UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Douglas Diniz Landim

DEMAND ANALYSIS OF UNIVERSITY REFECTORY AT ICT UNIFESP VIA NEURAL NETWORKS.

São José dos Campos, SP

Douglas Diniz Landim

DEMAND ANALYSIS OF UNIVERSITY REFECTORY AT ICT UNIFESP VIA NEURAL NETWORKS.

Monograph presented to the Institute of Science and Technology – UNIFESP for obtaining of the title of bachelor's degree in Computer Science.

Federal University of São Paulo – UNIFESP

Institute of Science and Technology

Bachelor's Degree in Computer Science

Orientador: Prof. Dr. Marcos Gonçalves Quiles

São José dos Campos, SP

October 2020

Douglas Diniz Landim

DEMAND ANALYSIS OF UNIVERSITY REFECTORY AT ICT UNIFESP VIA NEURAL NETWORKS.

Monograph presented to the Institute of Science and Technology – UNIFESP for obtaining of the title of bachelor's degree in Computer Science.

Prof. Dr. Marcos Gonçalves Quiles
Adviser

Prof.
Dr. Daniela Leal Musa

Prof.
Dr. Regina Célia Coelho

São José dos Campos, SP

October 2020

This monograph is dedicated to my parents for their endless support and sacrifices they gave my way along this Journey, to all my university mentors whom have provided me knowledge to develop my skills, given me hopes and spend their precious time for my progress and, above all, those who motivated me providing challenges I dared to go beyond my limits.

Agradecimentos

My Journey My journey through the gradution is characterized by a lot of persistence, difficulties and failings. First, I thank God for nourish me with faith, hope and not letting me succumb to pressure. Despite all the technical content of more than 40 disciplines of my course, what really provide me a vision and a deeper learning was the challenging environment of this University; which combined with the personal difficulties, accidents, health setback, career and family; constituted the perfect whole of challenges that made me stronger and fearless to deal with job market, more confident and braver to achieve my objectives. I also thank my family for always fully supported me, my mentors whose guided me in meetings, discussions and chats even on weekends. And finally, I thank all my university friends, colleagues and staff I met during graduation at UNIFESP.

Special thanks: To Mrs. Musa for being my first Coordinator and Counselor when I joined UNIFESP.

To the professors of differing subjects which constitute my knowledge base of Science of Computer that prepared me for academic and profissional activities whom helped me through the selection processes I was pleased to participate. To professors Reginaldo Kuroshu, Bruno Kimura, Àlvaro Fazenda, Regina Coelho, Arlindo Flavio, Antonio Chaves, Ana Luiza and Otavio Lemos.

To Mrs. Bertini who motivated me within the first studies discipline of System Simulation, which is the theme of this monograph, performing a realization of correlation with resampling of Consumption x Temperature in 2016.

To Mrs. Martins, who teaches Statistics and gave me insights about Theoretical Induction of Predictions with Neural Networks, her doctoral thesis was enriching in theoretical bases of this monograph.

To Mr. Veloso and Mr. Fabio, for introducing me the discipline of Artificial Intelligence but also provide good guidance in all the developing of theoretical foundation in the first part.

To Mr. Quiles, for guiding me in the second part of this monograph, introducing me since the beginning the work with Python, Tensorflow, Scikit Learning, in more specific guidelines for learning the Neural Network. Gratitude for guiding me in the evaluation metrics of models and about the whole experimental methodology, besides the revisions in the final text.

To the team of Data Science and Machine Learning Engineering from 2RP-NET, expert in fraud analysis, data and projects with Big-Data, of which recently I was hired for having these experiences and provide an extended meeting period during our office hour to discuss the results and metrics of this employment, team which is composed of professional Masters and Doctors in Data Science and Machine Learning.

To all others professors, colleagues, students and staff from the Science and Technology Institute – UNIFESP.

To my family for supporting me during this long period at UNIFESP.

Even though discredited and ignored by everyone, I can't give up, because for me, winning is never giving up.

(Albert Einstein)

Resumo

The main objective of this research is the study of sales forecasting methods for the UNIFESP university refectory, to avoid overprojection of demand and the consequences of teachers or students without meals. In a previous investigation, performed as a discipline work, the author used statistical methods to analyze the behavior of meal consumption. In this work, machine learning models were developed, more specifically, Neural Networks perceptron multilayer and nets *gated recurrent units*, with methods of analysis of collected data, preparation and pre-processing of informations, selection and evaluation of the best models and final conclusions.

Keywords: Artificial Neural Networks, Demand Forecasting, Machine Learning, Artificial Intelligence, Perceptron Multiple layers.

Abstract

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

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

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

ICT	Institute of Science and Technology
R.U.	University Refectory
UNIFESP	Federal University of São Paulo
UFV	Federal University of Viçosa
UNESP	State University Paulista Júlio de Mesquita Filho
BDMEP	Meteorological Database for Teaching and Research
RNA	Artificial Neural Network
MLP	Multi Layer Perceptron
GRU	Gated Recurrent Unit
RMSE	Root Mean Squared Error

Sumário

1 Introduction

Among the most accepted definitions of food security is the version coined during the 1996 World Food Summit (SHAW, 2007), which enunciates it as a situation where all people, at all times, have physical, social and economic conditions to access safe, healthy and nutritious food for a healthy and active life. However, according to the World Health Organization (ORGANIZATION et al., 2009) over 1 billion people in the world have a nutritionally insufficient diet and more than twice as many people necessitate micronutrients.

According Webb et al. (2006)food security can be decompose into 3 essential pillars, availability, possibility of access and rational utilization. These concepts are intrinsically hierarchical since the availability does not guarantee access, which in turn does not guarantee its rational use. With the advances in agricultural production and the opening of world economic markets, although it has not solved the problem, it has made possible strides in the first two pillars of food security above-mentioned. Thus, in the first decades of the 21st century, the third pillar of the rational use of resources, has proved to be an important factor to be considered in this new globalized world. (BARRETT, 2010).

At the end of 2019, a highly communicable disease was identified in China's Wuhan province, commonly called the new Coronavirus (COVID-19), which was declared in a pandemic state in March 2020, and has so far infected more than 40 million people and caused 1 million deaths.[1]. Despite global efforts to contain and remedy the COVID-19 pandemic, there are no scientifically proven vaccines or cures and the need for *lockdown* is still pertinent to avoid new waves of contamination. The effects of the pandemic have been devastating in virtually all economic areas, especially in the food sector which is essential for maintaining life and has no room for *lockdown* (GALANAKIS, 2020).

On 22 July 2020, the Brazilian Association of Food Industry and Consumption, held a conference on the impacts of the COVID-19 pandemic on the Brazilian food production chain[2]. In this conference were brought together experts and executives from the agro-industrial sector to analyze changes in consumption behavior in the tertiary sectors that deal directly with the supply of food to the final consumer. From this analysis of the new consumer routine, new forecasting methods are being looked for the current format of consumption and the new demand for these sectors. Finally, methods of production and supply of agribusiness for the tertiary sector

1 https://www.worldometers.info/coronavirus/

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

were also discussed.

Thus, in this chaotic period, when the global labor has reduced approximately 25% during the first months of the pandemic (HUFF et al., 2015), comes up the importance of developing processes that use food resources in the most rational way possible. Neste contexto, uma área que chama a atenção é a de previsão de demandas indústrias or from establishments that deal with perishable food in large quantities, since the food has an expiration date and production on demand can avoid storage for long periods or unwanted disposal of ready-to-eat foods.

In this sense, forecasting demands for meals in industrial or large restaurants, such as university refectory (RU) becomes viable, which meals are offered at affordable prices to the student community of the institution, through subsidies in the amount passed on to students. Generally, (RUs) are outsourced through bids in which the state government provides a percentage of each meal produced for the refectory management company. Due to health regulations, meals prepared but not consumed until the end of the working day should be discarded to avoid contamination and ensure the food safety of the consumer[3]. In this way, the excess production (not consumed) of meals by the (RU's) generates not only an exaggerated expenditure of public financial resources but also anunreasonable use of food resources.

In this context, Unifesp's university refectory in São José dos Campos is a good case study since it sells approximately 90 thousand meals annually, in which each meal has a fixed value of R$2,50 for students and an subsidy of approximately R$9,00 per meal in the last decade (2011-2019). With this, more than 6 million reais were subsidized during the years 2011 to 2019.

One way to reduce this waste is in the development of consumption prediction methods (LOPES, 2008; ROCHA; MATOS; FREI, 2011).Usually, the approaches for forecasting the consumption demand of a university restaurant are summarized in the exploratory analysis of collected data, for example the sales computed in the previous week or month. In addition, external information, denominated exogenous data, can be also considered in the prediction process, such as climate data, annual calendar data, holidays, among other information that may be relevant for consumption estimation.

The prediction of consumption of university refectory chains has already been discussed in earlier works.For example, in the work of Lopes (2008) held at the UFV, a model of a perceptron neural network with 1 hidden layer and 1 output neuron is used, using the previous 5 days as prediction parameters with 3% error over the total consumed, and in the work of Rocha, Matos e Frei (2011) performed at UNESP is also used a neural network model with 1 hidden

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

layer and 1 exit neuron, using only 1 parameter that informs the number of meals of the previous day, and other parameters informing averages of previous days and information regarding the date of consumption, this model obtained an error of 9,5% calculated on the total consumed.

So, this work consists in the application of Neural Network models to forecast the demand for the meals provided at the university refectory of Unifesp in the campus of São José dos Campos. For this, it will be used a set of historical data available in the Unifesp system and other external (exogenous) information that can impact the consumption behavior and its prediction process.

The general objectives of this work include the construction and comparison of Artificial Neural Networks models for forecasting the demand for meals at the ICT-UNIFESP university refectory. Specifically, it has as objectives:

a) Obtain and preprocess data of consumption and sale of meals from the refectory environment and data from the external environment to consumption, such as climate data, aiming their use as input of machine learning models to obtain the consumption predictions at the output of these models;

b) Execute exploratory and descriptive analysis of all data;

c) Build and evaluate predictive models of Neural Networks;

d) Execute analysis on the metrics of predictions of the models pointing their viable characteristics for a consumption prediction.

The rest of this document is organized as following. The chapter 2 introduces the necessary foundation for the understanding of this work; The literature related to the subject is described in chapter 3; The chapter 4 introduces the methodology used in this study; The results are summarized in chapter 5; The main conclusions of this project as well as indications of future investigations are presented in chapter 6; Lastly, chapter **??** presents some additional results generated during the experiments and the implemented codes.

2 Theoretical Foundations

In this chapter a quick review of the main characteristics of time series, as well as of Neural Networks, especially those that will be applied to the problem of interest, called Multilayer Perceptron Networks (MLP) and Gated Recurrent Unit (GRU).

2.1 Demand Forecasting Methods

Demand forecasting can be defined as a process of searching for information that produces sales of a product or a set of products.

In Moreira (1998) are defined as methods of demand forecasting that use such information to estimate future demand, and can fit this definition from subjective and intuitive methods to mathematical and computational methods. In the same article, the forecast models are classified into two major groups: qualitative methods and quantitative methods.

In Junior (2007) has been made the evaluation of several methods of demand forecasting, which are schematized in the Figure 1.

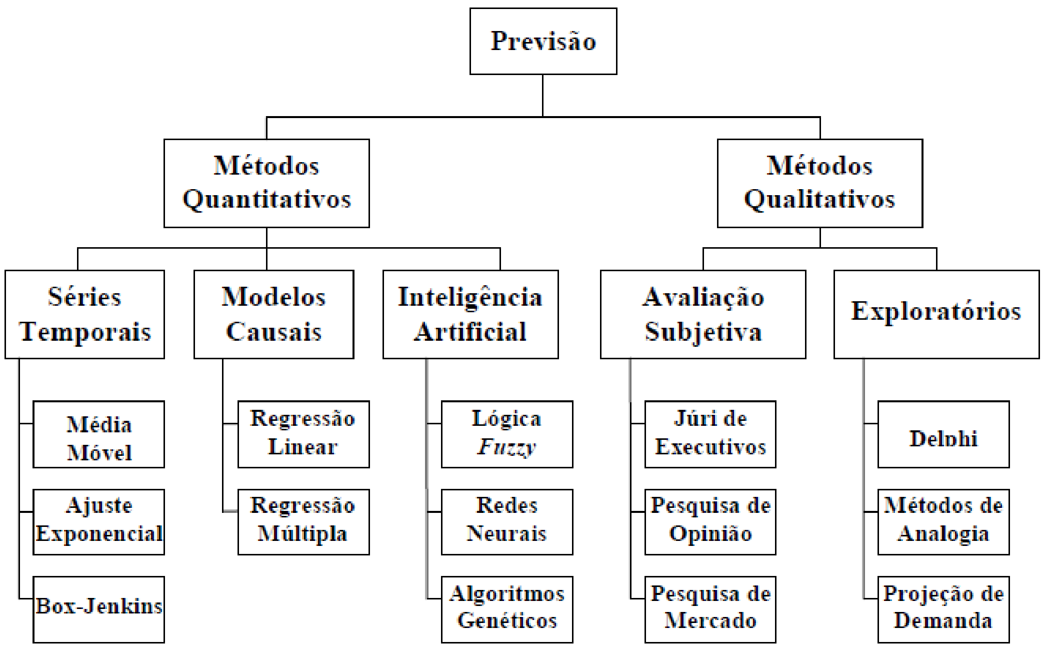

Figura 1: Demand forecasting methods.

Source: (JUNIOR, 2007).

Quality methods are defined in (JUNIOR, 2007) as a judgment of the data exposed without processing analytics, such as grouping and classification of data. These methods do not provide new numerical information or predict models.

On the other hand, quantitative methods are defined, in the same study, as being analytical and based on mathematical models for performing predictions. These methods analyze behavior patterns from a data historical, aiming to predict future behavior.

In Junior (2007) is also mentioned that the data collected for forecasting models, when graphically projected, show behaviors that can be generalized in a subjective way by managers of the data. For all cases, data analysis is necessary to select the parameters of the demand to make objective predictions. Yet, only data analysis may be insufficient, and if performed with incorrect criteria it may compromise the conclusions of the studies.

Data analysis is a large process that aims to treat data from its acquisition and preprocessing until its interpretation using complex mining tools. In different steps, diverse mathematical, probabilistic, statistical, computational and heuristic techniques are involved. In Section 2.1.1 will be commented the main characteristics of the time series, which will be the forecasting method applied to the practical problem of interest.

2.1.1 Time Series

According to Morettin (1987), a temporal series is a set of observations ordered in function of time, commonly the same, presenting a serial dependence among itself. Also it can be defined as a realization of a set of random variables $X = \{x_1, x_2, \ldots, x_T\}$, ordered in time, where T represents the length of the series, as is made in (DAVILA, 201-). The correlation between the variables is usually described by their collective distribution function or in other cases by the media and covariances.

Among the objectives of time series analysis, the following can be emphasized: i) describe the behavior of the series identifying trends and variations, ii) analyze and mold the dependency between the observations, and iii) fmake forecasts of future series values.

Time series have a deterministic component and a random component, they can be continuous or discrete, depending on the type of observation of the variables. They are called stationary when the properties of mean, variance and covariance are kept in time. The trend references the rate of growth or decrease, and can be linear, exponential or cushioned. Yet the oscillation of the trend is known as the cycle. In addition, they can present seasonality, in other words, exhibit behavior that has a tendency to repeat itself in a certain number of time periods.

Examples of each of these properties, as well as techniques applied to study and correct each of them can be seen in (EHLERS, 2009). These features, together with the random component study, provide information of interest for practical applications.

Because of the objective of this work, temporal series will be applied for events with seasonality in order to estimate forecasts. The techniques to do this are using regression methods and some practical applications can be seen to include forecasting of realized power consumption, as made in Almeida (2013), Ruas (2012) and Silva (2010), and of the demand forecast of cosmetic products, as it can be seen onJunior (2007).

2.2 Artificial Neural Networks

Artificial Neural Networks are tools with multidisciplinary basis, once they are nourished by knowledge of neuroscience, mathematics, statistics, physics, computer science and engineering (HAYKIN, 2007), and are part of the great area of knowledge called Artificial Intelligence (AI), term pointed out in (KAPLAN; HAENLEIN, 2019).

In a nutshell, it could be defined that "artificial intelligence is the branch of computer science that deals with intelligent behavior. (LUGER, 2004). Artificial Intelligence systems seek to solve functions and problems inspired by two human characteristics: the capacity for abstraction and learning from the error.

In the context of Artificial Intelligence, a neuron is a unit considered fundamental for information processing, and Neural Networks are sets of artificial neurons interconnected through connections, logical and mathematical functions (HAYKIN, 2007). Neurons of a network are capable of processing multiple values of inputs and react quickly producing a related response to these inputs, simulating the behavior of the human brain.

Inspired by a search for a computational model of the biological neuron, the first artificial neuron model, called MCP, was proposed in the article *A Logical Calculus of the Ideas Immanent in Nervous Activity.* (MCCULLOCH; PITTS, 1943), an illustration adapted for (LEMOS, 2003) of this model can be found in the figure 2.

McCuloch was a psychiatrist and neuroanatomist and spent about 20 years reflecting and studying about the representation of the nervous system, in 1942 he invited Pitts who was a mathematician to be part of his researches.

The structure of the artificial neuron reacts to an input vector and synapses are represented by numerical weights. A transfer function also called the activation function, measures a linear combination of the input values and the synapses weights, determining if or not the neuron is activated depending on the value obtained. If the neuron is activated an output value of 1 is emitted, otherwise an output value of 0 is emitted.

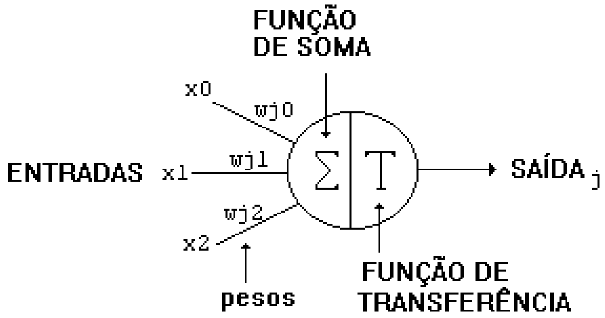

Figura 2: Artificial Neuron.

Source: (LEMOS, 2003)

The whole functionality of this model is then reduced to reply if the weighted sum received is greater than an established numerical value. Although, associated to this neuron was not proposed an automatic way to adjust the weights, that means no learning algorithm was given to train the neuron. This problem was later solved by Perceptron's formulation.

In Haykin (2007) in Chapter 1.9 entitled *Historical Notes*, is introduced in more detail the fascinating history of the development of Neural Networks from the initial conception of biological neuron studies until complex networks of supervised learning known as *Vector Support Machines*.

2.2.1 Perceptron

The proposed neuron model in Mcculloch e Pitts (1943), even though it simulated a biological neuron and solved some logical and mathematical tasks, it did not satisfy the main objective of Artificial Intelligence: The capacity of learning. To be able to use this structure, it was necessary to know how to adjust the weights of the inputs, which was not a trivial problem in many cases.

The first neuron with a learning algorithm was suggested in Rosenblatt (1958) and was named Perceptron. In this study, the weights of the connections are adjusted autonomously with the introduction of associated weights and a Bias value, in order to search for an autonomous recognition of standards. In the figure 3 a schematic of the operation of the structure is presented.

But in Minsky e Papert (1969) was proven that because of the learning model limited to a linear combination, the perceptron could only solve linearly separable problems. In the figure 4 two simple problems are presented, one that perceptron can solve *(a)*, and another does not *(b)*.

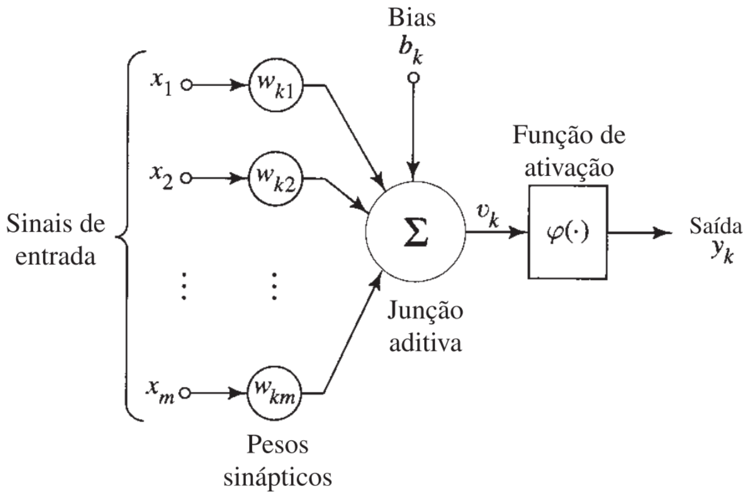

Figura 3: Artificial Neuron Perceptron.

Source: Haykin (2007)

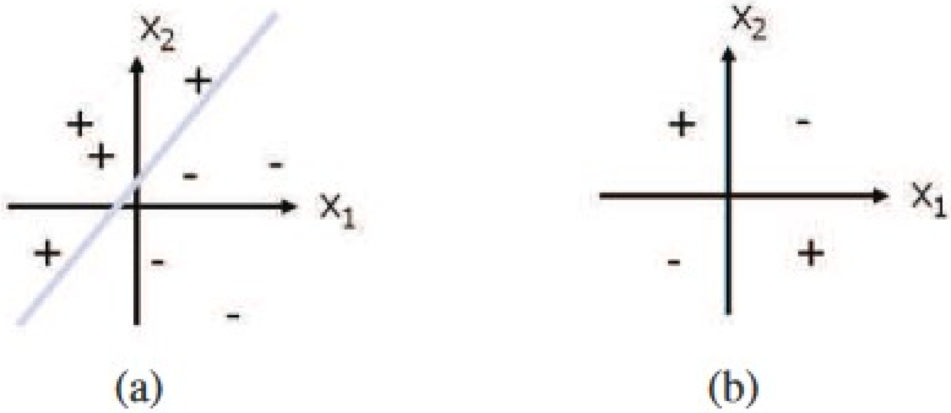

Figura 4: Problem linearly separable (a) and not separable (b).

Source: (MARIANO, 2014).

Almost two decades later, it was released in (RUMELHART; HINTON; MCCLELLAND, 1986) the first model of a Neural Net, called a perceptron net, applying the training by linear combinations to a set of interconnected perceptrons. This approach allowed solving more complex problems through a combination of solutions.

The network had just one input layer, one output and an activation function φ (HAYKIN, 2007). Perceptron network activation function could still be linear or non-linear. In the figure 5 some commonly used activation functions are illustrated.

In Almeida (2013) the learning process of the perceptron network is analyzed in a supervised way. During this process, the structure learns to relate an observed set of input variables in the network to one or more expected output values named as real values or truth values.

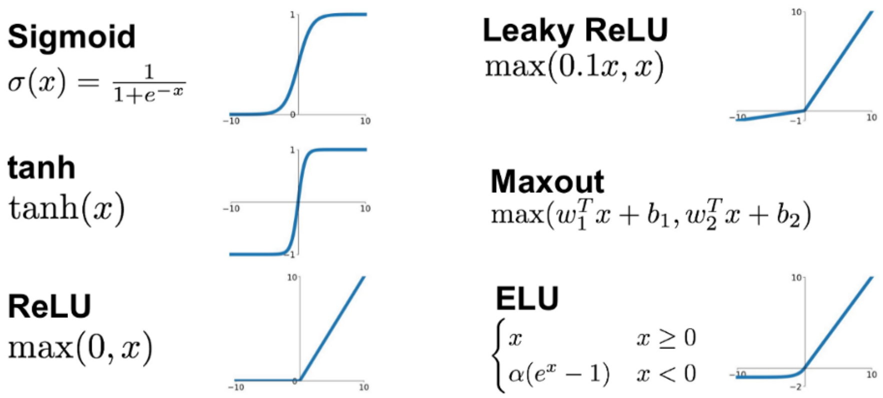

Figura 5: Examples of Activation functions.

Source: (MC AI, 2020)

Then the learning results are analyzed by comparing these values with the values generated by the perceptron net over the same set of data, and from this comparison is calculated the error measurement of the training.

A criterion for stopping training algorithm is to check whether the error is acceptable or not. If positive, the neural network maintains the values of the synapses weights obtained at the time. If not, a new training season is made trying to adjust the weights to obtain a smaller failure. The other criterion for stopping training is to reach a maximum number of training periods allowed. The adjustment of weights is called the rate of learning.

The next step in developing Neural Network models is related to the topology that determines the amount of perceptrons in the network and how they connect, generating multi-layered networks.

2.2.2 MultiLayer Network Perceptron (MLP)

The possibility of combining two or more layers of perceptrons was given by the use of an output signal combiner perceptron. With it the Neural Networks are scaled up to several interconnected perceptrons columns. Each column is denominated a hidden layer of the Neural Network. The last layer must have the number of perceptrons corresponding to the number of desired outputs. In the figure 6 a Neural Network with 2 hidden layers is presented, whose output layer has 3 neurons.

In Braga A. de P.; Carvalho (2000) is postulated that through an intermediate layer it is possible to approximate any continuous function and further that two intermediate layers are enough to approximate any mathematical function. If the use of two or more layers can facilitate the training of the network, it becomes unviable to use a large number of these, since in each hidden layer the error is estimated from the error in the previous layer, which generates loss of

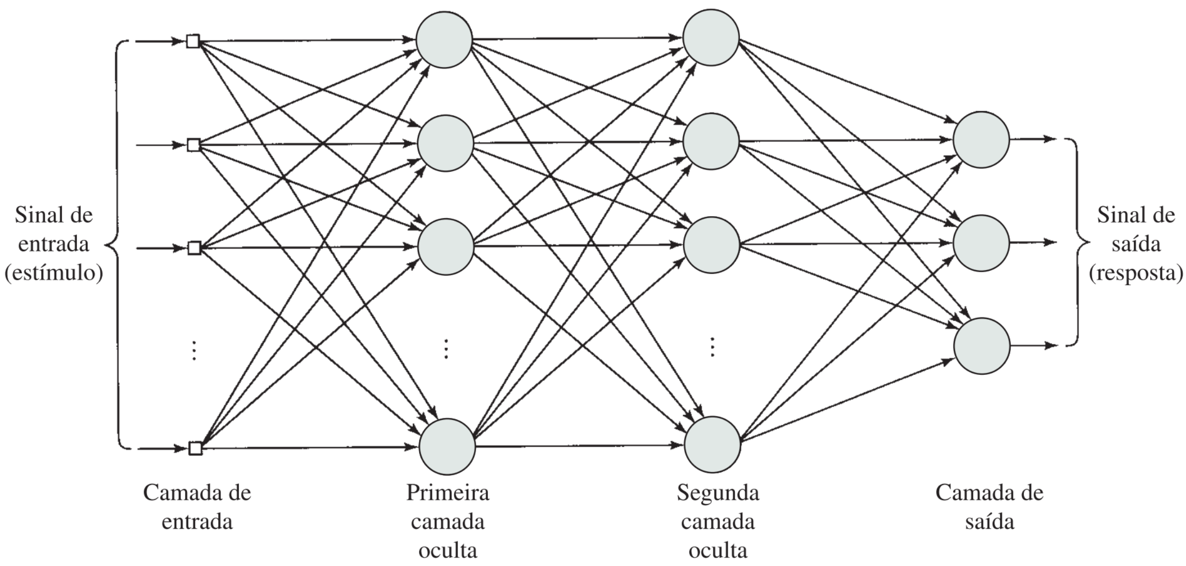

Figura 6: MultiLayer Network Perceptron.

Source: (HAYKIN, 2007)

accuracy.

In practical applications, it has been seen that in some cases, the capacity for abstraction and recognition of Neural Network patterns overcome human capabilities. In other cases a network may not produce an expected response by incorrectly solving a problem, as well as the human brain, because of learning limitations or training failures.

To train an MLP network in a supervised way, the input data set must be divided into two subsets, one of *training* and another of *validation*. These subsets can be separated with various techniques. Thus, for example in Data Science Academy (2020) a heuristic form of set separation in random order is presented, with 70% of the data for training and 30% of the data for validation.

About the maximum number of training attempts allowed, very long training sessions tend to memorize weights of the values observed in the training data. This implies a loss of network generalization capacity, resulting in a difficulty to evaluate entries outside the training data. This phenomenon is known as *overfitting*.

There is also another way of training an MLP network, called cross-validation, which was presented in (KOHAVI, 1995). This technique consists of exchanging the sets of training and validation at different times of training. In that case, the measurement of the validation error goes through an evaluation process taking into consideration the number of periods. By evaluating the average quadratic error of both sets it is possible to detect the beginning of the *overfitting*. For this reason, the optimal stopping point of the training is associated with the lower limit of this mean square error in the validation set illustrated in the figure 7 as an early stopping point for *overfitting*.

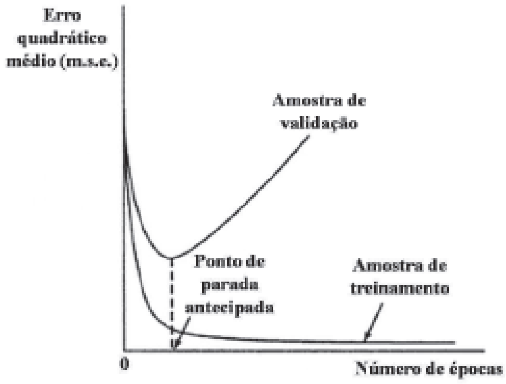

Figura 7: Ideal stop point of cross validation.

Source: (HAYKIN, 2007)

2.2.3 Multiple Layer Perceptron Network with *Backpropagation*.

The learning method for multi-layered Neural Networks known as *Backpropagation* was presented in (RUMELHART; HINTON; WILLIAMS, 1986), as an abbreviation of *backward propagation of errors*, in Portuguese, retro-propagação de erros. In this method, the training is conducted in two phases:

Feed-forward An input vector with known output vector is presented to the neurons of the first layer, and an output vector is calculated according to the natural flow of operations in the network.

Feed-backward The error gradient is calculated to obtain information that induces a decrease in function, given by the opposite direction to the gradient. With this, the weights of all layers are updated, starting with the last one and following the inverse flow of the network.

These two phases are schematized in Figure **??**.

In this method, as the error gradient is calculated from the lower to the upper layers, its norm decreases with exponential speed. This makes that in the layers closest to the entrance the weight adjustments are small, making the learning in them slower. This problem is known in the literature as *vanishing gradient problem*. Usually the values of the learning rates stay between 0.2 and 0.8.

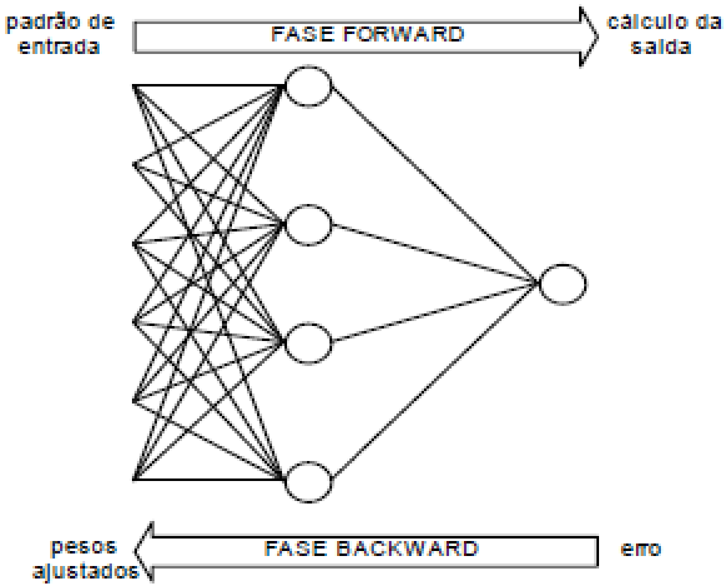

Figura 8: Training phases of the MLP-Back-Propagation

.

Source: (ALMEIDA, 2013)

In Neural Network training MLP with *backpropagation*, validation only takes place with the stage of *feedforward*, obtaining the quadratic errors of the output layer with the validation data observed.

As the ideal stopping point is a lower limit, the same is only discovered when it is exceeded after a few training periods, since in practical procedures the obtaining of an error is oscillatory and by which it is necessary to keep saved the parameters obtained during these periods.

Weight readjustment optimizer There are several algorithms known that optimize the convergence of the readjustment of the weights in the training of *backpropagation*, uch as those mentioned below. The 'Momentum' optimizer speeds up the readjustment of the weights in search of minimum global errors, and 'RMSProp' prevents the search in the direction of oscillations. A third optimizer, called 'Adam' for the abbreviation 'Adaptive Moment optimization', combines these two characteristics. For the 'ADAM' algorithm, the learning rate can be arbitrated but inBrownlee (2020) it is mentioned that a constant with a value of 0.001 has produced positive results in prediction problems.

In the figure 9 is compared the behavior of some optimizers. In this figure it is possible to see that the lower the cost of training, the greater the speed of convergence to the ideal readjustment of weights. Also, it is noticeable the computational advantage of 'ADAM' compared to other optimizers when increasing the number of iterations.

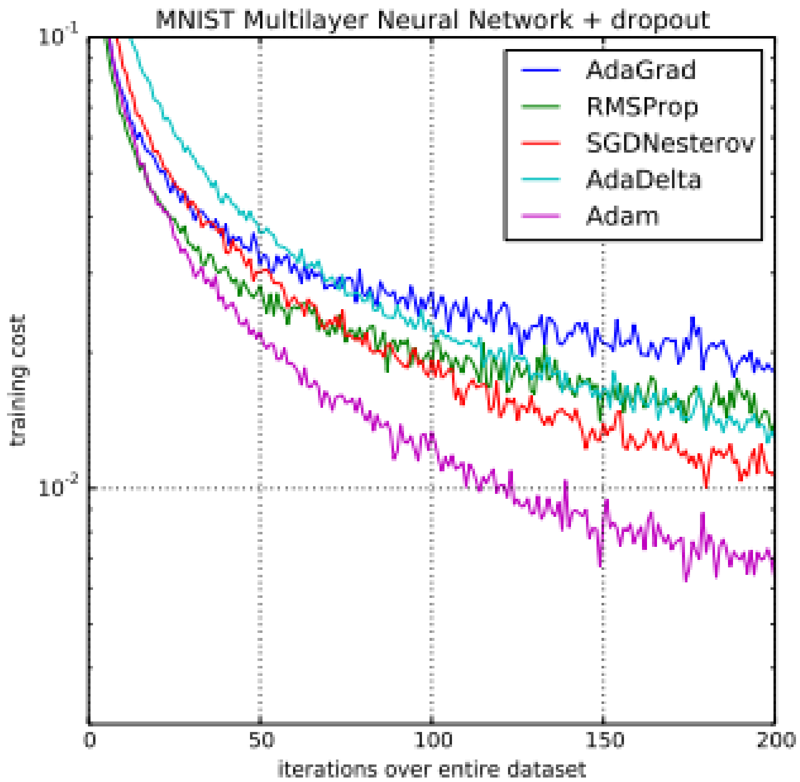

Figura 9: Behavior of optimizers for MLP trained with *Backpropagation*.

Source: (BROWNLEE, 2020)

2.2.4 Recurring Networks: The GRU model

The networks GRU, abbreviation of *Gated Recurrent Unit*,were first introduced in (KYUNGHYUN et al., 2014),being an adaptation of LSTM networks (Long Short-Term Memory).

The LSTM networks were presented in (HOCHREITER; SCHMIDHUBER, 1997), and use memory blocks called cells, which allow certain information to be kept on the network. The manipulation of information is made by *gates* (by which the procedure is often called *gating*). For these networks, there are three types of gates: i) forgetting gate, to remove information that is no longer useful, ii) input gate, to add useful information to the cell condition, and iii) output gate, to extract useful information from the cell condition.

LSTM networks allowed the resolution of more complex problems, but still had the problem of dissipating the gradient by which the memory could not maintain information of long sequences using the term of short-term memory.

The GRU recurring networks solved this problem by changing the use of cell state to a hidden state with two new gates. These gates, called *update gate* and *reset gate* determine which information should be passed to the output and can be trained to maintain long sequence information without being dissipated of the values.

In (Data Science Academy, 2020), these gates are cited as the useful structures for solving prediction problems. In the Figure 10 are presented two network models, an LSTM and a GRU, indicating the gates in each one.

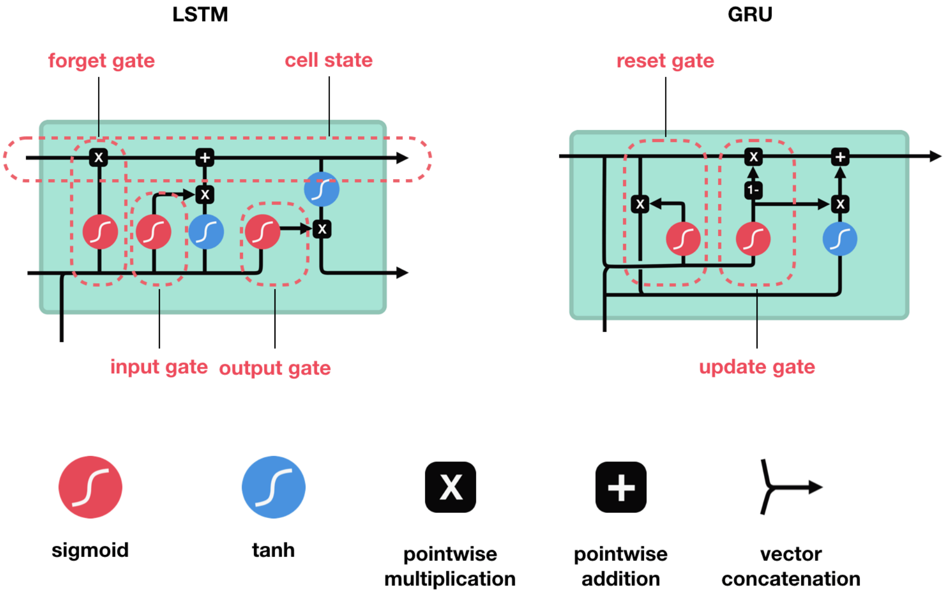

Figura 10: Architecture of the GRU model.

Source: (Data Science Academy, 2020)

The GRU networks then allowed the solution of problems with long sequences of data, solving the problem of short-term memory. So the biggest difficulty that could still arise in supervised trials was due to overfitting, by which another tool was proposed to prevent the network from memorizing beyond what was desired: the method *dropout*.

2.2.4.1 Dropout

The *Dropout* method has been introduced in (HINTON et al., 2012), a term translated into Portuguese in the literature as abandonment, and proposes the temporary removal of some cell from the network.

In (SRIVASTAVA et al., 2014) has been applied *dropout* was applied in the training of a neural net, in which, in each training period some cells of the entry layer are randomly disconnected and some of the hidden layers are all reconnected at the end of the training period. Thus, in each training season only a sample of the data is processed by a subset of the hidden cells.

This method seeks that the randomness of the choice of cells in each period induces a reduction in the dependence between them in the process of adjustment, making each unit generate patterns that do not depend on those learned by others. At the time of the test, all weights are multiplied by the probability that your cell has been turned off.

In (SRIVASTAVA et al., 2014) successful network training results are obtained with *dropout* when in each season 50% of the cells are disconnected in hidden layers and 20% of the cells in the entrance layer.

3 Related work

This chapter describes the main related projects found in the literature referenced in this work. The first section mentions the first literary reference found in the execution of this work that performs comparisons of demand forecasting methods. The second section mentions research related to the theme of this work, the prediction of demand in university refectory, containing studies related to a part of the methods performed in this work which are the models of MLP neural networks and the differential of this work in relation to others is the inclusion of modern recurrent neural networks called GRU based on the topic 2.2.4. The last section mentions the largest volume of references found during general demand prediction surveys, and which did not correspond to the topic of this monograph.

3.1 Comparison work of demand forecasting methods.

Junior (2007) develops a comparison between stochastic methods (Exponential Smoothing Method, Box-Jenkins models) and machine learning models (Neural Networks), illustrated in Figures 1, which are used to forecast the demand for cosmetic products distributed in time series. Among the Neural Networks, we find networks such as *feedforward* with the *backpropagation* training algorithm that was the main focus of the R.U. prediction project at the Federal University of Viçosa and the State University of Paulista Júlio de Mesquita Filho that also underpinned part of the development of this prediction work at ICT Unifesp. In this work of the author, we also analyze several predictive performance measures and make a final comparative analysis of these measures between the cited methods.

3.2 Forecast demand in university refectory

In the statistical study by Landim (2016), In the study by Dougkf, was analyzed the correlation between temperature and meal consumption on sales days of the university restaurant of the ICT-Unifesp campus, where the data contained only a small sample of sales of the second half of 2016. Because of low volume of occurrences, the data was submitted to re-sampling via bootstrap. According to the sample graphs, it was identified that the correlation shown in the graphs of the first half of the semester and the total semester formed bimodal distributions. However, in the second half of the semester it formed a unimodal distribution. The conclusion was that other variables and other analysis models should be used for this demand forecasting.

Lopes (2008) makes the same study of this scenario of ICT-Unifesp applied at the Federal University of Viçosa (UFV). In this study, the data used were only the university refectory's sales history, and no environment variables were collected such as temperature, precipitation, number of students enrolled, etc. The algorithm used was *Traincgp (Conjugate gradient back-propagation with Polak-Ribiere updates)* in software Matlab. This algorithm does not involve the calculation of second derivatives of variables and converges at least the quadratic function into a finite number of iterations as quoted by the author. For each node of the neural net, the day of the week (such as Monday, Tuesday, Wednesday, Thursday and Friday) and each layer of this net using the previous 5 days for each node (the previous 5 seconds, 5 Tuesdays and so on) were then considered and, finally, a model was obtained by the net that presented a maximum error of 3. The neural net applied in this work is represented in Figure 11.

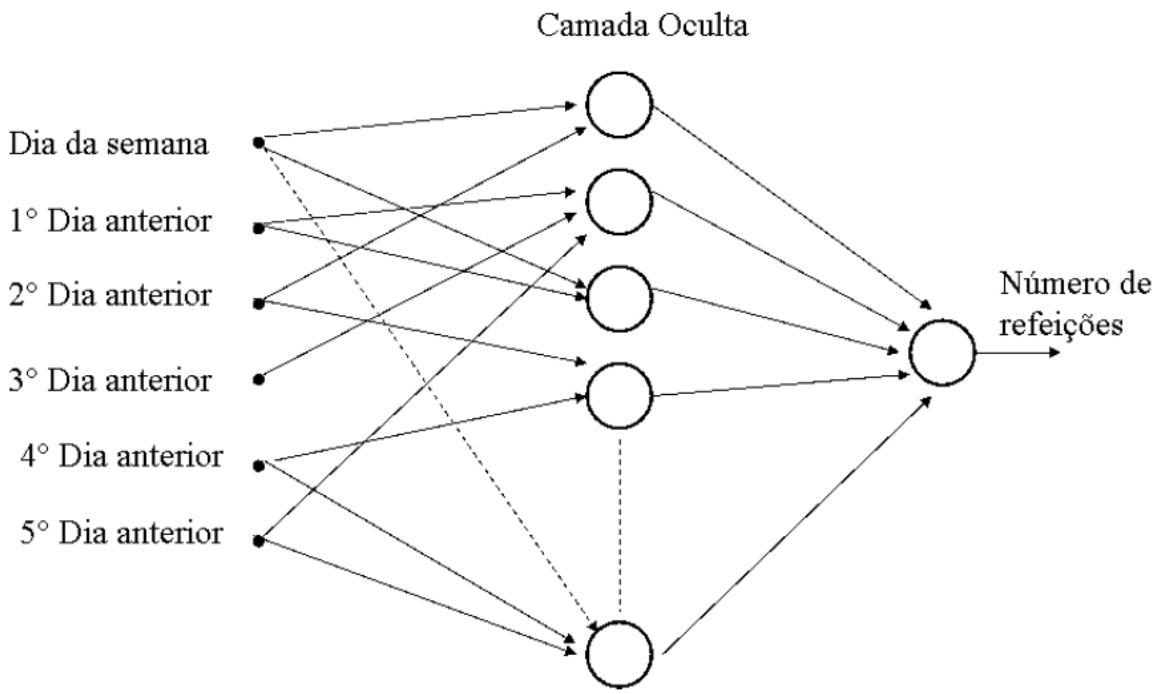

Figura 11: Multiple Layer Perceptron Neural Network.

Fonte:(LOPES, 2008).

Rocha, Matos e Frei (2011) also performs the demand study at the Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP) university restaurant, again with the methods of artificial neural networks with backpropagation and using only as data source (the numerical history of sales made), and other intermediate variables obtained from this, as means of subset of observations (Monday averages). The only environment variable collected was the number of holidays close to the sales observation. In the study of the total number of days analyzed, it can be seen that in 73% (187 days), the simple average method provided a greater error in relation to RNA, which in turn caused a greater error in the remaining 23% (69 days). In the case of less waste, it can be observed that RNA presents errors greater than 50 meals in 13 days, while the simple average method presents errors greater than 50 meals in 58 days, concluding then that the RNA method was much more efficient than the simple average calculation used by the administration of the university restaurant. The Figure 12 shows a Neural Network scheme applied in this research.

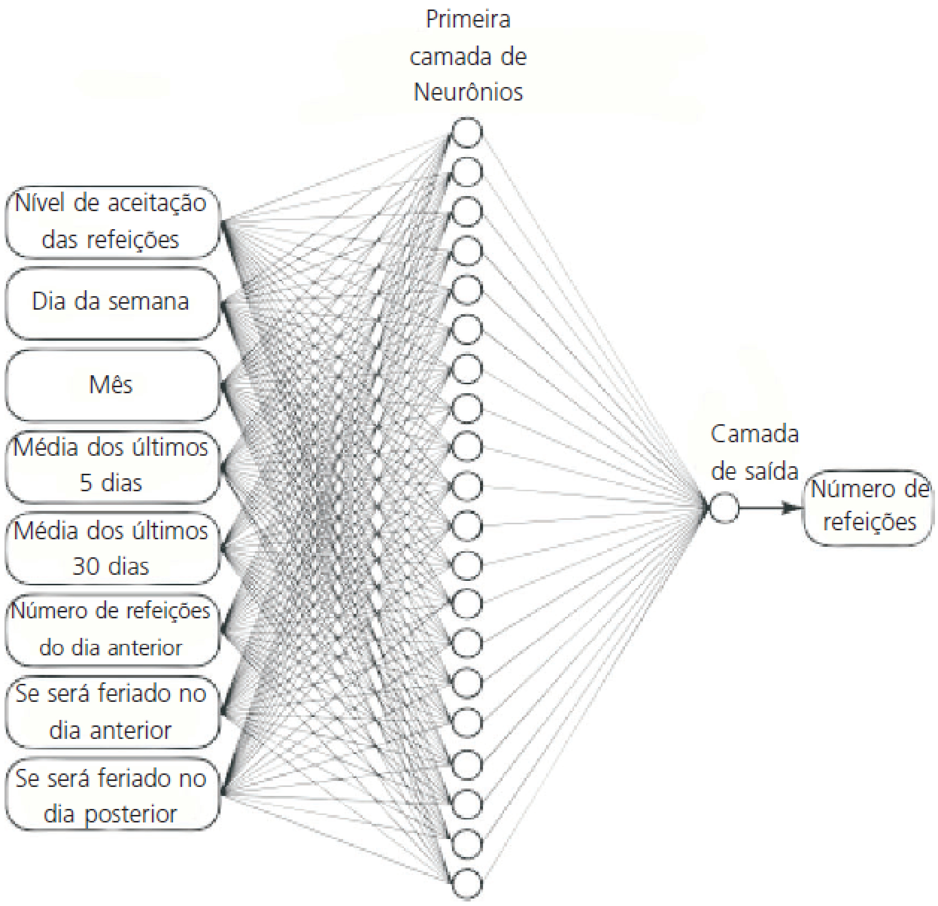

Figura 12: Multiple Layer Perceptron Neural Network

Fonte: (ROCHA; MATOS; FREI, 2011)

Both the model presented in (ROCHA; MATOS; FREI, 2011) as in (LOPES, 2008) have a single hidden layer.

3.3 Demand forecasting in other scenarios.

Ruas (2012)makes an analysis of demand forecasting for electricity in the state of Paraná, between the years 2004 and 2006, using Artificial Neural Networks and support vector machines. Although it is not the same example of the ICT-Unifesp university refectory scenario, we have the distribution of the consumption data collected as a time series. In this research for forecasting electricity demand a partially recurring Elman network was used, which allows the prediction of a step forward. To be possible to make the forecast for several points ahead it is necessary to use the values already predicted, that means the output of the network, as inputs to it.

Almeida (2013) analyzes a similar scenario of electric power demand, but using demand forecasting techniques with Artificial Multilayer Perceptron type Neural Network combined with fuzzy logic that allows placing temperature variables (among others) in a set of rules that impact the problem.

Silva (2010) also applies Neural Network techniques to predict electric power demand, with the study of climatic variables, but through a SOM MAP model - (Self-Organizing Map) which is a type of neural network developed for pattern recognition. Despite being an unsupervised model, the model is ideal for organizing the main impacting and disposable variables in the forecast. The sound map used by the author presents the data associated to his neurons so that similar patterns are found in contiguous neurons, having a topological organization. That way it is possible to extract abstract relations between the variables of the data vector through their position in the component maps, which through a color scale show the amount of a specific variable in each neuron of the map.

4 Methodology

This chapter describes the experimental methodology of this study, which consists of the following steps:

- Collection of endogenous and exogenous data.
- Transformation of each endogenous data record (the consumption and sales data), into a time series with an interval of previous five days.
- Exploratory analysis of endogenous and exogenous data sets with the data set to be predicted.
- Construction and training of exclusively endogenous and mixed models, duplicated in two experimental phases with different time domains.
- Comparative analysis of the model results.

4.1 Field of Study

The field of study of this project is the university refectory of the Science and Technology Institute of Unifesp (ICT-Unifesp) in São José dos Campos. The ICT campus was inaugurated in 2007 aiming to supply the scientific and technological demands of the Vale do Paraíba region. The Figure 13 presents a common day of use of the physical space of the university refectory in ICT.

Figura 13: ICT-Unifesp University Refectory

4.2 Data description

This study will make use of two types of data, endogenous and exogenous, which are described below.

Endogenous data: are the data of the prediction domain in this case.For this problem, are the daily amounts of meals consumed at lunch and dinner at the University Restaurant (RU) of ICT-Unifesp, quantified daily by the number of students passing the access point, the turnstile. Also endogenous data are considered the daily amount of meal tickets sold by the Refectory. In both cases, the information is taken from the class days. This data are transformed into Neural Networks input, in time series format.

Exogenous data: are all other data outside the prediction domain. For this problem, they are parameters derived from the dates of observations, such as the categorical data representing the day of the week (Monday to Friday), and the climate data.

4.3 Data acquisition and treatment

The data are obtained from two different sources, the endogenous data being entirely provided through the information technology sector of ICT Unifesp, part of the exogenous data from the date of registration of the endogenous data collected, and the remaining part of the exogenous data, the climate data, are obtained through a weather station closer to ICT Unifesp, located in the city of Taubaté-SP.

4.3.1 Endogenous data

The historical data of consumption in the restaurant were taken from the current database of subsidized meals at Hospital São Paulo, which manages the data of the cafeterias of all Unifesp units. Only a few authorized employees have access to the institution's database, so to collect such data the present work obtained an authorization with the ICT-Unifesp campus management. Consumption data were only requested for undergraduate students, since the database still contains the consumption information of teachers, graduate students and visitors, but of course with less relevance in quantitative terms. In addition, the consumption pattern of these other strata of the academic environment may influence the process of predicting demands by bringing different trends. In table 1.

After the collection, the Refectory's consumption data were transformed in a process of approximation by a temporal series, for an interval of five days, and in each sales record five new attributes were added containing the past values of this same attribute in an interval of five past days. This process adapts the data set to the process of memorizing the entries, structuring

Tabela 1: Original format of the original data obtained by the university refectory

DATE	(12/19/2017)	(12/18/2017)
BREAKFAST SALES	0	0
LUNCH SALES	24	71
DINNER SALES	0	0
MEAL SALES	24	71
TOTAL SALES	24	71
BREAKFAST ENTRY.	0	0
LUNCH ENTRY	42	70
DINNER ENTRY	3	24
TOTAL ENTRY MEAL	45	94
TOTAL ENTRANCE	45	94

the compatible data reading format in the applied Neural Networks models.

The table 2 exemplifies the new structure of a restaurant data record, with a time interval of five days prior. It is noted that the consumption value of the date 04/20/2017 was removed from the data set, because it is the supervised value to be predicted, since the learning process of neural networks use only data in the past since a previous day.

DATE	(12/19/2017)
1 DAY PREVIOUS	500
2 PREVIOUS DAYS	00
3 PREVIOUS DAYS	300
4 PREVIOUS DAYS	200
5 PREVIOUS DAYS	100

Tabela 2: Transformation of the refectory records into a time series.

4.3.2 Exogenous data

The exogenous data correlated with consumption are divided into two main types, climate data collected from weather stations near the ICT-Unifesp, and data derived from the dates of consumption records.

In terms of the climatic variables used as exogenous data, parameters that may affect the consumption of meals indirectly were considered, such as average ambient temperature, atmospheric pressure, humidity and wind speed. Such parameters can be obtained free of charge by BDMEP - Meteorological Database for Teaching and Research, belonging to the public institution INMET - National Institute of Meteorology, belonging to the Ministry of Agriculture, Livestock and Supply. t is necessary to register in the INMET platform [1]. The institution contains data registered in digital form since 1961 in the whole country, the historical data referring

[1] http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep

to periods before 1961 are not yet in digital form and, therefore, are unavailable in BDMEP. It is important to emphasize that the BDMEP takes 90 days to register each new date.

Besides the environmental data, exogenous data were also generated from the consumption data collected. The date information contained in the indices of the records of endogenous data was derived from various information that represent the consumption behavior in relation to the seasonality of student attendance influenced by the agendas of academic activities.
The following parameters have been defined:

- Semester 1 or 2 in categorical and binary format;
- Day of the week in categorical and binary format;
- Distance in days to previous and subsequent registration;
- Semester progress in percentage scale;
- Advance of the month in percentage scale.

The distributed consumption in a five-day window for entry into MLP networks followed a similar pattern with the demand forecasting work in R.U performed by Lopes (2008) and Rocha, Matos e Frei (2011), pictured in Figure 11. Finally, the Table 3 represents the data set structured and prepared for the process of division into training domains, validation and testing for model training.

Finally, the Table 3 represents the set of data structured and prepared for the process of division into training domains, validation and testing for model training.

4.3.3 Pre-processing

In the pre-processing stage, the endogenous data is transformed into five-day length time series, normalized with *outliers* removal, and application of scale 0 to 1 so that all the data correspond to the same learning domain. After the completion of these steps, the data set was prepared for the experimental phases 1 and 2, which performed a division of the final data set into distinct time intervals. Although the distribution of the data are presented in dates as a function of time classifying itself in a time series model, it is assumed that their behavior is also impacted by causal relationships with other exogenous variables, such as academic recess, holidays, events, intense rainfall that cause local traffic and impact on the logistics and frequency of the public, among other variables of less apparent causes.

Final structure of the data set indexed by date:		
identifier	variable name	variable type
0	SEMESTER_1	int64
1	SEMESTER_2	int64
2	MONDAY	int64
3	TUESDAY	int64
4	WEDNESDAY	int64
5	THURSDAY	int64
6	FRIDAY	int64
7	DISTANCE_DAY_PREVIOUS	int64
8	DISTANCE_DAY_POSTER	int64
9	PERC_CONCLUSION_SEM	float64
10	PERC_CONCLUSION_MONTH	float64
11	ATMOSPHERIC _PRESSURE	float64
12	TEMPERATURE	float64
13	HUMIDITY	int64
14	WIND	float64
15	SALES_LUNCH	int64
16	SALES_LUNCH_1	int64
17	SALES_LUNCH_2	int64
18	SALES_LUNCH_3	int64
19	SALES_LUNCH_4	int64
20	SALES_LUNCH_5	int64
21	ENTR_LUNCH	int64
22	ENTR_LUNCH_1	int64
23	ENTR_LUNCH_2	int64
24	ENTR_LUNCH_3	int64
25	ENTR_LUNCH_4	int64
26	ENTR_LUNCH_5	int64
27	ENTR_DINNER	int64
28	ENTR_DINNER_1	int64
29	ENTR_DINNER_2	int64
30	ENTR_DINNER_3	int64
31	ENTR_DINNER_4	int64
32	ENTR_DINNER_5	int64

Tabela 3: Final structure of the data set indexed by date

4.3.4 Data treatment for model entry

The endogenous data, after structured in the final table of the data sets, goes through the following transformations:

of the standard deviation of each vector of attributes, and normalization of the maximum values for the ceiling of 3x the standard deviation and minimum of 0;Data transformation in scale of 0 and 1.

Exogenous data do not be converted into time series, therefore they were treated according to the steps:

- Data transformation in scale of 0 and 1;
- The categorical binary parameters (days of the week and semester) are already scaled because they are binary categories.

4.3.5 Experimental Stages

The experimental process was carried out in two distinct routes of division of the temporal domain of the data set, and the results obtained between the two phases were compared.

The data set covering the period 2017 to 2019 was divided into a training, validation and testing set as follows:

1st Phase with validation in the 1st semester of 2018 and test in the 1st semester of 2019 In this script, the validation semester that makes up the data set for *backpropagation* training of neural networks contemplates the first semester of 2018 and the test set contemplates the first semester of 2019. Data from 2017 contemplating the 1st and 2nd semesters, and 2018 contemplating the 2nd semester, were used for training. The results obtained in this division were used to validate the hypothesis that the models specifically learn the seasonality of consumption in the first semester, doing better in the tests carried out in the first semester of 2019, compared to the other trained models with validation throughout the year 2018. Therefore, the data set of the first phase includes the following domain:

- Model training set, contemplating the first and second half of 2017 and second half of 2018;
- Validation set of the models, contemplating the first half of 2018;
- Test set of models, contemplating the first half of 2019.

2nd phase with training in 2017, validation in 2018 and test in 2019 In this phase, the sets were divided according to their description and the best model found goes through a final test phase in the domain of the first phase (test only in the first half of 2019). The metrics obtained in this test were compared with the best model of the first phase.

4.4 Model definition and training

In the data set of this work, the data obtained are divided into temporal and endogenous data (such that each record of consumption and sale brings the information of its domain in a time interval of five days prior) and discrete and exogenous data, being categorical date variables for each record and climate variables.

Therefore it was necessary to implement specific models for time entries and specific models for discrete entries. For the final output a committee of endogenous and exogenous Neural Networks was implemented, with a perceptron neuron at the output, receiving the two values of the endogenous and exogenous models for the regression of the outputs of the two networks to the value that will be the prediction of consumption.

Endogenous models

- Development of low depth perceptron networks to evaluate network learnability;
- Increase network depth and evaluate changes in RMSE loss function;
- Implementation and evaluation of models with GRU recurring networks, as shown in Figure 10 that are specially developed for learning with data memorization, and in the case of this work, can memorize the weekly seasonality of consumption (in a five-day interval).

Mixed Models : Endogenous and Exogenous

- For the time data (consumption and sales) we used the best endogenous models from previous experiments for endogenous inputs.
- For the discrete and categorical data the input of this data was adapted to perceptron network
- The exit of the two Neural Networks was concentrated in a perceptron, creating a committee of Neural Networks to obtain the expected final exit.

4.4.1 Hyper parameters : Activation and optimizer function

As based on the Chapter 2 in section 2.2.1 the activation function gives perceptron the ability, when connected to a network, to solve linear and non-linear problems, adding adaptation and improvisation when solving programs that are not contained in its power data. So for the hidden layers of the developed MLP neural networks the ReLu function will be applied and for the output neuron the linear function will be applied, and the training optimizer performs the function of optimizing the convergence time of the readjustment of the weights to the ideal values, being chosen the ADAM optimizer with learning rate set at 0.001.

4.5 Test and Evaluation Metrics

The main metrics for the evaluation of the models are the *Root Mean Squared Error* (RMSE), The Pearson correlation coefficient (R), and the ”chi-square coefficient” defined as R^2. These statistical metrics were used in the test steps to evaluate the proximity of the model predictions to the actual consumption behavior. The equations 4.1 and 4.2 present the formulation for the RMSE and the Pearson coefficient (R), respectively.

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

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

Where "est"are the estimated values; "obs"are the actual values; n is the number of samples ; σ is the default deviation; R is the linear correlation; $\bar{x}$ is the mean of x. We also evaluated the positive and negative errors between the predicted and real values, to represent how many meals would be discarded and how many would be missing if meal production were according to the model predictions.

5 Results

This chapter describes the main experimental results obtained during this research. As described in Chapter 4, the experiments were conducted in two phases. However, we have chosen to present in this chapter only the main results obtained. The other results are available in the Annexes to this document. The first sections introduce the organization of the data, a brief analysis of the variables and the experimental protocol, respectively.

5.1 Organization of the data set

The data collection procedure was carried out according to the methodology of Section 4.3.1 for endogenous data, and according to the steps of Section 4.3.2 the exogenous data were obtained. Both sets of data collected were structured according to the Table 3, containing a time interval of records, from April 12, 2017 (2017-04-12) for the first record to December 16, 2019 (2019-12-16) for the last record, totaling 514 records of meal consumption on school days.

According to the methodology defined in Section 4.3.5, this data set with a total of 514 records has been duplicated to 2 distinct experimental phases, each with a specific organization of the data set. The data set of the first experimental phase was organized according to the Figure 14. This phase has the validation set exclusively contemplating the first semester of 2018, indicating that the first semester of 2018 could present a consumption and sales movement similar to the first semester of 2019, being the ideal one for tests involving only the first semester of the test set.

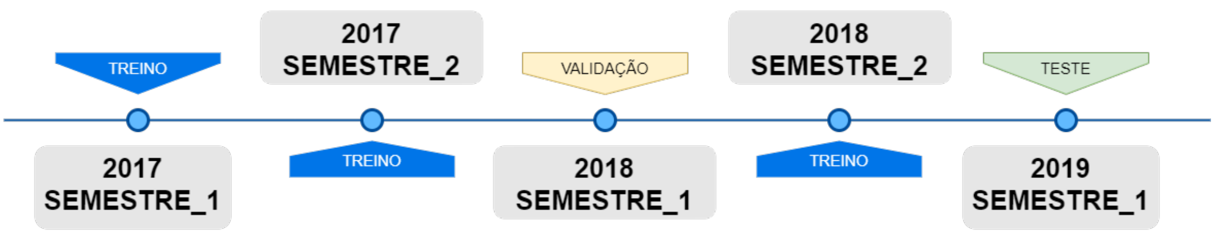

Figura 14: Time domain of phase 1.

For the second phase the data set was organized according to the Figure 15. For this phase the selected validation set contemplated the entire school year of 2018, while for the model testing experiments the selected data contemplated the entire school year of 2019.

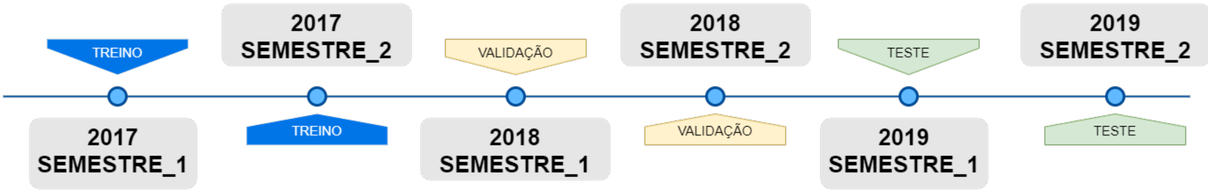

Figura 15: Time domain of phase 2

5.1.1 Manipulation and pre-processing of the data set

Seeking to organize the raw data obtained for its subsequent application in the models, some peculiarities were found. The first difficulty found in the experiments was an anomalous behavior of the forecast results for model RNN_ENDO_2. The blue line in Figure 16 represents a meal forecast of model RNN_ENDO_2, and the red line real consumption values of the first half of 2019. In this way, it was possible to observe that in both sets (real and predicted) this behavior was present.

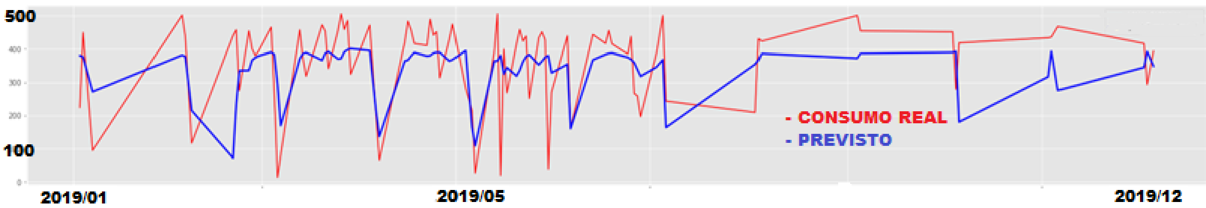

Figura 16: Result of RNN_ENDO_2 model obtained on the data set randomly ordered over time.

After an exploratory analysis it was discovered that the records contained an error in the indexing by date, where the date stamp was changed, that is, the days by months and vice versa. After the correction of this indexing problem the actual consumption and forecast data produced realistic results and within the expected format, as presented in Figure 17, that presents an example of prediction by model RNN_ENDO_2 and the real data for the lunch period.

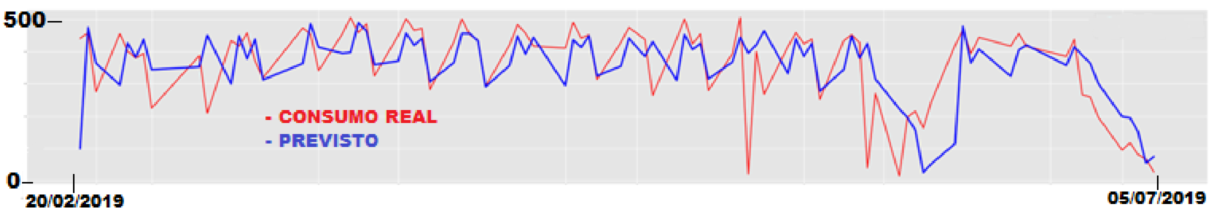

Figura 17: Result of RNN_ENDO_2 model obtained on the data set with corrected order.

5.2 Variable Evaluation

During this section some comments will be made about the variables characteristics that were most important to the problem, and some graphs will be presented from the statistical study done to evaluate the relationships among them.

5.2.1 Refectory consumption estimates

The analysis of the estimation technique of consumption, performed subjectively in relation to the consumption of the previous week, uses the calculation of 30% of production above the consumption of the fifth previous day. This estimation method is adopted to tolerate discards due to the existence of a contractual fine for lack of meals. Still, it is possible to observe that this model of 30% more produces a linear behavior, represented by the blue line in the Figure 18, being distant from the real consumption behavior, indicated by the red line.

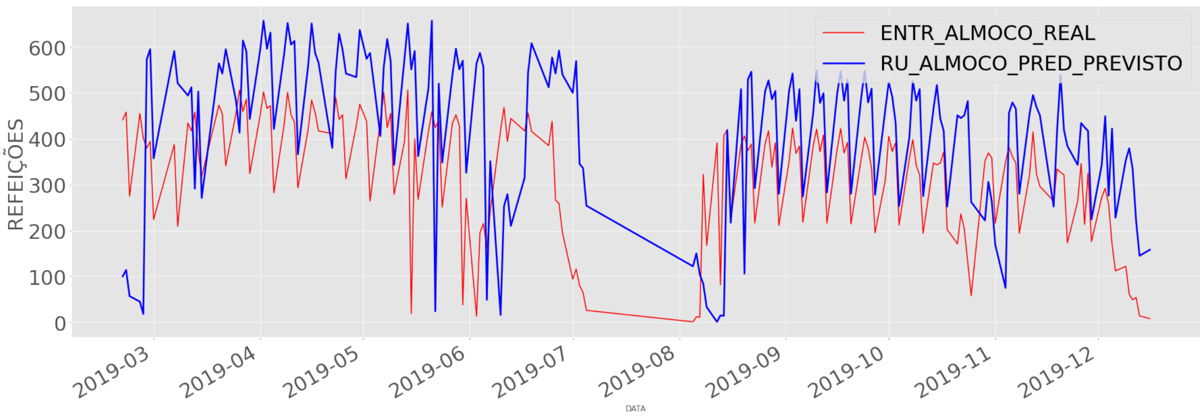

Figura 18: Refectory estimate for 2019.

Although the estimate follows the trends of falls and increases in consumption, the Figure 19 shows the dispersion generated between the estimate of the R.U and the actual consumption in the year 2019, showing that the linear regression (red line) has the axis totally decentralized with the function identity of the ideal estimate (represented by the imaginary diagonal formed between the origin of the graph and the top right vertex).

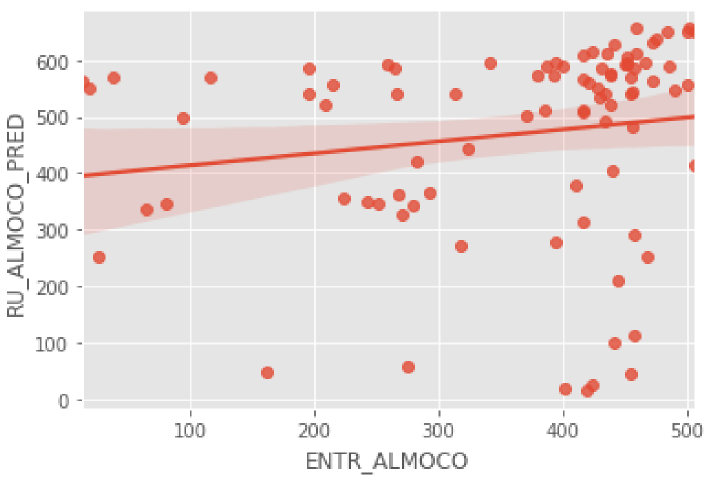

Figura 19: Scatter plot of the refectory's estimated consumption for the year 2019.

Thus, this format of prediction also generates an error higher than 30% in the total amount of meals discarded during the semester, caused by the oscillatory behavior of consumption, according to the table 4.

Consumption with margin 30% above the previous day 5	
TOTAL MEALS CONSUMED	58653
TOTAL ESTIMATED MEALS	76262
CORRELATION (r)	0.4006
P-value	2.0845e-08
RMSE	191.7620
SUM OF POSITIVE ERRORS	23412
SUM OF NEGATIVE ERRORS	-5803
AVERAGE ABSOLUTE ERROR	133.0
ABSOLUTE ERROR AVERAGE PERCENTAGE	205.6113%

Tabela 4: Metrics of the refectory's estimated consumption for the year 2019

5.2.2 Analysis of endogenous variables

The endogenous variables are the input time parameters in the MLP and GRU models, corresponding to the consumption domain in the restaurant.

5.2.2.1 Consumption of the current day in relation to the previous day's ticket sales

It is possible to notice in the Figure 20 that ticket sales in the lunch period showed a different behavior in 2017 compared to the following years due to a limitation imposed by the refectory, from 2018 onwards students could purchase only 2 tickets per day. Possibly, this limitation was given to approximate the consumption behavior of 1 or 2 days after the ticket sale. This limitation can be interpreted as a management aid method for meal production and waste treatment.

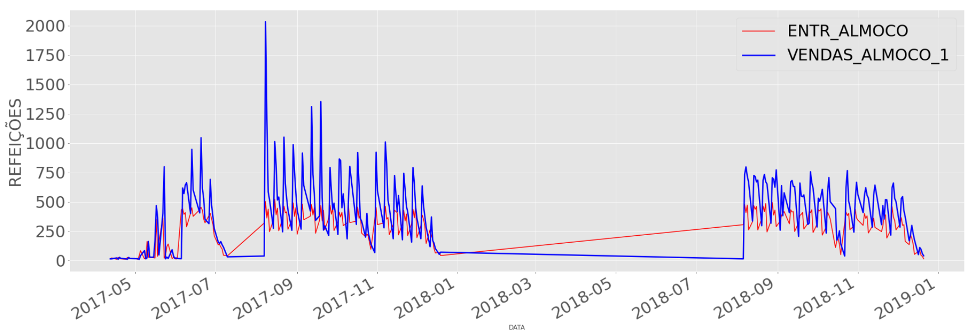

Figura 20: Correlation between consumption and lunch sales.

Even with the *outlier* value of 2000 sales in a single day, and with the new limitation on ticket purchases from 2018 onwards, consumption at lunchtime is strongly related to ticket sales in the previous day's lunch period. It is also noted that students have adapted to the limitation imposed for the use of tickets with a validity period of only two days, as the value of the correlation coefficient is approximately 72%, as shown in the table 5.

Tabela 5: Comparison of consumption with a previous day

CONSUMPTION IN RELATION TO SALES OF 1 DAY BEFORE	
CORRELATION (r)	0.7255528038157009
P-value	5.399561176138223e-41
RMSE	260.5399426736619
AVERAGE ABSOLUTE ERROR	139.0
ABSOLUTE ERROR AVERAGE PERCENTAGE	90.18

There are other unscheduled factors involved, such as possible failure of sales records in the system, as well as the *outlier* value of 2000 sales can be interpreted with the system and meal database migration that occurred in 2017 from the Talim unit of ICT-Unifesp to the Hospital São Paulo database. Possibly also sales were imported from the old system without the differentiation of dates.

The total of meal sales at lunchtime was 242282 for all 514 entries in the data set. In this same period the actual consumption value, meaning students who effectively entered the refectory (passed through the turnstile) totaled 163752 meals. Although notorious, the difference of 78,530 meals sold above the actual consumption was not able to be investigated in this work. It should be noted that these figures were obtained from the original set of data provided by the ICT-Unifesp R.U contract inspector through an e-mail request to this inspector.

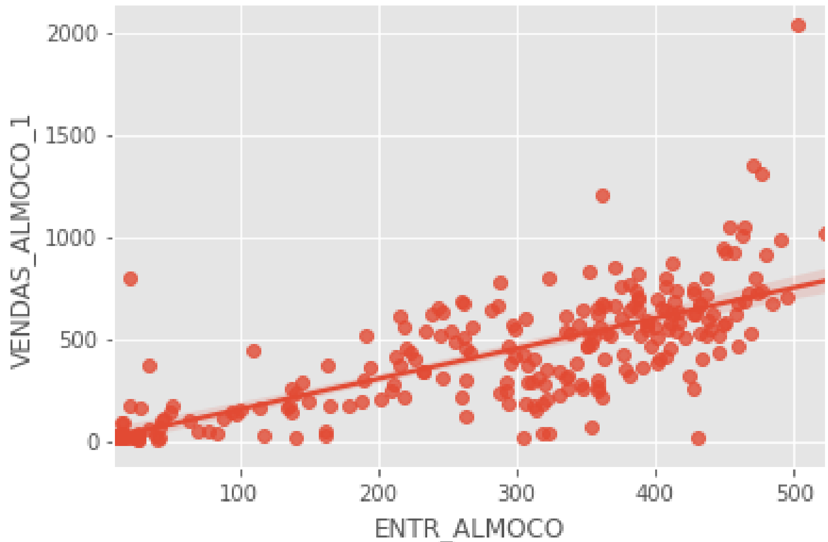

Figura 21: Dispersion graph between consumption and lunch sales.

5.2.2.2 Normalization and scale of *features*

The process of normalization and scale is demonstrated in this Section with a *feature* of sales of *tickets* from 1 previous day, because among all this is the one that has produced *outliers* of greater prominence. The data normalization is done with the ceiling of 3x the average standard deviation, so the peak of 2000 sales was normalized to the rounded value of 1356 meals. Even with the normalization, the linear behavior of this *feature* was maintained, as presented in the Figure 22.

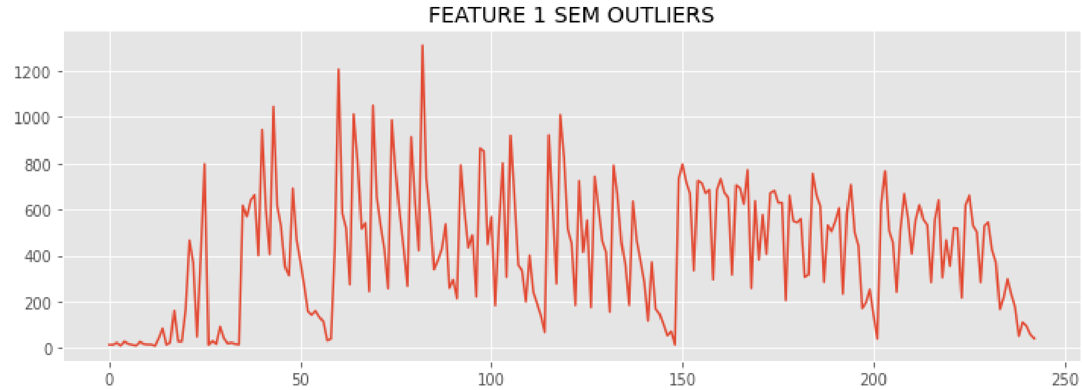

Figura 22: Sales of normalized *tickets* with 3x the standard deviation ceiling.

After normalization, the scale was standardized from 0 to 1 in this one and as it is observed in Figure 23, the linear behavior was maintained again.

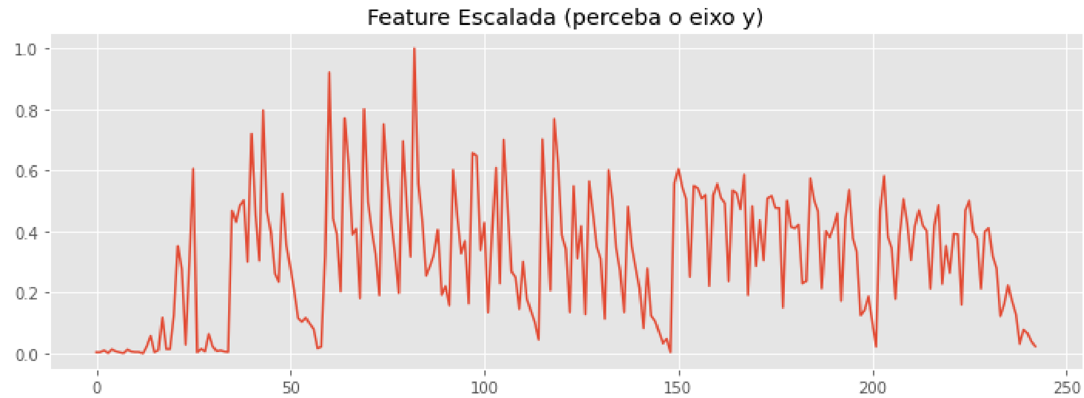

Figura 23: Sales of *tickets* climbing from 0 to 1.

This process of normalization and scale has been carried out for all endogenous metrics and also for climate metrics.

5.2.2.3 The current consumption in relation to the previous day's dinner.

Looking to find and evaluate the possible relationship between the various metrics used, this analysis gained relevance as an anomalous and probably casual effect found among the data. Although the curricular schedules of students consuming meals at lunch are usually fired at students consuming dinner the night before, a clear relationship between these 2 variables can be detected. This behavior can be evidenced by the congruence between the parameters, as shown in figure 24 and by the high correlation obtained in the linear regression (R = 0,7655) between these 2 consumptions, presented in figure 25. Even though it presents a relevant correlation, it was not possible to determine an evident cause for this anomalous effect found.

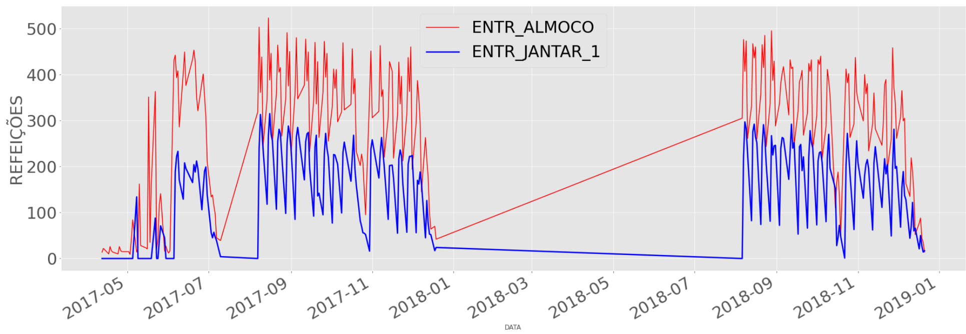

Figura 24: Correlation of lunch and dinner consumption from 1 previous day.

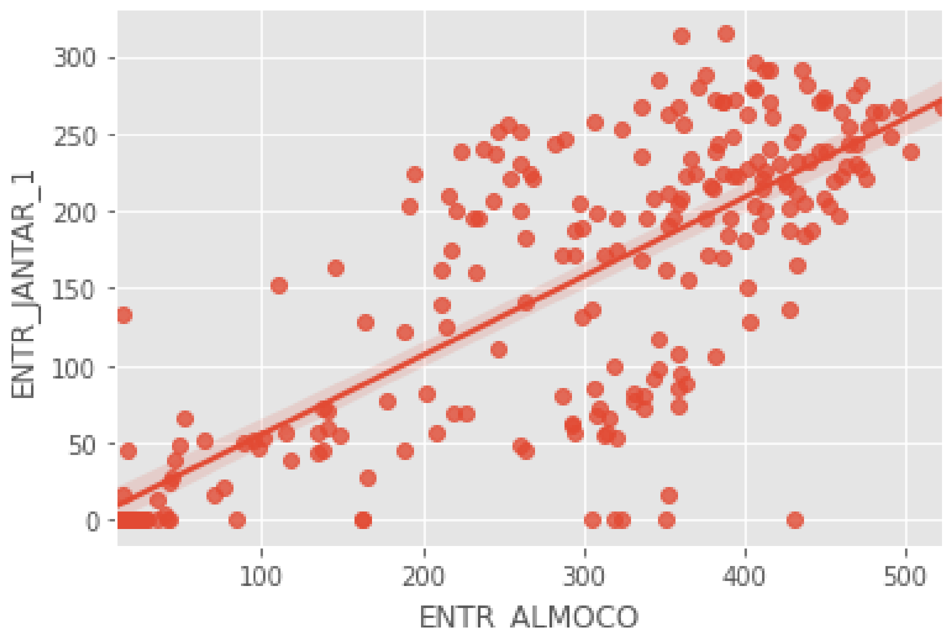

Figura 25: Graph of dispersion between consumption and dinner of 1 previous day.

5.2.3 Weekly Seasonality Analysis

The following consumption graphs from Figure 5 26, representing the Monday, up to figure 30, representing Friday, are generated for the binary categorical *features*, with the *violin-plot* functionality in the library *seaborn*, appropriate for distribution of binary categorical variables in a data set.

The blue violin with the value 1 represents the distribution of consumption across the total data set. The violin with value zero can be ignored and is a standard return on the tool graph, representing the consumption complement for the day of the week considered. On Fridays, the consumption had a smaller distribution scale for the whole set 2019. It was noticeable that despite the alternation of time slots during the exchange of semesters in the year 2019, the days of Tuesday and Thursday concentrated the largest movement of consumption.

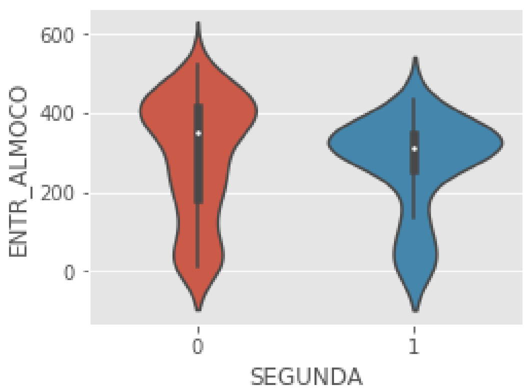

Figura 26: Gráfico de violino da distribuição do consumo na segunda feira.

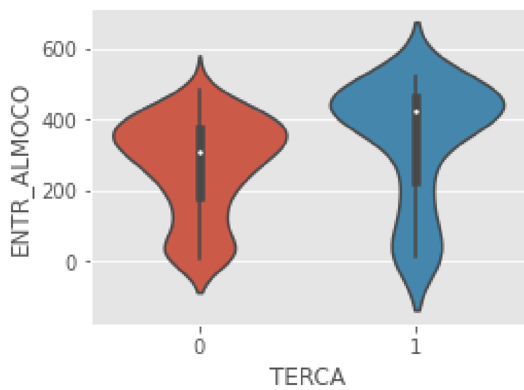

Figura 27: Violin graph of Tuesday's consumption distribution.

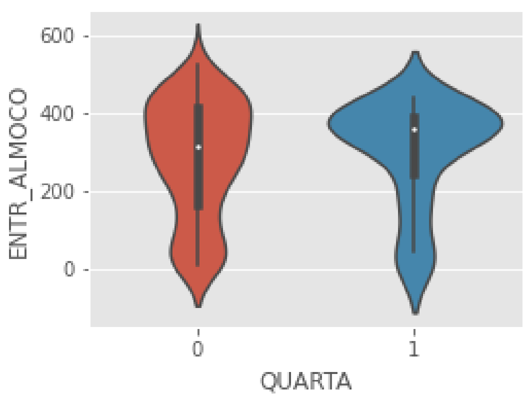

Figura 28: Violin graph of Wednesday's consumption distribution.

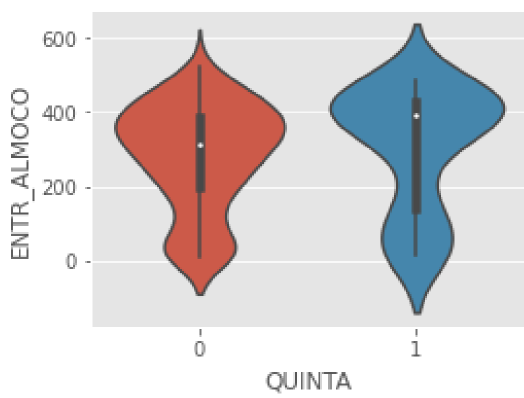

Figura 29: Violin graph of Thursday's consumption distribution.

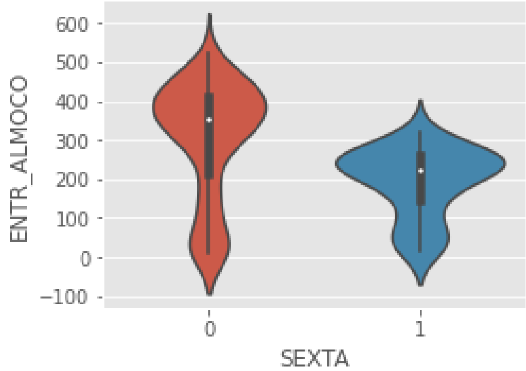

Figura 30: Violin graph of Friday's consumption distribution.

5.2.4 Exogenous variables analysis

The exogenous variables correspond to the discrete domain parameters, which are used exclusively in mixed neural network models and are read by the MLP layers of these models.

5.2.4.1 The current consumption in relation to the advance of the semester.

For this analysis it was necessary to restrict the domain of analysis to 1 semester, the consumption in relation to the advance of the semester had an abrupt fall in the last days of the semester, and thus the correlation of the data sets of figures 31 and 32 obtained negative value.

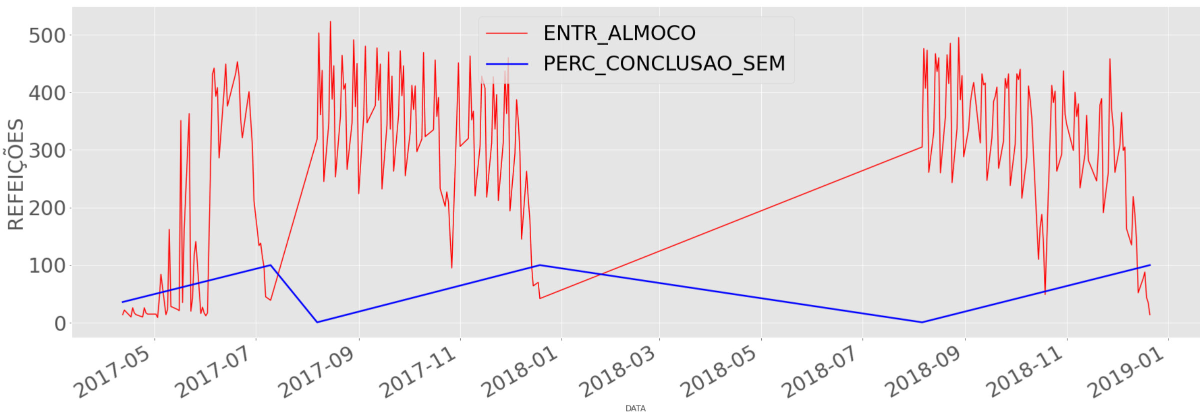

Figura 31: 1st Phase : Relation between consumption distribution and semester's advance, Correlation (r) = -0.35.

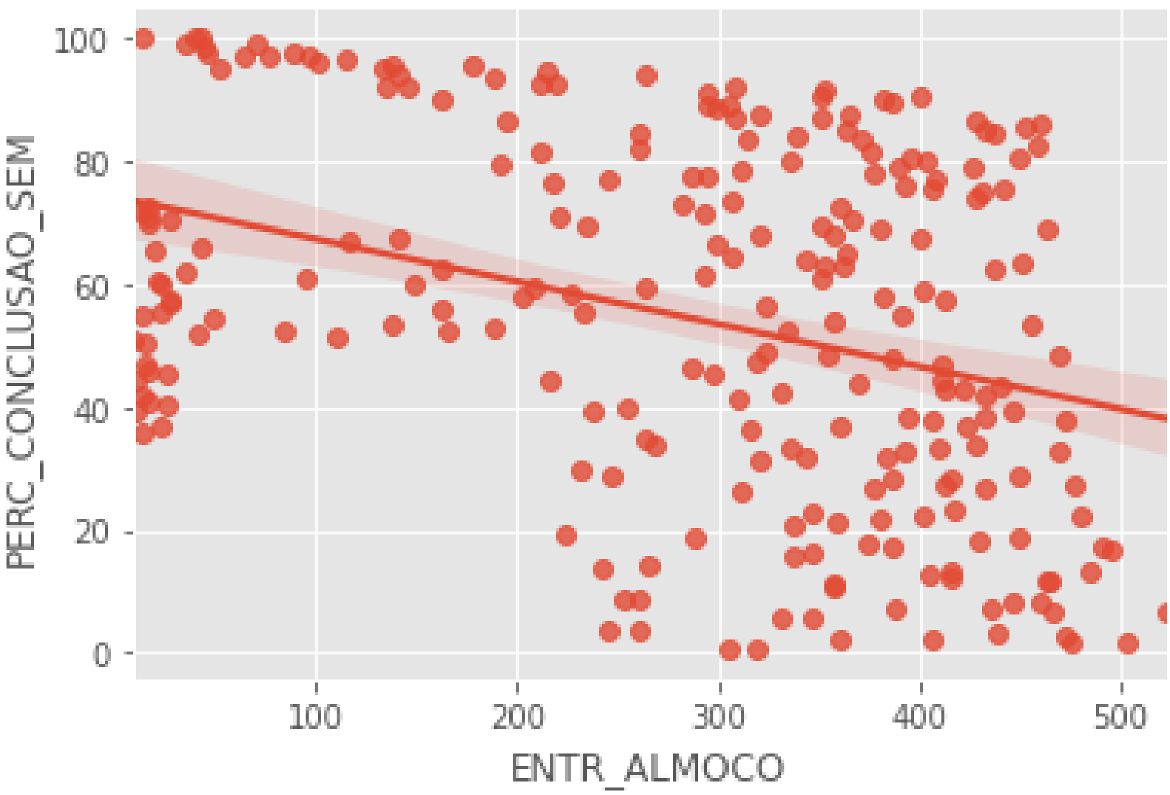

Figura 32: Graph of dispersion of consumption distribution as the semester progresses.

5.3 Experimental Protocol

In this section will be started the experiments with the Neural Network models. The first experiment evaluates the learning potential of the models in predicting R.iU consumption, through a basic neural network model, by concluding that the most basic neural network topologies have learning potential about the problem. Then, the topology was selected and the experimental phase that brought the best results and finally it is analyzed and indicated as the final solution of the project.

5.3.1 Evaluation of the learning problem of meal prediction through MLP Neural Networks

5.3.1.1 Empirical topology adjustment of the first perceptron model

The first neural net experiment performed in the first experimental phase evaluated the learning capacity of the perceptron model on the seasonality of endogenous data, referring to the domain of meal consumption in the R.U., verifying if the consumption behavior in the refectory could be learned by this type of neural net, therefore 1 initial perceptron net was defined with only 1 hidden layer containing 1 neuron for 15 input parameters (same number of endogenous parameters) and with 1 output neuron, called MLP1.

The parameters endogenous correspond to an interval of 5 days prior to the consumption of meals in the lunch period, dinner and ticket sales in the lunch period. The model was named MLP1, its illustration can be seen in figure 33 obtained by using the NETRON tool.

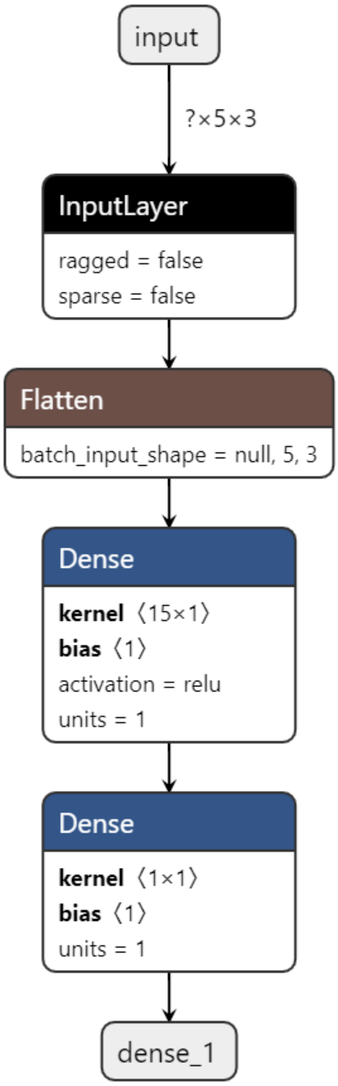

Figura 33: Topology of MLP1 model, NETRON tool.

Each layer of the MLP1 model corresponds to a block with title **Dense** in this Figure, the first edge of the figure between the *input* and InputLayer shows the 3 time series of the input parameters, with an interval of 5 days each. **ENTR_LUNCH, ENTR_DINNER e SALES_LUNCH**. The Flatten block converts each day of time series input into an MLP neural network input parameter, according to the conceptual model of the Lopes (2008)work illustrated in figure 11 which uses only 1 endogenous parameter with an interval of 5 days also. The first hidden layer of this network can be visualized in the first dense block of the figure that demonstrates the ReLu activation function of this neuron, and the number of units of this layer being 1. The output layer is the last block of the figure, also with 1 unit and since the activation function is linear, it is not displayed in the block description.

The training of this model was executed obtaining RMSE with the value 130.62 over the validation set, in this case the first phase being the data of the first semester of 2018, and it is possible to notice in figure 34 that the lines of the loss and training function demonstrated an adequate training until the last training season, and as there was no application of simultaneous Validation Loss with Train Loss, the training did not produce an overfitting.

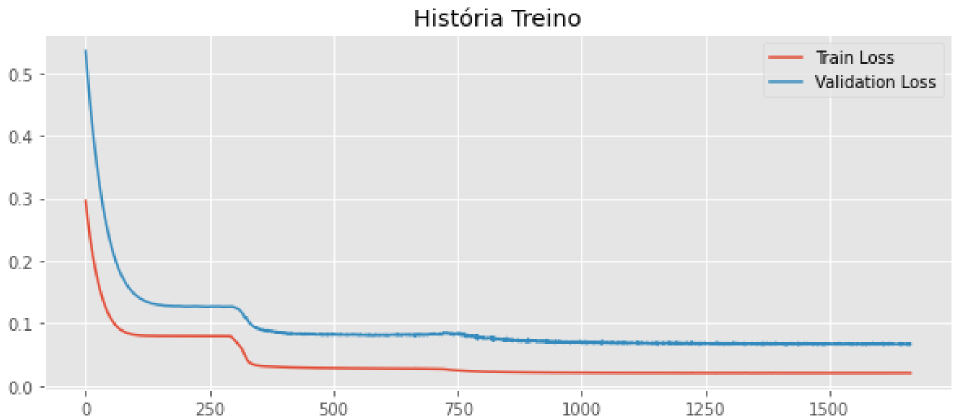

Figura 34: Model training graph MLP1, RMSE = 130,62

Therefore, the depth of the MLP1 model was increased by obtaining the MLP2 model with topology illustrated in figure 35 and after training this model it was possible to notice the decrease of the RMSE to the value of 107.97, as observed in figure 36. It is possible to realize that from the 300 training season on, the Train Loss line began to suffer a decrease in error as the Validation Loss began to gain error from the 400 season on. As the reconfiguration of the network topology produced improved results until an overfitting saturated this improvement for this topology, it was validated the hypothesis that the prediction of consumption in the refectory can be learned by simple models of Neural Networks, and therefore the research followed with the definition of new models demonstrated in the next section.

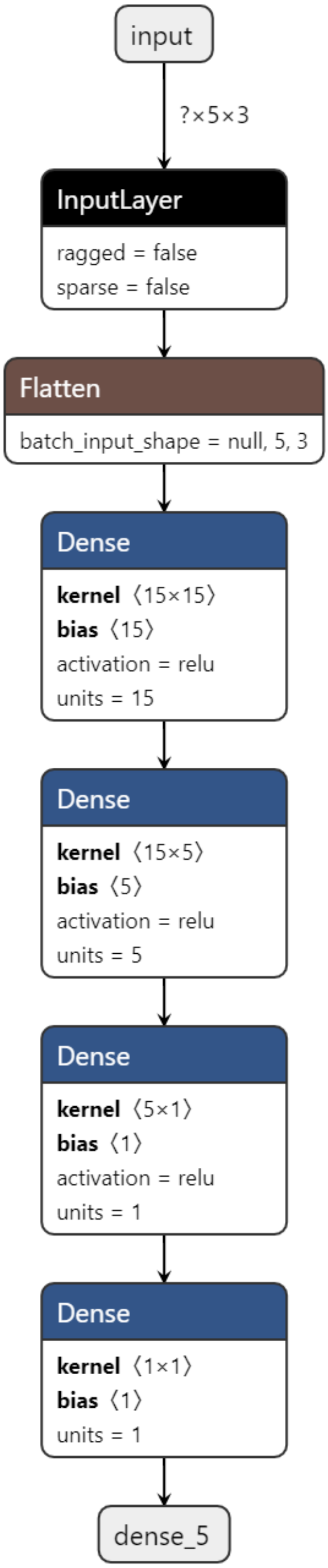

Figura 35: Model Topology MLP2.

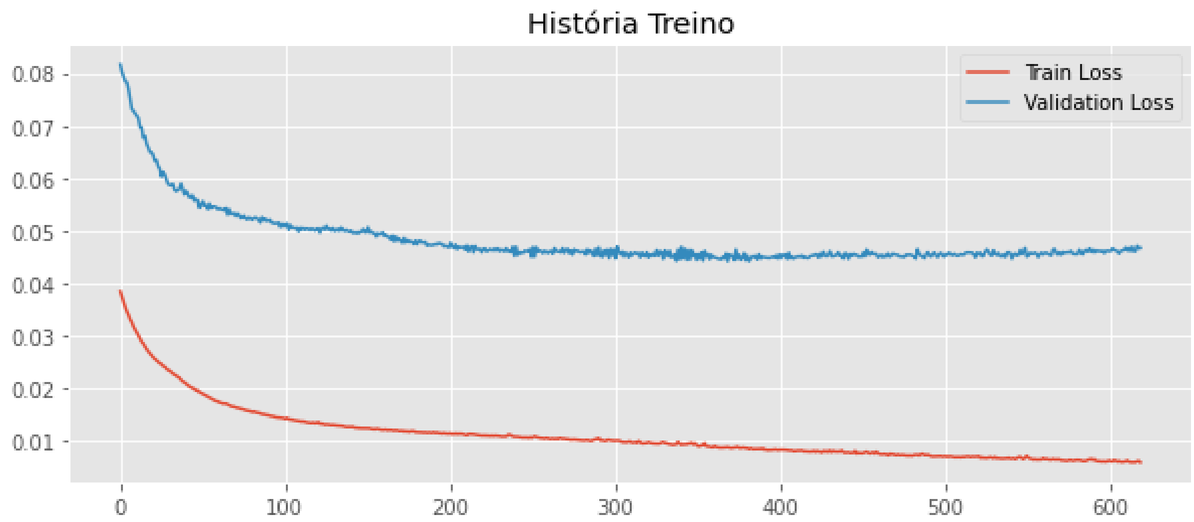

Figura 36: Model training graph MLP2. RMSE = 107,97.

5.3.2 Topologies of best models

In this section some comments are made about the topologies of the models that obtained results presented with schematic figures of the network. For the rest of the trained models, the same procedure is done in annex **??**.

5.3.2.1 Mixed Model RNN_EXO_1, best result in the second phase and in all the whole work

Interpreting the digram of the first mixed model, RNN_EXO_1, in figure 37 the block with the GRU title on the left in the topology figure of the model deals with endogenous (temporal) inputs, as exemplified in endogenous GRU models. The block with title **Dense** is an MLP network that receives a input with 10 parameters of 1 dimension, consequently all discrete, corresponding to the 4 climatic parameters (temperature, humidity, pressure and wind), 1 parameter for the current day of the week, 1 for the current semester and 4 calendar control parameters (distance from the previous date, later, semester advance, month advance). The output of the GRU and MLP blocks are concatenated and treated by the output MLP block.

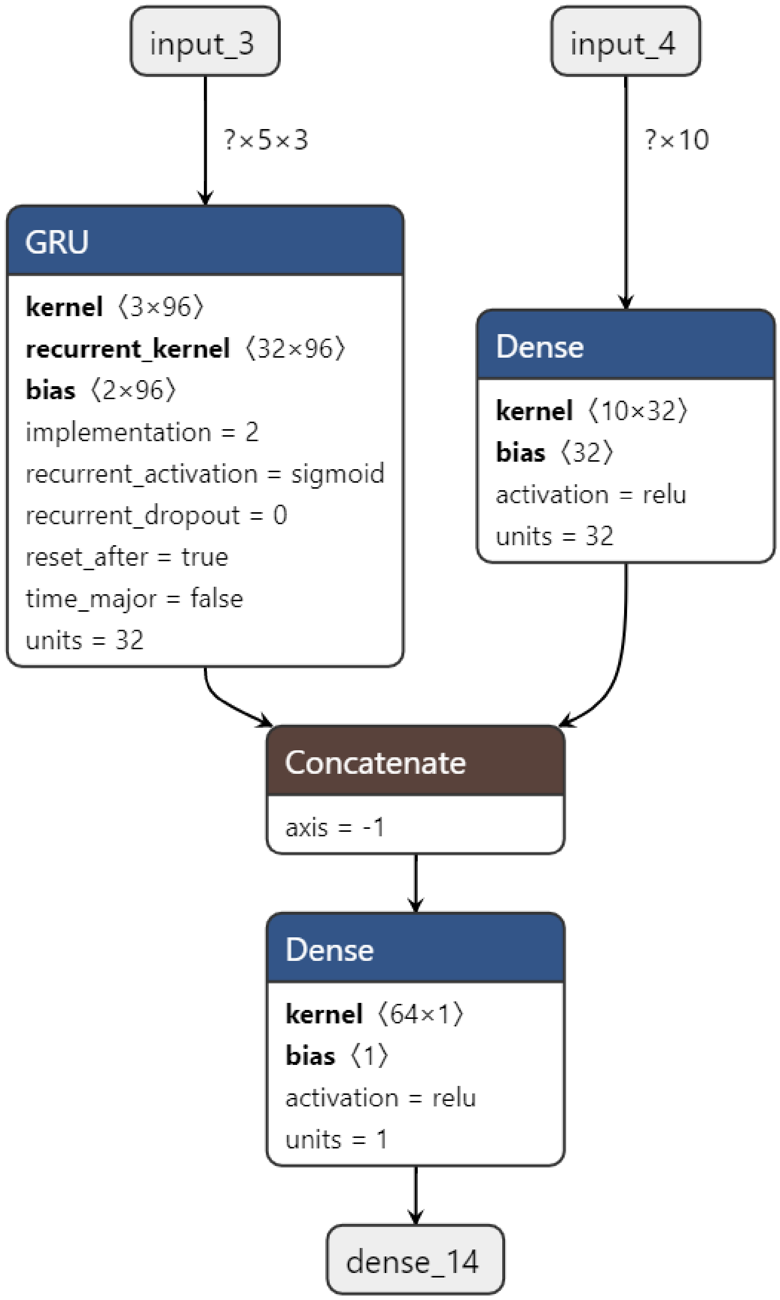

Figura 37: Model Topology RNN_EXO_1.

5.3.2.2 Endogenous model GRU RNN_ENDO_2, best result in the first phase

This model was obtained through a second reconfiguration of the first GRU model, RNN_ENDO_1 detailed in the figure **??** with the increase in unit depth of this previous model, in regressive format of 16 units in the first layer, 8 units in the second and 4 units in the third, and with the inclusion of the dropout resource based on the topic 2.2.4.1, the final topology is observed in figure 38. This model produced the best results in the first experimental phase, detailed in figure 39 in the next section.

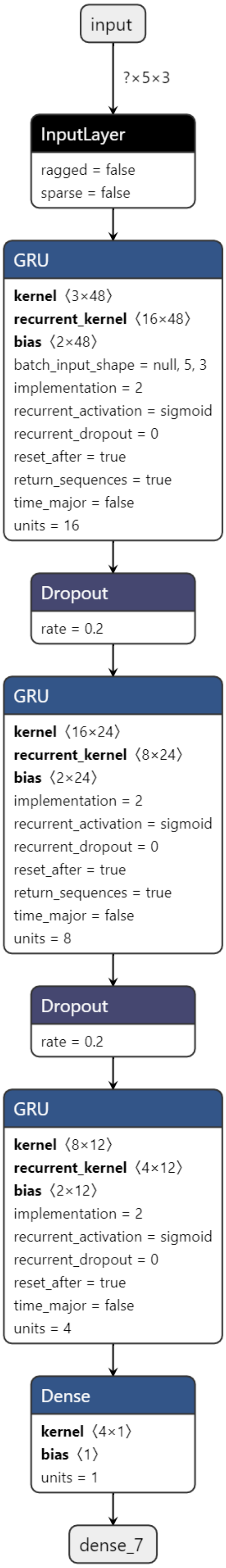

Figura 38: Model Topology RNN_ENDO_2.

5.3.3 Main differences of results between the experimental phases

Differences between the best models For the experiments of the 1st phase, the model that produced the smallest RMSE in the set of tests with advantage in all the other metrics was the endogenous model RNN_ENDO_2, with some prediction anomalies, as observed in figure 39.

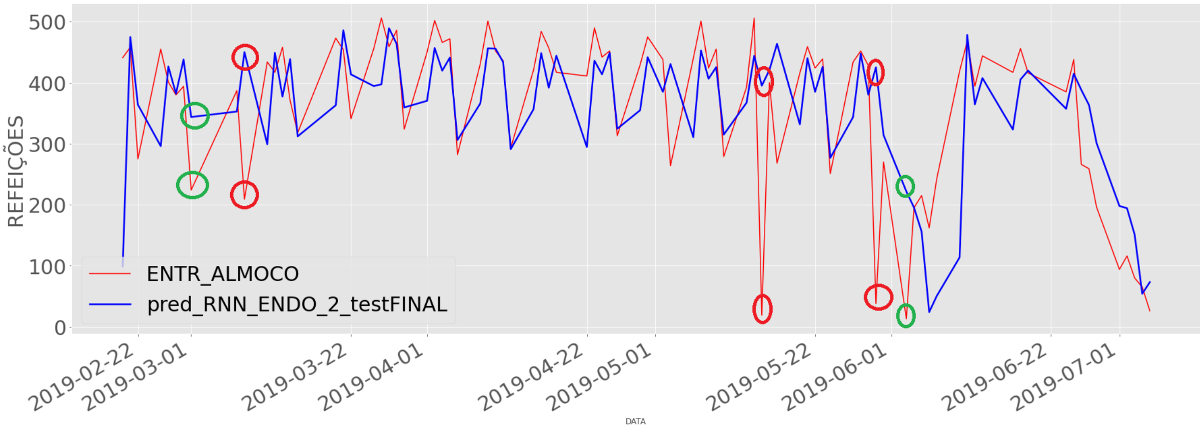

Figura 39: Analysis of predictive anomalies of RNN_ENDO_2

The green outliers represent predictions that corresponded to an upward or downward trend in consumption but with discrepant errors, and the red outliers represent predictions with an inverse trend in consumption. The justifications for the prediction within the trend, are found in the table 6, denoting special dates that could not match the learning process of the model.

Tabela 6: Model prediction errors RNN_ENDO_2 in the 1st phase

Date	Consumption	Justification
03/01/2019 (Friday)	224	Friday pre-carnival
06/03/2019 (Monday)	13	Monday after the student paralyzation

The justifications for forecasts where the model followed the opposite trend to the consumption also corresponded to the special dates, conferred in table 7.

Tabela 7: Anomalies of model predictions RNN_ENDO_2 in the 1st phase

Date	Consumption	Justification
03/08/2019 (Friday)	209	Friday after carnival
05/15/2019 (Wednesday)	19	Student paralyzation in Afonso Pena Square
30/05/2019 (Thursday)	38	Student paralyzation in Afonso Pena Square

In the metrics of this model, it is observed in table 8 that the total of positive errors, corresponded to a discard of approximately 3479 meals and the average quadratic error of forecast was approximately 108 meals.

Tabela 8: Metrics from the best model: RNN_ENDO_2

Best model:	RNN_ENDO_2:
Total_Consumed	31962
Total_Foreseen	31465,61133
Error_Total_Forecast	-496,3886719
Percentage_Error_Total	-1,5530%
Correlation	0,595439895
P-value	9,42215E-10
RMSE	108,0663015
Sum of negative errors.	2982,567947
Sum of positive errors.	3478,957266
ERROR_ABS_MEDIAN	46,70721436
ERROR_ABSOLUTE_AVERAGE_PERCENTUAL	74,93539002

Meanwhile, in the second phase, all the models obtained improvements in the training error over the validation set, and the model with the best predictions of the work was obtained, the mixed model **RNN_EXO_1** detailed in its own section to follow. It is important to note that the 2 phases have produced the best models of different classes, the first with a model that uses only endogenous data, and that includes a validation and test set with a range of only 1 semester, and the second with a model that uses temporal and discrete data and that uses a validation and test set with a range of 1 year. This denotes that better results were achieved without any change in parameters and hyper-parameters in the models, changing only the temporal organization of the data sets.

During the testing of all models, only the first semester contemplated special dates where these models produced forecast anomalies, illustrated as an example in figure 39.

5.4 Results with the best model, RNN_EXO_1

This model, represented in the figure 37 obtained the best results of all this work, when it was training in the second experimental phase. It is noticeable its improvement of results with a single change in the organization of the data sets between the experimental phases.

5.4.1 Comparison of training between the two phases

It is observed that this model produced a slower overfitting in the first phase, illustrated in the figure 40 ompared to the training of the second phase demonstrating a convergence to a fast and sharp overfitting from season 100 onwards, bringing better results with the early stopping of training, illustrated in the figure 41. The difference of the RMSE metric values between these 2 trainings, producing a better result in the second phase, of RMSE = 109.97 compared to the result of the first phase of RMSE = 132.94.

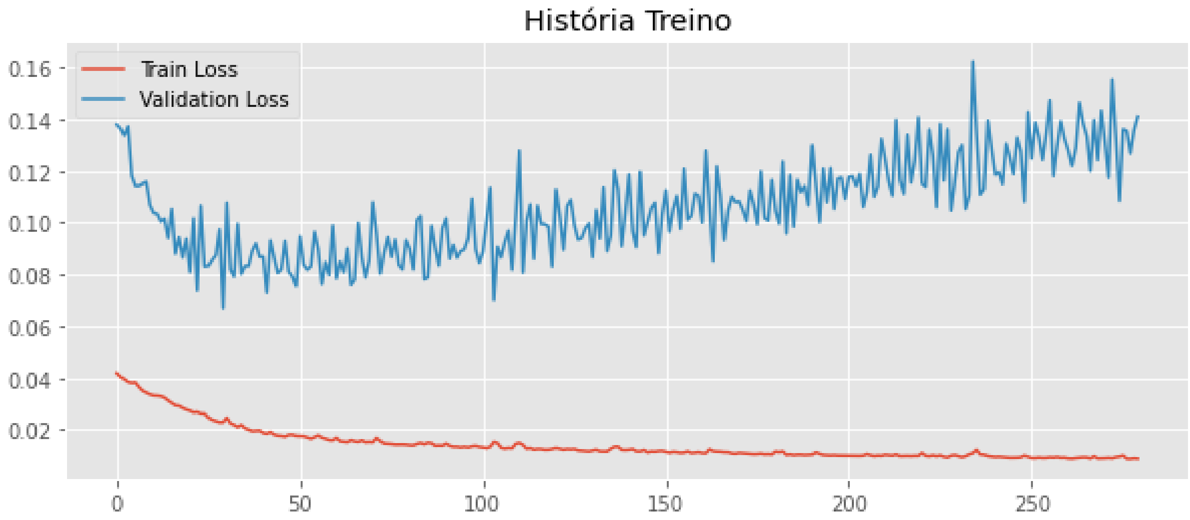

Figura 40: Model training RNN_EXO_1 in the 1st phase, RMSE = 132.94

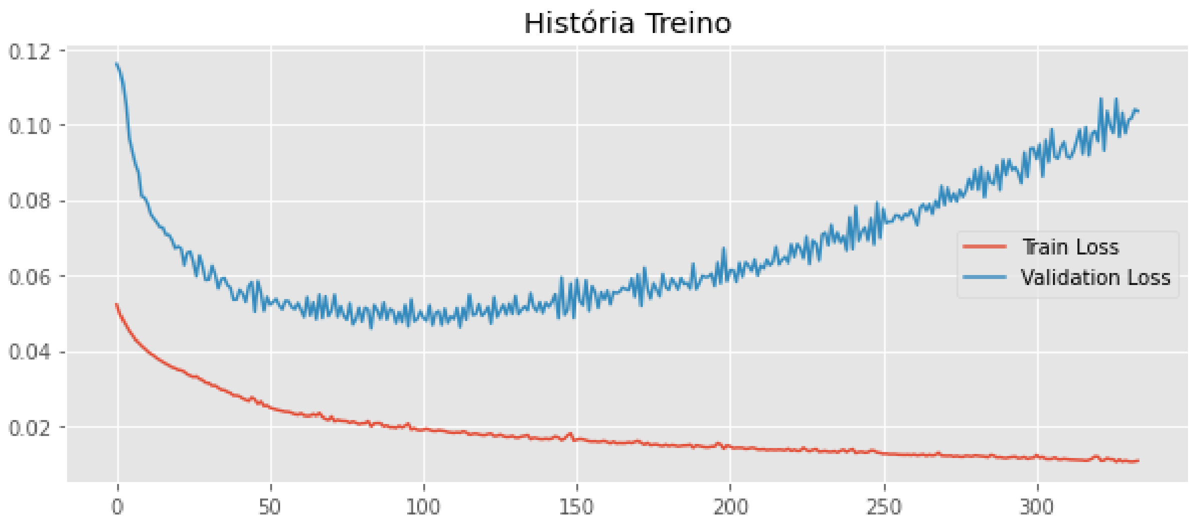

Figura 41: Model training graph RNN_EXO_1 in second phase, RMSE = 109.97

5.4.2 Comparative test of the model in the first semester between the two phases

As this model trained in the second phase obtained the best results of the work, the metrics for the model testing within the domain of the first semester of 2019 were recalculated to make an equivalent comparison with its version trained and tested also in the first semester of 2019 in the first phase.

The table 9 shows that the RMSE testing model trained in the first phase performed less well than the RMSE trained in the second phase according to the table 10. The RMSE being smaller in the training of this model in the second phase brings an improvement in all other metrics compared to its training in the first phase. The sum of the positive prediction errors was also lower in the second phase, which impacts less on meal disposal. The RMSE of this model tested in the first phase and reaching RMSE = 106.208 also did better than the best model of the first phase, the RNN_ENDO_2 which reached RMSE = 108.06.

Tabela 9: RNN_EXO_1 TRAINED IN PHASE 1,TEST FIRST SEMESTER 2019

RNN_EXO_1 TRAINED IN PHASE 1,TEST FIRST SEMESTER 2019	
RMSE	124.49
TOTAL MEALS CONSUMED	31962
TOTAL OF PROJECTED MEALS	28728.816
FORECAST ERROR	-3233.1839
PERCENTAGE ERROR	-10.11%
CORRELATION (r)	0.41
P-value	6.59e-05
R2	0.16
SUM OF POSITIVE ERRORS	2709.17
SUM OF NEGATIVE ERRORS	5942.35
MEDIAN ABSOLUTE ERROR	85.59
ABSOLUTE ERROR AVERAGE PERCENTAGE	90.98%

Tabela 10: RNN_EXO_1 TRAINED IN PHASE 2,TEST FIRST SEMESTER 2019

RNN_EXO_1 TRAINED IN PHASE 2,TEST FIRST SEMESTER 2019	
RMSE	106.2080
TOTAL MEALS CONSUMED	31962
TOTAL OF PROJECTED MEALS	32170.24
FORECAST ERROR	208.2460
PERCENTAGE ERROR	0.6515%
CORRELATION (r)	0.59
P-value (p)	1.4143e-09
R2	0.3485
SUM OF POSITIVE ERRORS	3454.8698
SUM OF NEGATIVE ERRORS	3246.6228
MEDIAN ABSOLUTE ERROR	59.5414
ABSOLUTE ERROR AVERAGE PERCENTAGE	83.2671%

The scatter plot of the trained model in the first phase as shown in figure 42 also came out worse, farther from the upper right edge of the plot, than the scatter plot of the trained model in the second phase as shown in figure 43.

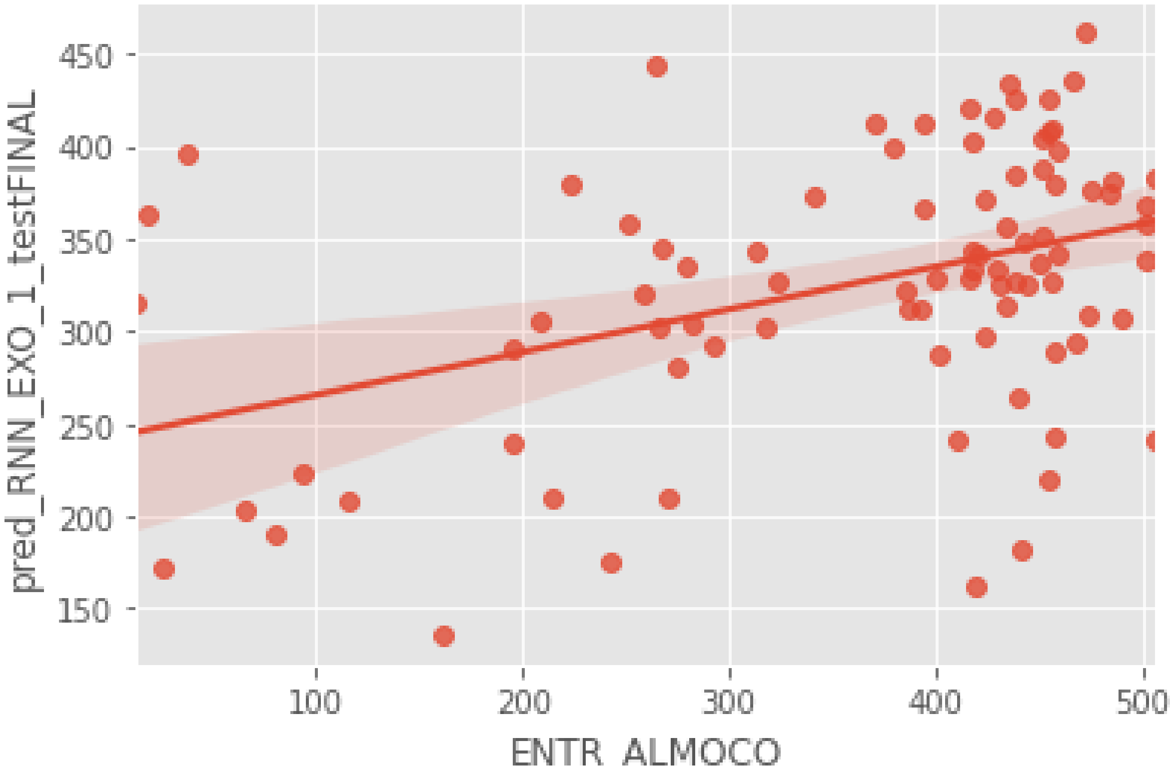

Figura 42: Model test dispersion graph RNN_EXO_1, first phase

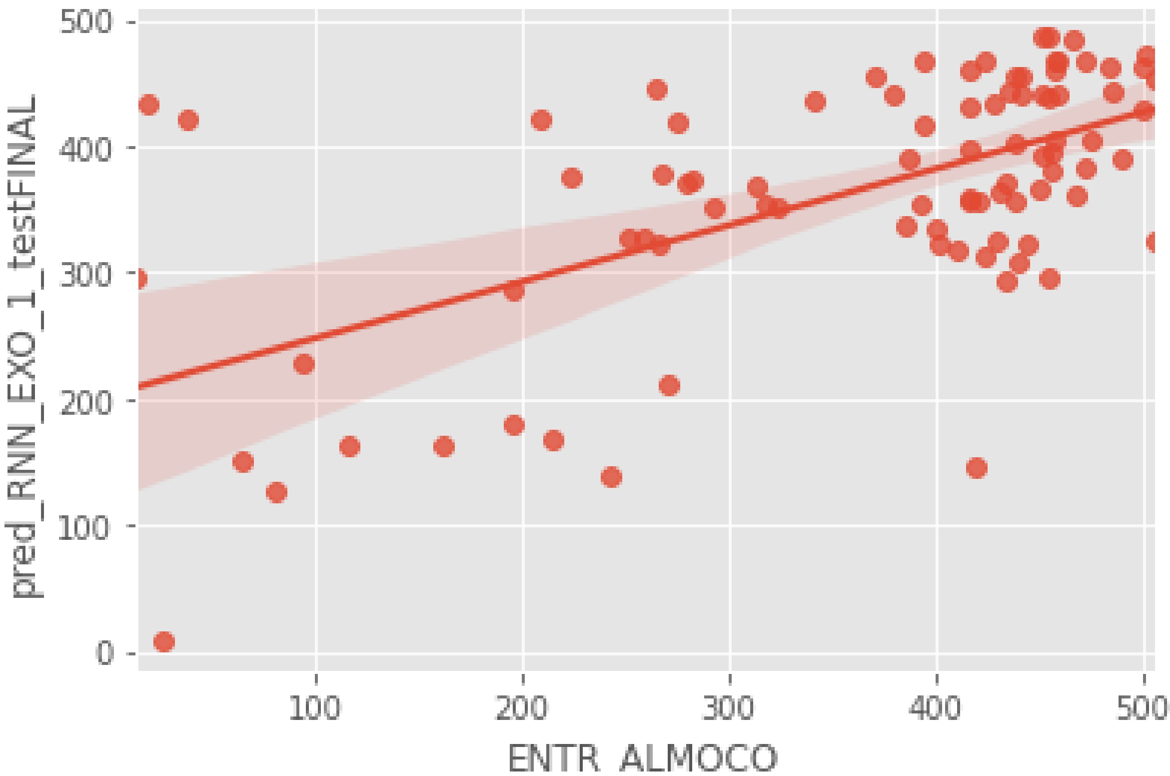

Figura 43: Test scatter plot of the first semester, RNN_EXO_1 trained in the second phase.

In conclusion, when comparing the prediction graphs, the RNN_EXO_1 model trained in the first phase produced worse predictions, and did not learn the weekly seasonality of consumption, as can be seen in figure 44, it is also possible to notice in table 9, that the correlation between the predicted values and actual consumption, as well as the value R^2 was lower compared to the metrics of the model trained in the second phase, and that this model trained in the second phase learned better the weekly and monthly seasonality of consumption as can be seen in the figure 45.

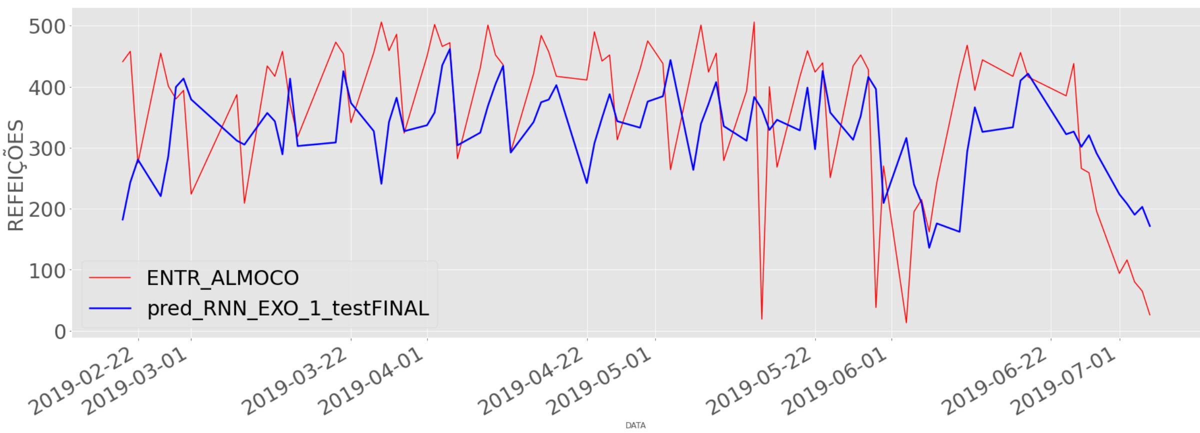

Figura 44: Model Test RNN_EXO_1, 1 Phase

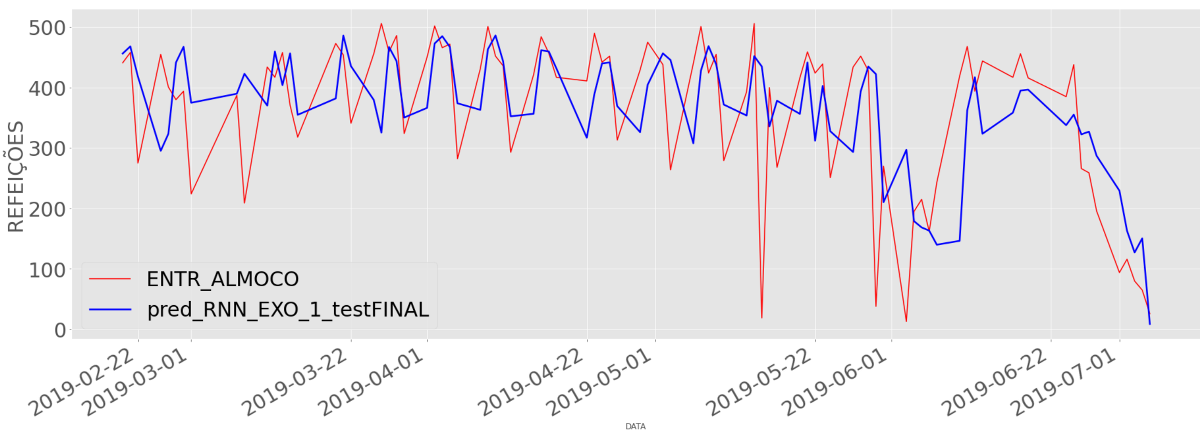

Figura 45: Test of the first semester of RNN_EXO_1 trained in the second phase.

5.4.3 Final test of the RNN_EXO_1 model for the ICT-Unifesp R.U predictions

The final test of RNN_EXO_1 produced the best results with its training in the second phase and being tested for the entire year 2019. The RMSE observed in table 11 was notoriously inferior to all the models trained and tested in all the research.

Tabela 11: RNN_EXO_1 TRAINED IN THE 2ND PHASE, TEST YEAR 2019

RNN_EXO_1 TRAINED IN THE 2ND PHASE, TEST YEAR 2019	
RMSE	99.36
TOTAL MEALS CONSUMED	58653
TOTAL OF PROJECTED MEALS	62048.04
FORECAST ERROR	3395.04
PERCENTAGE ERROR	5.78%
CORRELATION (r)	0.67
P-value (p)	3.29e-25
R2	0.45
SUM OF POSITIVE ERRORS	8163.18
SUM OF NEGATIVE ERRORS	4768.13
MEDIAN ABSOLUTE ERROR	55.23
ABSOLUTE ERROR AVERAGE PERCENTAGE	83.2671%

The disposal of meals was obtained by the sum of the positive errors reached the value of 4768 meals. In the figure46 it is possible to notice that the model learned well the monthly and weekly seasonality of the consumption, but obtained a discrepant error for the first predicted value of the second semester, the error was justifiable because its training set contemplates only 1 year with 1 alternation of semester making impossible a better learning about this behavior. The scatter plot illustrated in the figure 47 lso demonstrates a good linear regression over the predicted values of the model and the actual consumption value, approaching the identity function of an ideal forecast.

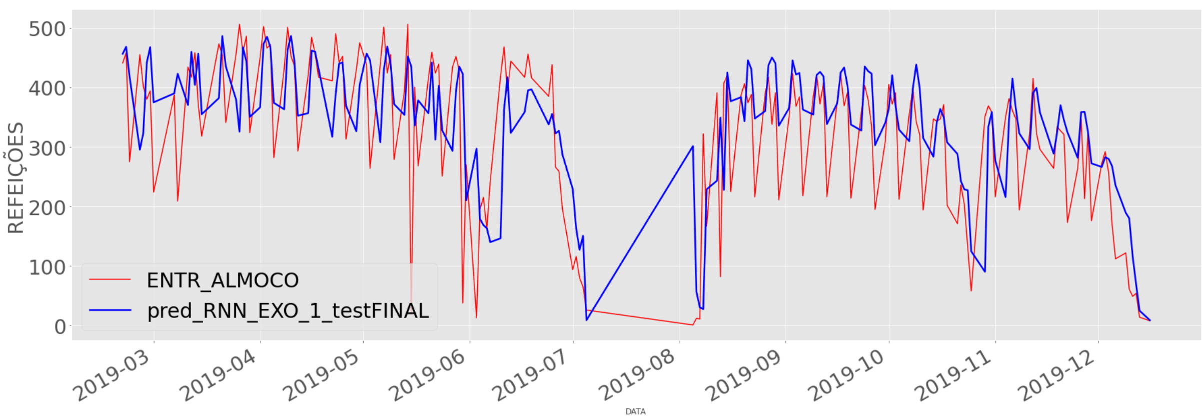

Figura 46: Final Test Graph of the Model RNN_EXO_1.

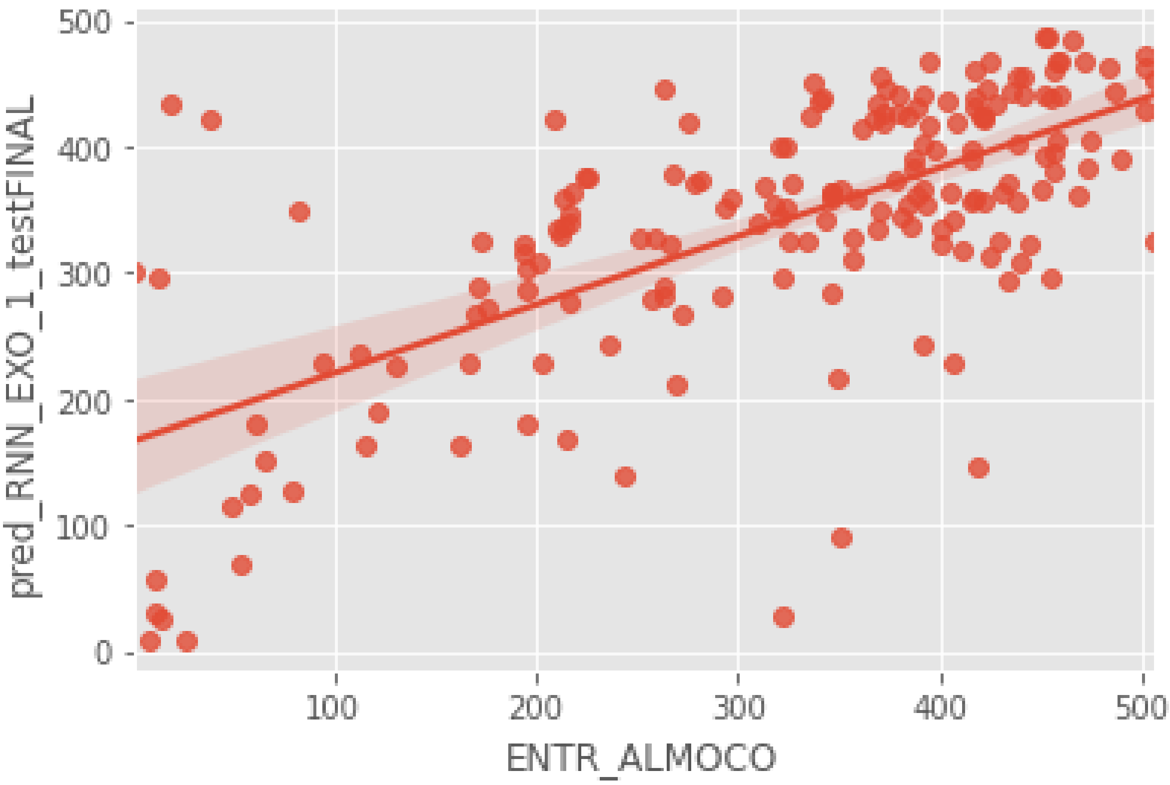

Figura 47: Model Dispersion Graph RNN_EXO_1.

Results obtained by the other tested methods are available in the **??**.

6 Conclusion

First, in this monograph it was possible to evaluate the importance of the methodology to divide the data set into time series. Since with a data set of temporal seasonality it becomes clear that the ordering of the data in the separation of training, testing and validation sets must follow a chronological order for the models to learn from the past and make predictions for the future. The anomalous behavior of the prediction identified in the chapter of division of the phase 1 data set, demonstrated in the figure 16, and its elimination after the correct organization of the time series, as shown in the figure 17, emphasizes the importance of the temporal sequence of the data for the correct learning of the tested models.

Regarding the method of production of meals with margin of error and analysis of the previous week it was possible to observe that even with the production 30% above the consumption in the previous week, at the end of each semester, the ICT-Unifesp restaurant more than 30% are thrown out. This oscillatory behavior of consumption and the addition of outliers ends up amplifying the error of the models in predicting meals. In 2019, following this method, 23 thousand meals were discarded. In this project, the RNN_EXO_3, Model, which presented the largest disposal among all 12 models tested, obtained a maximum value of 8914 disposals. This highlights the need to implement efficient methods for the production and planning of meals at the university refectory of Unifesp.

In relation to the empirical adjustments of the topology of the models during the validation stage of the first developed models, it was possible to notice the reduction of the RMSE along the deepening of the Perceptron network for training and evaluation under the validation set, validating the hypothesis that the models are capable of learning the problem in relation to the adjustment of their topology.

Although the data set contains 2 characteristics that inform the distance in days to the next record and the previous record for the models to identify holidays and long recesses, some events in the calendar, such as stoppages, are not very well represented, indicating the need for further exploration of these characteristics that can better represent this behavior.

Reviewing the best performing model in the first phase, with validation restricted to the first half of 2018, the endogenous models did better than the mixed models. This may mean that the exogenous attributes were noisy during learning. These attributes are mostly composed of annual seasonality such as weather, limited to the seasons. The model RNN_EXO_1 of the 2nd phase, had the best performance among all the models evaluated in this work, but some improvements are indicated:

- Expand the data set for the model to adjust to semester seasonalities and the exchange of semesters. Categorical attributes that indicate semesters, day of the week, as well as those that quantify recesses (previous and subsequent recording distance) have the potential to add learning to this issue. A greater diversification of the data set is still needed, since this model was trained with only 1 annual seasonality period (1 year for training, another year for validation and a third year for testing).
- Add important event attributes to identify events and shutdowns.
- A menu attribute has the potential to increase the quality of the prediction.
- An attribute representing the number of students registered in each period of each day of the week has great potential to increase prediction.
- Surveys can be done for a better transformation of the input data in the perceptron model, because they are discrete data, while the data entering the GRU layer are temporal (with an interval of 5 days).

6.1 General conclusions

The most evident phenomenon in this study was the significant improvement in all the evaluated artificial neural network models, only changing the organization of the data set between the first and second phase, without interfering in any parameter or hyper parameter of these models. This means that more studies and experiments related to data set organization are needed for predictions of consumption time series. The diverse analysis of demand forecasting for the subject requires extensive methods of implementation and data structuring.

The application of training methods with backward propagation, considering the diversity of parameters in machine learning within just one analysis where infinite different topologies can be assembled based on the structure of the collected data has the potential to increase the performance of the models in predicting the demands for the RU.

The heuristics on the definition of topology, although diverse, are not deterministic, and the process requires exploratory, subjective and empirical analysis on the theme or problem to be addressed. However, the efficiency of the machine learning models in work related to university restaurants has been noted. As in the case of the ICT-Unifesp RU there is no current forecasting model and the lack of a model causes food waste and restaurant damage, the approach of this research and its continuation with new methods becomes viable and important to support an assertive management of resources.

Anexos

Referências

ALMEIDA, V. A. d. *Previsão De Carga Através De Modelos Neuro-fuzzy.* 91 p. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2013. Disponível em: <http:/-/monografias.poli.ufrj.br/monografias/monopoli10007620.pdf>. Citado 4 vezes nas páginas 27, 29, 33 e 40.

BARRETT, C. B. Measuring food insecurity. *Science*, American Association for the Advancement of Science, v. 327, n. 5967, p. 825–828, 2010. Citado na página 21.

BRAGA A. DE P.; CARVALHO, A. P. L. F. L. *Redes neurais artificiais: teoria e aplicações*. [S.l.]: Livros técnicos e científicos, 2000. Citado na página 30.

BROWNLEE, J. Ebook machine learning mastery. 2020. Access date: 3 oct. 2020. Disponível em: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>. Citado 2 vezes nas páginas 33 e 34.

Data Science Academy. Deep learning book. 2020. Access date: 3 oct. 2020. Disponível em: <http://deeplearningbook.com.br/arquitetura-de-redes-neurais-gated-recurrent-unit-gru/>. Citado 2 vezes nas páginas 31 e 35.

DAVILA, V. H. *Introdução às Séries Temporais*. 201–. Diapositivas. Disponível em: <https://www.ime.unicamp.br/~hlachos/MaterialSeries.pdf>. Citado na página 26.

EHLERS, R. *Análise de Séries Temporais*. 2009. Disponível em: <https://sites.icmc.usp.br-/ehlers/stemp/cap2.pdf>. Citado na página 26.

GALANAKIS, C. M. The food systems in the era of the coronavirus (covid-19) pandemic crisis. *Foods*, Multidisciplinary Digital Publishing Institute, v. 9, n. 4, p. 523, 2020. Citado na página 21.

HAYKIN, S. *Redes Neurais, Princípios e Práticas, 2nd Ed.* [S.l.]: Bookman, 2007. Citado 5 vezes nas páginas 27, 28, 29, 31 e 32.

HINTON, G. E. et al. *Improving neural networks by preventing co-adaptation of feature detectors*. 2012. Citado na página 35.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 34.

HUFF, A. G. et al. How resilient is the united states' food system to pandemics? *Journal of environmental studies and sciences*, Springer, v. 5, n. 3, p. 337–347, 2015. Citado na página 22.

JUNIOR, A. M. *Análise de métodos de previsão de demanda baseados em séries temporais em uma empresa do setor de perfumes e cosméticos.* Dissertação (Mestrado) — Pontifícia Universidade do Paraná. Curitiba, 2007. Citado 4 vezes nas páginas 25, 26, 27 e 37.

KAPLAN, A.; HAENLEIN, M. Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, Elsevier, v. 62, n. 1, p. 15–25, 2019. Citado na página 27.

KOHAVI, R. Study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. [S.l.: s.n.], 1995. p. 1137–1143. Citado na página 31.

KYUNGHYUN, C. et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. [S.l.: s.n.], 2014. Citado na página 34.

LANDIM, D. D. Análise de demanda no restaurante universitário do instituto de ciência e tecnologia da unifesp - universidade federal de são paulo. *Projeto da Disciplina de Simulação de Sistemas*, Universidade Federal de São Paulo, v. 1, n. 1, p. 6, 2016. Disponível em: <https://github.com/ddlandim/monografy-ann-demand-prediction/tree/master/referencias>. Citado na página 37.

LEMOS, E. *Análise de crédito bancário com uso de data mining: Redes Neurais e Árvores de Decisão*. Dissertação (Mestrado) — Universidade Federal do Paraná, 2003. Citado 2 vezes nas páginas 27 e 28.

LOPES, L. Modelos para previsão de demanda no restaurante universitário utilizando técnicas de redes neurais. *Congresso Nacional De Excelência Em Gestão.Universidade Federal De Viçosa.*, Universidade Federal De Viçosa, IV, n. 1, p. 14, 2008. Disponível em: <http://www.novoscursos.ufv.br/projetos/ufv/gesep/www/wp-content/uploads/5.> Citado 6 vezes nas páginas 22, 37, 38, 39, 44 e 59.

LUGER, G. F. *Inteligência Artificial. Estruturas e Estratégias para a Solução de Problemas Complexos,4a ed.* [S.l.]: Porto Alegre: Bookman, 2004. Citado na página 27.

MARIANO, F. C. M. Q. *Redes Neurais na predição de valores energéticos de alimentos para aves*. 130 p. Tese (Doutorado) — Universidade Federal de Lavras, 2014. Citado na página 29.

MC AI. Overview. 2020. Access date: 3 oct. 2020. Disponível em: <https://mc.ai/complete-guide-of-activation-functions/>. Citado na página 30.

MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 127–147, 1943. Citado 2 vezes nas páginas 27 e 28.

MINSKY, M.; PAPERT, S. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press, 1969. Citado na página 28.

MOREIRA, D. A. Produção e operações. *São Paulo: Pioneira*, 1998. Citado na página 25.

MORETTIN, P. A. *Análise de séries temporais*. [S.l.]: Editora Blucher, 1987. Citado na página 26.

ORGANIZATION, W. H. et al. Global prevalence of vitamin a deficiency in populations at risk 1995-2005: Who global database on vitamin a deficiency. World Health Organization, 2009. Citado na página 21.

ROCHA, J. C.; MATOS, F. D.; FREI, F. Utilização de redes neurais artificiais para a determinação do números de refeições diárias de um restaurante universitário. *Rev. Nutr., Campinas*, Universidade Estadual Paulista Júlio de Mesquita Filho., n. 1, p. 8, 2011. Citado 4 vezes nas páginas 22, 38, 39 e 44.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, p. 386–408, 1958. Citado na página 28.

RUAS, G. I. e. S. Previsão de demanda de energia elétrica utilizando redes neurais artificiais e support vector regression. *Copel Distribuição S/A (COPEL)*, Universidade Federal do Paraná(UFPR), v. 1, n. 1, p. 10, 2012. Disponível em: <http://www.dcc.fc.up.pt/~ines-/enia07_html/pdf/27927.pdf>. Citado 2 vezes nas páginas 27 e 39.

RUMELHART, D.; HINTON, G.; MCCLELLAND, J. A general framework for parallel distributed processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, v. 1, 1986. Citado na página 29.

RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. *Nature*, v. 323, p. 533–536, 1986. Citado na página 32.

SHAW, D. J. World food summit, 1996. In: *World Food Security*. [S.l.]: Springer, 2007. p. 347–360. Citado na página 21.

SILVA, C. e. a. Metodologia para previsão de demanda de energia elétrica em curtíssimo prazo baseada em um modelo probabilístico discreto. *Desenvolvimento de Algoritmos e Software para Previsão de Carga na Área de Concessão da MuxEnergia - Horizonte de Curtíssimo Prazo*, Universidade Federal do Pampa - UNIPAMPA, v. 1, n. 1, p. 4, 2010. Citado 2 vezes nas páginas 27 e 40.

SRIVASTAVA, N. et al. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, v. 15, p. 1929–1958, 06 2014. Citado 2 vezes nas páginas 35 e 36.

WEBB, P. et al. Measuring household food insecurity: why it's so important and yet so difficult to do. *The Journal of nutrition*, Oxford University Press, v. 136, n. 5, p. 1404S–1408S, 2006. Citado na página 21.