1.1.3 Análise da sazonalidade semanal

Os gráficos a seguir da figura 53, representando a segunda-feira, até a figura 57 , 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 dataset. O violino azul com o valor 1 representa a distribuição do consumo ao logo do dataset. O violino com valor zero pode ser ignorado e é um retorno padrão da ferramenta, representando a distribuição da ausência de consumo no dia da semana considerado. Nas sextas feiras, o consumo teve escala de distribuição menor para o primeiro semestre de 2019.

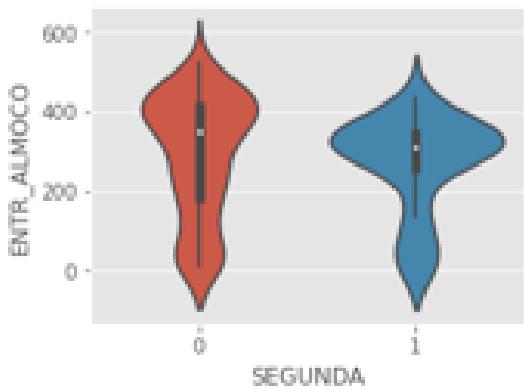


Figura 53: Gráfico violino da distribuição do consumo na segunda feira.

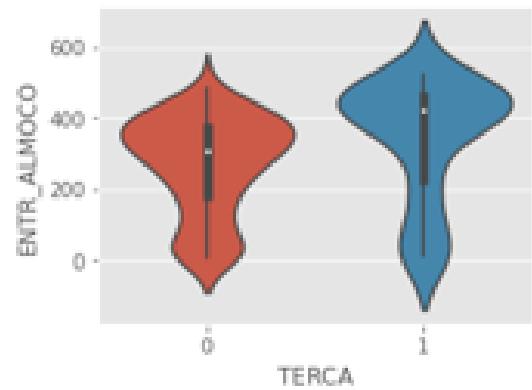


Figura 54: Gráfico violino da distribuição do consumo na terça feira.

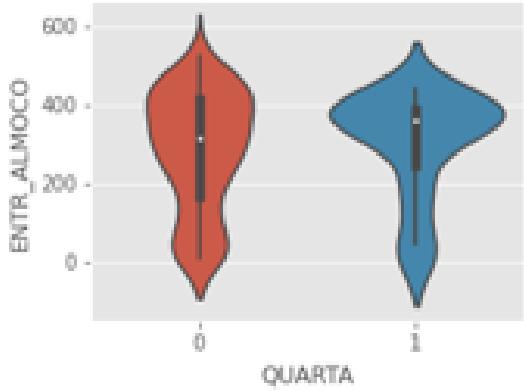


Figura 55: Gráfico violino da distribuição do consumo na quarta feira.

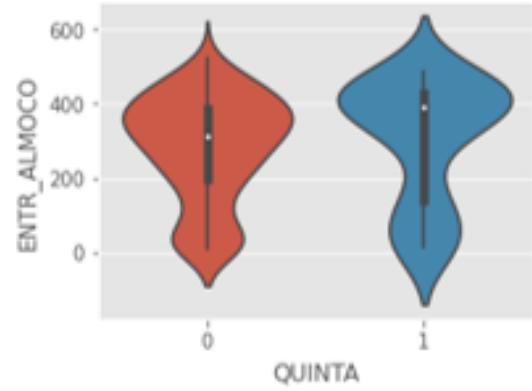


Figura 56: Gráfico violino da distribuição do consumo na quinta feira.

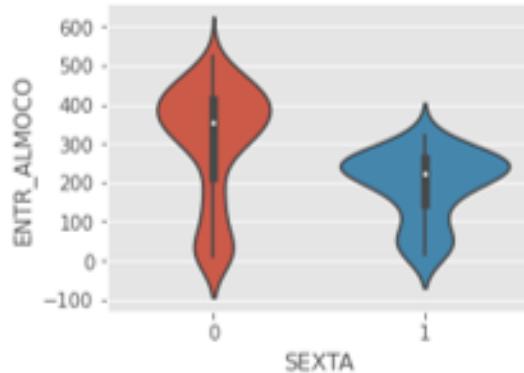


Figura 57: Gráfico violino da distribuição do consumo na sexta feira.

5.1.1.4 Análise das variáveis exógenas

Consumo atual em relação ao avanço do semestre

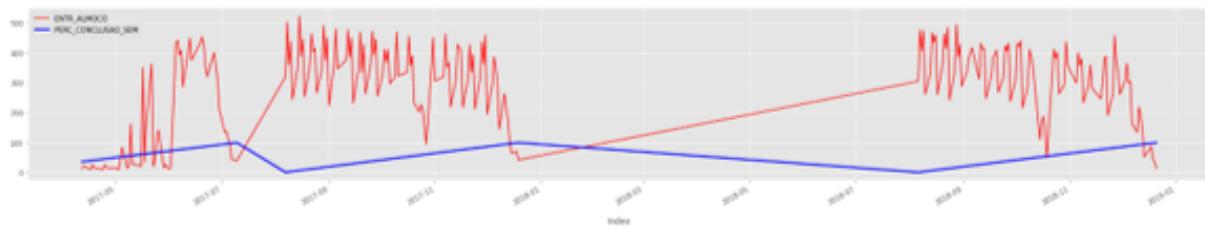


Figura 58: Relação da distribuição do consumo com o avanço do semestre.

Correlação (r) = -0.3513, p-value = 1.8029942608003656e-08

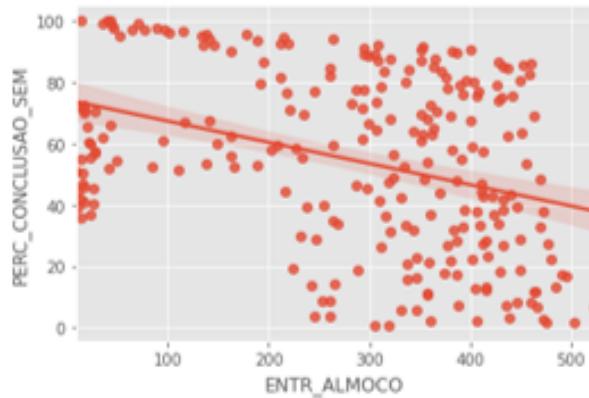


Figura 60: Gráfico scatter da distribuição do consumo com o avanço do semestre.

Na correlação do consumo em relação ao avanço do semestre, para os últimos dias do semestre o consumo teve queda abrupta, a correlação dos conjuntos de dados das figuras 59a e 60 obtém valor negativo mas devido às oscilações de consumo, a correlação não se torna significativa.

Consumo atual em relação ao avanço do mês

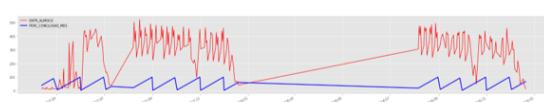


Figura 61: Relação da distribuição do consumo com o avanço do mês.

Correlação (r) = 0.04867

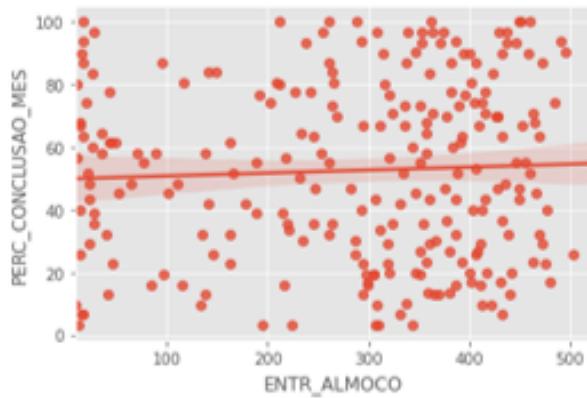


Figura 63: Gráfico scatter da distribuição do consumo com o avanço do mes.

Já o consumo em relação ao avanço do mês é inconclusivo. É possível notar o gráfico em forma de serra do avanço do mês em relação ao semestre na figura 62a, onde este comportamento oscilatório e característico das alternâncias do mês em relação ao ano pode ter gerado ruído para esta análise. O indicado para análises futuras é que o domínio de obtenção destas métricas seja reduzido à duração do mês. Mesmo sendo reduzido é possível observar que não há um padrão gráfico de consumo para todos os fragmentos de serra. Esta hipótese validaria que os alunos tivessem preferência de realizar refeições fora do restaurante em períodos típicos de recebimento de salário. A figura 63 conclui também que a regressão linear entre os 2 conjuntos é insignificante. Como o restaurante do ICT Unifesp não enfrenta concorrência dado o isolamento geográfico de seu público, a sazonalidade do consumo se mantém correlata com as atividades da universidade e não da concorrência com outros restaurantes.

Variáveis climáticas As correlações do consumo de refeições com as variáveis climáticas não obtiveram correlações evidentes com o consumo de refeições, sendo observadas nas figuras 64 à 71, todos os gráficos lineares não apresentaram padrões correlatos e evidentes, bem como os gráficos de scatter não apresentaram regressões lineares significantes.

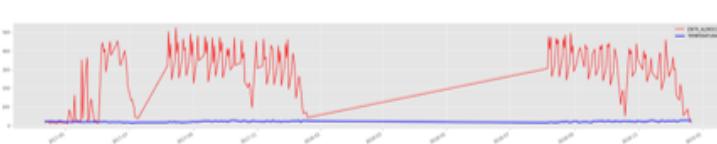


Figura 64: Correlação da distribuição do consumo com a temperatura.

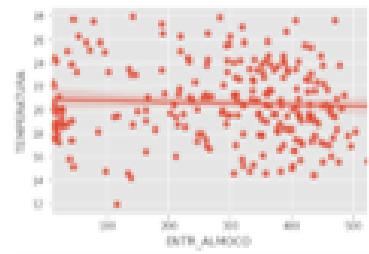


Figura 65: Gráfico scatter da distribuição do consumo com a temperatura.

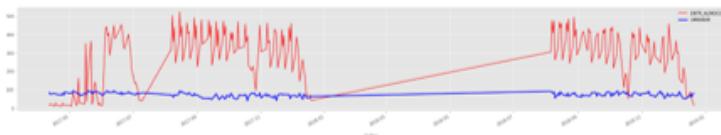


Figura 66: Correlação da distribuição do consumo com a umidade.

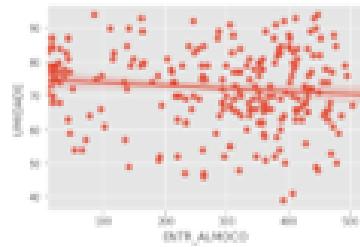


Figura 67: Gráfico scatter da distribuição do consumo com a umidade.

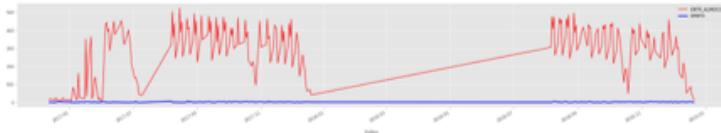


Figura 68: Correlação da distribuição do consumo com a velocidade do vento em m/s.

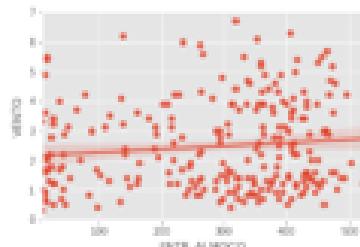


Figura 69: Gráfico scatter da distribuição do consumo com a velocidade do vento.

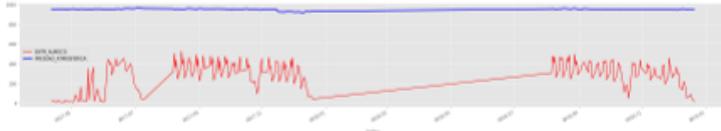


Figura 70: Correlação da distribuição do consumo com a pressão atmosférica.

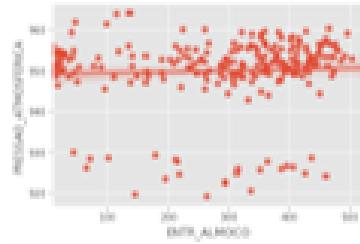


Figura 71: Gráfico scatter da distribuição do consumo com a pressão atmosférica.

5.1.2 Definição e treino dos modelos

Nesta seção os modelos de redes neurais foram desenvolvidos, treinados, avaliados sobre o conjunto de validação para obtermos métricas preliminares, e por fim foram testados. Na próxima subseção, 1 modelo inicial MLP é desenvolvido para avaliação da aprendizagem do problema. Como a avaliação de aprendizagem é um método comparativo com a profundidade em camadas do modelo, o mesmo modelo é redefinido com mais camadas de neurônios e novamente testado. Após os testes bem sucedidos sobre a avaliação de aprendizagem do problema nos modelos MLP, outros modelos com maior profundidade do tipo endógenos e exógenos foram desenvolvidos.

5.1.2.1 Ajuste empírico de topologia com os modelos perceptron

MLP1 O primeiro modelo perceptron multilayer foi definido com uma camada de 15 unidades (mesmo número de features endógenas sendo 5 para consumo almoço , 5 para vendas de tickets e 5 para consumo de jantar, com intervalo de 5 dias). O modelo será denominado MLP1, para futuras referências. A figura abaixo foi obtida com o comando `keras.utils.plot_model(MLP1, show_shapes=True)`. Cada camada da rede corresponde à um bloco na figura, nos blocos são demonstrados o formato matricial dos parâmetros de entrada. Nota-se que no primeiro bloco "InputLayer" são denotados o tamanho do intervalo temporal de cada parâmetro correspondente à 5 dias (de 1 dia anterior à 5 dias anteriores), e o número total de parâmetros que são 3 (ENTR_ALMOCO, ENTR_JANTAR e VENDAS_ALMOCO).

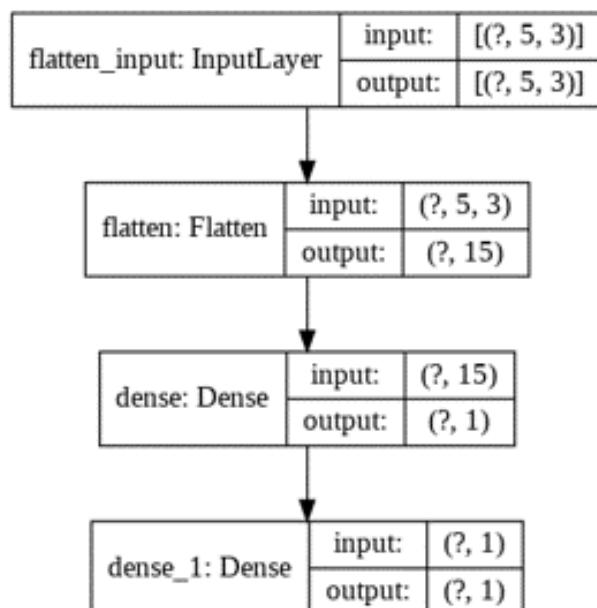


Figura 72: Topologia do modelo MLP1

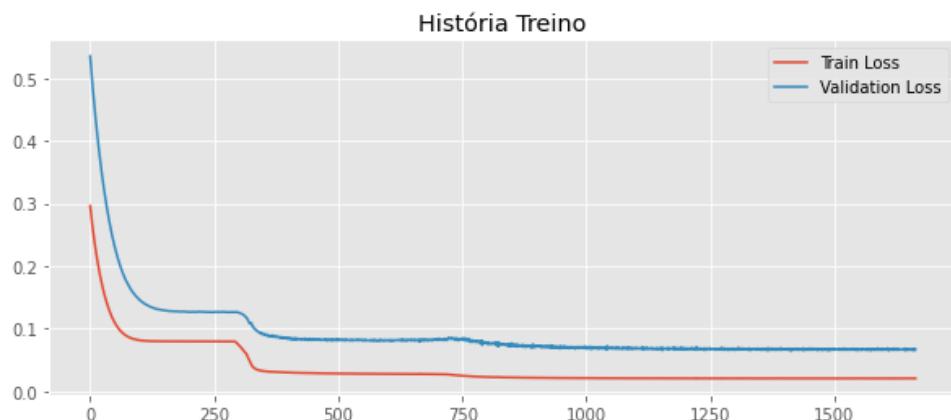


Figura 73: Gráfico de treino do modelo MLP1.

Loss values, RMSE = 130.6207420543897

MLP2 Foi definido um novo modelo MLP2, aumentando o número de neurônios do MLP1 para a comparação do RMSE de treino.

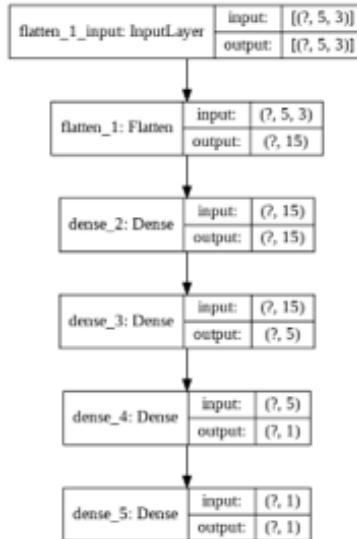


Figura 75: Topologia do modelo MLP2

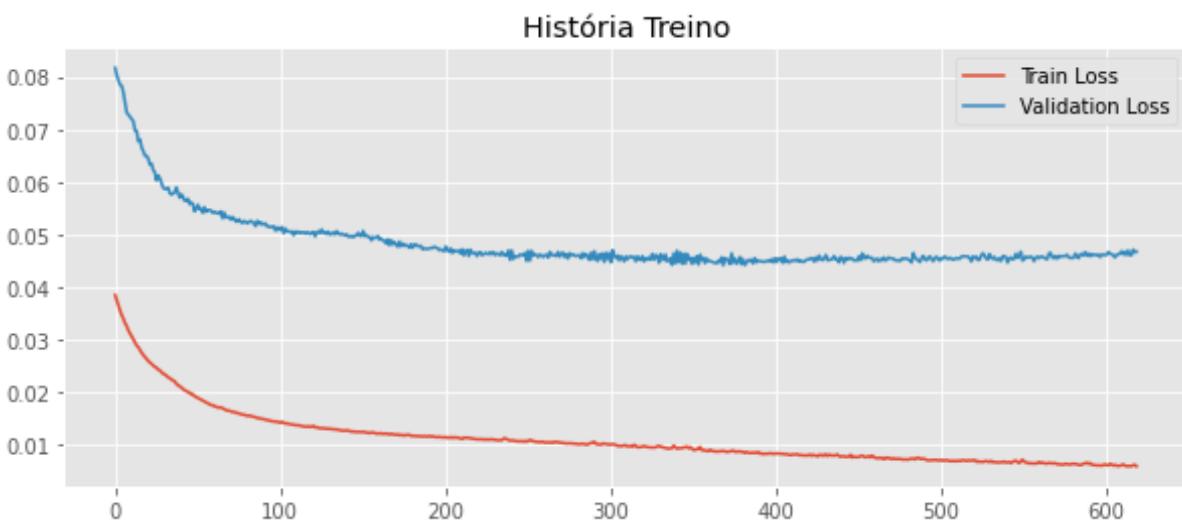


Figura 76: Gráfico de treino do modelo MLP2.

$$\text{RMSE} = 107.97413966672336$$

É possível notar a diminuição do RMSE (Raiz do erro quadrático médio) ao aumentar a profundidade da rede perceptron para treino e avaliação sob o conjunto de validação. Validando a hipótese de que a predição do consumo no restaurante, pode ser aprendida por modelos simples de redes neurais, sendo possível avançar nas pesquisas com modelos que tragam melhores resultados.

5.1.2.2 Modelos Endógenos

Como o ajuste empírico obteve avanço nos resultados, foi definido outro modelo perceptron MLP_ENDO_1 com maior profundidade, e depois um modelo recorrente GRU com 2 reajustes (RNN_ENDO_1 e RNN_ENDO_2). Para a avaliação prévia dos modelos, os mesmos foram testados no conjunto de validação.

MLP_ENDO_1 Definido primeiro modelo endógeno exclusivamente perceptron, MLP_ENDO_1, com o aumento da profundidade de camadas de neurônios, baseado no modelo anterior MLP2.

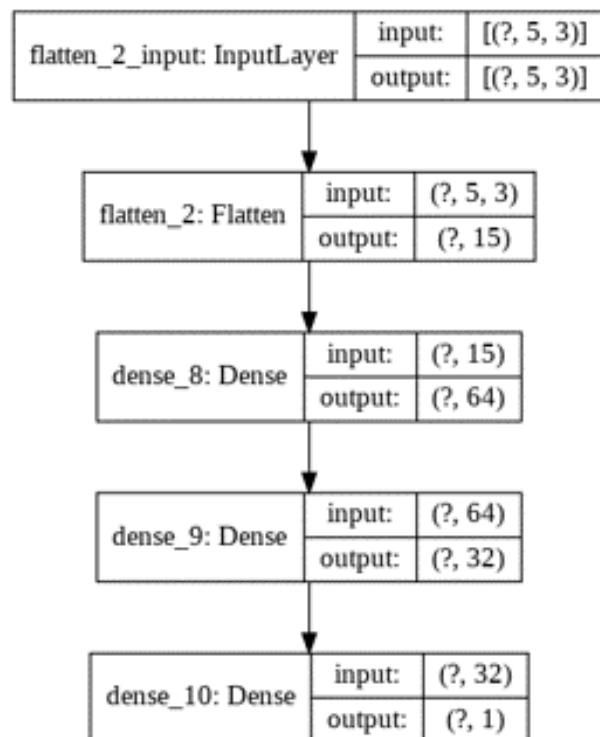


Figura 78: Topologia do modelo MLP_ENDO_1

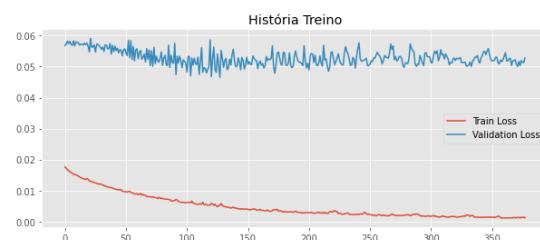


Figura 79: Treino do modelo
MLP_ENDO_1

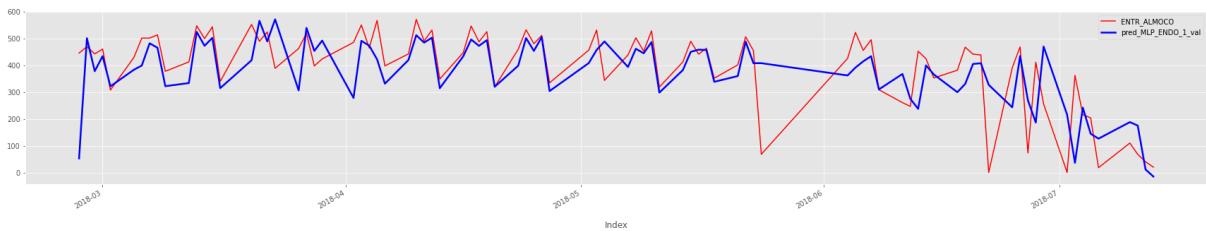


Figura 80: Avaliação do modelo MLP_ENDO_1

METRICAS DO MODELO MLP_ENDO_1 :

RMSE	110.92902359567118
TOTAL DE REFEIÇÕES PROJETADAS	89 : 33611.37430477142
TOTAL DE REFEIÇÕES CONSUMIDAS	89 : 35555
TOTAL DE REFEIÇÕES SUB PROJETADAS	-4328.862397193909
TOTAL DE REFEIÇÕES SUPER PROJETADAS	2385.236701965332

Tabela 6: Métricas de avaliação do modelo MLP_ENDO_1.

RNN_ENDO_1 Definido primeiro modelo com a utilização das redes recorrentes GRU, RNN_ENDO_1, para comparação de resultados com o modelo anterior perceptron.

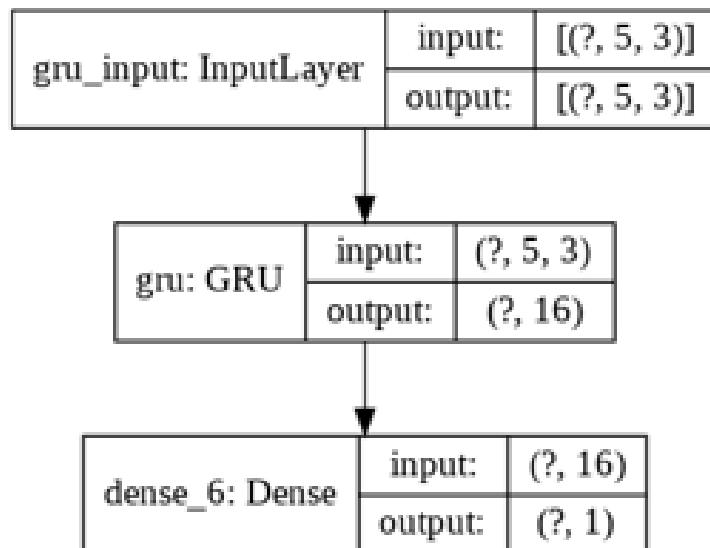


Figura 81: Topologia do modelo RNN_ENDO_1

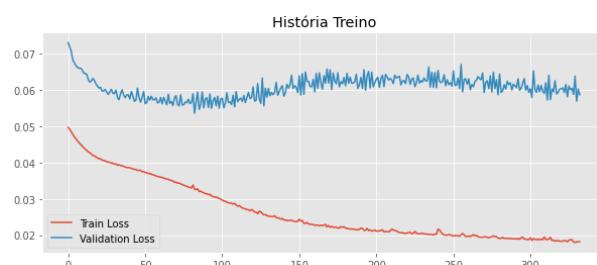


Figura 82: Treino do modelo RNN_ENDO_1

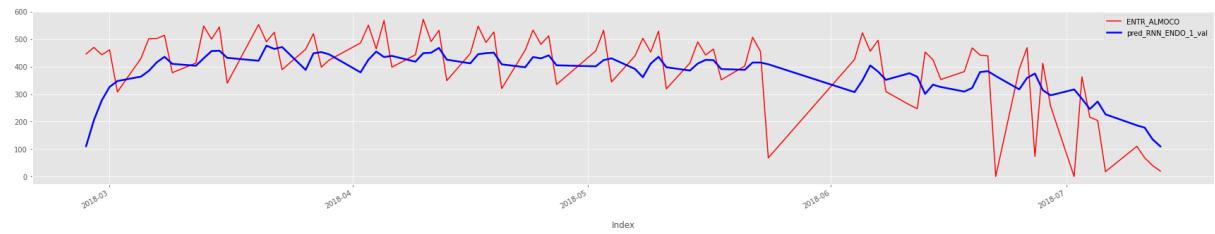


Figura 83: Avaliação do modelo RNN_ENDO_1

Tabela 7: Métricas de avaliação do modelo RNN_ENDO_1.

METRICAS DO MODELO RNN_ENDO_1 :	
RMSE	118.98273426840373
TOTAL DE REFEIÇÕES PROJETADAS	33423.96739196777
TOTAL DE REFEIÇÕES CONSUMIDAS	35555
TOTAL DE REFEIÇÕES SUB PROJETADAS	-5248.0078125
TOTAL DE REFEIÇÕES SUPER PROJETADAS	3116.9752044677734

RNN_ENDO_2 Definido segundo modelo GRU, RNN_ENDO_2, com o aumento da profundidade de unidades do modelo anterior RNN_ENDO_1 e com a inclusão do recurso Dropout para eliminar unidades com pesos negativos no grafo denso da rede, realizando a poda de unidades não relevantes.

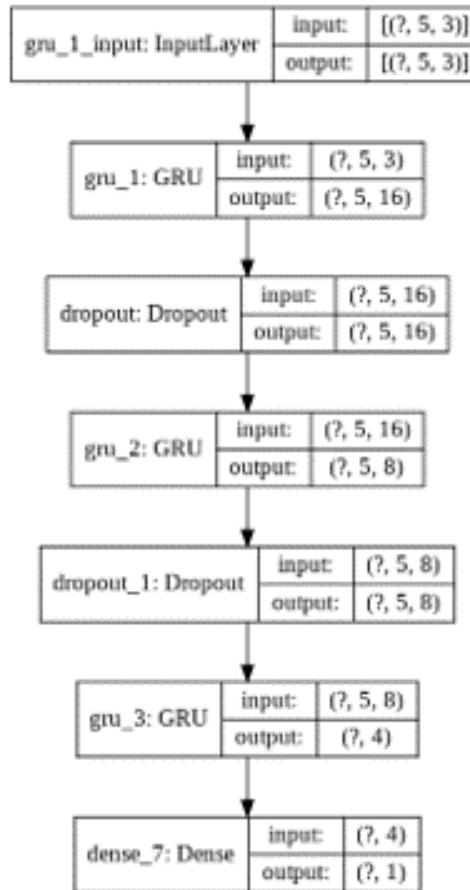


Figura 84: Topologia do modelo RNN_ENDO_2

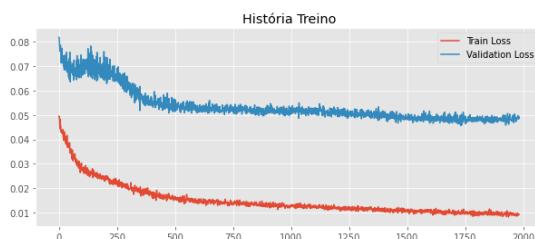


Figura 85: Treino do modelo
RNN_ENDO_2

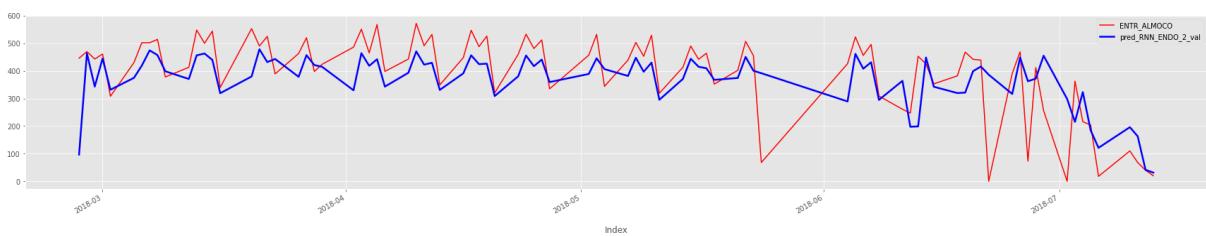


Figura 86: Avaliação do modelo RNN_ENDO_2

Comparativo gráfico da etapa de avaliação entre RNN_ENDO_2 e MLP_ENDO_1 É possível notar, conforme figura 87 que mesmo com correlação e RMSE muito próximo nos 2

Tabela 8: Métricas de avaliação do modelo RNN_ENDO_2

METRICAS DO MODELO RNN_ENDO_2 :	
RMSE	109.81020024143946
TOTAL DE REFEIÇÕES PROJETADAS	32984.970529556274
TOTAL DE REFEIÇÕES CONSUMIDAS	35555
TOTAL DE REFEIÇÕES SUB PROJETADAS	-4821.583518981934
TOTAL DE REFEIÇÕES SUPER PROJETADAS	2251.554048538208

modelos, o modelo RNN_ENDO_2 apresentou melhor valor p-value, no calculo de correlação de Pearson*. Em Junho após uma ocorrência outlier, o modelo GRU teve melhor recuperação de resultados em comparação ao MLP. De acordo com a referência da documentação da biblioteca scipy, o valor p-value indica aproximadamente a probabilidade de um sistema não correlacionado produzir conjuntos de dados que têm uma correlação de Pearson pelo menos tão extrema quanto a calculada a partir desses conjuntos de dados.

. Podemos notar também que o recurso dropout realizou a poda de unidades de forma bem sucedida gerando melhores resultados, em comparação ao modelo RNN_ENDO_1 sem a poda.

RNN_ENDO_2 :

CORRELAÇÃO (r): 0.6753384458677313

Pi (p) :3.903678433027079e-13

MLP_ENDO_1 :

MLP: CORRELAÇÃO (r): 0.6791684214084684

Pi (p) :2.5590689588089596e-13

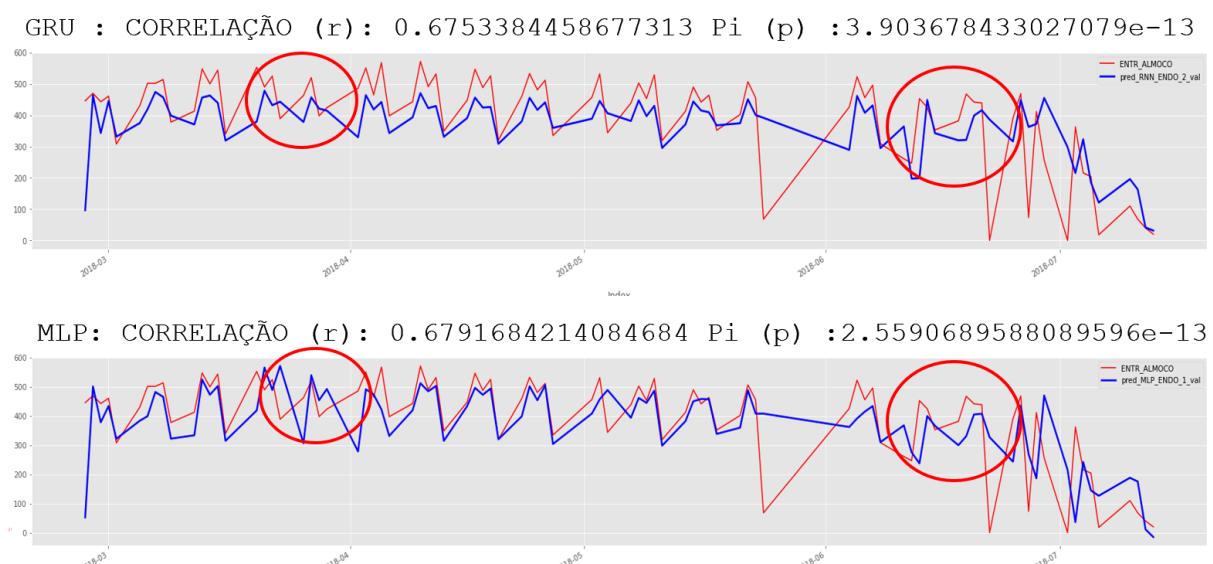


Figura 87: Comparação da etapa de validação entre MLP_ENDO_1 e RNN_ENDO_1

5.1.2.3 Modelos mistos endógenos e exógenos

Nos modelos desta subseção, a coluna de blocos à esquerda na figura de topologia do modelo, trata as entradas endógenas (temporais) com camadas GRU. A coluna à direita trata as entradas exógenas como variáveis climáticas e de calendário com camadas do tipo MLP. A saída de ambas as distintas camadas é concatenada e tratada por uma camada MLP de saída.

RNN_EXO_1 Primeiro modelo misto, com camada GRU à esquerda e MLP à direita.

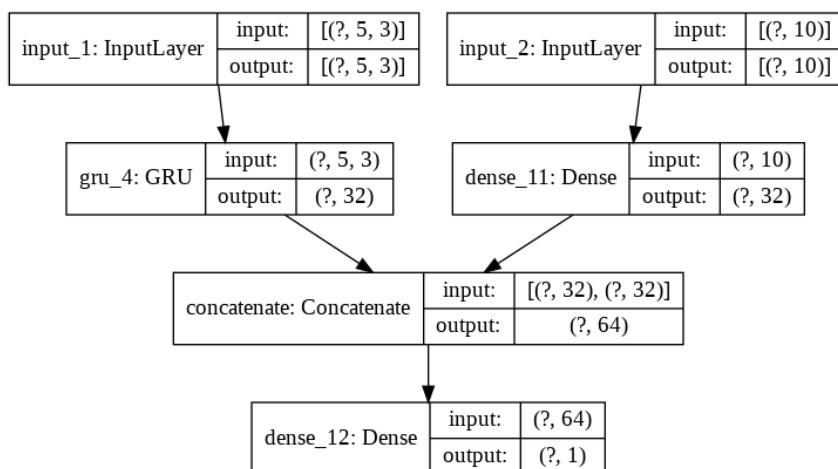


Figura 88: Topologia do modelo RNN_EXO_1

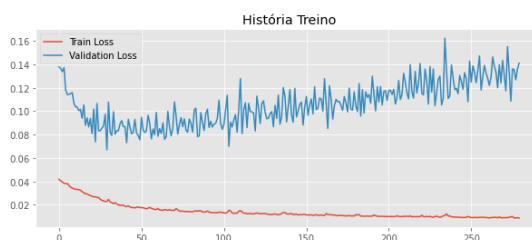


Figura 89: Treino do modelo
RNN_EXO_1

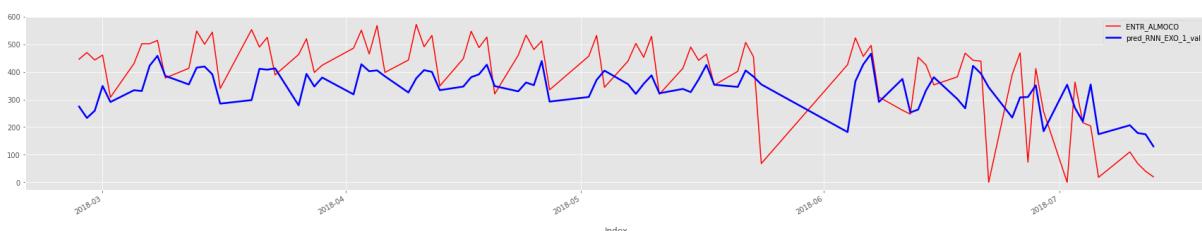


Figura 90: Avaliação do modelo RNN_EXO_1

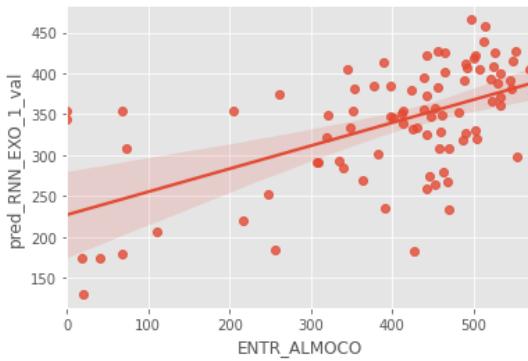


Figura 91: Grafico scatter de avaliação do modelo RNN_EXO_1

Tabela 9: Métricas do modelo RNN_EXO_1

METRICAS DO MODELO RNN_EXO_1 :	
RMSE	132.9496
TOTAL DE REFEIÇÕES PROJETADAS	30223.764709472656
TOTAL DE REFEIÇÕES CONSUMIDAS	35555
TOTAL DE REFEIÇÕES SUB PROJETADAS	-7585.754409790039
TOTAL DE REFEIÇÕES SUPER PROJETADAS	2254.5191192626953

RNN_EXO_2 Segundo modelo misto definido com o aumento da profundidade das camadas GRU e MLP do modelo anterior.

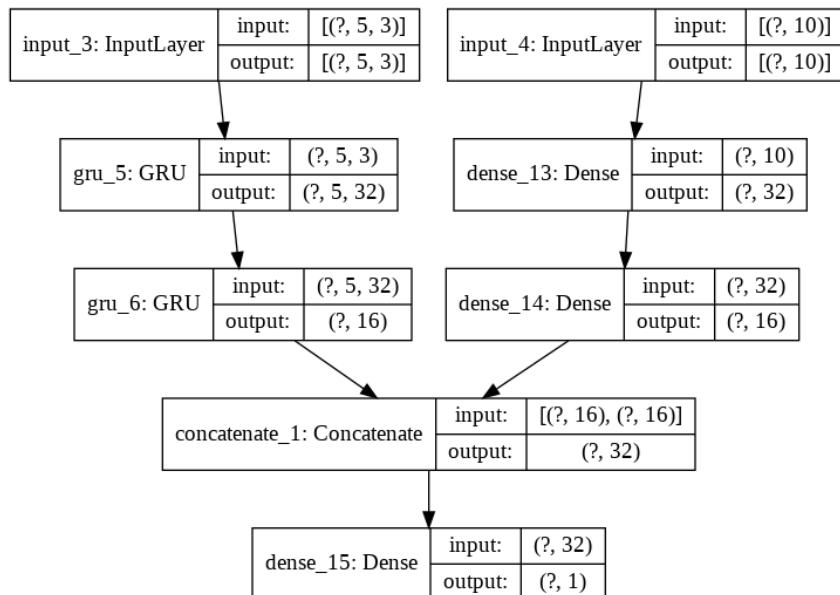


Figura 92: Topologia do modelo RNN_EXO_2

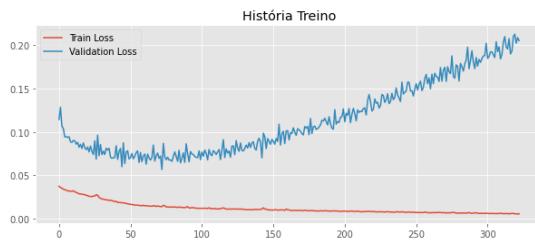


Figura 93: Treino do modelo
RNN_EXO_2

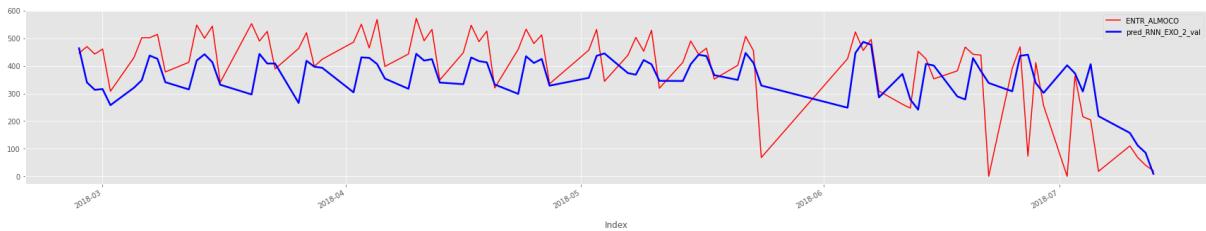


Figura 94: Avaliação do modelo RNN_EXO_2

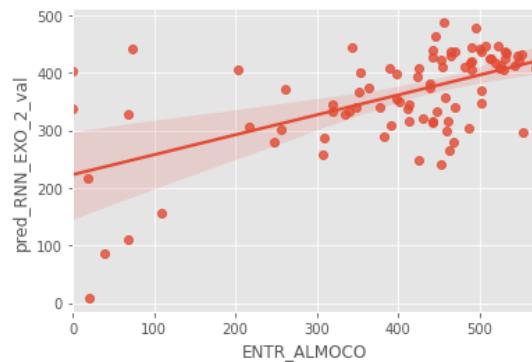


Figura 95: Gráfico scatter de avaliação do
modelo RNN_EXO_2

Tabela 10: Métricas do modelo RNN_EXO_2

METRICAS DO MODELO RNN_EXO_2 :	
RMSE	122.32142544582184
TOTAL DE REFEIÇÕES PROJETADAS	32146.112915039062
TOTAL DE REFEIÇÕES CONSUMIDAS	35555
TOTAL DE REFEIÇÕES SUB PROJETADAS	-5875.795913696289
TOTAL DE REFEIÇÕES SUPER PROJETADAS	2466.9088287353516

RNN_EXO_3 Terceiro modelo misto, com a utilização de dropout na camada GRU baseada no melhor modelo endógeno RNN_ENDO_2, mantendo a camada MLP à direita sem alterações.

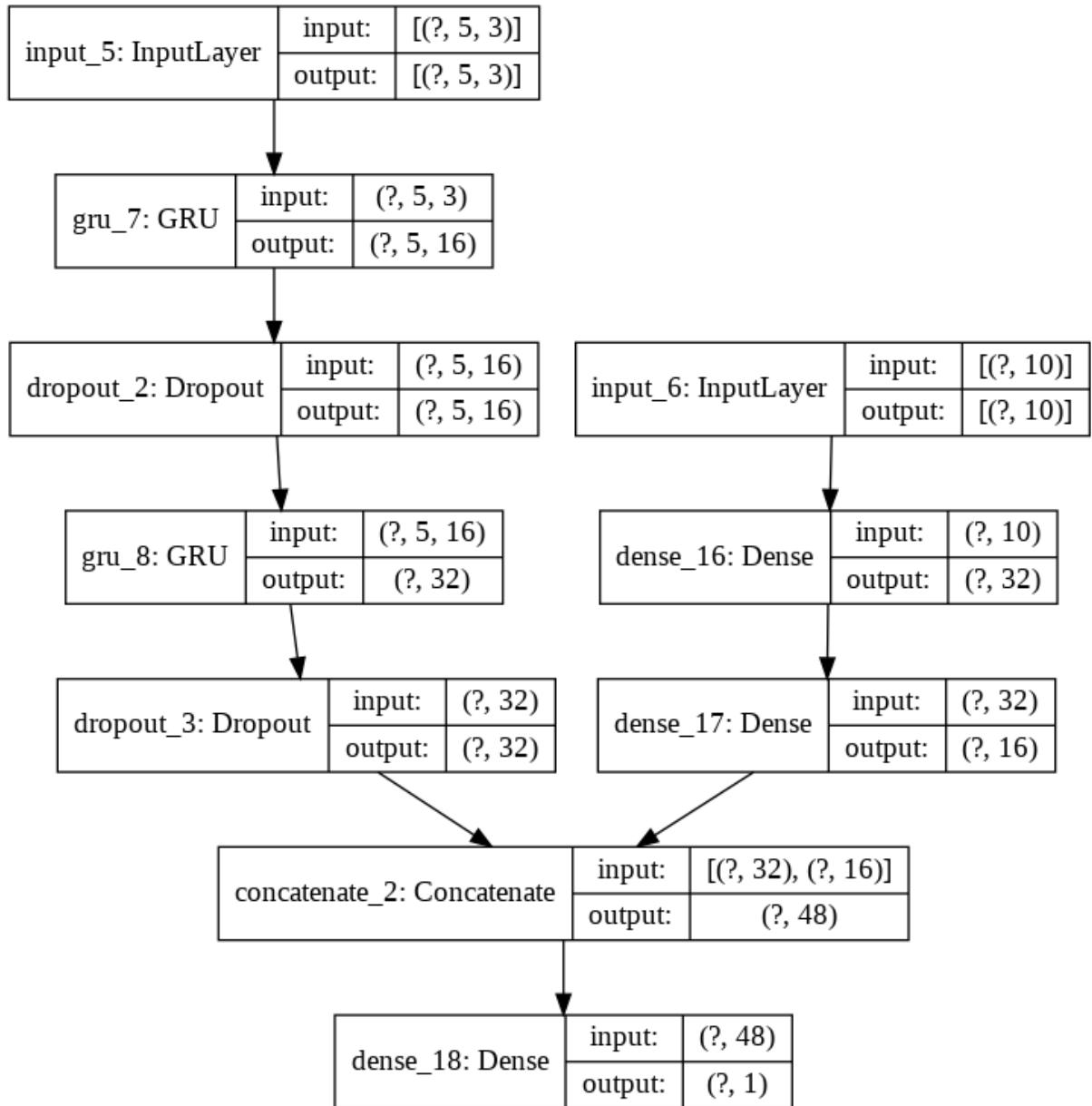


Figura 96: Topologia do modelo RNN_EXO_3

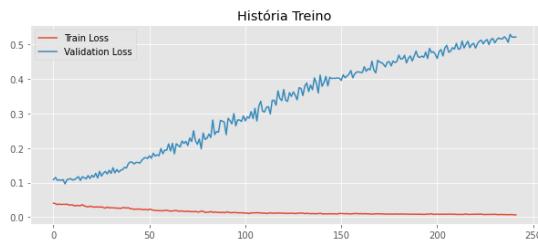


Figura 97: Treino do modelo
RNN_EXO_3

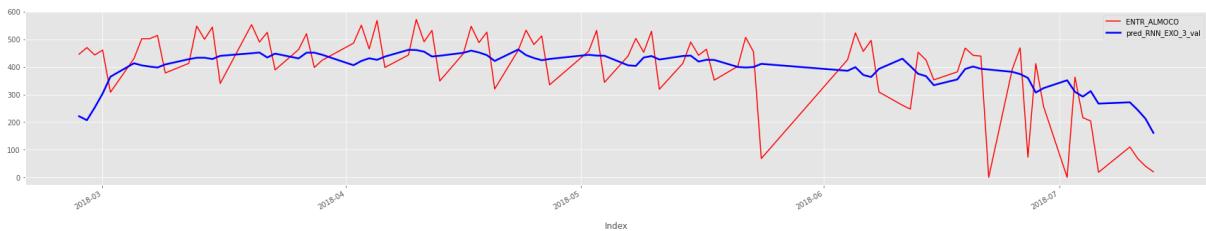


Figura 98: Avaliação do modelo RNN_EXO_3

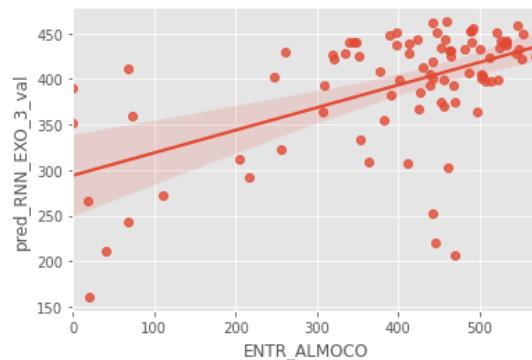


Figura 99: Grafico scatter de avaliação do
modelo RNN_EXO_3

Tabela 11: Métricas do modelo RNN_EXO_3

METRICAS DO MODELO RNN_EXO_3 :	
RMSE	121.25903250027721
TOTAL DE REFEIÇÕES PROJETADAS	35003.56196594238
TOTAL DE REFEIÇÕES CONSUMIDAS	35555
TOTAL DE REFEIÇÕES SUB PROJETADAS	-4471.480010986328
TOTAL DE REFEIÇÕES SUPER PROJETADAS	3920.041976928711

5.1.3 Teste e métricas

Nesta subseção todos os modelos foram testados no conjunto de teste, 1º semestre de 2019. Em práticas com maior volume de modelos, pode-se selecionar apenas os modelos com

os melhores resultados na etapa de validação. Porém, para um número de 6 modelos foi considerado o teste em todos os modelos a fim de documentação dos resultados em proveito à trabalhos futuros. Entre todos os 6 modelos, é selecionado o melhor modelo endógeno, o melhor exógeno, e ambos foram selecionados para comparações com a fase 2.

5.1.3.1 Endógenos : GRU x MLP

Modelo	Correlação	P-value	RMSE:
RNN_ENDO_2	0.5954398951050144	9.422151772316392e-10	108.0662
MLP_ENDO_1	0.5212411617024483	1.9211576129056944e-07	128.0541

Tabela 12: Métricas dos melhores modelos RNN_ENDO_2 x MLP_ENDO_1

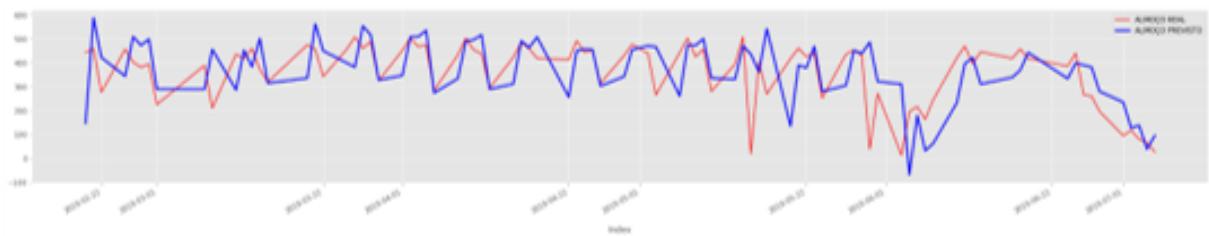


Figura 100: Teste MLP_ENDO_1

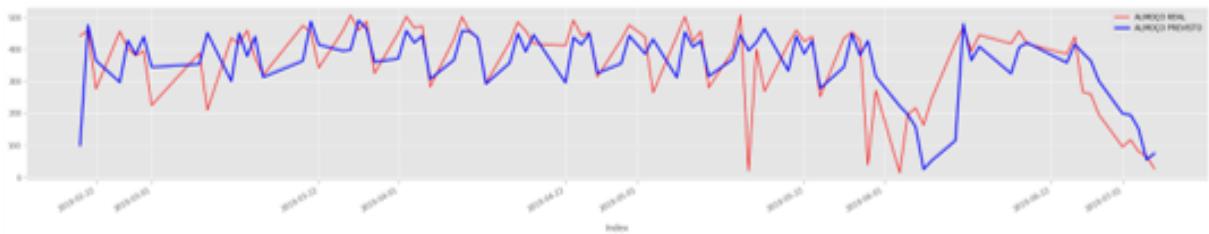


Figura 101: Teste RNN_ENDO_2

Podemos notar na figura 100 e 101 que apesar dos coeficientes de correlação serem próximos no teste entre os modelos RNN_ENDO_2 e MLP_ENDO_1, as métricas da tabela 12 para o valor p-value e RMSE (métrica utilizada no treino dos modelos) é discrepante e seleciona o modelo GRU como o melhor modelo nesta comparação.

5.1.3.2 Melhor modelo Endógeno

De acordo com as métricas de teste, o melhor modelo endógeno foi o RNN_ENDO_2, e suas anomalias de previsão foram devidamente justificadas por datas especiais no calendário do primeiro semestre:

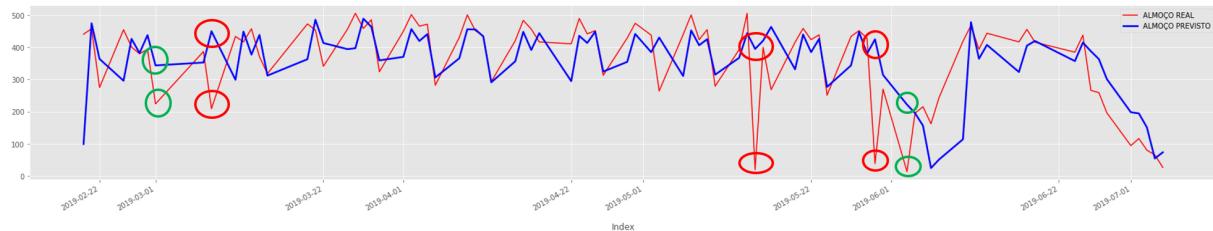


Figura 102: Analise de anomalias preditivas do RNN_ENDO_2

As datas que produziram outliers e anomalias de previsão no melhor modelo endógeno, podem ser observadas na figura 102, sendo os pontos verdes outliers onde o modelo seguiu a tendência de alta ou baixa de consumo mas obteve erro discrepante, e nos pontos vermelhos seguiu tendência inversa:

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

Datas com anomalias de previsão onde o modelo seguiu tendência oposta ao consumo, pontos vermelhos:

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

Métricas do melhor modelo:

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
Total de Refeições em falta	-2982,567947
Total Descartadas	3478,957266
ERRO_ABS_MEDIANO	46,70721436
ERRO_ABSOLUTO_PERCENTUAL_MEDIO	74,93539002

Tabela 13: Métricas do melhor modelo: RNN_ENDO_2

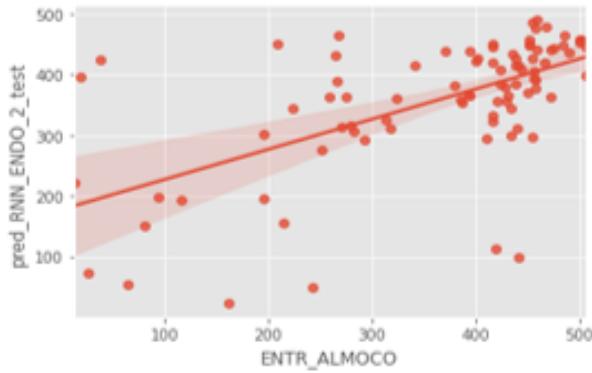


Figura 103: Grafico scatter de teste do RNN_ENDO_2

5.1.3.3 Modelos mistos

Todos os modelos mistos compõem uma mesma classe, que combina camadas GRU com MLP, portanto as métricas são registradas à seguir, por paragrafo, para cada modelo. Diferentemente nos endógenos onde haviam 2 modelos GRU e 1 MLP compondo classes diferentes e sendo necessário um paragrafo extra de comparação entre o melhor GRU e o MLP. Os parágrafos a seguir apenas reúnem todas as informações entre os modelos para extrair o melhor modelo com a melhor métrica de RMSE.

RNN_EXO_1

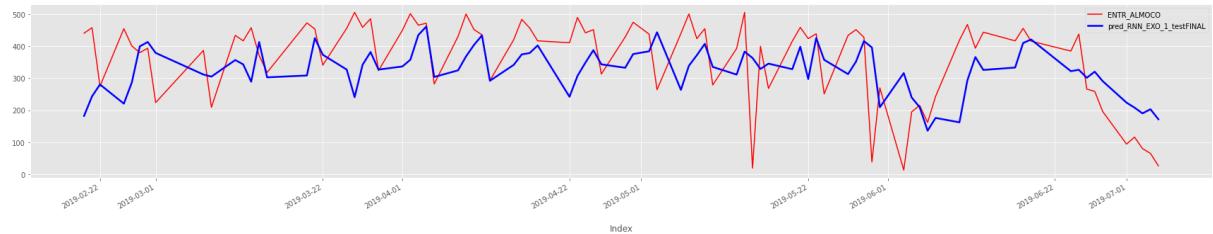


Figura 104: Teste do modelo RNN_EXO_1

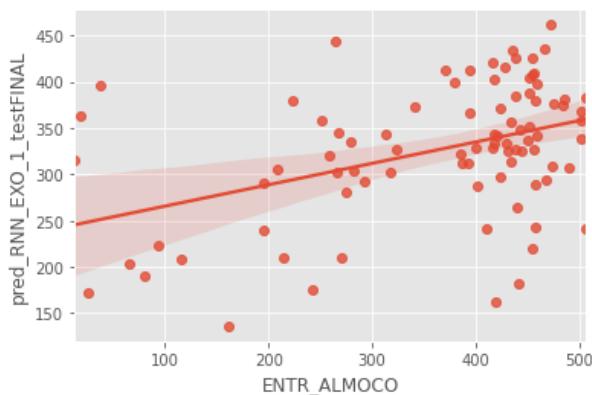


Figura 105: Grafico scatter de teste do modelo RNN_EXO_1

Tabela 14: Erros do modelo RNN_EXO_1

RNN_EXO_1:	
TOTAL DE REFEIÇÕES CONSUMIDAS	31962 : 88
TOTAL DE REFEIÇÕES PROJETADAS	28728.816 : 88
ERRO DE PREVISÃO	-3233.18359375
PERCENTAGEM DE ERRO	-10.11571113744466%
CORRELAÇÃO (r)	0.4122158648305426
Pi (p)	6.590759241637016e-05
R2	0.16992191921799218
RMSE	124.49076120428357
ERRO TOTAL DE REFEIÇÕES SUB PROJETADAS	-2709.1732788085938
ERRO TOTAL DE REFEIÇÕES SUPER PROJETADAS	5942.356475830078
ERRO ABSOLUTO MEDIANO	85.59107971191406
ERRO ABSOLUTO PERCENTUAL MÉDIO	90.98694558104974%

RNN_EXO_2

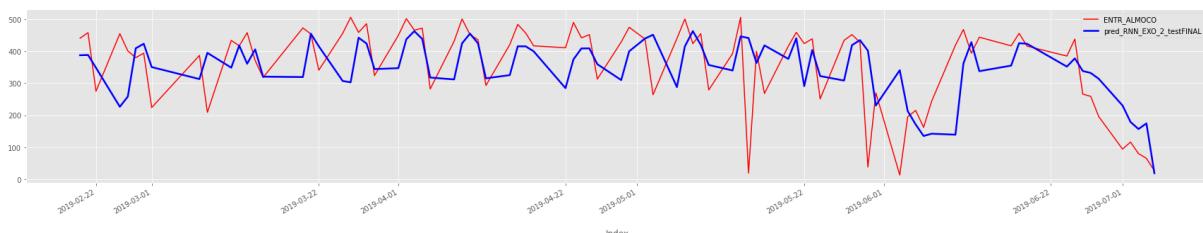


Figura 106: Teste do modelo RNN_EXO_2

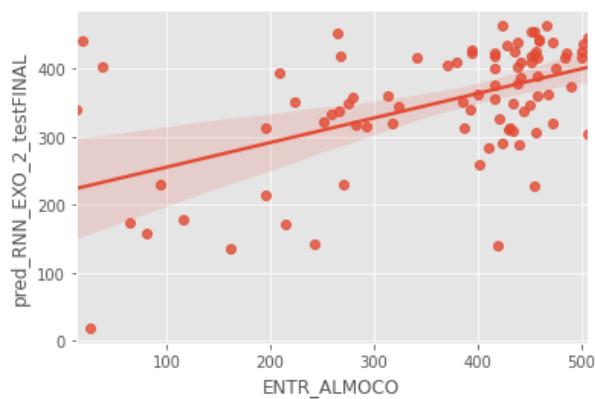
Figura 107: Gráfico scatter de teste do modelo
RNN_EXO_2

Tabela 15: Erros do modelo RNN_EXO_2

RNN_EXO_2:	
TOTAL DE REFEIÇÕES CONSUMIDAS	31962 : 88
TOTAL DE REFEIÇÕES PROJETADAS	30823.148 : 88
ERRO DE PREVISÃO	-1138.8515625
PERCENTAGEM DE ERRO	-3.563142364370189
CORRELAÇÃO (r)	0.5206433612135913
Pi (p)	1.995208420251518e-07
R2	0.27106950957578635
RMSE	112.99211377165491
ERRO TOTAL DE REFEIÇÕES SUB PROJETADAS	-3044.8834533691406
ERRO TOTAL DE REFEIÇÕES SUPER PROJETADAS	4183.734128952026
ERRO ABSOLUTO MEDIANO	63.59894561767578
ERRO ABSOLUTO PERCENTUAL MÉDIO	88.26875085831726%

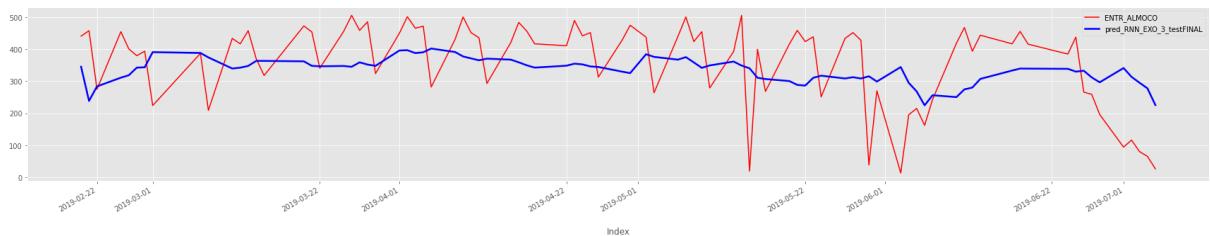


Figura 108: Teste do modelo RNN_EXO_3

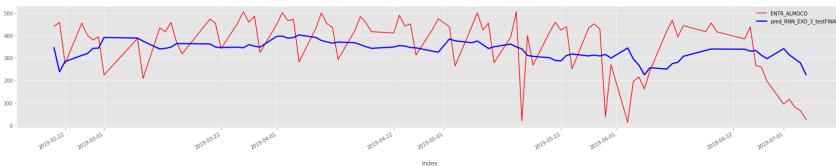


Figura 109: Grafico scatter de teste do modelo RNN_EXO_3

Tabela 16: Erros do modelo RNN_EXO_3

RNN_EXO_3:	
TOTAL DE REFEIÇÕES CONSUMIDAS	31962 : 88
TOTAL DE REFEIÇÕES PROJETADAS	29425.791 : 88
ERRO DE PREVISÃO -2536.208984375	
PERCENTAGEM DE ERRO	-7.935075978896815%
CORRELAÇÃO (r)	0.33963283016931
Pi (p)	0.0012067360859947137
R2	0.11535045932881538
RMSE	124.65810037942255
ERRO TOTAL DE REFEIÇÕES SUB PROJETADAS	-3417.6602630615234
ERRO TOTAL DE REFEIÇÕES SUPER PROJETADAS	5953.868759155273
ERRO ABSOLUTO MEDIANO	100.44429016113281
ERRO ABSOLUTO PERCENTUAL MÉDIO	98.79064922123037%

RNN_EXO_3

5.1.4 Melhor modelo da 1a fase

Entre todos os modelos, o melhor modelo produzindo o menor RMSE e com vantagem em todas as outras métricas foi o modelo endógeno, RNN_ENDO_2, ressaltando que o gráfico e as métricas de previsões deste modelo se encontram na figura 101. Na ultima seção deste capítulo são apresentadas todas as métricas de todos os modelos de todas as fases experimentais.

5.2 2a Fase Experimental

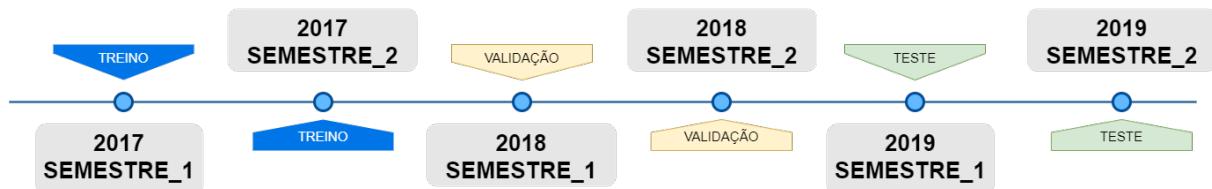


Figura 110: Domínio temporal da 2a fase

Nesta segunda fase, todos os modelos da fase anterior foram importados, treinados e testados sobre um novo domínio temporal de acordo com a figura 110.

5.2.1 Pré-Processamento

Nesta subseção para a etapa de pré-processamento foram avaliados apenas os resultados obtidos da divisão dos conjuntos e da técnica aplicada do restaurante.

Divisão dos conjuntos

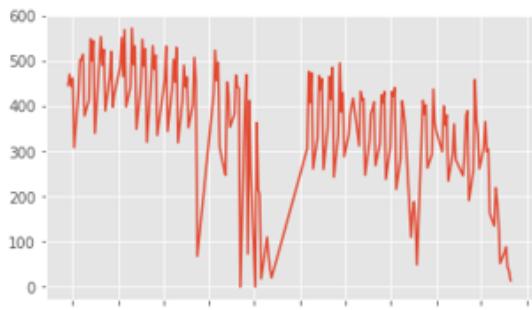


Figura 111: Conjunto de dados de treino,
ano de 2017 completo

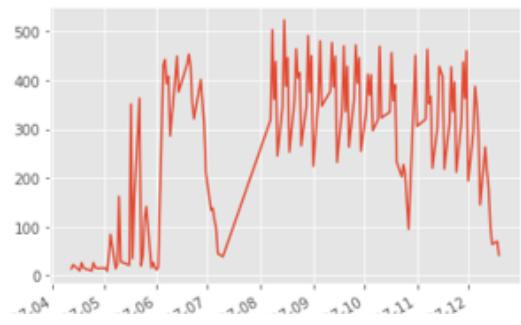


Figura 112: Conjunto de dados de
validação, ano de 2018 completo

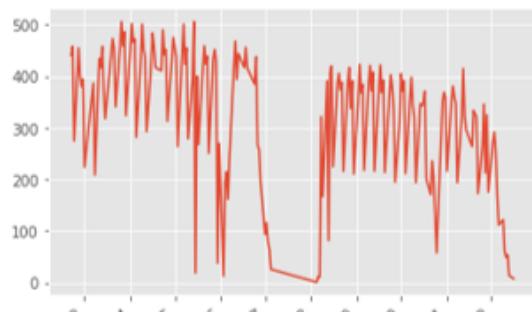


Figura 113: Conjunto de dados de teste,
ano de 2019 completo

A nova divisão dos conjuntos segue conforme as figuras 111 para o conjunto de treino, 111 para o conjunto de validação e 111 para o conjunto de teste.

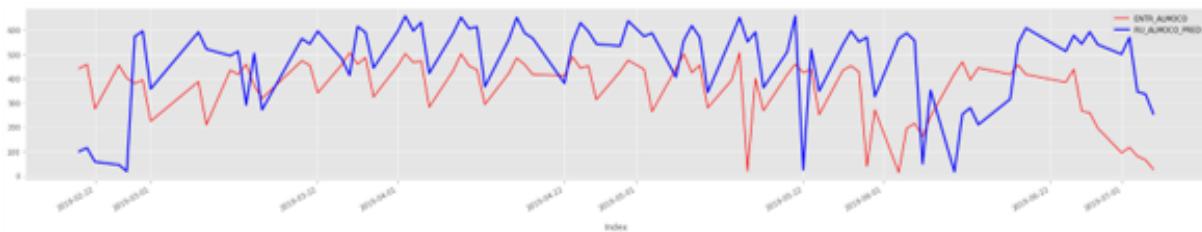


Figura 114: Técnica do restaurante para a 2a fase.

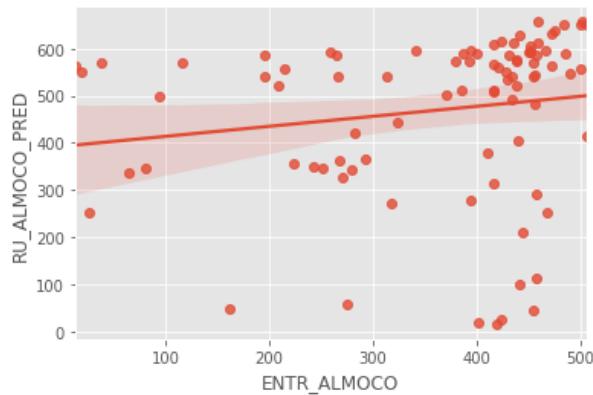


Figura 115: Gráfico scatter da técnica do restaurante para a 2a fase

Para a produção de refeições com margem de 30% acima do consumo do 5º dia anterior, conforme figura 114 e 115 obteve-se as seguintes métricas de acordo com a tabela 17. É possível notar que o descarte e anomalias de predição tomam grandes proporções a partir do segundo semestre até o fim do ano.

Tabela 17: Métricas da técnica do restaurante para a 2a fase.

Consumo com margem 30% acima do 5º dia anterior	
TOTAL DE REFEIÇÕES CONSUMIDAS	58653
TOTAL DE REFEIÇÕES PROJETADAS	76262
CORRELAÇÃO (r)	0.40067947341844423
Pi (p)	2.0845891721642294e-08
RMSE	191.7620291511743
TOTAL DE REFEIÇÕES SUPER PROJETADAS	23412
TOTAL DE REFEIÇÕES SUB PROJETADAS	-5803
ERRO ABSOLUTO MEDIANO	133.0
ERRO ABSOLUTO PERCENTUAL MÉDIO	205.61135949728225%

5.2.2 Treino e Validação dos modelos

Nesta subseção são registrados os resultados de treino, validação e teste dos modelos de forma sucinta, pois os mesmos já foram detalhados na 1a fase.

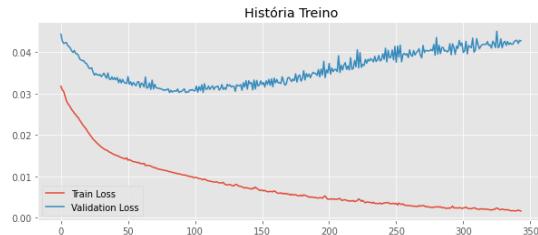
MLP_ENDO_1

Figura 116: Gráfico de treino do modelo MLP_ENDO_1 na 2a fase.

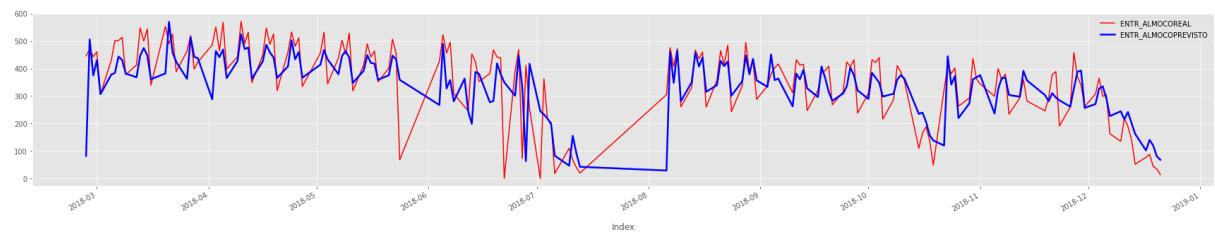


Figura 117: Gráfico de validação do modelo MLP_ENDO_1 na 2a fase.

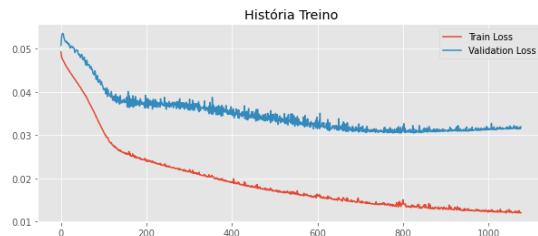
RNN_ENDO_1

Figura 118: Gráfico de treino do modelo RNN_ENDO_1 na 2a fase.

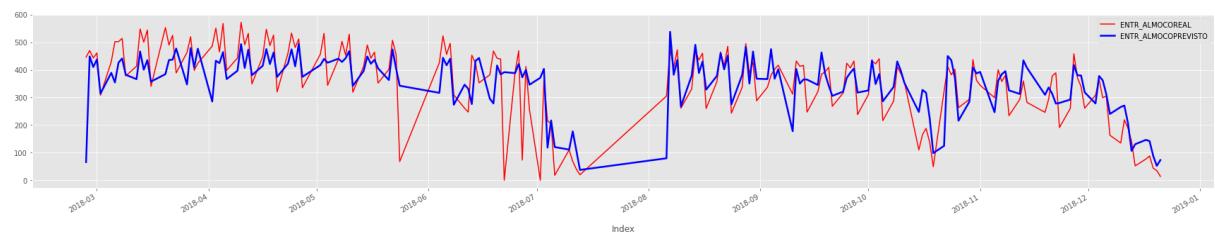


Figura 119: Gráfico de validação do modelo RNN_ENDO_1 na 2a fase.

RNN_ENDO_2

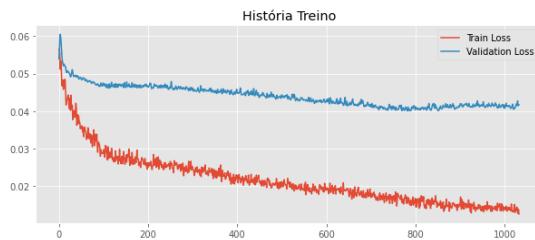


Figura 120: Gráfico de treino do modelo RNN_ENDO_2 na 2a fase.

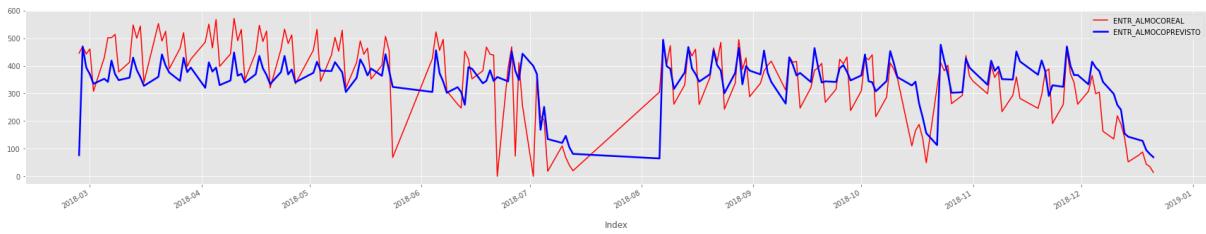


Figura 121: Gráfico de validação do modelo RNN_ENDO_2 na 2a fase.

RNN_EXO_1

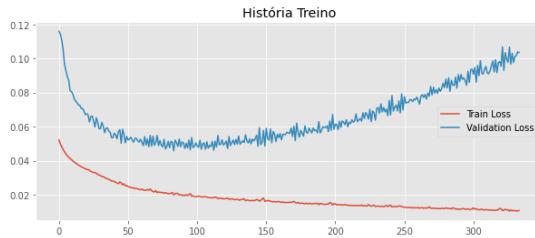


Figura 122: Gráfico de treino do modelo RNN_EXO_1 na 2a fase.

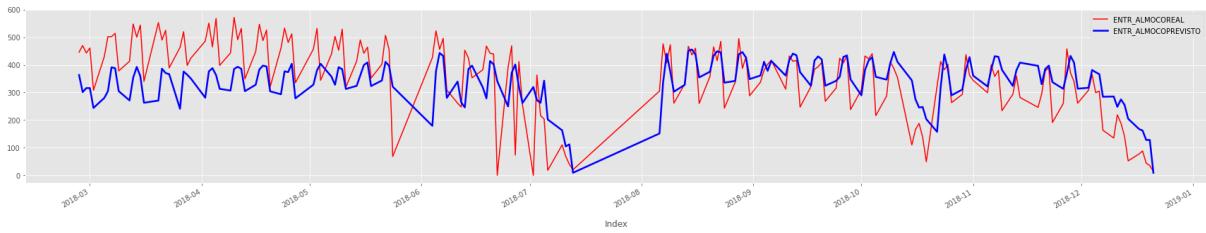


Figura 123: Gráfico de validação do modelo RNN_EXO_1 na 2a fase.

RNN_EXO_2

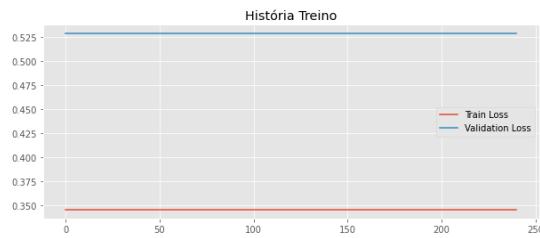


Figura 124: Gráfico de treino do modelo RNN_EXO_2 na 2a fase.

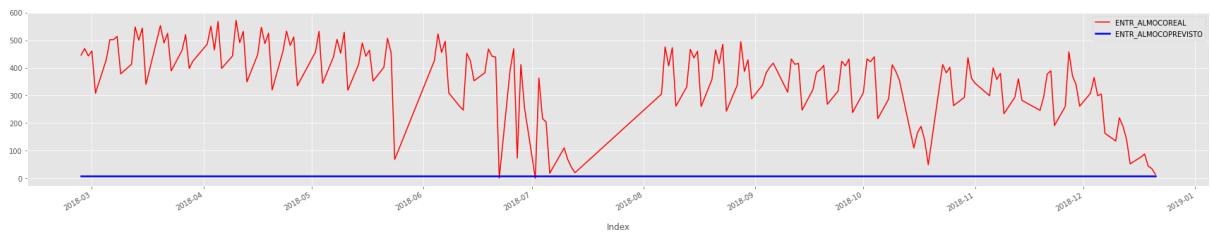


Figura 125: Gráfico de validação do modelo RNN_EXO_2 na 2a fase.

RNN_EXO_3

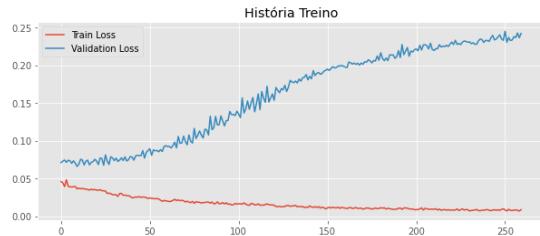


Figura 126: Gráfico de treino do modelo RNN_EXO_3 na 2a fase.

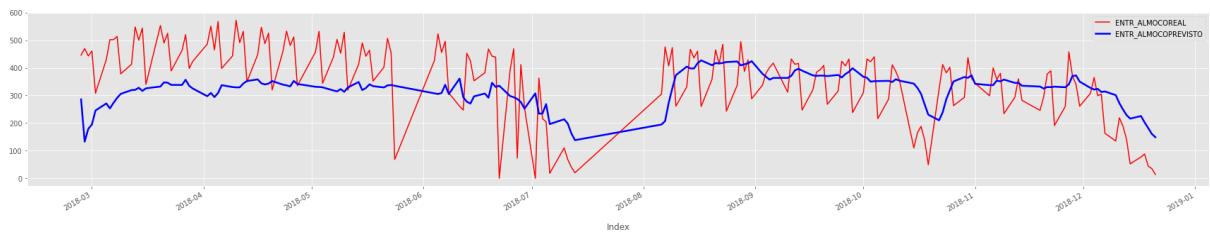


Figura 127: Gráfico de validação do modelo RNN_EXO_3 na 2a fase.

Conclusão sobre a validação dos modelos Após a análise de todos os gráficos desde a figura 116 à figura 127 de validação dos modelos desta segunda fase, nota-se que todos os modelos endógenos obtiveram melhoria nos resultados avaliados sobre o conjunto de validação. Já para os modelos exógenos, apenas 1 obteve melhoria significativa nos resultados, o modelo RNN_EXO_1 de acordo com a figura 123, e os outros obtiveram resultados anômalos no plot

dos gráficos de treino e validação, como o modelo _EXO_2 que obteve uma falha no algoritmo de treino, conforme figura 125, e o modelo _EXO_3 que convergiu à um overfitting evidenciado pela figura 127.

5.2.3 Teste

A seguir os modelos da segunda fase são testados para a seleção do melhor modelo. As figuras abaixo são para fins de registro dos resultados, a seguir é selecionado o melhor modelo com menor RMSE e para critério de desempate são selecionadas outras métricas como quantia de refeições descartadas e em falta, ou valor R^2 para mensuração do comportamento de consumo (tendencias de altas e baixas) com o comportamento previsto. O restaurante também pode selecionar o melhor modelo para fins particulares da gestão como por exemplo o modelo que tem menor erro de descarte ou menor erro de falta de refeições.

MLP_ENDO_1

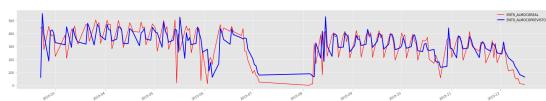


Figura 128: Gráfico de teste do modelo MLP_ENDO_1.

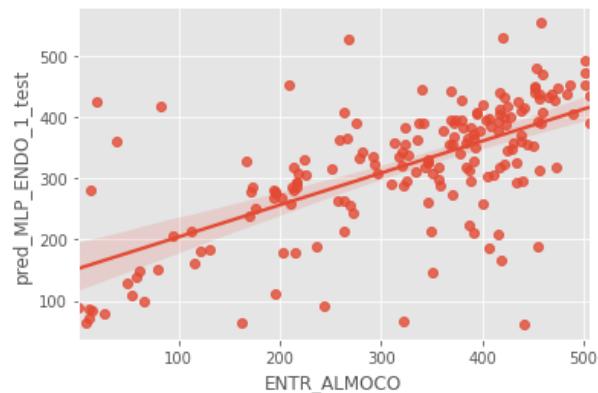


Figura 129: Gráfico de scatter do modelo MLP_ENDO_1.

RNN_ENDO_1

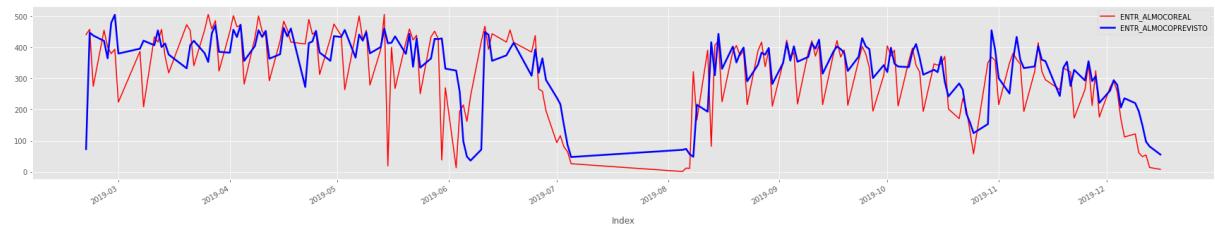


Figura 130: Gráfico de teste do modelo RNN_ENDO_1.

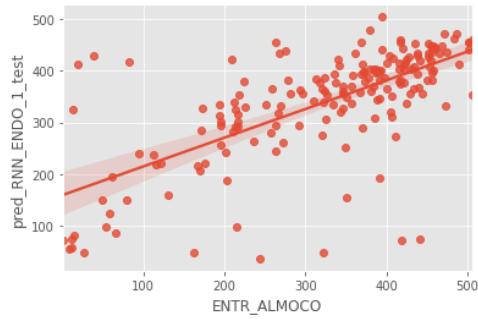


Figura 131: Gráfico de scatter do modelo RNN_ENDO_1.

RNN_ENDO_2

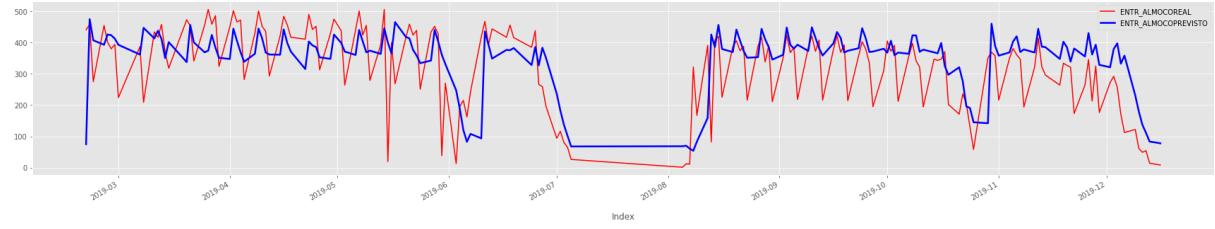


Figura 132: Gráfico de teste do modelo RNN_ENDO_2.

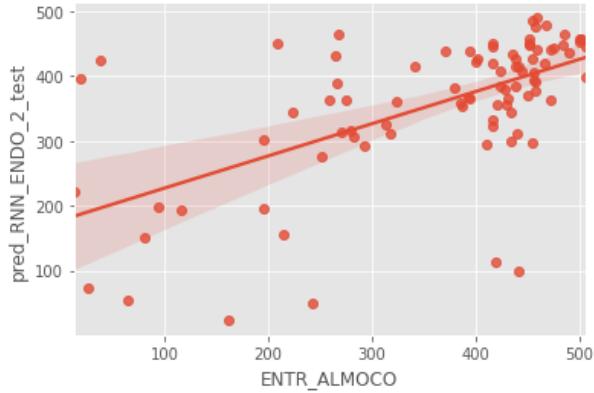


Figura 133: Gráfico de scatter do modelo RNN_ENDO_2.

RNN_EXO_1

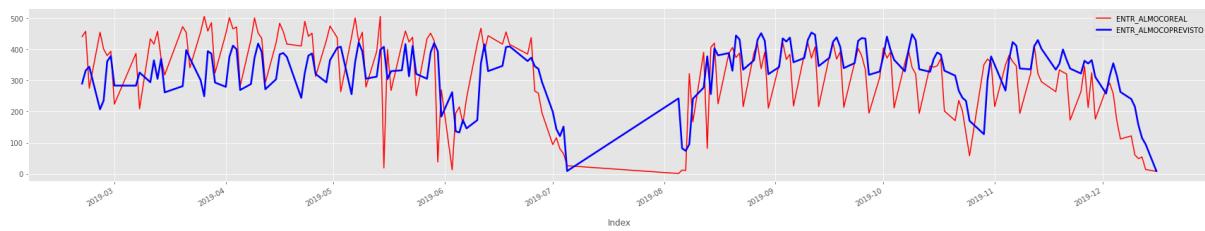


Figura 134: Gráfico de teste do modelo RNN_EXO_1.

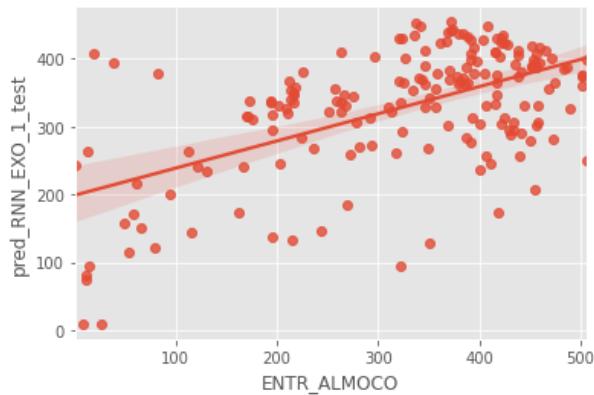


Figura 135: Gráfico de scatter do modelo RNN_EXO_1.

RNN_EXO_2

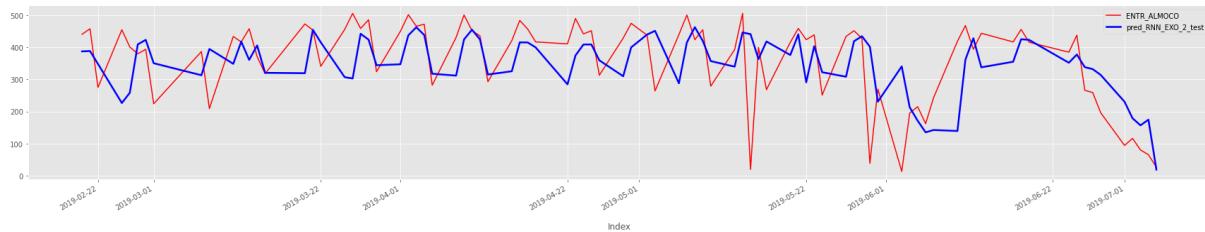


Figura 136: Gráfico de teste do modelo RNN_EXO_2.

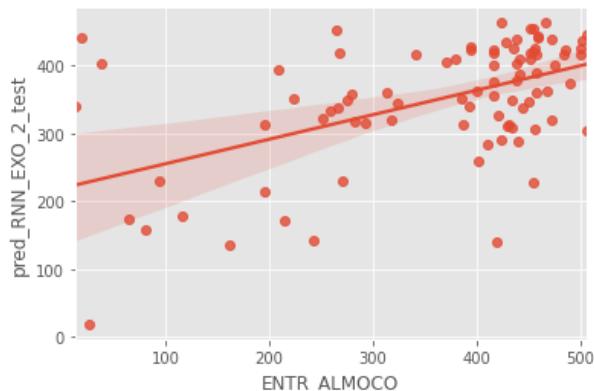


Figura 137: Gráfico de scatter do modelo RNN_EXO_2.

RNN_EXO_3

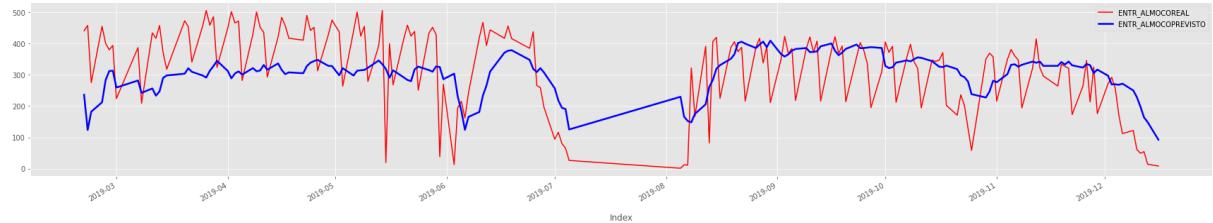


Figura 138: Gráfico de teste do modelo RNN_EXO_3.

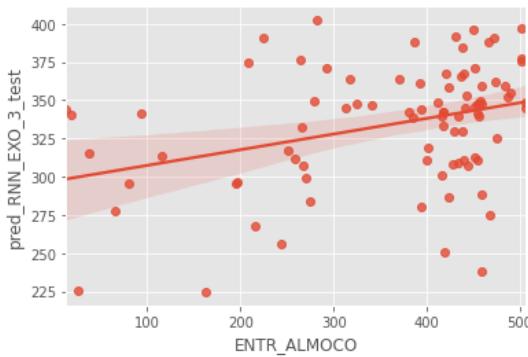


Figura 139: Gráfico de scatter do modelo RNN_EXO_3.

5.2.4 Melhor Modelo

Após o teste de todos os modelos da 2a fase, o modelo que obteve as melhores métricas entre todos, foi RNN_EXO_1, onde o gráfico de previsões pode ser conferido na figura 134. E o gráfico scatter com a representação da regressão linear entre o conjunto real de teste e as previsões pode ser conferido na figura 135.

Todos os gráficos de teste dos modelos da segunda fase estão entre a figura 128 à 139

Teste do melhor modelo da 2a fase no domínio da primeira fase Como o melhor modelo obtido na 2a fase, se saiu melhor do que o melhor modelo da primeira fase, foi feito um novo teste do modelo RNN_EXO_1 com o domínio de testes restrito para o primeiro semestre (domínio da primeira fase) a fim de uma comparação justa com o melhor modelo da primeira fase, o RNN_ENDO_2.

Tabela 18: Tabela de teste do RNN_EXO_1 segunda fase no domínio da primeira fase

Melhor modelo da segunda fase testado no domínio do primeiro semestre da primeira fase	
TOTAL DE REFEIÇÕES CONSUMIDAS	31962
TOTAL DE REFEIÇÕES PROJETADAS	32170.246
ERRO DE PREVISÃO	208.2460 0.6515% do total consumido
CORRELAÇÃO (r)	0.5903766574738285
P-value (p)	1.4143e-09 R2 :0.3485
RMSE	106.2080
ERRO TOTAL DE REFEIÇÕES SUB PROJETADAS	-3454.8698
ERRO TOTAL DE REFEIÇÕES SUPER PROJETADAS	3246.6228
ERRO ABSOLUTO MEDIANO	59.5414
ERRO ABSOLUTO PERCENTUAL MÉDIO	83.2671%

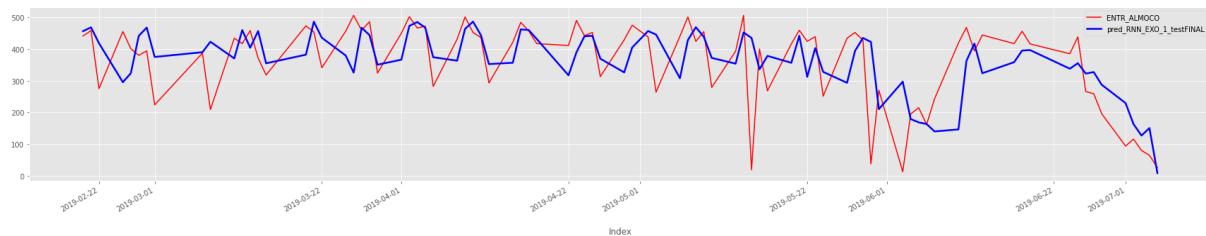


Figura 140: Gráfico de teste do RNN_EXO_1 segunda fase no domínio da primeira fase

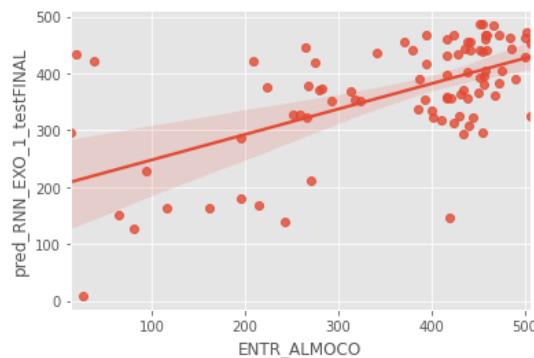


Figura 141: Gráfico scatter Gráfico de teste do RNN_EXO_1 segunda fase no domínio da primeira fase

5.2.5 Tabela completa de todos os modelos

A seguir são listadas as tabelas com todos resultados experimentais.

6 Conclusão

6.1 A importância da metodologia de divisão do conjunto de dados em séries temporais

No trabalho com um conjunto de dados de sazonalidade temporal, é notório que a ordenação dos dados na separação do conjunto em 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 37, e depois corrigido conforme a figura 38 valida esta afirmação.

6.2 Sobre o método de produção de refeições com margem de erro e análise da semana anterior

Mesmo com a produção de 30% acima do consumo na semana anterior, no fim de cada semestre, o restaurante do ICT Unifesp descarta mais do que 30%, pois o comportamento oscilatório do consumo e o acréscimo de outliers, acaba ampliando o erro. No ano de 2019, seguindo este método, 23 mil refeições foram descartadas. O Modelo RNN_EXO_3 que trouxe o maior descarte entre todos os 12 modelos testados, realizou 8914 descartes. Isso evidencia a necessidade de se implementar métodos eficientes para a produção e planejamento de refeições no restaurante universitário da Unifesp.

6.3 Sobre a sazonalidade semanal

Em todo o conjunto de dados, nos dias da semana, a sexta-feira é o dia de menor consumo independente do período do ano. Já as datas de terça e quinta concentram a maior movimentação de consumo.

6.4 Sobre o ajuste empírico da topologia dos modelos

Na etapa de validação dos primeiros modelos desenvolvidos, demonstrado na subseção "Ajuste empírico de topologia da 1a fase", conforme figuras 74a e 77a do capítulo de resultados da 1a fase, foi possível notar a diminuição do RMSE (Raiz do erro quadrático médio) ao aumentar a profundidade da rede Perceptron para treino e avaliação sob o conjunto de valida-

ção. Validando a hipótese de que os modelos tem capacidade de aprendizado do problema em relação ao ajuste da topologia dos mesmos.

6.5 Sobre os resultados preliminares da etapa de validação

Apesar de valores de correlação muito próximos, Foi possível notar na 1a fase algumas diferenças gráficas entre o modelo RNN_ENDO_2 e o modelo ‘MLP_ENDO_1 na etapa de validação dos modelos. Porém na etapa de testes, o modelo RNN_ENDO_2 obteve RMSE 108 e correlação de 0,59 enquanto o MLP obteve RMSE de 128 e correlação de 0,52. Evidenciando que métricas muito próximas, entre modelos, na etapa de teste sobre o conjunto de validação pode evidenciar que não devemos realizar uma seleção previa destes modelos para a etapa de testes, sendo necessário o teste dos 2 modelos para comparações finais. Já os modelos que apresentaram diferenças de RMSE discrepantes na etapa de validação, mantiveram essas diferenças (para o pior e o melhor) na etapa de testes.

6.6 Sobre os erros anômalos de predição

Apesar do conjunto de dados conter 2 features 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 por tais features, indicando a necessidade de mais pesquisa de features que possam representar melhor este comportamento.

6.7 Sobre o modelo com melhor resultado

No 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 as features exógenas (onde a maioria tem sazonalidade anual como as climáticas limitadas às estações do ano) foram ruidosas no aprendizado. Para o modelo RNN_EXO_1 da 2a fase, se saindo melhor que todos os modelos deste trabalho, algumas melhorias são indicadas para trabalhos futuros.

- Aumentar o conjunto de dados para o modelo se ajustar as sazonalidades semestrais e à troca de semestres. As features categóricas que indicam os semestres e dia da semana, bem com as features que quantificam recessos (distância registro anterior e posterior) tem potencial de agregar aprendizado nessa questão, mas é necessário uma diversificação maior do conjunto de dados, pois é treinado apenas com 1 período sazonal (1 ano para treino, 1 para validação e 1 para teste).

- Acrescentar features de eventos importantes para identificar paralisações e eventos do tipo.
- Uma feature de cardápio tem potencial de aumentar a qualidade da predição.
- Uma feature 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).

6.8 Conclusões gerais

As análises diversas de previsão de demanda para o tema abordado requerem extensos métodos de implementação e estruturação de dados. Uma das etapas mais importantes do trabalho, é o método de coleta de dados. Muitos podem ser de acesso burocrático, ou de difícil busca, e são o requisito primordial para o inicio de qualquer análise.

A diversidade de métodos de aprendizado de máquina é imensurável, e dentro de apenas uma análise, que é o treino com retropropagação, pode-se montar infinitas topologias diferentes com base na estrutura dos dados coletados. 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 e problema a ser abordado.

Todavia foi notório a eficiência dos modelos de aprendizado de máquina em trabalhos relacionados à restaurantes universitários.

Como no 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 dessa pesquisa e sua continuação com novos métodos após a conclusão deste trabalho torna-se viável.

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 7 vezes nas páginas 28, 34, 35, 40, 41, 44 e 52.

BRYAN KOLB; ARIF, M. Harnessing the power of neuroplasticity for intervention. In: *FRONTIERS IN HUMAN NEUROSCIENCE*. [S.l.: s.n.], 2014. p. 13. Citado na página 36.

CLARICE G.B. DEMÉTRIO; ZOCCHI, S. S. *Modelos de Regressão*. [S.l.]: Departamento de Ciências Exatas, ESALQ, USP, 2011. Citado na página 28.

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 28 e 48.

FERNANDO., R. J. C. M. F. D. F. Utilização de redes neurais artificiais para a determinação do números de refeições diárias de um restaurante universitário. *Rev. Nutr., Campinas*, Universidade Estadual Paulista Júlio de Mesquita Filho., n. 1, p. 8, 2011. Citado 6 vezes nas páginas 28, 43, 44, 45, 51 e 57.

HAYKIN, S. *Neural networks, a comprehensive foundation*. [S.l.]: Macmillan, 1994. Citado 2 vezes nas páginas 28 e 37.

Jason Brownlee. Ebook machine learning mastery. 2020. Access date: 3 oct. 2020. Disponível em: <<https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>>. Citado na página 62.

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 9 vezes nas páginas 27, 28, 31, 32, 33, 35, 39, 53 e 65.

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 51.

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.novoscursos.ufv.br/projetos/ufv/gesep/www/wp-content/uploads/5.>> Citado 5 vezes nas páginas 28, 43, 44, 51 e 57.

LUGER, G. F. *Inteligência Artificial. Estruturas e Estratégias para a Solução de Problemas Complexos*, 4^a ed. [S.l.]: Porto Alegre: Bookman, 2004. Citado na página 35.

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 6 vezes nas páginas 29, 38, 39, 40, 43 e 45.

MC AI. Overview. 2020. Access date: 3 oct. 2020. Disponível em: <<https://mc.ai/complete-guide-of-activation-functions/>>. Citado na página 61.

MORETTIN, P. A. *Análise de séries temporais.* [S.l.]: Editora Blucher, 1987. Citado na página 34.

RAGA 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 45.

Rebecca Vickery. The art of finding the best features for machine learning. Access date: 7 oct. 2020. Citado na página 65.

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>. Citado 3 vezes nas páginas 28, 35 e 52.

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. Disponível em: <http://alegrete.unipampa.edu.br/gesep/wp-content/uploads/2010/05/Previs%C3%A3o-de-demanda-utilizando-m%C3%A9todo-de-Markov_Final.pdf>. Citado 3 vezes nas páginas 28, 35 e 52.

ZORZETTO, R. Recontagem de neurônios põe em xeque ideias da neurociência. *Revista Fapesp* 192, Fapesp, v. 1, n. 1, p. 6, 2012. Citado na página 36.

7 Anexos

7.1 1a Fase Experimental

7.1.1 Pré-Processamento e treino dos modelos

<https://colab.research.google.com/drive/1spSZDJk1EwK9UK-jfYbWinKEq0EEYIuM?usp=sharing>

7.1.2 Importação e aplicação de métricas dos modelos

https://colab.research.google.com/drive/1vLx4TmsDy0jxVF0BFLng9XFdxgR_IAPL?usp=sharing

7.2 2a Fase Experimental

7.2.1 Pré-Processamento e treino dos modelos

<https://colab.research.google.com/drive/1U4pTAAzBQDS3s5LFZqK1-isa3pHJQ9D8?usp=sharing>

7.2.2 Importação e aplicação de métricas dos modelos

<https://colab.research.google.com/drive/1bBMTP9YQcy6sRNHZTr3NRSt1YCxJyu5t?usp=sharing>