



Douglas Diniz Landim

**ANÁLISE DE DEMANDA DO RESTAURANTE
UNIVERSITÁRIO DO I.C.T UNIFESP VIA REDES
NEURAIS**

São José dos Campos, SP

Douglas Diniz Landim

ANÁLISE DE DEMANDA DO RESTAURANTE UNIVERSITÁRIO DO I.C.T 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 I.C.T 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 DEZEMBRO/2018:

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 amigos universitários, e a todos os professores que me acompanharam e ofereceram desafios em todos esses anos 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 Vinicius Veloso e Fabio Faria, por me apresentarem a disciplina de inteligencia artificial. E ao Vinicius 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 como uma regressão em um hiperplano. Em me orientar nas métricas de avaliação dos modelos e sobre a toda a metodologia experimental.

À equipe de Data Science e Machine Learning Engineering da empresa 2RP-NET, especialista em análise de dados e 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 deste trabalho empregou métodos estatísticos para a construção do modelo de previsão. Neste trabalho, será empregado um método de aprendizado de máquina, mais especificamente, uma rede neural artificial do tipo Perceptron de Múltiplas Camadas. Na investigação anterior, verificou-se que a temperatura da região onde se localiza o campus do restaurante foi analisada por recorrência via *bootstrap* como um fator que exerce impacto sobre as vendas do restaurante em certos períodos do semestre. Assim, neste trabalho de conclusão de curso serão obtidas novas variáveis e um intervalo maior de amostragem na análise da previsão de demanda.

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

Abstract

This work has the objective of studying methods to forecast meals demand on Unifesp's university restaurant to avoid overprojection of demand as a result of food wastage, or underproduction with the consequence of teachers or students without meals. In an earlier research, carried out as a discipline work, the author of this paper used statistical methods for the construction of the prediction model. In this work, a machine learning method will be employed, more specifically, an artificial neural network of the Perceptron type of Multiple Layers. In the previous investigation, it was verified that the temperature of the region where the restaurant's campus is located was analyzed by recurrence via bootstrap as a factor that has an impact on restaurant sales in certain periods of the semester. Thus, in this work of conclusion of course will be obtained new variables and a greater interval of sampling in the analysis of the forecast of demand.

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

Listas de ilustrações

Figura 1 – Tipos de métodos quantitativos.	30
Figura 3 – Métodos de previsão de demanda.	31
Figura 5 – Dados de demanda de sazonalidade diária, em função do horário.	32
Figura 7 – Neuronio Biológico	36
Figura 9 – Neuronio Artificial.	37
Figura 11 – Neuronio Artificial - Operação AND.	37
Figura 13 – Neuronio Artificial Perceptron.	39
Figura 15 – Funções de Ativação.	39
Figura 17 – Problemas Linearmente Separáveis e Não Separáveis.	46
Figura 19 – Problemas Linearmente Não Separáveis. Múltiplas Soluções.	47
Figura 20 – Rede de perceptrons com múltiplas camadas.	47
Figura 22 – Ponto ótimo de parada da validação cruzada.	49
Figura 24 – Fases de treino da MLP-Back-Propagation.	50
Figura 26 – Rede Neural Perceptron de Múltiplas Camadas.	51
Figura 28 – Rede Neural Perceptron de Múltiplas Camadas.	51
Figura 30 – Apresentação da i-ésima observação do conjunto de treino à rede	54
Figura 31 – Cálculo do valor estimado de saída da rede.	54
Figura 32 – Aprendizado de um neurônio na camada de saída	55
Figura 33 – Reajuste dos pesos de saída	55
Figura 34 – Reajuste dos pesos de camadas ocultas	56
Figura 35 – Entradas de dia da semana em cofatores.	62
Figura 37 – Registros de consumo de refeições com intervalo temporal de 5 dias.	63
Figura 38 – Arquitetura do modelo GRU	66
Figura 40 – Funções de Ativação	67
Figura 42 – Comparison of Adam to Other Optimization Algorithms Training a Multi-layer Perceptron	68
Figura 44 – Resultado de um dos melhores modelos endógenos obtido sobre o conjunto de dados aleatoriamente ordenado sobre o tempo	71
Figura 45 – Resultado de um dos melhores modelos endógenos obtido sobre o conjunto de dados com ordenação corrigida	72
Figura 46 – Conjunto verdade de treino, dados de 2017 1o e 2o semestre, e 2018 2o semestre.	73
Figura 47 – Conjunto verdade de validação, dados de 2018 1o semestre.	73
Figura 48 – Conjunto verdade de teste, dados de 2019 1o semestre.	74
Figura 49 – Correlação entre consumo e vendas de almoço	74

Figura 50 – Gráfico Scatter entre consumo e vendas de almoço.	75
Figura 51 – Vendas de tickets normalizados com teto de 3x o desvio padrão.	76
Figura 52 – Vendas de tickets escalada.	76
Figura 53 – Correlação entre consumo e produção com 30% acima do consumo da semana anterior.	77
Figura 54 – Gráfico Scatter entre consumo e produção com margem de 30%.	77
Figura 55 – Correlação de consumo de almoço e jantar de 1 dia anterior.	78
Figura 56 – Gráfico Scatter entre consumo e jantar de 1 dia anterior.	78
Figura 57 – Gráfico violino da distribuição do consumo na segunda feira.	79
Figura 58 – Gráfico violino da distribuição do consumo na terça feira.	79
Figura 59 – Gráfico violino da distribuição do consumo na quarta feira.	80
Figura 60 – Gráfico violino da distribuição do consumo na quinta feira.	80
Figura 61 – Gráfico violino da distribuição do consumo na sexta feira.	81
Figura 62 – Correlação da distribuição do consumo com o avanço do semestre.	81
Figura 63 – Gráfico scatter da distribuição do consumo com o avanço do semestre.	82
Figura 64 – Correlação da distribuição do consumo com o avanço do mes.	82
Figura 65 – Gráfico scatter da distribuição do consumo com o avanço do mes.	83
Figura 66 – Correlação da distribuição do consumo com a temperatura.	83
Figura 67 – Gráfico escatter da distribuição do consumo com a temperatura.	83
Figura 68 – Correlação da distribuição do consumo com a umidade.	84
Figura 69 – Gráfico escatter da distribuição do consumo com a umidade.	84
Figura 70 – Correlação da distribuição do consumo com a velocidade do vento em m/s.	84
Figura 71 – Gráfico escatter da distribuição do consumo com a velocidade do vento.	85
Figura 72 – Correlação da distribuição do consumo com a pressão atmosférica.	85
Figura 73 – Gráfico escatter da distribuição do consumo com a pressão atmosférica.	85
Figura 74 – Topologia do modelo MLP1	86
Figura 75 – Gráfico de treino do modelo MLP1.	86
Figura 76 – Topologia do modelo MLP2	87
Figura 77 – Gráfico de treino do modelo MLP2.	88
Figura 78 – Topologia do modelo MLP_ENDO_1	89
Figura 79 – Treino do modelo MLP_ENDO_1	89
Figura 80 – Avaliação do modelo MLP_ENDO_1	90
Figura 81 – Topologia do modelo RNN_ENDO_1	90
Figura 82 – Treino do modelo RNN_ENDO_1	90
Figura 83 – Avaliação do modelo RNN_ENDO_1	91
Figura 84 – Topologia do modelo RNN_ENDO_2	92
Figura 85 – Treino do modelo RNN_ENDO_2	93
Figura 86 – Avaliação do modelo RNN_ENDO_2	93
Figura 87 – Comparaçao da etapa de validação entre MLP_ENDO_1 e RNN_ENDO_1	94

Figura 88 – Topologia do modelo RNN_EXO_1	94
Figura 89 – Treino do modelo RNN_EXO_1	94
Figura 90 – Avaliação do modelo RNN_EXO_1	95
Figura 91 – Grafico scatter de avaliação do modelo RNN_EXO_1	95
Figura 92 – Topologia do modelo RNN_EXO_2	96
Figura 93 – Treino do modelo RNN_EXO_2	96
Figura 94 – Avaliação do modelo RNN_EXO_2	96
Figura 95 – Grafico scatter de avaliação do modelo RNN_EXO_2	97
Figura 96 – Topologia do modelo RNN_EXO_3	98
Figura 97 – Treino do modelo RNN_EXO_3	98
Figura 98 – Avaliação do modelo RNN_EXO_3	98
Figura 99 – Grafico scatter de avaliação do modelo RNN_EXO_3	99
Figura 100 –Teste MLP_ENDO_1	100
Figura 101 –Teste RNN_ENDO_2	100
Figura 102 –Analise de anomalias preditivas do RNN_ENDO_2	100
Figura 103 –Grafico scatter de teste do RNN_ENDO_2	101
Figura 104 –Teste do modelo RNN_EXO_1	102
Figura 105 –Grafico scatter de teste do modelo RNN_EXO_1	102
Figura 106 –Teste do modelo RNN_EXO_2	103
Figura 107 –Grafico scatter de teste do modelo RNN_EXO_2	103
Figura 108 –Teste do modelo RNN_EXO_3	104
Figura 109 –Grafico scatter de teste do modelo RNN_EXO_3	104
Figura 110 –Conjunto de dados de treino, ano de 2017 completo	105
Figura 111 –Conjunto de dados de validação, ano de 2018 completo	105
Figura 112 –Conjunto de dados de teste, ano de 2019 completo	106
Figura 113 –Gráfico entre o consumo de refeições e a produção com margem de 30% para o ano de 2019 completo.	106
Figura 114 –Gráfico scatter entre o consumo de refeições e a produção com margem de 30% para o ano de 2019 completo.	106
Figura 115 –Gráfico de treino do modelo MLP_ENDO_1.	107
Figura 116 –Gráfico de validação do modelo MLP_ENDO_1.	107
Figura 117 –Gráfico de treino do modelo RNN_ENDO_1.	108
Figura 118 –Gráfico de validação do modelo RNN_ENDO_1.	108
Figura 119 –Gráfico de treino do modelo RNN_ENDO_2.	108
Figura 120 –Gráfico de validação do modelo RNN_ENDO_2.	108
Figura 121 –Gráfico de treino do modelo RNN_EXO_1.	109
Figura 122 –Gráfico de validação do modelo RNN_EXO_1.	109
Figura 123 –Gráfico de treino do modelo RNN_EXO_2.	109
Figura 124 –Gráfico de validação do modelo RNN_EXO_2.	109

Figura 125 –Gráfico de treino do modelo RNN_EXO_3.	110
Figura 126 –Gráfico de validação do modelo RNN_EXO_3.	110
Figura 127 –Gráfico de teste do modelo MLP_ENDO_1.	112
Figura 128 –Gráfico de scatter do modelo MLP_ENDO_1.	113
Figura 129 –Gráfico de teste do modelo RNN_ENDO_1.	113
Figura 130 –Gráfico de scatter do modelo RNN_ENDO_1.	113
Figura 131 –Gráfico de teste do modelo RNN_ENDO_2.	113
Figura 132 –Gráfico de scatter do modelo RNN_ENDO_2.	114
Figura 133 –Gráfico de teste do modelo RNN_EXO_1.	114
Figura 134 –Gráfico de scatter do modelo RNN_EXO_1.	114
Figura 135 –Gráfico de teste do modelo RNN_EXO_2.	114
Figura 136 –Gráfico de scatter do modelo RNN_EXO_2.	115
Figura 137 –Gráfico de teste do modelo RNN_EXO_3.	115
Figura 138 –Gráfico de scatter do modelo RNN_EXO_3.	115

Lista de tabelas

Tabela 1 – Representação de dados para regressão múltipla	33
Tabela 2 – Operação Lógica AND	38

Lista de abreviaturas e siglas

ICT	Instituto Central de 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
PMC	Perceptron de Múltiplas Camadas

Sumário

1	Introdução	25
1.1	Contextualização e Motivação	25
1.2	Definição do problema	27
1.3	Justificativas	27
1.4	Objetivos	27
1.4.1	Objetivo geral	27
1.4.2	Objetivos específicos	27
1.5	Metodologia	27
2	Fundamentação Teórica	29
2.1	Análises Estatísticas	29
2.1.1	Análise Exploratória dos Dados	29
2.1.2	Métodos de Previsão	29
2.1.3	Métodos de previsão de Demanda	30
2.1.4	Séries Temporais	31
2.1.5	Componentes Temporais	32
2.1.6	Previsão com Regressão Linear Múltipla	32
2.1.7	Algoritmo de Regressão Linear Multipla	34
2.2	Inteligencia Artificial	34
2.2.1	Neuronio Artificial	35
2.2.2	Redes Neurais Articiais	38
2.2.3	Funcionamento e treino do neurônio perceptron	40
2.2.4	Limite de um Perceptron	46
2.2.5	Rede Perceptrons Múltiplas Camadas - MLP	47
2.2.6	Perceptrons Múltiplas Camadas com Retro Propagação de Erro	49
2.2.7	Treino da MLP - BackPropagation para 1 sinal de saída	52
2.2.8	Algoritmo de treino e validação backpropagation	57
3	Trabalhos relacionados	59
3.0.1	Previsão de demanda em restaurantes universitários	59
3.0.2	Previsão de demanda em outros ambientes	60
3.0.3	Revisão bibliográfica de diversos métodos de previsão de demanda	60
4	Metodologia	61
4.1	Pré-processamento	61
4.1.1	Coleta de dados endógenos.	61

4.1.2	Dados Climáticos	61
4.1.3	Enriquecimento dos Dados	62
4.1.4	Temporização dos dados endógenos	62
4.1.5	Divisão dos Conjuntos	64
4.1.6	Tratamento das Features	65
4.2	Definição e treino dos modelos	65
4.2.1	Avaliação de modelos modernos	65
4.2.2	Topologia	66
4.2.2.1	Modelos endógenos	66
4.2.2.2	Modelos Mistas : Endógenos e Exógenos	67
4.2.3	Função de ativação	67
4.2.3.1	Para a entrada e camadas ocultas	67
4.2.3.2	Para a saída	67
4.2.4	Otimizador	68
4.3	Teste e Métricas	69
5	Resultados	71
5.1	1a Fase Experimental	71
5.1.1	Pré-Processamento	71
5.1.1.1	Divisão dos conjuntos	71
5.1.1.2	Análise das variáveis endógenas	74
5.1.1.3	Análise da sazonalidade semanal	78
5.1.1.4	Análise das variáveis exógenas	81
5.1.2	Definição e treino dos modelos	86
5.1.2.1	Ajuste empírico de topologia com os modelos perceptron	86
5.1.2.2	Modelos Endógenos	88
5.1.2.3	Modelos mistos endógenos e exógenos	94
5.1.3	Teste e métricas	100
5.1.3.1	Endógenos : GRU x MLP	100
5.1.3.2	Melhor modelo Endógeno	100
5.1.3.3	Modelos mistos	101
5.1.3.4	Melhor modelo da 1a fase	104
5.2	2a Fase Experimental	105
5.2.1	Pré-Processamento	105
5.2.1.1	Divisão dos conjuntos	105
5.2.1.2	Análise da técnica do restaurante	106
5.2.2	Treino e Validação dos modelos	107
5.2.3	Teste e métricas	112
5.2.3.1	Teste do melhor modelo da 2a fase no domínio da primeira fase	115
5.3	Tabela de resultados	116

5.3.1	Melhores Resultados	116
5.3.2	Tabela completa de todos os modelos	116
6	Conclusão	117
6.1	A importância da metodologia de divisão do conjunto de dados em séries temporais	117
6.2	Sobre o método de produção de refeições com margem de erro e análise da semana anterior	117
6.3	Sobre a sazonalidade semanal	117
6.4	Sobre o ajuste empírico da topologia dos modelos	117
6.5	Sobre os resultados preliminares da etapa de validação	118
6.6	Sobre os erros anômalos de predição	118
6.7	Sobre o modelo com melhor resultado	118
6.8	Conclusões gerais	119
Referências		121
7	Anexos	123
7.1	1a Fase Experimental	123
7.1.1	Pré-Processamento e treino dos modelos	123
7.1.2	Importação e aplicação de métricas dos modelos	123
7.2	2a Fase Experimental	123
7.2.1	Pré-Processamento e treino dos modelos	123
7.2.2	Importação e aplicação de métricas dos modelos	123

1 Introdução

1.1 Contextualização e Motivação

A previsão da demanda é o fator principal da eficiência de qualquer modelo de processamento do tipo entrada-saída, onde a sua saída deve atender uma demanda não determinística. É necessário prever a demanda para projetar e aperfeiçoar o processamento e a entrada deste modelo. Portanto, a previsão de demanda é um ponto de extrema importância para qualquer empresa e, sendo adequada, permite o ajuste de todo o mecanismo de operações da empresa 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 .Portanto, a escolha dos parâmetros não será influenciada por muitos fatores externos como concorrência, acessibilidade do ponto, entre outros.

Outro ponto importante é a obtenção dos valores de venda; não foram escolhidas as vendas diretas do ponto de venda de tickets de refeição, e sim os dados de coleta da entrada do restaurante, que demonstram a real movimentação de público no restaurante em determinado dia.

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 a ser analisado neste trabalho é o comportamento dos consumidores de um restaurante universitário (R.U), o qual precisa projetar sua compra de insumos e alocação de recursos na entrada de seu modelo de negócio, e projetar 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 humano, de que forma estes elementos exercem tal influência e em qual intensidade ela ocorre.

E para explorar tais elementos, é necessário explorar como este consumo ocorreu his-

toricamente a fim de se encontrar padrões de comportamento, que no cenário deste trabalho acontece com informações distribuídas através do tempo.

Neste capítulo o principal objetivo é definir todos os conceitos que serão 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 a definir 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 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 cruzam outras informações 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

1.4.1 Objetivo geral

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

1.4.2 Objetivos específicos

- Estruturar o conjunto de dados, e realizar análises exploratórias sobre os dados;
- Construir realizar análise e comparação da qualidade dos modelos de redes neurais;

1.5 Metodologia

Análises dos dados

Por se tratar de um trabalho de previsão de demanda, este trabalho irá realizar a coleta e tratamento dos dados de consumo do restaurante universitário da Unifesp, a coleta de dados

climáticos que influenciam em tal demanda, calcular os 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, será realizada uma análise exploratória fundamentada por ([JUNIOR, 2007](#)), Capítulo 2.4, para previsões de demanda em geral. O modelo de dados de venda estruturados receberá o acréscimo de dados de variáveis climáticas, como possíveis fatores de influencia no consumo, conforme ocorre no trabalho de previsão de demanda de energia elétrica fundamentado por ([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 usará a análise estatística de regressão linear múltipla fundamentada por ([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 plots dos gráficos, uma regressão linear simples entre 2 valores supervisionados, também são retornadas outras métricas como medidas de correlação, p-value, R^2 , entre outras métricas que serão documentadas para trabalhos futuros.

Estudo de 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 método de Perceptron de Múltiplas Camadas (PMC) 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.

Após a modificação da topologia, os dados serão estruturados em 2 fases experimentais, sendo a primeira com os experimentos focados para validação e teste no primeiro semestre de 2019, e na segunda fase os experimentos contemplarão o ano de 2018 inteiro para validação e o ano de 2019 inteiro para teste.

Análise dos resultados De acordo com ([MARIANO, 2014](#)), a comparação e a discussão dos resultados deve ser feita com base em medidas de avaliação de erro absoluto médio (EAM), erro quadrado médio (EQM) e raiz do erro quadrado médio (REQM).

2 Fundamentação Teórica

2.1 Análises Estatísticas

2.1.1 Análise Exploratória dos Dados

(JUNIOR, 2007) Cita 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 se enquadra 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.

2.1.2 Métodos de Previsão

(JUNIOR, 2007) Realiza 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.

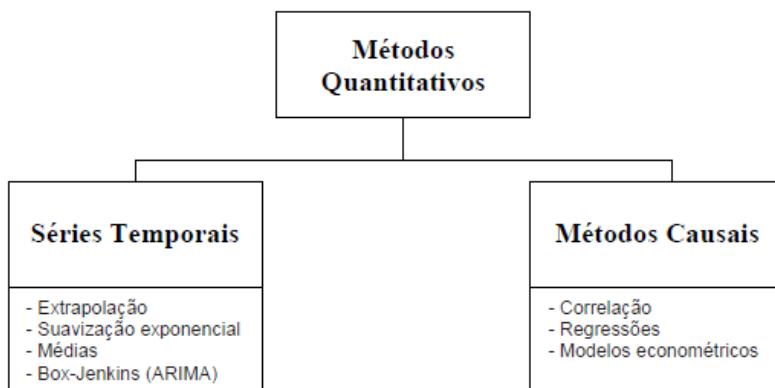


Figura 1: Tipos de métodos quantitativos.

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

2.1.3 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.

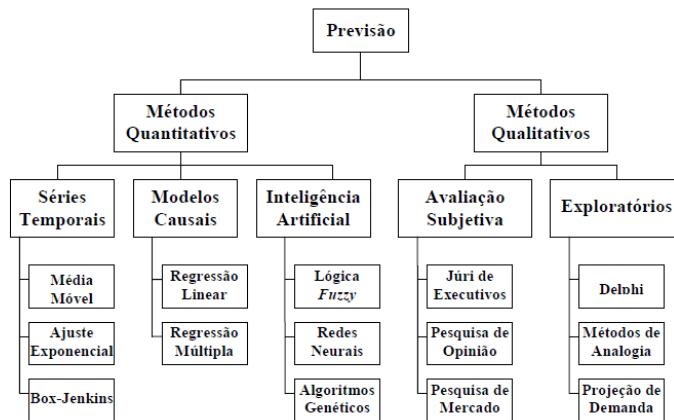


Figura 3: Métodos de previsão de demanda.

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

2.1.4 Séries Temporais

De acordo com ([MORETTIN, 1987](#)) uma série temporal é um conjunto de observações ordenadas em função do tempo, comumente iguais, apresentando uma dependência serial entre 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, sendo ideal as análises de regressão.

([ALMEIDA, 2013](#)) também cita 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 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.

2.1.5 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, conforme usam métodos de regressão, que pode ser observados nos trabalhos de previsão de consumo de energia elétrica realizados por ([ALMEIDA, 2013](#)), ([RUAS, 2012](#)), ([SILVA, 2010](#)) e de previsão de demanda de produtos cosméticos em ([JUNIOR, 2007](#))

O cenário deste estudo analisa a frequência de alunos dentro do ICT - UNIFESP e consequentemente dentro do R.U tem sazonalidade anual a partir do momento em que a grade de disciplinas tender a ter sazonalidade anual.

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.

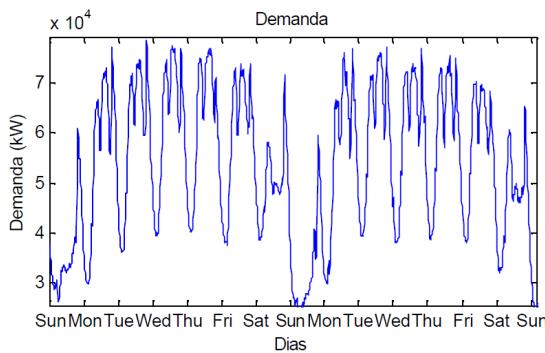


Figura 5: Dados de demanda de sazonalidade diária, em função do horário.

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

2.1.6 Previsão com Regressão Linear Múltipla

A teoria da regressão se iniciou no século XIX com Galton, que analisou a altura dos pais e dos filhos (X_i e Y_i) e buscou a influencia da altura do pai no filho, notando que a altura do filho tendia à media da altura do pai, e chamou esta técnica de regressão pois existe uma tendência dos dados regredirem à média.

([CLARICE G.B. DEMÉTRIO; ZOCCHI, 2011](#)) Informa que em estudos de regressão aplica-se o método relacionar uma variável aleatória Y com uma variável X , e no caso da regressão múltipla com múltiplas variáveis X , representando causas diferentes que se combinam para a ocorrência de um valor Y .

No exemplo abaixo de regressão linear simples, podemos analisar o total de venda Y de um dia X , sendo relacionado nas variações $(\beta_0, \beta_1, \dots, \beta_k)$ de datas de X .

Ou seja, $X, Y \rightarrow Y \simeq f(X)$

Na regressão simples Y depende apenas de uma variável X , isto é, $Y \simeq f(\beta_0, \beta_1, \dots, \beta_k) + E_y$. Sendo E_y um erro aplicado à esta função para se chegar no valor Y . Logo a função $f(\cdot)$ pode ser linear nos parâmetros $(\beta_0, \beta_1, \dots, \beta_k)$, se a derivada da função em relação às derivadas dos parâmetros, corresponderem à $h(X)$, i , para i variando de 0 à k . Sendo $h(X)$, dependente apenas de X . $\frac{\partial f}{\partial \beta_i} = h(X), i = 0, 1, \dots, k$

Na regressão múltipla, onde temos múltiplas relações causais X_1 à X_k , em ocorrências β_0 à β_k , temos a função: $Y \simeq f(X_1, X_2, \dots, X_k, \beta_0, \beta_1, \dots, \beta_k) + E_y$. Também sendo linear nos parâmetros se a derivada da função em relação às múltiplas derivadas dos parâmetros, corresponderem à $h(X_1, X_2, \dots, X_k)$, tendo $h(\cdot)$, dependente apenas de X_1, X_2, \dots, X_k . Em caso contrário, $f(\cdot)$ é uma função não linear dos parâmetros.

Logo, $Y = f(X_1, X_2, \dots, X_k, \beta_0, \beta_1, \dots, \beta_k) + E_y$ é linear nos parâmetros se: $\frac{\partial f}{\partial \beta_i} = h(X_1, X_2, \dots, X_k)$

Representação dos dados Na tabela abaixo representa-se a estrutura de dados para regressão linear múltipla, com n observações da variável resposta Y e p variáveis explicativas. Assim, y_i é o valor da i -ésima resposta e x_{ij} na é o valor da variável x_j do conjunto de j variáveis correspondente à resposta y_i . Para que seja possível aplicar o método dos quadrados mínimos (explicado adiante) de resolução do sistema de variáveis X em relação à Y , é importante que se obtenha p observações Y , onde $n \geq (p + 1)$.

Tabela 1: Representação de dados para regressão múltipla.

y	x_1	x_2	...	x_p
y_1	x_{11}	x_{12}	...	x_{1p}
y_2	x_{21}	x_{22}	...	x_{2p}
\vdots	\vdots	\vdots	...	\vdots
y_n	x_{n1}	x_{2n}	...	x_{np}

Satisfazendo o modelo final $Y_i = \beta_0 + \beta_1 * x_{i1} + \beta_2 * x_{i2} + \beta_p * x_{ip} + \epsilon_i$
 $i = 1, \dots, n$

Método dos Mínimos Quadrados O método dos mínimos quadrados, é uma técnica utilizada popularmente em estimações econômicas, que busca encontrar o melhor ajuste de um conjunto de dados conforme tabela e modelo acima, para minimizar a soma dos erros quadráticos ϵ . A técnica foi publicada por Adrien-Marie Legendre em 1809, em *Nouvelles méthodes pour la détermination des orbites des comètes*, realizando a hipótese de que a distribuição do ϵ seja aleatória e normal, embora tal dedução já existisse indiretamente nos teoremas de Gauss-Markov.

A minimização da soma dos erros quadráticos busca então minimizar a seguinte equação L :

$$L = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 * x_{i1} - \dots - \beta_p * x_{ip})^2$$

A técnica segue o princípio de ser linear nos parâmetros, já demonstrado acima conforme ([CLARICE G.B. DEMÉTRIO; ZOCCHI, 2011](#)). Logo deriva-se L em função dos parâmetros β .

$$\begin{aligned}\frac{\partial L}{\partial \beta_0} &= -2 \sum_{i=1}^n [Y_i - \beta_0 - \beta_1 * x_{i1} - \dots - \beta_p * x_{ip}] \\ \frac{\partial L}{\partial \beta_j} &= -2 \sum_{i=1}^n [Y_i - \beta_0 - \beta_1 * x_{i1} - \dots - \beta_p * x_{ip}] x_{ji}\end{aligned}$$

Aplicando-se álgebra, igualando as derivadas parciais a zero e organizando o sistema em forma matricial, o modelo final pode ser dado por: $Y = X\beta + \varepsilon$.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \text{ e } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

Logo,

$$\frac{\partial L}{\partial \beta} = -2X'Y + 2X'X\beta.$$

Aplicando-se álgebra matricial, substituindo o vetor de coeficientes β por $\hat{\beta}$ que é o objetivo da relação busca entre as variáveis X e Y : $\hat{\beta} = (X'X)^{-1}X'Y$

Portanto o modelo final é dado por $Y_{estimado} = X\hat{\beta}$.

2.1.7 Algoritmo de Regressão Linear Multipla

1. Estruturação dos dados em forma matricial, $Y_{1,1\dots n}$ e $X_{1\dots n,1\dots j}$, sendo $n \geq (j+1)$
2. Obtenção do vetor de coeficientes $\hat{\beta}$ das variáveis X , através da operação $\hat{\beta} = (X'X)^{-1}X'Y$
3. Construção do modelo final por $Y_{estimado} = X\hat{\beta}$.

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 humano. 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 computacio-

nalmente seu elemento biológico principal de aprendizagem: O neurônio biológico.

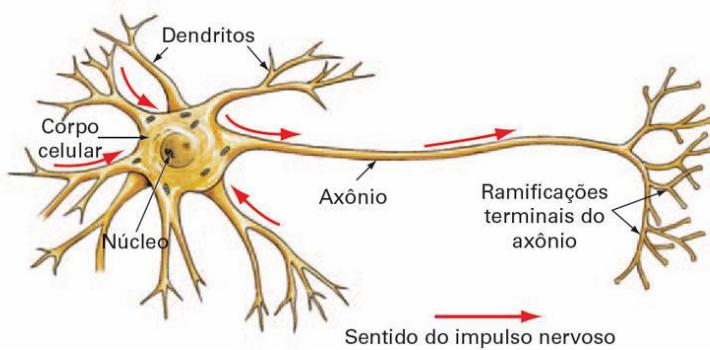


Figura 7: Neurônio Biológico

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

Um neurônio biológico (Figura 04) processa 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 permitem a transmissão de sinal elétrico pelo axônio até o dendrito de um neurônio vizinho, formando um ciclo de aprendizado em uma rede neural biológica.

Neurônio Artificial 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:

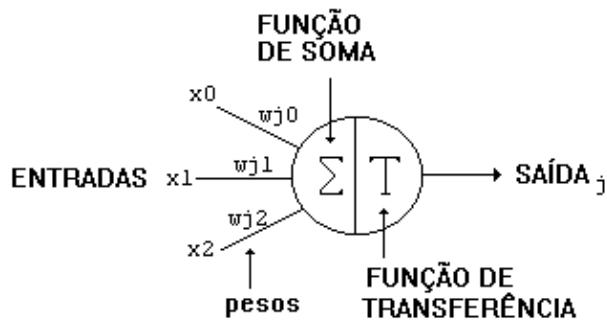


Figura 9: Neurônio Artificial.

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

Assim como no neurônio biológico, este modelo reage à um vetor de entradas X_0, X_1, \dots, X_n onde tem suas sinapses representadas por pesos numéricos, existe 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.

Neste modelo simplista o neurônio consegue realizar operações lógicas. Na figura 06 os pesos w_1 e w_2 e o limiar T estão ajustados para responder à operação AND que resumidamente à uma operação lógica que somente produz um valor positivo 1 na resposta de saída se ambos os valores de entrada forem 1.

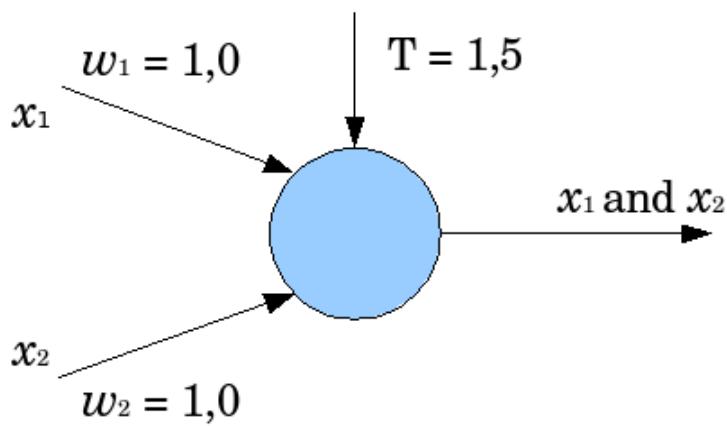


Figura 11: Neurônio Artificial - Operação AND.

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

O neurônio irá comparar se a soma das entradas X multiplicadas pelos seus devidos pesos W , será menor que o limiar T . Ele somente irá produzir uma resposta positiva (1) se essa

soma for maior. É fácil verificar então na imagem, que o resultado dessa soma só será positivo se ambas as entradas forem 1, onde a soma será $2 > 1, 5$.

Tabela 2: Operação Lógica AND

X_1	X_2	Saída
0	0	0
0	1	0
1	0	0
1	1	1

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, este ajuste não se torna trivial.

Em 1958, o cientista da computação Frank Rosenblatt desenvolveu o modelo de neurônio artificial 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 uma 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.

O perceptron ([MARIANO, 2014](#)) possui apenas uma camada de entrada e saída, a saída utiliza como função de ativação a função degrau u 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: $y = \delta(\sum_{i=1}^n X_i W_i + b)$

- X_i - sinais de entrada do neurônio;

- W_i - pesos sinápticos do neurônio;

- b - bias ou limiar de ativação;

- $\delta(\cdot)$ - função de ativação;

- y - sinal de saída do neurônio.

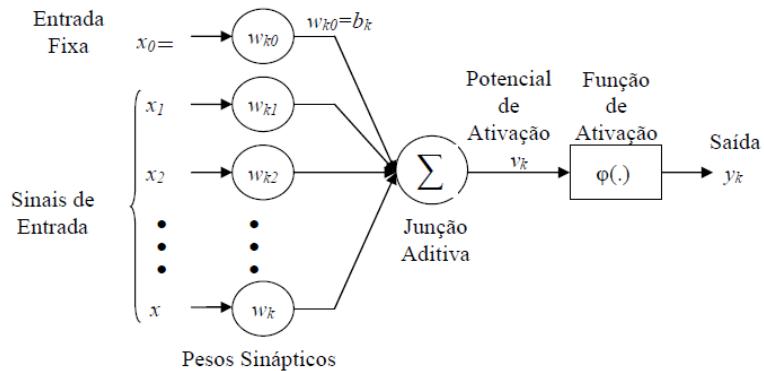
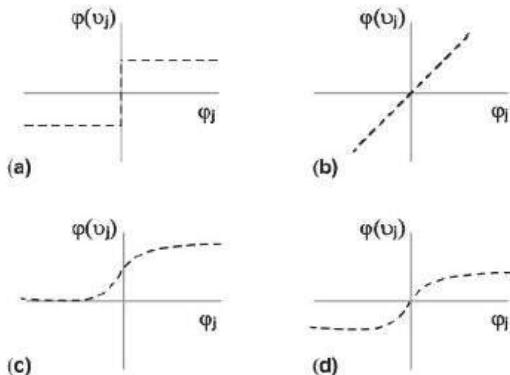


Figura 13: Neurônio Artificial Perceptron.

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. As funções de ativação podem ser observadas na figura 08, sendo a de maior popularidade de utilização, a sigmoide.



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

Figura 15: Funções de Ativação.

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

As funções de ativação dão a capacidade do perceptron quando conectado em rede (explicado no próximo capítulo) 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 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 Funcionamento e treino do neurônio perceptron

Em ([ALMEIDA, 2013](#)) verifica-se de forma procedural, o processo do treino supervisionado do perceptron, com m observações $Y = (Y_1, Y_{i+1}, \dots, Y_m)$ com n entradas X relacionadas à cada Y_m , ocorre da seguinte maneira: Matematicamente o neurônio é representado por um vetor global de n pesos sinápticos (W_0, W_1, \dots, W_n) , e um valor de bias b . O processo do algoritmo abaixo ajustará esses pesos e bias.

1. Coleta do valor supervisionado inicial Y_0 , que é a primeira observação histórica do conjunto de dados.
2. Coleta do vetor de variáveis $(X_{00}, X_{01}, \dots, X_{0n})$ relacionados à saída supervisionada Y_i .
3. Inicialização dos pesos sinápticos (W_0, W_1, \dots, W_n) com valores aleatoriamente pequenos.
4. Definição de um valor real entre 0 e 1 para a taxa de aprendizado α .

5. Definição do valor inicial de bias b .
6. Definição da função de ativação δ
7. Definição do limite de épocas de treino l_e , se não houver, definir l_e para -1.
8. LAÇO: Percorrer e atribuir em (Y_d) de Y_0 até Y_m , inicializando o contador de épocas $c_e = 0$, enquanto $CP(e)$ for verdadeiro E/OU (l_e) for menor que c_e :
9. Nova coleta do vetor de variáveis $(X_{d0}, X_{d1}, \dots, X_{dn})$ relacionados à saída supervisionada Y_d .
10. Incrementar 1 em c_e .
11. Cálculo do sinal $u = \sum_{i=1}^n X_{di} * W_i + b$
12. Aplicação de valor de U na função de ativação δ obtendo $Y_u = \delta(u)$
13. Obtenção do erro $e = Y_d - Y_u$
14. Aplicação do critério de parada CP em relação ao erro e ; $CP(e)$
15. SE $CP(e)$ for APROVADO, o algoritmo continua, SE REPROVADO, a aprendizagem é feita ajustando $b = b + (\alpha * (e))$, e ajustando o vetor $W_i = W_i + (\alpha * e * X_{di})$.

Funcionamento do perceptron para logica AND .

Este exemplo aborda a aplicação do problema do neurônio de Warren McCulloch e Walter Pitts, o aprendizado para a tabela AND, que busca fornecer uma resposta verdadeira apenas se a primeira E todas as demais entradas forem verdadeiras.

A utilização do perceptron com o modelo de aprendizado supervisionado, busca então conseguir de forma autônoma os valores de pesos W_i e desta forma aprender o funcionamento da logica E, sem que seja necessário a tarefa humana de ajustar o mesmo.

linha	X_0	X_1	Y
0	0	0	0
1	0	1	0
2	1	0	0
3	1	1	1

Para este aprendizado vamos definir o critério de parada de 0 para o erro. Definir o valor 0 para o *bias*. 1 para α e função $\delta(u) = degrau(1, se u \geq 0, 0 se u < 0)$. O número de épocas terá o mesmo número de valores supervisionado, então o treino executará 1 vez para cada observação.

$$\text{linha } d = 0$$

1. Coleta do valor supervisionado $Y_d = Y_0 = 0$.
2. Coleta do vetor de variáveis ($X_{00} = 0, X_{01} = 0$).
3. Inicialização dos todos os pesos com valores aleatoriamente pequenos: $W_0 = 3, W_1 = 3$
4. Definição da taxa de aprendizado $\alpha = 1$.
5. Definição do valor de bias $b = 3$.
6. Definição da função de ativação $\delta = degrau$
7. Cálculo do sinal $u = x_{00} * w_0 + x_{01} * w_1 + b = (0 * 3 + 0 * 3 + 0) = 0$
8. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(0) = 1$
9. Obtenção do erro $e = Y_0 - Y_u = 0 - 1 = -1$
10. Aplicação do critério de parada CP em relação ao erro e REPROVADO, pois o erro esperado é 0.
11. Como CP foi reprovado, a aprendizagem é realizada $b = b + (\alpha * e)$ e $W_i = W_i + (\alpha * e) * X_{0i}$

Aprendizagem:

1. $b = b + (\alpha * e) = 0 + 1 * (-1) = -1$
2. $W_0 = W_0 + (\alpha * e) * X_{00} = 3 + 1 * (-1) * 0 = 3$
3. $W_1 = W_1 + (\alpha * e) * X_{01} = 3 + 1 * (-1) * 0 = 3$

Reinicio dos pesos reajustados pela aprendizagem e apresentação da próxima linha ao neurônio: linha $d = 1$.

linha	X_0	X_1	Y
1	0	1	0

1. Coleta do valor supervisionado $Y_d = Y_1 = 0$.
2. Cálculo do sinal $u = X_{10} * w_0 + X_{10} * W_1 + b = (0 * 3 + 1 * 3 - 1) = 0$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(0) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 1 = -1$

5. Aplicação do critério de parada CP em relação ao erro e reprovado, pois o erro esperado é 0.
6. Como CP foi reprovado, a aprendizagem é realizada $b = b + (\alpha * e)$ e $W_i = W_i + (\alpha * e) * X_{1i}$

Aprendizagem:

1. $b = b + (\alpha * e) = -1 + 1 * (-1) = -2$
2. $W_0 = W_0 + (\alpha * e) * X_{10} = 3 + 1 * (-1) * 0 = 3$
3. $W_1 = W_1 + (\alpha * e) * X_{11} = 3 + 1 * (-1) * 1 = 2$

Reinicio dos pesos reajustados pela aprendizagem e apresentação da próxima linha ao neurônio: linha $d = 2$

linha	X_0	X_1	Y
2	1	0	0

1. Coleta do valor supervisionado $Y_d = Y_2 = 0$.
2. Cálculo do sinal $u = X_{20} * w_0 + X_{20} * W_1 + b = (1 * 3 + 0 * 2 - 2) = 1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(1) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 1 = -1$
5. Aplicação do critério de parada CP em relação ao erro e reprovado, pois o erro esperado é 0.
6. Como CP foi reprovado, a aprendizagem é realizada $b = b + (\alpha * e)$ e $W_i = W_i + (\alpha * e) * X_{2i}$

Aprendizagem:

1. $b = b + (\alpha * e) = -2 + 1 * (-1) = -3$
2. $W_0 = W_0 + (\alpha * e) * X_{20} = 3 + 1 * (-1) * 1 = 2$
3. $W_1 = W_1 + (\alpha * e) * X_{21} = 2 + 1 * (-1) * 0 = 2$

Reinicio dos pesos reajustados pela aprendizagem e apresentação da próxima linha ao neurônio: linha $d = 3$

linha	X_0	X_1	Y
3	1	1	1

1. Coleta do valor supervisionado $Y_d = Y_3 = 1$.
2. Cálculo do sinal $u = X_{30} * w_0 + X_{31} * W_1 + b = (1 * 2 + 1 * 2 - 3) = 1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_d , $Y_u = \delta(1) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 1 - 1 = 0$
5. Aplicação do critério de parada CP em relação ao erro e_n APROVADO, então conclui-se que os pesos sinápticos estão corretos.

Pesos finais e bias aprovados pelo processo de treino:

1. $b = -3$
2. $W_1 = 2$
3. $W_2 = 2$

Fim do processo de treino.

Validação do treino Após todo o processo de treino e apresentação dos conjuntos, é necessário realizar um novo ciclo no perceptron, para ver se o neurônio responde corretamente ao modelo de dados inicial.

Pesos finais e bias aprovados pelo processo de treino:

1. $b = -3$
2. $W_1 = 2$
3. $W_2 = 2$

linha $d = 0$

linha	X_0	X_1	Y
0	0	0	0

1. Coleta do valor supervisionado $Y_d = Y_0 = 0$.
2. Cálculo do sinal $u = X_{00} * w_0 + X_{01} * W_1 + b = (0 * 2 + 0 * 2 - 3) = -3$

3. Aplicação de valor de U na função de ativação δ obtendo Y_u , $Y_u = \delta(-3) = 0$
4. Obtenção do erro $e = Y_i - Y_u = 0 - 0 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

linha $d = 1$

linha d	X_0	X_1	Y
1	0	1	0

1. Coleta do valor supervisionado $Y_d = Y_1 = 0$.
2. Cálculo do sinal $u = X_{10} * w_0 + X_{11} * W_1 + b = (0 * 2 + 1 * 2 - 3) = -1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(-1) = 0$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 0 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

linha $d = 2$

linha d	X_0	X_1	Y
2	1	0	0

1. Coleta do valor supervisionado $Y_d = Y_2 = 0$.
2. Cálculo do sinal $u = X_{20} * w_0 + X_{21} * W_1 + b = (1 * 2 + 0 * 2 - 3) = -1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(-1) = 0$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 0 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

linha $d = 3$

linha d	X_0	X_1	Y
3	1	1	1

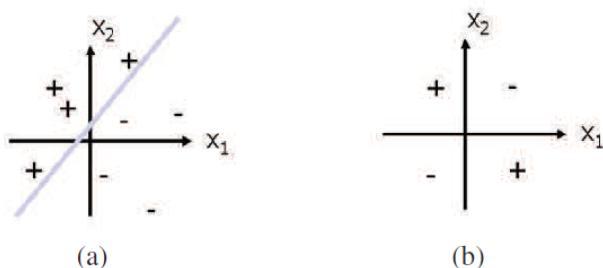
1. Coleta do valor supervisionado $Y_d = Y_3 = 0$.
2. Cálculo do sinal $u = X_{30} * w_0 + X_{31} * W_1 + b = (1 * 2 + 1 * 2 - 3) = 1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(1) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 1 - 1 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

Como o neurônio não teve nenhum critério de parada REPROVADO em todas as interações de todas as linhas, podemos dizer que o mesmo está devidamente treinado e validado para a solução do problema.

2.2.4 Limite de um Perceptron

É possível reparar que o processo de aprendizagem de um perceptron, se forma por uma combinação linear que é o valor u dos pesos e entradas x_i . Essa combinação pode ser aprovada ou reprovadas no processo repetitivo de treino controlado pela função δ , e pela aprovação do erro medido entre o valor da combinação linear e o valor esperado.

Então em 1969, Marvin Mjinsky 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 10, abaixo, 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 17: 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, tangindo 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.

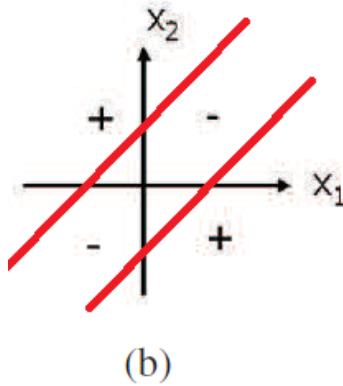


Figura 19: Problemas Linearmente Não Separáveis. Múltiplas Soluções.

2.2.5 Rede Perceptrons Múltiplas Camadas - MLP

A solução base para se combinar 2 ou mais 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.

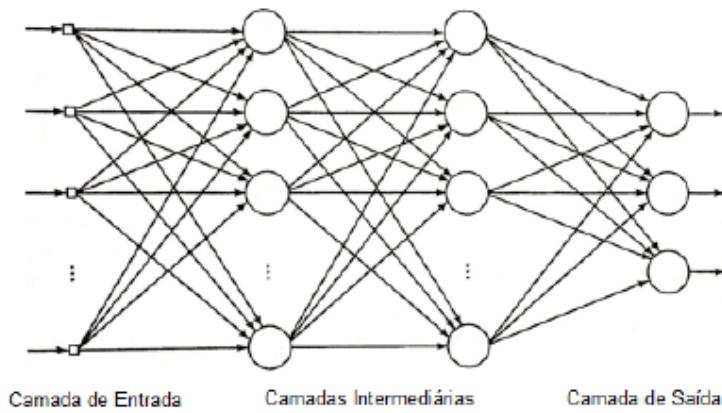


Figura 20: Rede de perceptrons com múltiplas camadas.

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 limiar de 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.

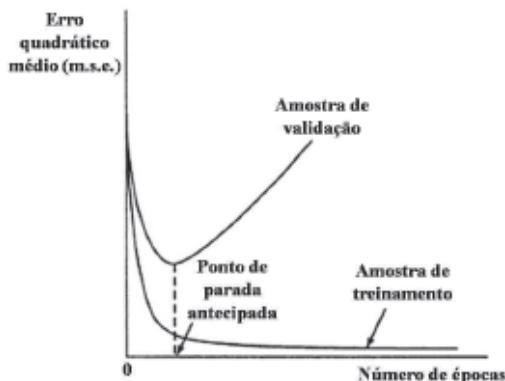


Figura 22: Ponto ótimo de parada da validação cruzada.

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.

2.2.6 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.

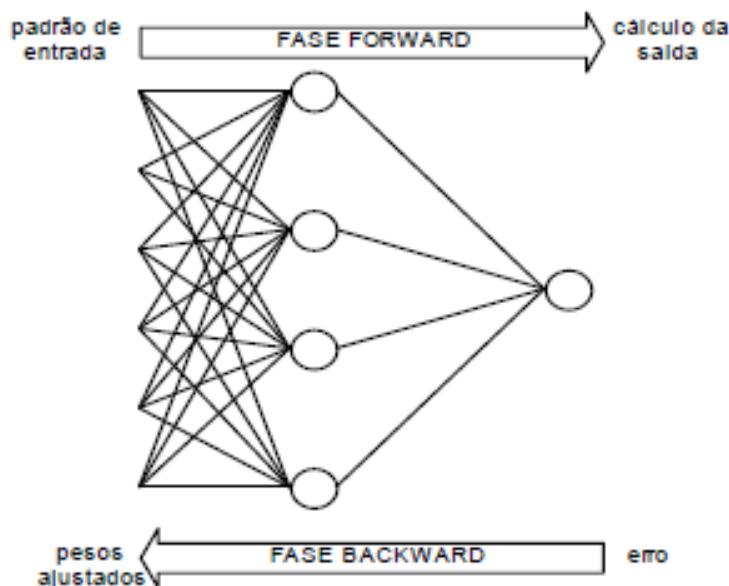


Figura 24: Fases de treino da MLP-Back-Propagation.

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

MLP Backpropagation em R.U em trabalhos relacionados Em ([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 ultimas observações anteriores ao dia a se analisar, e como variáveis qualitativas o dia da semana variando de segunda a sexta, em valores binários.

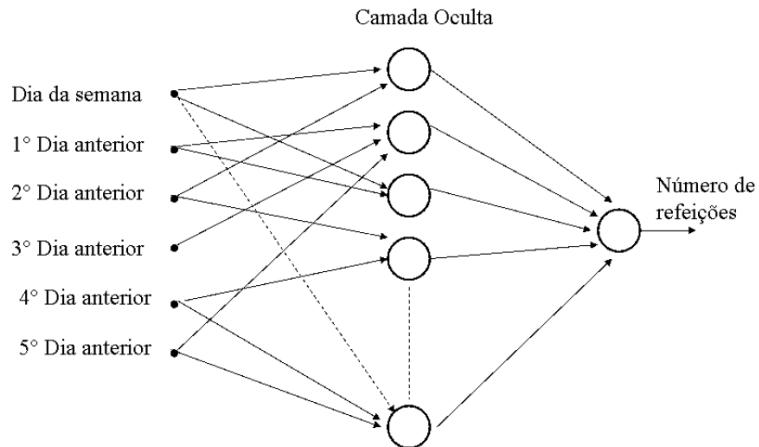


Figura 26: Rede Neural Perceptron de Múltiplas Camadas.

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

Em ([FERNANDO., 2011](#)) o trabalho realizado 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.

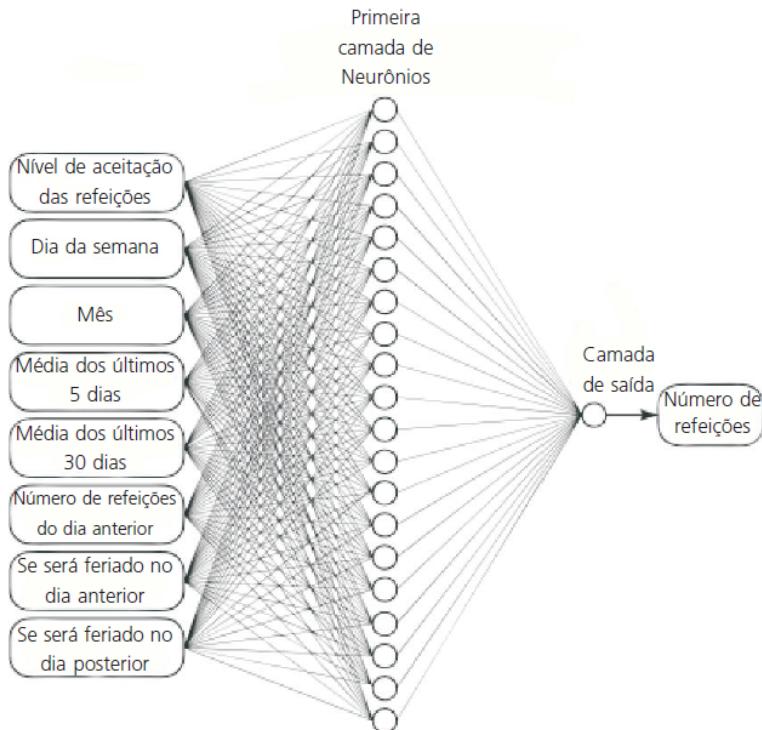


Figura 28: Rede Neural Perceptron de Múltiplas Camadas.

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 ([RAGA A. DE P.; CARVALHO, 2000](#)) cita 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.

([MARIANO, 2014](#)) cita que 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.

2.2.7 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, porém o somatório no cálculo do erro pode envolver mais neurônios de saída em outras aplicações.

Um vetor i do conjunto de dados, de tamanho n de variáveis de entrada $X_{i0}(j), X_{i1}(j), \dots, X_{in}(j)$, que são os sinais de entrada, são apresentados à rede, relacionados à um valor supervisionado Y_i .

Em resumo 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 e deste sinal de saída em s com o valor Y_i , 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

Neste capítulo, será denominado a função de ativação do neurônio por $F(\cdot)$ para uma

melhor coerência, devido ao símbolo δ utilizado para esta função nas seção anterior sobre o neurônio artificial ser amplamente utilizado para outros fins na regra de reajuste dos pesos sinápticos da rede MLP. Esta função na MLP - Backpropagation deve ser uma função não-linear diferenciável em todos os pontos, popularmente é usada a função sigmoide.

O erro quadrático do treino de uma rede MLP com Backpropagation, com 1 sinal de saída, no neurônio de saída s , em uma i -ésima iteração de treino em uma determinada época k será:

$$e(i) = \frac{1}{2} * ((Y_i - Y_{i\mu}(s))^2)$$

onde $Y_{i\mu}(s)$ é o sinal de saída deste neurônio de saída, e Y_i é o valor de saída da observação do conjunto de dados.

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 deste ponto ótimo exibida anteriormente, 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 $\mu = (\sum_{i=1}^n (X_{i\mu}(j) * W_{i\mu}(j)) + b)$ 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 $X_{i\mu}(j)$ são $W_{i\mu}(j)$. A saída estimada pela rede será $Y_{i\mu}(s)$.

Se a iteração i da época k for a primeira, todos os pesos $W_{i\mu}(j)$ em todos os neurônios, são inicializados com valores aleatórios pequenos.

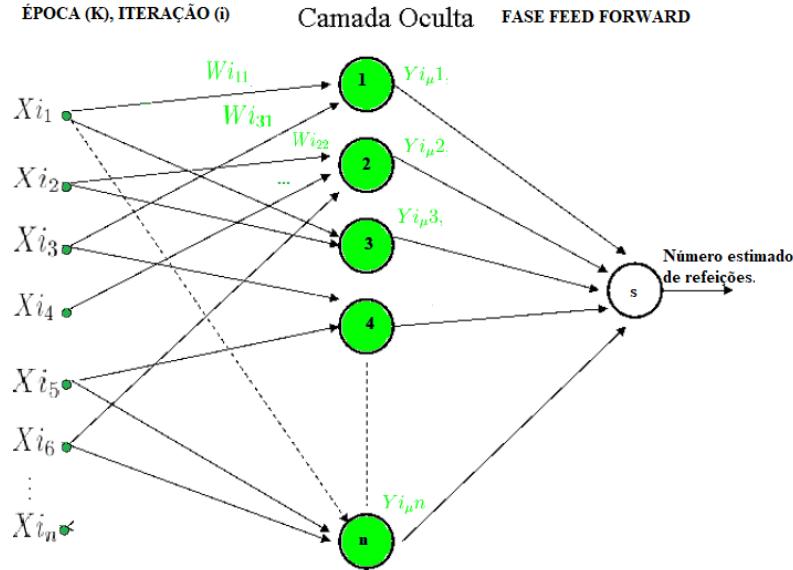


Figura 30: Apresentação da i-ésima observação do conjunto de treino à rede

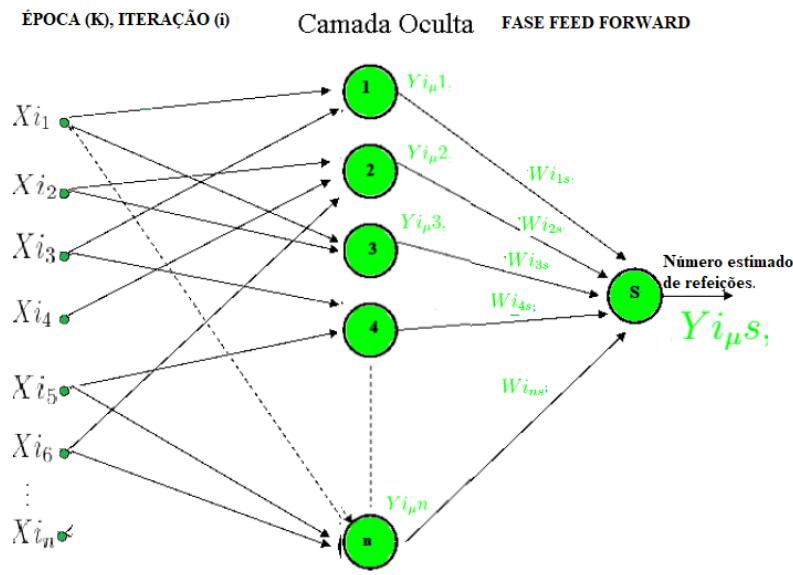


Figura 31: Cálculo do valor estimado de saída da rede.

O valor de $Y_{i_\mu}(j)$. dos neurônios da primeira camada oculta, são obtidos de forma que $X_{i_n}(j)$ são os sinais de entrada das variáveis de 1 à n .

O valor de $Y_{i_\mu}(j)$. dos neurônios das próximas camadas, inclusive a de saída, utiliza o sinal de saída Y_{i_μ} dos neurônios da camada anterior conectados à j como um sinal de entrada X_{i_n} .

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

$$\delta_s = (Y_i - Y_{i_\mu}(s)) * F'(\mu(s)).$$

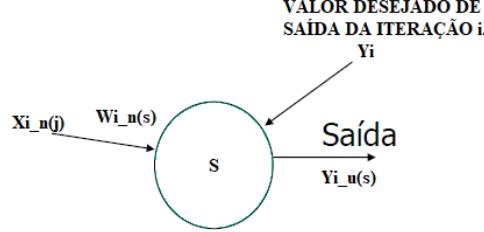


Figura 32: Aprendizado de um neurônio na camada de saída

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 $\Delta Wi_n(s) = \eta * \delta_s * Y_{i_\mu}(j)$

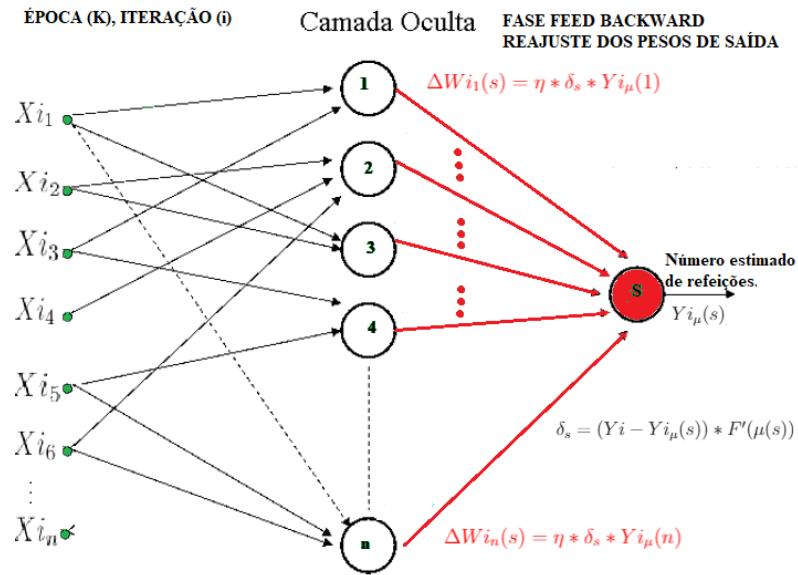


Figura 33: Reajuste dos pesos de saída

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:

$$\delta_n = F'(\mu(n)) * (\delta_s * W i_n s)$$

Por fim os pesos das n entradas $W i_n(n)$ 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.

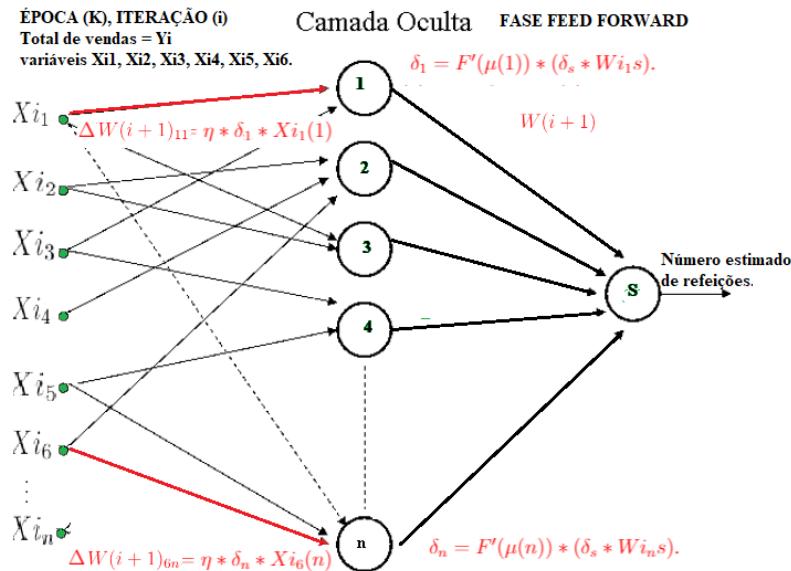


Figura 34: Reajuste dos pesos de camadas ocultas

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.

2.2.8 Algoritmo de treino e validação backpropagation

Parâmetros de entrada:

- 'rna'[][][], Grafo de nós do tipo neurônio, da rede neural, onde rna[m][n] corresponde ao nó "n" da camada "m".
- valor real η , Taxa de aprendizado .
- Yt[][] e Yv[][][], Matrizes de valores reais Yt do conjunto de treino e Yv do conjunto de validação, onde Y[0][0] corresponde ao primeiro valor supervisionado, e Y[0][m] corresponde ao último parâmetro X relacionado ao valor supervisionado.
- L, Limite do contador de épocas, a cada novo erro mínimo o contador m.s.e é zerado, após o limite L do aumento de erros do m.s.e o treino é interrompido.

O principal critério de parada do treino, será o contador L. Durante o treino os valores de m.s.e sofrerão uma curva com similaridade parabólica, concavidade apontada para cima, conforme Figura 12 da seção 2.3.5. Busca-se então obter o valor mínimo desta curva. O treino trabalhará com 2 cópias do grafo rna[][]. Sendo rnaAtual[][] atual, e rnaMinimo[][]. A cada novo valor mínimo encontrado, o rnaMinimo[][] receberá o rnaAtual[], ao fim do treino quando o limite L de incremento de erros for superado, o treino será interrompido retornado o grafo rnaMinimo[][],

1. Instancie rnaAtual[][] e rnaMinimo[][].
2. Instancie um contador inteiro de erros ce=0.
3. Copie rna[][] para rnaAtual[][] e rnaMinimo[][].
4. Inicialize uma variável inteira, contador de épocas, ck = 0.
5. Inicialize uma variável real, errok.
6. Insira uma cópia de rna[][] para rnaAtual[][]
7. LAÇO: conte i1, Enquanto ce<L faça:
 8. -LAÇO: conte i2, Percorra it[] em Yt[i2].
 9. —Se i2=0, inicialize os pesos de K[atual] com valores aleatoriamente pequenos.
 10. —Execute a fase feedforward apresentando it[] em rnaAtual[][] e guarde o erro em errok.
 11. —Execute a fase feedbackward reajustando os pesos de rnaAtual[][] com errok.

12. –Incialize um vetor temporario ev[] de tamanho Yv[].tamanho, a quantidade de observações de treino.
13. –LAÇO: Conte i3, Percorra iv[] em Yv[i3].
14. ——Execute a fase feedforward apresentando iv[] em rnaAtual[][] e insira o erro quadrático em ev[i].
15. –Obtenha a média dos valores quadráticos do vetor ev[] e insira em mse[i1]
16. –Destrua ev[].
17. –Se mse[i1] < mse[i1 -1], rnaMinimo[][] recebe rnaAtual[][][], ce = 0.
18. –incremente ck, ck=ck+1.
19. Retorne rnaMinimo[][]

3 Trabalhos relacionados

3.0.1 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 métodos de redes neurais artificiais com backpropagation, e utilizando apenas como fonte de dados o histórico numérico das vendas realizadas, e outras variáveis intermediárias obtidas a partir deste como médias de subconjunto de observações (médias de segundas-feiras), 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.

3.0.2 Previsão de demanda em outros ambientes

(RUAS, 2012) fazem 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 desenvolvida 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.

3.0.3 Revisão bibliográfica de diversos métodos de previsão de demanda

(JUNIOR, 2007) realiza um trabalho de comparação entre diversos métodos estocásticos e Redes Neurais Artificiais (RNA) 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.

4 Metodologia

4.1 Pré-processamento

4.1.1 Coleta de 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. É importante ressaltar que tais dados são apenas de funcionários, docentes e discentes da instituição. Eventuais compras de refeições realizadas por visitantes que não possuem vínculo com a instituição não são registradas pelo sistema.

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. Os parâmetros dos quais o mesmo consegue realizar a exportação de dados são diversos, porém foi solicitado o formato de valores separados por vírgula que compreendem o ano de 2017 a 2018.

O modelo exportado segue o seguinte da seguinte forma abaixo, contendo o exemplo de 2 datas:

1 "CONSULTA POR PERÍODO ",,,,"CAMPUS: SÃO JOSÉ DOS CAMPOS ",,,,"RU: TODOS ",,,,"VINCULO: ALUNO ",,"PERÍODO DE: 01/01/2017 A 31/10/2018 ",,, DATA,VENDAS CAFÉ,VENDAS ALMOÇO,VENDAS JANTAR,VENDAS
REFEIÇÃO*,TOTAL VENDAS,ENTR. CAFÉ,ENTR. ALMOÇO,ENTR.
JANTAR,TOTAL ENTR. REFEIÇÃO*,TOTAL ENTRADA
(31/10/2018),0,395,0,395,395,0,362,0,362,362
(30/10/2018),0,667,0,667,667,0,437,256,693,693

4.1.2 Dados Climáticos

Também serão analisadas variáveis climáticas juntos com os dados de consumo, de forma a analisar a influência de fatores externos como temperatura média ambiente e precipitação. 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.

O BDMEP pode fornecer os registros de Precipitação(mm), Temp Máxima($^{\circ}$ C), Temp Mínima($^{\circ}$ C), Insolação(horas) Evaporação do Piche(mm), Temperatura Compensada Média($^{\circ}$ C), Umidade Relativa Média(%), Velocidade Vento Média(mps).

Importante ressaltar que o BDMEP leva 90 dias para registrar cada nova data.

Serão analisadas a temperatura máxima do dia que tem registro em momento próximo à frequência de refeições no horário de almoço, umidade, pressão atmosférica e velocidade do vento em metros por segundo.

4.1.3 Enriquecimento dos Dados

A informação de data, irá gerar informações derivadas da data tais como: 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.

O dia da semana seguirá 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				
	1	0	0	0	0
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 35: Entradas de dia da semana em cofatores.

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

4.1.4 Temporização dos dados endógenos

Os dados endógenos obtidos de forma discreta no conjunto de dados fornecidos pela Unifesp devem passar pelo processo de temporização em um intervalo de 5 dias, em cada registro de venda deverão ser acrescentados 5 novos atributos contendo os dados de consumo dos 5 dias anteriores. Este processo adapta o conjunto de dados para um processo de memorização das entradas que será o formato compatível de leitura nos modelos de redes neurais a serem desenvolvidos.

W	X	Y	Z	AA	AB
ENTR_ALMOÇO	ENTR_ALMOÇO_1NTR_ALMOÇO_2NTR_ALMOÇO_3NTR_ALMOÇO_4NTR_ALMOÇO_5				
17	12	17	27	18	141
431	17	12	17	27	16
442	431	17	12	17	27
393	442	431	17	12	17
408	393	442	431	17	12
286	408	393	442	431	17
409	286	408	393	442	431
449	409	286	408	393	442

Figura 37: Registros de consumo de refeições com intervalo temporal de 5 dias.

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

4.1.5 Divisão dos Conjuntos

O processo experimental será realizado em 2 roteiros distintos de divisão do conjunto de dados, e os resultados obtidos entre os 2 roteiros serão comparados. O conjunto de dados

contemplando o período de 2017 a 2019, será dividido em conjunto de treino, validação e teste da seguinte maneira: - 1º Roteiro com validação e teste de semestres iguais: Neste roteiro o semestre de validação será o primeiro semestre de 2018 e o conjunto de teste será no primeiro semestre de 2019, os dados de 2017, e 2018 – 2º semestre serão usados para treino. Os resultados obtidos nesta divisão irão validar a hipótese de que os modelos aprenderão especificamente a sazonalidade de consumo no primeiro semestre, se saindo melhor nos testes no primeiro semestre em comparação aos outros modelos treinados com validação no ano todo.

- 2º Roteiro com treino, validação e teste de anos completos.
- 2º Roteiro com treino, validação e teste de anos completos.
- Validação com o ano de 2018
- Teste com o ano de 2019

4.1.6 Tratamento das Features

Nas features endógenas (dados de venda e consumo):

- Normalizar os dados de venda e consumo para o máximo e mínimo de 3x o desvio padrão.
- Escalar os dados entre 0 e 1

Nas features exógenas:

- Escalar os dados entre 0 e 1
- As features categóricas binárias (dias da semana e semestre) já estarão escaladas
- Plotar gráficos de comportamento da feature com correlação do comportamento de consumo de refeições.

4.2 Definição e treino dos modelos

4.2.1 Avaliação de modelos modernos

De acordo com ([Data Science Academy, 2020](#)) no capítulo de arquitetura de redes neurais gated recurrent unit gru, as redes recorrentes GRU são uma melhoria das redes recorrentes padrão 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. Logo serão indicadas na aplicação deste trabalho para a memorização da sazonalidade semestral e anual dos dados do restaurante universitário.

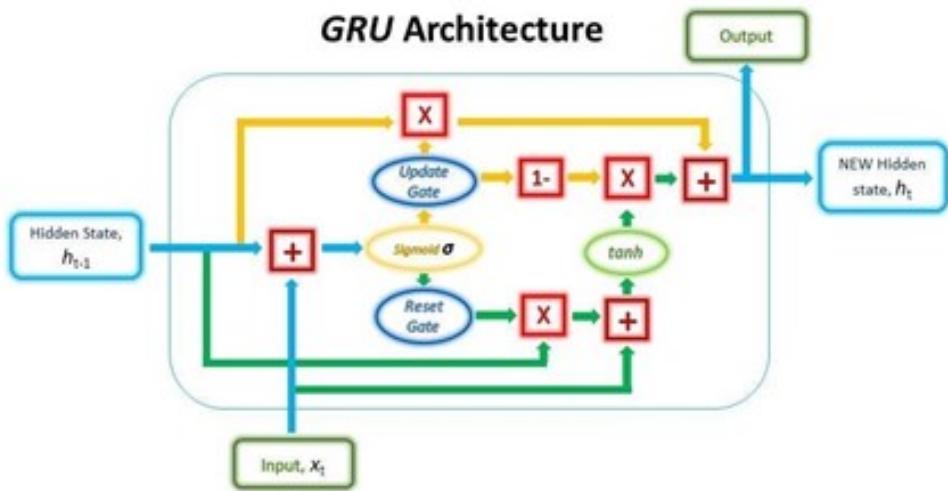


Figura 38: Arquitetura do modelo GRU

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

4.2.2 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 contínuos. Portanto é necessária a implementação de modelos específicos para entradas temporais e discretas. Para a saída final pode ser implementado um comitê de redes neurais endógenas e exógenas, com um neurônio perceptron na saída, recebendo os 2 valores dos modelos endógenos e exógenos e regredindo ao valor de saída que será a previsão do consumo.

4.2.2.1 Modelos endógenos

- Adaptar o conjunto de entrada para entradas compatíveis do modelo de rede neural GRU e do modelo fundamentado em (LOPES, 2008) conforme a figura 26 da rede neural para leitura dos valores temporais, onde cada linha de entrada trará os dados de consumo das 5 linhas anteriores.
- Desenvolver as redes perceptron de baixa profundidade para avaliar o aprendizado da rede
- Aumentar a profundidade da rede e avaliar as mudanças da função de perda rmse
- Implementar e avaliar modelos com redes recorrentes GRU, conforme a figura 38 que são especialmente desenvolvidos para o aprendizado com memorização de dados, e no caso

deste trabalho, poderão memorizar as sazonalidades semanais de consumo (de até 5 dias, dado a temporização do conjunto em 5 dias).

4.2.2.2 Modelos Mistas : Endógenos e Exógenos

- Para os dados temporais (consumo e venda) utilizar os melhores modelos endógenos dos experimentos anteriores para as entradas endógenas.
- Para os dados discretos e categóricos adaptar a entrada destes dados para rede perceptron
- Concatenar a saída das 2 redes neurais em um neurônio perceptron criando um comitê de redes neurais para obter a saída final prevista.

4.2.3 Função de ativação

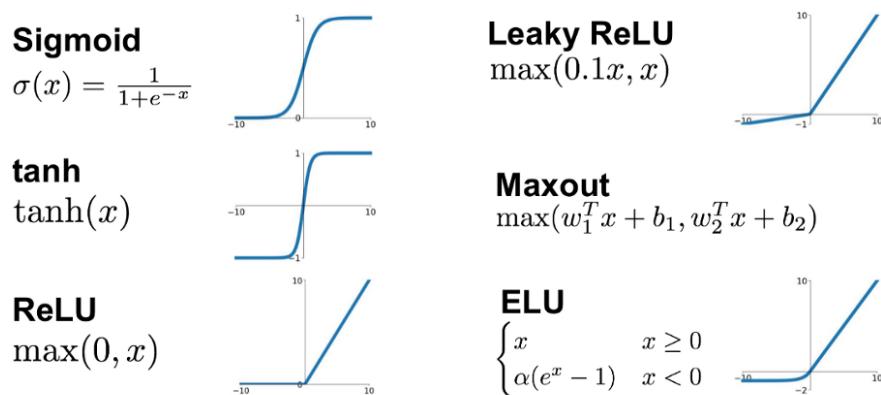


Figura 40: Funções de Ativação

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

4.2.3.1 Para a entrada e camadas ocultas

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

4.2.3.2 Para a saída

Para os valores previstos a função escolhida será a linear, pois na fase feed-backward do reajuste dos pesos pelo algoritmo de treino backpropagation, a derivada da função linear será zero, portanto os valores e peso de saída não sofrerão reajustes, mantendo a execução do algoritmo de treino apenas nos valores de entrada e nas camadas ocultas das redes.

4.2.4 Otimizador

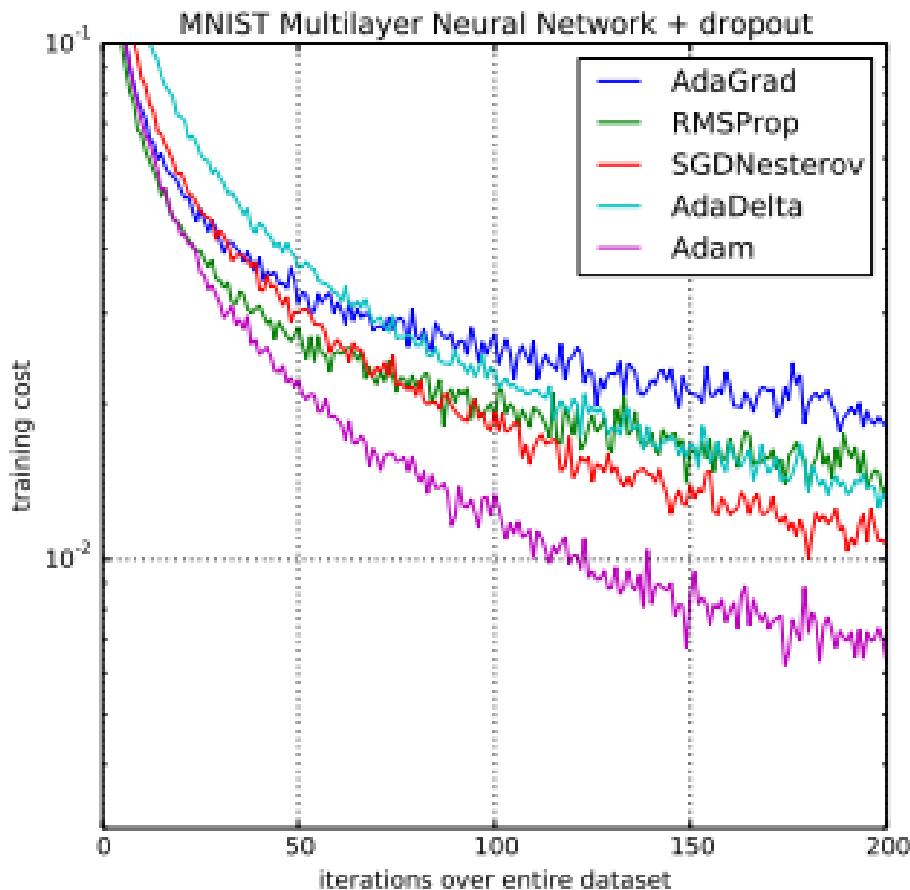


Figura 42: Comparison of Adam to Other Optimization Algorithms Training a Multilayer Perceptron

Fonte : ([Jason Brownlee, 2020](https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/)) <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>

Para este trabalho será 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 optimization combina estas 2 heurísticas. O coeficiente de aprendizado escolhido para o otimizador, ou conhecido como "learning rate", será definido para a constante alpha, com valor 0.001. Esta constante tem produzido resultados positivos em problemas de forecast (previsões) de acordo com ([Jason Brownlee, 2020](https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/))

4.3 Teste e Métricas

As 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. Citado como "Loss Function" nos diversos frameworks de machine learning como TensorFlow, Matlab, Weeka entre outros) do treino dos modelos.

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`" será 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).

Serão documentados, para trabalhos futuros, outras métricas estatísticas obtidas por esta chamada de função como os valores de slope, intercept, p_value, std_err.

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

Serão avaliados 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

5.1 1a Fase Experimental

5.1.1 Pré-Processamento

5.1.1.1 Divisão dos conjuntos

Dificuldades encontradas e resolvidas A primeira dificuldade encontrada foi um comportamento anormal dos gráficos 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 divisões fora do formato esperado.

```

1 df_file = "https://raw.githubusercontent.com/ddlandim/
            /monografia-
            /ann-demand-prediction/master/
            /experimentos_monografia/input/Todos.csv"
2 df = pd.read_csv(df_file, index_col='DATA',
                  parse_dates=True)
3 df.info()

```

Resultado obtido com um dos melhores modelos endógenos GRU, figura: [44](#)

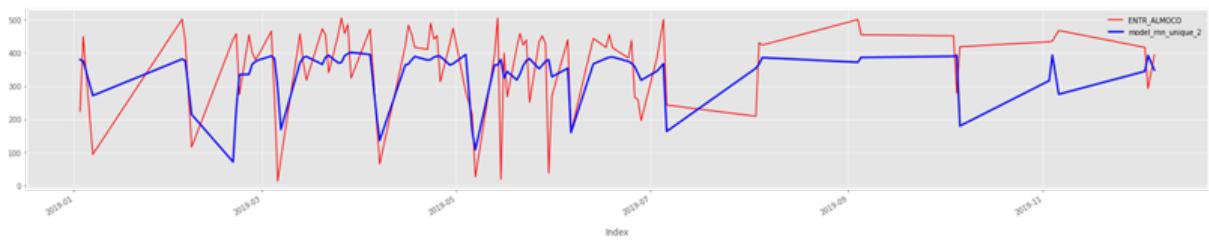


Figura 44: Resultado de um dos melhores modelos endógenos obtido sobre o conjunto de dados aleatoriamente ordenado sobre o tempo

Corrigindo a importação dos registros para o formato correto de dados, os resultados foram melhores e condizentes com o comportamento real de consumo, além disso podemos observar que a primeira fase filtra dados obtidos apenas para o 1o semestre de 2019 e no caso da indexação incorreta na figura 44 a ordenação ultrapassou o mês de julho.

```

1      infile = "https://raw.githubusercontent.com/ddlandim/
           monografy-ann-demand-
           prediction/master/experimentos_monografia
           /input/Todos.csv"
2
3      dateparse = lambda x: datetime.strptime(x, '%d/%m/%Y')
4      df = pd.read_csv(infile, parse_dates=[ 'DATA' ],
           date_parser=dateparse, index_col='DATA')

```

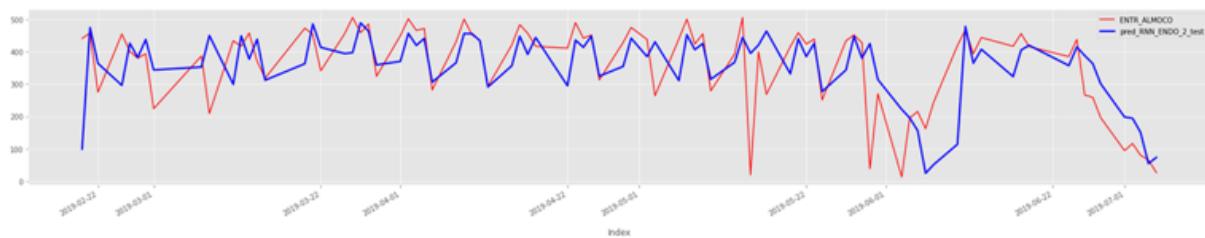


Figura 45: Resultado de um dos melhores modelos endógenos obtido sobre o conjunto de dados com ordenação corrigida

Nas figuras 46 , 47 e 48 são exibidos os valores reais de consumo nos conjuntos de treino (2017 1o e 2o semestre, 2018 2o semestre), validação (2018 - 1o semestre) e teste (2019 - 1o semestre).

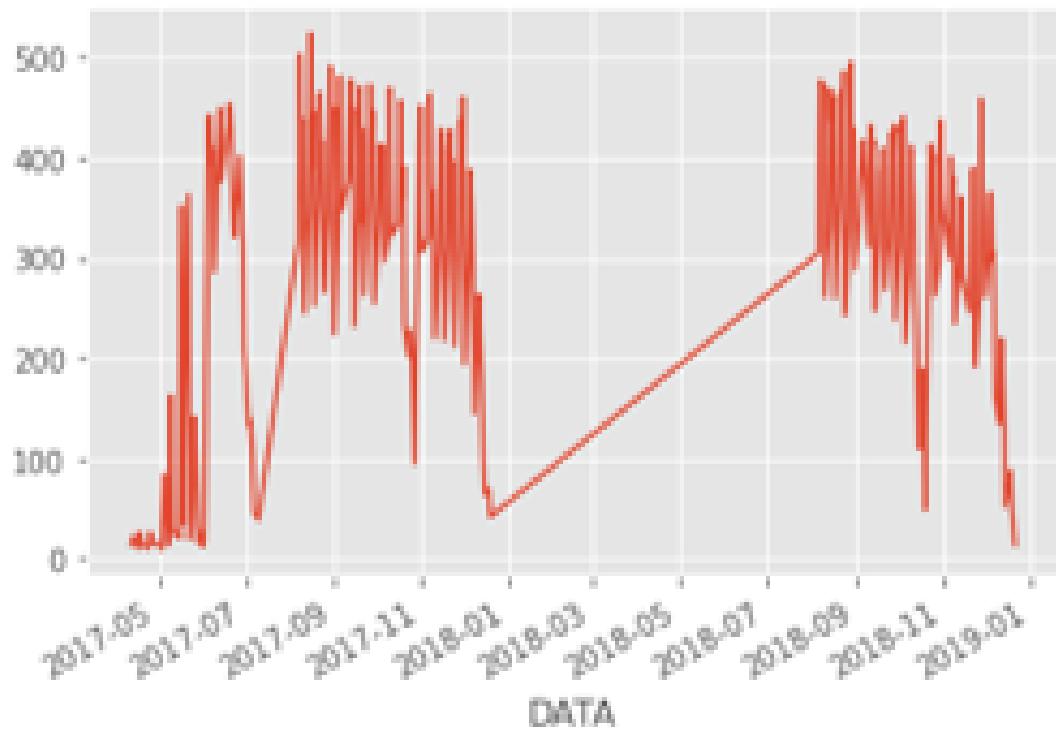


Figura 46: Conjunto verdade de treino, dados de 2017 1o e 2o semestre, e 2018 2o semestre.

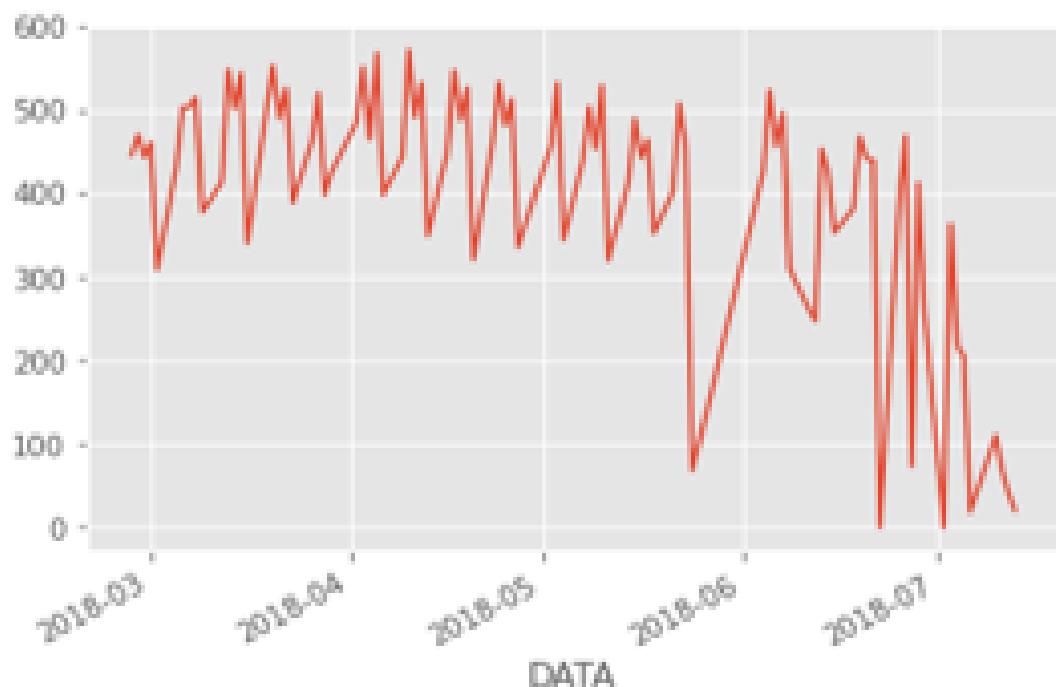


Figura 47: Conjunto verdade de validação, dados de 2018 1o semestre.

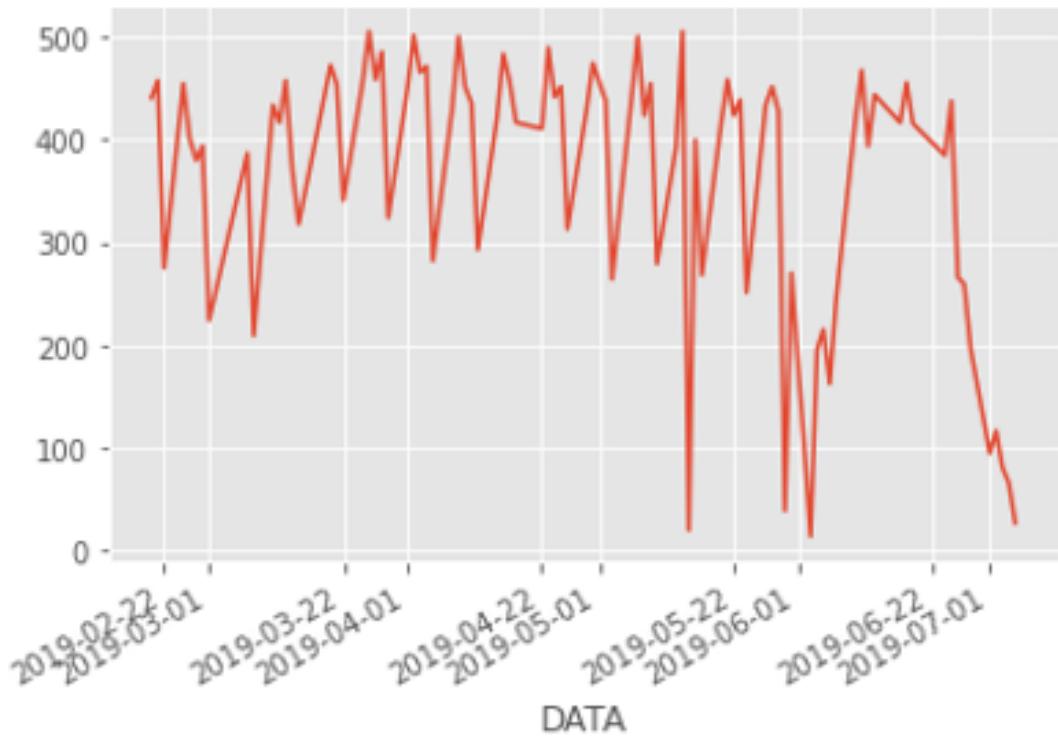


Figura 48: Conjunto verdade de teste, dados de 2019 1º semestre.

5.1.1.2 Análise das variáveis endógenas

Consumo atual em relação às vendas do 1º dia anterior

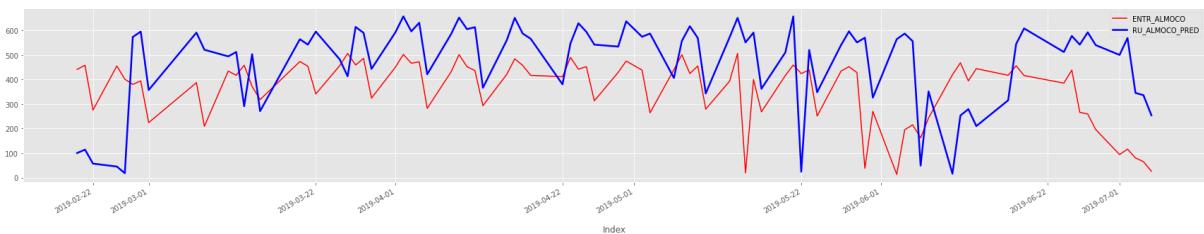


Figura 49: Correlação entre consumo e vendas de almoço.

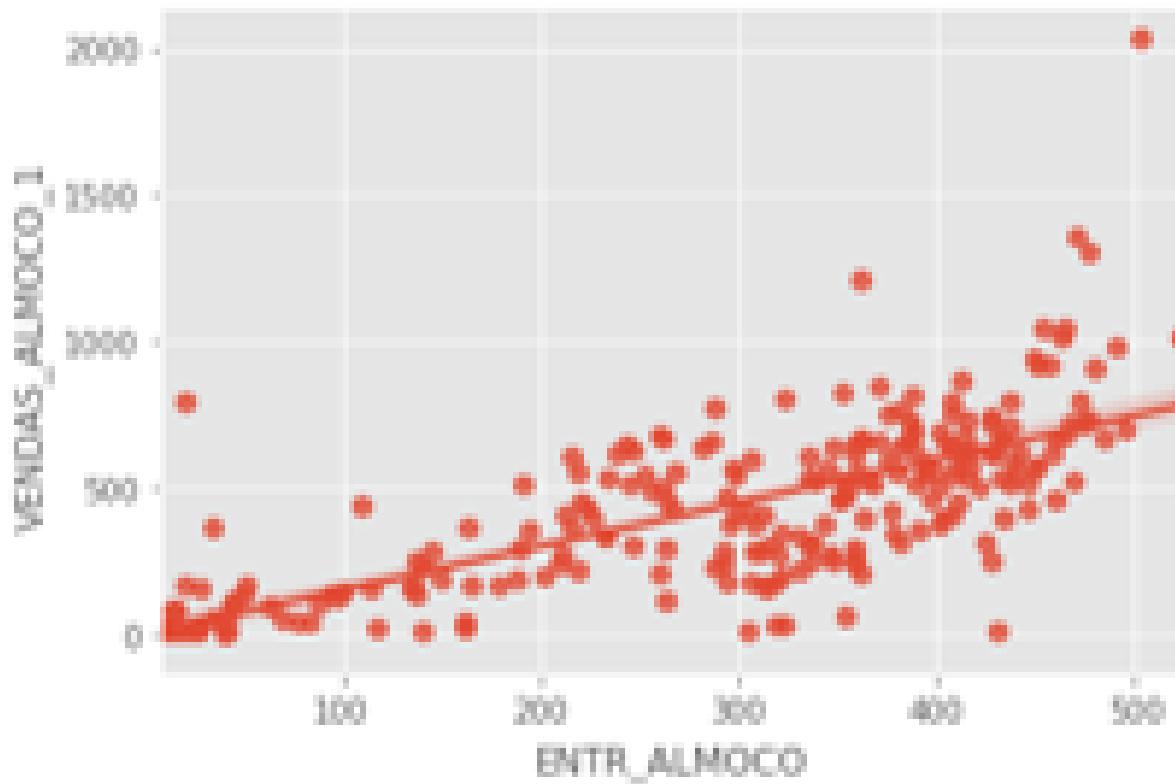


Figura 50: Gráfico Scatter entre consumo e vendas de almoço.

Métricas:

CONSUMO EM RELAÇÃO ÀS VENDAS DE 1 DIA ANTERIOR

CORRELAÇÃO (r)	0.7255528038157009
Pi (p)	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 49 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 presente 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.

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

Após a normalização dos dados para 3x o desvio padrão médio, 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 51, se manteve.

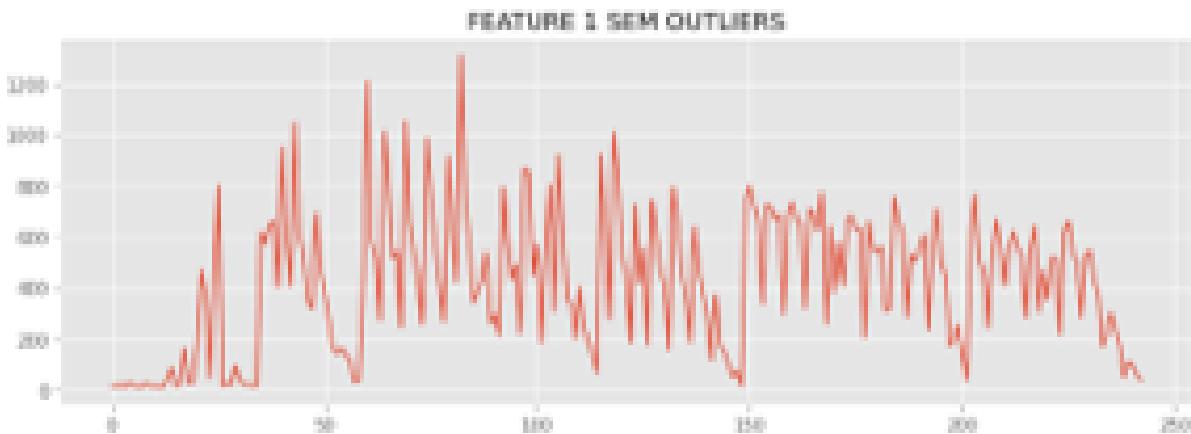


Figura 51: Vendas de tickets normalizados com teto de 3x o desvio padrão.

E após a aplicação da escala na feature 52, o comportamento linear também se manteve.

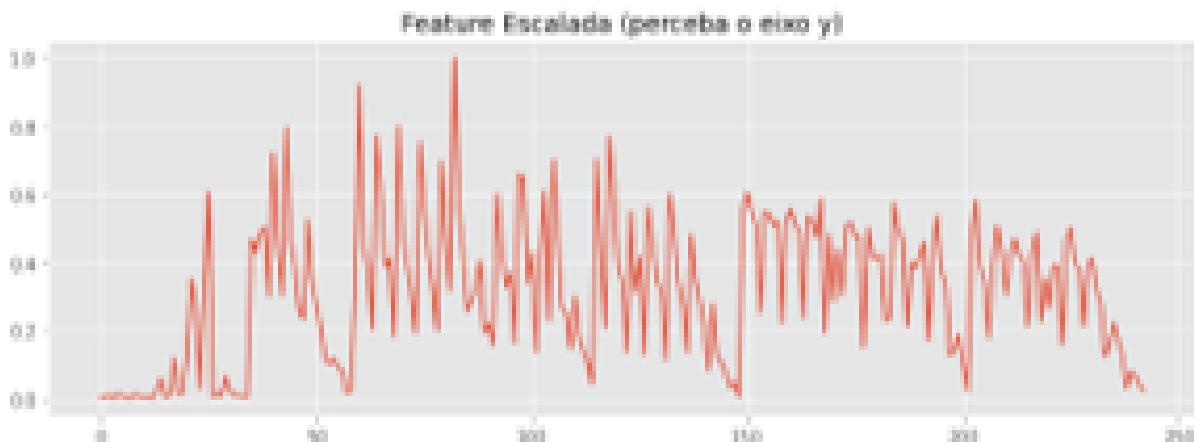


Figura 52: Vendas de tickets escalada.

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

Análise da técnica do restaurante A análise da técnica 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 53, gera um erro muito maior do que 30% devido ao comportamento oscilatório do consumo, conforme a figura 54.

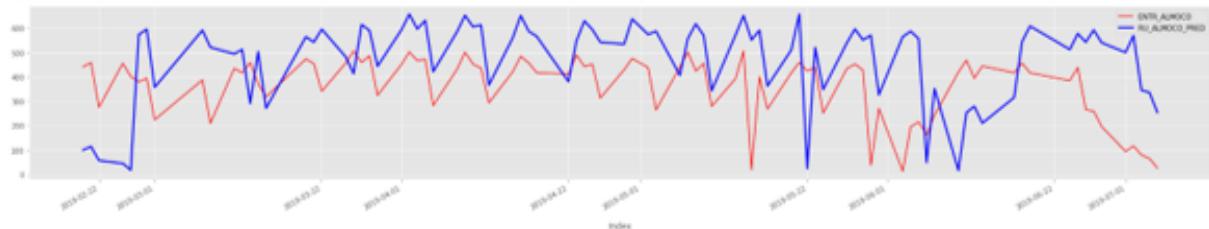


Figura 53: Correlação entre consumo e produção com 30% acima do consumo da semana anterior.

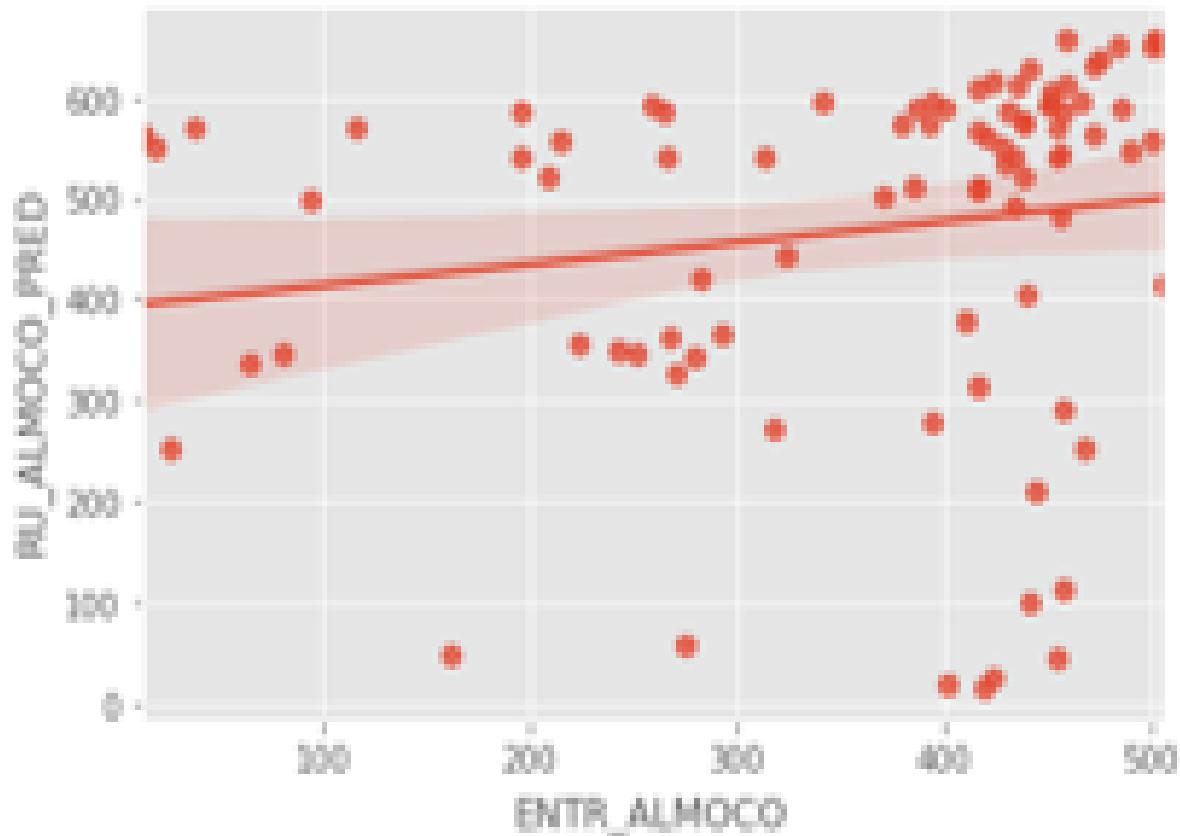


Figura 54: Gráfico Scatter entre consumo e produção com margem de 30%.

Consumo atual em relação ao consumo do jantar de 1 dia anterior. Mesmo se tratando de turmas diferentes, os alunos que consomem refeições no almoço são de período e grade horária diferente dos alunos no período da noite, mas há uma relação evidente entre os 2 consumos, conforme as figuras 55 e 56.

COMPORTAMENTO DE CONSUMO EM RELAÇÃO À JANTA DE 1 DIA ANTERIOR

CORRELAÇÃO (r): 0.765599143585576

P-value :4.526981505307197e-48

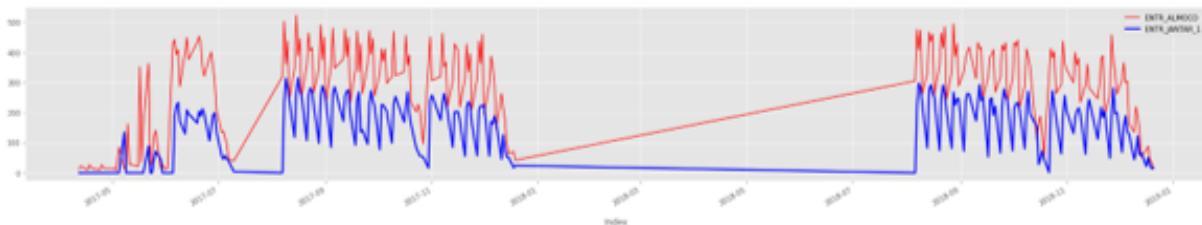


Figura 55: Correlação de consumo de almoço e jantar de 1 dia anterior.

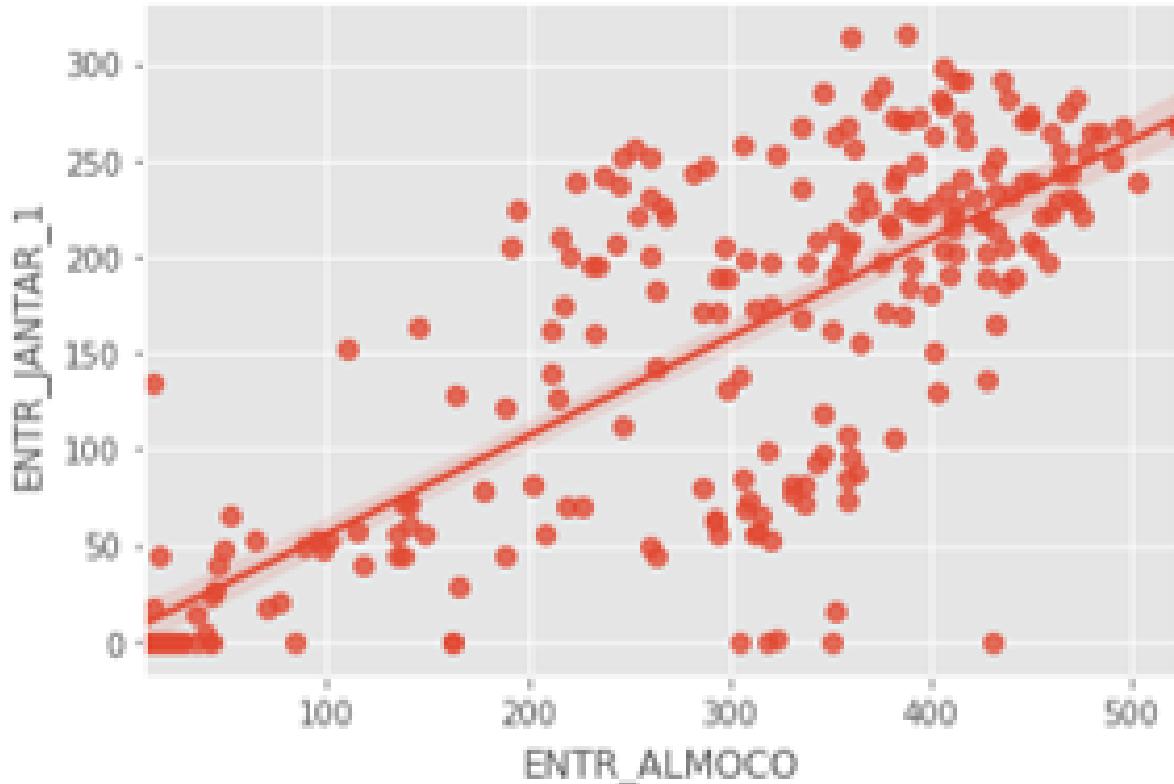


Figura 56: Gráfico Scatter entre consumo e jantar de 1 dia anterior.

5.1.1.3 Análise da sazonalidade semanal

Os gráficos a seguir da figura 57, representando a segunda-feira, até a figura 61 , 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 longo do dataset. O violino com valor zero pode ser ignorado e é um retorno padrão da ferramenta, re-

presentando 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.

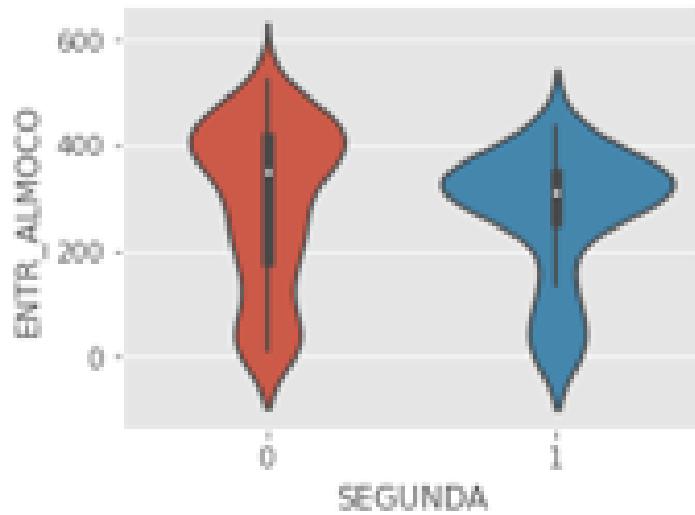


Figura 57: Gráfico violino da distribuição do consumo na segunda feira.

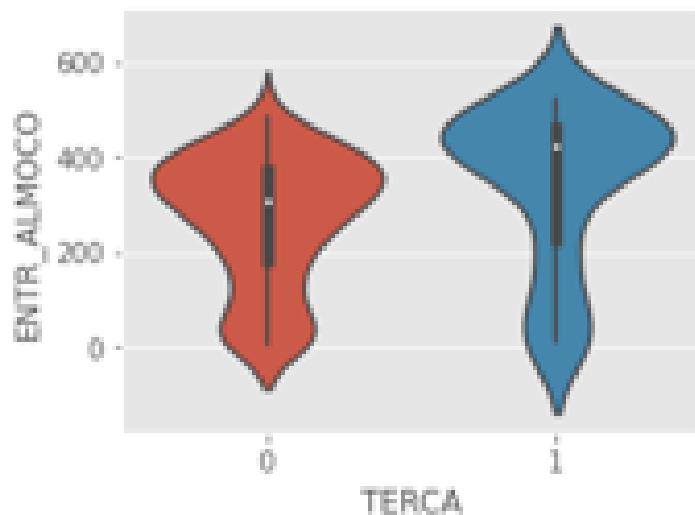


Figura 58: Gráfico violino da distribuição do consumo na terça feira.

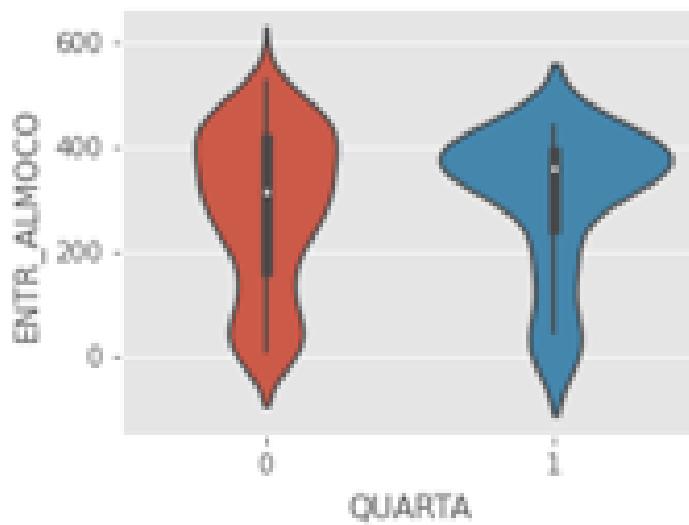


Figura 59: Gráfico violino da distribuição do consumo na quarta feira.

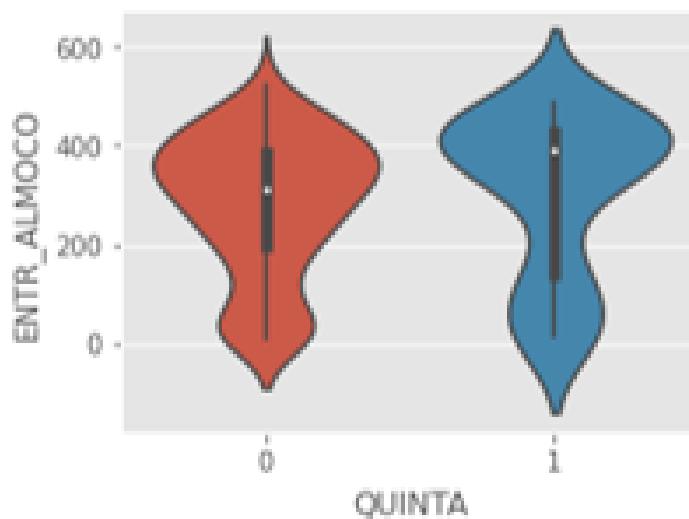


Figura 60: Gráfico violino da distribuição do consumo na quinta feira.

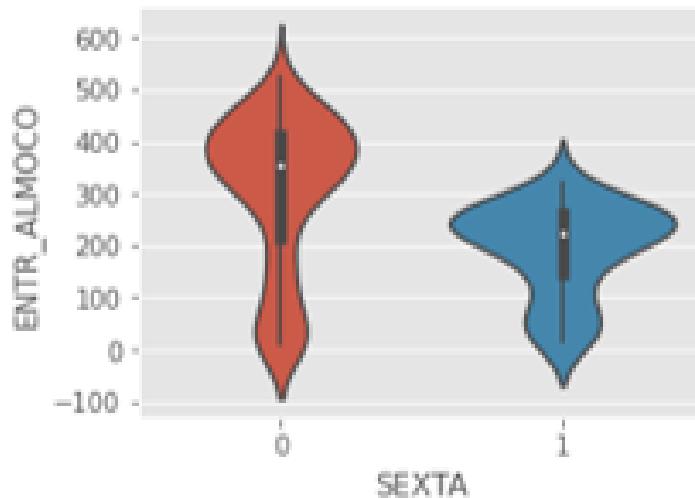


Figura 61: Gráfico violino da distribuição do consumo na sexta feira.

5.1.1.4 Análise das variáveis exógenas

Consumo atual em relação ao avanço do semestre A correlação do consumo em relação ao avanço do semestre, nos últimos dias do semestre o consumo cairam abruptamente, nota-se uma correlação negativa, mas devido às oscilações de consumo durante o semestre a correlação não se torna significativa.

COMPORTAMENTO DE CONSUMO EM RELAÇÃO AO AVANÇO DO SEMESTRE

CORRELAÇÃO (r): -0.351388192148681

Pi (p) :1.8029942608003656e-08

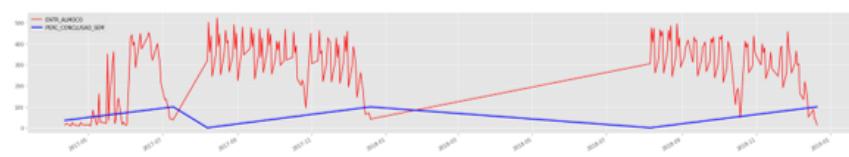


Figura 62: Correlação da distribuição do consumo com o avanço do semestre.

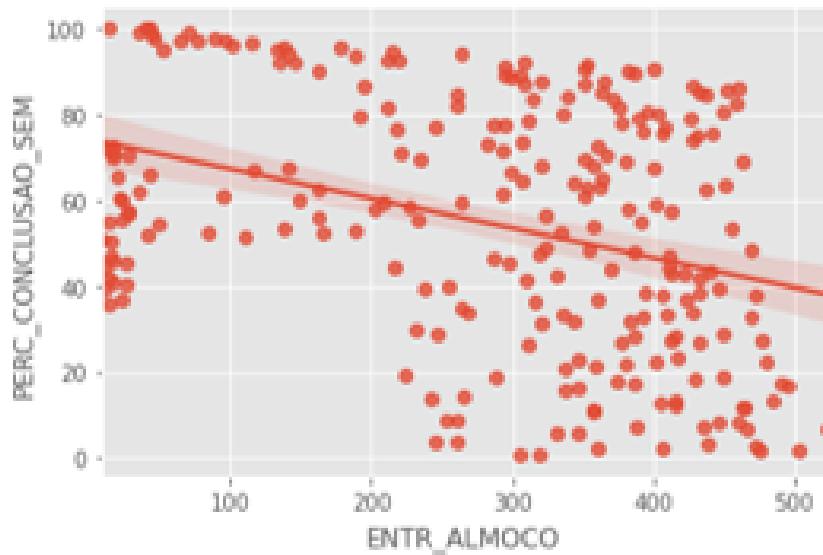


Figura 63: Gráfico scatter da distribuição do consumo com o avanço do semestre.

Consumo atual em relação ao avanço do mês Já o consumo em relação ao avanço do mês é inconclusivo. Esta hipótese validaria que os alunos tivessem preferência de realizar refeições fora do restaurante em períodos típicos de dias de pagamentos geralmente no 5º dia útil, ou todo dia 20. 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.

COMPORTAMENTO DE VENDAS EM RELAÇÃO AO AVANÇO DO MÊS

CORRELAÇÃO (r): 0.04867329355640767

Pi (p) :0.4500827452374414

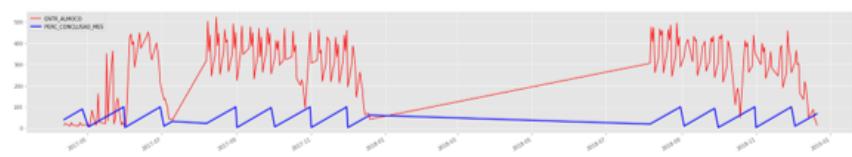


Figura 64: Correlação da distribuição do consumo com o avanço do mes.

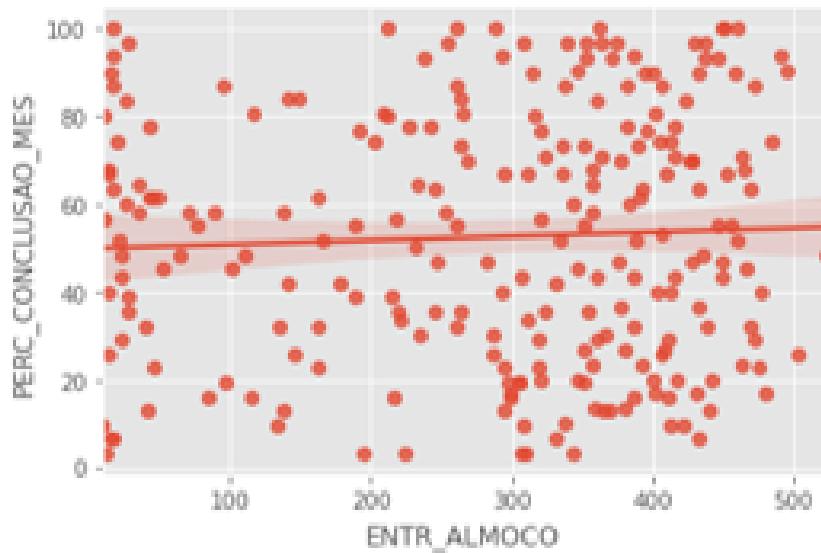


Figura 65: Gráfico scatter da distribuição do consumo com o avanço do mes.

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 66 à 73.

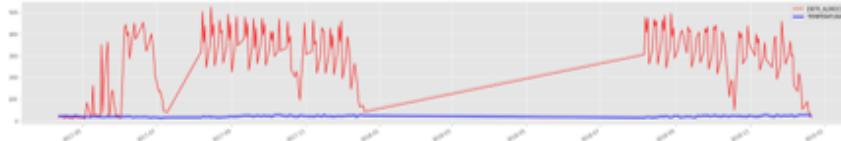


Figura 66: Correlação da distribuição do consumo com a temperatura.

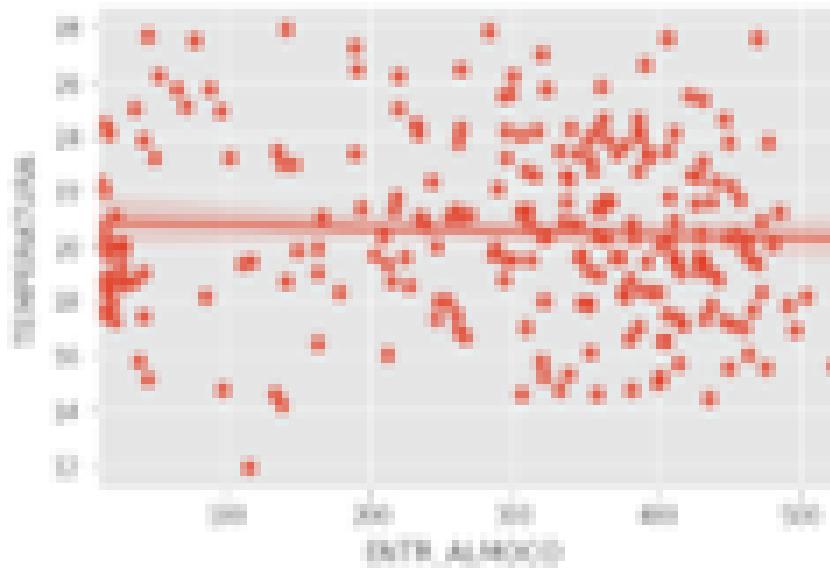


Figura 67: Gráfico escatter da distribuição do consumo com a temperatura.

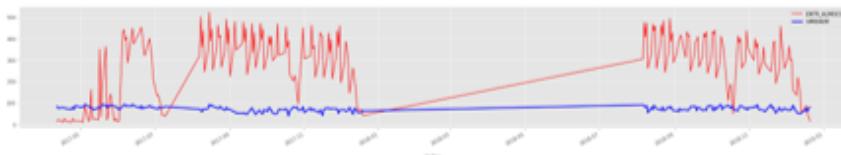


Figura 68: Correlação da distribuição do consumo com a umidade.

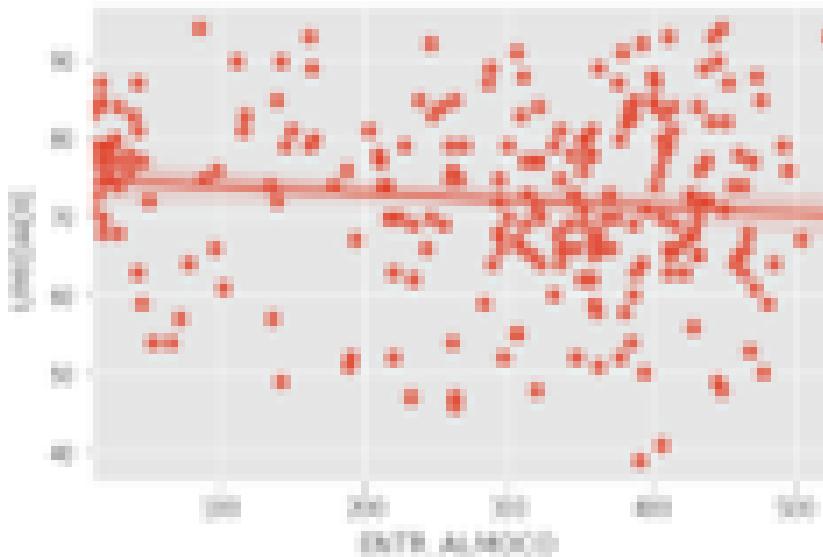


Figura 69: Gráfico escatter da distribuição do consumo com a umidade.

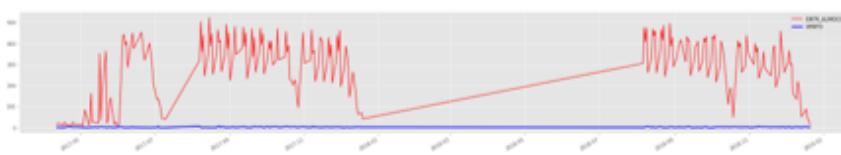


Figura 70: Correlação da distribuição do consumo com a velocidade do vento em m/s.

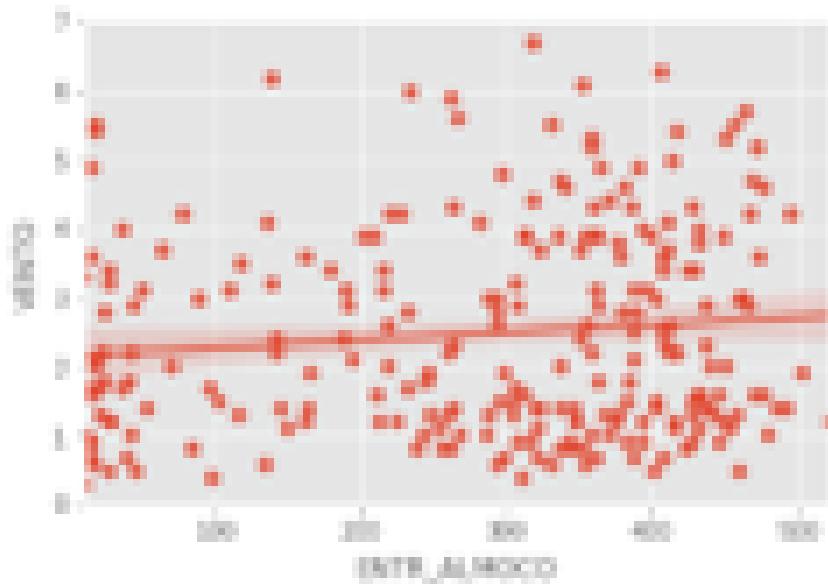


Figura 71: Gráfico escatter da distribuição do consumo com a velocidade do vento.

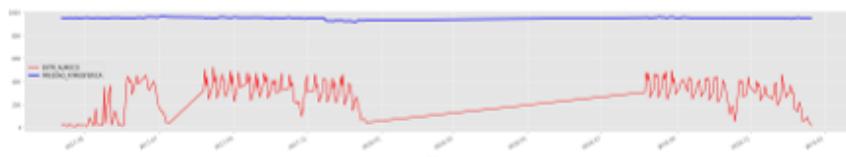


Figura 72: Correlação da distribuição do consumo com a pressão atmosférica.

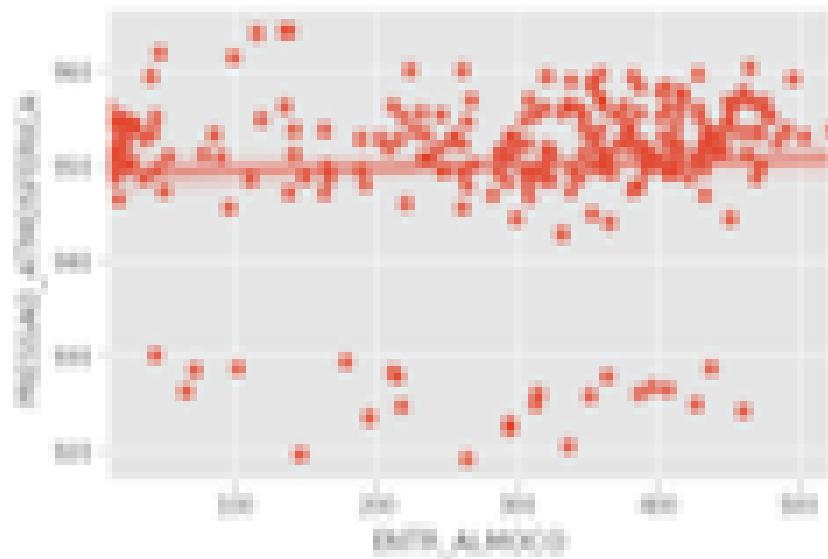


Figura 73: Gráfico escatter da distribuição do consumo com a pressão atmosférica.

5.1.2 Definição e treino dos modelos

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 (de 1 a 5 dias anteriores para cada entrada), 5 para vendas de tickets e 5 para consumo de jantar. O modelo será denominado MLP1 para futuras referências.

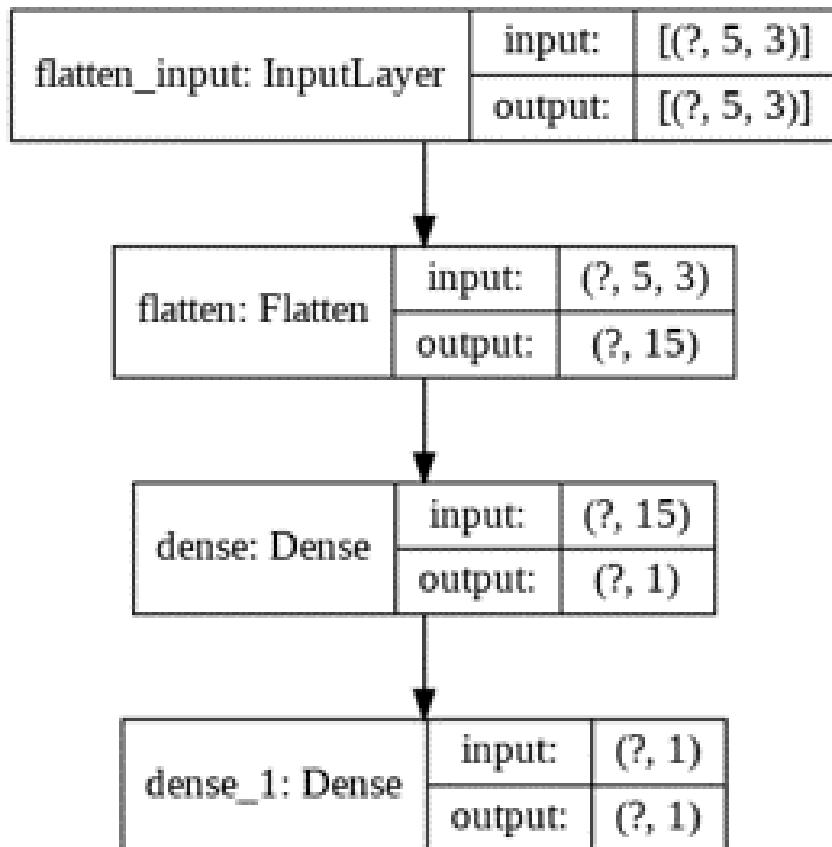


Figura 74: Topologia do modelo MLP1

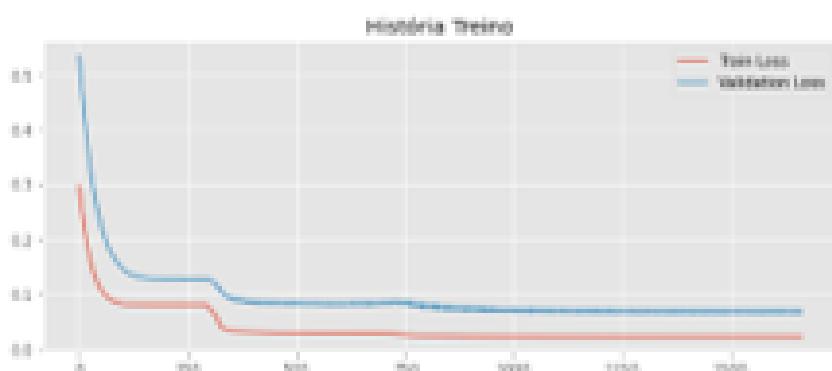


Figura 75: Gráfico de treino do modelo MLP1.

A raiz do erro quadrático médio do modelo foi, RMSE = 130.6207420543897

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

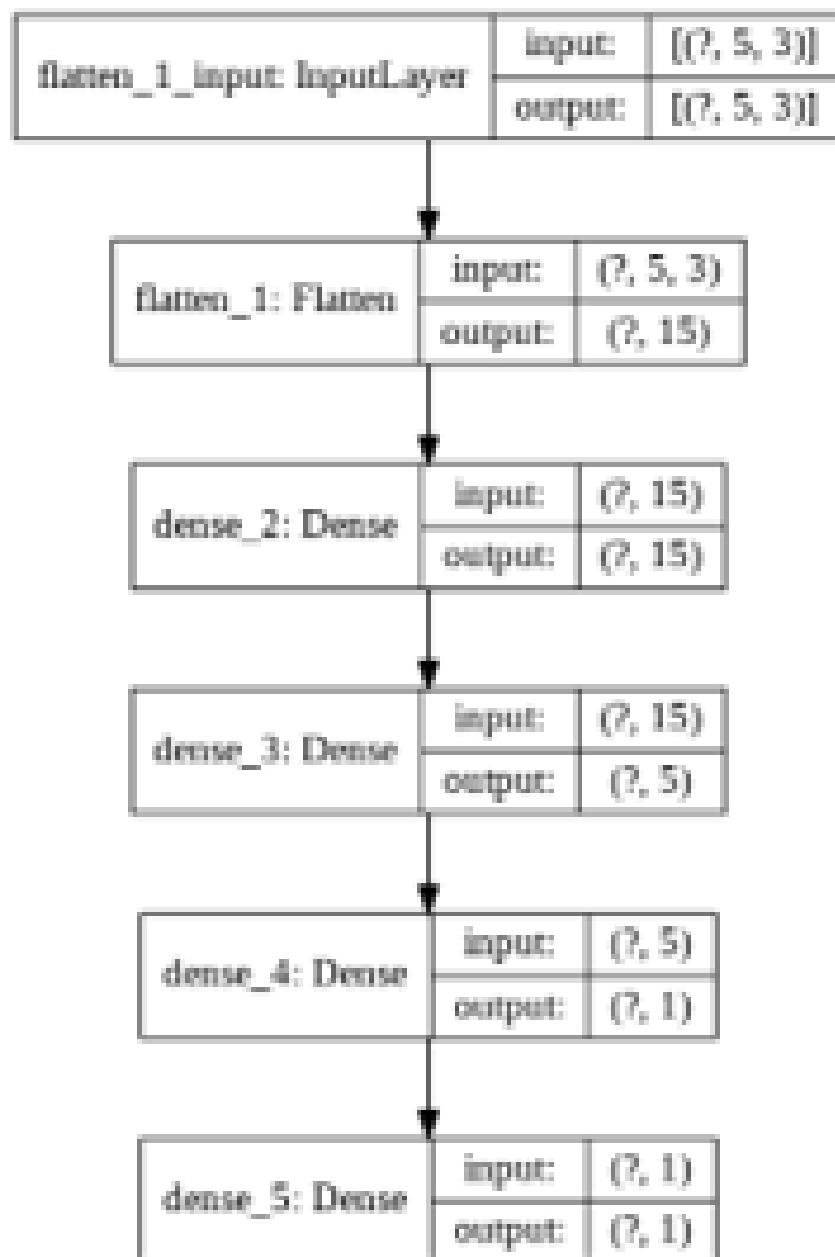


Figura 76: Topologia do modelo MLP2

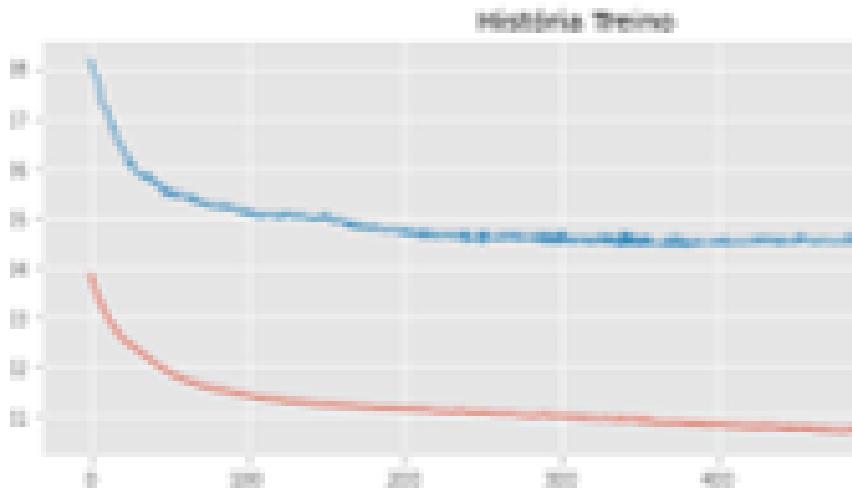


Figura 77: Gráfico de treino do modelo MLP2.

A raiz do erro quadrático médio do modelo MLP2 foi, $\text{RMSE} = 107.97413966672336$. 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 os modelos tem capacidade de aprendizado do problema em relação ao ajuste da topologia dos mesmos. As figuras de topologia e treino dos modelos vão de 74 à 77.

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

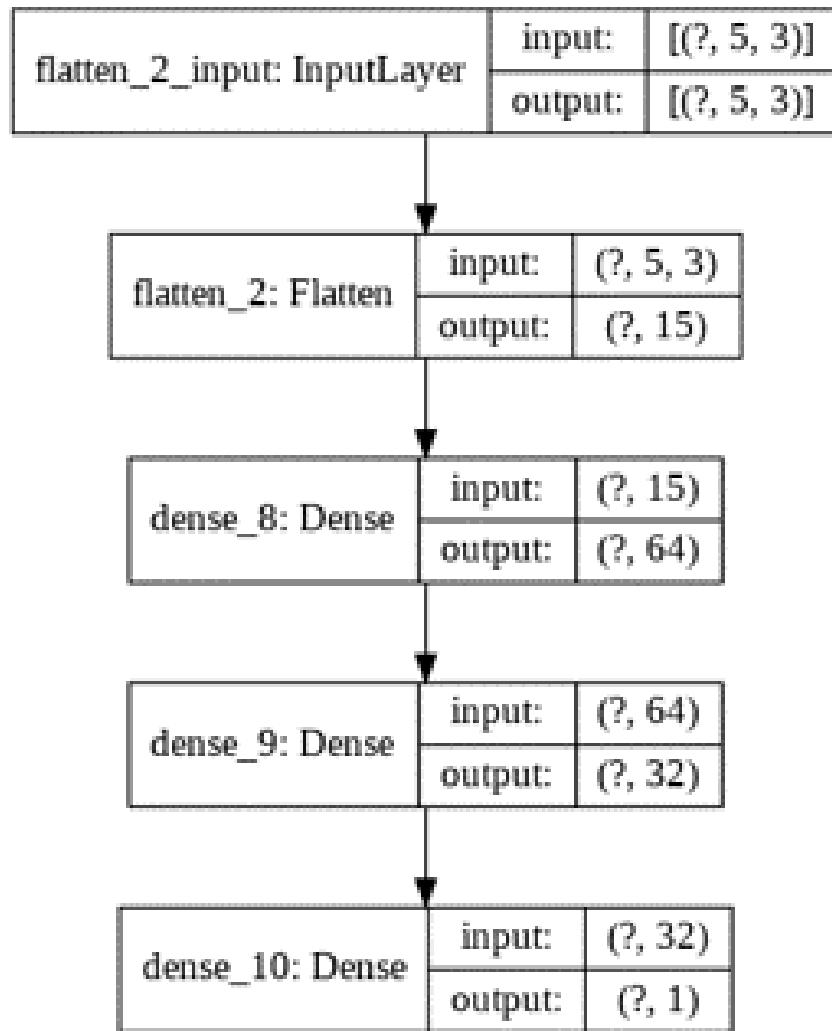


Figura 78: Topologia do modelo MLP_ENDO_1

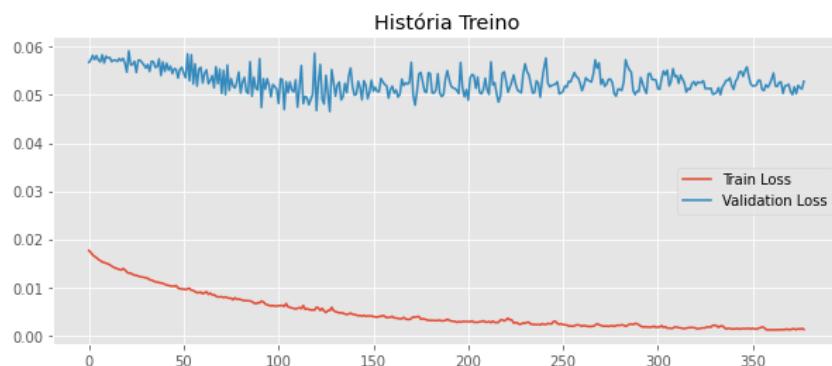


Figura 79: Treino do modelo MLP_ENDO_1



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

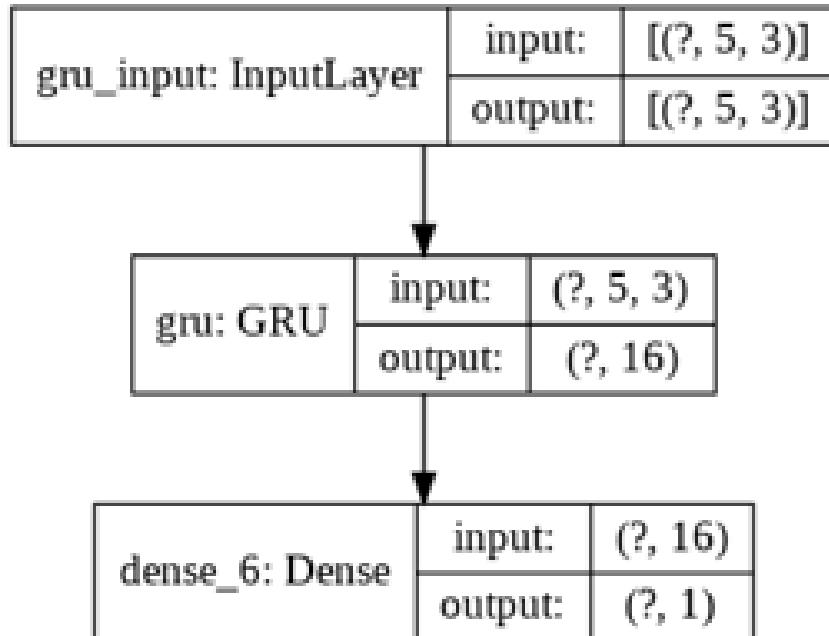
RNN_ENDO_1

Figura 81: Topologia do modelo RNN_ENDO_1

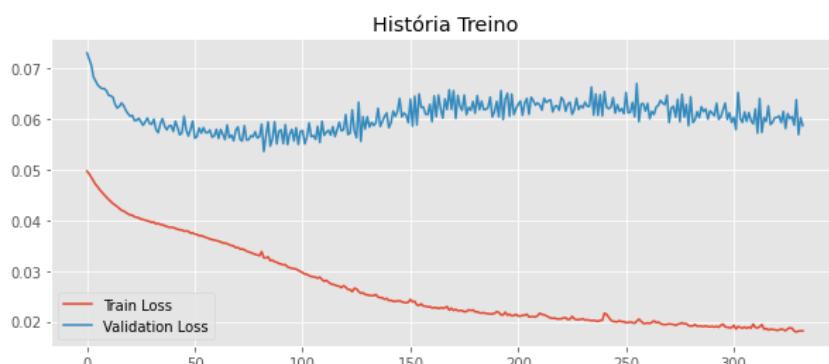


Figura 82: Treino do modelo RNN_ENDO_1

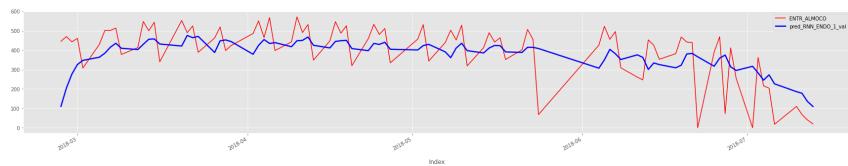


Figura 83: Avaliação do modelo RNN_ENDO_1

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

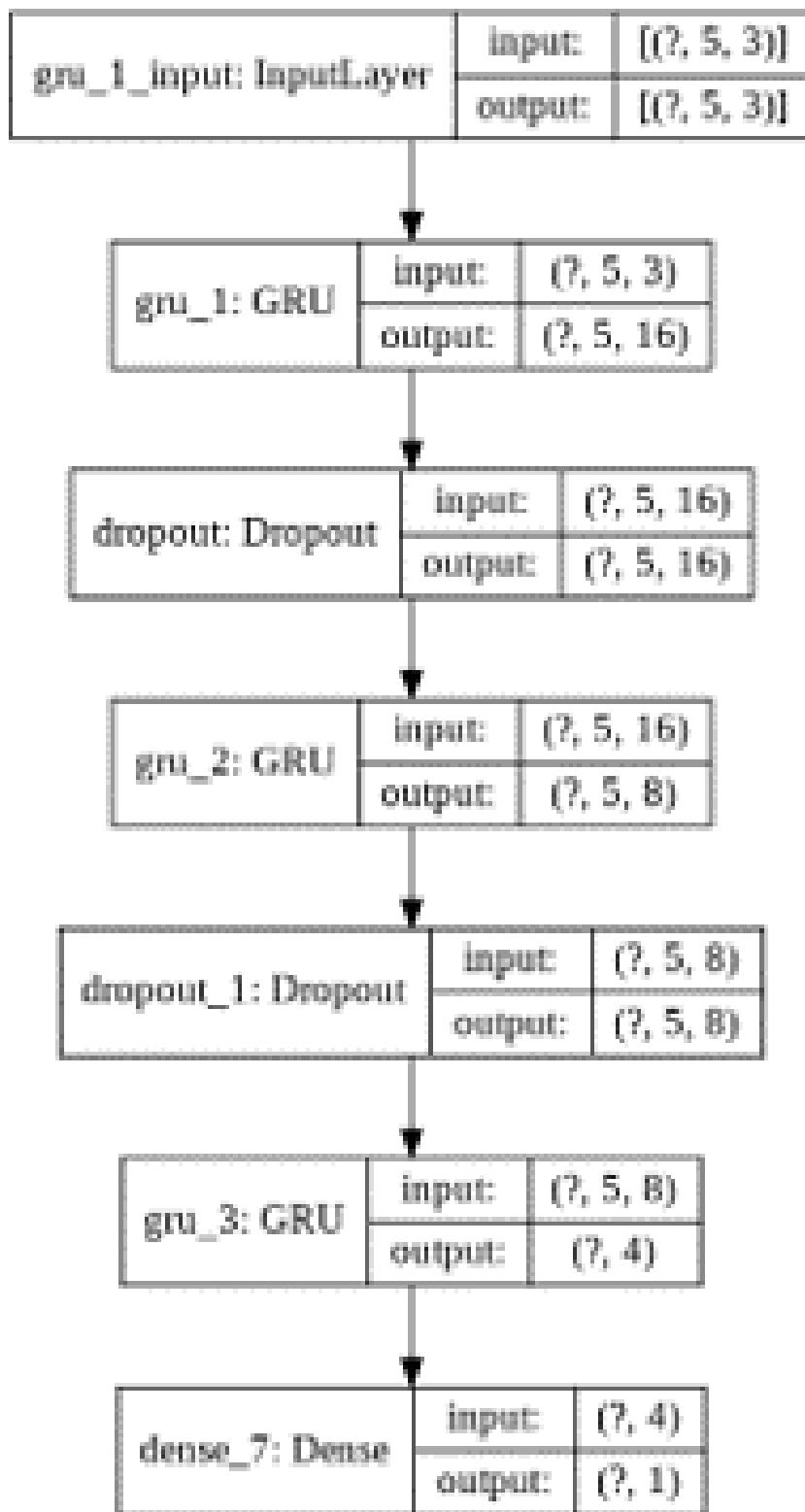


Figura 84: Topologia do modelo RNN_ENDO_2

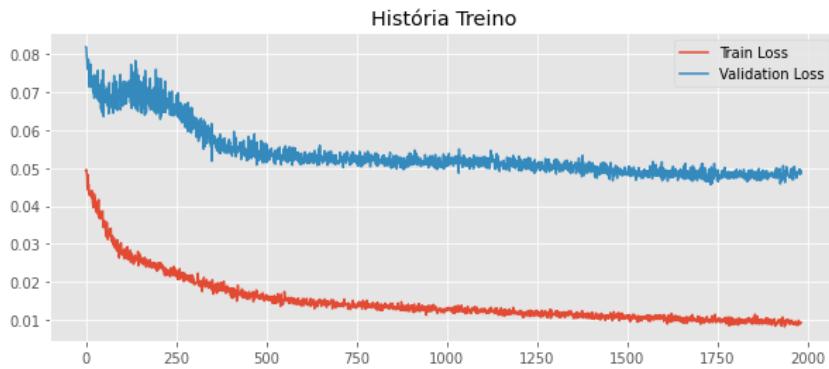


Figura 85: Treino do modelo RNN_ENDO_2

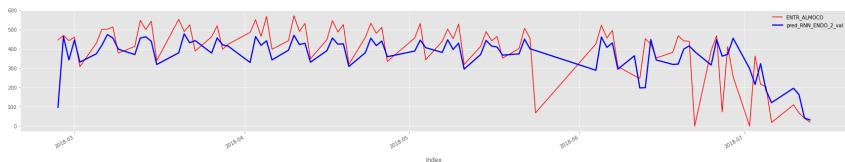


Figura 86: Avaliação do modelo RNN_ENDO_2

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

Comparativo gráfico da etapa de avaliação no teste no conjunto de validação, entre os melhores modelos endógenos treinados, RNN_ENDO_2 e MLP_ENDO_1

É possível notar, conforme figura 87 que mesmo com correlação e RMSE muito próximo nos 2 modelos, o modelo GRU 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 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.

RNN_ENDO_2 :

CORRELAÇÃO (r): 0.6753384458677313

Pi (p) :3.903678433027079e-13

MLP_ENDO_1 :

MLP: CORRELAÇÃO (r): 0.6791684214084684

$\text{Pi (p)} : 2.5590689588089596e-13$

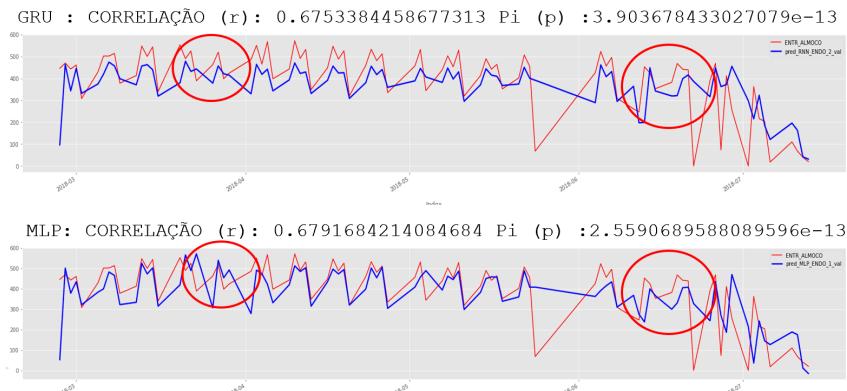


Figura 87: Comparação da etapa de validação entre MLP_ENDO_1 e RNN_ENDO_1

5.1.2.3 Modelos mistos endógenos e exógenos

RNN_EXO_1

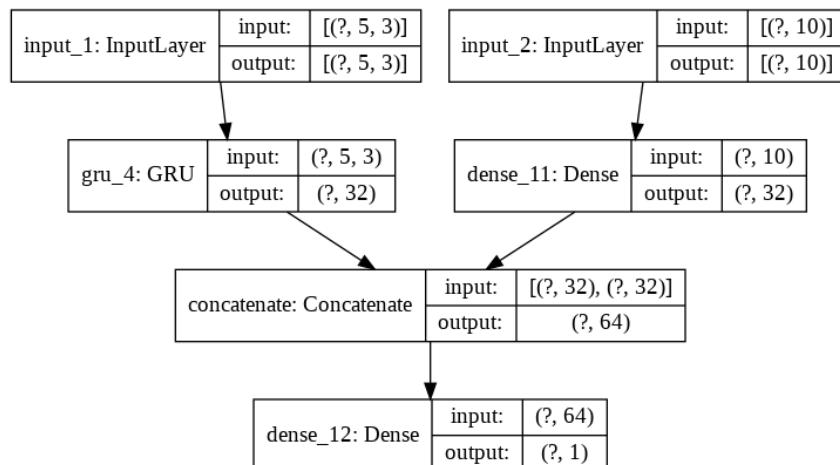


Figura 88: Topologia do modelo RNN_EXO_1

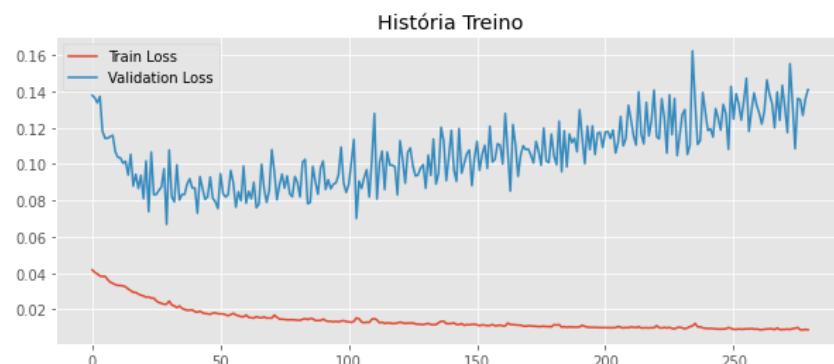


Figura 89: Treino do modelo RNN_EXO_1

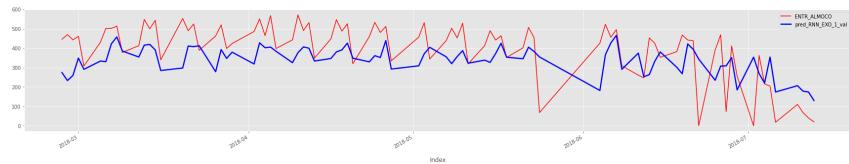


Figura 90: Avaliação do modelo RNN_EXO_1

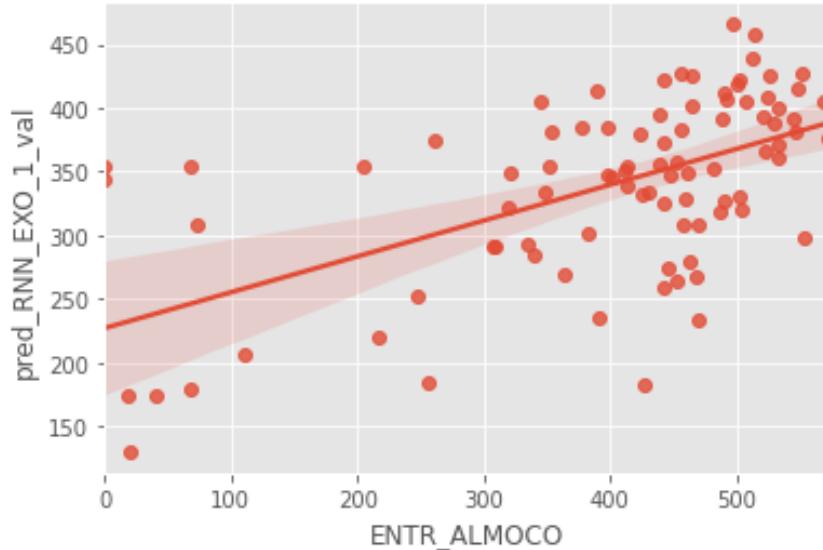


Figura 91: Grafico scatter de avaliação do modelo RNN_EXO_1

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

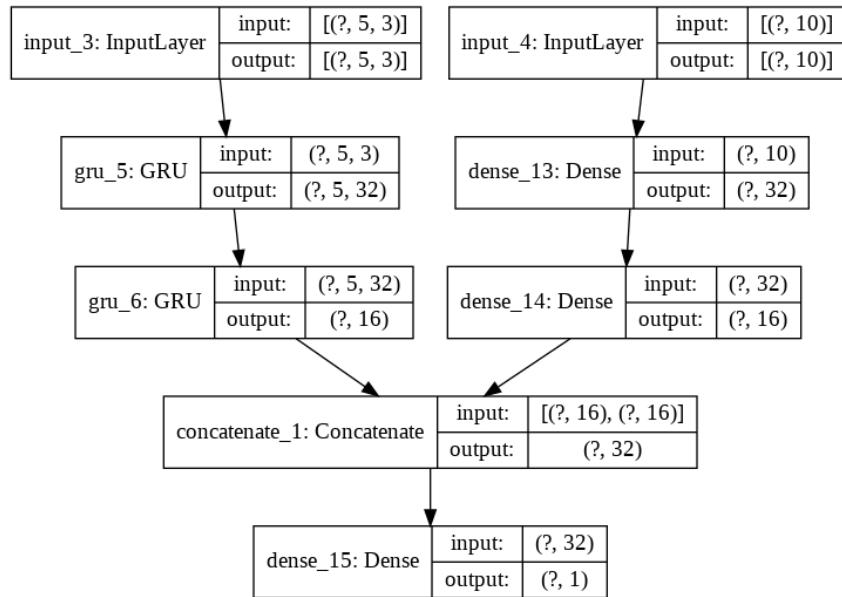


Figura 92: Topologia do modelo RNN_EXO_2

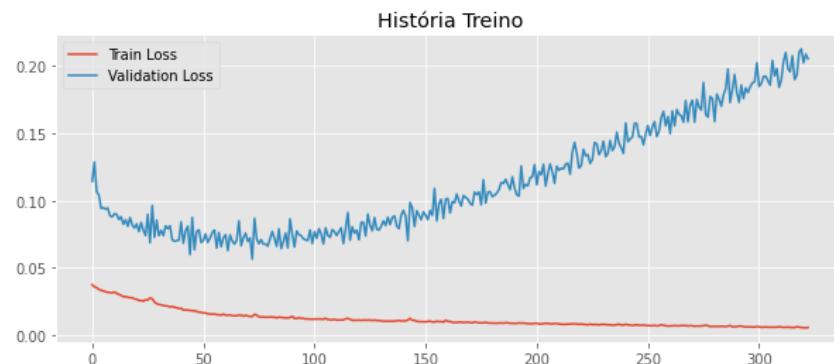


Figura 93: Treino do modelo RNN_EXO_2

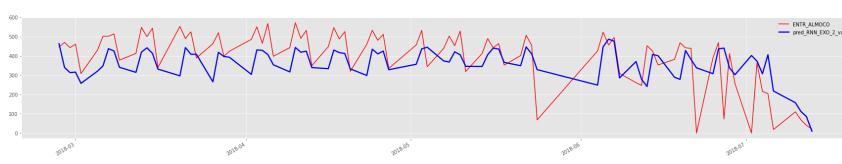


Figura 94: Avaliação do modelo RNN_EXO_2

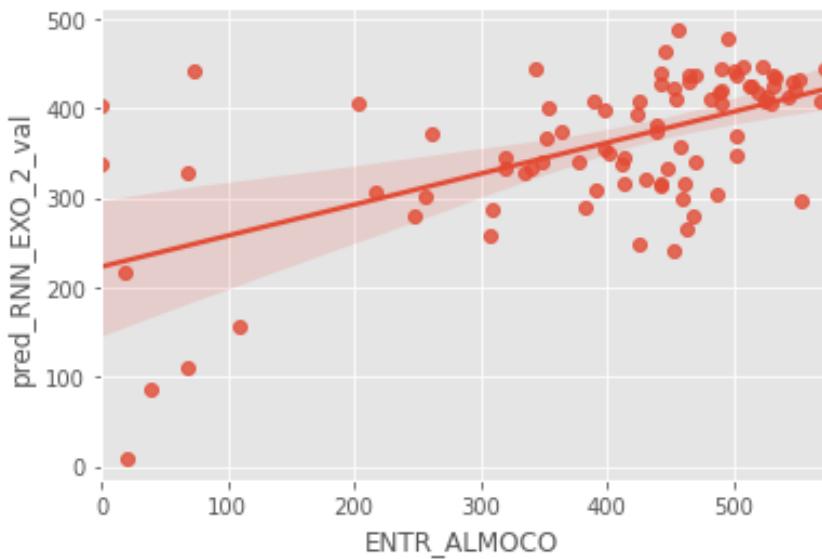


Figura 95: Grafico scatter de avaliação do modelo RNN_EXO_2

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

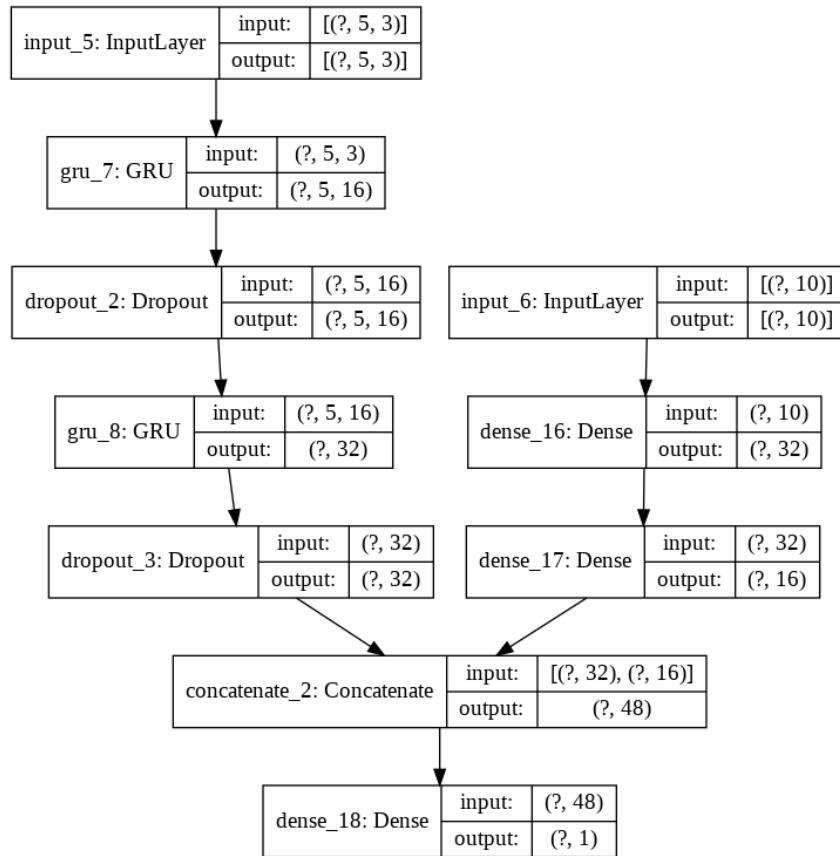


Figura 96: Topologia do modelo RNN_EXO_3

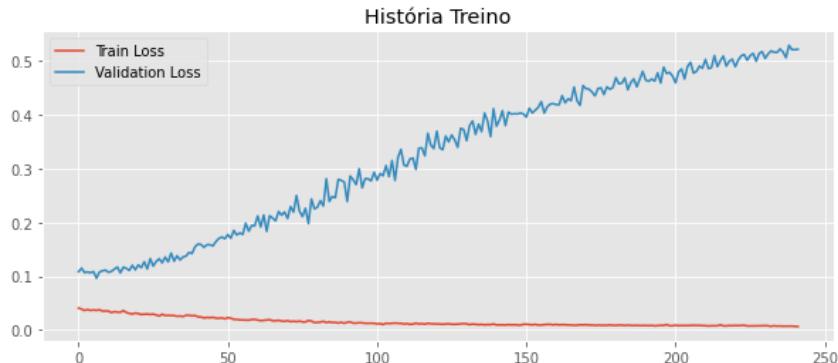


Figura 97: Treino do modelo RNN_EXO_3

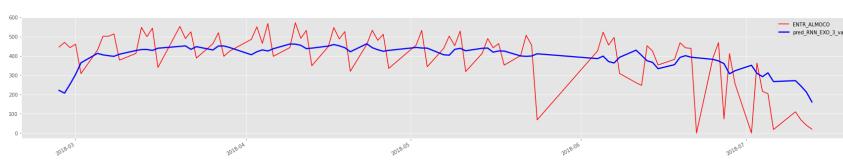


Figura 98: Avaliação do modelo RNN_EXO_3

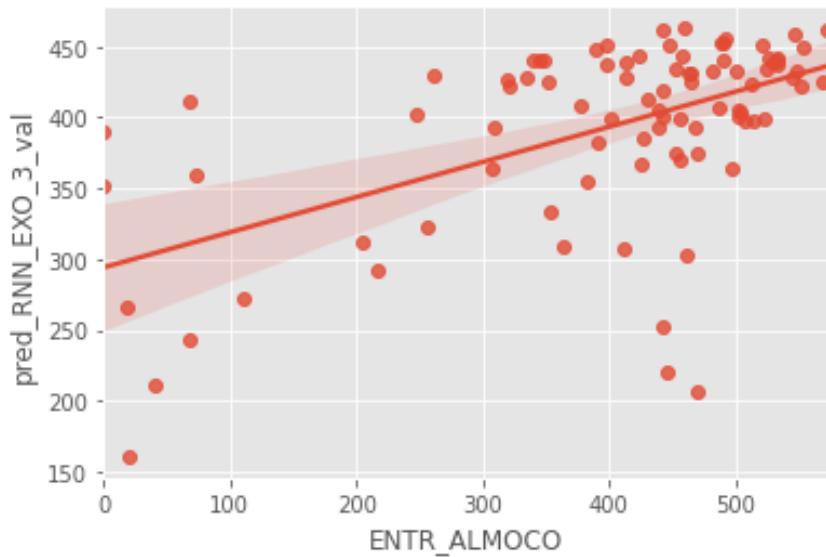


Figura 99: Grafico scatter de avaliação do modelo RNN_EXO_3

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

5.1.3.1 Endógenos : GRU x MLP

Podemos notar na figura 100 e 101 que apesar dos coeficientes de correlação próximos no teste entre os modelos RNN_ENDO_2 e MLP_ENDO_1 impactando na proximidade de outras métricas como RMSE, MAPE, etc. O valor p-value é notoriamente discrepante.

Resultados dos modelos RNN_ENDO_2 X MLP_ENDO_1:

RNN_ENDO_2 : CORRELAÇÃO (r): 0.5954398951050144 Pi (p) :9.422151772316392e-10

MLP_ENDO_1 : MLP: CORRELAÇÃO (r): 0.5212411617024483 Pi (p) :1.9211576129056944e-07

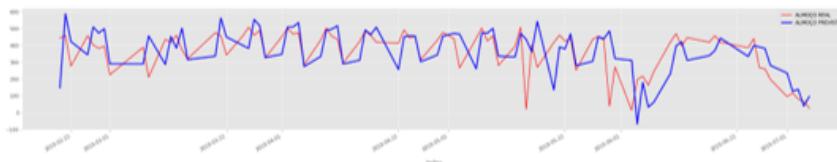


Figura 100: Teste MLP_ENDO_1

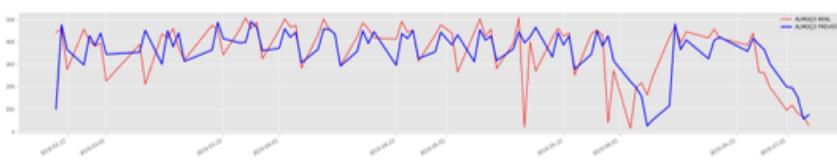


Figura 101: Teste RNN_ENDO_2

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 são devidamente justificadas por datas especiais no calendário do primeiro semestre:

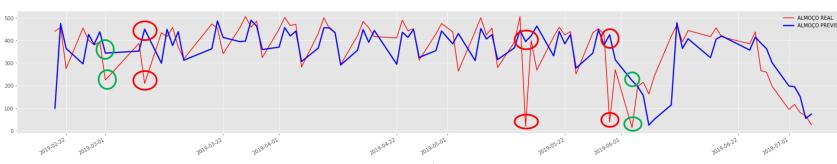


Figura 102: Analise de anomalias preditivas do RNN_ENDO_2

As datas que produziram outliers e anomalias de predição no melhor modelo exógeno, 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

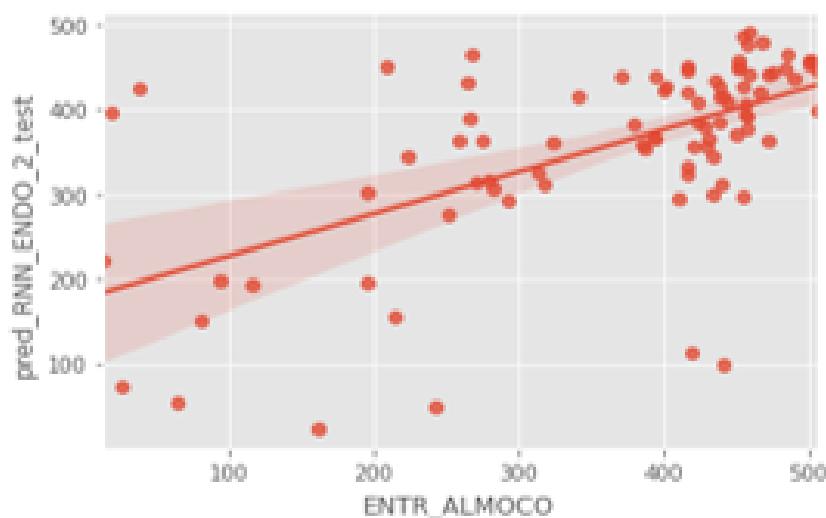


Figura 103: Grafico scatter de teste do RNN_ENDO_2

5.1.3.3 Modelos mistos

RNN_EXO_1

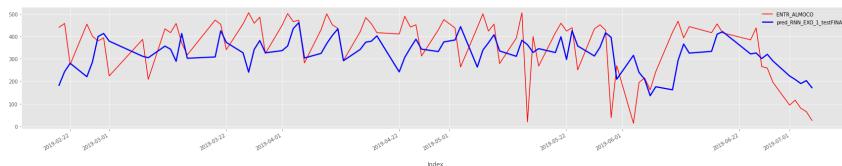


Figura 104: Teste do modelo RNN_EXO_1

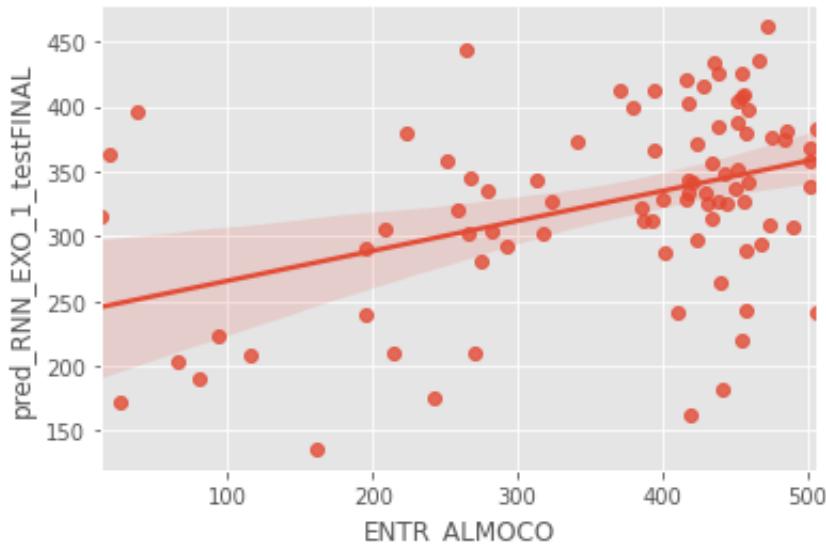


Figura 105: Grafico scatter de teste do modelo RNN_EXO_1

RNN_EXO_1:	
TOTAL DE REFEIÇÕES CONSUMIDAS	31962 : 88
TOTAL DE REFEIÇÕES PROJETADAS	28728.816 : 88
ERRO DE PREVISÃO	-3233.18359375
PERCENTAGEM DE ERRO	-10.115711137444466%
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%

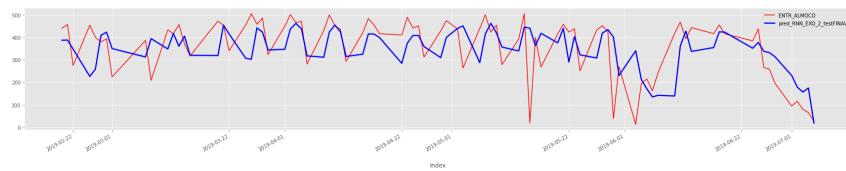
RNN_EXO_2

Figura 106: Teste do modelo RNN_EXO_2

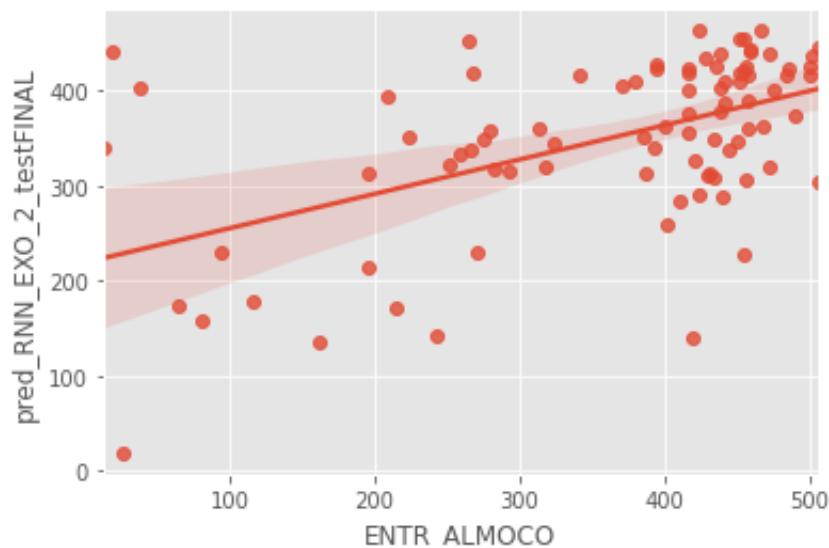


Figura 107: Grafico scatter de teste do modelo RNN_EXO_2

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%

RNN_EXO_3

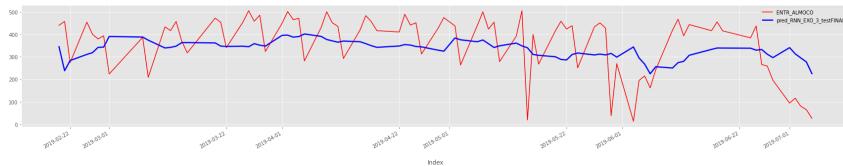


Figura 108: Teste do modelo RNN_EXO_3

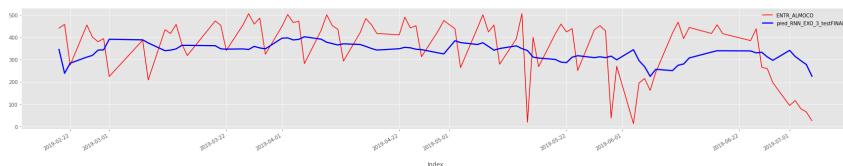


Figura 109: Grafico scatter de teste do modelo RNN_EXO_3

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%

5.1.3.4 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 de previsões deste modelo se encontra na figura 101 Comparativos finais:

5.2 2a Fase Experimental

5.2.1 Pré-Processamento

5.2.1.1 Divisão dos conjuntos

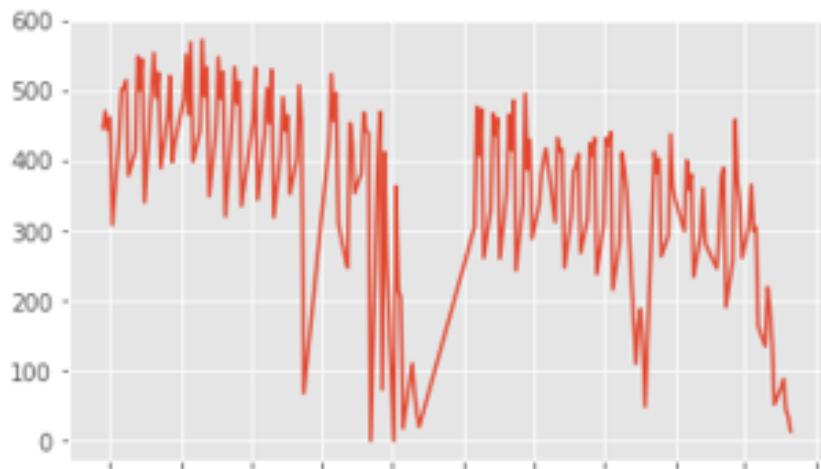


Figura 110: Conjunto de dados de treino, ano de 2017 completo

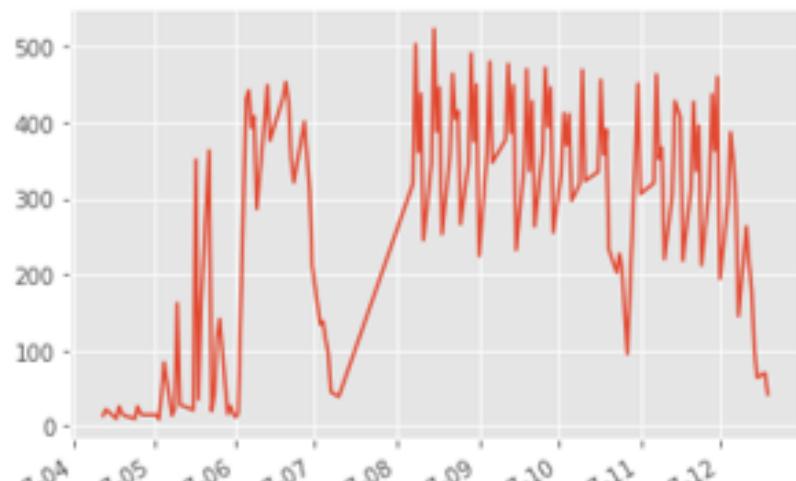


Figura 111: Conjunto de dados de validação, ano de 2018 completo

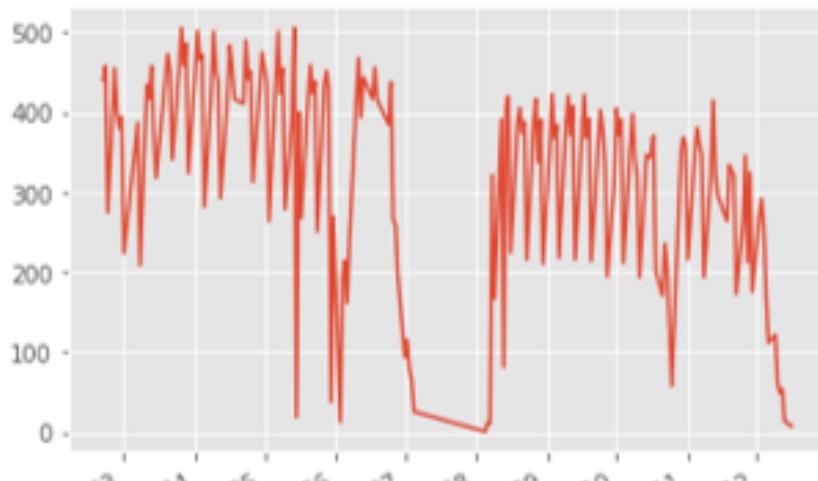


Figura 112: Conjunto de dados de teste, ano de 2019 completo

A nova divisão dos conjuntos segue conforme as figuras 110 para o conjunto de treino, 110 para o conjunto de validação e 110 para o conjunto de teste.

5.2.1.2 Análise da técnica do restaurante

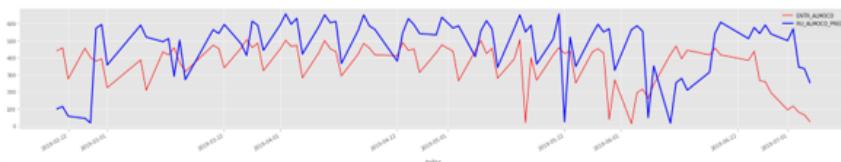


Figura 113: Gráfico entre o consumo de refeições e a produção com margem de 30% para o ano de 2019 completo.

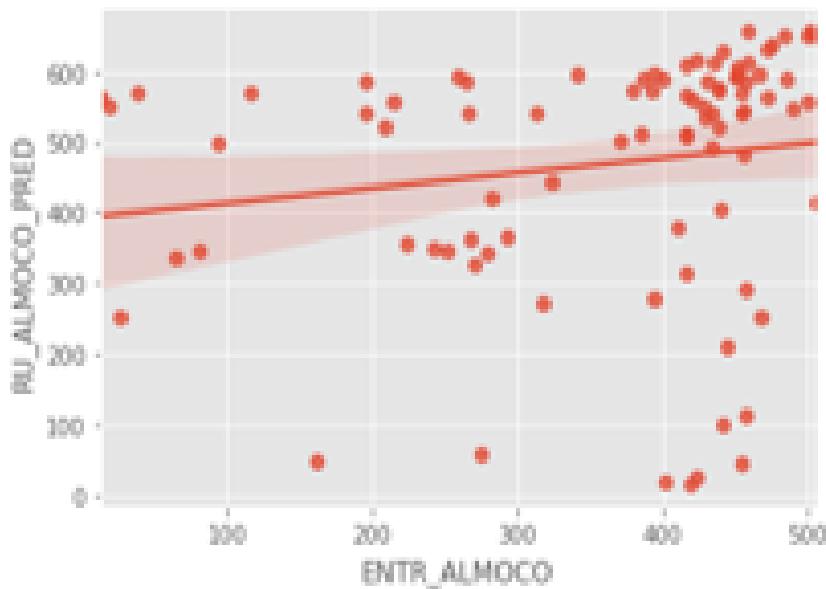


Figura 114: Gráfico scatter entre o consumo de refeições e a produção com margem de 30% para o ano de 2019 completo.

Para a produção de refeições com margem de 30% acima do consumo do 5o dia anterior, conforme figura 113 e 114 obtemos as seguintes métricas:

Consumo com margem 30% acima do 5o dia anterior	
TOTAL DE REFEIÇÕES CONSUMIDAS	58653 : 182
TOTAL DE REFEIÇÕES PROJETADAS	76262 : 182
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

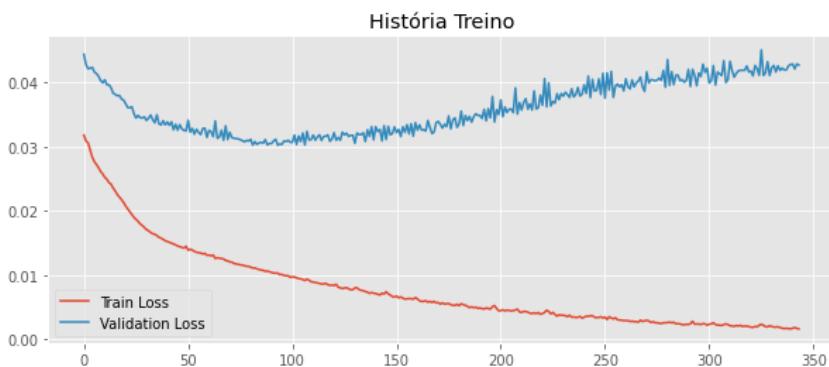


Figura 115: Gráfico de treino do modelo MLP_ENDO_1.

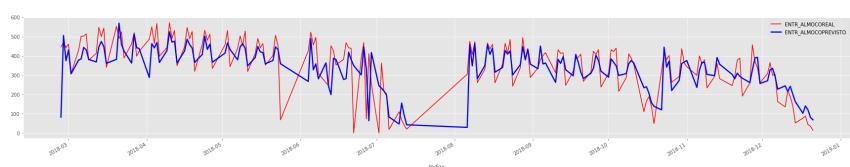


Figura 116: Gráfico de validação do modelo MLP_ENDO_1.

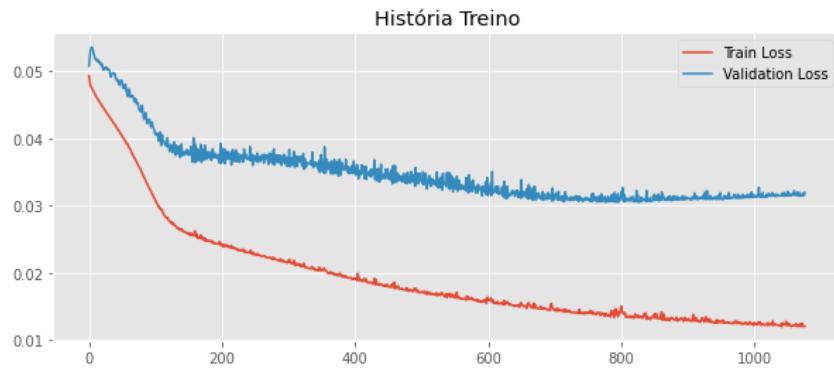


Figura 117: Gráfico de treino do modelo RNN_ENDO_1.

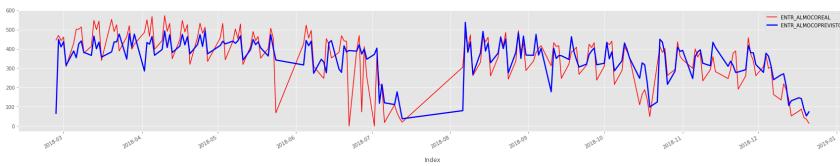


Figura 118: Gráfico de validação do modelo RNN_ENDO_1.

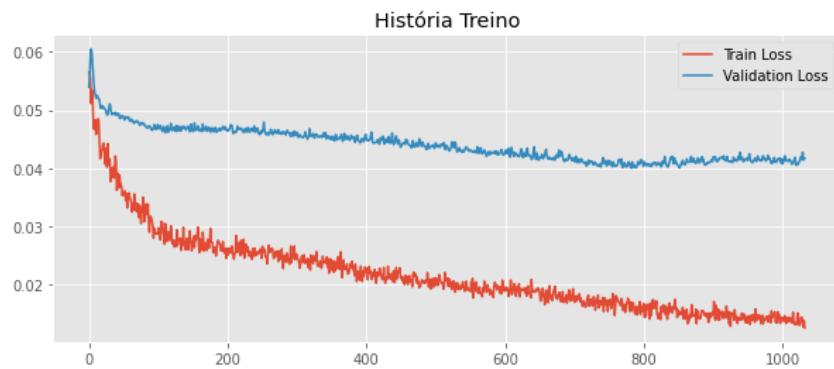


Figura 119: Gráfico de treino do modelo RNN_ENDO_2.

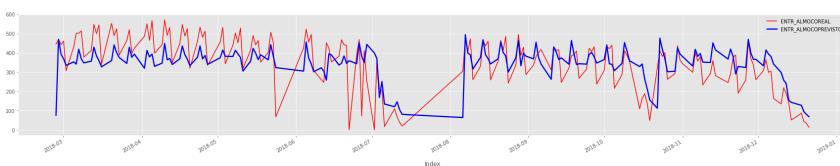


Figura 120: Gráfico de validação do modelo RNN_ENDO_2.

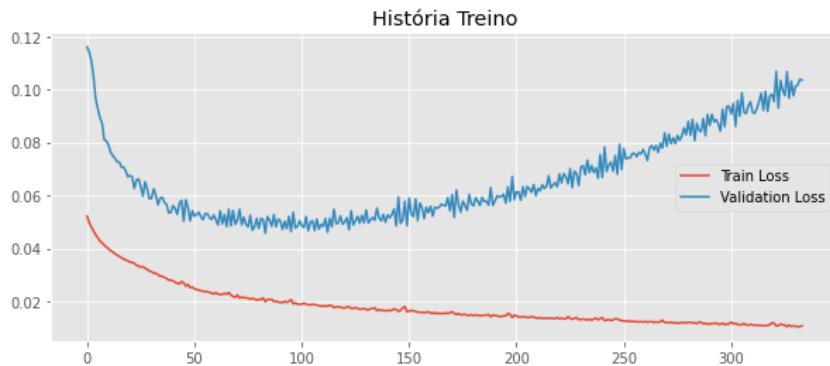


Figura 121: Gráfico de treino do modelo RNN_EXO_1.

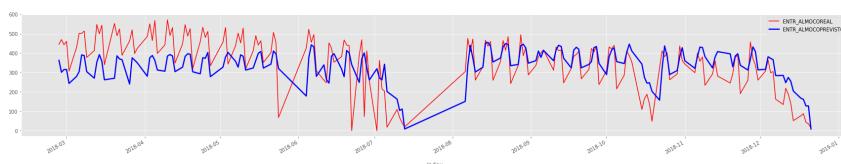


Figura 122: Gráfico de validação do modelo RNN_EXO_1.

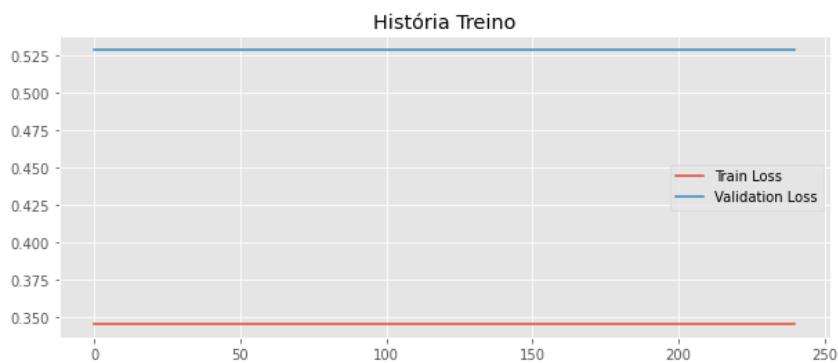


Figura 123: Gráfico de treino do modelo RNN_EXO_2.

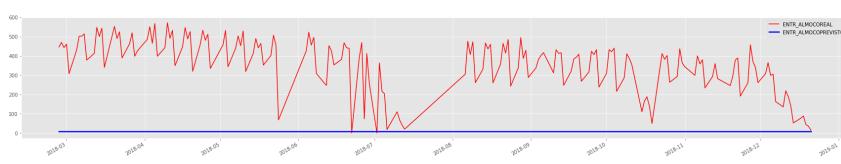


Figura 124: Gráfico de validação do modelo RNN_EXO_2.

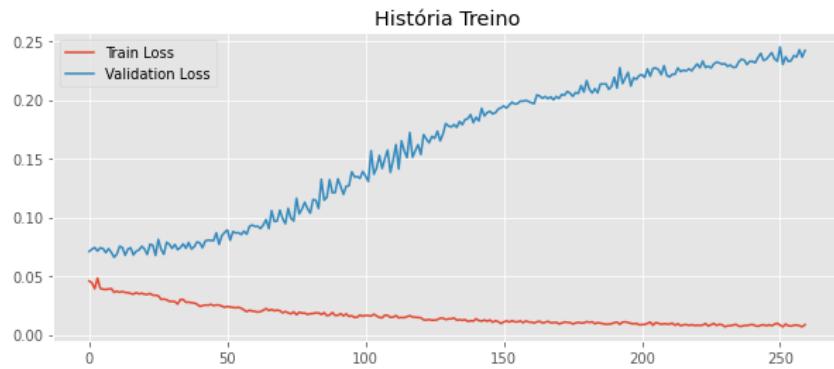


Figura 125: Gráfico de treino do modelo RNN_EXO_3.

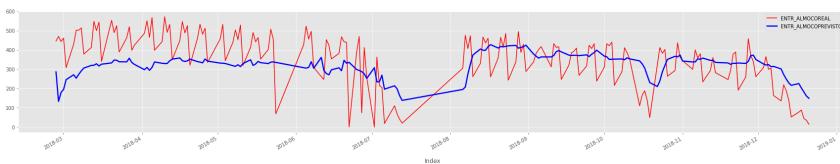


Figura 126: Gráfico de validação do modelo RNN_EXO_3.

Nesta etapa experimental, todos os modelos foram definidos com a mesma topologia da primeira fase. Os modelos endógenos obtiveram melhoria nos resultados avaliados sobre o conjunto de validação. Já os modelos exógenos, apenas 1 obteve melhoria significativa nos resultados, o modelo RNN_EXO_1, porém os outros obtiveram resultados anômalos no plot dos gráficos de treino e validação. Os gráficos de treino e validação podem ser conferidos da figura 115 à figura 126

MÉTRICAS

MLP_ENDO_1	
TOTAL DE REFEIÇÕES CONSUMIDAS	65637 : 184
TOTAL DE REFEIÇÕES PROJETADAS	62781.324 : 184
CORRELAÇÃO (r)	0.7649396511306203
Pi (p)	1.3054591862307924e-36
RMSE	89.32526704475531
TOTAL DE REFEIÇÕES SUB PROJETADAS	-4257.537273406982
TOTAL DE REFEIÇÕES SUPER PROJETADAS	7113.216419219971
ERRO ABSOLUTO MEDIANO	47.23118591308594
ERRO ABSOLUTO PERCENTUAL MÉDIO	3234942340.243013%

RNN_ENDO_1	
TOTAL DE REFEIÇÕES CONSUMIDAS	65637 : 184
TOTAL DE REFEIÇÕES PROJETADAS	65219.613 : 184
CORRELAÇÃO (r)	0.7538102329191316
Pi (p)	5.014834155014456e-35
RMSE	89.73281442880939
TOTAL DE REFEIÇÕES SUB PROJETADAS	-5666.187889099121
TOTAL DE REFEIÇÕES SUPER PROJETADAS	6083.576965332031
ERRO ABSOLUTO MEDIANO	48.55305480957031
ERRO ABSOLUTO PERCENTUAL MÉDIO	4145884336.345922%

RNN_ENDO_2	
TOTAL DE REFEIÇÕES CONSUMIDAS	65637 : 184
TOTAL DE REFEIÇÕES PROJETADAS	64170.63 : 184
CORRELAÇÃO (r)	0.6606462451035008
Pi (p)	1.9184607118463233e-24
RMSE	102.80909876605612
TOTAL DE REFEIÇÕES SUB PROJETADAS	-6437.710876464844
TOTAL DE REFEIÇÕES SUPER PROJETADAS	7904.081573486328
ERRO ABSOLUTO MEDIANO	61.39728546142578
ERRO ABSOLUTO PERCENTUAL MÉDIO	4124092635.50105%

RNN_EXO_1	
RMSE	109.97634427462593
TOTAL DE REFEIÇÕES CONSUMIDAS	65637 : 184
TOTAL DE REFEIÇÕES PROJETADAS	62530.465: 184
CORRELAÇÃO (r)	0.6052852633603059
Pi (p)	8.961737817855194e-20
RMSE	109.97634427462593
TOTAL DE REFEIÇÕES SUB PROJETADAS	-6378.557273864746
TOTAL DE REFEIÇÕES SUPER PROJETADAS	9485.092361450195
ERRO ABSOLUTO MEDIANO	67.952392578125
ERRO ABSOLUTO PERCENTUAL MÉDIO	3590946402.38917%

RNN_EXO_2	
RMSE	373.5611599422393
TOTAL DE REFEIÇÕES CONSUMIDAS	65637 : 184
TOTAL DE REFEIÇÕES PROJETADAS	1656.0 : 184
CORRELAÇÃO (r)	nan
Pi (p)	nan
RMSE	373.5611599422393
TOTAL DE REFEIÇÕES SUB PROJETADAS	-18.0
TOTAL DE REFEIÇÕES SUPER PROJETADAS	63999.0
ERRO ABSOLUTO MEDIANO	380.0
ERRO ABSOLUTO PERCENTUAL MÉDIO	97826180.62563947%

RNN_EXO_3	
RMSE	131.93314687496996
TOTAL DE REFEIÇÕES CONSUMIDAS	65637 : 184
TOTAL DE REFEIÇÕES PROJETADAS	59270.867 : 184
CORRELAÇÃO (r)	0.3675645280265282
Pi (p)	2.8557023726066666e-07
RMSE	131.93314687496996
TOTAL DE REFEIÇÕES SUB PROJETADAS	-6869.128311157227
TOTAL DE REFEIÇÕES SUPER PROJETADAS	13235.260711669922
ERRO ABSOLUTO MEDIANO	101.04127502441406
ERRO ABSOLUTO PERCENTUAL MÉDIO	3493320353.99382%

5.2.3 Teste e métricas

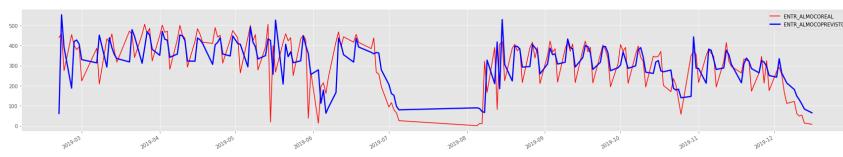


Figura 127: Gráfico de teste do modelo MLP_ENDO_1.

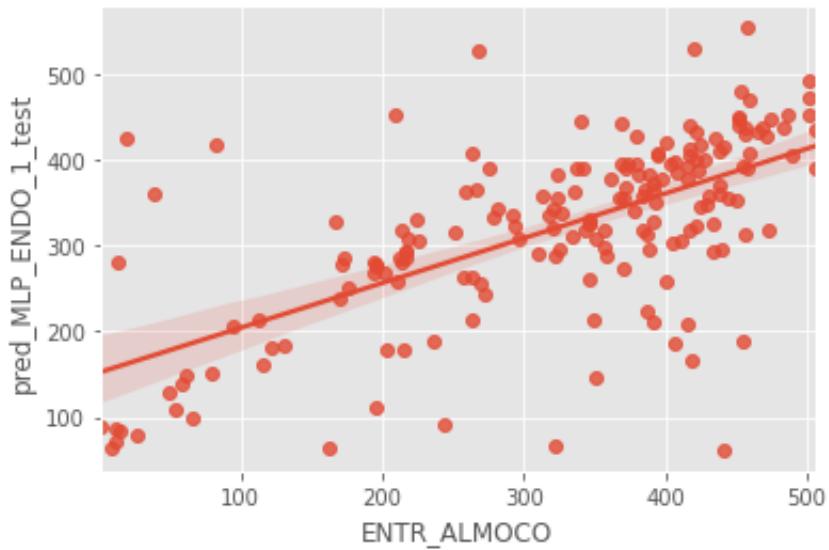


Figura 128: Gráfico de scatter do modelo MLP_ENDO_1.

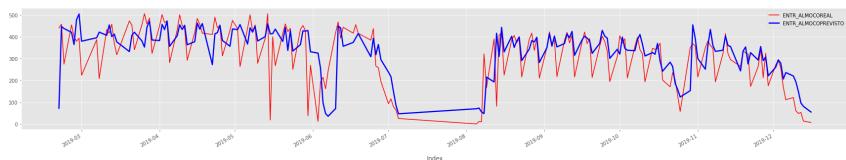


Figura 129: Gráfico de teste do modelo RNN_ENDO_1.

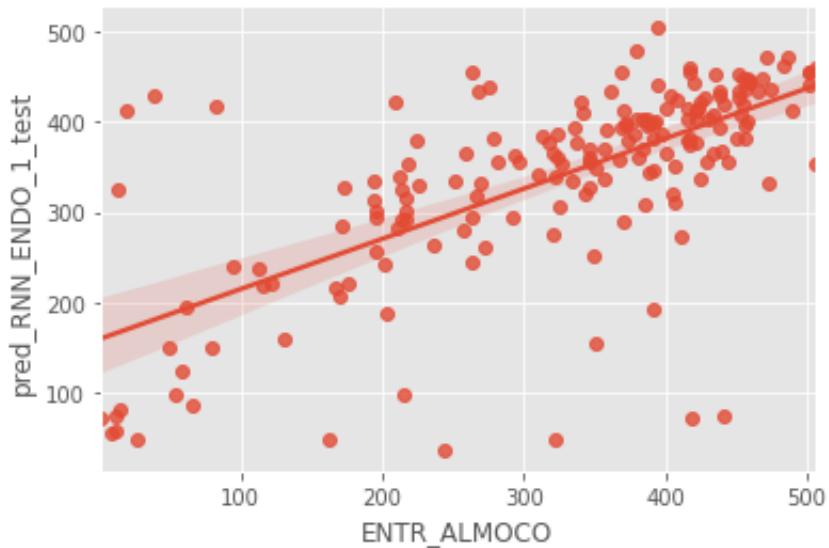


Figura 130: Gráfico de scatter do modelo RNN_ENDO_1.

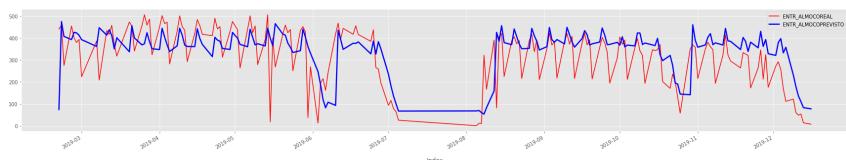


Figura 131: Gráfico de teste do modelo RNN_ENDO_2.

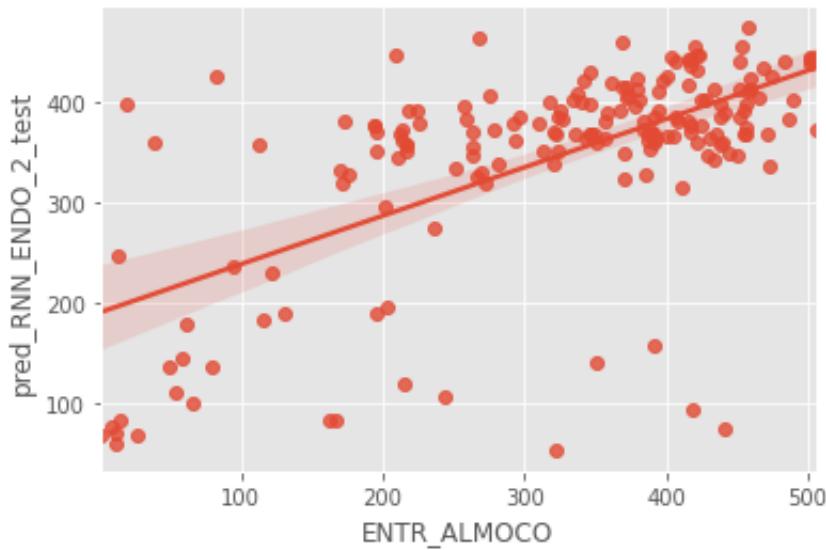


Figura 132: Gráfico de scatter do modelo RNN_ENDO_2.

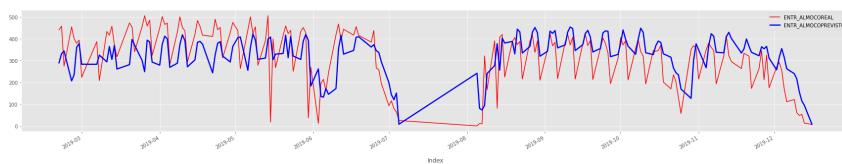


Figura 133: Gráfico de teste do modelo RNN_EXO_1.

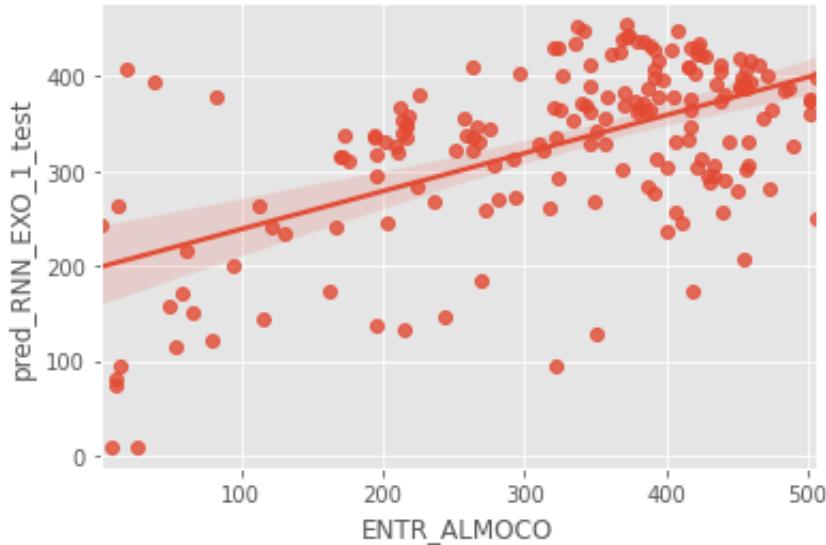


Figura 134: Gráfico de scatter do modelo RNN_EXO_1.

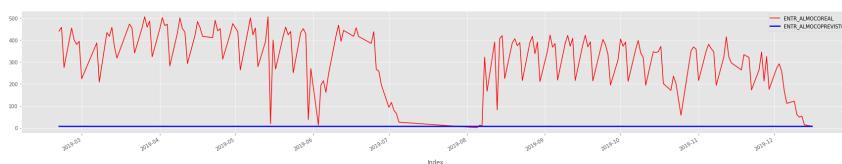


Figura 135: Gráfico de teste do modelo RNN_EXO_2.

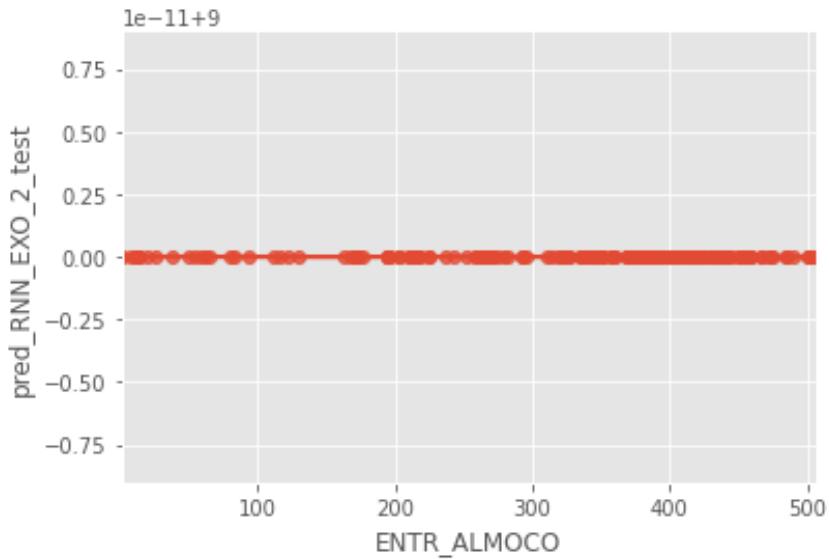


Figura 136: Gráfico de scatter do modelo RNN_EXO_2.

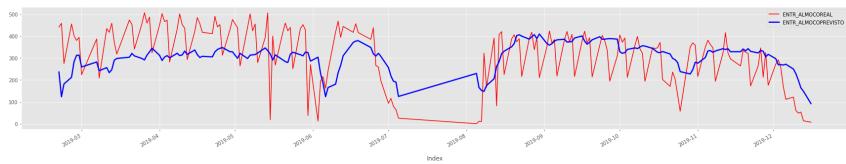


Figura 137: Gráfico de teste do modelo RNN_EXO_3.

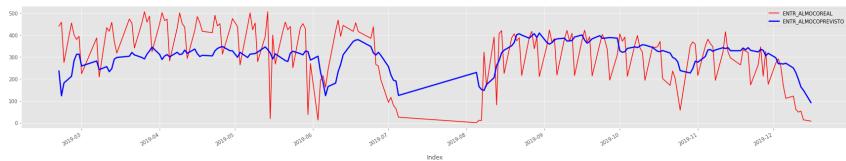


Figura 138: Gráfico de scatter do modelo RNN_EXO_3.

Após o teste de todos os modelos da 2a fase, o modelo que obteve as melhores métricas entre todos, foi RNN_EXO_2, onde o gráfico de previsões e scatter podem ser conferidos na figura 127. Todos os gráficos de teste estão entre a figura 127 à 138

5.2.3.1 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.

5.3 Tabela de resultados

5.3.1 Melhores Resultados

5.3.2 Tabela completa de todos os modelos

6 Conclusão

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

Como estamos trabalhando com um conjunto de dados com todas as features e targets de sazonalidade temporal, é notório que a ordenação dos dados na separação do conjunto em treino, teste e validação deve seguir uma ordem cronológica para os modelos aprenderem com o passado e realizarem previsões para o futuro. O erro na divisão identificado no capítulo de divisão do conjunto de dados, demonstrado com erros na figura 44 e depois corrigido conforme a figura 45 demonstra esta conclusão.

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 75 e 77 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 validação. 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 distancia 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 primeiro split, 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 os 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 (distancia registro anterior e posterior) tem potencial de agregar aprendizado nessa questão, mas é necessário uma diversificação maior do conjunto de dados pois treinamos 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, 31, 32, 40, 47, 50 e 60.

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 35.

CLARICE G.B. DEMÉTRIO; ZOCCHI, S. S. *Modelos de Regressão*. [S.l.]: Departamento de Ciências Exatas, ESALQ, USP, 2011. Citado 3 vezes nas páginas 28, 32 e 34.

Data Science Academy. *Deep Learning Book*. 2020. Access date: 3 oct. 2020. Disponível em: <<http://deeplearningbook.com.br/arquitetura-de-redes-neurais-gated-recurrent-unit-gru/>>. Citado na página 65.

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 5 vezes nas páginas 28, 49, 51, 59 e 62.

HAYKIN, S. *Neural networks, a comprehensive foundation*. [S.l.]: Macmillan, 1994. Citado na página 28.

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 68.

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 7 vezes nas páginas 28, 29, 30, 31, 32, 39 e 60.

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 59.

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 7 vezes nas páginas 28, 49, 50, 51, 59, 62 e 66.

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 34.

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 28, 38, 39, 46, 49 e 52.

MC AI. *Overview.* 2020. Access date: 3 oct. 2020. Disponível em: <<https://mc.ai/complete-guide-of-activation-functions/>>. Citado na página 67.

MORETTIN, P. A. *Análise de séries temporais.* [S.l.]: Editora Blucher, 1987. Citado na página 31.

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 52.

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, 32 e 60.

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, 32 e 60.

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 35.

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>